

ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE CIENCIAS

OPTIMIZACIÓN MULTI-ECHELON DEL INVENTARIO BASADO EN  
PRONÓSTICOS DE LA DEMANDA POR EL MÉTODO RANDOM FOREST

TRABAJO DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL TÍTULO DE  
INGENIERO MATEMÁTICO

PROYECTO DE INVESTIGACIÓN

MARTÍN EDUARDO LARCO ÁLVAREZ  
martin\_larco@hotmail.com

Directora: ADRIANA UQUILLAS ANDRADE, PhD  
adriana.uquillas@epn.edu.ec

QUITO, Octubre 2021

## DECLARACIÓN

Yo MARTÍN EDUARDO LARCO ÁLVAREZ, declaro bajo juramento que el trabajo aquí escrito es de mi autoría bajo la asesoría y tutela de AMBACAR ECUADOR y de mi directora de tesis; además, que no ha sido previamente presentado para ningún grado o calificación profesional; y que he consultado las referencias bibliográficas que se incluyen en este documento.

A través de la presente declaración cedo mis derechos de propiedad intelectual mas no la información confidencial de AMBACAR ECUADOR, correspondientes a este trabajo, a la Escuela Politécnica Nacional, según lo establecido por la Ley de Propiedad Intelectual, por su reglamento y por la normativa institucional vigente.

## CERTIFICACIÓN

Certifico que el presente trabajo fue desarrollado por MARTIN EDUARDO LARCO  
ÁLVAREZ, bajo mi supervisión.

Adriana Uquillas Andrade

Directora del Proyecto

## AGRADECIMIENTOS

A Brownie y Emma mis fieles compañeros de estudio, trabajo y apoyo emocional.

A Ambacar Ecuador, en especial a Christian Diaz por su ayuda y colaboración dentro del desarrollo de la tesis.

A Adriana Uquillas por su ayuda y apoyo durante el proceso de desarrollo de la tesis.

Finalmente, a mi abuelito César que fue mi mejor amigo y pieza fundamental durante toda mi carrera.

# Índice

<b>1. Introducción</b>	<b>12</b>
<b>2. Justificación</b>	<b>14</b>
<b>3. Objetivos</b>	<b>15</b>
3.1. Objetivo general . . . . .	15
3.2. Objetivos específicos . . . . .	15
<b>4. Metodología</b>	<b>16</b>
<b>5. Nociones Generales y Conocimientos Previos</b>	<b>17</b>
5.1. Análisis de las series de tiempo . . . . .	17
5.2. Historia del pronóstico de la demanda . . . . .	18
5.3. Machine Learning . . . . .	19
5.4. Machine Learning en pronóstico empresarial . . . . .	20
5.5. Beneficios de usar métodos Machine Learning . . . . .	20
5.6. Qué es optimización del inventario . . . . .	21
5.7. Importancia de la optimización de inventario . . . . .	22
5.8. Cómo optimizar el inventario . . . . .	24
5.9. Forecast Pro . . . . .	24
5.10. Teoría de grafos . . . . .	25
5.10.1. Concepto de un grafo . . . . .	25
5.10.2. Representaciones de un grafo . . . . .	25
<b>6. Pronóstico de la Demanda</b>	<b>27</b>
6.1. Introducción a los Random Forest . . . . .	27
6.2. Cualidades generales de los Random Forest . . . . .	28
6.2.1. Caracterización de la eficacia de los Random Forest . . . . .	28
6.3. Random Forest para regresión . . . . .	28
6.3.1. Caracterización de la eficacia de los Random Forest en regresión . . . . .	29
6.3.2. Correlación y robustez en los Random Forest . . . . .	29

6.3.3.	Principios básicos . . . . .	32
6.3.4.	Criterio CART . . . . .	33
6.3.5.	Algoritmo Random Forest . . . . .	33
6.3.6.	Mecanismos alternativos de remuestreo . . . . .	35
6.3.7.	Decisión de partición y propiedades de los nodos hijos . . . . .	36
6.3.8.	Optimización de Parámetros . . . . .	38
6.3.9.	Importancia de las variables . . . . .	39
6.3.10.	Uso de variables aleatorias . . . . .	39
6.3.11.	Pronóstico de las series de tiempo con Random Forest . . . . .	41
6.3.12.	Medidas de error de pronóstico . . . . .	42
<b>7.</b>	<b>Optimización del Inventario</b>	<b>44</b>
7.1.	Inventario Multi-Echelon . . . . .	44
7.1.1.	Estructura de un sistema Multi-Echelon . . . . .	44
7.1.2.	Características de los sistemas Multi-Echelon . . . . .	46
7.2.	Optimización del Inventario usando metodología Multi-Echelon . . . . .	47
7.2.1.	Suposiciones para la Optimización de Inventario . . . . .	49
7.2.2.	Beneficios del algoritmo de Optimización de Inventario . . . . .	50
7.2.3.	Módulo de Simulación . . . . .	50
7.2.4.	Inicialización de la Simulación . . . . .	54
7.2.5.	Módulo de Optimización . . . . .	55
7.2.6.	Procesos Gaussianos . . . . .	56
7.2.7.	Propiedades básicas de los procesos Gaussianos . . . . .	57
7.2.8.	Propiedades de la función de autocovarianza . . . . .	58
7.2.9.	Procesos lineales Gaussianos . . . . .	59
7.2.10.	Predicción . . . . .	59
7.2.11.	Optimización Bayesiana usando procesos de Gauss . . . . .	60
<b>8.</b>	<b>Implementación: Caso Ambacar</b>	<b>63</b>
8.1.	Sobre Ambacar . . . . .	63
8.2.	Red de concesionarios . . . . .	63

8.3. Análisis importancia de marcas . . . . .	68
8.3.1. Ventas Haval-Great Wall en el año 2020 . . . . .	70
8.4. Estimación kilometraje recorridos . . . . .	71
8.5. Categorización de los repuestos . . . . .	72
8.6. Pronóstico de la demanda aplicado a Ambacar . . . . .	74
8.7. Implementación y resultados del modelo de pronóstico Random Forest . . .	77
8.7.1. Recopilación y depuración de la información: . . . . .	77
8.8. Análisis de datos: . . . . .	79
8.8.1. Construcción del modelo y resultados . . . . .	86
8.8.2. Comparación modelo Random Forest vs ventas reales vs Forecast Pro	94
8.9. Optimización del inventario aplicado a Ambacar . . . . .	98
8.10. Resultados . . . . .	102
8.11. Niveles de servicio . . . . .	110
8.12. Impacto del modelo de optimización a nivel de ítem y punto de venta . . .	112
8.13. Puntos de Reaprovisionamiento . . . . .	120
<b>9. Conclusiones</b>	<b>126</b>
<b>10. Anexos</b>	<b>128</b>
<b>11. Referencias Bibliográficas</b>	<b>144</b>

## Índice de figuras

1.	Ejemplo representación gráfica de un grafo . . . . .	26
2.	Particiones árboles de decisión vs Random Forest.) . . . . .	41
3.	Ejemplo de un Sistema Multi-Echelon. . . . .	44
4.	Ejemplo de un Sistema Multi-Echelon Serial. . . . .	45
5.	Ejemplo de un Sistema Multi-Echelon con estructura de ensamblaje. . . . .	45
6.	Ejemplo de un Sistema Multi-Echelon con múltiples clientes. . . . .	46
7.	Ejemplo de un Sistema Multi-Echelon Combinado. . . . .	46
8.	Mapa completo del Ecuador con las ubicaciones de los distintos concesionarios y centro de distribución. . . . .	64
9.	Mapa de la provincia de Tungurahua que muestra las ubicaciones de los concesionarios en Ambato y el centro de distribución principal. . . . .	65
10.	Mapa de la provincia de Pichincha que muestra las ubicaciones de los concesionarios en Quito y el centro de distribución regional. . . . .	66
11.	Promedio que un vehículo nuevo retorna al taller por agencia. . . . .	68
12.	Ventas de vehículos de Ambacar por segmento . . . . .	69
13.	Unidades vendidas por mes en el año 2020 . . . . .	70
14.	Unidades vendidas por agencia en el año 2020 . . . . .	70
15.	Ejemplo de la data proporcionada de venta de repuestos. . . . .	78
16.	Ejemplo de la matriz con las variables consolidadas. . . . .	79
17.	Distribución de la venta total de repuestos. . . . .	80
18.	Ventas anuales de vehículos de Great Wall-Haval. . . . .	81
19.	Ventas anuales de vehículos de Great Wall-Haval por agencia. . . . .	82
20.	Ventas mensuales de vehículos de Great Wall-Haval. . . . .	82
21.	Ventas mensuales de vehículos por agencia de Great Wall-Haval. . . . .	83
22.	Histórico anual de venta de repuestos a nivel nacional. . . . .	84
23.	Histórico anual de venta de repuestos a nivel nacional por concesionario. . . . .	85
24.	Consolidado mensual de la venta de repuestos. . . . .	85
25.	Error en el pronóstico vs número de árboles. . . . .	87
26.	Importancia relativa de las variables dentro del modelo Random Forest. . . . .	88
27.	Ejemplo de un árbol en el modelo Random Forest. . . . .	89



28.	Comparativa conjunto de entrenamiento ventas reales vs predicción, Año-Concesionario. . . . .	90
29.	Comparativa conjunto de entrenamiento ventas reales vs predicción, Año-Código. . . . .	91
30.	Comparativa conjunto de entrenamiento ventas reales 2020 vs predicción 12 primera referencias, Año-Código. . . . .	91
31.	Comparativa conjunto de prueba ventas reales vs predicción, Año-Concesionario.	92
32.	Comparativa conjunto de prueba ventas reales vs predicción, Año-Código. .	93
33.	Comparativa conjunto de prueba ventas reales 2020 vs predicción 12 primera referencias, Año-Código. . . . .	94
34.	Comparativa resultados Forecast Pro vs Real. . . . .	95
35.	Estadísticos obtenidos del software Forecast Pro. . . . .	95
36.	Comparación de los modelos para los 5 concesionarios más importantes. . .	96
37.	Comparación de los modelos para los 10 repuestos más importantes. . . . .	96
38.	Grafo que muestra el sistema de distribución de Ambacar. . . . .	98
39.	Cantidad de ítems disponibles en cada uno de los puntos de atención post-venta. . . . .	104
40.	Cantidad de ítems Actuales vs Propuestos . . . . .	104
41.	Cantidad de ítems Actuales vs Propuestos en Ciauto. . . . .	105
42.	Cantidad de ítems Actuales vs Propuestos en Matriz Ambato. . . . .	106
43.	Cantidad de ítems Actuales vs Propuestos en Juan Tanca Marengo. . . . .	106
44.	Cantidad de ítems Actuales vs Propuestos en Granados. . . . .	107
45.	Cantidad de ítems Actuales vs Propuestos en Guayaquil. . . . .	107
46.	Cantidad de ítems Actuales vs Propuestos en Latacunga. . . . .	108
47.	Cantidad de ítems Actuales vs Propuestos en Manta. . . . .	108
48.	Cantidad de ítems Actuales vs Propuestos en Quicentro Sur. . . . .	109
49.	Cantidad de ítems Actuales vs Propuestos en Galo Plaza Lasoo. . . . .	109
50.	Cantidad de ítems Actuales vs Propuestos en San Rafael. . . . .	110
51.	Cantidad de ítems Actuales vs Propuestos en Chone. . . . .	139
52.	Cantidad de ítems Actuales vs Propuestos en El Coca. . . . .	139
53.	Cantidad de ítems Actuales vs Propuestos en Esmeraldas. . . . .	140
54.	Cantidad de ítems Actuales vs Propuestos en Ibarra. . . . .	140

55.	Cantidad de ítems Actuales vs Propuestos en Machala. . . . .	140
56.	Cantidad de ítems Actuales vs Propuestos en Seminuevos Ambato. . . . .	141
57.	Cantidad de ítems Actuales vs Propuestos en El Puyo. . . . .	141
58.	Cantidad de ítems Actuales vs Propuestos en Seminuevos Quito. . . . .	141
59.	Cantidad de ítems Actuales vs Propuestos en Riobamba. . . . .	142
60.	Cantidad de ítems Actuales vs Propuestos en Salinas. . . . .	142
61.	Cantidad de ítems Actuales vs Propuestos en Santo Domingo. . . . .	142
62.	Cantidad de ítems Actuales vs Propuestos en Bolivariana. . . . .	143

## Índice de cuadros

1.	Código, nombre comercial, ciudad y provincia de cada uno de los centros de distribución postventa . . . . .	67
2.	Unidades vendidas por Marca que comercializa Ambacar en el año 2019. .	68
3.	Modelos y versiones más vendidos de la marca Haval-Great Wall en el año 2019 . . . . .	69
4.	Promedio de Kilometraje recorrido de los vehículos de la marca Haval-Great Wall por línea de negocio. . . . .	71
5.	Ítems disponibles por código A001-A009 en los puntos de atención postventa	71
6.	Top 50 de los repuestos más importantes para el servicio postventa de Ambacar . . . . .	73
7.	Ejemplo de la variable: Vehículos Vendidos . . . . .	75
8.	Ejemplo de la variable: Promedio de días desde la última visita. . . . .	76
9.	Ejemplo de la variable: Kilometraje promedio. . . . .	77
10.	Ejemplo de las variables Venta de repuestos promedio y Venta de vehículos promedio. . . . .	86
11.	Primeros resultados estadísticos del modelo Random Forest y la variabilidad explicada. . . . .	88
12.	Resultados del conjunto de entrenamiento, medidas de error MAPE Y RMSE.	89
13.	Comparación de las ventas reales y lo pronóstico para los concesionarios más importantes en el conjunto de entrenamiento. . . . .	90
14.	Resultados del conjunto prueba, medidas de error MAPE Y RMSE. . . . .	92
15.	Comparación de las ventas reales y lo pronóstico para los concesionarios más importantes en el conjunto de prueba. . . . .	93
16.	Estadísticos de los modelos usando el software Forecast Pro. . . . .	95
17.	Comparativa de los modelos Random Forest vs Forecast Pro. . . . .	95
18.	Comparativa de los errores de los modelos Random Forest vs Forecast Pro.	97
19.	Días de espera en los envíos del centro de distribución a cada provincia. . .	99
20.	Ítems disponibles por código A001-A009 en los puntos de atención postventa.	99
21.	Ítems disponibles por código A010-A018 en los puntos de atención postventa.	100
22.	Ítems disponibles por código A019-A027 en los puntos de atención postventa.	100
23.	Ítems disponibles por código A028-A036 en los puntos de atención postventa.	101
24.	Ítems disponibles por código A037-A045 en los puntos de atención postventa.	101

25.	Ítems disponibles por código A046-A050 en los puntos de atención postventa.	102
26.	Nivel de Servicio . . . . .	111
27.	Análisis de los ítems A001 a A004 por concesionario. . . . .	112
28.	Análisis de los ítems A005 a A008 por concesionario. . . . .	113
29.	Análisis de los ítems A009 a A012 por concesionario. . . . .	113
30.	Análisis de los ítems A013 a A016 por concesionario. . . . .	114
31.	Análisis de los ítems A017 a A020 por concesionario. . . . .	114
32.	Análisis de los ítems A021 a A024 por concesionario. . . . .	115
33.	Análisis de los ítems A025 a A028 por concesionario. . . . .	116
34.	Análisis de los ítems A029 a A032 por concesionario. . . . .	116
35.	Análisis de los ítems A033 a A036 por concesionario. . . . .	117
36.	Análisis de los ítems A037 a A040 por concesionario. . . . .	117
37.	Análisis de los ítems A041 a A044 por concesionario. . . . .	118
38.	Análisis de los ítems A045 a A048 por concesionario. . . . .	119
39.	Análisis de los ítems A049 a A050 por concesionario. . . . .	120
40.	Puntos de Reaprovisionamiento ítems A001-A010. . . . .	121
41.	Puntos de Reaprovisionamiento ítems A011-A020. . . . .	122
42.	Puntos de Reaprovisionamiento ítems A021-A030. . . . .	123
43.	Puntos de Reaprovisionamiento ítems A031-A040. . . . .	124
44.	Puntos de Reaprovisionamiento ítems A041-A050. . . . .	125

# 1. Introducción

En la actualidad, los cambios en el consumo y la mayor exigencia de los clientes han ocasionado que el contexto de los negocios varíe y se cuestione el estatus quo. La opinión de un cliente sobre un negocio puede cambiar rápidamente debido a una demora en la entrega o la falta de un ítem específico. Todo esto sumado al gran catálogo de productos que una empresa debe almacenar para permanecer competitiva en el mercado, o simplemente por la necesidad de mantener un alto número de ítems en inventario para cumplir obligaciones legales ya sean estas garantías, compromisos de servicio o atención al cliente. Todos los factores mencionados han ocasionado una serie de problemas, tales como una baja demanda y demandas intermitentes en gran parte del catálogo de productos. Todo esto se puede resumir como la dificultad y complejidad del pronóstico de la demanda. Además, muchos de estos problemas cotidianos mencionados, sumado a la ineffectividad y dificultad de los métodos tradicionales de pronóstico aplicados a la vida real. Ya sean estos ARIMA, Suavización Exponencial (SE), Modelos Lineales, e inclusive métodos de modelamiento más complejos; han impulsado el estudio de nuevos métodos e innovaciones dentro del campo del pronóstico (Seeger et al., 2016).

Los impactos negativos en la credibilidad del negocio se producen por la incapacidad de cumplir con la demanda lo cual genera una experiencia no satisfactoria en los clientes, lo que resulta en una pérdida de ventas. Existen varias razones por las cuales el pronóstico puede ser inexacto, entre las principales están que los planificadores suponen que las tendencias continuarán en el futuro. Puede que esto sea válido para un corto periodo de tiempo, pero eventualmente las tendencias van a cambiar. Identificar y predecir esos puntos críticos de cambio es uno de los aspectos más difíciles en el pronóstico empresarial. Todos estos problemas incrementan los riesgos de falta o exceso de inventario, inclusive inventario muerto. Un manejo incorrecto de las políticas de inventario lleva no solo a pérdidas económicas, pero también a una sobreproducción y exceso de inventario, lo cual es extremadamente dañino para la continuidad de cualquier negocio industrial o comercial (Böse et al., 2017).

Uno de los errores más frecuentes en el área de logística empresarial es mantener inventario en lugares erróneos dentro de su sistema de almacenaje, sobre todo en negocios en los cuales existen múltiples bodegas, centros de distribución y puntos atención al público, lo que ha ocasionado que los niveles de stock de seguridad bajen tanto que causen problemas en el nivel de servicio ofrecido a los clientes, generando continuos desabastecimientos, aumentando así los costos de transportación, incluso puede existir la necesidad de realizar órdenes de compra de emergencia que elevan inesperadamente los costos. Por otro lado, una mala aplicación de las políticas de inventario podría generar que la empresa termine con una gran cantidad de stock que se vuelva obsoleto, incrementando el costo de almacenamiento para el resto de los productos, además de tener que lidiar con el problema de cómo reducir el inventario y mitigar pérdidas para la compañía.

Los errores que se cometen en el manejo de inventarios representan oportunidades perdidas en la optimización del negocio por los recursos que se destinan para almacenar productos. Fallas en las predicciones para el manejo del inventario han reducido en un

10 % las ganancias de las organizaciones (DeHoratius y Ton, 2001). Por otro lado, reportes de International Data Corporation (IDC) confirman que numerosas compañías que han usado o implementado soluciones inteligentes para optimización de inventarios han logrado reducir sus niveles de stock hasta en un 25 % (Esther, 2012). Este hecho enfatiza que la optimización de inventario es, sin duda, una tarea a priorizarse en cualquier organización, sean empresas privadas o públicas.

Consecuentemente, el trabajo propuesto tiene como meta brindar sugerencias a la empresa Ambacar para establecer políticas útiles con el fin de mantener un inventario sano, basadas en un pronóstico de la demanda realizado por modelos Random Forest con la ayuda de variables internas y externas que nos permitan entender mejor la variación que existe en la demanda, lo que permitirá en una siguiente etapa mediante un modelo Multi-Echelon, establecer niveles de reaprovisionamiento y de stock de seguridad en cada uno de los concesionarios mediante el modelamiento del sistema de almacenamiento y distribución que tiene la compañía.

## 2. Justificación

La constante búsqueda de metodologías que permitan un pronóstico de demanda más preciso, con un menor grado de error, ha llegado a convocar no solo a desarrolladores de software, fabricantes, distribuidores, vendedores, si no que ha involucrado a la academia en el objetivo de reducir la incertidumbre. Predicciones de la demanda más precisas conllevan a mejores decisiones en compras, manejo de inventario, manejo de capacidades y planeación en general (Punia et al., 2020).

Los métodos comúnmente usados para el pronóstico de la demanda se basan principalmente en tratar de identificar tendencias y estacionalidades que usualmente existen en las series de tiempo, para esto se utilizan métodos univariados y multivariados que establecen relaciones entre la variable de interés y otras variables de apoyo independientes. Modelos estadísticos, ya sean ARIMA, ARIMAX o de Suavización Exponencial (SE) han demostrado que son efectivos cuando las series temporales muestran tendencias claras y un comportamiento compatible con las condiciones iniciales. Sin embargo, estos métodos tienen dificultades para realizar buenas predicciones, a pesar de que la variable dependiente pudiera tener todas las características antes mencionadas, cuando la variabilidad que se genera en la demanda proviene del comportamiento humano, por lo tanto, las variables que contribuyen al análisis contienen tal complejidad que los modelos clásicos no pueden aproximar fácilmente (Jiang et al., 2019).

Consecuentemente, la academia y algunos desarrolladores de software se han enfocado en métodos de Machine Learning para lograr pronósticos más exactos, uno de estos modelos es Random Forest, por cuanto por su estructura de árboles tienen la facilidad de expandirse para acomodar a los datos históricos y variables auxiliares que se pueden incluir en el pronóstico. Random Forest es un buen modelo adaptativo y versátil (Mei et al., 2014). Esta gran versatilidad ha popularizado su uso ya que funcionan eficientemente trabajando con una gran cantidad de data y proporcionan una gran exactitud (Kumar et al., 2016).

Los modelos Random Forest poseen una ventaja en comparación al resto de modelos ya que estos requieren una menor cantidad de hiper parámetros que deban optimizarse (Scornet y Erwan, 2017). Por lo que logran evitar la sobre especificación que conduciría a resultados más imprecisos en el pronóstico.

Hoy en día las organizaciones han desarrollado redes de distribución más complejas para llegar a sus clientes de una manera más rápida y efectiva con el objetivo de servir mejor. Esto significa que los productos pasan por distintos niveles de almacenamiento antes de llegar a los puntos de venta finales, éstos son centros de distribución, bodegas regionales o bodegas internas, a esta metodología de almacenamiento se denomina Multi-Echelon o Multi Nivel. En la optimización de inventario se tiene el objetivo de eliminar el problema que ocasiona una optimización simple de un nivel ya que esta no considera el impacto que puede tener otros centros de distribución y puntos de atención dentro de todo el sistema de logística que tiene una organización. Las consecuencias de no considerar un acercamiento sistemático multinivel pueden llevar a desperdicios al establecer stocks de

seguridad erróneos y a fallas en el servicio al cliente a pesar de tener un inventario correcto (Agarwal, 2019).

Los beneficios de implementar un sistema de optimización de inventario considerando un acercamiento Multi-Echelon pueden significar una disminución hasta un 30 % del inventario, además ha logrado mejorar el flujo de efectivo hasta un 50 % (Bondestab, 2015). La metodología presentada en este trabajo de titulación se enfoca en mejorar el servicio al cliente porque elimina el riesgo de pérdida de clientes y ventas en el mediano plazo por quiebre de inventario.

El trabajo propuesto busca una alternativa a los modelos clásicos para el pronóstico de la demanda, considerando variables exógenas, y que además son modelos que pueden considerar el factor humano para los pronósticos cuando las variables iniciales son complementadas por componentes expresadas como variables de apoyo, en particular, el algoritmo Random Forest muestra una mejora sobre los modelos básicos en predecir el factor humano (Bourgin, 2019). De esta manera y con los resultados obtenidos dar un enfoque más global y sobre todo empresarial a los modelos de manejo de inventarios, con esto trataremos de entregar fundamentos técnicos para la toma de decisiones, buscando brindar más información y alternativas a lo que hoy en día se maneja en la compañía. Los resultados serán políticas de inventarios más acordes a la actualidad nacional y el presente de la empresa, que reduzcan los altos costos que conlleva mantener un gran inventario de repuestos para cumplir los niveles de satisfacción que los clientes esperan.

## 3. Objetivos

### 3.1. Objetivo general

- Establecer un modelo de optimización de inventario para los productos categoría A de la marca más vendida de Ambacar mediante una optimización Multi-Echelon que nos permita determinar stocks de seguridad y nivel de inventario para el reaprovisionamiento óptimo.

### 3.2. Objetivos específicos

- Realizar una categorización y segmentación de datos que nos permita encontrar los productos que tienen mayor importancia para Ambacar de acuerdo con su nivel de ventas.
- Implementar un modelo Random Forest que pronostique la demanda de los principales productos de Ambacar, estudiar la variabilidad de está y comparar los resultados con el software comercial Forecast Pro.
- Sugerir una política de inventarios que minimice el número de ítems en cada uno de los puntos de venta cumpliendo con los niveles de servicio al 95 %.



## 4. Metodología

El primer paso es la obtención de datos que serán proporcionados por la empresa Ambacar, los datos serán sobre el sistema de almacenamiento de repuestos a nivel nacional, la información contemplará ventas históricas, tiempos de espera de los repuestos, tiempo de traslado hasta las distintas locaciones y datos sobre el negocio que serán útiles para el pronóstico de ventas.

El segundo paso es la categorización ABC de los datos (Análisis de Pareto) para determinar, por ejemplo, los productos más vendidos, los de muy poca demanda o los ítems muertos, de acuerdo con su importancia para la compañía y sobre todo para que el pronóstico sea más preciso al considerar ítems que se comporten de manera similar considerando su tendencia y estacionalidad.

Posteriormente, en el tercer paso, se procederá a realizar el pronóstico de ventas por medio de la metodología Random Forest (RF). Los resultados serán analizados por medio de métricas tales como MAPE, SMAPE, MAE para evaluar su exactitud, además los resultados se compararán con los resultados obtenidos por el software Forecast Pro, que actualmente es la herramienta usada por la empresa para el pronóstico de la demanda. Cabe recalcar que para esta fase el modelo Random Forest como todos sus parámetros serán implementados y optimizados en el programa estadístico R.

Cuarto paso, se realizará un modelamiento a través de grafos del sistema de logística y abastecimiento de la empresa donde los valores de las aristas son los tiempos de espera y transporte de los ítems más importantes de la marca más vendida de Ambacar que actualmente posee la empresa en inventario. Se procederá a recrear gráficamente el grafo obtenido modelando en un mapa virtual del Ecuador para que haya un mejor entendimiento del proceso y el sistema de logística y entrega de la empresa.

Inmediatamente, en el quinto paso, se procederá a implementar el sistema de optimización Multi-Echelon usando la data histórica de ventas, niveles de servicio deseados y tiempos de espera y transporte como argumentos iniciales. Este paso se realizará en el programa Google Colab usando lenguaje de programación Python.

Sexto paso, finalmente, mediante el aplicativo Tableau se realizará un programa de visualización de datos que ayude a las personas encargadas del manejo del inventario tener un mejor control y consecuentemente una mejor toma de decisiones.

## 5. Nociones Generales y Conocimientos Previos

### 5.1. Análisis de las series de tiempo

Es difícil poder determinar una fecha exacta en la cual el análisis de las series de tiempo tomó realmente su nombre como ciencia. Desde el inicio de la humanidad como especie y sobre todo el inicio de comercio entre civilizaciones ha habido nociones basadas más en el conocimiento empírico y poco almacenado de lo que ha pasado para tratar de analizar comportamientos futuros.

El análisis de las series de tiempo empieza con la recolección de datos, serían los británicos (Royal Statistical Society) que desde hace más de 175 años empezaron a recolectar data donde las primeras series de tiempo empiezan a tomar forma. Los análisis estadísticos de series de tiempo empezaron hace mucho tiempo con los modelos autorregresivos (Yule 1927). Sin embargo, el análisis seguía siendo muy básico y no muy popular debido a la complejidad que presentaban estos modelos.

Durante la década de los 20s y 30s el concepto de media móvil fue introducido para eliminar las fluctuaciones periódicas en las series de tiempo. Sobre todo, las fluctuaciones que se producen por la estacionalidad. Posteriormente, se introdujeron las primeras nociones de los modelos ARMA para series de tiempo estacionarias. Sin embargo, no se logró aún crear una función de máxima verosimilitud para la estimación de los parámetros.

La publicación de *Time Series Analysis: Forecasting and Control* de Box y Jenkins en 1970 ayudó a la popularización del análisis de las series de tiempo, ya que provee una aproximación sistemática que permite a los analistas aplicar métodos de series de tiempo en el pronóstico (Tsay, 2000). Box y Jenkins popularizaron el modelo Media Móvil Autorregresivo (ARIMA) por su procedimiento iterativo y por los avances tecnológicos. El desarrollo de la computación a finales de la década sobre todo con la creación de computadoras personales permitió que todos estos modelos sean explotados de una mejor manera.

El éxito de los modelos ARIMA impulsó posteriores investigaciones en los análisis de las series de tiempo. Sin embargo, el análisis de estas ha tenido sus dificultades y problemas. Al inicio, el análisis era dividido en dos fases: la primera en el dominio de la frecuencia y la segunda enfocado en el dominio del tiempo. Los proponentes de estas tendencias generaron debates intensos y varias críticas entre las dos escuelas (Tsay, 2000). Los enfocados en el dominio del tiempo usando una aproximación con la función de autocorrelación y modelos paramétricos, tales como los modelos ARIMA, para describir la dependencia dinámica de estas series (Box, 2013). Al contrario, la aproximación por la frecuencia del dominio se enfoca en el análisis espectral para estudiar la teoría y aplicaciones del análisis de las series de tiempo.

En tiempos recientes las tendencias de los análisis de las series de tiempo y las predicciones de estas se han inclinado por metodologías tipo Machine Learning (ML) e Inteligencia Artificial (IA) que se mencionaran más adelante en el texto.

## 5.2. Historia del pronóstico de la demanda

Desde el inicio de la humanidad los pobladores han buscado maneras de anticipar y prever lo que pasará en el futuro. El punto de vista de los pobladores antiguos sobre la incertidumbre se basa más en la inexistencia de la necesidad de saber que pasará o podría suceder ya que los dioses tienen un control completo sobre todo lo que acontecerá en el futuro.

Durante el Renacimiento la filosofía de las personas empezó a cambiar. Las personas empezaron a entender conceptos sobre incertidumbre y eventos probabilísticos por la noción de que estos comportamientos y modelos fueron puestos por obra de la naturaleza o algún ser divino (Lapide y LaPiDe, 2009). Durante los siglos posteriores estas nociones fueron mezcladas con conocimientos empíricos que la población obtuvo sobre ciertos elementos, trabajos o procesos. No sería hasta finales del siglo XIX que con la llegada de teorías que sirven de base para futuras innovaciones como: teoría de juegos (Equilibrio de Nash), que se empieza a notar que aparentemente tenemos más control de lo que puede suceder en el futuro. Desde la década de los 50 los avances relacionados al pronóstico se han multiplicado sobre todo por los avances en computación.

En los primeros intentos de pronósticos de la demanda que se dieron en los finales de las décadas de los 50 se inclinaron y desarrollaron más hacia los modelos de suavización exponencial (Lapide y LaPiDe, 2009). Desde entonces los modelos de pronóstico han evolucionado incluyendo un amplio abanico de métodos estadísticos de series de tiempo.

Al pasar de los años se entendió la necesidad de desarrollar métodos más sofisticados. Estos fueron desarrollados para intentar comprender las variaciones de tendencias y estacionalidad; Sin embargo, estos aún estaban basados en la creencia que los datos siguen tendencias recurrentes que una vez que sean analizadas, predecir el futuro es algo muy sencillo (Lapide y LaPiDe, 2009). Las últimas crisis financieras generaron la necesidad de implementar mejores métodos de simulación para pronosticar y predecir la demanda, que dejaron de asumir que la demanda solo aparecía mágicamente y la única forma precisa de ser estimada es entendiendo que ha pasado previamente. Con esto, se empezaron a incorporar métodos de regresión múltiple y métodos de análisis de series de tiempo usando efectos de mercado como, por ejemplo: las actividades promocionales. Por tanto, este tipo de variables fueron incorporadas al estudio y pronóstico de las series de tiempo. Modelos tipo ARIMA fueron los más populares en el inicio de esas tendencias (Lapide y LaPiDe, 2009). Sin embargo, existen desarrolladores, tal como Business Forecast Systems (desarrolladores de Forecast Pro) que incorporaron matemáticamente los efectos del mercado a los modelos de pronóstico.

Los efectos de la globalización y la expansión de los mercados han convertido al pronóstico de la demanda y en general a los procesos que poseen las entidades en grandes sistemas con una gran cantidad de datos, sobre todo en instituciones que necesitan hacer pronósticos en niveles multinacionales. El avance de la globalización ha ocasionado que los planes de negocios adquieran una gran complejidad ya que todos los productos deben tener un pronóstico, y dependiendo su importancia dentro del contexto del negocio,

su pronóstico debe ser analizado con más rigurosidad. Todos estos problemas se suman al hecho que muchos SKU (Stock-Keeping-Unit) poseen ciclos de vida muy cortos y sus demandas pueden ser muy bajas o intermitentes; y sea necesario el uso de métodos empresariales para su pronóstico.

Los avances tecnológicos sumados a los crecimientos en la escala de los negocios han ocasionado la necesidad de buscar métodos generales que puedan encargarse de los problemas empresariales de hoy en día; considerando variables de los negocios y que puedan captar de manera eficientes los efectos de estacionalidad y tendencias. Modelos como Redes Neuronales (NNS), Random Forest (RF), Gradient Boosting (GB) han tomado un papel protagónico en los últimos tiempos en el pronóstico de la demanda.

### 5.3. Machine Learning

Machine Learning es un subconjunto de la inteligencia artificial en donde las computadoras son usadas para de manera automática aprender de la data e información. En Machine Learning las computadoras no tienen que ser explícitamente programadas, pero pueden cambiar y mejorar los algoritmos por ellos mismos (Bernard, 2016).

Los primeros pasos de Machine Learning (ML) que se pueden encontrar datan de 1943, cuando el neuropsicólogo Warren McCulloch y el matemático Walter Pitts escribieron un documento sobre neuronas y como funcionaban. Ellos decidieron crear un modelo de aquello usando un circuito eléctrico, la combinación de eso generaría que la Red Neuronal (NN) sea creada.

En el año de 1950, Alan Turing crea la “Turing Test” para determinar si una computadora tiene verdadera inteligencia. La prueba consistía que la computadora debe ser capaz de engañar a un ser humano, el éxito de la prueba se determinaba de que el sujeto de prueba piense que la computadora es humana.

Dos años más tarde, en 1952. Arthur Samuel escribió el primer programa de aprendizaje de computación. El programa era el juego de ajedrez. Consistía que entre más se jugaba el programa iba aprendiendo las mejores jugadas y estrategias para incorporarlas en el programa, con eso, logrando el objetivo de ganar la partida y convertirse en el mejor jugador de ajedrez.

Posteriormente, en 1957. Frank Rosenblatt diseñó la primera red neuronal para computadoras basado en el trabajo de McCulloch y Pitts para simular el proceso de un cerebro humano. Diez años más tarde, el algoritmo de “Nearest Neighbor” fue escrito, esto permitió que las computadoras pueden reconocer patrones básicos en la información. Uno de los principales usos que se le dio fue el problema del agente viajero (TSP).

En el año de 1981, Gerald Dejong introdujo los conceptos iniciales de aprendizajes por medio de explicación (EBL). Donde las computadoras eran capaces de analizar la data de entrenamiento y tienen la capacidad de generar una regla que pueda descartar informa-

ción no importante.

En la década de los 90s podemos ver unos los avances más importantes de Machine Learning (ML). Se genera un cambio en el enfoque que se centraba primordialmente en el conocimiento, se empieza a tomar un acercamiento más centrado en la data. Los científicos de datos empezaron a crear programas que sean capaces de analizar gran cantidad de datos y aprender de los resultados.

Durante las dos siguientes décadas los avances se realizaron más en el ámbito privado con el objetivo de resolver problemas puntuales de los negocios e inclusive ofertarse como servicio empresarial como Amazon que en el año 2015 lanzo su propia plataforma de Machine Learning (ML). Empresas como Microsoft y Google han seguido el mismo camino.

## 5.4. Machine Learning en pronóstico empresarial

El éxito de Machine Learning está basado en la utilización de los algoritmos que son capaces de aprender por el proceso de prueba y error, mejorando su rendimiento a través del tiempo, no es un simple proceso paso a paso de programación basado en lógica, reglas condicionales de Si y árboles de decisión (Makridakis et al., 2018).

Los modelos ML por su gran versatilidad han encontrado un gran campo de desarrollo en el área de pronóstico siendo su principal misión mejorar las predicciones de las series de tiempo. El objetivo de los métodos ML es el mismo que los modelos estadísticos, ambos apuntan a mejorar la exactitud de la predicción al minimizar alguna función de pérdida, típicamente la suma de los errores al cuadrado (Makridakis et al., 2018).

Los primeros registros de aplicaciones de ML se datan alrededor de mediados de la década de los 60, pero no generaron mucho interés en la comunidad científica hasta que la técnica de propagación hacia atrás (Backpropagation) fue introducida 20 años más tarde. (Zhao, 2009).

Métodos ML han ganado importancia y protagonismo en el tiempo conforme el interés en Inteligencia Artificial (AI) ha ido aumentando, se han usado para predecir series financieras (Wang y Wang, 2017). La dirección del Mercado de Valores (Qui y Song, 2016). Variables Macroeconómicas (Kock y Terasvirta, 2016). Además; un gran número de otras aplicaciones.

## 5.5. Beneficios de usar métodos Machine Learning

Uno de los principales beneficios que muestran los modelos Machine Learning es la poca o nula interferencia humana. Siendo esa relativa independencia uno de los elementos que más destacan dentro de la utilización de los modelos ML. El método de aprendizaje de los modelos se basa en el reconocimiento y práctica que se obtiene de la minería de

datos que actualmente la mayoría de las entidades poseen. La analítica exploratoria y algoritmos, como Redes Neuronales (NN), Redes Neuronales Recurrentes (RNN), Support Vector Machines (SVM) Y Random Forests (RF) han sido los métodos que más avances y más aplicaciones han tenido dentro del pronóstico, una de las principales ventajas es que son autocorrectivos y capaces de aprender de sus propios errores (Premkumar, 2000).

Hoy en día, cada vez es más fácil para una organización implementar modelos ML dentro de sus procesos de cadenas de suministro; en especial en el área de pronóstico y abastecimiento. Esto se logra gracias a los avances tecnológicos como computación en la nube, regresión dinámica y reconocimiento de patrones. Como consecuencia de todos estos avances son evidentemente, que el pronóstico con Machine Learning mejore en sus predicciones y en muchos casos superando los resultados obtenidos por métodos tradicionales (Teboul, 2018).

Una de las principales ventajas que tienen los modelos ML es su capacidad de combinar y conectarse con Big Data de manera sencilla por medio de computación en la nube y el uso de múltiples factores al mismo tiempo.

Los resultados de la aplicación de los modelos ML nos permiten obtener la importancia relativa de las distintas fuentes de datos. Así, logrando obtener una mejor interpretación de todas las variables que forman parte del análisis. Además; hacen una gran recopilación de toda los datos que son significativos y que deben continuar siendo usados para futuros análisis. Todo esto, genera una gran combinación de elementos que usualmente conllevan a un mejor pronóstico.

## 5.6. Qué es optimización del inventario

Podemos definir la optimización de inventario como un método para lograr las mejores soluciones al problema de la cantidad de inventario necesario que requiere cada locación y como abastecer a las mismas. Esto se puede lograr con la búsqueda de equilibrar las restricciones o los objetivos de inversión de capital y los objetivos de nivel de servicio en una gran variedad de referencias (SKU) a la vez que se tiene en cuenta la volatilidad de la demanda y del suministro.

Buscando una explicación más precisa y amigable podemos traducirlo a un idioma más sencillo como la práctica de tener el inventario adecuado para satisfacer los niveles de servicio mientras que se trata de minimizar el capital invertido en inventario. El objetivo del inventario óptimo es mantener en stock la cantidad de producto que se va a vender (demanda) más el stock de seguridad que puede ser determinado ya sea por políticas de la empresa, o puede ser calculado mediante el uso de intervalos de confianza.

Para lograr eso, se deben considerar los diversos factores de volatilidad de la demanda en sus diversos factores ya sean internos o externos.

Si podemos citar de las tareas que componen Manejo y Optimización de Inventario estas serían:

- Mantener niveles de inventario al menor costo posible para la organización.
- Lograr un nivel alto de Fill Rates (Número de productos despechados dentro de una orden de compra, completos y a tiempo).
- Reducir la cantidad de puntos de reabastecimiento fuera de la planificación.
- Establecer controles predictivos sobre la efectividad del manejo del inventario usando la información y datos disponibles.

Durante las últimas décadas, la Optimización de Inventario ha logrado moverse desde la teoría impartida en las aulas de clases a un enfoque más práctico, las empresas han logrado tomar provecho y mejorar sus procesos de almacenamiento y control.

Cuatro factores han sido los pilares esenciales para el crecimiento que se han tenido en los últimos años; el éxito de iniciativas de mejoras en el área de operaciones; el reconocimiento de la importancia del pronóstico; la implementación de medidas para evaluar rendimientos; existencia y comercialización de software de Optimización de Inventario.

## 5.7. Importancia de la optimización de inventario

El mundo escondido y cansado de recopilación de datos, verificación y seguimiento de inventario en sus diferentes canales, el pronóstico de las demandas y análisis de resultados nunca han sido tareas muy populares dentro de las organizaciones. En la actualidad con los crecimientos exponenciales y popularización del internet se pueden apreciar patrones, actitudes y políticas autodestructivas que algunas compañías tienen. A continuación; citaremos algunos ejemplos de malas prácticas de planificación.

En el lanzamiento de Target Canadá En el año 2015. Un sobre apilamiento de carros de juguete de Barbie de la fábrica de MATTEL, que fueron mal colocados por exceso de inventario en los centros de distribución de Target, afectaron en el plan de descuentos y lanzamiento de la compañía en su entrada al mercado canadiense ya que el distribuidor se vio en la necesidad de seguir acumulando en las bodegas los carros de juguete ocasionando que el manejo y control dentro de la misma sea muy difícil. El problema del inventario fue tan grande que Target Canadá tuvo que cerrar (Trujillo,2016).

Lo ocurrido fue un claro ejemplo de que cuando se quiere crecer muy rápido generalmente cometes muchos errores en el manejo del inventario. Desde entonces, Target ha anunciado que ha reducido y mejor organizado su stock para evitar tener esos inconvenientes.

Walmart es protagonista de una de las historias más famosas de un mal manejo de inventario. No es ningún misterio que Walmart ha tenido problemas por años para mantener sus perchas abastecidas (Trujillo, 2016). En 2011, La empresa Acosta Inc en asociación con empleados de Walmart y analistas de la compañía caminaron literalmente por los pasillos de los distintos supermercados para realizar un seguimiento del tránsito de los

productos. Se realizaron algunas mejoras puntuales que ayudaron a un reabastecimiento más eficaz. Sin embargo, en 2013 los problemas volvieron y de hecho empeoraron, de acuerdo con Bloomberg News, los responsables del proceso dijeron: “Nos quedamos muy rápido sin productos para abastecer las perchas y los nuevos productos no vienen”. Ese año, una mala ubicación de los productos, ya sea por tenerlos en puntos de venta incorrectos o no ser despachados a tiempo, representaron a la empresa alrededor de 3 billones de dólares, sumado al problema que las ventas no crecían al mismo ritmo que las pérdidas y las despensas no podían mantenerse abastecidas (Trujillo, 2016).

En los años siguientes, los gerentes se enfocaron en mejorar los sistemas de información y tener más control en cada una de las bodegas de los almacenes. En posteriores años, Walmart lanzó "My productivity app", enfocado a dar a los gerentes una visión más general de toda la cadena de suministros.

Por último, Best Buy, una compañía enfocada a la venta de artefactos electrónicos. En diciembre del 2011, en medio de las festividades Navideñas un sorprendente anuncio por parte de Best Buy dejó a muchos clientes atónitos. Best Buy no pudo con las órdenes de compra puestas por los clientes durante festividades y al no poder satisfacer la demanda Best Buy tuvo que disculparse por lo sucedido, los inconvenientes causados y notificar a los clientes que se han visto perjudicados (Trujillo, 2016).

Posteriormente, Best Buy tuvo que cancelar las órdenes puestas. Los clientes estuvieron enojados por la decisión tomada en lugar de retrasar las entregas. Aunque públicamente Best Buy nunca lo ha admitido se asume que se quedaron sin stock. En el artículo titulado “Como Best Buy se robó la navidad”. Se estima que una de las consecuencias de lo sucedido fue el daño de reputación y pérdida de mercado contra Amazon (Trujillo, 2016).

Las historias contadas en los párrafos anteriores naturalmente nos llevan a preguntarnos los beneficios hipotéticos que conllevarían implementar medidas y políticas que favorezcan el manejo y optimización de inventario. Uno de los principales beneficios es proveer a los negocios un mejor flujo tanto de productos como de efectivo. El resultado a mediano y largo plazo son las mejoras en competitividad, mejoras en el servicio y la flexibilidad en el manejo de procesos de operaciones.

La explicación de como el uso de métodos de manejo y optimización de inventario logran mejorar los niveles de servicio se puede relacionar a la necesidad de como en hoy en día se manejan los negocios y el empoderamiento de los clientes. Es decir, las organizaciones actualmente se enfocan en brindar una mejor experiencia a los consumidores siendo estos cada día más exigentes y críticos al momento de comprar o solicitar un servicio. En el caso de los negocios que cuenten con inventarios, es crítico que se implementen políticas y medidas sanas de reabastecimiento y control de inventario con el objetivo de satisfacer de manera apropiada y correcta la demanda de sus consumidores.

Los beneficios de establecer medidas de control y manejo de inventario también implican una reducción de costos de almacenaje (costos de infraestructura, costos de gestión, costos de operación). Lo mencionado se logra al mantener niveles adecuados de inventario.



Así, se puede evitar que algunos productos críticos como lo son los ítems de baja rotación, baja demanda se acumule en los anaqueles y bodegas causando costos adicionales que las empresas tienen que amortiguar. Al implementar esas soluciones existe la posibilidad de usar los recursos adicionales en otras áreas del negocio.

## 5.8. Cómo optimizar el inventario

Existen diferentes técnicas empleadas por las distintas organizaciones para garantizar que el control del inventario sea eficiente, rentable y con gastos mínimos. El objetivo de la optimización de inventario es el de generar el mayor beneficio con la menor inversión en inventario sin perjudicar la satisfacción del cliente (Fritsch, 2015).

Las estrategias y prácticas que son comúnmente usadas para reducir los niveles de inventario usualmente constan de pasos y fases, ya que optimizar el inventario es un proceso que requiere prestar atención en muchas áreas

Uno de los aspectos más importantes que se debe realizar es mantener un control sobre cada ítem almacenado es decir saber cantidad, ubicación y valor de cada uno. La creación de medidas para controlar el rendimiento y ciclos de vida de los productos y materiales tienen en el sistema de bodegaje. Medidas como ITR (Inventory Turnover Ratio) que se encargan de medir cuantas veces una compañía ha vendido y reemplazado el inventario en un tiempo dado. Entre más alto sea el indicador, este, nos mostrará que el tiempo en bodegas es menor y el ítem es más rentable para la compañía.

Debemos señalar que la planeación y análisis toman un papel importante dentro del proceso de optimización y manejo del inventario, siendo el pronóstico de la demanda (basado en datos históricos y variables del mercado) el componente inicial y al que más atención se debe prestar. Un pronóstico lo más preciso posible permite evitar acumulaciones de inventario o en su defecto quiebres de inventario en las distintas locaciones. En ocasiones, es necesario realizar un análisis riguroso de los productos, sobre todo cuando se muestran la existencia de eventos atípicos que afectan la tendencia y comportamiento del producto o simplemente cuando el producto muestre ventas bajas o intermitentes. Para el segundo caso, es necesario realizar una categorización de los productos que nos permita determinar si el producto es obsoleto (ya sea porque la tecnología lo has superado o porque existen productos más baratos en el mercado). De igual manera, todo ítem que presente una demanda con tendencia a la baja debería ser marcado en el sistema y con ello sus stocks de seguridad y puntos de reaprovisionamiento deben ser recalculados considerando el riesgo de obsolescencia (Fritsch, 2015).

## 5.9. Forecast Pro

Forecast Pro es un software comercial para el pronóstico de la demanda dentro de las organizaciones. Capaz de realizar múltiples pronósticos al mismo tiempo, cuenta con más de 32 años en el mercado con más de 40000 organizaciones usando las soluciones brindadas

por el software, entre los principales clientes podemos encontrar Anheuse-Bush (empresa madre de Cervecería Nacional), Bacardi, Coca Cola Enterprises Inc, etc.

La gran variedad de familias de modelos para el pronóstico ha hecho que se convierta en una herramienta muy versátil y útil en su uso en el entorno empresarial. Podemos hallar métodos tradicionales como: Suavización Exponencial (SE), modelos Box-Jenkins, Regresión Dinámica, Modelización de eventos, Modelo de Demanda Intermitente de Croston e inclusive modelos simples como "Lo mismo que el año pasado". (The right Tool for the Job, s.f.).

## 5.10. Teoría de grafos

La teoría de grafos es una de las áreas de las matemáticas cuyo desarrollo se ha dado por la necesidad de encontrar soluciones a problemas cotidianos. Por tanto, se ha convertido en una rama cuyas aplicaciones son ampliamente investigadas y usadas.

El primer artículo que hace referencia a la misma fue escrito por el matemático Euler con el famoso problema de los puentes de Königsberg; cuya historia y resolución se pueden encontrar en el artículo "Siete puentes, un camino: Königsberg" de Nuñez.

### 5.10.1. Concepto de un grafo

Los grafos pueden ser considerados formalmente como diagramas o dibujos, o bien algebraicamente como un par de conjuntos (Velázquez, 1998).

De forma geométrica podemos definir a un grafo  $G = (V, E)$  como un conjunto de puntos, los cuales, tienen la particularidad de que pueden estar unidos entre sí.

Por tanto, un grafo  $G = (V, E)$  es un par ordenado tal que,  $V$  es un conjunto de vértices también nodos o puntos y  $E \subseteq \{ \{ x, y \} | x, y \in V, x \neq y \}$  también conocido como aristas.

Estas simples definiciones nos ayudaran a entender y lograr formar un grafo que nos permita conectar todos los concesionarios a su respectiva bodega de aprovisionamientos y con eso modelar el sistema logístico deseado.

### 5.10.2. Representaciones de un grafo

Hay varias maneras de representar grafos, cada forma posee ventajas y características que permiten un mejor entendimiento tanto del sistema o ya sea para almacenarlo de manera más eficiente.

Una de las formas más sencillas para interpretar y entender la estructura, relación, vértices y aristas de un grafo es la representación gráfica; la ventaja de representar un grafo

de manera gráfica está ampliamente relacionado con el análisis, entendimiento y reconocimiento de patrones que se pueden dar al momento de expresarlo de esta manera. Si bien, visualmente en sistemas complejos es difícil encontrar las características antes mencionadas. Siempre se puede usar su forma gráfica para validación e interpretación de resultados.

Una representación gráfica debe mostrar claramente los elementos de los grafos mencionados en la sección anterior. El siguiente ejemplo es un grafo sencillo que se compone de  $V = \{1, 2, 3, 4, 5\}$ .

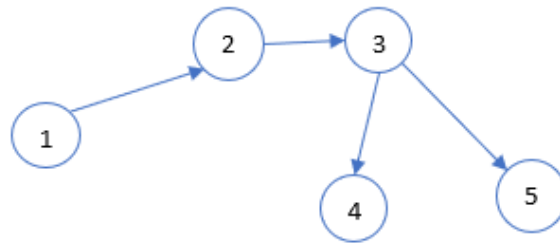


Figura 1: Ejemplo representación gráfica de un grafo

Otra de las formas para representar un grafo, más eficiente computacionalmente es la matriz de adyacencias. Para un grafo de  $|V|$  vértices, una matriz de adyacencia es una matriz de  $|V| \times |V|$  compuesta de ceros y unos, donde la entrada de la fila  $i$  en la columna  $j$  es 1 si y solo si la arista  $(i, j)$  está en el grafo. Tomando el ejemplo de la gráfica anterior la matriz de adyacencia del grafo es:

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix} \quad (1)$$

La representación matricial será usada para el modelamiento del sistema logístico de Ambacar.

## 6. Pronóstico de la Demanda

### 6.1. Introducción a los Random Forest

Los primeros intentos de implementación y creación de métodos Random Forest, es con el uso de métodos bagging (Breiman, 1996). Recordando que el principal objetivo de estos métodos es de reducir la varianza, esto se basaba en el crecimiento de los árboles usando un selector aleatorio.

Los modelos Random Forest tuvieron sus inicios con los avances significativos que se generaron al dejar crecer los conjuntos de árboles ya sea con métodos bagging o algún otro método de crecimiento y dejar que los mismos a través de procesos aleatorios tomen la mejor decisión en base a la información que tienen disponible. Para lograr dicho cometido, usualmente se generaron vectores aleatorios que controlan el crecimiento de cada árbol en el conjunto (Breiman, 2001).

Con el paso del tiempo otras técnicas para el crecimiento de árboles fueron puestas en práctica, como lo es la partición aleatoria (Dietterich, 1998). Donde en cada nodo la nueva partición es seleccionada de manera aleatoria de un conjunto  $\mathbf{K}$  de las mejores particiones. Muchos métodos incluyen la generación de nuevos conjuntos de entrenamiento al aleatorizar los resultados obtenidos del conjunto de entrenamiento original. Otras aproximaciones consisten en seleccionar el conjunto de entrenamiento de un conjunto aleatorio de pesos, haciendo una selección aleatoria de subconjuntos de atributos que se pueden usar para hacer crecer un árbol (Ho, 1998).

El elemento común entre todos estos procedimientos es que para el  $k$ -ésimo árbol, un vector aleatorio  $\Theta_k$  es generado, teniendo como principal característica la independencia del resto de vectores aleatorios  $\Theta_1, \dots, \Theta_{k-1}$ , pero conservando la misma distribución; posteriormente, los árboles crecerán usando el conjunto de entrenamiento y el vector  $\Theta_k$ , esto genera un clasificador  $h(x, \Theta_k)$  donde  $\mathbf{x}$  es un vector de inicialización (Breiman, 2001). Por ejemplo, en el caso de la metodología bagging el vector aleatorio  $\Theta$  es seleccionado como la generación de los conteos en  $N$  cajas como resultados de  $N$  lanzamientos al azar en las cajas, podemos interpretar a  $N$  como el número de datos en el conjunto de entrenamiento. La naturaleza y dimensión del vector  $\Theta$  depende de su uso en la construcción del árbol (Breiman, 2001).

Con los antecedentes presentados, cuando se genera un número lo suficientemente grande de árboles, estos tienen la capacidad de elegir la mejor dirección dentro su clase. este procedimiento se ha denominado Random Forest.

Formalmente, podemos definir a los Random Forest de la siguiente manera.

*Definición .1.* Los Random Forest son un clasificador que consiste en una colección de clasificadores con estructura de árboles.  $(h(x, \Theta_k), k = 1, \dots)$ , donde los  $\Theta_k$  son vectores aleatorios idénticamente distribuidas, donde cada árbol tiene derecho a un voto unitario

para seleccionar a la clase más popular de la entrada  $x$ .

## 6.2. Cualidades generales de los Random Forest

### 6.2.1. Caracterización de la eficacia de los Random Forest

Sea un conjunto de clasificadores  $h_1(x), h_2(x), \dots, h_k(x)$  y un conjunto de entrenamiento seleccionado aleatoriamente de la distribución  $Y, X$ , definimos la función marginal de la misma como:

$$mg(X, Y) = av_k I(h_k(X) = Y) - \max_{j \neq Y} av_k I(h_k(X) = j)$$

Donde  $I(\cdot)$  es la función indicatriz. La medida marginal mide la extensión en la cual el número promedio de votos en  $(X, Y)$  para la clase que supere la votación de cualquier otra clase.

Mientras más amplio sea el margen, más confianza se obtiene en la selección (Breiman, 2001).

El error de generalización viene dado por:

$$PE = P_{X,Y}(mg(X, Y) < 0) \tag{2}$$

La ecuación anterior se refiere a la probabilidad de que la función marginal sea menor que 0 en el espacio  $(X, Y)$ . Como consecuencia podemos obtener el siguiente teorema sobre la convergencia de los Random Forest.

*Teorema 1.* Conforme aumenta el número de árboles, casi seguramente toda las sucesiones de vectores aleatorios  $\Theta_1, \dots$ , el error de generalización  $PE$  converge a:

$$P_{X,Y}(P_{\Theta}(h(X, \Theta) = Y) - \max_{j \neq Y} P_{\Theta}(h(X, \Theta) = j) < 0)$$

La demostración del teorema se puede encontrar en el artículo “Random Forest” de Leo Breiman citado anteriormente.

Lo más importante que produce el teorema anterior es el resultado que explica que los Random Forest no causan una sobre especificación conforme se añaden más árboles, pero producen un valor limitante al error de generalización (Breiman, 2001).

## 6.3. Random Forest para regresión

Para el caso de regresión, los Random Forest se generan cuando crecen los árboles que los componen; siempre y cuando dependan del vector aleatorio  $\Theta$  tales que el predictor  $h(x, \Theta)$  tome valores numéricos en lugar de escoger clases como fue mencionado en la sección anterior.

Los valores que se obtienen al realizar esta metodología deben ser numéricos y debe provenir de un conjunto de entrenamiento que sea independiente de los vectores aleatorios  $(X, Y)$ .

El predictor para los Random Forest en regresión se forma al tomar el promedio sobre el  $k$  número de árboles.

### 6.3.1. Caracterización de la eficacia de los Random Forest en regresión

Para estudiar e inferir la convergencia de los Random Forest empezamos estimando el error cuadrático de la generalización para cualquier predictor  $h(x)$  es  $E_{x,y}(Y - h(x))^2$  (Breiman, 2001).

Al igual que en el caso de clasificación que fue el primer inicio de los Random Forest tenemos un teorema que nos garantiza que al aumentar el número de árboles en los modelos estos no causan una sobre especificación. A continuación; el teorema que su demostración puede ser encontrada en el artículo “Random Forest” de Leo Breiman.

*Teorema 2.* Conforme aumenta el número de árboles, casi seguramente

$$E_{X,Y}(Y - \text{av}_k h(x, \theta_k))^2 \rightarrow E_{X,Y}(Y - E_\theta h(x, \theta))^2$$

Del resultado del teorema anterior podemos inferir que el error de la generalización del modelo, es decir,  $PE = E_{X,Y}(Y - E_\theta h(x, \theta))^2$ , de igual manera podemos generalizar el error para cada árbol.

$$PE^* = E_\Theta E_{X,Y}(Y - E_\theta h(x, \theta))^2$$

Como siguiente paso, debemos hallar una relación entre los errores de generalización para todo el modelo como para cada árbol basado en la baja correlación entre los residuos y los errores generados por la adición de árboles.

### 6.3.2. Correlación y robustez en los Random Forest

Uno de los beneficios del uso de los modelos Random Forest es su facilidad para medir el impacto y la precisión que tiene los clasificadores y la dependencia entre ellos.

Para ello, haremos uso de la siguiente definición.

*Definición .2.* La función marginal para un Random Forest es

$$mr(X, Y) = (P_\Theta(h(X, \Theta) = Y) - \max_{j \neq Y} P_\Theta(h(X, \Theta) = j) < 0)$$

y definimos la robustez del conjunto de clasificadores  $\{h(X, \Theta)\}$  como

$$s = E_{X,Y}mr(X, Y)$$

Podemos interpretar la robustez  $s$  como una medida de precisión para cada árbol dentro del conjunto de árboles, asumiendo que  $s \geq 0$ ; por la desigualdad de Chebyshev tenemos que

$$PE_* \leq var(mr)/s^2$$

De manera similar podemos expresar la varianza de la función marginal de la siguiente manera.

*Definición .3.* Sea

$$\hat{j}(X, Y) = argmax_{j \neq Y}(P_{\Theta}(h(X, \Theta) = j))$$

entonces

$$\begin{aligned} mr(X, Y) &= P_{\Theta}(h(X, \Theta) = Y) - P_{\Theta}(h(X, \Theta) = \hat{j}(X, Y)) \\ &= E_{\Theta}(I(h(X, \Theta) = Y) - I(h(X, \Theta) = \hat{j}(X, Y))) \end{aligned}$$

De igual forma, podemos definir una segunda función marginal de la siguiente manera.

*Definición .4.* Sea

$$rmg(\Theta, X, Y) = I(h(X, \Theta) = Y) - I(h(X, \Theta) = \hat{j}(X, Y))$$

Por tanto,  $mr(X, Y)$  es la esperanza de  $rmg(\Theta, X, Y)$  respecto a  $\Theta$ . Entonces, para cualquier función  $f$  se cumple la identidad

$$[E_{\Theta}f(\Theta)]^2 = E_{\Theta, \Theta^*}f(\Theta)f(\Theta^*)$$

cuando  $\Theta$  y  $\Theta^*$  sean independientes e idénticamente distribuidos, lo cual significa que

$$mr(X, Y)^2 = E_{\Theta, \Theta^*}rmg(\Theta, X, Y)rmg(\Theta^*, X, Y)$$

Realizando operaciones algebraicas y usando la función marginal cuadrada anterior podemos obtener

$$var(mr) = E_{\Theta, \Theta^*}(cov_{X,Y}(rmg(\Theta, X, Y)rmg(\Theta^*, X, Y)))$$

$$var(mr) = E_{\Theta, \Theta^*}(\rho(\Theta, \Theta^*)sd(\Theta)sd(\Theta^*))$$

Si tomamos  $\Theta$  y  $\Theta^*$  fijas entonces  $\rho(\Theta, \Theta^*)$  es la correlación entre  $rmg(\Theta, X, Y)$  y  $rmg(\Theta^*, X, Y)$  y  $sd(\Theta)$  es la desviación estándar de  $rmg(\Theta, X, Y)$ , con los resultados obtenidos podemos expresar la varianza de la función marginal de la siguiente manera

$$var(mr) = \rho^*(E_{\Theta}sd(\Theta))^2$$

$$var(mr) \leq \rho^*E_{\Theta}var(\Theta)$$

donde  $\rho^*$  es el promedio de la correlación; eso es,

$$\rho^* = E_{\Theta}E_{\Theta^*}(\rho(\Theta, \Theta^*)sd(\Theta)sd(\Theta^*)) / (E_{\Theta}sd(\Theta))^2$$

Con todas las definiciones presentadas podemos expresar el siguiente teorema que nos ayudará a mostrar la necesidad para que el modelo Random Forest sea preciso, es decir, una baja correlación entre los residuos y árboles de bajo error. Por tanto, se debe cumplir el siguiente teorema.

*Teorema 3.* Para todo  $\Theta$ ,  $EY = E_Xh(X, \theta)$ . Entonces:

$$PE^*(RandomForest) \leq \rho^*PE^*(\text{árbol})$$

donde  $\rho^*$  es un peso de la correlación entre los residuos  $Y - h(X, \Theta)$  y  $Y - h(X, \Theta^*)$ , y además  $\Theta$  y  $\Theta^*$  son independientes.

### **Demostración:**

Por la definición de error de generalización de cada árbol tenemos:

$$\begin{aligned} PE^* &= E_{\Theta}E_{X,Y}(Y - E_*h(X, \Theta))^2 \\ &= E_{\Theta}E_{\Theta^*}E_{X,Y}(Y - h(x, \Theta))(Y - h(x, \Theta^*)) \end{aligned}$$

El último término concuerda con la definición de covarianza. Por tanto, podemos escribirlo como:

$$E_{\Theta}E_{\Theta^*}(\rho(\Theta, \Theta^*)sd(\Theta)sd(\Theta^*)) \tag{3}$$

donde  $sd(\Theta) = \sqrt{E_{X,Y}(Y - h(X, \Theta))^2}$ , usando álgebra y despejando términos podemos definir el parámetro de correlación como:

$$\rho^* = E_{\Theta}E_{\Theta^*}(\rho(\Theta, \Theta^*)sd(\Theta)sd(\Theta^*)) / (E_{\Theta}sd(\Theta))^2 \tag{4}$$

Obtenemos reemplazando la ecuación anterior.

$$PE^*(RandomForest) = \rho^*(E_{\Theta}sd(\Theta))^2 \leq \rho^*PE^*(\text{árbol})$$

□

El resultado del teorema anterior nos indica que los Random Forest disminuyen el error de los árboles usados por el factor  $\rho^*$  (Breiman, 2001). La aleatorización empleada debe enfocarse a bajar la correlación.

Con las caracterizaciones presentadas el objetivo es buscar y entrenar un árbol con una robustez  $s$  lo más cercana a 1 posible y un valor  $\rho^*$  bajo.



### 6.3.3. Principios básicos

El objetivo de esta sección es proveer de una representación matemática precisa como una antesala para la presentación del algoritmo de Random Forest para regresión.

El marco general de trabajo es una estimación no paramétrica, en donde las variables iniciales son un vector aleatorio  $X \in \chi \subset \mathbb{R}^p$ . La tarea principal es predecir una respuesta aleatoria  $Y \in \mathbb{R}$  estimando una función de regresión  $h(x) = E\{Y|X=x\}$ .

En base a lo presentado en el párrafo anterior asumimos una muestra de entrenamiento  $D_n = \{(X_1, Y_1), \dots, (X_n, Y_n)\}$  que tiene como propiedades su formación compuesta por variables aleatorias independientes que heredan la distribución del par  $(X, Y)$ . El siguiente paso es partir de la muestra de entrenamiento  $D_n$  y construir un estimador  $h_n : \chi \rightarrow \mathfrak{R}$  de la función  $h$ . Es importante señalar que llamaremos al estimador  $h_n$  consistente si se verifica que:

$$E[h_n(X) - h(X)]^2 \rightarrow 0 \text{ cuando } n \rightarrow \infty \quad (5)$$

La esperanza precedente es evaluada en  $X$  y  $D_n$ .

Los Random Forest son predictores que consisten en una colección de  $M$  árboles aleatorios (Biau y Scornet, 2016). Para el  $j$ -ésimo árbol en la familia, el valor pronosticado en el punto  $x$  es denotado por  $h_n(x; \Theta_j, D_n)$ , donde los  $\Theta_1, \dots, \Theta_M$  son variables aleatorias independientes, la distribución es heredada de  $\Theta$  e independiente de  $D_n$ .

Usaremos la variable  $\Theta$  para generar una nueva muestra y se obtiene del conjunto de entrenamiento, este paso es primordial para generar los crecimientos de los árboles individuales y generar las direcciones para las siguientes divisiones. En términos generales definimos al estimador  $h_n$  del  $j$ -ésimo árbol de la siguiente manera:

$$h_n(x, \Theta_j, D_n) = \sum_{i \in D_n^*(\Theta_j)} \frac{\mathbb{1}_{X_i A_n(x; \Theta_j, D_n)} Y_i}{N_n(x; \Theta_j, D_n)} \quad (6)$$

Donde  $D_n^*(\Theta_j)$  es el conjunto de datos seleccionados antes de la construcción de los árboles,  $A_n(x; \Theta_j, D_n)$  es la celda que contiene a  $x$ , y  $N_n(x; \Theta_j, D_n)$  es el número de puntos que son preseleccionados que caen sobre  $A_n(x, \Theta_j, D_n)$ .

En base a la ecuación anterior, podemos inferir el estimador finito del conjunto de árboles que denominaremos: estimador del bosque como:

$$h_{M,n}(x, \Theta_j, \dots, \Theta_M, D_n) = \frac{1}{M} \sum_{j=1}^M h_n(x; \Theta_j, D_n) \quad (7)$$

Ya que como  $M$  puede ser escogida arbitrariamente, desde un punto de vista de modelamiento, podemos permitir que  $M$  tiende a  $\infty$ , y considerar en lugar de la ecuación anterior para el estimador del bosque  $h_{\infty,n}(x, D_n) = E_{\Theta} [h_n(x, \Theta, D_n)]$  (Biau y Scornet, 2016).

A través de la definición expuesta,  $E_{\Theta}$  representa la esperanza respecto al parámetro  $\Theta$ , condicional a  $D_n$ . De hecho, debemos notar que la operación  $M \rightarrow \infty$  se justifica por la ley de los grandes números (Biau y Scornet, 2016).

$$\lim_{M \rightarrow \infty} (h_{M,n}(x, \Theta_1, \dots, \Theta_M, D_n) = h_{\infty,n}(x, D_n)$$

Notemos que es una convergencia casi segura (Breiman, 2001).

### 6.3.4. Criterio CART

El criterio CART es un algoritmo para la construcción de árboles de decisión basado en el índice de Gini como criterio de partición (Breiman, 1983). Por lo tanto, el criterio CART es un proceso repetitivo de construcción de árboles binarios en donde se busca encontrar la mejor división del árbol posible que maximice el criterio de partición.

Recordemos, además, que estos criterios representan una complejidad media dependiendo del modelo que se trata de construir, la complejidad de encontrar la mejor partición posible está ligado al número de variables o también llamado direcciones *mtry*.

### 6.3.5. Algoritmo Random Forest

Para inferir un algoritmo de regresión con el método Random Forest debemos citar como es la construcción de los árboles y el papel que tiene la aleatoriedad dentro del proceso. En primer lugar; cada nodo de cada árbol es asociado con una celda hiperrectangular donde la raíz de cada árbol es  $\chi$ . En los pasos siguientes de la construcción de los árboles los nodos son particionados en dos partes. Los nodos terminales (u hojas) si son coleccionadas juntas, estas forman una partición de  $\chi$ .

El algoritmo funciona cuando los  $M$  diferentes árboles crecen de forma aleatoria siguiendo un proceso de elaboración por etapas. En la etapa de construcción de cada árbol, las observaciones son seleccionadas aleatoriamente con la posibilidad de ser con o sin reposición respecto a la data original. Estas observaciones  $a_n$  que han sido seleccionadas son las únicas tomadas en cuenta para la edificación del árbol. Entonces, en cada celda del árbol, una partición se genera al maximizar el criterio CART sobre las direcciones *mtry* elegidas uniformemente de manera aleatoria sobre las  $w$  originales (el subconjunto resultante de las direcciones seleccionadas se denominará *Mtry*). Por último, la construcción de los árboles individuales se detiene cuando las celdas contienen menos puntos que el parámetro *nodesize*. Para cualquier  $x \in \chi$ , cada árbol de regresión predice el promedio de los  $Y_i$  dentro de su conjunto de observaciones  $a_n$  y para sus respectivos  $X_i$  que recaen en la celda correspondiente a  $x$ . Notemos que los árboles crecientes y la predicción final solo dependen de los  $a_n$  puntos seleccionados de la data (Biau y Scornet, 2016).

Empezamos notando que el algoritmo tiene tres importantes componentes:

- $a_n \in \{1, \dots, n\}$ , representa el número de datos muestreados en cada árbol.

- $mtry \in \{1, \dots, w\}$ , representa el número de posibles direcciones para la partición de los nodos de cada árbol.
- $nodesize \in \{1, \dots, a_n\}$ , representa el número de ejemplos de cada celda por el cual al estar más abajo de ese nivel no se realizan más particiones.

Para los criterios de partición estos siguen una metodología CART, para ejemplificar su uso supondremos todo un conjunto de datos  $D_n$ ,  $A$  una celda genérica y denotemos por  $N_n(A)$  el número de puntos en la data que son contenidos en  $A$ . El corte de  $A$  en  $A$  es un par  $(j, z)$ , donde  $j$  es algún valor entre  $\{1, \dots, w\}$  y  $z$  es la posición del corte lo largo de la  $j$ -ésima coordenada, dentro de los límites de  $A$ . Sea  $C_A$  el conjunto de todos los posibles cortes de  $A$ . Entonces, usamos la notación  $X_i = (X_i^{(1)}, \dots, X_i^{(w)})$ , para cualquier par  $(j, z) \in C_A$ . El criterio de partición CART toma la siguiente forma:

$$L_{reg,n} = \frac{1}{N_n(A)} \sum_{i=1}^n (Y_i - \bar{Y}_A)^2 \mathbb{1}_{x_i \in A} - \frac{1}{N_n(A)} \sum_{i=1}^n ((Y_i - \bar{Y}_{A_L} \mathbb{1}_{X_i^{(j)} < z} - \bar{Y}_{A_R} \mathbb{1}_{X_i^{(j)} \geq z})^2) \mathbb{1}_{x_i \in A} \quad (8)$$

Donde  $A_L = \{x \in A : x^{(j)} < z\}$ ,  $A_R = \{x \in A : x^{(j)} \geq z\}$ , además sean  $\bar{Y}_A, \bar{Y}_{A_L}, \bar{Y}_{A_R}$  las medias respectivas de los  $Y_i$  con respecto a  $A, A_L, A_R$  con la suposición necesaria que le media es igual a 0 cuando  $x_i \notin (A, A_L, A_R)$ . Para cada celda  $A$ , el mejor corte que se puede obtener es el par  $(j_n^*, z_n^*)$ , este se obtiene al maximizar  $L_{reg,n}(j, z)$  sobre  $mtry$  y  $C_A$ , en términos matemáticos podemos expresar el problema de maximización como:

$$(j_n^*, z_n^*) \in \underset{\substack{j \in Mtry \\ (j,z) \in C_A}}{\operatorname{argmax}} L_{reg,n}(j, z)$$

Entonces, en cada celda de cada uno de los árboles, el algoritmo escoge de manera uniforme y aleatoria  $mtry$  en las coordenadas  $(1, \dots, w)$ . Posteriormente se realizan criterios de evaluación de todos los posibles cortes a lo largo de todas las direcciones y se retorna la mejor dirección (Biau y Scornet, 2016).

Por tanto, bajo los antecedentes presentados podemos presentar un algoritmo para Random Forest. Los datos iniciales que se necesitan para la utilización del procedimiento son un conjunto de entrenamiento  $D_n$ , número de árboles  $M > 0$ ,  $mtry \in \{1, \dots, p\}$ ,  $nodesize \in \{1, \dots, a_n\}$ , y  $x \in \chi$ . Adicionalmente, haremos uso de la función  $Concatenar(\cdot)$ . la función realiza la tarea de coleccionar y unir las celdas dentro de una lista asociada a un árbol.

---

**Algorithm 1** RandomForest( $x, D_n, M, mtry, nodesize$ )

---

**Result:** Predicción del Random Forest en  $x$ **for**  $j = 1, \dots, M$  **do**    Seleccionar  $a_n$  puntos de manera aleatoria y uniforme de  $D_n$ . En los siguientes pasos, solo estas observaciones serán usadas.    Seleccionamos  $P = (\chi)$ , la lista que contiene todas las celdas asociadas a la raíz del árbol.    Seleccionamos  $P_{final} = \emptyset$ .    **while**  $P \neq \emptyset$  **do**        Sea  $A$  el primer elemento de  $P$         **if**  $A$  contiene menos puntos que  $nodesize$  o si todos los  $x_i \in A$  son iguales **then**            Remover la celda  $A$  de la lista  $P$              $P_{final} \leftarrow Concatenar(P_{final}, A)$         **else**            Seleccionar uniformemente, sin reposición, un subconjunto  $Mtry \subset \{1, \dots, p\}$  de cardinalidad  $mtry$ .            Seleccionar la mejor partición de  $A$  al optimizar el criterio de partición CART a lo largo de las coordenadas  $Mtry$ .            Cortar la celda  $A$  de acuerdo a la mejor partición, llamaremos a estas dos celdas  $A_L, A_R$ .            Remover la celda  $A$  de la lista  $P$              $P \leftarrow Concatenar(P, A_L, A_R)$         **end**        Calcular el valor pronosticado  $h_n(x, \Theta_j, D_n)$  en  $x$  que es igual al promedio de  $Y_i$  que caen sobre  $x$  en la partición  $P_{final}$     **end**    Calcular el estimado del Random Forest  $h_{M,n}(x, \Theta_1, \dots, \Theta_M)$ **end**

---

### 6.3.6. Mecanismos alternativos de remuestreo

Los problemas de regresión y pronóstico de la demanda con el método Random Forest requieren un mecanismo de remuestreo, siguiendo las ideas del trabajo realizado por Breiman (2001), el algoritmo Random Forest trabaja al escoger igual número de puntos por la cantidad de veces que se corre el procedimiento con substitución para calcular los estimados del árbol. Este procedimiento adquiere el nombre de bootstrap.

Naturalmente, la idea de generar muchas muestras bootstrap y promediar los predictores se llama bagging o agregación bootstrap. El objetivo de esta técnica es mejorar la eficacia de los mecanismos de aprendizaje débiles o inestables (Biau y Scornet, 2016). La complejidad del mecanismo bootstrap proviene del hecho que la distribución del conjunto  $D_n^*$  es distinta al del conjunto original de data  $D_n$ . Esto se puede ver fácilmente con un ejemplo, Si asumimos que  $X$  tiene una densidad cualquiera, si los puntos de la data son muestreados con reposición, con una probabilidad positiva, se va a seleccionar más de una vez a al menos una observación. Esto ocasiona, que existan dos puntos idénticos en  $D_n^*$  y

que la distribución de la misma no sea absolutamente continua.

Con el paso de los años el rol que tiene bootstrap dentro de Random Forest sigue siendo poco comprendido y en muchas ocasiones subestimado. Por ende; muchos de los análisis que se realizan tienen a reemplazar la metodología bootstrap por un esquema basado en submuestras. Asumiendo que cada árbol crece con  $a_n < n$  observaciones escogidas aleatoriamente con reposición de la muestra original (Wager, 2014). La mayor parte del tiempo, la tasa de las submuestras  $a_n/n \rightarrow 0$  a una velocidad predicha por el modelo. Cuando se cumple esa condición se puede decir que el modelo es consistente.

En consecuencia, a pesar de que la eficacia del modelo está netamente ligada al conjunto de observaciones  $D_n$ , existen muchos estudios que nos pueden asegurar que con un poco de complejidad nos podemos desligar de la dependencia al conjunto inicial por medio de estadística y convergencias de subconjuntos.

### 6.3.7. Decisión de partición y propiedades de los nodos hijos

La división coordinada del algoritmo de Random Forest es un proceso con alta dificultad, ya que desde un inicio usa las variables  $X_i, Y_i$  para realizar sus decisiones. Una de las interrogantes que surgieron fue establecer límites para la decisión de partición en el contexto del modelo de regresión de la forma  $Y = h(x) + \epsilon$  (Bühlmann y Yu, 2002). Donde  $X$  es un valor real y  $\epsilon$  es un ruido blanco. En esencia, el resultado se puede interpretar de la siguiente manera. Si asumimos que la distribución de  $(X, Y)$  es conocida, y la partición óptima  $d^*$  obtenida por el criterio CART en un nodo cualquiera. Para las expresiones de los hijos tanto derecho como izquierdo toman la siguiente forma:

$$\beta_{l,n}^* = E[Y|X \leq d^*] \text{ y } \beta_{r,n}^* = E[Y|X > d^*] \quad (9)$$

En el caso que la distribución de  $X, Y$  no son conocidas, evidentemente  $\beta_{l,n}^*, \beta_{r,n}^*$  y  $d^*$  también son desconocidas y deben ser estimadas por sus contrapartes empíricas.

$$(\widehat{\beta}_{l,n}^*, \widehat{\beta}_{r,n}^*, \widehat{d}^*) \in \underset{(\beta_l^*, \beta_r^*, d^*)}{\operatorname{argmin}} \sum_{i=1}^n [Y_i - \beta_l \mathbb{1}_{(x_i \leq d)} - \beta_r \mathbb{1}_{(x_i > d)}]^2 \quad (10)$$

Si agregamos condiciones de regularidad, es decir que  $X$  tiene una densidad  $f$ , y además  $f$  y  $h$  son continuamente diferenciables, se verifica:

$$n^{\frac{1}{3}} \begin{pmatrix} \widehat{\beta}_{l,n}^* - \beta_l^* \\ \widehat{\beta}_{r,n}^* - \beta_r^* \\ \widehat{d}^* - d^* \end{pmatrix} \xrightarrow{d} \begin{pmatrix} c_1 \\ c_2 \\ 1 \end{pmatrix} \underset{t}{\operatorname{argmax}} (aW(t) - bt^2) \quad (11)$$

La demostración se puede hallar en el trabajo de Banerjee y McKeague (2007), notemos que  $W$  es un movimiento browniano estándar, con  $a$  y  $b$  constantes positivas que dependen de los parámetros del modelo y las cantidades  $\beta_l^*, \beta_r^*, d^*$ . El beneficio que nos otorga la

convergencia anterior es que nos permite construir intervalos de confianza para la posición de partición usando criterio CART (Biau y Scornet, 2016).

El resultado anterior nos indica que los puntos de partición no solo sirven para la creación de árboles y bosques, pero que también pueden proveer información importante sobre la distribución de la estructura subyacente de los mismos.

Con el paso de los años se han realizado muchos análisis con respecto al comportamiento en la partición del bosque. Sobre todo; respecto a la preferencia del corte final (ECP) del criterio de partición CART. ECP hace referencia a la ubicación dentro del bosque de las particiones que se generan dentro del mismo con las variables no informativas y que no aportan de manera significativa al modelo, una propiedad muy deseada es que se ubiquen en los filos y extremos del bosque.

El mecanismo de aleatorización cuando es puesto en práctica dentro del bosque, hay la probabilidad positiva que ninguna de las variables seleccionadas dentro del nodo sea significativa (Biau y Scornet, 2016). En el caso de que ocurra, y si el corte tiene lugar, digamos, en el centro del nodo, consecuentemente, el tamaño de los dos nodos que resultan de esa partición se ve drásticamente reducido por un factor 2 lo que puede resultar muy dañino para la tarea de predicción. Por lo tanto, la propiedad ECP nos asegura que una división a lo largo de una variable que no es de tanta relevancia se realiza cerca de los filos, así se logra maximizar el tamaño de la muestra del nodo y hace posible para el árbol que se recupere de la mala partición (Biau y Scornet, 2016). Inclusive, esta propiedad puede ser un gran beneficio inclusive cuando la partición se realiza sobre variables informativas (Ishwaran, 2015).

Lo expuesto en el párrafo anterior, nos indica que los modelos Random Forest tiene una eficacia asintótica, con una alta probabilidad, que las particiones las realice dentro de un conjunto  $S$  de variables informativas (Scornet et al, 2015). Si fijamos el número posible de direcciones para la partición de los nodos de cada árbol  $mtry = t$ , donde  $t$  es el número total de variables disponibles para la construcción del modelo, para todo  $k \leq t$  notemos como  $j_{n,1}(X), \dots, j_{n,k}(X)$  a los primeros  $k$  direcciones de corte usados en la construcción de la celda de  $X$ , con la suposición de que  $j_{n,k}(X) = \infty$  si la celda ha sido cortada menos que  $q$  veces. Asumiendo algunas condiciones de regularidad del modelo de regresión se puede probar que la probabilidad  $1 - \epsilon$  para  $n$  suficientemente grande y para todo  $1 \leq q \leq k$  se verifica que

$$j_{n,q}^* \in \{1, \dots, S\} \tag{12}$$

El resultado anterior ofrece un argumento solido de porque los Random Forest se adaptan de manera favorable ante una configuración escasa, de hecho; muestra que el algoritmo presentado selecciona en gran mayoría particiones a largo de las  $S$  variables informativas. En resumen, todo el proceso en su mayoría se realiza como si la fuera proyectada dentro del espacio vectorial generado por esas variables (Biau y Scornet, 2016).

### 6.3.8. Optimización de Parámetros

La literatura enfocada y especializada a la optimización de parámetros  $M$ ,  $mtry$ ,  $nodesize$  es desafortunadamente rara (Biau y Scornet, 2016). Existen excepciones, como las que podemos encontrar en los trabajos realizados por Díaz-Uriarte y De Andrés (2006), Bernard et al. (2008), y Genuer et al. (2010), la optimización de los hiperparámetros puede resultar ser un trabajo muy complejo, en particular para data muy grande, con miles de muestras y variables.

Los objetivos de optimizar los parámetros es la reducción de la varianza, es fácil apreciar que la varianza de los bosques decrece conforme el número de los árboles  $M$  aumenta (Biau y Scornet, 2016). Entonces, para pronósticos y predicciones más precisas son comúnmente obtenidas con un mayor número de árboles. Como ya fue mencionado en secciones anteriores, un largo número de árboles no causa una sobrestimación.

Sin embargo, el costo computacional de inducir un bosque aumenta linealmente con  $M$  (Biau y Scornet, 2016). Una buena elección de  $M$  se puede considerar de la combinación entre la complejidad computacional y la eficacia del modelo; al considerar un  $M$  grande nos llevaría a que el modelo no termine de realizar los cálculos en un tiempo computacional razonable, por otra parte, se necesita un  $M$  lo suficientemente grande para que las predicciones sean estables.

Hay muchos investigadores que tienen un diferente pensamiento de como elegir  $M$ , muchos de ellos consideran, el valor de  $M$  es totalmente irrelevante cuando los datos son sustanciosos (Díaz-Uriarte y De Andrés, 2006). Otra aproximación interesante que se considera es la posibilidad de hallar gráficamente un mínimo número de árboles para obtener una predicción con niveles de precisión similares que al usar una cantidad grande árboles, estos resultados muestran que es posible hallar un límite al número de árboles que se necesitan (Latinne et al., 2001).

Para el parámetro  $nodesize$ , a pesar de que no hay una evidencia sólida y una teoría sustancial que la respalde se considera que 5 es un buen valor para regresión, este valor ha sido considerado como una buena elección (Díaz-Uriarte y De Andrés, 2006). El efecto que puede tener el parámetro  $mtry$  es de poco impacto en la eficacia del método, aunque valores muy altos están asociados a la reducción de la eficacia del modelo (Díaz-Uriarte y De Andrés, 2006). Una aproximación racional considera tomar  $mtry = p$ , donde  $p$  es la dimensión de  $X_i$ .

Finalmente, debemos notar que no hay garantía basada en la teoría que nos asegure que los parámetros de *default* sean los correctos para el modelo. Sin embargo, estos parámetros son fáciles de optimizar sin que se requiera un conjunto de validación independiente (Biau y Scornet, 2016).

### 6.3.9. Importancia de las variables

Uno de los beneficios de usar una metodología Random Forest es que por medio de diferentes medidas nos permite conocer que tan importante es una variable dentro del modelo de regresión. El método más usado se denomina MDA (Disminución Eficacia Promedio). Surge de la idea que, si la variable no es importante, entonces reordenar sus valores no debería degradar la eficacia de la predicción (Biau y Scornet, 2016).

El método MDA confía en el uso de la estimación del error OBB. De manera abreviada para medir la importancia de la  $j$  -ésima variable, aleatoriamente se permutan los valores de la variable  $X^{(j)}$  en los valores de las observaciones OBB y se colocan esos valores permutados en la parte de abajo del árbol. El MDA de  $x^{(j)}$  se obtiene al promediar la diferencia entre la estimación del error OBB antes y después de la permutación sobre todos los árboles. En términos matemáticos, podemos considerar la variable  $x^{(j)}$  Y  $D_{l,n}$  como el conjunto de datos OBB del  $l$  -ésimo árbol y  $D_{l,n}^j$  el mismo conjunto de datos pero con la diferencia que los valores  $x^{(j)}$  han sido permutados aleatoriamente dentro de ese conjunto. Recordando el estimador para el  $l$  -ésimo es  $h_n(\cdot, \Theta_l)$ . Entonces, podemos definir a la medida MDA como:

$$MDA(\widehat{X}^{(j)}) = \frac{1}{M} \sum_{l=1}^M [R_n [h_n(\cdot; \Theta_l), D_{l,n}^j] - R_n [h_n(\cdot; \Theta_l), D_{l,n}]] \quad (13)$$

Donde  $R_n$  se define para  $D = D_{l,n}^j$  y  $D = D_{l,n}$  por:

$$R_n [h_n(\cdot, \Theta_l), D] = \frac{1}{|D|} \sum_{i:(X_i, Y_i) \in D} ((Y_i - h_n(X_i, \Theta))^2) \quad (14)$$

Resumiendo, las dos ecuaciones anteriores podemos encontrar una expresión más reducida para MDA.

$$MDA(X^{(j)})^* = E [Y - h_n(X_j^*, \Theta)]^2 - E [Y - h_n(X, \Theta)]^2 \quad (15)$$

Donde  $X_j^* = (X^{(1)}, \dots, X^{(j)*}, \dots, X^{(p)})$  y  $X^{(j)*}$  es una copia idéntica e independiente de  $X^{(j)}$ .

### 6.3.10. Uso de variables aleatorias

Para mejorar la exactitud, el efecto de la aleatorización debe minimizar la correlación  $\rho$  mientras mantiene la fortaleza característica de los Random Forest. En general, los bosques que se han generado usan variables elegidas aleatoriamente en cada nodo para así, lograr que cada árbol crezca (Breiman, 2001). Este tipo de procedimientos tiene características muy favorables y deseadas. A continuación, citaremos alguna de estas características.



- La exactitud de los Random Forest es tan buena como lo de los métodos Adaboost e inclusive mejor en muchos casos.
- Es robusto con puntos extremos y ruido dentro de la data.
- Es relativamente más rápido por su baja complejidad computacional.
- La información de retorno que brinda nos permite analizar la información sobre estimación de los errores, correlación e importancia de las variables.

Al momento de agregar estas variables aleatorias es necesario controlar el error, correlación y verificar la importancia de la variable. Para ello, muchas técnicas se han usado, siendo de las más populares la creación de nuevos conjuntos de entrenamiento con reemplazo; estos creados a partir del conjunto de entrenamiento original. Posteriormente, un árbol crece dentro del nuevo conjunto de entrenamiento usando una selección de variables aleatorias. Es importante recalcar que los árboles que han sido creados bajo esta metodología no son “recortados”.

Métodos más sofisticados se han acoplado a los Random Forest, como lo son las metodologías out-of-bag (Wolperd y Macready, 1996). Ambos propusieron usar out-of-bag (OBB). Basado en el uso de una agregación por medio de bootstrap obteniendo sub-muestras del conjunto de entrenamiento, siendo OBB el promedio del error de las predicciones en cada sub-muestra.

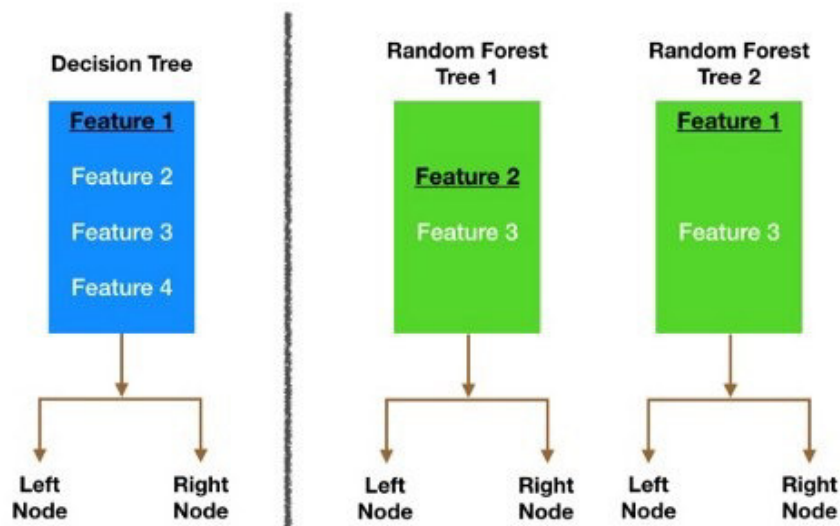
La evidencia empírica hallada por Breiman sugiere que OBB logra estimar de manera tan precisa que al usar un conjunto de entrenamiento de tamaño original. Por lo tanto, usar el estimador de error OBB remueve la necesidad de un conjunto de prueba de reserva (Breiman, 2001).

Al momento de usar la técnica bootstrap, en cada conjunto de entrenamiento creado por la misma, al menos el 30 % del conjunto es dejado de lado. Consecuentemente, la estimación que realiza la metodología OBB se basa en combinar las variables (tantas como sean posibles) con las combinaciones de variables usadas previamente. Ya que como la tasa de error desciende conforme al número de combinaciones aumenta, la estimación OBB tendra a sobrestimar la tasa actual de error. Para lograr estimaciones más precisas, es necesario llegar al punto crítico donde el error del conjunto de entrenamiento converge. La correlación puede también ser estimada usando métodos OBB (Breiman, 2001). Esto nos da un estimado de como el modelo está funcionando, cuál es su eficacia y como se puede mejorar.

En comparación a los árboles de decisión tradicionales, en los mismos se considera todas las posibles variables y se elige la que más “distancia” genera entre las observaciones restantes en el nodo izquierdo con las que se seleccionan para el nodo derecho. En contraste, cada árbol dentro del Random Forest puede escoger de entre un subconjunto de variables, esto genera aún más variación dentro de los árboles del modelo y por ende una menor correlación entre los árboles generando una mayor diversificación.

En la siguiente gráfica podemos ver una clara ejemplificación de como el árbol tradicional

puede seleccionar entre distintas variables para particionar el nodo y su comparación de cómo funcionan los Random Forest.



(a) Nota. Tomado de Node splitting in a random forest model is based on a random subset of features for each tree (Gráfico), Yiu, T, 2012, towardsdatascience (<https://towardsdatascience.com/understanding-random-forest-58381e0602d2>)

Figura 2: Particiones árboles de decisión vs Random Forest.)

De manera más explicativa podemos ver como el árbol puede elegir entre 4 variables para particionar el nodo, decide ir por la variable 1 y separa los nodos hijos lo más posible uno del otro. Ahora bien, en el caso Random Forest, por simplicidad solo vamos a considerar dos árboles de todos los  $M$  árboles posibles que puede tener un bosque. En este ejemplo, cuando analizamos el árbol 1, observamos que solo considera las variables 2 y 3 para la decisión de partición de los nodos. Del árbol de decisión tradicional sabemos que la variable 1 es la mejor para partir el árbol, así que es necesario usar la variable 2 como criterio de partición. Para el árbol 2, sabemos que como la variable 1 es la mejor para generar más diferencia entre el conjunto de datos. Por tanto, la elegimos como el criterio de partición.

Por lo mencionado en los párrafos anteriores, terminamos con árboles que además de ser entrenados con diferentes subconjuntos de datos, también estos árboles poseen diferentes variables para la toma de decisiones.

### 6.3.11. Pronóstico de las series de tiempo con Random Forest

El uso de los algoritmos Machine Learning en modelos predictivos y de manera más específica en el pronóstico de las series de tiempo ha incrementado de manera considerable en los últimos tiempos (Bontempi et al., 2012).

En particular el método Random Forest puede ser aplicado a un gran abanico de problemas de predicción ya que tiene una menor cantidad de parámetros a ser optimizados, es simple de usar y versátil con un gran abanico de aplicaciones exitosas en el nivel práctico (Tyralis y Papacharalampous, 2017). Además, puede lidiar con tanto pequeños y grandes tamaños de datos, gran cantidad de variables y capacidad lidiar correctamente con estructuras complejas que puede haber en los datos.

Para de series de tiempo que el horizonte de planificación  $h$  es mayor a 1, el pronóstico se realiza por medio de un procedimiento recursivo iniciando en el pronóstico que se realiza cuando  $h = 1$ . Los pronósticos obtenidos son usados en los posteriores pasos como un dato adicional, sin cambiar el modelo predictor (Tyralis y Papacharalampous, 2017).

La precisión del algoritmo Random Forest depende de cómo se optimicen los parámetros y sobre todo la selección de variables que se usen para realizar el modelamiento. La eficacia de los procedimientos como ya fue mencionado anteriormente depende netamente de las variables seleccionadas, el número de los árboles  $M$ , el número de posibles direcciones en los particionamientos  $mtry$  y el número de ejemplos en cada celda  $nodesize$ , en donde si llega por debajo de cierto número no será particionado.

En los problemas usuales de regresión, la muestra que consisten en observaciones de la variable dependiente y las correspondientes observaciones de las variables que tienen el rol de predictoras es dada. El modelo de regresión es entrenado usando esa muestra. A continuación, se realiza la predicción cuando se obtienen las nuevas observaciones de las variables predictoras. En el caso que decidamos usar una menor cantidad de variables que se encuentran originalmente en la muestra, entonces el tamaño del conjunto de entrenamiento no cambia. Sin embargo, es importante señalar que la inclusión de variables no importantes como predictores no afecta seriamente la exactitud con la cual actúan los Random Forest (Díaz-Uriarte y De Andres, 2006). Por otro lado, el uso de Random Forest para el pronóstico de las series de tiempo no es idéntico al de una simple regresión. En este caso, el rol que cumplen las variables predictoras es tomado por los valores previos de la serie de tiempo (datos históricos), lo mencionado tiene las siguientes consecuencias en los conjuntos de entrenamiento ya que al incrementar el número de variables predictoras inevitablemente reducirá la longitud del conjunto. Usar menos predictores puede reducir la calidad de la información obtenido por la dependencia temporal (Tyralis y Papacharalampous, 2017).

### 6.3.12. Medidas de error de pronóstico

El análisis de la efectividad de los modelos de pronóstico se basa en los resultados de las métricas, en la actualidad hay muchas maneras y técnicas para evaluar las mismas, las más populares son error porcentual absoluto medio (MAPE), error cuadrático medio (RMSE), error medio absoluto (MAE). Estas medidas son las más populares y son bien conocidas por representar la eficacia del modelo (Moon et al, 2018).

- **Error porcentual absoluto medio:** MAPE es una medida de predicción de la

eficacia para valores construidos con series de tiempo, específicamente en la estimación de tendencia. Usualmente presenta el resultado como un porcentaje del error. Es más fácil para comprender que otras medidas ya que el número que devuelve es un porcentaje (Moon et al, 2018). Es bien sabido que el resultado de MAPE es bastante grande si los valores reales de la serie se acercan a 0. La fórmula para MAPE se presentara en la siguiente ecuación donde  $y_t$  y  $\hat{y}_t$  son los valores reales y pronosticados respectivamente.

$$MAPE = \frac{100}{n} \sum_{t=1}^n \frac{|y_t - \hat{y}_t|}{|y_t|} \quad (16)$$

- **Error cuadrático medio:** RMSE es usado para agregar los residuos en una simple medida de habilidad predictiva. La raíz del error cuadrático medio, mostrado en la ecuación siguiente, representa la resta del valor real  $y_t$  menos el valor pronosticado  $\hat{y}_t$ . RMSE es la raíz cuadrada de la varianza, la que denota el error estándar.

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{n}} \quad (17)$$

- **Error medio absoluto:** El estadístico MAE es usado para evaluar que tan cerca está el pronóstico respecto a los valores reales de la serie. Es calculado al promediar la diferencia absoluta entre la predicción y lo real. MAE es definido en la siguiente ecuación donde  $y_t$  y  $\hat{y}_t$  son los valores reales y pronosticados respectivamente.

$$MAE = \frac{1}{n} \sum_{t=1}^n (|y_t - \hat{y}_t|) \quad (18)$$

Usaremos estas medidas para evaluar el modelo desarrollado por la metodología Random Forest.

## 7. Optimización del Inventario

### 7.1. Inventario Multi-Echelon

Una cadena Multi-Echelon o Multi Nivel tiene múltiples etapas y distintos protagonistas dentro del sistema de almacenaje. Los productos en una primera fase son entregados a una bodega central para posteriormente ser transportados a los centros de distribución regionales para finalmente abastecer a los puntos de venta o llegar al consumidor final. Un sistema Multi-Echelon es una red compuesta de niveles que son agrupados en echelons. Los niveles pueden representar locaciones físicas, ítems en inventario o actividades en proceso (Sbai y Berrado, 2018). La figura 3 representa un ejemplo de un sistema Multi-Echelon.



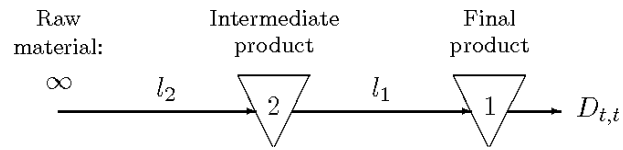
(a) Nota. Tomado de: A multi-echelon supply chain example (Gráfico), por Agarwal, A, 2019, semanticscholar (<https://www.semanticscholar.org/paper/Multi-echelon-Supply-Chain-Inventory-Planning-using-Agarwal/12ebceba1d30f9ac625b625fd606ee77e94e7e16/figure/0>)

Figura 3: Ejemplo de un Sistema Multi-Echelon.

#### 7.1.1. Estructura de un sistema Multi-Echelon

Un sistema de inventario Multi-Echelon involucra múltiples proveedores y clientes para cada nivel de almacenaje dentro del sistema. Existen diversos tipos de estructuras que son consideradas sistemas multiniveles de almacenaje debido a que siguen jerarquías marcadas dentro un plan de almacenamiento.

El sistema serial es la forma Multi-Echelon más sencilla, cada locación de inventario tiene a lo mucho un nodo predecesor y un nodo sucesor, en general trabajar con sistemas así es más simple que otras formas Multi-Echelon (Sbai y Berrado, 2018).

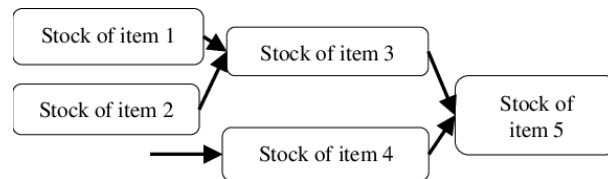


(a) Nota. Tomado de The serial, two-echelon production/inventory system (Gráfico), 2006, semanticscholar (<https://www.semanticscholar.org/paper/Multi-echelon-production2Finventory-systems-A-and-Houtum/a058600f89d2d7886126661e3dcb1dd60d7dc719/figure/0>)

Figura 4: Ejemplo de un Sistema Multi-Echelon Serial.

Podemos apreciar en la figura 3 que fácilmente los ítems pasan de un nivel a otro de manera directa, siempre pasando por el nodo de mayor jerarquía inmediata.

El segundo tipo de sistema multinivel se asemeja a una línea de ensamblaje, cada instalación tiene un inmediato sucesor, pero múltiples componentes pueden agruparse en un solo nivel lo que puede resultar en varios proveedores para un ítem (Sbai y Berrado, 2018).

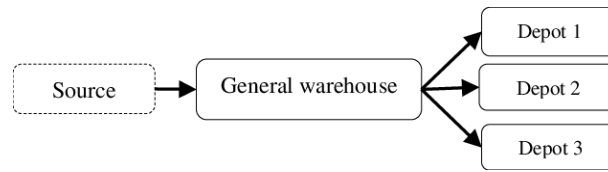


(a) Nota. Tomado de Example of assembly system (Gráfica), por Sbai, N, Barredo, A. 2018, matec-conferences ([https://www.matec-conferences.org/articles/mateconf/pdf/2018/59/mateconf\\_iwtsce2018\\_00013.pdf](https://www.matec-conferences.org/articles/mateconf/pdf/2018/59/mateconf_iwtsce2018_00013.pdf))

Figura 5: Ejemplo de un Sistema Multi-Echelon con estructura de ensamblaje.

En el gráfico de la figura 4 se puede apreciar que un proveedor puede entrar a abastecer en la cadena en cualquier nivel del sistema.

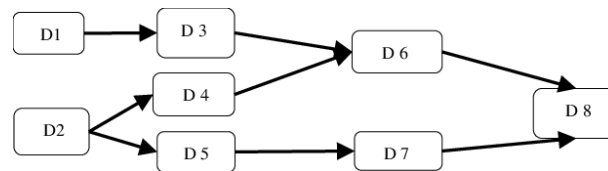
Una tercera estructura en el sistema de distribución involucra el incremento del número de instalaciones paralelas juntamente con el flujo de ítems, finalmente se añade la posibilidad de que haya varios puntos de venta al cliente o consumidores finales (Sbai y Berrado, 2018).



(a) Nota. Tomado de Distribution system (4) (Gráfica), por Sbai, N, Barredo, A. 2018, matec-conferences ([https://www.matec-conferences.org/articles/mateconf/pdf/2018/59/mateconf\\_iwtsce2018\\_00013.pdf](https://www.matec-conferences.org/articles/mateconf/pdf/2018/59/mateconf_iwtsce2018_00013.pdf))

Figura 6: Ejemplo de un Sistema Multi-Echelon con múltiples clientes.

Por último, un sistema que abarque las tres formas precedentes es la manera más clásica en que las empresas hoy en día conforman su sistema de distribución.



(a) Nota. Tomado de Illustration of a general multi-echelon inventory system of 8 depots (Gráfica), por Sbai, N, Barredo, A. 2018, matec-conferences ([https://www.matec-conferences.org/articles/mateconf/pdf/2018/59/mateconf\\_iwtsce2018\\_00013.pdf](https://www.matec-conferences.org/articles/mateconf/pdf/2018/59/mateconf_iwtsce2018_00013.pdf))

Figura 7: Ejemplo de un Sistema Multi-Echelon Combinado.

En el nivel final es posible que haya varios puntos de venta finales.

Un sistema de inventario Multi-Echelon puede también ser representando por un grafo directo donde los nodos representan varias actividades y locaciones dentro del sistema y las aristas representan el flujo de productos (Clark, 1972).

### 7.1.2. Características de los sistemas Multi-Echelon

Los sistemas Multi-Echelon dependen fuertemente en las capas y distintos niveles de proveedores distribuidos a través de múltiples centros de distribución. En tales sistemas, el nuevo inventario es primero almacenado en una facilidad central o regional de distribución (Sbai y Berrado, 2018).

En base a los antecedentes mencionados, la interacción entre las distintas etapas en el proceso de almacenamiento permite que los sistemas Multi-Echelon repartan correctamente el inventario a través de toda la cadena de suministro mientras toman en cuenta la compleja interdependencia que existe entre los distintas niveles e inclusive las variables que causan exceso de inventario (Sbai y Berrado, 2018).

Diferentes características y variables son consideradas para el análisis y optimización.

- **Demanda:** El elemento más importante dentro de la cadena de suministro y el factor más importante a considerar dentro de los sistemas Multi-Echelon, la demanda es estocástica, generalmente se estima con algún modelo estadístico.
- **Inventario/Políticas de Inventario:** Hay dos políticas comúnmente usadas en la práctica, modelo de Revisión Continuo y modelo de Revisión Periódico. En los modelos de Revisión Continuo, la posición de inventario dentro del sistema es revisado constantemente durante un periodo de tiempo y una cantidad  $Q$  es ordenada cuando los niveles de inventario llegan a una cantidad determinada. En los modelos de Revisión Periódica el estatus del inventario es revisado en intervalos de tiempo y los niveles de inventario son elevados a un punto predefinido.
- **Inventario centralizado o descentralizado:** Son decisiones de reaprovisionamiento de inventario, en el caso de centralización el objetivo es encontrar el costo óptimo para todo el sistema, para los inventarios descentralizados; cada locación ya sean estos centros de distribución, bodegas locales o regionales su objetivo es optimizar sus propios costos independientes sin considerar el resto del sistema.
- **Clasificación de productos:** Este criterio ayuda a la clasificación de modelos dentro del sistema que mejor se adapten a la necesidad y patrones de consumo de los consumidores.
- **Número de niveles/echelons:** Forma en la que el sistema está estructurado.

## 7.2. Optimización del Inventario usando metodología Multi-Echelon

La implementación de métodos Multi-Echelon permite el uso de múltiples locaciones de almacenamiento y elementos dentro de la cadena de abastecimiento con el objetivo de no solo minimizar los costos logísticos, si no reducir el costo de mantener inventario en bodegas mediante un manejo eficiente del riesgo operativo.

La optimización tradicional de un sistema puede ser conseguida mediante un acercamiento simple de un solo nivel, es decir, optimizar cada uno de los niveles y bodegas donde los productos son almacenados o distribuidos sin considerar los efectos de interacción que hay en el sistema. Sin embargo, esta forma de optimización falla en capturar una verdadera aproximación sistemática; al despreciar una vista global del sistema y aplicar estrategias de reaprovisionamiento de un solo nivel sin considerar el impacto de otros niveles y locaciones, puede llevar a tener niveles de stock de seguridad erróneos o llegar a fallas en el servicio al cliente sin importar que se tenga el inventario adecuado (Lee, 2003).

Las configuraciones Multi-Echelon pueden con facilidad guiar y ayudar a que las empresas mejoren sus interacciones entre los distintos niveles de almacenaje, es decir; optimizar la relación entre los nodos superiores de abastecimiento y los nodos inferiores que son generalmente de venta o atención al cliente. Ciertas relaciones se pueden derivar respecto a la importancia de mantener niveles adecuados de inventario tanto en todos los nodos, la falta de inventario en los nodos de abastecimiento puede incrementar el tiempo de espera para las locaciones inferiores (Agarwal,2019). De igual manera, mantener inventario en



los nodos de abastecimiento permite que los nodos inferiores tengan un menor stock de seguridad siempre y cuando el tiempo de espera no sea muy largo entre nodos.

Consecuentemente, muchas dificultades se pueden presentar cuando los nodos de abastecimiento también tienen la función de servir a los clientes al igual que a los múltiples nodos inferiores que le suceden y surja la necesidad de priorizar entre las múltiples órdenes que existan. En el caso de la existencia de órdenes de compra grandes, pero esporádicas por parte de los nodos inferiores, estas generan un gran desafío para los nodos de abastecimiento para seguir un manejo constante de las políticas de inventario, generando en promedio un incremento del inventario a través del tiempo (Agarwal,2019). Por otro lado, en términos de niveles de servicio al cliente se pueden generar muchas alternativas en el manejo, los nodos de abastecimiento pueden escoger bajar sus niveles de servicio siempre y cuando las demandas de los clientes se cumplan en los nodos inferiores.

Basado en los antecedentes presentados, las aproximaciones sistemáticas para la optimización de toda la cadena de suministro son un desafío que requiere un acercamiento holístico (enfatar la importancia del todo considerando su globalidad) que logre explotar y maximizar la sinergia que existe en las interacciones entre los distintos niveles.

Muchos artículos enfrentan el problema de optimización Multi-Echelon basado en los primeros modelos presentados por Sherbrooke (Sheerbrooke, 1964)., Ganeshan (Ganeshan, 1999)., y el modelo presentado por Hopp (Hopp et al.,1999 ). Los investigadores mencionados desarrollaron modelos de dos niveles. Sin embargo, ellos asumieron políticas de reaprovisionamiento y principalmente que la demanda era un proceso de Poisson para cada nivel, de tal manera, cuantificaron los pedidos pendientes esperados usando series infinitas (Agarwal,2019). Otros avances se produjeron posteriormente cuando Glasserman y Tayur desarrollaron un modelo multi periodo (Glasserman y Tayur, 1994). El modelo presentado resultó ser un gran avance, pero asumía una política de inventario base con periodos de revisión de 1 día y generaba órdenes de reaprovisionamiento constantes que ocasionaban que los niveles de inventario se mantengan altos durante grandes periodos de tiempo.

Posteriormente, los modelos propuestos se enfocaron en las políticas de inventario medio usando parámetros como: niveles de servicio y cumplimiento que se deben verificar en cada etapa y nivel (Graves et al, 2000). Todas estas suposiciones lograron que los modelos se simplifiquen pero que erróneamente representen la verdadera variabilidad. Además, muchos de los modelos que anteriormente se estudiaron hacen suposiciones respecto a las interacciones de los modelos, políticas de almacenamiento, o bien el flujo del producto a través de los nodos para poder resolver tales problemas con modelos y trabajo matemático (Agarwal,2019).

Para corregir las hipótesis que llevan a errores y lograr que se capture de manera adecuada la verdadera variabilidad y complejidad de los sistemas Multi-Echelon realizaremos las siguientes suposiciones.

- En lugar de asumir una función de distribución para modelar la demanda en cada

uno de los centros de distribución y sobre todo en los puntos de venta vamos a modelar y pronosticar la demanda por medio de la técnica Machine Learning Random Forest. Con esto, podemos cuantificar la demanda y tiempo de espera realizando una muestra bootstrap.

- El modelado del sistema se hará mediante una simulación de eventos discretos. Esto nos permitirá capturar de una mejor manera todas las complejas interacciones del sistema.

Recordemos que el propósito de realizar esta metodología es minimizar el costo total del sistema esto lo logramos minimizando el promedio de inventario disponible. Incluyendo variables del negocio como: costo de realizar la orden, costo de mantener inventario en cada una de las locaciones, además otras variables incluyendo costos que incurren una locación. En otras palabras, el objetivo primordial es minimizar el inventario promedio disponible para todas las locaciones mientras aseguramos niveles de servicio alto y satisfactorios para los clientes.

### 7.2.1. Suposiciones para la Optimización de Inventario

Notemos en primer lugar que las distintas ubicaciones no comparten sus inventarios con las otras bodegas del mismo nivel y no se comparte información sobre el nivel de inventario que posee cada una exceptuando al centro que provee. Cada uno de los centros de distribución, bodegas regionales o locales operan independientemente, cada uno de ellos son optimizados simultáneamente.

Las suposiciones que usaremos para modelar el problema son:

- Cada bodega o depósito de inventario es abastecida por solo un nodo superior de abastecimiento.
- Cada unidad de almacenamiento sigue una combinación entre las políticas niveles de reaprovisionamiento y la cantidad mínima de inventario. En el caso de que los niveles de inventario estén por debajo del punto de reaprovisionamiento, se realizara una orden para abastecer la localidad.
- Las órdenes de compra siempre son completas, es decir, que no son parcialmente enviadas. En el caso de que una unidad de almacenamiento no tenga suficiente inventario disponible, esperara hasta que haya suficiente inventario para satisfacer esa orden de compra.
- Cuando la bodega o centro de distribución tiene la doble función de abastecer al resto de la red y servir al cliente, se prioriza el servicio al cliente respecto a abastecer un nodo menor.
- Si hay muchas órdenes de compra en espera por parte de los nodos inferiores, las órdenes serán atendidas de acuerdo con el orden que fueron recibidas.

### 7.2.2. Beneficios del algoritmo de Optimización de Inventario

En este trabajo el método para optimizar el inventario se basa en un proceso de optimización y simulación. Notemos que en general esta forma de trabajo no garantiza la optimalidad absoluta ya que como muchos algoritmos de optimización estos no incluyen un criterio fuerte de convergencia (Rios y Sahinidis, 2013).

Como resultado de lo anterior, para asegurarnos de obtener resultados positivos y en cierto grado óptimos realizaremos un procedimiento iterativamente, ocasionando que el algoritmo sea computacionalmente pesado. Sin embargo, los beneficios de escoger un procedimiento basado en la optimización con simulación son amplios, ya que llevados a la práctica este método es superior al momento de capturar la dinámica de interacción del sistema y además al incluir un módulo de simulación permite flexibilizar el modelo con distintas reglas y políticas del inventario para representar mejor a los mismos (Agarwal, 2019).

### 7.2.3. Módulo de Simulación

El modelo de simulación de eventos discretos consiste en 4 fases basados primordialmente en la forma en la que los pedidos son realizados y como se abastecen los nodos inferiores de los nodos superiores.

Las 4 fases son:

1. **Órdenes de reaprovisionamiento:** Este proceso realiza la orden de reaprovisionamiento de los nodos inferiores que tienen que ser abastecidos por los nodos superiores una vez que los niveles de inventario bajen el límite del punto de reaprovisionamiento.
2. **Cumplimiento de las órdenes de reaprovisionamiento:** Este proceso es usado por los centros de abastecimiento para preparar y enviar las órdenes que fueron realizadas en el proceso anterior a los nodos inferiores. Una vez que la orden es preparada en su totalidad, se procede a enviar.
3. **Enviar la orden de reaprovisionamiento:** Este proceso es una continuación directa del proceso anterior. Cada envío es contemplado en esta fase y se basa principalmente en el tiempo de espera que tienen los nodos inferiores para que sus órdenes lleguen a sus respectivas bodegas. Además; este procedimiento se encarga de actualizar el inventario disponible de cada locación una vez que llega el nuevo inventario.
4. **Atender la demanda del cliente:** Este proceso se utiliza para atender y satisfacer las demandas de los clientes de cada una de las distintas locaciones donde se realiza venta al público.

Los procedimientos serán a continuación detallados con más exactitud con su correspondiente algoritmo.

## Órdenes de reaprovisionamiento

El procedimiento representa la lógica de cada establecimiento  $f=1,2,\dots,n$  para generar órdenes reposición de inventario. El trabajo del algoritmo es realizar el seguimiento del nivel de inventario  $I_f$  de cada locación. Una vez que el nivel de inventario se encuentra por debajo del punto de reabastecimiento  $R_f \in \mathbb{R}_0^+$ , una orden  $q \in \mathbb{R}_0^+$  se prepara para poder llevar a niveles normales de inventario ( $B_f \in \mathbb{R}_0^+$ ) el inventario disponible  $O_f \in \mathbb{R}_0^+$ , es decir, *Volumen de la Orden = Nivel base del Inventario - Inventario Disponible*. Posteriormente, la orden  $q$  de la locación  $f$  es insertada a la cola de espera de pedidos de la instalación de reaprovisamiento  $Q_{S_f}$ . El nivel de inventario también se actualiza en esta fase.

El algoritmo para esta fase es el siguiente, los inputs necesarios son: Locación  $f$ , Nivel base de inventario  $B_f$ , Locación de Reabastecimiento  $S_f$  y Horizonte de Simulación  $H \in \mathbb{N}$ . Adicionalmente, el algoritmo hace uso de la función *ColocarEnCola(.)* que hace referencia a agrupar las ordenes de reposición en la locación de reaprovisionamiento.

---

**Algorithm 2** ÓrdenesDeReaprovisionamiento( $f, R_f, B_f, S_f, H$ )

---

**Result:** Nivel del Inventario  $I_f$  y una cola de pedidos de reaprovisionamiento  $Q_{S_f}$  actualizada.

```
while  $t \leq H$  do
  if  $I_{f,t} \leq R_f$  then
     $q \leftarrow B_f - O_{f,t}$ 
    ColocarEnCola( $Q_{S_f}, (q, f)$ )
     $I_{f,t+1} \leftarrow I_{f,t+q}$ 
  end
   $t \leftarrow t + 1$ 
end
```

---

## Cumplimiento de las órdenes de reaprovisionamiento

El procedimiento es una continuación directa del algoritmo 2 presentado anteriormente, el algoritmo representa el proceso que deben realizar los nodos de abastecimiento para enviar nuevos inventarios a los nodos inferiores. Como se ha mencionado antes, existe una cola de pedidos  $Q_{S_f}$  que mantiene el registro de todas las órdenes que fueron puestas por sus respectivos nodos inferiores. La cola sigue un mecanismo FIFO (primero en llegar primero en salir). Es decir, que las órdenes son completadas de acuerdo con el orden que fueron recibidas.

Si la cola de pedidos  $Q_{S_f}$  de la locación  $f$  es no vacía, Se procede a atender la orden, la cual nos brindara información sobre la cantidad del pedido  $q$  y la locación  $f$  al cual se dirigirá. Es importante notar que la orden debe ser completada en su totalidad para que se empiece a considerar la distribución de la siguiente orden. Dependiendo del inventario disponible  $O_f \in \mathbb{R}_0^+$  en la locación de reabastecimiento, si hay un volumen  $r > 0$  de

inventario que debe ser adquirido para completar el volumen  $q$  de la respectiva orden.

Si  $r > 0$  significa que el establecimiento no tiene suficiente inventario disponible para completar la orden y tiene que esperar hasta que haya suficiente del mismo para completar el pedido. Una vez que se realizan las asignaciones que deben ser despachadas, los niveles de inventario y el inventario disponible son actualizados. El siguiente proceso que se debe realizar es la entrega del pedido a las distintas bodegas.

Para el algoritmo de este proceso es necesario información sobre la locación  $f = 1, 2, \dots, n$  y el horizonte de planificación  $H \in \mathbb{N}$ . Llamaremos  $Q_{f^*}$  a la orden generada por la locación  $f^*$ . Adicionalmente, el algoritmo hace uso de las funciones  $ColaNoVacía(.)$ ,  $QuitarDeCola(.)$ ,  $Min(.)$  y  $EnviarOrden(.)$ . La primera hace relación a verificar si la orden  $Q_{f^*}$  es no vacía. Es decir, que existe un pedido de reposición de la locación de reabasteciendo a la locación  $f^*$ . La segunda función, hace referencia a que si el pedido es atendido entonces el pedido  $Q_{f^*}$  es retirado de la cola de pedidos  $Q_{s_f}$ . La función  $Min(.)$  se encarga de evaluar si el pedido  $q$  es menor que el inventario disponible  $O_f$  para así asegurarnos de abastecer solo lo que tenemos posible. Por último, la función  $EnviarOrden(.)$  consiste en enviar la cantidad  $q$  a la locación solicitante  $f^*$ .

---

**Algorithm 3** CumplimientoDeLasÓrdenesDeReaprovisionamiento( $f, f^*, H$ )

---

**Result:** Nivel del Inventario  $I_f$ , Inventario Disponible  $O_f$  y una actualización de la cola  $Q_{s_f}$ . Preparación de la orden  $q$  para la locación  $f^*$ .

```

while  $t \leq H$  do
  if  $ColaNoVacía(Q_{s_f})$  then
     $(q, f^*) \leftarrow QuitarDeCola(Q_{f^*})$ 
     $s \leftarrow Min(q, O_{f,t})$ 
     $O_{f,t} \leftarrow O_{f,t} - s$ 
     $I_{f,t} \leftarrow I_{f,t} - s$ 
     $r \leftarrow q - s$ 
    if  $r > 0$  then
      while  $O_{f,t} < r$  do
         $t \leftarrow t + 1$ 
         $O_{f,t} \leftarrow O_{f,t} - r$ 
         $I_{f,t} \leftarrow I_{f,t} - r$ 
      end
       $EnviarOrden(f^*, q)$ 
    end
  end
  else
     $t \leftarrow t + 1$ 
  end
end

```

---

Enviar la orden de reaprovisionamiento

El procedimiento de envío de nuevo inventario es en esencia similar a una cinta de transportación. Cada proceso de aprovisionamiento sigue los principios de pasar por diferentes etapas hasta llegar a su destino final. Esta fase se caracteriza porque tiene que ser llamada directamente por la fase anterior y su principal función y objetivo es esperar que transcurra el tiempo de espera que hay en el envío de los nuevos productos del nodo de abastecimiento hasta el nodo que solicitó la orden. Al momento de que el envío se completa se actualiza el inventario disponible en la locación de destino.

Definiremos como  $L_f \in \mathbb{R}_0^+$  como el tiempo que transcurre entre que la orden  $Q_f$  es despachada y la misma llega a la locación de destino.

---

**Algorithm 4** EnviarOrden( $f^*, q$ )

---

**Result:** Actualización del Inventario Disponible  $O_{f^*}$

$l \leftarrow L_{f^*}$

$t \leftarrow t + l$

$O_{f^*,t} \leftarrow O_{f^*} + q$

---

### Atender la demanda del cliente

El procedimiento de cumplir con la demanda del cliente en esencia lo que realiza es seguir la demanda total observada  $M_f \in \mathbb{R}_0^+$  para así, estimar el nivel de servicio  $\beta$ . Como en muchos negocios de venta de repuestos de vehículos de marcas generalistas, en el caso de que los materiales necesarios para realizar el trabajo o mantenimiento no estén disponibles, los clientes optarán por buscar los repuestos o servicios requeridos en otro establecimiento, en ese caso se considera como una venta pérdida que es muy común en cualquier tipo de negocio y sobre todo en el negocio automovilístico.

En el caso anteriormente mencionado, la venta es pérdida cuando el ítem solicitado no se encuentra disponible. Aquí, si la demanda  $d$  del cliente puede ser satisfecha con el inventario disponible  $O_f \in \mathbb{R}_0^+$ , en ese caso se le entrega el ítem al cliente, caso contrario, no hay inventario disponible para entregar. El procedimiento también guarda los pedidos entregados  $P_f \in \mathbb{R}_0^+$ . Conjuntamente  $P_f$  con  $M_f$  son usados para estimar el nivel de servicio  $\beta$ . El nivel de servicio se calcula usando la cantidad de productos entregados  $P_f$  sobre la demanda que hay por los mismos  $M_f$  en un horizonte de tiempo  $H$ .

$$\beta = P_{f,H}/M_{f,H}, \quad H \in \mathbb{N}, \quad \forall f = 1, 2, \dots, n$$

Dependiendo de la cantidad de volumen que fue enviado se actualiza el inventario disponible y el nivel de inventario  $I_f$ . En este procedimiento se trata de capturar la variabilidad de la demanda por medio de bootstrap. Así, nos aseguramos de que en la simulación la variabilidad sea correctamente representada.

El correspondiente algoritmo para esta fase requiere información sobre la locación  $f$  donde se realiza la venta y un horizonte de planificación  $H$ .

---

**Algorithm 5** *AtenderLaDemandaDelCliente( $f, H$ )*

---

**Result:** Posición de Inventario  $I_f$ , Inventario Disponible  $O_f$ , Demanda Total  $M_f$  actualizada, Total Ítem Entregados  $P_f$  actualizado.

**while**  $t \leq H$  **do**

```
     $d \leftarrow \text{Bootstrap}(D_f)$   
     $M_{f,t+1} \leftarrow M_{f,t} + d$   
     $s \leftarrow \text{Min}(d, O_{f,t})$   
     $P_{f,t+1} \leftarrow P_{f,t} + s$   
     $O_{f,t+1} \leftarrow O_{f,t} - s$   
     $I_{f,t+1} \leftarrow I_{f,t} - s$   
     $t \leftarrow t + 1$ 
```

**end**

---

#### 7.2.4. Inicialización de la Simulación

El siguiente objetivo se basa en la simulación del sistema por medio de una función que conecte todos los procedimientos anteriormente descritos. Al momento de iniciar con el proceso, este calcula la media del inventario disponible y el nivel de servicio  $\beta$  de cada uno de los establecimientos.

Al momento de llamar al módulo de optimización que será descrito posteriormente, el mismo utilizará el módulo de simulación para evaluar las funciones que contiene, al momento de ser llamada, el procedimiento inicia usando los algoritmos descritos en la sección anterior para cada una de las locaciones en el sistema, Esto correrá en paralelo para todo el horizonte de planificación  $H \in \mathbb{N}$ .

Para poder iniciar el módulo de simulación, este requiere una lista de las distintas locaciones  $F$ , puntos iniciales de reaprovisionamiento  $R_f$ , puntos iniciales de los niveles bases de inventario  $B_f$ , el sistema organizado en una matriz que indique que nodos cumplirán con las funciones de abastecimiento y puntos de venta  $S_f$  para todas las locaciones  $f \in F$ , horizonte de simulación  $H$ . Cabe recalcar que algunos de estos inputs se modificarán o serán proporcionadas por el módulo de optimización.

Una vez que se ha completado la simulación, el procedimiento calcula el inventario medio disponible  $A_f \in \mathbb{R}_0^+$  a lo largo de sistema, al tomar una media de la muestra del horizonte  $H$ . Se procede entonces a calcular el nivel de servicio  $\beta$ .

La importancia del cálculo del nivel de servicio cae en la necesidad de emplearlo como criterio de parada en el módulo de optimización. El algoritmo para iniciar el proceso de simulación descrito en los párrafos anteriores es el siguiente.

---

**Algorithm 6** SimilarSistema( $F, R, B, S, H$ )

---

**Result:** Lista de Locaciones  $F$ , Puntos Iniciales Reaprovisionamiento  $R_f$ , Niveles Bases de inventario  $B_f$ , Lista de Locaciones de Abastecimiento  $S_f$ , Horizonte de Simulación  $H$

```
foreach  $f \in F$  do
  ÓrdenesDeReaprovisionamiento( $f, R_f, B_f, S_f, H$ )
  CumplimientoDeLasÓrdenesDeReaprovisionamiento( $f, f^*, H$ )
  AtenderLaDemandaDelCliente( $f, H$ )
end
foreach  $f \in F$  do
   $A_f \leftarrow 0$ 
  while  $t \leq H$  do
     $A_f \leftarrow A_f + O_{f,t}$ 
     $A_f \leftarrow A_f / H$ 
     $B_f \leftarrow P_{f,H} / M_{f,H}$ 
  end
end
```

---

### 7.2.5. Módulo de Optimización

El módulo de optimización centra su énfasis en el uso de una función objetivo, con dos componentes, que se busca minimizar. El primer componente es el inventario promedio disponible dentro del sistema. Es decir, de cada una de las locaciones, mientras que el segundo componente enfocado en el cumplimiento de los niveles de servicio objetivo; este se logra generando una penalización por cada vez que no se logre los niveles deseados a través del sistema. Ambos componentes son calculados en múltiples simulaciones y promediados en todas las repeticiones. Logrando así crear una función objetivo compuesta.

En cada una de las réplicas que se realizan dentro de esta fase, el módulo de optimización hace uso de una simulación para obtener el cálculo de la media de inventario disponible en cada locación del sistema  $A_f \in \mathbb{R}_0^+$  y el nivel de servicio  $\beta_f$ .

Definamos  $AA \in \mathbb{R}_0^+$  como el promedio de todos los promedios  $A_f$  para todos las locaciones  $f = 1, 2, \dots, n$ . Por tanto, el promedio del inventario disponible de cada locación es agregado y contabilizado en  $AA$ , Definimos adicionalmente a  $\beta_f^T$  como el nivel de servicio objetivo deseado, el siguiente paso es comparar  $\beta_f$  con el nivel de servicio deseado  $\beta_f^T$ . Todo nivel de servicio que no ha sido satisfecho, es decir,  $\beta_f^T - \beta_f$  es almacenado en  $A\beta \in \mathbb{R}_0^+$ . Este proceso es repetido  $N$  cantidad de veces. Finalmente,  $AA$  y  $A\beta$  son promediadas en todas las repeticiones  $N$ , Finalmente,  $A\beta$  es multiplicado por un factor de penalización  $pe^* \in \mathbb{R}_0^+$  para obtener una función objetivo  $Z$ .

Los inputs necesarios para el funcionamiento del módulo, al igual que en el módulo de simulación son: una lista de locaciones  $F$ , puntos de reaprovisionamiento  $R_f$ , niveles medio de inventario  $B_f$ , locaciones que sirven de abastecimiento  $S_f$  y un horizonte de simulación



$H$ .

Por tanto, obtenemos un algoritmo con todo lo descrito anteriormente.

---

**Algorithm 7** OptimizarSistema( $F, R, B, S, H$ )

---

**Result:** Lista de Locaciones  $F$ , Puntos Iniciales Reaprovisionamiento  $R_f$ , Puntos Iniciales Niveles Medio de Inventario  $B_f$ , Lista de Locaciones de Abastecimiento  $S_f$ , Horizonte de Simulación  $H$

$AA \leftarrow 0$

$A\beta \leftarrow 0$

**foreach**  $n \in N$  **do**

$(A_f, \beta_f) \leftarrow \text{SimilarSistema}(F, R, B, S, H)$

**foreach**  $f \in F$  **do**

$AA \leftarrow AA + A_f$

$A\beta \leftarrow A\beta + \text{Max}(0, \beta_f^T - \beta_f)$

**end**

**end**

$Z \leftarrow AA/N + pe * A\beta/N$

---

El resultado de este algoritmo es una función objetivo  $Z$  para el algoritmo de optimización de tipo caja negra basada en el promedio del promedio del inventario disponible  $AA$  y el nivel de servicio  $\beta$ .

Para realizar la optimización usaremos un acercamiento usando optimización Bayesiana. En resumen, la función objetivo es considerada una función aleatoria en donde su distribución previa es definida usando un proceso Gaussiano. La evaluación de las funciones es tratada como data y usadas para actualizar la distribución previa a una post distribución para la función objetivo. Posteriormente, es usada para maximizar una función de utilidad (Agarwal, 2019).

### 7.2.6. Procesos Gaussianos

La clase de procesos Gaussianos es una de las familias más usadas de procesos estocásticos para modelar data dependiente del tiempo, espacio e inclusive ambos (Davis, 2014). La popularidad de este tipo de procesos radica principalmente en dos propiedades. La primera, en la composición del proceso, este se determina por su media y su respectiva función de covarianza. La segunda, la facilidad que posee al resolver el problema de predicción. Siendo una función lineal el mejor predictor para un proceso Gaussiano, en la mayoría de los casos, estas funciones pueden ser calculadas de manera rápida usando formulas recursivas (Davis, 2014).

La caracterización fundamental de los procesos Gaussianos es que logran generalizar la distribución multivariada de Gauss sobre dimensiones finitas de vectores a dimensión infinita. Específicamente, un proceso Gaussiano es un proceso estocástico que tiene distri-

bución marginal de dimensión finita con distribución de Gauss. Al hacer eso, se define la distribución sobre funciones, es decir, cada punto de un proceso Gaussiano es una función (Zhang et al., 2011).

La principal característica de los procesos Gaussianos es que, en lugar de trabajar sobre espacios paramétricos, estos toman lugar en un espacio de funciones sin parametrizar la función, es decir, no paramétricas (Zhang et al., 2011).

Formalmente podemos definir a los procesos Gaussianos como:

*Definición .5.* Un proceso Gaussiano es un proceso estocástico, es decir, una colección de variables aleatorias en donde las distribuciones de dimensión finita son Gauss multivariada para cualquier elección de variables finitas.

En general, un proceso Gaussiano es usado para definir una distribución de probabilidad sobre funciones  $f : X \rightarrow \mathbb{R}$  tales que el conjunto de valores de  $f$  evaluados en un conjunto arbitrario de puntos de  $\{x_i\}_i^N \in X$  tendrá una distribución N-variada (Zhang et al., 2011).

### 7.2.7. Propiedades básicas de los procesos Gaussianos

Sea un proceso estocástico de valores en  $\mathbb{R} \{x_t, t \in T\}$ , donde  $T$  es un conjunto de índices, el proceso estocástico es además un proceso Gaussiano si todas las distribuciones finitas tienen una distribución normal multivariada. Eso es, para cualquier colección de valores  $t_1, \dots, t_k \in T$ , el vector aleatorio  $X = (x_{t_1}, \dots, x_{t_k})$  tienen una distribución normal multivariada con un vector de medias  $\mu = E[X]$  y matriz de covarianza  $\sigma = cov(X, X)$ , lo denotaremos como:

$$X \sim N(\mu, \Sigma)$$

Si la matriz  $\Sigma$  es no singular, el vector aleatorio  $X$  tiene una función de probabilidad dado por:

$$f_x(x) = (2\pi)^{-\frac{n}{2}} (det(\Sigma))^{-\frac{1}{2}} exp(-\frac{1}{2}(x - \mu)' \Sigma^{-1}(x - \mu)) \quad (19)$$

En el caso de que nos compete (series de tiempo),  $t$  toma valores  $\{0, 1, \dots\}$ .

La media y función de covarianza de un proceso Gaussiano son definidos de la siguiente manera:

$$\mu(t) = E[X_t]$$

y

$$\gamma(s, t) = cov(X_s, X_t)$$

respectivamente. A pesar de que los procesos Gaussianos dependen de solo dos parámetros, el modelamiento puede ser difícil sin que se añaden simplificaciones a la forma de la media y la función de covarianza (Davis, 2014). La suposición de estacionalidad frecuentemente provee un nivel significativo de simplificación para los procesos Gaussianos sin sacrificar el potencial de generalización. Además, al momento de aplicar transformaciones simples a la data, la suposición de estacionalidad de la data transformada se consigue fácilmente.

Una serie de tiempo Gaussiana  $\{x_i\}$  se dice estacionaria si:

1.  $\mu(t) = E[X_t] = \mu$  es independiente de  $t$ , y
2.  $\gamma(t+h, t) = cov(X_{t+h}, X_t)$  es independiente de  $t$  para todo  $h$ .
3.  $X_t \sim N(\mu, \gamma(0))$  para todo  $t$ , y
4.  $(X_{t+h}, X_t)'$  tiene una distribución normal bivariada con matriz de covarianza:

$$\begin{bmatrix} \gamma(0) & \gamma(h) \\ \gamma(h) & \gamma(0) \end{bmatrix} \quad (20)$$

para todo  $t$  y  $h$ .

Un proceso estocástico  $\{x_t\}$  que satisface las condiciones 1 y 2 es débilmente estacionario o estacionario de segunda orden. Los procesos estacionarios de primer y segundo orden son invariantes respecto a las translaciones del tiempo. Un proceso estocástico  $\{x_t\}$  se dice estrictamente estacionario si la distribución de  $(x_{t_1}, \dots, x_{t_n})$  es la misma para  $(x_{t_1+s}, \dots, x_{t_n+s})$  para cualquier  $s$ . Es decir, que las propiedades de la distribución de las series de tiempo son invariantes ante cualquier tipo de translación del tiempo.

Para las series de tiempo Gaussianas, los conceptos de estacionalidad estricta o débil se unen (Davis, 2014). El resultado precedente se sustenta al hecho que los procesos estacionarios estrictos  $(x_{t_1}, \dots, x_{t_n})$  y  $(x_{t_1+s}, \dots, x_{t_n+s})$  tienen el mismo vector de medias y matriz de covarianza. Ya que ambos vectores tienen distribución normal multivariada, estas deben ser idénticamente distribuidas.

### 7.2.8. Propiedades de la función de autocovarianza

Una función de autocovarianza  $\gamma(\cdot)$  tiene las siguientes propiedades:

1.  $\gamma(0) \geq 0$ ,
2.  $|\gamma(h)| \leq \gamma(0)$  para todo  $h$ ,
3.  $\gamma(h) = \gamma(-h)$ , es decir,  $\gamma(\cdot)$  es una función par.

La autocovarianza tiene otra propiedad fundamental; la no negatividad de la función.

$$\sum_{i,j=1}^n a_i \gamma(t_i - t_j) a_j \geq 0 \quad (21)$$

notemos que si  $a_1, \dots, a_n$  son números reales y por las propiedades antes mencionadas es muy fácil apreciar que se verifica la condición de no negatividad.

### 7.2.9. Procesos lineales Gaussianos

Sea  $\{x_t, t = 0, 1, 2, \dots\}$  un proceso estacionario Gaussiano con media 0, entonces la descomposición de Wold nos permite expresar  $X_t$  como la suma de dos componentes independientes (Davis, 2014). Por tanto, expresamos de la siguiente manera:

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} + V_t \quad (22)$$

donde  $\{Z_t\}$  es una sucesión independiente normal distribuida idénticamente con media 0 y varianza  $\sigma^2$ ,  $\{\psi_t\}$  es una secuencia de coeficientes tales que  $\psi_0 = 1$ , y  $\{V_t\}$  es un proceso determinista que es independiente de  $\{Z_t\}$ . Los mismos que son referidos como innovaciones y son definidos de la siguiente manera  $Z_t = X_t - E(x_t | x_{t-1} x_{t-2}, \dots)$ .

Es necesario definir el significado que un proceso sea determinístico, se dice que el proceso  $\{V_t\}$  es determinístico si  $V_t$  está completamente determinado por sus valores pasados  $\{V_s, s < t\}$  (Davis, 2014).

Los procesos Gaussianos puramente no deterministas no poseen un componente determinista y pueden ser representado como un proceso lineal Gaussiano de la siguiente manera:

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} \quad (23)$$

La función de autocovarianza correspondiente a  $\{X_t\}$  tiene la forma

$$\gamma(h) = \sum_{j=0}^{\infty} \psi_j \psi_{j+h} \quad (24)$$

### 7.2.10. Predicción

Sean dos vectores aleatorios  $\{X_1\}$  y  $\{X_2\}$  que tienen una distribución normal conjunta, es decir, su distribución viene dado de la siguiente forma:

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N \left( \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right) \quad (25)$$

donde  $\Sigma_{22}$  es no singular, entonces la distribución condicional de  $X_1$  dado  $X_2$  tiene una distribución normal multivariada con media:

$$\mu_{X_1|X_2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(X_2 - \mu_2) \quad (26)$$

y la matriz de covarianza:

$$\Sigma_{X_1|X_2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

Es importante notar que el mejor predictor en términos del error cuadrático medio de  $X_1$  en términos de  $X_2$  (es decir, la función multivariada  $g(X_2)$  que minimiza  $E\|X_1 - g(X_2)\|^2$ , donde  $\|\cdot\|$  es la distancia euclidiana) es  $E(X_1|X_2) = \mu_{X_1|X_2}$  (Davis, 2014). Señalemos que el predictor obtenido es una función lineal de  $X_2$ . Además, la matriz de covarianza del error de predicción  $\Sigma_{X_1|X_2}$ , no depende de los valores de  $X_2$ . Estos resultados se pueden extender directamente al problema de predicción de los procesos Gaussianos (Davis, 2014)

Supongamos que  $\{X_t, t = 1, 2, \dots\}$  es un proceso estacionario Gaussiano con media  $\mu$  y función de autocovarianza  $\gamma(\cdot)$ , y que el vector aleatorio que consta de las primeras  $n$  observaciones,  $X_n = (X_1, \dots, X_n)$ , el objetivo es tratar de pronosticar la siguiente observación  $X_{n+1}$ , el mejor predictor para un paso viene dado por:

$$\hat{X}_{n+1} = E(X_{n+1}|X_1, \dots, X_n) = \mu + \phi_{n1}(X_n - \mu) + \dots + \phi_{nn}(X_1 - \mu)$$

Donde

$$(\phi_{n1}, \dots, \phi_{nn}) = \Sigma_n^{-1}\gamma_n$$

además,  $\Sigma_n = cov(X_n, X_n)$ , y  $\gamma_n = cov(X_{n+1}, X_n)$ , el error medio cuadrático de predicción es:

$$\nu_n = \gamma_0 - \gamma_n'\Sigma_n^{-1}\gamma_n \quad (27)$$

Todas las fórmulas presentadas asumen la no singularidad de  $\Sigma_n$ , Por lo tanto, existe una relación lineal entre los  $X_1, \dots, X_n$  y el problema de predicción que puede ser resuelto al escoger un subconjunto de los predictores generales que consisten de variables lineales independientes. Cabe recalcar que la matriz de no covarianza que resulta de ese subconjunto también debe ser no singular.

### 7.2.11. Optimización Bayesiana usando procesos de Gauss

Para usar procesos Gaussianos en la optimización Bayesiana, debemos establecer el dominio del proceso de Gauss  $X$  que sea el espacio de hiperparámetros, usualmente el kernel se puede convertir en otro parámetro que necesita ser definido. Por tanto, el proceso Gaussiano puede ser usado como aproximación previa para los datos que no se conocen de la función de pérdida.

Consecuente, el uso de optimización Bayesiana requiere el uso de ideas previas que detallaremos a continuación. Supongamos que queremos optimizar una función  $f : X \rightarrow \mathfrak{R}$ , donde  $X$  es un conjunto de hiperparámetros que queremos optimizar.

- Modelar  $f$  como una función de distribución.
- Si evaluamos  $x_1, x_2, \dots, x_D$  en  $f$ , consideraremos a  $f(x_1), f(x_2), \dots, f(x_D)$  como variables observadas del modelo.
- Si hay alguna  $x$  que no ha sido evaluada, entonces se considera  $f(x)$  como variable oculta.
- A pesar de no haber calculado  $f(x)$  se puede calcular la distribución condicional  $P(f(x)|f(x_1), f(x_2), \dots, f(x_D))$
- Podemos usar la distribución condicional del punto anterior para estimar  $f(x)$  para valores de  $x$  que no son observados.
- La distribución condicional también puede ser usada para escoger el siguiente valor de  $x$  al que vamos a calcular  $f(x)$ .

Para poder realizar el procedimiento hay ciertos parámetros y decisiones que se necesita tomar, el primero la distribución inicial que dotara de muchas propiedades a la función  $f$ , como se ha mencionado antes se asume un proceso Gaussiano. La siguiente es determinar la función de adquisición, basado en cómo vamos a seleccionar el siguiente punto, dado una distribución condicional de valores de  $f(x)$ .

### **Función de Adquisición:**

La función principal como fue mencionada anteriormente es la búsqueda de nuevos puntos. Algunos de las funciones clásicas de adquisición que son usados para inteligencia artificial son para la suposición distribución de Gauss. Además, es generalmente una función que considera tres aspectos: la media de la variable oculta  $f(x)$ , la desviación estándar de  $f(x)$ , y el mejor valor que se ha encontrado hasta el momento en la optimización.

### **Probabilidad de Mejora:**

La probabilidad de mejora de la función de adquisición nos pide que calculemos la probabilidad de que el siguiente punto que se escoja sea una mejoría respecto al punto que hemos observado. Esto es, tratar de maximizar la probabilidad de que el  $x_i$  actual sea nuestro nuevo mejor  $x_i$ . Definimos esta probabilidad como:

$$P(f(x^* < y_{mejor})) = P\left(\frac{f(x^*) - \mu(x^*)}{\sigma(x^*)} < \frac{y_{mejor} - \mu(x^*)}{\sigma(x^*)}\right) = \Phi\left(\frac{y_{mejor} - \mu(x^*)}{\sigma(x^*)}\right) \quad (28)$$

Donde  $\Phi$  es la función de distribución acumulativa de la distribución estándar de Gauss.

### **Mejora Esperada:**

La mejora esperada de la función de adquisición nos pide minimizar en el valor del nuevo  $y_{mejor}$  antes de que se realice la búsqueda de un nuevo punto. Este valor esperado es:

$$\mathbf{E}[(\min(f(x^*) - y_{mejor}, 0))] = \mathbf{E} \left[ \frac{f(x^*) - \mu(x^*)}{\sigma(x^*)} - \frac{y_{mejor} - \mu(x^*)}{\sigma(x^*)} \right] \sigma(x^*) \quad (29)$$

sí suponemos que  $\mu \sim \mathcal{N}(0, 1)$ , podemos obtener una expresión simplificada para  $E_\mu[(\min(\mu - c, 0))]$  al resolver la integral, donde  $\phi$  es la función de probabilidad de distribución de Gauss.

$$E_\mu[(\min(\mu - c, 0))] = \int_{-\infty}^c (\mu - c)\phi(\mu) d\mu = -\phi(c) - c\Phi(c) \quad (30)$$

Por tanto, podemos inferir que la función de adquisición es:

$$a_{EI}(y_{mejor}, \mu, \sigma) = -\left(\phi\left(\frac{y_{mejor} - \mu}{\sigma}\right) + \frac{y_{mejor} - \mu}{\sigma} * \Phi\left(\frac{y_{mejor} - \mu}{\sigma}\right)\right) \quad (31)$$

Por tanto, podemos establecer un algoritmo para la optimización Bayesiana usando proceso de Gauss.

---

### Algorithm 8 Optimización Bayesiana

---

**Result:** Optimización de la función objetivo Z

---

```

ymejor ← ∞
for i=1 a Ninicial do
  Seleccionar xi
  Calcular la función de pérdida, yi ← f(xi)
  if yi ≤ ymejor then
    xmejor ← xi
    ymejor ← yi
  end
end
for i=Ninicial + 1 a N do
  Actualizar la matriz de Kernel Σ
  Sea μ(x) y σ(x) denotan la media y la desviación estándar respectivamente de f(x)
  bajo el supuesto del proceso Gaussiano.
  xi ← argminx, a(μ(x), σ(x), ymejor)
  Calcular la función de pérdida, yi ← f(xi)
  if yi ≤ ymejor then
    xmejor ← xi
    ymejor ← yi
  end
end

```

---

Posteriormente, se procederá a implementar estos algoritmos y analizar los resultados obtenidos.

## **8. Implementación: Caso Ambacar**

### **8.1. Sobre Ambacar**

Ambacar es una empresa orgullosamente ecuatoriana creada en los años 70 para competir en el creciente mercado automotriz ecuatoriano. Siendo sus inicios la ciudad de Ambato. En la actualidad cuenta con una amplia red de concesionarios y talleres de servicio técnico y repuestos a nivel nacional (s.f.).

Durante el boom de la industria automovilística china a inicios del nuevo milenio, Ambacar vio la oportunidad de ser pionera en el mercado ecuatoriano con el ingreso de marcas chinas que en la actualidad ya son consolidadas como: Great Wall, Haval, Zotye, Soueast. Siendo las primeras marcas constantes entre los vehículos más vendidos a nivel nacional.

A inicio de la década pasada, Ambacar vio el potencial y calidad de la mano de obra ecuatoriana y comenzó la fabricación nacional de alguno de los modelos más populares y un gran catálogo de repuestos. Ambacar potencia la generación de empleo en el Ecuador, a través de la creación de su empresa aliada, Ciauto (Ciudad del auto), la cual se encarga de ensamblar los vehículos más solicitados por miles de ecuatorianos con una gran pasión y dedicación por la excelencia en su elaboración (s.f.).

La búsqueda de la excelencia y atención al detalle ha llevado a la empresa a la búsqueda de un camino donde la atención al cliente sea el pilar más importante dentro de la filosofía y concepción de Ambacar. Por esta razón, el objetivo primordial es brindar un gran servicio post venta donde el cliente sin importar su ubicación en el país sepa que puede encontrar uno de los más de veinte puntos especializados dentro de la red, con repuestos originales, equipos de última generación para garantizar un buen desempeño de su vehículo.

En base a los buenos resultados obtenidos en el Ecuador durante la última década y gracias al fiel deseo de expandir el modelo a nivel internacional, con el gran esfuerzo, dedicación, compromiso y el trabajo en equipo de cada uno de nuestros colaboradores, ahora también nos encontramos presentes en Costa Rica, Colombia y Perú (s.f.).

### **8.2. Red de concesionarios**

Ambacar en su objetivo de brindar un mejor servicio a sus clientes y acercarse a una cobertura a nivel nacional, cuenta la actualidad con más de 20 concesionarios que brindan atención postventa y venta de repuestos.

Consecuentemente, el servicio postventa de Ambacar está presente en 13 provincias y 15 ciudades diferentes (Esmeraldas, Ibarra, Quito, Santo Domingo, Latacunga, Coca, Ambato, Puyo, Machala, Guayaquil, Salinas, Manta, Portoviejo y Quevedo) y en algunas ciudades como Quito, Guayaquil y Ambato con múltiples concesionarios.



El proceso de almacenamiento de Ambacar consta de dos niveles, como primer nivel, un centro de distribución principal centralizado ubicado a las afueras de la ciudad de Ambato que se encarga de abastecer a todos los concesionarios y, como segundo nivel se encuentran todos los concesionarios que poseen atención postventa o bien venta de repuestos. Cabe aclarar que Ciauto (centro de distribución) también realiza venta de repuestos a concesionarios autorizados.

A continuación, mostraremos algunas gráficas generadas por la aplicación Tableau para mostrar los distintos concesionarios y el centro de distribución ubicado en el mapa del Ecuador.

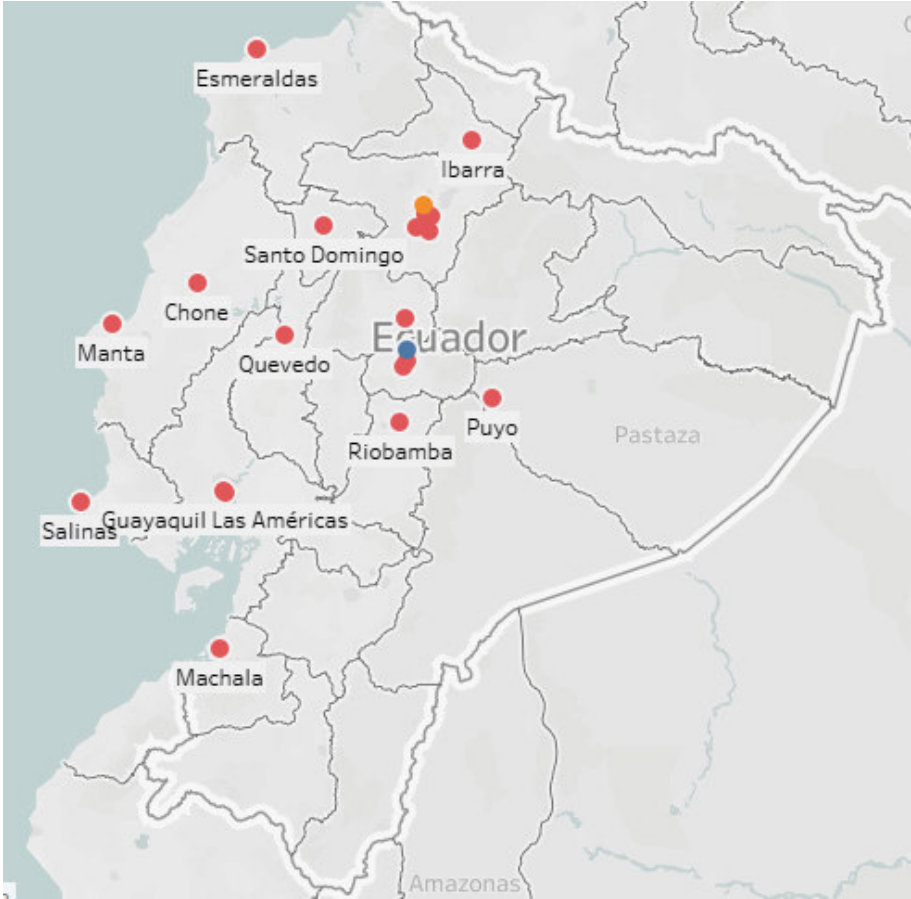


Figura 8: Mapa completo del Ecuador con las ubicaciones de los distintos concesionarios y centro de distribución.

En la figura anterior podemos observar que los puntos rojos y naranja representan los concesionarios con centros de atención postventa (es decir, que tengan taller y una bodega de almacenamiento). Marcando con un punto azul ubicado en la provincia de Tungurahua para señalar el centro de distribución principal de toda la red de distribución de Ambacar.

Es importante señalar que el centro de distribución principal (Ciauto) es la ensambladora de algunos de los repuestos y vehículos de algunas de las marcas comercializadas por

Ambacar.

A continuación, mostraremos los mapas de Quito y Ambato que son las dos ciudades más importantes para la empresa a nivel nacional tanto en venta de vehículos como venta de repuestos.



Figura 9: Mapa de la provincia de Tungurahua que muestra las ubicaciones de los concesionarios en Ambato y el centro de distribución principal.

En la figura anterior se muestran los dos concesionarios de la ciudad de Ambato (los primeros concesionarios de Ambacar a nivel nacional, Ambato Indoamérica y Ambato Manuelita Sáenz) y la ubicación de Ciauto al norte de la provincia.

La información proporcionada por la gerencia de compras y planificación de repuestos de Ambacar, en la ciudad de Ambato es casi inmediato el despacho de repuestos por parte de Ciauto a los dos concesionarios de la ciudad siendo Ambato uno de los pilares fuertes de ventas dentro de la compañía por lo tanto mantener altos estándares de servicio al cliente es primordial dentro de la ciudad.

Con el paso de los años, Quito se convirtió en el bastión de ventas de la empresa debido a su mayor población y vehículos más enfocados al uso dentro de la ciudad. En consecuencia, Quito es la ciudad con más concesionarios y servicios postventa, tal y como lo indica el siguiente gráfico.

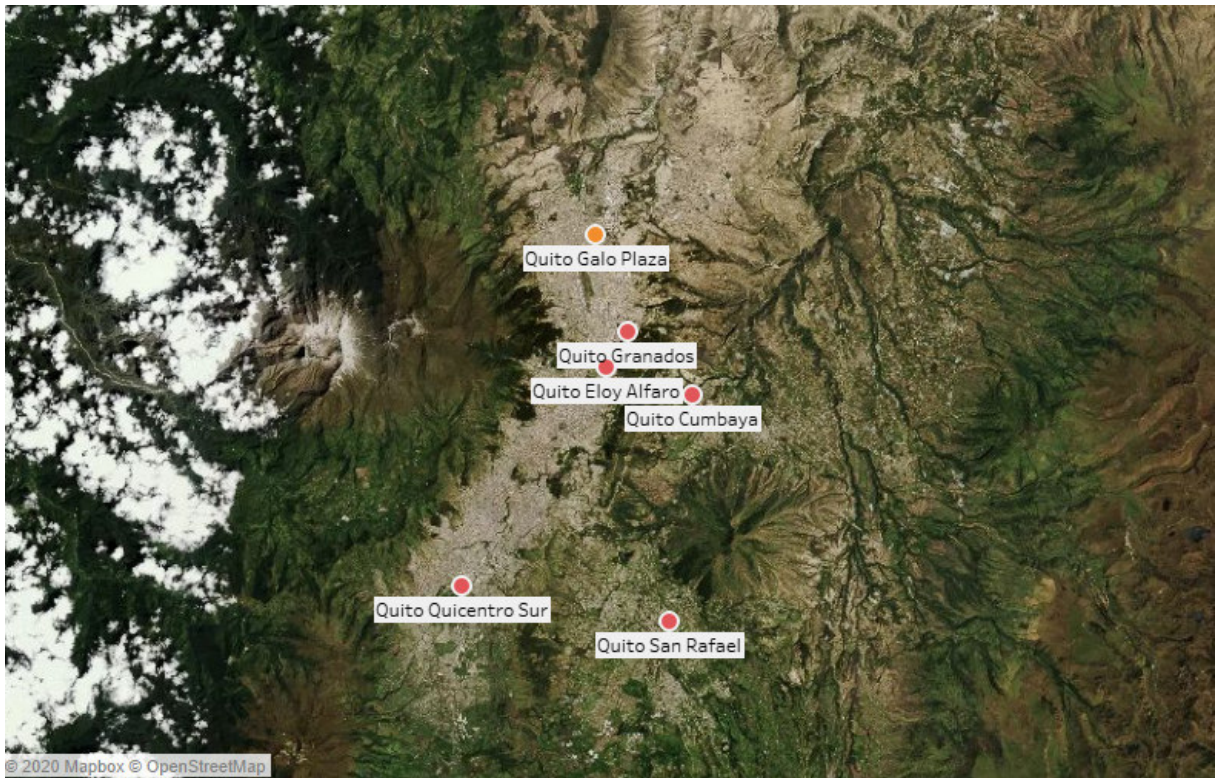


Figura 10: Mapa de la provincia de Pichincha que muestra las ubicaciones de los concesionarios en Quito y el centro de distribución regional.

Al momento de realizar el análisis de venta de los repuestos a nivel de punto de venta, es importante hacer una diferenciación respecto a las ciudades; ya que las ventas de Ambarcar están mayormente concentradas en 3 provincias: Pichincha, Guayas y Tungurahua (AEADE, 2020).

Todos los conceptos previos antes de realizar un análisis más profundo nos sugieren que los comportamientos de los diferentes concesionarios serán distintos entre sí, es por eso por lo que es necesario realizar análisis tanto del comportamiento de los clientes y de los vehículos asociados a cada concesionario. Analizaremos medidas el cálculo promedio de veces que un vehículo vendido en la agencia retorna al taller en los años 2017, 2018, 2019 y 2020. Kilometraje promedio recorrido en los mismos años.

Usaremos la siguiente nomenclatura para diferenciar las distintas agencias en las distintas ciudades y provincias.

Agencia	Bodega	Ciudad	Provincia
RI	Riobamba	Riobamba	Chimborazo
LT	Latacunga	Latacunga	Cotopaxi
MC	Machala	Machala	El Oro
ES	Esmeraldas	Esmeraldas	Esmeraldas
GY	Guayaquil	Guayaquil	Guayas
GJ	Juan Tanca Marengo	Guayaquil	Guayas
GH	GYQ Haval	Guayaquil	Guayas
IB	Ibarra	Ibarra	Imbabura
VD	Quevedo	Quevedo	Los Ríos
CH	Chone	Chone	Manabí
MT	Manta	Manta	Manabí
CO	Coca	Coca	Orellana
QB	Los Bancos	Los Bancos	Pichincha
GR	Granados	Quito	Pichincha
QC	Quicentro	Quito	Pichincha
QN	Quito Norte	Quito	Pichincha
QE	Seminuevos	Quito	Pichincha
QA	Eloy Alfaro	Quito	Pichincha
GP	Guayaquil Perimetral	Guayaquil	Guayas
SR	San Rafael	San Rafael	Pichincha
SD	Santo Domingo	Santo Domingo	Santo Domingo
MA	Matriz	Ambato	Tungurahua
QP	Galo Plaza	Quito	Pichincha
IN	Bolivariana	Ambato	Tungurahua
SA	Salinas	Salinas	Santa Elena
MS	Manuelita Saenz	Ambato	Tungurahua
MU	Seminuevos Amba	Ambato	Tungurahua
PV	Portoviejo	Portoviejo	Manabí

Cuadro 1: Código, nombre comercial, ciudad y provincia de cada uno de los centros de distribución postventa

Para el primer análisis. El resultado se obtiene de la data de cada uno de los concesionarios con servicio postventa, la identificación de cada vehículo se realiza por medio de su código único (número de chasis) y se realizan los cálculos en base a la fecha de facturación del servicio. Posteriormente, se realizará un conteo de cuantas veces determinado vehículo fue al taller por año calendario y finalmente se hace un promedio entre todos los vehículos que fueron vendidos por ese concesionario y se encuentren activos.

Finalmente, obtenemos la siguiente gráfica de las distintas agencias.

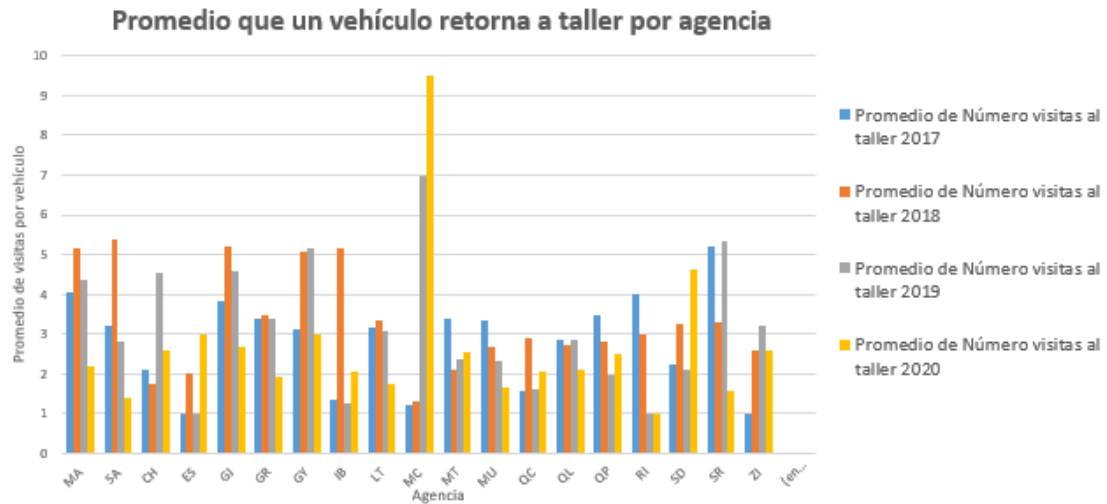


Figura 11: Promedio que un vehículo nuevo retorna al taller por agencia.

Podemos notar por la gráfica anterior que en términos generales existe un decrecimiento en el número de vehículos nuevos que acuden al taller. Esto se puede explicar de manera sencilla por el efecto de la pandemia y las restricciones de movilidad. La mayoría de los vehículos de uso personal o familiar recorrieron menos kilómetros durante el 2020 y solo los vehículos de uso comercial rodaron más durante el último año.

### 8.3. Análisis importancia de marcas

Al cierre del año 2019 Ambacar cuenta con 4 marcas dentro de su catálogo de vehículos: Haval-Great Wall, Soueast, Shineray y Zotye. A mediados del año 2020 se incorporó la marca SWM con un solo modelo de vehículo por lo que será excluido de los análisis.

Para el análisis de la marca más importante de Ambacar usaremos los datos proporcionados tanto por la AEADE (Asociación de empresas automotrices del Ecuador) como por Ambacar (Los datos proporcionados serán al cierre del año 2019, ya que se considera el 2020 un año atípico en venta de vehículos).

MARCA	UNIDADES VENDIDAS 2019
HAVAL-GREAT WALL	7305
SHINERAY	762
SOUEAST	1779
ZOTYE	521

Cuadro 2: Unidades vendidas por Marca que comercializa Ambacar en el año 2019.

La marca Haval-Great Wall al contar dentro de su catálogo con 13 modelos (entre vehículos de pasajeros y camionetas) y ser la marca más consolidada dentro de Ambacar. Es la marca más importante con una participación del 71 % dentro de Ambacar y el 5.53 % a nivel nacional (AEADE, 2020).

De manera informativa mostraremos los vehículos más vendidos de la marca durante el año 2019:

MODELO	UNIDADES VENDIDAS 2019
M4 CONFORT	1571
WINGLE 5 2.2 CD 4X2	919
VOLEEX C30	596
HAVAL H6 SUPREME	569
WINGLE 5 2.4 CD 4.2	506

Cuadro 3: Modelos y versiones más vendidos de la marca Haval-Great Wall en el año 2019

El análisis de importancia de la marca nos permite determinar que Haval-Great Wall es la marca más importante dentro de todo el portafolio de Ambacar. Además, nos ayudará a segregar y filtrar la información para extraer la mayor cantidad de estadísticos posibles sobre los vehículos de la marca. Así, tener un mejor pronóstico sobre que ítems necesitan cada vehículo y lograr planear con antelación.

Por otro lado, Haval-Great Wall cuenta en su catálogo de productos tanto vehículos de uso comercial como de uso personal o familiar. La siguiente gráfica mostrará como estuvieron repartidas las ventas durante el año 2019. Consideraremos a las camionetas como vehículos de trabajo y uso comercial.

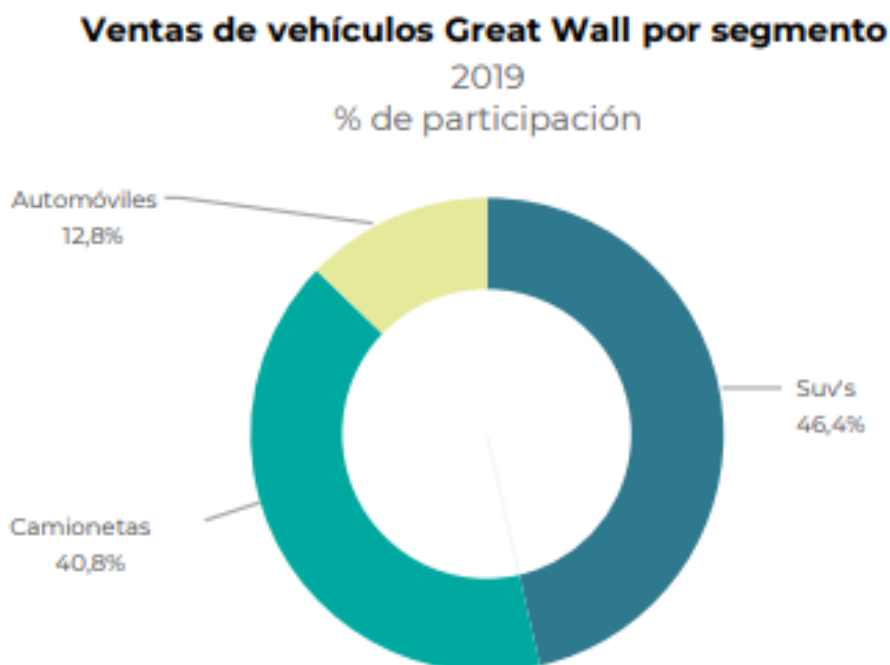


Figura 12: Ventas de vehículos de Ambacar por segmento

Adicionalmente, es importante realizar una diferencia dentro del comportamiento del consumo de los vehículos (Kilometraje recorrido), ya que al tener una combinación del 59.2% vehículos de pasajeros y un 40.8% de vehículos que son en su mayoría usados con

finés laborales. Los vehículos de pasajeros por su naturaleza recorren menos kilómetros anualmente que los vehículos de trabajo. Por lo tanto, los vehículos de trabajo tienen más probabilidad que un componente falle y visite de manera más regular un taller.

### 8.3.1. Ventas Haval-Great Wall en el año 2020

La venta de vehículos en el año 2020 fue a nivel mundial afectada por la pandemia causada por el coronavirus. Para Ambacar los niveles de venta llegaron a apenas techos un poco superiores a los vistos en el año 2015-2016 en la marca Haval-Great Wall.

En el año 2020 se vendieron 3738 unidades, con la mayoría de las ventas acumuladas en los primeros dos meses del año y una recuperación en el último trimestre. A continuación; mostramos un gráfico de como progresaron las ventas en el transcurso del 2020.

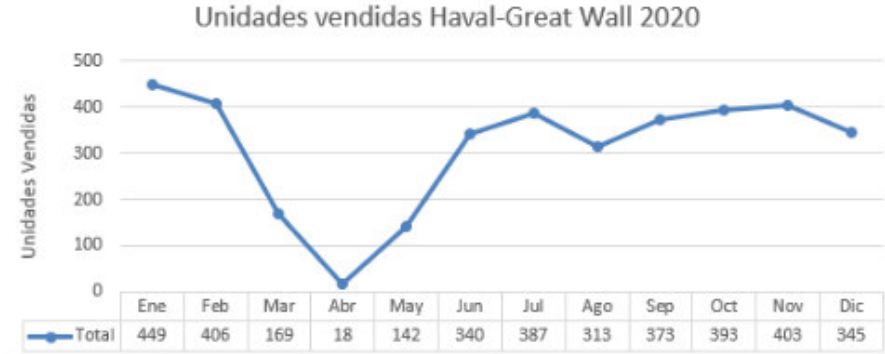


Figura 13: Unidades vendidas por mes en el año 2020

De manera similar es necesario realizar una diferenciación entre las unidades vendidas por las distintas agencias. El siguiente gráfico podemos ver las unidades vendidas por las distintas agencias.

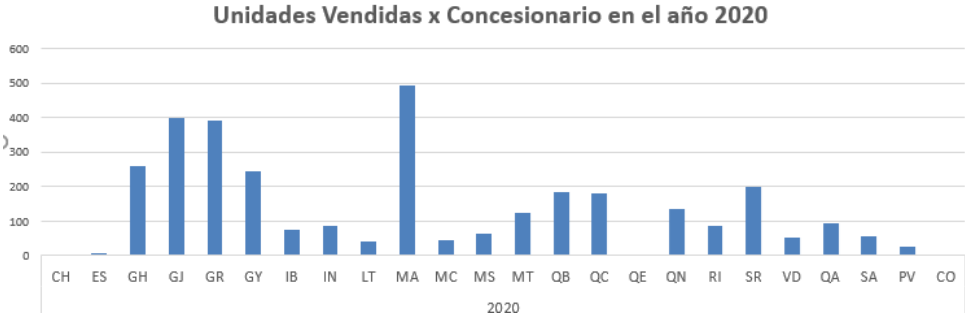


Figura 14: Unidades vendidas por agencia en el año 2020

El gráfico anterior nos muestra que los concesionarios ubicados en las ciudades de Quito, Guayaquil y Ambato son los que mayor venta de vehículos concentran, en especial los concesionarios Matriz Ambato, Juan Tanca Marengo y Granados.

## 8.4. Estimación kilometraje recorridos

Para el cálculo del estadístico de la estimación del kilometraje recorrido por año, usaremos la información proporcionada de la base de datos del área de taller. Usaremos la información que se recopila cada vez que un vehículo entra al servicio postventa. Uno de los datos que se registran es el kilometraje del vehículo. Haremos un registro por auto tomando la primera y última vez que visita el taller durante un año calendario. A demás compararemos el último kilometraje marcada en el año  $n - 1$  con el primer kilometraje marcado en el año  $n$ .

En base a los cálculos realizados obtenemos los siguientes resultados. Notemos que se hace una diferenciación respecto al uso del vehículo.

Línea	Tipo de uso	Km rec. 2018	Km rec. 2019	Km rec. 2020
GWM	Comercial	13411	25558	58191
GWM	Particular	11820	17468	13814

Cuadro 4: Promedio de Kilometraje recorrido de los vehículos de la marca Haval-Great Wall por línea de negocio.

Podemos notar el significativo descenso en el kilometraje recorrido en el año 2020 debido a la pandemia y el incremento en el uso de los vehículos comerciales que también se explica con la introducción en el año 2018 y 2019 de camionetas exitosas de la marca en el mercado ecuatoriano.

De manera similar, podemos realizar una estimación del kilometraje promedio y edad promedio de los autos activos (es decir, que su última visita al taller no ha superado los 400 días).

Para la marca Haval-Great Wall tenemos los siguientes resultados.

Kilometraje promedio	km
Particular	53451
Comercial	83745
Edad promedio	años
Particular	2.5
Comercial	1.83

Cuadro 5: Ítems disponibles por código A001-A009 en los puntos de atención postventa

Los resultados anteriores muestran que la edad promedio de los vehículos que actualmente se encuentran en circulación es de 2.5 años, este resultado concuerda con los datos de ventas de la AEADE; en el año 2017 y 2018 Haval-Great Wall registran sus mejoras ventas con 6792 y 8390 unidades respectivamente (AEADE,2020).

Todos estos estadísticos obtenidos nos ayudarán como variables de apoyo al momento de realizar el pronóstico de la demanda con el método Random Forest.



## 8.5. Categorización de los repuestos

Ambacar cuenta actualmente con más de 23000 ítems dentro de su catálogo de repuestos, desde elementos básicos como aceites para motor y bujías, hasta piezas complejas de cajas de cambios y elementos del motor.

Para categorizar los repuestos vamos a realizar un análisis de Pareto que nos permita clasificar los ítems más importantes por su rotación y unidades vendidas ya sea en taller o en la caja (las ventas de repuestos se dan de dos maneras en Ambacar. La primera por taller, es decir, que taller de acuerdo con la necesidad de los vehículos que requieren mantenimiento compran los repuestos a la bodega; y la segunda, por caja, cuando un cliente particular se acerca al concesionario a comprar cierto repuesto).

Los cálculos simples de categorización nos arrojaron los ítems más importantes que serán detallados a continuación. Debemos aclarar que el código real de los artículos ha sido clasificado por confidencialidad y protección de los intereses corporativos de Ambacar y parte de la descripción ha sido cortada por las mismas razones antes mencionadas.

Los ítems seleccionados todos son repuestos, aceites, bujías, partes o accesorios que son usadas en los distintos modelos de la marca Haval-Great Wall. Para la selección de los datos para este cálculo se usaron los datos desde el primero de enero del 2020 hasta el 31 de diciembre del 2020. Se excluyeron ventas puntuales realizadas por Ciauto hacia los concesionarios autorizados.

A continuación, mostraremos la lista de los 50 ítems más importantes de la marca Haval-Great Wall.

LÍNEA	CÓDIGO	DESCRIPCIÓN	TIPO PROCEDIMIENTO
GWM	A001	GUARDACHOQUE DELT WINGLE 7	NO COMÚN
GWM	A002	PLATO DE EMBRAGUE (H6-1.5T)	NO COMÚN
GWM	A003	GUARDACHOQUE DEL.	NO COMÚN
GWM	A004	NEBLINERO DELANTERO RH	NO COMÚN
GWM	A005	SOPORTE METAL. SUP. GUARD. DLT WINGLE 5E	NO COMÚN
GWM	A006	MODULO PROXIMIDAD- PUERTA POST LH	NO COMÚN
GWM	A007	RODAMIENTO DE EMBRAGUE (H5-2.0TCI)	NO COMÚN
GWM	A008	NEBLINERO DELANTERO LH	NO COMÚN
GWM	A009	OVERFENDER DELT RH WINGLE 5E	NO COMÚN
GWM	A010	OVERFENDER DELT LH WINGLE 5E	NO COMÚN
GWM	A011	TEMPLADOR DISTRIBUCION MOTOR 4D20	NO COMÚN
GWM	A012	SOPORTE GUARDACHOQUE DELT RH	NO COMÚN
GWM	A013	DISCO DE EMBRAGUE NUEVO GW4G15	COMÚN
GWM	A014	BUJIAS IKH16TT/ IKH20TT	COMÚN
GWM	A015	ANILLO SINCRONIZADOR 3-4	COMÚN
GWM	A016	PLATO DE EMBRAGUE ANTIGUO 4G15	COMÚN
GWM	A017	RODAMIENTO BANDA DISTRIBUCION 4D20	COMÚN
GWM	A018	NIQUELADO POST DE COMPUERTA M4	COMÚN
GWM	A019	DISCO FRENO DEL H5T/4X2 - NUEVOH3 - 4D20	NO COMÚN
GWM	A020	SELLO DE VALVULAS 4D20 - H6D	NO COMÚN
GWM	A021	VINCHA POSTERIOR	NO COMÚN
GWM	A022	NEBLINERO DELT LH WINGLE 5E	NO COMÚN
GWM	A023	SENSOR VELOCIDAD ABS LH	COMÚN
GWM	A024	BUJIA DE ENCENDIDO H6 COUPE ALL NEW - H9	COMÚN
GWM	A025	DISCO DE EMBRAGUE ANTIGUO 4G15	COMÚN
GWM	A026	BARRA ANTICOLISION	NO COMÚN
GWM	A027	BANDA DE DISTRIBUCION (4D20)	NO COMÚN
GWM	A028	CORREDIZA PLASTICA DELT RH (H5-2.0TCI)	NO COMÚN
GWM	A029	LOGO WINGLE	COMÚN
GWM	A030	KIT CABLES DE BUJIAS 4G63-4G69	COMÚN
GWM	A031	TAPA CUBO	COMÚN
GWM	A032	DISCO DE EMBRAGUE (H6-1.5T)	COMÚN
GWM	A033	MANIJA INT PUERTA DELT LH H5 ELITE - H3	NO COMÚN
GWM	A034	ROTULA DE SUSPENSION INFERIOR DELT	NO COMÚN
GWM	A035	MULTIPLE DE ESCAPE (H5-2.0TCI)	NO COMÚN
GWM	A036	ROTULA DE SUSPENSION SUPERIOR DELT	NO COMÚN
GWM	A037	RELE FLASH	COMÚN
GWM	A038	RETENEDOR POST CIGUENAL	COMÚN
GWM	A039	BANDA ALTERNADOR MOTOR 4D20	COMÚN
GWM	A040	TAPA DE LLENADO DEL RADIADOR	COMÚN
GWM	A041	PLATO DE EMBRAGUE NUEVO GW4G15	COMÚN
GWM	A042	TERMOSTATO (H5-2.0TCI)	NO COMÚN
GWM	A043	MASCARILLA DELT HAVAL M4	NO COMÚN
GWM	A044	NEBLINERO DELT RH WINGLE 5E	NO COMÚN
GWM	A045	EMPAQUE TAPA VALVULAS	NO COMÚN
GWM	A046	RODAMIENTO EMBRAGUE 2.8	COMÚN
GWM	A047	PLATO DE EMBRAGUE (2.8TC)	COMÚN
GWM	A048	TRABADOR DEL SELECTOR MARCHAS	NO COMÚN
GWM	A049	BUJIA DE PRECALENTAMIENTO	COMÚN
GWM	A050	CABLES DE BUJIAS	COMÚN

Cuadro 6: Top 50 de los repuestos más importantes para el servicio postventa de Ambacar

En la tabla anterior podemos apreciar a los 50 ítems más importantes en base a las ventas del 2020, alrededor del 50% de los repuestos de la lista pertenecen a procesos cotidianos de revisiones de rutina ya sean estas revisiones por cumplimiento de determinado kilometraje o cambios de aceite. El porcentaje restante se debe a piezas que fallan inesperadamente por desgaste del vehículo o por algún accidente de este.

Con los estadísticos e información obtenida el paso siguiente es el pronóstico de la demanda a nivel ítem-concesionario. La siguiente sección trata de la construcción del modelo y los resultados obtenidos.

## 8.6. Pronóstico de la demanda aplicado a Ambacar

Para el módulo del pronóstico de la demanda, Ambacar nos proveyó los datos mensuales de venta en los distintos talleres desde su fecha de apertura hasta el 31 de diciembre del año 2020. Es decir, datos desde el 2010 para su concesionario GR y MA (Granados Quito, Matriz Ambato) y conforme han ido creciendo la red de concesionarios, los datos también han sido proporcionados para este estudio.

Recordemos que los datos usados para este análisis son para la línea GMW (Great Wall Motors-Haval) y solo son considerados los 50 artículos que en base a las ventas del año 2020 son los más importantes para Ambacar. Se obtienen aproximadamente 13000 datos de los 22 concesionarios y de las ventas puntuales que se han generado en el centro de distribución. Notemos que estas ventas se realizan a concesionarios aliados de la red Ambacar como es el caso de Maxximundo en la ciudad de Cuenca que desde mediados de la década pasada es el representante de las marcas de Ambacar y se encarga del cumplimiento de garantía de los vehículos vendidos en Cuenca.

Para la construcción del modelo de pronóstico de la demanda usando el método Random Forest, en primer lugar, se ha usado Power Query para la consolidación de la data. En los siguientes párrafos se explicará el origen de las variables usadas y la construcción del modelo. A continuación, mostraremos las variables usadas.

- **Año:** Se ha decidido introducir la variable Año ya que Ambacar es un negocio que constantemente ha ido creciendo con el paso del tiempo, ya sea abriendo nuevos concesionarios o impulsando el posicionamiento de sus marcas a nivel nacional; logrando hoy en día ser una de las marcas más reconocidas a nivel nacional. Para los conjuntos de entrenamiento y prueba los datos se encuentran desde el año 2010 al 2020.
- **Código Confidencial:** Está variable por las diferencias en el uso y aplicación que tienen cada uno de los distintos repuestos se ha decidido que es importante para la diferenciación al momento de realizar el pronóstico de la demanda. La variable consta de factores desde 1 a 50. Código Confidencial se usará como variable cualitativa.
- **Venta de Vehículos por Agencia:** La variable consta de vehículos de la marca GMW (Great Wall Motors-Haval) vendidos en los distintos concesionarios por año y por mes. Esta variable guarda una alta relación respecto al número de repuestos vendidos ya que a mayor número de vehículos facturados es natural que estos requieran más repuestos y es más probable que acudan a la red de concesionarios.

De manera de ejemplo, mostraremos las ventas de los dos primeros meses del año 2015 para mostrar la forma en la que está estructurado esta variable.

Año	Mes de Venta	Agencia	Unidades Vendidas (GWM)
2015	Ene	GJ	30
2015	Ene	GR	61
2015	Ene	GY	55
2015	Ene	IB	18
2015	Ene	IN	24
2015	Ene	MA	55
2015	Ene	MT	1
2015	Ene	QC	26
2015	Ene	QN	26
2015	Ene	QP	3
2015	Ene	SD	19
2015	Ene	SR	36
2015	Feb	GJ	25
2015	Feb	GR	58
2015	Feb	GY	48
2015	Feb	IB	12
2015	Feb	IN	15
2015	Feb	MA	49
2015	Feb	MT	1
2015	Feb	QC	23
2015	Feb	QN	13

Cuadro 7: Ejemplo de la variable: Vehículos Vendidos

- **Solo repuestos?:** Muchos de los actuales concesionarios empezaron su vida siendo solamente talleres de servicio postventa. Es por eso, que en muchos casos no se registran ventas de vehículos y solo constan registros de talleres y venta de repuestos. Identificaremos como 1 o 0, donde 1 representa que el concesionario solo prestaba servicio postventa y venta de repuestos; 0 el caso contrario. Los clientes prefieren los talleres que se encuentran en concesionarios a los que solo son talleres.
- **Promoción:** En muchos negocios el uso de estrategias comerciales como las promociones se han convertido en claves para llamar la atención de los clientes y traer recursos a las empresas. Ambacar no es ajena al uso de tácticas promocionales para llamar la atención de sus clientes.

Adicionalmente, el uso de variables exógenas como promociones nos ayudarán a explicar la variabilidad de la data de ventas y puntos atípicos que se generan por los impulsos que los eventos comerciales ocasionan.

Usaremos 1 para identificar si un determinado ítem en algún mes y en uno o varios concesionarios se aplicó una promoción; 0 caso contrario.

- **Covid Cierre:** La industria automotriz ecuatoriana fue uno de los sectores más afectadas durante la pandemia ocasionada por el virus del coronavirus. Durante el semáforo en rojo impuesto por el gobierno ecuatoriano entre los meses de marzo y

julio del año 2020 la movilidad ecuatoriana se vio afectada, sobre todo la movilidad particular que fue reemplazada por el teletrabajo en muchos casos.

En consecuencia, tanto la venta de vehículos como la de repuestos durante estos meses se vio afectada. Por tanto, es importante diferenciar y señalar estos meses del resto de la base de datos. Usaremos 1 para identificar a los meses antes mencionados; caso contrario usaremos 0.

- **Covid No Cierre:** A partir del mes de agosto del 2020 las restricciones de movilidad se suavizaron y algunas personas optaron por volver a una normalidad relativa. De igual manera, la venta de vehículos se recuperó respecto a meses anteriores y la actividad de los talleres incrementó. Sin embargo; los niveles en los que se están desarrollando los mismos son reducidos respecto a años anteriores. Por tanto; es importante identificar y señalar estos meses de actividades atípicas en la data.

Usaremos 1 para identificar a los meses de agosto del 2020 en adelante; 0 caso contrario.

- **Promedio de días desde la última visita:** Al momento que uno de los vehículos vendidos por la empresa Ambacar entra a alguno de los concesionarios de la red, se registra al inicio dos factores primordiales, fecha de ingreso y kilometraje del vehículo.

En base a la información proporcionada por Ambacar de la base de datos de talleres. Podemos realizar una agrupación de la data de año-concesionario-vehículo para calcular en promedio a los cuantos días un vehículo de la marca GWM (Great Wall Motors-Haval) regresa al taller. Recordemos que existen ciertos factores para determinar si un vehículo está activo dentro de la red Ambacar que fueron explicados en secciones anteriores.

Posteriormente, obtenemos los resultados deseados al realizar una resta de la fecha del último ingreso al taller o en caso de que sea su primera vez en el taller, la información se compara con la fecha de venta del vehículo.

Mostraremos una muestra de los resultados obtenidos.

Año	PDV	Días desde última visita
CH	2018	34
CH	2019	56.80
CH	2020	57.48
CO	2020	55.27
ES	2019	29.75
ES	2020	84.57
GJ	2017	56.22
GJ	2018	69.40
GJ	2019	39
GR	2017	57.52

Cuadro 8: Ejemplo de la variable: Promedio de días desde la última visita.

- **Kilometraje Promedio:** En base a los datos proporcionados de los talleres de la red Ambacar y un análisis similar a la variable anterior, se busca determinar el kilometraje promedio recorrido de los clientes ya sean estos particulares o comerciales durante el año y por concesionario ya que a mayor kilometraje representa un mayor número de visitas y componentes dentro del vehículo que pueden necesitar refacción.

A continuación, mostraremos una tabla que muestra como está estructurada la variable Kilometraje promedio.

AGENCIA	2018	2019	2020
CH	6595.94	27864.24	12043.53
CO	861.88	5253.15	15997.97
ES	4298.52	17963.71	17608.50
GJ	20604.01	14110.90	7622.29
GR	12116.46	12219.26	6659.98
GY	12423.30	23275.66	16420.24
IB	27879.00	8847.65	12382.52
LT	9837.08	24283.68	10973.36
MA	15664.77	14191.47	7146.26

Cuadro 9: Ejemplo de la variable: Kilometraje promedio.

- **Venta total:** La variable muestra las ventas históricas de los distintos ítems en la red de concesionarios a nivel nacional.

## 8.7. Implementación y resultados del modelo de pronóstico Random Forest

La implementación del modelo de pronóstico por medio del método Random Forest consta de tres etapas. La primera de recopilación y depuración de la data. Una segunda fase de análisis de datos en donde visualizaremos la interacción de las variables entre sí y, por último; la construcción del modelo Random Forest.

### 8.7.1. Recopilación y depuración de la información:

La construcción de la matriz de datos necesita de una combinación de fuentes. La primera, los datos históricos de la venta de repuestos tanto de ventas generadas por el taller como de las ventas por ventanilla. De esta base de datos obtendremos la variable venta total. También en la base de datos se encuentra la información de las promociones de los repuestos.

De igual manera, Ambacar nos proporcionó la información de su base de venta de vehículos, en donde, extraemos la información de los vehículos vendidos de la marca GWM

(Great Wall Motors-Haval) que fueron vendidos en cada uno de los concesionarios durante los distintos meses y años que la marca ha estado presente dentro de Ambacar.

De las bases de datos mencionadas se realiza un trabajo de consolidación, obtenemos la información de los 22 actuales concesionarios y la cantidad de vehículos vendidos, no tenemos que no se realizará una diferenciación por modelo vendido, sino que usaremos la suma del consolidado del conjunto de modelos de la marca. Esta variable nos indicará el crecimiento de la marca y el potencial retorno de vehículos a los talleres.

Por último, los datos sobre registro de talleres y servicios realizados en el taller fueron usados con el propósito de obtener información del kilometraje promedio de los vehículos por año de cada uno de los concesionarios y el promedio de cuantos días sus vehículos retornan a los talleres. Al final, a pesar de que se puede diferenciar entre vehículos de uso particular y comercial por el kilometraje que recorren en el año y su frecuencia de visita a los talleres al momento de consolidar la data no se realiza ninguna diferenciación sobre el tipo de uso de los vehículos y se realiza una media ponderada para obtener los distintos promedios.

Consecuentemente, el siguiente paso es consolidar y combinar todas las variables. Para eso, usaremos el editor de Power Query (el código fuente de consolidación de las bases de datos se encontrará en los anexos). En la actual etapa también se crean las variables de Covid cierre, Covid no cierre y solo repuestos de acuerdo con las fechas de los datos.

Primero, debes mencionar a que a diciembre del 2020; la empresa Ambacar ha comercializado 56696 unidades a nivel nacional. Cada vehículo vendido cuenta con su código único (número de chasis). Por tanto, es muy fácil asociar a cada vehículo vendido: Taller de preferencia, año vendido, y ciertas cualidades de uso.

A continuación, la base de datos de la venta de repuestos en los talleres se encuentra estructurada por 9 columnas que nos indican año, concesionario, mes, línea (marca del vehículo), código del repuesto, descripción, mostrador, taller y total. A continuación, mostraremos una captura de la base de datos.

ANIO	AGE	MES	LINEA	CODIGO	DESCRIP	Mostrador	Taller	Total Ventas
2012	GY	12	GWM		FILTRO DE ACEITE	36	0	36
2012	QL	12	GWM		JUEGO DE PASTILLAS	0	3	3
2012	MA	12	GWM		CRUCETA	1	1	2
2012	GR	12	GWM		FILTRO COMBUSTIBLE	6	43	49
2012	GR	12	GWM		FOCO TIPO H3	4	3	7
2012	MA	12	GWM		PLATO DE EMBRAGUE	1	5	6
2012	QC	12	GWM		FILTRO DE COMBUSTIBLE	2	12	14
2012	GY	12	BYD		MANIJA EXTERIOR POST RH	1	0	1
2012	GY	12	GWM		PROTECTOR DE CAUCHO PARA PUERTA DEL RH	1	0	1
2012	QL	12	PINTU		BARNIS BRILLO CLEAR 923-155	0	2600	2600
2012	SD	12	GWM		VIDRIO PUERTA POST RH	0	1	1
2011	QS	12	MAZDA		GRAPAS	4	0	4
2011	MA	12	LUBRI		ACEITE TECTION 15W40	7	622	629
2011	MA	12	OTROS		FILTRO DE COMBUSTIBLE	0	9	9
2011	MA	12	OTROS		SINCRONIZADO C/C	0	2	2

Figura 15: Ejemplo de la data proporcionada de venta de repuestos.

La información de la columna código fue cubierta por acuerdo de confidencialidad. Los datos que se proporcionaron son mensuales.

Por lo tanto, mostraremos la información final que obtenemos por medio del editor Power Query. La data obtenida cuenta con 14 columnas: año, concesionario, mes, código codificado, descripción, total venta repuestos, venta de vehículos por agencia, solo repuestos, promoción, covid no cierre, covid cierre, días promedio desde la última visita, km promedio.

AN	Conces	M	Categoría	Códig	DESCRIP	Total	Venta de Vehículo	Solo rep	Prom	Covid	Covid N	Comb	Días última visita	Días	km por
2010	GR	12	31	TAPACUBO		1	7	0	0	0	0	0	2010-GF	85	600
2010	GR	10	15	ANILLO SINCRONIZADOR 3-4		2	6	0	0	0	0	0	2010-GF	85	600
2011	GR	9	33	MANIJA INT. PUERTA DEL LH		1	15	0	0	0	0	0	2011-GF	85	800
2011	MA	11	33	MANIJA INT. PUERTA DEL LH		1	35	0	0	0	0	0	2011-M/	85	800
2011	GR	11	45	EMPAQUE TAPA VALVULAS		1	13	0	0	0	0	0	2011-GF	85	800
2011	GR	11	46	RODAMIENTO EMBRAGUE		1	13	0	0	0	0	0	2011-GF	85	800
2011	GR	12	49	BUJIA DE PRECALENTAMIE'		6	43	0	0	0	0	0	2011-GF	85	800
2011	GR	12	31	TAPACUBO		8	43	0	0	0	0	0	2011-GF	85	800
2011	GR	12	25	DISCO DE EMBRAGUE		2	43	0	0	0	0	0	2011-GF	85	800
2011	GR	12	47	PLATO DE EMBRAGUE		2	43	0	0	0	0	0	2011-GF	85	800
2011	GR	12	46	RODAMIENTO EMBRAGUE		5	43	0	0	0	0	0	2011-GF	85	800
2011	GR	1	31	TAPACUBO		2	29	0	0	0	0	0	2011-GF	85	800
2011	MA	4	49	BUJIA DE PRECALENTAMIE'		1	85	0	0	0	0	0	2011-M/	85	800
2011	GR	9	29	LOGO WINGLE		1	15	0	0	0	0	0	2011-GF	85	800
2011	QL	8	15	ANILLO SINCRONIZADOR 3-4		2	0	1	0	0	0	0	2011-QL	85	800
2011	QL	10	46	RODAMIENTO EMBRAGUE		1	0	1	0	0	0	0	2011-QL	85	800

Figura 16: Ejemplo de la matriz con las variables consolidadas.

Con la matriz de datos consolidada procedemos a la segunda fase de análisis de datos.

## 8.8. Análisis de datos:

En esta etapa visualizaremos la información de las variables de la matriz consolidada, las distintas interacciones de las variables entre sí y su comportamiento.

Por medio del programa estadístico R analizaremos la distribución de la venta total de repuestos. El siguiente gráfico muestra dicha distribución.





Figura 17: Distribución de la venta total de repuestos.

Podemos apreciar que la mayoría de las ventas de los ítems son solo de una unidad, es decir, que los clientes solo compran el repuesto que necesita su vehículo y en muy pocas ocasiones compran para negocio.

Adicionalmente, en la siguiente gráfica podemos ver el crecimiento sostenido que tenía Ambacar hasta el año 2019, para está gráfica consolidamos las ventas anuales de los vehículos.

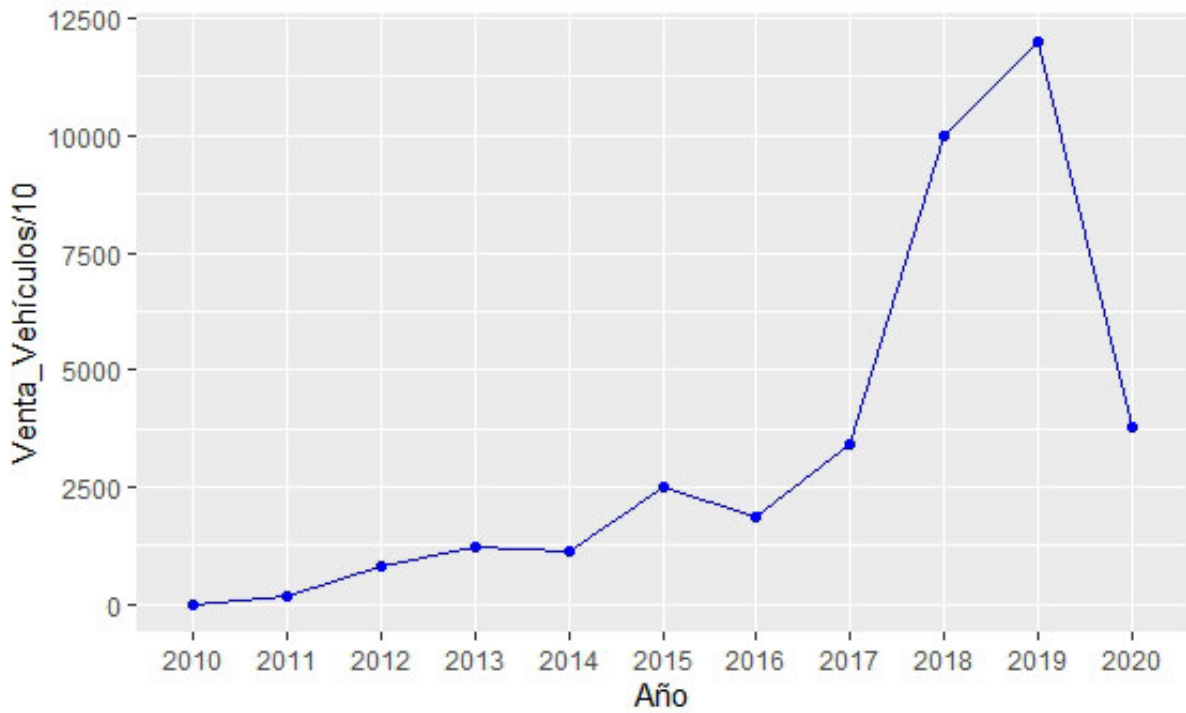


Figura 18: Ventas anuales de vehículos de Great Wall-Haval.

Podemos notar que hubo un crecimiento sostenido hasta el año 2018 y un crecimiento notable en el año 2019 debido al lanzamiento de modelos exitosos y renovaciones de la mitad de la flota de modelos. En el año 2020 por efectos de la pandemia la venta de vehículos es similar a la que podemos ver en el año 2017.

Posteriormente, los concesionarios tienen diferencias en el comportamiento de ventas, los concesionarios más antiguos y en las ciudades más grandes en población y en importancia económica tienden a llevarse la mayor cantidad de ventas. En el siguiente gráfico mostraremos la venta de vehículos de la marca GMW por cada año y concesionario desde su fecha de apertura.

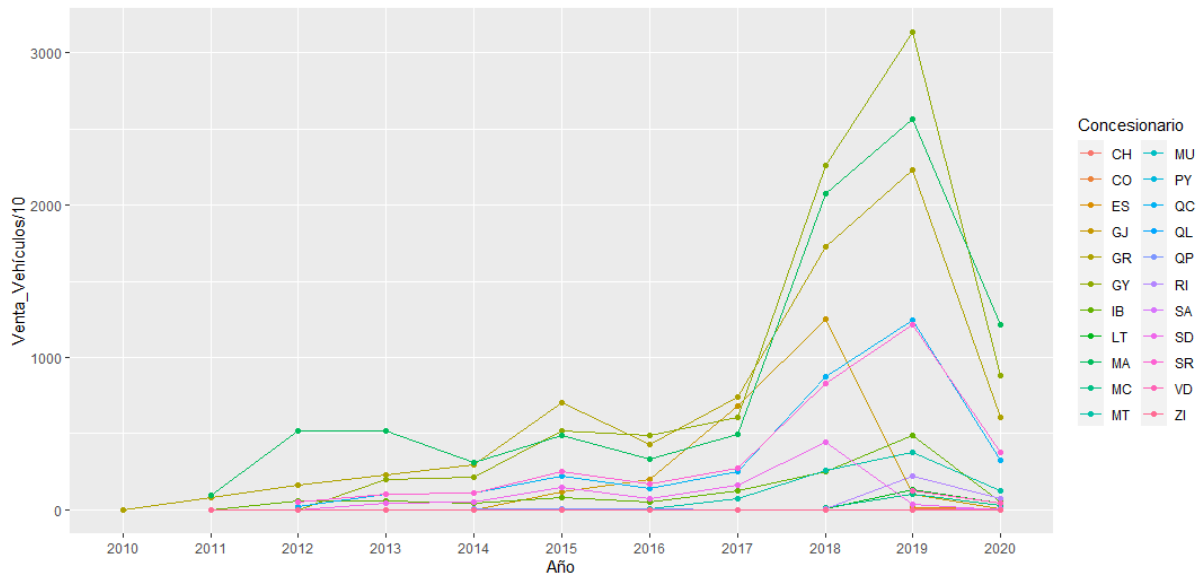


Figura 19: Ventas anuales de vehículos de Great Wall-Haval por agencia.

Podemos observar que claramente la pandemia ha afectado a todos los concesionarios bajando las ventas en un 60 % en los mismos. Además, podemos observar que los puntos de venta más grandes, Guayaquil (Juan Tanga Marengo), Quito (Granados) y Ambato (Matriz) son los mejores concesionarios en términos de venta vehicular.

La venta de vehículos también muestra diferentes patrones de acuerdo con el mes del año, efectos como las promociones de mes de enero de modelos del año pasado o pago de utilidades y decimos también influyen que los meses se comporten diferente. Adicionalmente, la llegada de los nuevos modelos o la versión de este con nomenclatura del siguiente año (generalmente se da a mediados de año) también influye en incrementos de ventas. La siguiente gráfica muestra que el comportamiento de todos los meses no es homogéneo.

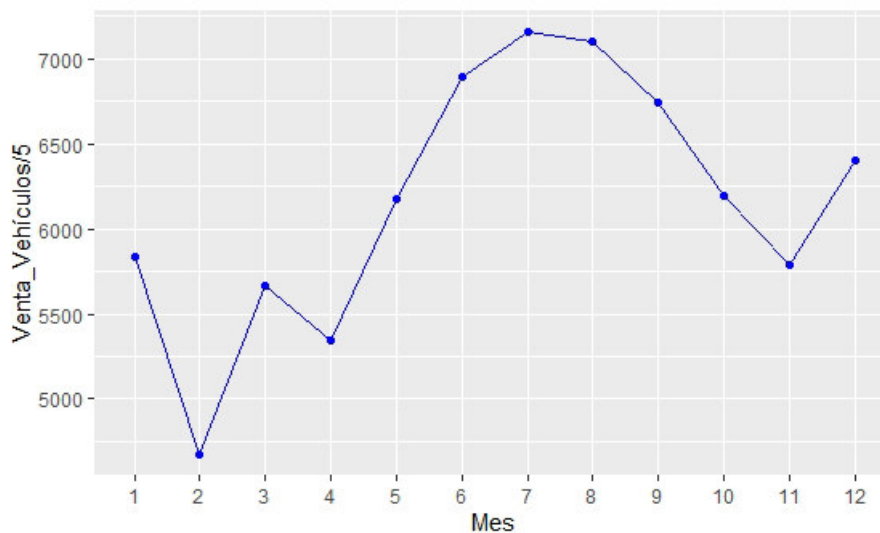


Figura 20: Ventas mensuales de vehículos de Great Wall-Haval.

En la gráfica anterior podemos apreciar que los meses más populares para la compra de vehículos son los meses de junio, julio, agosto y septiembre, mientras que el menos popular es el mes de febrero.

De igual manera, analizaremos la data consolidada mensual para ver si existen comportamientos similares entre los puntos de venta. En la siguiente gráfica podemos analizar los consolidados mensuales de la venta de vehículos.

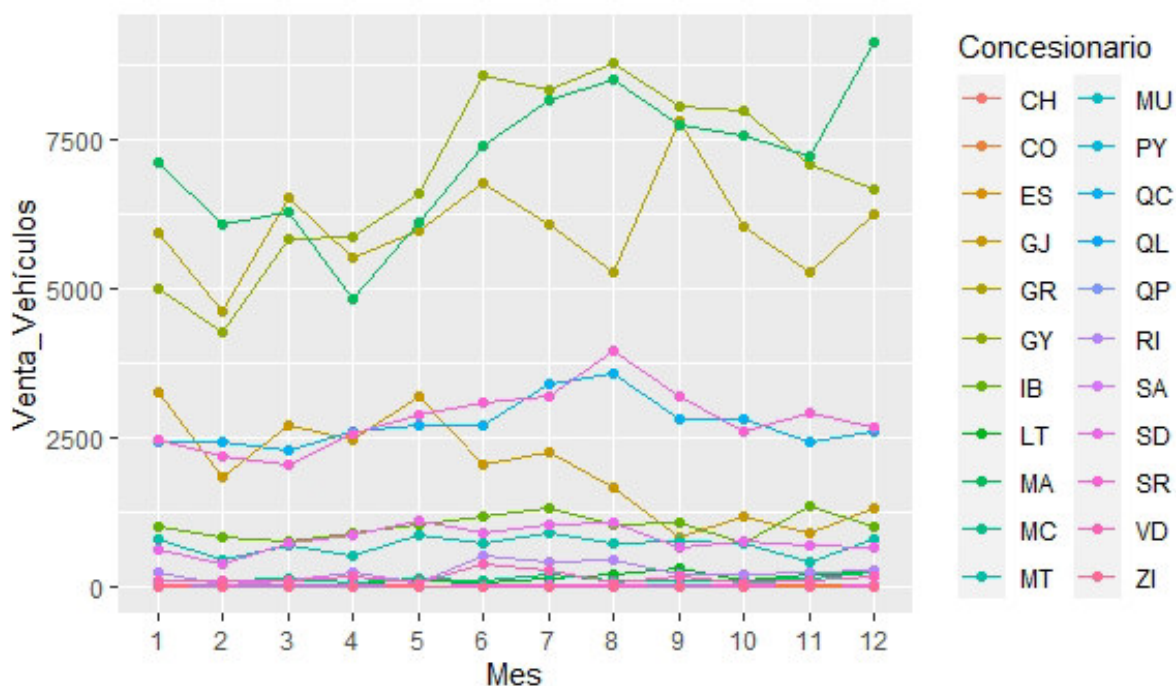


Figura 21: Ventas mensuales de vehículos por agencia de Great Wall-Haval.

En base a los resultados de la gráfica anterior podemos notar que en la mayoría de los concesionarios la tendencia es bastante homogénea, exceptuando los puntos de venta antes mencionados que son los que concentran la mayoría de las ventas a nivel nacional, el resto de los concesionarios tienen comportamientos muy estables a través de los meses del año.

Una vez analizados la venta de los vehículos y mostrado que existen impactos negativos por la pandemia, es pertinente analizar que tanto la misma ha afectado a la venta de repuestos tanto a nivel nacional como en cada uno de los concesionarios. En primer lugar, veamos un análisis histórico de la venta de repuestos.

El siguiente gráfico se muestra un histórico anual de la venta de repuestos ya sean estas por medio de taller o ventanilla.

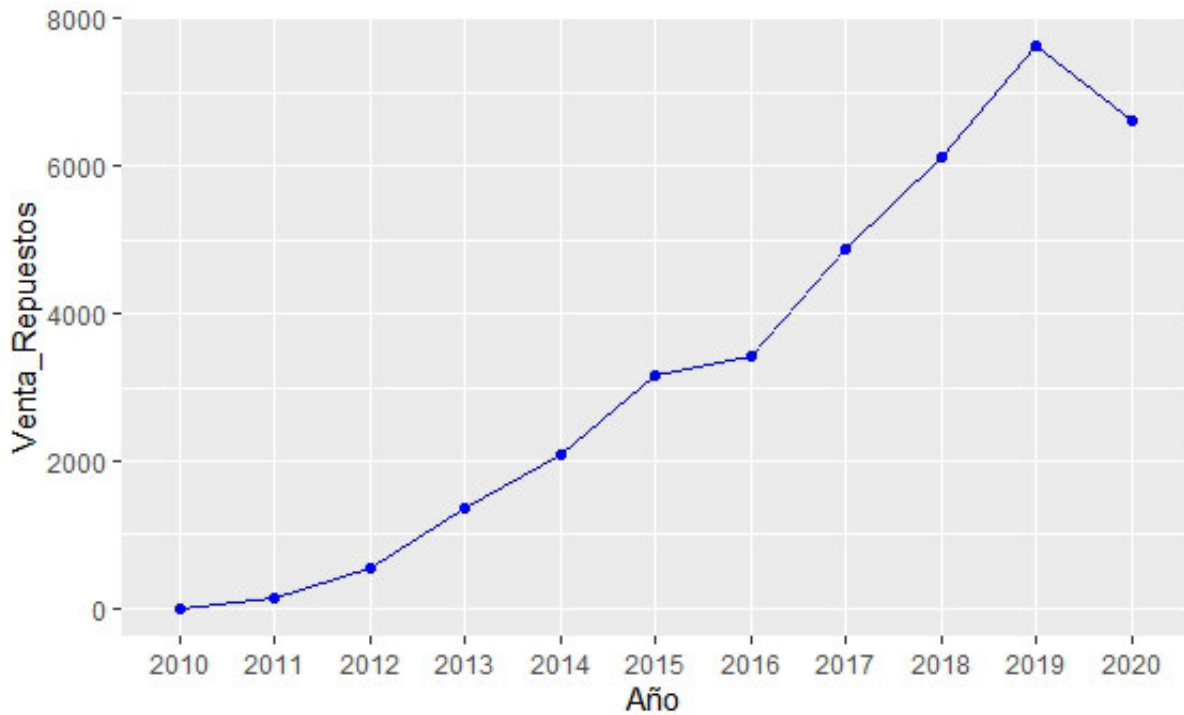


Figura 22: Histórico anual de venta de repuestos a nivel nacional.

Basados en el gráfico anterior, las ventas muestran un comportamiento creciente sostenido hasta el año 2019, durante el año 2020 hay una leve caída debido a efectos de la pandemia. Sin embargo, si consideramos que la venta ha ido creciendo de manera sostenida durante 10 años, el efecto de la pandemia ha afectado en gran medida a esta parte del negocio. Además, consideremos que los ítems que estamos considerando son los que mayor rotación tienen y que actualmente son los más importantes; si se considerase un grupo más amplio de productos los resultados negativos de la pandemia serían más dramáticos y negativos de lo que se puede apreciar con la pequeña muestra de 50 ítems.

Consecuentemente, es necesario analizar si en los ítems que estamos estudiando existen diferencias claras y visuales entre los distintos concesionarios y si se puede apreciar gráficamente si existe un impacto de la pandemia en las distintas ciudades donde están ubicados los concesionarios y talleres de servicio postventa.

La siguiente imagen muestra los consolidados anuales de la venta de repuestos de los 50 mejores ítems de la marca GMW a nivel nacional separado por concesionario.

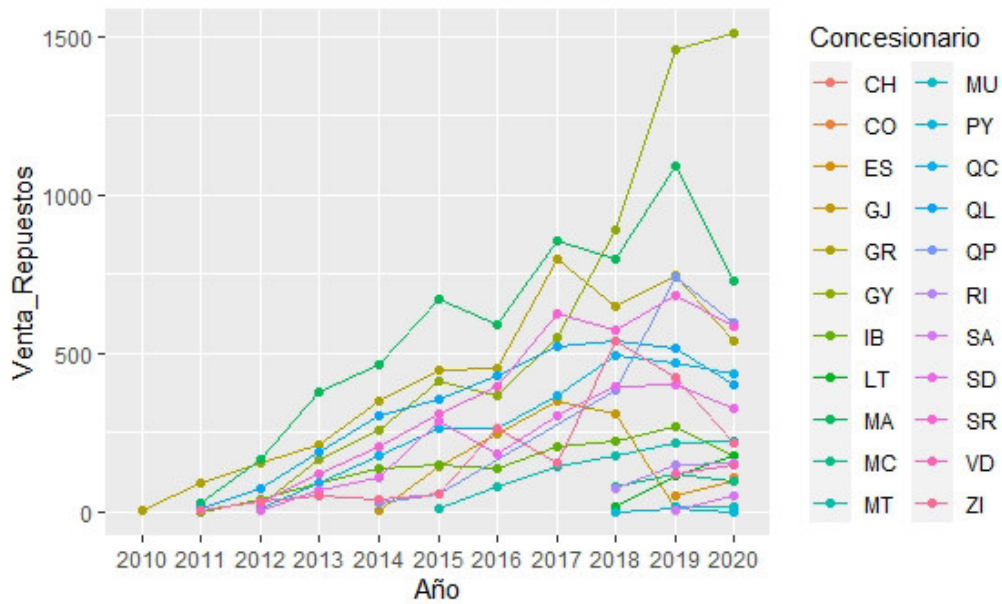


Figura 23: Histórico anual de venta de repuestos a nivel nacional por concesionario.

Basado en la gráfica anterior, 21 de los 22 concesionarios mostraron caídas en sus ventas en el año 2020 con respecto al año anterior, el caso del concesionario GY (Guayaquil) es un caso atípico con respecto al resto de la data. Las altas ventas registradas por este taller ocasionan que el efecto de la pandemia de manera global en la muestra de los 50 ítems no parezca tan abrupto como en realidad lo es.

El siguiente paso es analizar si existen diferencias significativas entre los meses en la venta de repuestos, el siguiente gráfico muestra un consolidado de la venta por meses de los 50 repuestos de la muestra.

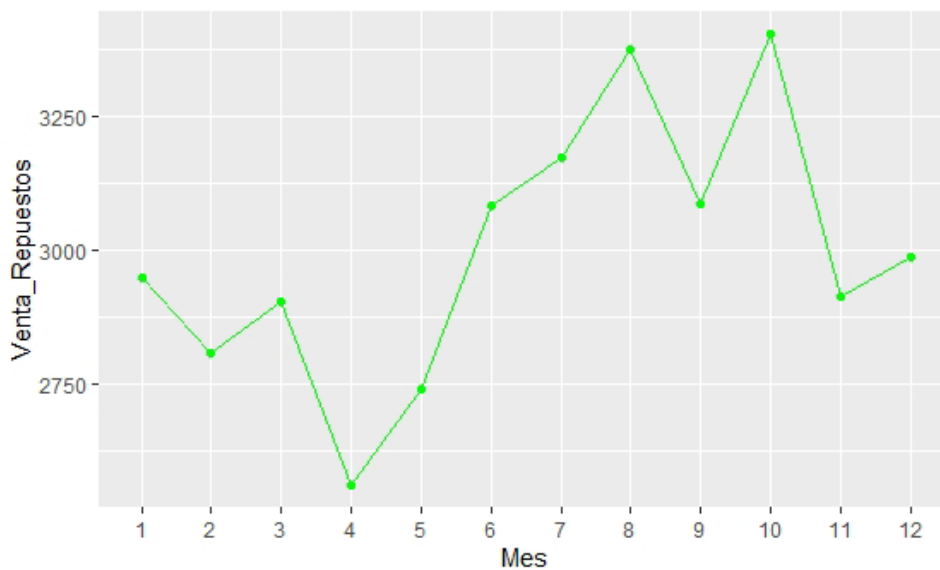


Figura 24: Consolidado mensual de la venta de repuestos.

Es fácil apreciar que existen amplias diferencias en el comportamiento de cada mes, en base a explicaciones brindadas por el área comercial sobre el comportamiento de los clientes se puede evidenciar que estos tienden a asistir a los talleres con mayor frecuencia cuando se acerca la época invernal y en una menor escala al caso anterior en meses donde existen feriados y vacaciones. Estos cambios en el comportamiento de consumo ocasionan que los meses tengan distintos comportamientos entre sí.

Los análisis precedentes nos enseñan que es necesario diferenciar tanto las agencias, los meses y que el uso de variables exógenas nos ayudara a explicar mejor la variabilidad que tiene la venta de repuestos en los distintos concesionarios.

### 8.8.1. Construcción del modelo y resultados

El modelo Random Forest como se ha mencionado en capítulos anteriores tiene la finalidad de ayudar a los planificadores de Ambacar a predecir y anticipar los comportamientos de todos los participantes en el proceso de toma de decisiones y pronóstico de la demanda.

El modelo Random Forest propuesto hará uso de las variables ya antes mencionadas y dos variables extras necesarias para la implementación del modelo basados en la comparación de tanto la venta de vehículos como la venta de repuestos que ayudaran a que los árboles dentro de su estructura puedan tomar mejores decisiones. Ambas medidas son calculadas usando una agrupación de los datos por concesionario y por mes.

En el siguiente recuadro se muestra un pequeño encabezado de como esas variables están conformadas.

Concesionario	Mes	VentaDeRepuestosPromedioXM	VentaDeVehículosPromedioXM
CH	1	1.166	3
CH	2	1	0
CH	3	1.75	0
CH	4	1.66	10
CH	6	1	5
CH	9	2	3
CH	10	4	11
CH	12	1.25	4
CO	5	1	0
CO	6	4	0
CO	7	6	0
CO	8	2	0
CO	9	16.66	0
CO	10	1	1
CO	11	1	0
CO	12	4.75	0

Cuadro 10: Ejemplo de las variables Venta de repuestos promedio y Venta de vehículos promedio.

Cabe recalcar que todo el modelo Random Forest fue programado en lenguaje R. En

la tabla anterior podemos notar cual será la medida usada para que los árboles dentro del modelo puedan realizar una mejor categorización y les sea más fácil para tomar mejores decisiones.

El siguiente paso que se realizó es generar dos subconjuntos de la data; el primer subconjunto compuesto aleatoriamente con el 70 % de la data que lo llamaremos Train Data, el otro subconjunto compuesto del 30 % de la data lo llamaremos Test Data. Para la construcción y entrenamiento del modelo usaremos el subconjunto Train Data. Usaremos para el entrenamiento del modelo todas las variables antes mencionadas; es decir usaremos: “Año”, “Concesionario”, “Mes”, “Código Confidencial”, “Venta Vehículos Agencia x Mes”, “Solo Repuestos?”, “Promoción”, “Covid Cierre”, “Covid No cierre”, “Días promedio visita”, “Km Promedio”, “Venta de Repuestos promedio”, “Venta de Vehículos Promedio”.

Con las variables seleccionadas nos basta seleccionar los hiper parámetros; en base a la teoría sabemos que el parámetro  $mtry$  en cualquier elección que tomemos no afectará de sobremanera al modelo. Sin embargo, en base a recomendaciones que podemos encontrar en la sección de optimización de parámetros tomaremos  $mtry$  como el número de variables del modelo. Es decir,  $mtry = 13$ .

De manera adicional debemos escoger el número de árboles a usar en el modelo, la teoría nos indica que el número a escoger debe ser lo suficientemente grande para bajar la variabilidad del modelo, pero siempre considerando que entre más alto sea el número de este la complejidad computacional también aumentará.

Por tanto, para asegurarnos que la menor variabilidad posible tomaremos  $M = 500$ ; en el siguiente gráfico podemos ver como el modelo con la muestra de entrenamiento tiene un comportamiento estable.

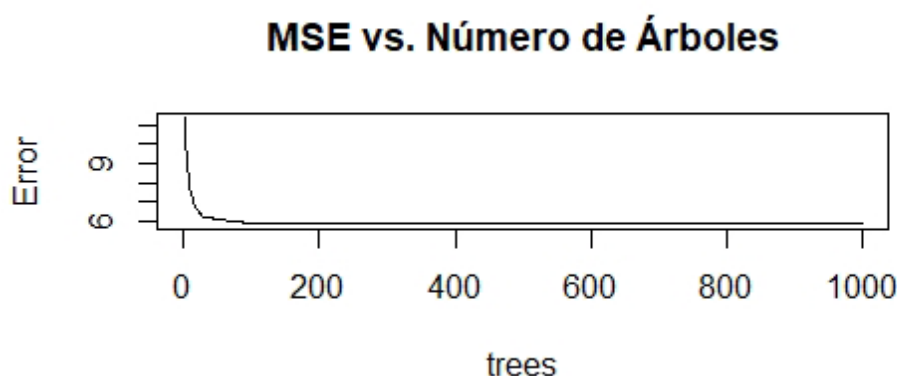


Figura 25: Error en el pronóstico vs número de árboles.

En la gráfica podemos notar que a partir de 100 árboles el error de la muestra se estabiliza. Sin embargo, para asegurarnos que la muestra sea estable sin agregar una complejidad computacional excesiva usaremos  $M = 500$ .



De igual manera, la ventaja del método Random Forest es que nos permite visualizar de manera sencilla la importancia relativa de las variables. La siguiente gráfica muestra que tan importante dentro del conjunto de entrenamiento.

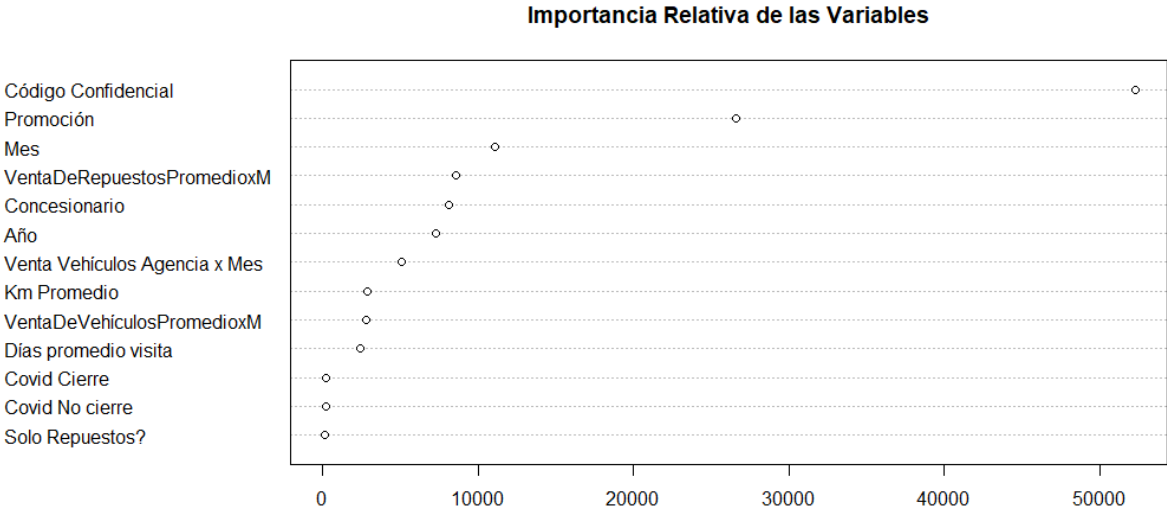


Figura 26: Importancia relativa de las variables dentro del modelo Random Forest.

Podemos ver que hay una interacción en las variables donde lo más importante es la diferencia que hay entre los distintos ítems, marcándose así que cada procedimiento es diferente entre sí y además denotando la existencia de patrones en los datos que marcan comportamientos cíclicos en el uso vehicular.

Por otro lado, notemos que las diferencias que ya habíamos mencionado en el análisis de los datos muestran que si se realiza una diferenciación efectiva entre cada mes y concesionario se obtienen mejores resultados. De igual manera vemos que la venta de vehículos es un factor para tomar en cuenta al momento de pronosticar.

El siguiente paso es analizar la eficacia del modelo; para ello usaremos distintas medidas para analizar los resultados. Primero mostraremos los estadísticos obtenidos por el modelo y la variabilidad explicada.

Tipo de Random Forest	Número de árboles	Número de variables probadas en cada división	MSE	% Var explicada
Regresión	500	13	5.880217	59.83

Cuadro 11: Primeros resultados estadísticos del modelo Random Forest y la variabilidad explicada.

La tabla precedente obtenida corriendo el modelo Random Forest en el programa estadístico R muestra dos estadísticos; el primero, MSE usado para medir el promedio de los errores al cuadrado que en nuestro modelo  $RMS = 5.880217$ . El segundo, el porcentaje de

la variabilidad explicado que nos indica que tanto las variables seleccionadas explican el comportamiento de la venta de repuestos en los distintos concesionarios de la red Ambacar.

Adicionalmente, usaremos la medida RMSE que es una medida que nos mostrará que tan bien se desempeña el modelo, en pocas palabras, el RMSE es la desviación estándar del error. Obtenemos en el modelo  $RMSE = 1.131148$ . De igual manera, podemos medir el error porcentual absoluto medio (MAPE) que nos ayudara a medir el error en términos porcentuales que para la muestra de entrenamiento es  $MAPE = 28.41636\%$ . La siguiente tabla muestra el resumen de los estadísticos arrojados por el programa R.

Conjunto de data	RMSE	MAPE
Conjunto de entrenamiento	1.131148	28.41636 %

Cuadro 12: Resultados del conjunto de entrenamiento, medidas de error MAPE Y RMSE.

Por la gran cantidad de árboles y variables que han sido usados en el modelo, el paquete Random Forest en R nos permite extraer un árbol, la siguiente gráfica muestra un ejemplo de los resultados que arroja el modelo.

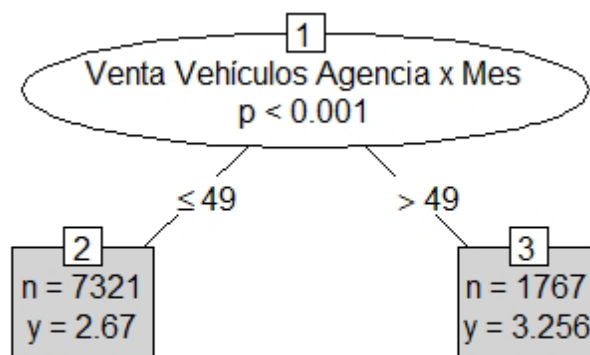


Figura 27: Ejemplo de un árbol en el modelo Random Forest.

En la gráfica podemos ver que, si el número de vehículos vendidos en la agencia es menor a 49 y el código del ítem confidencial es 2, entonces aproximadamente se venderán 3 unidades de ese repuesto; caso contrario si se venden más de 49 unidades y el código del ítem confidencial es 3, entonces aproximadamente se venderan 4 unidades de ese repuesto.

Analizaremos gráficamente los resultados de las predicciones realizadas por el modelo Random Forest y las ventas reales de los repuestos, primero analizaremos por año y concesionarios. La siguiente gráfica muestra la comparación entre las mismas.

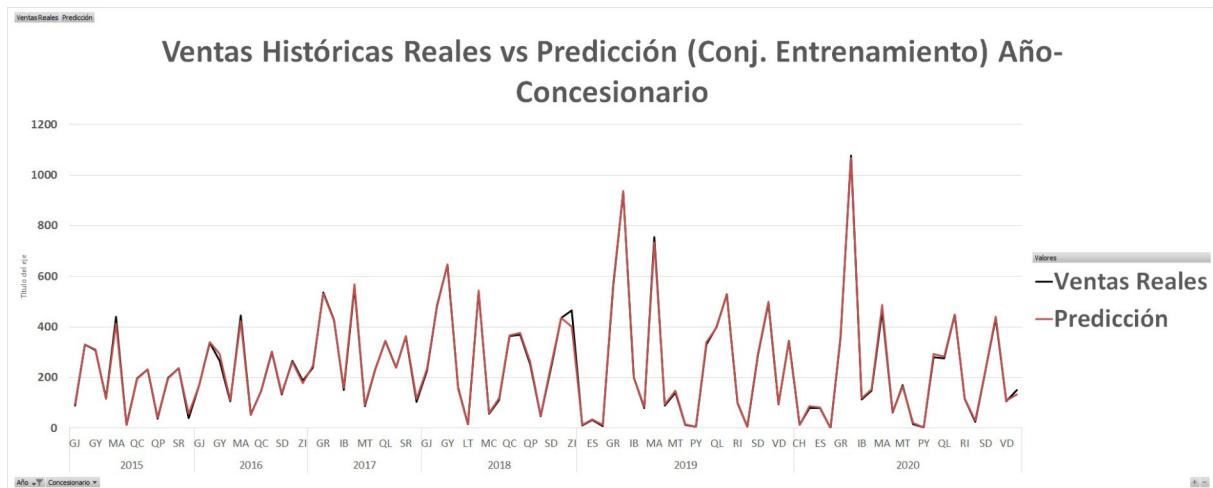


Figura 28: Comparativa conjunto de entrenamiento ventas reales vs predicción, Año-Concesionario.

La comparativa nos muestra que predice bien en términos generales. Sin embargo, falla en el pronóstico de ciertos picos. En general los resultados del modelo se comportan bien en para el consolidado de concesionarios.

La siguiente tabla nos mostrara las diferencias entre lo pronosticado y las ventas reales para los principales concesionarios a nivel nacional.

Concesionario	Año	Ventas Reales	Predicción
GJ	2019	9	12
GJ	2020	1	1
GR	2010	3	45
GR	2011	47	51
GR	2012	108	111
GR	2013	137	144
GR	2018	484	479
GR	2019	566	573
GR	2020	360	367
QC	2012	13	14
QC	2013	53	55
QC	2019	329	340
QC	2020	282	293

Cuadro 13: Comparación de las ventas reales y lo pronóstico para los concesionarios más importantes en el conjunto de entrenamiento.

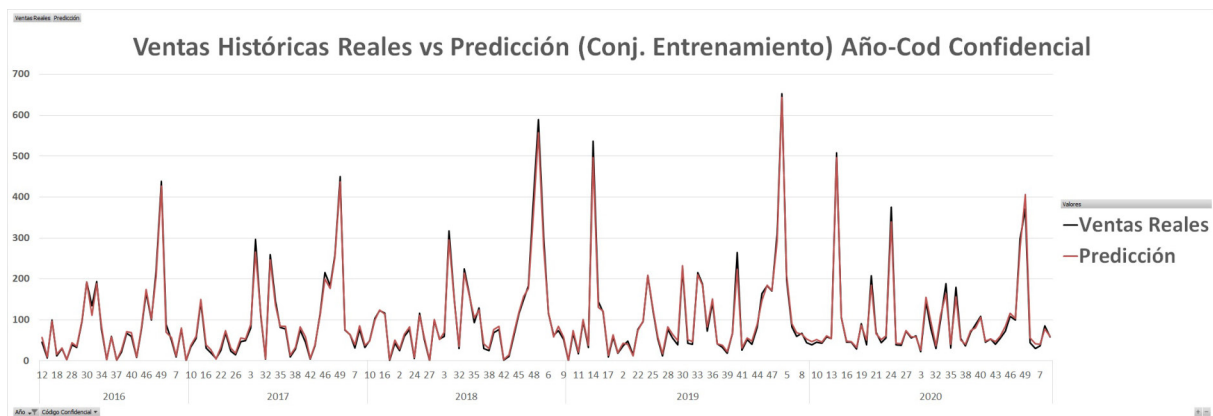


Figura 29: Comparativa conjunto de entrenamiento ventas reales vs predicción, Año-Código.

La gráfica precedente nos nuestro un panorama global del comportamiento del modelo; para comparar de manera más específica nos enfocaremos en las predicciones para el año 2020 de las 12 primeras referencias.

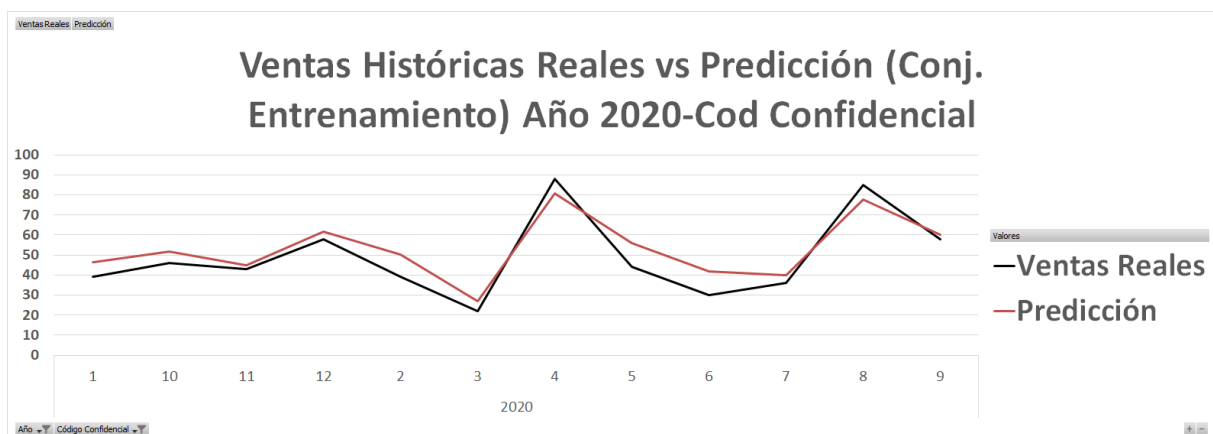


Figura 30: Comparativa conjunto de entrenamiento ventas reales 2020 vs predicción 12 primera referencias, Año-Código.

En base a la gráfica anterior de la comparativa de ventas históricas y predicción de ventas acumulado por código, notamos que tiende a sobre estimar los picos y subestimar los valles. Pero el comportamiento en general del modelo se aproxima bastante bien a lo que en realidad sucede en la vida real.

Después de comprobar la eficacia del modelo en el conjunto de entrenamiento debemos probar los datos con el conjunto de prueba, usando el mismo modelo Random Forest obtenemos los estadísticos de error de pronóstico RMSE y MAPE. La siguiente tabla nos mostrará los resultados obtenidos por medio del programa estadístico R.

Conjunto de data	RMSE	MAPE
Conjunto de prueba	1.208569	28.77512%

Cuadro 14: Resultados del conjunto prueba, medidas de error MAPE Y RMSE.

En la tabla anterior podemos notar que las diferencias obtenidas en las medidas de error en ambos conjuntos son ínfimas, lo que nos muestra que el modelo no está sobre especificado. Para el conjunto de prueba obtenemos  $MAPE = 28.77512\%$  y  $RMSE = 1.208569$ .

Mostremos gráficamente la comparación entre las ventas históricas de repuestos y las predicciones del modelo Random Forest que obtuvimos. En primer lugar, analizaremos consolidados de ventas anuales y concesionario. La línea color naranja representa las ventas históricas y la línea celeste la predicción del modelo.

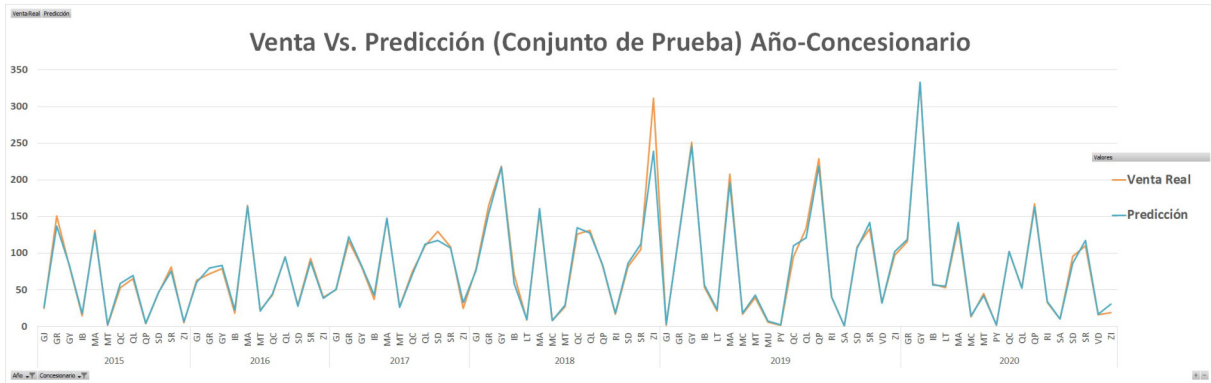


Figura 31: Comparativa conjunto de prueba ventas reales vs predicción, Año-Concesionario.

Basado en la figura anterior notemos que el modelo Random Forest que entrenamos con el conjunto de entrenamiento obtiene una predicción aceptable tanto en los estadísticos de error de medida como gráficamente, existen ciertos picos en los cuales el modelo tiende a sobrestimar los resultados. A continuación, mostraremos una tabla que muestra los resultados obtenidos en los principales concesionarios.

Concesionario	Año	Venta Real	Predicción
GJ	2015	25	26
GJ	2016	63	60
GJ	2017	50	51
GJ	2018	78	76
GJ	2019	2	3
GR	2016	72	79
GR	2017	117	122
GR	2018	165	154
GR	2019	126	124
GR	2020	116	118
QC	2015	53	58
QC	2016	43	44
QC	2017	75	71
QC	2018	126	135
QC	2019	95	110
QC	2020	102	102
QP	2014	4	51
QP	2015	4	4
QP	2018	83	84
QP	2019	229	218
QP	2020	167	162

Cuadro 15: Comparación de las ventas reales y lo pronóstico para los concesionarios más importantes en el conjunto de prueba.

De igual manera, podemos analizar de manera gráfica las diferencias entre el consolidado anual y código confidencial.

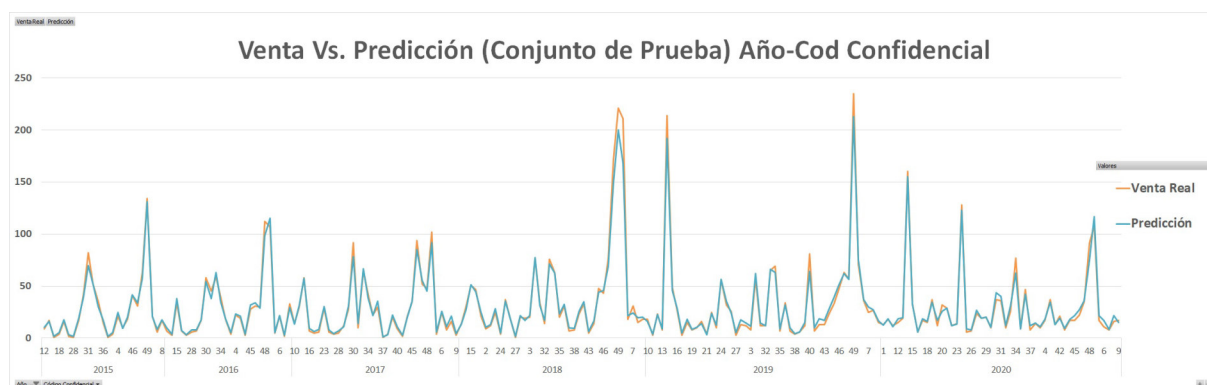


Figura 32: Comparativa conjunto de prueba ventas reales vs predicción, Año-Código.

De manera similar al conjunto de entrenamiento, nos enfocaremos para analizar los resultados del año 2020 de las 12 primeras referencias.

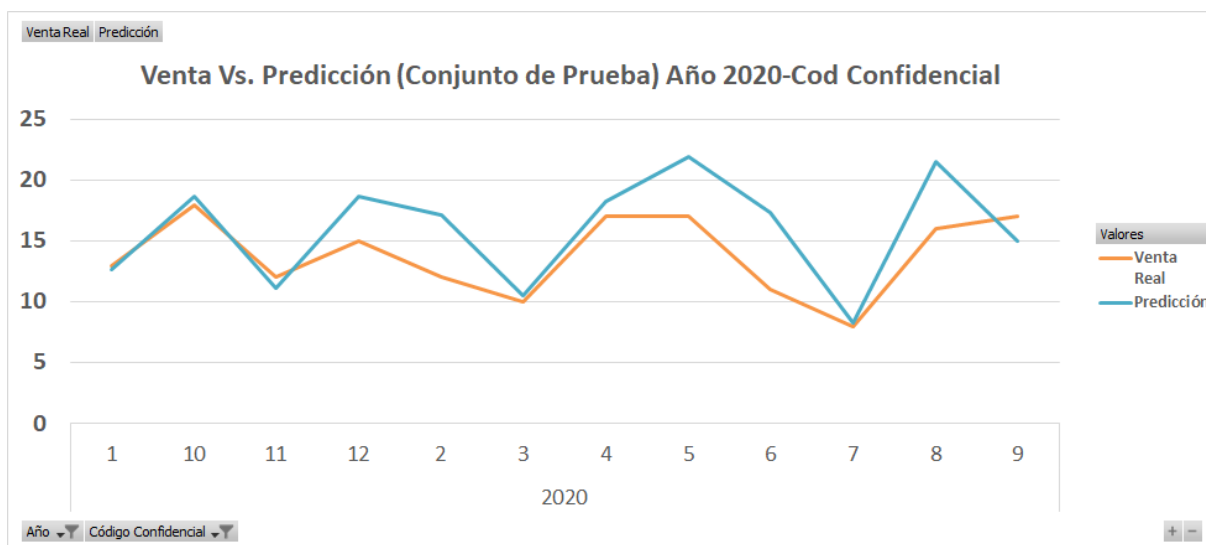


Figura 33: Comparativa conjunto de prueba ventas reales 2020 vs predicción 12 primera referencias, Año-Código.

Al igual que en el caso del consolidado anual de concesionarios el modelo Random Forest entrenado tiende a sobrestimar los picos en la serie de datos. Sin embargo; el modelo tiene un comportamiento aceptable tanto en los estadísticos de error como gráficamente.

### 8.8.2. Comparación modelo Random Forest vs ventas reales vs Forecast Pro

En la actualidad Ambacar cuenta con un software comercial para el pronóstico de la demanda (Forecast Pro). Para Ecuador Forecast Pro es distribuido por la empresa NET-CONTROL; la cual amablemente nos proveyó una licencia temporal para la realización de este trabajo.

Para el uso del software de pronóstico usaremos data desde enero del 2016 hasta diciembre del 2020. Realizaremos el pronóstico para código en cada concesionario, en total obtenemos 808 series temporales a pronosticar en donde el software encontrara el mejor método (menor error en el pronóstico). Adicionalmente nos proporcionara con medidas de error para control de los resultados.

La interfaz de Forecast Pro nos permite tener un control visual y estadísticos de los resultados arrojados por los diversos modelos e inclusive tener a mayor detalle por concesionario e ítem. Mostraremos un detalle del consolidado de datos tanto por concesionario y repuesto que nos arroja el software.

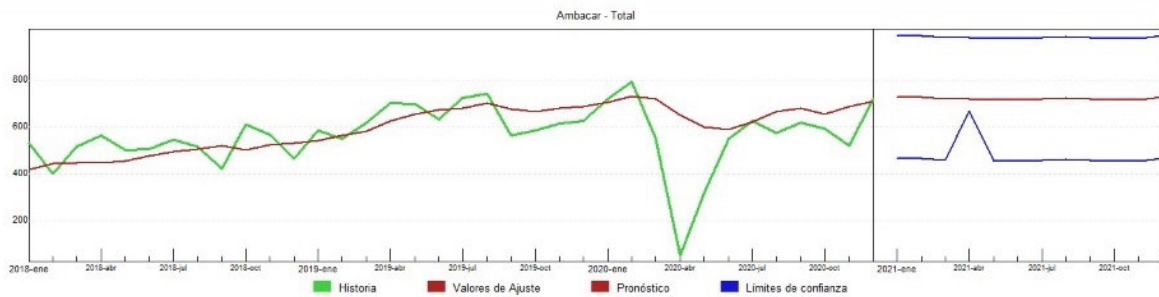


Figura 34: Comparativa resultados Forecast Pro vs Real.

Reporte de Pronóstico			
jul - sep	72,03	21,05	65,04
oct - dic	41,04	-33,95	168,0
Estadísticas de la muestra			
Tamaño muestra	60	No. parámetros	1
Media	477,83	Desv. estándar	161,04
R-Cuadrada	0,07	R-Cuadrada Aj.	0,07
Durbin-Watson	0,78	Ljung-Box(18)	71,5 P=1,00
Error de pronóstico	155,49	BIC	159,54
MAPE	40,68%	SMAPE	24,23%
RMSE	154,19	MAD	109,74
Relación MAD/Media	0,23		

Figura 35: Estadísticos obtenidos del software Forecast Pro.

En la gráfica anterior podemos ver representado en color verde la gráfica histórica de datos y en la línea roja el pronóstico del modelo. Además, en la parte inferior se muestran información sobre los estadísticos de los modelos que serán resumidos en la siguiente tabla.

R-Cuadrado	RMSE	MAPE	SMAPE
7 %	154.19	40.68 %	24.23 %

Cuadro 16: Estadísticos de los modelos usando el software Forecast Pro.

Por medio de los estadísticos que nos arrojaron tanto el modelo Random Forest como los modelos generados por Forecast Pro podemos realizar la primera comparación de eficacia de los modelos. La siguiente gráfica muestra un consolidado de los estadísticos obtenidos.

Modelo	R-Cuadrado	RMSE	MAPE
Forecast Pro	7 %	154.19	40.68 %
Random Forest	59.83 %	1.298	28.78 %

Cuadro 17: Comparativa de los modelos Random Forest vs Forecast Pro.

De igual manera podemos realizar una comparación gráfica de como los modelos se comparan respecto a las ventas históricas, usaremos los meses de octubre, noviembre y



diciembre para una comparativa. Primero analizaremos los resultados obtenidos para los 5 concesionarios más importantes.

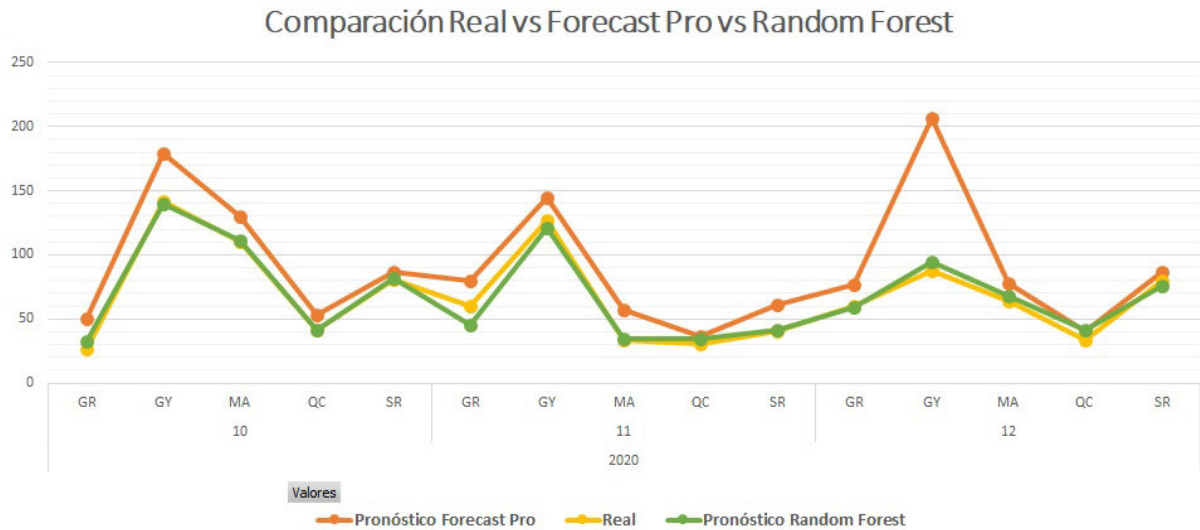


Figura 36: Comparación de los modelos para los 5 concesionarios más importantes.

Basado en los resultados obtenidos en la gráfica anterior, podemos deducir que tanto el modelo de Forecast Pro y Random Forest predicen de manera acertada la tendencia y nivel. Sin embargo, los resultados proporcionados por Forecast Pro muestran una sobrestimación respecto a la venta real. Una primera explicación de este efecto es que el software no está considerando el efecto de la pandemia para estos meses y considera de gran manera las ventas del año pasado de esos meses.

La siguiente comparación que haremos es para los 10 repuestos más populares a la hora del hacer el análisis, es decir, los ítems con código confidencial del 1 al 10. El análisis se realiza considerando todos los concesionarios.

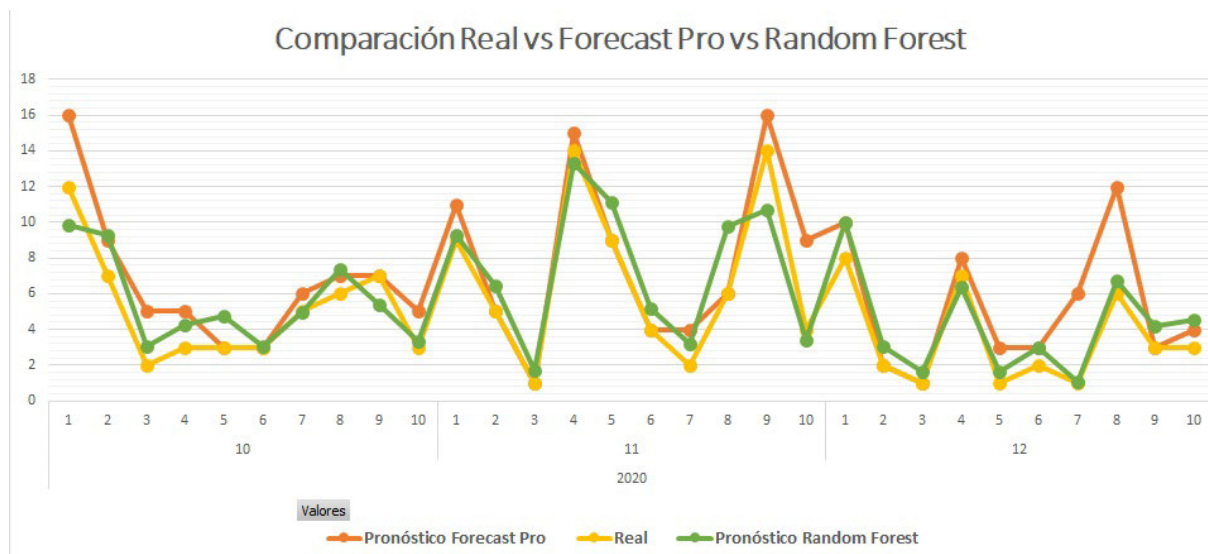


Figura 37: Comparación de los modelos para los 10 repuestos más importantes.

Podemos observar un patrón entre los modelos un poco más caótico. No obstante, los resultados obtenidos con el modelo Random Forest muestran un comportamiento más acercado a la realidad, mientras que los resultados obtenidos con el software de pronóstico tienden a sobrestimar la venta en una gran mayoría de casos.

En base a los resultados obtenidos tanto en las medidas de error como en los gráficos presentados se puede concluir que el modelo Random Forest tiene un comportamiento aceptable al momento de pronosticar las ventas de los 50 repuestos más importantes de la marca más vendida.

Finalmente, la siguiente tabla muestra un comparativo entre los errores y estadísticos obtenidos de los modelos comparados. Para el análisis usaremos tanto los estadísticos arrojados por el modelo Random Forest con el conjunto de entrenamiento como con el conjunto de prueba.

Modelo	R-Cuadrado	RMSE	MAPE	Error de Pronóstico
Forecast Pro	7%	154.19	40.68%	155.49
Random Forest (Conj. Prueba)	59.83%	1.298	28.78%	31.84
Random Forest (Conj. Entren)	59.83%	1.131	28.41%	212.21

Cuadro 18: Comparativa de los errores de los modelos Random Forest vs Forecast Pro.

En base a los resultados de la tabla anterior, podemos apreciar que los resultados obtenidos con el modelo Random Forest explican de mejor manera la variabilidad del pronóstico de la demanda. Además obtienen errores de pronóstico muy aceptables considerando la gran cantidad de datos que fueron usados para el entrenamiento de los modelos a comparación que el software Forecast Pro solo uso los últimos 3 años de datos.

A continuación, usaremos los resultados de los pronósticos obtenidos con el modelo Random Forest y los históricos de ventas a nivel ítem concesionario para la optimización del inventario y establecer tanto mínimos e inventario recomendable a tener en los distintos puntos de venta en la red de concesionarios de Ambacar.

### 8.9. Optimización del inventario aplicado a Ambacar

Con los datos presentados en las secciones anteriores y juntamente con los datos de inventario que actualmente posee la compañía en cada bodega de los concesionarios; procederemos a realizar la optimización de cada uno de los concesionarios con atención post-venta y el centro de distribución Ciauto ubicado en la provincia de Tungurahua.

Para modelar el sistema logístico y de distribución procedemos, con la ayuda de un grafo a representar en cada uno de los nodos los puntos de servicio postventa y el centro de distribución. Donde el primer nodo representará el centro de distribución y el resto de los nodos los concesionarios que poseen servicio post-venta. Con la información proporcionada obtenemos el siguiente grafo que demuestra las relaciones dentro del sistema.

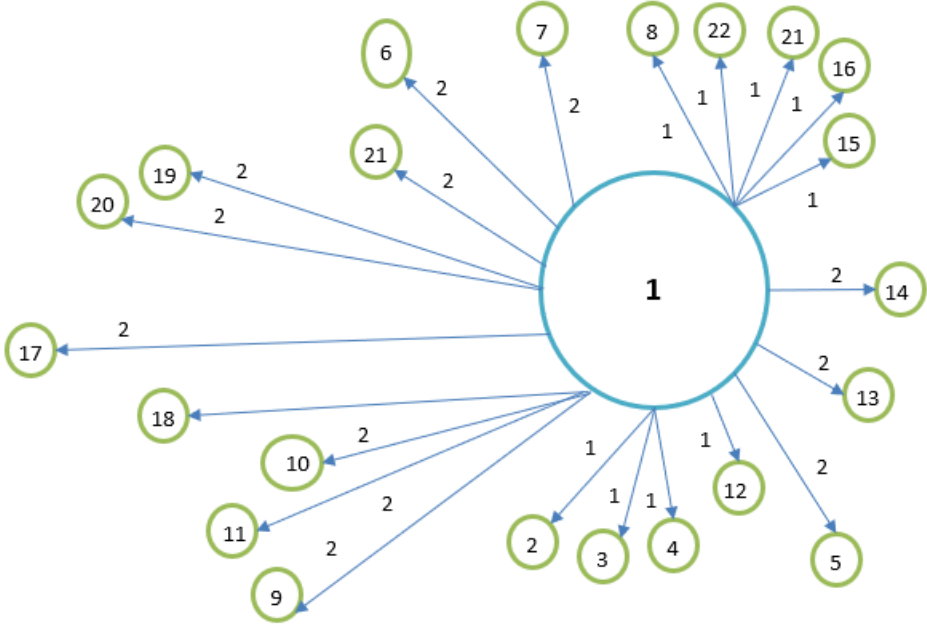


Figura 38: Grafo que muestra el sistema de distribución de Ambacar.

Como podemos ver en el sistema, el centro de distribución es el principal eje en donde todo el flujo de ítems se distribuye, el número en las aristas representa el número de días que se demora el proceso logístico de mover el inventario entre el centro de distribución y las agencias con servicio postventa. Cabe recalcar que en casos extremos la comunicación entre puntos de servicio postventa es posible más uno de los objetivos de la tesis es reducir lo más posible esas interacciones ya que los costos asociados a los mismos son altos; por tanto, se ignora la posibilidad de que haya interacciones entre los mismos.

Como puntos iniciales necesitamos los tiempos de espera a cada uno de los concesionarios; por indicaciones de Christian Diaz (Gerente Nacional de Repuestos), los ítems que se destinan tanto para las provincias de Quito, Tungurahua, Cotopaxi tienen un día de espera desde que se genera la orden de abastecimiento y el resto de las provincias donde está presente Ambacar toma en general dos días. En la siguiente tabla se resumen los días de espera por provincia.

Provincia	Días de Espera
Pichincha	1
Tungurahua	1
Cotopaxi	1
Santo Domingo	2
Manabí	2
Santa Elena	2
Guayas	2
Orellana	2
Los Ríos	2
Chimborazo	2
El Oro	2
Orellana	2
Imbabura	2

Cuadro 19: Días de espera en los envíos del centro de distribución a cada provincia.

Adicionalmente, es necesario poseer la información sobre la cantidad de inventario que se mantiene en cada agencia y por cada uno de los ítems que entran dentro del análisis. A continuación, mostraremos el inventario disponible por cada ítem y concesionario.

Centro	A001	A002	A003	A004	A005	A006	A007	A008	A009
CH	0	0	0	0	0	0	0	0	0
CO	2	0	0	1	0	0	2	2	2
ES	0	2	0	1	0	1	2	2	0
GJ	0	0	0	1	0	0	2	0	0
GR	0	2	0	1	0	2	1	1	0
GY	1	2	0	1	1	2	7	1	2
IB	0	2	0	4	0	2	1	0	1
LT	0	1	0	2	0	1	4	1	3
MA	0	2	1	2	2	4	2	3	0
MC	1	0	2	5	1	0	2	4	0
MT	0	1	0	2	1	1	7	2	2
MU	0	0	0	0	0	2	0	0	0
PY	0	0	0	0	0	1	0	0	0
QC	1	3	0	2	1	3	2	1	0
QL	1	2	0	0	1	1	2	1	0
QP	0	2	1	2	26	15	3	11	0
RI	0	1	1	2	1	7	6	3	0
SA	0	0	0	0	0	0	0	0	0
SD	0	1	1	4	0	3	0	5	0
SR	0	1	0	1	0	1	1	3	0
VD	0	3	1	4	1	1	9	2	2
ZI	5	24	260	547	260	129	92	402	201

Cuadro 20: Ítems disponibles por código A001-A009 en los puntos de atención postventa.

Centro	A010	A011	A012	A013	A014	A015	A016	A017	A018
CH	0	0	0	0	0	0	0	0	0
CO	2	1	0	0	0	2	1	1	0
ES	0	0	5	0	4	0	1	0	0
GJ	1	3	0	0	0	0	0	0	0
GR	0	1	1	0	31	2	0	0	0
GY	1	7	1	4	8	0	0	7	1
IB	1	1	2	0	12	2	1	1	0
LT	3	1	3	2	8	6	1	1	1
MA	0	4	4	3	20	11	1	3	0
MC	1	2	0	2	8	7	2	0	0
MT	1	2	3	2	20	5	2	2	1
MU	0	0	0	0	0	0	1	0	0
PY	0	0	0	0	0	0	0	0	0
QC	2	3	2	1	20	2	1	2	0
QL	1	1	1	1	24	2	0	0	0
QP	1	4	2	0	20	7	0	5	2
RI	1	2	1	1	12	5	1	2	0
SA	0	3	0	2	0	3	0	0	0
SD	5	4	1	0	20	7	4	5	1
SR	0	0	3	0	0	7	1	0	0
VD	1	2	2	0	0	4	5	2	1
ZI	189	26	136	16	120	157	63	6	58

Cuadro 21: Ítems disponibles por código A010-A018 en los puntos de atención postventa.

Centro	A019	A020	A021	A022	A023	A024	A025	A026	A027
CH	0	0	0	0	0	0	0	0	0
CO	0	0	0	1	2	0	1	0	1
ES	2	0	0	2	2	0	1	0	0
GJ	0	0	0	1	1	0	0	0	0
GR	2	0	3	0	3	20	0	0	0
GY	6	16	10	1	6	16	0	2	3
IB	0	0	0	2	1	4	1	0	2
LT	2	0	0	1	2	4	2	1	2
MA	2	8	8	0	2	0	1	1	6
MC	0	0	0	1	1	0	2	0	3
MT	0	0	4	2	0	0	1	0	1
MU	0	0	0	0	0	8	1	0	0
PY	0	0	0	0	0	0	0	0	1
QC	2	0	15	2	2	12	2	0	0
QL	0	0	0	2	5	0	0	0	0
QP	0	16	4	2	6	0	0	1	3
RI	0	0	4	1	1	4	3	0	1
SA	0	0	0	0	0	0	0	0	0
SD	0	0	12	3	3	12	2	1	2
SR	4	0	12	1	4	22	4	0	0
VD	0	0	0	4	2	0	6	0	0
ZI	13	52	126	87	80	0	12	80	0

Cuadro 22: Ítems disponibles por código A019-A027 en los puntos de atención postventa.

Centro	A028	A029	A030	A031	A032	A033	A034	A035	A036
CH	0	0	0	0	0	0	0	0	0
CO	0	0	0	0	2	0	0	0	0
ES	0	4	0	0	2	4	0	0	0
GJ	0	0	0	0	1	0	3	2	2
GR	4	1	5	0	2	3	6	2	6
GY	1	1	7	0	2	0	12	1	12
IB	1	2	10	0	1	1	1	1	2
LT	4	3	6	1	1	3	8	0	8
MA	4	5	4	3	2	2	1	0	2
MC	1	0	2	1	0	1	2	0	4
MT	1	0	3	3	1	2	2	2	2
MU	0	0	3	0	0	0	0	0	0
PY	0	0	0	0	0	1	0	0	0
QC	1	4	3	0	3	9	0	2	2
QL	1	0	4	0	2	0	1	1	1
QP	5	3	29	0	11	3	0	1	5
RI	1	0	3	0	2	4	2	2	2
SA	0	0	0	0	1	4	2	0	2
SD	1	4	1	0	1	1	0	0	2
SR	0	1	2	0	1	2	2	1	0
VD	0	3	0	1	1	2	10	0	11
ZI	98	40	132	0	32	172	62	63	115

Cuadro 23: Ítems disponibles por código A028-A036 en los puntos de atención postventa.

Centro	A037	A038	A039	A040	A041	A042	A043	A044	A045
CH	0	0	0	0	0	0	0	0	0
CO	0	0	1	0	0	0	0	0	2
ES	0	2	3	9	0	1	0	2	5
GJ	8	2	0	4	0	0	0	1	0
GR	0	4	2	6	0	1	1	0	1
GY	2	2	18	0	2	3	3	1	2
IB	0	1	1	0	0	0	0	0	4
LT	1	5	1	0	2	1	1	2	2
MA	2	0	1	7	3	0	1	0	1
MC	2	1	2	0	1	1	0	0	1
MT	5	1	2	0	2	1	0	0	3
MU	0	0	0	0	0	0	0	0	0
PY	0	0	2	0	0	0	0	0	0
QC	1	1	2	0	1	1	1	0	10
QL	2	1	5	5	1	2	2	0	3
QP	1	6	5	0	9	1	0	1	5
RI	2	1	1	3	1	0	1	2	2
SA	4	0	2	0	2	0	0	0	3
SD	4	0	2	0	0	2	1	3	10
SR	4	1	2	19	0	1	1	0	10
VD	2	0	3	0	0	0	1	2	5
ZI	121	11	0	82	22	0	11	87	211

Cuadro 24: Ítems disponibles por código A037-A045 en los puntos de atención postventa.

Centro	A046	A047	A048	A049	A050
CH	0	0	0	0	0
CO	1	2	14	3	2
ES	2	2	0	0	4
GJ	3	4	0	0	0
GR	2	2	0	9	6
GY	6	5	13	11	11
IB	13	2	11	6	4
LT	4	2	13	16	6
MA	3	1	26	15	6
MC	2	3	0	12	1
MT	5	4	5	10	2
MU	0	2	0	3	5
PY	0	0	0	0	0
QC	2	2	13	8	5
QL	0	0	0	19	3
QP	2	4	10	10	4
RI	2	3	5	10	3
SA	2	2	0	0	1
SD	5	4	39	6	6
SR	2	3	26	39	8
VD	3	3	13	12	4
ZI	0	26	47	96	10

Cuadro 25: Ítems disponibles por código A046-A050 en los puntos de atención postventa.

Las tablas precedentes muestran la cantidad de ítems disponibles por punto de atención postventa y el centro de distribución expresado como ZI (CIAUTO).

## 8.10. Resultados

La implementación del sistema Multi-Echelon fue realizada en Python con programación en la nube por medio del programa Colab. Como ya ha sido mencionado en las secciones anteriores, el programa se alimenta tanto de la información de tiempos de espera y la información de la cantidad de ítems disponibles en cada una de las bodegas.

En base a la información de las tablas anteriores sobre la cantidad actual de inventario que tenemos en cada uno de las agencias de cada uno de los 50 ítems más importante. Lo primero que fijamos por políticas de abastecimiento es un nivel de servicio deseado; se ha llegado a un acuerdo que la satisfacción del cliente o nivel de servicio debería ser de al menos el 95 % en cada uno de las agencias y sobre todo en las más importantes ubicadas en las ciudades de Quito, Guayaquil y Ambato.

Como ayuda para obtener mejores resultados supondremos que el centro de distribución tiene inventario suficiente para cubrir cualquier necesidad de las agencias. A pesar, de que en la práctica no es una condición que deba considerarse ya que el inventario disponible en el centro de distribución es uno de los principales factores que se deben considerar para priorizar y enviar inventario a los puntos de venta. Sin embargo, en el proceso que realizamos de búsqueda de políticas de reaprovisionamiento e inventario medio necesario podemos adoptar la suposición del inventario suficiente en el centro de distribución. Por último, lo que haremos referente al inventario en el mismo es determinar qué cantidad de ítems debemos tener para ante cualquier eventualidad mantener a las agencias.

El siguiente paso que consideramos consiste en simular el sistema logístico, para eso lo primero que tenemos que hacer es expresar el grafo (presentado en la sección 8.9) de manera matricial. Es decir, obtendremos la matriz de adyacencia que nos permitirá encontrar dentro del modelo logístico las relaciones entre los nodos. La siguiente parte será usar los resultados obtenidos en el modelo Random Forest juntamente con los datos históricos de ventas y en base a esos datos y resultados usarlos en la simulación.

Se procede a crear el modelo de simulación, en primer se deben simular las 4 fases que se han considerado para la forma de realizar y abastecer pedidos. Estas son: Órdenes de reaprovisionamiento, Cumplimiento de las órdenes de reaprovisionamiento, Enviar la orden de reaprovisionamiento, Atender la demanda del cliente.

Con las funciones determinadas y explicadas en secciones anteriores, el siguiente paso consiste en simular el sistema, esto se logra realizando el proceso de simulación en cada una de las agencias de servicio postventa generando órdenes de reaprovisionamiento con el fin que el centro de distribución recepte, genere y cumpla esa orden y para posteriormente atender la demanda del cliente. En cada uno de esos pasos actualizando el inventario medio disponible y el nivel de servicio en cada una de las agencias.

Posteriormente, todo lo explicado en párrafos anteriores se tiene que consolidar en el módulo de optimización llamado *OptimizarSistema(.)*. Este módulo se encarga de simular el sistema y funciones expresados anteriormente y recalcular el inventario medio en cada uno de los pasos y el nivel de servicio. Para así obtener la siguiente función objetivo.

$$Z = \sum_{i=1}^N \sum_{f \in F} A_f + pe \sum_{i=1}^N \sum_{f \in F} \max(0, B_f^T - B_f)$$

Donde  $A_f$  es el inventario promedio en cada una de las agencias,  $pe$  es el factor de penalización,  $B_f^T$  el nivel de servicio deseado en cada una de las agencias y  $B_f$  es el nivel de servicio en cada una de las agencias.

Finalmente se buscará optimizar la función descrita anteriormente.

A continuación, haremos un breve detalle de la situación actual de Ambacar Ecuador en referencia a la cantidad de ítems que poseen en bodegas y en el centro de distribución. Al día 31 de diciembre del 2020 existen en inventario disponible 6885 unidades. Cabe



recaltar que las 6885 unidades que se encuentran disponibles son de los 50 ítems más importantes de acuerdo con el análisis del Pareto.

En cada uno de los 22 centros de servicio postventa la distribución de los ítems no es uniforme entre ellos; es decir, existen distintos factores que ocasionan que la distribución de repuestos sea desproporcional. Como una de las principales razones podemos citar es la centralización de venta de vehículos nuevos en las ciudades de Quito, Guayaquil y Ambato.

De manera general vamos a mostrar la distribución de repuestos dentro del sistema.

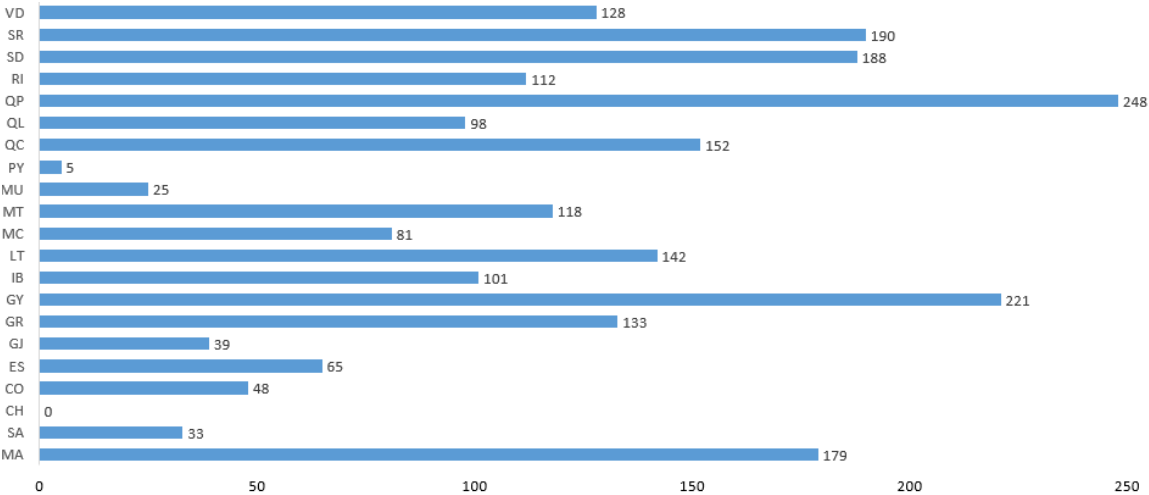


Figura 39: Cantidad de ítems disponibles en cada uno de los puntos de atención postventa.

Es fácil notar que las grandes ciudades son las que mayor número de ítems concentran y además la que mayor número de concesionarios con servicio postventa obtienen; por tanto, los resultados que se obtengan en los centros antes mencionados serán revisados con más atención y minuciosidad.

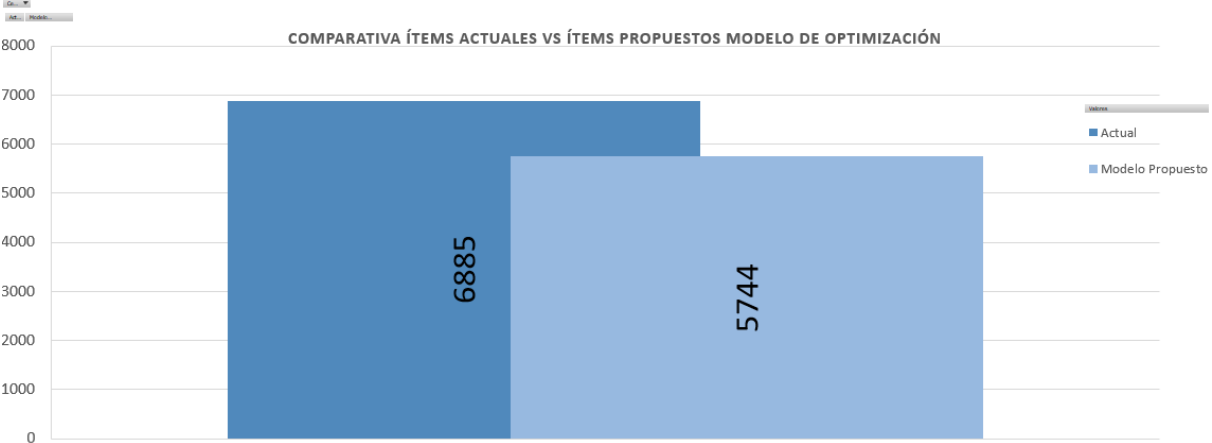


Figura 40: Cantidad de ítems Actuales vs Propuestos

Con el modelo propuesto podemos lograr una reducción de 1141 ítems; es decir, una mejora del 16.57%. La mayoría de la mejora se obtiene bajando los niveles de inventario en el centro de distribución principal (Ciauto) y aumentando ligeramente el nivel de inventario en algunas de las agencias. A continuación, mostraremos el impacto que tendría el modelo propuesto en el centro de distribución principal y en los concesionarios con servicio postventa ubicados en las ciudades más importantes del país y de otros concesionarios grandes. El resto de las gráficas sobre el impacto del modelo propuesto se pueden encontrar en la sección de anexos.

En la siguiente gráfica mostraremos el impacto dentro del centro de distribución principal. Como mencionamos en el párrafo anterior el mayor impacto que se genera en el modelo es en el mismo, la razón por la cual el número de ítems es alto es porque en muchas ocasiones el nivel de inventario para repuestos es también compartido con el inventario de ensamblaje de autos.

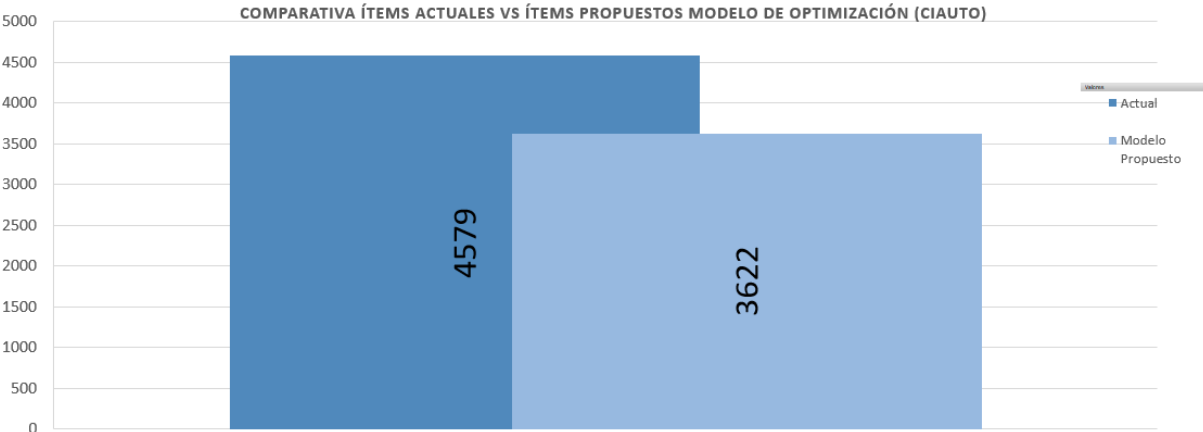


Figura 41: Cantidad de ítems Actuales vs Propuestos en Ciauto.

El modelo propuesto disminuye claramente el exceso de inventario en el centro de distribución; mejorando un 16.5% lo que actualmente existe en inventario. En el caso de los puntos de atención postventa los resultados son mixtos como veremos a continuación.

En el caso del concesionario MA (matriz Ambato), los resultados son los siguientes para el nivel de inventario recomendado.

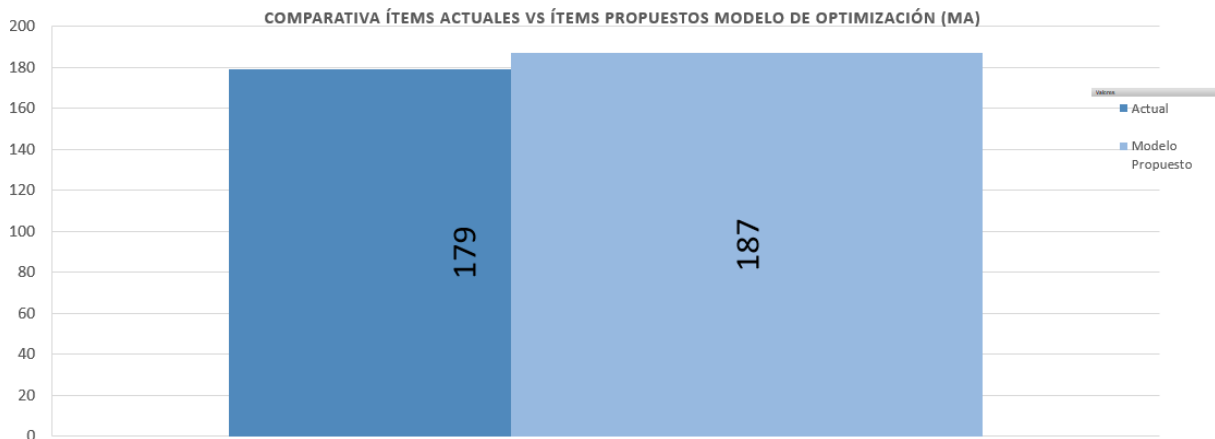


Figura 42: Cantidad de ítems Actuales vs Propuestos en Matriz Ambato.

Una explicación más detallada en que ítems se generan las variaciones se puede encontrar en la siguiente sección. En el caso específico de Matriz Ambato el tiempo de espera para que un repuesto llegue al concesionario es casi mínimo por tanto no existe una necesidad de realizar un incremento grande en los niveles de inventario; sin embargo, hay que mantener un número de ítems razonables para mecánica y procedimientos del día como cambio de aceite y bujías.

En el caso de la ciudad de Guayaquil, específicamente el concesionario GJ (Juan Tanca Marengo) obtenemos los siguientes resultados.

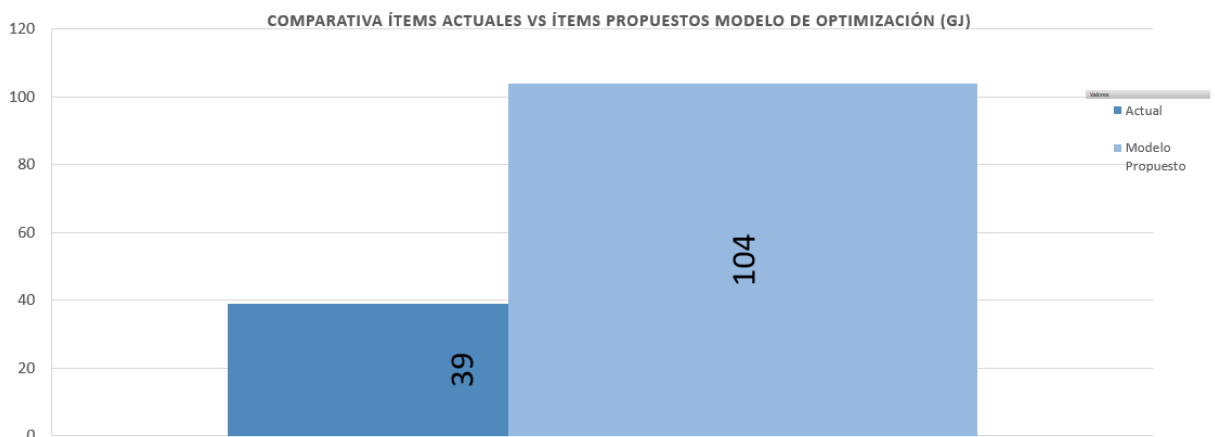


Figura 43: Cantidad de ítems Actuales vs Propuestos en Juan Tanca Marengo.

El incremento de inventario es notorio, potencialmente, el incremento de niveles de inventario es del 166 %. La mayoría del incremento de ítems se registra en procedimientos que son poco convencionales y aumento en procedimientos del día a día. Una de las razones del gran salto, es, además, contrarrestar los efectos de los 2 días de espera que hay entre el centro de distribución a la ciudad de Guayaquil.

Uno de los concesionarios más importante a nivel nacional, ubicado en la ciudad de Quito es GR (Granados) a continuación mostraremos los resultados.

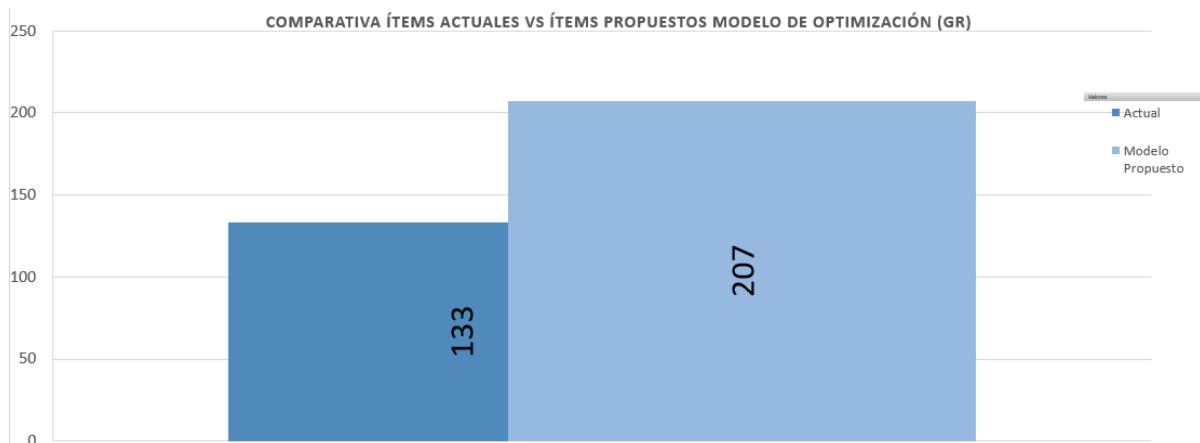


Figura 44: Cantidad de ítems Actuales vs Propuestos en Granados.

En el caso de Granados el incremento se da porque con el modelo pretendido no existe una conexión entre las distintas bodegas locales de Quito y los valles, es, por tanto, que los niveles de inventario se aumentan en distintos ítems ya sean estos de procedimientos diarios u ocasionales.

Volviendo a la ciudad de Guayaquil, uno de los más importantes concesionarios y el primero en la ciudad. El concesionario GY obtiene los siguientes resultados con el modelo propuesto.

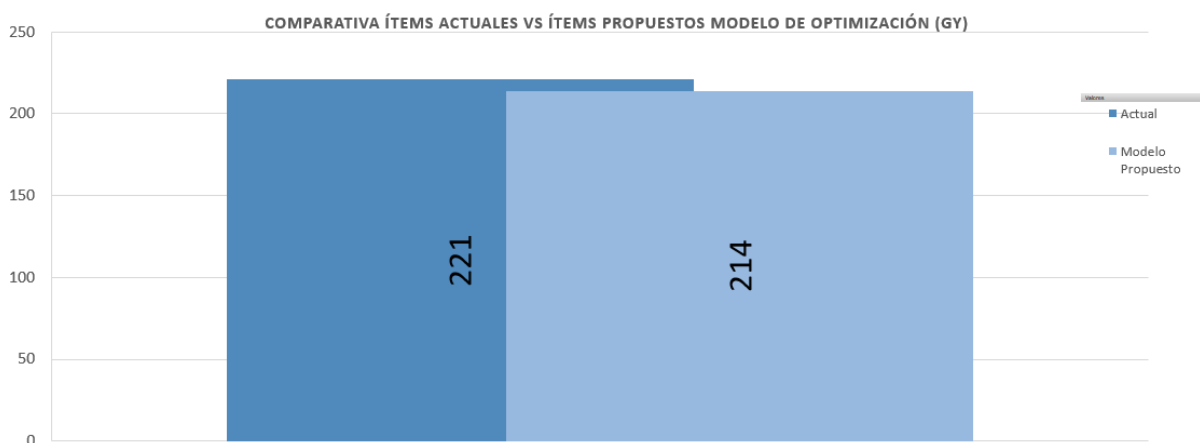


Figura 45: Cantidad de ítems Actuales vs Propuestos en Guayaquil.

La diferencia entre ambos montos no es estadísticamente significativa, la leve disminución se debe a una mejor utilización en el servicio postventa de la ciudad.

Ambacar cuenta con una presencia importante dentro las provincias de la sierra, uno de los concesionarios más importantes fuera de las 3 ciudades más importantes es el de Latacunga (LT). Este concesionario cuenta con la particularidad de estar entre Ambato y Quito, por tanto, es uno de los más rápidos en ser abastecidos en caso de requerir un mayor inventario en mano. A continuación, mostraremos los resultados.

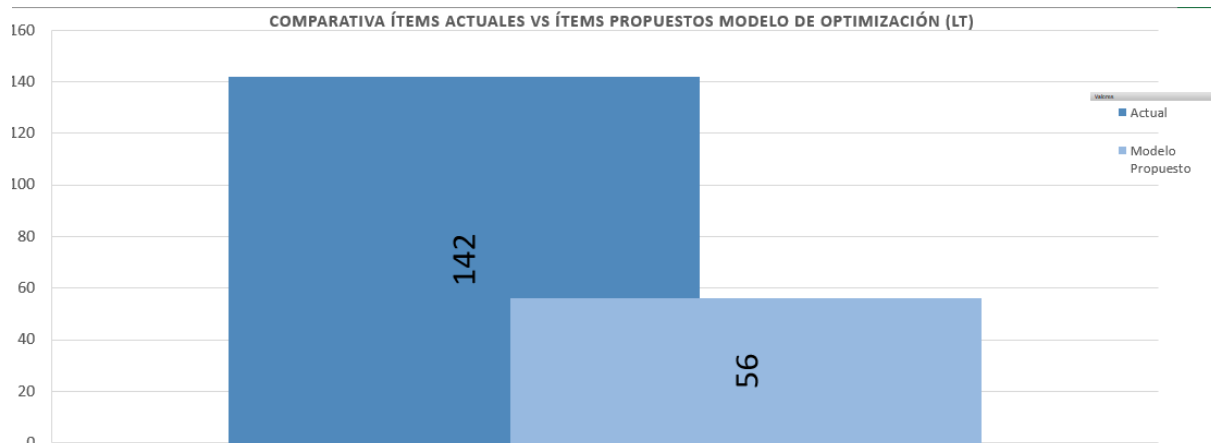


Figura 46: Cantidad de ítems Actuales vs Propuestos en Latacunga.

Una de las principales diferencias que se puede notar es la disminución de ítems que son necesarios para procedimientos que no son cotidianos, ya que, al tener un bajo tiempo de espera desde el centro de distribución, no es necesario tenerlo en las bodegas locales.

Ambacar en su visión de expansión a nivel nacional ha buscado incrementar su presencia en ciudades como Manta, siendo el año 2020 donde abren otra agencia dedicada solo a la venta de sus vehículos y potenciando su concesionario con servicio postventa MT para cubrir la demanda que generaran las nuevas ventas.

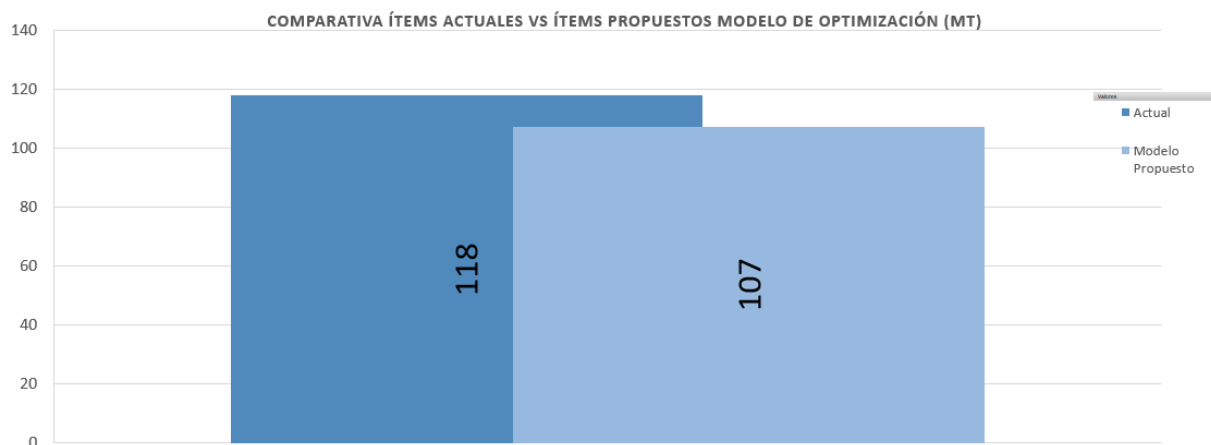


Figura 47: Cantidad de ítems Actuales vs Propuestos en Manta.

El modelo propuesto disminuye el inventario en alrededor del 9.32 %, bajando los niveles en servicios poco ocasionales y manteniendo coberturas para procedimientos comunes y frecuentes.

El concesionario más importante de la ciudad de Quito (QC), Quicentro sur goza de una posesión estratégica y a demás son donde las oficinas del área de planificación de re-

puestos y reaprovisionamiento se encuentran. Al momento se encuentra son 152 unidades de los 50 ítems más importantes de la categoría A. Con el modelo propuesto obtenemos los siguientes resultados.

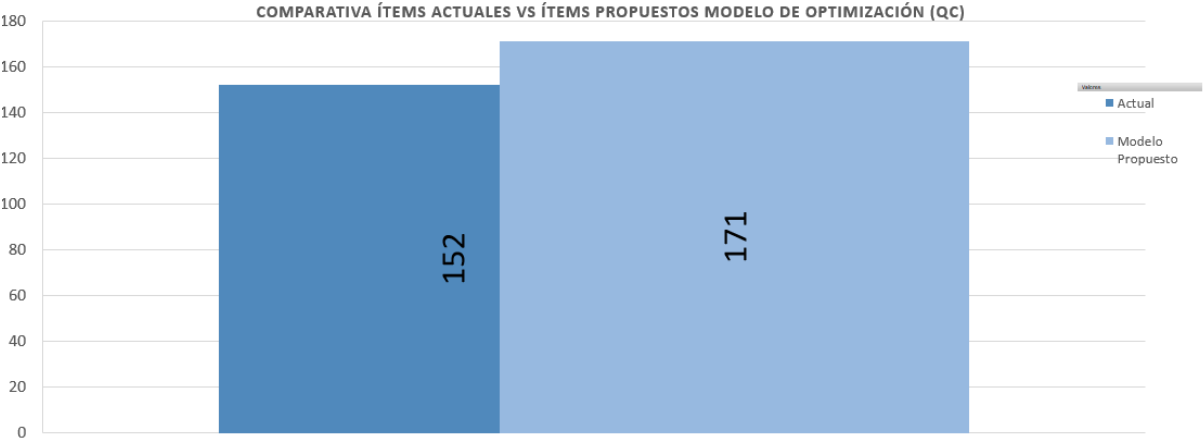


Figura 48: Cantidad de ítems Actuales vs Propuestos en Quicentro Sur.

Los resultados evidencian un incremento en la cantidad de unidades, los mismos responden a la creciente afluencia de vehículos en este punto estratégico y la constante venta de vehículos nuevos en este concesionario.

Continuando en la ciudad de Quito, el segundo concesionario más importante de Quito (QP), Ubicado en la calle Galo Plaza Lasso. Actualmente una calle tradicional para concesionarios y talleres de reparación de vehículos. Ambacar cuenta con una fuerte presencia en este sector de la capital, siendo el mismo el concesionario más grande de Ambacar de toda la ciudad. Actualmente existen 248 unidades de los 50 ítems más importante de la categoría A. Obtenemos los siguientes resultados con el modelo propuesto.

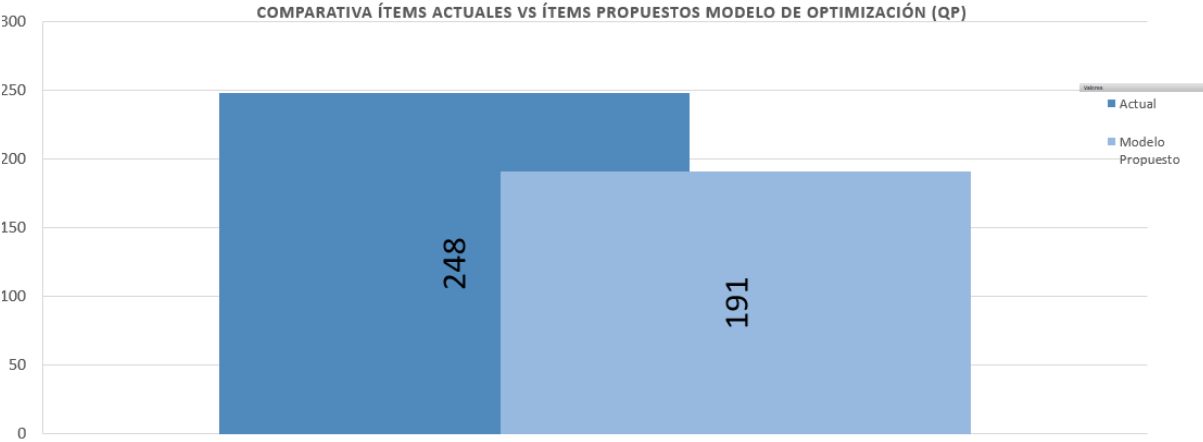


Figura 49: Cantidad de ítems Actuales vs Propuestos en Galo Plaza Lasoo.

Podemos apreciar una disminución significativa de las unidades necesarias, eso se debe

al poco de tiempo de espera entre el centro de distribución.

Por último, uno de los concesionarios con mayor crecimiento ubicado en San Rafael (SR). Debido al gran crecimiento poblacional del Valle de los Chillos este concesionario se ha convertido en uno de los más importantes para Ambacar a nivel nacional. Actualmente cuenta con 194 unidades de los 50 ítems más importantes de la categoría A. obtenemos los siguientes resultados con el modelo propuesto.

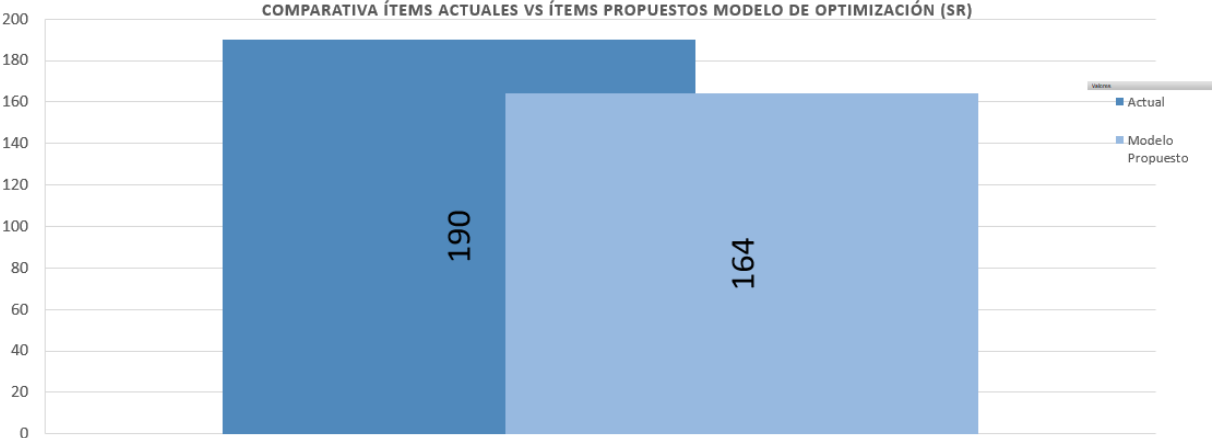


Figura 50: Cantidad de ítems Actuales vs Propuestos en San Rafael.

Con el modelo propuesto se puede lograr una mejora del 13.68%. La disminución se logra principalmente enfocándose en mantener un nivel de servicio bueno en artículos de mantenimiento rápido con un 95.87% y con los ítems ocasionales depender del corto tiempo de espera que poseen los concesionarios de la ciudad de Quito. Sin embargo, en los puntos de venta donde se ha registrado ventas de esos ítems se mantiene un nivel de servicio aceptable del 72.21%.

### 8.11. Niveles de servicio

Para el cálculo de los niveles de servicio nos enfocaremos en la característica del repuesto si es usado en un procedimiento común o si uso implica que el vehículo pase más de un día dentro del taller de la agencia (no común) nos enfocaremos en los puntos de venta donde el ítem tiene ha registrado ventas.

En base a las definiciones anteriores tenemos el siguiente nivel de servicio por tipo de repuesto:

Tipo de Mantenimiento	Nivel de Servicio
Común	95.87 %
No Común	72.21 %

Cuadro 26: Nivel de Servicio

En base a la tabla anterior y los resultados obtenidos al calcular el nivel de servicio promedio de cada uno de los ítems que han registrado ventas en las agencias que tienen servicio post-venta. Para el caso de los repuestos que se usan en procedimiento comunes se tiene una buena cobertura a nivel nacional sobrepasando el nivel de servicio deseado; mientras que en el caso de los repuestos que son usados en procedimientos no comunes se tiene un nivel de servicio aceptable considerando que con el bajo tiempo de espera se puede satisfacer las necesidades de los clientes de manera efectiva.



## 8.12. Impacto del modelo de optimización a nivel de ítem y punto de venta

A continuación, analizaremos los impactos por ítem y punto de venta para ver la variación que se genera con el modelo propuesto.

Centro	A001	Actualmente A001	A002	Actualmente A002	A003	Actualmente A003	A004	Actualmente A004
CH	1	0	1	0	0	0	2	0
CO	2	2	2	0	1	0	4	1
ES	2	0	2	2	4	0	1	1
GJ	0	0	0	0	1	0	5	1
GR	0	0	2	2	3	0	1	1
GY	2	1	2	2	2	0	0	1
IB	2	0	1	2	2	0	4	4
LT	1	0	1	1	3	0	2	2
MA	3	0	2	2	4	1	0	2
MC	2	1	0	0	4	2	2	5
MT	2	0	2	1	2	0	2	2
MU	0	0	2	0	0	0	2	0
PY	0	0	1	0	0	0	2	0
QC	3	1	3	3	4	0	3	2
QL	3	1	1	2	4	0	3	0
QP	2	0	2	2	3	1	3	2
RI	1	0	3	1	1	1	1	2
SA	0	0	0	0	0	0	1	0
SD	3	0	3	1	3	1	2	4
SR	1	0	2	1	4	0	3	1
VD	0	0	3	3	3	1	4	4
ZI	8	5	16	24	210	260	496	547
TOTAL:	38	11	51	49	258	267	543	582

Cuadro 27: Análisis de los ítems A001 a A004 por concesionario.

Los ítems A001-A004 son ítems que son usados en procedimientos no comunes dentro de los concesionarios, es decir, elementos que son usados cuando los vehículos sufren algún accidente o tienen alguna imperfección mecánica. Por factores externos a Ambacar, como fue mencionado en secciones anteriores el kilometraje promedio recorrido bajo en el año 2020 debido a la pandemia. Por eso, no es sorprendente que los ítems más populares sean estos, por tanto, en la mayoría de los concesionarios se establece un incremento en el nivel de inventario que debe existir en cada concesionario.

Centro	A005	Actualmente A005	A006	Actualmente A006	A007	Actualmente A007	A008	Actualmente A008
CH	1	0	0	0	1	0	1	0
CO	1	0	1	0	0	2	2	2
ES	6	0	1	1	1	2	0	2
GJ	6	0	2	0	1	2	1	0
GR	7	0	4	2	0	1	2	1
GY	3	1	1	2	4	7	2	1
IB	2	0	2	2	2	1	1	0
LT	1	0	0	1	0	4	1	1
MA	9	2	4	4	2	2	2	3
MC	4	1	0	0	0	2	4	4
MT	2	1	1	1	2	7	2	2
MU	0	0	0	2	1	0	0	0
PY	0	0	0	1	2	0	1	0
QC	2	1	9	3	0	2	0	1
QL	3	1	5	1	0	2	2	1
QP	4	26	2	15	2	3	10	11
RI	3	1	1	7	3	6	1	3
SA	0	0	0	0	1	0	0	0
SD	3	0	3	3	1	0	3	5
SR	2	0	3	1	2	1	1	3
VD	0	1	2	1	2	9	0	2
ZI	212	260	86	129	69	92	378	402
TOTAL:	271	295	127	176	96	145	414	444

Cuadro 28: Análisis de los ítems A005 a A008 por concesionario.

De igual manera que con los repuestos de la tabla anterior, los ítems A005-A008 son repuestos de procedimientos no comunes que han ganado espacio por la complejidad del año 2020. por tanto, se hace una mejor distribución de los ítems a través de los concesionarios para mejorar los niveles de servicio y la satisfacción de los clientes.

Centro	A009	Actualmente A009	A010	Actualmente A010	A011	Actualmente A011	A012	Actualmente A012
CH	0	0	0	0	0	0	0	0
CO	0	2	0	2	0	1	0	0
ES	0	0	0	0	0	0	1	5
GJ	2	0	2	1	0	3	2	0
GR	2	0	2	0	1	1	3	1
GY	3	2	3	1	3	7	4	1
IB	2	1	2	1	1	1	1	2
LT	1	3	1	3	1	1	1	3
MA	4	0	4	0	2	4	3	4
MC	1	0	0	1	1	2	0	0
MT	1	2	1	1	1	2	2	3
MU	0	0	0	0	0	0	0	0
PY	0	0	0	0	0	0	0	0
QC	2	0	1	2	0	3	4	2
QL	1	0	1	1	0	1	3	1
QP	2	0	3	1	2	4	4	2
RI	1	0	1	1	0	2	1	1
SA	0	0	0	0	0	3	0	0
SD	1	0	2	5	1	4	1	1
SR	0	0	2	0	0	0	1	3
VD	0	2	1	1	1	2	0	2
ZI	156	201	158	189	15	26	94	136
TOTAL:	179	213	184	210	29	67	125	167

Cuadro 29: Análisis de los ítems A009 a A012 por concesionario.

Los ítems A009 a A012 siguen el patrón de los ítems que los preceden, siendo estos usados especialmente cuando hay algún accidente y se deben reemplazar componentes exteriores de los vehículos. Una mejor distribución de estos productos es lo que se propone

para obtener un nivel de servicio del 95 %. Especialmente los vehículos que necesitan estas piezas son vehículos comerciales que han rodado más durante el año 2020.

Centro	A013	Actualmente A013	A014	Actualmente A014	A015	Actualmente A015	A016	Actualmente A016
CH	0	0	4	0	0	0	0	0
CO	0	0	0	0	0	2	0	1
ES	0	0	8	4	0	0	0	1
GJ	0	0	0	0	3	0	1	0
GR	0	0	12	31	5	2	4	0
GY	2	4	14	8	6	0	3	0
IB	0	0	8	12	3	2	4	1
LT	1	2	4	8	1	6	0	1
MA	2	3	8	20	6	11	1	1
MC	0	2	0	8	1	7	0	2
MT	1	2	16	20	3	5	1	2
MU	0	0	0	0	0	0	0	1
PY	0	0	0	0	0	0	0	0
QC	0	1	12	20	4	2	4	1
QL	3	1	16	24	3	2	3	0
QP	1	0	32	20	2	7	1	0
RI	1	1	4	12	1	5	0	1
SA	0	2	4	0	0	3	0	0
SD	0	0	12	20	4	7	1	4
SR	0	0	12	0	4	7	3	1
VD	0	0	0	0	0	4	2	5
ZI	8	16	95	120	110	157	24	63
TOTAL:	19	34	261	327	156	229	52	85

Cuadro 30: Análisis de los ítems A013 a A016 por concesionario.

Los ítems A013-A016 son repuestos más comunes sobre todo en revisiones de kilometraje que se realizan en los distintos talleres a nivel nacional. Especialmente, los ítems A014 (Bujías) representan una mejor distribución en los talleres por su alta necesidad y poco tiempo de instalación requerido, por tanto, es necesario que es ten presentes de manera más homogénea a lo largo de la red.

Centro	A017	Actualmente A017	A018	Actualmente A018	A019	Actualmente A019	A020	Actualmente A020
CH	0	0	0	0	0	0	0	0
CO	0	1	0	0	0	0	16	0
ES	0	0	0	0	1	2	0	0
GJ	0	0	1	0	0	0	0	0
GR	1	0	3	0	0	2	16	0
GY	3	7	3	1	7	6	16	16
IB	0	1	1	0	1	0	0	0
LT	1	1	0	1	2	2	0	0
MA	2	3	3	0	2	2	0	8
MC	1	0	0	0	0	0	0	0
MT	1	2	0	1	1	0	0	0
MU	0	0	0	0	0	0	0	0
PY	0	0	0	0	0	0	0	0
QC	0	2	2	0	1	2	0	0
QL	2	0	2	0	0	0	0	0
QP	2	5	2	2	1	0	16	16
RI	0	2	0	0	2	0	0	0
SA	0	0	0	0	0	0	0	0
SD	2	5	0	1	0	0	16	0
SR	0	0	0	0	1	4	0	0
VD	2	2	0	1	0	0	0	0
ZI	8	6	34	58	11	13	40	52
TOTAL:	25	37	51	65	30	33	120	92

Cuadro 31: Análisis de los ítems A017 a A020 por concesionario.

Los ítems A017-A020 presentan una combinación de ítems comunes y dentro de los servicios de mantenimiento y repuestos que son usados en caso de siniestros, el ítem A017 (Banda de rodamiento) es un repuesto bastante común en los servicios de mantenimiento que se ofrecen y, por tanto, el método propuesto propone que este artículo esté presente en más concesionarios y optimizado en los que ya estaba presente. Los otros ítems son mejor distribuidos considerando el tiempo de espera del centro de distribución a los distintos concesionarios.

Centro	A021	Actualmente A021	A022	Actualmente A022	A023	Actualmente A023	A024	Actualmente A024
CH	0	0	0	0	0	0	0	0
CO	0	0	0	1	2	2	0	0
ES	1	0	1	2	2	2	0	0
GJ	0	0	3	1	2	1	0	0
GR	1	3	2	0	7	3	16	20
GY	5	10	5	1	6	6	16	16
IB	1	0	3	2	8	1	4	4
LT	0	0	2	1	5	2	0	4
MA	3	8	5	0	4	2	9	0
MC	0	0	2	1	0	1	0	0
MT	2	4	2	2	2	0	0	0
MU	0	0	0	0	0	0	0	8
PY	0	0	0	0	0	0	0	0
QC	2	15	3	2	3	2	8	12
QL	0	0	2	2	6	5	6	0
QP	2	4	3	2	5	6	0	0
RI	0	4	2	1	2	1	4	4
SA	0	0	0	0	0	0	0	0
SD	7	12	1	3	3	3	4	12
SR	3	12	1	1	4	4	8	22
VD	0	0	2	4	1	2	0	0
ZI	44	126	62	87	37	80	50	0
TOTAL:	71	198	101	113	99	123	125	102

Cuadro 32: Análisis de los ítems A021 a A024 por concesionario.

Los ítems A021-A023 son repuestos considerados que se usan en casos de siniestros (Vincha Posterior, Neblinero), los resultados propuestos muestran una optimización que consiste en una mejor distribución de los repuestos a través de la red, el artículo A024 (Bujía de encendido) siendo un repuesto más común y considerando la popularidad del vehículo que lo ocupa (HAVAL H6) se necesita un incremento en general de las unidades en todo el sistema.

Centro	A025	Actualmente A025	A026	Actualmente A026	A027	Actualmente A027	A028	Actualmente A028
CH	0	0	0	0	0	0	0	0
CO	0	1	0	0	0	1	0	0
ES	1	1	1	0	0	0	1	0
GJ	2	0	1	0	0	0	2	0
GR	4	0	2	0	2	0	2	4
GY	4	0	4	2	3	3	4	1
IB	3	1	0	0	1	2	1	1
LT	2	2	0	1	1	2	1	4
MA	4	1	2	1	2	6	4	4
MC	0	2	1	0	2	3	1	1
MT	3	1	2	0	2	1	2	1
MU	0	1	0	0	0	0	0	0
PY	0	0	0	0	0	1	0	0
QC	4	2	3	0	0	0	2	1
QL	3	0	2	0	1	0	5	1
QP	2	0	2	1	3	3	2	5
RI	2	3	1	0	0	1	1	1
SA	0	0	0	0	0	0	0	0
SD	2	2	1	1	0	2	3	1
SR	4	4	1	0	0	0	3	0
VD	3	6	1	0	0	0	0	0
ZI	9	12	54	80	8	0	72	98
TOTAL:	52	39	78	86	25	25	106	123

Cuadro 33: Análisis de los ítems A025 a A028 por concesionario.

Los artículos A025-A028 son repuestos que son comúnmente usados en reparaciones en caso de siniestro (Disco de Embrague, Barra Anticolisión, Corrediza Plástica). El modelo propuesto mejora el flujo de ítems a través del sistema llegando incluso a concesionarios que actualmente no tienen en inventario disponible esos repuestos.

Centro	A029	Actualmente A029	A030	Actualmente A030	A031	Actualmente A031	A032	Actualmente A032
CH	0	0	0	0	0	0	0	0
CO	0	0	0	0	0	0	0	2
ES	0	4	0	0	1	0	0	2
GJ	1	0	8	0	3	0	0	1
GR	4	1	4	5	6	0	2	2
GY	3	1	6	7	7	0	3	2
IB	2	2	2	10	2	0	0	1
LT	0	3	3	6	0	1	0	1
MA	4	5	9	4	4	3	2	2
MC	1	0	2	2	1	1	0	0
MT	1	0	5	3	5	3	0	1
MU	0	0	0	3	0	0	0	0
PY	0	0	0	0	0	0	0	0
QC	4	4	10	3	5	0	2	3
QL	3	0	3	4	4	0	2	2
QP	1	3	9	29	2	0	2	11
RI	0	0	3	3	4	0	0	2
SA	0	0	0	0	0	0	0	1
SD	2	4	4	1	4	0	0	1
SR	2	1	7	2	4	0	2	1
VD	0	3	0	0	0	1	0	1
ZI	27	40	93	132	17	0	26	32
TOTAL:	55	71	168	214	69	9	41	68

Cuadro 34: Análisis de los ítems A029 a A032 por concesionario.

Los ítems A029-A032 (Logo Wingle, Kit Cables de Bujías, Tapa Cubo, Disco de Embrague) no son ítems comunes ni ítems que son propios de un siniestro, el modelo propuesto para este grupo de repuestos mejora la presencia dentro de algunos concesionarios, mejo-

rando la cobertura a nivel nacional. Inclusive en algunos casos, solicitando más unidades que están próximos a terminarse.

Centro	A033	Actualmente A033	A034	Actualmente A034	A035	Actualmente A035	A036	Actualmente A036
CH	0	0	0	0	0	0	1	0
CO	0	0	0	0	0	0	0	0
ES	0	4	0	0	0	0	0	0
GJ	8	0	6	3	3	2	7	2
GR	6	3	6	6	3	2	9	6
GY	4	0	4	12	2	1	5	12
IB	2	1	0	1	1	1	10	2
LT	3	3	0	8	0	0	2	8
MA	5	2	7	1	2	0	4	2
MC	2	1	2	2	0	0	3	4
MT	4	2	2	2	3	2	2	2
MU	0	0	0	0	0	0	0	0
PY	0	1	0	0	0	0	0	0
QC	10	9	2	0	4	2	6	2
QL	6	0	2	1	3	1	5	1
QP	2	3	4	0	3	1	2	5
RI	3	4	0	2	0	2	0	2
SA	0	4	0	2	0	0	0	2
SD	4	1	4	0	2	0	6	2
SR	7	2	7	2	3	1	6	0
VD	0	2	0	10	0	0	0	11
ZI	137	172	44	62	45	63	72	115
TOTAL:	203	214	90	114	74	78	140	178

Cuadro 35: Análisis de los ítems A033 a A036 por concesionario.

Los ítems A033-A036 (Manija Int Puerta, Rotula de Suspensión Inferior, Múltiple de Escape, Rotula de Suspensión Delt). Son repuestos comunes en el caso de siniestros o mantenimientos de alto kilometraje de los vehículos. el modelo propuesto propone un mejor flujo y repartición de repuestos para una mejor cobertura a nivel nacional, bajando el inventario disponible en algunos concesionarios que la demanda histórica no justifica un alto número de inventario disponible.

Centro	A037	Actualmente A037	A038	Actualmente A038	A039	Actualmente A039	A040	Actualmente A040
CH	0	0	0	0	0	0	0	0
CO	0	0	0	0	0	1	0	0
ES	0	0	0	2	0	3	0	9
GJ	2	8	2	2	2	0	0	4
GR	2	0	5	4	3	2	7	6
GY	2	2	3	2	2	18	5	0
IB	0	0	0	1	4	1	5	0
LT	0	1	0	5	0	1	0	0
MA	3	2	4	0	6	1	9	7
MC	11	2	0	1	0	2	0	0
MT	6	5	0	1	0	2	0	0
MU	0	0	0	0	0	0	0	0
PY	0	0	0	0	0	2	0	0
QC	1	1	4	1	3	2	3	0
QL	4	2	3	1	5	5	6	5
QP	2	1	3	6	3	5	0	0
RI	0	2	0	1	0	1	0	3
SA	0	4	0	0	0	2	0	0
SD	4	4	2	0	4	2	1	0
SR	6	4	3	1	4	2	3	19
VD	0	2	0	0	0	3	0	0
ZI	90	121	8	11	10	0	55	82
TOTAL:	133	161	37	39	46	55	94	135

Cuadro 36: Análisis de los ítems A037 a A040 por concesionario.

Los ítems A037-A040 (Rele Flash, Retenedor Post Cigüeñal, Banda Alternador, Tapa de Llenado del Alternador) son repuestos bastante comunes dentro de los mantenimientos preventivos que se realizan en todos los talleres de la red de concesionarios de Ambacar. El modelo propuesto propone una mejor distribución de los repuestos bajando el nivel de inventario disponible en exceso de algunos concesionarios y mejorando la cobertura en otros concesionarios elevando la cobertura a nivel nacional.

Centro	A041	Actualmente A041	A042	Actualmente A042	A043	Actualmente A043	A044	Actualmente A044
CH	0	0	0	0	0	0	1	0
CO	0	0	0	0	0	0	1	0
ES	0	0	0	1	0	0	0	2
GJ	0	0	1	0	0	0	1	1
GR	0	0	2	1	2	1	1	0
GY	2	2	3	3	3	3	2	1
IB	0	0	0	0	0	0	2	0
LT	3	2	0	1	1	1	1	2
MA	3	3	0	0	3	1	2	0
MC	0	1	0	1	3	0	0	0
MT	0	2	0	1	2	0	1	0
MU	0	0	0	0	0	0	0	0
PY	0	0	0	0	0	0	0	0
QC	0	1	0	1	1	1	3	0
QL	2	1	3	2	3	2	2	0
QP	3	9	0	1	3	0	3	1
RI	0	1	0	0	0	1	3	2
SA	0	2	0	0	0	0	0	0
SD	0	0	0	2	1	1	2	3
SR	0	0	0	1	3	1	2	0
VD	0	0	0	0	0	1	0	2
ZI	14	22	8	0	10	11	59	87
TOTAL:	27	46	17	15	35	24	86	101

Cuadro 37: Análisis de los ítems A041 a A044 por concesionario.

Los ítems A041-A044 (Plato de Embrague, Termostato, Mascarilla Delt M4, Neblinero Delt) son una mezcla de repuestos usados en mantenimientos de alto kilometraje como de repuestos en caso de siniestros. El modelo propuesto eleva los niveles de inventario general en los casos del termostato y Mascarilla ya que la demanda histórica muestra patrones de crecimiento y sobre todo es necesario una mayor cobertura a nivel nacional.

Centro	A045	Actualmente A045	A046	Actualmente A046	A047	Actualmente A047	A048	Actualmente A048
CH	0	0	0	0	1	0	0	0
CO	0	2	0	1	0	2	0	14
ES	0	5	1	2	1	2	0	0
GJ	4	0	3	3	1	4	9	0
GR	6	1	3	2	3	2	22	0
GY	7	2	3	6	3	5	6	13
IB	0	4	2	13	3	2	22	11
LT	2	2	2	4	2	2	0	13
MA	3	1	3	3	5	1	0	26
MC	1	1	3	2	3	3	20	0
MT	3	3	2	5	5	4	3	5
MU	0	0	0	0	0	2	0	0
PY	0	0	0	0	0	0	0	0
QC	4	10	3	2	4	2	9	13
QL	2	3	2	0	5	0	4	0
QP	5	5	2	2	3	4	14	10
RI	2	2	2	2	1	3	0	5
SA	1	3	0	2	0	2	0	0
SD	2	10	2	5	4	4	10	39
SR	8	10	2	2	5	3	7	26
VD	0	5	2	3	0	3	0	13
ZI	191	211	10	0	16	26	40	47
TOTAL:	241	280	47	59	65	76	166	235

Cuadro 38: Análisis de los ítems A045 a A048 por concesionario.

Los ítems A045-A048 (Empaque Tapas Válvulas, Rodamiento Embrague, Plato de Embrague, Trabador del Selector de Marchas). Estos repuestos son usados en procedimientos de mantenimiento de alto kilometraje. El modelo propuesto mejora la distribución a nivel nacional optimizando los niveles de inventario disponible en concesionarios que no requieren un alto número de estos repuestos y dando los mismos a concesionarios que ya no poseen estos ítems en sus respectivas bodegas.



Centro	A049	Actualmente A049	A050	Actualmente A050
CH	1	0	0	0
CO	1	3	1	2
ES	0	0	0	4
GJ	5	0	1	0
GR	4	9	3	6
GY	6	11	3	11
IB	5	6	2	4
LT	2	16	2	6
MA	10	15	3	6
MC	2	12	1	1
MT	2	10	3	2
MU	0	3	0	5
PY	0	0	0	0
QC	12	8	2	5
QL	16	19	1	3
QP	8	10	2	4
RI	8	10	1	3
SA	0	0	0	1
SD	8	6	4	6
SR	16	39	2	8
VD	0	12	1	4
ZI	76	96	10	10
TOTAL:	182	285	42	91

Cuadro 39: Análisis de los ítems A049 a A050 por concesionario.

Los dos últimos ítems A049-A050 (Bujía de Pre calentamiento, Cables de Bujías). Estos repuestos son comunes en todo tipo de mantenimiento preventivo; el modelo propuesto baja los niveles de inventario en concesionarios que no requieren tener una gran cantidad de estos y dando inventario a los concesionarios que históricamente han mostrado tener demanda.

### 8.13. Puntos de Reaprovisionamiento

En la actualidad Ambacar no cuenta con una política de mínimos o puntos de reaprovisionamiento. Por tanto; el modelo propuesto en base a las demandas históricas y el modelo de optimización de inventario obtiene los siguientes resultados por ítem y punto de venta.

Centro	A001	A002	A003	A004	A005	A006	A007	A008	A009	A010
CH	0	0	0	1	0	0	0	0	0	0
CO	1	0	0	1	0	0	0	0	0	0
ES	1	0	1	0	3	0	0	0	0	0
GJ	0	0	0	1	3	1	0	0	1	1
GR	1	1	1	0	3	2	0	1	1	1
GY	1	1	1	0	1	1	1	1	1	1
IB	0	0	1	1	1	1	1	0	1	1
LT	0	0	1	0	0	0	0	1	0	0
MA	1	1	2	1	3	2	0	1	2	2
MC	1	0	1	1	1	0	0	1	0	0
MT	1	1	1	1	1	1	1	1	0	0
MU	0	1	0	1	0	0	0	0	0	0
PY	0	0	0	1	0	0	1	0	0	0
QC	1	1	2	1	1	3	0	0	1	0
QL	1	1	1	1	1	2	0	1	0	0
QP	1	0	1	1	2	1	1	5	1	1
RI	0	1	0	1	1	0	1	0	0	0
SA	0	0	0	0	0	0	1	0	0	0
SD	1	1	1	1	1	1	1	1	0	1
SR	1	1	2	1	1	1	1	1	0	1
VD	0	1	1	0	0	0	1	0	0	0
ZI	4	9	100	350	125	45	39	201	98	70

Cuadro 40: Puntos de Reaprovisionamiento ítems A001-A010.

El punto de reaprovisionamiento de los ítems A001-A010 nos aseguran que el nivel de servicio al 95 % se cumpla, generalmente en los puntos de venta que no son centrales tienden a tener puntos de reaprovisionamiento 1 o 0. Esto debido a que las demandas mensuales de esos productos no superan una unidad. Además, recordemos que los puntos de reaprovisionamiento se calculan en torno a la demanda y es la unidad necesaria para evitar que haya un quiebre de inventario de acuerdo con la misma.

En los ítems A001 y A002 podemos evidenciar que no existe una necesidad de mantener altos puntos de reaprovisionamiento ya que son muy pocos los puntos de venta que tiene un flujo de rotación alto. Otros repuestos como A004 y A005 muestran una mayor rotación y actividad en los concesionarios. Es por eso por lo que hasta los puntos de reaprovisionamiento de Ciauto es alto en comparación a los primeros repuestos mencionados.

Centro	A011	A012	A013	A014	A015	A016	A017	A018	A019	A020
CH	0	0	0	2	0	0	0	0	0	0
CO	0	0	0	0	0	0	0	0	0	8
ES	0	0	0	4	0	0	0	0	0	0
GJ	0	1	0	0	1	0	0	0	0	0
GR	0	1	0	6	2	1	0	1	0	8
GY	1	2	1	8	3	0	1	1	3	8
IB	0	0	0	4	1	2	0	0	0	0
LT	0	0	0	2	0	0	0	0	1	0
MA	1	1	1	4	3	0	1	1	1	0
MC	0	0	0	0	0	0	0	0	0	0
MT	0	1	0	4	1	0	0	0	0	0
MU	0	0	0	0	0	0	0	0	0	0
PY	0	0	0	0	0	0	0	0	0	0
QC	0	2	0	4	2	2	0	1	0	0
QL	0	1	1	8	1	1	1	1	0	0
QP	1	1	0	12	1	0	1	1	0	8
RI	0	0	0	2	0	0	0	0	1	0
SA	0	0	0	2	0	0	0	0	0	0
SD	0	0	0	4	2	0	1	0	0	8
SR	0	0	0	4	2	1	0	0	0	0
VD	0	0	0	0	0	1	1	0	0	0
ZI	9	60	5	32	53	7	6	17	5	16

Cuadro 41: Puntos de Reaprovisionamiento ítems A011-A020.

Los puntos de reaprovisionamiento de los puntos A011-A020 nos aseguran que el nivel de servicio sea de al menos el 95 %, estos puntos fueron determinados tales que la demanda mensual de los ítems por concesionario sea cubierta y se considere además los tiempos de espera que tiene cada concesionario.

Los ítems A011, A013, A016, A017 y A019 son ítems en los cuales no es necesario tener un punto de reaprovisionamiento alto ya que la demanda de los puntos venta es bajo y son muy pocos los concesionarios que requieren tener más de una unidad de inventario.

Los repuestos A012, A014 y A015 por la demanda histórica y la cantidad de concesionarios donde se encuentran activos y es necesario que la cobertura de estos sea mayor sobre todo en el centro de distribución.

Centro	A021	A022	A023	A024	A025	A026	A027	A028	A029	A030
CH	0	0	0	0	0	0	0	0	0	0
CO	0	0	1	0	0	0	0	0	0	0
ES	0	0	1	0	0	0	0	0	0	0
GJ	0	1	1	0	0	0	0	0	0	3
GR	0	1	3	4	2	0	0	0	2	2
GY	2	2	2	4	2	1	1	2	1	3
IB	0	1	3	4	2	0	0	0	1	1
LT	0	1	1	0	0	0	0	0	0	1
MA	1	2	1	4	2	0	0	2	2	4
MC	0	1	0	0	0	0	0	0	0	1
MT	0	1	0	0	1	0	0	1	0	2
MU	0	0	0	0	0	0	0	0	0	0
PY	0	0	0	0	0	0	0	0	0	0
QC	1	1	1	4	2	1	0	1	2	5
QL	0	0	2	2	1	1	0	2	1	1
QP	1	1	2	0	0	0	1	1	0	3
RI	0	0	0	0	0	0	0	0	0	1
SA	0	0	0	0	0	0	0	0	0	0
SD	3	0	1	0	0	0	0	1	1	2
SR	1	0	2	4	2	0	0	1	1	3
VD	0	0	0	0	1	0	0	0	0	0
ZI	24	36	21	26	5	30	5	49	11	50

Cuadro 42: Puntos de Reaprovisionamiento ítems A021-A030.

Los puntos de reaprovisionamiento de los repuestos A021-A030 muestran en su conjunto una combinación de ítems que no son necesarios mantener niveles altos de inventario y por lo tanto, no es necesario que sus niveles de reaprovisionamiento sean innecesariamente altos (A025, A027, A029).

Los repuestos A021, A022, A026, A028 y A030 necesitan un punto de reaprovisionamiento más alto debido a que a los concesionarios muestran una mayor demanda y sobre todo es necesario que haya en inventario estos repuestos en una mayor cantidad.

Centro	A031	A032	A033	A034	A035	A036	A037	A038	A039	A040
CH	0	0	0	0	0	0	0	0	0	0
CO	0	0	0	0	0	0	0	0	0	0
ES	0	0	0	0	0	0	0	0	0	0
GJ	1	0	4	2	1	3	1	1	1	0
GR	2	0	3	2	1	4	1	2	1	4
GY	3	1	2	1	1	2	1	1	1	2
IB	1	0	1	0	0	5	0	0	2	2
LT	0	0	1	0	0	1	0	0	0	0
MA	2	0	2	3	1	2	1	2	3	5
MC	0	0	1	1	0	1	5	0	0	0
MT	2	0	2	1	2	1	2	0	0	0
MU	0	0	0	0	0	0	0	0	0	0
PY	0	0	0	0	0	0	0	0	0	0
QC	2	1	6	1	2	2	0	2	1	1
QL	2	1	2	1	1	2	2	1	3	3
QP	1	1	1	2	1	1	1	1	1	0
RI	2	0	1	0	0	0	0	0	0	0
SA	0	0	0	0	0	0	0	0	0	0
SD	2	0	2	2	1	3	2	1	2	0
SR	2	1	3	3	2	3	3	1	2	2
VD	0	0	0	0	0	0	0	0	0	0
ZI	10	10	83	20	27	46	69	3	6	35

Cuadro 43: Puntos de Reaprovisionamiento ítems A031-A040.

Los puntos de reaprovisionamiento de los ítems A031-A040 muestran una variedad de repuestos que necesitan tener un alto punto de reaprovisionamiento y repuestos que no es necesario mantener altos para mantener un nivel de servicio de al menos 95 %.

Los ítems A031, A032, A038 y A039 son repuestos que por la demanda histórica que poseen no es necesario que los puntos de reaprovisionamiento sean altos. En especial para los ítems A038 y A039 que los puntos de venta donde es necesario que estén presentes son menos de la mitad de los concesionarios que se manejan en el portafolio de Ambacar.

Centro	A041	A042	A043	A044	A045	A046	A047	A048	A049	A050
CH	0	0	0	0	0	0	0	0	0	0
CO	0	0	0	0	0	0	0	0	0	0
ES	0	0	0	0	0	0	0	0	0	0
GJ	0	0	0	0	2	2	0	6	2	0
GR	1	1	1	0	3	2	2	9	2	1
GY	0	1	2	1	4	1	2	1	4	1
IB	0	0	0	1	0	1	1	14	3	1
LT	1	0	0	0	1	1	1	0	1	1
MA	1	0	2	1	2	1	3	0	6	1
MC	0	0	2	0	0	2	1	9	1	0
MT	0	0	1	0	2	0	3	1	1	2
MU	0	0	0	0	0	0	0	0	0	0
PY	0	0	0	0	0	0	0	0	0	0
QC	0	0	0	1	3	1	2	6	8	0
QL	1	1	2	1	1	1	3	1	8	0
QP	1	0	2	1	2	1	2	5	4	1
RI	0	0	0	1	1	1	1	0	4	0
SA	0	0	0	0	0	0	0	0	0	0
SD	0	0	0	1	0	1	2	5	4	3
SR	0	0	1	1	4	1	3	4	8	1
VD	0	0	0	0	0	1	0	0	0	0
ZI	8	4	7	34	107	6	9	30	48	8

Cuadro 44: Puntos de Reaprovisionamiento ítems A041-A050.

Los puntos de reaprovisionamiento para los ítems A041-A050 en su mayoría son repuestos que no es necesario que sean altos, ya que por la baja rotación y demanda que tienen los mismos en los distintos concesionarios a nivel nacional no es necesario un alto número de inventario que asegure un nivel de servicio de al menos el 95 %.

Los ítems A041, A042, A043, A046, A047 y A050 son ejemplos que por su baja rotación en los puntos de venta no es necesario que se mantengan mínimos de inventario altos que ocuparan espacio innecesario en la bodega. Además, podemos notar que la mayoría de los puntos de venta no es necesaria la existencia de un mínimo.

Por otro lado; los ítems A044, A045, A048 y A049 son repuestos que muestran la necesidad por la alta rotación que los puntos de reaprovisionamiento sean altos ya sea por la cantidad actualmente existe en inventario o la gran cantidad de concesionarios donde es necesario la existencia de mínimos.

## 9. Conclusiones

Los tiempos difíciles que afrontamos, la necesidad de cambio y pronta repuesta a factores externos han generado que las empresas tengan que innovar, cambiar los procesos y buscar de manera constante mejorar. Trabajos como este proyecto de titulación brinda a Ambacar una mejor herramienta que le permite ahorrar dinero en una situación donde cada aspecto del negocio debe ser revisado y optimizado.

Los resultados obtenidos muestran una mejora en la predicción de la demanda alrededor del 20 %, lo que implica una reducción tanto en compra y manufactura de los repuestos estudiados en este caso. Durante el proceso de levantamiento de información y construcción del modelo nos pudimos dar cuenta que hay una gran variedad de información que Ambacar no estaba aprovechando en sus análisis ya sean cosas muy simples como la cantidad de vehículos vendidos de la marca hasta cosas mucho más específicas como el número de vehículos que aún se encuentran vigentes dentro de sus agencias y la cantidad promedio de kilómetros recorridos por año de los mismos. Todas estas investigaciones no solo ayudarán a obtener un mejor pronóstico sino que también aportaran a tener un mejor control e interpretación de toda las bases de datos que fueron proporcionadas para el proyecto de titulación. En particular, uno de los principales resultados obtenidos fue poder determinar al edad promedio y el kilometraje promedio por su uso (particular-comercial) de los vehículos que aun hacen uso de los talleres y servicios que ofrece Ambacar en su red de concesionarios.

Respecto al método usado, el método Random Forest nos permitió considerar una gran cantidad de variables que fueron el producto de analizar, estudiar e investigar dentro de todo el conjunto de información que tiene Ambacar dentro de sus sistemas e información proporcionada por sus colaboradores. Con la ayuda de esas variables proporcionadas el modelo fue mejorando en los resultados obtenidos y cada vez más acercándose a encontrar una diferencia real con el software que actualmente usa la empresa para las pronósticos de la demanda. Si bien el modelo provee en los resultados una mejor precisión que los obtenidos con el software Forecast Pro una de las principales desventajas es el tiempo de cálculo computacional. El modelo propuesta toma alrededor de 12 minutos en obtener los pronósticos mientras que Forecast Pro toma alrededor de 30 segundos. Sin embargo, con el modelo obtenido nos podemos enfocar netamente en los ítems más importantes por sus ventas y rotación con la finalidad de obtener resultados más precisos y mejorar la satisfacción del cliente en cada una de las agencias.

A pesar de que los cambios de metodología y nuevos métodos implementados dentro de cualquier empresa generalmente son medidos en los beneficios económicos o la reducción de costos que generan; también existen beneficios físicos sobre todo en el almacenamiento de las bodegas, poseer repuestos que no rotan o quedarse corto de repuestos necesarios es perjudicial tanto para la perspectiva de los clientes como para los planificadores puesto que sus indicadores son medidos en su precisión en los pronósticos y la rotación de los inventarios de repuestos en los distintos concesionarios. Liberar espacio y ocuparlo de una manera más eficiente es uno de los beneficios que muestra el modelo de optimización multi-echelon, bajando el nivel de inventario en un 15 % respecto a lo que actualmente hay

en bodega. Adicionalmente, política de mínimos ayudará de gran manera a la automatización de procesos. Debido al gran número de repuestos y concesionarios es imposible para un humano analizar de manera rápida y eficiente las necesidades de cada concesionario y repuesto; al establecer políticas de mínimos ayudará que el análisis de los ítems lleguen a esos niveles su reposición y requerimiento sea más eficiente.

Con la metodología propuesta basado en la simulación del sistema nos permite que sin la necesidad de comprometer inventario ver como esto afectaría en términos de niveles de servicio o si el incremento en los niveles de inventario es realmente necesarios para cubrir la necesidad de los clientes. Adicionalmente, en un futuro el modelo puede ser usado para estudiar la necesidad y pertinencia de centros de distribución intermedios en grandes ciudades como Quito y Guayaquil que se encarguen de abastecer únicamente a las agencias ubicadas en las ciudades antes mencionadas; respecto a la implementación del modelo se consideraron factores que si bien eran conocidos por los analistas y planeadores de la demanda de la empresa, en mucho de los casos en sus análisis eran pasados por alta, como ya sean los tiempos de espera y la interacción de cada una de las agencias en el sistema. Por tanto, se depende mucho de las transacciones de ítems entre agencias de la misma ciudad y a que en mucho de los casos haya una demora dentro de los trabajos comunes que se realizan en las agencias. Adicionalmente, la sugerencia de políticas de inventario promedio y puntos de reaprovisionamiento permitirán tener un mayor control del inventario y así evitar tener cantidades innecesarias de productos que pueden ser consideradas como sobrestock y además nos permite mantener niveles de servicio altos que cumplan con la exigencia de los clientes. Actualmente, el proceso de planeación y abastecimiento se realiza en base en calcular la demanda con el software forecast Pro y en base a los resultados garantizar en las agencias 60 días de inventario en los ítems prioritarios; con el modelo se pretende que se usen las políticas sugeridas para gestionar el reabastecimiento en base a esos indicadores. Si bien el tiempo computacional es mayor al que actualmente se usa, se sugiere que el uso del modelo no sea diario, sino que haya un recalcu de las políticas en la periodicidad que la empresa considere pertinente.

Los objetivos de brindar alternativas a la manera tradicional que se está manejando fue cumplida. Los beneficios que se pueden obtener son claros y necesarios sobre todo en épocas como las actuales en donde cada dólar que se pueda ahorrar cuenta y puede hacer la diferencia. Sobre todo, si el enfoque principal es mejorar el servicio que se ofrece a los clientes que han depositado su confianza y dinero en una marca orgullosamente ecuatoriana.



## 10. Anexos

Código de la consolidación de los datos para la matriz de variables del modelo Random Forest.

top 50 ventas:

```
let
  Origen = Excel.Workbook(File.Contents("C:\Users\melarcoa\
    Documents\Pruebas\top 50.xlsm"), null, true),
  MATRIZ_Sheet = Origen{[Item="MATRIZ",Kind="Sheet"]}[Data],
  #"Tipo cambiado" = Table.TransformColumnTypes(MATRIZ_Sheet,{{"
    Column1", type text}, {"Column2", type text}, {"Column3",
    type text}, {"Column4", type text}}),
  #"Encabezados promovidos" = Table.PromoteHeaders(#"Tipo
    cambiado", [PromoteAllScalars=true]),
  #"Tipo cambiado1" = Table.TransformColumnTypes(#"Encabezados
    promovidos",{{"LINEA", type text}, {"CODIGO", type text}, {"
    DESCRIPCION", type text}, {"CATEGORIA GENERAL", type text}})
  ,
  #" ndice agregado" = Table.AddIndexColumn(#"Tipo cambiado1", "
    ndice ", 1, 1, Int64.Type),
  #"Personalizada agregada" = Table.AddColumn(#" ndice agregado
    ", "Top 50", each if [ ndice ]<50 then 1 else 0),
  #"Filas filtradas" = Table.SelectRows(#"Personalizada agregada
    ", each ([Top 50] = 1)),
  #"Columnas con nombre cambiado" = Table.RenameColumns(#"Filas
    filtradas",{{" ndice ", "Categor a C digo Confidencial"}})
in
  #"Columnas con nombre cambiado"
```

Ventas por años

```
let
  Origen = Excel.Workbook(File.Contents("C:\Users\melarcoa\
    Documents\Pruebas\5. VEHICULOS VENDIDOS.xlsx"), null, true),
  #"Ventas por A o_Sheet" = Origen{[Item="Ventas por A o",Kind
    ="Sheet"]}[Data],
  #"Encabezados promovidos" = Table.PromoteHeaders(#"Ventas por
    A o_Sheet", [PromoteAllScalars=true]),
  #"Encabezados promovidos1" = Table.PromoteHeaders(#"Encabezados
    promovidos", [PromoteAllScalars=true]),
  #"Consultas combinadas" = Table.NestedJoin(#"Encabezados
    promovidos1", {"ME S DE VENTA"}, Mes, {"Mes"}, "Mes",
    JoinKind.LeftOuter),
  #"Se expandi Mes" = Table.ExpandTableColumn(#"Consultas
    combinadas", "Mes", {"Cod_Mes"}, {"Mes.Cod_Mes"}),
```

```

#"Tipo cambiado" = Table.TransformColumnTypes("#Se expandi
    Mes",{{"ANIO", Int64.Type}, {"AGENCIA", type text}, {"UND
    VENTA AGENCIA", Int64.Type}}),
#"Filas filtradas" = Table.SelectRows("#Tipo cambiado", each
    true),
#"Columnas combinadas" = Table.CombineColumns(Table.
    TransformColumnTypes("#Filas filtradas", {{"ANIO", type text
    }, {"Mes.Cod_Mes", type text}}, "es-EC"),{"ANIO", "Mes.
    Cod_Mes", "AGENCIA"},Combiner.CombineTextByDelimiter("-",
    QuoteStyle.None),"A o -Mes-Agencia")
in
#"Columnas combinadas"

```

### Km recorridos:

```

let
    Origen = Excel.Workbook(File.Contents("C:\Users\melarcoa\
        Documents\Pruebas\Reporte.xlsx"), null, true),
    #"Km Recorridos_Sheet" = Origen{[Item="Km Recorridos",Kind="
        Sheet"]}[Data],
    #"Encabezados promovidos" = Table.PromoteHeaders("#Km
        Recorridos_Sheet", [PromoteAllScalars=true]),
    #"Columna de anulaci n de dinamizaci n" = Table.
        UnpivotOtherColumns("#Encabezados promovidos", {"AGENCIA"},
        "Atributo", "Valor"),
    #"Columnas combinadas" = Table.CombineColumns("#Columna de
        anulaci n de dinamizaci n",{ "Atributo", "AGENCIA"},
        Combiner.CombineTextByDelimiter("-", QuoteStyle.None),"A o
        Agencia"),
    #"Columnas con nombre cambiado" = Table.RenameColumns("#
        Columnas combinadas",{ "Valor", "Km recorrido promedio"})
in
#"Columnas con nombre cambiado"

```

### Días última visita:

```

let
    Origen = Excel.Workbook(File.Contents("C:\Users\melarcoa\
        Documents\Pruebas\Reporte.xlsx"), null, true),
    #"D as ltima visita_Sheet" = Origen{[Item="D as ltima
        visita",Kind="Sheet"]}[Data],
    #"Encabezados promovidos" = Table.PromoteHeaders("#D as
        ltima visita_Sheet", [PromoteAllScalars=true]),
    #"Tipo cambiado" = Table.TransformColumnTypes("#Encabezados
        promovidos",{{"A o", type text}, {"PDV", Int64.Type}, {"
        D as desde ltima visita", type number}, {"Column4", type
        any}, {"Column5", type any}, {"Column6", type any}, {"
        Column7", type any}, {"Column8", type any}, {"Column9", type

```

```

    any}, {"Column10", type any})),
#"Columnas quitadas" = Table.RemoveColumns("#Tipo cambiado",{"
    Column4", "Column5", "Column6", "Column7", "Column8", "
    Column9", "Column10"}),
#"Columnas combinadas" = Table.CombineColumns(Table.
    TransformColumnTypes("#Columnas quitadas", {"PDV", type
    text}}, "es-EC"),{"PDV", "A o"},Combiner.
    CombineTextByDelimiter("-", QuoteStyle.None),"PDV.1")
in
#"Columnas combinadas"

```

## Matriz Consolidada

```

let
    Origen = Excel.Workbook(File.Contents("C:\Users\melarcoa\
        Documents\Pruebas\ConsumosRedAmbacarpor Agencia_17022020.
        xlsx"), null, true),
    Hoja1_Sheet = Origen{[Item="Hoja1",Kind="Sheet"]}[Data],
    #"Encabezados promovidos" = Table.PromoteHeaders(Hoja1_Sheet, [
        PromoteAllScalars=true]),
    #"Tipo cambiado" = Table.TransformColumnTypes("#Encabezados
        promovidos",{"ANIO", Int64.Type}, {"AGE", type text}, {"MES
        ", Int64.Type}, {"LINEA", type text}, {"CODIGO", type text},
        {"DESCRIP", type text}, {"Mostrador", Int64.Type}, {"Taller
        ", Int64.Type}, {"Total_Ventas", Int64.Type})),
    #"Columnas quitadas" = Table.RemoveColumns("#Tipo cambiado",{"
        Mostrador", "Taller"}),
    #"Consultas combinadas" = Table.NestedJoin("#Columnas quitadas
        ", {"CODIGO"}, #"Top 50 ventas", {"CODIGO"}, "Top 50 ventas
        ", JoinKind.LeftOuter),
    #"Se expandi Top 50 ventas" = Table.ExpandTableColumn("#
        Consultas combinadas", "Top 50 ventas", {"Categor a C digo
        Confidencial", "Top 50"}, {"Top 50 ventas.Categor a
        C digo Confidencial", "Top 50 ventas.Top 50"}),
    #"Filas filtradas" = Table.SelectRows("#Se expandi Top 50
        ventas", each ([Top 50 ventas.Top 50] = 1)),
    #"Filas ordenadas" = Table.Sort("#Filas filtradas",{"ANIO",
        Order.Ascending})),
    #"Columnas con nombre cambiado" = Table.RenameColumns("#Filas
        ordenadas",{"AGE", "Concesionario"}),
    #"Columnas quitadas1" = Table.RemoveColumns("#Columnas con
        nombre cambiado",{"LINEA"}),
    #"Columna duplicada" = Table.DuplicateColumn("#Columnas
        quitadas1", "Concesionario", "Concesionario - Copia"),
    #"Columna duplicada1" = Table.DuplicateColumn("#Columna
        duplicada", "ANIO", "ANIO - Copia"),
    #"Columna duplicada2" = Table.DuplicateColumn("#Columna
        duplicada1", "MES", "MES - Copia"),
    #"Columnas combinadas" = Table.CombineColumns(Table.

```

```

TransformColumnTypes("#Columna duplicada2", {"ANIO - Copia", type text}, {"MES - Copia", type text}, "es-EC"), {"ANIO - Copia", "MES - Copia", "Concesionario - Copia"}, Combiner.CombineTextByDelimiter("-", QuoteStyle.None), "A o -Mes-Agencia"),
#"Consultas combinadas1" = Table.NestedJoin("#Columnas combinadas", {"A o -Mes-Agencia"}, #"Ventas por A o", {"A o -Mes-Agencia"}, "Ventas por A o", JoinKind.LeftOuter),
#"Se expandi Ventas por A o" = Table.ExpandTableColumn("#Consultas combinadas1", "Ventas por A o", {"UND VENTA AGENCIA"}, {"Ventas por A o.UND VENTA AGENCIA"}),
#"Personalizada agregada" = Table.AddColumn("#Se expandi Ventas por A o", "Solo repuestos?", each if [Ventas por A o.UND VENTA AGENCIA]=null then 1 else 0),
#"Personalizada agregada1" = Table.AddColumn("#Personalizada agregada", "Promoci n", each if [Total_Ventas]>40 then 1 else 0),
#"Personalizada agregada2" = Table.AddColumn("#Personalizada agregada1", "Covid Cierre ", each if [ANIO]=2020 and [MES]>3 and [MES]<7 then 1 else 0),
#"Personalizada agregada3" = Table.AddColumn("#Personalizada agregada2", "Covid No cierre", each if [ANIO]=2020 and [MES ]>6 then 1 else 0),
#"Valor reemplazado" = Table.ReplaceValue("#Personalizada agregada3", null, 0, Replacer.ReplaceValue, {"Ventas por A o .UND VENTA AGENCIA"}),
#"Columna duplicada3" = Table.DuplicateColumn("#Valor reemplazado", "ANIO", "ANIO - Copia"),
#"Columna duplicada4" = Table.DuplicateColumn("#Columna duplicada3", "Concesionario", "Concesionario - Copia"),
#"Columnas combinadas1" = Table.CombineColumns(Table.TransformColumnTypes("#Columna duplicada4", {"ANIO - Copia", type text}, "es-EC"), {"ANIO - Copia", "Concesionario - Copia"}, Combiner.CombineTextByDelimiter("-", QuoteStyle.None), "Combinada"),
#"Consultas combinadas3" = Table.NestedJoin("#Columnas combinadas1", {"Combinada"}, #"D as ltima visita", {"PDV .1"}, "D as ltima visita", JoinKind.LeftOuter),
#"Se expandi D as ltima visita" = Table.ExpandTableColumn("#Consultas combinadas3", "D as ltima visita", {"D as desde ltima visita"}, {"D as ltima visita.D as desde ltima visita"}),
#"Valor reemplazado2" = Table.ReplaceValue("#Se expandi D as ltima visita", null, 85, Replacer.ReplaceValue, {"D as ltima visita.D as desde ltima visita"}),
#"Consultas combinadas2" = Table.NestedJoin("#Valor reemplazado2", {"Combinada"}, #"Km Recorridos", {"A o Agencia"}, "Km Recorridos", JoinKind.LeftOuter),
#"Se expandi Km Recorridos" = Table.ExpandTableColumn("#

```

```

    Consultas combinadas2", "Km Recorridos", {"Km recorrido
    promedio"}, {"Km Recorridos.Km recorrido promedio"}),
#"Personalizada agregada4" = Table.AddColumn("#Se expandi Km
    Recorridos", "km ponderado", each if [ANIO] = 2010 then 600
    else if [ANIO] = 2011 then 800 else if [ANIO] = 2012 then
    1200 else if [ANIO] = 2013 then 1600 else if [ANIO] = 2014
    then 1700 else if [ANIO] = 2015 then 1900 else if [ANIO] =
    2016 then 2000 else if [ANIO] = 2017 then 2000 else if [Km
    Recorridos.Km recorrido promedio] = null then 7900 else [Km
    Recorridos.Km recorrido promedio]),
#"Columnas quitadas2" = Table.RemoveColumns("#Personalizada
    agregada4",{"Km Recorridos.Km recorrido promedio"}),
#"Columnas con nombre cambiado1" = Table.RenameColumns("#
    Columnas quitadas2",{"Ventas por A o .UND VENTA AGENCIA", "
    Venta de Veh culos por Agencia"}),
#"Columnas quitadas3" = Table.RemoveColumns("#Columnas con
    nombre cambiado1",{"Top 50 ventas.Top 50", "A o -Mes-Agencia
    "}),
#"Columnas con nombre cambiado2" = Table.RenameColumns("#
    Columnas quitadas3",{"Top 50 ventas.Categor a C digo
    Confidencial", "Categor a C digo Confidencial"}),
#"Columnas quitadas4" = Table.RemoveColumns("#Columnas con
    nombre cambiado2",{"CODIGO"}),
#"Columnas reordenadas" = Table.ReorderColumns("#Columnas
    quitadas4",{"ANIO", "Concesionario", "MES", "Categor a
    C digo Confidencial", "DESCRIP", "Total_Ventas", "Venta de
    Veh culos por Agencia", "Solo repuestos?", "Promoci n", "
    Covid Cierre ", "Covid No cierre", "Combinada", "D as
    ltima visita.D as desde ltima visita", "km ponderado"})
in
    #"Columnas reordenadas"

```

## Código Simulación Optimización del inventario

```

!pip install simpy

import simpy
import numpy as np

class new_order(object):

    def __init__(self, requester, order_qty):
        self.requester = requester
        self.orderQty = order_qty

class stocking_facility(object):

    # CREACI N DE SISTEMA DE ABASTECIMIENTO

```

```

# Definiciones y variables
def __init__(self, env, node_id, is_source, initial_inv, ROP,
             base_stock,
             upstream, hist_demand, default_lead_time,
             lead_time_delay):
    self.env = env
    self.name = "node" + str(node_id)
    self.isSource = is_source
    self.on_hand_inventory = initial_inv
    self.inventory_position = initial_inv
    self.ROP = ROP
    self.baseStock = base_stock
    self.upstream = upstream
    self.histDemand = hist_demand
    self.defaultLeadTime = default_lead_time
    self.leadTimeDelay = lead_time_delay
    self.order_q = []
    self.totalDemand = 0.0
    self.totalShipped = 0.0
    self.serviceLevel = 0.0
    self.avgOnHand = 0.0
    self.onHandMon = []

# Funciones
self.env.process(self.check_inventory())
self.env.process(self.prepare_replenishment())
self.env.process(self.serve_customer())

# rdenes de reaprovisionamiento
def check_inventory(self):
    while True:
        yield self.env.timeout(1.0)
        if self.inventory_position <= 1.00 * self.ROP:
            order_qty = self.baseStock - self.on_hand_inventory
            order = new_order(self, order_qty)
            self.upstream.order_q.append(order)
            self.inventory_position += order_qty

# Cumplimiento de las ordenes de reaprovisionamiento
def prepare_replenishment(self):
    while True:
        if len(self.order_q) > 0:
            order = self.order_q.pop(0)

            shipment = min(order.orderQty, self.
                           on_hand_inventory)
            if not self.isSource:
                self.inventory_position -= shipment

```

```

        self.on_hand_inventory -= shipment

        # En casa de falta de stock en el nodo de
        # reaprovisionamiento esperar
        # hasta que haya stock disponible para enviar
        remaining_order = order.orderQty - shipment
        if remaining_order:
            while not self.on_hand_inventory >=
                remaining_order:
                yield self.env.timeout(1.0)
            if not self.isSource:
                self.inventory_position -= remaining_order
                self.on_hand_inventory -= remaining_order
            self.env.process(self.ship(order.orderQty, order.
                requester))
        else:
            yield self.env.timeout(1.0)

# Enviar la orden de reaprovisionamiento
def ship(self, qty, requester):
    lead_time = requester.defaultLeadTime + \
        np.random.choice(requester.leadTimeDelay,
            replace=True) # bootstrap sample lead time
            delay
    yield self.env.timeout(lead_time)
    requester.on_hand_inventory += qty

# Atender la demanda del cliente
def serve_customer(self):
    while True:
        self.onHandMon.append(self.on_hand_inventory)
        yield self.env.timeout(1.0)
        demand = np.random.choice(self.histDemand, replace=True
            )
        self.totalDemand += demand
        shipment = min(demand, self.on_hand_inventory)
        self.totalShipped += shipment
        self.on_hand_inventory -= shipment
        self.inventory_position -= shipment

# Simulaci n del sistema
def simulate_network(seedinit, num_nodes, network, initial_inv, ROP
    ,
        base_stock, demand, lead_time, lead_time_delay
    ):

    env = simpy.Environment() # inicilizaci n del proceso de
        simulaci n Simpu

```

```

np.random.seed(seedinit)

nodes = []

for i in range(num_nodes):
    if i == 0: # Definici n del centro de abastecimiento con
               stock suficiente
        s = stocking_facility(env, i, 1, initial_inv[i], ROP[i]
                               ], base_stock[i],
                               None, np.zeros(100), lead_time[i]
                               ], lead_time_delay)
    else:
        # Ubicaci n del nodo que abastece a la agencia
        for j in range(num_nodes):
            if network[j][i] == 1: # Entonces j abastece a i
                s = stocking_facility(env, i, 0, initial_inv[i]
                                       ], ROP[i], base_stock[i],
                                       nodes[j], demand[:, i -
                                                           1], lead_time[i],
                                       lead_time_delay)
                break

        nodes.append(s)

env.run(until=360)

# Encontrar el nivel de servicio en cada nodo
for i in range(num_nodes):
    nodes[i].serviceLevel = nodes[i].totalShipped / (nodes[i].
                                                       totalDemand + 1.0e-5)

# Encontrar el nivel de inventario promedio en cada nodo
for i in range(num_nodes):
    if i == 0: # Suponer que el centro de distribuci n tiene
               siempre inventario suficiente
        nodes[i].avgOnHand = 0.0
    else:
        nodes[i].avgOnHand = np.mean(nodes[i].onHandMon)

return nodes # Retorna informaci n sobre el inventario en
              los nodos

!pip install scikit-optimize
import numpy as np
from skopt import gp_minimize, forest_minimize
import csv
import time
import pandas as pd

```



```

# Leer la demanda hist rica + predicciones Random Forest
demandAllNodes = []
with open('/content/drive/MyDrive/data/data2/demanddataa050.csv', '
r') as f:
    next(f) # Quitar encabezado
    reader = csv.reader(f)
    for a in reader:
        demandAllNodes.append([float(j) for j in a])

demandAllNodes = np.array(demandAllNodes) # No considerar el nodo
de abastecimiento
# Tiempo de espera de abastecimiento del centro de distrbuci n a
las agencias
leadTimeDelay = []
with open('/content/drive/MyDrive/data/data2/leadtimeextradays.csv'
, 'r') as f:
    reader = csv.reader(f)
    for a in reader:
        leadTimeDelay.append(int(a[0]))

leadTimeDelay = np.array(leadTimeDelay)

# Expresar el grafo de manera matricial
numNodes = 23
nodeNetwork = np.zeros((numNodes, numNodes))
nodeNetwork[0, 1] = 1
nodeNetwork[1, 2] = 1
nodeNetwork[1, 3] = 1
nodeNetwork[1, 4] = 1
nodeNetwork[1, 5] = 1
nodeNetwork[1, 6] = 1
nodeNetwork[1, 7] = 1
nodeNetwork[1, 8] = 1
nodeNetwork[1, 9] = 1
nodeNetwork[1, 10] = 1
nodeNetwork[1, 11] = 1
nodeNetwork[1, 12] = 1
nodeNetwork[1, 13] = 1
nodeNetwork[1, 14] = 1
nodeNetwork[1, 15] = 1
nodeNetwork[1, 16] = 1
nodeNetwork[1, 17] = 1
nodeNetwork[1, 18] = 1
nodeNetwork[1, 19] = 1
nodeNetwork[1, 20] = 1
nodeNetwork[1, 21] = 1
nodeNetwork[1, 22] = 1

```

```

# OBJETIVOS NIVEL DE SERVICIO

serviceTarget = np.array([0, 0.95, 0.95, 0.95, 0.95, 0.95, 0.95,
    0.95, 0.95, 0.95, 0.95, 0.95, 0.95, 0.95, 0.95, 0.95, 0.95,
    0.95, 0.95, 0.95, 0.95, 0.95, 0.95])

def getObj(initial_guess):

    # stock inicial
    excess_inventory_guess = initial_guess[: (numNodes - 1)]
    ROP_guess = initial_guess[(numNodes - 1):]
    base_stock_guess = np.add(excess_inventory_guess, ROP_guess)

    # Base de stock del nodo de reaprovisionamiento
    baseStock = np.insert(base_stock_guess, 0, 5)

    # Punto de reaprovisionamiento del centro de distribuci n debe
    ser cero
    ROP = np.insert(ROP_guess, 0, 0)

    #Inicializaci n del nivel de inventatio
    initialInv = 0.9*baseStock

    replications = 100
    totServiceLevel = np.zeros(numNodes)
    totAvgOnHand = 0.0
    for i in range(replications):
        nodes = simulate_network(i, numNodes, nodeNetwork, initialInv,
            ROP, baseStock, \
                                demandAllNodes, defaultLeadTime,
                                leadTimeDelay)

        totServiceLevel = np.array([totServiceLevel[j] + \
            nodes[j].serviceLevel for j in
            range(len(nodes))]) #convert
            list to array

        totAvgOnHand += np.sum([nodes[j].avgOnHand for j in range(
            len(nodes))])

    servLevelPenalty = np.maximum(0, serviceTarget -
        totServiceLevel/replications) # element-wise max
    objFunValue = totAvgOnHand/replications + 1.0e6*np.sum(
        servLevelPenalty)
    return objFunValue
niter = 1
def callbackF(res):

```

```

global niter
print('{0:4d} {}{1:6.6f}'.format(niter, res.fun))
niter += 1

##### Informaci n para la optimizaci n , inventataio
disponible y punto de reaprovisionamiento inicial #####
excess_inventory_initial_guess = [10, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5,
5, 5, 5, 5, 5, 5, 5, 5, 5]
ROP_initial_guess = [5, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3,
3, 3, 3, 3, 3, 3, 3]
guess_vec = excess_inventory_initial_guess + ROP_initial_guess #
concatenate lists
guess = []
for j in range(len(guess_vec)):
    g = (0, guess_vec[j])
    guess.append(g)

NUM_CYCLES = 70
TIME_LIMIT = 1440
start_time = time.time()
print("\nMax {}time {}limit: {}".format(str(TIME_LIMIT) + " {}minutes")
print("Max {}algorithm {}cycles: {}".format(str(NUM_CYCLES) + " {}(20 {}iterations
{}per {}cycle)")
print("The {}algorithm {}will {}run {}either {}for {}".format(str(TIME_LIMIT) + " {}
minutes {}or {}".format(str(NUM_CYCLES) + " {}cycles")
ctr = 1
elapsed_time = (time.time() - start_time)/60.0
bestObj = 1e8
bestSoln = []
bestCycle = 0
while ctr <= NUM_CYCLES and elapsed_time <= TIME_LIMIT:
    print('\nCycle: {}'.format(ctr))
    print('{0:4s} {}{1:9s}'.format('Iter', 'Obj'))
    """
    opt = forest_minimize(func=getObj
                        , dimensions=guess
                        , n_calls=20
                        , n_random_starts=10
                        , random_state=707
                        , verbose=False
                        , callback=callbackF
                        , kappa=50)
    """
    opt = gp_minimize(func=getObj
                    , dimensions=guess
                    , n_calls=20
                    , n_random_starts=10
                    , random_state=ctr
                    , verbose=False

```

```

, callback=callbackF
, kappa=50)

if opt.fun < bestObj:
    bestObj = opt.fun
    bestSoln = opt.x
    bestCycle = ctr
ctr += 1
elapsed_time = (time.time() - start_time)/60.0

print("\nFinal objective: " + "{0:10.3f}".format(bestObj))
print("\nFinal solution: " + str(bestSoln))
print("\nBest cycle: " + str(bestCycle))
print("\nTotal time: " + "{0:3.2f}".format(elapsed_time) + "
minutes")

```

### Optimización del inventario

Comparativas de los concesionarios entre el actual del inventario contra el modelo propuesto.

#### Chone

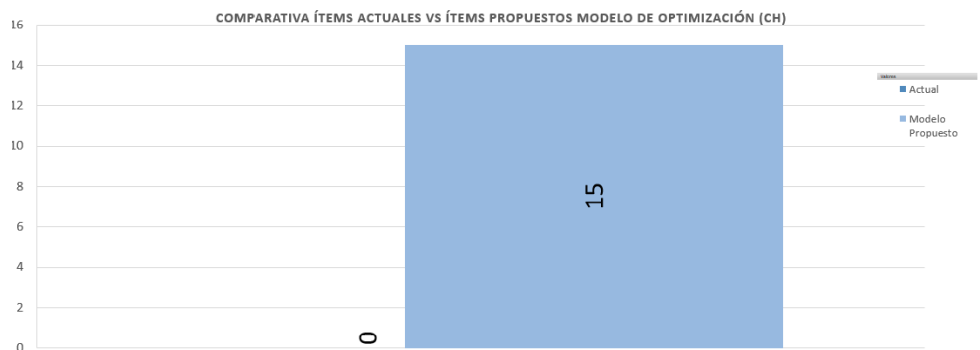


Figura 51: Cantidad de ítems Actuales vs Propuestos en Chone.

#### El Coca

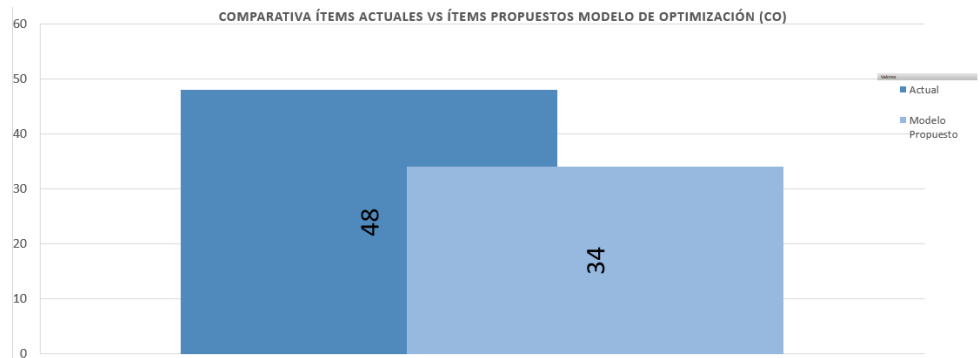


Figura 52: Cantidad de ítems Actuales vs Propuestos en El Coca.

## Esmeraldas

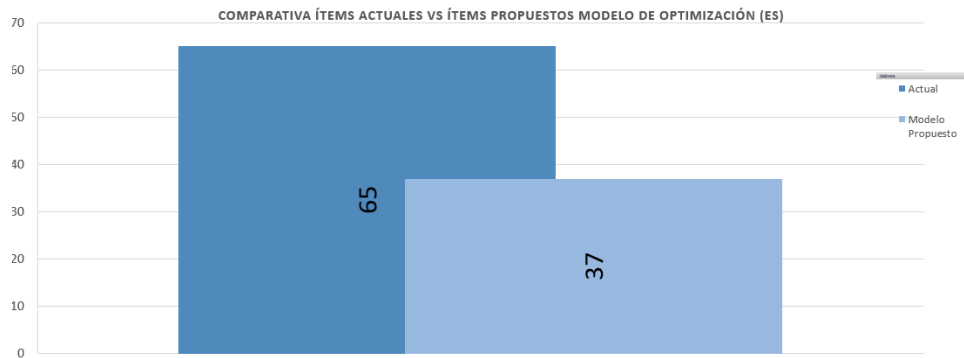


Figura 53: Cantidad de ítems Actuales vs Propuestos en Esmeraldas.

## Ibarra

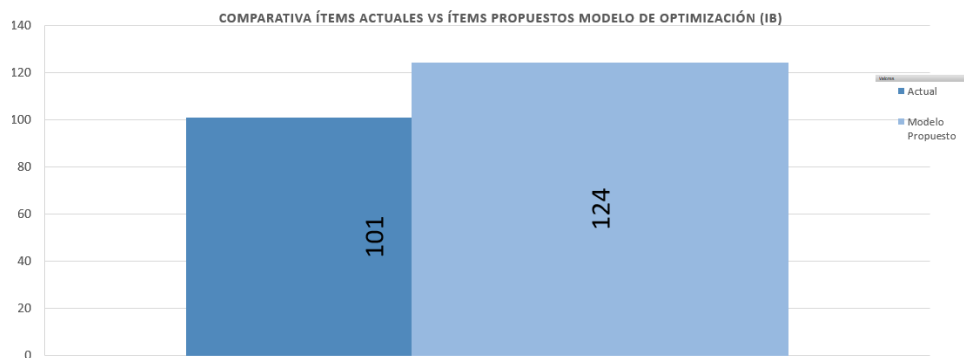


Figura 54: Cantidad de ítems Actuales vs Propuestos en Ibarra.

## Machala

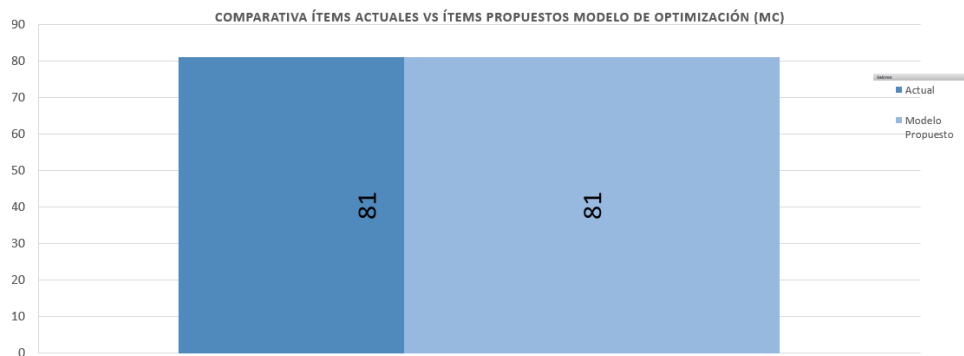


Figura 55: Cantidad de ítems Actuales vs Propuestos en Machala.

## SemiNuevos Ambato

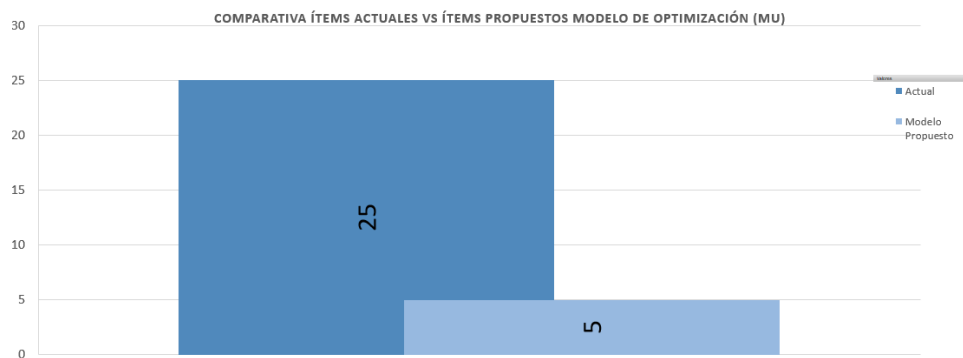


Figura 56: Cantidad de ítems Actuales vs Propuestos en Seminuevos Ambato.

### El Puyo

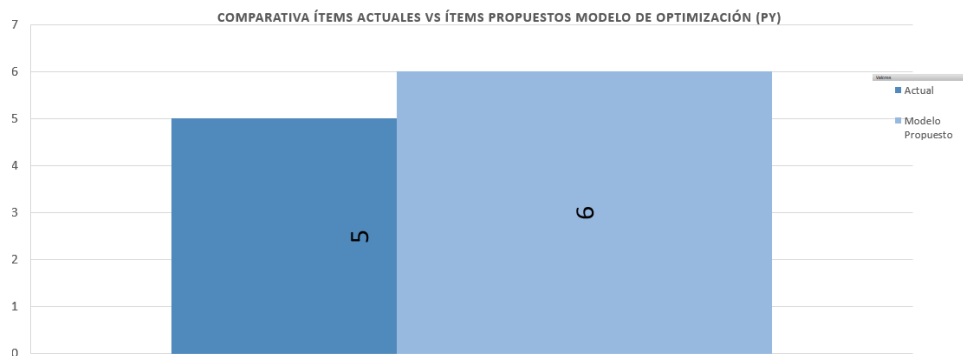


Figura 57: Cantidad de ítems Actuales vs Propuestos en El Puyo.

### SemiNuevos Quito

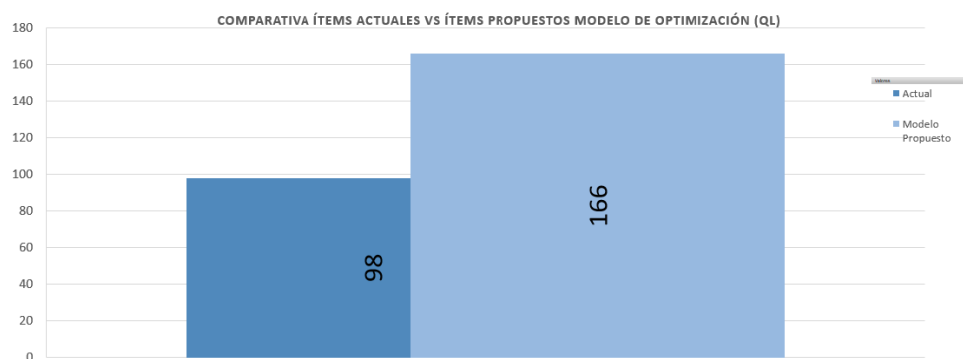


Figura 58: Cantidad de ítems Actuales vs Propuestos en Seminuevos Quito.

### Riobamba

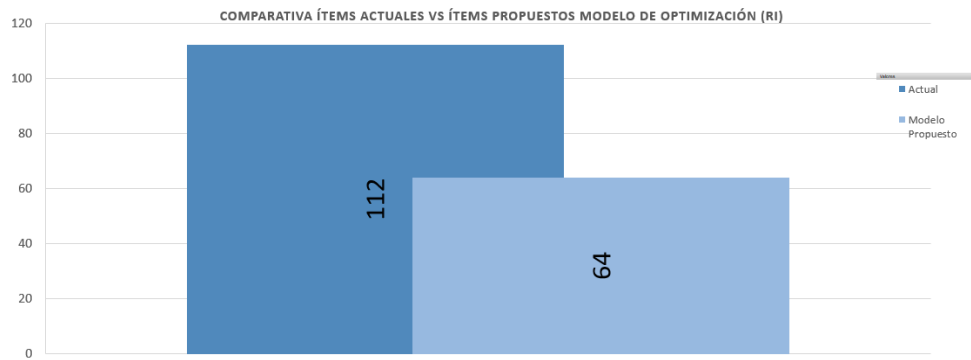


Figura 59: Cantidad de ítems Actuales vs Propuestos en Riobamba.

### Salinas

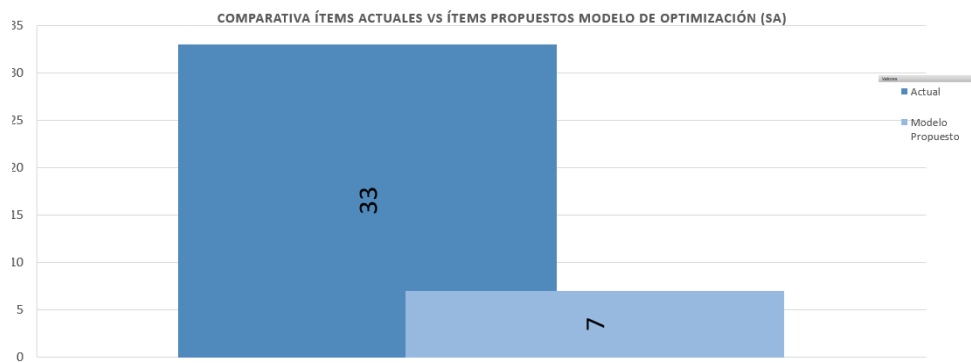


Figura 60: Cantidad de ítems Actuales vs Propuestos en Salinas.

### Santo Domingo

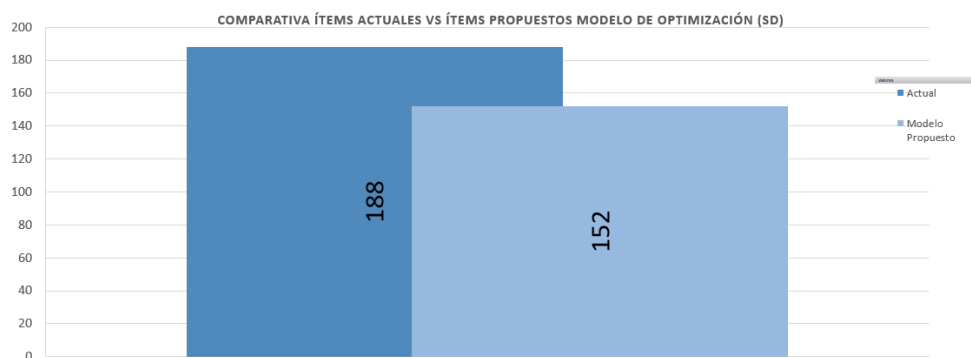


Figura 61: Cantidad de ítems Actuales vs Propuestos en Santo Domingo.

### Bolivariana

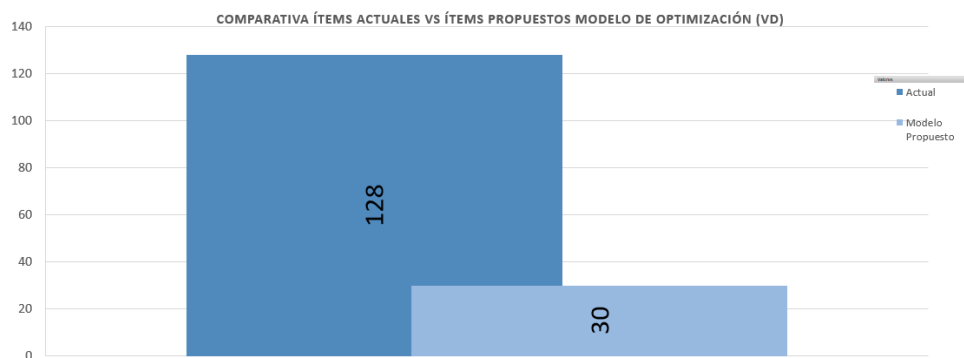


Figura 62: Cantidad de ítems Actuales vs Propuestos en Bolivariana.



## 11. Referencias Bibliográficas

- Seeger, M. W., Salinas, D., Flunkert, V. (2016). Bayesian intermittent demand forecasting for large inventories. In *Advances in Neural Information Processing Systems* (pp. 4646-4654).
- Böse, J.-H., Flunkert, V., Gasthaus, J., Januschowski, T., Lange, D., Salinas, D., Schelter, S., Seeger, M., Wang, Y. (2017). Probabilistic demand forecasting at scale. *Proceedings of the VLDB Endowment*, 10(12), 1694-1705. <https://doi.org/10.14778/3137765.3137775>
- Raman, A., DeHoratius, N., Ton, Z. (2001). Execution: The Missing Link in Retail Operations. *California Management Review*, 43(3), 136-152.
- Esther, U. U. (2012). Effectiveness of Inventory Management in a Manufacturing Company. Thesis, Caritas University, Amorji-Nike, Department of Accountancy.
- Punia, S., Nikolopoulos, K., Singh, S. P., Madaan, J. K., Litsiou, K. (2020). Deep learning with long short-term memory networks and random forests for demand forecasting in multi-channel retail. *International Journal of Production Research*, 1-16.
- Jiang, S., Xiao, R., Wang, L., Luo, X., Huang, C., Wang, J. H., ... Nie, X. (2019). Combining Deep Neural Networks and classical time series regression models for forecasting patient flows in Hong Kong. *IEEE Access*, 7, 118965-118974.
- Mei, J., He, D., Harley, R., Habetler, T., Qu, G. (2014, July). A random forest method for real-time price forecasting in New York electricity market. In *2014 IEEE PES General Meeting| Conference Exposition* (pp. 1-5). IEEE.
- Kumar, M., Thenmozhi, M. (2006, January). Forecasting stock index movement: A comparison of support vector machines and random forest. In *Indian institute of capital markets 9th capital markets conference paper*.
- Scornet, E. (2017). Tuning parameters in random forests. *ESAIM: Proceedings and Surveys*, 60, 144-162.
- Bondestab, J. (2015). Multi-Echelon Inventory Optimization and Fast Time-to-Benefit | ToolsGroup. Recuperado de: <https://www.toolsgroup.com/blog/multi-echelon-inventory-optimization-and-fast-time-to-benefit/>
- Agarwal, A. (2019). Multi-echelon Supply Chain Inventory Planning using Simulation-Optimization with Data Resampling. arXiv preprint arXiv:1901.00090.
- Bourgin, D. D., Peterson, J. C., Reichman, D., Russell, S. J., Griffiths, T. L. (2019, May). Cognitive model priors for predicting human decisions. In *International conference on machine learning* (pp. 5133-5141).
- Yule, G. U. (1927). VII. On a method of investigating periodicities disturbed series, with special reference to Wolfer's sunspot numbers. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 226(636-646), 267-298.
- Tsay, R. S. (2000). Time series and forecasting: Brief history and future research. *Journal of the American Statistical Association*, 95(450), 638-643.

- Box, G. (2013). Box and Jenkins: time series analysis, forecasting and control. In *A Very British Affair* (pp. 161-215). Palgrave Macmillan, London.
- Lapide, L., LaPiDe, L. (2009). History to demand-driven forecasting. *Journal of Business Forecasting Methods and Systems*, 28(2), 18.
- Bernard, m. (2016, 19 de febrero). A Short History of Machine Learning – Every Manager Should Read. *Forbes*. Recuperado de: <https://www.forbes.com/sites/bernardmarr/2016/02/19/a-short-history-of-machine-learning-every-manager-should-read/#53ca2bea15e7/>
- Makridakis, S., Spiliotis, E., Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PloS one*, 13(3), e0194889.
- Zhao, L. (2009). Neural networks in business time series forecasting: benefits and problems. *Review of Business Information Systems (RBIS)*, 13(3).
- Wang, J., Wang, J. (2017). Forecasting stochastic neural network based on financial empirical mode decomposition. *Neural Networks*, 90, 8-20.
- Ahmed, N. K., Atiya, A. F., Gayar, N. E., El-Shishiny, H. (2010). An empirical comparison of machine learning models for time series forecasting. *Econometric Reviews*, 29(5-6), 594-621.
- Qiu, M., Song, Y. (2016). Predicting the direction of stock market index movement using an optimized artificial neural network model. *PloS one*, 11(5), e0155133.
- Bredahl Kock, A., Teräsvirta, T. (2016). Forecasting macroeconomic variables using neural network models and three automated model selection techniques. *Econometric Reviews*, 35(8-10), 1753-1779.
- Teboul, W. (2018, July 22). Why use Machine Learning Instead of Traditional Statistics. Recuperado de: <https://towardsdatascience.com/why-use-machine-learning-instead-of-traditional-statistics-334c2213700a/>
- Trujillo, P. (20 de septiembre 2016). 6 Times Horrific Inventory Control Almost Killed These Companies. Recuperado de <https://www.business2community.com/product-management/6-times-horrific-inventory-control-almost-killed-companies-01659644>
- Fritsch, D. (3 de agosto 2015). 6 Inventory Control Techniques for Stock Optimization. Recuperado de <https://www.eazystock.com/blog/6-inventory-control-techniques-for-stock-optimization/>
- The Right Tool for the Job. (s.f). Forecastpro. Recuperado de: <https://www.forecastpro.com/solutions/forecast-pro/forecasting-methods/>
- Álvarez R. (2014). "Modelo de Scoring Crediticio en una empresa financiera", Montevideo, Uruguay.
- Brown, R. G, Meyer, R.F, (1961). "The fundamental theory of exponential smoothing, *Operations Research*", p. 673-685.
- Ostertagova, E, Ostertag, O. (2011). The Simple Exponential Smoothing Model.

- Martinez, J. (2017). Modelación Box Jenkins ARIMA. *Econometría Modelación ARIMA*.
- Nash, S. G., Nocedal, J. (1991). A numerical study of the limited memory BFGS method and the truncated-Newton method for large scale optimization. *SIAM Journal on Optimization*, 1(3), 358-372.
- Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the society for Industrial and Applied Mathematics*, 11(2), 431-441.
- Makridakis, S., Hibon, M. (1997). ARMA models and the Box–Jenkins methodology. *Journal of Forecasting*, 16(3), 147-163.
- Zhang, X., Quadrianto, N., Kersting, K., Xu, Z., Engel, Y., Sammut, C., Reid, M., Liu, B., Webb, G. I., Sammut, C., Sipper, M., Saitta, L., Sebag, M., Aggarwal, C. C., Gärtner, T., Horváth, T., Wrobel, S., Chakrabarti, D., McAuley, J., Getoor, L. (2011). Gaussian Process. In *Encyclopedia of Machine Learning* (pp. 428–439). Springer US. [https://doi.org/10.1007/978-0-387-30164-8\\_324](https://doi.org/10.1007/978-0-387-30164-8_324)
- Davis, R. A. (2014). *Gaussian Process: Theory*. Wiley StatsRef: Statistics Reference Online.
- Andrews, B. H., Dean, M. D., Swain, R., Cole, C. (2013). Building ARIMA and ARIMAX models for predicting long-term disability benefit application rates in the public/private sectors. *Society of Actuaries*, 1-54.
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.
- Breiman, L. (1996). Bagging predictors. *Machine learning*, 24(2), 123-140.
- Menéndez Velázquez, A. (1998). Una breve introducción a la teoría de grafos. *Suma*.
- Breiman, L. (1983). *Classification Regression Trees*. Wadsworth Pub Co.
- Dietterich, T. G. (2000). An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. *Machine learning*, 40(2), 139-157.
- Ho, T. K. (1998). The random subspace method for constructing decision forests. *IEEE transactions on pattern analysis and machine intelligence*, 20(8), 832-844.
- Wolpert, D. H., Macready, W. G. (1999). An efficient method to estimate bagging's generalization error. *Machine Learning*, 35(1), 41-55.
- Bontempi, G., Taieb, S. B., Le Borgne, Y. A. (2012, July). Machine learning strategies for time series forecasting. In *European business intelligence summer school* (pp. 62-77). Springer, Berlin, Heidelberg.
- Tyralis, H., Papacharalampous, G. (2017). Variable selection in time series forecasting using random forests. *Algorithms*, 10(4), 114.
- Díaz-Uriarte, R., De Andres, S. A. (2006). Gene selection and classification of microarray data using random forest. *BMC bioinformatics*, 7(1), 3.
- Biau, G., Scornet, E. (2016). A random forest guided tour. *Test*, 25(2), 197-227.

- Yiu, T. (2012, 12 junio). Node splitting in a random forest model is based on a random subset of features for each tree. [Gráfico]. towardsdatascience. <https://towardsdatascience.com/understanding-random-forest-58381e0602d2>
- Wager, S. (2014). Asymptotic theory for random forests. arXiv preprint arXiv:1405.0352.
- Bühlmann, P., Yu, B. (2002). Analyzing bagging. *The Annals of Statistics*, 30(4), 927-961.
- Ishwaran, H. (2015). The effect of splitting on random forests. *Machine Learning*, 99(1), 75-118.
- Scornet, E., Biau, G., Vert, J. P. (2015). Consistency of random forests. *The Annals of Statistics*, 43(4), 1716-1741.
- Bernard, S., Heutte, L., Adam, S. (2008, September). Forest-rk: A new random forest induction method. In *International conference on intelligent computing* (pp. 430-437). Springer, Berlin, Heidelberg.
- Genuer, R., Poggi, J. M., Tuleau-Malot, C. (2010). Variable selection using random forests. *Pattern recognition letters*, 31(14), 2225-2236.
- Latinne, P., Debeir, O., Decaestecker, C. (2001, July). Limiting the number of trees in random forests. In *International workshop on multiple classifier systems* (pp. 178-187). Springer, Berlin, Heidelberg.
- Moon, J., Kim, Y., Son, M., Hwang, E. (2018). Hybrid short-term load forecasting scheme using random forest and multilayer perceptron. *Energies*, 11(12), 3283.
- Sbai, N., Berrado, A. (2018). A literature review on multi-echelon inventory management: the case of pharmaceutical supply chain. In *MATEC Web of Conferences* (Vol. 200, p. 00013). EDP Sciences.
- The serial, two-echelon production/inventory system. (2006). [Gráfico]. semanticscholar. <https://www.semanticscholar.org/paper/Multi-echelon-production-inventory-systems-3A-and-Houtum/a058600f89d2d7886126661e3dcb1dd60d7dc719/figure/0>
- Sbai, N., Barredo, A. (2018). Example of assembly system [Gráfica]. matec-conferences. [https://www.matec-conferences.org/articles/mateconf/pdf/2018/59/mateconf\\_iwtsce2018\\_00013.pdf](https://www.matec-conferences.org/articles/mateconf/pdf/2018/59/mateconf_iwtsce2018_00013.pdf)
- Clark, A. J. (1972). An informal survey of multi-echelon inventory theory. *Naval research logistics quarterly*, 19(4), 621-650.
- Lee, C. B. (2003). *Evant White Paper Series*.
- Agarwal, A. (2019). Multi-echelon Supply Chain Inventory Planning using Simulation-Optimization with Data Resampling. arXiv preprint arXiv:1901.00090.
- Agarwal, A. (2019). A multi-echelon supply chain example [Gráfico]. semanticscholar. <https://www.semanticscholar.org/paper/Multi-echelon-Supply-Chain-Inventory-Planning-using-Agarwal/12ebceba1d30f9ac625b625fd606ee77e94e7e16>
- Sbai, N., Berrado, N. (2018). Distribution system [4] [Gráfico]. matec-conferences. [https://www.matec-conferences.org/articles/mateconf/pdf/2018/59/mateconf\\_iwtsce2018\\_](https://www.matec-conferences.org/articles/mateconf/pdf/2018/59/mateconf_iwtsce2018_)

00013.pdf

Sbai, N., Barredo, A. (2018). Illustration of a general multi-echelon inventory system of 8 depots [Gráfica]. matec-conferences. [https://www.matec-conferences.org/articles/mateconf/pdf/2018/59/mateconf\\_iwtsce2018\\_00013.pdf](https://www.matec-conferences.org/articles/mateconf/pdf/2018/59/mateconf_iwtsce2018_00013.pdf)

Sherbrooke, C. C. (1968). METRIC: A multi-echelon technique for recoverable item control. *Operations research*, 16(1), 122-141.

Ganeshan, R. (1999). Managing supply chain inventories: A multiple retailer, one warehouse, multiple supplier model. *International Journal of Production Economics*, 59(1-3), 341-354.

Hopp, W. J., Zhang, R. Q., Spearman, M. L. (1999). An easily implementable hierarchical heuristic for a two-echelon spare parts distribution system. *IIE transactions*, 31(10), 977-988.

Glasserman, P., Tayur, S. (1994). The stability of a capacitated, multi-echelon production-inventory system under a base-stock policy. *Operations Research*, 42(5), 913-925.

Graves, S. C., Willems, S. P. (2000). Optimizing strategic safety stock placement in supply chains. *Manufacturing Service Operations Management*, 2(1), 68-83.

Rios, L. M., Sahinidis, N. V. (2013). Derivative-free optimization: a review of algorithms and comparison of software implementations. *Journal of Global Optimization*, 56(3), 1247-1293.

Quienes Somos - Great Wall. (s. f.). Ambacar Ecuador. <https://www.ambacar.ec/sobre-ambacar>

AEADE. (2020, enero). ANUARIO 2019. [https://www.aeade.net/wp-content/uploads/2020/07/AEADE-ANUARIO-2019-OK\\_1.pdf](https://www.aeade.net/wp-content/uploads/2020/07/AEADE-ANUARIO-2019-OK_1.pdf)