ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE CIENCIAS

ANÁLISIS DEL RIESGO TRIBUTARIO EN LA DEVOLUCIÓN DEL IMPUESTO AL VALOR AGREGADO A PERSONAS ADULTAS MAYORES, MEDIANTE HERRAMIENTAS DE MINERÍA DE DATOS, PARA EL AÑO FISCAL 2015

TRABAJO DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL TÍTULO DE INGENIERO EN CIENCIAS ECONÓMICAS Y FINANCIERAS

PROYECTO DE INVESTIGACIÓN

JUAN JAVIER HARO QUILUMBA
juanharo42@hotmail.com

JOHANNA ALEXANDRA HIDALGO PARRA
joko_flk_1712 @hotmail.com

DIRECTOR: JOSÉ FERNANDO RAMÍREZ ÁLVAREZ, MSc.
jose.ramirez@epn.edu.ec

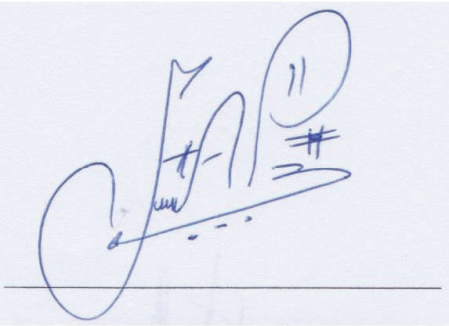
CODIRECTOR: MARCO PATRICIO NARANJO CHIRIBOGA, PhD.
marco.naranjo@epn.edu.ec

Quito, Junio 2017

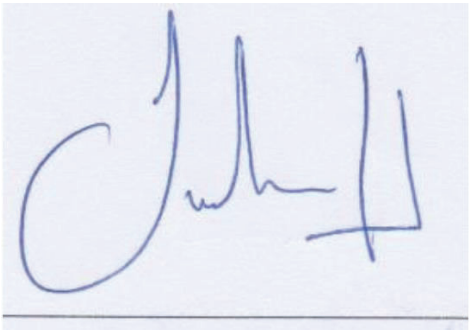
DECLARACIÓN

Nosotros, Juan Javier Haro Quilumba y Johanna Alexandra Hidalgo Parra, declaramos bajo juramento que el trabajo aquí descrito es de nuestra autoría; que no ha sido previamente presentado para ningún grado o calificación profesional; y, que hemos consultado las referencias bibliográficas que se incluyen en este documento.

La Escuela Politécnica Nacional puede hacer uso de los derechos correspondientes a este trabajo, según lo establecido por la Ley de Propiedad Intelectual, por su Reglamento y por la normatividad institucional vigente.



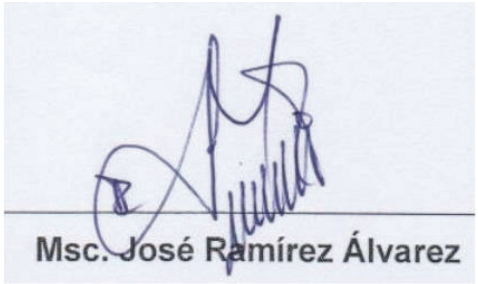
Juan Javier Haro Quilumba



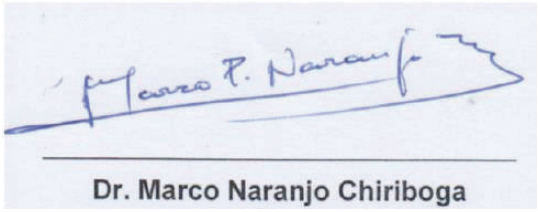
Johanna Alexandra Hidalgo Parra

CERTIFICACIÓN

Nosotros, José Ramírez y Marco Naranjo, certificamos que el presente trabajo fue desarrollado por el señor Juan Javier Haro Quilumba y la señorita Johanna Alexandra Hidalgo Parra, bajo nuestra supervisión.



DIRECTOR



CODIRECTOR

AGRADECIMIENTOS

A mis padres, hermanas y amigos, gracias por todo el apoyo.

Juan Haro

AGRADECIMIENTOS

A Dios por bendecirme cada día.

A toda mi familia por estar siempre conmigo y brindarme todo su cariño. Los amo. ¡Gracias por todo!

A mis amigos, por hacer de este tiempo algo inolvidable, gracias por todas las experiencias compartidas.

A nuestro tutor y al Servicio de Rentas Internas, por la apertura y confianza depositada en nosotros para llevar a cabo este proyecto.

Johanna Hidalgo

DEDICATORIA

A todas las personas que hicieron posible cumplir esta meta.

Juan Haro

DEDICATORIA

A Alexandra, Geovanny, Michelle y Andrés, para ustedes todos mis logros.

Johanna Hidalgo

ÍNDICE DE CONTENIDO

LISTA DE FIGURAS ……i

LISTA DE TABLAS ……ii

LISTA DE ANEXOS ……iii

RESUMEN……iv

ABSTRACT……v

INTRODUCCIÓN 1

1 FUNDAMENTOS TEÓRICOS DE LA GESTIÓN Y POLÍTICA TRIBUTARIA 5

1.1 Intervención del Estado como ente regulador del mercado 5

1.2 Política tributaria 6

1.3 Gestión tributaria 9

1.3.1 Gestión de riesgos tributarios 12

2 DEVOLUCIÓN DEL IVA A PERSONAS ADULTAS MAYORES EN ECUADOR ... 16

2.1 Análisis comparativo de proyectos tributarios en beneficio de las personas adultas mayores, entre Ecuador y países de América Latina. 16

2.2 Análisis de la devolución del IVA a personas adultas mayores 21

2.2.1 Algunas estadisticas internacionales 21

2.2.2 Contexto legal en Ecuador 23

2.2.3 Normas para solicitar la devolución del IVA a personas adultas mayores 24

2.2.4 Proceso de devolución del IVA 26

2.3 Riesgo tributario en la devolución del IVA a personas adultas mayores 27

3 MINERÍA DE DATOS Y PRINCIPALES HERRAMIENTAS 29

3.1 Conceptualización de la minería de datos 29

3.2 Clasificación de las técnicas de minería de datos 31

3.2.1 Técnicas predictivas. Los modelos de elección binaria 32

3.2.1.1 Modelos de probabilidad lineal (M.P.L.) 32

3.2.1.2 Modelo probit 33

3.2.1.3 Modelo logit 34

3.2.1.4 Estimación de máxima verosimilitud (V.M.) 36

3.2.1.5 Comparación del modelo logit con otros modelos de elección binaria 37

3.2.1.6 Bondad de ajuste 37

3.2.1.7 Métodos de busqueda secuencial 41

3.3 Metodología Cross Industry Standard Process for Data Mining (CRISP-DM) 44

3.3.1 Fases de la metodología (CRISP-DM) 45
3.3.1.1 Comprensión del negocio 45
3.3.1.2 Comprensión de los datos 46
3.3.1.3 Preparación de los datos 46
3.3.1.4 Modelado 47
3.3.1.5 Evaluación 48
3.3.1.6 Implementación o distribución 48

4 ANÁLISIS DEL RIESGO TRIBUTARIO EN LA DEVOLUCIÓN DEL IVA A PERSONAS ADULTAS MAYORES EN ECUADOR 50

4.1 Panorama de devolución del IVA 50
4.2 Análisis exploratorio de datos 51
4.2.1 Principales estadísticos 52
4.3 Preparación de datos 53
4.3.1 Datos eliminados 53
4.3.2 Transformación de variables 54
4.3.3 Datos de prueba y entrenamiento 55
4.4 Desarrollo del modelo 57
4.4.1 Variable objetivo “Riesgo” 57
4.4.2 Especificación del modelo 58
4.4.3 Estimación del modelo 58
4.4.4 Bondad de ajuste del modelo seleccionado 60
4.5 Interpretacion de resultados 60
4.5.1 Interpretación 60
4.5.2 Capacidad clasificatoria 65
4.6 Evaluación e implementación 68

5 CONCLUSIONES Y RECOMENDACIONES 69

5.1 Conclusiones 69
5.2 Recomendaciones 71

BIBLIOGRAFÍA 73

ANEXOS 77

LISTA DE FIGURAS

FIGURA 1: PIRÁMIDE DE CUMPLIMIENTO 10
FIGURA 2: PROCESO DE GESTIÓN DE RIESGO..... 12
FIGURA 3: MODELO DE GESTIÓN DE RIESGOS DEL SRI 15
FIGURA 4: TOTAL DE INGRESOS TRIBUTARIOS COMO PORCENTAJE DEL PIB..... 21
FIGURA 5: ESTRUCTURAS IMPOSITIVAS EN AMÉRICA LATINA Y OCDE, 2014 22
FIGURA 6: ETAPAS MINERÍA DE DATOS 30
FIGURA 7: FUNCIÓN DE DISTRIBUCIÓN ACUMULATIVA LOGÍSTICA..... 34
FIGURA 8: TIPOS DE CURVA ROC 41
FIGURA 9: CICLO DE LA METODOLOGÍA CRISP-DM 44
FIGURA 10. EVOLUCIÓN DEL MONTO DEVUELTO Y DEL NÚMERO DE BENEFICIARIOS, POR AÑO DE NOTIFICACIÓN..... 50
FIGURA 11: DISTRIBUCIÓN DE LA VARIABLE "MONTO SOLICITADO" (VARIABLE ORIGINAL)... 55
FIGURA 12: DISTRIBUCIÓN DE LA VARIABLE "MONTO SOLICITADO" (LOGARITMO NATURAL)55
FIGURA 13: EFECTO MARGINAL DE LA VARIABLE "NÚMERO DE PERIODOS"..... 64
FIGURA 14: EFECTO MARGINAL DE LA VARIABLE "MONTO SOLICITADO (LN)" 65
FIGURA 15: CURVA ROC (DATOS ENTRENAMIENTO) 67
FIGURA 16: CURVA ROC (DATOS PRUEBA) 68

LISTA DE TABLAS

TABLA 1: BENEFICIOS TRIBUTARIOS EN AMÉRICA LATINA 19
TABLA 2: COMPARACIÓN DE MODELOS LOGIT VS PROBIT 37
TABLA 3: TABLA DE CLASIFICACIÓN 39
TABLA 4: LISTA DE VARIABLES CLASIFICADAS POR PERIODO, TRÁMITE E INDIVIDUO 52
TABLA 5: PRINCIPALES ESTADÍSTICOS. VARIABLES CONTINUAS 53
TABLA 6: PRINCIPALES ESTADÍSTICOS. VARIABLES CATEGÓRICAS 53
TABLA 7: DATOS GENERALES PROCESO DEVOLUCIÓN IVA AÑOS FISCAL 2015 54
TABLA 8: CARACTERÍSTICAS DATOS DE ENTRENAMIENTO 56
TABLA 9: CARACTERÍSTICAS DATOS DE PRUEBA 56
TABLA 10: CLASIFICACIÓN DE TRÁMITES SEGÚN VARIABLE OBJETIVO "RIESGO" 58
TABLA 11: TABLA DE CLASIFICACIÓN 59
TABLA 12: ESTADÍSTICOS BONDAD DE AJUSTE 59
TABLA 13: RESULTADOS DEL MODELO ESTIMADO 60
TABLA 14: EFECTOS MARGINALES 63
TABLA 15: CARACTERÍSTICAS DE UN TRÁMITE PROMEDIO 63
TABLA 16: TABLA DE CLASIFICACIÓN - DATOS DE ENTRENAMIENTO 66
TABLA 17: TABLA DE CLASIFICACIÓN - DATOS DE PRUEBA 66

LISTA DE ANEXOS

ANEXO A: MODELO SOLICITUD IVA PERSONAS ADULTAS MAYORES..... 78
ANEXO B: DEFINICIÓN DE VARIABLES 80
ANEXO C: ANÁLISIS EXPLORATORIO..... 81
ANEXO D: TRÁMITES DUPLICADOS DE UN MISMO CONTRIBUYENTE..... 88
ANEXO E: CONTRASTE DEL MODELO CON LA VARIABLE "MONTO SOLICITADO" VS "MONTO SOLICITADO EN LOGARITMO NATURAL"..... 89
ANEXO F: VARIABLES QUE NO ESTÁN EN LA ECUACIÓN 91

RESUMEN

El presente proyecto de investigación tiene como objetivo, evaluar el riesgo tributario asociado a la devolución del IVA a personas adultas mayores, a través de técnicas predictivas. Parte de una revisión de la intervención del Estado como ente regulador del mercado; profundiza el estudio de la política y gestión tributaria, con el fin de conocer la estructura de la Administración, y el proceso de gestión de riesgos. Luego, mediante un análisis comparativo de los principales proyectos tributarios enfocados a adultos mayores a nivel de América Latina, se examina a fondo el proceso de devolución del IVA a personas adultas mayores en Ecuador, para identificar las principales causas del riesgo tributario. Un amplio marco metodológico brinda las bases para desarrollar un modelo de elección binaria, que posee un alto poder de clasificación de los trámites de devolución, y determina las principales variables que influyen en la probabilidad de que un trámite sea clasificado en el grupo de riesgo. La parte final del trabajo, muestra las variables y su grado de influencia al momento de catalogar un trámite como riesgoso; información que constituye una herramienta objetiva para que el SRI identifique contribuyentes con comportamiento fraudulento, y tome acciones oportunas.

Palabras clave: Riesgo tributario, IVA, Minería de datos, Política tributaria, Gestión tributaria, Administración tributaria, Tercera edad, Devolución IVA, Impuestos, Modelo logit, Adulto Mayor, CRISP-DM, Gestión de riesgos.

ABSTRACT

The present research project aims to evaluate the tax risk associated with the value added tax (VAT) refund for older adults, through predictive techniques. It begins with a review of the State intervention as market regulator, and deepens the study of tax policy and management, in order to know the structure of the Administration, and the process of risk management. Then, through a comparative analysis of the main tax projects focused on older adults in Latin America, the process of refunding VAT to older adults in Ecuador is examined in depth to identify the main causes of tax risk. The methodological framework provides the basis for developing a binary choice model, which has a high classification power, and determines the main variables that influence the probability of a process being classified in the risk group. The final part of the paper, shows the variables and their level of influence in the moment of cataloging a process as risky; information that constitutes an objective tool for the SRI to identify taxpayers with fraudulent behavior, and take timely action.

Keywords: Tax risk, Value Added Tax, Data Mining, Tax Policy, Tax Management, Tax Administration, Old age, Return VAT, Taxes, Logit model, CRISP-DM, Risk management.

INTRODUCCIÓN

Stiglitz (2000) menciona que el Estado se encuentra presente en la mayoría de actividades que realizan los individuos dentro de la sociedad, por lo que adopta mecanismos para intervenir, controlar y regular la actividad económica; mismos que permitan generar un equilibrio entre el sector público y privado, para mitigar de manera oportuna los fallos de mercado. En este sentido, la política y gestión tributaria, brindan al Estado los instrumentos necesarios para realizar dichas actividades en el marco de los impuestos.

La política tributaria al ser parte de la política fiscal, según el Centro Interamericano de Administraciones Tributarias CIAT (2016) tiene como principales objetivos: realizar una adecuada distribución de la carga impositiva, y analizar los efectos derivados de la aplicación de tributos dentro del sistema económico. Para ello se apoya en la gestión tributaria, concebida como un conjunto de medidas, que motiven y coercionen a los contribuyentes al cumplimiento eficiente de sus obligaciones tributarias; además de propiciar la recaudación de ingresos fiscales, que permitan al Gobierno cumplir con sus objetivos.

Bajo este contexto, en el marco internacional, se distinguen dos grupos de países de acuerdo a su estructura impositiva. Por un lado, se encuentran los países que poseen un mayor porcentaje de ingresos provenientes de impuestos indirectos, como América Latina y el Caribe, siendo cerca del 50% de los ingresos fiscales correspondientes a impuestos sobre el consumo para el año 2014. Y, por otro lado, se encuentran los países miembros de la Organización para la Cooperación y Desarrollo (OCDE), que recaudan mayores ingresos provenientes de impuestos directos, representando éstos el 32,83% del total de ingresos fiscales para el mismo año (OCDE, et al., 2017, p. 70). Sin embargo, el hecho que en América Latina los sistemas tributarios estén enfocados en la recaudación de tributos indirectos, y al ser el IVA uno de los que mayor recaudación genera, puede ser un síntoma de que el sistema tributario sea regresivo[1], por lo que el

[1] El IVA es regresivo bajo la perspectiva del ingreso, puesto que los estratos de ingresos bajos destinan un porcentaje mayor de sus ingresos para cubrir dicho impuesto, a comparación de los estratos de ingresos altos que destinan un porcentaje menor. Sin embargo el IVA es progresivo si se lo analiza desde el consumo.

Estado busca propuestas tributarias para ayudar a los grupos de atención prioritaria, y mitigar dicha regresividad.

En Ecuador, el Estado busca consolidar un sistema tributario progresivo, lo que conlleva a brindar apoyo a grupos vulnerables mediante programas como la devolución del IVA a personas adultas mayores[2]. Este beneficio tributario se encuentra fuertemente argumentado en la Constitución del Ecuador, la Ley del Anciano y la Ley de Régimen Tributario Interno.

La devolución del IVA a personas adultas mayores, además de contribuir a mitigar la regresividad de los sistemas tributarios, mejora la calidad de vida de las personas de la tercera edad, e incentiva la cultura tributaria tanto en los proveedores como en los consumidores, al momento de emitir y solicitar la elaboración de facturas.

Sin embargo, en los últimos años se ha presentado un incremento significativo del número de beneficiarios que solicitan dicha devolución, llegando al punto de casi triplicarse en el periodo 2012-2015, mientras que el monto devuelto para el mismo periodo se sextuplicó. Es decir, para el año 2012 existieron casi 30 mil beneficiarios, mientras que para el año 2015 llegaron a ser 84 mil. Este incremento en el número de beneficiarios provocó que el monto devuelto pase de 9 millones para el 2012, a más de 61 millones para el año 2015 (SRI, 2016).

Estos cambios significativos tanto en el número de beneficiarios como en el monto devuelto, responden en parte al mandato presidencial dictado en el año 2015, donde se estableció que la acreditación de la devolución del IVA debe efectuarse en un plazo no mayor a 48 horas; hecho que obligó a la Administración Tributaria a cambiar de enfoque, y simplificar el proceso de devolución del IVA. Se pasó de un proceso con una extensa y minuciosa etapa de verificación de la información presentada por parte del contribuyente; a un proceso donde se acredita la cantidad de dinero que solicita el adulto mayor, sin que los expertos tributarios realicen ninguna verificación previa de la documentación presentada como justificativo para la devolución; al mismo

[2] Según el artículo 36 de la Constitución del Ecuador se considera como Adulto Mayor a partir de los 65 años de edad.

tiempo se tomaron medidas para agilitar aún más el proceso, como la realización del trámite vía internet, o por terceras personas.

Este cambio lleva consigo algunos riesgos tributarios implícitos, por ejemplo, devolver un monto que no se encuentre debidamente justificado, o que no corresponda a bienes y servicios de primera necesidad. Así también, el hecho de que el trámite sea realizado por internet, no permite que se verifique la veracidad de la información presentada como justificación de los valores solicitados. En la práctica, resulta casi imposible realizar un control y evaluación de cada uno de los documentos presentados por todos los solicitantes, y por lo tanto llegar a determinar el nivel de riesgo que existe en cada trámite presentado.

Con estos antecedentes y dada la importancia de la devolución del IVA a personas adultas mayores dentro del sistema tributario ecuatoriano, es necesario la construcción de un modelo que, a través de herramientas de minería de datos permita identificar niveles de riesgo para cada trámite, y, de ser el caso, alerte a la Administración Tributaria a realizar un seguimiento posterior adecuado, con el fin de conocer si toda la documentación presentada es válida y amerita efectivamente una devolución.

Por otra parte, las técnicas de minería de datos, según Gorunescu (2011), son herramientas que conjugan la estadística, la inteligencia artificial y los sistemas de bases de datos, para identificar patrones y extraer conocimiento útil de grandes bases de datos, que apoye en la toma de decisiones. En este estudio, dichas técnicas permiten conocer las características que definen a los trámites riesgosos, con el fin de identificarlos y, posteriormente tomar las medidas correctivas necesarias. Para ello, se utiliza la metodología CRISP-DM, que permitirá construir, basado en buenas prácticas, un modelo supervisado para clasificar los trámites de acuerdo a su nivel de riesgo, y gestionarlo en el contexto del riesgo tributario.

El presente trabajo tiene como objetivo principal evaluar el riesgo tributario asociado a la devolución del IVA a personas adultas mayores. Para lograr esto, se tiene como objetivos secundarios: analizar la teoría sobre gestión y política tributaria, estudiar el proceso de devolución del IVA a personas adultas mayores en el caso de Ecuador, y finalmente proponer un modelo que, mediante la

metodología CRISP-DM, permita clasificar los trámites en dos grupos, dependiendo de su valoración de riesgo.

Se busca además comprobar tres hipótesis: primero, el riesgo tributario es mayor cuando el monto solicitado es cercano al máximo permitido; segundo, el riesgo tributario es mayor cuando el trámite posee un alto número de facturas; y tercero, el riesgo tributario es mayor cuando el solicitante posee RUC activo.

1 FUNDAMENTOS TEÓRICOS DE LA GESTIÓN Y POLÍTICA TRIBUTARIA

1.1 INTERVENCIÓN DEL ESTADO COMO ENTE REGULADOR DEL MERCADO

Según Stiglitz (2000), el Estado es un conjunto de instituciones caracterizadas por poseer legitimidad y derecho de coerción[3]. Éste desarrolla diversas actividades en áreas como: salud, educación, consumo, recaudación de impuestos; entre otras, efectuando una intervención que propicie un adecuado funcionamiento del mercado. Para ello, desempeña tres principales funciones: asignación, distribución y estabilización.

La *función de asignación* analiza la dotación, provisión y producción de bienes públicos[4] por parte del Estado. La *función de distribución* busca la manera de repartir la riqueza o renta entre los miembros de la sociedad. Por su parte, la *función de estabilización* garantiza que se mantengan condiciones macroeconómicas estables, como es un nivel adecuado de empleo y un continuo crecimiento económico (Musgrave & Musgrave, 1992, pp. 6-13).

Cada una de dichas funciones proporcionan las herramientas necesarias al Estado, para intervenir y corregir las fallas de mercado[5] que pudieran presentarse; sin embargo, el grado de tal intervención depende del tipo y la magnitud del fallo, así como de la capacidad del Estado para mitigarlo; tomando en cuenta que él mismo posee información limitada y desconoce efectos secundarios que pudieran derivarse de la aplicación de determinada política. Además, posee un control limitado de la burocracia, ya que, al delegar

---

[3]La legitimidad denota la existencia directa o indirecta de un proceso electoral para la designación de las autoridades encargadas de dirigir las distintas instituciones públicas. Mientras que el derecho de coerción faculta al Estado a imponer algún tipo de sanción sobre otras instituciones (Stiglitz , 2000, p. 21).

[4] Los bienes públicos se caracterizan por ser no excluyentes y no rivales, es decir, no se puede exigir un pago por ellos; y además, los beneficios de consumir dichos bienes no se limitan al consumidor en concreto, sino que están disponibles para otros consumidores al mismo tiempo (Stiglitz , 2000, p. 150).

[5] Los fallos de mercado hacen referencia a situaciones generalmente negativas, ocasionadas porque los mercados no son eficientes en el sentido de Pareto; entre ellos constan las externalidades, los bienes públicos, monopolios, etc. (Stiglitz , 2000, p. 92).

determinadas funciones, depende de la eficiencia con que las instituciones encargadas actúen. Por tanto, se busca propiciar un equilibrio entre ambos sectores, dando lugar a una mutua cooperación (Stiglitz , 2000, pp. 17-19).

Una manera de lograr dicha intervención, es a través de la política tributaria, que, conjugada con una adecuada gestión tributaria, ofrece al Estado la posibilidad de ejercer control y regulación de la actividad económica, entre otras cosas.

1.2 POLÍTICA TRIBUTARIA

La política tributaria puede definirse como un conjunto de "medidas que dicen relación con el establecimiento de distintas categorías y volúmenes de recaudación de impuestos, de acuerdo con los objetivos de la política fiscal[6]" (Gallo Escobar & Cuartas Mejia , 2006, p. 369).

De acuerdo con Alink y Kommer (2011), la política tributaria analiza la carga impositiva destinada a cada contribuyente, así como los ingresos tributarios totales a ser recaudados; no únicamente se enfoca en establecer y aumentar el número de impuestos ya existentes, sino que constituye una herramienta para una adecuada distribución de ingresos y riqueza dentro de la sociedad.

De esta manera, la política tributaria tiene como principales objetivos: estructurar un adecuado sistema impositivo, analizar la distribución de la carga impositiva, así como los efectos que puedan causan los diferentes tributos sobre los agentes económicos, y finalmente, brindar herramientas útiles al Estado para influir en el proceso de desarrollo (CIAT, Diplomado en Tributación, 2016).

Un pilar fundamental dentro de la política tributaria, es contar con una adecuada estructura del sistema tributario, mismo que según Stiglitz (2000), debe caracterizarse por poseer:

---

[6] Política fiscal: Medidas implementadas por el gobierno, tendientes a orientar la economía hacia el logro de ciertas metas. Las herramientas fundamentales con las que cuenta el gobierno para ello, son el manejo del volumen y el destino del gasto público. La política fiscal también incluye la forma de financiar los gastos del gobierno (Gallo Escobar & Cuartas Mejia , 2006, p. 369).

- **Eficiencia económica. -** demanda que el sistema tributario sea lo menos distorsionador posible, que los impuestos a implementarse causen el menor impacto dentro de la sociedad, y no influyan en la asignación de recursos, ni en el incentivo para realizar actividades económicas.

- **Sencillez administrativa. -** busca que tanto los costos del funcionamiento de la Administración Tributaria, como los costos en los cuales incurre el contribuyente al momento de cumplir con sus obligaciones, sean los más bajos posibles.

- **Flexibilidad. -** analiza la facilidad de los tributos para adaptarse a los cambios del ciclo económico; y la velocidad de creación e implementación de nuevos tributos por parte de la Administración Tributaria.

- **Responsabilidad política. -** exige transparencia en las actividades del sistema tributario, en cuanto a la creación e implementación de tributos. En otras palabras, se requiere que exista claridad sobre quién debe pagar el tributo, por qué motivo lo paga, y quién será beneficiado.

- **Justicia. -** sostiene que el sistema tributario sea lo más equitativo posible. Para ello, se basa en dos principios: Primero, el *principio de equidad vertical*, donde personas en diferentes circunstancias no deben ser tratadas por igual, es decir, los individuos deben contribuir de acuerdo a su capacidad de pago, dando lugar a los impuestos progresivos. Segundo, el *principio de equidad horizontal,* el cual plantea un mismo tratamiento entre los contribuyentes con similares características; es decir, los individuos con ingresos similares deben tener una carga impositiva similar (Stiglitz , 2000, pp. 495-496).

El cumplimiento de las características antes enunciadas, contar con un sistema tributario de base amplia que grave adecuadamente al ingreso, consumo y riqueza, bases impositivas bien definidas, justicia tributaria, entre otros aspectos, fortalecen al sistema tributario, permitiendo que el mismo contribuya a la generación de ingresos tributarios permanentes y alcance objetivos sociales, económicos y políticos (Alink & Kommer, 2011, pp. 44-50).

Según Alink y Kommer (2011), los impuestos constituyen “una tarifa involuntaria pagada por las personas o empresas al Gobierno; son cargos obligatorios amparados por la ley, impuestos a un individuo o a una entidad jurídica por el Estado”. Los tributos son considerados como pagos no correspondidos, ya que el beneficio que brinda el Estado a los contribuyentes, no guarda una relación proporcional con el monto de la carga impositiva que los mismos pagan.

Los impuestos según Stiglitz (2000), influyen en la conducta de los contribuyentes, alterando su comportamiento en ahorro, educación, consumo, trabajo, etc. Modifican la asignación de recursos, ya que alteran los precios relativos[7]. También, el hecho de que un sistema tributario este compuesto por un alto o bajo número de tributos, causa alteraciones en la economía, por ejemplo incentivando o no las inversiones nacionales o extranjeras.

Dado los efectos positivos y negativos que provocan los impuestos, Sevilla (2004) estable dos principios para determinar el tipo de impuesto que debe asumir cada contribuyente: el principio de beneficio y el principio de capacidad de pago.

El *principio de beneficio* establece que la carga impositiva que recae sobre los individuos, guarda una relación directa con los beneficios que obtienen por el uso y/o consumo de bienes públicos provistos por el Estado; mientras más beneficiados sean los individuos por las actividades públicas, mayor será la cantidad de impuestos que deben asumir. Sin embargo, este principio excluye la capacidad de desarrollar políticas redistributivas, dado que el Estado tiende a brindar apoyo a personas de bajos ingresos que, según este principio, tendrían que asumir una mayor carga impositiva. Un instrumento creado bajo este principio es el precio público, que es utilizado por el Estado cuando suministra bienes y servicios demandados voluntaria e individualmente por los ciudadanos, tal es el caso del servicio de energía eléctrica.

Por otro lado, el *principio de capacidad de pago* se encuentra fuertemente relacionado con los criterios de equidad (horizontal y vertical), redistribución y

[7] El precio relativo de un bien está expresado en términos de otro bien, y es igual a la razón de los precios absolutos entre ambos bienes. El precio absoluto de un bien es aquel expresado de unidades monetarias (Sepúlveda, 2004, p. 146).

progresividad, puesto que fija una carga impositiva en función a la capacidad de pago de los individuos, independientemente del beneficio que puedan obtener del Estado. Este principio es el que rige en mayor medida la creación y clasificación de los impuestos. Bajo este marco, existen dos grandes grupos de impuestos: los impuestos directos, que recaen sobre las personas físicas o sobre las sociedades, por ejemplo el impuesto sobre la renta; y los impuestos indirectos, dirigidos a una amplia variedad de bienes y servicios, siendo el IVA el impuesto más representativo dentro de éste grupo (Stiglitz , 2000, pp. 477 - 488).

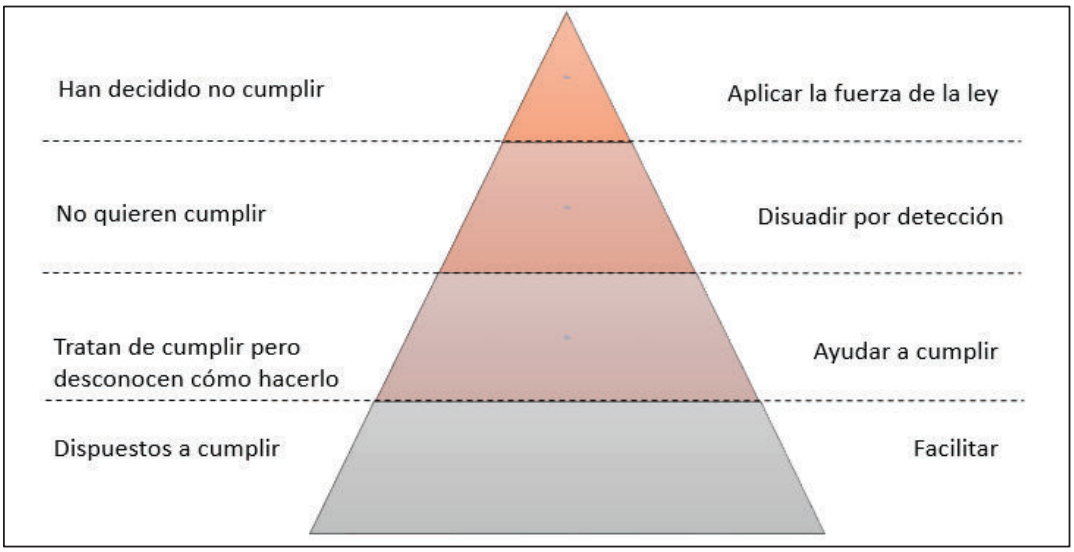
Otra manera de clasificar los diversos tipos de impuestos, es mediante la categorización entre instrumentos financieros y reguladores. Los instrumentos financieros cuyo objetivo es captar la mayor cantidad de ingresos tributarios, abarcan: tasas, índice de renta, índice de patrimonio, e índice de consumo. Mientras que los instrumentos reguladores, a más de contribuir a la recaudación, buscan generar un cambio en el comportamiento de los contribuyentes, siendo los aranceles aduaneros y tributos medio-ambientales, parte de este grupo (Sevilla Segura, 2004, pp. 57-58).

1.3 GESTIÓN TRIBUTARIA

La gestión tributaria abarca una serie de acciones para recaudar, controlar y gestionar los tributos y leyes impositivas, dictadas a través de la política tributaria. Según Alink y Kommer (2011), “implica el diseño, construcción y mantenimiento de una infraestructura que permita y facilite a los contribuyentes el cumplimiento de sus obligaciones, y les aliente a cumplir con las mismas”.

La gestión tributaria analiza el comportamiento de los contribuyentes, lo que implica la comprensión del comportamiento humano; por tanto, una forma de generalizar y agrupar el comportamiento de los contribuyentes, es a través de la pirámide de cumplimiento tributario, que establece los siguientes niveles de cumplimiento (OCDE, 2004):

Figura 1: Pirámide de cumplimiento



Fuente: OCDE, Compliance Risk Management: Managing and Improving Tax Compliance, 2004
Elaboración: Los autores

En la base de la pirámide está el grupo de contribuyentes cuyo comportamiento facilita a la Administración Tributaria su gestión, dado que cumplen adecuadamente con sus obligaciones. El siguiente nivel, agrupa a los contribuyentes que requieren un fortalecimiento de su cultura tributaria a través de la educación, asesoría y capacitación, debido a que están dispuestos a cumplir con sus obligaciones, pero se les dificulta realizarlas. En el tercer nivel se encuentran los contribuyentes que se resisten a cumplir sus obligaciones, puesto que perciben erróneamente el accionar de la Administración Tributaria, por lo que ésta debe identificarlos y emplear diversas acciones para disuadir su comportamiento. Finalmente en el pico de la pirámide se hallan los contribuyentes que por voluntad propia han decidido no contribuir; ante los cuales la Administración Tributaria debe ejercer su derecho a sancionar y tomar las medidas legales correspondientes, influyendo presencia y control fiscal (OCDE, 2004, p. 41).

Uno de los deberes de las Administraciones Tributarias, a través de su gestión, es analizar y mitigar el incumplimiento tributario. Según Trigueros, et al. (2012) el incumplimiento tributario se define como “todas aquellas acciones deliberadas o no, que resultan en una eliminación o reducción indebida de las obligaciones tributarias”. Estas acciones pueden deberse a múltiples factores, tales como: una

alta carga tributaria, poca confianza en el sistema tributario, baja cultura tributaria, etc.

El incumplimiento tributario desencadena en la medición de brechas tributarias. Según Trigueros et al. (2012) existen cuatro tipos de brechas:

- **Brecha de registro:** Resulta de comparar el universo de contribuyentes potenciales, con los contribuyentes realmente registrados.

- **Brecha de declaración:** diferencia entre el número de contribuyentes registrados (inscritos) en el sistema de la Administración Tributaria, y el número de contribuyentes que efectivamente realizan sus declaraciones.

- **Brecha de pago:** cociente entre el impuesto no pagado en los plazos establecidos y el impuesto declarado voluntariamente por parte del contribuyente.

- **Brecha de veracidad:** es el cociente entre el impuesto no declarado y el impuesto que realmente debió declararse; es decir, muestra la discrepancia en la información que el contribuyente presenta, con la información de su realidad económica.

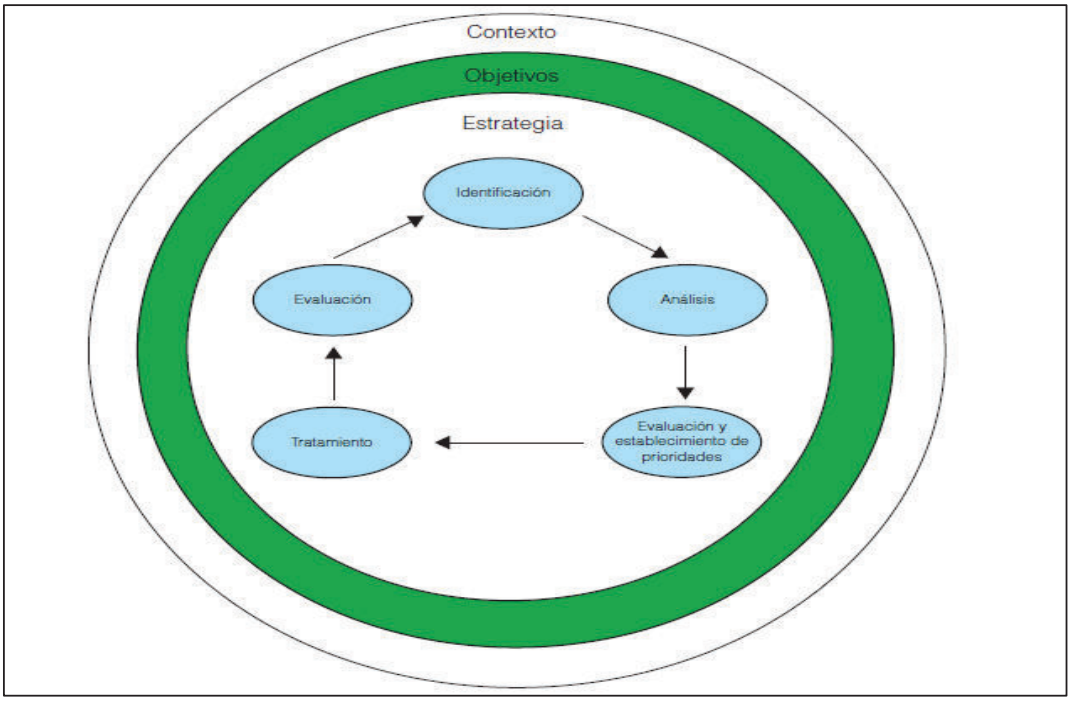
Incentivar una adecuada cultura tributaria, fortalecer la confianza de los contribuyentes ante la Administración, un adecuado sistema sancionador que transmita presencia y control tributario, brindar adecuados servicios y asistencia a los contribuyentes, mantener una oportuna actualización de la información de los contribuyentes registrados, realizar controles rápidos y masivos, efectuar una adecuada gestión de riesgo; entre otras medias, pueden contribuir a la reducción de las brechas tributarias, y aumentar los niveles de cumplimiento tributario. (Pecho Trigueros, et al., 2012, pp. 30-32).

1.3.1 GESTIÓN DE RIESGOS TRIBUTARIOS

El riesgo se define como la "amenaza o probabilidad de que una acción o evento, afecten negativamente a la capacidad de una organización para lograr sus objetivos" (European Commission, 2010). En esta línea, el riesgo tributario hace referencia a la probabilidad que tiene la Administración Tributaria de incurrir en pérdidas, o de que ocurra algún evento desfavorable, como consecuencia de un comportamiento inusual por parte de los contribuyentes, lo que ocasiona que su capacidad para cumplir plenamente sus objetivos se vea afectada. Por lo tanto, los programas de gestión de riesgo ayudan a la Administración Tributaria, a mitigar los distintos tipos de riesgos a los que se encuentre expuesta.

Un sistema de gestión de riesgos a seguir, a nivel de Administraciones Tributarias es el desarrollado por la Comisión Europea (2010). Consta de cinco etapas: identificación, análisis, evaluación, establecimiento de prioridades, tratamiento y evaluación. Tal como se muestra en la figura 2, dichas etapas o pasos son considerados como estrategias alineadas con los objetivos y el contexto de la Administración Tributaria (European Commission, 2010, p. 8).

Figura 2: Proceso de Gestión de riesgo



Fuente: European Commission, Compliance Risk Management Guide, 2010

Según la guía de gestión de riesgos de la Comisión Europea (2010), el *contexto* hace referencia a las reglas sobre las cuáles las Administraciones Tributarias desempeñan sus funciones, en el constan la legislación, las condiciones económicas, la opinión pública, la estructura misma de la Administración Tributaria, entre otras. Los o*bjetivos* indican las metas, los resultados que se esperan y/o la situación futura que desean alcanzar las Administraciones; y, las *estrategias,* por su parte, muestran el cómo alcanzar los objetivos planteados mediante un plan de acción.

Dentro de las estrategias, el primer paso es la *identificación del riesgo*; mismo que analiza eventos externos, cooperación mutua con otras instituciones, el intercambio de información con otras Administraciones Tributarias, la legislación, las auditorías al azar y las constantes investigaciones, para la detección oportuna de los riesgos (a nivel de contribuyente o de área) que puedan existir.

Una vez realizada la identificación, se procede a efectuar un *análisis de riesgos*. Para ello, es necesario agrupar los diversos tipos de riesgos y conocer su frecuencia (número de riesgos detectados en la fase anterior), la probabilidad de que dichos riesgos lleguen a cumplirse, y las posibles consecuencias que producirían (relacionadas con las pérdidas económicas para la Administración Tributaria); es decir, el objetivo principal de esta etapa es conocer a fondo las características de los riesgos determinados en el paso anterior.

En la tercera etapa, se realiza una *priorización de los riesgos*, con el objetivo de determinar el grupo de contribuyentes o el tipo específico de riesgo que se va a tratar de minimizar; para lograr esto, se realiza una evaluación y un ranking de los mismos, considerando su probabilidad y su impacto. Además, se estudia las posibles formas de tratamiento, para de esta manera elaborar un plan de acción.

La etapa de *tratamiento de riesgos* busca neutralizar o disminuir el impacto negativo de los riesgos, mediante la utilización de la transferencia, la reducción o la cobertura de riesgos. La transferencia es útil en determinados casos, pero se debe considerar que el riesgo sigue existiendo, únicamente ha pasado de una institución a otra. La reducción, busca limitar al máximo las posibilidades de cometer errores involuntarios y/o intencionales, para lo cual, la educación tributaria brindada por las Administraciones es sumamente importante. La

cobertura ante el riesgo, implica la ejecución de medidas correctivas y/o auditorias ante los contribuyentes, que de una u otra manera estuvieran realizando algún tipo de fraude.

Finalmente, como quinto y último paso, la *evaluación de riesgos* construye un círculo de aprendizaje, además de conocer si las medidas implementadas en pasos anteriores han causado el resultado. Se utilizan técnicas de evaluación de impacto, para estimar en qué grado un programa, medidas correctivas y/o políticas, ayudaron en la realización de los objetivos planteados al inicio de la gestión de riesgos; y las técnicas de evaluación de procesos, con la finalidad de saber cómo y porqué funcionó o no determinada acción.

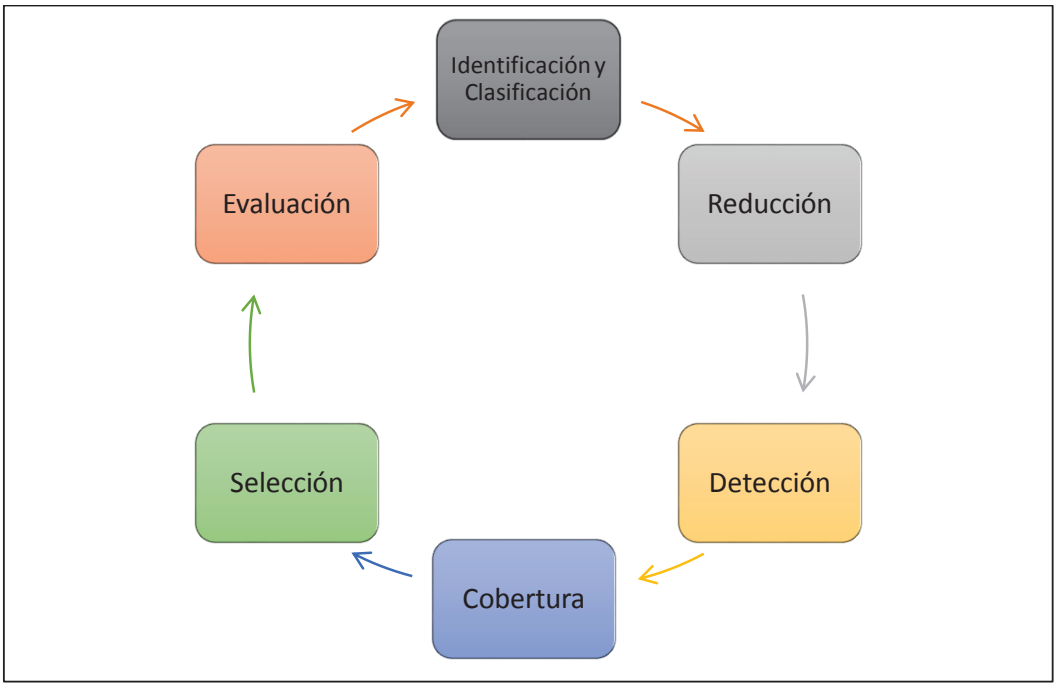
1.3.1.1 MODELO DE GESTIÓN DE RIESGOS DEL SERVICIO DE RENTAS INTERNAS

Basados en el modelo general de gestión de riesgos tributarios elaborado por la Comisión Europea (2010), el Servicio de Rentas Internas, al ser la entidad encargada de realizar una adecuada ejecución de la política y gestión tributaria, establece un modelo de gestión de riesgos acorde al sistema tributario nacional.

El proceso de gestión de riesgos inicia con la identificación y clasificación de los posibles riesgos, esta etapa abarca los tres primeros pasos (identificación, análisis y priorización) del modelo de gestión de la comisión europea. Luego, las etapas de reducción, detección[8], cobertura y selección son una segregación de la cuarta etapa (tratamiento) del modelo general de riesgos. Finalmente, el proceso concluye con la etapa de evaluación de riesgos (SRI, 2012, p. 124).

[8] La detección de riesgos puede ser proactiva (antes de que ocurra el riesgo), corriente (inmediatamente después de que ocurra el riesgo), reactiva (luego de la declaración de impuestos), activa (visita al contribuyente), o pasiva (intercambio de información con otras instituciones) (SRI, 2012, p. 125).

Figura 3: Modelo de Gestión de Riesgos del SRI



Fuente: SRI, Plan Estratégico 2012-2015, 2012
Elaboración: Los autores

2 DEVOLUCIÓN DEL IVA A PERSONAS ADULTAS MAYORES EN ECUADOR

2.1 ANÁLISIS COMPARATIVO DE PROYECTOS TRIBUTARIOS EN BENEFICIO DE LAS PERSONAS ADULTAS MAYORES, ENTRE ECUADOR Y PAÍSES DE AMÉRICA LATINA.

Los proyectos, leyes y políticas enfocados a brindar apoyo a personas adultas mayores, consideran áreas de atención como: salud, seguridad social, vivienda, transporte, incentivos tributarios, entre otros. Estos proyectos se establecen de acuerdo a las normas y principios de cada país, con el objetivo de cumplir con los derechos de dicho grupo y garantizarles un adecuado nivel de vida. En cuanto a los beneficios tributarios, que incluyen la atención prioritaria a personas adultas mayores entre los principales países de América Latina, se ha realizado una recopilación de las siguientes leyes y programas:

- **Argentina. -**

De acuerdo a la Ley N° 27.253, establecida en 13 de junio de 2016, se decreta la devolución parcial del IVA por concepto de compras de bienes muebles realizadas con tarjetas de débito, a personas que perciban jubilaciones, pensiones por fallecimiento, bonos por embarazos, y bonos relacionados con protección social. Esta normativa establece que los beneficiarios recuperarán el 15% del IVA[9], teniendo como monto máximo de devolución 300 pesos mensuales, que representan el 15% de la canasta básica (Ley 27253, 2016).

- **Bolivia. -**

La Ley 1886, creada el 14 de Agosto de 1998, establece un régimen de privilegios y descuentos enfocados a las personas adultas mayores[10]. Dicho

[9] En Argentina, el IVA corresponde al 21%
[10] En Bolivia, se considera como personas adultas mayores a partir de los 60 años.

régimen considera diversas deducciones en cuanto a transporte, salud y servicios públicos; además se plantea que tendrán una rebaja del 20% en el impuesto anual sobre propiedades de bienes inmuebles de interés social (Ley 1886, 1998).

- **Chile. -**

La ley 20.732, creada el 05 de Marzo del 2014, plantea la rebaja del impuesto territorial correspondiente a propiedades de adultos mayores vulnerables económicamente, mediante la imposición de una cuota máxima de pago (Ley 20732, 2014).

- **Colombia. -**

Dado que Colombia, posee una gran población de personas adultas mayores, aproximadamente 5 millones, de los cuales 3,5 millones no poseen un sistema de apoyo económico para su vejez, (Fundación para la Educación Superior y el Desarrollo; Fundación Saldarriaga Concha, 2015, p. 95) se establecen las siguientes leyes y programas:

La ley 155 del 2010, conocida como “Ley último empleo”, incentiva la inclusión de personas adultas mayores al mercado laboral de manera indirecta, puesto que otorga exenciones (disminución del 12% en el impuesto a la renta) a quienes contraten personas mayores de 50 años (Ley 155, 2010).

El programa “Colombia Mayor” conocido anteriormente como “Programa de Protección Social a los Adultos Mayores”, brinda un subsidio económico para personas adultas mayores en estado de vulnerabilidad, desamparo o en extrema pobreza. Dicho programa se encuentra a cargo del Ministerio de Trabajo y se accede al mismo cada dos meses (Consorcio Colombia Mayor, 2016) .

Y, el programa “Beneficios Económicos Periódicos” creado bajo la Ley 1328 del 2009, en su artículo 87, establece que el principal objetivo de dicho programa es incentivar a la realización de un ahorro voluntario para la vejez (Ley 1328, 2009).

- **Perú. -**

La ley de tributación municipal en su decreto legislativo N° 776, en su artículo 19, establece que las personas adultas mayores, tendrán una disminución de 50 UIT (unidad impositiva tributaria) en el pago del impuesto predial (Decreto Legislativo N° 776, 1994).

Perú, al igual que Colombia, ofrece algunos beneficios económicos. Por ejemplo, el programa “Gratitud” que brinda un subsidio de 100 soles mensuales a personas adultas mayores de 75 años que se encuentran en extrema pobreza (Decreto N°050-2010, 2010).

Y finalmente el programa “Pensión 65” que otorga subvenciones a adultos mayores de 65 años; este programa busca atender a los sectores más pobres de Perú (Decreto Supremo N° 081-2011-PCM, 2011) .

- **México. -**

México a través del art. 186 de la ley de impuesto sobre la renta, ofrece un beneficio indirecto a las personas adultas mayores, ya que incentiva la contratación de dicho grupo mediante un estímulo fiscal a las empresas, correspondiente al 25% del salario efectivamente pagado a las personas mayores de 65 años (Ley del Impuesto sobre la Renta, 2016).

En resumen, se tiene:

Tabla 1: Beneficios tributarios en América Latina

PAÍS	LEY	FECHA	NOMBRE	DESCRIPCIÓN	BENEFICIARIOS
Ecuador	LORTI	2004	LORTI	Devolución del IVA	Personas adultas mayores de 65 años.
	Ley del Anciano	2006	Ley del Anciano Codificación 2006-2007	Exoneración tributarias	
Argentina	N° 27253	2016	Régimen de reintegro por compras en Comercios de venta Minorista.	Devolución parcial del IVA	Personas que reciban: Jubilaciones, Pensiones, Bonos
Bolivia	N° 1886	1998	S/N	Descuentos en salud, servicios públicos y transporte etc.	Adultos Mayores (a partir de 60 años)
Chile	N° 20732	2014	Rebaja de Impuesto territorial a propiedades de adultos mayores vulnerables económicamente	Rebaja impuesto territorial	Adultos mayores vulnerables económicamente
Colombia	N° 155	2010	Ley último empleo	Exención tributaria a empleadores de personas adultas mayores	Personas mayores a 50 años

	Contrato de Fiducia Pública N° 216	2013	Programa Colombia Mayor	Subsidio económico	Adultos mayores en estado de vulnerabilidad, desamparo o extrema pobreza
	N°1328	2009	Beneficios económicos periódicos	Incentivo a ahorro voluntario	• Adultos mayores • Trabajadores de sector informal
Perú	Decreto legislativo N°776	1994	Ley de tributación municipal	Descuento en el pago del impuesto predial	Adultos mayores
		2010	Programa Gratitud	Subsidio mensual	Adultos mayores (a partir de 75 años)
	Decreto Supremo N° 081-2011-PCM	2011	Pensión 65	Subvención	Adultos mayores (a partir de 65 años)
México	Ley Impuesto sobre Renta.	2016	Ley del Impuesto sobre la renta	Estímulo fiscal a empresa que contraten personas adultas mayores	Adultos mayores (a partir de 65 años)

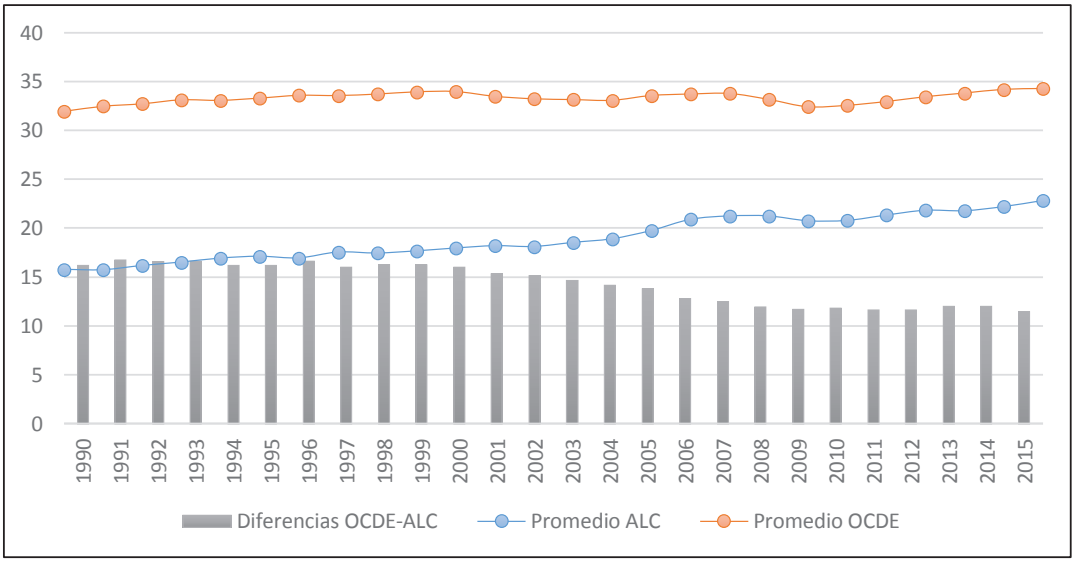
Elaboración: Los autores

2.2 ANÁLISIS DE LA DEVOLUCIÓN DEL IVA A PERSONAS ADULTAS MAYORES

2.2.1 ALGUNAS ESTADISTICAS INTERNACIONALES

A nivel de América Latina, los avances en tributación han sido significativos. El nivel de ingresos tributarios como porcentaje del PIB ha aumento paulatinamente desde el año 1990 hasta la actualidad, respondiendo a "condiciones macroeconómicas favorables, cambios en el diseño de los sistemas tributarios y al fortalecimiento de las Administraciones Tributarias" (OCDE, et al., 2016, p. 40). Pese a que la recaudación tributaria es menor en Latinoamérica (ALC) en comparación con los países pertenecientes a la Organización para la Cooperación y el Desarrollo Económicos (OCDE), en la figura 4, se observa que desde del año 2000, la brecha comienza a disminuir gracias a las nuevas políticas de recaudación implantadas en cada país, el uso de nuevos sistemas tecnológicos que facilitan la recaudación y su control, además de brindar capacitación continua tanto a los funcionarios como a los contribuyentes. (OCDE, et al., 2016, p. 44).

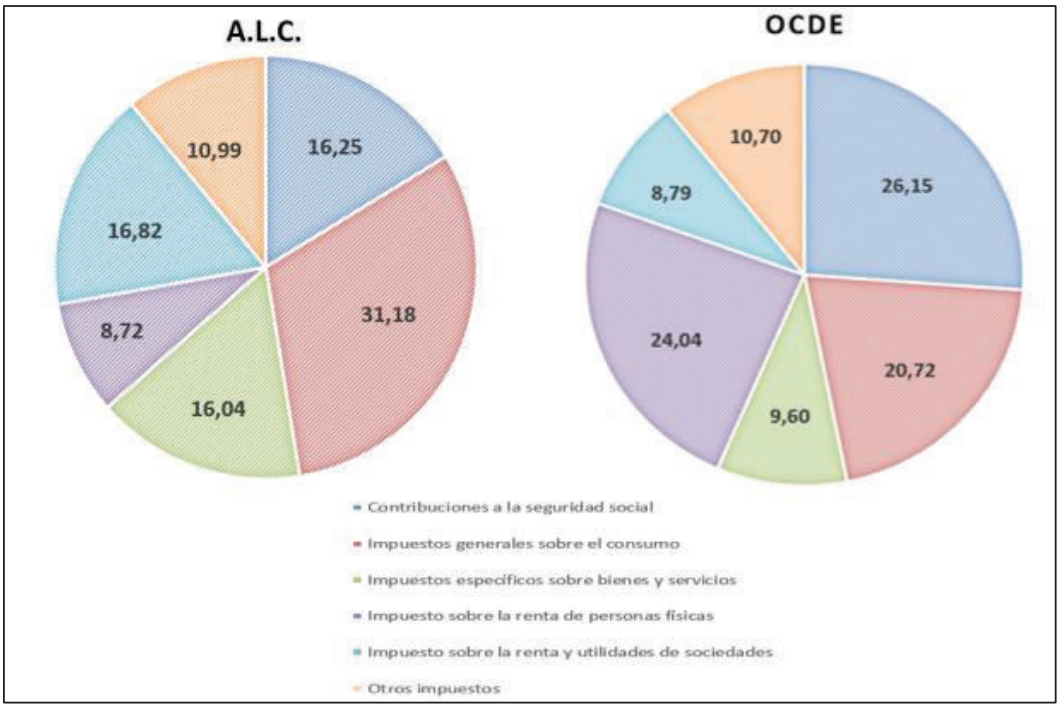
Figura 4: Total de ingresos tributarios como porcentaje del PIB



Fuente: CIAT, Estadísticas tributarias en América Latina y el Caribe 1990-2015, 2017
Elaboración: Los autores

En América Latina, el mayor porcentaje de recaudación proviene de impuestos indirectos, siendo los impuestos al consumo los más representativos, con un porcentaje de 47,22% del total de los ingresos fiscales para el año 2014, y representando un 10,7% del PIB para el mismo año (OCDE, et al., 2017, p. 57); en comparación con los países pertenecientes a la OCDE, donde los impuestos directos (impuesto a la renta y utilidades) poseen un mayor peso, representando el 32,83% del total de ingresos fiscales (Figura 5).

Figura 5: Estructuras impositivas en América Latina y OCDE, 2014



Fuente: CIAT, Estadísticas tributarias en América Latina y el Caribe 1990-2015.
Elaboración: Los autores

Para el caso de Ecuador, la recaudación tributaria como porcentaje del PIB ha ido evolucionando de 7,12% en el año 1999 a 12,8% en el año 2016. Los impuestos indirectos representan el 49% del total recaudado para el mismo año; y, únicamente el IVA representa el 34,8% del total recaudado, lo que evidencia la importancia de éste impuesto dentro del sistema tributario (SRI, 2017).

Bajo este escenario, y dada la importancia del IVA dentro de los sistemas tributarios en América Latina; Corbacho, Fretes y Lora (2012) mencionan que es necesario analizar su regresividad, ya que este tributo posee dicha característica

si se lo analiza desde el ingreso, puesto que, al ser una tarifa única que no distingue entre los distintos niveles de ingresos de los individuos, los estratos de ingresos bajos destinan un porcentaje mayor de su renta para cubrir éste impuesto, a diferencia de los estratos más altos, que en relación con su nivel de ingreso, el porcentaje que destinan para el pago del IVA es menor. Sin embargo, dicha regresividad disminuye notablemente, si se realiza un análisis desde el consumo, debido a que existe una relación directa entre el monto pagado de IVA, y el volumen de bienes consumidos por los individuos (Corbacho, et al., 2012).

Por tal motivo, las Administraciones Tributarias adoptan medidas[11] cuyo objetivo es mitigar la regresividad del impuesto, de tal manera que exista apoyo a los sectores de menos ingresos.

2.2.2 CONTEXTO LEGAL EN ECUADOR

En Ecuador, uno de los programas creados para atenuar la regresividad del IVA, y apoyar a grupos vulnerables, específicamente a las personas adultas mayores, es la devolución del IVA. Este proceso empezó a implementarse a partir del 2002 a través del Servicio de Rentas Internas (SRI) y se encuentra plenamente justificado en la legislación ecuatoriana.

La Constitución del Ecuador, para garantizar los derechos de las personas y grupos vulnerables, establece en su artículo 36, que “las personas adultas mayores recibirán atención prioritaria y especializada en los ámbitos público y privado, en especial en los campos de inclusión social y económica,..., considerando personas adultas mayores a aquellas que hayan cumplido los sesenta y cinco años de edad” (Constitución del Ecuador, 2008).

De igual manera en su artículo 37, se mencionan los derechos que el Estado garantizará a las personas adultas mayores, entre los que consta, el derecho a exenciones en el régimen tributario. Éste último artículo, se vincula fuertemente con la Ley del Anciano que, en su artículo 14 exonera del pago de impuestos fiscales y municipales a los adultos mayores (Ley del Anciano, 2006).

---

[11] Entra las medidas que se aplican para reducir la regresividad del IVA constan las exenciones y la fijación de impuestos reducidos para cierto tipo de bienes (Corbacho, et al., 2012, p. 373).

Así mismo, la Ley Orgánica de Régimen Tributario Interno, en su artículo innumerado posterior al artículo 74, estable las normas para la devolución del IVA a personas adultas mayores, de la siguiente manera:

> **IVA pagado por personas adultas mayores**. - Las personas adultas mayores tendrán derecho a la devolución del IVA pagado en la adquisición de bienes y servicios de primera necesidad de uso o consumo personal. La base imponible máxima de consumo mensual a la que se aplicará el valor a devolver será de hasta cinco remuneraciones básicas unificadas del trabajador, vigentes al 1 de enero del año en que se efectuó la adquisición (LORTI, 2014).

Al referirse a "bienes y servicios de primera necesidad", el Reglamento para la Aplicación de la Ley de Régimen Tributario Interno, en su artículo innumerado posterior al artículo 181, considera a aquellos bienes y servicios relacionados con vestimenta, vivienda, salud, comunicación, alimentación, educación, cultura, deporte, movilidad y transporte; y excluye a aquellos relacionados con actividades comerciales (Reglamento LORTI, 2014).

La devolución del IVA debe cumplir además con los principios que rigen el código tributario (legalidad, generalidad, igualdad, proporcionalidad e irretroactividad), con el fin de lograr una recaudación eficiente y un buen manejo de los ingresos fiscales (Código Tributario, 2005).

2.2.3 NORMAS PARA SOLICITAR LA DEVOLUCIÓN DEL IVA A PERSONAS ADULTAS MAYORES

Al ser el SRI la entidad encargada de administrar todo el proceso de devolución del IVA, según la resolución N° NAC-DGERCGC13-00636, se establecen las siguientes normas para el desarrollo del proceso (Resolución, 2013):

- Para acceder al derecho, el beneficiario debe haber cumplido sesenta y cinco años de edad al momento de producirse la adquisición de bienes o servicios de uso o consumo personal y de primera necesidad.

- El beneficiario puede realizar una solicitud mensual o anual de devolución. Así mismo, puede presentar varias solicitudes mensuales dentro de un

mismo trámite, siempre y cuando las justifique debidamente. No se permite realizar nuevas peticiones por el mismo periodo[12].

- El monto máximo mensual a devolverse para el año fiscal 2015, es el correspondiente al Impuesto al Valor Agregado y/o al Impuesto a los Consumos Especiales, calculados sobre una única base imponible máxima de consumo de hasta cinco[13] remuneraciones básicas unificadas del trabajador, vigentes al primero de enero del año en que se efectuó la adquisición. Dichos valores deberán sustentarse en comprobantes de venta válidos (facturas) al adquirir o importar bienes o servicios, gravados con uno solo de estos impuestos o con ambos.

$$\text{Monto máximo permitido (IVA 12\%)} = RMU \times IVA \times 5$$
$$= \$\,354 \times 12\% \times 5 \;= \$\,212{,}40$$

$$\text{Monto máximo permitido (ICE 15\%)} = RMU \times ICE \times 5$$
$$= \$\,354 \times 15\% \times 5 \;= \$\,265{,}50$$

- El SRI emitirá una resolución al solicitante respecto al proceso de devolución, para luego realizar la acreditación respectiva.

- La solicitud de devolución podrá realizarse vía internet a través del portal web del SRI, o de manera personal en cualquiera de sus oficinas.

- El SRI está plenamente facultado para realizar un proceso de control posterior de los montos devueltos entregados a los solicitantes.

- El solicitante será responsable por toda la información presentada ante la Administración Tributaria.

---

[12] El SRI considera como periodo, al mes del cual se solicita la devolución.

[13] El monto de la base imponible correspondiente a cinco remuneraciones básicas, estuvo vigente hasta abril de 2016; a partir de dicha fecha, la base imponible se modificó a dos remuneraciones básicas unificadas y no considera la devolución de impuestos a consumos especiales (ICE) (SRI, 2016).

2.2.4 PROCESO DE DEVOLUCIÓN DEL IVA

El proceso inicia con la solicitud[14] realizada por parte del adulto mayor, misma que al ser presentada por primera vez[15], obligatoriamente debe ser entregada de manera personal en las oficinas del SRI. Una vez que el beneficiario entrega la solicitud, la misma ingresa al sistema de devoluciones, y se emite una resolución al adulto mayor indicándole que su solicitud será validada.

Cada una de las solicitudes pasa por un proceso de control previo, donde expertos del SRI, verifican la veracidad de la información presentada. Si los productos sobre los cuales se solicita la devolución del IVA no corresponden a bienes de uso y consumo personal y de primera necesidad, se descuenta dicho valor del monto total solicitado por el adulto mayor. Esta etapa del proceso es en parte subjetiva, ya que depende en gran medida del criterio del funcionario para identificar y aprobar o no, los productos sobre los cuales se devolverá el IVA. Es en esta etapa que el proceso se extiende, puesto que se revisa de manera detallada todas y cada una de las facturas[16] que presenta el adulto mayor.

Finalmente, una vez que se ha verificado y determinado el monto a devolverse, se realiza la acreditación al solicitante en su cuenta bancaria registrada.

Por tanto, el proceso puede tardar varios meses, desde que el adulto mayor presenta la solicitud, hasta que el valor solicitado es acreditado en su cuenta bancaria.

Sin embargo, al ser dicho grupo uno de atención prioritaria, por mandato presidencial establecido el año 2015, se aceleró el proceso de devolución de IVA, con el fin de que los valores solicitados por los adultos mayores, estén acreditados en su cuenta en máximo 48 horas laborables. Dicha situación obligó

---

[14] La solicitud del IVA, debe cumplir con los requisitos expuestos en el artículo 178 del Reglamento para la aplicación de la Ley de Régimen Tributario Interno (ver anexo A).

[15] A partir de la segunda vez que el adulto mayor solicita la devolución del IVA, el proceso se lo puede realizar mediante internet o ventanilla.

[16] Se aplica la "prescripción" para las facturas que presente el adulto mayor; es decir que, si han transcurrido más de 5 años desde la fecha de emisión del comprobante de venta o factura, el valor del IVA correspondiente, ya no será devuelto (SRI, 2016).

a que el SRI realice un cambio de enfoque y se eliminen ciertas etapas del proceso. Debido a esto, actualmente el programa consta de tres pasos:

- El beneficiario solicita la devolución del IVA ya sea por ventanilla o por internet, justificando dicho valor con los comprobantes de venta o facturas correspondientes.
- Se ingresa la información al sistema de devoluciones del SRI.
- Finalmente, se acreditan dichos valores a la cuenta bancaria del adulto mayor.

Resulta evidente que al suprimir el control previo, el proceso se torna corto, sin embargo lleva inmersas algunas desventajas, por lo que surge la necesidad de analizar el riesgo tributario implícito en la devolución del IVA.

2.3 RIESGO TRIBUTARIO EN LA DEVOLUCIÓN DEL IVA A PERSONAS ADULTAS MAYORES

Al eliminar el control previo en el proceso de devolución del IVA, el SRI se encuentra expuesto a algunos riesgos. Primero, el hecho de que se devuelva al adulto mayor todo el dinero que solicita sin ningún tipo de verificación previa, provoca que la Administración Tributaria destine cada vez más ingresos a cubrir dicho programa; además de que se propicie un entorno de baja percepción de control tributario.

Segundo, el hecho de que el proceso de devolución del IVA se realice por internet, implica otro riesgo, ya que no se verifica realmente que las facturas ingresadas como justificativo del monto solicitado, sean válidas, respondan a bienes de primera necesidad, no se encuentren relacionadas con actividades comerciales del solicitante, pertenezcan al adulto mayor, entre otras características.

Tercero, Stiglitz hace referencia a que el comportamiento de los individuos tiende a la maximización de su utilidad; basados en esta idea, la Administración Tributaria debe tener presente que los contribuyentes pueden presentar cierto

tipo de comportamiento que los lleve a obtener un mayor provecho económico individual.

Una última desventaja se refiere a la capacidad operativa del SRI, puesto que en la práctica resulta casi imposible realizar un control y evaluación de los documentos presentados por todos y cada uno de los solicitantes de la devolución. Por lo tanto, se complica el hecho de llegar a determinar el nivel de riesgo presente en cada trámite, lo que da como resultado que existan inconsistencias entre el monto de devolución que exige el solicitante y el monto real que se le debería devolver.

Para mitigar dichas desventajas, el Reglamento para la Aplicación de la Ley de Régimen Tributario Interno, en su artículo 181, faculta plenamente al SRI para verificar la concordancia entre los valores devueltos al adulto mayor, y los ingresos que el mismo percibe; dado el caso que se encuentren inconsistencias o anomalías en el proceso de devolución, el SRI es el encargado de tomar y ejecutar las acciones correspondientes.

3 MINERÍA DE DATOS Y PRINCIPALES HERRAMIENTAS

3.1 CONCEPTUALIZACIÓN DE LA MINERÍA DE DATOS

Según Gorunescu (2011), la minería de datos es un área de estudio multidisciplinario que combina tres ramas: la estadística, la inteligencia artificial y los sistemas de bases de datos. Su propósito es reconocer, buscar, identificar y extraer información implícita u oculta en grandes bases de datos, que pueda ser modelada, y permita obtener el conocimiento requerido.

La estadística es la principal rama sobre la que se fundamenta la minería de datos, puesto que mediante el análisis exploratorio, permite identificar las relaciones entre las variables utilizando técnicas de visualización[17] y métodos computacionales[18]. Por otro lado, la inteligencia artificial basada en la heurística, tiene por objetivo diseñar programas de procesamiento de información, que se basen en el conocimiento humano a través del aprendizaje automático[19], es decir, busca incorporar la experiencia de los investigadores en el desarrollo y mejoramiento de distintos programas computacionales. Finalmente, los sistemas de bases de datos proporcionan la información a ser estudiada y sobre la cual se aplicarán las herramientas tanto de la estadística como de la inteligencia artificial (Gorunescu, 2011, pp. 1-4).

Los datos por sí solos no proporcionan ningún tipo de información útil o ayuda directa para la toma de decisiones, ya que su valor reside en la cantidad de conocimiento o información que se pueda extraer de ellos (Molina López & García Herrero, 2006, p. 2).

Si bien la minería de datos emplea herramientas bastantes útiles para extraer información, se pueden presentar ciertos inconvenientes, por ejemplo: dificultad

---

[17] Las técnicas de visualización de datos, consideran histogramas, gráficos de dispersión, diagramas de cajas, etc.

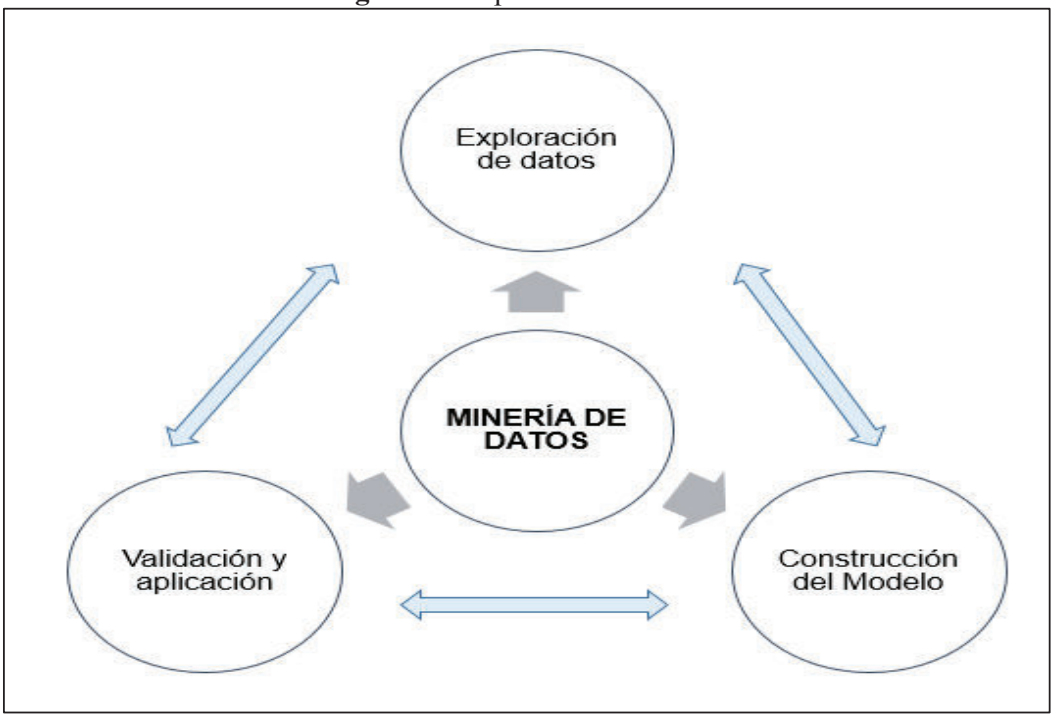
[18] Los métodos computacionales abarcan las técnicas de estadística descriptiva, como son regresiones, cluster, análisis de clasificación, etc.

[19] El aprendizaje automático, es el mecanismo que utiliza la experiencia del investigador, para crear y mejorar programas informáticos (Mitchell, 1997, p. 1).

en la comprensión de datos por parte del investigador, utilización de técnicas de forma errónea, violación de los supuestos, o interpretación errada de resultados (Hair, et al., 1999, pp. 29-30).

Un proceso de minería de datos tiene tres etapas fundamentales: exploración de datos, construcción del modelo y, validación y aplicación. La primera etapa considera operaciones de tratamiento de datos, como trasformaciones, limpieza, análisis de dimensión, identificación de datos atípicos, análisis gráficos, etc.; con el fin de analizar, conocer sus características y verificar la calidad de los mismos. En la segunda etapa se analiza los posibles modelos y se elige aquel que tenga la mejor capacidad predictiva, cumpla con los supuestos matemáticos, se asemeje lo mayor posible a la realidad, y satisfaga los objetivos planteados al inicio de la investigación. Finalmente como tercera y última etapa, se realiza la validación y aplicación del modelo, mediante la utilización de un nuevo conjunto de datos para verificar si el modelo posee efectivamente una buena capacidad predictiva, y si los resultados obtenidos están acorde a los esperados (Gorunescu, 2011, pp. 7-14).

Figura 6: Etapas minería de datos



Fuente: Gorunescu; Data Mining, concepts, models, and techniques; 2011
Elaboración: Los autores

3.2 CLASIFICACIÓN DE LAS TÉCNICAS DE MINERÍA DE DATOS

Según Pérez López (2004), en minería de datos existen generalmente dos técnicas de análisis: técnicas descripticas y técnicas predictivas.

Las técnicas descriptivas[20] no distinguen a las variables entre dependientes y explicativas, es decir analizan todas las variables de manera simultánea, de tal manera que las observaciones puedan ser clasificadas en grupos homogéneos, en base a determinadas características que posean en común. Dentro de este grupo se consideran los métodos de clustering, segmentación, técnicas de asociación, de análisis exploratorio, y de reducción de la dimensión.

Las técnicas predictivas[21] realizan una identificación previa de las variables dependientes e independientes a utilizarse en el modelo en base a los fundamentos teóricos. Su objetivo es estimar valores futuros con ayuda de las variables existentes. Dentro de este grupo constan regresiones, análisis discriminante, árboles de decisión, entre otras (Pérez López, 2004, pp. 1-4).

$$\textit{técnicas de minería de datos}\begin{cases}\textit{descriptivas}\begin{cases}\textit{segmentación}\\ \textit{técnicas de asociación}\\ \textit{análisis exploratorio}\\ \textit{reducción de la dimensión}\end{cases}\\ \textit{predictivas}\begin{cases}\textit{Regresiones}\\ \textit{Árboles de decisión}\\ \textit{Redes neuronales}\\ \textit{Modelos de elección discreta}\end{cases}\end{cases}$$

Fuente: Pérez López César. Técnicas de dependencia y modelos predictivos, Panorámica de modelos, 2004.
Elaboración: Los autores

[20] Conocidas también como aprendizaje no supervisado o técnicas estadísticas de la interdependencia.
[21] Conocidas también como aprendizaje supervisado o técnicas estadísticas de la dependencia.

Tanto las técnicas descriptivas como las técnicas predictivas pueden ser implementadas de manera independiente o conjunta, ya que su utilización depende de la calidad, cantidad y tipo de datos que se dispone, además del objetivo de la investigación.

Debido a que el objetivo principal del presente estudio es explicar el riesgo tributario representado por una variable cualitativa, se dará un mayor enfoque a las técnicas predictivas, especialmente en modelos de elección discreta.

3.2.1 TÉCNICAS PREDICTIVAS. LOS MODELOS DE ELECCIÓN BINARIA

Los modelos de elección binaria o de respuesta dicotómica, se caracterizan por poseer una variable endógena de naturaleza cualitativa, que únicamente toma dos valores: 1 si se cumple determinada condición (éxito), y 0 caso contrario (fracaso). Son conocidos también como modelos probabilísticos, debido a que su objetivo es conocer la probabilidad de que un determinado evento ocurra (Gujarati, 2003, pp. 560-562).

Según Amemiya (2013), se asume que la probabilidad de que un evento ocurra depende de una función compuesta por un vector de variables independientes X, y de los parámetros desconocidos (β) de la siguiente manera.

$$P(y_i = 1) = F\left(x_i'\beta\right) \quad donde, \;\; i = 1, \dots, n \tag{1}$$

Dependiendo del tipo de función F se tienen los modelos de probabilidad lineal, probit y logit.

3.2.1.1 MODELOS DE PROBABILIDAD LINEAL (MPL)

Gujarati (2003), establece que un modelo de probabilidad lineal, es un modelo de regresión lineal múltiple, caracterizado por poseer una variable endógena binaria. El objetivo de estos modelos es estimar la probabilidad de que ocurra un evento, dado un conjunto de variables independientes. Su forma general es:

$$F(x_i', \beta) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \mu \quad (2)$$

Donde:

y: Variable binaria, que únicamente toma dos valores posibles 0 y 1.
β_k: Coeficientes de las variables exógenas.
x_k: Variables exógenas de naturaleza cualitativa o cuantitativa
μ: Término de error.

Al ser un modelo de regresión múltiple, puede ser estimado bajo el método de mínimos cuadrados; donde los estimadores β_i miden la variación de la probabilidad de éxito, al variar x_i, manteniendo los demás factores constantes (Wooldridge, 2010, p. 247).

Sin embargo, Gujarati (2003) y Wooldrige (2010), mencionan que el MPL, posee ciertas desventajas: la probabilidad de éxito depende linealmente de las variaciones de x_i situación que no se asemeja a la realidad; se tiene probabilidades estimadas fuera del rango permitido y, los errores son heterocedasticos ya que la varianza depende de las x_i.

Por ello, se plantea los modelos a continuación, que son una alternativa ante los problemas que presenta el modelo de probabilidad lineal.

3.2.1.2 MODELO PROBIT

Si la función $F(x_i'\beta)$ de la ecuación (1) sigue una función de distribución normal estándar se tiene como resultado los modelos Probit. Estos modelos poseen una variable endógena de respuesta binaria; se estiman mediante un proceso de máxima verosimilitud; y la función de distribución tiene la siguiente estructura (Amemiya, 1981, p. 1486).

$$F(I_i) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{I_i} e^{\frac{-Z^2}{2}} dz \quad (3)$$

Donde:

$I_i = \beta_1 + \beta_2 X_i$
$Z \sim N(0, \sigma^2)$

3.2.1.3 MODELO LOGIT

Según Amemiya (2013) los modelos logit se obtienen cuando la función $F(x_i'\beta)$ de la ecuación (1) sigue una función de distribución acumulada logística, de la siguiente forma.

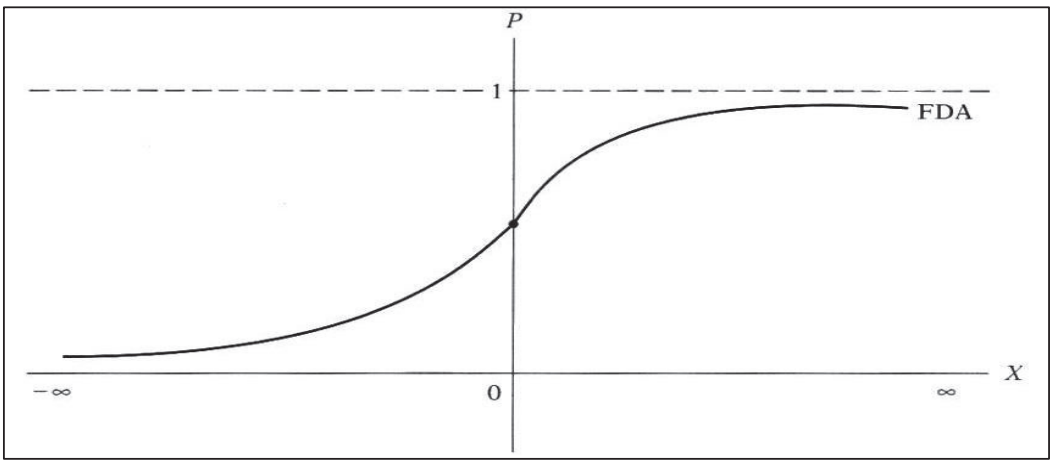
$$F(x_i'\beta) = \frac{e^{x_i'\beta}}{1+e^{x_i'\beta}} \tag{4}$$

Esta distribución se caracteriza por tener una forma de "S", lo que quiere decir que, la probabilidad de la variable endógena se acerca al valor de 1, a medida que el valor de las variables explicativas aumenta; mientras que se acerca a 0, a medida que el valor de las variables explicativas disminuye, tal como se puede notar en la figura 7 (Gujarati, 2003, pp. 572-574).

A diferencia de la distribución normal, la distribución logística posee extremos ligeramente más pesados. Esta propiedad acota la probabilidad de pertenecer a cualquiera de los dos grupos dentro del intervalo de 0 a 1.

Según Gujarati (2003), el modelo Logit no requiere cumplir el supuesto de linealidad[22] y, a diferencia del MPL estima probabilidades dentro del rango permitido.

Figura 7: Función de distribución acumulativa logística



Fuente: Gujarati, Econometría, 2003
Elaboración: Los autores

[22] El supuesto de linealidad establece que a medida que la variable exógena aumenta, aumenta la variable endógena en la misma proporción (Gujarati, 2003, p. 40).

Según Gujarati (2003), la probabilidad de éxito en un modelo Logit viene dada por:

$$P_i(Y = 1) = \frac{1}{1+e^{-Z_i}} = \frac{e^{Z_i}}{1+e^{Z_i}} \quad (5)$$

Con $Z_i = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k$

Por tanto, la probabilidad de fracaso o de tener Y=0 viene dado por:

$$P_i(Y = 0) = 1 - P_i(Y = 1) = \frac{1}{1+e^{Z_i}} \quad (6)$$

Si se divide la ecuación (5) para la (6), se obtiene la **razón de probabilidades u odd ratio**, que muestra el número de veces que es más probable que ocurra el evento, frente a que no ocurra; puesto que divide la probabilidad de éxito para la probabilidad de fracaso de la siguiente manera (Hair, et al., 1999, p. 282):

$$Odds_i = \frac{P_i}{1-P_i} = \frac{1+e^{Z_i}}{1+e^{-Z_i}} \quad (7)$$

Si el odd ratio es igual a 1, significa que la probabilidad de que ocurra el evento es igual a la probabilidad de que no ocurra. Si el odd ratio es mayor que 1, significa que aumenta la probabilidad de que el evento ocurra (probabilidades mayores a 50%), es decir, aumenta la probabilidad de que la variable endógena tome el valor de 1. Caso contrario, si el odd ratio es menor que 1, disminuye la probabilidad de que el evento suceda (probabilidades menores a 50%), es decir, provoca que la variable endógena se acerque a 0 (Hair, et al., 2009, p. 322) .

Al aplicar logaritmo natural a la ecuación (7), se tiene como resultado el **logit.**

$$L_i = \ln\left(\frac{P_i}{1-P_i}\right) = Z_i = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k \quad (8)$$

Dado que los modelos logit incumplen el supuesto de linealidad, el mejor método de estimación es por máxima verosimilitud, mismo que busca obtener los parámetros desconocidos (β_i) para maximizar la probabilidad de que el evento ocurra.

3.2.1.4 ESTIMACIÓN DE MÁXIMA VEROSIMILITUD (VM)

Para explicar la estimación de máxima verosimilitud, Gujarati (2003) parte del modelo logit, donde la probabilidad de pertenencia al grupo de éxito (Y=1), viene dado por la fórmula (5).

Puesto que Y_i es una variable aleatoria que sigue una distribución de Bernoulli, la probabilidad de éxito y la probabilidad de fracaso vienen expresadas respectivamente de la siguiente manera:

$$Pr(Y_i = 1) = P_i \tag{9}$$

$$Pr(Y_i = 0) = (1 - P_i) \tag{10}$$

Si se supone que se tiene una muestra aleatoria de *n* observaciones, la probabilidad conjunta para las n observaciones; es decir, $f_i(Y_1, Y_2, \dots, Y_n)$ se expresa como:

$$f_i(Y_1, Y_2, \dots, Y_n) = \prod_1^n f_i(Y_i) = \prod_1^n P_i^{Y_i}(1 - P_i)^{1-Y_i} \tag{11}$$

Donde

$\prod$: Operador del producto

La probabilidad conjunta de la ecuación (11) es la función de verosimilitud (FV).

En la estimación de máxima verosimilitud (VM) se tiene como objetivo maximizar la función de verosimilitud, es decir, obtener los valores de los parámetros desconocidos, de tal forma que la probabilidad de observar el valor de las Y sea tan grande (máximo) como sea posible, dada la observación de las variables explicativas.

3.2.1.5 COMPARACIÓN DEL MODELO LOGIT CON OTROS MODELOS DE ELECCIÓN BINARIA

De lo mencionado anteriormente, se tiene que las principales características de los tres tipos de modelos de elección binaria son las siguientes:

Tabla 2: Comparación de modelos logit vs probit

CARACTERÍSTICAS	LOGIT	PROBIT	M.P.L.
Función de Distribución	Logística[23]	Normal[24]	Bernoulli
Varianza	$\frac{\pi^2}{3}$	1	$p(1-p)$
Factor de Conversión	$\widehat{\beta_L} \approx 1{,}6\widehat{\beta_P}$[25]	$\widehat{\beta_P} \approx \frac{5}{8}\widehat{\beta_L}$	$\widehat{\beta_{PL}} \approx 0{,}4\widehat{\beta_P}$ $\widehat{\beta_{PL}} \approx 0{,}25\widehat{\beta_L}$
Media	0	0	p
Método de Estimación	Máxima Verosimilitud		MCO
Signo β_i estimados	El signo de los coeficientes indica el sentido del efecto de las variables exógenas sobre la variable endógena.		
Magnitud β_i estimados	Es necesario cálculo de Odds Ratios y efectos marginales		Interpretación Directa

Fuente: Gujarati, Econometría, 2003.
Elaboración: Los autores

3.2.1.6 BONDAD DE AJUSTE

Los principales criterios de bondad de ajuste en modelos con variable endógena binaria según Hair (1999) son:

---

[23] Función de distribución con extremos ligeramente más anchos
[24] Requiere cálculo de integrales para obtener su distribución de probabilidad
[25] T. Amemiya, "Qualitative Response Model: A Survey" Journal of Economic Literature, Vol. 19, 1981
$\widehat{\beta_L}$: Coeficiente estimado Logit.
$\widehat{\beta_p}$ Coeficiente estimado Probit.
$\widehat{\beta_{Lp}}$ Coeficiente estimado Modelo de Probabilidad Lineal.

- -2 Log de Verosimilitud (-2LL)

Es la medida global de ajuste del modelo. Un modelo que tiene un buen ajuste, posee un valor pequeño de este estadístico, siendo 0 el valor mínimo, mismo que se consigue cuando el ajuste es perfecto.

- Test de Hosmer y Lemeshow

El test de Hosmer y Lemeshow permite conocer si el modelo posee o no un adecuado ajuste, al comparar los valores predichos con los observados, de tal manera que mientras menor sea la diferencia entre dichos valores, mejor ajuste tendrá el modelo. Para ello se realiza una prueba chi-cuadrado, donde el objetivo es que el p-valor de dicha prueba sea no significativo, lo que implica aceptar la hipótesis nula, y concluir que el modelo posee un adecuado ajuste (Hair, et al., 1999, p. 321).

$$H_0: el\ modelo\ posee\ un\ buen\ ajuste$$

$$H_A: el\ modelo\ no\ posee\ un\ buen\ ajuste$$

- R^2

En los modelos de respuesta binaria, el R^2 no es un buen indicador del ajuste del modelo, debido al método de estimación de dichos modelos, por lo que Hair (1999) menciona dos medidas comparables con el R^2 de la regresión múltiple.
EL R^2 de Cox y Snell, indica que a valores más altos, mayor será el ajuste del modelo, sin embargo esta medida no posee un límite superior, es decir, no alcanza el valor de 1 para indicar un ajuste perfecto; es por ello que el R^2 de Nagelkerke es una versión mejorada del R^2 de Cox y Snell, debido a que establece un rango entre 0 y 1 para indicar el ajuste del modelo, siendo los valores cercanos a 1, indicadores de un buen ajuste.

- Tabla de Clasificación

Según Hosmer y Lemeshow (2000), la tabla de clasificación muestra de manera resumida los resultados obtenidos mediante la estimación del modelo. Esta tabla

cruza los valores observados de la variable endógena, versus, los valores estimados de dicha variable por el modelo. Para obtener ésta última variable binaria, es necesario establecer un punto de corte[26].

La tabla de clasificación considera criterios de sensibilidad y especificidad de la siguiente manera (Fawcett, 2005, p. 862):

Tabla 3: Tabla de clasificación

PRONOSTICADO	**OBSERVADO**		
	Positivo	**Negativo**	
Positivo	VP	FP	
Negativo	FN	VN	
Total	VP + FN = T_1	FP + VN = T_2	$T_1 + T_2$

VP: verdaderos positivos; FP: falsos positivos; FN: falsos negativos; VN: verdaderos negativos
Fuente: Fawcett, 2005
Elaboración: Los autores

La sensibilidad muestra la proporción de observaciones que siendo "positivas", el modelo las clasificó como "positivas", es decir, la sensibilidad hace referencia a la tasa de *verdaderos positivos*.

$$Sensibilidad = \frac{VP}{T_1}$$

La especificidad muestra la capacidad que tiene el modelo para discriminar adecuadamente los casos que no poseen la característica de éxito; muestra la diferencia entre 1 menos el valor de los *falsos positivos*. Siendo éstos últimos, la proporción de observaciones que siendo "negativas", el modelo las clasificó como "positivas".

$$Especificidad = \frac{VN}{FP + VN} = 1 - tasa\ de\ falsos\ positivos = 1 - \frac{FP}{T_2}$$

[26] El punto de corte generalmente toma el valor de 0,5. De esta manera, las probabilidades estimadas que superen dicho valor, se les asignará el valor de 1, caso contrario tomarán el valor de 0 (Hosmer & Lemeshow, 2000, p. 156).

Por lo tanto, mientras más cercanos sean los valores de sensibilidad y especificidad al 100%, mejor será el poder de clasificación del modelo estimado.

Otros criterios a tener en cuenta son:

Tasa de Verdaderos positivos: muestra la proporción de observaciones positivas correctamente clasificadas del total de observaciones positivas.

$$Tasa\ de\ Verdaderos\ Positivos = \frac{VP}{T_1}$$

Tasa de Falsos Positivos: muestra la proporción de observaciones negativas, clasificadas incorrectamente del total de observaciones negativas.

$$Tasa\ de\ Falsos\ Positivos = \frac{FP}{T_2}$$

Precisión: El valor predictivo positivo.

$$Precisión = \frac{VP}{VP + FP}$$

Exactitud: muestra el valor predictivo global del modelo, es decir la proporción de observaciones verdaderas positivas y verdaderas negativas del total de observaciones.

$$Exactitud = \frac{VP + VN}{T_1 + T_2}$$

- Curva ROC

La curva ROC, según Hosmer y Lemeshow (2000), es una medida gráfica del ajuste del modelo, puesto que considera en el eje de las abscisas el valor de los falsos positivos, es decir 1- especificidad; y en el eje de las ordenadas el valor de la sensibilidad.

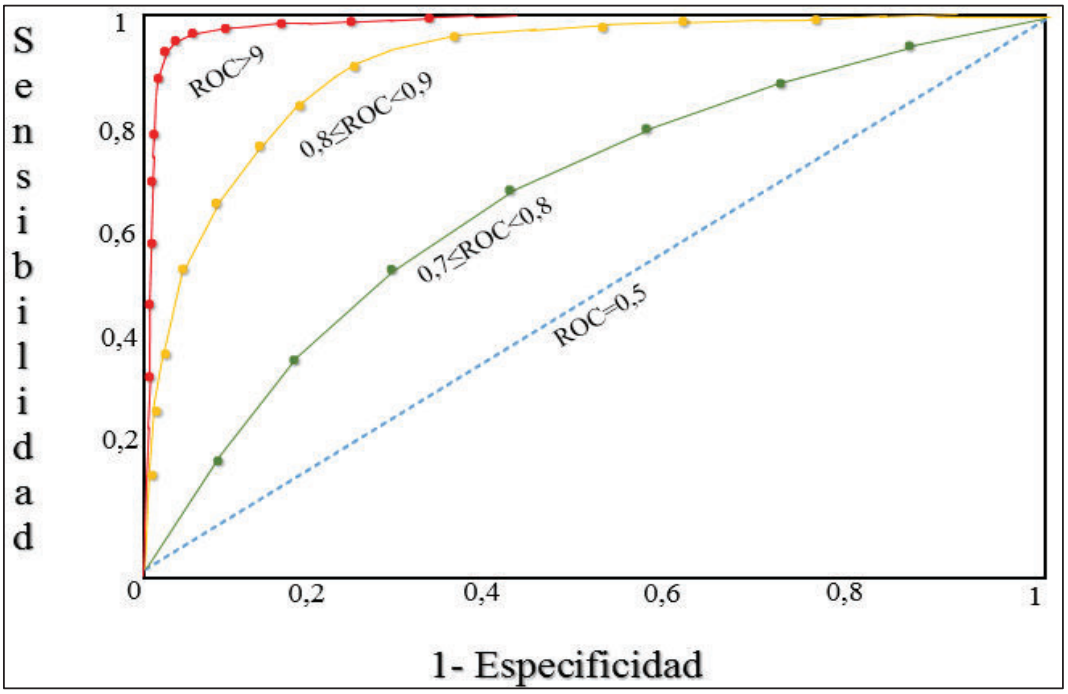
El área bajo la curva ROC muestra el porcentaje de discriminación del modelo, por lo que mientras más alto sea dicho porcentaje, el modelo presentará un mejor poder de clasificación de las observaciones entre los grupos.

De esta manera se tiene que:

- $Curva\ ROC = 0{,}5$ no existe discriminación
- $0{,}7 \leq Curva\ ROC < 0{,}8$ discriminación aceptable
- $0{,}8 \leq Curva\ ROC < 0{,}9$ discriminación excelente
- $Curva\ ROC > 0{,}9$ discriminación excepcional

Bajo este escenario, la curva ROC puede presentar las siguientes formas:

Figura 8: Tipos de Curva ROC



Fuente: Hosmer y Lemeshow (2000)
Elaboración: Los autores

3.2.1.7 MÉTODOS DE BUSQUEDA SECUENCIAL

Los métodos de búsqueda secuencial son útiles para la selección de variables independientes dentro del modelo, ya que al ir añadiendo o descartando dichas variables de acuerdo a determinados criterios, da como resultado el mejor

modelo posible, respondiendo de esta manera al criterio de parsimonia (IBM, 2012, p. 259).

De acuerdo con Hair (1999), los tres principales métodos de búsqueda secuencial son: Forward, Backward y Stepwise.

- Forward (Método hacia Adelante)

El método forward incluye variables en la estimación del modelo una a la vez; es decir, empieza realizando una estimación del modelo considerando únicamente la variable explicativa que presenta una mayor correlación con la variable objetivo; para luego, una vez estimado el modelo con la variable seleccionada, analizar si existe o no una mejora en los estadísticos de bondad de ajuste, es decir, si el valor de -2LL disminuye y, si los R cuadrados de Cox y Snell, y de Nagelkerke aumentan; también se examina si aumenta el poder de clasificación del modelo. En caso de que la variable seleccionada no mejore los estadísticos mencionados, se descarta la inclusión de la misma dentro de la especificación y estimación del modelo. Este proceso se repite con cada una de las variables explicativas hasta que todas hayan sido evaluadas (IBM, 2012, p. 260).

La inclusión de variables explicativas en cada paso dentro de la estimación del modelo debe cumplir dos requisitos: el primero es el estadístico de puntuación o score statistic, puesto que representa una medida de asociación de las variables en los modelos logístico; la variable que posea el mayor puntaje del score statistic y que al mismo tiempo dicho score sea estadísticamente significativo, es la indicada para ingresar en el modelo. El segundo, es la mejora del estadístico -2LL, lo que implica que la variable que ingresa en el modelo debe disminuir el valor de dicho estadístico. Únicamente la variable que cumpla estos dos requisitos de manera simultánea es la que ingresará en al modelo (Hair, et al., 2009, pp. 332-335).

La mejora del estadístico -2LL se basa en el test de razón de verosimilitud, mismo que analiza si una variable explicativa debe o no ser incluida en la especificación del modelo, al examinar su significancia mediante la estimación y comparación de dos modelos, uno incluida la variable y otro sin ella. El test

considera como hipótesis nula que el parámetro estimado es igual a 0, y usa el valor de la función de verosimilitud de la siguiente manera:

$$D = -2ln\left[\frac{Verosimilitud\ del\ modelo\ construido}{Verosimilitud\ del\ modelo\ saturado}\right]$$ [27]

Para evaluar la significancia de una variable independiente, se compara el valor de D con y sin dicha variable independiente en la ecuación, de esta manera, el cambio en D para la inclusión de la variable en el modelo está dado por:

$$G = D(Modelo\ \sin\ la\ variable) - D(Modelo\ con\ la\ variable)$$

Este estadístico al diferenciarse puede ser expresado como:

$$G = -2ln\left[\frac{Verosimilitud\ \text{del modelo}\ \sin\ la\ variable}{verosimilitud\ del\ modelo\ con\ la\ variable}\right]$$

El estadístico G sigue una distribución chi- cuadrado con un grado de libertad (Hosmer & Lemeshow, 2000, pp. 12-14).

- Backward (Método hacia atrás)

El método backward es similar al método hacia delante; se diferencia en que realiza la primera estimación incluyendo todas las variables explicativas al mismo tiempo para posteriormente eliminar una por una aquellas que no aporten al modelo, tomando como criterio de exclusión de variables al estadístico de Wald o al test de razón de verosimilitud.

- Stepwise (Paso a Paso)

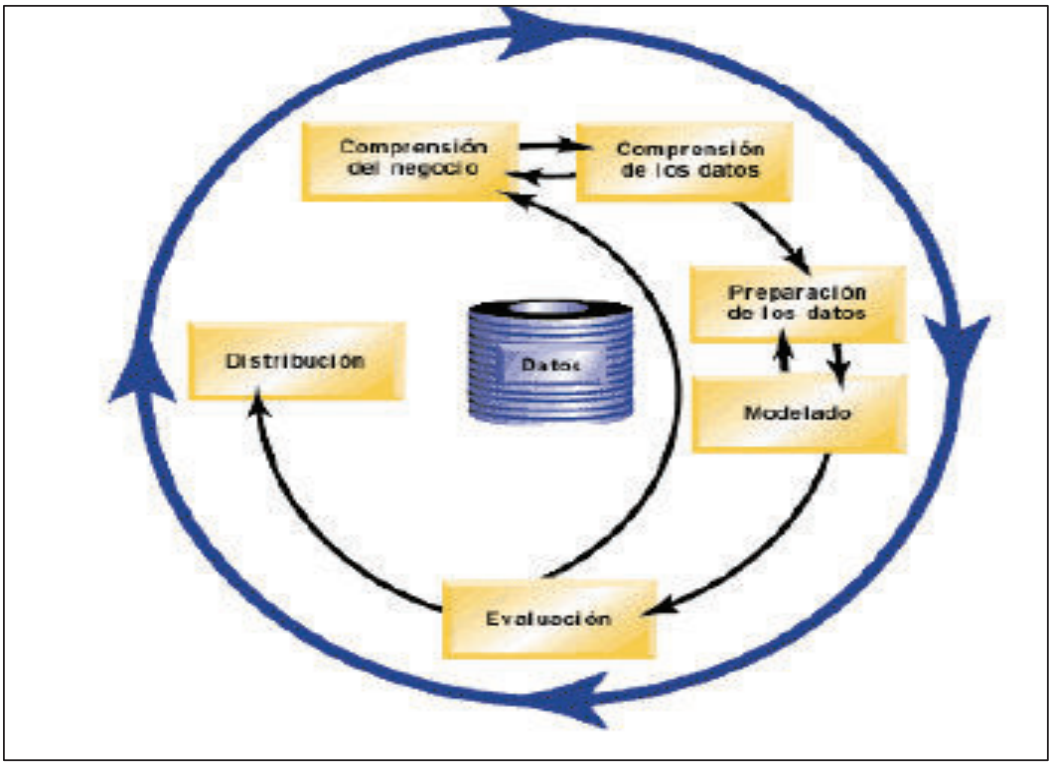
El método paso a paso o stepwise es similar al método forward, con la única diferencia que las variables ya incluidas en el modelo pueden ser eliminadas en pasos posteriores.

[27] "Un modelo saturado es un modelo que contiene tantos parámetros como datos" (Hosmer & Lemeshow, 2000, p. 12).

3.3 METODOLOGÍA CROSS INDUSTRY STANDARD PROCESS FOR DATA MINING (CRISP-DM)

Según Chapman et al (1999), la metodología Cross-Industry Standard Process for Data Mining o CRISP-DM nace en 1996, con el objetivo de estructurar un proceso estándar para la realización de proyectos con herramientas de minería de datos, enfocado a la industria. Es considerado como un modelo interactivo y flexible, para el estudio de grandes bases de datos. Toma como base al menos las tres etapas fundamentales enunciadas anteriormente (exploración de datos, construcción del modelo y, validación y aplicación), para formular un ciclo compuesto de seis pasos: comprensión del negocio, comprensión de datos, preparación de datos, modelado, evaluación y, distribución, como se muestra en la figura 9. Estos pasos brindan una guía completa al investigador para desarrollar de manera efectiva cualquier tipo de proyecto.

Figura 9: Ciclo de la metodología CRISP-DM



Fuente: Manual CRISP-DM, IBM, 2012

El orden de las fases es flexible, puesto que al ser una metodología de minería de datos, se basa en el principio de aprendizaje automático; es decir, se puede personalizar el ciclo, retrocediendo o avanzando entre las distintas fases, dependiendo de la investigación que se realice y los objetivos planteados, con el fin de generar un modelo que se adapte a necesidades concretas.

A diferencia de otras metodologías para minería de datos, una de las principales ventajas de utilizar CRISP DM es que dicha metodología da prioridad a que el investigador tenga un conocimiento amplio sobre el problema que se intenta solucionar y el contexto bajo el cual se desarrolla el accionar de la industria (comprensión del negocio). Esta etapa es fundamental para llevar a cabo adecuadamente el proceso de minería de datos y responder eficientemente a las necesidades y objetivos de la industria.

3.3.1 FASES DE LA METODOLOGÍA CRISP-DM

3.3.1.1 COMPRENSIÓN DEL NEGOCIO

La compresión del negocio proporciona al investigador una idea clara y precisa del funcionamiento del negocio y define el área donde existe el problema. Para ello, se debe contar con la mayor cantidad de personas expertas en el área de aplicación del proyecto (no precisamente personas de carácter técnico, matemático o estadístico), puesto que con su ayuda, el analista puede entender cómo funciona el negocio, conocer los objetivos, las restricciones, los recursos, el alcance de la investigación, las hipótesis a probar, la situación actual del negocio, y los posibles resultados. Por tanto, es necesario estructurar un plan de trabajo preliminar que guíe al investigador y permita trazar un camino a seguir para la buena ejecución del proyecto (Chapman, et al., 1999, pp. 16-19).

Es necesario en ésta etapa que todos los participantes dentro de la investigación estandaricen el significado de cada término del negocio, por lo que la creación de un glosario de términos es indispensable. De igual manera, se debe discutir si se cuenta con la información suficiente y de calidad para realizar la modelización, ya que de no ser el caso, se estaría invirtiendo recursos económicos y humanos en una investigación que no arrogará resultados

eficientes. Finalmente, se debe generar un bosquejo de las herramientas de minería de datos que posiblemente vayan a ser utilizadas, considerando las técnicas descriptivas y predictivas mencionadas anteriormente, y teniendo presente los requerimientos del cliente (IBM, 2012, pp. 4-13).

3.3.1.2 COMPRENSIÓN DE LOS DATOS

El investigador se encarga de recolectar todos los datos necesarios para empezar a realizar el proyecto. Una vez constituida la base de datos, se procede como primer paso a realizar un análisis exploratorio de la misma, para lo cual, las técnicas de visualización, los análisis gráficos, los análisis estadísticos simples, entre otros son sumamente útiles, puesto que permiten conocer la calidad de los datos, el significado de cada variable, los valores que toman, el tipo de datos, y en general llegar a tener una visión clara de la información con la que se cuenta (Chapman, et al., 1999, pp. 20-22).

Si bien los gráficos facilitan la comprensión del tipo de datos, los resultados deben estar siempre validados por tests o pruebas matemáticas para que los mismos sean lo más objetivos posibles. Por esta razón, es conveniente dentro de esta etapa calcular y conocer los estadísticos básicos de cada variable a incluir en el análisis. Dichos estadísticos comprenden la media, desviación típica, mediana, percentiles, valores mínimos, máximos, entre otros. Finalmente, una vez que se tenga una idea clara acerca de la posible distribución que siguen los datos, se debe analizar la posibilidad de realizar transformaciones de los mismos para ajustarlos a una distribución que facilite el proceso de modelado (Pérez López, 2004, pp. 21-39).

3.3.1.3 PREPARACIÓN DE LOS DATOS

La preparación de los datos es la etapa más importante para realizar una investigación mediante minería de datos, ya que la misma tiene como objetivo principal la estructuración de una base de datos final que será utilizada directamente en el proceso de modelado. Esta estructuración se fundamenta en un proceso de limpieza de datos, donde se analiza si es conveniente realizar

transformaciones, crear nuevas variables, realizar operaciones aritméticas entre variables, fusionar datos, implementar técnicas para la detección de datos atípicos y agregar variables, para obtener información útil y de calidad (Chapman, et al., 1999, pp. 23-26).

La limpieza de datos ayuda a distinguir si existen datos ausentes, atípicos, extremos, erróneos, etc. La presencia de datos faltantes o ausentes puede deberse a varios motivos, por ejemplo, a un mal registro en cuanto al levantamiento de datos, a la negativa de los encuestados a responder, o simplemente a la naturaleza de la información buscada. Para dar solución oportuna a estos problemas, se recurren a diversas técnicas de supresión[28] e imputación[29] de datos.

Por otra parte, los datos atípicos presentan características únicas y distintas al resto de datos, y pueden influenciar en los resultados de la estimación. Para identificar este tipo de datos de manera gráfica y/o estadística, es conveniente realizar un análisis exploratorio (Pérez López, 2004, pp. 39-55).

3.3.1.4 MODELADO

Esta etapa primero toma en cuenta todas las posibles técnicas matemáticas que podrían ser útiles en el proyecto, luego, considera el requerimiento del cliente, así como las restricciones que puedan existir, para finalmente llegar a tener un número limitado de posibles técnicas a ser probadas.

Inmediatamente que se ha optado por una o algunas técnicas, se debe particionar los datos, de tal manera que un conjunto de ellos sea utilizado para realizar todo el proceso de modelado, y la parte restante, para el proceso de prueba. Una vez hecho esto, se estima el modelo en la herramienta informática disponible, para finalmente obtener los resultados, los mismos que deben ser validados y estar acorde a los objetivos planteados en el primer paso.

---

[28] La supresión de datos consiste en eliminar los datos faltantes.

[29] La imputación de datos considera un conjunto de métodos para estimar valores que sirvan para sustituir los datos ausentes, entre los principales métodos de imputación están: imputación por sustitución de media, por sustitución de mediana, por interpolación, por un valor constante, por regresión (Pérez López, 2004, pp. 46-48).

Dentro de esta etapa, se deben realizar diversos tests técnicos para comprobar, en primer lugar, si se cumplen los supuestos básicos que rigen la técnica matemática utilizada; y en segundo lugar, para verificar la validez de los resultados obtenidos. No basta con estimar uno o varios modelos, ya que el aporte del investigador radica en ser capaz de entender los resultados, interpretarlos y transmitirlos en un formato sencillo y fácilmente entendible, además de contar con un dominio de conocimientos matemáticos para escoger aquel modelo que realice la mejor estimación y cumpla con los objetivos planteados en el negocio (Chapman, et al., 1999, pp. 53-56).

3.3.1.5 EVALUACIÓN

Esta etapa evalúa el grado en que el modelo cumple con los objetivos del negocio. Para ello, se considera el criterio del experto, que ayuda a interpretar los resultados obtenidos en términos del negocio.

Durante la evaluación se debe revisar nuevamente cada acción efectuada a lo largo del proceso de minería de datos, para constatar que cada tarea realizada haya sido justificada correctamente, y verificar si existe algún tipo de factor que haya sido excluido y que podría utilizarse en investigaciones futuras para una actualización del modelo (Chapman, et al., 1999, pp. 57-59).

Una manera de validar los resultados obtenidos es a través de una base de datos externa al proceso, de tal manera que se pueda realizar un seguimiento y monitoreo de los resultados (IBM, 2012, pp. 34-37).

3.3.1.6 IMPLEMENTACIÓN O DISTRIBUCIÓN

Una vez que se han evaluado los resultados del modelo de manera técnica, y desde el punto de vista del negocio en cuanto al alcance de los objetivos planteados, el cliente debe elaborar las estrategias a ejecutar para implementar el proyecto dentro de su empresa. Aquí, la construcción de un informe final con toda la investigación del proyecto de minería de datos es sumamente útil para elaborar las estrategias de acción y, poder de ésta manera, implementar el modelo obtenido o sus resultados dentro del sistema de la empresa.

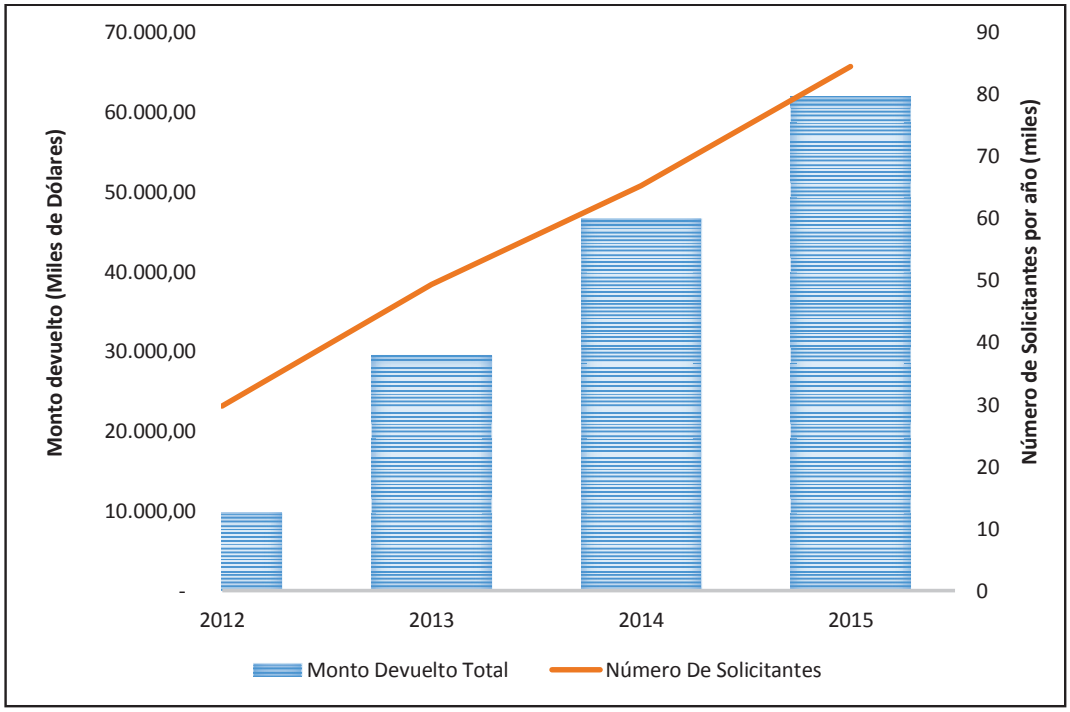
Dentro de las estrategias elaboradas debe constar un plan de acción para la supervisión y mantenimiento de los resultados del modelo, es decir, se debe definir quién será el responsable de la implementación del mismo, hasta cuando los resultados obtenidos pueden ser fiables, cuándo es necesario actualizar el modelo, etc. Finalmente, ésta etapa concluye con una recopilación de las experiencias tanto del cliente, como de cada uno de los involucrados en el proceso de minería de datos, con el objetivo de realizar una retroalimentación que brinde información útil para proyectos futuros (Chapman, et al., 1999, pp. 60-62).

4 ANÁLISIS DEL RIESGO TRIBUTARIO EN LA DEVOLUCIÓN DEL IVA A PERSONAS ADULTAS MAYORES EN ECUADOR

4.1 PANORAMA DE DEVOLUCIÓN DEL IVA

A partir del año 2012, el monto destinado al proceso de devolución del IVA a personas adultas mayores se ha incrementado de manera acelerada, pasando de 29.714 beneficiarios a quienes se les devolvió un monto de $9´843.605, a 84.443 beneficiarios con un monto de devolución equivalente a $61.914.882 en el año 2015; lo que significó que entre dichos años, casi se triplique el número de beneficiarios y se multiplique por seis la cantidad del monto devuelto, ocasionado que en promedio se realicen cerca de 28.000 trámites mensuales (SRI, 2016).

Figura 10. Evolución del monto devuelto y del número de beneficiarios, por año de notificación.



Fuente: SRI
Elaboración: Los autores

En la práctica[30], resulta casi imposible realizar un proceso de control y evaluación de cada trámite previo a la acreditación, debido al extenso volumen de trámites y beneficiarios que existen, lo que dificulta la determinación del nivel de riesgo presente en cada trámite. Por tal motivo, las estrategias de gestión de riesgo tributario contemplan la implementación de un modelo de riesgo tributario, para determinar con anticipación las características de trámites riesgosos, a los que el SRI debe realizar un proceso previo de verificación de información.

Mediante la utilización de minería de datos, se logra pasar de una revisión subjetiva (basada en criterio de experto) a una objetiva (basada en tramas de riesgo encontradas por medio de un análisis exploratorio de datos), con la que se consigue conocer e identificar de una mejor manera el nivel de riesgo tributario presente en cada trámite.

4.2 ANÁLISIS EXPLORATORIO DE DATOS

La base de datos[31] proporcionada por el SRI, cuenta con información recolectada según el año fiscal[32] 2015. Se tiene variables socioeconómicas, geográficas y propias del proceso de devolución del IVA, gran parte de ellas agregadas por trámites e individuos, y otras desagregadas por periodo y monto por el que se solicita la devolución. Estas variables se encuentran clasificadas de la siguiente manera (ver anexo B para la definición de las variables).

[30] La capacidad operativa con la cuenta el SRI para realizar un control posterior es de aproximadamente 130 trámites mensuales a nivel nacional.

[31] Si bien el SRI posee información del proceso de devolución del IVA de años anteriores, dicha base de datos no se encuentra desglosada completamente, razón por la cual, se decide realizar el análisis únicamente considerando el año fiscal 2015, puesto que es el año en el que se cuenta con la mayor cantidad de información.

[32] En este estudio, el año fiscal se refiere al año en el cual el beneficiario realiza el trámite de devolución del IVA; en dicho trámite pueden constar solicitudes de meses o años anteriores a la fecha en la que se realiza el trámite.

Tabla 4: Lista de variables clasificadas por periodo, trámite e individuo

PERIODO Y MONTO	TRÁMITES	INDIVIDUO
Mes fiscal	Tipo físico	Número de RUC
Monto Devuelto	Número de trámite	Correo repetido
Monto Solicitado	Número de facturas	Fecha de nacimiento
	Tipo ingreso o Sistema Nacional de Trámites (Snt) o Ventanilla o Internet	Género o Hombre o Mujer
	Oficina Resuelve	RUC activo

Fuente: SRI
Elaboración: Los autores

Considerar las variables antes mencionadas en la especificación del modelo de riesgo, responde a dos criterios: primero, son variables que intervienen directamente en el proceso de devolución del IVA a personas adultas mayores; y segundo, la experiencia de expertos tributarios del SRI, ratifican su inclusión.

El análisis empírico realizado por expertos del SRI a través de los años, permite tener un panorama claro de cuáles son las variables de interés para la Administración Tributaria dentro del proceso de devolución del IVA a personas adultas mayores; es así que, contar con un techo permitido de devolución, exigir respaldos que avalen las compras realizadas y que justifiquen el monto solicitado por el adulto mayor, contar con múltiples canales de recepción de solicitudes de devolución del IVA, entre otros aspectos, permiten al SRI llevar a cabo un proceso adecuado de gestión de riesgos.

4.2.1 PRINCIPALES ESTADÍSTICOS

Siguiendo la metodología de minería de datos CRIPS-DM, se realiza un análisis exploratorio de cada una de las variables antes mencionadas, con el propósito de tener un amplio panorama sobre el tipo de datos a utilizarse (Ver anexo C).

A continuación, se muestra en resumen, los principales estadísticos de cada variable.

Tabla 5: Principales estadísticos. Variables continuas

VARIABLE	MEDIA	DESV. ESTAN-DAR	MODA	MÍNIMO	MÁXIMO	SUMA TOTAL
Monto Devuelto	147,64	130,34	212,4	0	2543,06	49554071,94
Monto Solicitado	163,61	233,80	212,4	0,01	15267,86	54´913.770,51
N° de Facturas	45	43,8	4	1	999	14´941.327
Edad	73	5	69	65	105	--
Número de Periodos	3	2	1	1	24	335.636

Fuente: SRI
Elaboración: Los autores

Tabla 6: Principales estadísticos. Variables categóricas

VARIABLE	CATEGORÍA	NÚMERO DE TRÁMITES	% DEL TOTAL DE TRÁMITES
TIPO INGRESO	Ventanilla	248933	74,17 %
	SNT	5921	1,76 %
	Internet	80782	24,07 %
OFICINA RESUELVE	Quito- Sucursal Amazonas	115.912	34,54%
	Guayas	89.290	26,6 %
	Manabí	39.423	11,75 %
	Otros	171.372	27,31 %
CORREO REPETIDO	SI	150692	44,9 %
	NO	184944	55,1 %
GÉNERO	Hombre	143.456	42,74 %
	Mujer	192.180	57,26 %
RUC ACTIVO	SI	69.495	20,71 %
	NO	266.141	79,29 %

Fuente: SRI
Elaboración: Los autores

4.3 PREPARACIÓN DE DATOS

4.3.1 DATOS ELIMINADOS

El análisis exploratorio de datos indica la presencia de 4.332 registros duplicados según la variable “Tipo Físico” mismos que al ser registrados dos veces, modifican el monto devuelto total real, por lo que es necesario eliminarlos.

Se detecta 35 periodos que no cumplen con los requerimientos establecidos para el proceso de devolución del IVA, es decir, periodos con montos solicitados extremadamente altos o con valores nulos, y periodos donde el solicitante no se encuentra en el rango permitido de edad (ver anexo D).

Una vez depurada la información, se tiene una base de datos con las siguientes características:

Tabla 7: Datos generales proceso devolución IVA años fiscal 2015

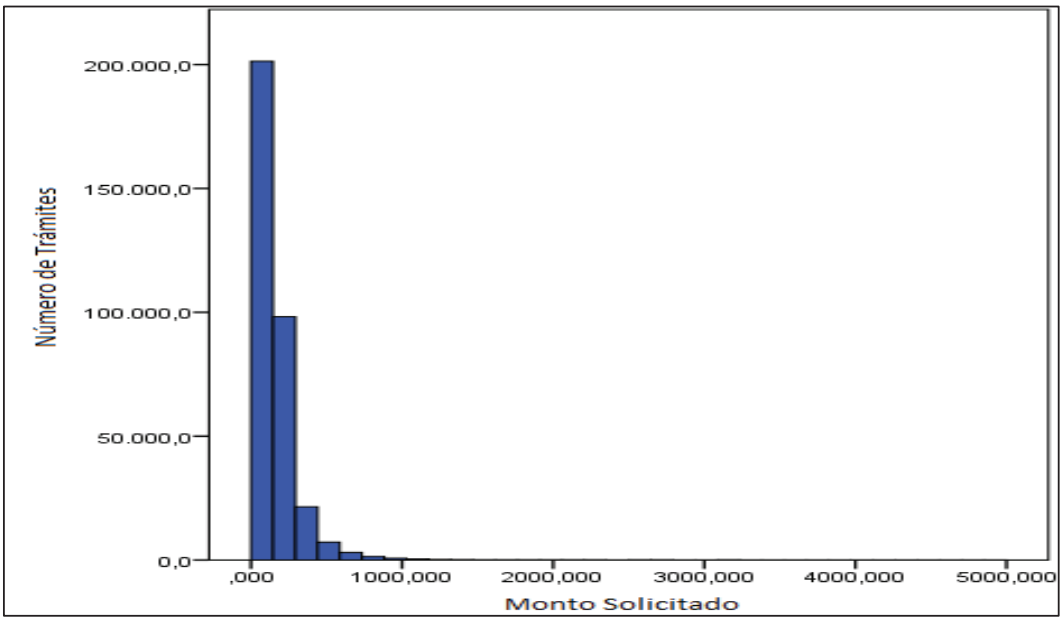
VARIABLES	CANTIDAD
Monto Devuelto Total	$ 49'554.071,94
Monto Solicitado promedio	$ 163,61
Facturas promedio por trámite	45
Total Individuos	80.347
% RUC activo	22,28%
% Correos repetidos	45,63%
% Hombres	42,74%
% Mujeres	57,26%
Total Trámites	335.636
Total Periodos	977.843

Fuente: SRI
Elaboración: Los autores

4.3.2 TRANSFORMACIÓN DE VARIABLES

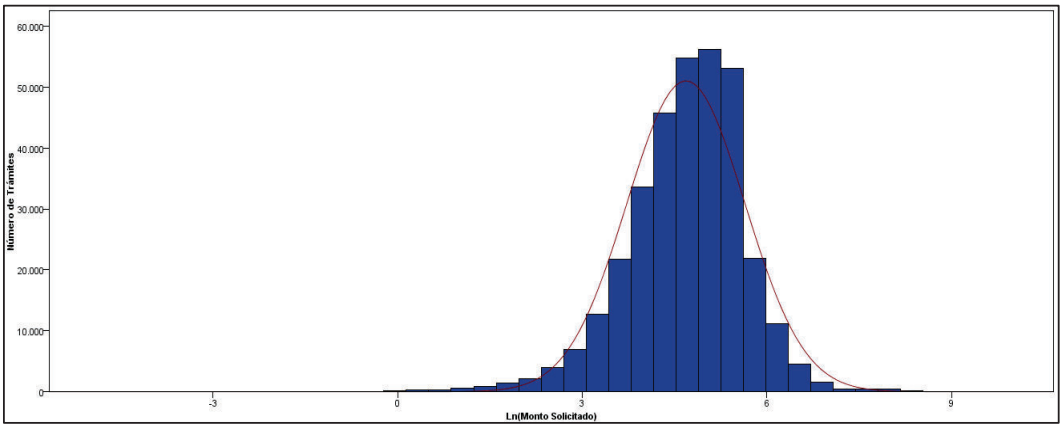
La distribución de la variable "Monto Solicitado" no se asemeja a una normal, y presenta asimetría positiva como se muestra en la figura 11. Si bien los modelos de regresión logística son flexibles en sus supuestos y no exigen normalidad en las variables (Hair, et al., 2009, p. 280), el poder de clasificación del modelo para trámites riesgosos aumenta si se considera dicha variable aplicada el logaritmo natural (ver anexo E). Por lo tanto, se opta por transformar la variable.

Figura 11: Distribución de la variable "Monto Solicitado" (variable original[33])



Fuente: SRI
Elaboración: Los autores

Figura 12: Distribución de la variable "Monto Solicitado" (Logaritmo Natural)



Fuente: SRI
Elaboración: Los autores

4.3.3 DATOS DE PRUEBA Y ENTRENAMIENTO

Para la construcción del modelo predictivo, se divide aleatoriamente los datos en dos subconjuntos independientes para las fases de desarrollo o entrenamiento,

[33] Se modifica los intervalos del gráfico, para una mejor visualización de la distribución de los datos; el gráfico con el intervalo completo se muestra en el anexo C.

y comprobación o prueba. Es decir, se selecciona un subconjunto (aproximadamente el 80% de los datos) para generar el modelo, y otro subconjunto (aproximadamente el 20% restante) para probarlo. Los datos para el desarrollo del modelo, y para la etapa de prueba, tienen las siguientes características.

Tabla 8: Características datos de entrenamiento

VARIABLES	SUBCONJUNTO ENTRENAMIENTO
Monto Devuelto Total	$ 39´589.108,55
Monto Devuelto Promedio	$ 147,50
Monto Solicitado Total	$ 43´897.177,53
Monto Solicitado promedio	$ 163,55
Factura promedio por trámite	45
Total individuos	75.519
% hombres	42,76 %
% mujeres	57,24 %
% RUC activo	22,14 %
% Correo repetido	45,57 %
Total trámites	268.388
Total Periodos	781.006

Fuente: SRI
Elaboración: Los autores

Tabla 9*:* Características datos de prueba

VARIABLES	SUBCONJUNTO PRUEBA
Monto Devuelto Total	$ 9’964.963,39
Monto Devuelto Promedio	$ 148,18
Monto Solicitado Total	$ 11’016.592,98
Monto Solicitado promedio	$ 163,83
Factura promedio por trámite	45
Total individuos	41.149
% hombres	42,50 %
% mujeres	57,49 %
% RUC activo	21,18 %
% Correo repetido	45,07 %
Total trámites	67.248
Total Periodos	196.837

Fuente: SRI
Elaboración: Los autores

Se nota que tanto los datos seleccionados para la construcción del modelo, como los datos destinados para pruebas, mantienen características similares en cada una de las variables mencionadas.

4.4 DESARROLLO DEL MODELO

4.4.1 VARIABLE OBJETIVO "RIESGO"

La variable objetivo llamada "RIESGO", es una variable dummy[34] que clasifica los trámites en dos grupos; por un lado se encuentran los trámites con alta probabilidad de fraude, ya que cumplen con determinadas características de riesgo tributario; y por otro lado, aquellos que no cumplen con dichas características y que por tanto no son considerados riesgosos.

El proceso de identificación y construcción de dicha variable se realizó internamente en el SRI, conjuntamente con ayuda del conocimiento de expertos en temas de devolución del IVA, por lo que su publicación en el presente trabajo está sujeta a políticas de confidencialidad.

Formalmente, esta variable toma dos valores:

$$RIESGO = \begin{cases} 1 & el\ trámite\ posee\ riesgo \\ 0 & caso\ contrario \end{cases}$$

El valor 1 indica que el trámite cumple con alguna característica de riesgo, lo que quiere decir que existe una gran posibilidad de que el monto devuelto por dicho trámite no corresponda al que realmente se debería devolver al adulto mayor. Mientras que, el valor 0 muestra que el trámite no posee características que le cataloguen como riesgoso. Según esta variable, el 17,6% del total de trámites tienen riesgo de fraude, mientras que el resto no.

[34] Una variable dummy es una variable no métrica o cualitativa, que se transforma a métrica, asignando el valor de 1 o 0, dependiendo si se cumple o no determinada característica (Hair, et al., 2009, p. 2)

Tabla 10: Clasificación de trámites según variable objetivo "RIESGO"

RIESGO	FRECUENCIA	PORCENTAJE
0	276.448	82,4%
1	59.188	17,6%
Total	335.636	100,0%

Fuente: SRI
Elaboración: Los autores

4.4.2 ESPECIFICACIÓN DEL MODELO

El modelo que se escoge para explicar el comportamiento de los trámites de devolución con alto riesgo de fraude es un modelo logit. Según Hair (1999), los modelos logit son una técnica recomendable para explicar el comportamiento de una variable binaria ya que descansan sobre supuestos flexibles a diferencia de otras técnicas de minería de datos. De igual manera, el modelo logit es el más sencillo al momento de su implementación en el sistema del SRI.

Este modelo posee la siguiente especificación:

$$P_i(RIESGO = 1) = \frac{1}{1 + e^{-Z_i}} = \frac{e^{Z_i}}{1 + e^{Z_i}}$$

Con $Z_i = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k$

Donde:

P_i: Probabilidad de que el trámite pertenezca el grupo de trámites con riesgo

X_k: Variables explicativas (dadas en la tabla 4)

β_k: Coeficientes de las variables explicativas.

4.4.3 ESTIMACIÓN DEL MODELO

Como se mencionó en el capítulo 3, los métodos de búsqueda secuenciales permiten estimar el mejor modelo posible, al analizar la inclusión de variables explicativas, de acuerdo a determinados criterios. En este estudio se considera como criterio de inclusión de variables el método forward o de adición progresiva;

incluyendo en cada paso la variable que presente el mayor score statistic y un α menor a 0,05. Se busca además que la inclusión de cada variable reduzca el valor del estadístico -2LL y mejore los demás estadísticos de bondad de ajuste.

En este caso, el método forward realizó siete pasos; el mejor modelo es el resultante del último paso, con las variables explicativas: Snt, Ventanilla, Género, Edad, Número de Periodos, Monto solicitado (Ln), Número de Facturas y RUC activo, todas estadísticamente significativas a un nivel de confianza del 95%; La variable "Correo Repetido" no fue incluida en el modelo. (Ver anexo F).

El modelo final brinda un poder de clasificación global de 92,4%, y de 72,4% para el grupo de riesgo (tabla 11); además los estadísticos de bondad de ajuste son los mejores a comparación de los pasos anteriores (tabla 12).

Tabla 11: Tabla de Clasificación

Observado			Pronosticado		
			RIESGO		Corrección de porcentaje
			0	1	
Paso 7	RIESGO	0	213766	7351	96,7 %
		1	13051	34220	72,4 %
	Porcentaje global				92,4 %

Fuente: SRI
Elaboración: Los autores

Tabla 12: Estadísticos Bondad de Ajuste

Paso	Logaritmo de la verosimilitud -2	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	99633,668[a]	42,9 %	70,8 %
2	91719,850[b]	44,5 %	73,5 %
3	89554,793[b]	45,0 %	74,2 %
4	89044,882[b]	45,1 %	74,4 %
5	88835,740[b]	45,1 %	74,5 %
6	88825,261[b]	45,1 %	74,5 %
7	88817,460[b]	45,1 %	74,5 %
a. La estimación ha terminado en el número de iteración 8 porque las estimaciones de parámetro han cambiado en menos de 0,001.			
b. La estimación ha terminado en el número de iteración 9 porque las estimaciones de parámetro han cambiado en menos de 0,001.			

Fuente: SRI
Elaboración: Los autores

Cabe mencionar que los resultados obtenidos mediante el método forward son idénticos si se emplea el método backward, debido a que en ambos métodos el mejor modelo es aquel que incluye a las 7 de las 8 variables explicativas, siendo todas ellas estadísticamente significativas.

4.4.4 BONDAD DE AJUSTE DEL MODELO SELECCIONADO

Como se nota, el modelo escogido es aquel que posee las mejores medidas de bondad de ajuste entre los demás. Según la tabla 12, este modelo tiene el logaritmo de verosimilitud (-2LL) más bajo, lo cual evidencia su buen ajuste.

El R cuadrado de Cox y Snell indica que el 45,1% de la variación de la variable "RIESGO", es explicado por las variables explicativas. Por otro lado, la corrección de escala del R cuadrado de Nagelkerke arroja un valor 74,5%, que al aproximarse a 100%, muestra que el modelo posee un ajuste adecuado.

4.5 INTERPRETACION DE RESULTADOS

4.5.1 INTERPRETACIÓN

Una vez estimado y seleccionado el mejor modelo, se tienen los siguientes resultados:

Tabla 13: Resultados del modelo estimado

VARIABLES	β	Error estándar	gl	Sig.	Exp(B)
Número de Facturas	0,003	0,000	1	0,000	1,003
Tipo Ingreso (SNT)	-0,579	0,068	1	0,000	0,561
Tipo Ingreso (Ventanilla)	-0,452	0,020	1	0,000	0,636
Ruc Activo (si)	0,923	0,020	1	0,000	2,516
Género (Mujer)	-0,048	0,017	1	0,005	0,953
Edad	-0,004	0,001	1	0,002	0,996
Número de Periodos	-0,338	0,004	1	0,000	0,713
Monto_Solicitado_Ln	7,133	0,040	1	0,000	1.252,722
Constante	-37,862	0,232	1	0,000	0,000

β: Coeficiente estimado logístico; gl: grados de libertad; Sig: significación o p-valor; Exp (β): coeficiente exponenciado.
Fuente: SRI
Elaboración: Los autores

Formalmente, el modelo logit estimado está dado por:

$$\mathrm{P}(Y_i = 1) = \frac{1}{1 + \ e^{-(\mathrm{Z})}}$$

Donde:

$$\begin{aligned} Z = -37,862 - 0{,}579\ Snt - 0{,}452\ Ventanilla - 0{,}048\ Mujer - 0{,}004\ Edad \\ - 0{,}338\ Número\ de\ Periodos + \ 7{,}133\ Monto\ Solicitado\ (Ln) \\ + 0{,}003\ Número\ de\ Facturas + 0{,}923\ RUC\ Activo \end{aligned}$$

Se observa que las 7 variables independientes logran explicar el comportamiento riesgoso de los trámites, ya que son estadísticamente significativas a un nivel de confianza del 95% (poseen un p-valor[35] menor a 0,05) (tabla 13).

El signo de los coeficientes estimados β_i indica la dirección del efecto que tienen las variables explicativas sobre la variable endógena. De esta manera, las variables “Snt”, “Ventanilla”, “Género”, “Edad” y “Número de periodos”, al tener signo negativo, poseen una relación inversa con la variable “RIESGO”; es decir, a medida que el valor de estas variables aumenta, la probabilidad de riesgo de cada trámite disminuye. Por otro lado, las variables “Monto Solicitado (Ln)”, “RUC Activo” y “Número de Facturas” al tener un coeficiente estimado de signo positivo guardan una relación directa, es decir que mientras dichas variables aumentan, aumenta al mismo tiempo la probabilidad de riesgo que posee el trámite.

Para conocer ceteris-paribus la magnitud del impacto que causan cada una de las variables explicativas sobre la variable “RIESGO”, se analiza la razón de momios, o por nombre en inglés odds ratio; mismo que representa la probabilidad de que ocurra un trámite riesgoso divido para la probabilidad de que no ocurra. En este sentido, se tiene:

- El odd ratio de la variable “Monto Solicitado (Ln)”, indica que por cada punto porcentual adicional en el monto solicitado por trámite, la proporción

[35] El p-valor es el menor nivel de significancia al que se habría rechazado la hipótesis nula en un test de significancia individual (Ho: β_j=0). Se suele comparar el p-valor con el nivel de significancia (α) escogido por parte del investigador, que generalmente toma el valor de 0,05. Si el p-valor es menor que α, se rechaza la hipótesis nula y por la tanto el parámetro β_j es estadísticamente significativo (Wooldridge, 2010, p. 776).

de trámites con riesgo aumenta en 125.172% manteniendo los demás factores constantes.

- La variable “Edad” muestra que por cada año de edad adicional, la proporción de trámites con riesgo disminuye en (1-0,996) 0,4% manteniendo los demás factores constantes.
- En cuanto a la variable “Número de Periodos” se tiene que por cada periodo adicional presente en cada trámite, la proporción de trámites con riesgo se reduce en (1 – 0,713) 28,7% manteniendo los demás factores constantes. Se nota que la variable “Número de Periodos” posee una mayor influencia dentro del modelo, a comparación de la variable “Edad”.
- La variable “Número de Facturas”, indica que por cada factura adicional presente en cada trámite, la proporción de trámites con riesgo aumenta en (1,003 – 1) 0,3% manteniendo los demás factores constantes. Ésta variable al igual que la variable “edad”, no causan un gran impacto sobre la variable objetivo.
- La variable “Tipo Ingreso” indica que la probabilidad de que un trámite realizado por el “Sistema Nacional de Trámites” sea riesgoso es 0,56 veces menor, que la probabilidad de que el trámite sea riesgoso cuando es realizado por internet. Mientras que, la probabilidad de que un trámite realizado por “Ventanilla” sea riesgoso es 0,64 veces menor, que la probabilidad de que el trámite sea riesgoso cuando es realizado por internet, manteniendo los demás factores constantes.
- La variable "Género" indica que la probabilidad de que un trámite sea riesgoso al ser realizado por una mujer es 0,95 veces menor, que la probabilidad de que el trámite sea riesgoso al ser realizado por un hombre, manteniendo los demás factores constantes.
- La variable “RUC Activo” indica que la probabilidad de que un trámite sea riesgoso al ser realizado por un beneficiario que posee RUC activo es 2,52 veces mayor, que la probabilidad de que un trámite sea riesgoso al ser realizado por un beneficiario que no posee RUC activo.

Otra manera de conocer la magnitud del impacto que tienen las variables explicativas sobre la dependiente, es a través de los efectos marginales. Mismos que muestran como varía la probabilidad de pertenecer al grupo de riesgo, ante una variación de una unidad de la variable independiente, manteniendo todos los demás términos constantes y considerando las características de un trámite promedio.

Tabla 14: Efectos marginales

VARIABLE	EFECTO MARGINAL	ERROR ESTÁNDAR	z	p > |z|	95% INTERVALO DE CONFIANZA	
Número de periodos	-0,09679 %	0,0000514	-18,85	0,0000	-0,0010685	-0,0008672
Número de Facturas	0,0000089 %	0,00000126	7,10	0,0000	0,00000649	0,0000114
Edad	-0,00133 %	0,00000827	-1,60	0,109	-0,0000295	0,0000029
Monto solicitado (ln)	2,069 %	0, 0010322	20,05	0,0000	0,0186763	0,0227224
Ventanilla	-0,13036 %	0,0001582	-8,24	0,0000	-0,0016136	-0,0009935
SNT	-0,14719 %	0,0003615	-4,07	0,0000	-0,0021805	-0,0007633
Género (mujer)	-0,00306 %	0,0001008	-0,30	0,762	-0,0002281	0,0001669
RUC Activo	0,37538 %	0,0002855	13,15	0,0000	0,0031943	0,0043134

Fuente: SRI
Elaboración: Los autores

Tabla 15: Características de un trámite promedio

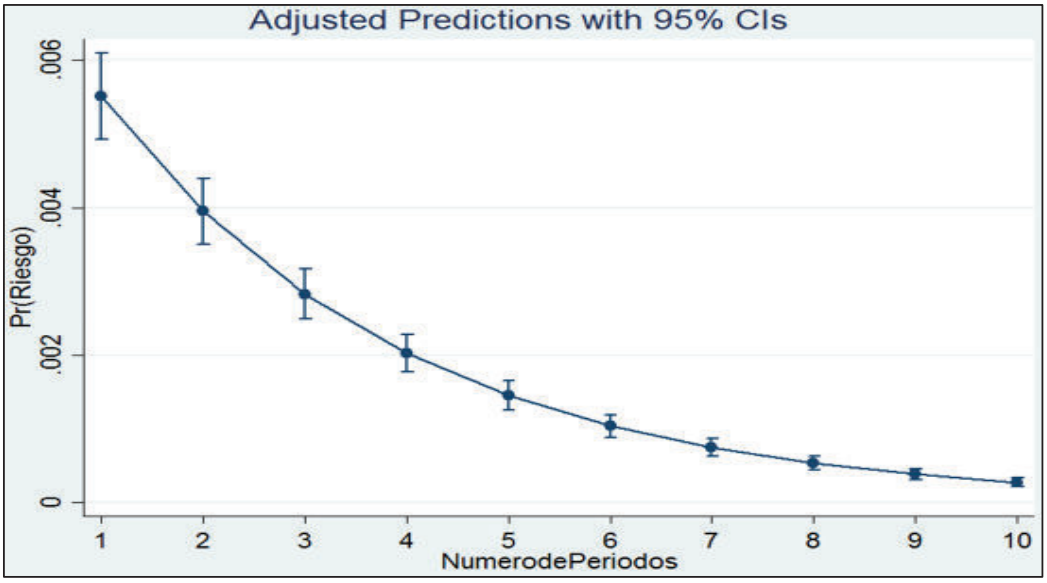
VARIABLE	PROMEDIO
Edad	73
Número de Periodos	3
Monto Solicitado	$ 108,31
Número de Facturas	45
Ventanilla	74,19 %
SNT	1,8 %
Género (mujer)	57,33 %
RUC Activo (si)	20,64 %

Fuente: SRI
Elaboración: Los autores

Si se analiza la variable “Número de Periodos”, se tiene que ante un incremento de una unidad, la probabilidad de riesgo que posee el trámite disminuye en 0,09%, considerando un individuo que realiza en promedio un trámite con 3

periodos. La figura 13 muestra dicho efecto marginal y además la relación inversa existente entre ambas variables.

Figura 13: Efecto marginal de la variable "Número de periodos"



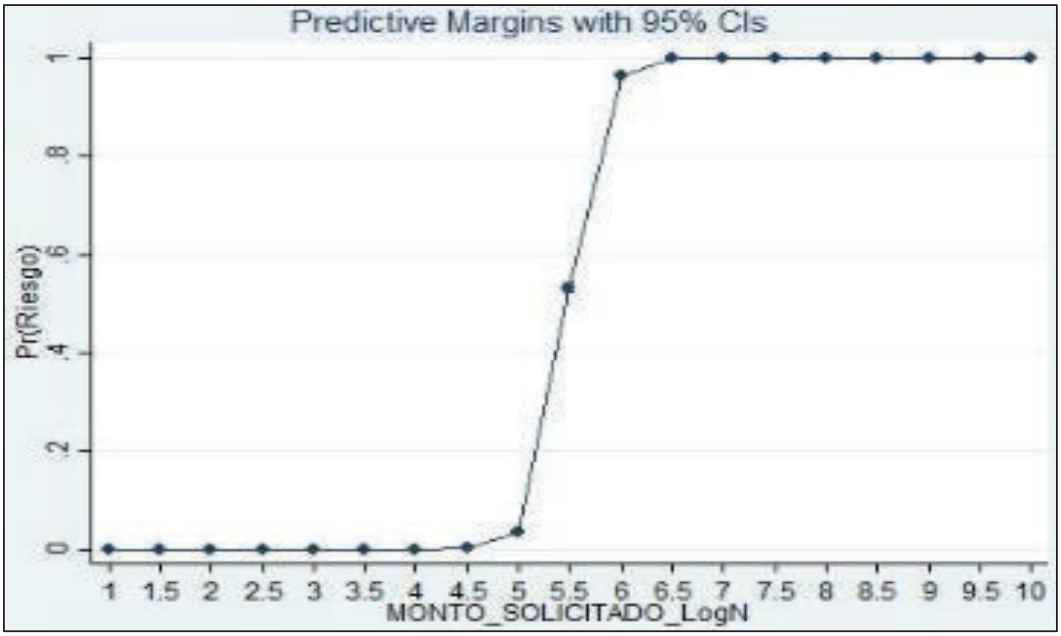
Fuente: SRI
Elaboración: Los autores

La variable "Número de Facturas", muestra que ante un incremento de una factura dentro del trámite, la probabilidad de riesgo aumenta en 0,0000089%, si se considera que en promedio un trámite presenta 45 facturas. Se nota que a pesar de que la variable es estadísticamente significativa, no causa un gran impacto en la probabilidad de riesgo de un trámite.

En cuanto a la variable "Edad", se tiene que la misma al igual que la variable "Número de Facturas", no influye de manera significativa en el nivel de riesgo de cada trámite, ya que su efecto marginal es de apenas -0,00133 %.

La variable "Monto solicitado (Ln)", muestra que ante un aumento de una unidad porcentual en el monto solicitado, el porcentaje de riesgo aumenta en 2,07% para un trámite, considerando que en promedio, un trámite posee un monto solicitado de 4,68 puntos porcentuales ($108,31); así lo muestra la figura 14, donde además se nota la relación directa entre ambas variables.

Figura 14: Efecto marginal de la variable "Monto solicitado (Ln)"



Fuente: SRI
Elaboración: Los autores

El hecho de que un trámite sea realizado por "Ventanilla" o por "SNT", reduce la probabilidad de riesgo en 0,13% y en 0,14% respectivamente, a comparación de realizar el trámite vía internet, y considerando las características de un trámite promedio.

Si un trámite es realizado por una mujer, la probabilidad de riesgo disminuye en 0,0031%, en comparación de cuando es realizado por un hombre, y considerando las características de un trámite promedio.

Finalmente, en cuanto a la variable "RUC Activo", se tiene que cuando el beneficiario tiene RUC activo, la probabilidad de riesgo aumenta en 0,38%, en comparación de cuando el beneficiario no tiene RUC activo, y considerando las características de un trámite promedio.

4.5.2 CAPACIDAD CLASIFICATORIA

El poder de clasificación del modelo para datos de entrenamiento, a nivel global es de 92,4%, mostrando que de cada 100 trámites que ingresan al SRI solicitando devolución del IVA, 92 son correctamente clasificados. Si se enfatiza

el análisis para el grupo de trámites riesgosos, se tiene un poder de clasificación del 72,4%, es decir, de cada 100 trámites riesgosos, el modelo clasifica correctamente 72 de ellos (tabla 16).

Tabla 16: Tabla de clasificación - datos de entrenamiento

DATOS ENTRENAMIENTO				
		PRONOSTICADO		PORCENTAJE CORRECTO
OBSERVADO		RIESGO		
		0	1	
Riesgo	0	213766	7351	96,70%
	1	13051	34220	72,40%
Porcentaje Global				92,40%

Fuente: SRI
Elaboración: Los autores

Por otro lado, si se considera los datos de prueba, se tiene que existe un poder clasificatorio del 92,4% de manera global, y para el grupo de trámites de riesgo un 72,4% (tabla 17). Estos porcentajes no difieren de los obtenidos con los datos de entrenamiento, lo que indica que el modelo realiza una adecuada clasificación y puede ser generalizado a toda la población.

Tabla 17: Tabla de clasificación - datos de prueba

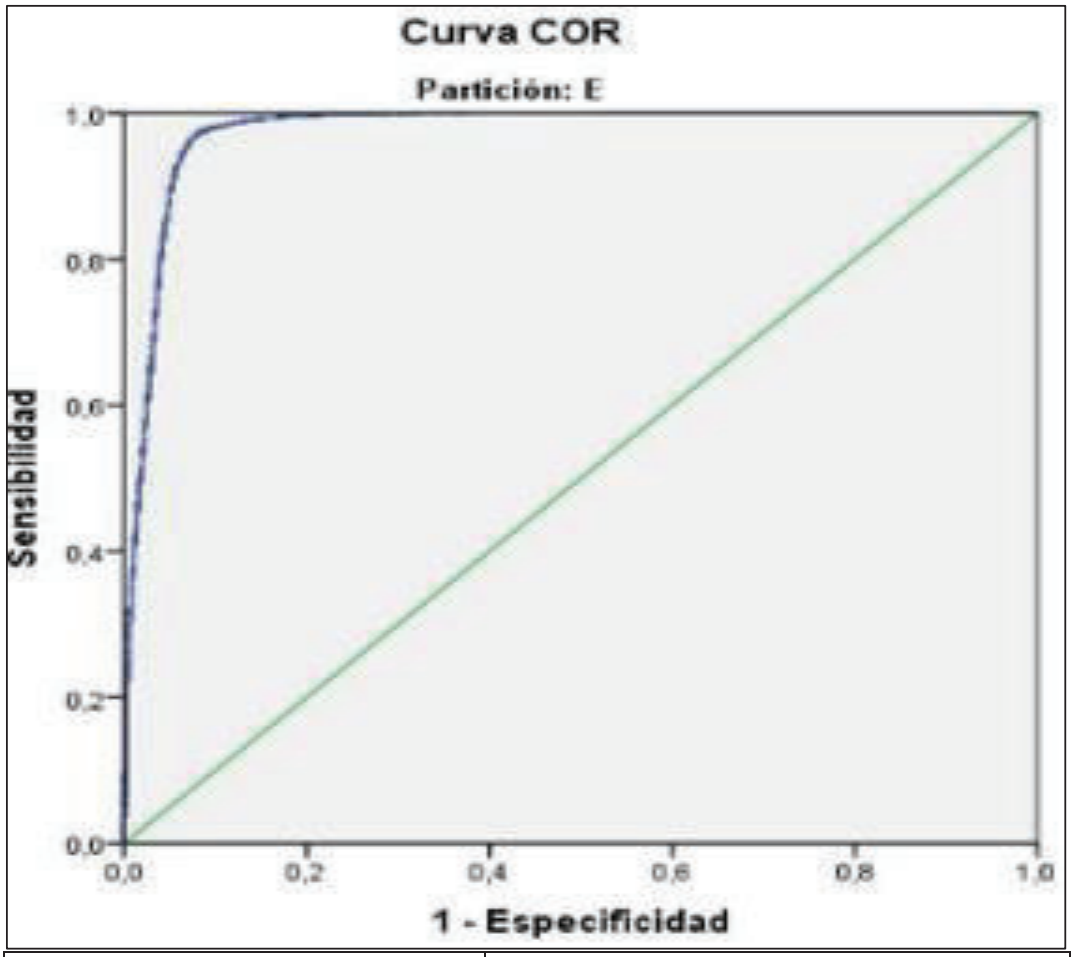
DATOS PRUEBA				
		PRONOSTICADO		PORCENTAJE CORRECTO
OBSERVADO		RIESGO		
		0	1	
Riesgo	0	53506	1825	96,70%
	1	3292	8625	72,40%
Porcentaje Global				92,40%

Fuente: SRI
Elaboración: Los autores

Otra manera de medir la capacidad de clasificación del modelo, es mediante la CURVA ROC, la misma que permite conocer la relación entre la *especificidad* y la *sensibilidad*. Mientras mayor sea el área bajo la curva, se tendrá una mayor capacidad clasificatoria entre los grupos de riesgo y no riesgo. Por tanto, se puede observar en las figuras 15 y 16, que el área bajo la curva ROC para datos

de entrenamiento y prueba es del 0,98 y 0,97 respectivamente, lo que demuestra que es un modelo que clasifica correctamente los trámites de devolución.

Figura 15: Curva ROC (datos entrenamiento)

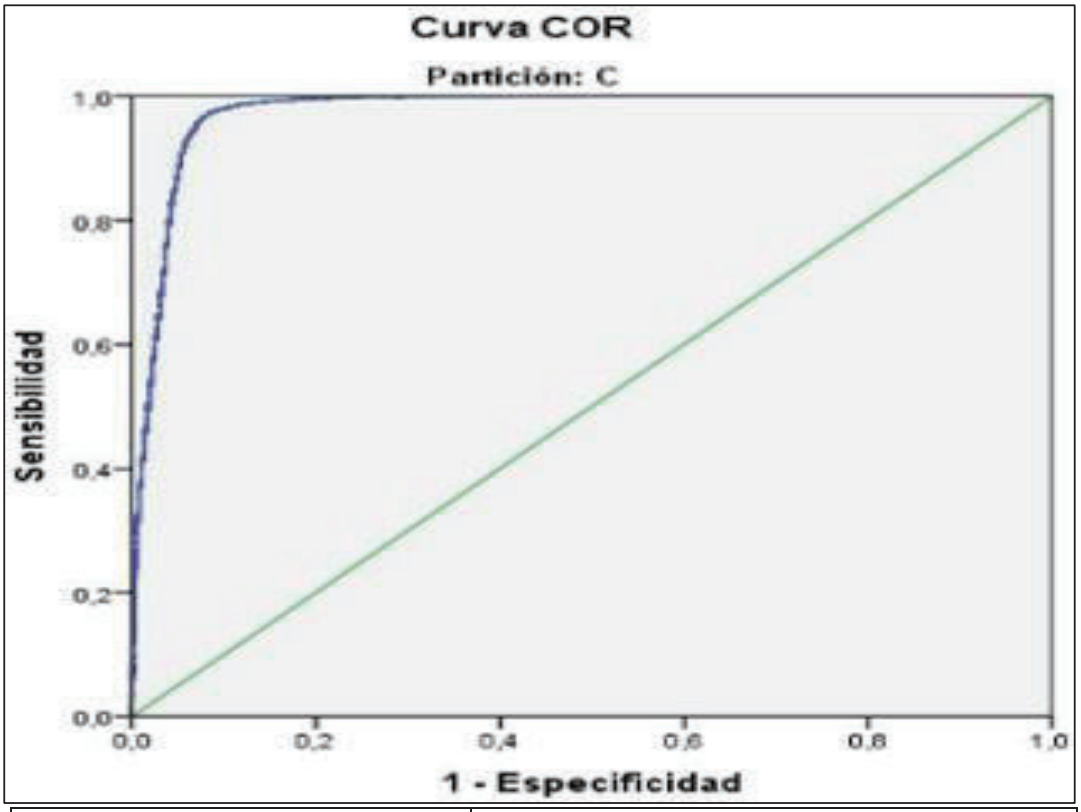


		95% de Intervalo de Confianza Asintótico	
ÁREA	**SIGNIFICACIÓN**	**LÍMITE INFERIOR**	**LÍMITE SUPERIOR**
0,976	0.00	0,975	0,976

Fuente: SRI
Elaboración: Los autores

De igual forma, la curva ROC para datos de prueba muestra resultados similares.

Figura 16: Curva ROC (datos prueba)



		95% de Intervalo de Confianza Asintótico	
ÁREA	**SIGNIFICACIÓN**	**LÍMITE INFERIOR**	**LÍMITE SUPERIOR**
0,975	0.00	0,974	0,976

Fuente: SRI
Elaboración: Los autores

4.6 EVALUACIÓN E IMPLEMENTACIÓN

Una vez estimado el modelo, se lo puede evaluar con un conjunto externo de datos para su posterior implementación dentro de los procesos del SRI. Dicha acción se debe realizar principalmente con la participación de expertos en devolución de IVA, quienes validarán los resultados de la aplicación del modelo desarrollado, con el fin de definir estrategias de control diferenciadas, control previo y control posterior. Por tanto esta última etapa será desarrollada por el SRI. Cabe mencionar que es fundamental recopilar todos los resultados generados dentro de esta etapa, para obtener información que sirva de retroalimentación para el modelo, y por tanto llegar a tener una mejor estimación futura.

5 CONCLUSIONES Y RECOMENDACIONES

5.1 CONCLUSIONES

Para lograr estimar el modelo, y siguiendo la metodología CRISP DM, se empezó realizando una breve indagación acerca de la gestión y política tributaria, y la manera en cómo el Estado utiliza éstas herramientas para realizar una oportuna intervención, que, entre otras cosas, genere un sistema tributario progresivo, brinde atención a grupos prioritarios, y aliente a los contribuyentes a cumplir con sus obligaciones tributarias.

La comparación de la recaudación de impuestos entre América Latina y los países miembros de la OCDE, mostró la influencia de los impuestos directos en especial el impuesto a la renta en el total de recaudación de los países de la OCDE, a diferencia de la recaudación tributaria en los países de América Latina, donde el IVA es el tributo con mayor peso.

Se realizó una comparación entre algunos países de la región para tener una visión global acerca de los principales beneficios otorgados a los adultos mayores a nivel de América Latina; en general, la mayoría de estos países poseen programas de atención a adultos mayores, vinculados a seguridad social, salud, vivienda, tributación, entre otros. Muchos de éstos países brindan subsidios económicos o bonos a adultos mayores; mientras que otros, como Ecuador brindan exenciones o devoluciones tributarias.

En cuanto al proceso de devolución del IVA, el cambio de enfoque en el procedimiento es sin duda un elemento clave en el proceso de gestión de riesgos dentro del SRI, puesto que al pasar de proceso con un arduo control previo, a uno con un mínimo control de verificación de información, se facilita el hecho que los contribuyentes tiendan a realizar algún tipo de fraude con el objetivo de obtener un mayor beneficio económico personal; provocando que se destinen mayores ingresos estatales a cubrir el funcionamiento de dicho programa, y se distorsione de esta manera su objetivo principal.

En lo que respecta al modelo, el modelo logit estimado es consistente en sus resultados, ya que tanto para los datos de prueba como para los de

entrenamiento, la capacidad clasificatoria se mantiene; lo que indica que los resultados del modelo son consistentes, y pueden generalizarse.

Los resultados obtenidos muestran que todas las variables independientes utilizadas en la estimación del modelo logit, contribuyen a explicar el nivel de riesgo que existe en cada trámite al ser estadísticamente significativas. Se evidenció que la mayor influencia en el nivel de riesgo que posee cada trámite, está determinado por el monto solicitado; siendo dicha variable una de las más importantes al momento de estimar el modelo, resultado que corrobora la primera hipótesis planteada; mientras más alto sea el valor solicitado por trámite, mayor será su probabilidad de pertenecer al grupo de riesgo.

Se notó que el transformar la variable “monto solicitado” a logaritmo natural, hace que la capacidad clasificatoria del modelo mejore significativamente, tanto para el grupo de los riesgosos como de manera global del modelo, cumpliendo de esta manera con el objetivo planteado conjuntamente con la Administración Tributaria, de realizar un modelo que clasifique adecuadamente los trámites.

La variable “Edad” si bien es estadísticamente significativa, no causa un gran efecto en el aumento o disminución de la probabilidad de riesgo de cada trámite, puesto que su efecto marginal es sumamente pequeño; razón por la cual, se concluye que el hecho de que un beneficiario haga algún tipo de fraude, no se encuentra ligado a una determinada edad. Por lo que la inclusión o no de esta variable dentro del modelo queda a criterio de la Administración Tributaria.

Las variables “Número de facturas” y “RUC activo”, al mantener una relación directa con la variable objetivo “RIESGO”, evidencian el cumplimiento de las dos hipótesis restantes; por lo que se concluye que mientras más facturas se presenten por trámite, y el hecho que el trámite sea realizado por un individuo que posea RUC activo, son factores que provocan un aumento en la probabilidad de que el trámite sea clasificado en el grupo de riesgo.

Finalmente el modelo logit estimado al poseer una capacidad predictiva del 92,4% a nivel global muestra que el modelo clasifica de una manera adecuada a los trámites en los dos grupos; lo que indica que el modelo logra evidenciar de manera objetiva los patrones de conducta identificados de manera empírica por

los expertos del SRI. De esta manera, el modelo puede ser generalizado para la población de trámites de otros años.

5.2 RECOMENDACIONES

Luego de haber realizado la presente investigación se recomienda, para futuros trabajos o actualizaciones del modelo de riesgo tributario, en primer lugar se consideren variables que muestren el nivel de ingreso económico que poseen los adultos mayores solicitantes de la devolución, ya que dicha variable sería de gran ayuda para realizar un contraste entre lo que el adulto mayor solicita por mes por concepto de devolución del IVA, y lo que su capacidad adquisitiva efectivamente le permite solicitar.

En segundo lugar, otra de las variables que serviría para enriquecer el presente modelo, es la fecha de defunción del adulto mayor, ya que se evidenció en el análisis exploratorio de datos, que las bases de datos entregadas por el SRI consideraban en sus registros a adultos mayores de más de 90 años, llegando inclusive a tener casos de beneficiarios con más de 100 años que efectivamente solicitaban cada mes el monto máximo permitido.

Se presentó el caso de incumplimiento de edad mínima para acceder al beneficio de devolución, es decir, existieron algunos trámites realizados por individuos de menos de 65 años; por lo que se recomienda que la Administración Tributaria mantenga constantemente actualizadas sus bases de datos y realice un proceso de cruce de información, con el fin de verificar si se cumplen con los requisitos básicos para acceder al proceso de devolución.

En cuanto a la calidad de los datos, las bases reflejan la información de los beneficiarios, sin embargo la explotación de dicha información está limitada a los requerimientos de la Administración Tributaria, por lo que en ocasiones se necesita información que el contribuyente indirectamente proporciona pero que la Administración no la actualiza en sus bases de datos. Esto ocasiona que la investigación se vea limitada a la información disponible en ese momento y que, por lo tanto, se estime un modelo que tenga un alcance limitado, pues al no

contar con la información real de las variables, se deberá hacer uso de variables proxy para cubrir la información faltante. Es el caso de las variables “número de facturas” y “número de correos repetidos”, mismas que no reflejan los valores reales en cada uno de los trámites realizados, sino que son una aproximación. Se recomienda por tanto que el SRI estandarice o implemente un proceso de recolección de información adjudicada por los beneficiarios.

En cuanto a la cantidad de los datos, se recomienda recolectar la mayor cantidad de información posible en lo que se refiere a años posteriores al 2015, para analizar el comportamiento de los individuos no solo en un periodo específico, sino a través de los años, con el fin de definir de mejor manera la evolución de su comportamiento tributario, y poder realizar una estimación que sea lo más cercana a la realidad.

Una vez que el SRI posee la estimación del modelo, se recomienda realizar un agrupamiento de dichas probabilidades de riesgo estimadas, con el fin de establecer un score de riesgo, que mediante la definición de rangos de puntuaciones, se clasifique a los trámites en probabilidades de riesgo alta, media y baja, con el propósito de caracterizar dichos grupos y establecer perfiles de riesgo, que permitan identificar no solo a los trámites con riesgo, sino ademas al contribuyente que se encuentra realizando fraude a la Administración.

BIBLIOGRAFÍA

- Alink, M. & Kommer, V. V., 2011. *Manual de Administración Tributaria.* Amsterdam: IBFD.
- Amemiya, T., 1981. Qualitative Response Models: A Survey. *Journal of Economic Literature ,* Volumen 19, pp. 1483-1536.
- Chapman, P. y otros, 1999. *CRISP-DM 1.0 Step-by-step data mining guide.* U.S.A.: SPSS Inc..
- CIAT, Diplomado en Tributación, 2016. *Diplomado en Tributación.* s.l.:s.n.
- CIAT, 2015. *CIAT Data Recaudación.* [En línea] https://ciatorg-public.sharepoint.com/_layouts/15/WopiFrame.aspx?sourcedoc=%7B2164CE86-B0D3-4356-9071-6243B79A8B1C%7D&file=CIATData-Recaudacion.xlsx&action=default
- Código Tributario, 2005.
- Consorcio Colombia Mayor, 2016. *Colombia Mayor Consorcio, 2013.* [En línea] https://colombiamayor.co/quienes_somos.html
- Constitución del Ecuador, 2008. *Constitución de la República del Ecuador.*
- Corbacho, A., Fretes, V. & Lora, E., 2012. *Recaudar No Basta, Los impuestos como instrumento de desarrollo.* Washington D.C.: Pórtico Bookstore.
- Decreto Legislativo N° 776, 1994. *Ley de Tributación Municipal.* Lima
- Decreto N°050-2010, 2010. *Programa Piloto de Asistencia Solidaria: Gratitud.* Lima
- Decreto Supremo N° 081-2011-PCM, 2011. *Programa Social denominado Programa Nacional de Asistencia Solidaria "Pensión 65".* Lima

- European Commission, 2010. *Compliance Risk Management Guide for Tax Administrations,* European Union.
- Fawcett, T., 2005. An introduction to ROC analysis. *Pattern Recognition Letters,* pp. 861-862.
- Fundación para la Educación Superior y el Desarrollo; Fundación Saldarriaga Concha, 2015. *Misión Colombia Envejece. Cifras, Retos y Recomendaciones.,* Bogotá: Fundación Saldarriaga Concha.
- Gallo Escobar , H. & Cuartas Mejia , V., 2006. *Diccionario Económico Financiero.* [En línea] https://books.google.com.ec/books?id=O2U5GgwgjKsC&pg=PP2&lpg=PP2&dq=Diccionario+Econ%C3%B3mico+Financiero+gallo+cuartas&source=bl&ots=cxtPkeqZ_7&sig=iM415ZYH9aS1-9PDeKOAapYW3sc&hl=es&sa=X&ved=0ahUKEwir2aG_udnRAhXD5CYKHfakCasQ6AEIIjAC#v=onepage&q=Diccionar
- Gorunescu, F., 2011. *Data Mining. Concept, Models and Techniques.* 12 ed. Berlin: Springer.
- Gujarati, D., 2003. *Econometría.* 4 ed. México D.F.: McGraw-Hill Interamericana.
- Hair, J., Anderson, R., Tatham, R. & Black, W., 1999. *Análisis Multivariante.* 5 ed. Madrid: Prentice Hall Iberia.
- Hair, J., Black, W., Babin, B. & Anderson, R., 2009. *Multivariate Data Analysis.* Séptima ed.Prentice Hall.
- Hosmer, D. W. & Lemeshow, S., 2000. *Applied Logistic Regression.* 2° ed. John Wiley & Sons INC..
- IBM, 2012. *Manual de CRISP-DM de IBM SPPSS MODELER.* Chicago: IBM Corporation.
- Ley 1328, 2009. *Normas en materia financiera, de seguros, del mercado de valores, y otras disposiciones.* Bogotá: Congreso de la República.

- Ley 155, 2010. [En línea] http://www.imprenta.gov.co/gacetap/gaceta.mostrar_documento?p_tipo=05&p_numero=155&p_consec=26855
- Ley 1886, 1998. *Ley de 14 de Agosto de 1998.* La Paz
- Ley 20732, 2014. *Rebaja del Impuesto Territorial correspondiente a propiedades de Adultos Mayores Vulnerables Económicamente.* Santiago de Chile: Biblioteca del Congreso Nacional de Chile.
- Ley 27253, 2016. *Régimen de Reintegro por compras en comercios de venta minorista..* Buenos Aires: Boletín Oficial de la República Argentina.
- Ley del Anciano, 2006.
- Ley del Impuesto sobre la Renta, 2016. *Ley del Impuesto sobre la Renta.* Mexico D.F.
- LORTI, 2014. *Ley Orgánica de Régimen Tributario Interno.*
- Mitchell, T., 1997. *Machine Learning.* New York: McGraw-Hill.
- Molina López, J. M. & García Herrero, J., 2006. *Técnica de análisis de datos. Aplicaciones prácticas utilizando Microsoft, Excel y Weka.* Madrid.
- Musgrave, R. & Musgrave, P., 1992. *Hacienda Pública. Teórica y Aplicada.* Quinta ed. Madrid: McGraw-Hill.
- OCDE, 2004. *Compliance Risk Management: Managing and Improving Tax Compliance. Guidance Note,* Centre for Tax Policy and Administration.
- OCDE, BID, CEPAL & CIAT, 2016. *Estadísticas Tributarias en América Latina y el Caribe, 1990 - 2014.* [En línea] http://www.keepeek.com/Digital-Asset-Management/oecd/taxation/revenue-statistics-in-latin-america-and-the-caribbean-2016_rev_lat_car-2016-en-fr#page42
- OCDE, BID, CEPAL & CIAT, 2017. *Estadísticas tributarias para América Latina y el Caribe. 1990-2015.* París.

- Paz y Miño Cepeda, J. J., 2015. *Historia de los impuestos en Ecuador. Vision sobre el regimen impositivo en la historia económica nacional.* Quito.
- Pecho Trigueros, M., Peláez Longinotti, F. & Sánchez Vecorena, J., 2012. *Estimación del Incumplimiento Tributario en América Latina: 2000-2010.* Dirección de Estudios e Investigaciones Tributarias, CIAT.
- Pérez López, C., 2004. *Técnicas de Análisis Multivariante de Datos. Aplicaciones con SPSS.* Madrid: Pearson Educación.
- Reglamento LORTI, 2014. *Reglamento para la aplicación de la Ley de Régimen Tributario Interno.* Quito.
- Resolución, 2013. *Normas para la devolución de los valores pagados por personas adultas mayores, del impuesto al valor agregado (IVA) e impuesto a los consumos especiales (ICE), en la adquisición de bienes y/o servicios, para su uso y consumo personal.* Quito.
- Sepúlveda, C., 2004. *Diccionario de términos económicos.* [En línea] https://books.google.es/books?hl=es&id=UDcOcMhyU0MC&q=precio+relativo#v=snippet&q=precio%20relativo&f=false
- Sevilla Segura, J., 2004. *Política y Técnicas Tributarias.* Madrid: Instituto de Estudios Fiscales.
- SRI, 2012. *Plan Estratégico 2012-2015.*
- SRI, 2016. *Servicio de Rentas Internas.*
  Available at: http://www.sri.gob.ec/
- SRI, 2017. *Rendición de Cuentas. Período 2016.*
- Stiglitz , J., 2000. *La economía del sector público.* Tercera ed. Barcelona: Bosch Antoni.
- Wooldridge, J., 2010. *Introducción a la econometría. Un enfoque moderno.* 4a. ed. México D.F.: Cengage Learning Editores S.A..

ANEXOS

ANEXO A

REQUISITOS Y MODELO DE SOLICITUD DEL IVA A PERSONAS ADULTAS MAYORES

El Reglamento para la Aplicación de la Ley de Régimen Tributario Interno, en su artículo 178 establece que:

> ***Art. 178.-*** *Contenido de la solicitud de devolución de IVA.- Toda solicitud de devolución de IVA deberá ser formulada por escrito, en los formatos establecidos por la Administración Tributaria para el efecto, los que contendrán:*
> *1. La designación de la autoridad administrativa ante quien se presenta la solicitud.*
>
> *2. El nombre y apellido del solicitante y su número de cédula de identidad; en el caso de sociedades, la razón social respectiva y el número del Registro Único de Contribuyentes.*
>
> *3. Fundamentos de derecho.*
>
> *4. La petición concreta indicando mes, base imponible y valor de IVA sobre el cual se solicita devolución.*
>
> *5. La indicación del domicilio tributario, y para notificaciones, el que señalare;*
>
> *6. La indicación de la forma de pago respectiva o acreditación en cuenta; y,*
>
> *7. La firma del solicitante (persona natural o su representante legal en caso de sociedades).*
>
> *Una vez solicitada la devolución por un período determinado, no se aceptarán nuevas peticiones respecto de ese mismo período, salvo el caso de que se hayan presentado las respectivas declaraciones sustitutivas y únicamente se devolverá el impuesto por adquisiciones no consideradas en la petición inicial.*

Basados en dicho artículos, el SRI presente el siguiente formato para la devolución del IVA a personas adultas mayores.

Anexo A: Modelo solicitud IVA personas adultas mayores

SOLICITUD PARA LA APLICACIÓN DE BENEFICIOS TRIBUTARIOS A PERSONAS DE LA TERCERA EDAD

Ciudad _____, día_____ de _____ del 20____

Señor
Director Zonal / Provincial del Servicio de Rentas Internas

Yo, _____ portador de la cédula de ciudadanía N° _____, con dirección tributaria o domiciliaria en la ciudad de _____, en pleno conocimiento de las responsabilidades que podría incurrir por falsedad, simulación, ocultación o engaño, certifico que el **Listado de comprobantes físicos y electrónicos para devolución de IVA e ICE** (adjunto a la presente petición), corresponde a las adquisiciones de bienes y/o servicios que efectué para mi uso personal y que se emitieron a mi nombre.
A partir de enero de 2015, los comprobantes de venta detallados corresponden a la **adquisición de bienes y servicios de primera necesidad, de uso o consumo personal**, que fueron emitidos a mi nombre y que guardan relación con:

Vestimenta	Vivienda	Salud	Alimentación	Educación	Comunicación	Cultura	Deporte	Transporte y movilidad

Por lo expuesto, solicito que de conformidad a los límites y condiciones que se establecen en el artículo 14 de la Ley del Anciano, artículo 181 del Reglamento para la Aplicación de la Ley de Régimen Tributario Interno y el artículo innumerado que está a continuación del artículo 181 del mismo cuerpo legal, se devuelvan los valores del Impuesto al Valor Agregado – IVA e Impuesto a los Consumos Especiales – ICE.

AÑO (Ingrese el año (uno por solicitud))	MES (Marque con una X el mes o los meses que solicita)	TOTAL DE COMPROBANTES (Ingrese la cantidad total de comprobantes de cada mes)	VALOR MENSUAL DE IVA (Ingrese la suma total de IVA de las facturas de cada mes)	VALOR MENSUAL DE ICE (Ingrese la suma total de ICE de las facturas de cada mes)
20___	ENERO ☐			
	FEBRERO ☐			
	MARZO ☐			
	ABRIL ☐			
	MAYO ☐			
	JUNIO ☐			
	JULIO ☐			
	AGOSTO ☐			
	SEPTIEMBRE ☐			
	OCTUBRE ☐			
	NOVIEMBRE ☐			
	DICIEMBRE ☐			
	TOTAL			

Solicito que los valores sean acreditados mediante lo señalado a continuación y autorizo se descuente la comisión por transferencia bancaria.

DATOS PARA ACREDITACIÓN	NOMBRE DE LA INSTITUCIÓN FINANCIERA			
	NÚMERO DE CUENTA	TIPO DE CUENTA	AHORROS ☐	CORRIENTE ☐
	NOMBRE DEL TITULAR DE LA CUENTA (Como se encuentra registrado en la institución financiera)			
	NÚMERO DE IDENTIFICACIÓN C.C. / RUC / PASAPORTE DEL TITULAR DE LA CUENTA (Como se encuentra registrado en la institución financiera)			

En caso de existir problemas con la acreditación, por favor contactarme a:

DIRECCIÓN DOMICILIARIA O TRIBUTARIA	TELÉFONO		CELULAR		E-MAIL	
	CIUDAD		PARROQUIA		BARRIO	
	SECTOR		CALLE PRINCIPAL		INTERSECCIÓN	
	CASA N°.		LOTE N°		EDIFICIO	
	REFERENCIA DE UBICACIÓN					

Atentamente,

_____ _____
Firma del beneficiario / representante legal / apoderado N°. C.C.

(Para su respaldo presentar una copia del presente formulario lleno). Pag. 1/2

Fuente: SRI

LLENAR ÚNICAMENTE SI LA PRESENTACIÓN DE LA SOLICITUD LA REALIZA UNA TERCERA PERSONA

Autorizo para que realice la presentación de la solicitud para la aplicación de beneficios tributarios a personas de la tercera edad, en mi representación, el (la) Sr. (a)

_____

Portador (a) de la cédula de ciudadanía N° _____

Con quien mantengo el parentesco o relación de: _____

_____	_____
Firma del beneficiario / representante legal / apoderado	Firma de quien presenta la solicitud por el beneficiario
N°. C.C. _____	N°. C.C. _____

REQUISITOS (documentos que debe adjuntar)

REQUISITOS BÁSICOS	1. Copia de la presente solicitud para respaldo de recepción. 2. Presentación de la cédula de ciudadanía o pasaporte original del beneficiario. 3. Listado de comprobantes físicos y electrónicos para devolución del IVA e ICE a personas de la tercera edad. 4. Copia legible de la cédula de ciudadanía del beneficiario (si es primera solicitud).* 5. Copia del certificado bancario (si es primera solicitud o cambio de cuenta).*
REQUISITOS ESPECIALES	**En caso de presentación de la solicitud por una tercera persona:** - Copia legible de la cédula de ciudadanía de la persona beneficiaria. - Copia legible de la cédula de ciudadanía de la persona que ingresa la solicitud.
	En caso de que la solicitud sea firmada por un tercero: - Copia del poder general especial (representante legal / apoderado).
	En caso de que la solicitud sea presentada por herederos (si tiene devoluciones pendientes antes del fallecimiento del beneficiario): - Copia de posesión efectiva notariada e inscrita en el Registro de la Propiedad. - Copia de la partida de la defunción.
	En caso de que la solicitud sea presentada por una persona de nacionalidad extranjera no naturalizada: - Si solicita la devolución de un período anterior a la expedición de su cédula de ciudadanía, presentar la copia del documento que acredite la fecha de su inscripción en el Registro de Extranjeros del Departamento Consular del Ministerio de Relaciones Exteriores.

*Se entiende por primera solicitud, si es la primera vez que presenta este formulario.

SOLICITAR LA DEVOLUCIÓN POR INTERNET

1. Obtener la clave de ingreso a Servicios en Línea en cualquier agencia del SRI.
2. Ingresar a **www.sri.gob.ec**, escoger la opción Devolución de impuestos / IVA / Tercera edad.
3. Descargar la "Guía para llenar y enviar el formato de registro de comprobantes de venta para solicitudes enviadas por internet".
4. Llenar el "Formato para registro de comprobantes de venta para solicitudes enviadas por internet".
5. Ingresar a **www.sri.gob.ec** y escoger la opción: Servicios en Línea / Devoluciones.
6. Una vez registrada la solicitud en línea, verificar 24 horas después con el número de trámite.
7. Comprobar la acreditación en la cuenta bancaria.

1700 SRI SRI (774 -774)
en Cuenca al 04-2598-441

www.SRI.gob.ec

@SRIoficialEc

SRIECUADOR

Pag. 2/2

Fuente: SRI

ANEXO B

DEFINICIÓN DE VARIABLES

Anexo B: Definición de variables

VARIABLES	DEFINICIÓN
Mes fiscal (Periodo)	Mes del que se solicita la devolución en el trámite.
Monto Devuelto	Monto devuelto por el SRI, por cada periodo incluido en cada uno de los trámites presentados por los beneficiarios.
Monto solicitado	Monto que solicita el adulto mayor en cada periodo, incluido en cada uno de los trámites presentados por los beneficiarios.
Tipo físico	Documentación que ingresa con el trámite (hojas, cd, etc.)
Número de trámite	Identificación única del trámite realizado para el beneficiario.
Número de facturas	Número de hojas presentadas por trámite.
Tipo ingreso	Canal de ingreso del trámite en el sistema del SRI (sistema nacional de trámites “SNT”, ventanilla , internet)
Número de RUC	Identificación del beneficiario
Número de correo repetido	Indica el número de veces que el correo electrónico del beneficiario se repite en otros contribuyentes, independientemente del tipo de trámite que realicen en el SRI.
Fecha de nacimiento	Fecha de nacimiento del beneficiario
RUC activo	El Registro Único de Contribuyentes (RUC), permite conocer si un beneficiario de la tercera edad tiene actividad económica, tomando en cuenta que el beneficio de la devolución de IVA se circunscribe a las compras de uso y consumo personal de primera necesidad.
Género	Género del beneficiario
Oficina Resuelve	Agencia del SRI donde el contribuyente realiza el trámite de devolución del IVA.

Fuente: SRI
Elaboración: Los autores

La variable “Número de Facturas” es una variable proxy del valor real de número de facturas presentadas como justificación del monto solicitado por cada trámite de devolución del IVA, ya que el SRI no cuenta con dicha información a nivel tan detallado.

La variable “Correo Repetido” es una variable dummy que se crea, tomando como base a la variable “Número de Correos Repetidos”; si un individuo posee al menos un correo repetido, se le asigna la marca de 1, caso contrario 0.

La variable “Edad” se calcula mediante la resta entre la fecha actual del análisis (01 febrero 2016) y la variable “Fecha de Nacimiento”.

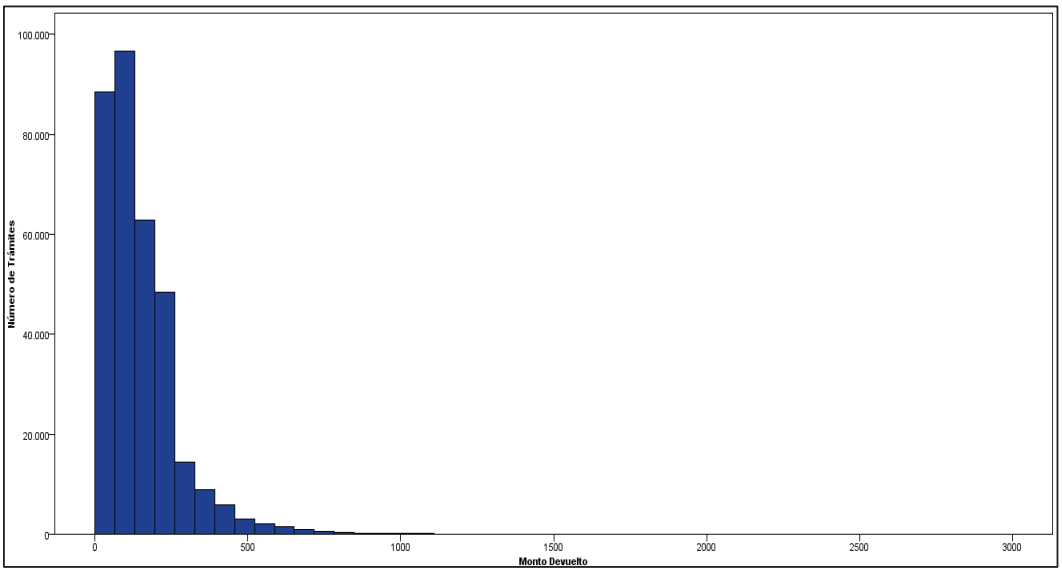
ANEXO C

ANÁLISIS EXPLORATORIO DE VARIABLES

A continuación se presentan los histogramas y las principales estadísticas de cada una de las variables agregadas a trámites, que serán utilizadas en la etapa de modelado.

- MONTO DEVUELTO

Anexo C: Análisis Exploratorio



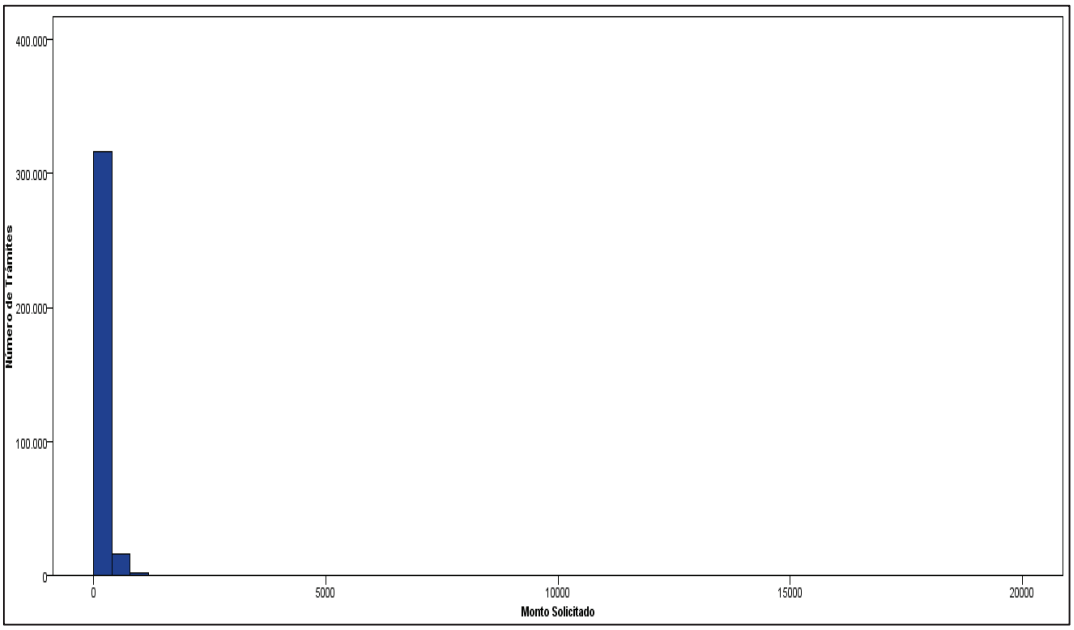
Fuente: SRI
Elaboración: Los autores

OBS	MEDIA	VARIANZA	DESVIANCIÓN ESTANDAR	KURTOSIS
335636	$ 147,64	16.988,95	130,34	21,285

MODA	MÍNIMO	MÁXIMO	SUMA TOTAL
$ 212,40	$ -	$ 2.543,06	$ 49´554.071,94

PERCENTILES							
1%	5%	10%	25%	50%	75%	90%	99%
1	5,72	20,7	62,5	116,7	199,97	282,89	645,01

- MONTO SOLICITADO



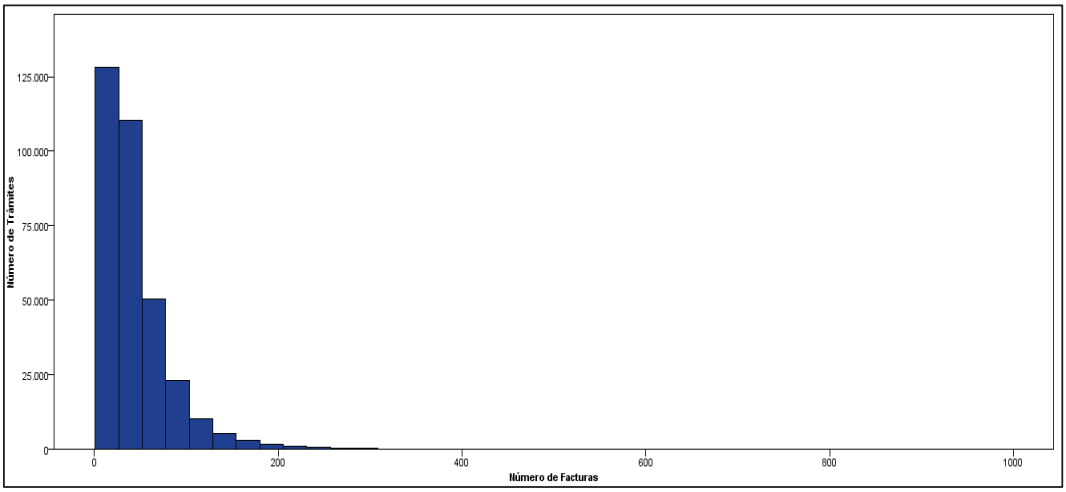
Fuente: SRI
Elaboración: Los autores

OBS	MEDIA	VARIANZA	DESVIANCIÓN ESTANDAR	KURTOSIS
335636	$ 163,61	54.662,52	233,80	340,01

MODA	MÍNIMO	MÁXIMO	SUMA TOTAL
$ 212,40	$ 0,01	$ 15.267,86	$ 54´913.770,51

PERCENTILES							
1%	5%	10%	25%	50%	75%	90%	99%
6,57	21,33	33,47	63,29	118,3	202,15	304,89	819,19

- NÚMERO DE FACTURAS



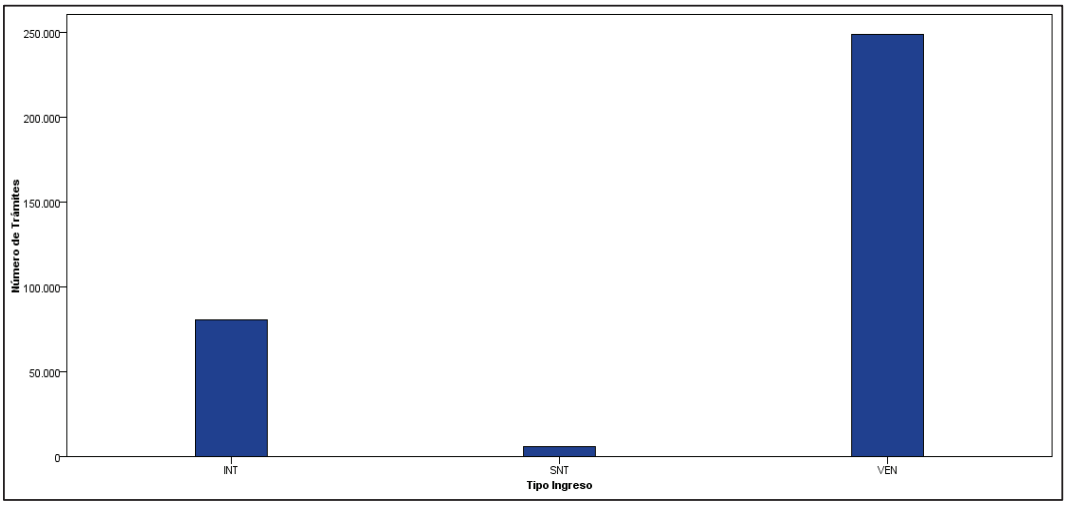
Fuente: SRI
Elaboración: Los autores

OBS	MEDIA	VARIANZA	DESVIANCIÓN ESTANDAR	KURTOSIS
335636	44,51	1.920,80	43,8	44,28

MODA	MÍNIMO	MÁXIMO	SUMA TOTAL
4	1	999	14´941.327

PERCENTILES							
1%	5%	10%	25%	50%	75%	90%	99%
2	4	7	18	34	57	90	205

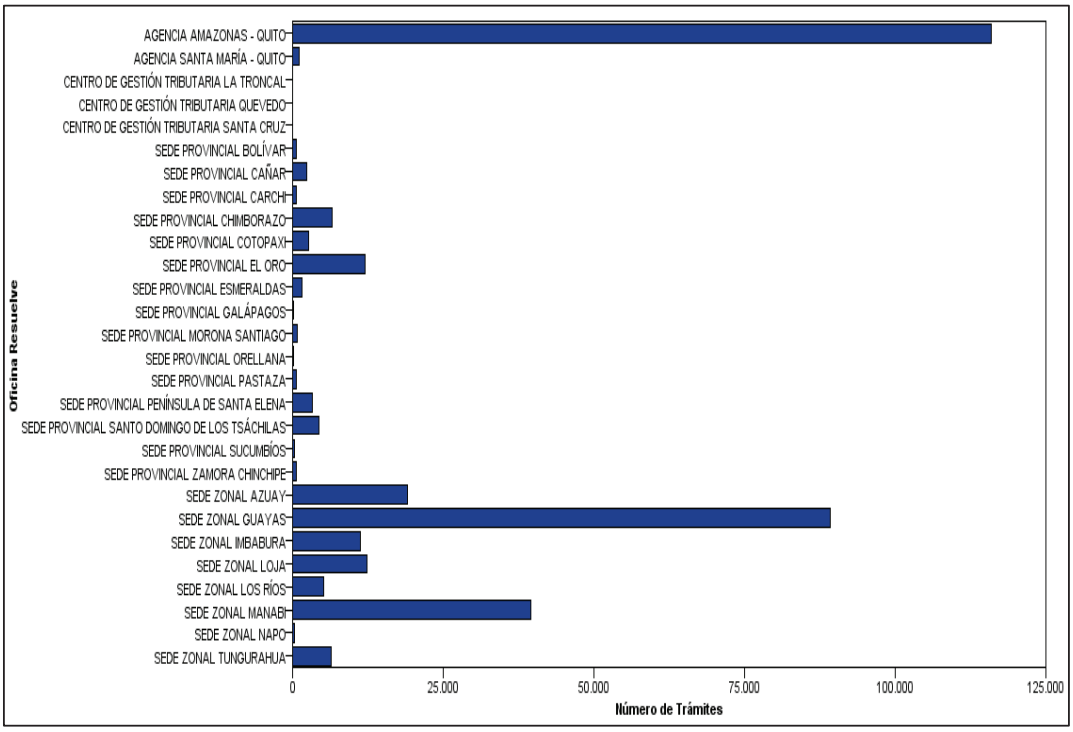
- TIPO INGRESO



Fuente: SRI
Elaboración: Los autores

CATEGORÍA_ TIPO INGRESO	NÚMERO DE TRÁMITES	% DEL TOTAL DE TRÁMITES
Ventanilla	248933	74,17
SNT	5921	1,76
Internet	80782	24,07

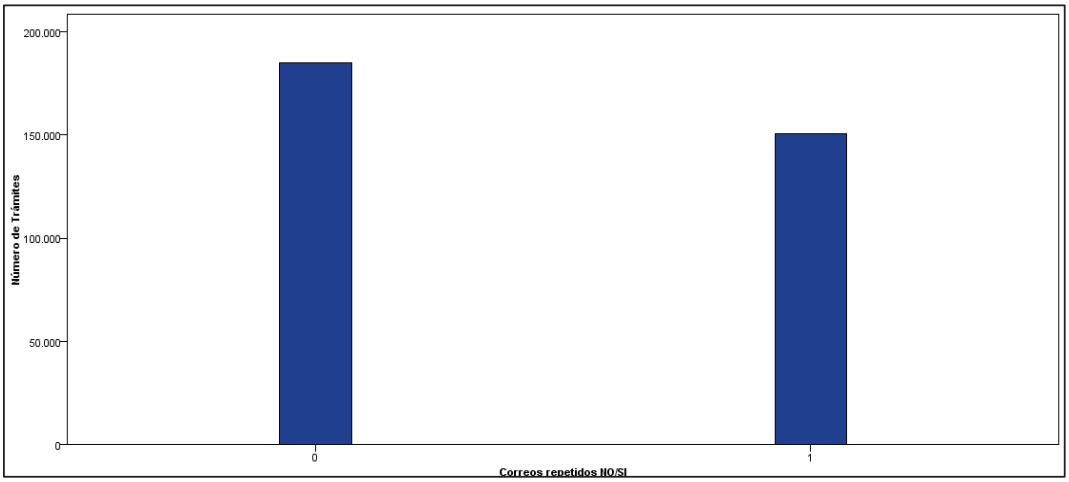
- OFICINA RESUELVE



Fuente: SRI
Elaboración: Los autores

CATEGORÍA_ OFICINA RESUELVE	NÚMERO DE TRÁMITES	% DEL TOTAL DE TRÁMITES
Quito- Amazonas	115.912	34,54
Guayas	89.290	26,6
Manabí	39.423	11,75
Otros	171.372	27,31

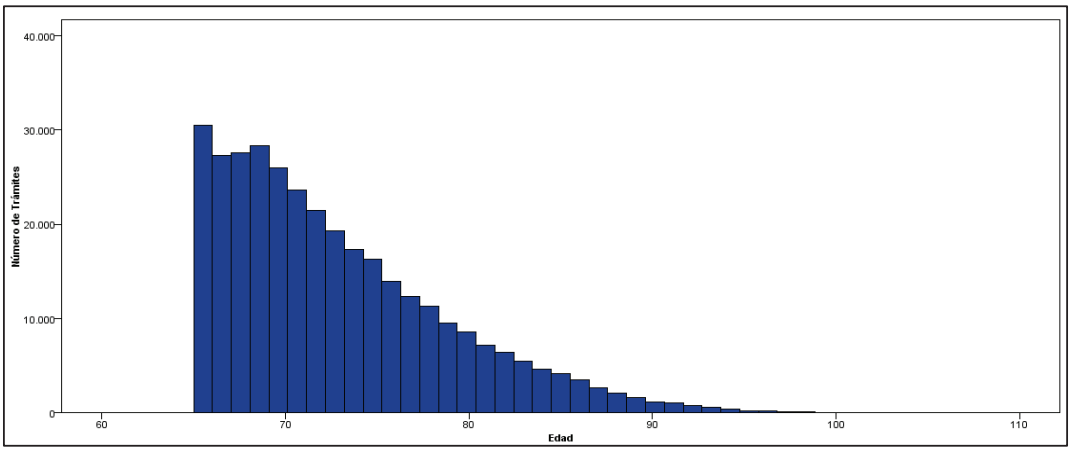
- CORREO REPETIDO



Fuente: SRI
Elaboración: Los autores

CORREO REPETIDO	NÚMERO DE TRÁMITES	% DEL TOTAL DE TRÁMITES
SI =1	150692	44,9
NO=0	184944	55,1

- EDAD



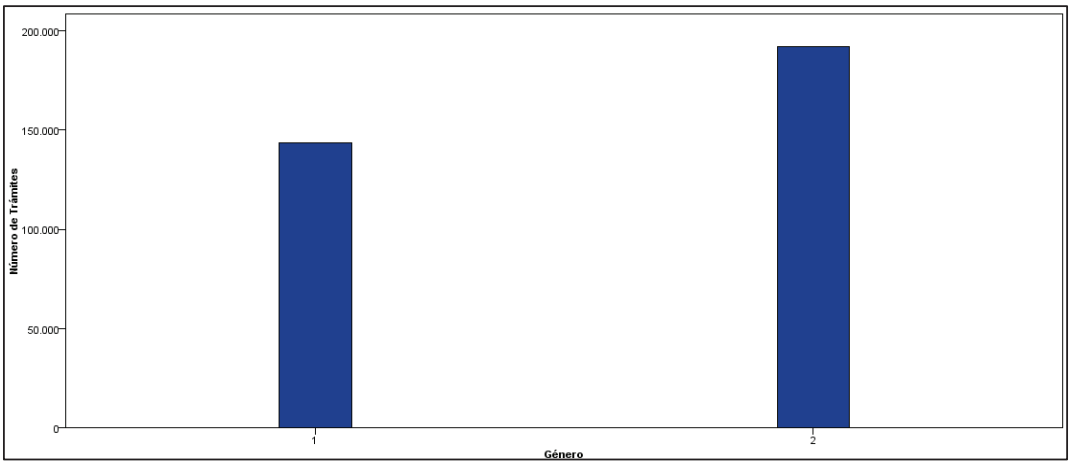
Fuente: SRI
Elaboración: Los autores

OBS	MEDIA	VARIANZA	DESVIANCIÓN ESTANDAR	KURTOSIS
335636	73	35,60	5,96	3,70

MODA	MÍNIMO	MÁXIMO
69	65	105

PERCENTILES							
1%	5%	10%	25%	50%	75%	90%	99%
65	66	67	68	72	77	82	91

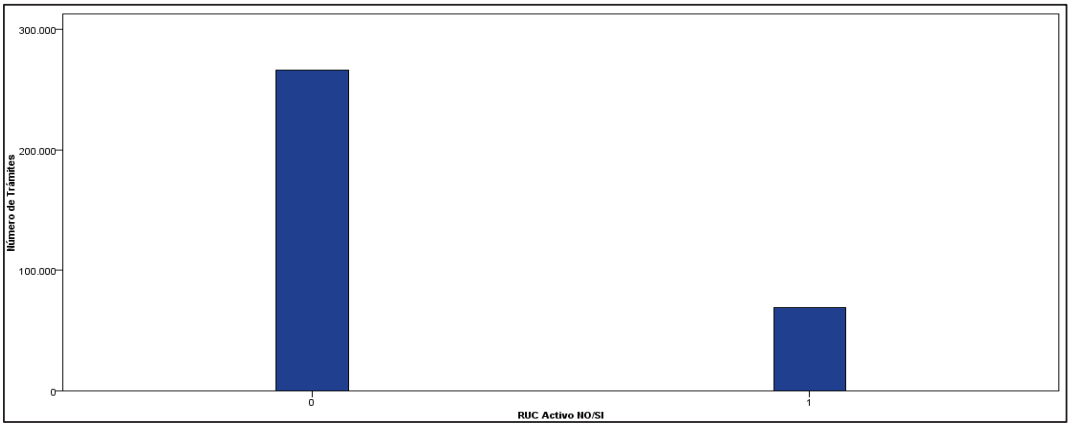
- GÉNERO



Fuente: SRI
Elaboración: Los autores

GÉNERO	NÚMERO DE TRÁMITES	% DEL TOTAL DE TRÁMITES
Hombre =1	143.456	42,74
Mujer=2	192.180	57,26

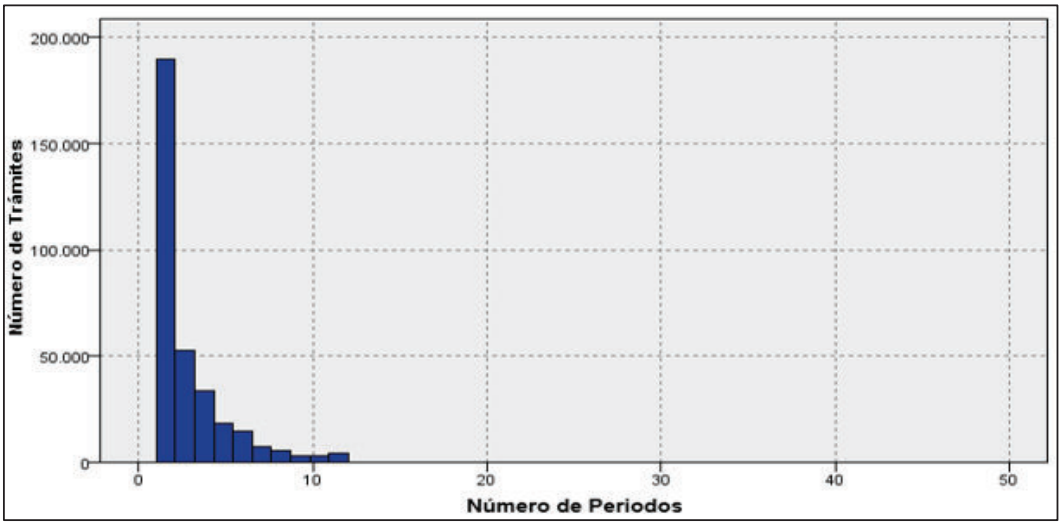
- RUC ACTIVO



Fuente: SRI
Elaboración: Los autores

RUC ACTIVO	NÚMERO DE TRÁMITES	% DEL TOTAL DE TRÁMITES
SI = 1	69.495	20,71
NO = 0	266.141	79,29

- NÚMERO DE PERIODOS



Fuente: SRI
Elaboración: Los autores

OBS	MEDIA	VARIANZA	DESVIANCIÓN ESTANDAR	KURTOSIS
335636	2,91	5,74	2,39	11,87

MODA	MÍNIMO	MÁXIMO
1	1	24

PERCENTILES							
1%	5%	10%	25%	50%	75%	90%	99%
1	1	1	1	2	4	6	12

ANEXO D

DEPURACIÓN BASE DE DATOS

Mediante el análisis exploratorio de cada una de las variables, la variable “Tipo Físico” presenta 4.332 casos de periodos duplicados, lo que indica que un mismo trámite fue registrado en el sistema del SRI bajo distintos tipos físico (hoja, carpeta y/o cd), manteniendo el resto de información igual. A través de expertos del SRI, se confirma que son registros duplicados que no necesariamente indican que la devolución el IVA se haya efectuado más de una vez; sin embargo su no eliminación mostraría un monto devuelto total para el año fiscal 2015 que no corresponde al que realmente se devolvió. Por tal motivo, no se consideran los

periodos que presentan una opción de la variable "Tipo Físico" diferente de hoja[36].

Un ejemplo de registro duplicado para el individuo con número de RUC "anon_S7386" es el siguiente:

Anexo D: Trámites duplicados de un mismo contribuyente

Número RUC	Número Trámite	Mes o periodo	Oficina	Monto Solicitado ($)	Monto Devuelto ($)	Tipo Físico
anon_S7386	anon_S7390	Junio	Azuay	2,44	2,44	carpeta
		Junio	Azuay	2,44	2,44	CD
		Junio	Azuay	2,44	2,44	hoja
		Abril	Azuay	73,92	18,41	CD
		Abril	Azuay	73,92	18,41	hoja
		Abril	Azuay	73,92	18,41	carpeta

Fuente: SRI
Elaboración Los autores

En la tabla del anexo D, se nota claramente que el individuo con número de RUC "anon_S7386", realiza un trámite (anon_7390) compuesto de dos periodos, correspondientes a los meses abril y junio; sin embargo, en dichos meses los registros se triplican, diferenciándose entre ellos, únicamente en la variable "tipo físico". Por ejemplo, el registro por el mes de junio se ingresó tres veces, manteniendo toda la información igual, y cambiando únicamente la forma en cómo se presentó la información en ventanilla (carpeta, cd y hoja), por lo que, en teoría se solicita tres veces por el mismo mes; sin embargo en realidad, solo fue devuelto un valor al beneficiario, por tal motivo, se decide trabajar con el tipo físico "hoja".

De igual manera, se elimina 35 periodos, al ser datos que presentan algún tipo de inconsistencia:

- 3 periodos con un monto solicitado[37] mayor a $15.000.
- 14 y 12 periodos correspondientes a individuos con una edad menor a 65 años y mayor a 114 años respectivamente.

[36] El tipo físico "hoja" está presente en el 99,5% de los trámites realizados para el año fiscal 2015.
[37] El monto máximo que un beneficiario puede solicitar por periodo o mes es de $212,40 o $265,50 considerando el IVA o ICE.

- 6 periodos en donde el monto solicitado y devuelto fue $0.

Todos estos casos encontrados y eliminados, representan menos del 0,01% del total de periodos considerados para el análisis.

ANEXO E

CONTRASTE DEL MODELO CON LA VARIABLE "MONTO SOLICITADO" VS "MONTO SOLICITADO (LN)"

Como se muestra en la tabla del anexo E, si se estima el modelo con la variable "Monto Solicitado" en su forma original, se obtiene una bondad de ajuste menor, y un porcentaje de clasificación menor (especialmente en el grupo de riesgo), por lo que para la estimación del modelo de riesgo, es preferible utilizar la variable "Monto Solicitado" transformada a su logaritmo natural.

Anexo E: Contraste del modelo con la variable "monto solicitado" vs "monto solicitado en logaritmo natural"

	MODELO	
	Monto Solicitado	**Monto Solicitado (LN)**
BONDAD DE AJUSTE		
-2LL	137.998,75	113.456,55
R cuadrado Cox y Snell	40,6 %	44,8 %
R cuadrado Nagelkerke	66,9 %	73,9 %
TEST DE HOSMER Y LEMESHOW		
Chi cuadrado	16.492,64	9.838,72
gl	8	8
Sig. (p-valor)	0	0
PODER DE CLASIFICACIÓN		
Global	91,2 %	92,2 %
Grupo de riesgo	63,1 %	71,5 %
Grupo de no riesgo	97,2 %	96,6 %

Fuente: SRI
Elaboración Los autores

ANEXO F

MÉTODO FORWARD

Tal como se revisó en el marco teórico, el método Forward incluye en cada paso nuevas variables dentro de la especificación del modelo dependiendo de la puntuación del score statistic, es decir, en cada paso ingresa al modelo la variable que tenga mayor correlación con la variable objetivo (puntaje del score de riesgo más alto), reduzca el valor del estadístico -2LL y, mejore los demás estadísticos de bondad de ajuste.

En este caso, el método forward realizó siete pasos, hasta llegar a la especificación del mejor modelo posible. Se comienza estimado un modelo únicamente con la inclusión de la constante, y dejando de lado a las variables explicativas (paso 0), la constante al ser estadísticamente significativa se incluye en el modelo. Para realizar el primer paso, al analizar la tabla del anexo F, se observa que la variable "Monto Solicitado (Ln)" presenta la mayor puntuación del score statistic (83719,178), y al mismo tiempo un valor de α (Sig) menor a 0,05, indicando que dicha variable presenta la mayor correlación con la variable dependiente; se realiza la estimación del modelo incluida dicha variable (tabla F.1), y se obtiene los estadísticos de bondad de ajuste de la tabla F.2, donde se observa que el modelo con la constante y la variable "Monto Solicitado" presenta un R cuadrado de Nagelkerke de 70,8%; además posee un poder de clasificación global de 91%, y para el grupo de riesgo de 66% (tabla F.3).

En el paso 2, se incluye en el modelo la variable "Número de Periodos", puesto que es la que mayor correlación presenta con la variable "RIESGO"; su inclusión disminuye el estadístico -2LL, en 7.913,818 (99633,668 - 91719,850), aumenta los R cuadro de Cox y Snell y Nagelkerke, y mejora la capacidad predictiva del modelo. De esta manera el algoritmo se sigue desarrollando para la inclusión de las demás variables.

Anexo F: Variables que no están en la ecuación

			PUNTUACIÓN	GL	SIG.
Paso 0	Variables	Número de Facturas	22462,211	1	0,000
		Tipo_Ingreso	37,604	2	,000
		Tipo_Ingreso(SNT)	19,783	1	,000
		Tipo_Ingreso(Ventanilla)	9,652	1	,002
		RUC Activo (si)	2919,359	1	0,000
		Género (mujer)	142,151	1	,000
		Edad	77,582	1	,000
		Número de Periodos	8274,280	1	0,000
		Monto_Solicitado_Ln	83719,178	1	0,000
		Correo Repetido (si)	1107,614	1	,000
	Estadísticos Globales		88139,248	9	0,000
Paso 1	Variables	Número de Facturas	116,276	1	,000
		Tipo_Ingreso	1180,389	2	,000
		Tipo_Ingreso(SNT)	20,887	1	,000
		Tipo_Ingreso(Ventanilla)	1024,607	1	,000
		RUC Activo (si)	1859,641	1	0,000
		Género (mujer)	96,849	1	,000
		Edad	78,164	1	,000
		Número de Periodos	7216,695	1	0,000
		Correo Repetido (si)	412,503	1	,000
	Estadísticos Globales		9925,784	8	0,000
Paso 2	Variables	Número de Facturas	181,802	1	,000
		Tipo_Ingreso	497,904	2	,000
		Tipo_Ingreso(SNT)	3,023	1	,082
		Tipo_Ingreso(Ventanilla)	451,555	1	,000
		RUC Activo (si)	2211,269	1	0,000
		Género (mujer)	89,617	1	,000
		Edad	73,508	1	,000
		Correo Repetido (si)	512,553	1	,000
	Estadísticos Globales		2936,452	7	0,000
Paso 3	Variables	Número de Facturas	185,319	1	,000
		Tipo_Ingreso	515,677	2	,000
		Tipo_Ingreso(SNT)	14,803	1	,000
		Tipo_Ingreso(Ventanilla)	435,791	1	,000
		Género (mujer)	11,579	1	,001
		Edad	11,288	1	,001
		Correo Repetido (si)	,861	1	,354
	Estadísticos Globales		741,684	6	,000
Paso 4	Variables	Número de Facturas	208,887	1	,000
		Género (mujer)	8,123	1	,004
		Edad	8,545	1	,003
		Correo Repetido (si)	2,304	1	,129

	Estadísticos Globales		227,486	4	,000
Paso 5	Variables	Género (mujer)	8,696	1	,003
		Edad	10,461	1	,001
		Correo Repetido (si)	2,245	1	,134
	Estadísticos Globales		18,746	3	,000
Paso 6	Variables	Género (mujer)	7,806	1	,005
		Correo Repetido (si)	1,231	1	,267
	Estadísticos Globales		8,283	2	,016
Paso 7	Variables	Correo Repetido (si)	,477	1	,490
	Estadísticos Globales		,477	1	,490

Fuente: SRI
Elaboración Los autores

Anexo F.1: Variables en la ecuación

		B	Error estándar	Wald	gl	Sig.	Exp(B)
Paso 1[a]	Monto_Solicitado_Ln	5,802	,030	36602,430	1	0,000	331,061
	Constante	-32,055	,164	38155,409	1	0,000	,000
Paso 2[b]	Número de Periodos	-,321	,004	6645,208	1	0,000	,726
	Monto_Solicitado_Ln	7,014	,039	33160,785	1	0,000	1111,660
	Constante	-37,594	,203	34389,295	1	0,000	,000
Paso 3[c]	RUC Activo(si)	,928	,020	2151,778	1	0,000	2,530
	Número de Periodos	-,333	,004	6879,605	1	0,000	,716
	Monto_Solicitado_Ln	7,138	,039	32676,818	1	0,000	1258,599
	Constante	-38,443	,209	33968,120	1	0,000	,000
Paso 4[d]	Tipo_Ingreso			512,561	2	,000	
	Tipo Ingreso(SNT)	-,595	,068	76,077	1	,000	,551
	Tipo Ingreso(Vent)	-,442	,020	497,676	1	,000	,643
	RUC Activo(si)	,936	,020	2166,766	1	0,000	2,550
	Número de Periodos	-,323	,004	6385,808	1	0,000	,724
	Monto_Solicitado_Ln	7,174	,040	32627,301	1	0,000	1305,597
	Constante	-38,341	,209	33573,096	1	0,000	,000
Paso 5[e]	Número de Facturas	,003	,000	206,935	1	,000	1,003
	Tipo_Ingreso			536,540	2	,000	
	Tipo Ingreso(SNT)	-,577	,068	71,421	1	,000	,562
	Tipo Ingreso(Vent)	-,455	,020	524,503	1	,000	,635
	RUC Activo(si)	,937	,020	2169,462	1	0,000	2,553
	Número de Periodos	-,338	,004	6439,898	1	0,000	,713
	Monto_Solicitado_Ln	7,131	,040	32071,183	1	0,000	1250,380
	Constante	-38,203	,210	33234,271	1	0,000	,000
Paso 6[f]	Número de Facturas	,003	,000	208,816	1	,000	1,003
	Tipo_Ingreso			533,536	2	,000	
	Tipo Ingreso(SNT)	-,578	,068	71,718	1	,000	,561

	Tipo Ingreso(Vent)	-,453	,020	521,240	1	,000	,635
	RUC Activo(si)	,930	,020	2108,250	1	0,000	2,534
	Edad	-,005	,001	10,460	1	,001	,995
	Número de Periodos	-,338	,004	6439,953	1	0,000	,713
	Monto_Solicitado_Ln	7,133	,040	32067,334	1	0,000	1252,362
	Constante	-37,874	,232	26553,471	1	0,000	,000
Paso 7[g]	Número de Facturas	,003	,000	209,264	1	,000	1,003
	Tipo_Ingreso			530,193	2	,000	
	Tipo Ingreso(SNT)	-,579	,068	71,857	1	,000	,561
	Tipo Ingreso(Vent)	-,452	,020	517,708	1	,000	,636
	RUC Activo(si)	,923	,020	2043,780	1	0,000	2,516
	Género (mujer)	-,048	,017	7,805	1	,005	,953
	Edad	-,004	,001	9,573	1	,002	,996
	Número de Periodos	-,338	,004	6437,633	1	0,000	,713
	Monto_Solicitado_Ln	7,133	,040	32063,992	1	0,000	1252,722
	Constante	-37,862	,232	26528,289	1	0,000	,000
a. Variables especificadas en el paso 1: Monto Solicitado (Ln)							
b. Variables especificadas en el paso 2: Número de Periodos							
c. Variables especificadas en el paso 3: RUC Activo							
d. Variables especificadas en el paso 4: Tipo ingreso.							
e. Variables especificadas en el paso 5: Número de Facturas							
f. Variables especificadas en el paso 6: Edad							
g. Variables especificadas en el paso 6: Género							

Fuente: SRI
Elaboración Los autores

Anexo F.2: Estadísticos de Bondad de Ajuste.

Paso	Logaritmo de la verosimilitud -2	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	99633,668[a]	42,9 %	70,8 %
2	91719,850[b]	44,5 %	73,5 %
3	89554,793[b]	45,0 %	74,2 %
4	89044,882[b]	45,1 %	74,4 %
5	88835,740[b]	45,1 %	74,5 %
6	88825,261[b]	45,1 %	74,5 %
7	88817,460[b]	45,1 %	74,5 %
a. La estimación ha terminado en el número de iteración 8 porque las estimaciones de parámetro han cambiado en menos de ,001.			
b. La estimación ha terminado en el número de iteración 9 porque las estimaciones de parámetro han cambiado en menos de ,001.			

Fuente: SRI
Elaboración Los autores

Anexo F.3: Tabla de Clasificación

OBSERVADO			PRONOSTICADO		
			RIESGO		CORRECCIÓN DE PORCENTAJE
			0	1	
Paso 1	RIESGO	0	213088	8029	96,4
		1	16075	31196	66,0
	Porcentaje global				91,0
Paso 2	RIESGO	0	213486	7631	96,5
		1	14609	32662	69,1
	Porcentaje global				91,7
Paso 3	RIESGO	0	213780	7337	96,7
		1	13181	34090	72,1
	Porcentaje global				92,4
Paso 4	RIESGO	0	213736	7381	96,7
		1	13059	34212	72,4
	Porcentaje global				92,4
Paso 5	RIESGO	0	213757	7360	96,7
		1	13042	34229	72,4
	Porcentaje global				92,4
Paso 6	RIESGO	0	213769	7348	96,7
		1	13049	34222	72,4
	Porcentaje global				92,4
Paso 7	RIESGO	0	213766	7351	96,7
		1	13051	34220	72,4
	Porcentaje global				92,4
El valor de corte es 0,5.					

Fuente: SRI
Elaboración Los autores