

**UNIVERSIDAD MAYOR DE SAN ANDRÉS**  
**FACULTAD DE CIENCIAS PURAS Y NATURALES**  
**CARRERA DE INFORMÁTICA**



**TESIS DE GRADO**

**DETECCIÓN DE PERSONALIDAD EN RED SOCIAL**  
**TWITTER**

**Tesis de Grado para obtener el Título de Licenciatura en Informática**  
**Mención Ingeniería de Sistemas Informáticos**

**POR: EBER EDGAR QUENTA LOPEZ**

**TUTOR METODOLÓGICO: M.SC. FRANZ CUEVAS QUIROZ**

**ASESOR: M.SC. ALDO RAMIRO VALDEZ ALVARADO**

**NUESTRA SEÑORA DE LA PAZ – BOLIVIA**

**2021**

HOJA DE CALIFICACIONES  
UNIVERSIDAD MAYOR DE SAN ANDRÉS  
FACULTAD DE CIENCUAS PURAS Y NATURALES  
CARRERA DE INFORMÁTICA

Tesis de grado:

DETECCIÓN DE PERSONALIDAD EN RED SOCIAL TWITTER

Presentado por: Eber Edgar Quenta Lopez

Para optar el grado Académico de Licenciado en Informática

Mención Ingeniería de Sistemas Informáticos

Nota Numeral: .....

Nota Literal: .....

Ha sido: .....

Director de la carrera de Informática: Ph.D. Jose Maria Tapia Baltazar

Tutor: M.Sc. Franz Cuevas Quiroz

Asesor M.Sc. Aldo Ramiro Valdez Alvarado

Tribunal: M.Sc. Menfy Morales Rios

Tribunal: Ph.D. Yohoni Cuenca Sarzuri

Tribunal: Lic. Eufren Llanque Quispe



**Universidad Mayor de San Andrés**  
**Facultad de Ciencias Puras y Naturales**  
**Carrera de Informática**



**LA CARRERA DE INFORMÁTICA DE LA FACULTAD DE CIENCIAS PURAS Y NATURALES PERTENECIENTE A LA UNIVERSIDAD MAYOR DE SAN ANDRÉS AUTORIZA EL USO DE LA INFORMACIÓN CONTENIDA EN ESTE DOCUMENTO SI LOS PROPÓSITOS SON ESTRICTAMENTE ACADÉMICOS.**

**LICENCIA DE USO**

El usuario está autorizado a:

- a) Visualizar el documento mediante el uso de un ordenador o dispositivo móvil.
- b) Copiar, almacenar o imprimir si ha de ser de uso exclusivamente personal y privado.
- c) Copiar textualmente parte(s) de su contenido mencionado la fuente y/o haciendo la referencia correspondiente respetando las normas de redacción e investigación.

El usuario no puede publicar, distribuir o realizar emisión o exhibición alguna de este material, sin la autorización correspondiente.

**TODOS LOS DERECHOS RESERVADOS EL USO NO AUTORIZADO DE LOS CONTENIDOS PUBLICADOS EN ESTE SITIO DERIVA EN EL INICIO DE**

**ACCIONES LEGALES CONTEMPLADOS EN LEY DERECHOS DE AUTOR.**

### **DEDICATORIA**

El presente proyecto de tesis está dedicada a Dios por haberme permitido realizar mis estudios. A mi familia que ha sido un pilar fundamental en mi formación como profesional y persona, por brindarme un apoyo constante en mis estudios. A mis amigos por apoyarme en continuar con mis estudios. Y por último a todas las personas que he tenido gusto de intercambiar ideas y mostrar una perspectiva diferente de algo.

## **AGRADECIMIENTOS**

Agradezco a todas aquellas personas que me han acompañado en este periodo de mi vida y a pesar de mis errores siempre han estado conmigo, representando mi voluntad en los momentos de duda y por brindarme la oportunidad de seguir.

Le doy gracias a mi familia por haberme apoyado en muchos aspectos de mi vida no solo en mi formación profesional, y en mi formación como persona.

[donEber98@gmail.com](mailto:donEber98@gmail.com)

## RESUMEN

Las redes sociales como han tomado gran importancia en la sociedad y son un medio de comunicación y difusión de información accesible y rápida.

El uso de la red social Twitter puede llegar a tener tanta repercusión que una publicación de una persona en esta red social puede llegar a cambiar inclusive el precio de la bolsa de valores en el mercado. Además el uso cotidiano de esta red social llega ser de tal magnitud que los usuarios van dejando muchos datos de ellos mismos, como sus gustos, ideas o pensamientos e incluso rasgos de su personalidad.

Hay tanta información pública que este puede llegar a ser aprovechada por personas de áreas como de reclutamiento, enseñanza, entre otros.

El presente proyecto de investigación tiene por objeto explorar la relación que existe entre la actividad de un usuario en Twitter con su personalidad, así también brindando una herramienta para el apoyo a las áreas de interés en contratar, conocer o enseñar a otras personas.

Durante el proceso de este proyecto también se va utilizando y adaptando una metodología para el análisis de datos y otra para el desarrollo de software.

Finalmente se muestran algunos resultados e interpretaciones de la investigación.

**Palabras clave:** Personalidad, Twitter, Software, Framework, API

**Metodologías:** Knowledge Discovery in Databases y Waterfall refactorizados

## ABSTRACT

Social networks as they have taken great importance in society and are a means of communication and dissemination of accessible and fast information.

The use of the social network Twitter can have such an impact that a publication by a person on this social network can even change the price of the stock market in the market. In addition, the daily use of this social network becomes of such magnitude that users leave a lot of data about themselves, such as their tastes, ideas or thoughts and even their personality traits.

There is so much public information that it can be used by people in areas such as recruiting, teaching, among others.

The present research project aims to explore the relationship that exists between the activity of a user on Twitter and her personality, thus also providing a tool to support areas of interest in hiring, meeting or teaching other people.

During the process of this project, a methodology for data analysis and another for software development are also being used and adapted.

Finally, some results and interpretations of the investigation are shown.

**Keywords:** Personality, Twitter, Software, Framework, API

**Methodologies:** Knowledge Discovery in Databases and Waterfall refactorized



# Contenido

<b>CAPÍTULO I INTRODUCCIÓN.....</b>	<b>1</b>
1.1. ANTECEDENTES .....	2
1.2. PLANTEAMIENTO DEL PROBLEMA .....	4
1.3. JUSTIFICACIÓN.....	4
1.4. HIPÓTESIS .....	5
1.5. OBJETIVOS .....	6
1.5.1. Objetivo general .....	6
1.5.2. Objetivos específicos .....	6
1.6. ALCANCES Y LÍMITES.....	6
1.6.1. Alcances .....	6
1.6.2. Límites .....	7
<b>CAPÍTULO II MARCO TEÓRICO.....</b>	<b>8</b>
2.1. KNOWLEDGE DISCOVERY IN DATABASES .....	8
2.1.1. Etapa de selección .....	9
2.1.2. Etapa de pre-procesamiento .....	10
2.1.3. Etapa de transformación/reducción.....	10
2.1.4. Etapa de minería de datos.....	11
2.1.5. Etapa de interpretación/Evaluación de datos .....	12
2.2. CICLO DE VIDA DEL DESARROLLO DE SOFTWARE.....	12

2.2.1.	Las fases de Ciclo de Vida del Desarrollo de Software .....	13
2.3.	<i>MODELO WATERFALL</i> .....	16
2.3.1.	Análisis.....	17
2.3.2.	Diseño.....	18
2.3.3.	Implementación.....	18
2.3.4.	Verificación .....	18
2.3.5.	Mantenimiento.....	19
2.4.	<i>API REST</i> .....	19
2.4.1.	API.....	19
2.4.2.	REST .....	19
2.5.	<i>ARQUITECTURA SOFEA</i> .....	20
2.5.1.	Ventajas .....	20
2.6.	<i>RED SOCIAL</i> .....	21
2.6.1.	Tipos de redes sociales .....	21
2.6.2.	Principales redes sociales .....	22
2.6.3.	Twitter .....	23
2.7.	<i>LA PERSONALIDAD</i> .....	24
2.8.	<i>COMPONENTES DE LA PERSONALIDAD</i> .....	24
2.8.1.	Temperamento .....	24
2.8.2.	Carácter .....	24
2.9.	<i>TEORIAS DE LA PERSONALIDAD</i> .....	25
2.9.1.	Teoría de la personalidad de Freud .....	26
2.9.2.	Teoría de la personalidad de Jung .....	26
2.9.3.	Teoría de la personalidad de Eysenck.....	27

<b>CAPITULO III DETECCIÓN DE PERSONALIDAD .....</b>	<b>29</b>
3.1. <i>DESARRLLO DEL ANALISIS DE PERSONALIDAD .....</i>	31
3.1.1.    SELECCIÓN .....	31
3.1.2.    RECOLECCIÓN .....	33
3.1.3.    PRE-PROCESAMIENTO .....	39
3.1.4.    ANÁLISIS DE DATOS .....	44
3.1.5.    INTERPRETACIÓN DE RESULTADOS .....	48
3.2. <i>DESARROLLO DE LA HERRAMIENTA SOFTWARE .....</i>	52
3.2.1.    REQUERIMIENTOS .....	52
3.2.2.    PLANIFICACIÓN .....	52
3.2.3.    Diseño de la arquitectura .....	55
3.2.4.    IMPLEMENTACIÓN .....	55
3.2.5.    PRUEBAS .....	58
<b>CAPÍTULO IV PRUEBA DE HIPÓTESIS .....</b>	<b>59</b>
4.1. <i>DEFINICIÓN DE HIPÓTESIS .....</i>	59
4.2. <i>NIVEL DE SIGNIFICANCIA .....</i>	60
4.3. <i>ESTADÍSTICO DE PRUEBA .....</i>	60
4.4. <i>REGLA DE DECISIÓN .....</i>	62
4.5. <i>TOMA DE DECISIÓN .....</i>	62
<b>CAPITULO IV CONCLUSIONES Y RECOMENDACIONES .....</b>	<b>64</b>
5.1. <i>CONCLUSIONES .....</i>	64
5.2. <i>RECOMENDACIONES .....</i>	65

**BIBLIOGRAFÍA..... 66**

## Indice figuras

Figura	2.1: Etapas del proceso KDD .....	9
Figura	2.2: Ciclo de vida del desarrollo de software .....	13
Figura	2.3: Etapas del modelo en Waterfall .....	17
Figura	Figura 2.4: Componentes de la personalidad .....	25
Figura	3.1: Etapas de la Metodología KDD para el proyecto .....	29
Figura	3.2: Modelo en cascada .....	30
Figura	3.3: Esquema del KDD' con una metodología tradicional .....	31
Figura	3.4: Esquema de la obtención de datos con la API de twitter .....	35
Figura	3.5: Visualización del vocabulario de una persona .....	43
Figura	3.6: Distribución del vocabulario de un usuario .....	44
Figura	3.7: Distribución de la cantidad de re-tweets .....	46
Figura	3.8: Distribución de los tweets de los usuarios .....	46
Figura	3.9: Distribución de reacciones de los usuarios .....	47
Figura	3.10: Distribución de la personalidad de la muestra .....	48
Figura	3.11: Comparación de la actividad con la personalidad .....	49
Figura	3.12: Relación entre extroversión y los retweets de los usuarios .....	50
Figura	3.13: Relación entre extroversión y los tweets de los usuarios .....	50

Figura	3.14 Relación entre extroversión y las reacciones de los usuarios .....	51
Figura	3.11 Diagrama de contexto.....	53
Figura	3.12 Mockup de la vista home .....	54
Figura	3.13 Mockup del estado de analizando.....	54
Figura	3.14 Mockup de la visualización de resultados del perfil .....	55
Figura	3.15 Arquitectura SOFEA .....	55
Figura	3.16 Vista principal del sitio TweetTest con un usuario .....	56
Figura	3.17: Vista del análisis de extracción y procesado de datos .....	57
Figura	3.18 Vista del resultado del análisis del usuario.....	57

## Tablas

Tabla	3.1 Extrovertidos e Introvertidos.....	32
Tabla	3.2 Atributos a considerar en Twitter .....	33
Tabla	3.3 Credenciales para consumir la API de Twitter.....	34
Tabla	3.4: Ejemplo de los últimos tweets de un determinado usuario .....	35
Tabla	3.5: Lista de datos obtenidos mediante la API de Twitter.....	36
Tabla	3.6 Datos del comportamiento de usuarios en Twitter.....	37
Tabla	3.7 Niveles de extroversión e introversión .....	38
Tabla	3.8 Descripción de las expresiones regulares para el pre-procesamiento de datos.....	39
Tabla	3.9. Ejemplo de entrada y salida en la limpieza de texto .....	39
Tabla	3.10. Descripción de las expresiones regulares para el pre-procesamiento de datos.....	40
Tabla	3.11. Entrada y salida en la transformación de datos .....	41
Tabla	3.12 Convención universal de categorías gramaticales.....	41
Tabla	3.13 Ejemplo de categorización.....	42
Tabla	3.12 Listado de módulos .....	56
Tabla	3.13 Listado de pruebas manuales .....	58
Tabla	3.14 Datos de una muestra de usuarios encuestados.....	60

# **CAPÍTULO I**

## **INTRODUCCIÓN**

Todos ejercemos una influencia en nuestro entorno más cercano y de manera inversa, el entorno tiene influencia en nosotros mismos. Las personas y las cosas que nos rodean provocan un efecto en nuestro bienestar. Pueden ser buenos aliados como grandes enemigos, afectando el ambiente donde convivimos, afectando incluso nuestras creencias y pensamientos.

El crecimiento de las redes sociales en la era digital es indudable; son los medios de comunicación más utilizados por las personas para mantener relaciones las unas con las otras, compartir ideas, noticias, entre otros. Las estadísticas que se hacen con respecto a su uso, revelan un gran crecimiento y muchas empresas modifican su estrategia de marketing.

Según Alfonso Alcántara (2012) "Las redes sociales son ventanas que muestran de forma más natural cómo se comportan las personas, ya sea para bien o para mal".

Las redes sociales se han convertido en una especie de ventana a través de la cual se divulga gran cantidad de información sobre nuestra vida y personalidad.

La información en formatos no estructurados o semiestructurados son cada vez más utilizados, como mensajes de correo electrónico, notas de los centros de servicio al cliente, respuestas de encuestas con final abierto, fuentes de noticias, formularios web, etc. Esta



abundancia de información se presenta como un problema para muchas empresas a la hora de preguntarse cómo recopilar, explorar y aprovechar toda esta información.

Es en estas redes sociales donde podemos estimar la personalidad de los individuos, analizar comportamientos y detectar perfiles, incluso psicólogos y profesionales dedicados al coaching y recursos humanos se han aficionado a hurgar en la red con el fin de evaluar la personalidad del futuro empleado. Twitter se ha vuelto en un medio de información masivo de gran influencia en la sociedad. Además de jugar un papel importante tanto en la política como en otros aspectos de la sociedad.

### 1.1. ANTECEDENTES

**Título:** Redes sociales, Big Data y su influencia.  
**Autor:** Martin Hilbert.  
**Año:** 2017.  
**Institución:** Sociedad de la Información de la Comisión Regional de la CEPAL  
 Resumen

El doctor en Ciencias Sociales y PhD en Comunicación Martin Hilbert, en una charla sobre Big Data, redes sociales, y su influencia en la sociedad y hasta en la democracia. "Con 100 clicks los algoritmos de Facebook pueden conocer tu personalidad. Con 150 likes pueden conocerla mejor que tu pareja; y con 250 likes, incluso conocen tu personalidad mejor que tú mismo"

**Título:** Modelo de Personalidad  
**Autor:** Hans Jürgen Eysenck  
**Año:** 1952  
**Institución:** Revista Personality and Individual Differences Resumen

El Inventario de Personalidad de Eysenck (EPI) mide dos dimensiones omnipresentes e independientes de la personalidad, Extraversión-Introversión y Neuroticismo-Estabilidad, que explican la mayor parte de la varianza en el dominio de la personalidad. Cada formulario contiene 57 elementos "Sí-No" sin repetición de artículos. La inclusión de una

escala de mentira (respuestas falsas) proporciona una detección de distorsión de respuestas. Los rasgos medidos son Extraversión-Introversión y Neuroticismo.

**Título:** Escrutinio de depresión y ansiedad en Twitter  
**Autor:** Estudiantes de Medicina, UAM  
**Año:** 2015  
**Institución:** Universidad Autónoma Metropolitana Resumen

Un grupo de especialistas diseñó un software que promete analizar los enunciados publicados en Twitter y detectar diversos componentes de la personalidad. Esta investigación se enfocó en conocer si es posible utilizar estas publicaciones como elementos para detectar depresión o ansiedad, y comparar la congruencia entre la interpretación del software y las pruebas in vivo a través de inventarios de diagnóstico estandarizados.

**Título:** The Secret Life of Pronouns  
**Autor:** James W. Pennebaker  
**Año:** 2013

En parte un viaje de investigación, el libro rastrea el descubrimiento de los vínculos entre las palabras funcionales y los estados sociales y psicológicos. Escrito para una audiencia general, el libro lleva al lector a un viaje extraordinario y a menudo inesperado en la mente de autores, poetas, letristas, políticos y gente común a través del uso de las palabras. En el corazón de este libro está la idea de que nuestras palabras dejan huellas indelebles de la personalidad, nuestras relaciones y antecedentes, e incluso nuestros planes para el futuro. Una vez que vea el poder de los pronombres, artículos y otras palabras funcionales. (Pennebaker, 2013)

Es importante conocer la personalidad para detectar al tipo de personas en nuestro grupo social, trabajo o comunidad.

## 1.2. PLANTEAMIENTO DEL PROBLEMA

Ante diferentes problemas como:

- Los usuarios en twitter como en el cyberspacio de internet desconocen a los demás usuarios con los que interactua.
- En el proceso de reclutamiento de personal en el área de RRHH en las empresas necesitan de las redes sociales para perfilar mejor idea del postulante.
- La interacción en ambientes educativos como colegios, la relacion entre estudiantes y profesores llega a ser distante.

Los problemas señalados anteriormente nos llevan a formular la siguiente pregunta:

¿Cómo detectar rasgos de la personalidad de una persona mediante su perfil en la red social Twitter?

## 1.3. JUSTIFICACIÓN

La repercusión del comportamiento del individuo en la sociedad resulta de gran impacto, así pues, es de gran importancia poder conocer las personas con quien nos relacionamos. Utilizando esta esté software hacemos un análisis rápido sobre el posible comportamiento de determinada persona. En presente proyecto de tesis adquiere gran importancia en la sociedad, ya que estamos analizando lo que en sociología se conoce como la realidad social.

La implementación de sistemas de detección de personalidad pueden llegar a ayudar al área de RRHH de una empresa. Llegando a disminuir el tiempo de entrevista y ayuda a elegir un mejor personal para aumentar el desempeño del equipo.

Así tambien el interés científico en la presente investigación se basa avances de tecnología de software e investigación y métodos de detección del comportamiento humano entre otros.

Haciendo enfoque en la aplicación de la información que nos brinda la plataforma de Twitter funcionamos tanto elementos de psicología como de informática.

#### **1.4. HIPÓTESIS**

La interacción de un usuario en la red social Twitter permite identificar rasgos de personalidad del mismo.

**Variable Independiente:** La interacción en Twitter

**Concepto:** La actividad de un usuario en la red social

**Indicador:** Información significativa y estructurada

**Instrumento:** Técnicas de extracción de datos

**Valor:** Cantidad de información

**Variable Dependiente:** Personalidad

**Concepto:** Comportamiento, forma de expresarse o manera de actuar

**Indicador:** Modelo de personalidad

**Instrumento:** Cuestionario de evaluación de personalidad

**Valor:** Extrovertido, Introverso

## **1.5. OBJETIVOS**

### **1.5.1. Objetivo general**

Proponer un prototipo herramienta web para la de detección de rasgos de personalidad de un individuo mediante su perfil en la red social Twitter

### **1.5.2. Objetivos específicos**

- Diferenciar los rasgos de personalidades de un individuo.
- Analizar la información de un individuo en Twitter.
- Determinar los diferentes aspectos presentes en la personalidad.
- Desarrollar una interfaz web para visualizar y analizar la infomación de un perfil en la red social Twitter.

## **1.6. ALCANCES Y LÍMITES**

### **1.6.1. Alcances**

El proyecto ocupará los siguientes alcances:

- Extraer información de perfiles en Twitter.
- Análizar la información extraída de un perfil de Twitter.
- Desplegar infromación de la actividad de un determinado perfil de Twitter.

### **1.6.2. Límites**

- El proyecto solo analizará la información pública en un perfil de Twitter.
- No se considerará las conversaciones u otro tipo de información privada unos con otros usuarios
- Se considera la teoría de personalidad limitada a 2 aspectos de la personalidad.

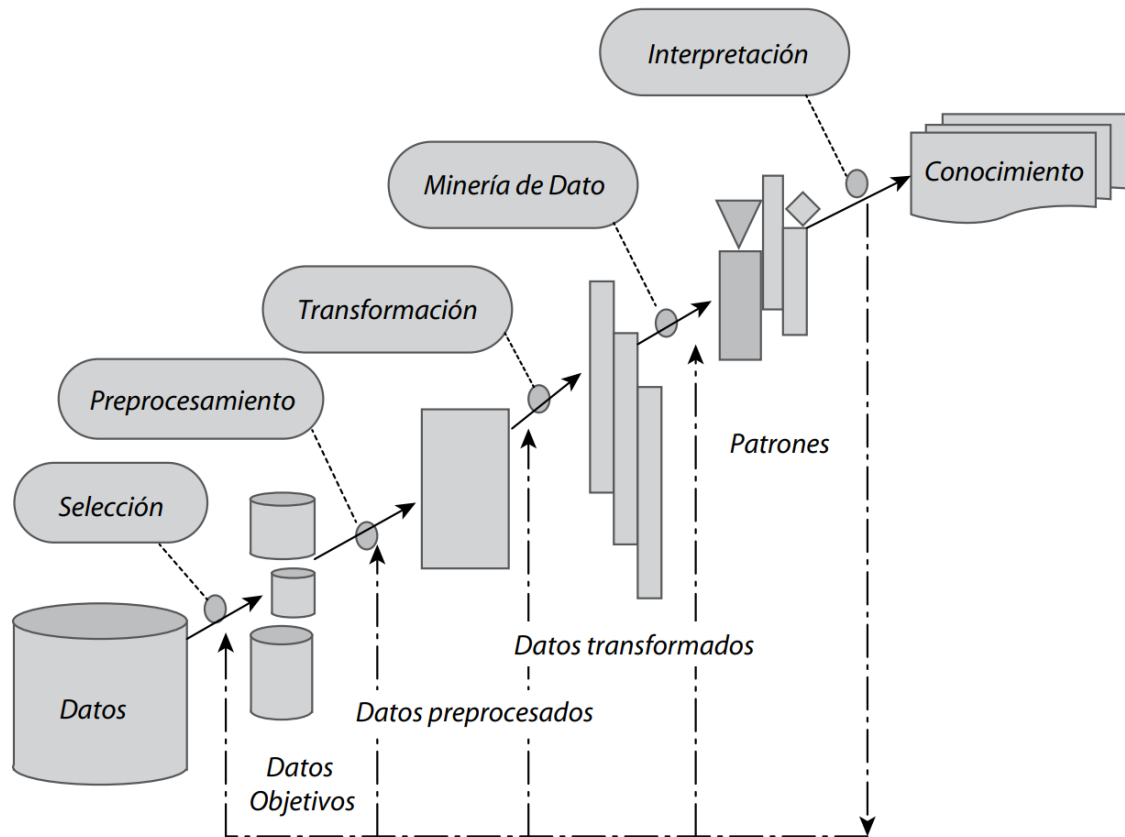
## **CAPÍTULO II**

### **MARCO TEÓRICO**

#### **2.1. KNOWLEDGE DISCOVERY IN DATABASES**

El proceso de extraer conocimiento a partir de grandes volúmenes de datos ha sido reconocido por muchos investigadores como un tópico de investigación clave en los sistemas de bases de datos, y por muchas compañías industriales como una importante área y una oportunidad para obtener mayores ganancias (Timarán, 2009). Autores como Fayyad, Piatetsky-Shapiro y Smith (1996, p. 89) lo definen como “El proceso no trivial de identificación de patrones válidos, novedosos, potencialmente útiles y fundamentalmente entendibles al usuario a partir de los datos”. El Descubrimiento de conocimiento en bases de datos (kdd, del inglés Knowledge Discovery in Databases) es básicamente un proceso automático en el que se combinan descubrimiento y análisis. El proceso consiste en extraer patrones en forma de reglas o funciones, a partir de los datos, para que el usuario los analice. Esta tarea implica generalmente preprocesar los datos, hacer minería de datos (data mining) y presentar resultados (Agrawal y Srikant, 1994) (Chen, Han y Yu, 1996) (Piatetsky Shapiro, Brachman y Khabaza, 1996) (Han y Kamber, 2001). kdd se puede aplicar en diferentes dominios, por ejemplo, para determinar perfiles de clientes fraudulentos (evasión de impuestos), para descubrir relaciones implícitas existentes entre síntomas y enfermedades, entre características técnicas y diagnóstico del estado de equipos y máquinas, para determinar perfiles de estudiantes “académicamente exitosos” en términos

de sus características socioeconómicas y para determinar patrones de compra de los clientes en sus canastas de mercado.



**Figura 2.1: Etapas del proceso KDD**  
(Timaran et al., 2016)

### 2.1.1. Etapa de seleccion

En la etapa de seleccion, una vez identificado el conocimiento relevante y prioritario y definidas las metas del proceso KDD, desde el punto de vista del usuario final, se crea un conjunto de datos objetivo, seleccionando todo el conjunto de datos o una muestra representativa de este, sobre el cual se realiza el proceso de descubrimiento. La seleccion de los datos vara de acuerdo con los objetivos del negocio



### **2.1.2. Etapa de pre-procesamiento**

En la etapa de preprocesamiento se analiza la calidad de los datos, se aplican operaciones básicas como la remoción de datos ruidosos, se seleccionan estrategias para el manejo de datos desconocidos (*missing* y *empty*), datos nulos, datos duplicados y técnicas estadísticas para su reemplazo. En esta etapa, es de suma importancia la interacción con el usuario o analista. Los datos ruidosos son valores que están significativamente fuera del rango de valores esperados; se deben principalmente a errores humanos, a cambios en el sistema, a información no disponible a tiempo y a fuentes heterogéneas de datos. Los datos desconocidos *empty* son aquellos a los cuales no les corresponde un valor en el mundo real y los *missing* son aquellos que tienen un valor que no fue capturado. Los datos nulos son datos desconocidos que son permitidos por los sistemas gestores de bases de datos relacionales (sgbdr). En el proceso de limpieza todos estos valores se ignoran, se reemplazan por un valor por omisión, o por el valor más cercano, es decir, se usan métricas de tipo estadístico como media, moda, mínimo y máximo para reemplazarlos.

### **2.1.3. Etapa de transformación/reducción**

En la etapa de transformación/reducción de datos, se buscan características útiles para representar los datos dependiendo de la meta del proceso. Se utilizan métodos de reducción de dimensiones o de transformación para disminuir el número efectivo de variables bajo consideración o para encontrar representaciones invariantes de los datos (Fayyad et al., 1996).

Los métodos de reducción de dimensiones pueden simplificar una tabla de una base de datos horizontal o verticalmente. La reducción horizontal implica la eliminación de tuplas idénticas como producto de la sustitución del valor de un atributo por otro de alto nivel, en una jerarquía definida de valores categóricos o por la discretización de valores continuos (por ejemplo, edad por un rango de edades). La reducción vertical implica la eliminación de atributos que son insignificantes o redundantes con respecto al problema, como la eliminación de llaves, la eliminación de columnas que dependen funcionalmente (por

ejemplo, edad y fecha de nacimiento). Se utilizan técnicas de reducción como agregaciones, compresión de datos, histogramas, entre otras (Han y Kamber, 2001).

#### **2.1.4. Etapa de minería de datos**

El objetivo de la etapa minería de datos es la búsqueda y descubrimiento de patrones insospechados y de interés, aplicando tareas de descubrimiento como clasificación (Quinlan, 1986) (Wang, Iyer y Scott, 1998), clustering (Ng y Han, 1994), (Zhang, Ramakrishnan, Livny, 1996), patrones secuenciales (Agrawal y Srikant, 1995) y asociaciones (Agrawal y Srikant, 1994), (Srikant y Agrawal, 1996), entre otras. Las técnicas de minería de datos crean modelos que son predictivos o descriptivos. Los modelos predictivos pretenden estimar valores futuros o desconocidos. El proceso de descubrimiento de conocimiento en bases de datos de variables de interés, que se denominan variables objetivo, dependientes o clases, usando otras variables denominadas independientes o predictivas, como por ejemplo predecir para nuevos clientes si son buenos o malos basados en su estado civil, edad, género y profesión, o determinar para nuevos estudiantes si desertan o no en función de su zona de procedencia, facultad, estrato, género, edad y promedio de notas. Entre las tareas predictivas están la clasificación y la regresión. Los modelos descriptivos identifican patrones que explican o resumen los datos; sirven para explorar las propiedades de los datos examinados, no para predecir nuevos datos, como identificar grupos de personas con gustos similares o identificar patrones de compra de clientes en una determinada zona de la ciudad. Entre las tareas descriptivas se cuentan las reglas de asociación, los patrones secuenciales, los clustering y las correlaciones.

Por lo tanto, la escogencia de un algoritmo de minería de datos incluye la selección de los métodos por aplicar en la búsqueda de patrones en los datos, así como la decisión sobre los modelos y los parámetros más apropiados, dependiendo del tipo de datos (categóricos, numéricos) por utilizar

### **2.1.5. Etapa de interpretación/Evaluación de datos**

En la etapa de interpretación/evaluación, se interpretan los patrones descubiertos y posiblemente se retorna a las anteriores etapas para posteriores iteraciones. Esta etapa puede incluir la visualización de los patrones extraídos, la remoción de los patrones redundantes o irrelevantes y la traducción de los patrones útiles en términos que sean entendibles para el usuario. Por otra parte, se consolida el conocimiento descubierto para incorporarlo en otro sistema para posteriores acciones o, simplemente, para documentarlo y reportarlo a las partes interesadas; también para verificar y resolver conflictos potenciales con el conocimiento previamente descubierto.

## **2.2. CICLO DE VIDA DEL DESARROLLO DE SOFTWARE**

El ciclo de vida del software puede tener distintas metodologías pero una única normativa es la ISO/IEC/IEEE 12207:2017.

Esta Norma Internacional establece un marco común para los procesos del ciclo de vida del software, con una terminología bien definida, que puede ser referenciada por la industria del software. Contiene procesos, actividades y tareas que deben aplicarse durante la adquisición de un sistema, producto o servicio de software y durante el suministro, el desarrollo, la explotación, el mantenimiento y la eliminación de los productos de software. Esto se logra mediante la participación de las partes interesadas, con el objetivo final de lograr la satisfacción del cliente. Esta Norma Internacional se aplica a la adquisición de sistemas, productos y servicios de software, al suministro, desarrollo, operación, mantenimiento y eliminación de productos de software y a la parte de software de cualquier sistema, ya sea realizada interna o externamente a una organización. El software incluye la parte de software del firmware. Se incluyen aquellos aspectos de la definición del sistema necesarios para proporcionar el contexto de los productos y servicios de software. Esta Norma Internacional también proporciona procesos que pueden emplearse para definir, controlar y mejorar los procesos del ciclo de vida del software dentro de una organización o un proyecto. Los procesos, actividades y tareas de esta Norma Internacional también

pueden aplicarse durante la adquisición de un sistema que contenga software. (IEEE/ISO/IEC 12207, 2017)

### 2.2.1. Las fases de Ciclo de Vida del Desarrollo de Software

La metodología para el desarrollo de software es un modo sistemático de realizar, gestionar y administrar un proyecto para llevarlo a cabo con grandes posibilidades de éxito. Esta sistematización indica cómo se divide un proyecto en módulos más pequeños para normalizar cómo se administra el mismo.



**Figura 2.2: Ciclo de vida del desarrollo de software**  
(Fuente propia )

Así, una metodología para el desarrollo de software son los procesos a seguir sistemáticamente para idear, implementar y mantener un producto de software desde que

surge la necesidad del producto hasta que se cumple el objetivo por el cual fue creado(*Ciclo de vida del software*, s. f.) .

De esta forma, las etapas del desarrollo de software son las siguientes:

#### **2.2.1.1. Planificación**

Antes de empezar un proyecto de desarrollo de un sistema de información, es necesario hacer ciertas tareas que influirán decisivamente en el éxito del mismo. Dichas tareas son conocidas como el *fuzzy front-end* del proyecto, puesto que no están sujetas a plazos.

Algunas de las tareas de esta fase incluyen actividades como la determinación del ámbito del proyecto, la realización de un estudio de viabilidad, el análisis de los riesgos asociados, la estimación del coste del proyecto, su planificación temporal y la asignación de recursos a las diferentes etapas del proyecto.

#### **2.2.1.2. Análisis**

Por supuesto, hay que averiguar qué es exactamente lo que tiene que hacer el software. Por eso, la etapa de análisis en el ciclo de vida del software corresponde al proceso a través del cual se intenta descubrir qué es lo que realmente se necesita y se llega a una comprensión adecuada de los requerimientos del sistema (las características que el sistema debe poseer).

#### **2.2.1.3. Diseño**

En esta fase se estudian posibles opciones de implementación para el software que hay que construir, así como decidir la estructura general del mismo. El diseño es una etapa compleja y su proceso debe realizarse de manera iterativa.

Es posible que la solución inicial no sea la más adecuada, por lo que en tal caso hay que refinarla. No obstante, hay catálogos de patrones de diseño muy útiles que recogen errores que otros han cometido para no caer en la misma trampa.

#### **2.2.1.4. Implementación**

En esta fase hay que elegir las herramientas adecuadas, un entorno de desarrollo que facilite el trabajo y un lenguaje de programación apropiado para el tipo de software a construir. Esta elección dependerá tanto de las decisiones de diseño tomadas como del entorno en el que el software deba funcionar.

Al programar, hay que intentar que el código no sea indescifrable siguiendo distintas pautas como las siguientes:

- Evitar bloques de control no estructurados.
- Identificar correctamente las variables y su alcance.
- Elegir algoritmos y estructuras de datos adecuadas para el problema.
- Mantener la lógica de la aplicación lo más sencilla posible.
- Documentar y comentar adecuadamente el código de los programas.
- Facilitar la interpretación visual del código utilizando reglas de formato de código previamente consensuadas en el equipo de desarrollo.

También hay que tener en cuenta la adquisición de recursos necesarios para que el software funcione, además de desarrollar casos de prueba para comprobar el funcionamiento del mismo según se vaya programando.

#### **2.2.1.5. Pruebas**

Como errar es humano, la fase de pruebas del ciclo de vida del software busca detectar los fallos cometidos en las etapas anteriores para corregirlos. Por supuesto, lo ideal es hacerlo antes de que el usuario final se los encuentre. Se dice que una prueba es un éxito si se detecta algún error.

### **2.2.1.6. Uso y mantenimiento**

Esta es una de las fases más importantes del ciclo de vida de desarrollo del software. Puesto que el software ni se rompe ni se desgasta con el uso, su mantenimiento incluye tres puntos diferenciados:

- Eliminar los defectos detectados durante su vida útil (mantenimiento correctivo).
- Adaptarlo a nuevas necesidades (mantenimiento adaptativo).
- Añadirle nuevas funcionalidades (mantenimiento perfectivo).

Aunque suene contradictorio, cuanto mejor es el software más tiempo hay que invertir en su mantenimiento. La principal razón es que se usará más (incluso de formas que no se habían previsto) y, por ende, habrá más propuestas de mejoras.

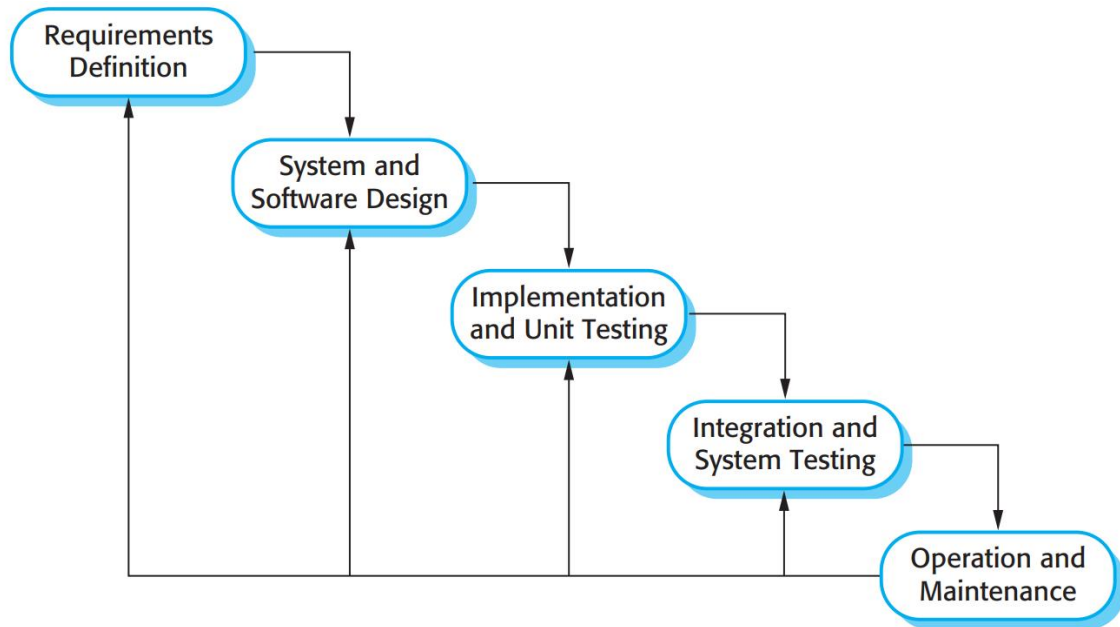
## **2.3. MODELO WATERFALL**

El modelo Waterfall es el modelo SDLC más antiguo y conocido. Este modelo se utiliza ampliamente en proyectos gubernamentales y en muchas empresas importantes. La característica especial de este modelo son sus pasos secuenciales. Va hacia abajo a través de las fases de análisis de requisitos, diseño, codificación, pruebas y mantenimiento. Además, asegura los defectos de diseño antes del desarrollo de un producto. Este modelo funciona bien para proyectos en los que el control de calidad es una preocupación importante debido a su intensa documentación y planificación. Las etapas que construyen este modelo no son etapas superpuestas, lo que significa que el modelo en cascada comienza y termina una etapa antes de comenzar la siguiente. (Adel y Abdullah, 2015)

Existen distintas metodologías que ayudan a definir este proceso para el desarrollo de un software, una de las más comunes y conocidas es el modelo en cascada. Este método,

originalmente fue propuesto en 1970 por Winston W. Royce, y es también conocido como modelo lineal o modelo de ciclo de vida de un programa.

El nombre de este modelo se debe a la manera en la que se dividen y se llevan a cabo cada una de las fases de su proceso, ya que se realiza de manera escalonada, siguiendo una secuencia ordenada desde la primera hasta la última etapa.



**Figura 2.3: Etapas del modelo en Waterfall**  
(Sommerville, 2011)

Se puede adaptar estas fases según lo requiera el proyecto de software que se este desarrollando, siempre y cuando se tomen en cuenta todos los puntos importantes dentro de las 5 fases principales, que se explica a continuación:

### 2.3.1. Análisis

Esta es la etapa de preparación del proyecto, el trabajo será determinar cuáles son las necesidades y los objetivos a cumplir del proyecto, y posteriormente reunir todos los requisitos que se deben cumplir en el desarrollo del software para llevar a cabo todo el proceso y cumplir con los objetivos.



Por lo general, esta también es la fase del modelo de cascada en la que se debe presentar una propuesta del proyecto en caso de estar trabajando para un cliente. En caso de que se tenga un trabajo como freelance, no deja de ser una buena práctica el realizar esta propuesta y añadirla a tu portafolio de trabajo.

### **2.3.2. Diseño**

Durante esta fase, se debe definir la organización de la estructura y de todos los elementos que se necesita para el desarrollo de tu software. El proceso de diseño de sistemas asigna los requisitos a los sistemas de hardware o software mediante el establecimiento de una arquitectura general del sistema. El diseño de software implica identificar y describir las abstracciones fundamentales del sistema de software y sus relaciones.

### **2.3.3. Implementación**

En la etapa de implementación se debe realizar una traducción de todos los elementos del diseño que se preparó en la etapa previa al lenguaje del software.

Posteriormente, se debe integrar cada uno dentro del código y programación del software, realizando pruebas para verificar que no existan errores y dando forma poco a poco al producto terminado.

### **2.3.4. Verificación**

En esta fase del modelo de cascada se debe probar y ejecutar el código final y verificar su funcionamiento. Es necesario que se compare los resultados finales con los objetivos iniciales y se compruebe si se cumplió con cada uno de ellos.

Se recomienda realizar pruebas de cada uno de los elementos que se haya utilizado y que se tome en cuenta comentarios que se reciba con alguna encuesta del resultado ante el cliente.

### **2.3.5. Mantenimiento**

Finalmente, se debe analizar los resultados del paso anterior y realizar los cambios pertinentes, si es que fueran necesarios, así para dar por concluido el proyecto.

Es probable que se tenga que regresar a esta fase del modelo en cascada más de una vez cada cierto tiempo para comprobar que se adapta a los cambios de su entorno, ya que con lo rápido que se dan los cambios en la tecnología hoy en día, es esencial que se haga un mantenimiento del software constantemente actualizado para que siga siendo relevante.

## **2.4. API REST**

### **2.4.1. API**

Una Interfaz de Programación de Aplicaciones por sus siglas del inglés *representational state transfer* (API) es un término común en Ciencias de la Computación. Es un término genérico que es usado para cubrir todos los requerimientos técnicos para tener varios componentes de software capaces de funcionar con los demás. En un contexto web, su uso es un poco más restrictivo, ya que es usado casi sólo con JavaScript. En ese contexto, una API es usualmente un conjunto de métodos, propiedades y eventos para con el fin lograr ciertas tareas. Por ejemplo, el API 2D de Canvas provee los métodos necesarios para dibujar gráficos 2D dentro de un elemento Canvas de HTML (*API - Glosario / MDN*, s. f.).

### **2.4.2. REST**

El término Transferencia de Estado Representacional (REST) representa un conjunto de características de diseño de arquitecturas software que aportan confiabilidad, eficiencia y escalabilidad a los sistemas distribuidos. Un sistema es llamado RESTful cuando se ajusta a estas características (*REST - Glosario / MDN*, s. f.).

La idea básica de REST es que un recurso, por ejemplo un documento, es transferido con su estado y su relaciones (hipertexto) mediante formatos y operaciones estandarizadas bien definidas.

Como HTTP, el protocolo estandar de la Web, también transfiere documentos e hipertexto, las APIs HTTP a veces son llamadas APIs RESTful, servicios RESTful, o simplemente servicios REST, aunque no se ajusten del todo a la definción de REST. Los principiantes pueden pensar que una API REST es un servicio HTTP que puede ser llamado mediante librerías y herramientas web estandar.

## 2.5. ARQUITECTURA SOFEA

SOFEA es el acrónimo de *Service Oriented Front-End Architecture*. Estas nuevas arquitecturas están orientadas a configurar la parte de Backend como un conjunto de Servicios REST que se exponen a un cliente . Este cliente puede ser de cualquier tipo pero es cierto que en muchas ocasiones no encontraremos con librerías como React o Frameworks como Angular que realizan estas labores (Caules, 2020).

### 2.5.1. Ventajas

- Fundamentalmente tienen la ventaja de separar las responsabilidades y poder conectar de forma flexible cualquier tipo de cliente . Esta es una de sus mayores ventajas. Eso sí, existen otras ventajas a tener muy en cuenta.
- El cliente suele estar construido con HTML y JavaScript utilizando un framework MVC de tal forma que una vez hemos cargado la aplicación el resto de peticiones serán a url REST y solo nos traerán los datos que necesitamos mejorando el rendimiento HTTP.
- División de responsabilidades muy clara a nivel del equipo de desarrollo ya que unos desarrolladores se encargarán de diseñar el BackEnd y otros el FrontEnd.

- Simplicidad en cuanto a futuras evoluciones de la parte Cliente o la parte Servidora de forma totalmente independiente. Es decir si en un momento tenemos desarrollado el lado Servidor en Java y lo queremos cambiar por Node . Es algo viable sin afectar a todo el código desarrollado de Cliente . Simplemente mantenemos las mismas URL REST
- Mayor facilidad para encontrar especialistas tanto un desarrollador de Angular como un desarrollador de Spring encajan en la Arquitectura.
- Es una Arquitectura Moderna y nos permitirá usar framework y soluciones actuales con lo cual su ciclo de vida se extenderá.

## 2.6. RED SOCIAL

Servicio de la sociedad de la información que ofrece a los usuarios una plataforma de comunicación a través de internet para que estos generen un perfil con sus datos personales, facilitando la creación de comunidades con base en criterios comunes y permitiendo la comunicación de sus usuarios, de modo que pueden interactuar mediante mensajes, compartir información, imágenes o vídeos, permitiendo que estas publicaciones sean accesibles de forma inmediata por todos los usuarios de su grupo. (RAE, 2019)

### 2.6.1. Tipos de redes sociales

Las redes sociales se pueden clasificar en dos tipos:

- a) **Redes sociales horizontales o genéricas.** Son aquellas redes sociales que no poseen una temática determinada, sino que apuntan a todo tipo de usuarios. Estas redes funcionan como medios de comunicación, información o

entretenimiento. Son muy numerosas y populares, por ejemplo: Facebook o Twitter.

- b) **Redes sociales verticales.** Son aquellas redes sociales que relacionan personas con intereses específicos en común, como música, hobbies, deportes. Por ejemplo: Flickr, red social cuya temática es la fotografía. Dentro de estas redes se encuentran las redes verticales profesionales, como LinkedIn, que involucra individuos que comparten el ámbito laboral o que buscan ampliar sus fronteras laborales.

En general, ingresar en una red social es muy sencillo ya que simplemente implica rellenar un cuestionario con datos personales básicos y así obtener un nombre de usuario y una contraseña que le servirán al usuario para ingresar de manera privada a la red. Mientras el usuario cumpla los requisitos para el registro en dicha red (por ejemplo, mayoría de edad), podrá hacerlo de manera simple y rápida (Raffino, Redes Sociales, 2020).

### 2.6.2. Principales redes sociales

En la actualidad hay numerosas redes sociales, pero las más usadas a nivel laboral y personal son las siguientes:

- Facebook: Es la más famosa en la actualidad y supera los 2000 millones de usuarios activos al mes.
- Whatsapp: Mantiene en contacto a millones de usuarios repartidos por todo el mundo a través del móvil con mensajes instantáneos. La publicidad en esta plataforma está en auge.
- YouTube: Es la red de vídeos por excelencia. Las empresas la utilizan para introducir su publicidad, siendo también actores clave los youtuber.
- Instagram: Una de las redes más visuales y más utilizadas en la actualidad por los usuarios.

- Twitter: Está basada en el formato microblogging a través de mensajes cortos de 280 caracteres máximo.
- Pinterest: Otra red visual donde se incluyen tableros de fotografías que son compartidos por los usuarios.
- LinkedIn: Es una red social para intercambiar ofertas laborales y hacer conexión con contactos profesionales.

### **2.6.3. Twitter**

Twitter, un término inglés que puede traducirse como *gorjear* o *trinar*, es el nombre de una red de microblogging que permite escribir y leer mensajes en Internet que no superen los 280 caracteres. Estas entradas son conocidas como tweets.

El microblogging es una variante de los blogs (las bitácoras o cuadernos digitales que nacieron como diarios personales online). Su diferencia radica en la brevedad de sus mensajes y en su facilidad de publicación (pueden enviarse desde el móvil, ordenador o dispositivos con software de mensajería instantánea).

Cuando un usuario publica un mensaje en su página de Twitter, es enviado automáticamente a todos los usuarios que hayan escogido la opción de recibirlos. Dicho mensaje también puede ser visto de forma inmediata en el perfil del usuario.

La plataforma de Twitter nació en octubre de 2006 en San Francisco (Estados Unidos) como una red social con ciertas características auténticas. Entre las mismas, podemos mencionar que permite un servicio absolutamente gratuito y sin publicidades (se financia con las inversiones de empresas de capital riesgo), es muy sencillo de utilizar y estableció una nueva forma de comunicación, en la que lo público y lo privado se fusionan. (Julián Pérez Porto y Ana Gardey, 2010)

## **2.7. LA PERSONALIDAD**

La personalidad es la organización relativamente estable de características estructurales y funcionales, innatas y adquiridas bajo las especiales condiciones de su desarrollo, que conforman el equipo peculiar y definitorio de la conducta con lo que cada individuo afronta las distintas situaciones. (Bermúdez, 1996)

La personalidad se encuentra dentro del estudio del campo de la psicología, y se puede explicar desde varias funciones, en primer lugar permite conocer de forma aproximada los motivos que llevan a un individuo a actuar, a sentir, a pensar y a desenvolverse en un medio; por otro lado, la personalidad permite conocer la manera en la cual una persona puede aprender del entorno. La personalidad se puede definir como la estructura dinámica que tiene un individuo en particular; se compone de características psicológicas, conductuales, emocionales y sociales. Los griegos atribuyeron el término de persona a la máscara que utilizaban los actores en los teatros para identificarse con alguien en particular. Cuando los actores utilizaban esta máscara, actuaban de acuerdo al personaje que representaban. Del concepto de persona deriva, por lógica, el término personalidad, esto significa que la personalidad es lo que determina quién es cada quien y cómo nos manifestamos en el entorno. (Seelbach Gonzáles, 2013)

## **2.8. COMPONENTES DE LA PERSONALIDAD**

### **2.8.1. Temperamento**

Todos los seres humanos poseen una herencia genética, es decir, las características que se heredan de los padres, como por ejemplo, el color de ojos, el tono de piel, e incluso la propensión a determinadas enfermedades. El temperamento es un componente de la personalidad porque determina, de alguna manera, ciertas características de cognición, comportamiento y emocionales.

### **2.8.2. Carácter**

Se denomina carácter a las características de la personalidad que son aprendidas en el medio, por ejemplo, los sentimientos son innatos, es decir, se nace con ellos, pero la manera en cómo se expresan forma parte del carácter. Las normas sociales, los comportamientos y el lenguaje, son sólo algunos componentes del carácter que constituyen a la personalidad. Ver figura 2.1.

**Figura**



**Figura 2.4: Componentes de la personalidad**  
(Seelbach Gonzáles, 2013)

Algunos autores y ciertas corrientes tienen diferentes puntos de vista sobre la importancia de uno u otro factor, es decir, algunos autores aseguran que el temperamento es más importante en el despliegue de la personalidad, mientras que otros autores sostienen que el factor ambiental —el carácter— determina la personalidad (Seelbach Gonzáles, 2013)

## **2.9. TEORIAS DE LA PERSONALIDAD**

En psicología, se conoce como teorías de la personalidad a las distintas aproximaciones teóricas propuestas por los estudiosos de la personalidad en sus respectivas épocas, es decir, los intentos psicológicos formales por definir y clasificar las personalidades humanas en base a algún tipo de rasgos mínimos compartidos. (Raffino, Teorías de la Personalidad, 2020)



### **2.9.1. Teoría de la personalidad de Freud**

Propuesta por el célebre padre del psicoanálisis, el austríaco Sigmund Freud (1856-1939), esta teoría propone que la personalidad de los individuos se forma a lo largo de su historia de vida, mediante la sumatoria de todos los objetos amados y perdidos.

Dichos “objetos” vendrían a ser, en primera instancia, los padres, por quienes se produce inicialmente un vínculo de amor, que el llamado “Complejo de Edipo” nos hará superar mediante la renuncia. Pero posteriormente serán otros quienes ocupen ese lugar de objeto amado y luego perdido, como los amigos, parejas, colegas, etc.

Esta dinámica de amor y renuncia va formando al “yo”, una de las tres instancias básicas de la psique para Freud (junto con el “superyó” o la ley, y el “ello” o el inconsciente), a medida que éste asimila como propios algunos de los rasgos de cada objeto perdido. Así, de un profesor muy querido podemos “heredar” nuestra vocación, o ciertos gustos de algún amigo, etc.

En todo caso, según Freud la personalidad vendría a ser una suerte de “colección” de objetos perdidos, que nos confieren un recorrido afectivo único, pero con muchos puntos de encuentro con los demás

### **2.9.2. Teoría de la personalidad de Jung**

Elaborada por uno de los discípulos de Freud, el psiquiatra y psicoanalista suizo Carl Gustav Jung (1875-1961), esta teoría de la personalidad de 1921 propone que ciertos arquetipos determinan la conformación de nuestra mente, los cuales permiten la existencia de ocho perfiles posibles de la personalidad, que son:

- Pensamiento-introvertido. Personalidades centradas en su mundo interior, mucho más que en el afuera, e interesadas por el pensamiento abstracto, reflexivo y teórico.

- Sentimental-introvertido. Personalidades empáticas, que valoran su vínculo con los demás, aunque no sean muy dadas a expresarlo de manera abierta y franca.
- Sensación-introvertido. Personalidades enfocadas en fenómenos subjetivos, introspectivas, pero más vinculadas con lo que captan sus sentidos, o sea, con su propia sensibilidad.
- Intuitivo-introvertido. Personalidades soñadoras, que se desapegan de lo real inmediato y son dadas a la fantasía.
- Pensamiento-extrovertido. Personalidades que disfrutan de la explicación, o sea, de registrar lo que ocurre a su alrededor y constituir con ello un sistema mental, abstracto.
- Sentimental-extrovertido. Personalidades muy sociables, que disfrutan de la compañía de los demás y presentan baja propensión al pensamiento y la reflexión de tipo abstracto, al ser más inmediatistas en sus intereses.
- Sensación-extrovertido. Personalidades que anhelan nuevas sensaciones provenientes del afuera y de los demás, por lo que suelen ser dadas a la búsqueda del placer y ser muy abiertas a lo nuevo.
- Intuición-extrovertido. Personalidades aventureras, carismáticas y dotadas de liderazgo, que tienden a ocupar roles protagónicos en su comunidad y a liderar causas sociales, políticas o comunitarias, dado que se realizan ante los demás.

### **2.9.3. Teoría de la personalidad de Eysenck**

Hans Eysenck (1916-1997) es el psicólogo inglés autor de esta teoría centrada en lo biológico, para el cual ideó el modelo PEN, una explicación de las motivaciones de la personalidad en base a elementos internos del organismo. Así, Eysenck determina tres factores centrales para definir la personalidad:

- **Psicoticismo.** O tendencia a actuar con dureza, que dependería de la activación del Sistema de Activación Reticular Ascendente (SARA).
- **Neuroticismo.** O estabilidad de las emociones, que dependería del sistema límbico.
- **Introversión/extraversión.** O tendencia a focalizarse en el mundo interno o externo, el cual se vincula con los niveles de andrógenos y de neurotransmisores como la dopamina y la serotonina.

En base a los niveles de estos factores, las personalidades podrán ser de una u otra manera, según Eysenck.

## CAPITULO III

### DETECCIÓN DE PERSONALIDAD

Para el desarrollo del proyecto se procede a elaborar una adaptación de las metodologías de desarrollo adaptando la metodología **Knowledge Discovery in Databases** a nuestras necesidades.



**Figura 3.1: Etapas de la Metodología KDD para el proyecto**  
(Elaboración propia)

En la figura 3.1 para la fase de análisis de los datos de los usuarios de Twitter en las etapas del KDD se agrega una etapa de selección de datos e cambios.

Para la metodología de desarrollo del software se toma como base el modelo en cascada **Waterfall**. En la figura 3.2 se observa las etapas para el desarrollo del software



**Figura 3.2: Modelo en cascada**  
(Elaboración propia)

Así podemos combinar ambas metodologías para elaborar un flujo que se adapte a las necesidades del proyecto.

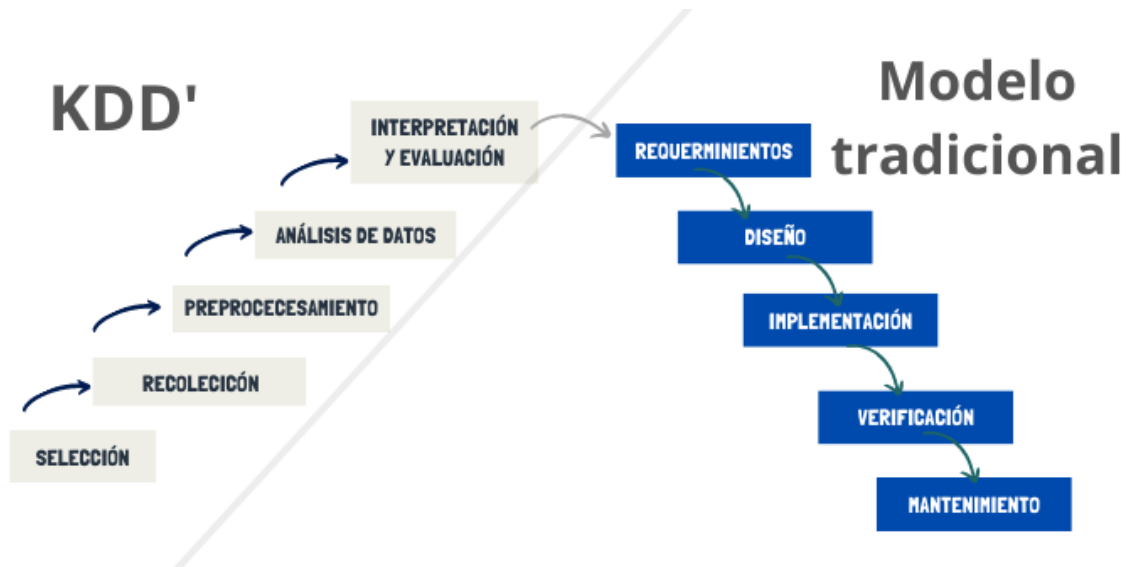


Figura 3.3: Esquema del KDD' con una metodología tradicional  
(Elaboración propia)

### 3.1. DESARROLLO DEL ANALISIS DE PERSONALIDAD

#### 3.1.1. SELECCIÓN

##### 3.1.1.1. Definición de conceptos

En esta etapa se define los tipos de personalidad que pretendemos encontrar.

Como se pudo observar en el capítulo 2 existen una variedad de definiciones de la personalidad de un individuo, y cada uno lo clasifica de una forma u otra. En este punto se puede tomar una gran variedad de opiniones respecto a la personalidad. En este caso se procede de Carl Jung y tomar 2 aspectos de la personalidad de un individuo; introvertidos y extrovertidos, Así entonces definidos como:

**Personalidad Extraversión:** Se define al extrovertido como `la persona cuyo interés en general se enfoca hacia la vida social y el mundo externo y no hacia sus experiencias internas

**Personalidad Intraversión:** Se define al introvertido como 'una persona que en general se enfoca hacia sus pensamientos y sentimientos interiores'.

También se considerara que para recolección de atributos de los usuarios de Twitter así que se procede a tomar en cuenta diferentes factores en el comportamiento del usuario en la red social, como su vocabulario, las palabras mas frecuentes, palabras negativas, palabras positivas, tipos de palabras gramaticalmente clasificadas, interacción con la red social entre otros.

Además de la personalidad de Carl Jung también se debe considerara otras investigaciones que puedan reforzar los parámetros que pueden intervenir respecto a la personalidad que demuestran los individuos en relación a su ambiente social. Por ejemplo a continuación se muestra características que contrastan las dimensiones de extroversión e introversión que se aprecia en la tabla 3.1

**Tabla 3.1 Extrovertidos e Introvertidos  
(Lo que las palabras que escoges dicen de ti, Christian Jarrett)**

<i>Extrovertidos</i>	<i>Introvertidos</i>
<ul style="list-style-type: none"> <li>• Hacen mayor ruido</li> <li>• Propensos a ser mas abstracto y "relajados"</li> <li>• Los extrovertidos también están listos para enfrentar riesgos con la precisión, espontaneidad y alcance de lo que dicen</li> <li>• Abordan una gama muy diversa de tópicos, en particular los "placenteros"</li> <li>• Tienden a hacer referencias a emociones <b>positivas</b> y situaciones <b>sociales</b> con bastante frecuencia.</li> <li>• Los extrovertidos dicen:</li> </ul>	<ul style="list-style-type: none"> <li>• No interactuan tanto</li> <li>• Se expresan con términos mas concretos</li> <li>• Utilizan más artículos al hablar (el, la, un, etc.) por definición, se refieren a objetos o eventos.</li> <li>• Son propensos a ser cautelosos al expresarse: utilizan "quizás", "a lo mejor"</li> <li>• Usan mas términos cuantificables, haciendo referencias a números específicos.</li> <li>• Los neuróticos dicen:</li> </ul>

"Estamos tan felices"	"Estoy pasándolo muy bien".
-----------------------	-----------------------------

En base a la tabla 3.1 tomaremos en cuenta diferentes aspectos que son útiles para determinar la personalidad de un individuo en la red social Twitter, así considerando diferentes atributos que pueden caracterizar a la persona en la red social Twitter, esta se listan en la tabla 3.2

**Tabla 3.2 Atributos a considerar en Twitter  
(Elaboración propia)**

<i>Atributos</i>
Likes por día
Retweets
Tweets

### 3.1.2. RECOLECCIÓN

#### 3.1.2.1. Recopilación de información

**Api de twitter:** Para usar la API de twitter es necesario una cuenta de desarrollador, para esto debemos solicitar una cuenta tal como indica en la documentación web de desarrolladores de Twitter.

A continuación se enumeran los pasos para solicitar una cuenta de desarrollador en Twitter:

- **Información de la cuenta:** En este paso simplemente llenamos algunos datos básicos de la cuenta, como nuestra ubicación o nuestro.



- **Intensión de uso:** En este paso debemos describir las razones por las que solicitamos una cuenta de desarrollador, en nuestro caso explicamos que por fines academicos usaremos la api de twitter para analizar el texto de los tweets.
- **Verificación de datos:** En este paso verificamos la información brindada
- **Terminos de uso:** Finalmente leemos los terminos y condiciones de uso y aceptamos.

Una vez terminada y aceptada nuestra solicitud podremos crear un proyecto en nuestra cuenta de twitter donde se nos otorgara 4 claves de acceso para consumir la api, que son las siguientes:

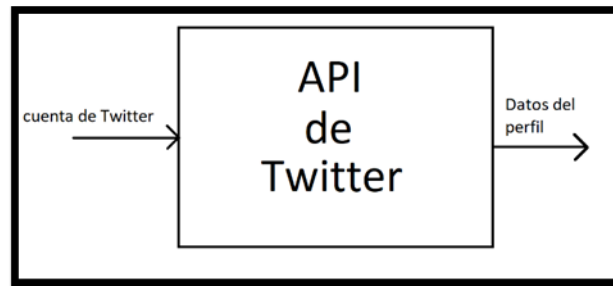
**Tabla 3.3 Credenciales para consumir la API de Twitter  
(Elaboración propia)**

<i>Nombre</i>	<i>Descripción</i>
consumer_key	Clave del consumidor
consumer_secret	Secreto del consumidor
access_token	Token de acceso
access_token_secret	Token de acceso secreto

Ya con estas credenciales se tiene acceso a la API de twitter.

### **3.1.2.2. Obtención de datos del perfil de twitter**

Para obtener la información de un perfil de Twitter, se necesita hacer una petición a la API de twitter con las claves de acceso que se obtuvo con la cuenta de desarrollador y con se obtiene datos de un usuario como se observa en la figura 3.2



**Figura 3.4: Esquema de la obtención de datos con la API de twitter**  
(Elaboración propia)

Para recolectar los datos utilizaremos la biblioteca *Tweepy* en el lenguaje de programación Python y junto con las credenciales podemos extraer información de los perfiles en Twitter. Los datos obtenidos devolverán los últimos 100 tweets y otros datos del perfil del usuario en un array tal como se muestra en la tabla 3.4

**Tabla 3.4: Ejemplo de los últimos tweets de un determinado usuario**  
(Elaboración propia)

Nº	Tweet
1	<i>Las cosas como son. Python tiene una sintaxis demasiado sencilla como para llamarse código :D #Python #programming <a href="https://t.co/ZqANzn51Ut">https://t.co/ZqANzn51Ut</a></i>
2	<i>Es muy importante entender la diferencia, sobre todo en POO programming <a href="https://t.co/SsIBTe8fcG">https://t.co/SsIBTe8fcG</a></i>
3	<i>#javascript <a href="https://t.co/FJI2IDp8yy">https://t.co/FJI2IDp8yy</a></i>
4	<i>3 años resumidos. FULL STACK DEVELOPER #developer #programming <a href="https://t.co/wfraSU4Lzn">https://t.co/wfraSU4Lzn</a></i>
5	<i>Esto si que explica como funcionan #programaciónFuncional #Javascript <a href="https://t.co/ExNDaYaPWz">https://t.co/ExNDaYaPWz</a></i>

	Amor de madre
6	#love #LoveMother <a href="https://t.co/cyTiCQGa2u">https://t.co/cyTiCQGa2u</a>
7	@DotCSV "Incluso podria estar con este cachivache que no sé para que sirve..." 🤔👉👍
8	@Maza_Design It's so beautiful &lt;3 ❤️
...	...
100	Omg <a href="https://t.co/jTv0tECLTB">https://t.co/jTv0tECLTB</a>

Además también mediante la API de twitter se solicita los datos del usuario donde obtendremos un objeto JSON como se puede ver en la tabla 3.5

**Tabla 3.5: Lista de datos obtenidos mediante la API de Twitter**

Descripción	Contenido
id'	1212419499485323264
id_str'	1212419499485323264'
name'	Eber'
screen_name'	donEber98'
location'	La Paz, Bolivia'
profile_location'	None
description'	I like turtles 🐢'
url'	<a href="https://t.co/feVe9lCst6">https://t.co/feVe9lCst6</a>

protected'	FALSO
followers_count'	14
friends_count'	68
listed_count'	0
created_at'	Wed Jan 01 17:05:27 +0000 2020'
...	...

Como se puede apreciar en la tabla 3.5 tenemos diferentes atributos que propios de un perfil de twitter, como su nombre, cuenta, likes, entre otros.

### 3.1.2.3. Test de personalidad mediante encuesta

Para obtener la interacción de un usuario con la red social Twitter se realizó una encuesta con un test para obtener los rasgos de personalidad de los usuarios. En esta encuesta se recolecto información de alrededor 30 usuarios de la carrera de Informática

Para obtener las cantidad de likes, replies, publicaciones entre otros del usuario procedemos a utilizar las funciones correspondientes de la biblioteca de Tweepy, la cual nos ofrece una variedad de atributos, sin embargo a nosotros tomamos las anteriormente mencionadas de nuestro interés.

Ahora procedemos a recolectar toda la información posible del timeline del usuario, en el cual se encuentran los tweets y comentarios. De esta forma obtendremos una lista de tweets por parte del usuario. En la tabla 3.9 podemos observar los primeros 10 datos del comportamiento de los usuarios estudiados.

**Tabla 3.6 Datos del comportamiento de usuarios en Twitter**

<i>Cantidad de</i>	<i>Cantidad de</i>	<i>Cantidad de</i>
--------------------	--------------------	--------------------

<i>Re tweets</i>	<i>tweets</i>	<i>reacciones</i>
3	3	40
6	4	66
0	0	14
2	0	35
1	0	8
5	0	9
4	0	34
5	4	38
0	0	8
1	0	9

De los mismos usuarios mediante el test de personalidad, se obtuvo los resultados como se puede ver las primeras 10 filas en la tabla 3.7

**Tabla 3.7 Niveles de extroversión e introversión**

<i>Porcentaje de Introversión (%)</i>	<i>Porcentaje de Extroversión (%)</i>
80	20
100	0
80	20
80	20
40	60
40	60
80	20
80	20
60	40
60	40

### 3.1.3. PRE-PROCESAMIENTO

A continuación a partir de la información indagaremos sobre los tweets de los usuarios, así intentando obtener información del texto de los tweets del usuario, así para su posterior análisis y consideración de atributos.

En este punto se tiene las publicaciones e información del perfil de los usuarios de twitter, sin embargo este puede contener información irrelevante, para esto procedemos a hacer una limpieza de datos.

Las publicaciones de los usuarios de twitter pueden contener información irrelevante como URLs, cadenas extra que no aportan a nuestra investigación, por lo que a través de expresiones regulares podemos eliminar los caracteres innecesarios de los tweets

Para la aplicación del preprocesamiento de datos se usa la biblioteca de expresiones regulares **re** (*regular expressions*) para python. En la tabla 3.8 se observa el proceso

**Tabla 3.8 Descripción de las expresiones regulares para el pre-procesamiento de datos**

<i>Expresión Regular</i>	<i>Descripción</i>
<code>https?://\S+ www\.\S+</code>	Encuentra las coincidencias de URLs
<code>\n</code>	Encuentra saltos de línea
<code>+</code>	Elimina varios espacios juntos
<code>^</code>	Elimina si hay espacios al final

Una vez aplicada las expresiones regulares anteriores en un tweet se observa la diferencia del ejemplo de la tabla 3.9

**Tabla 3.9. Ejemplo de entrada y salida en la limpieza de texto**

<i>Entrada</i>	<i>Salida</i>
<p>Mi papá acaba de publicar este video para que repasen geometría/matemáticas e inglés al mismo tiempo.</p> <p>¡Disfrútenlo!</p> <p><a href="https://youtu.be/uNkYf1Dc6Uk">https://youtu.be/uNkYf1Dc6Uk</a></p>	<p>Mi papá acaba de publicar este video para que repasen geometría matemáticas e inglés al mismo tiempo Disfrútenlo</p>

Así como en el ejemplo de la tabla 3.9 se aplica el mismo proceso en todos los tweets recolectados del contenido de los tweets de los usuarios.

Apartir de las publicaciones recolectadas, sea esta información desestructurada, procederemos a obtener información mejor estructurada así analizando las palabras de las publicaciones del usuario de Twitter.

En este paso usaremos el modulo de la librería NLTK llamando `regexp_tokenize` que nos ayuda a para hacer el tokenizado de las palabras de acuerdo a una expresión regular dada. En la tabla 3.10 se muestra las expresiones regulares y una descripción de lo que hace en el contenido.

**Tabla 3.10. Descripción de las expresiones regulares para el pre-procesamiento de datos**

<i>Expresión Regular</i>	<i>Descripción</i>
<code>(?:[A-Z]\.)*</code>	Encuentra abreviaciones como U.S.A.
<code>\w+(?:-\w+)*</code>	Encuentra palabras que pueden tener un guión interno
<code>\\$(?:\d+(?:\.\d+)? %)?</code>	Encuentra dinero o porcentajes como \$15.5 o 100%
<code>\.\.\.</code>	Encuentra puntos suspensivos

En la tabla 3.11 se puede ver la diferencia de un tweet antes y después de aplicar las expresiones regulares y el tokenizado. Esto nos ayuda a tener las palabras que usa el usuario de una forma más limpia.

**Tabla 3.11. Entrada y salida en la transformación de datos**

<i>Entrada</i>	<i>Salida</i>
Mi papá acaba de publicar este video para que repasen geometría matemáticas e inglés al mismo tiempo Disfrútenlo	['Disfrútenlo', 'Mi', 'acaba', 'al', 'de', 'e', 'este', 'geometría', 'inglés', 'matemáticas', 'mismo', 'papá', 'para', 'publicar', 'que', 'repasen', 'tiempo', 'video']

Una vez hecho la limpieza de los tweets del usuario, a partir de ese resultado se puede obtener otros datos de su contenido, como las palabras y su vocabulario.

A continuación indagaremos algunos otros datos que podemos obtener a partir de la información ya procesada, en este caso la gramática para obtener los diferentes tipos de palabras con las que se relaciona el usuario. Se procede a usar las categorías gramaticales universales que se muestran en la tabla 3.12

**Tabla 3.12 Convención universal de categorías gramaticales**

<i>Sigla</i>	<i>Descripción</i>
ADJ	adjetivo
ADV	adverbio
AUX	auxiliar
CCONJ	conjunción de coordinación
DET	determinante
INTJ	interjección



NOUN	nombre
NUM	numeral
PART	particula
PRON	pronombre
SYM	símbolo
VERB	verbo
X	otro

Para esto haremos uso del categorizador gramatical para el idioma español STANZA. En la tabla 3.13 podemos ver la categorización de una oración

**Tabla 3.13 Ejemplo de categorización**

<i>Tweet</i>	<i>Categoría gramatical</i>
Mi	DET
papá	NOUN
acaba	VERB
de	ADP
publicar	VERB
este	DET
video	NOUN
para	ADP
que	SCONJ
repasen	VERB
geometría	NOUN

Como se muestra en la tabla 3.13 se obtiene algunos datos del contenido y su estructura gramatical.

Algo que también se puede obtener con los tweets del usuario es obtener el vocabulario que maneja en su cuenta de Twitter. Para esto podemos analizar y recolectar el vocabulario

de un usuario. Se procede a tomar todas las publicaciones tanto *tweets* como *replies* sin considerar los retweets, ya que esa información no corresponde propiamente al del usuario.

Para recolectar el vocabulario haremos uso de diferentes estructuras de datos en python, como *diccionarios*, *conjuntos* y *pares*, así obteniendo el vocabulario que maneja.

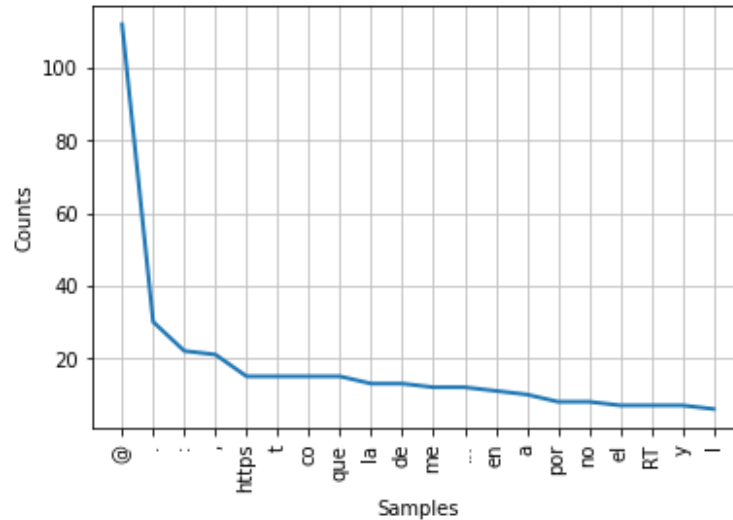
Para obtener el vocabulario es suficiente con unir todas las publicaciones del usuario, separar las palabras, introducirlos en la estructura de datos denominada **conjunto**, el cual ayuda a ordenar los elementos y elimina elementos duplicados.

Para visualizar mejor el vocabulario también se puede hacer uso de la librería de *WordCloud* la cual nos permite apreciar de una forma sencilla el vocabulario de un usuario.



**Figura 3.5: Visualización del vocabulario de una persona  
(Wordcloud)**

Algo que también se puede hacer para observar las distribución del lenguaje que maneja un usuario es utilizar el modulo *FreqDist* que una vez tokenizado nos servira para visualizar la cantidad palabras del usuario. En este caso podemos ver una grafica en la figura 3.6:



**Figura 3.6: Distribución del vocabulario de un usuario  
(Elaboración propia)**

Para tener información relevante en las palabras mas importantes que usa el usuario se procederá a eliminar las palabras mas comunes del idioma, estas palabras son llamadas las “STOPWORDS”, que son palabras tan comunes que pierden relevancia en el lenguaje y no nos sirven como para considerarlos importantes en el analisis del usuario.

Existen muchas otros metadatos que analizar, sin embargo resulta poco practico seguir indagando en otros atributos.

### 3.1.4. ANÁLISIS DE DATOS

Para el análisis de los datos recolectados y ya pre-procesados se toma en consideración algunos aspectos que dificultan una interpretación significativa de la información ya obtenida.

- Existe un alto nivel de ambigüedad en los textos obtenidos de las publicaciones de los usuarios.

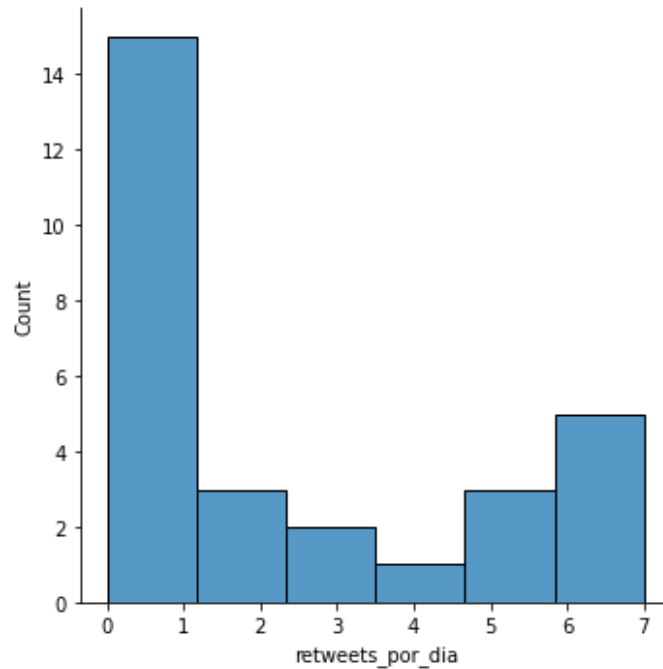
- El vocabulario de un usuario muchas veces diferentes lenguajes o incluso mezclas entre ellos, como inglés, español o spanglish
- El contenido del usuario de twitter contiene jergas propias del lenguaje, nombres propios, entre otros.

De todas formas la información ya procesada será utilizada en la aplicación de software para visualizar el comportamiento del usuario.

A continuación se puede ver los datos del usuario recolectados algunos datos en cuestión

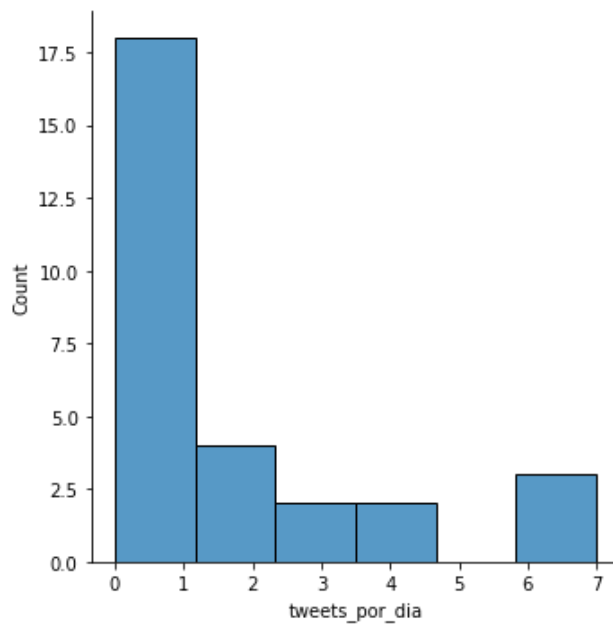
- Reacciones
- Retweets
- Tweets
- Vocabulario

En la figura 3.7 se puede ver la distribución de retweets por los usuarios de la muestra tomada en la etapa de recolección.



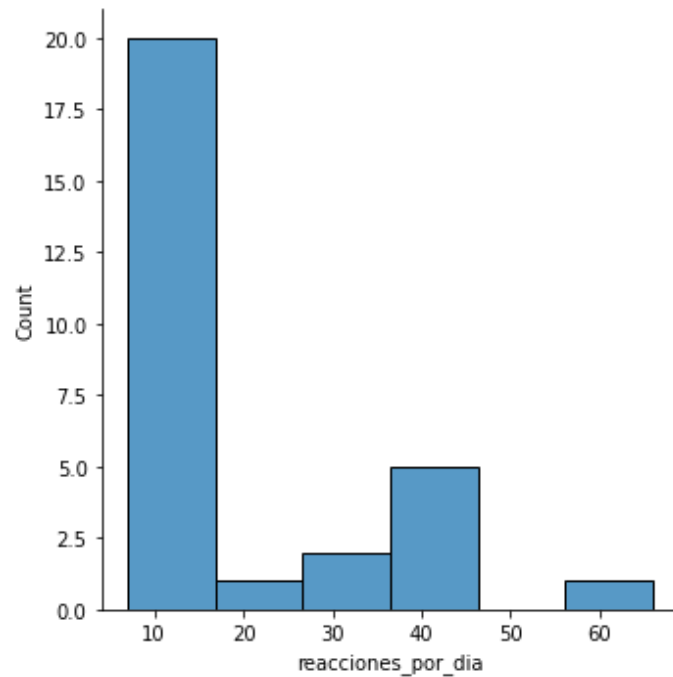
**Figura 3.7: Distribución de la cantidad de re-tweets**

En la figura 3.8 se muestra la distribución de publicaciones de los usuarios, obteniendo así una mayoría de la muestra sin muchas publicaciones



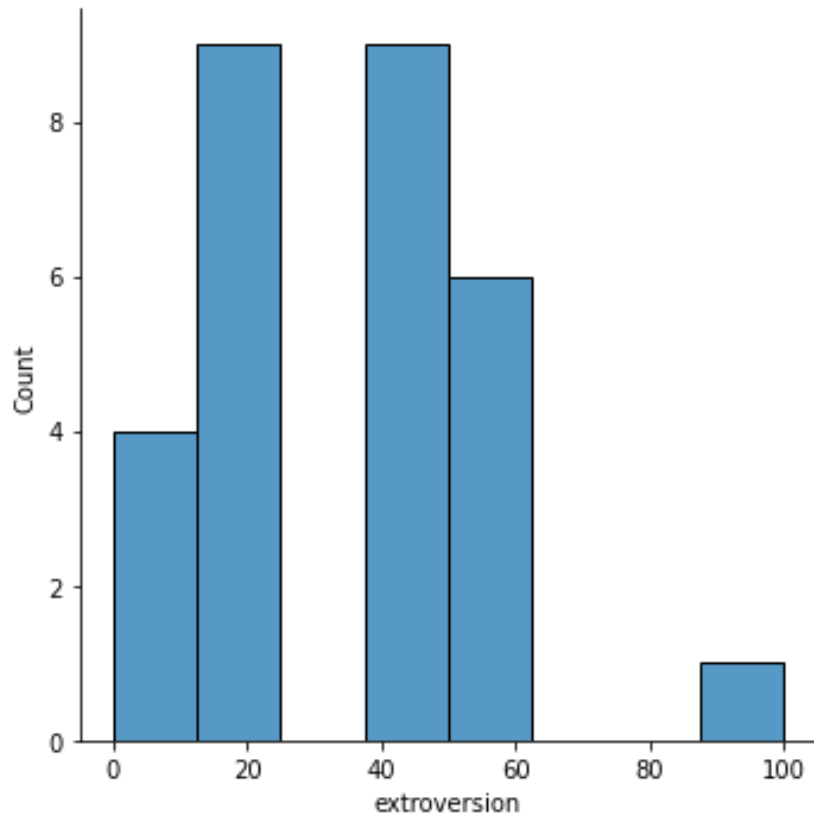
**Figura 3.8: Distribución de los tweets de los usuarios**

En la figura 3.8 se distingue la cantidad de reacciones (likes) por usuario, donde se evidencia una mayoría con usuarios sin reacciones



**Figura 3.9: Distribución de reacciones de los usuarios**

Para ver como se distribuye la cantidad de personas entre extrovertidos e introvertidos solo tomaremos a los extrovertidos para no redundar en los datos

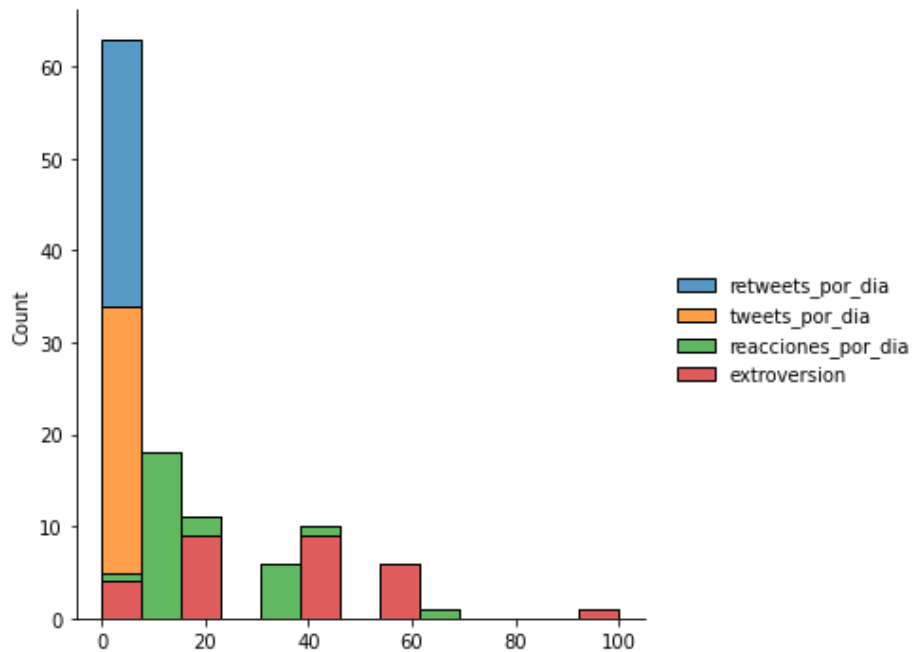


**Figura 3.10: Distribución de la personalidad de la muestra**

Con los datos de las figuras anteriores se hará un análisis de la relación entre los parámetros anteriormente recolectados y el resultado de la personalidad, así haciendo una regresión múltiple apoyándonos de una herramienta estadística, en este caso Excel.

### 3.1.5. INTERPRETACIÓN DE RESULTADOS

A continuación se compara puede ver la cantidad de usuarios que se diagnostican desde extremo introvertidos hasta extremo extrovertidos, así comparando sus diferentes tipos de actividad

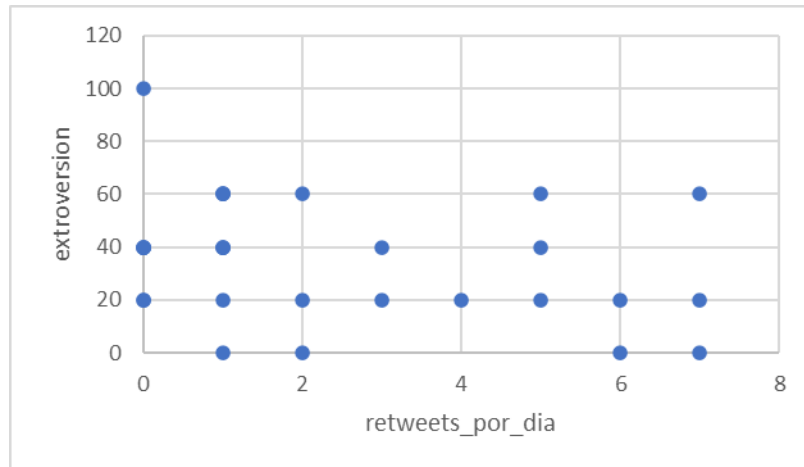


**Figura 3.11: Comparación de la actividad con la personalidad**

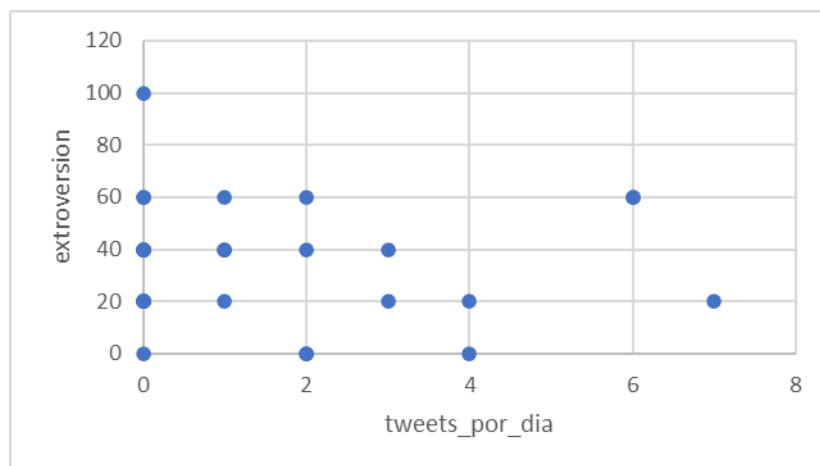
Se puede observar en el figura 3.11 la relación de los usuarios con su interacción con la red social.

Para observar mejor las relaciones entre la actividad de los usuarios con su personalidad se procede a realizar un analisis individual de los parametros con respecto a su personalidad como se puede ver en la figuras 3.12, 3.13 y 3.14

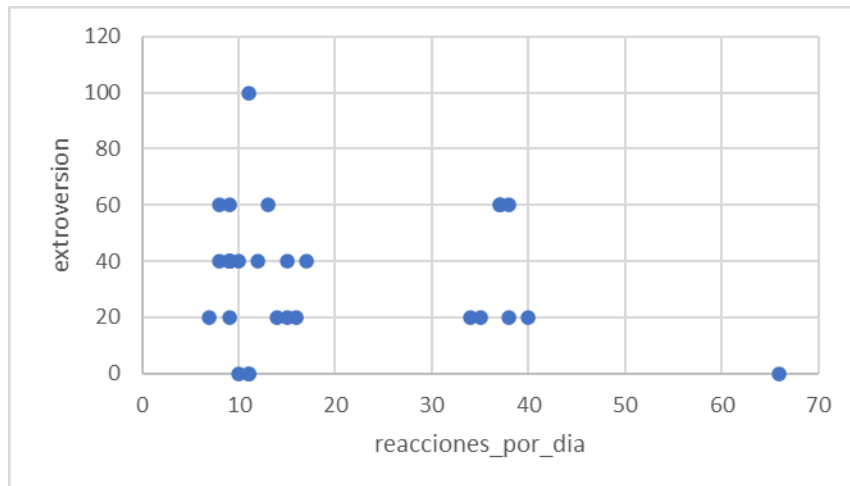




**Figura 3.12: Relación entre extroversión y los retweets de los usuarios**



**Figura 3.13: Relación entre extroversión y los tweets de los usuarios**



**Figura 3.14 Relación entre extroversión y las reacciones de los usuarios**

Así obtenidos las gráficas que apoyandonos en el analisis de datos que nos brinda Excel Si bien no se observa una relación clara entre la comparación individual, en relación a la muestra y los parametros analizados con respecto a su nivel de extraversión e introversión.

## **3.2. DESARROLLO DE LA HERRAMIENTA SOFTWARE**

### **3.2.1. REQUERIMIENTOS**

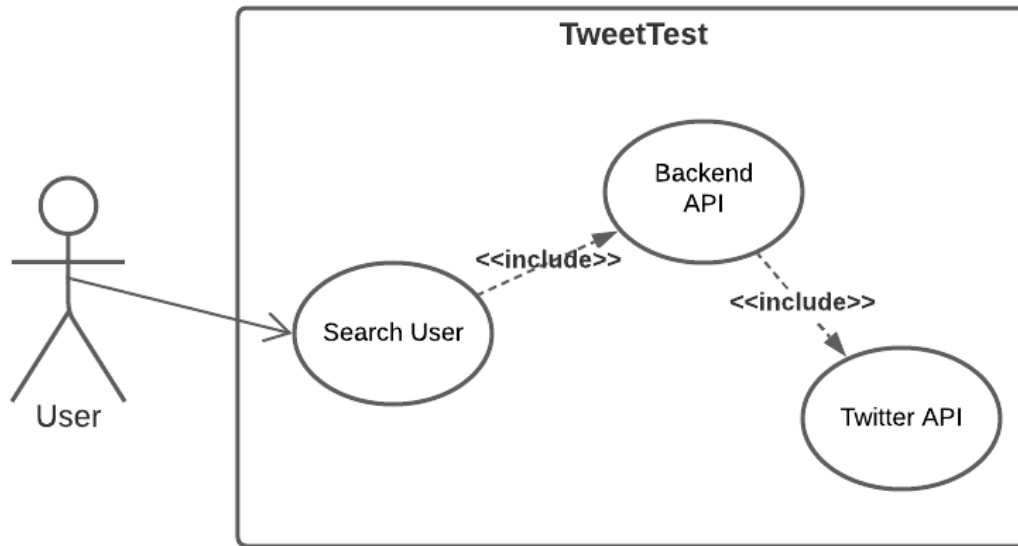
En este punto se procede a analizar los requisitos para lograr obtener diferente información a partir del perfil en base a los aspectos (parametrso analizados a en la anterior parte del proyecto). La idea es que ayude tanto a estudiantes y reclutments

#### **Requisitos funcionales**

- Crear una intefaz web
  - Crear un buscador para encontrar un determinado usuario en Twitter
  - Desarrollar una interfaz simple y amigable
- Datos del usuario
  - Ya encontrado el usuario visualizar información básica del mismo
  - Mostrar datos generales acerca del comportamiento del mismo
- Personalidad
  - Mediante la información objetidoa intuir la personalidad del usuario buscado

### **3.2.2. PLANIFICACIÓN**

#### **Diagrama de casos de uso**



**Figura 3.11 Diagrama de contexto**

Fuente: Elaboración propia

### 3.2.2.1. Diseño de pantallas

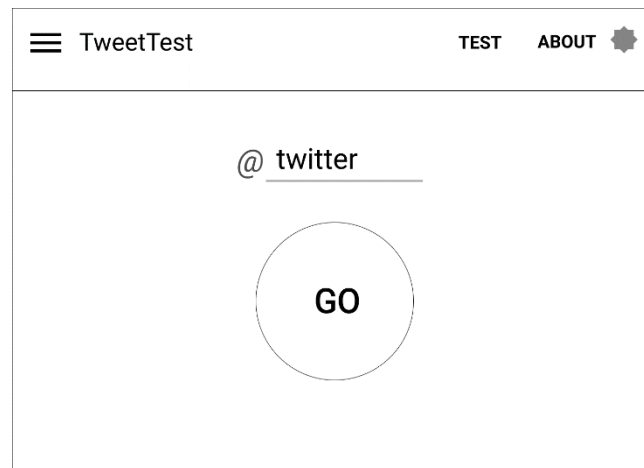
Para el desarrollo de la interfaz con un usuario final desarrollaremos un sitio web que llamaremos “TweetTest”, donde crearemos 2 vistas principales:

**Home:** Es la vista inicial del sitio web, es esta se podrá instetar un usuario de twitter para poder hacer un analisis del perfil

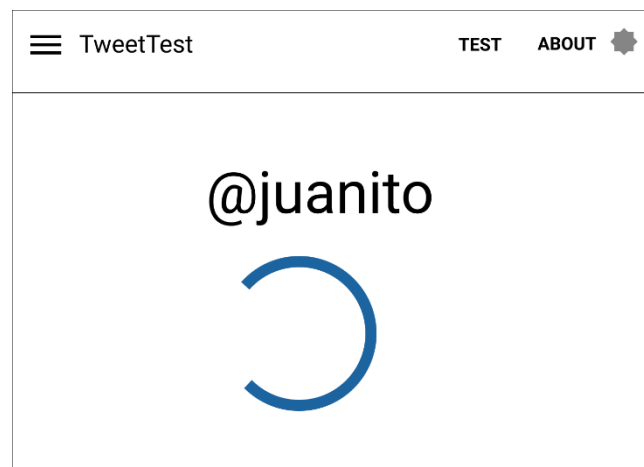
**Test:** En esta vista podremos observar el resultado final, donde se información relacionada al perfil

### 3.2.2.2. Diseño de los mockups

A continuación se observa como se puede obtener la personalidad de un individuo mediante la interfaz web



**Figura 3.12** Mockup de la vista home



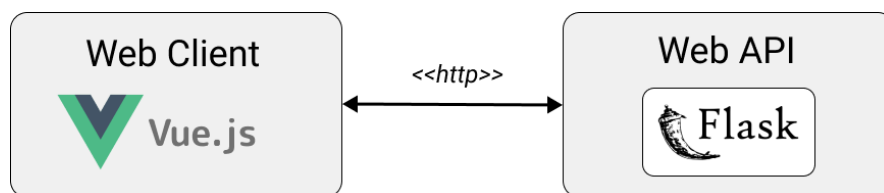
**Figura 3.13** Mockup del estado de analizando



**Figura 3.14 Mockup de la visualizacion de resultados del perfil**

### 3.2.3. Diseño de la arquitectura

Para el desarrollo de la herramienta software web usaremos una arquitectura orientada al frontend, así teniendo un backend que mediante petición HTTP envíe en formato JSON la información solicitada. Para esto se procede a escoger herramientas como son el framework para el frontend en este caso Vue.js y para el lado del backend se elige al microframework **Flask**. Así nuestro diseño arquitectónico se aprecia en la figura:



**Figura 3.15 Arquitectura SOFEA**

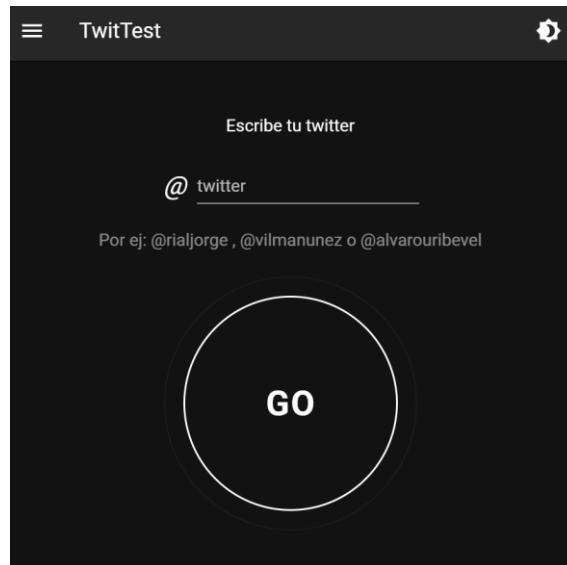
### 3.2.4. IMPLEMENTACIÓN

En este punto procedemos a elaborar los siguientes módulos basados en los requerimientos seleccionados en la etapa de requerimientos:

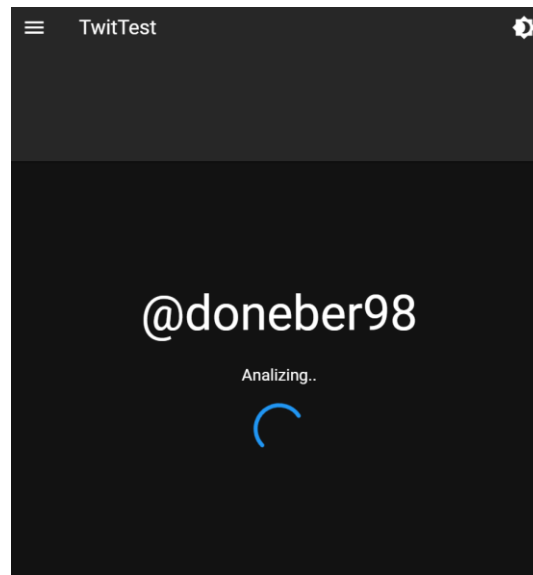
**Tabla 3.12 Listado de módulos**

Maquetar las vistas basadas en el diseño de pantallas ya desarrollado (mockups)
Desarrollar las vistas del frontend
Desarrollar la API del backend

A continuación se observa la las vistas

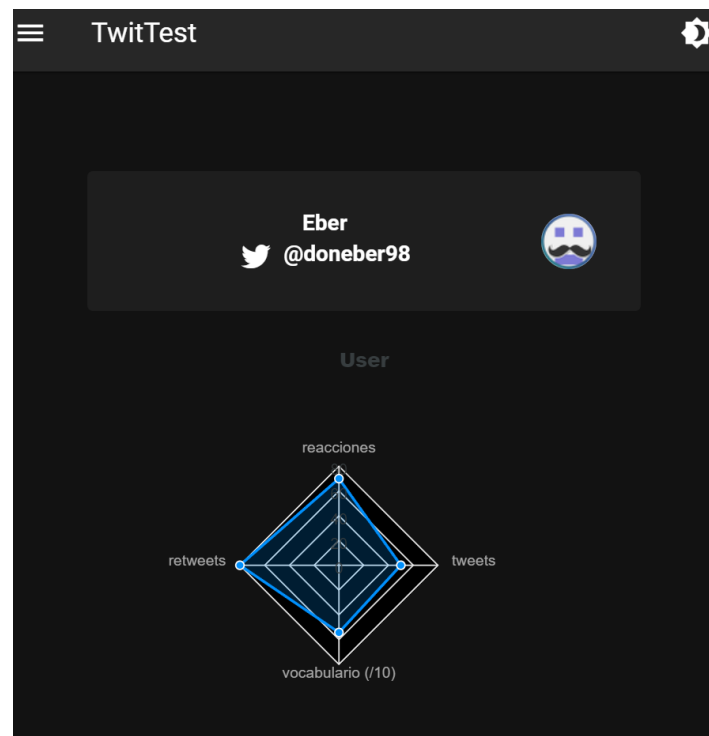
**Figura 3.16 Vista principal del sitio TweetTest con un usuario**

En la figura 3.16se observa la preparación de la cuenta de twitter a analizar



**Figura 3.17: Vista del analisis de extracción y procesado de datos**

Una vez ingresado el usuario se procede a cargar los datos



**Figura 3.18 Vista del resultado del analisis del usuario**



Finalmente obtendremos los resultados donde podremos observar una variedad de datos obtenidos del usuarios y una tentativa determinación de tipo de personalidad.

### 3.2.5. PRUEBAS

En este punto se procede a desarrollar las pruebas para determinar el funcionamiento correcto del producto software. Así usaremos las pruebas manuales o *manual tests* para corroborar el funcionamiento adecuado el producto software:

**Tabla 3.13 Listado de pruebas manuales**

Probar si el cliente frontend esta corriendo correctamente
Probar si el servidor backend esta funcionando correctamente
Probar la API del backend
Probar la visualización correcta de los datos solicitados por el usuario
Probar los datos obtenidos de la API del backend sean los esperados

Como se muestra en la tabla 3.13 se realizaron los diferentes tests para probar la funcionalidad de la herramienta de software.

## **CAPÍTULO IV**

### **PRUEBA DE HIPÓTESIS**

En este capítulo se realiza la prueba de hipótesis planteada. Para la misma se efectuará la prueba de rachas de Wald-Wolfowitz y nos apoyaremos en la distribución normal dada la muestra considerada

#### **4.1. DEFINICIÓN DE HIPÓTESIS**

A continuación procedemos a definir la hipótesis nula  $H_0$  y la hipótesis alternativa  $H_1$  para su posterior demostración:

**$H_0$ :** La interacción de un usuario en la red social Twitter no permite identificar rasgos de su personalidad

**$H_1$ :** La interacción de un usuario en la red social Twitter permite identificar rasgos de su personalidad

## 4.2. NIVEL DE SIGNIFICANCIA

Como se tiene un tamaño muestral de 20, para este caso utilizando una prueba de rachas, nos apoyamos en la distribución normal y para el nivel de significancia se toma un  $\alpha = 0.05$ .

## 4.3. ESTADÍSTICO DE PRUEBA

Para paso se tomó una muestra de 20 usuarios donde evaluamos con un test de personalidad y donde consideramos a una persona como extrovertida en caso de tener un nivel de extroversión mayor, caso contrario lo consideramos como introvertido. Así también evaluamos la aceptación con el signo + en caso de ser acertada y – en caso de ser rechazada tal como se muestra en la tabla 3.14:

**Tabla 3.14 Datos de una muestra de usuarios encuestados**

N°	Personalidad	Aceptación
1	Introvertido	+
2	Introvertido	+
3	Extrovertido	-
4	Extrovertido	-
5	Introvertido	+
6	Introvertido	-
7	Introvertido	+
8	Introvertido	+
9	Extrovertido	-
10	Introvertido	+
11	Introvertido	+
12	Extrovertido	-
13	Introvertido	+
14	Introvertido	-
15	Introvertido	+
16	Introvertido	+
17	Introvertido	-
18	Extrovertido	+
19	Introvertido	+
20	Extrovertido	-

Para el estadístico de prueba  $T$  (número total de rachas). Para esto calculamos el número de rachas que se presenta en la anterior tabla 3.14 en la columna de **Aceptación** donde el número de rachas es:

$$T = 14$$

Así también tenemos como  $n_1 = 12$  y un  $n_2 = 8$  que son los resultados positivos y negativos respectivamente.

Para obtener  $\mu$  tenemos:

$$\mu = \frac{2n_1n_2}{n_1 + n_2} + 1$$

Así remplazando los datos obtenemos un

$$\mu = 25$$

Para obtener  $\sigma$  lo calculamos con:

$$\sigma = \sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}}$$

Donde obtenemos un resultado de

$$\sigma = 39.6$$

Para obtener  $Z_c$  tenemos:

$$Z_c = \frac{T - \mu}{\sigma}$$

Y finalmente obtenemos:

$$Z_c = -0.277$$

#### 4.4. REGLA DE DECISIÓN

La regla de decisión es aceptar la hipótesis alternativa  $H_1$  y rechazar la hipótesis nula

$H_0$  si  $z_c$  se encuentra en el intervalo:

$$-z_t < z_c < z_t$$

#### 4.5. TOMA DE DECISIÓN

Obteniendo  $z_t$  con un  $\alpha = 0.05$  en la tabla normal se tiene un

$$z_t = -1.645$$

Entonces tenemos que:

$$-1.645 < -0.277 < 1.645|$$

Ya que el resultado de  $Z_c$  se encuentra dentro del intervalo  $-z_t$  y  $z_t$  se rechaza la hipótesis nula  $H_0$  y se acepta la hipótesis alternativa  $H_1$

## **CAPITULO IV**

### **CONCLUSIONES Y RECOMENDACIONES**

#### **5.1. CONCLUSIONES**

- La clasificación de personalidades basados en la teoría de la personalidad de Carl Jung llega a ser útil para distinguir y diferenciar el nivel de extroversión e introversión, y para el caso de estudio nos rasgos para categorizar la personalidad de un individuo usuario de Twitter.
- El análisis de los datos obtenidos en la red social permite obtener parámetros de usuarios de Twitter, sin embargo existen otros factores que considerar para una precisión y análisis más profundo.
- Los parámetros tomados para definir la actividad e interacción de los usuarios de Twitter no muestran un grado de correlación preciso con los tests de personalidad realizados.
- La refactorización de las metodologías aplicadas a lo largo del proyecto junto con las necesidades de la investigación llegan a ser útiles para el caso de estudio del presente proyecto.

## 5.2. RECOMENDACIONES

- Considerar otras teorías propias de la psicometría y explorar otros parámetros para una mejor inferencia de la personalidad a través de la actividad de los usuarios de Twitter.
- Aplicar técnicas de Machine Learning para detectar patrones ocultos de la actividad de los usuarios de Twitter y una mejor inferencia de su personalidad.
- Considerar utilizar a profundidad procesamiento de lenguaje natural así elaborar una investigación sobre el texto en redes sociales.
- Investigar y profundizar el estudio en otras redes sociales ya sean redes sociales horizontales o verticales además de contrastar los resultados entre las mismas.



## BIBLIOGRAFÍA

- *Anahí Aradas, (23 marzo 2012), Lo que dicen de ti las redes sociales. Recuperado de [https://www.bbc.com/mundo/noticias/2012/03/120323\\_tecnologia\\_secretos\\_redes\\_social\\_es\\_aa](https://www.bbc.com/mundo/noticias/2012/03/120323_tecnologia_secretos_redes_social_es_aa)*
- *Blanca Nogrady, (20 noviembre 2016), Lo que tu muro de Facebook y otras redes sociales revelan de tu salud mental. Recuperado de <https://www.bbc.com/mundo/vert-fut-37821742>*
- *Martin Hilbert, (22 agosto 2017), Big Data, redes sociales, y su influencia en la sociedad. Recuperado de <https://oceano.uy/justiciainfinita/entrevistas/8175-martin-hilbert-en-justicia-infinita>*
- *Rafael Caballero, Enrique Martín, Adrián Riesco, (2019), Big Data con Python. Recolección, almacenamiento y proceso*
- *Estudiantes de Medicina, Universidad Autónoma Metropolitana, 2015, Escrutinio de depresión y ansiedad en Twitter*
- *Timarán-Pereira, S. R., Hernández-Arteaga, I., Caicedo-Zambrano, S. J., Hidalgo-Troya, A. y AlvaradoPérez, J. C. (2016). El proceso de descubrimiento de conocimiento en bases de datos. En Descubrimiento de patrones de desempeño académico con árboles de decisión en las competencias genéricas de la formación profesional (pp. 63-86). Bogotá: Ediciones Universidad Cooperativa de Colombia. doi: <http://dx.doi.org/10.16925/9789587600490>*
- *Javier Landa, (19 febrero 2016), KDD y Minería de Datos. Recuperado de Manuel Montes y Gómez, (2001), Minería de texto: Un nuevo reto computacional*
- *Salesforce Latinoamérica, (2017), Recuperadodeht*

- RAE, 2019, *Real Academia de la Lengua Española*. Recuperado de <https://definicion.de/twitter/>
- Julián Pérez Porto y Ana Gardey, (2010), *Definición de Twitter*. Recuperado de <https://definicion.de/twitter/>
- Pereira, S. R. T., Arteaga, I. H., Zambrano, S. J. C., Troya, A. H., & Pérez, J. C. A. (s. f.). *Descubrimiento de patrones de desempeño académico* [Text.Book]. Ediciones Universidad Cooperativa de Colombia. Recuperado 3 de septiembre de 2020, de <https://ediciones.ucc.edu.co/index.php/ucc/catalog/book/36>
- IEEE/ISO/IEC 12207-2017—ISO/IEC/IEEE International Standard—Systems and software engineering—Software life cycle processes. (s. f.). Recuperado 2 de septiembre de 2021, de <https://standards.ieee.org/standard/12207-2017.html>
- Adel Alshamranil y Abdullah Bahattab, (2015), *A Comparison Between Three SDLC Models Waterfall Model, Spiral Model, and Incremental/Iterative Model*. College of Telecommunications and Electronics.
- *Ciclo de vida del software: Todo lo que necesitas saber*. (s. f.). Intelequia. Recuperado 2 de septiembre de 2020, de <https://intelequia.com/blog/post/2083/ciclo-de-vida-del-software-todo-lo-que-necesitas-saber>
- Caules, C. Á. (2020, enero 31). *SOFEA , Arquitecturas y flexibilidad*. Arquitectura Java. <https://www.arquitecturajava.com/sofea-arquitecturas-y-flexibilidad/>
- Sommerville, Ian. *Software Engineering* 9th edition. 2011.
- *API - Glosario | MDN*. (s. f.). Recuperado 27 de septiembre de 2021, de <https://developer.mozilla.org/es/docs/Glossary/API>
- *REST - Glosario | MDN*. (s. f.). Recuperado 27 de septiembre de 2021, de <https://developer.mozilla.org/es/docs/Glossary/REST>

# DOCUMENTACIÓN

La Paz, Abril de 2021

Señor

Ph.D. José María Tapia Baltazar  
**Director**  
**Carrera de Informática**

Presente

**Ref. Aval para Defensa de Tesis de Grado**

De mi mayor consideración

Por intermedio de la presente, y en mi calidad de Tutor Metodológico, tengo a bien dirigirme a su autoridad, para darle a conocer que luego de efectuar el seguimiento a la estructura y contenido de la Tesis de Grado, titulado “**DETECCIÓN DE PERSONALIDAD EN RED SOCIAL TWITTER**”, elaborado por el postulante Eber Edgar Quenta Lopez con C.I. 9875596 LP, y habiendo presentado también el Aval de su Asesor, me corresponde **dar mi CONFORMIDAD Y AVAL**, para que el mismo proceda a la **DEFENSA PÚBLICA**, de acuerdo a normas y reglamentos universitarios vigentes.

Sin otro particular, me despido de usted con las consideraciones más distinguidas.

Atentamente



M.Sc. Franz Cuevas Quiroz|  
**Tutor Metodológico**

av/orac

cc/archivo

La Paz, Agosto de 2021

Señor

Ph.D. José María Tapia Baltazar  
**Director**  
**Carrera de Informática**

Presente

**Ref. Aval para Defensa de Tesis de Grado**

De mi mayor consideración

Por intermedio de la presente, y en mi calidad de Asesor Metodológico, tengo a bien dirigirme a su autoridad, para darle a conocer que luego de efectuar el seguimiento a la estructura y contenido de la Tesis de Grado, titulado “**DETECCIÓN DE PERSONALIDAD EN RED SOCIAL TWITTER**”, elaborado por el postulante Eber Edgar Quenta Lopez con C.I. 9875596 LP, me corresponde **dar mi CONFORMIDAD Y AVAL**, para que el mismo proceda a la **DEFENSA PÚBLICA**, de acuerdo a normas y reglamentos universitarios vigentes.

Sin otro particular, me despido de usted con las consideraciones más distinguidas.

Atentamente



M.Sc. Aldo Ramiro Valdez Alvarado  
**Asesor Metodológico**

av/orac  
cc/archivo