

# **ESCUELA POLITÉCNICA NACIONAL**

## **FACULTAD DE INGENIERÍA DE SISTEMAS**

**Estudio comparativo entre analítica de datos y análisis de datos  
mediante un caso de estudio.**

**TRABAJO DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL TÍTULO DE  
INGENIERO EN SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN**

**Sandy Dayana Simbaña Simbaña**  
**sandy.simbana@epn.edu.ec**


**DIRECTOR: MSC. Boris Astudillo.**  
**boris.astudillo@epn.edu.ec**

**CODIRECTOR: MARCO OSWALDO SANTÓRUM GAIBOR, PHD.**  
**marco.santorum@epn.edu.ec**

**Quito, enero de 20201**

## **AVAL**

Certificamos que el presente trabajo fue desarrollado por la estudiante Sandy Dayana Simbaña Simbaña, bajo nuestra supervisión.



---

**MSC. BORIS ASTUDILLO.  
DIRECTOR DEL TRABAJO DE  
TITULACIÓN**

---

**MARCO OSWALDO SANTÓRUM  
GAIBOR, PHD.  
CODIRECTOR DEL TRABAJO DE  
TITULACIÓN**

## DECLARACIÓN DE AUTORÍA

Yo Sandy Dayana Simbaña Simbaña, declaro bajo juramento que el trabajo aquí descrito es de mi autoría; que no ha sido previamente presentada para ningún grado o calificación profesional; y, que he consultado las referencias bibliográficas que se incluyen en este documento.

A través de la presente declaración cedo mis derechos de propiedad intelectual correspondientes a este trabajo, a la Escuela Politécnica Nacional, según lo establecido por la Ley de Propiedad Intelectual, por su Reglamento y por la normatividad institucional vigente.



---

SANDY DAYANA SIMBAÑA SIMBAÑA

## **AGRADECIMIENTO**

A mi Dios por haberme permitido la consecución de este mérito, por ser mi fortaleza, guía y compañía en todo momento.

A mis padres Hugo y Cecilia por todo el apoyo incondicional, amor y paciencia que siempre me han brindado.

A mi abuelito Gaspar por ser mi motivación de superación, ternura y cariño.

A mis hermanos Freddy y Soraya por ser mi inspiración y ejemplo a seguir.

A mis amigos por cada uno de sus consejos, risas y momentos vividos durante toda mi carrera universitaria.

A mi Director del proyecto de titulación Boris Astudillo, por ser mi gran apoyo con su conocimiento, disponibilidad y paciencia en esta última etapa de mi carrera.

## INDICE DE CONTENIDOS

AVAL .....	I
DECLARACIÓN DE AUTORÍA .....	II
AGRADECIMIENTO.....	III
INDICE DE CONTENIDOS .....	IV
INDICE DE FIGURAS .....	VI
INDICE DE TABLAS .....	VII
RESUMEN .....	VIII
ABSTRACT .....	IX
CAPITULO 1 .....	1
1. INTRODUCCIÓN.....	1
1.1. Objetivo general .....	2
1.2. Objetivos específicos.....	3
1.3. Hipótesis.....	3
1.4. Conclusiones del capítulo.....	3
CAPITULO 2 .....	4
2. METODOLOGÍA.....	4
2.1. Búsqueda de la literatura.....	5
2.1.1. Identificación de los parámetros de búsqueda.....	5
2.1.2. Selección de estudios .....	10
2.1.3. Filtración de estudios .....	12
2.1.4. Extracción de información relevante .....	14
2.2. Comparación de criterios.....	16
2.2.1. Definición de criterios.....	16
2.2.2. Resultados iniciales de la extracción .....	17
2.2.3. Resultados finales de la extracción.....	33
2.3. Conclusiones del capítulo.....	36
CAPITULO 3 .....	37
3. PROYECCIÓN DEL CASO DE ESTUDIO Y RESULTADOS .....	37
3.1. Proyección del caso de estudio.....	37
3.1.1. Contexto del caso de estudio .....	37
3.1.2. Definición de Herramientas.....	37

3.1.3. Método de la Proyección del caso de estudio .....	38
3.2. Resultados y discusión del caso de estudio .....	44
3.2.1. Resultados .....	44
3.2.2. Discusión.....	48
3.3. Conclusiones del Capitulo .....	49
CAPITULO 4 .....	50
4. CONCLUSIONES Y RECOMENDACIONES.....	50
4.1. CONCLUSIONES.....	50
4.2. RECOMENDACIONES .....	51
5. REFERENCIAS BIBLIOGRÁFICAS .....	52

## INDICE DE FIGURAS

<b>Figura 1:</b> Metodología de Investigación propuesta por los Autores .....	5
<b>Figura 2:</b> Organización de Palabras Clave .....	6
<b>Figura 3:</b> Organización Final de Palabras Clave .....	7
<b>Figura 4:</b> Proceso de Extracción de Información Relevante .....	15
<b>Figura 5:</b> Proceso Final de Extracción de Información Relevante .....	16
<b>Figura 6:</b> Ventana para la Creación de la Aplicación .....	39
<b>Figura 7:</b> Ventana de Asignación de Permisos .....	39
<b>Figura 8:</b> Captura de Claves y Tokens .....	40
<b>Figura 9:</b> Captura de la Autenticación con el API de Twitter .....	40
<b>Figura 10:</b> Captura de Asignación de Tokens para Acceso a la API de Twitter ..	40
<b>Figura 11:</b> Captura de la Lista de Tweets .....	41
<b>Figura 12:</b> Captura del Conteo del Tweets .....	41
<b>Figura 13:</b> Captura de la Función del Cálculo de Polaridad .....	42
<b>Figura 14:</b> Captura de la Función que Retorna Resultados de Polaridad .....	42
<b>Figura 15:</b> Captura de la Función para Graficar Resultados .....	43
<b>Figura 16:</b> Captura de la Función de Cálculo de Porcentajes .....	43
<b>Figura 17:</b> Nivel de Aceptación del Candidato Gerson Almeida .....	44
<b>Figura 18:</b> Nivel de Aceptación del Candidato Guillermo Celi .....	44
<b>Figura 19:</b> Nivel de Aceptación del Candidato Lucio Gutiérrez .....	45
<b>Figura 20:</b> Nivel de Aceptación del Candidato Cesar Montufar .....	45
<b>Figura 21:</b> Nivel de Aceptación del Candidato Paul Carrasco .....	45
<b>Figura 22:</b> Nivel de Aceptación del Candidato Isidro Romero .....	46
<b>Figura 23:</b> Nivel de Aceptación del Candidato Guillermo Lasso .....	46
<b>Figura 24:</b> Nivel de Aceptación del Candidato Andrés Arauz .....	46
<b>Figura 25:</b> Nivel de Apoyo de Usuarios Twitter a Candidatos Electorales .....	47

## INDICE DE TABLAS

<b>Tabla 1:</b> Cuadro para la Estrategia de Búsqueda V1.....	8
<b>Tabla 2:</b> Cuadro para la Estrategia de Búsqueda V2.....	8
<b>Tabla 3:</b> Cadenas de Búsqueda .....	9
<b>Tabla 4:</b> Matriz de Búsqueda V1 .....	10
<b>Tabla 5:</b> Matriz de Búsqueda V2 .....	11
<b>Tabla 6:</b> Criterios de Inclusión y Exclusión .....	12
<b>Tabla 7:</b> Evaluación de Calidad .....	13
<b>Tabla 8:</b> Documentación Final Recolectada.....	14
<b>Tabla 9:</b> Artículos Académicos Extras .....	15
<b>Tabla 10:</b> Tabla Comparativa entre Análisis y Analítica de Datos .....	18
<b>Tabla 11:</b> Tabla Comparativa Final entre Análisis y Analítica de Datos .....	34
<b>Tabla 12:</b> Herramientas para el Caso de Estudio.....	38



## RESUMEN

El presente trabajo muestra un estudio comparativo a partir de las diferencias entre análisis y analítica de datos. Para ello, se propone realizar una búsqueda y revisión de la información en las diferentes fuentes de datos que servirán como base teórica para efectuar el estudio comparativo. A continuación, se realiza la comparación de ambos términos mediante la creación de una tabla comparativa completa y eficaz que permita distinguir claramente las diferencias entre ambos términos, con el objetivo de utilizar este nuevo conocimiento en la ejemplificación de un caso de estudio como es la predicción del candidato ganador en las elecciones 2021 mediante el análisis de sentimientos de las personas expresadas en tweets.

Este trabajo se encuentra organizado en 4 capítulos: en el primer capítulo se introduce a la investigación, se define algunos conceptos importantes para la misma, se plantea el objetivo general, los objetivos específicos y la hipótesis.

En el segundo y tercer capítulo respectivamente se detalla la metodología propia que realizamos, la cual incluye 4 etapas a seguir comenzando con la búsqueda de la literatura, seguida de la comparación de criterios, la proyección del caso de estudio y la presentación de los resultados del caso de estudio.

Finalmente, en el cuarto capítulo se presentan las conclusiones y recomendaciones obtenidas en base a los objetivos planteados para la investigación.

**PALABRAS CLAVE:** diferencias, analítica de datos, análisis de datos, tabla comparativa, caso de estudio.

## **ABSTRACT**

The present work shows a comparative study based on the differences between analysis and data analytics. To do this, it is proposed to carry out a search and review of the information in the different data sources that will serve as a theoretical basis to carry out the comparative study. Next, the comparison of both terms is made by creating a complete and effective comparative table that allows to clearly distinguish the differences between both terms, with the aim of using this new knowledge in the exemplification of a case study such as prediction of the winning candidate in the 2021 elections by analyzing people's sentiments expressed in tweets.

This work is organized into 4 chapters: the first chapter introduces the research, defines some important concepts for it, sets out the general objective, the specific objectives, and the hypothesis.

In the second and third chapters, respectively, the own methodology that we carry out is detailed, which includes 4 stages to follow, starting with the search of the literature, followed by the comparison of criteria, the projection of the case study and the presentation of the results of the study. case study.

Finally, the fourth chapter presents the conclusions and recommendations obtained based on the objectives set for the research.

**KEY WORDS:** differences, Data Analytics, Data Analysis, comparative table, case study.

# CAPITULO 1

## 1. INTRODUCCIÓN

Con los recursos informáticos, la información y el conocimiento en rápido crecimiento, ha surgido un nuevo mundo digital. La información se crea, se almacena a un ritmo rápido y es accesible mediante una amplia gama de aplicaciones a través de la informática científica, cargas de trabajo comerciales y redes sociales. En 2020, más de 50 mil millones de dispositivos inteligentes estarán conectados en todo el mundo y el flujo de tráfico de Internet será 92 veces mayor que en 2005 [1].

En todo el mundo, las organizaciones de todas las formas y tamaños luchan entre sí para tener en sus manos datos de alta calidad con el fin de extraer información cuantificada que pueda conducir a mejores resultados comerciales. Sin embargo, los datos, incluso los macro datos, no son valiosos en sí mismos. Más bien, es la información procesable que se puede extraer de ella donde reside el valor, y aquí es donde entran en juego el análisis de datos y la analítica de datos [2].

Mientras que el análisis de datos se refiere al proceso de examinar, transformar y organizar un conjunto de datos dado de formas específicas para estudiar sus partes individuales y extraer información útil [3]; la analítica de datos, por otro lado, se refiere a la utilización de varias herramientas y habilidades que involucran métodos cualitativos y cuantitativos, que emplean estos datos recopilados y producen un resultado que se utiliza para mejorar la eficiencia, la productividad, reducir el riesgo y aumentar las ganancias comerciales [4]. Por lo tanto, la analítica de datos es un subconjunto del análisis de datos que implica el uso de diversas herramientas y técnicas que ayudarán en el proceso de análisis de datos para tomar una decisión comercial [5].

Tanto el análisis de datos como la analítica de datos se utilizan para descubrir patrones, tendencias y anomalías que se encuentran dentro de los datos y, por lo tanto, brindar la información que las empresas necesitan para permitir la toma de decisiones basada en evidencia, aunque desde una perspectiva diferente [2].

A pesar de que los dos términos están interconectados, proporcionan resultados y enfoques diferentes, utilizan métodos, técnicas y herramientas que en la actualidad no se distinguen, ya sea por falta de conocimiento o información, incluso la complejidad en el uso de estos términos ha llevado al uso de un tercero, como es minería de datos.

El incorrecto enfoque usado para implementar estas áreas genera varias desventajas significativas ya que, al no disponer de los datos correctos, los riesgos comerciales aumentan a medida que los directivos toman malas decisiones, proporcionando inadecuados servicios a los clientes y socios, y desperdiciando las oportunidades de ventas debido al equivocado conocimiento generado por información errónea [6].

Por esta razón es necesario un estudio para comprender las diferencias entre análisis de datos y analítica de datos, analizando sus procesos, métodos, técnicas, beneficios y limitaciones, permitiendo una eficaz aclaración de los conceptos para aplicar en futuras investigaciones y su aprovechamiento de nuevos conocimientos en el caso de estudio planteado.

Para ello es necesario definir los conceptos más importantes dentro de la investigación, para que al momento de profundizar con el estudio mediante el análisis comparativo se pueda identificar y clasificar de manera correcta términos relacionados con análisis y analítica de datos.

### **Análisis de Datos**

Se describe al proceso de análisis de datos como una acción compleja de ir y venir entre datos y conceptos, entre descripción e interpretación, utilizando tanto el razonamiento inductivo como el deductivo. Describe el análisis como basado en "conocimientos acordados, el reconocimiento de propiedades o estándares mutuamente reconocidos" que son "inherentemente conservadores, cuidadosos, sistemáticos" [7].

### **Analítica de Datos**

El término Data Analytics se hizo popular a principios de la década de 2000. Data Analytics se define como la aplicación de sistemas informáticos al análisis de grandes conjuntos de datos para respaldar decisiones. Data Analytics es un campo muy interdisciplinario que ha adoptado aspectos de muchas otras disciplinas científicas como la estadística, el aprendizaje automático, el reconocimiento de patrones, la teoría de sistemas, la investigación de operaciones, o inteligencia artificial [8].

## **1.1. Objetivo general**

Implementar un caso de estudio que permita la aplicación práctica de los conceptos de analítica y análisis de datos y que permita evidenciar sus diferencias.

## **1.2. Objetivos específicos**

- Elaborar un marco conceptual de los dominios en estudio.
- Analizar y realizar un estudio comparativo con las características claves y diferencias entre análisis y analítica de datos.
- Ejemplificar los resultados de su utilidad.

## **1.3. Hipótesis**

Mediante un caso de estudio es posible identificar elementos claves diferenciadores entre los conceptos de análisis y analítica de datos, optimizando los procesos necesarios que permiten al investigador obtener información para convertirla en conocimiento, poder tomar decisiones con menor riesgo y poder anticiparse a los cambios.

## **1.4. Conclusiones del capítulo**

En el capítulo 1 “INTRODUCCIÓN”, se establece la problemática, objetivos e hipótesis que promueven el presente trabajo. A continuación, se procederá a detallar el desarrollo de la metodología propuesta para la búsqueda de literatura y comparación de criterios.

## **CAPITULO 2**

### **2. METODOLOGÍA**

Para este trabajo de investigación se propone realizar una metodología propia, basada en la metodología de Kitchenham, y que consta de 4 etapas principales como se muestra en la Figura 1. Se decide utilizar la metodología de Kitchenham para la Etapa 1 con el fin de lograr un proceso de búsqueda de la literatura que sea de calidad y ordenado (sistemático) para obtener la mejor información en el proceso de extracción de conocimiento, a partir de esto observamos la necesidad de fusionar nuevas etapas que permitan enfocar el objetivo principal de este proyecto de investigación, la cual es la creación de una tabla comparativa que refleje las diferencias fundamentales entre Data Analysis y Data Analytics, y la proyección del caso de estudio donde se ilustre claramente los beneficios de utilizar ambos conceptos correctamente.

Para la etapa 2, si bien es cierto que la metodología de Kitchenham también presenta resultados de su revisión sistemática, en esta etapa decidimos incorporar un valor agregado a la metodología con la presentación de resultados en forma de una tabla comparativa en la cual es evidente que nuevos pasos y componentes fueron agregados, por esta razón al hablar de un reestructuración de pasos no se podría continuar hablando de una metodología de Kitchenham como tal, más bien se detalla una nueva metodología que parte de esta pero que nos permite optimizar procedimientos.

Finalmente en las etapas 3 y 4 evidenciamos claramente la mejora o evolución de la metodología, añadiendo pasos para el desarrollo de una aplicación y mostrando sus resultados fruto de la búsqueda de la literatura y el estudio comparativo anteriormente realizado.



**Figura 1:** Metodología de Investigación propuesta por los Autores

## 2.1. Búsqueda de la literatura

En esta etapa se lleva a cabo la búsqueda y revisión de la información en las diferentes fuentes de datos que servirán como base teórica para efectuar el estudio comparativo.

### 2.1.1. Identificación de los parámetros de búsqueda

Esta etapa tiene como propósito específico definir los parámetros más importantes que serán tomados en cuenta cuando se lleve a cabo la búsqueda.

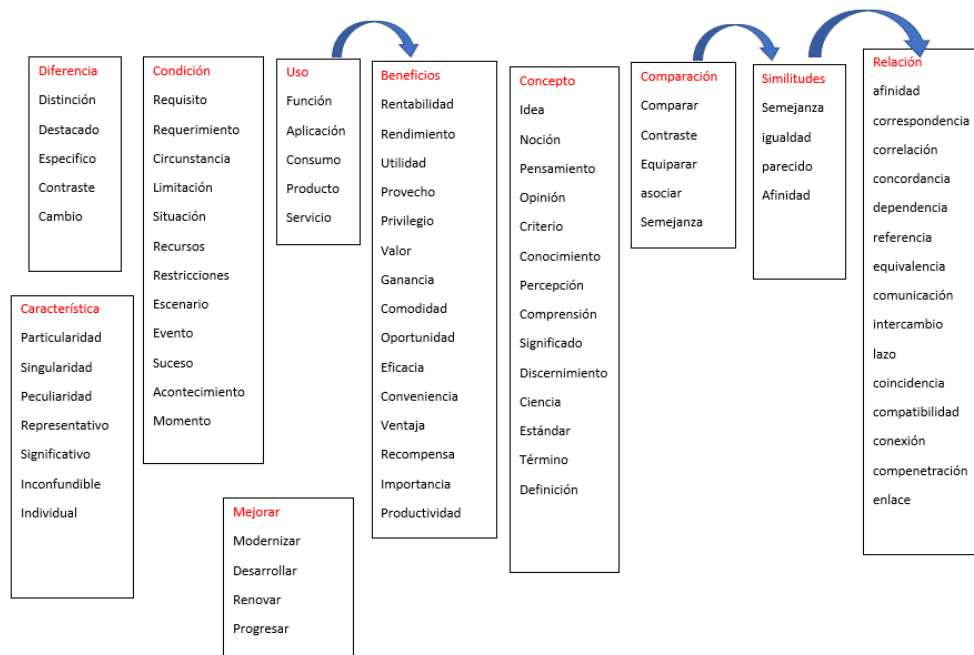
Para esto se detalla las siguientes subetapas:

1. Identificación de la necesidad de la búsqueda. Antes de emprender una búsqueda de la literatura, se debe asegurar de que ésta es necesaria, por ello se investigó de manera general el tema de análisis y analítica de datos, se realizó una breve discusión del problema con el experto (tutor) y por último se formuló las siguientes preguntas de investigación:

- ¿Cuáles son las diferencias con respecto a técnicas y metodologías de los ámbitos de estudio?
- ¿Cuál es la relación entre análisis de datos y analítica de datos?
- ¿En qué condiciones es necesario aplicar analítica de datos o análisis de datos?

2. Definición de un protocolo de búsqueda. En este proceso se definió las normas que seguirá la investigación respecto del proceso de búsqueda.

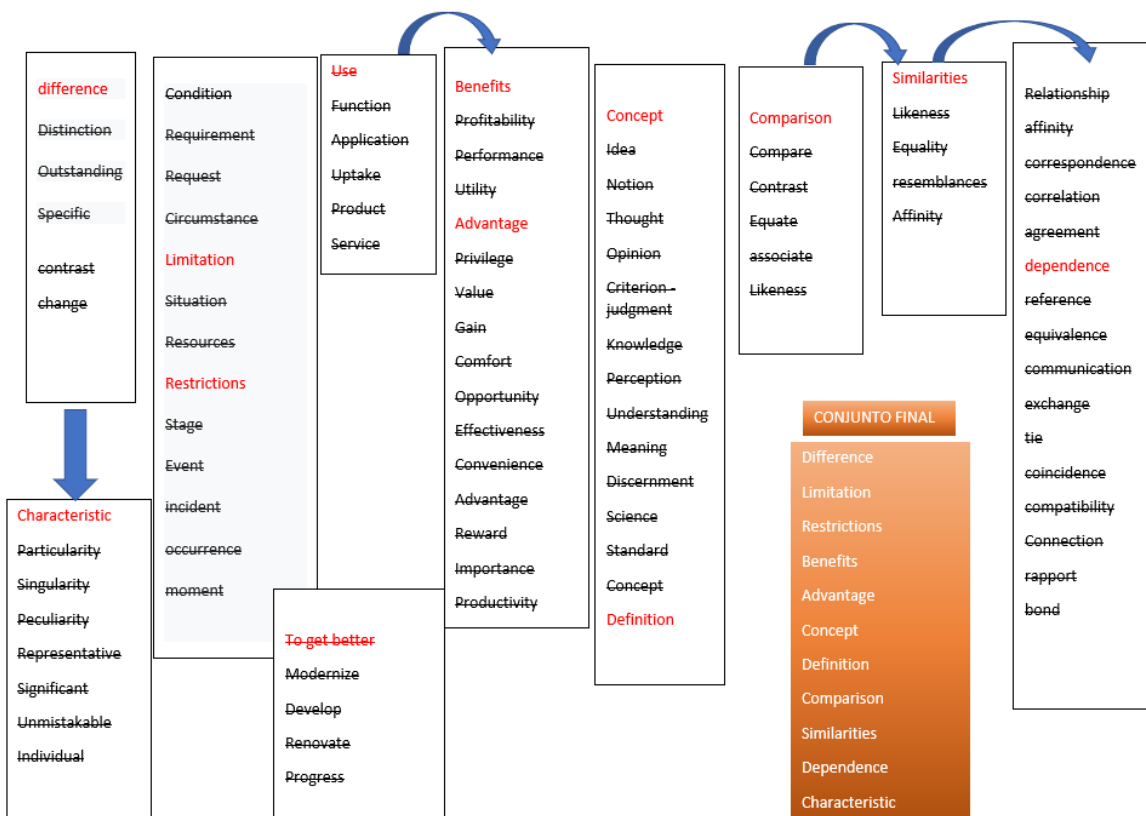
Primero definimos palabras clave que se buscarán en las bases de datos, para ello nos basamos en las preguntas de investigación antes mencionadas, con lo cual obtuvimos un conjunto de palabras a las cuales se les buscaron sinónimos y se las organizó de la siguiente manera:



**Figura 2:** Organización de Palabras Clave

Para mejorar el proceso de búsqueda, traducimos todas estas palabras al idioma inglés y con ayuda del experto (tutor) obtuvimos el conjunto final de palabras clave.





**Figura 3:** Organización Final de Palabras Clave

Luego de obtener el conjunto final de palabras clave procedemos a organizarlas en un cuadro para obtener la estrategia de búsqueda, Tabla 1: donde la columna A y C contiene las palabras clave relacionadas con el objetivo principal del estudio (Data Analytics y Data Analysis), la columna B contiene los términos relacionados a la investigación (methodology, method, process) y la columna D son el conjunto final de palabras clave que obtuvimos anteriormente en la Figura 3.

**Tabla 1:** Cuadro para la Estrategia de Búsqueda V1

CUADRO PARA LA ESTRATEGIA DE BÚSQUEDA V1							
A		B		C		D	
A1	Data Analytics	B1	Methodology	C1	Data Analysis	D1	Diferencia
		B2	Method			D2	Condición
		B3	Process			D3	Uso
						D4	Beneficios
						D5	Concepto
						D6	Similitudes
						D7	Comparación
						D8	Relación
						D9	Característica
						D10	Modernizar
						D11	Requisito
						D12	Requerimiento
						D13	Afinidad
						D14	Correspondencia
						D15	Correlación
						D16	Modernizar
						D17	Desarrollar
						D18	Renovar
						D19	Particularidad
						D20	Singularidad
						D21	Producto
						D22	Servicio

**Tabla 2:** Cuadro para la Estrategia de Búsqueda V2

CUADRO PARA LA ESTRATEGIA DE BÚSQUEDA v2							
A		B		C		D	
A1	Data analytics	B1	Methodology	C1	Data Analysis	D1	Difference
		B2	Method			D2	Limitation
		B3	Process			D3	Restriction
						D4	Benefits
						D5	Advantage
						D6	Concept
						D7	Definition
						D8	Comparison
						D9	Similarities
						D10	Dependence
						D11	Characteristic

Mediante las combinaciones de estas palabras claves y dependiendo de la base de datos a utilizar se definió conectores y signos para realizar la búsqueda, a continuación, se presenta las cadenas de búsqueda que utilizamos para esta investigación:

**Tabla 3:** Cadenas de Búsqueda

N	A1		B1		C1		D1 OR D2 OR ..D11
1	Data Analytics	AND	Methodology	AND	Data Analysis	AND	Difference
2	Data Analytics	AND	Methodology	AND	Data Analysis	AND	Limitation
3	Data Analytics	AND	Methodology	AND	Data Analysis	AND	Restriction
4	Data Analytics	AND	Methodology	AND	Data Analysis	AND	Benefits
5	Data Analytics	AND	Methodology	AND	Data Analysis	AND	Advantage
6	Data Analytics	AND	Methodology	AND	Data Analysis	AND	Concept
7	Data Analytics	AND	Methodology	AND	Data Analysis	AND	Definition
8	Data Analytics	AND	Methodology	AND	Data Analysis	AND	Comparison
9	Data Analytics	AND	Methodology	AND	Data Analysis	AND	Similarities
10	Data Analytics	AND	Methodology	AND	Data Analysis	AND	Dependence
11	Data Analytics	AND	Methodology	AND	Data Analysis	AND	Characteristic
	A1		B2		C1		D1 OR D2 OR D3 OR.... D11
12	Data Analytics	AND	Method	AND	Data Analysis	AND	Difference
13	Data Analytics	AND	Method	AND	Data Analysis	AND	Limitation
14	Data Analytics	AND	Method	AND	Data Analysis	AND	Restriction
15	Data Analytics	AND	Method	AND	Data Analysis	AND	Benefits
16	Data Analytics	AND	Method	AND	Data Analysis	AND	Advantage
17	Data Analytics	AND	Method	AND	Data Analysis	AND	Concept
18	Data Analytics	AND	Method	AND	Data Analysis	AND	Definition
19	Data Analytics	AND	Method	AND	Data Analysis	AND	Comparison
20	Data Analytics	AND	Method	AND	Data Analysis	AND	Similarities
21	Data Analytics	AND	Method	AND	Data Analysis	AND	Dependence
22	Data Analytics	AND	Method	AND	Data Analysis	AND	Characteristic
	A1		B3	AND	C1		D1 OR D2 D3 OR.... D11
23	Data Analytics	AND	Process	AND	Data Analysis	AND	Difference
24	Data Analytics	AND	Process	AND	Data Analysis	AND	Limitation
25	Data Analytics	AND	Process	AND	Data Analysis	AND	Restriction
26	Data Analytics	AND	Process	AND	Data Analysis	AND	Benefits
27	Data Analytics	AND	Process	AND	Data Analysis	AND	Advantage
28	Data Analytics	AND	Process	AND	Data Analysis	AND	Concept
29	Data Analytics	AND	Process	AND	Data Analysis	AND	Definition
30	Data Analytics	AND	Process	AND	Data Analysis	AND	Comparison
31	Data Analytics	AND	Process	AND	Data Analysis	AND	Similarities
32	Data Analytics	AND	Process	AND	Data Analysis	AND	Dependence
33	Data Analytics	AND	Process	AND	Data Analysis	AND	Characteristic

## 2.1.2. Selección de estudios

Una vez que se definen las cadenas de búsqueda procedemos a introducirlas dentro de las siguientes bases de datos: IEEE (I), Scopus (S), Taylor and Francis (T) y ACM (A) dando como resultado la siguiente matriz de búsqueda, donde se detalla el número total de 48661 papers encontrados:

**Tabla 4:** Matriz de Búsqueda V1

N	IEEE	Scopus	Taylor & Francis	ACM	TOTAL
1	3	3	681	647	1334
2	11	16	628	601	1256
3	3	1	244	175	423
4	24	19	536	391	970
5	13	9	452	613	1087
6	24	15	552	542	1133
7	4	5	605	551	1165
8	14	5	594	555	1168
9	13	7	562	517	1099
10	1	0	355	227	583
11	23	18	578	415	1034
12	28	17	954	1113	2112
13	46	22	863	1015	1946
14	6	3	319	303	631
15	62	43	723	658	1486
16	82	37	623	1065	1807
17	74	61	750	916	1801
18	8	10	844	938	1800
19	43	30	824	940	1837
20	44	14	778	885	1721
21	6	5	478	360	849
22	101	57	779	684	1621
23	43	13	1010	1283	2349
24	92	21	908	1174	2195
25	5	3	342	301	651
26	184	48	776	776	1784
27	138	30	659	1242	2069
28	179	55	805	1052	2091
29	24	15	895	1088	2022
30	86	13	867	1065	2031
31	59	9	811	1016	1895
32	3	1	496	398	898
33	163	51	821	778	1813
<b>TOTAL</b>	<b>1609</b>	<b>656</b>	<b>22112</b>	<b>24284</b>	<b>48661</b>

Del total de 48661 artículos científicos encontrados procedemos con la primera exclusión donde descartamos títulos de los papers que se encuentren duplicados o irrelevantes es decir que no posean alguna palabra clave dentro del mismo; quedando un total de 3670 artículos, como lo muestra la Tabla 5.

**Tabla 5:** Matriz de Búsqueda V2

<b>N</b>	<b>IEEE</b>	<b>Scopus</b>	<b>Taylor &amp; Francis</b>	<b>ACM</b>	<b>TOTAL</b>
1	3	2	671	616	1292
2	11	12	4	27	54
3	3	1	0	2	6
4	23	12	2	3	40
5	11	4	0	4	19
6	18	10	2	1	31
7	3	1	1	0	5
8	10	1	0	0	11
9	10	3	0	0	13
10	1	0	0	0	1
11	15	10	0	0	25
12	28	13	306	496	843
13	39	15	5	43	102
14	3	3	1	2	9
15	52	24	2	2	80
16	61	13	1	5	80
17	57	32	1	0	90
18	5	6	0	0	11
19	24	16	0	1	41
20	26	5	0	0	31
21	2	0	0	0	2
22	59	17	0	0	76
23	29	3	43	130	205
24	54	4	0	6	64
25	2	0	0	5	7
26	124	22	3	0	149
27	69	12	0	0	81
28	107	15	1	0	123
29	14	9	0	0	23
30	47	6	0	0	53
31	29	3	0	0	32
32	1	0	0	0	1
33	57	13	0	0	70
<b>TOTAL</b>	<b>997</b>	<b>287</b>	<b>1043</b>	<b>1343</b>	<b>3670</b>

De los 3670 artículos académicos procedemos con la ejecución de los siguientes criterios de inclusión y exclusión que detallamos a continuación:

**Tabla 6:** Criterios de Inclusión y Exclusión

<b>Criterios de inclusión</b>	Artículos científicos escritos en idioma inglés.
	Papers, libros, revistas, publicaciones de doctorados, ingenierías y maestrías.
	Artículos que contengan palabras clave en el título.
<b>Criterios de exclusión</b>	Artículos duplicados extraídos de las distintas bases de datos.
	Artículos incompletos en su estructura.
	Artículos que no sean de interés que no aporten con la investigación

Una vez ejecutados los criterios de inclusión y exclusión obtenemos el primer conjunto de 139 artículos académicos para ser analizados.

### **2.1.3. Filtración de estudios**

Para la filtración de estudios procedemos a la revisión de los Abstract o Resumen de cada uno de los 139 artículos seleccionados anteriormente con el fin de descartar los artículos que no sean de interés, es decir que no aporten con la investigación, Para ello realizamos una evaluación de calidad que consta en responder las siguientes preguntas con Sí o No:

**Tabla 7:** Evaluación de Calidad

EVALUACIÓN DE LA CALIDAD	
EC	PREGUNTA
EC1	¿El documento define el concepto de análisis de datos?
EC2	¿El documento define el concepto de analítica de datos?
EC3	¿El documento describe algún proceso/método/metodología de análisis de datos?
EC4	El documento describe algún proceso/método/metodología de analítica de datos?
EC5	¿El documento compara características entre análisis datos y analítica de datos?
EC6	¿El documento describe similitudes o dependencias entre análisis de datos y analítica de datos?
EC7	¿El documento describe los beneficios o ventajas de utilizar analítica de datos o análisis de datos?
EC8	¿El documento describe alguna restricción o limitación al momento de usar analítica datos o análisis de datos?

Una vez respondidas las preguntas y en colaboración del experto (tutor) obtenemos el grupo final de 29 artículos necesarios para leer, analizar y comenzar con la extracción de información necesaria para la investigación.

**Tabla 8:** Documentación Final Recolectada

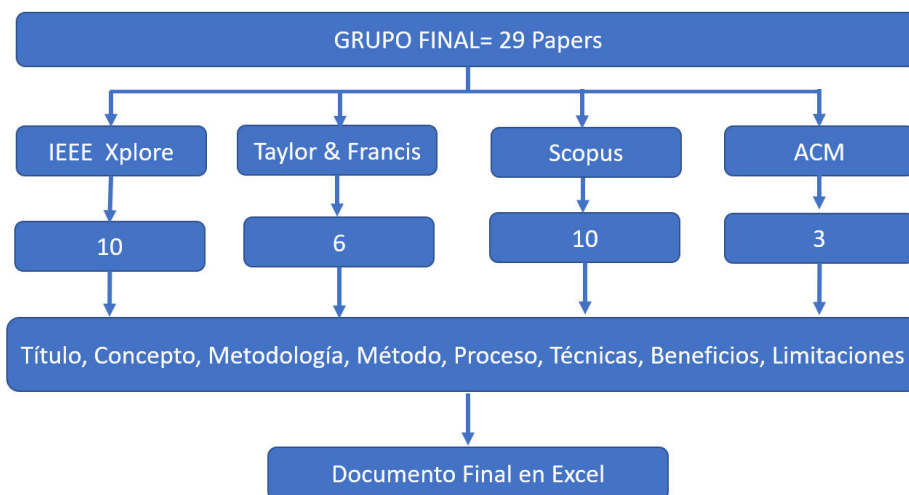
N	BASE	TITULO DE LA DOCUMENTACION RECOLECTADA
1	T	A Case Study on Privacy Threats and Research Challenges in Privacy Preserving Data Analytics
2	I	A Generalized Methodology for Data Analysis
3	I	A Performance Study of Big Data Analytics Platforms
4	I	A Survey on Real-time Big Data Analytics: Applications and Tools
5	T	A SWOT analysis of big data
6	T	Advancing manufacturing systems with big-data analytics: A conceptual framework
7	S	Big Data Analytics Concepts, Technologies Challenges, and Opportunities
8	I	Big Data Analytics Frameworks
9	I	Big data analytics using agile model
10	T	Big data analytics: a literature review
11	T	Big data contributions to human resource management: a systematic review
12	I	Big Data: A Review
13	S	Data Analysis in the Cloud: Models, Techniques and Applications
14	I	Data Analytics Methods for Anomaly Detection: Evolution and Recommendations
15	S	Data analytics: Models and algorithms for intelligent data analysis
16	S	Data science for business: benefits, challenges and opportunities
17	T	Envisioning the power of data analytics
18	S	How is data analytics changing entrepreneurial opportunities?
19	S	Perspectives on big data analytics and techniques
20	S	Sentiment analysis-A tool for data mining in big data analytics
21	S	Smart City Data Analysis
22	I	Smart Meter Data Analysis Issues: A Data Analytics Perspective
23	S	The intersection of data analytics and data-driven innovation
24	S	Towards intelligent data analysis: The metadata challenge
25	I	Yield Improvement Using Advanced Data Analytics
26	A	Big Data Analytics Architecture for Security Intelligence
27	A	Clarifying Big Data: The Concept and Its Applications
28	A	The 10 Vs, Issues and Challenges of Big Data
29	I	A Literature Review on Big Data Analytics Capabilities

#### 2.1.4. Extracción de información relevante

En esta etapa se extraerá la información de interés para el estudio del grupo final de 29 artículos, ya sean resúmenes, ideas o partes de los documentos, para ello se agrupará dicha información según la base de datos utilizada, el título del paper, conceptos, metodologías, método, proceso, técnicas, uso, beneficios y limitaciones, de cada uno de los papers como se muestra en la Figura 4 de tal forma que el registro de la información de los documentos permitirá realizar búsquedas y agruparlos ordenadamente, en general,



hacer una adecuada gestión de la información. Se usará Excel para mantener los datos más relevantes de cada estudio revisado.



**Figura 4:** Proceso de Extracción de Información Relevante

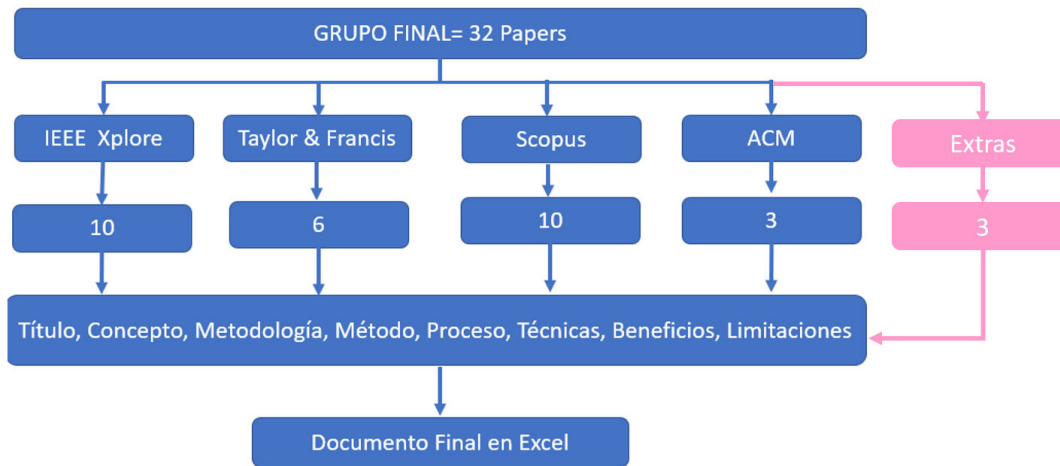
Una vez obtenido el documento final con la información más relevante, pudimos observar que la mayoría de información recopilada contenía únicamente datos sobre Data Analytics por lo cual para proceder con una comparación de criterios completa y de calidad entre ambos términos es necesario realizar una consulta académica adicional que permita complementar ámbitos de la investigación sobre Data Analysis que no se encontraron anteriormente con la búsqueda de la literatura, esta consulta académica se realizará en las bases de datos Google Académico (G), Scielo (S) y Latindex (L).

Para ello todos los artículos encontrados fueron sometidos al mismo proceso de inclusión y exclusión donde los títulos duplicados y que no contengan palabras claves fueron eliminados, así mismo los artículos que no tenían un Abstract relevante o de importancia para la investigación fueron excluidos, quedando para el final el siguiente grupo de artículos académicos extras:

**Tabla 9:** Artículos Académicos Extras

N	BASE	TITULO DE LA DOCUMENTACION RECOLECTADA
30	G	Data Analysis Techniques in Qualitative Research.
31	L	Intelligent Data Analysis: An Introduction
32	Sc	The Future of Data Analysis

Este grupo de artículos académicos extras se añaden al proceso de extracción de información relevante como se muestra en la Figura 5:



**Figura 5:** Proceso Final de Extracción de Información Relevante

## 2.2. Comparación de criterios

Con la información relevante extraída de los 32 papers totales, en esta etapa se realiza la comparación de ambos términos mediante la creación de una tabla, para lo cual se efectuarán los siguientes pasos:

### 2.2.1. Definición de criterios

Se definirán los aspectos que se emplearán para realizar la comparación entre ambos términos. Para este trabajo de investigación se decidió comparar Data Analytics y Data Analysis según los siguientes criterios: concepto, metodología, método, proceso, técnicas, beneficios y limitaciones. Con este fin encontramos esencial definir estos conceptos para clasificar de manera clara la información encontrada anteriormente y así evitar posibles confusiones.

#### Metodología

Como metodología se denomina la serie de métodos y técnicas de rigor científico que se aplican sistemáticamente durante un proceso de investigación para alcanzar un resultado teóricamente válido. En este sentido, la metodología funciona como el soporte conceptual que rige la manera en que aplicamos los procedimientos en una investigación [9].

## **Proceso**

Proceso es un conjunto de fases sucesivas o serie de pasos organizados y sistematizados cuyo fin es alcanzar un objetivo determinado, se trata de una planificación científica, técnica, social, política, de empresa, o simplemente de la vida cotidiana. Un proceso es un mecanismo diseñado por el ser humano para establecer un ordenamiento o mejora para servicio del hombre [10].

## **Método**

Método es un modo, manera o forma de realizar algo de forma sistemática, organizada y/o estructurada. Hace referencia a una técnica o conjunto de tareas para desarrollar una tarea [11].

## **Técnicas**

Como técnica se define la manera en que un conjunto de procedimientos, materiales o intelectuales es aplicado en una tarea específica, con base en el conocimiento de una ciencia o arte, para obtener un resultado determinado [12].

## **Beneficios**

La palabra beneficio se refiere a un bien que es dado o que es recibido. El beneficio siempre implica una acción o resultado positivo y que por consiguiente es buena y puede favorecer a una o más personas, así como satisfacer alguna necesidad [13].

## **Limitaciones**

La noción de límite es propia de muchas situaciones de la comunicación humana. Y la limitación consiste en el establecimiento de aquellos límites u obstáculos que impiden algo [14].

### **2.2.2. Resultados iniciales de la extracción**

A continuación, se presentará los resultados de la clasificación según los criterios definidos anteriormente en forma de una tabla comparativa completa y eficaz que permita distinguir claramente las diferencias entre Data Analytics y Data Analysis.

**Tabla 10:** Tabla Comparativa entre Análisis y Analítica de Datos

TABLA COMPARATIVA	
Data Analysis	<p>El proceso que utiliza un investigador para reducir los datos a una historia y su interpretación. El análisis de datos es el proceso de reducir grandes cantidades de datos recopilados para darles sentido.</p> <p>Merriam (1998) describe el proceso de análisis de datos como una acción compleja de ir y venir entre datos y conceptos, entre descripción e interpretación, utilizando tanto el razonamiento inductivo como el deductivo.</p> <p>Wolcott (1994) describe el análisis como basado en "conocimientos acordados, el reconocimiento de propiedades o estándares mutuamente reconocidos" que son "inherentemente conservadores, cuidadosos, sistemáticos". [7]</p>
Data Analytics	<p>Es el proceso de analizar datos en función de criterios de interés, obtener conocimientos más profundos de los datos y tomar decisiones importantes. [15]</p> <p>La analítica de Big Data es una herramienta importante que una organización puede aprovechar para utilizar la gran cantidad de datos a los que tiene acceso. Se trata de recopilar, almacenar y analizar grandes volúmenes de datos de manera eficiente con el objetivo de extraer información invaluable, pero a menudo oculta. [16]</p> <p>Big data analytics puede percibirse como un marco amplio para extraer el valor de datos tan grandes y complejos. [17]</p> <p>Big data analytics es el proceso de examinar big data para descubrir patrones ocultos, correlaciones desconocidas y otra información útil que se puede utilizar para tomar mejores decisiones. Big Data analytics es un canal de adquisición, extracción, limpieza, integración, agregación y visualización, análisis y modelado e interpretación. [18]</p> <p>Big data analytics: proceso de investigación de cantidades masivas de datos para revelar patrones ocultos y correlaciones secretas. [19]</p> <p>El término Data analytics se hizo popular a principios de la década de 2000. Data analytics se define como la aplicación de sistemas informáticos al análisis de grandes conjuntos de datos para respaldar decisiones. Data analytics es un campo muy interdisciplinario que ha adoptado aspectos de muchas otras disciplinas científicas como la estadística, el aprendizaje automático, el reconocimiento de patrones, la teoría de sistemas, la investigación de operaciones, o inteligencia artificial. [8]</p>
CONCEPTO	

Data analytics: es el gran desafío para identificar tendencias y patrones de comportamiento en el mercado. [20]

Analytics ha surgido como un término general para una variedad de iniciativas relacionadas con aplicaciones y BI. [21] Para algunos, es el proceso de analizar información de un dominio en particular, como el análisis de sitios web. Para otros, está aplicando la amplitud de las capacidades de BI a un área de contenido específica (por ejemplo, ventas, servicio, cadena de suministro, etc.). [21]

Existen diferentes metodologías de Data Analysis como se describe a continuación [7]:

1. NARRATIVA

LeCompte y Schensul: Revisar las preguntas de investigación, Escribir algo de historia, Describir un proceso social, Crear resúmenes de entrevistas, Crear colecciones de cotizaciones, Describir funciones / estructuras del grupo, Escribir los eventos críticos cronológicamente, Hacer una lista de hechos importantes.

Wolcott: conectarse a su propia experiencia.

Beck: Leer descripciones escritas.

Doucet y Mauthner: Relacionar la historia del participante con su propia experiencia, Ubicarse en la historia en relación con los participantes, Observar cómo los participantes hablan sobre sí mismos y su mundo.

Miles y Huberman: Realizar metáforas, Tener en cuenta las reflexiones sobre los datos recopilados.

2. CODIFICACION

LeCompte y Schensul: Crear viñetas, Crear un marco conceptual.

Wolcott: Identificar patrones de datos, Ampliar el análisis haciendo preguntas derivadas de los datos.

Beck: Desarrollar significado a partir de las declaraciones, Organizar significados, no grupos de temas.

Doucet y Mauthner: Dividir las transcripciones de texto en temas y subtemas superpuestos, Organice los datos de diferentes formas.

**METODOLOGIA** Data Analysis

3. INTERPRETACION

LeCompte y Schensul: Desarrollar una metáfora, Mirar el marco teórico; revisar las teorías relevantes, Participar en la especulación, Buscar relevancia para el programa / política, Evaluar el proyecto.

Wolcott: Relacionarse con la teoría, centrarse en la base de su tradición / disciplina, Evaluar contra la interpretación de los participantes, Posicionar los resultados en un marco analítico más amplio, Hacer inferencias usando razonamiento inductivo, Desarrollar el marco analítico.

Beck: Extraer enunciados significativos relacionados con el fenómeno en estudio.

Doucet y Mauthner: Cambiar el enfoque de casos individuales a grupos.

Miles y Huberman: Busque plausibilidad, Construya una cadena lógica de evidencia, Hacer coherencia conceptual / teórica, Evidencia de peso, Compruebe el significado de los valores atípicos, Usa casos

## METODOLOGIA

extremos, Hacer pruebas if- then, Descartar relaciones falsas, Desarrollar la interpretación de los hallazgos, Compare los datos para determinar qué se ajusta a sus suposiciones u otros hallazgos, Desarrollar premoniciones, Repetir la pregunta para que se ajuste a los datos.

### 4. CONFIRMACION

LeCompte y Schensul: Contrastar las vistas internas con las externas.

Wolcott: Criticar el proceso de investigación, Informar los procedimientos sistemáticos del trabajo de campo, Proponer un rediseño del estudio, Detenerse cuando llegue al final y preguntar qué debe hacerse a continuación, Comparar con un caso conocido, Analizar el proceso interpretativo.

Beck: Utilizar verificaciones de miembros para validar la descripción escrita, Utilizar el número para documentar, verificar y probar interpretaciones.

Doucet y Mauthner: Mirar las suposiciones.

Miles y Huberman: Triangular, Contar, Contrastes y comparaciones, Compruebe la representatividad, Compruebe los efectos de las investigaciones, Busque evidencia negativa, Replicar un hallazgo, Consulte las explicaciones del rival, Obtenga comentarios de los participantes, Verificar interpretaciones mediante comprobaciones de miembros, revisión por pares, triangulación, Compare constantemente datos anteriores con datos posteriores utilizando diferentes bases de comparación.

### 5. PRESENTACION

LeCompte y Schensul: Considerar la audiencia, Dibujar pantalla visual, Escribir en forma narrativa, tomando prestado de los participantes.

Wolcott: Enfatizar los datos importantes, Acepte sugerencias de editores / comités / colegas, Mostrar los resultados de forma gráfica, Explorar formatos alternativos para la presentación.

Beck: Utilice los resultados del análisis de datos para escribir una descripción exhaustiva.

Doucet y Mauthner: Escriba los resultados en forma de estudio de caso.

Miles y Huberman: Usar pantallas visuales.

Se han desarrollado varios modelos conceptuales como herramientas para ayudar a la introducción de data analytics en los sistemas de fabricación. Por ejemplo [17]:

Lechevalier, Narayanan y Rachuri (2014) proponen un marco específico de dominio para las aplicaciones de la analítica predictiva en la producción.

### Data Analytics

Las principales contribuciones de O'Donovan (2015b) son un conjunto de requisitos de datos y sistemas para implementar aplicaciones de mantenimiento de equipos en entornos industriales, y un modelo de sistema de información que proporciona una canalización de big data escalable y tolerante a fallas para integrar, procesar y analizar datos de equipos industriales.

	<p>Un marco para la conceptualización, Dutta y Bose (2015) presentan la planificación e implementación de proyectos de big data en empresas.</p> <p>Zhang (2017) proponen una arquitectura general para big-data analytics con el fin de tomar mejores decisiones de gestión del ciclo de vida del producto y de producción más limpia basadas en big data.</p> <p>Zhang (2017) proponen un marco para gestión de big data del ciclo de vida del producto impulsada para abordar desafíos como la falta de datos confiables y conocimientos valiosos que se pueden emplear para respaldar la toma de decisiones optimizada de la gestión del ciclo de vida del producto.</p> <p>Tao (2018) proponen un marco de fabricación inteligente basado en datos que consta de cuatro módulos: el módulo de fabricación, el módulo controlador, módulo de monitorización en tiempo real y módulo de procesamiento de problemas.</p> <p>Jun, Lee y Kim (2019) proponen una plataforma de análisis de big data basada en la nube para la industria manufacturera.</p>
<p><b>METODO</b></p>	<p><b>Data Analysis</b></p> <p>Merriam (1998) discutió varios enfoques para el análisis de datos, incluido el análisis etnográfico, el análisis narrativo, el análisis fenomenológico y el método comparativo constante.</p> <p>Bernard (2000) también sugiere varios enfoques para el análisis de datos, incluida la hermenéutica o el análisis interpretativo, el análisis narrativo y de desempeño, el análisis del discurso, el análisis de la teoría fundamentada, el análisis de contenido y el análisis transcultural. [7]</p> <p>Los conceptos de Big Data analytics se pueden clasificar en tres etapas principales: analítica descriptiva, analítica predictiva y analítica prescriptiva. El análisis descriptivo ayuda a comprender lo que ya sucedió con los datos. El análisis predictivo ayuda a anticipar lo que le sucederá, el análisis prescriptivo ayuda a responder ahora qué y para qué.</p> <p>Los métodos analíticos de big data se pueden clasificar en métodos clásicos y modernos. Los métodos clásicos son el análisis de texto, el análisis de audio y el análisis de video. Métodos modernos de análisis de datos como bosques aleatorios, redes neuronales artificiales, y las máquinas de vectores de soporte son ideales para identificar patrones lineales y no lineales en los datos. [18]</p> <p>Para hacer frente a desafíos de Big Data Analytics, los investigadores han propuesto muchos enfoques [22]: La mejora de la eficiencia de los métodos de descubrimiento de conocimiento de fuente única (Chang, Bai y Zhu, 2009), los métodos de extracción de datos dinámicos (R. Chen, Sivakumar y Kargupta , 2004;</p>



<p><b>METODO</b></p>	<p><b>Data Analytics</b></p>	<p>Domingos y Hulten, 2000), el método de distancia de información dual (DID) para la construcción de árboles de decisión (Ben-Gal, Dana, Shkolnik y Singer, 2014), análisis de relaciones no obvias de referencia (Kogge y Bayliss, 2013), restricciones de asociación máxima para extracción de patrones (Soysal, 2015), los datos modelo de creación de ejes impulsados para la medición de correlación (Nakanishi, 2014), un descenso de gradiente incremental paralelo con E / S escalables para análisis de big data (Qin y Rusu, 2013) y un enfoque orientado a servicios para aplicaciones de razonamiento (Cheptsov y Koller, 2013) ).</p> <p>La mayoría de los enfoques anteriores se pueden agrupar en tres niveles de análisis de datos según lo definido por Blackett [22]:</p> <p>Blackett (2013) clasifica data analytics en analítica descriptiva, analítica predictiva y analítica prescriptiva. En su taxonomía, el análisis descriptivo utiliza datos históricos para resumir lo que sucedió. Un informe comercial es un ejemplo de análisis descriptivo; puede proporcionar información histórica relacionada con las operaciones, ventas y producción de una empresa. Según la estimación de M. Wu (2013), la mayoría de las analíticas comerciales son descriptivas.</p> <p>El segundo nivel es el análisis predictivo; utiliza varias técnicas para predecir tendencias futuras. Dos ejemplos de análisis predictivo son la investigación de Shin y la investigación de Johansson. Al utilizar la infraestructura de big data, Shin, Woo y Rachuri (2014) diseñaron un modelo analítico para predecir el desempeño de la sustentabilidad, especialmente el consumo de energía. Johansson, Sonstrod, Linusson y Bostrom (2014) aplican la predicción conforme de Mondrian sobre el modelo resultante para obtener el aprendizaje en línea de árboles de regresión de datos de transmisión.</p> <p>El nivel final, analítica prescriptiva, se centra en la recomendación y asistencia para la toma de decisiones. En el mundo real, el análisis de predicciones no prescribe ninguna acción específica; sin embargo, la analítica prescriptiva puede ayudar a los usuarios finales a identificar problemas y encontrar soluciones óptimas para las empresas.</p>
	<p><b>Data Analysis</b></p>	<p>El proceso de análisis implica cinco pasos principales: integración de datos, gestión de datos, preprocesamiento de datos, minería de datos, y presentación de conocimientos. [18]</p> <p>Los proyectos típicos de data analysis se pueden dividir en varias fases. Los datos se evalúan y seleccionan, limpian y filtran, visualizan y analizan, y los resultados del análisis finalmente se interpretan y evalúan. El proceso de descubrimiento de conocimiento en bases de datos (KDD) comprende las seis fases de selección, preprocesamiento, transformación, minería de datos, interpretación y evaluación. El proceso estándar de la industria para la minería de datos (CRISP-DM) comprende las seis fases de comprensión empresarial, comprensión de datos, preparación de datos, modelado, evaluación e implementación. [8]</p> <p>Describen al análisis de campo como que incluye inscripción, descripción y transcripción. También sugieren que el análisis puede realizarse tanto de arriba hacia abajo como de abajo hacia arriba.</p> <p>Patton (1987) indica que durante el análisis ocurren tres cosas: los datos se organizan, los datos se</p>



<p style="text-align: center;"><b>PROCESO</b></p>		<p>reducen mediante el resumen y la categorización, y los patrones y temas en los datos se identifican y vinculan. [7]</p> <p>El análisis de datos incluye, entre otras cosas: procedimientos para analizar datos, técnicas para interpretar los resultados de dichos procedimientos, formas de planificar la recopilación de datos para hacer su análisis más fácil, más preciso o exacto, y toda la maquinaria y resultados de la estadística (matemática) que se aplican al análisis de datos. [23]</p>
	<p style="text-align: center;"><b>Data Analytics</b></p>	<p>La investigación y aplicaciones de la analítica de big data en la fabricación se puede dividir en (1) investigación teórica sobre modelos generales para introducir big-data analytics en la fabricación, exámenes de las situaciones existentes en la industria y el desarrollo de soluciones conceptuales, y (2) la investigación aplicada y el desarrollo de soluciones dedicadas. [17]</p> <p><b>ARQUITECTURA:</b> Se basa en el concepto de marco de ciclo de vida de datos que comienza con la captura de datos, procede a través de la transformación de datos y culmina con el consumo de datos. Se compone libremente de cinco capas arquitectónicas principales (1) datos, (2) agregación de datos, (3) análisis, (4) exploración de información y (5) gobernanza de datos. [18]</p> <p>A continuación, se describe las 3 fases de Data Analytics [24] :</p> <ol style="list-style-type: none"> <li>1.Fase de planificación: Implica actividades relacionadas con el análisis y obtención de requisitos. La fuente de recopilación de big data también se planificará en esta fase. La fuente de datos se puede utilizar de forma incremental para recopilar datos en función de la necesidad.</li> <li>2.Fase de desarrollo: Los datos recopilados de diferentes fuentes se analizan en esta fase. Los datos recopilados de múltiples fuentes de datos se combinarán en una sola hoja de trabajo. Los datos se unirán en dimensiones comunes y se presentarán en una sola vista. El analista de datos filtrará los datos no deseados. Esta técnica de eliminación ahorrará mucho tiempo en el procesamiento de datos no deseados que no son necesarios en absoluto. La reunión diaria de pie tendrá lugar en esta fase. Esto ayudará a realizar un seguimiento del desarrollo correcto para lograr el objetivo final. Las herramientas que se planifican durante la fase de planificación se utilizarán para el desarrollo. En esta fase también se desarrollará un algoritmo genérico. El ajuste fino del algoritmo ocurrirá en cada iteración. El uso de herramientas de análisis predictivo también se puede utilizar para predecir o pronosticar los datos. Las revisiones frecuentes también se encuentran en esta fase. Al final de cada una de las iteraciones, se recibirán los datos necesarios para la toma de decisiones clave.</li> <li>3. Fase de cierre: El cierre se produce cuando se cumplen todos los requisitos. Esta es la última fase antes</li> </ol>

		<p>de lanzar el producto con toda la documentación. Las pruebas finales se realizan en este punto y al final, el lanzamiento en sí es definitivo. En esta fase, el algoritmo desarrollado y los datos entregados por este se utilizarán para tomar decisiones clave para el negocio.</p> <p>PASOS INVOLUCRADOS EN BIG DATA ANALYTICS 1. Conozca su objetivo 2. Recopile de la fuente correcta 3. Filtre los no deseados 4. Verifique con frecuencia 5. Genere algoritmos 6. Herramientas de análisis predictivo del usuario. [24]</p>
<p><b>HERRAMIENTAS / TÉCNICAS</b></p>	<p>Data Analysis</p>	<p>Algunas de las herramientas de Data Analysis son [7]:</p> <p>Narrativa: estas técnicas incluyen acciones sugeridas que son preparatorias para el proceso de codificación o que ayudan al investigador a derivar significado a través de enfoques narrativos, como la lectura narrativa o la escritura de datos textuales.</p> <p>Codificación: estas técnicas se aplican a aquellas acciones que implican organizar y reorganizar los datos en categorías que permiten al investigador identificar relaciones entre categorías.</p> <p>Interpretación: Estas técnicas son aquellas acciones que permiten al investigador dar significado a las actividades narrativas y de codificación y facilitan la comprensión del investigador del marco conceptual generado a través del proceso de codificación.</p> <p>Confirmación: Estas técnicas permiten a los investigadores estar seguros de que las interpretaciones que han derivado son de los datos y no de la construcción del investigador.</p> <p>Presentación: Estas técnicas son aquellas acciones que los investigadores pueden utilizar para presentar los hallazgos a una audiencia en particular de una manera convincente y que sirven como una herramienta para un análisis y explicación adicionales.</p> <p>Algunas de las herramientas de data analytics son: Hadoop, como base para el almacenamiento de datos y el procesamiento distribuido, Spark es una poderosa herramienta para la computación en memoria que analiza datos en tiempo real. Además, para la analítica de datos en tiempo real, se necesitarán herramientas de ingestión de datos como Kafka, Storm, y Flume, que utilizan patrones para importar datos de forma que estén listos para ser analizados en grupos. [25]</p> <p>Se introdujeron muchas plataformas para el procesamiento de datos a gran escala para respaldar la analítica de big data. Las plataformas se pueden dividir en (procesamiento, almacenamiento, analíticas).</p>

## Data Analytics

Las plataformas analíticas deben ser escalables para adaptarse a la mayor cantidad de datos desde la perspectiva de la frontera, como Hadoop, Map Reduce, Hive, Pig, WibiData, PLATFORMA y Rapidminer. Las herramientas de análisis son escalables en dos direcciones: escala horizontal y escala vertical. El escalado vertical ayuda a instalar más procesadores, más memoria y hardware más rápido, por lo general, dentro de un solo servidor e involucra solo una instancia única de un sistema operativo. El escalado horizontal distribuye la carga de trabajo entre muchos servidores. Típicamente, varias instancias del sistema operativo se ejecutan en máquinas independientes. Map Reduce, Hadoop, Spark y Apache Flink son ejemplos de plataformas de escalado horizontal analítico de big data. [18]

Las herramientas de aprendizaje automático y análisis para data analytics de los entornos Hadoop y Elasticsearch, así como los algoritmos de minería de la biblioteca de Python, se utilizan para la predicción de datos. [26]

La necesidad de descubrir conocimientos está aumentando enormemente. Esto es más notable hoy en día gracias a las plataformas de procesamiento y almacenamiento de datos distribuidos y de bajo costo (por ejemplo, Hadoop). Permiten almacenar y procesar enormes conjuntos de datos en grandes grupos de hardware básico. [27]

Análisis de Sentimiento como herramienta de data analytics puede interpretarse como la tarea de detectar la perspectiva de varios autores sobre conjuntos de datos particulares.

El análisis de sentimientos intenta clasificar las opiniones y declaraciones como positivas, negativas o neutrales. También conocido como minería de opiniones, comprende el comportamiento humano y categoriza las opiniones expresadas. Utiliza el procesamiento del lenguaje natural (PNL) para clasificar e interpretar los datos disponibles en línea.

Utiliza una combinación de minería de datos y PNL para explorar, extraer y purificar los abundantes datos que se encuentran en la World Wide Web para que se adapten al propósito en cuestión. La PNL utiliza métodos computacionales para captar, percibir y crear contenido vernáculo humano. [28]

Análisis de Sentimiento consiste en los siguientes Pasos [28]:

## Data Analytics

Paso 1: Es importante identificar el objetivo de realizar el análisis en primer lugar. El propósito podría ser diferente según la industria que utilice la herramienta.

Paso 2: Una vez establecido el objetivo del estudio, es imperativo decidir sobre el canal del que se deben extraer los datos para el análisis. La información se puede obtener directamente de la Web. Una vez acumulada la información requerida, se ingresa en el marco del análisis de sentimientos. El texto debe estar libre de palabras o frases no deseadas. El texto está hecho para ser más preciso y se resaltan las emociones expresadas a través de palabras.

		<p>Paso 3: El siguiente paso consiste en diferenciar las palabras en función de su polaridad, lo que significa el nivel de positividad o negatividad expresada por una palabra. Implica analizar los datos en términos de gramática y partes del discurso.</p> <p>Paso 4: El análisis se vuelve más ingenioso al refinarlo en términos de sinónimos, verbos y sustantivos.</p> <p>Paso 5: El paso final implica puntuar los datos así analizados. Las frases y palabras que transmiten los sentimientos se resaltan y puntúan según su polaridad. La puntuación es un procedimiento mediante el cual se puntúa la magnitud de la emoción.</p>
	<p>Data Analysis</p>	<p>Actualmente, la generación de datos ha aumentado a pasos agigantados. Las empresas utilizan todos los datos generados para tomar decisiones cruciales que pueden obstaculizar o impulsar su negocio. Por lo tanto, el análisis de datos es un proceso para extraer información significativa de datos como patrones ocultos, correlaciones desconocidas, tendencias del mercado y preferencias de los clientes. Esto se hace mediante el estudio y análisis de los datos adquiridos. Cuando se trata de los beneficios del análisis de datos, se puede mencionar su uso para la gestión de riesgos, para diversas innovaciones de productos, también ayuda a las organizaciones a tomar decisiones más rápidas y mejores y también se utiliza para mejorar la experiencia del cliente. Todo esto es posible solo cuando se analizan los datos. Los datos desordenados y desestructurados deben procesarse y analizarse para obtener información y descubramos patrones ocultos a partir de ellos. [29]</p>
<p><b>BENEFICIOS/ IMPORTANCIA</b></p>	<p>Data Analytics</p>	<p>Al clasificar los métodos de analítica de datos en tiempo real propuestos por sus diferentes áreas de aplicación y herramientas utilizadas, los futuros investigadores y científicos de datos tendrán la capacidad de mejorar sus negocios y tecnologías con el beneficio de la analítica de datos en tiempo real mientras utilizan las herramientas adecuadas para sus circunstancias. [25]</p> <p>Los médicos almacenan una gran cantidad de datos sobre el historial médico de los pacientes, la medicación y otros detalles. La empresa farmacéutica está almacenando una gran cantidad de datos. Estos datos son de naturaleza muy compleja y, a veces, los profesionales no pueden correlacionarse con otra información, por lo tanto, la información importante permanece oculta. Aplicando técnicas de análisis avanzadas, esta información oculta se puede extraer, lo que da como resultado una medicación personalizada. Las técnicas de análisis avanzadas también pueden ayudar a comprender las causas genéticas y ambientales de las enfermedades. [30]</p> <p>Después de que una organización aplicó alguna forma de big data analytics, se producen estos beneficios:</p>

marketing mejor dirigido, conocimientos comerciales más directos, segmentación basada en el cliente, reconocimiento de las ventas y oportunidades de mercado. [19]

Data analytics se presenta como rápido, accesible, revelador, panorámico, profético e inteligente [31]:

**RAPIDO:** Cuanto más rápida sea la analítica, entonces la lógica dice, mejor y más útil es. El análisis rápido conduce a organizaciones receptivas, flexibles y exitosas, está implícito. Parte de esta velocidad está ligada a la visualización del poder de estos análisis. Estos análisis no solo se describen como rápidos, también son poderosos. La idea es que las empresas que utilizan data analytics puedan acelerar sus prácticas y mantenerse al día con lo que se percibe como un mundo en aceleración y un conjunto de competidores.

**ACCESIBLE:** Junto con la velocidad, la accesibilidad de estos análisis es un segundo tema clave. Una idea dominante aquí es que el conocimiento se puede obtener sin ningún conocimiento técnico real. Por lo tanto, estos son análisis instantáneos en el sentido de que el análisis de datos produce resultados instantáneos, pero también son instantáneos en el sentido de que los usuarios pueden interpretar y comprender fácilmente los análisis que encuentran. Los datos pueden ser incomprensibles, pero los análisis son intuitivos y, por lo tanto, se puede acceder y comprender fácilmente.

**REVELADOR:** La analítica se presenta como el medio por el cual se puede desenterrar el valor "oculto" o donde se pueden aprovechar nuevos tipos de valor. Se trata entonces de crear o extraer valor oculto de información, comunicación y sociedad. Las imágenes son de una gran cantidad de datos no utilizados desperdiciados que podrían ser potencialmente lucrativos.

## Data Analytics

### BENEFICIOS/ IMPORTANCIA

**PANORAMICO:** Data analytics se presenta como casi omnipotente. Data analytics permiten al usuario "ver las esquinas y el futuro". Data analytics arroja luz sobre los puntos ciegos, aquellas partes de la organización escondidas en las esquinas de repente se vuelven visibles. Nada está en las sombras. El enfoque aquí está en la capacidad de reunir diferentes flujos de datos para aumentar la cobertura de los análisis. En este sentido, términos como 'integrado' o 'integración', 'mezcla' y 'armonización' se utilizan para sugerir el uso continuo de datos variados para producir un conjunto coherente y completo de conocimientos. Además de proporcionar una vista panorámica del funcionamiento interno de las organizaciones, estos Data analytics también se describen como una vista panorámica perfecta del contexto exterior.

**PROFETICO:** La visibilidad que brinda data analytics, como ya se ha insinuado, no se trata solo del momento en que ocurre. Se dice que actúan en tiempo real y proporcionan una visión completa de las condiciones internas y externas, pero más allá de esto, también se dice que tiene propiedades proféticas. Data analytics abre un mundo en el que es posible anticipar lo que sucederá y responder en consecuencia. Se trata de ver el futuro para proteger y mantener el valor y una ventaja competitiva. Se nos dice que podemos anticiparnos en lugar de responder a través de estos análisis.



## BENEFICIOS/ IMPORTANCIA

### Data Analytics

**INTELIGENTE:** Lo anterior muestra que las capacidades predictivas de data analytics a menudo se asocian con una inteligencia latente que reside en estos sistemas. Pueden aprender para predecir. La noción de aprendizaje es importante aquí. Las analíticas no son pasivas, más bien, se presentan como dispositivos inteligentes y activos que pueden aprender, adaptarse y desarrollarse en los conocimientos que producen activamente. De esta forma, responden a necesidades particulares, aprendiendo lo que se requiere, al mismo tiempo que aprenden de los datos que se acumulan. En efecto, la palabra "inteligente" se usa mucho para evocar esta inteligencia latente y poder de aprendizaje. Se trata tanto de la inteligencia de los análisis como de la inteligencia de los conocimientos que crean, y luego de la inteligencia implícita y heredada de las organizaciones que los utilizan.

La integración de data analytics puede generar muchas ventajas para los emprendedores, tales como [21]:

**Decisión de respaldo:** como se mencionó anteriormente, los emprendedores pueden hacer uso de la gran cantidad de datos relevantes para su negocio en particular. Por lo tanto, tendrían que filtrar los datos de acuerdo con sus necesidades específicas y derivar el significado de los datos que mejor se ajusten a ellos. Esto no solo ampliará su comprensión de su propio dominio, sino que también facilitará una mejor toma de decisiones, lo que a su vez mejorará la eficiencia operativa.

**Reducción de costos:** se ha descubierto que los macrodatos pueden ser extremadamente instrumentales para aumentar las arquitecturas existentes de las empresas. Además, cuando se toman decisiones más precisas, también se alivia la posibilidad de incurrir en pérdidas. Por lo tanto, con el uso correcto de la analítica, las empresas emergentes y los emprendedores pueden tener éxito en la reducción de sus costos operativos, que suele ser uno de los mayores desafíos para cada nueva empresa.

**Perspectivas del cliente:** El crecimiento de cualquier empresa depende de cómo mantener las preferencias, gustos de sus clientes en cuenta para diseñar sus productos y servicios. Big Data analytics puede ayudar a las empresas a obtener acceso a la información necesaria y relevante. Por ejemplo, las redes sociales presentan una gran herramienta para adquirir y asimilar enormes volúmenes de información de los clientes y se pueden utilizar de forma eficaz para recopilar datos con este fin.

**Usos de datos abiertos:** durante el último año, ha habido un aumento en el uso percibido de datos abiertos por parte de los empresarios para crear nuevos productos y servicios. Los datos abiertos, además de su potencial económico y de creación de nuevas actividades, también se encuadran dentro de una elección filosófica o ética. Cifran el comportamiento humano colectivo, y por lo tanto también pertenecen a quienes medimos estos comportamientos. La cultura de este fenómeno se basa en la disponibilidad de datos para una orientación comunicativa.

El marketing ha utilizado data analytics durante años. Las firmas financieras pueden comprender mejor las necesidades de gasto de los consumidores y los intereses personales para orientar mejor sus productos crediticios. Los datos de las redes sociales enriquecen la comprensión de un ejecutivo al complementar los datos de preferencias y recomendaciones casi en tiempo real. El precio se puede calibrar mejor en función de la popularidad de un producto o destino. Complementar el marketing con datos mientras lo ejecuta con un toque y cuidado personal real realmente cambia la mente de los consumidores. [32]

Desafortunadamente, como ocurre con la buena salud, las desviaciones del análisis de datos inteligente pueden ocurrir en muchas direcciones. Datos distorsionados. elección incorrecta de preguntas, aplicación incorrecta de herramientas de análisis de datos. el sobreajuste, un modelo demasiado idealizado, un modelo que va más allá de las diversas fuentes de incertidumbre y ambigüedad en los datos, etc., representan todas las posibilidades para un análisis de datos poco inteligente.

A menudo, solo en retrospectiva podemos ver que un análisis no fue tan buena idea después de todo. Esta es una de las razones por las que el dominio del análisis inteligente de datos es tan interesante. No se trata en gran medida de aplicar simplemente un directorio de herramientas a un problema dado, sino más bien de una evaluación crítica, exploración, prueba y evaluación. Es un dominio que requiere inteligencia y cuidado. así como la aplicación de conocimientos y experiencia sobre los datos. Es una disciplina desafiante y exigente. Además, es fundamentalmente interdisciplinar, tomando ideas de varios campos de actividad. Y es una disciplina que sigue evolucionando. [29]

Data Analysis

Beck (2003) señaló los errores comunes que hacen los investigadores principiantes al analizar datos cualitativos. Estos errores incluyen la mezcla de datos, el cierre prematuro y el cierre demasiado retrasado. Explicó que a veces es más fácil para los principiantes clasificar los datos que pasar por el proceso de análisis e interpretación. Los errores típicos que señaló estaban relacionados con el cierre, como cuando el investigador deja de recopilar datos antes de que se saturen todas las categorías o antes de que se recopilen suficientes datos o "se cree un resultado conceptual sólido". Señaló que el análisis de datos cualitativos es complejo y requiere que los datos se organicen, reorganicen, presenten y representen. Lincoln y Guba (1985) señalaron que saber cuándo dejar de recolectar datos y concentrarse en simplemente analizar los datos recolectados está determinado por el hecho de que el investigador haya agotado las fuentes, la saturación de categorías, la aparición de regularidades en los datos y la sobre extensión. [7]

Amenazas a la privacidad en la analítica de datos  
 Las amenazas clave a la privacidad incluyen: Vigilancia, Divulgación, Discriminación, Aceptación personal, Abuso, Robo de información, Ataque de suplantación de identidad, Seguridad e integridad nacional. [15]

LIMITACIONES /  
 RESTRICCIONES

Data Analytics

Como las consultas de Big Data analytics tienden a ser complejas, la calidad de los planes físicos para procesarlas juega un papel importante en el rendimiento general: Importancia de la optimización basada en costos, Importancia del formato de almacenamiento y el paralelismo de E / S, Importancia del procesamiento del flujo de datos, Esquema desnormalizado frente a esquema normalizado. [16]

Al implementar big data analytics, estos problemas son barreras potenciales: cosas inexpertas, costo, privación de patrocinio empresarial, difícil de diseñar sistemas analíticos, falta de software de base de datos actual en análisis. [19]

Algunas de las limitaciones de Data Analytics son [30]:

-Privacidad, seguridad y confianza: La organización que utiliza Big Data, debe estar comprometida a proteger la privacidad y seguridad de sus usuarios y debe garantizar que la organización debe cumplir con todas las leyes relacionadas con la privacidad y la seguridad para mejorar la protección y establecer límites claros para el uso de información personal. La confianza en la organización debe mantenerse a medida que aumenta el volumen de almacenamiento de datos. La confianza que los usuarios tienen en estas agencias y su capacidad para conservar de forma segura la información personal puede verse fácilmente afectada por la filtración de datos o información al dominio público.

-Gestión e intercambio de datos: Las agencias se dan cuenta de que para que los datos tengan algún valor, deben ser detectables, accesibles y utilizables. Las agencias deben cumplir con estos requisitos, pero seguir adhiriéndose a las leyes de privacidad. Las tendencias actuales hacia los datos abiertos se han centrado en poner los conjuntos de datos a disposición del público. Las agencias deben centrarse en hacer que los datos estén disponibles, abrir y estandarizar dentro y entre agencias de tal manera que les permita a las agencias usar y colaborar en la medida que sea posible por las leyes de privacidad.

-Tecnología y habilidades analíticas: Big Data Analytics ponen mucho estrés en los proveedores de TIC para desarrollar nuevas herramientas y tecnología para manejar datos complejos. Las herramientas y tecnologías actuales no pueden almacenar, procesar y analizar una gran cantidad de datos diversos. Los proveedores y desarrolladores de sistemas y soluciones de Big Data, incluido el software de código abierto, están desarrollando herramientas más capaces para simplificar los desafíos de Big Data Analytics.

-Almacenamiento y recuperación de datos: las tecnologías actualmente disponibles pueden manejar la entrada y el almacenamiento de datos. Pero las herramientas diseñadas para el procesamiento de transacciones que agregarán actualizarán y buscarán cantidades pequeñas o enormes de datos, no pueden manejar big data. Aún se desconoce cómo manejar datos semiestructurados o no estructurados para procesarlos.



## LIMITACIONES / RESTRICCIONES

### Data Analytics

- Calidad frente a cantidad: cuando se trata de una gran cantidad de datos, a veces es difícil decidir: · ¿Qué datos son inapropiados y cómo seleccionamos los más adecuados? · ¿Cómo se asegura la autenticidad de los datos? · ¿Cómo estimar el valor de los datos?
  - Crecimiento y expansión de datos: a medida que las organizaciones aumentan sus servicios, también se espera que sus datos crezcan. Pocas organizaciones también consideran la expansión de datos debido a que los datos crecen en riqueza, los datos evolucionaron con nuevas técnicas.
  - Velocidad y escala: cuando aumenta el volumen de datos, es difícil obtener información sobre los datos dentro de un período de tiempo. Obtener información sobre los datos es más importante que procesar un conjunto completo de datos. El procesamiento de datos casi en tiempo real siempre requerirá un intervalo de procesamiento para producir una salida satisfactoria.
  - Datos estructurados y no estructurados: la transición entre datos estructurados, almacenados en tablas bien definidas y datos no estructurados (imágenes, videos, texto) necesarios para el análisis afectará el procesamiento de los datos de un extremo a otro. La invención de nuevas tecnologías no relacionales proporcionará cierta flexibilidad en la representación y el procesamiento de datos.
  - Propiedad de los datos: una gran cantidad de datos reside en los servidores de los proveedores de servicios de redes sociales. Estos datos no son realmente de su propiedad, pero almacenan datos de sus usuarios. El propietario real de la página es quien ha creado la página o la cuenta. Este es un gran desafío continuo en el área de las redes sociales.
- Durante el proceso de data analytics, se encontraron varios desafíos, pero los 3 principales a mencionar son [33]:
- Computación de potencia. Nuestro esquema de recopilación de datos nos permite recibir datos de diferentes fuentes y muestra un gran potencial como herramienta de toma de decisiones, pero también tiene una gran complejidad de datos que hace que el análisis sea intensivo en computadoras. Se necesita un servidor con suficiente memoria para asignar grandes cantidades de datos y un procesador de gran peso para calcular los algoritmos de manera oportuna para entregar la potencia de procesamiento informática masiva necesaria.
  - Calidad de los datos. Es importante asegurar la calidad de los datos. Con ese propósito, se toman diferentes enfoques para asegurar la integridad, singularidad, precisión y coherencia de los datos. Todas estas actividades dan confianza en el análisis y las conclusiones.
  - Visualización de datos. Aunque se procesan datos complejos, la representación de los hallazgos debe ser en un formato sencillo que lo haga comprensible para la audiencia de manera que la comunicación de

## Data Analytics

dichos datos sea efectiva. Para realizar la tarea, se utilizan una variedad de gráficos y tablas.

En un esfuerzo por aprovechar datos para el análisis de seguridad, surgen varios desafíos relacionados con lo siguiente [34]:

- Volumen de datos: a medida que los conjuntos de datos aumentan de tamaño, también lo hace la capacidad de almacenamiento requerida para albergar los datos.
  - Incoherencia de datos: datos recopilados de fuentes heterogéneas (por ejemplo, registros de red, registros de aplicaciones de software, y registros de eventos biométricos) varían en su estructura y formato. Sin una capa de preprocesamiento, los datos suelen ser inconsistentes, incompletos y ruidosos, lo que dificulta la realización de análisis y correlaciones adicionales.
  - Análisis de datos: obtener conocimientos a partir de los datos se vuelve más difícil a medida que aumenta el tamaño de los conjuntos de datos. Las aplicaciones de big data, particularmente en análisis de seguridad, desean calcular en tiempo real para detectar y reportar amenazas y anomalías de manera efectiva. Sin embargo, realizar correlaciones en tiempo real en datos de gran volumen y, a menudo, inconsistentes presenta una tarea complicada.
  - Visualización de datos: el objetivo de la visualización de datos en el análisis de seguridad de big data es conferir información oculta en complejos y grandes conjuntos de datos y el estado de seguridad en tiempo real de una infraestructura a un humano utilizando la representación visual más eficaz. Esto ayuda al administrador o analista de seguridad a interpretar fácilmente los datos y los resultados; Armandando al individuo con el conocimiento necesario para tomar decisiones de seguridad proactivas. Sin embargo, la dificultad surge al intentar visualizar grandes conjuntos de datos debido al volumen y la naturaleza multifacética de los datos.
- La visualización de big data actual tiene una funcionalidad deficiente y baja escalabilidad, tiempo de respuesta e interactividad.
- Otra limitación importante en el desarrollo de la visualización de datos para big data es la escasez de talento humano. Los analistas de big data deben dominar habilidades matemáticas complejas que son difíciles de educar y que requieren un largo período de tiempo para capacitarse.

Debilidades: Falta de profesionales capacitados, Infraestructura inadecuada, Falta de visión clara, Calidad de los datos y análisis, Preocupaciones sobre la privacidad. Amenazas: Los problemas de privacidad pueden causar oposición pública / privada a los macrodatos, datos aislados. [35]

### **2.2.3. Resultados finales de la extracción**

En esta etapa del proyecto de investigación es necesario la creación de una tabla final con un resumen práctico de la información más importante que se destaca dentro de la Tabla 10, para ilustrar de mejor manera los dominios de estudio de los conceptos de análisis y analítica de datos.

**Tabla 11:** Tabla Comparativa Final entre Análisis y Analítica de Datos

TABLA COMPARATIVA FINAL	
<b>CONCEPTOS</b>	Data Analysis Para la definición de Data Analysis se decidió utilizar el concepto de [7]: Proceso de reducir grandes cantidades de datos recopilados para darles sentido, basado en "conocimientos acordados, el reconocimiento de propiedades o estándares mutuamente reconocidos" que son "inherentemente conservadores, cuidadosos y sistemáticos".
	Data Analytics La definición más importante y completa de este término lo obtuvimos de los autores Shehab N, M Badawy y Arafat H, donde definen a Data Analytics como el proceso de examinar big data para descubrir patrones ocultos, correlaciones desconocidas y otra información útil que se puede utilizar para tomar mejores decisiones. [18]
<b>METODOLOGIA</b>	Data Analysis Existen diferentes metodologías de Data Analysis según [7] las cuales son clasificadas principalmente en: Narrativa, Codificación, Interpretación, Confirmación y Presentación; cada uno de estos autores detallan sus diferentes metodologías con pasos a desarrollar dentro de las mismas.
	Data Analytics Encontramos diferentes metodologías de Data Analytics para sistemas de fabricación, entre las que destacamos como importantes [17]: Un marco de dominio para las aplicaciones de analítica predictiva, Una arquitectura general para Big Data Analytics con el fin de tomar mejores decisiones de gestión del ciclo de vida de un producto y Un modelo de sistema de información que proporciona una canalización de Big Data escalable y tolerante a fallas para integrar, procesar y analizar datos de equipos industriales.
	Data Analysis Se encontró varios enfoques para el análisis de datos según [7] como: análisis etnográfico, análisis narrativo, el análisis fenomenológico, el método comparativo constante, análisis estadístico, análisis interpretativo, análisis del discurso, análisis de la teoría fundamentada, el análisis de contenido y el análisis transcultural.
	Data Analytics Big Data analytics se pueden clasificar en tres etapas principales según Shehab N, Badawy M y Arafat H: analítica descriptiva, analítica predictiva y analítica prescriptiva. El análisis descriptivo ayuda a comprender lo que ya sucedió con los datos. El análisis predictivo ayuda a anticipar lo que le sucederá, el análisis prescriptivo ayuda a responder ahora qué y para qué. [18]
<b>METODOS</b>	Data Analysis El proceso de análisis de datos según [18] implica cinco pasos principales: integración de datos, gestión de datos, preprocesamiento de datos, minería de datos, y presentación de conocimientos.
	Data Analytics Para el proceso de Data Analytics se obtuvieron 3 fases importantes, que son [36]: 1. Fase de planificación: Implica actividades relacionadas con el análisis y obtención de requisitos. 2. Fase de desarrollo: Los datos recopilados de diferentes fuentes se analizan en esta fase. El analista de datos filtrará los datos no deseados. Se desarrollará un algoritmo genérico.
<b>PROCESO</b>	

		<p>3. Fase de cierre: El cierre se produce cuando se cumplen todos los requisitos. Las pruebas finales se realizan en este punto y al final, el lanzamiento en sí es definitivo. En esta fase, el algoritmo desarrollado y los datos entregados por este se utilizarán para tomar decisiones clave para el negocio.</p> <p>Algunas de las herramientas de Data Analysis según [7] son:</p> <ul style="list-style-type: none"> <li>-Narrativa: estas técnicas ayudan al investigador a derivar significado a través de enfoques narrativos, como la lectura narrativa o la escritura de datos textuales.</li> <li>-Codificación: estas técnicas se aplican a aquellas acciones que implican organizar y reorganizar los datos en categorías que permiten al investigador identificar relaciones entre categorías.</li> <li>-Interpretación: Estas técnicas son aquellas acciones que permiten al investigador dar significado a las actividades narrativas y de codificación.</li> <li>-Confirmación: Estas técnicas permiten a los investigadores estar seguros de que las interpretaciones que han derivado son de los datos y no de la construcción del investigador.</li> <li>-Presentación: Estas técnicas son aquellas acciones que los investigadores pueden utilizar para presentar los hallazgos y explicación adicionales.</li> </ul>
<p><b>HERRAMIENTAS</b></p>	<p>Data Analysis</p>	<p>Algunas de las herramientas de Data Analytics son divididas como escalables en dos direcciones, escala horizontal y escala vertical [18]; En [28] realizan una descripción general de la herramienta de análisis de sentimientos donde se detallan su principales fases para su desarrollo.</p>
<p><b>BENEFICIOS</b></p>	<p>Data Analysis</p>	<p>Cuando se trata de los beneficios del análisis de datos, se puede mencionar su uso para la gestión de riesgos, para diversas innovaciones de productos, también ayuda a las organizaciones a tomar decisiones más rápidas y mejores y también se utiliza para mejorar la experiencia del cliente. Todo esto es posible solo cuando se analizan los datos. Los datos desordenados y desestructurados deben procesarse y analizarse para obtener información y descubramos patrones ocultos a partir de ellos. [29]</p>
<p><b>LIMITACIONES</b></p>	<p>Data Analysis</p>	<p>Dentro de los beneficios más importantes de Data Analytics encontramos que se presenta como rápido, accesible, revelador, panorámico, profético e inteligente[31] y además encontramos algunas ventajas que pueden generar los emprendedores al momento de utilizar Data Analytics. [21]</p> <p>Algunos de los desafíos encontrados según Berthold Michael y Hand David J son: Datos distorsionados. elección incorrecta de preguntas, aplicación incorrecta de herramientas de análisis de datos. el sobreajuste, un modelo demasiado idealizado, un modelo que va más allá de las diversas fuentes de incertidumbre y ambigüedad en los datos. [29]</p> <p>Para finalizar los autores Chandarana y Vijayalakshmi describen algunas de las más importantes limitaciones de utilizar Data Analytics al momento de hablar sobre: Privacidad, seguridad y confianza, Gestión e intercambio de datos, Tecnología y habilidades analíticas, Almacenamiento y recuperación de datos, Calidad frente a cantidad, Crecimiento y expansión de datos, Velocidad y escala, Datos estructurados y no estructurados, Propiedad de los datos. [30]</p>

### **2.3. Conclusiones del capítulo**

En el capítulo 2 “METODOLOGÍA”, se desarrolla la metodología para la Búsqueda de la literatura y la Comparación de criterios, dando como resultado una Tabla Comparativa Final que contiene la información más relevante de toda la investigación. A continuación, se procederá a detallar la metodología aplicada al caso de estudio.

## **CAPITULO 3**

### **3. PROYECCIÓN DEL CASO DE ESTUDIO Y RESULTADOS**

En esta etapa se aplicará todo el conocimiento adquirido anteriormente sobre los términos de análisis y analítica de datos, a través de la ejemplificación de un caso de estudio. En esta sección se presenta el desarrollo de esta aplicación.

#### **3.1. Proyección del caso de estudio**

##### **3.1.1.Contexto del caso de estudio**

Para el siguiente caso de estudio se propone realizar un análisis predictivo que consta de la extracción y análisis de sentimientos de tweets, para determinar el futuro candidato ganador de las elecciones del 2021, según el nivel de aceptación del candidato en las 5 provincias más pobladas del Ecuador.

Para ello, durante un intervalo de tiempo aproximado de 3 días, se decide extraer tweets de las 5 provincias más pobladas del Ecuador como son: Guayas, Pichincha, Manabí, los Ríos y Azuay [37] ; esta muestra consta de 3000 tweets por candidato, cada uno de estos tweets son almacenados en diferentes listas respectivamente. Una vez que se terminen de almacenar cada uno de los tweets mencionados anteriormente se procede al análisis de sentimientos, donde se determinan porcentajes de polaridad que se transmite a través de las palabras o frases que se extraen de los tweets, para finalmente presentar resultados de este porcentaje como nivel de aceptación de cada candidato utilizando graficas que permitan ilustrar dichos resultados de mejor manera y determinar el ganador.

##### **3.1.2.Definición de Herramientas**

Para este caso de estudio se determinó necesarias las siguientes herramientas:



**Tabla 12:** Herramientas para el Caso de Estudio

Nombre	Descripción
Twitter	Es una red social y un servicio de microblogging para la comunicación en tiempo real utilizado por millones de personas y organizaciones. [38]
Python	Es un lenguaje de programación versátil multiplataforma y multiparadigma que se destaca por su código legible y limpio. [39]
Tweepy	Es una librería Python fácil de usar para acceder a la API de Twitter. [40]
Textblob	Es una librería Python para procesar datos textuales. Proporciona una API simple para bucear en tareas comunes de procesamiento de lenguaje natural (NLP). [40]
Matplotlib	Es una librería de gráficos 2D de Python que produce figuras de calidad de publicación en una variedad de formatos impresos y entornos interactivos en todas las plataformas. [40]

### 3.1.3. Método de la Proyección del caso de estudio

Para el desarrollo del aplicativo decidimos seguir las siguientes etapas con sus respectivos pasos:

#### **Etapas Iniciales: Extracción de la Información**

Para tener acceso y descargar los diferentes tweets que nos servirán como fuente de información para comenzar con el análisis, necesitamos:

1. Tener una cuenta como desarrollador en Twitter.
2. Crear una aplicación en la página de desarrolladores de Twitter, para ello:
  - a) Nos dirigimos a la dirección web en <https://apps.twitter.com/> e iniciamos Twitter con la cuenta desarrollador.
  - b) Luego nos dirigimos a la pestaña "Apps" y comenzamos llenando el formulario "Create an application" con algunos campos requeridos para la creación de nuestra App, como lo son el nombre, dirección web, una descripción e incluso un pequeño texto donde describimos para que utilizaremos nuestra App. Finalizamos el proceso pulsando la opción de "Create your Twitter application"





### 3. Importar la librería Tweepy

Una vez obtenidos el API key y el API secret ya podemos iniciar con la importación de la librería Tweepy. Esta librería nos permitirá acceder a las publicaciones, no solo de comentarios pasados, sino también en tiempo real.

Para su funcionamiento necesitamos instalar la librería utilizando el siguiente comando:

```
pip install tweepy
```

Una vez instalada importamos la librería desde nuestro código de la siguiente manera:

```
import tweepy
```

Ahora necesitamos las claves y tokens que nos proporcionó Twitter al momento de crear nuestra aplicación, guardamos estos permisos en cuatro variables para hacerlo un poco más sencillo, como se muestra a continuación.

```
consumerKey = "OJx1KY5ODi5MAxq6fDWRyFaDQ"  
consumerSecret = "XM4RpNhQAhciam0SIPAvowhys3WjFDmS7zVwrHiAykRabrYHzH"  
accessToken = "67497207-A6VvGqqrchy1SOy2ff8LM9KaJCeczEr2opVWpJZqb"  
accessTokenSecret = "dMV9pDuVLcHAFCvTceQCL2jhuzaPVR1d3yWHke0fsNM00"
```

**Figura 8:** Captura de Claves y Tokens

Ya con las variables listas realizamos la autenticación con el API de Twitter. Creamos una instancia al objeto OAuthHandler donde asignamos las llaves como parámetros, esto lo guardaremos en una variable que utilizaremos después para hacer la conexión al API.

```
auth = tweepy.OAuthHandler(consumer_key=consumerKey, consumer_secret=consumerSecret)
```

**Figura 9:** Captura de la Autenticación con el API de Twitter

Una vez concluido con lo anterior, asignamos los tokens de acceso utilizando el objeto de OAuthHandler (en nuestro caso es la variable auth) y con el método set\_access\_token () asignaremos los valores de los tokens.

```
auth.set_access_token(accessToken, accessTokenSecret)  
api = tweepy.API(auth, wait_on_rate_limit=True)
```

**Figura 10:** Captura de Asignación de Tokens para Acceso a la API de Twitter

A continuación, determinamos los criterios de búsqueda, para este caso de estudio hemos decidido utilizar el nombre y apellido de cada candidato como el término a buscar

en los diferentes tweets, tomamos una muestra de 3000 tweets por cada candidato, cada tweet registrado es añadido a una lista propia del candidato.

```
searchTerm = 'Guillermo Lasso'  
numBusqueda = 3000  
tweets_Lasso = list(tweepy.Cursor(api.search, q=searchTerm).items(numBusqueda))
```

**Figura 11:** Captura de la Lista de Tweets

### **Eta de Procedimiento: Análisis de Sentimientos**

Una vez distribuidos los diferentes tweets en las respectivas listas por cada candidato, procedemos a guardar el número total de tweets registrados en sus respectivas variables como se muestra a continuación.

```
numBusqueda_Almeida = len(tweets_Almeida)  
numBusqueda_Andrade = len(tweets_Andrade)  
numBusqueda_Arauz = len(tweets_Arauz)  
numBusqueda_Carrasco = len(tweets_Carrasco)
```

**Figura 12:** Captura del Conteo del Tweets

Para el siguiente paso del procesamiento de los datos, se utilizará la librería de TextBlob de Python que es el encargado de tratar y estimar el valor de sentimiento de cada tweet obteniendo una métrica de **polaridad** y **subjetividad**. La polaridad es el sentimiento mismo, que va de -1 a +1, donde 1 representa el sentimiento positivo, 0 el sentimiento neutral y -1 el sentimiento negativo que expresa el tweet. La subjetividad que hace referencia al juicio u opinión (objetivo o subjetivo) que contiene ese tweet y va desde los rangos de 0 a 1. Para este caso de estudio se decidió calcular el nivel de aceptación de cada candidato, para ello se creó la siguiente función, donde se toma cada comentario registrado y se analiza según las palabras que contenga el mismo, para el caso de la polaridad la librería TextBlob relaciona algunas de estas palabras como positivas: bueno, excelente, inspirador, honesto, transparente, capacitado, competente, solidario, generoso, servicial. Y como palabras negativas: corrupto, corrupción, robo, ladrón, inseguridad, mentiroso, injusticia; a lo cual dará un porcentaje para el cálculo de la polaridad.

```

def calcular_apoyo(tw=[], numBusqueda=0):
    positive = 0
    negative = 0
    neutral = 0
    polarity = 0

    for tweet in tw:
        analysis = TextBlob (tweet.text)
        polarity = analysis.sentiment.polarity

        if (analysis.sentiment.polarity == 0):
            neutral += 1
            neutral += porcentaje(neutral, numBusqueda)
        elif (analysis.sentiment.polarity < 0):
            negative += 1
            negative += porcentaje(negative, numBusqueda)
        else:
            positive += 1
            positive += porcentaje(positive, numBusqueda)

    return [positive, negative, neutral, polarity]

```

**Figura 13:** Captura de la Función del Cálculo de Polaridad

Aquí llamamos la función calcular apoyo donde nos retornara el nivel aceptación que poseen cada uno de los candidatos, la cantidad de comentarios negativos, la cantidad de comentarios neutrales y la polaridad del ultimo tweet por cada candidato.

```

resultado_Almeida = calcular_apoyo(tweets_Almeida)
resultado_Andrade = calcular_apoyo(tweets_Andrade)
resultado_Arauz = calcular_apoyo(tweets_Arauz)
resultado_Carrasco = calcular_apoyo(tweets_Carrasco)
resultado_Celi = calcular_apoyo(tweets_Celi)

```

**Figura 14:** Captura de la Función que Retorna Resultados de Polaridad

### **Eta de Presentación: Graficas de Resultados**

Para presentar los resultados del análisis de sentimientos, se necesita de la ayuda de las librerías Matplotlib, que posee una gran variedad de figuras para hacer un gráfico de los resultados, a continuación, se presenta la función que utilizamos para graficar la aceptación de cada candidato con la muestra de las 5 provincias más pobladas del Ecuador.

```

def graficas_individuales(positive=0, negative=0, neutral=0, polarity=0, candidato=''):

    if negative+positive ==0:
        positive+=1
    labels = [
        'positive: '+str(round(100*positive/(negative+positive), 3))+ '% ]',
        'negative: '+str(round(100*negative/(negative+positive), 3))+ '% ]'
    ]

    sizes = [
        positive/(negative+positive),
        negative/(negative+positive)
    ]

    colors = ['blue', 'red']
    patches, texts = pyplot.pie(sizes, colors=colors, startangle=90)
    pyplot.legend(patches, labels, loc="best")
    pyplot.title("ACEPTACION DEL CANDIDATO: " + candidato)
    pyplot.axis('equal')
    pyplot.tight_layout()
    pyplot.show()

```

**Figura 15:** Captura de la Función para Graficar Resultados

Como se puede observar para este caso de estudio decidimos omitir la cantidad de comentarios neutrales debido a no aportan con el cálculo de apoyo, así que todos estos comentarios son considerados como un margen de error y por tal motivo no son considerados en la gráfica.

Para la gráfica final tomamos el nivel de apoyo que posee cada candidato y lo transformamos a puntaje de tal forma que nos permita ilustrar de manera comparativa los diferentes puntajes de cada candidato.

```

def grafica_general(candidatos=[], puntajes=[]):

    puntajes_normalizados=[0.0 for _ in range(len(puntajes))]

    for i in range(len(puntajes)):
        puntajes_normalizados[i] = round(puntajes[i]/sum(puntajes), 2)

    pyplot.bar(candidatos, puntajes_normalizados)
    pyplot.title('Apoyo a los candidatos en Twitter')
    pyplot.show()

```

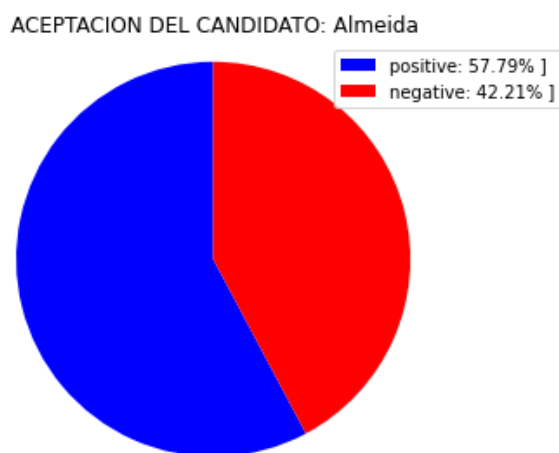
**Figura 16:** Captura de la Función de Cálculo de Porcentajes

## 3.2. Resultados y discusión del caso de estudio

### 3.2.1. Resultados

A continuación, presentamos parte de las gráficas obtenidas donde se ilustra el nivel de aprobación individual de cada candidato de la muestra tomada de las 5 provincias más pobladas del Ecuador, en el periodo señalado y con la cantidad de tweets analizados.

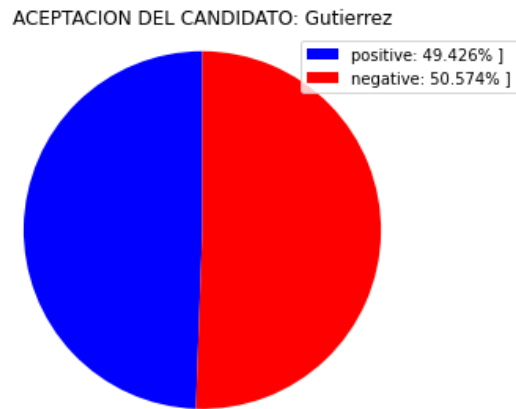
**Figura 17:** Nivel de Aceptación del Candidato Gerson Almeida



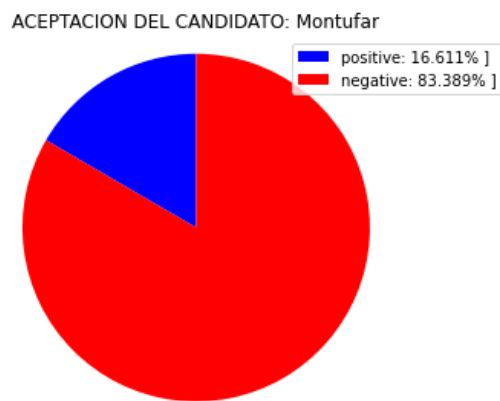
**Figura 18:** Nivel de Aceptación del Candidato Guillermo Celi



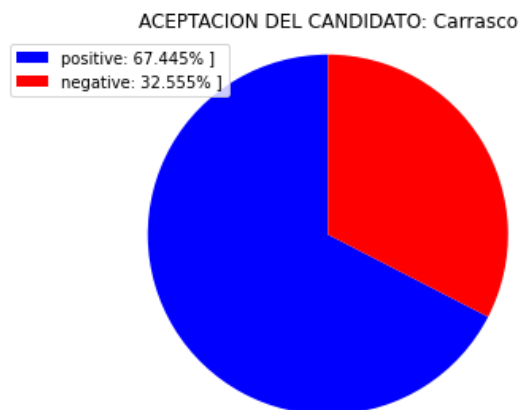
**Figura 19:** Nivel de Aceptación del Candidato Lucio Gutiérrez



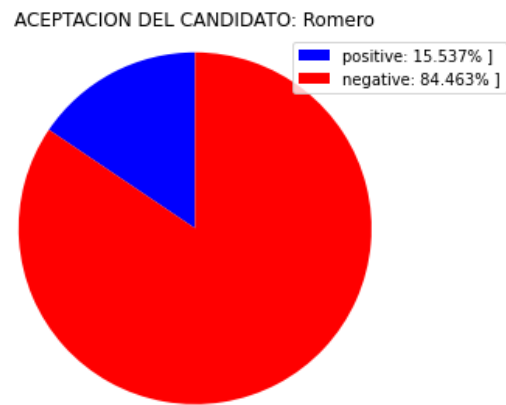
**Figura 20:** Nivel de Aceptación del Candidato Cesar Montufar



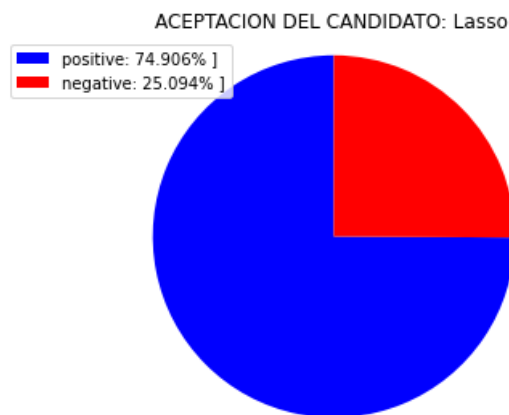
**Figura 21:** Nivel de Aceptación del Candidato Paul Carrasco



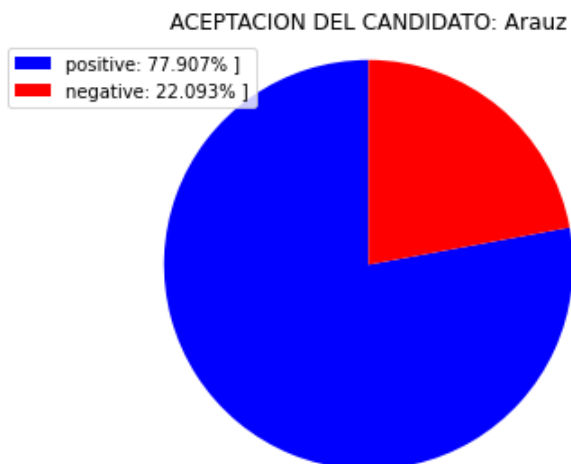
**Figura 22:** Nivel de Aceptación del Candidato Isidro Romero



**Figura 23:** Nivel de Aceptación del Candidato Guillermo Lasso



**Figura 24:** Nivel de Aceptación del Candidato Andrés Arauz

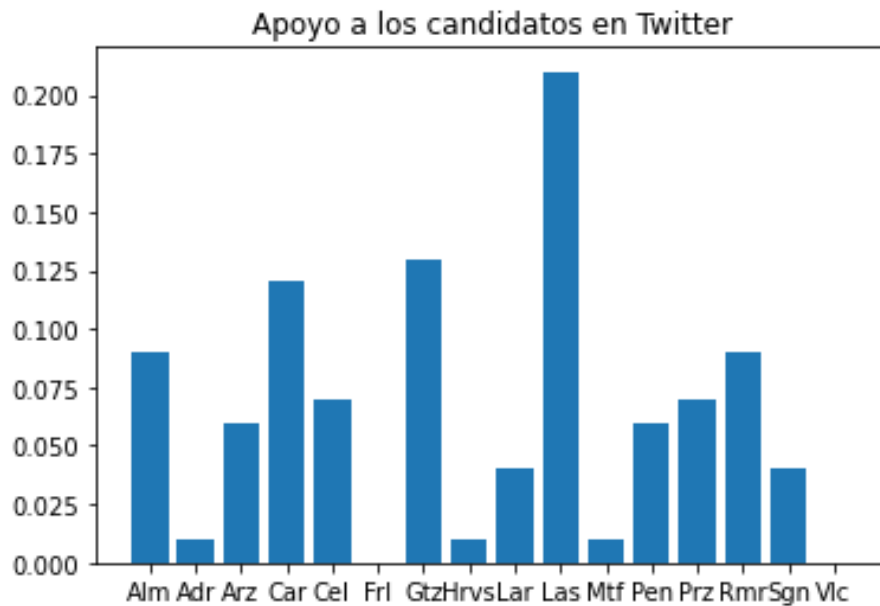




A continuación, presentamos la gráfica final donde se muestra que el candidato con mayor aceptación de las 5 provincias más pobladas de Ecuador es el candidato por el partido CREO, Guillermo Lasso.

**Figura 25:** Nivel de Apoyo de Usuarios Twitter a Candidatos Electorales

Donde las siglas de los candidatos son: Gerson Almeida (Alm), Giovanni Andrade (Adr), Andrés Arauz (Arz), Paul Carrasco (Car), Guillermo Celi (Cel), Pedro Freile (Frl), Lucio Gutiérrez (Gtz), Xavier Hervas (Hrvs), Gustavo Larrea (Lar), Guillermo Lasso (Las), Cesar Montufar (Mtf), Ximena Peña (Pen), Yaku Pérez (Prz), Isidro Romero (Rmr), Carlos



Sagñay (Sgn) y Juan Fernando Velasco (Vlc).

Para concluir con el análisis de sentimientos de los tweets extraídos debemos tener en cuenta los siguientes criterios:

- Si bien es cierto que el candidato con mayor aceptación en el ejercicio fue Guillermo Lasso, se debe tener en cuenta que influye el nivel de popularidad del candidato a la hora de la presentación de resultados.
- El estudio es netamente dirigido a la población que posee una cuenta en Twitter.
- El estudio se realizó a menos de dos meses de la elección primaria y antes del debate presidencial.

- La muestra tomada fue equitativamente tratada.
- La calidad de inferencia se mejoró dando porcentajes de cada provincia según su número de habitantes, además de normalizar la escala mediante la extracción del margen de error que en este caso fueron los comentarios neutrales.

Por tanto, se puede concluir que el análisis de sentimientos realizado da una idea cercana de la tendencia del posible ganador de las elecciones del 2021, aunque para mayor precisión los periodos de tiempo, la cantidad de tweets, el número de provincias y otros factores adicionales deben considerarse.

### **3.2.2. Discusión**

Es claro que los beneficios en tiempo y recursos al utilizar herramientas de data analytics como el análisis de sentimientos son enormes, como se describen en la tabla comparativa resultante del análisis, en donde se menciona que Data Analytics se presenta como rápido, y entre más rápida sea la analítica, entonces la lógica se dice mejor y más útil. [31] Desde el momento de la extracción de información desde la red social Twitter, se evidencia la rapidez con la que se obtuvo los datos necesarios para comenzar con el estudio utilizando esta herramienta de Data Analytics, como algoritmo de minería la biblioteca de Python (Tweepy) [26], esto con otro tipo de análisis como el estadístico nos hubiera llevado varias semanas o incluso meses el terminar con la recolección de información.

Otro de los beneficios que se comprueban acerca de utilizar Data Analytics para esta etapa del estudio es el lugar, como se menciona en la tabla comparativa, Data Analytics se presenta como casi omnipotente. El enfoque está en la capacidad de reunir diferentes flujos de datos para aumentar la cobertura de los análisis [31]. Por lo tanto, reduce los costos, optimiza el tiempo y agiliza los resultados, evitando la ejecución de encuestas físicas empleando cuestionarios y entrevistas.

Adicionalmente, cuando se toman decisiones más precisas, se alivia la posibilidad de incurrir en pérdidas. Por lo tanto, con el uso correcto de la analítica, las empresas emergentes y los emprendedores pueden tener éxito en la reducción de sus costos operativos [21].

En sí, el proceso completo de este caso de estudio es un análisis predictivo, que utiliza una herramienta Data Analytics como es el análisis de sentimientos que intenta clasificar las opiniones y declaraciones como positivas, negativas o neutrales, en donde se

comprende el comportamiento humano y se categoriza las opiniones expresadas, utilizando el procesamiento del lenguaje natural (PNL) para clasificar e interpretar los datos disponibles en línea y la utilización una combinación de minería de datos y PNL para explorar, extraer y purificar los abundantes datos que se encuentran en la red social Twitter para que se adapten al propósito en cuestión [28].

Este proceso termina con un análisis de interpretación que se define como aquellas acciones que permiten al investigador dar significado a las actividades narrativas y de codificación que facilitan la comprensión del investigador del marco conceptual generado a través del proceso de codificación [7].

Por lo cual terminamos interpretando el posible resultado del candidato ganador según el análisis de sentimientos realizado mediante un proceso de codificación, el cual permitió extraer la polaridad y la subjetividad de cada comentario y así distribuir los resultados negativos y positivos de cada tweet, con la finalidad de ilustrar de manera colectiva cada uno de los resultados de los diferentes candidatos y mostrar como ganador el candidato con mayor porcentaje de comentarios positivos.

### **3.3. Conclusiones del Capítulo**

En este capítulo se presenta la metodología utilizada para el desarrollo de la aplicación que consta de 2 fases importantes: 1. Proyección del caso de estudio y 2. Resultados; dentro de esta primera fase se describió el contexto, herramientas y método para desarrollo del caso de estudio, este último se pudo subdividir en tres subetapas principales como son: Etapa Inicial: Extracción de información, Etapa de procesamiento: Análisis de Sentimientos y la Etapa final: Presentación de graficas.

En la segunda fase (Resultados) se pudo mostrar los resultados del análisis de sentimientos realizado y mostrar una discusión sobre los beneficios del uso de Data Análisis y Data Analytics dentro del caso de estudio.

A continuación, en el capítulo 4 se procede a exponer las conclusiones obtenidas al culminar con este proyecto de investigación. Así mismo, se proponen las recomendaciones pertinentes.

## **CAPITULO 4**

### **4. CONCLUSIONES Y RECOMENDACIONES**

#### **4.1. CONCLUSIONES**

Una vez realizada la investigación en las diferentes fuentes de información científicas e implementado el experimento respectivo podemos concluir que:

- La diferencia más notable entre los términos de estudio es que Data Analysis se enfoca en el tratamiento de datos históricos, mientras que Data Analytics toma estos mismos datos y se enfoca en lo que se puede hacer con ellos en el futuro, convirtiéndolos en información útil y que aportará en una mejor toma de decisiones para obtener ventajas competitivas en una organización.
- En base al análisis de los 36 artículos académicos del total de los 3670 investigados, todos elaborados por expertos en las áreas de estudio, se logró determinar las marcadas diferencias entre los términos de análisis y analítica de datos mediante la comparación de los conceptos, las metodologías, técnicas, herramientas, procesos, beneficios y limitaciones.
- La elaboración de los marcos conceptuales y la tabla comparativa, permitieron agrupar en un solo documento un macro resumen científico de toda la información redactada por varios autores que plasmaron en los 36 artículos como producto de sus investigaciones previas, para lo cual se empleó criterios de inclusión/exclusión relacionados a los términos de estudio, con lo que se espera que dicho documento sirva como una fuente de consulta sólida y veraz para futuras investigaciones en el campo de estudio.
- La ejecución del caso de estudio, en la cual se incorporan procesos de análisis de sentimientos, permite ejemplificar más claramente la diferencia de la aplicabilidad de estos dos términos y presenta una serie de beneficios al utilizarlos correctamente.

## 4.2. RECOMENDACIONES

- Para la utilización de herramientas de Data Analytics y Data Analysis es preciso identificar la finalidad a esperarse en el proceso y la evaluación de la aplicabilidad de uso de la herramienta, para lo cual es necesario saber el segmento de usuarios al que se dirige el análisis, el tiempo necesario para efectuar un análisis de calidad, los recursos tecnológicos disponibles y la capacidad de incorporar nuevos procedimientos y hasta conocimientos ante cualquier cambio.
- En el momento de ejecutar el proceso de la búsqueda de artículos académicos, es preciso una revisión rápida para verificar que existan los artículos suficientes a la temática de estudio, para evitar incurrir en búsquedas académicas adicionales, también se recomienda seleccionar adecuadamente las fuentes de información.
- Para que el caso de estudio presente resultados más eficaces y que permitan obtener cifras más reales acerca de los niveles de aceptación de los candidatos, es necesario efectuar seguimientos periódicos y más completos, es decir ejecutar el mismo procedimiento, pero incluyendo factores adicionales como la ampliación de los lapsos de tiempo analizados, incorporar un mayor número de fuentes, aumentar la cantidad de usuarios y evaluar otros factores relevantes.
- Data Analysis, en general, se puede recomendar para encontrar correlaciones anónimas, preferencias de los clientes, tendencias del mercado y otra información necesaria basada en hechos históricos que puede ayudar a tomar decisiones fundamentales con fines comerciales. Por otro lado, Las aplicaciones de Data Analytics en cambio se enfoca en nuevas herramientas tecnológicas que permiten realizar análisis del comportamiento del cliente, servicios de valor agregado para los clientes, mantener una ventaja competitiva, buen gobierno y sistemas de soporte, atención médica preventiva, predicción y pronóstico.

## 5. REFERENCIAS BIBLIOGRÁFICAS

- [1] M. R. Hussain, Q. N. Naveed, N. Khan, A. Naim, N. Ahmad y S. Qamar, «The 51 V's Of Big Data: Survey, Technologies, Characteristics, Opportunities, Issues and Challenges,» de *COINS '19: Proceedings of the International Conference on Omni-Layer Intelligent Systems*, May 2019.
- [2] T. Brown, «IT Chronicles,» 2019. [En línea]. Available: <https://itchronicles.com/big-data/data-analytics-vs-data-analysis-whats-the-difference/>. [Último acceso: 17 August 2020].
- [3] «GetSmarter,» 23 January 2020. [En línea]. Available: <https://www.getsmarter.com/blog/career-advice/difference-data-analytics-data-analysis/#:~:text=Essentially%2C%20the%20primary%20difference%20between,data%20analysis%20is%20a%20subcomponent.&text=It's%20the%20role%20of%20the,data%20into%20information%20tha>. [Último acceso: 9 September 2020].
- [4] «EDUCBA,» [En línea]. Available: <https://www.educba.com/data-analytics-vs-data-analysis/>. [Último acceso: 9 September 2020].
- [5] O. Akinawonu, «LOGINWORKS,» 27 June 2018. [En línea]. Available: <https://www.loginworks.com/blogs/top-10-small-differences-between-data-analyticsdata-analysis-and-data-mining/>. [Último acceso: 9 September 2020].
- [6] «IT USER,» 15 January 2019. [En línea]. Available: <https://discoverthenew.ituser.es/predictive-analytics/2019/01/analitica-de-datos-no-estructurados-por-que-es-tan-importante>. [Último acceso: 9 September 2020].
- [7] B. Kawulich, « Data Analysis Techniques in Qualitative Research,» *Journal of Research in Education* , p. 113, 2004.
- [8] Runkler, *Data analytics: Models and algorithms for intelligent data analysis*, 2012.
- [9] F. Coelho, «Significados,» 17 Mayo 2019. [En línea]. Available: <https://www.significados.com/metodologia/> . [Último acceso: 2 Diciembre 2020].
- [10] M. Porporatto, «Que Significado,» [En línea]. Available: <https://quesignificado.com/proceso/>. [Último acceso: 3 Diciembre 2020].

- [11] «Significados,» 02 febrero 2018. [En línea]. Available: <https://www.significados.com/metodo/> . [Último acceso: 3 Diciembre 2020].
- [12] «Significados,» 24 03 2017. [En línea]. Available: <https://www.significados.com/tecnica/> . [Último acceso: 3 Diciembre 2020].
- [13] «SignificadosSignificados,» 07 Junio 2017. [En línea]. Available: <https://www.significados.com/beneficio/> . [Último acceso: 3 Diciembre 2020].
- [14] J. Navarro, «Definición ABC,» 3 mayo 2014. [En línea]. Available: <https://www.definicionabc.com/deporte/limitacion.php>. [Último acceso: 3 Diciembre 2020].
- [15] S. M. K. A. S. K. P. Ram Mohan Rao, «A Case Study on Privacy Threats and Research,» *International Conference of Electronics, Communication and Aerospace Technology (ICECA)*., 2017.
- [16] P. Pirzadeh, M. Carey y T. Westmann, «A Performance Study of Big Data Analytics Platforms,» *IEEE International Conference on Big Data (BIGDATA)*, 2017. [En línea]. Available: [sci-hub.se/10.1109/BigData.2017.8258260](https://sci-hub.se/10.1109/BigData.2017.8258260).
- [17] D. Kozjek, R. Vrabič, B. Rihtaršič, N. Lavrač y P. Butala, «Advancing manufacturing systems with big-data analytics: A conceptual framework,» *International Journal of Computer Integrated*, 2020. [En línea]. Available: [sci-hub.se/10.1080/0951192X.2020.1718765](https://sci-hub.se/10.1080/0951192X.2020.1718765).
- [18] N. Shehab, MBadawy y H. Arafat, «Big Data Analytics Concepts, Technologies Challenges, and Opportunities,» *5th International Conference on Advanced Intelligent Systems and Informatics*, October 2019. [En línea]. Available: [https://www.scopus.com/inward/record.uri?eid=2-s2.0-85075594326&doi=10.1007%2f978-3-030-31129-2\\_9&partnerID=40&md5=5154f5cf3b28cbd316c4fb1e02e1c604](https://www.scopus.com/inward/record.uri?eid=2-s2.0-85075594326&doi=10.1007%2f978-3-030-31129-2_9&partnerID=40&md5=5154f5cf3b28cbd316c4fb1e02e1c604).
- [19] S. D. Sagiroglu S, «Big Data: A Review,» de *2013 International Conference on Collaboration Technologies and Systems (CTS)*, 2013.
- [20] M. M. d. Medeiros, N. Hoppen y A. C. G. Maçada, «Data science for business: benefits, challenges and opportunities,» *Emerald Insight*, 2020. [En línea]. Available: [sci-hub.se/10.1108/bl-12-2019-0132](https://sci-hub.se/10.1108/bl-12-2019-0132).

- [21] S. Sedkaoui, «How data analytics is changing entrepreneurial opportunities?,» *International Journal of Innovation Science*, 12 mayo 2018. [En línea]. Available: [sci-hub.se/10.1108/IJIS-09-2017-0092](https://sci-hub.se/10.1108/IJIS-09-2017-0092).
- [22] D. Chong y H. Shi, «Big data analytics: a literature review,» *Journal of Management Analytics*, 2015. [En línea]. Available: [sci-hub.se/10.1080/23270012.2015.1082449](https://sci-hub.se/10.1080/23270012.2015.1082449).
- [23] J. W. Tukey, *The Future of Data Analysis*, Institute of Mathematical Statistics.
- [24] S. R. Dharmapal y K. Sikamani, «Big Data Analytics Using Agile Model,» *International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*, 2016. [En línea]. Available: [sci-hub.se/10.1109/iceeot.2016.7754854](https://sci-hub.se/10.1109/iceeot.2016.7754854).
- [25] B. Yadranjiaghdam, N. Pool y N. Tabrizi, «A Survey on Real-time Big Data Analytics: Applications and Tools,» *International Conference on Computational Science and Computational Intelligence*, 2016. [En línea]. Available: [sci-hub.se/10.1109/csci.2016.0083](https://sci-hub.se/10.1109/csci.2016.0083).
- [26] M. N. Mouchili, S. Aljawarneh y W. Tchouati, «Smart city data analysis,» *ACM International Conference Proceeding Series*, 1 October 2018. [En línea]. Available: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85058177143&doi=10.1145%2f3279996.3280029&partnerID=40&md5=681537215ffe4d2fa440ad2683bc5b6>.
- [27] B. Bilalli, A. Abelló, T. Aluja-Banet y R. Wrembel, «Towards intelligent data analysis: The metadata challenge,» *Proceedings of the International Conference on Internet of Things and Big Data*, 2016. [En línea]. Available: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84979524709&partnerID=40&md5=1fc860f4e6147296d72e60ee0c0d60cb>.
- [28] G. Moorjani y L. Sadath, «Sentiment analysis-A tool for data mining in big data analytics,» *International Journal of Innovative Technology and Exploring Engineering*, July 2019. [En línea]. Available: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85084725597&doi=10.35940%2fijitee.i8005.078919&partnerID=40&md5=0a629f11739bb702c59531640b387723>.
- [29] M. Berthold y D. J. Hand, *Intelligent Data Analysis: An Introduction*, Heidelberg,



Germany: Springer, 2007.

- [30] P. Chandarana y M. Vijayalakshmi, «Big Data Analytics Frameworks,» *International Conference on Circuits, Systems, Communication and Information Technology Applications (CSCITA)*, 2014. [En línea]. Available: [sci-hub.se/10.1109/CSCITA.2014.6839299](https://doi.org/10.1109/CSCITA.2014.6839299).
- [31] D. Beer, «Envisioning the power of data analytics,» *Information, Communication & Society*, 16 February 2017. [En línea]. Available: [sci-hub.se/10.1080/1369118x.2017.1289232](https://doi.org/10.1080/1369118x.2017.1289232).
- [32] M. Zhou, M. Cao y J.-H. P. Taeho Park\*, «Clarifying Big Data: The Concept and Its Applications,» *International Conference on Big Data Applications and Services - BigDAS*, 2015. [En línea]. Available: [sci-hub.se/10.1145/2837060.2837068](https://doi.org/10.1145/2837060.2837068).
- [33] A. Anaya, W. Henning, N. Basantkumar y J. Oliver, «Yield Improvement Using Advanced Data Analytics,» *30th Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC)*, 2019. [En línea]. Available: [sci-hub.se/10.1109/ASMC.2019.8791752](https://doi.org/10.1109/ASMC.2019.8791752).
- [34] A. Dauda, S. Mclean, A. Almeahadi y K. El-Khatib, «Big Data Analytics Architecture for Security Intelligence,» *Proceedings of the 11th International Conference on Security of Information and Networks*, 2018. [En línea]. Available: <https://doi.org/10.1145/3264437.3264474>.
- [35] M. Ahmadi, P. Dileepan y K. K. Wheatley, «A SWOT analysis of big data,» *Journal of Education for Business*, 2016. [En línea]. Available: [sci-hub.se/10.1080/08832323.2016.1181045](https://doi.org/10.1080/08832323.2016.1181045).
- [36] S. R. Dharmapal y K. Sikamani, «Big Data Analytics Using Agile Model,» *International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*, 2016. [En línea]. Available: [sci-hub.se/10.1109/iceeot.2016.7754854](https://doi.org/10.1109/iceeot.2016.7754854).
- [37] «INEC,» [En línea]. Available: <https://www.ecuadorencifras.gob.ec/censo-de-poblacion-y-vivienda/>. [Último acceso: 17 Diciembre 2020].
- [38] «Lenovo,» [En línea]. Available: <https://www.lenovo.com/ec/es/faqs/pc-vida-faqs/que-es-twitter/>. [Último acceso: 9 Febrero 2021].

- [39] Á. Robledano, «Openwebinars,» 23 Septiembre 2019. [En línea]. Available: <https://openwebinars.net/blog/que-es-python/>. [Último acceso: 9 Febrero 2021].
- [40] F. G. López, «El Observatorio de Big Data & Analytics.,» [En línea]. Available: <http://spaceanalytics.blogspot.com/2018/04/tweepy-python.html#:~:text=Tweepy%3A%20es%20una%20librer%C3%ADa%20Python,int%20eractivos%20en%20todas%20las%20plataformas..> [Último acceso: 9 Febrero 2021].