

**UNIVERSIDAD INTERNACIONAL SEK**

**FACULTAD DE ARQUITECTURA E INGENIERÍAS**

**TRABAJO DE FIN DE CARRERA TITULADO:**

DESARROLLO DE UNA HERRAMIENTA DE MINERÍA DE DATOS  
PARA EL ANÁLISIS DE INFLUENCIA DE CUENTAS AUTOMATIZADAS  
EN TEMAS DE TENDENCIA SOBRE LA OPINIÓN DE LOS USUARIOS  
DE TWITTER EN ECUADOR

**Realizado por:**

MARIO EMILIANO GORDON PICO

**Director del proyecto:**

DANIEL ANDRÉS RIOFRÍO ALMEIDA, Ph.D

Como requisito para la obtención del título de:

MÁSTER EN TECNOLOGÍAS DE LA INFORMACIÓN

QUITO, JUNIO DEL 2018

## **DECLARACIÓN JURAMENTADA**

Yo, MARIO EMILIANO GORDON PICO, con cédula de identidad # 172270539-7, declaro bajo juramento que el trabajo aquí desarrollado es de mi autoría, que no ha sido previamente presentado para ningún grado a 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 UNIVERSIDAD INTERNACIONAL SEK, según lo establecido por la Ley de Propiedad Intelectual, por su reglamento y por la normativa institucional vigente.

Mario Emiliano Gordon Pico

**CC: 1722705397**

## **DECLARATORIA**

El presente trabajo de investigación titulado:

**“DESARROLLO DE UNA HERRAMIENTA DE MINERÍA DE DATOS  
PARA EL ANÁLISIS DE INFLUENCIA DE CUENTAS  
AUTOMATIZADAS EN TEMAS DE TENDENCIA SOBRE LA  
OPINIÓN DE LOS USUARIOS DE TWITTER EN ECUADOR”**

Realizado por:

**MARIO EMILIANO GORDON PICO**

como Requisito para la Obtención del Título de:  
**MÁSTER EN TECNOLOGÍAS DE LA INFORMACIÓN**

ha sido dirigido por el profesor:

**DANIEL ANDRÉS RIOFRÍO ALMEIDA, Ph.D**

quien considera que constituye un trabajo original de su autor

**Daniel Andrés Riofrío Almeida, Ph.D**

**DIRECTOR**

## **LOS PROFESORES INFORMANTES**

Los Profesores Informantes:

ING. RODRIGUEZ ARBOLEDA VERÓNICA ELIZABETH, MBA.

SEGURA MORALES MARCO ANTONIO, Ph.D.

Después de revisar el trabajo presentado,

lo han calificado como apto para su defensa oral  
ante el tribunal examinador

ING. RODRIGUEZ ARBOLEDA VERÓNICA ELIZABETH, MBA.

SEGURA MORALES MARCO ANTONIO, Ph.D.

Quito, Junio del 2018

## **DEDICATORIA**

Dedico el presente trabajo de investigación a Dios, a mis padres quienes supieron inculcarme valores, principios y dedicación que han guiado mi vida, a Estefanía Alexandra Zúñiga que con su apoyo, palabras de ánimo y amor fue clave para la culminación de este trabajo de investigación. Gracias por estar siempre junto a mí.

## **AGRADECIMIENTOS**

A mis padres, hermanas y novia Estefanía Zúñiga por su apoyo y respaldo en el transcurso de los estudios de maestría y el desarrollo del presente trabajo.

A Daniel Riofrío por su acertada dirección de la tesis. Su profesionalismo y entrega fueron determinantes a la hora de conformar este documento.

A los profesores, quienes con sus lecturas aportaron una visión diferente e integradora a mi trabajo de investigación.

A la Universidad Internacional SEK, por su esfuerzo de formar profesionales íntegros.

# ÍNDICE GENERAL DE CONTENIDOS

CAPÍTULO I. INTRODUCCIÓN	1
1.1 EL PROBLEMA DE INVESTIGACIÓN	1
1.1.1 Planteamiento del Problema	1
1.1.1.1 Diagnóstico del problema	3
1.1.1.2 Pronóstico	4
1.1.1.3 Control del Pronóstico	4
1.1.2 Formulación de problema	5
1.1.3 Sistematización del problema	5
1.1.4 Objetivo General	5
1.1.5 Objetivos Específicos	5
1.1.6 Justificación	6
1.2 MARCO TEÓRICO	7
1.2.1 Estado actual del conocimiento sobre el tema	7
1.2.2 Adopción de una perspectiva teórica	9
1.2.3 Marco conceptual	10
Minería de datos	10
Aprendizaje de Máquina ( <i>Machine Learning</i> )	10
Técnicas de aprendizaje de máquina supervisado para clasificación	11
Clasificación multi-etiqueta	12
Algoritmo <i>Naïve Bayes</i>	12
Algoritmo <i>Multinomial Naïve Bayes</i> (MNB)	12
Algoritmo <i>Support Vector Machines</i> (SVM)	13
Algoritmo <i>Logistic Regression</i>	13
Algoritmo <i>K-Nearest Neighbor</i> (KNN)	14
Detección de cuentas automatizadas	14
1.2.4 Hipótesis	16
CAPÍTULO II. MÉTODO	17
2.1 Tipo de estudio	17
2.2 Modalidad de investigación	17
2.3 Método	18
2.4 Población y Muestra.	18
2.5 Selección de instrumentos de Investigación	19
2.6 Validez y Confiabilidad de los Instrumentos	19
CAPÍTULO III. DISEÑO Y DESARROLLO DEL PROTOTIPO	20
	VII

3.1 Funcionalidad	20
3.2 Diseño de la arquitectura tecnológica	20
3.2.1 Infraestructura de desarrollo	20
3.2.2 Descripción de hardware y software	20
3.2.3 Configuración del entorno de desarrollo	22
3.2.4 Desarrollo de los módulos de la herramienta	25
Módulo de adquisición de datos de Twitter	25
Módulo de almacenamiento de datos	26
Módulo de procesamiento de datos	30
Módulo de análisis de cuentas automatizadas	32
Módulo de análisis de datos	33
Módulo de presentación de datos	33
CAPÍTULO IV. RESULTADOS	34
4.1 Levantamiento de datos	34
4.1.1 Resultados de la clasificación de los datos	34
4.1.2 Resultados del análisis de cuentas automatizadas	38
4.2 Presentación y análisis de resultados	39
4.2.1 Resultados del entrenamiento del algoritmo de clasificación	39
4.2.2. Resultados de clasificación e índices de influencia	40
4.3 Aplicación práctica	52
CAPÍTULO V. DISCUSIÓN	53
5.1 Conclusiones	53
5.2 Recomendaciones y trabajo futuro	55
LISTA DE REFERENCIAS	57
ANEXOS	61

## ÍNDICE DE FIGURAS

<b>Figura 1.</b> Interfaz Google Cloud para creación de máquinas virtuales.....	22
<b>Figura 2.</b> Elección de parámetros de la máquina virtual.....	23
<b>Figura 3.</b> Permisos de conexiones de red de la máquina virtual.....	24
<b>Figura 4.</b> Selección de la máquina virtual en el dashboard de Google Cloud. ....	25
<b>Figura 5.</b> Arquitectura de la herramienta de minería de datos.....	25
<b>Figura 6.</b> Diseño de base de datos en MongoDB.....	28
<b>Figura 7.</b> a) Tweet de enviado por la cuenta @eloritocom. b) Tweet de enviado por la cuenta @vitotvo.....	39
<b>Figura 8.</b> Página de inicio de la interfaz gráfica de la herramienta de minería de texto.....	40
<b>Figura 9.</b> Valores de índice de influencia última hora y promedio en política.....	41
<b>Figura 10.</b> Valores de índice de influencia última hora y promedio en economía. ....	41
<b>Figura 11.</b> Valores de índice de influencia última hora y promedio en deportes. ....	42
<b>Figura 12.</b> Valores de índice de influencia última hora y promedio en otros.....	43
<b>Figura 13.</b> Valores de índice de influencia última hora y promedio en entretenimiento.....	43
<b>Figura 14.</b> Página de número de tweets por cada hora. ....	44
<b>Figura 15.</b> Número de Tweets de política.....	45
<b>Figura 16.</b> Número de tweets otros.....	46
<b>Figura 17.</b> Número de tweets de deportes.....	48
<b>Figura 18.</b> Número de tweets de economía.....	49
<b>Figura 19.</b> Número de Tweets de entretenimiento.....	51

## ÍNDICE DE TABLAS

<b>Tabla 1.</b> Detalle de software base .....	22
<b>Tabla 2.</b> Descripción de las colecciones de la base de datos Twitterdb.....	29
<b>Tabla 3.</b> <i>Scores</i> de entrenamiento de los algoritmos de aprendizaje supervisado. ....	32
<b>Tabla 4.</b> Resultados de las pruebas de clasificación con LSVM. ....	35
<b>Tabla 5.</b> Resultados de las pruebas de clasificación con Multinomial NB. ....	35
<b>Tabla 6.</b> Resultados de las pruebas de clasificación con Logistic Regression.....	36
<b>Tabla 7.</b> <i>Scores</i> de entrenamiento de los 3 algoritmos de mejor rendimiento. ....	36
<b>Tabla 8.</b> Resultados pruebas de clasificación con Multinomial NB. ....	37
<b>Tabla 9.</b> Resultados pruebas de clasificación con Logistic Regression.....	37
<b>Tabla 10.</b> Resultados pruebas de clasificación con LSVM.....	37
<b>Tabla 11.</b> Valores de tweets e índices de influencia por hora (política). ....	45
<b>Tabla 12.</b> Valores de tweets e índices de influencia por hora (otros). ....	47
<b>Tabla 13.</b> Valores de tweets e índices de influencia por hora (deportes).....	48
<b>Tabla 14.</b> Valores de tweets e índices de influencia por hora (economía).....	50
<b>Tabla 15.</b> Valores de tweets e índices de influencia por hora (entretenimiento). ....	51

## ÍNDICE DE ECUACIONES

( 1 ) Ecuación Naïve Bayes.....	12
( 2 ) Ecuación Multinomial Naïve Bayes.....	12
( 3 ) Ecuación Logistic Regression .....	12
( 4 ) Ecuación de índices de influencia .....	32

## RESUMEN

El crecimiento de las redes sociales y su popularidad en el mundo entero han sido razones para ser utilizadas con el fin de influenciar a los usuarios mediante contenido viral, noticias falsas o mensajes de interés político con el uso de *bots*. Estas técnicas han ido tomando popularidad para varias campañas en los últimos años, siendo una importante herramienta de difusión de información.

En el presente trabajo se desarrolló una herramienta para la clasificación y análisis de tweets con el uso de técnicas de *Machine Learning* para mensajes en español, utilizando información de portales de noticias locales para el entrenamiento de varios modelos. Durante el procesamiento de datos se encontró que el algoritmo Multinomial Naïve Bayes tiene los mejores resultados de clasificación. Por otro lado, con el uso del algoritmo para detección de social *bots* de DeBot se almacenaron las cuentas automatizadas para un posterior análisis del índice de influencia de estas cuentas en 5 categorías: política, economía, deportes, entretenimiento y otros. Los resultados obtenidos reflejan la influencia de *bots* en las 5 categorías seleccionadas, cuyos niveles varían dependiendo de las horas del día o las noticias de interés que se susciten.

El prototipo desarrollado permite el acceso a información condensada de la presencia e influencia de cuentas automatizadas en Twitter de los eventos cotidianos del país. Adicionalmente, abre la puerta para que se realicen más estudios sobre estas plataformas en los que se beneficie la comunidad.

Palabras clave: Machine learning, cuentas automatizadas, minería de datos.

## ABSTRACT

The growth of social media and its popularity around the world are the main reasons they have been used to influence its users through viral content, fake news or messages of public interest using bots. These techniques have gained popularity and been used in numerous campaigns in recent years, becoming an important tool for disseminating information.

This work shows the development of a tool for classification and analysis of tweets using Machine Learning for messages written in Spanish using information gathered from local online news to train a few models. While processing the data, we found that Multinomial Naïve Bayes has achieved the best classification results. On the other hand, we used DeBot to detect social bots and save automated accounts detected for a later analysis of the influence index of these accounts in 5 categories: politics, economy, sports, entertainment and other. Our results show bots' influence in those 5 categories varies depending on the hour of the day or the current local news of interest.

The prototype developed allows access to a summarized way of visualizing the influence of automated accounts in Twitter in a daily basis on issues of the country. Additionally, this work opens the door for studies to come on these platforms which will benefit our community.

Keywords: Machine learning, social bots, data mining.

# CAPÍTULO I. INTRODUCCIÓN

## 1.1 EL PROBLEMA DE INVESTIGACIÓN

### 1.1.1 Planteamiento del Problema

Según Shannon et al. (2016) las redes sociales se han convertido en una de las principales herramientas de comunicación de la sociedad a nivel mundial, desde su inicio estas han cambiado la manera de relacionarse entre las personas y han llegado a ser una herramienta capaz de ser la voz del usuario. Su influencia se evidencia en hitos históricos como: la primavera árabe, las elecciones populares de salida del Reino Unido de la Unión Europea, las elecciones presidenciales de los Estados Unidos de América, el referéndum de Cataluña; así como también las noticias falsas y mensajes de odio que se viralizan en las redes sociales.

Varios medios de comunicación elogiaron al internet y su papel “libertador” durante la primavera árabe, según *The New York Times* (2011), los manifestantes “disparaban tweets” frente a las balas; al mismo tiempo los términos “revolución Twitter” y “revolución Facebook” se pusieron de moda debido a los sucesos que marcaron al mundo árabe. Revueltas en Egipto, Túnez y Libia causaron la caída de sus gobiernos, por lo que Twitter y Facebook se convirtieron en cajas resonantes, transmitiendo y amplificando la desilusión de los manifestantes (Reporteros sin Fronteras, 2016). La gran diferencia de conflictos anteriores es que las redes sociales ya estaban consolidadas en la sociedad, por consiguiente, las revoluciones tunecina y egipcia revelaron ser revoluciones humanas impulsadas por el internet y las redes sociales (Warda, 2012). Dichos cambios políticos y sociales en la región se originaron debido al aislamiento al que estaban sometidos y concluyeron cuando el internet y las redes sociales se expandieron a todo el mundo. Por esta razón hay que tener en cuenta las capacidades comunicativas que ofrecen el internet y las redes sociales para comprender el desarrollo del proceso que ha derivado en las revueltas y en la posterior caída de los regímenes de Túnez, Egipto y Libia. En adición, estos medios de comunicación no

solo son usados por ciudadanos comunes, sino también por las autoridades, para divulgar propaganda oficial y fortalecer la vigilancia y el control de la población (Soengas, 2013).

El análisis realizado por Polonski (2016) refiere que la campaña a favor del *Brexit* tomó impulso *online* y marcó el tono del debate en las principales plataformas de redes sociales. Los datos obtenidos demostraron que los seguidores del *Brexit* recibieron un mensaje poderoso y emocional. Esto llevó a la agrupación de un mayor número de partidarios a favor de la campaña de salida de la Unión Europea, permitiéndoles dominar completamente en plataformas como Facebook, Twitter e Instagram, influenciando a grupos de votantes indecisos. Esta tecnología en aceleración está abierta a todos y puede utilizarse para cambiar la agenda pública e impulsar el cambio social para bien o para mal.

Lobo (2017) refiere que Donald Trump aseguró que su triunfo lo debía a las redes sociales, y manifestó: “Creo que las redes sociales son más poderosas que el dinero de campaña”. Asimismo, las manipulaciones de la popularidad de diferentes temas o las alteraciones de las preferencias del público repercuten luego en el pensamiento de las personas. Según una investigación realizada por la Universidad de Oxford existe evidencia de manipulación de información en redes sociales, con el propósito de influenciar la opinión de audiencias nacionales e internacionales, mediante el uso de cuentas automatizadas (Bradshaw & Howard, 2017).

Otro acontecimiento de importancia se registró en España en el referéndum de Cataluña, según afirmó un medio de comunicación español llamado “El País”, el cual acusó a Rusia de intromisión en el llamado *Procés Català* y señaló a “RT en español” cuenta de Twitter, como una de las principales herramientas de las que se sirve el Kremlin para desgastar a los Estados Unidos y la Unión Europea, es así como tuiteó: “La máquina de injerencias rusa penetra la crisis catalana. La red global que actuó contra Trump y el Brexit se dedica ahora en España”. (El País, 2017)

Este medio también se justificó diciendo que Julian Assange se encuentra a favor de la celebración de este referéndum y que la cuenta “RT en español” se ha convertido en eco de sus publicaciones y aún más importante afirmó que Assange cuenta con un ejército de robots que retwitean continuamente sus publicaciones en esta red social. Según *TwitterAudit* (2017) un análisis realizado en una muestra de 5000 seguidores de Assange en Twitter dio como resultado que el 51% de estas cuentas son falsas. Por otro lado, Lesaca (2017) utilizó un software avanzado de medición y análisis de *big data* y refirió:

Lo más sorprendente de la investigación ha sido descubrir todo un ejército de cuentas zombis perfectamente coordinadas que se dedican a compartir contenidos generados por RT y Sputnik en diversas conversaciones digitales, que van desde Siria y Estados Unidos hasta Cataluña. (Lesaca, 2017)

En su investigación pudo finalizar que solo un 3% de la conversación obedeció a perfiles reales.

Por otra parte, el linchamiento mediático, los discursos de odio y el *cyberbullying* son problemas que se difunden debido a la acción multiplicadora de las redes sociales las mismas que en muchos casos son causadas por el uso de *social bots*, noticias falsas o simplemente por la conmoción social que provoca en los usuarios, en la que, con poca o nula reflexión; se comparten o publican videos, fotografías o expresiones, que tienen consecuencias profundas en la sociedad y en la vida de las personas (Lowry et al., 2016).

Las redes sociales han cambiado la naturaleza de las campañas políticas y seguirán desempeñando un papel clave en las futuras elecciones. Cuantas más personas pasen gran parte de su vida cotidiana *online*, las redes sociales seguirán convirtiéndose en una de las fuerzas más poderosa para influenciar la difusión de ideas y mensajes políticos (Polonski, 2016). Con estos antecedentes queda en evidencia el poder de influencia de las redes sociales en la opinión de la población moderna, el potencial impacto social que conlleva dicha influencia, los peligros de promover sus contenidos y el desconocimiento que existe sobre los temas en los que la población ecuatoriana está siendo influenciada ya sea por cuentas automatizadas o por tendencias que se han viralizado. Es por esto por lo que se ha visto la importancia de proponer un estudio automatizado de la red social Twitter en el Ecuador que permita analizar el nivel de influencia de cuentas automatizadas.

#### **1.1.1.1 Diagnóstico del problema**

La manipulación de redes sociales, muchas veces propiciada por cuentas automatizadas genera tendencias que pueden influenciar la opinión de sus usuarios, provocando reacciones negativas tales como: acoso, linchamiento mediático o discursos de odio. En Twitter estas tendencias virales pueden ser generadas por *bots* que apoyan una publicación o la republican, motivando así el interés de usuarios reales y por consiguiente provocando un impacto en la sociedad. En Ecuador existen un promedio de 11 millones de usuarios activos en las redes sociales, conformando el 66% de la población ecuatoriana que

mantiene activa una o más cuentas, por otro lado existe un promedio del 73% de usuarios móviles, siendo estos usuarios más propensos al uso diario de esos medios digitales. En el caso específico de Twitter existen alrededor de 1 millón de cuentas activas de usuarios en Ecuador (We Are Social, 2018).

### **1.1.1.2 Pronóstico**

El uso de las redes sociales continuará creciendo a nivel mundial, así como también el uso de las cuentas automatizadas, lo que tendrá como consecuencia la generación de tendencias en las redes sociales (Rodríguez et al., 2017). Debido a que la influencia de Twitter es bastante significativa en hechos en tiempo real, se correrá el riesgo de impactar en las emociones de los usuarios, causando diversos problemas como linchamiento mediático, *cyberbullying* e incluso manipulación de la opinión de la población, con el propósito de influenciar en el pensamiento de un mayor número de personas.

Se incrementará el uso de *bots* electorales, como ya se vio en las elecciones de Estados Unidos que emitían mensajes positivos sobre Donald Trump (Lobo, 2017) y en Francia con un 20% aproximadamente de cuentas automatizadas que hacían menciones negativas al candidato Emmanuel Macron (Ferrara, 2017). Además, aumentarán los escenarios de cuentas automatizadas que comparten información incorrecta o noticias falsas las cuales no tienen un emisor válido, que afectan el interés de las personas y engañan a los usuarios.

A medida que siguen evolucionando las tecnologías como la inteligencia artificial, será menos perceptible diferenciar entre una publicación realizada por un humano o por un *bot* en discusiones en tiempo real, y esto ocasionará que las interacciones de las cuentas automatizadas tengan mayor credibilidad. Es así como el desconocimiento de las personas sobre los efectos de las redes sociales en la opinión de la población, y el uso de *bots* para la influencia de los usuarios continuará en aumento.

### **1.1.1.3 Control del Pronóstico**

Al contar con indicadores de tendencias de temas en la red social Twitter se puede conocer el índice de influencia sobre la opinión de los usuarios, y constatar la interacción de usuarios reales versus cuentas automatizadas en temas de tendencia, así como también comprender la realidad sobre los temas de interés público en la sociedad ecuatoriana.

### **1.1.2 Formulación de problema**

El incremento en el uso de Twitter en acontecimientos importantes como una elección presidencial, eventos deportivos o fenómenos naturales, refleja el tipo de uso que le dan los usuarios para obtener información en tiempo real de varias fuentes, el 64% de sus usuarios declaran usar a Twitter como una red social informativa (IAB Ecuador, 2017). Dichos acontecimientos provocan la influencia de temas de tendencia, por noticias que se viralizan o por interacción de cuentas automatizadas, que pueden generar conmoción social entre los usuarios de Twitter, quienes al no estar al tanto de estos fenómenos podrían ser manipulados.

### **1.1.3 Sistematización del problema**

Una vez planteado el problema objeto de esta investigación, es importante entender desde diferentes perspectivas por lo que se proponen las siguientes preguntas como pauta para el desarrollo de la investigación:

- a) ¿Cuál es la técnica más apropiada para analizar los datos en Twitter?
- b) ¿Cómo se analizará la información para el análisis de los temas de tendencia de Ecuador en Twitter?
- c) ¿De qué manera se identificarán las cuentas automatizadas?
- d) ¿Qué índice de intervención de cuentas automatizadas existe en los temas de interés de Ecuador en Twitter?
- e) ¿Cómo se presentarán los datos obtenidos de la investigación?

### **1.1.4 Objetivo General**

Desarrollar una herramienta de minería de datos para el análisis de influencia de cuentas automatizadas, procesando los temas que generan tendencia sobre la red social Twitter en Ecuador.

### **1.1.5 Objetivos Específicos**

- Seleccionar la técnica de análisis de datos mediante una investigación documental del estado del arte para la categorización de textos en español.
- Construir la arquitectura tecnológica de la herramienta para la extracción, procesamiento y categorización de los datos obtenidos de Twitter mediante un algoritmo de clasificación.

- Identificar las cuentas automatizadas con mayor número de interacciones mediante un algoritmo de detección escogido de la investigación documental.
- Analizar los índices de intervención de cuentas automatizadas de Twitter en los temas de tendencia en el país, utilizando los datos obtenidos.
- Implementar un prototipo web de visualización para el observatorio digital de la UISEK, a través de gráficas y tablas demostrativas.

### 1.1.6 Justificación

Con 12 años de trayectoria y alrededor de 328 millones de usuarios activos mensualmente, Twitter se ha consolidado como una de las redes sociales más importantes y populares de los últimos años, cuya fortaleza es la información en tiempo real, y es el medio oficial de empresas, instituciones públicas e inclusive gobiernos. La red social también se convirtió en un medio de comunicación masivo y un puente entre personajes públicos que ahora pueden interactuar con sus seguidores de forma directa (Twitter, 2017).

En esta investigación se identificaron los temas de mayor interés que se tratan diariamente en el territorio nacional ecuatoriano mediante los datos extraídos de la red social Twitter, utilizando técnicas de minería de texto y *Machine Learning*, para identificar los temas que generan un gran volumen de entradas. Estos resultados ayudaron a mostrar de forma resumida los temas de tendencia, así como ofrecer información para evidenciar el comportamiento social y el tipo de interacción que se da en la red social de forma diaria. Adicionalmente, con el desarrollo de este proyecto se pudo analizar el nivel de influencia de las cuentas automatizadas detectadas en Twitter en los temas de interés.

En Ecuador, la Ley Orgánica de Comunicación (2013), Sección 1 acerca de los derechos de la libertad artículo 17, menciona que: “Todas las personas tienen derecho a expresarse y opinar libremente de cualquier forma y por cualquier medio, y serán responsables por sus expresiones”. Además, el artículo 18 determina que está prohibida la censura previa por cualquier persona que revise, apruebe o desapruebe los contenidos previos a su difusión a través de cualquier medio de comunicación, a fin de obtener de forma ilegítima un beneficio propio, favorecer una tercera persona y/o perjudicar a un tercero. La omisión deliberada y recurrente de la difusión de temas de interés público constituye un acto de censura previa (Ley Orgánica de Comunicación, 2013). En lo internacional, el derecho fundamental a la libertad de expresión fue reconocido por la Resolución 59 de la Asamblea

General de las Naciones Unidas (1946), así como por el Artículo 19 de la Declaración Universal de Derechos Humanos (1948), dispone que: “El derecho fundamental a la libertad de expresión incluye el derecho de investigar y recibir informaciones y opiniones, y el de difundirlas, sin limitación de fronteras, por cualquier medio de expresión”.

Se considera de importancia el desarrollo de esta investigación porque aporta al estudio de las aplicaciones de análisis de datos digitales tanto en Ecuador como en proyectos internacionales que se puedan llevar a cabo en torno a las áreas que involucran la información y la transformación digital, las cuales constituyen el ámbito fundamental del Observatorio Digital de la Universidad Internacional SEK en Ecuador.

El desarrollo de una herramienta tecnológica de minería de datos para la clasificación de contenidos y análisis de influencia de cuentas automatizadas permitirá dar a conocer el índice de participación en temas de tendencia con miras a prevenir efectos nocivos que la viralización de noticias falsas o mal intencionadas pueden tener sobre nuestra sociedad.

## **1.2 MARCO TEÓRICO**

### **1.2.1 Estado actual del conocimiento sobre el tema**

En la actualidad existe una gran cantidad de información sobre diversos temas que están disponibles en las redes sociales, en las cuales se puede conocer lo que hace la gente y cómo se siente la gente, para saber el comportamiento y la opinión de los usuarios (Fernández, 2016). El acceso a la información permite predecir y gestionar diferentes sucesos aprovechados social o económicamente. Se tiene como ejemplo las aplicaciones descritas en las investigaciones de Sang et al. (2016) o de Moro et al. (2016). En el estudio de Birjali et al. (2017) proponen construir un vocabulario para analizar los sentimientos y opiniones expresadas en las redes sociales para la prevención de suicidios. Sus resultados experimentales demostraron que su método basado en *Machine Learning* puede extraer predicciones de pensamientos suicidas usando un *dataset* de 892 tweets, y con el uso del algoritmo *Naïve Bayes* se obtuvo una precisión del 87,5% en tweets con sospecha de riesgo de suicidio. Este tipo de estudios se encuentran más desarrollados en el idioma inglés.

En la investigación realizada por van Zoonen y van der Meer (2016) categorizan más de medio millón de Tweets de 430 empleados para una compañía de Países Bajos, en donde comparan el rendimiento de un grupo de clasificadores de aprendizaje supervisado evaluados

mediante indicadores de precisión. Usando 4000 tweets para entrenar los modelos de clasificación cuyos resultados fueron: *Linear Support Vector Machine (LSVM)* 85%, *Naïve Bayes* 84%, *K-Nearest Neighbors* 59% y *Logistic Regression* 80%. Adicionalmente, se determinó que el rendimiento de evaluación aumenta más lentamente para el algoritmo *Logistic Regression* que para el *LSVM* y *Naïve Bayes* al ampliar la cantidad de tweets de entrenamiento.

En relación al idioma castellano, existe una herramienta que permite evaluar el rendimiento de clasificación de tweets escritos en español con tópicos multi-etiqueta llamado Taller de Análisis del Sentimiento en la SELPN (TASS) (Villena & García, 2013); este *corpus* se trata de una colección de tweets que están etiquetados en 10 temáticas: política, entretenimiento, economía, música, fútbol, tecnología, deportes, cine, literatura y una categoría llamada otros. Por otro lado, Vilares (2014) realiza un análisis de contenidos de twitter para clasificar los mensajes e identificar la tendencia política de los usuarios, usando técnicas de minería de texto desde una perspectiva de procesamiento de lenguaje natural que permiten identificar los temas que se tratan en los tweets en idioma castellano con el uso del TASS. El modelo bigramas de lemas mostró un mejor rendimiento en los resultados experimentales, con los siguientes valores de clasificación en sus categorías seleccionadas: política 75,4%, economía 37,2%, deportes 50,3%, entretenimiento 44,3%, música 36,1%, tecnología 34,6%, fútbol 50,3% y otros 57,8%. Además, su clasificador de tendencia política es capaz de diferenciar entre usuarios conservadores, progresistas y de centro. En otra investigación realizada por Fernández (2016) muestra el análisis de contenidos en redes sociales, clasificando los mensajes e identificando las personas influyentes para el Banco Central Europeo. En este trabajo se realiza un rastreo web utilizando diferentes fuentes de información como Twitter, blogs, periódicos, e investiga a las cuentas de usuarios influyentes. De esta manera se desarrolló la capacidad de monitorear mediante los *trending topics* el comportamiento de sus usuarios.

De acuerdo a una investigación de las universidades de Carolina del Sur e Indiana (2016), hasta un 15% de las cuentas de Twitter son cuentas automatizadas (*bots*). En este trabajo desarrollaron una herramienta llamada BotorNot, que usa el método de aprendizaje supervisado *Random Forest*, y un conjunto de tweets de 5,6 millones para entrenar el modelo de clasificación de cuentas, entre humanos y *bots*. Los investigadores de la Universidad de Carolina del Sur resaltan que los *bots* tienen muchos beneficios como la difusión de noticias y publicaciones importantes, pero también tienen desventajas ya que hay un creciente registro

de aplicaciones maliciosas de *bots* sociales (Davis et al., 2016). En otro estudio realizado por Chavoshi, Hamooni y Mueen (2017) desarrollan un indicador de actividades entre los usuarios de las redes sociales, el cual reconoce el comportamiento de las cuentas automatizadas. Esta herramienta es llamada DeBot, la misma que identifica los *bots* de Twitter utilizando un método de aprendizaje no supervisado, que informa y archiva miles de cuentas todos los días. En febrero de 2017, recolectó más de 710 mil *bots* únicos desde agosto de 2015 (Chavoshi et al., 2017). Además, esta herramienta correlaciona las actividades de cuenta de millones de usuarios en tiempo real, puede detectar cuentas automatizadas en lapsos de 2 a 3 horas desde su inicialización y es capaz de identificar el comportamiento sincronizado en clúster de *bots*. En un estudio realizado por Kusen y Strembeck (2018), usando DeBot para analizar 1'317.555 cuentas de usuario distintas de Twitter, se identificaron 35,247 cuentas como *bots*, dando un valor del 2,67% de cuentas automatizadas.

La creciente popularidad de las redes sociales en los últimos años ha convertido estas plataformas en objetivos atractivos para los ataques de *Sybil*. Se estima que existen decenas de millones de cuentas *Sybil* en las redes sociales más populares como Twitter y Facebook (Kelly, 2012). Un atacante puede usar estas cuentas de *Sybil* para interrumpir las campañas electorales e influir en bolsas de valores mediante la difusión de noticias falsas (von Hoffman, 2013). La detección de actividades sospechosas en las redes sociales puede convertirse en un elemento importante de investigación, cuyo papel garantizará la privacidad de los usuarios, de igual modo el detectar actividades maliciosas tales como *spamming*, clonación de cuentas, *sybils*, etc. (Savage et al., 2014).

### **1.2.2 Adopción de una perspectiva teórica**

En esta investigación se realizó la clasificación multi-clase de los tweets de los usuarios según los temas de tendencia así como también se identificaron cuentas automatizadas que influyen los temas de tendencia en el territorio ecuatoriano, por lo que se ha tomado la siguiente perspectiva teórica del estado del arte.

Los autores que han publicado acerca del análisis de textos en la red social Twitter más relevantes son: David Vilares Calvo, quien utiliza la técnica de análisis de Lenguaje Natural para los tweets y Javier Mateo Fernández quien hace uso de la técnica de *Bag of Words*, en ambos casos se analizan tweets en español. Para la clasificación de los tweets se

decide utilizar la perspectiva teórica del autor Javier Fernández debido a que se entrena el modelo de aprendizaje supervisado con textos actuales provenientes de fuentes de internet con lenguaje común usado en medios digitales.

Respecto al análisis de cuentas automatizadas, se decide tomar la perspectiva teórica de Nikan Chavoshi, et al. (2016), por el modelo de aprendizaje no supervisado para detectar el comportamiento sincronizado de cuentas en la red social Twitter utilizado en su investigación, y los porcentajes de aciertos encontrados en su estudio.

### **1.2.3 Marco conceptual**

#### **Minería de datos**

Estudia métodos y algoritmos que permiten la extracción automática de información sintetizada para caracterizar las relaciones escondidas en una gran cantidad de datos, con el propósito de soportar los procesos de toma de decisiones con mayor conocimiento. La Minería de Datos combina técnicas semi-automáticas de inteligencia artificial, análisis estadístico, bases de datos y visualización gráfica, para la obtención de información que no esté representada explícitamente en los datos (Castelló & Santamaría, 2015).

#### **Aprendizaje de Máquina (*Machine Learning*)**

El aprendizaje automático es una rama de la inteligencia artificial cuyo objetivo es desarrollar técnicas que permitan a las computadoras aprender. Este trata de crear programas capaces de generalizar comportamientos a partir de una información no estructurada suministrada en forma de ejemplos. El aprendizaje automático es una técnica de la ciencia de datos que permite a los equipos utilizar datos existentes para prever tendencias, resultados y comportamientos futuros (Microsoft Azure, 2016). Divya (2016) define al también llamado *Machine Learning* (ML) como el diseño y estudio de artefactos de software que utilizan la experiencia pasada para tomar decisiones futuras, es decir programas que aprenden de los datos. El objetivo fundamental del aprendizaje automático es generalizar o inducir una regla desconocida a partir de ejemplos de la aplicación de la regla. Existen dos tipos de aprendizaje de máquina el supervisado y no supervisado.

#### **Aprendizaje Supervisado**

El método de aprendizaje automático también conocida como *Supervised Machine Learning* (SML), es aplicable a la clasificación supervisada en general y a las técnicas de clasificación de texto en particular. Por lo tanto, se llama "aprendizaje supervisado". En una clasificación basada en el aprendizaje automático, se requieren dos conjuntos de documentos: *training set* y *test set*. Un conjunto de datos de entrenamiento o *training set* es utilizado por un clasificador automático para aprender las características de diferenciación de los documentos, y un conjunto de datos de pruebas *test set* se utiliza para validar el rendimiento del clasificador automático (Alhojely, 2016).

El SML generalmente está conformado de tres etapas. En primer lugar, los documentos de texto en el conjunto de entrenamiento son convertidos en datos accesibles para el análisis computacional. Cada documento se representa como un vector de elementos de texto cuantificables denominados características o *features*. En segundo lugar, los vectores de características de todos los documentos del *dataset* junto con las etiquetas de contenido de los documentos se utilizan para formar un clasificador. Al hacerlo, un algoritmo de aprendizaje supervisado analiza las características de cada categoría del contenido y genera un modelo predictivo para clasificar los futuros documentos de acuerdo con las categorías de contenido (Burscher et al., 2014).

### **Aprendizaje no supervisado**

El aprendizaje no supervisado es un tipo de aprendizaje automático que utiliza información que no está clasificada ni etiquetada, y que permite que el algoritmo actúe sobre la información sin orientación. Por lo tanto, los algoritmos de aprendizaje automático no supervisados actúan sobre datos que no han sido entrenados previamente (Huang et al., 2014).

### **Técnicas de aprendizaje de máquina supervisado para clasificación**

Las técnicas de aprendizaje supervisado como *Naïve Bayes* (NB), *Logistic Regression* y *Support Vector Machines* (SVM) han logrado un gran éxito en la categorización de texto. Los otros métodos de aprendizaje de máquina más conocidos en el área de procesamiento de lenguaje natural son *K-Nearest Neighbor*, clasificador de centroides y el modelo N-gram (Alhojely, 2016).

## Clasificación multi-etiqueta

La clasificación es una de las principales áreas de interés en el campo del aprendizaje automático. Esencialmente, consiste en hallar una función, denominada clasificador, que reciba como entrada los atributos de un determinado patrón y entregue como salida una clase, de entre una serie de clases predefinidas en el problema, que indique la clase con la que se considera que está asociado dicho patrón (Fernández, 2016).

### Algoritmo *Naïve Bayes*

*Naïve Bayes* es un algoritmo de clasificación simple pero eficaz y es ampliamente utilizado para la clasificación de documentos. La idea básica es estimar las probabilidades de las categorías dado un documento de prueba (*test set*), usando las probabilidades conjuntas de palabras y categorías. La parte ingenua de tal modelo es la suposición de la independencia de la palabra. La simplicidad de este supuesto hace que el cálculo computacional del clasificador *Naïve Bayes* sea mucho más eficiente (Divya, 2016).

El algoritmo *Naïve Bayes* se puede interpretar matemáticamente de la siguiente manera: dado un conjunto de  $r$  vectores de documento  $D = \{d_1, \dots, d_r\}$ , clasificados a lo largo de un conjunto  $C$  de clases  $q$ ,  $C = \{c_1, \dots, c_q\}$ , los clasificadores bayesianos estiman las probabilidades de cada clase  $c_k$  dado un documento  $d_j$  como:

$$P(c_k|d_j) = \frac{P(c_k)P(\vec{d}_j|c_k)}{P(\vec{d}_j)} \quad (1)$$

En la ecuación 1,  $P(\vec{d}_j)$  es la probabilidad de que un documento tomado al azar tenga el vector  $\vec{d}_j$  como su representación, y  $P(c_k)$  es la probabilidad de que un documento tomado al azar pertenezca a  $c_k$ . Debido a que la cantidad de documentos posibles es muy alta, la estimación de  $P(\vec{d}_j|c_k)$  es problemático. *Naïve Bayes* presenta resultados que son muy competitivos con los obtenidos por métodos más elaborados (Divya, 2016).

### Algoritmo *Multinomial Naïve Bayes* (MNB)

Este algoritmo es una versión especializada de *Naïve Bayes*, que captura información de frecuencia de palabras en documentos, por lo que está diseñado para trabajar en la clasificación de documentos de texto (Zhao et al., 2016). En el modelo MNB, un documento

representa una secuencia ordenada de palabras, extraídos del mismo vocabulario  $V$ . Suponiendo que las longitudes de los documentos son independientes de la clase, y que la probabilidad de cada palabra en un documento es independiente del contexto y de la posición de la palabra en el documento. Por lo tanto, cada documento  $\vec{d}_j$  se extrae de una distribución multinomial de palabras con tantas pruebas independientes como la longitud de  $\vec{d}_j$ . Generando una representación de los documentos llamado *bag of words*. Definiendo  $N_{jt}$  como el conteo del número de ocurrencias de la palabra  $w_t$  en el documento  $\vec{d}_j$ . Por lo tanto la probabilidad de un documento dada su clase es la distribución multinomial de la Ecuación 2 (ChandraKala & Sindhu, 2012).

$$P(c_k | d_j) = P(|d_j|) |d_j|! \prod_{t=1}^{|V|} \frac{P(w_t | c_k)^{N_{jt}}}{N_{jt}!} \quad (2)$$

### **Algoritmo Support Vector Machines (SVM)**

*Support Vector Machines* es un algoritmo clasificador discriminativo, se considera el mejor método de clasificación de texto. La máquina de vectores de soporte es un método de clasificación estadística, basado en el principio de minimización del riesgo estructural de la teoría de aprendizaje computacional. El SVM busca una base de decisión para separar los puntos de datos de entrenamiento (*training set*) en dos clases y toma decisiones basadas en los vectores de soporte que se seleccionan como los únicos elementos efectivos en el conjunto de entrenamiento (*training set*). Se han desarrollado múltiples variantes de SVM en las que se utiliza SVM de clase múltiple para la clasificación de textos (Patil, 2016).

### **Algoritmo Logistic Regression**

El algoritmo *Logistic Regression* es un modelo lineal de clasificación, este algoritmo es también conocido por su clasificación de entropía máxima, o clasificador log-lineal. En los modelos de regresión logística, las probabilidades que describen los posibles resultados de un solo ensayo se modelan utilizando una función logística. La técnica de *Logistic Regression* multinomial, consiste en la estimación de la probabilidad que una observación  $c_k$  pertenezca a cada uno de los grupos. El modelo compara  $G-1$  categorías contra una categoría de referencia dada, donde  $d_j$  es un vector, y  $c_k$  es una variable aleatoria independiente multinomial con valores 1, 2, ...G, la cual indica el grupo a la que pertenece. La probabilidad

condicional de pertenencia de  $d_j$  a cada grupo está dada por la ecuación 3 (Yu, Huang, & Lin, 2016):

$$P(c_k | d_j) = \frac{e^{\alpha_c + \beta_c d_j}}{1 + \sum_{g=2}^G e^{\alpha_c + \beta_g d_j}} \quad (3)$$

### **Algoritmo *K-Nearest Neighbor* (KNN)**

El KNN es un típico ejemplo de un clasificador que no construye una representación explícita, declarativa de la categoría, pero se basa en las etiquetas de categoría adjunto a los documentos de formación similar al documento de prueba. Dado un documento de prueba, el sistema encuentra los  $k$  vecinos más próximos entre los documentos de entrenamiento. La puntuación de similitud de cada documento vecino más cercano al documento de prueba se utiliza como el peso de las clases del documento de vecino (Bijalwan et al., 2014).

### **Detección de cuentas automatizadas**

Una cuenta automatizada usada en una red social es también conocida como *bot* o *social bot*, que consiste en un algoritmo que automáticamente produce contenido e interactúa con humanos, y que además cumplen con ciertas funciones como manipular las noticias con rumores, *spam*, generan desinformación y difamación en las redes sociales. Algunos *bots* emulan el comportamiento humano con el objetivo de influenciar la opinión de los usuarios, fabricar falsos apoyos políticos o para promover la propaganda terrorista y el reclutamiento de personas (Davis, 2016).

En el estado del arte se mencionan los resultados obtenidos con dos herramientas de análisis de cuentas automatizadas en redes sociales, BotOrNot y DeBot, cuyo funcionamiento se explicará a detalle a continuación.

### **DeBot**

Esta herramienta recopila los datos de Twitter constantemente y detecta los *bots* en función de su sincronización. DeBot es una novedosa técnica casi en tiempo real que utiliza la correlación de actividad entre los usuarios como un indicador del comportamiento de las cuentas automatizadas. En esta investigación se analiza matemáticamente la importancia de los *bots* correlacionados y demuestra que el falso positivo del sistema es casi nulo. Este

motor de detección de *social bots* en Twitter consiste de cuatro pasos: en primer lugar, comienza a recopilar tweets al escuchar un conjunto de palabras clave de la zona referenciada geográficamente, en este caso se usan los temas de tendencia. En el segundo paso, se usan técnicas de *hashing* para filtrar aquellos usuarios que no están altamente correlacionados e informan a un conjunto de usuarios sospechosos para el siguiente paso. En el tercer paso, se escucha a los usuarios sospechosos detectados previamente durante un cierto tiempo de duración. Finalmente, en el último paso, DeBot calcula la correlación entre los usuarios sospechosos del paso tres y reporta los altamente correlacionados como cuentas automatizadas (Chavoshi et al. 2016).

Debot usa una señal de comportamiento del usuario, llamada correlación *warping* la cual es invariante en el tiempo, cuyo umbral de correlación *warping* usado para identificar bots es de 0.995, confirmando un comportamiento casi idéntico entre dos usuarios. Esta correlación es computacionalmente muy costosa de usar, por lo que dependiendo el procesador que se utilice los cálculos de este algoritmo serán casi en tiempo real (Chavoshi et al. 2016).

De acuerdo al estudio realizado por Chavoshi et al. (2016) este método detecta más bots que las técnicas de aprendizaje supervisado existentes. En su investigación desarrollan una nueva técnica de *hashing* para agrupar rápidamente usuarios en base a sus correlaciones de *warping*. Lo cual permite realizar una comparación cruzada de millones de series de actividad bajo deformación en el tiempo; demostrando empíricamente que DeBot tiene un 94% de precisión, capaz de detectar bots que otros métodos no detectan. Llegando a la conclusión que los bots se pueden agrupar funcionalmente y que su número crece a un ritmo elevado.

## **BotOrNot**

La herramienta de clasificación de BotOrNot genera más de 1,000 características usando metadatos disponibles e información extraída de patrones de interacción y contenido. Las cuales se pueden agrupar en 6 características principales que capturan varias dimensiones de los patrones de difusión de la información. Este modelo recolecta información de redes basadas en *retweets*, menciones y co-ocurrencias de *hashtags*, y extrae sus características estadísticas: grados de distribución, coeficiente de agrupamiento y medidas de centralidad. Las funciones del usuario se basan en los metadatos de Twitter relacionados con una cuenta,

incluido el idioma, las ubicaciones geográficas y el tiempo de creación de la cuenta. Las características de los amigos incluyen estadísticas descriptivas relativas a los contactos sociales de una cuenta, como la mediana, los momentos y la entropía de las distribuciones de su número de seguidores, seguidores, publicaciones, etc. Las características temporales capturan los patrones de sincronización de la generación y el consumo de contenido, como la tasa de tweets y la distribución del tiempo entre tweets. Las características del contenido se basan en pistas lingüísticas calculadas a través del procesamiento del lenguaje natural, especialmente el etiquetado de parte del habla. Las características de sentimiento se crean usando algoritmos de análisis de sentimientos, que incluyen emociones de felicidad, excitación-dominancia-valencia y emoticones (Davis et al. 2016).

El clasificador de aprendizaje supervisado usado por BotOrNot es el algoritmo *Random Forest*, en donde las características extraídas de este modelo se aprovechan para entrenar siete clasificadores diferentes, uno para cada subclase de características y otro para el puntaje general. Una validación cruzada de diez veces produce un rendimiento de 0.95 AUC (área bajo curva ROC) (Davis et al. 2016). Siendo ROC (curva característica operativa del receptor) una gráfica que muestra el rendimiento de un modelo de clasificación en todos los umbrales de clasificación, y AUC representa la medición de toda el área bidimensional debajo de toda la curva ROC. El AUC varía en valor de 0 a 1, en un modelo cuyas predicciones tiene un AUC de 1.0 son 100% correctas (Marsland, 2015).

#### **1.2.4 Hipótesis**

La experiencia en otros países demuestra que las redes sociales, en este caso particular Twitter, son afectadas por la presencia de cuentas automatizadas. La hipótesis de esta investigación es que al igual que en el resto de países estudiados, las redes sociales del Ecuador son víctimas de la presencia de cuentas automatizadas que buscan generar tendencia en temas de interés social. Por lo que, el desarrollo de una herramienta tecnológica de minería de datos para la clasificación de contenidos y análisis de influencia de cuentas automatizadas en los usuarios de Twitter en Ecuador, permitirá dar a conocer a la población ecuatoriana el índice de participación de cuentas automatizadas en temas de tendencia y prevenir el incremento de los efectos potencialmente nocivos que son provocados por la viralización de noticias.

## **CAPÍTULO II. MÉTODO**

### **2.1 Tipo de estudio**

#### **Exploratorio**

La presente investigación busca estudiar la influencia de cuentas automatizadas en los temas de tendencia de la red social Twitter en el Ecuador. Como ha sido descrito en el estado del arte este es un problema actual que puede tener repercusiones sociales ante todo por la poca información acerca de este tema.

Para este estudio la adquisición de los *tweets* y los temas de tendencia se utilizó la interfaz de programación provista por Twitter (Twitter API), con una cuenta gratuita de desarrollador. Para entrenar los métodos de aprendizaje supervisados se usó la información de diarios de suscripción gratuita mediante la herramienta RSS (*Really Simple Syndication*). A continuación, los tweets fueron procesados por los algoritmos de clasificación, permitiendo detectar automáticamente el tema que trata el tweet, y la herramienta DeBot aportó con la información si el tweet fue generado por una cuenta automatizada. Adicionalmente, se puede indicar que se contemplaron todas las fases de la analítica de datos en el proceso realizado.

Este es un estudio de campo (digital) que mide la interacción de personas y cuentas automatizadas alrededor de los temas de interés que se posicionan como tendencia en la red social Twitter.

#### **Descriptivo**

Este estudio busca describir relativamente el nivel de influencia que las cuentas automatizadas tienen sobre un tema (o categoría) tratado en Twitter. La aplicación agrupa de forma automatizada temas en común y se limita a señalar las características de ese conjunto estudiado contraponiendo cuentas reales y cuentas automatizadas.

### **2.2 Modalidad de investigación**

#### **De campo**

Los datos son recopilados directamente de Twitter a través de su interfaz de programación que permite el acceso al objeto de estudio que son los tweets que corresponden a una tendencia en Ecuador y que son generados tanto por cuentas reales como por cuentas automatizadas.

## **Proyecto de Desarrollo**

En esta investigación se desarrolló un prototipo funcional de una herramienta web que permite la recopilación, análisis y clasificación de *tweets* generados en Ecuador tanto por cuentas autónomas como por cuentas reales sobre temas de tendencia en la red social Twitter.

En el capítulo III de diseño y desarrollo del prototipo se evidencia la viabilidad de implementación de una herramienta similar para uso personal o institucional.

## **2.3 Método**

### **Método Hipotético – Deductivo**

El método para el presente trabajo es del tipo Hipotético – Deductivo, debido a que se aplica una hipótesis inicial con el fin de obtener predicciones a ser verificadas al finalizar la investigación. Cuyas conclusiones posibilita la propuesta y verificación de nuevas hipótesis de trabajo.

Partiendo de la hipótesis que en Ecuador las redes sociales se comportan como en los países estudiados en el estado del arte, se puede afirmar que nuestro país se encuentra vulnerable a todas las problemáticas suscitadas anteriormente.

## **2.4 Población y Muestra.**

**Población:** Todos los mensajes enviados por los usuarios de Twitter que generan tendencia en Ecuador.

**Muestra:** Todos los mensajes enviados a través de la interfaz de programación de Twitter por una cuenta de desarrollador gratuita (Twitter API).

Nótese que no hay una manera específica de cuantificar la población y la muestra, ya que se trata de poblaciones y muestras generadas y calculadas en tiempo real conforme los usuarios de la red social comparten información. Sin embargo, con el tipo de API usada

Twitter garantiza que envía el 1% de tweets en una ubicación. Adicionalmente, se trabajó con esa muestra para determinar la incidencia de los *bots* en la red social.

## 2.5 Selección de instrumentos de Investigación

### Observación

Las observaciones se realizaron de forma sistemática durante los 9 últimos meses de transcurso en la aplicación del proyecto, procesando los *tweets* obtenidos en tiempo real, mediante un algoritmo de clasificación para categorizarlos de acuerdo a los temas escogidos. Para el entrenamiento del modelo de clasificación se utilizaron noticias actualizadas para evitar un sesgo en los clasificadores y hacer pruebas de resultados. Adicionalmente, con la herramienta de búsqueda de cuentas automatizadas DeBot, se obtuvieron dichas cuentas para calcular los índices de influencias en los temas clasificados.

## 2.6 Validez y Confiabilidad de los Instrumentos

En un estudio realizado por van Zoonen y van der Meer (2016) se implementan algoritmos de multi-clasificación logrando un porcentaje del 80% para el procesamiento de tweets, teniendo una base de registros de 3000 tweets, por lo que se seleccionaron estos modelos de clasificación para el desarrollo de este proyecto. A partir de estos modelos se realizaron pruebas empíricas para comparar los valores de evaluación obtenidos.

DeBot como herramienta de búsqueda de bots no depende del lenguaje en el que el *tweet* está escrito, de esta manera las cuentas automatizadas presentan un comportamiento similar independiente del lenguaje. Por esta razón se utilizaron los resultados empíricos de DeBot (94%) como garantía del índice de influencia medido por categoría.

## **CAPÍTULO III. DISEÑO Y DESARROLLO DEL PROTOTIPO**

### **3.1 Funcionalidad**

La aplicación ha sido desarrollada en un sistema operativo Ubuntu 16.04, en el cual se encuentran alojados los códigos de minería de datos, y estará en línea para ser accedida mediante una interfaz web como una terminal gráfica de consultas. Esta aplicación busca recolectar los *tweets* enviados por la red social Twitter, y a su vez ser clasificados por temas, cuyo algoritmo de entrenamiento será alimentado por noticias descargadas a través de canales RSS de diarios locales. Por último, con el uso de la herramienta DeBot identificar los usuarios de cuentas automatizadas que han generado contenido en la red social.

Para la construcción de esta herramienta se propone el siguiente diseño modular describiendo la infraestructura donde se encuentra el detalle del hardware y software utilizado.

### **3.2 Diseño de la arquitectura tecnológica**

Para el diseño de la arquitectura se seleccionaron herramientas de programación, de almacenamiento de la información y de visualización de los datos procesados.

#### **3.2.1 Infraestructura de desarrollo**

Debido a la facilidad y a los costos relativamente bajos de los servicios en la nube hoy en día, se propone implementar la herramienta de minería de datos en un entorno virtual, haciendo uso de servidores que funciones 24/7, donde la herramienta desarrollada se encuentre almacenada en este entorno con todas las librerías necesarias y que trabaje recopilando datos todos los días. Ofreciendo así un servicio de última generación para la distribución de información a la comunidad.

#### **3.2.2 Descripción de hardware y software**

##### **Hardware**

Para el desarrollo del presente proyecto se ha utilizado el servicio de computación en la nube de Google, este servicio tiene el nombre de *Google Cloud Computing*, cuyos servicios en la nube son considerados como uno de los mejores en el mercado de *Cloud Computing*, y además provee una infraestructura de servicio 24/7, con seguridad *end-to-end*, configuración de entornos virtuales escalables y facilidad de desarrollo de prototipos de proyectos de forma gratuita por un tiempo de hasta un año o un cupo de \$300 para el desarrollo de proyectos informáticos. Pasado este tiempo se puede descargar la máquina virtual configurada y usarse en un computador con un software de virtualización como VMware.

Existen diversas opciones de configuración de hardware para las máquinas virtuales, en las cuales se puede elegir desde un procesador de baja capacidad hasta varios procesadores en una misma máquina virtual, de la misma manera se puede elegir una memoria RAM desde 600 MB hasta 64 GB, esto dependerá del costo de cada elemento, ya que se cuenta con un cupo limitado de prueba. El sistema operativo se puede elegir de una lista de varias distribuciones Linux como versiones de Windows

## Software

En la tabla 1, se presenta el detalle de software utilizado para el desarrollo del prototipo, con sus respectivas versiones y una breve descripción de sus funciones.

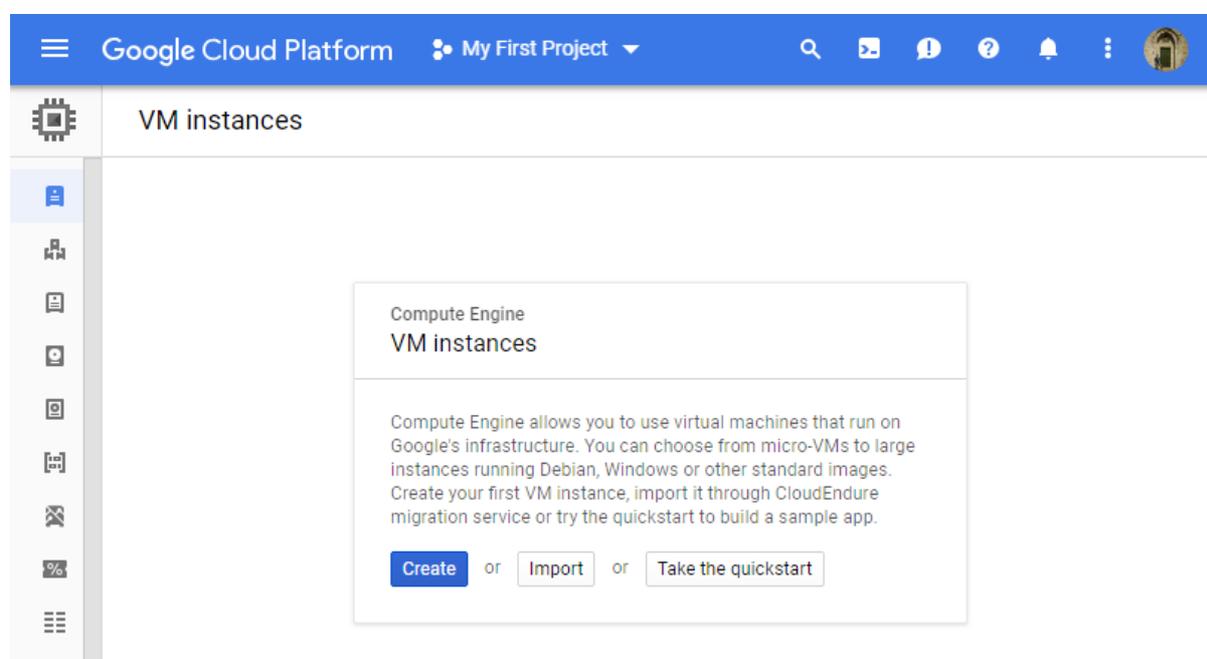
Software	Detalle
Python 2.7	Lenguaje de programación.
MongoDB 4.10	Base de datos no relacional orientada a los documentos.
Pymongo 3.5.0	Librería en Python para trabajar con MongoDB.
Tweepy 3.5.0	Librería en Python para trabajar con la API de Twitter.
Oauth 1.0.1	Librería en Python para trabajar con el registro de las credenciales del API de Twitter de forma segura.
Scikit-learn 0.19.0	Librería en Python para minería de datos y <i>Machine Learning</i> .
Numpy 1.13.3	Librería en Python de funciones matemáticas para trabajar con arreglos, vectores y matrices multidimensionales.
Pandas 0.21.0	Librería en Python para la manipulación y análisis de datos.
Matlab 2017a	Entorno de software para operaciones matemáticas.
Gunicorn 19.7.1	Herramienta WSGI HTTP para Python

Bokeh 0.12.10	Librería en Python para la visualización web.
---------------	---

**Tabla 1.** Detalle de software base. Elaborado por: Autor.

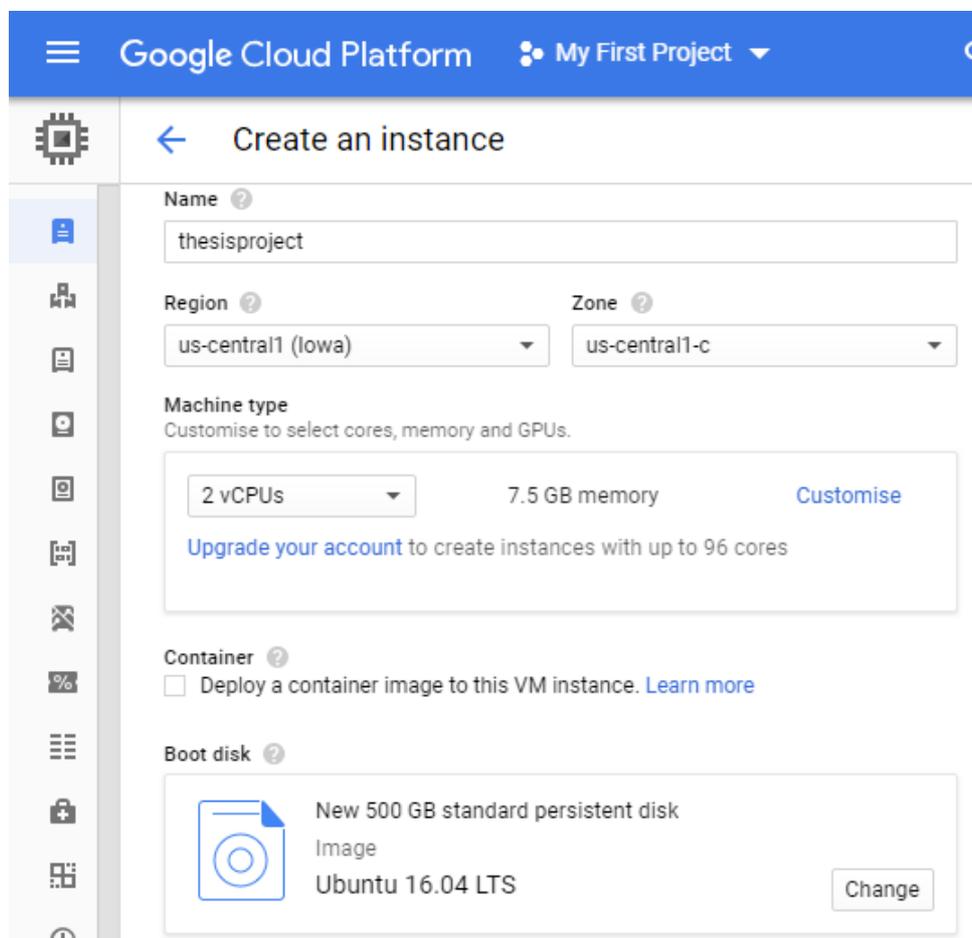
### 3.2.3 Configuración del entorno de desarrollo

Para iniciar con las configuraciones de la máquina virtual para alojar la herramienta de minería de datos se crea una instancia en *Google Cloud Computing*, en donde se trabajó con una cuenta gratuita por un año. En la Figura 1 se visualiza la interfaz del *dashboard* de la plataforma de Google, desde aquí se pueden crear una o más máquinas virtuales para el desarrollo del proyecto.



**Figura 1.** Interfaz Google Cloud para creación de máquinas virtuales. Elaborado por: Autor.

En la Figura 2 se observa el siguiente paso para crear y configurar los parámetros generales de la máquina virtual, como el procesador, la memoria RAM, el tamaño del disco para el almacenamiento de datos del proyecto, y el tipo de sistema operativo, que en este caso se utilizará la distribución de Linux llamada Ubuntu 14.04.



**Figura 2.** Elección de parámetros de la máquina virtual. Elaborado por: Autor.

En la Figura 6 se detallan los tipos de acceso de red a la máquina virtual, eligiendo los parámetros por defecto, y habilitando los puertos HTTP y HTTPS para trabajar con el prototipo de herramienta web; y finalmente crear la máquina virtual que estará alojada en la nube disponible las 24 horas del día procesando las tareas programadas.

Identity and API access ?

Service account ?  
Compute Engine default service account

Access scopes ?  
 Allow default access  
 Allow full access to all Cloud APIs  
 Set access for each API

Firewall ?  
Add tags and firewall rules to allow specific network traffic from the Internet.  
 Allow HTTP traffic  
 Allow HTTPS traffic

Management, disks, networking, SSH keys

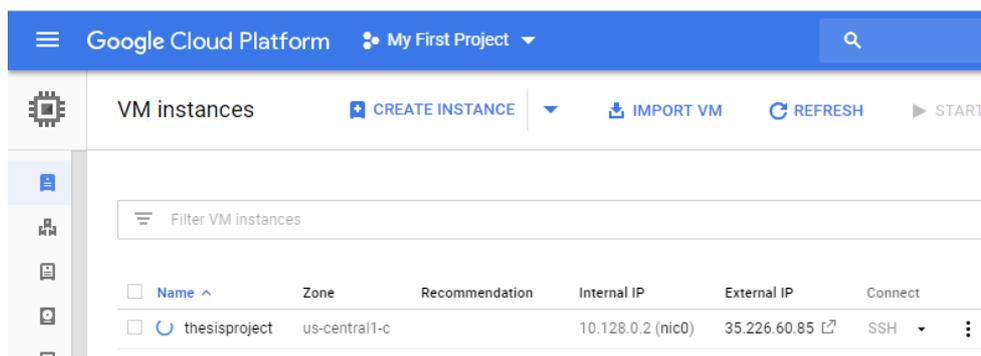
Your Free Trial credits, if available, will be used for this instance

Create Cancel

Equivalent REST or command line

**Figura 3.** Permisos de conexiones de red de la máquina virtual. Elaborado por: Autor.

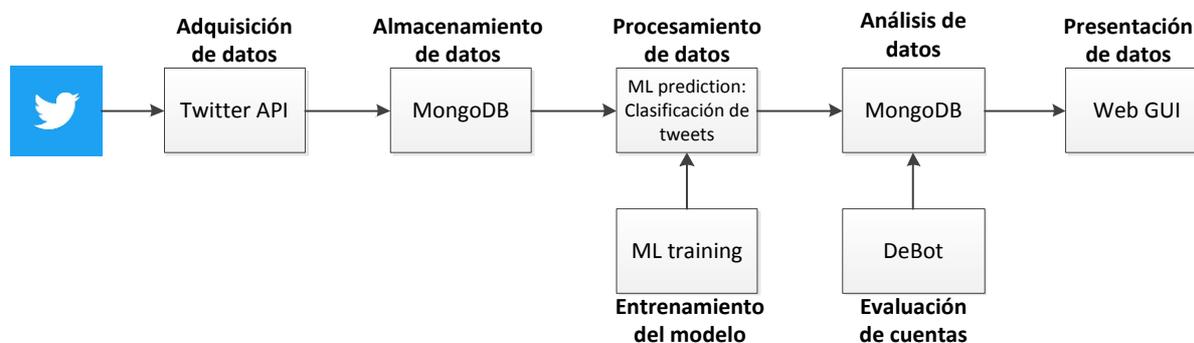
Una vez creada la máquina virtual, hay que realizar el procedimiento de verificación de llaves de acceso para la conexión remota SSH, o a su vez se puede usar la interfaz web SSH. En la figura 4, se muestra la máquina virtual en el *dashboard* de Google Cloud, donde se visualizan las IP's asignadas por Google automáticamente. La IP pública para la conexión remota es la 35.226.60.85.



**Figura 4.** Selección de la máquina virtual en el *dashboard* de Google Cloud. Elaborado por: Autor.

### 3.2.4 Desarrollo de los módulos de la herramienta

La herramienta de análisis está dividida en módulos que realizan diferentes funciones como la adquisición, almacenamiento, procesamiento, análisis y presentación de los datos, este procedimiento contempla las siguientes fases de la analítica de datos: obtención de datos, procesamiento de datos, planeación del modelo, construcción del modelo, comunicación o visualización de datos y operacionalización. En la figura 5 se muestra un diagrama de bloques con los respectivos módulos, desde la obtención de los datos de Twitter hasta la presentación en una interfaz gráfica, los que explican a detalle a continuación.



**Figura 5.** Arquitectura de la herramienta de minería de datos. Elaborado por: Autor.

#### Módulo de adquisición de datos de Twitter

Por medio de la API de Twitter se adquieren los tweets y los temas de tendencia usando el lenguaje de programación Python versión 2.7 apoyado de una librería llamada Tweepy versión 3.5.

Twitter proporciona una interfaz de desarrollo (API) gratuita para la creación de software y aplicaciones que permiten manejar datos de Twitter. Esta API tiene acceso limitado al número de solicitudes por hora. Para tener acceso a sus API's Twitter usa el modo de autenticación OAuth, el cual es un protocolo abierto para permitir acceso seguro de manera estándar. La documentación acerca de las API's ha sido tomada de Twitter en Septiembre del 2017 para proceder con el diseño de la herramienta del presente trabajo de investigación. Dos de las APIs más usadas que proporciona Twitter para la generación de aplicaciones son:

### **REST API**

Es una API web que funciona por HTTP a la cual se accede por URLs que devuelven contenido en formato JSON, XML, HTML, etc. Esta API no tiene limitación temporal de peticiones, pero sí una limitación del número de resultados devueltos establecido en 3200 tweets.

### **Streaming API**

Esta API permite recibir información casi en tiempo real, cuyos contenidos se encuentran en formato JSON. Además, se pueden obtener datos filtrados por palabras clave o usuarios, idioma, localización, *hashtags*, etc. Existe restricción en el envío de datos a través de esta interfaz, con un máximo de 50 tweets por segundo (Twitter, 2017).

Para el propósito de este proyecto se utilizó la Streaming API, ya que los datos se leen en tiempo real y se los almacena para analizarlos posteriormente. Los temas de tendencia que son palabras clave o *hashtags*, se los toma cada 30 minutos, durante este tiempo se descargan los *tweets*. A continuación, los *tweets* y los temas de tendencia se los almacena en una base de datos, con los parámetros originales en formato JSON proporcionados por Twitter. La adquisición de los datos se realiza de forma segura usando un API\_KEY y un Token para la autenticación, estos parámetros son generados mediante la cuenta de desarrollador gratuita proporcionada por Twitter (2017).

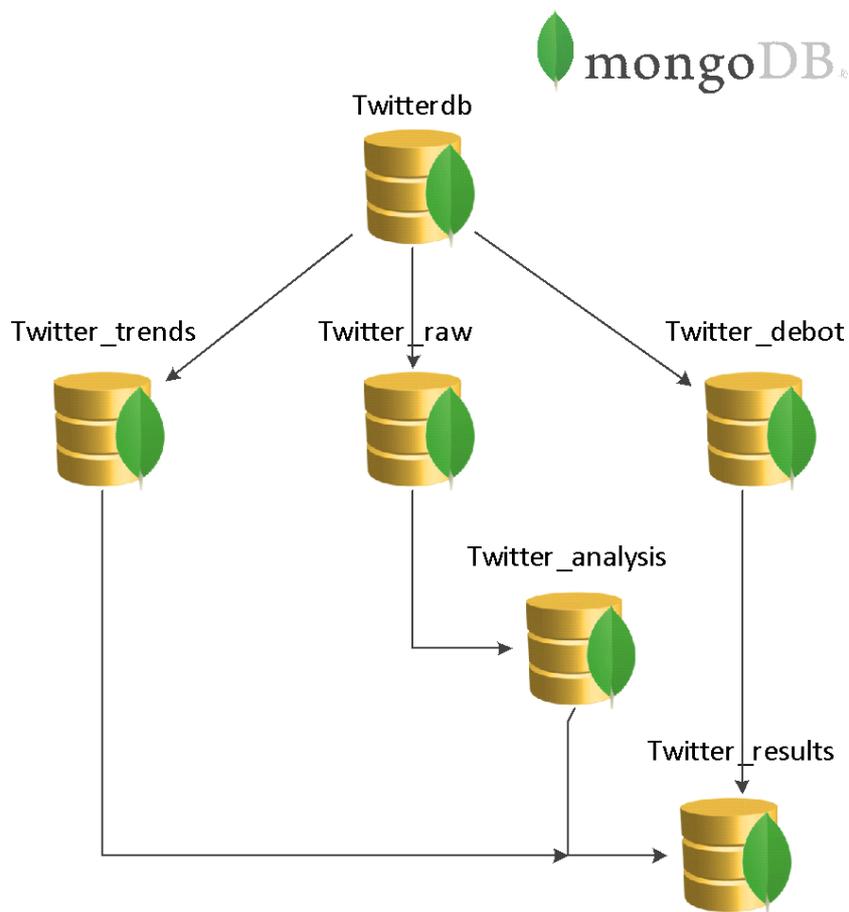
### **Módulo de almacenamiento de datos**

La base de datos seleccionada para el almacenamiento de la información ha sido MongoDB versión 4.10, debido a sus prestaciones de escalabilidad, rendimiento, seguridad y gran disponibilidad. Este tipo de base de datos permite a los esquemas cambiar rápidamente cuando las aplicaciones evolucionan, proporcionando una funcionalidad que los

desarrolladores esperan de las bases de datos tradicionales, tales como índices secundarios, un lenguaje completo de búsquedas y consistencia estricta.

MongoDB es una base de datos orientada a documentos NoSQL, que ha sido creada para el manejo de arquitecturas complejas de centros multidados. Además, brinda un elevado rendimiento, tanto para lectura como para escritura, potenciando la computación en memoria (in-memory). La replicación nativa de MongoDB y la tolerancia a fallos automática ofrece fiabilidad a nivel empresarial y flexibilidad operativa. Esta base de datos es de tipo NoSQL, que brinda la facilidad de almacenar datos utilizando un estilo de diccionario o un estilo orientado a columnas, o como documentos, objetos, gráficos, tuplas o una combinación de los mismos. Los documentos se almacenan en el formato BSON, que es similar a JSON (MongoDB, 2018).

La elección de la base de datos MongoDB es natural porque la API de Twitter envía objetos del tipo JSON. En esta base de datos se crearon colecciones para almacenar la información obtenida de los módulos de procesamiento de datos. En la figura 6, se muestran los nombres de las colecciones que forman parte de la jerarquía de la base de datos.



**Figura 6.** Diseño de base de datos en MongoDB. Elaborado por: Autor.

En la tabla 2 se detalla la descripción de las colecciones que forman parte de la base de datos Twitterdb.

<b>Colección</b>	<b>Descripción</b>
Twitter_raw	Almacena los datos completos de cada tweet que es enviado a través de la API de Twitter.
Twitter_trends	Almacena la lista de temas de tendencia de Ecuador solicitada cada 30 minutos, con la fecha de adquisición de estos datos.
Twitter_debot	Almacena las cuentas automatizadas encontradas, con el nombre de usuario y la fecha de interacción.
Twitter_analysis	Almacena los tweets de acuerdo a las categorías seleccionadas, con la fecha, usuario y texto del tweet.
Twitter_results	Almacena los índices de influencia y el total de tweets de cada uno de los temas.

**Tabla 2.** Descripción de las colecciones de la base de datos Twitterdb. Elaborado por: Autor.

### **Configuración de parámetros de seguridad**

MongoDB ofrece varias funciones para proteger la base de datos y manejar la información de manera segura. Las características de seguridad son: autenticación, control de acceso y cifrado. Para que la base de datos implementada esté segura y protegida se deben modificar parámetros de autorización y autenticación en el archivo de configuración de MongoDB, el cual se encuentra disponible en el siguiente directorio: /etc/mongod.conf.

### **Parámetros de autorización a la base de datos**

- La autenticación de usuarios

```
security:  
  authorization: "enabled"
```

- Sólo se permiten conexiones desde la IP indicada (en este caso la IP local)
- Se cambia el puerto por defecto al que deseemos

```
net:  
  bindIp: 127.0.0.1  
  port: 27019 (el que se desee)
```

- Se deshabilita cualquier acceso vía http tanto a la parte de administración como a la API Rest.

```
http:  
  enabled: false  
  RESTInterfaceEnabled: false
```

### **Parámetros de autenticación a la base de datos**

Se deben crear usuarios con acceso a la base de datos, y asignar roles a cada usuario dependiendo el propósito.

Estos cambios se deben realizar accediendo a la base de datos de MongoDB

- Crear usuario admin con el rol “root” en la base de datos admin, y colocar una contraseña.
- Crear usuario twitter con el rol “dwOwner” en la base de datos y colocar una contraseña.

### **Módulo de procesamiento de datos**

Este módulo está constituido por una aplicación de clasificación de textos, la cual está desarrollada con un algoritmo de aprendizaje supervisado de *Machine Learning*. Este tipo de aplicaciones son usadas para responder preguntas de manera correcta, enseñándolas previamente mediante un conjunto de datos seleccionados y un modelo matemático de clasificación multi-etiqueta, concepto detallado en marco teórico.

El lenguaje de programación utilizado para procesar los tweets es Python, usando la librería Scikit-learn de Machine Learning (ML), esta librería de Python permite el acceso a varios métodos parametrizables de aprendizaje de máquina que implementan una misma interfaz lo que facilita a un investigador probar diferentes algoritmos con bajo esfuerzo. Haciendo el uso de esta librería se puede entrenar el modelo de clasificación usando los datos conocidos y relacionados a los temas que se desean clasificar; de esta manera predecir la categoría de cada uno de los tweets generados por los usuarios de la red social.

Bastos et al. (2012) dice que los temas más populares en Twitter son los siguientes: política, economía, deportes, entretenimiento, estilo de vida, juegos, tecnología, eventos, cine, celebridades, idiomas y música. En el presente estudio se han seleccionado 4 categorías

de las antes mencionadas que son: política, deportes, economía, entretenimiento, y una categoría adicional llamada otros que abarca las temáticas restantes.

### **Entrenamiento del modelo de clasificación**

Para iniciar con el procedimiento de clasificación de textos es importante recopilar los datos para entrenar el algoritmo matemático, esto se realizó mediante la obtención de noticias de las 5 categorías seleccionadas previamente, tomadas de sitios web de los noticieros más importantes del país con el uso de los canales RSS (*Really Simple Syndication*), que se almacenaron en archivos individuales por cada tema. De esta manera se tomó la información y se procesaron estos textos para entrenar el modelo multi-clasificador.

Previo al entrenamiento, preparamos los datos en función del algoritmo que se vaya a utilizar. En este caso particular los algoritmos a ser utilizados se basan en el concepto de *bag of words*, en dónde los textos que pertenecen a cada categoría tienen que ser transformados a vectores numéricos en función de las características más importantes de cada categoría. Además, dentro de este proceso se tienen que remover aquellas palabras que no agregan significado a una oración o frase conocidas como *stopwords* que en español son: artículos, pronombres, números, preposiciones, conjunciones, símbolos gramaticales, números. Como se había discutido en el marco teórico este proceso genera una matriz de vectores que posteriormente se divide en dos partes: datos para el entrenamiento del modelo y datos para la evaluación o validación del modelo.

El modelo de clasificación LSVM utilizado por van Zoonen y van der Meer (2016) muestra un porcentaje de efectividad del 80% para 3.000 *tweets* recolectados en su investigación. Sin embargo se realizaron múltiples pruebas tomando varios algoritmos de aprendizaje supervisado para comprobar su asertividad en la clasificación de la información en el idioma español y así realizar pruebas de clasificación de *tweets* para obtener el *score* de evaluación

En la tabla 3 se muestran los valores obtenidos del entrenamiento con los 5 algoritmos de clasificación. Para esta prueba se tomaron 6.580 registros por cada categoría, dando un total de 32.900 registros de entrenamiento. Cada uno de los algoritmos ha sido configurado con parámetros estándar para obtener el *score* de evaluación y decidir cuál es el mejor método de clasificación.

Algoritmo	Registros	Score de evaluación (%)	Tiempo de cálculo
K Neighbors Classifier	32.900	65,25	1 hora
Multinomial NB	32.900	82,59	5 minutos
SVM	32.900	86,14	12 horas
LSVM	32.900	84,31	5 minutos
Logistic Regression	32.900	85,29	5 minutos

**Tabla 3.** Scores de entrenamiento de los algoritmos de aprendizaje supervisado. Elaborado por: Autor.

Tomando en cuenta la eficiencia en tiempo de cálculo y *score* de evaluación, los mejores métodos de clasificación son *Multinomial NB*, *LSVM* y *Logistic Regression*.

### Módulo de análisis de cuentas automatizadas

En este módulo se hace el uso de la herramienta de análisis y correlación de cuentas automatizadas DeBot, cuyo motor de búsqueda fue compartido por sus autores Chavoshi, Hamooni y Mueen, en una colaboración para extender el uso de su aplicación. La misma que para el presente trabajo fue adaptada para los temas de tendencia en Ecuador, cuyo funcionamiento se explica en el marco teórico.

DeBot ha sido configurado para recolectar los tweets de los temas de tendencia de Twitter de Ecuador, analizar las cuentas sospechosas y obtener las cuentas automatizadas confirmadas como *bots*; las mismas que serán almacenadas en la base de datos de MongoDB.

Para obtener los resultados más óptimos con la búsqueda y comprobación de las cuentas automatizadas, se realizaron pruebas con el algoritmo de DeBot. Uno de los factores que más influyó fue el tiempo de recolección de datos de los usuarios detectados como cuentas automatizadas y la ejecución del algoritmo de correlación ejecutado en *Matlab*. Este script toma aproximadamente entre 2 a 3 horas para obtener los mejores resultados de las búsquedas de bots. Después de este tiempo empieza nuevamente a recolectar nuevos tweets de la lista de temas de tendencia y los analiza. Una vez obtenidos estos datos de las cuentas detectadas como *bots*, se almacenan en la base de datos, que sirve de consulta para el análisis de índices de influencia en las categorías seleccionadas.

## Módulo de análisis de datos

De la información recolectada, clasificada y almacenada en la base de datos se realiza el análisis de los tweets, mediante una regla de tres se obtiene el índice de influencia contabilizando el número de tweets enviados por *bots* detectados de cada categoría para el total de tweets enviados en la última hora, y de esta manera obtener los índices de influencia de cada categoría. En la ecuación 4, se muestra el cálculo de los índices de influencia en la última hora de recolección de los tweets por cada categoría.

$$\text{índice de influencia} = \frac{\text{Tweets enviados por bots detectados}}{\text{Total de tweets de la última hora}} \times 100 \quad (4)$$

## Módulo de presentación de datos

Para realizar la presentación visual de la información obtenida, procesada y clasificada se desarrolló un prototipo de aplicación web. La herramienta utilizada para este desarrollo es Bokeh, la cual es una librería de visualización para navegadores web, que proporciona gráficos y tablas de alto rendimiento, de manera rápida bajo el lenguaje de programación Python y HTML. De esta manera presentar los resultados obtenidos de los índices de influencia en el Ecuador de forma gráfica.

## **CAPÍTULO IV. RESULTADOS**

### **4.1 Levantamiento de datos**

Para desarrollar el presente proyecto se necesitó recolectar dos clases de datos. La primera clase se trata de la información para entrenar el modelo de clasificación y la segunda clase de datos son los textos que van a ser clasificados en tiempo real, que en este caso se trata de los tweets.

La recolección de los datos de entrenamiento para la clasificación de la información se realizó mediante un programa para guardar las noticias de los principales periódicos del país a través de los canales de RSS. Esta información se acumula en carpetas separadas con el respectivo nombre de las categorías de interés para el desarrollo del proyecto. Los datos a ser utilizados en el clasificador de textos se debe realizar de manera correcta, la recolección de los datos es muy importante para entrenar el modelo de clasificación, porque la calidad y cantidad de los datos determinará directamente cuán bueno es el modelo de predicción.

Esta herramienta está pensada para operar las 24 horas del día los 7 días de la semana, recolectando los tweets para clasificarlos en categorías y para analizar las cuentas con el fin de encontrar usuarios con comportamiento automatizado, de esta manera obtener los índices de influencia de estas cuentas.

#### **4.1.1 Resultados de la clasificación de los datos**

Para verificar la óptima clasificación de datos en tiempo real obtenidos de Twitter, se realizaron dos pruebas con muestras distintas de tweets. La primera prueba se realizó con un total de 7.000 tweets etiquetados por varias personas, que leyeron cada *tweet* y eligieron la categoría de los textos. En la segunda prueba se tomó una muestra de *tweets* por cada categoría. A partir de esta segunda prueba se encontró la forma de clasificar los tweets.

## Primera prueba de clasificación

Para la realización de la primera prueba se usó una base de 7.000 tweets recolectados 6 meses antes y etiquetados por un grupo de 25 personas, los cuales fueron procesados por un clasificador entrenado con noticias actuales.

Los resultados obtenidos no fueron favorables para el objetivo de esta investigación. Las tablas 4, 5, y 6 muestran los resultados de clasificación utilizando LSVM, *Multinomial NB* y *Logistic Regression*, respectivamente. En la tabla 3 se muestra que la categoría de deportes recibe el mejor puntaje de desempeño, lo mismo que pasa que para *Multinomial NB* y *Logistic Regression* (ver tablas 5 y 6). Por lo que se realizó otro grupo de pruebas.

Categoría	Tweets	Clasificados	Desempeño (%)
Política	339	131	38,64
Deportes	797	714	89,59
Economía	3	2	66,67
Entretenimiento	584	263	45,03
Otros	5.277	2.264	42,90

**Tabla 4.** Resultados de las pruebas de clasificación con LSVM. Elaborado por: Autor.

Categoría	Tweets	Clasificados	Desempeño (%)
Política	339	73	21,53
Deportes	797	698	87,58
Economía	3	1	33,33
Entretenimiento	584	223	38,18
Otros	5277	2750	52,11

**Tabla 5.** Resultados de las pruebas de clasificación con Multinomial NB. Elaborado por: Autor.

<b>Categoría</b>	<b>Tweets</b>	<b>Clasificados</b>	<b>Desempeño (%)</b>
<b>Política</b>	339	121	35,69
<b>Deportes</b>	797	562	70,51
<b>Economía</b>	3	1	33,33
<b>Entretenimiento</b>	584	285	48,80
<b>Otros</b>	5.277	2.412	47,71

**Tabla 6.** Resultados de las pruebas de clasificación con Logistic Regression. Elaborado por:

Autor.

### Segunda prueba de clasificación

Para la segunda prueba se tomaron muestras actuales de tweets de cada una de las categorías durante ocho horas. Y de igual manera, se actualizaron los contenidos de los textos de entrenamiento para entrenar a los algoritmos de clasificación con información actual.

En particular, para el nuevo entrenamiento de los clasificadores se recolectó 586 registros por cada categoría, con un total de 2.930 registros para el entrenamiento del clasificador. En la tabla 7 se pueden observar los valores de puntuación de los 3 algoritmos de mejor rendimiento entrenados, siendo el algoritmo *Multinomial NB* el que obtuvo el *score* de evaluación más alto con 75,81%, seguido del LSVM con 72,76% y en tercer lugar el *Logistic Regression* con 72,25%.

<b>Algoritmo</b>	<b>Registros</b>	<b>Score de evaluación (%)</b>	<b>Tiempo aproximado</b>
Multinomial NB	2.930	75,81	< 5 minutos
Logistic Regression	2.930	72,25	< 5 minutos
LSVM	2.930	72,76	< 5 minutos

**Tabla 7.** Scores de entrenamiento de los 3 algoritmos de mejor rendimiento. Elaborado por:

Autor.

En la tabla 8 se muestran los resultados de clasificación de tweets recolectados empleando el algoritmo *Multinomial NB*, se puede notar que la puntuación de desempeño de cada categoría varía, a pesar de ser entrenado con el mismo número de registros. La que mejor puntuación tiene es la categoría de Entretenimiento con un 85,27%.

<b>Categoría</b>	<b>Tweets</b>	<b>Clasificados</b>	<b>Desempeño (%)</b>
<b>Política</b>	1.164	835	71,74
<b>Deportes</b>	669	476	71,15
<b>Economía</b>	162	116	71,61
<b>Entretenimiento</b>	638	544	85,27
<b>Otros</b>	217	162	74,65

**Tabla 8.** Resultados pruebas de clasificación con Multinomial NB. Elaborado por: Autor.

En la tabla 9 se muestran los resultados de clasificación empleando el algoritmo Logistic Regression, se observa en los valores de desempeño de la clasificación que tienen valores bajos a excepción de las categorías de Entretenimiento con 87,72% y Otros con 78,34%.

<b>Categoría</b>	<b>Tweets</b>	<b>Clasificados</b>	<b>Desempeño (%)</b>
<b>Política</b>	1.164	694	59,62
<b>Deportes</b>	669	404	60,47
<b>Economía</b>	162	96	59,25
<b>Entretenimiento</b>	638	551	87,32
<b>Otros</b>	217	170	78,34

**Tabla 9.** Resultados pruebas de clasificación con Logistic Regression. Elaborado por: Autor.

En la tabla 10 se muestran los resultados de clasificación empleando el algoritmo LSVM, se puede observar que en 3 de las 5 categorías tiene valores bajos de desempeño y sólo dos superan el 80%.

<b>Categoría</b>	<b>Tweets</b>	<b>Clasificados</b>	<b>Desempeño (%)</b>
<b>Política</b>	1.164	687	59,02
<b>Deportes</b>	669	407	60,93
<b>Economía</b>	162	102	62,96
<b>Entretenimiento</b>	638	549	87,00
<b>Otros</b>	217	180	82,94

**Tabla 10.** Resultados pruebas de clasificación con LSVM. Elaborado por: Autor.

Los resultados de clasificación de estos 3 algoritmos con datos de entrenamiento tomados el mismo día es muy similar, por esta razón se ha realizado una prueba con tweets de cada una de las categorías y de esta manera conocer su desempeño en la clasificación de textos en español.

De las pruebas realizadas anteriormente se comprueba que el mejor algoritmo para este caso de clasificación de tweets es el *Multinomial NB*. Este algoritmo además de tener un buen desempeño con el entrenamiento de 32.900 registros, también lo hizo con el entrenamiento de menor número de registros (2.930), tomados en las 5 categorías para las pruebas de comprobación. Se puede evidenciar que el modelo de clasificación *Multinomial NB* es el que mejor clasifica conjuntos de datos pequeños.

Debido a que el conjunto de datos entrenamiento es pequeño, los clasificadores de alto sesgo/baja varianza como por ejemplo, *Naïve Bayes*, tienen una ventaja sobre los clasificadores de bajo sesgo/alta varianza como por ejemplo el KNN, ya que este último se sobre ajustará. Sin embargo, los clasificadores de bajo sesgo/alta varianza se desempeñan mejor a medida que su conjunto de datos entrenamiento crece, ya que los clasificadores de alto sesgo no son lo suficientemente potentes como para proporcionar modelos precisos.

#### **4.1.2 Resultados del análisis de cuentas automatizadas**

Como se había mencionado en el apartado anterior se utiliza DeBot para la detección de cuentas automatizadas. Es así como el corazón de DeBot puede determinar los clústers de cuentas de Twitter que comparten el mismo comportamiento en el tiempo, es decir que estas cuentas tienen los enlaces, imágenes o videos en tiempo de publicación semejante. El algoritmo de Debot aproximadamente encuentra entre 6 a 8 *bots* por día, según las pruebas realizadas con el flujo de datos de Twitter en Ecuador. Actualmente se cuenta con una base de datos aproximadamente de 500 cuentas automatizadas detectadas por Debot.

Se puede apreciar en las figuras 7a y 7b el mismo tweet enviado a la misma hora por diferentes usuarios de la red social. Todas las cuentas encontradas como bots son almacenadas en la base de datos de MongoDB.



**Figura 7.** Publicación con el mismo contenido enviado por dos cuentas diferentes en tiempos similares. a) Tweet enviado por la cuenta @eloritocom. b) Tweet enviado por la cuenta @vitotvo. Fuente: Twitter.

## 4.2 Presentación y análisis de resultados

### 4.2.1 Resultados del entrenamiento del algoritmo de clasificación

En base a las pruebas realizadas se determinó que la información de entrenamiento del modelo debe ser alimentada constantemente con noticias actuales. Debido a que cada día hay nuevas noticias en varios temas y con diferente vocabulario, y que posiblemente esas palabras no se encontraban en el *bag of words* de un entrenamiento pasado.

En el segundo tipo de prueba realizada se tomaron tweets de un día durante 8 horas, relacionado con las temáticas seleccionadas anteriormente. Posteriormente, se aplicó el clasificador entrenado con noticias del mismo día, obteniendo los resultados mostrados en la tabla 10, y mejorando notablemente el resultado de clasificación obtenido en el primer tipo de prueba. Por lo que el entrenamiento debe ser constante con información actual, ya que habrá vocabulario que el modelo entrenado no lo tenga, sin embargo en tweets actuales existe, por lo que el clasificador no hará un buen trabajo.

Como ventajas de usar el algoritmo probabilístico Multinomial NB, se tiene su simplicidad y su efectividad. Si la suposición de independencia condicional de Naïve Bayes

realmente se cumple, un clasificador Naïve Bayes convergerá más rápido que los modelos discriminativos como el algoritmo Logistic Regression, por lo que necesita menos datos de entrenamiento. Incluso si la suposición no se cumple, un clasificador de Naïve Bayes todavía hace un gran trabajo en la práctica, como se ha visto en la presente investigación para la clasificación de tweets.

#### 4.2.2. Resultados de clasificación e índices de influencia

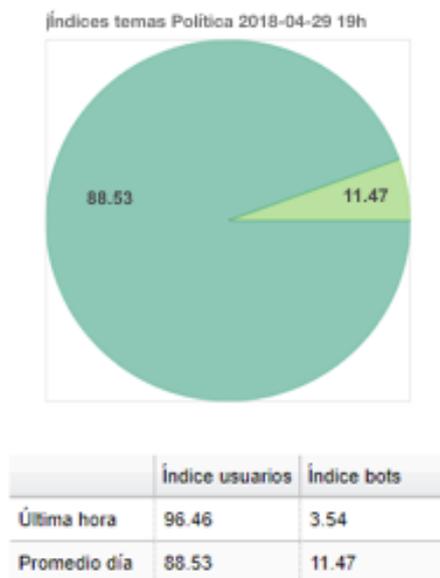
La presentación de la herramienta con interfaz web en su página inicial se visualiza los valores de la información obtenida de Twitter, mostrando los índices de influencia de las tendencias diarias en figuras en forma de pastel. En la figura 8 se muestra una captura de pantalla de la herramienta la cual se puede acceder mediante una IP generada en el explorador web. Los datos mostrados en las siguientes imágenes son tomados el jueves 12 de Abril del 2018, aproximadamente a las 19h15.



**Figura 8.** Página de inicio de la interfaz gráfica de la herramienta de minería de texto.

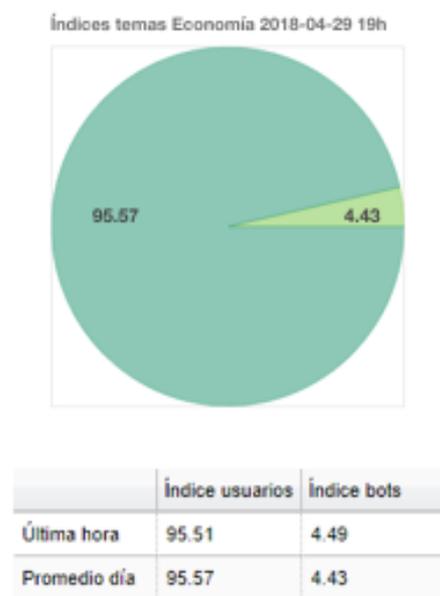
Elaborado por: Autor.

En la figura 9 se observa el resultado promedio obtenido de las 18 primeras horas del día 12 de Abril, con un índice de influencia de cuentas automatizadas del 11,47% en política, y un 88,53% de tweets generados por el resto de usuarios.



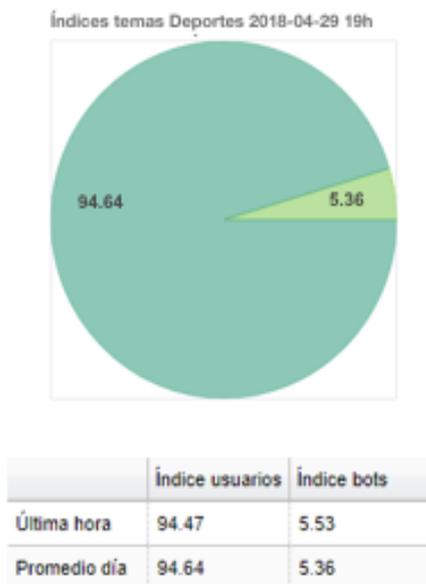
**Figura 9.** Valores de índice de influencia última hora y promedio en política. Elaborado por: Autor.

En la figura 10 se observa el resultado promedio obtenido de las 18 primeras horas del día 12 de Abril, con un índice de influencia de cuentas automatizadas del 4,43% en economía, y un 95,57% de tweets generados por el resto de usuarios.



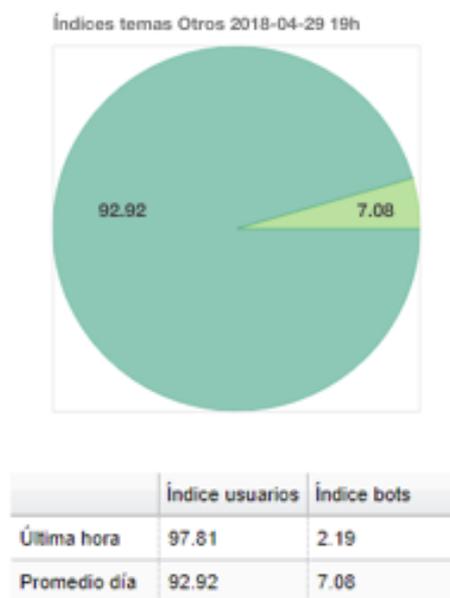
**Figura 10.** Valores de índice de influencia última hora y promedio en economía. Elaborado por: Autor.

En la figura 11 se observa el resultado promedio obtenido de las 18 primeras horas del día 12 de Abril, con un índice de influencia de cuentas automatizadas del 5,36% en deportes, y un 94,64% de tweets generados por el resto de usuarios.



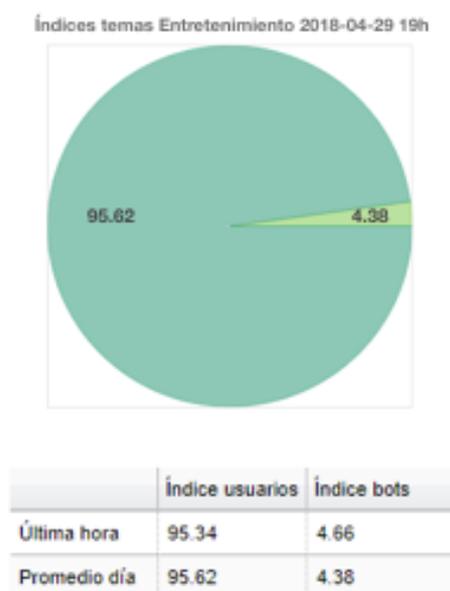
**Figura 11.** Valores de índice de influencia última hora y promedio en deportes. Elaborado por: Autor.

En la figura 12 se observa el resultado promedio obtenido de las 18 primeras horas del día 12 de Abril, con un índice de influencia de cuentas automatizadas del 71,08% en política, y un 88,53% de tweets generados por el resto de usuarios.



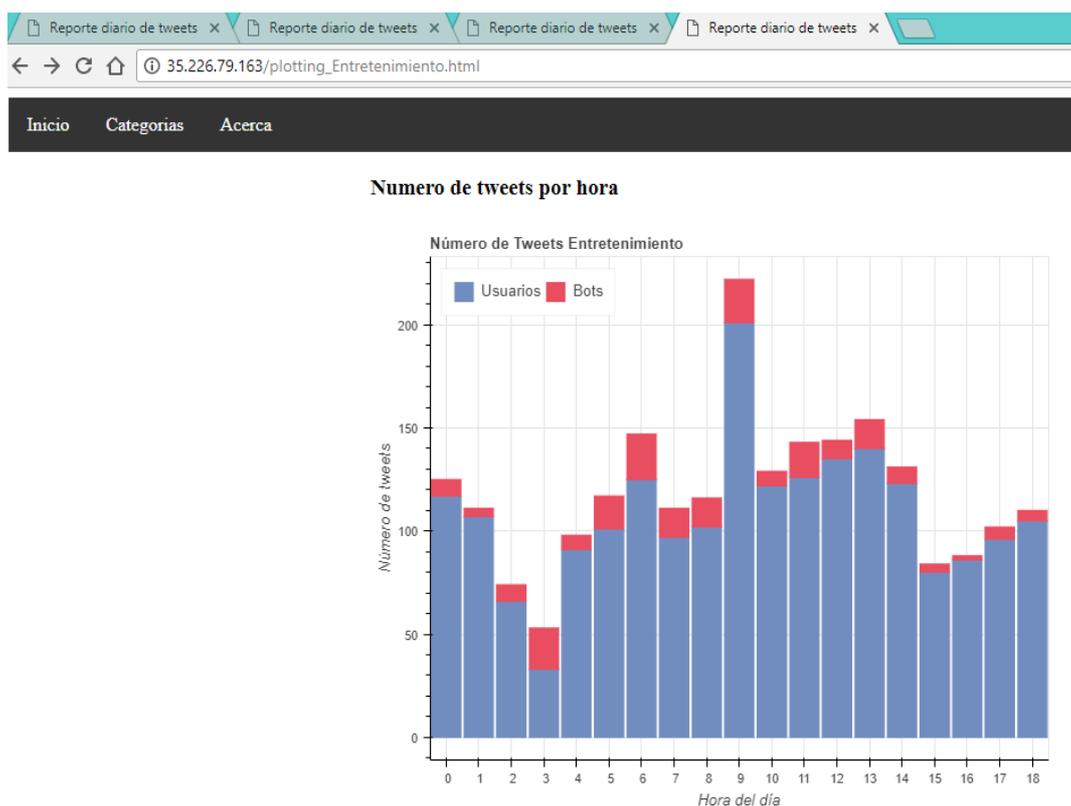
**Figura 12.** Valores de índice de influencia última hora y promedio en otros. Elaborado por:  
Autor.

En la figura 13 se observa el resultado promedio obtenido de las 18 primeras horas del día 12 de Abril, con un índice de influencia de cuentas automatizadas del 4,38% en política, y un 95,62% de tweets generados por el resto de usuarios.



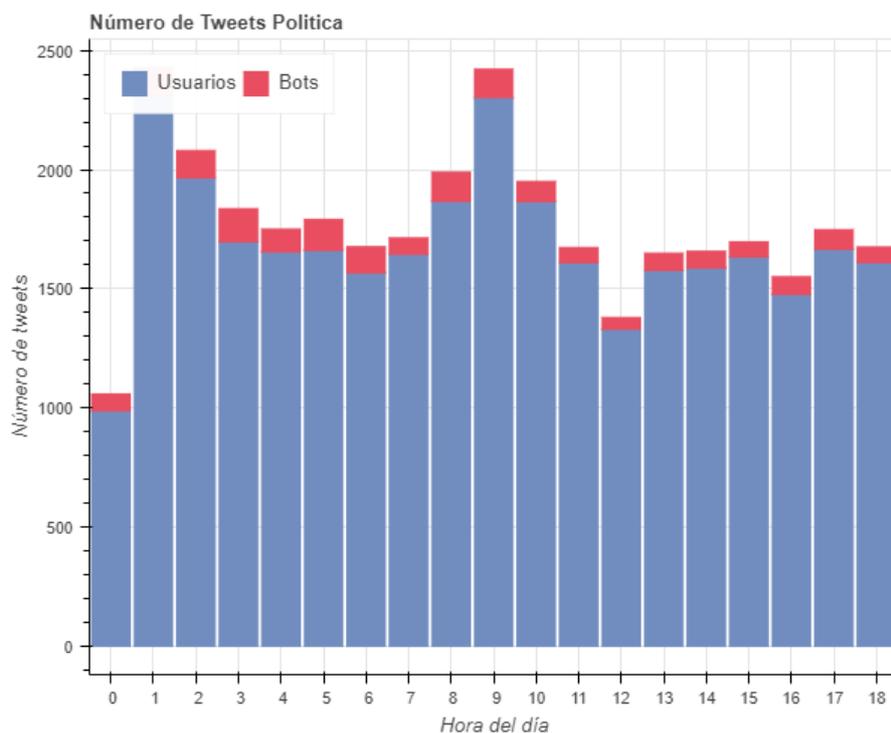
**Figura 13.** Valores de índice de influencia última hora y promedio en entretenimiento.  
Elaborado por: Autor.

A continuación se muestran los resultados obtenidos del análisis de tweets en cada hora del día mediante gráficos de barras y tablas de valores con el número de tweets y los índices de influencia. En la figura 14, se observa la página de presentación de cada una de las categorías, mismas que se seleccionan mediante el menú categorías, cada una de las 5 categorías actualizan sus valores cada hora de recolección y el resultado del análisis de esos tweets.



**Figura 14.** Página de número de tweets por cada hora. Elaborado por: Autor.

En la figura 15 se muestra el gráfico de barras de la categoría política, representando con color azul el número de tweets enviados por usuarios comunes, y de color rojo los tweets de usuarios detectados por DeBot como cuentas automatizadas. En la categoría política en las primeras 18 horas del día 12 de Abril se observa una fluctuación del número de tweets generados por usuarios comunes, y de la misma forma el número de tweets detectados como cuentas automatizadas. Dando porcentajes de influencia diferentes durante el transcurso del día. Se puede observar en la figura 15 a las 9 am existe un mayor número de tweets, que pudo ser causado por un tema de tendencia. De la misma manera en la tabla 11, se muestran los valores detallados del número de tweets de usuarios, número de tweets de bots y el nivel de influencia de los valores en cada una de las horas.

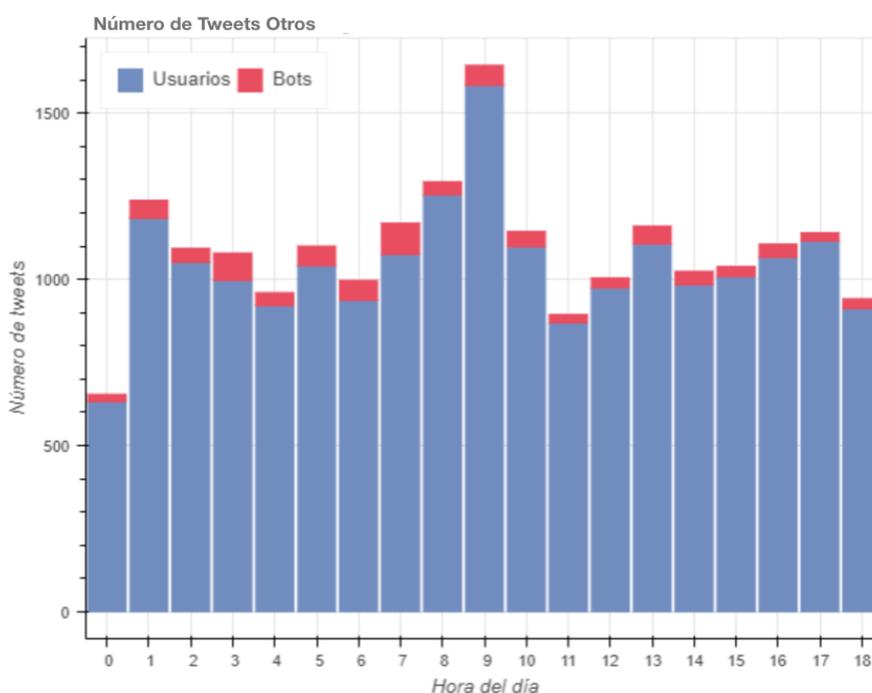


**Figura 15.** Número de Tweets de Política. Elaborado por: Autor.

#	Hora	Número de tweets por Usuarios	Número de tweets por Bots	Índice de influencia (%)
0	0	987	71	6.71
1	1	2311	116	4.78
2	2	1965	115	5.53
3	3	1697	139	7.57
4	4	1655	95	5.43
5	5	1661	130	7.26
6	6	1567	109	6.5
7	7	1644	69	4.03
8	8	1867	122	6.13
9	9	2302	119	4.92
10	10	1866	84	4.31
11	11	1609	64	3.83
12	12	1330	49	3.55
13	13	1577	72	4.37
14	14	1587	70	4.22
15	15	1633	64	3.77
16	16	1476	74	4.77
17	17	1666	82	4.69
18	18	1610	65	3.88

**Tabla 11.** Valores de tweets e índices de influencia por hora (Política). Elaborado por: Autor.

En la figura 16, se muestra el gráfico de barras de la categoría Otros, representando con color azul el número de tweets enviados por usuarios comunes, y de color rojo los tweets de usuarios detectados por DeBot como cuentas automatizadas. En la categoría Otros en las primeras 18 horas del día 12 de Abril se observa una fluctuación del número de tweets categorizados de este tema, y de la misma forma el número de tweets detectados como cuentas automatizadas. Dando porcentajes de influencia diferentes durante el día. Se puede observar en la figura 16 a las 9 horas existe un mayor número de tweets, que pudo ser causado por un tema de tendencia. De la misma manera en la tabla 12, se muestran los valores detallados del número de tweets de usuarios, número de tweets de bots y el nivel de influencia de los valores en cada una de las horas.

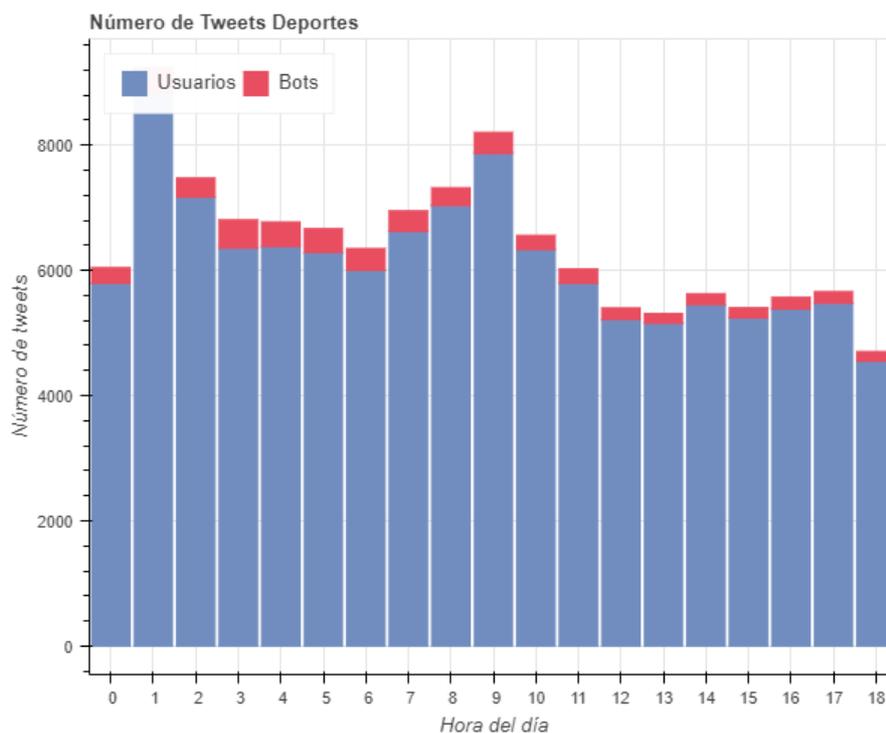


**Figura 16.** Número de tweets de Otros. Elaborado por: Autor.

#	Hora	Número de tweets por Usuarios	Número de tweets por Bots	Índice de influencia (%)
0	0	632	22	3.36
1	1	1184	54	4.36
2	2	1051	42	3.84
3	3	997	82	7.6
4	4	921	39	4.06
5	5	1041	59	5.36
6	6	937	60	6.02
7	7	1075	94	8.04
8	8	1254	40	3.09
9	9	1584	60	3.65
10	10	1098	46	4.02
11	11	869	25	2.8
12	12	974	30	2.99
13	13	1106	54	4.66
14	14	984	40	3.91
15	15	1008	31	2.98
16	16	1066	40	3.62
17	17	1115	25	2.19
18	18	912	30	3.18

**Tabla 12.** Valores de tweets e índices de influencia por hora (Otros). Elaborado por: Autor.

En la figura 17 se muestra el gráfico de barras de la categoría deportes, representando con color azul el número de tweets enviados por usuarios comunes, y de color rojo los tweets de usuarios detectados por DeBot como cuentas automatizadas. En la categoría Otros en las primeras 18 horas del día 12 de Abril se observa una fluctuación del número de tweets categorizados de este tema, y de la misma forma el número de tweets detectados como cuentas automatizadas. Dando porcentajes de influencia diferentes durante el día. Se puede observar en la figura que en horas de la madrugada hubo un flujo alto de tweets cuyo pico máximo se da a las 9 am, que pudo ser causado por un tema de tendencia. De la misma manera en la tabla 13, se muestran los valores detallados del número de tweets de usuarios, número de tweets de bots y el nivel de influencia de los valores en cada una de las horas.

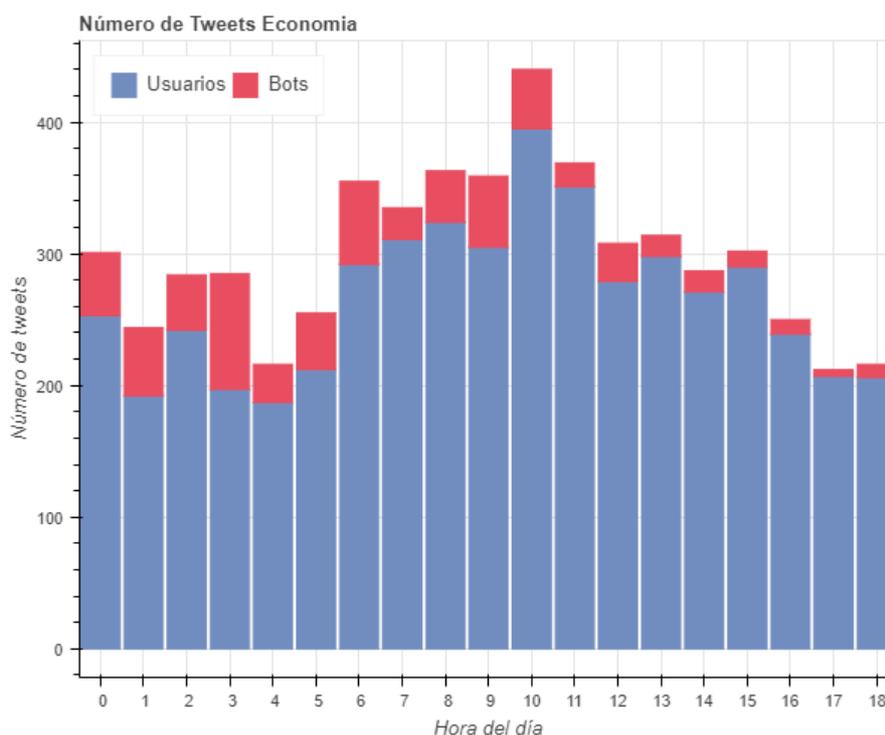


**Figura 17.** Número de tweets de Deportes. Elaborado por: Autor.

#	Hora	Número de tweets por Usuarios	Número de tweets por Bots	Índice de influencia (%)
0	0	5794	249	4.12
1	1	8843	394	4.27
2	2	7170	306	4.09
3	3	6355	452	6.64
4	4	6380	394	5.82
5	5	6285	381	5.72
6	6	6002	347	5.47
7	7	6624	327	4.7
8	8	7037	284	3.88
9	9	7868	337	4.11
10	10	6331	229	3.49
11	11	5793	230	3.82
12	12	5213	187	3.46
13	13	5155	156	2.94
14	14	5450	177	3.15
15	15	5242	162	3
16	16	5383	186	3.34
17	17	5476	184	3.25
18	18	4548	155	3.3

**Tabla 13.** Valores de tweets e índices de influencia por hora (Deportes). Elaborado por: Autor.

En la figura 18 se muestra el gráfico de barras de la categoría economía, representando con color azul el número de tweets enviados por usuarios comunes, y de color rojo los tweets de usuarios detectados por DeBot como cuentas automatizadas. En esta categoría las primeras 18 horas del día 12 de Abril se observa una fluctuación del número de tweets categorizados de este tema, y de la misma forma el número de tweets detectados como cuentas automatizadas. Dando porcentajes de influencia diferentes durante el día, en esta categoría en particular que se observa que en las primeras horas del día hay un índice de cuentas automatizadas mucho mayor y disminuye al transcurrir el día. Se puede observar que a las 10 am existe un mayor número de tweets, que pudo ser causado por un tema de tendencia. De la misma manera en la tabla 14, se muestran los valores detallados del número de tweets de usuarios, número de tweets de bots y el nivel de influencia de los valores en cada una de las horas.

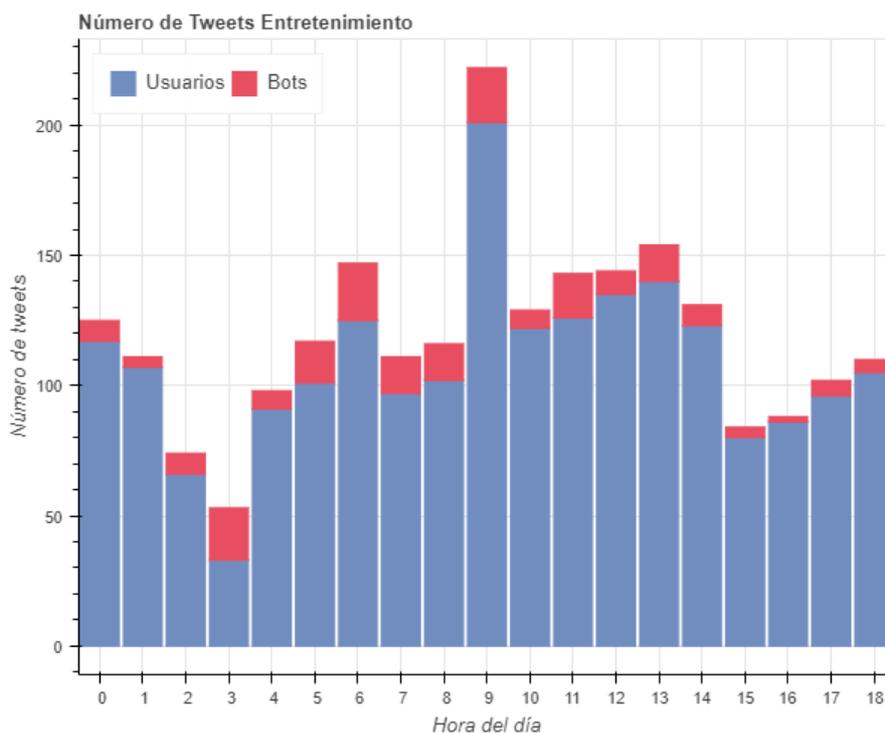


**Figura 18.** Número de tweets de Economía. Elaborado por: Autor.

#	Hora	Número de tweets por Usuarios	Número de tweets por Bots	Índice de influencia (%)
0	0	253	48	15.95
1	1	192	52	21.31
2	2	242	42	14.79
3	3	197	88	30.88
4	4	187	29	13.43
5	5	212	43	16.86
6	6	292	63	17.75
7	7	311	24	7.16
8	8	324	39	10.74
9	9	305	54	15.04
10	10	395	45	10.23
11	11	351	18	4.88
12	12	279	29	9.42
13	13	298	16	5.1
14	14	271	16	5.57
15	15	290	12	3.97
16	16	239	11	4.4
17	17	207	5	2.36
18	18	206	10	4.63

**Tabla 14.** Valores de tweets e índices de influencia por hora (Economía). Elaborado por:  
Autor.

En la figura 19 se muestra el gráfico de barras de la categoría deportes, representando con color azul el número de tweets enviados por usuarios comunes, y de color rojo los tweets de usuarios detectados por DeBot como cuentas automatizadas. En la categoría Entretenimiento en las primeras 18 horas del día 12 de Abril se observa una fluctuación pronunciada del número de tweets categorizados de este tema, y de la misma forma el número de tweets detectados como cuentas automatizadas. Dando porcentajes de influencia diferentes durante el día. En esta categoría a las 9 am hay un pico de tendencia con un número mayor de tweets e índice de influencia. De la misma manera en la tabla 15, se muestran los valores detallados del número de tweets de usuarios, número de tweets de bots y el nivel de influencia de los valores en cada una de las horas.



**Figura 19.** Número de Tweets de Entretenimiento. Elaborado por: Autor.

#	Hora	Número de tweets por Usuarios	Número de tweets por Bots	Índice de influencia (%)
0	0	117	8	5.76
1	1	107	4	2.48
2	2	66	8	7.41
3	3	33	20	17.39
4	4	91	7	5.38
5	5	101	16	10
6	6	125	22	11.89
7	7	97	14	7.33
8	8	102	14	9.86
9	9	201	21	8.05
10	10	122	7	4.17
11	11	126	17	11.26
12	12	135	9	5.45
13	13	140	14	7.22
14	14	123	8	4.91
15	15	80	4	3.6
16	16	86	2	1.59
17	17	96	6	4.96
18	18	105	5	3.7

**Tabla 15.** Valores de tweets e índices de influencia por hora (entretenimiento). Elaborado por: Autor.

### **4.3 Aplicación práctica**

La investigación realizada en el presente proyecto de maestría puede ser aplicada a varios ámbitos de interés público, ya sea social, empresarial, gubernamental o de investigación, para la clasificación o filtrado de información recolectada no solo de la red social Twitter, sino también de otras fuentes de información que se encuentran en internet, cuya clasificación de la información fue realizada por un algoritmo de aprendizaje supervisado, previamente entrenado con noticias actuales, de manera que la clasificación de los temas tenga mejores resultados.

Esta propuesta de herramienta se presenta de manera online para que se pueda acceder desde cualquier lugar mediante la web, presentando las estadísticas de los temas de tendencia en Twitter y los índices de influencia de cuentas automatizadas.

## CAPÍTULO V. DISCUSIÓN

### 5.1 Conclusiones

Luego del análisis del estado del arte se seleccionaron los algoritmos de clasificación: LSVM, SVM, KNN, *Multinomial Naïve Bayes* y *Logistic Regression*; para probarlos en esta investigación. Los algoritmos fueron entrenados utilizando textos en idioma español superando el 70% de precisión en su etapa de entrenamiento. El modelo de clasificación de mejor rendimiento fue *Multinomial NB* debido a sus características de entrenamiento con textos cortos con un *score* de 82,59%.

Se diseñó y desarrolló un prototipo de una aplicación basada en una arquitectura modular para la obtención de los datos, almacenamiento, procesamiento, y la categorización de los Tweets, utilizando herramientas de desarrollo modernas en la construcción de aplicaciones de minería de datos que facilita y optimizan el prototipado rápido con el fin de tener un óptimo desempeño de la aplicación

Para la detección de bots se utilizó DeBot, aprovechando el apoyo de los autores de la aplicación y su interés de extender su aplicación. Para un óptimo funcionamiento con la solución desarrollada se adaptó y configuró parámetros de la herramienta en los temas de tendencia de Twitter en Ecuador, encontrando cuentas como @vitotvo y @eloritocom entre otras, que generan contenido de forma automática en la red social.

La adaptación del buscador de cuentas automatizadas DeBot fue satisfactoria, lo que favoreció la búsqueda de cuentas que publican la misma información en tiempos muy similares, comprobando su efectividad en la correlación de cuentas automatizadas permitiendo construir una base de datos de estas cuentas para Ecuador con alrededor de 500 cuentas, y un estimado de 6 a 8 cuentas detectadas por día.

Del análisis realizado de los índices de intervención de cuentas automatizadas en la red social, se concluye que estos valores cambian por diversos factores, entre ellos están la época del año, la situación política del país o los eventos públicos a llevarse a cabo, entre otros, en general hechos que sean populares y que demandan la atención de un gran público.

En los días que existe campeonato nacional o partidos internacionales, la categoría deportes inunda las red de mensajes sobre estos acontecimientos, los cuales se convierten en tendencia y se recopilan más tweets de temas relacionados con deportes. En el resto de categorías existen días que hay una escasa publicación tweets, por ejemplo en economía o en política si existe algún acontecimiento en específico y se genera una tendencia entre la gente se tendrán muchos más tweets, o si existe un evento artístico o concierto de las misma manera esos temas se hacen tendencia en las redes sociales locales.

El modelo de clasificación fue entrenado basado en información de noticias actuales a través de un canal RSS Feed gratuito proporcionado por noticieros online locales, que si bien nos enfrentamos a problemas propios de no entrenar con datos similares a los que posteriormente se clasifica, los resultados obtenidos muestran se compensan por la variedad de contenidos agregados al modelo de clasificación automática.

Los resultados de clasificación de tweets con datos de entrenamiento tomados el mismo día son mucho mejores, debido al vocabulario que se utiliza o a temas nuevos no existieron para el entrenamiento del modelo de clasificación. Al realizar la prueba con tweets de cada una de las categorías se encontró un mejor desempeño en la clasificación de los tweets, dando como precisión un 75,81% para Multinomial Naïve Bayes, 72,25% para Logistic Regression y un 72,76% para LSVM. Valores superiores a los alcanzados por Vilares (2014) en su estudio de multi-clasificación de textos en español. Siendo el idioma español un lenguaje difícil de procesar para este tipo de herramientas.

El número de datos para el entrenamiento del algoritmo de clasificación es importante para conseguir una mejor clasificación de los textos de forma automática. Fue importante usar una muestra de datos de cada categoría para que mejoren los resultados de los clasificadores, subiendo su porcentaje de clasificación en algunos casos hasta un 30% (*Multinomial Naïve Bayes*).

El algoritmo *Multinomial Naïve Bayes* no sólo tuvo un buen desempeño con el entrenamiento de 32.900 registros, sino que también lo hizo con el entrenamiento de registros recolectados en un día, con esto se evidenció que es el mejor método de clasificación para tweets entrenado con un conjunto pequeño de datos.

Se implementó un prototipo de interfaz web para la visualización del contenido obtenido y el análisis de los índices de influencia mediante gráficas y tablas demostrativas, al ser una herramienta web se puede acceder por una IP a través del internet las 24 horas del día.

## 5.2 Recomendaciones y trabajo futuro

El proceso de aprendizaje supervisado es muy dependiente de la calidad de los datos de entrenamiento. Tomando en cuenta que en Twitter se comparten mensajes en un contexto que evoluciona según la dinámica social, se recomienda utilizar una fuente de datos actualizada o que se actualice constantemente estableciendo un proceso que permita pre-procesar adecuadamente estos datos para el entrenamiento del modelo de clasificación y así mejorar la categorización de los temas utilizando algoritmos de *Deep Learning* o inteligencia artificial.

Los datos para el entrenamiento del modelo pueden ser tomados de otras fuentes de información que proporcionen un vocabulario actualizado y acorde a los temas de tendencia de la red social, esta base de información puede ser alimentada por tweets recopilados por cada categoría.

Se puede trabajar con del modelo de *bag of words* para una mejora en el entrenamiento del algoritmo, realizando un código de adquisición de datos con la misma cantidad de tamaño para cada una de las categorías a entrenar. Este modelo no asume nada de los datos, ni de su estructura, además es bastante funcional para cualquier lenguaje y no requiere un análisis semántico del lenguaje. Adicionalmente, existen otros algoritmos que intentan armar árboles semánticos para definir el contenido de textos pero son más elaborados y sufren mucho en contextos como las redes sociales.

A partir del prototipo de herramienta desarrollada en este proyecto se puede ampliar y añadir más detalles a la presentación de los datos, como la visualización con otros colores, los tamaños de los textos, los campos de las tablas de resultados, agregar un acceso al listado de *bots*, añadir un acceso para permitir denunciar a los *bots*, e incluir una opción para generar reportes.

Se recomienda realizar más pruebas en días específicos como elecciones o eventos deportivos para verificar empíricamente su funcionamiento.

Se recomienda probar otros tipos de algoritmos como el *Reinforcement Learning*, el cual es un método de aprendizaje no supervisado que no tiene un agente de entrenamiento sino que su aprendizaje es mediante su entorno, de esta manera aprende continuamente en base a recompensas e interacción con los textos de cada día.

## LISTA DE REFERENCIAS

- Aggarwal, C. C. (2015). *Data mining: the textbook*. Springer.
- Alhojely, S. (2016). Sentiment Analysis and Opinion Mining: A Survey. *International Journal of Computer Applications*, 150(6).
- Bastos, M. T., Travitzki, R., & Puschmann, C. (2012, May). What sticks with whom? Twitter follower-followee networks and news classification. In *The Potential of Social Media Tools and Data for Journalists in the News Media Industry-Sixth International AAAI Conference on Weblogs and Social Media*, Dublin, Ireland.
- Batenburg, A., & Bartels, J. (2017). Keeping up online appearances: How self-disclosure on Facebook affects perceived respect and likability in the professional context. *Computers in Human Behavior*, 74, 265-276.
- Bayart, J. F. (2014). Retour sur les Printemps arabes. *Politique africaine*, (1), 153-175.
- Bijalwan, V., Kumar, V., Kumari, P., & Pascual, J. (2014). KNN based machine learning approach for text and document mining. *International Journal of Database Theory and Application*, 7(1), 61-70.
- Bradshaw, S., & Howard, P. N. (2017) *Troops, Trolls and Troublemakers: A Global Inventory of Organized Social Media Manipulation*.
- Castelló, J. S., & Santamaría, F. P. (2015). Desarrollo de un sistema de análisis de sentimiento sobre Twitter.
- Chavoshi, N., Hamooni, H., & Mueen, A. (2016, December). DeBot: Twitter Bot Detection via Warped Correlation. In *ICDM* (pp. 817-822).
- Chavoshi, N., Hamooni, H., & Mueen, A. (2017, April). On-Demand Bot Detection and Archival System. In *Proceedings of the 26th International Conference on World Wide Web Companion* (pp. 183-187). International World Wide Web Conferences Steering Committee.
- Davis, C. A., Varol, O., Ferrara, E., Flammini, A., & Menczer, F. (2016, April). Botornot: A system to evaluate social bots. In *Proceedings of the 25th International Conference Companion on World Wide Web* (pp. 273-274). International World Wide Web Conferences Steering Committee.

- Declaración Universal de Derechos Humanos. Asamblea General de las Naciones Unidas, 1948.
- Dickerson, J. P., Kagan, V., & Subrahmanian, V. S. (2014, August). Using sentiment to detect bots on twitter: Are humans more opinionated than bots?. In *Advances in Social Networks Analysis and Mining (ASONAM), 2014 IEEE/ACM International Conference on* (pp. 620-627). IEEE.
- Divya, G., & Suresh, R. (2016). Sentiment Analysis and Sentiment Classification using NLP.
- Duarte, J. I., Rodríguez, G. E., Lares, J., & Sosa, J. R. (2017). Venezolanos en Twitter: ¿Humanos, Bots o Ciborgs? Modelo de Clasificación.(P. 47-59). Tekhné, 1(19).
- EL PAIS (2017). Recuperado de: [https://twitter.com/el\\_pais](https://twitter.com/el_pais)
- Fandango, A. (2017). Python Data Analysis.
- Ferrara, E. (2017). Disinformation and social bot operations in the run up to the 2017 French presidential election.
- Folgado, Á. L. O., la Madrid, Ó. E. P., Rastrollo, E. V., Esteban, A. D., & Gilmartín, V. F. ESTUDIO DE TENDENCIAS DIARIAS EN TWITTER.
- Gao, P., Wang, B., Gong, N. Z., Kulkarni, S. R., Thomas, K., & Mittal, P. (2018). SybilFuse: Combining Local Attributes with Global Structure to Perform Robust Sybil Detection. arXiv preprint arXiv:1803.06772.
- Huang, G., Song, S., Gupta, J. N., & Wu, C. (2014). Semi-supervised and unsupervised extreme learning machines. *IEEE transactions on cybernetics*, 44(12), 2405-2417.
- IAB Ecuador. (2018). Estudio Digital 2017. Recuperado de: [http://iabecuador.com/doc/EstudioDigital2017\\_IAB.pdf](http://iabecuador.com/doc/EstudioDigital2017_IAB.pdf)
- Kelly, Heather. (2012). 83 million Facebook accounts are fakes and dupes. CNN. Recuperado de: <https://edition.cnn.com/2012/08/02/tech/social-media/facebook-fake-accounts/index.html>
- Khan, F. H., Bashir, S., & Qamar, U. (2014). TOM: Twitter opinion mining framework using hybrid classification scheme. *Decision Support Systems*, 57, 245-257.
- Kušen, E., & Strembeck, M. (2018). Why so Emotional? An Analysis of Emotional Bot-generated Content on Twitter.
- Lee, K., Palsetia, D., Narayanan, R., Patwary, M. M. A., Agrawal, A., & Choudhary, A. (2011, December). Twitter trending topic classification. In *Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on* (pp. 251-258). IEEE.
- Lesaca, J (2017). *La trama rusa empleó redes chavistas para agravar la crisis catalana: de noviembre del 2017.* Recuperado

de: [https://politica.elpais.com/politica/2017/11/10/actualidad/1510341089\\_316043.htm](https://politica.elpais.com/politica/2017/11/10/actualidad/1510341089_316043.htm)

1

Ley Orgánica De Comunicación. Registro Oficial Órgano del Gobierno del Ecuador, Quito, Ecuador, 25 de Junio del 2013.

Lobo, S. (2017). Cómo influyen las redes sociales en las elecciones. *Nueva sociedad*, (269), 40-44.

Lowry, P. B., Zhang, J., Wang, C., & Siponen, M. (2016). Why do adults engage in cyberbullying on social media? An integration of online disinhibition and deindividuation effects with the social structure and social learning model. *Information Systems Research*, 27(4), 962-986.

Marsland, S. (2015). *Machine learning: an algorithmic perspective*. CRC press.

Mateo Fernández, J. (2016). Análisis de contenidos en Social Media: Clasificación de mensajes e identificación de influyentes en el Banco Central Europeo (BCE).

McCANN, C. (2011). Arab Spring. *The New York Times*. Recuperado de: <https://www.nytimes.com/2011/12/25/opinion/sunday/arab-spring.html>

Microsoft Azure (2016). Machine Learning: What is machine learning?. Recuperado de: <https://docs.microsoft.com/es-es/azure/machine-learning/machine-learning-what-is-machine-learning>

MongoDB, Inc. (2017). What Is MongoDB? . Recuperado de: <https://www.mongodb.com/>

Patil, S. A., & Manwade, K. B. (2016). Opinion Classification from Online Reviews based on Support Vector Machine.

Pedregosa, F., Abraham, A., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., ... & Varoquaux, G. (2014). Machine learning for neuroimaging with scikit-learn. *Frontiers in neuroinformatics*, 8, 14.

Polonski, V. (2016). Impact of social media on the outcome of the EU referendum. EU Referendum Analysis 2016. Recuperado de: <https://www.referendumanalysis.eu/eu-referendum-analysis-2016/section-7-social-media/impact-of-social-media-on-the-outcome-of-the-eu-referendum/>

Reporteros sin fronteras (2016). Nuevos medios: entre revolución y represión, la solidaridad en la Red frente a la censura. Recuperado de: <https://rsf.org/es/noticias/nuevos-medios-entre-revolucion-y-represion-la-solidaridad-en-la-red-frente-la-censura>

Resolución 59. Asamblea General de las Naciones Unidas, 1946.

Rodríguez, G. P. P., Gómez, B. M., & Altamirano, K. A. A. (2017). Análisis Geográfico de la Red Social Twitter en Ecuador. *Revista Publicando*, 4(11 (1)), 513-525.

- RT en español. (2017). *Sobre los Bots del El País, el referendun de Cataluña y la máquinade injerencias rusa*: 23 de septiembre del 2017. Recuperado de: <https://actualidad.rt.com/actualidad/250999-pais-referendum-cataluna-maquina-injerencias>
- Savage, D., Zhang, X., Yu, X., Chou, P., & Wang, Q. (2014). Anomaly detection in online social networks. *Social Networks*, 39, 62-70.
- Shannon et al. (2016). 6 ways social media is changing the world. *World Economic Forum*. Recuperado de: <https://www.weforum.org/agenda/2016/04/6-ways-social-media-is-changing-the-world/>
- Soengas, X. (2013). El papel de Internet y de las redes sociales en las revueltas árabes: una alternativa a la censura de la prensa oficial. *Comunicar*, 21(41).
- Twitter, Inc. (2017). Twitter Developer Platform - Twitter Developers. Recuperado de: <https://dev.twitter.com/>
- TwitterAudit. (2017). Recuperado de: <https://www.twitteraudit.com/assange>
- Van Zoonen, W., & Toni, G. L. A. (2016). Social media research: The application of supervised machine learning in organizational communication research. *Computers in Human Behavior*, 63, 132-141.
- Vilares Calvo, D. (2014). Análisis de contenidos en Twitter: clasificación de mensajes e identificación de la tendencia política de los usuarios.
- Villena-Román, J., & Garcia-Morera, J. (2013). TASS 2013—workshop on sentiment analysis at SEPLN 2013: An overview. In *Proceedings of the TASS workshop at SEPLN* (pp. 112-125).
- Von Hoffman, Constantine. (2013). Fake tweets just one way hackers can disrupt stocks. CNN. Recuperado de: <https://www.cbsnews.com/news/fake-tweets-just-one-way-hackers-can-disrupt-stocks/>
- WeAreSocial. (2018). Digital In 2018: World's Internet Users Pass The 4 Billion Mark. Recuperado de: <https://wearesocial.com/blog/2018/01/global-digital-report-2018>
- Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.
- Zhao, L., Huang, M., Yao, Z., Su, R., Jiang, Y., & Zhu, X. (2016, February). Semi-Supervised Multinomial Naive Bayes for Text Classification by Leveraging Word-Level Statistical Constraint. In *AAAI* (pp. 2877-2884).

## ANEXOS

### Descripción de tecnologías empleadas

#### Python

Python es un lenguaje de programación ampliamente usado para el análisis de datos, ocupada para manipulación, procesamiento, limpieza y procesamiento de datos. Python se ha convertido en uno de los lenguajes y plataformas estándar de facto para el análisis de datos y la ciencia de datos. El siguiente cuadro muestra algunas de las numerosas bibliotecas disponibles en el ecosistema de Python que utilizan los analistas de datos (Fandango, 2017).

#### Tweepy

Tweepy es una biblioteca de código abierto que proporciona acceso a la API de Twitter para Python. Con tweepy, es posible utilizar cualquier método que ofrece la API oficial de Twitter. Mediante esta librería y la autenticación con el método OAuth, Twitter admite el acceso a su información, que ahora es la única forma de utilizar la API de Twitter. De esta manera no se muestra la contraseña de usuario, por lo que es más segura. Por lo que es más fácil administrar los permisos, por ejemplo, se puede generar un conjunto de tokens y llaves que solo permite leer desde las líneas de tiempo, por lo que en caso de que alguien obtenga esas credenciales, no podrá escribir o enviar mensajes directos, minimizando el riesgo. La aplicación no responde con una contraseña, por lo que incluso si el usuario la cambia, la aplicación seguirá funcionando (Tweepy, 2017).

#### Tweepy Streaming API

Uno de los principales casos de uso de tweepy es monitorear tweets y realizar ciertas acciones cuando ocurre algún evento. El componente clave de eso es el objeto StreamListener, que supervisa y captura los tweets en tiempo real.

StreamListener tiene varios métodos, siendo `on_data()` y `on_status()` los más útiles. Tweepy es una de las mejores librerías de Twitter para Python, especialmente cuando se considera la compatibilidad con la API de *Streaming* de Twitter. Existen otras librerías como `python-twitter` que también ofrecen muchas funciones, pero la librería `tweepy` tiene la comunidad más activa y que aporta con el código al desarrollo y mejoramiento de la librería.

## **Scikit-Learn**

Scikit-Learn es un módulo de aprendizaje general para el lenguaje de programación Python. Este módulo de Python integra una amplia gama de algoritmos de aprendizaje de máquina de última generación y proporciona una biblioteca de lenguaje de programación Python aplicado por los usuarios para implementar algoritmos de aprendizaje supervisados y no supervisados.

A medida que Scikit-learn expone una amplia variedad de algoritmos de aprendizaje automático, esto permite una fácil comparación de métodos para una aplicación dada. Los algoritmos, implementados en un lenguaje de alto nivel, se usan como bloques de construcción para enfoques específicos de casos, por ejemplo, uso de medios relacionados con el trabajo (Pedregosa et al., 2014).

## Glosario de términos

**Bots:** Es una aplicación informática que ejecuta tareas automatizadas.

**Tweet:** Es un mensaje que enviado a través de Twitter

**Machine Learning:** Es un algoritmo que revisa y aprende datos siendo capaz de predecir comportamientos futuros.

**RSS:** Son las siglas de Really Simple Syndication, un formato que cumple con el estándar XML para compartir contenido en la web.

**Spam:** Es la denominación del correo electrónico no solicitado que recibe una persona. Dichos mensajes, también llamados correo no deseado o correo basura.

**Trending topics:** Son las palabras clave más utilizadas en un plazo de tiempo concreto en Twitter. Se trata de keywords de moda (también llamados hashtags en Twitter), aquello que es tendencia y de lo que más se habla en ese momento en esta red social.

**Data set:** Es una representación de datos residente en memoria que proporciona una modelo de programación relacional coherente independientemente del origen de datos que contiene.

**Cyberbullying:** El uso de los medios telemáticos (Internet, telefonía móvil y videojuegos online principalmente) para ejercer el acoso psicológico entre iguales. No se trata aquí el acoso o abuso de índole estrictamente sexual ni los casos en los que personas adultas intervienen.

**Dashboard:** Tablero de información ordenado con representaciones gráficas y numéricas.