



ESPE

UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN

CARRERA DE INGENIERÍA EN SISTEMAS E INFORMÁTICA

**TRABAJO DE TITULACIÓN, PREVIO LA OBTENCIÓN DEL TÍTULO DE
INGENIERO EN SISTEMAS E INFORMÁTICA**

**TEMA: PREDICCIÓN DE PERSONALIDAD DE PERFILES DE
FACEBOOK UTILIZANDO APRENDIZAJE AUTOMÁTICO COMO
APOYO A LA SELECCIÓN DE PERSONAL DE TALENTO HUMANO**

AUTOR: SARCHI ALBUJA, ELVIS RAMIRO

DIRECTOR: ING. GUERRERO IDROVO, GRACIELA

SANGOLQUÍ

2018



DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN
CARRERA DE INGENIERÍA EN SISTEMAS E INFORMÁTICA

CERTIFICACIÓN

Certifico que el trabajo de titulación, "PREDICCIÓN DE PERSONALIDAD DE PERFILES DE FACEBOOK UTILIZANDO APRENDIZAJE AUTOMÁTICO COMO APOYO A LA SELECCIÓN DE PERSONAL DE TALENTO HUMANO" fue realizado por el señor Elvis Ramiro Sarchi Albuja, el mismo que ha sido revisado en su totalidad, analizado por la herramienta de verificación de similitud de contenido; por lo tanto cumple con los requisitos teóricos, científicos, técnicos, metodológicos y legales establecidos por la Universidad de Fuerzas Armadas ESPE, razón por la cual me permito acreditar y autorizar para que lo sustente públicamente.

Sangolquí, 6 de marzo de 2018.

Firma:
Ing. Graciela Guerrero MSc.
C. C. 1720513322



DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN

CARRERA DE TECNOLOGÍAS DE LA INFORMACIÓN (SISTEMAS E INFORMÁTICA)

AUTORÍA DE RESPONSABILIDAD

Yo, **SARCHI ALBUJA, ELVIS RAMIRO**, declaro que el contenido, ideas y criterios del trabajo de titulación: **PREDICCIÓN DE PERSONALIDAD DE PERFILES DE FACEBOOK UTILIZANDO APRENDIZAJE AUTOMÁTICO COMO APOYO A LA SELECCIÓN DE PERSONAL DE TALENTO HUMANO**, es de mi autoría y responsabilidad, cumpliendo con los requisitos teóricos, científicos, técnicos, metodológicos y legales establecidos por la Universidad de Fuerzas Armadas ESPE, respetando los derechos intelectuales de terceros y referenciando las citas bibliográficas.

Consecuentemente el contenido de la investigación mencionada es veraz.

Sangolquí, 06 de marzo del 2018.

ELVIS RAMIRO SARCHI ALBUJA

C.C.: 172231360-6



DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN
CARRERA DE TECNOLOGÍAS DE LA INFORMACIÓN (SISTEMAS E
INFORMÁTICA)
AUTORIZACIÓN

Yo, **SARCHI ALBUJA, ELVIS RAMIRO** autorizo a la Universidad De Las Fuerzas Armadas **ESPE**, publicar el trabajo de titulación: **PREDICCIÓN DE PERSONALIDAD DE PERFILES DE FACEBOOK UTILIZANDO APRENDIZAJE AUTOMÁTICO COMO APOYO A LA SELECCIÓN DE PERSONAL DE TALENTO HUMANO** en el repositorio Institucional, cuyo contenido, ideas y criterios son de mi responsabilidad.

Sangolquí, 06 de marzo del 2018.

ELVIS RAMIRO SARCHI ALBUJA

C.C.: 172231360-6

DEDICATORIA

A mi familia, que supo guiarme por el buen camino en especial a mi madre que me ayudo en cada momento de necesidad.

A mis profesores que me formaron, no solo en lo académico sino en lo personal para ser un profesional con valores y ética.

A todos mis amigos que estuvieron ahí ayudándome en los trabajos y enseñándome cosas nuevas cada día, sobre todo a aquella persona que estuvo a mi lado apoyando hasta en los momentos más difíciles.

Elvis Sarchi

AGRADECIMIENTO

A mi madre, por darme la vida y ser un ejemplo de esfuerzo, con el cual se puede conseguir los éxitos de la vida.

A todas las personas que me apoyaron y creyeron en mi para lograr cumplir este reto.

Elvis Sarchi

ÍNDICE

CERTIFICACIÓN	ii
AUTORÍA DE RESPONSABILIDAD.....	iii
AUTORIZACIÓN	iv
DEDICATORIA.....	v
AGRADECIMIENTO	vi
ÍNDICE DE TABLAS	x
ÍNDICE DE FIGURAS	xi
RESUMEN.....	xiii
ABTRACT	xiv
CAPÍTULO I	1
INTRODUCCIÓN.....	1
1.1 Antecedentes.....	1
1.2 Problemática.....	2
1.3 Justificación	3
1.4 Objetivos.....	4
1.5 Alcance.....	5
CAPÍTULO II	8
MARCO TEÓRICO Y TRABAJOS RELACIONADOS.....	8
2.1. Redes Sociales (Facebook).....	8
2.2. Cinco grupos de la Personalidad	11
2.2.1. Apertura a la experiencia	12
2.2.2. Cordialidad, Amabilidad	13
2.2.3. Extraversión	14

2.2.4. Responsabilidad	14
2.2.5. Neuroticismo	14
2.3. Palabras más usadas según la personalidad	15
2.4. Personalidad y Facebook	16
2.5. Inteligencia artificial	18
2.6. Api Graph	19
2.7. El inventario de los cinco grandes	21
2.8. Trabajos Relacionados	21
2.8.1. Apply Magic Sauce	21
CAPÍTULO III	23
DESARROLLO	23
3.1. Introducción	23
3.2. Metodología XP	23
3.3. Desarrollo aplicación Facebook	24
3.4. Desarrollo Página Web	25
3.4.1. Plataforma Heroku	25
3.4.2. Configuración de la plataforma	26
3.4.3. Configuración del Botón de Inicio de Facebook.....	26
3.4.4. Creación de la Base de datos.....	27
3.4.5. Extracción de datos del Usuario de Facebook.....	31
3.4.6. Inserción de las palabras según su personalidad.....	34
3.4.7. Análisis del texto de las publicaciones	43
3.5. Test de Personalidad	44
3.5.1. Resultado Test de Personalidad.....	46

3.5.2. Aplicación del test de personalidad a Usuarios	47
3.5.3. Resultados obtenidos del test de personalidad	49
CAPITULO IV.....	52
ANÁLISIS DE DATOS	52
4.1. Weka.....	52
4.2. Pre procesamiento de Datos.....	54
4.3. Clasificación.....	56
4.3.1. Random Forest.....	62
4.3.2. Kstar	64
4.4. Clusterización	66
4.4.1. Canopy.....	66
4.4.2. Filtered Clustered.....	68
4.4.3. Hierarchical Clusterer	70
4.4.4. Simple K means	71
4.5. Regla de Asociación.....	73
4.6. Discusión.....	79
CAPÍTULO V.....	81
CONCLUSIONES Y FUTURAS LÍNEAS DE INVESTIGACIÓN	81
5.1. Conclusiones	81
5.2. Recomendaciones.....	81
5.3. Futuras Líneas de Investigación	82
BIBLIOGRAFÍA.....	84

ÍNDICE DE TABLAS

Tabla 1 <i>Palabras más usadas según la personalidad</i>	16
Tabla 2 <i>Tabla Usuario</i>	28
Tabla 3 <i>Tabla de Palabras según su personalidad</i>	30
Tabla 4 <i>Palabras de la personalidad amabilidad</i>	35
Tabla 5 <i>Palabras de la personalidad abierto a la experiencia</i>	37
Tabla 6 <i>Palabras de la personalidad extrovertido</i>	38
Tabla 7 <i>Palabras de la personalidad neurótico</i>	40
Tabla 8 <i>Palabras de la personalidad responsable</i>	42
Tabla 9 <i>Preguntas del Test de Personalidad</i>	44
Tabla 10 <i>Atributos del archivo .cvs</i>	55
Tabla 11 <i>Clasificación de los algoritmos</i>	57
Tabla 12 <i>Funcionamiento de algoritmos de Weka</i>	58
Tabla 13 <i>Resultados Filtered Clustered</i>	69

ÍNDICE DE FIGURAS

Figura 1 Arquitectura del Sistema.....	6
Figura 2 Esquema de Clasificación	7
Figura 3 Usuarios de Internet y Redes Sociales en Ecuador 2017.....	10
Figura 4 Ranking Comparativo 2015 - 2017 Redes Sociales Web Ecuador.....	11
Figura 5 Inicio de sesión de la aplicación	31
Figura 6 Ventana de diálogo de Facebook	32
Figura 7 Ventana de información de Permisos	33
Figura 8 Test de Personalidad.....	46
Figura 9 Resultado del Test de Personalidad.....	47
Figura 10 Escenario de Aplicación del Test de Personalidad	48
Figura 11 Usuario realizando el Test de Personalidad	48
Figura 12 Visualización de los resultados del Test de Personalidad	49
Figura 13 Agrupamiento por Personalidad	50
Figura 14 Clasificación por género	50
Figura 15 Clasificación por Género y Personalidad.....	51
Figura 16 Técnicas de Minería de Datos	53
Figura 17 Numérico de los cinco tipos de personalidad	56
Figura 18: Atributo con relación a la personalidad de los usuarios.....	56
Figura 19 Ejecución de Algoritmos de 118 instancias aprendidas.....	59
Figura 20 Árbol de decisiones.	60
Figura 21 Escenario 2: Ejecución de Algoritmos de 35 instancias aprendidas.....	62
Figura 22 Errores generados de la ejecución de algoritmos Ramdom	63
Figura 23 Errores generados de la ejecución del algoritmo Kstar	65
Figura 24 Evaluación del algoritmo Canopy mediante el género.....	68
Figura 25 Resultado del Filtered Clustered.....	68
Figura 26 Filtered Clustered en agrupamiento por género	69
Figura 27 Clusterización Hierarchical por el género y la personalidad	71
Figura 28: Simple Kmeans (Abierto a la experiencia y Extroversión)	73

Figura 29 Personalidad con el número de post	74
Figura 30 Personalidad con numero de shares	74
Figura 31 Personalidad con el número de fotos	75
Figura 32 Personalidad con el número de actualizaciones.....	76
Figura 33 Personalidad con el peso de las palabras enmarcadas en neuróticas	76
Figura 34 Personalidad con el peso de las palabras enmarcadas en extrovertidas	77
Figura 35 Personalidad con el peso en abiertos a la experiencia.....	78
Figura 36 Personalidad con el peso en responsables	78
Figura 37 Personalidad según su género	79

RESUMEN

El presente trabajo de titulación tiene como objetivo predecir la personalidad del perfil de Facebook de una persona, mediante el aprendizaje automático para facilitar la elección de candidatos en los Recursos Humanos. Para ello se identificaron los atributos que se extraen de Facebook, con lo cual fue posible la predicción de la personalidad, los datos se extraen mediante la API Graph de Facebook, la cual se implementó en una página web. Para lograr la base de conocimientos del aprendizaje automático se implementó el test de personalidad BFI a los estudiantes próximos a egresar de la Universidad de las Fuerzas Armadas ESPE, con el fin de realizar el entrenamiento y la clasificación de modelos del aprendizaje automático mediante el uso de la herramienta Weka, para verificar el grado de precisión de los algoritmos usados en la predicción de la personalidad del usuario.

PALABRAS CLAVE:

- **APRENDIZAJE AUTOMÁTICO**
- **PERFILES DE FACEBOOK**
- **API GRAPH**
- **PREDICCIÓN DE PERSONALIDAD**
- **RECURSOS HUMANOS**

ABSTRACT

This work titration has the general objective to predict the personality of user Facebook profile, through machine learning to facilitate the election of applicants in Human Resources. To do this, the attributes taken from Facebook were identified for possible prediction of personality, the data is extracted through the Graph API Facebook, which was implemented on a web page. The knowledge base of machine learning was created from the BFI personality test to students near the graduation of the University of the Armed Forces ESPE, in order to perform the training and classification of machine learning models through the use Weka tool, to verify the degree of accuracy of the algorithms used in the prediction of the user's personality.

KEYWORDS:

- **MACHINE LEARNING**
- **FACEBOOK PROFILES**
- **API GRAPH**
- **PREDICTION PERSONALITY**
- **HUMAN RESOURCES**

CAPÍTULO I

INTRODUCCIÓN

1.1 Antecedentes

Las redes sociales han cambiado la manera en que la gente se comunica e interactúa, también ha impactado en los negocios. Según (Bachrach Y, 2015), una tendencia reciente es que las empresas utilizan las redes sociales como parte del proceso de contratación de nuevos aspirantes. Las redes sociales se han convertido en la carta de presentación para muchos, y también se ha convertido en una manera de averiguar acerca de alguien que acabas de conocer en especial la red social Facebook, gracias a los likes (“me gusta”) que realiza el usuario como son: libros, películas, música, intereses, política, etc. Todas estas interacciones que realiza el usuario en su perfil de Facebook y compartirlo al público, permite saber cuáles son sus gustos y tendencias políticas.

Facebook, desde su creación en el 2004 ha experimentado un crecimiento notable, ahora cuenta con más de seis millones de cuentas de estudiantes universitarios de USA, con un adicional de 20.000 nuevas cuentas creadas diariamente (Samuel D. Gosling, 2007) . De esta muestra el 60% de usuarios se conectan diariamente, la gente está cada vez más conectada y emplean su tiempo en las redes sociales dejando su huella digital.

Según (Kornblum, 2006), Facebook se ha ganado a los medios de comunicación para su uso en la ayuda de investigaciones criminales y las audiencias disciplinarias de las universidades, en eso se beneficia la investigación para el proceso legal y facilitar a la policía o personas que toman parte en la operación sobre la ley. También se ha usado para examinar a los solicitantes de empleo (Brady, 2006).

Otra de las aplicaciones para las redes sociales en la actualidad es el análisis forense, en muchos países desarrollados como: Japón, países europeos y los Estados Unidos han aplicado los conocimientos científicos y la tecnología existente para identificar evidencias en los juicios penales para obtener verdaderos resultados científicos, especialmente en Japón más del 90% de los asesinos son arrestados mediante la

aplicación de equipos científicos y la tecnología innovada para ser utilizados en ciencia forense (Mahasak Ketcham, 2016).

Según (Samuel D. Gosling, 2007) indica que la personalidad de un individuo puede expresarse en notificaciones de identidad, estas son un intento de transmitir cómo les gustaría ser vistos, por ejemplo, van desde pistas sutiles que se encuentran en la elección de la ropa de un individuo a demandas más directas como pegatinas de parachoques o declaraciones verbales. Otro tipo, es el residuo de la conducta esta se refiere a las pistas dadas por el comportamiento de una persona. Por ejemplo, una colección de películas bien organizadas refleja la tendencia de un individuo a ser organizado.

Por otro lado (Chiavenato, 2007), las empresas usan los test de personalidad los cuales sirven para analizar los distintos rasgos de la personalidad, sean estos determinados por el carácter o por el temperamento. Un rasgo de personalidad es una característica señalada del individuo capaz de distinguirlo de los demás. Los test de personalidad son genéricos y revelan rasgos generales de la personalidad, existen varios test como son: psicodiagnósticos, test proyectivos, de Rorschach, test de percepción temática, test de árbol (Koch, 1957), test de figura humana de (Machover, 1976), etcétera. A estos test de personalidad se los llama específicos cuando se investiga rasgos o aspectos determinados de la personalidad, como el equilibrio emocional, las frustraciones, los intereses, la motivación, etc.

1.2 Problemática

El departamento de recursos humanos es el responsable de la gestión del personal de la organización (Chiavenato, 2007) para ello tiene como objetivos seleccionar y formar a las personas que la empresa necesita. Por lo tanto, una persona calificada es una persona preparada, capaz de realizar un determinado trabajo, que dispone de todas las competencias profesionales que se requieren en ese puesto. Para ello existen distintos pasos los cuales son: i) determinar si el candidato cumple con las competencias mínimas predeterminadas para el puesto de trabajo, ii) evaluar las competencias tanto técnicas como psicológicas, iii) asignar un puntaje a las evaluaciones efectuadas anteriormente y

decidir a quién se le ofrece el puesto (Chiavenato, 2007).

La conducta empresarial es fundamental para las organizaciones de hoy, se debe crear grupos de trabajo armoniosos donde el jefe y los empleados tengan buenas relaciones, para ser trabajadores productivos y no conflictivos. La mayoría de los empleados nuevos no logran mantener su puesto de trabajo debido a su actitud y no por su formación o experiencia.

El problema existe en el reclutamiento a los candidatos cuando éstos tienen una trayectoria desconocida, a pesar de las técnicas de selección o de predicción, las empresas aceptan al personal con un contrato por un periodo experimental o de prueba debido a la inseguridad del proceso (Chiavenato, 2007).

En el presente documento da una propuesta para determinar la personalidad del individuo mediante aprendizaje automático en los perfiles de Facebook de los candidatos para un puesto de trabajo, siendo una herramienta de apoyo para el departamento de Recursos Humanos en la selección y reclutamiento del personal para la empresa.

1.3 Justificación

En la actualidad, el departamento de Recursos Humanos de una empresa se ha convertido en un eje fundamental, ya que su principal función es la del reclutamiento, selección, organización y planificación del personal. Debido a que el corazón de una empresa son los empleados estos deben ser elegidos cuidadosamente con el fin de que puedan adaptarse a las necesidades de la empresa tanto en su misión y visión de esta. Al tener empleados que puedan desenvolverse y relacionarse con el resto de compañeros, permite que la empresa obtenga mejores resultados por lo tanto más ganancias.

Sin embargo, el departamento de Recursos Humanos, para la selección de los nuevos aspirantes lo hacen mediante el uso del currículum y una entrevista para medir sus actitudes y conocimientos. Pero no llegan a conocer a la persona en su totalidad, si esta persona que ingresa a la empresa no se puede comunicar bien con sus compañeros o si en el área de trabajo que se encuentra no puede desenvolverse correctamente, esto podría producir pérdidas para la empresa y malestar en los compañeros de trabajo.

Según (Bachrach, 2015) los datos que provee Facebook son más ricos que los de un currículo, debido a que en este se incluyen datos como pensamientos, opiniones y sentimientos, lo que permite determinar la personalidad del usuario de Facebook. Con esta herramienta se facilita el trabajo del departamento de recursos humanos y brindarle un panorama más amplio del postulante al trabajo para la empresa y de esta manera ahorrar dinero a la empresa en la contratación de personas que no cumplan con los requisitos para trabajar en ella.

Según la página Keyword Research, Competitor Analysis, & Website Ranking | Alexa (Cooper, 2017), Facebook ocupa el cuarto lugar de las páginas más visitadas en el Ecuador, estando en primer lugar el motor de búsquedas de Google (google.com.ec), segundo la página de alojamiento de videos Youtube.com y en tercer lugar nuevamente se muestra el motor de búsquedas Google.com(google.com.ec). Sin embargo, se ha descartado a estos 3 primeros lugares por no ser redes sociales completas (fotografías, post, likes, videos, streaming, etc).

Facebook es la red social más visitada por los ecuatorianos, los cuales gastan un promedio de 11 minutos revisando y registrando sus actividades en esta red social, esto permite que la predicción de la personalidad mediante el uso del perfil de Facebook sea viable. Por lo tanto, este es el motivo que por lo que, en la presente investigación se ha tomado Facebook como red social, en la que se evaluará los 5 perfiles de personalidad orientado a brindar apoyo a las empresas de recursos humanos, en cuanto al estudio de postulantes que se encuentren aplicando a una oportunidad laboral.

1.4 Objetivos

1.4.1 Objetivo General

Predecir la personalidad del perfil de Facebook de una persona, mediante aprendizaje automático como herramienta de soporte para el proceso de selección de candidatos en el área de Recursos Humanos.

1.4.2 Objetivos Específicos

- Identificar los atributos que son necesarios para predecir la personalidad de usuario de Facebook mediante características obtenidas de análisis de los trabajos realizados.
- Obtener información de las publicaciones realizadas en el muro del usuario, mediante el desarrollo de una aplicación Facebook.
- Implementar el test de personalidad BFI a los estudiantes próximos a egresar, que permita llenar la base de conocimientos necesario para el aprendizaje automático.
- Aplicar el entrenamiento y clasificación mediante modelos de aprendizaje automático que permitan arrojar un análisis de la predicción de la personalidad del usuario de Facebook.
- Analizar los datos obtenidos mediante el uso de la herramienta Weka, para verificar el grado de precisión en las predicciones de la personalidad.

1.5 Alcance

El proyecto se llevará a cabo hasta la implementación de un prototipo teniendo como funcionalidades las siguientes actividades:

- Implementación de una aplicación Facebook que permita el acceso y recolección de datos del usuario mediante la API Graph
- Implementación del test de personalidad BFI que permita llenar la base de conocimientos necesarios para el aprendizaje automático.
- Realización del entrenamiento y clasificación mediante modelos de aprendizaje automático que permitan realizar el análisis de la predicción de la personalidad según las cinco grandes personalidades.
- Validación y análisis de los datos obtenidos mediante el uso de la herramienta Weka.

El prototipo será desarrollado en un ambiente web y los datos obtenidos de Facebook se los realizará a estudiantes próximos a egresar de la carrera de Ingeniería en Sistemas e Informática de la Universidad de las Fuerzas Armadas ESPE, debido a que son futuros candidatos de postulación en una empresa y

dedicarse al mundo laboral, la arquitectura del prototipo del sistema es la siguiente (ver figura 1).

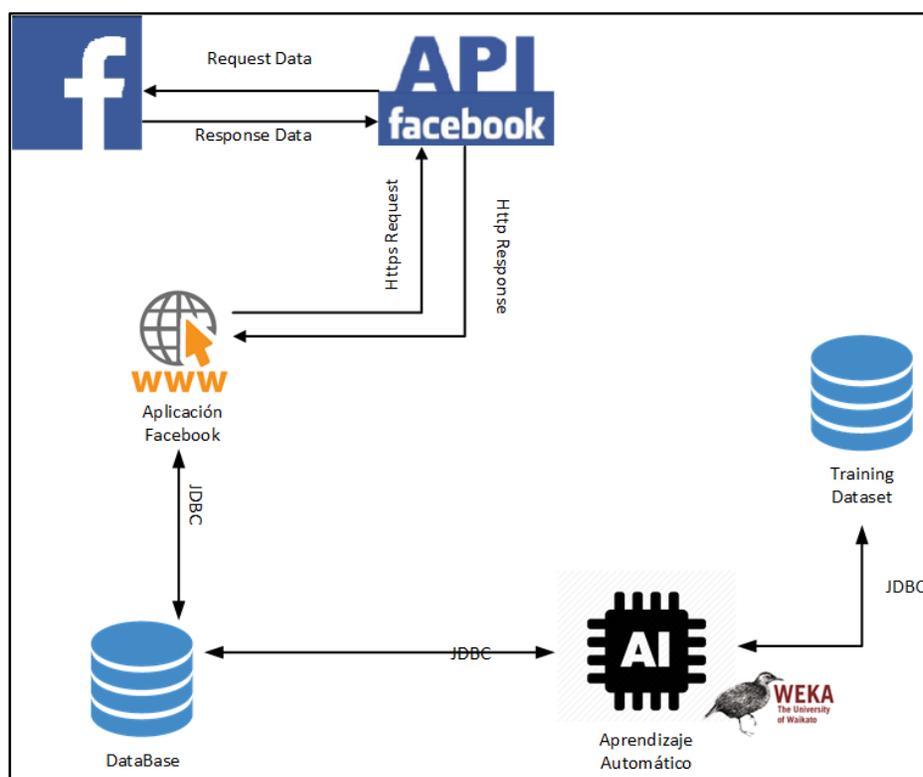


Figura 1 Arquitectura del Sistema

Se puede observar el esquema de clasificación (ver figura 2), en el cuál se va implementar el aprendizaje automático para la construcción de modelos, estos modelos serán entrenados según los datos de la cuenta del usuario Facebook, tomando sus comentarios y likes(me gusta) hasta la fecha, generando un sistema de clasificación según las cinco grandes personalidades, para la validación del modelo se va a someter a ejemplos de prueba lo cuales van a ser objeto de testeo para determinar el porcentaje al que pertenece según la clasificación del sistema.

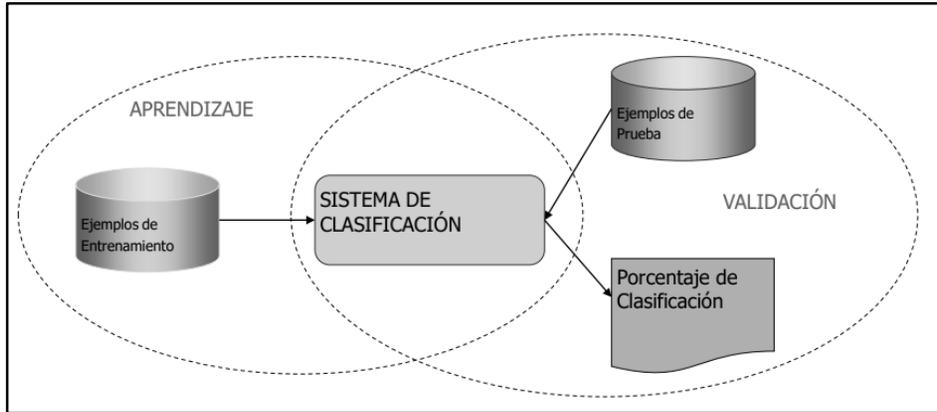


Figura 2 Esquema de Clasificación

Fuente: (García S., Luengo J., Herrera, F., 2015)

CAPÍTULO II

MARCO TEÓRICO Y TRABAJOS RELACIONADOS

2.1. Redes Sociales (Facebook)

Según (Castells M., 2013) menciona que en la última década del siglo XXI se ha presenciado una revolución tecnológica, morfológica y organización de la comunicación socializada, de tal forma que esta transformación ha logrado cambiar el modo en que la sociedad interactúa, a su vez se la define como el paso de la comunicación de las masas a la auto-comunicación de masas. Lo que realmente la auto-comunicación en masas se ve relacionada fuertemente con el internet y redes móviles. Además, es el resultado de una aplicación informática desplegada en plataformas capaces de almacenar y compartir información.

El análisis de Orihuela de 2008 menciona lo siguiente:

“Las redes sociales en línea, como LinkedIn, Facebook o Tuenti, son servicios basados en la web que permiten a sus usuarios relacionarse, compartir información, coordinar acciones y en general, mantenerse en contacto. Estas aplicaciones son la nueva forma en la que se representa nuestra red social, pero también son el modo en el que se construye nuestra identidad on-line y el cauce mediante el que se agrega y comparte nuestra actividad en la red”.

De este las redes sociales con el paso de los años desde su desarrollo han evolucionado, lo que se conocía entes en estado físico paso a estar en medios virtuales en línea. Cada persona que posee una red social toma el nombre de usuario, el cual se manifiesta como quiere que lo conozcan de forma virtual y todo se verá reflejado en el historial que toma el nombre de perfil de usuario. Es así que, el mismo es el encargado de actualizar su estado, compartir información en texto y en multimedia (imagen, video).

La red social de Facebook desde su origen tiene el propósito establecer un vínculo

virtual como se tuvo en el mundo físico en los cuales estaban compañeros de escuela, colegio, trabajo, grupos especializados entre otros. Lo que, en fin, Facebook permite encontrar viejas relaciones y por supuesto crear nuevas.

Esta misma red social de Facebook permite realizar numerosos análisis en cuanto al comportamiento de cada usuario activo en esta plataforma virtual respectivamente a la información que quiere mostrar. Este comportamiento del usuario en específico es prestado atención por el departamento de recursos humanos para el proceso de reclutamiento de nuevo personal, ya que los mismos identifican que el perfil de Facebook es una segunda vista de cómo la persona se desenvuelve en su entorno.

Para (Mababu Richard, 2016) la tendencia de analizar al usuario de Facebook para su reclutamiento lo ha denominado Reclutamiento 3.0, el cual consiste en identificar, captar y reclutar a candidatos que a las empresas les interesa para formar parte de estas. Se ha identificado esta necesidad debido a la responsabilidad que toma la correcta elección de una vacante en una empresa, siendo un éxito o fracaso. Por tanto, la aplicación de las TIC en la fase de reclutamiento y selección de personal está en sobre manera brindando su apoyo a esta necesidad.

Por lo consiguiente de acuerdo con Anand (2008), “el reclutamiento a través de redes sociales se posiciona como un proceso de reclutamiento eficaz, fiable y una manera de gestionar eficazmente el talento y de ser competitivo”.

Ecuador, acercamiento en el uso redes sociales

Según (Instituto Nacional de Estadística y Censos., 2017), la población actual es de 16´798 175 millones de habitantes en el territorio ecuatoriano. En lo que procede es una clasificación por preferencias de plataformas utilizadas por usuarios con acceso a internet en el Ecuador.

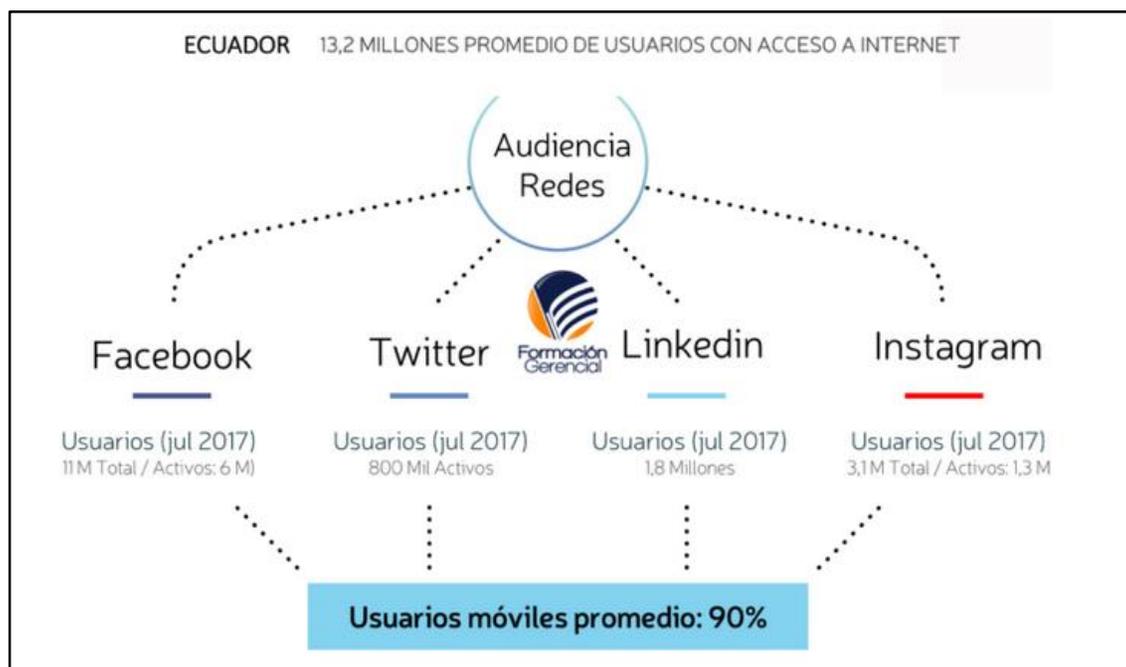


Figura 3 Usuarios de Internet y Redes Sociales en Ecuador 2017

Fuente: (Del Alcazar Juan, 2017)

Como se observa en el gráfico (ver figura 3), en el Ecuador hay alrededor 13.2 millones de usuarios que disponen de acceso al internet entre escritorio y dispositivos móviles que representa el 90%, lo cual se divide en: la plataforma de Facebook que tiene 11 millones de usuarios que poseen una cuenta distinguiéndose que 6 millones de usuarios son activos mensualmente, seguido de Twitter, LinkedIn e Instagram, datos correspondientes hasta Julio 2017. Además, (Del Alcazar Juan, 2017) menciona que las principales puertas de ingreso a la navegación es Google y Youtube.

Enero 2017		Posición Alexa Ecuador			
Posición Ecuador	Página	Enero 2015	Enero 2016	Enero 2017	Variación 2016 - 2017
1	Youtube.com	2	1	2	-50%
2	Facebook.com	1	2	9	-78%
3	Slideshare.net	31	33	22	50%
4	Instagram.com	25	27	24	13%
5	Ask.com	18	15	26	-42%
6	Scribd.com	80	81	49	65%
7	Twitter.com	15	17	59	-71%
8	Linkedin.com	58	48	101	-52%
9	Tumblr.com	115	68	131	-48%
10	Pinterest.com	86	56	252	-78%
Plataformas Mensajería Instantánea					
1	Whatsapp		45	102	-56%
2	Messenger		83	358	-77%

Figura 4 Ranking Comparativo 2015 - 2017 Redes Sociales Web Ecuador
Fuente: (Alexa Top Sites Service, 2017)

Por tanto, en la Figura 4 muestra según (Alexa Top Sites Service, 2017) la colocación de frecuencia de uso de las redes sociales en Ecuador. Tomando en cuenta que Facebook se posiciona en primer lugar dado que, YouTube es una plataforma que se caracteriza únicamente como sitio web dedicado a compartir videos sin contar con las características que posee Facebook.

Lo cual justifica la razón suficiente por la cual, para el presente trabajo se eligió la red social de Facebook como objetivo de análisis y extracción de datos. Dado que, entre más usuarios activos de Facebook, es posible obtener resultado con un alto porcentaje de precisión.

2.2. Cinco grupos de la Personalidad

Según (Sibel Adah, 2012) la personalidad de las personas se las puede dividir en cinco grandes ejes, las cuales son: la apertura, la conciencia, la extraversión, la

amabilidad y el neuroticismo. Estos 5 grandes grupos de personalidades expuestas han sido aceptadas por los profesionales del campo de estudio de la psicología, convirtiéndose en el modelo definitivo de la personalidad. Cabe señalar que la dependencia de este modelo se basa en un enfoque léxico a la medición de la personalidad. Los cinco rasgos de la personalidad se caracterizan por lo siguiente:

- **Apertura a la experiencia:** curioso, inteligente, imaginativo. Los de altos puntaje tienden a ser artísticos y sofisticados en el gusto y apreciar diversas opiniones, ideas y experiencias.
- **Responsabilidad:** responsable, organizado, perseverante. Los individuos conscientes son extremadamente confiables y tienden a ser grandes realizadores, trabajadores duros y planificadores.
- **Extraversión:** extrovertidos, amistosos, asertivos. Amigables y enérgicos, los extrovertidos se inspiran en situaciones sociales.
- **Cordialidad, Amabilidad:** cooperativo, servicial, educativo. Las personas que obtienen un alto grado de acogimiento son los que mantienen la paz y son generalmente optimistas y confían en los demás.
- **Neuroticismo:** ansioso, inseguro, sensible. Los neuróticos son malhumorados, tensos, y fácilmente sujetos a experimentar emociones negativas.

Según (Neal Andrew, 2011) las relaciones entre los rasgos de personalidad de los cinco grandes representan una propensión a comportarse de una manera identificable, y se expresan en respuestas a señales en el entorno de trabajo:

2.2.1. Apertura a la experiencia

La Persona Abierta a la experiencia es propensa a ser imaginativo, de mente abierta y curioso, este rasgo se expresa cuando hay una oportunidad para que un individuo sea creativo y contribuya a una cultura de innovación o aprenda nuevas formas de hacer las cosas y acepte las ideas de los demás. Este argumento debe influir en cómo responden los empleados en condiciones de incertidumbre, porque tales condiciones se caracterizan por cambios en sistemas, procesos o estructuras. Este rasgo debe predecir la adaptabilidad y la proactividad porque estas formas de comportamiento son

importantes cuando hay incertidumbre en los insumos, procesos o productos. La respuesta de un individuo a las expectativas con respecto a estas formas de comportamiento debe variar de acuerdo con su nivel de apertura. Por ejemplo, la inclinación a ser de mente abierta puede predisponer a los empleados a adaptarse a los cambios en sus roles de trabajo, mientras que la creatividad y la curiosidad intelectual pueden predisponer a los empleados a iniciar un cambio. Este tipo de comportamiento influye positivamente en las evaluaciones de los supervisores en la medida que el individuo cumpla con sus funciones.

Apertura a la experiencia está relacionado con la adaptabilidad, según (Thoresen, 2004) realizaron estudios y encontraron que esta personalidad tuvo un mayor número de ventas en tareas no rutinarias, pero no para los empleados que llevaban a cabo tareas rutinarias de ventas. Además, tiene una gran adaptabilidad a las demandas cambiantes de tareas y que su entrenamiento lo puede realizar con éxito.

2.2.2. Cordialidad, Amabilidad

La amabilidad es una afición a ser cooperativo, cortés y tolerante, este rasgo probablemente se exprese cuando un individuo trabaja en un equipo y tiene que depender de otros, o tiene la oportunidad de llevar a cabo actividades de ciudadanía que beneficien a la organización en su conjunto. Es decir, la amabilidad debe influir en la forma en que las personas interpretan y responden a las acciones de los demás. Las personas con altos niveles de amabilidad deben responder a estas señales cooperando con sus compañeros de trabajo, ayudando a otros miembros de la organización y adaptándose a los cambios en el contexto social.

En términos de la dirección del desempeño del rol del trabajo, los argumentos anteriores sugieren que la amabilidad tiene comportamientos que contribuyen a la efectividad del equipo o de la organización. Con respecto a la forma de desempeño del rol de trabajo, existen motivos para creer que la amabilidad puede impedir la proactividad. La amabilidad tiende a la competencia del equipo, la adaptabilidad del equipo, la competencia organizacional y la adaptabilidad organizativa; pero no a la proactividad individual.

2.2.3. Extraversión

La extraversión es un apego a ser sociable, gregario y asertivo, es probable que este rasgo aparezca cuando se requiere interactuar con otras personas y trabajar en equipo. Las personas con altos niveles de extraversión crean relaciones interpersonales efectivas con las personas en el trabajo y generan energía y unión. Estos argumentos sugieren que la extraversión contribuye a la eficacia. La extraversión predice positivamente la competencia, adaptabilidad y proactividad del equipo.

2.2.4. Responsabilidad

La responsabilidad refleja una inclinación a ser confiable y luchar por el logro, pero que puede haber variabilidad en la forma en que este rasgo predice el rendimiento. Los empleados que son responsables y orientados a los logros deben ser más propensos a gastar el esfuerzo requerido para cumplir y superar las expectativas que sus contrapartes. Estos esfuerzos deberían influir positivamente en las evaluaciones de los supervisores sobre la medida en que el individuo cumple con los requisitos de su función.

2.2.5. Neuroticismo

El neuroticismo se caracteriza por una predisposición a cogniciones negativas, pensamientos intrusivos y reactividad emocional según (Luke D Smillie, 2006). Se cree que este rasgo refleja una forma generalizada de responder en todas las situaciones porque es una predisposición para interpretar los estímulos personalmente relevantes como una amenaza potencial y a reaccionar emocionalmente. Si un individuo tiene un alto grado de neuroticismo, los procesos de tratar de cumplir con los requisitos del propio rol, adaptarse al cambio o iniciar el cambio, cada uno de ellos podría indicar cogniciones y emociones negativas, lo que debería ser perjudicial para el rendimiento. Del mismo modo, trabajar en las propias tareas o llevar a cabo el papel de uno como miembro de un equipo u organización podría indicar cogniciones y emociones negativas. De acuerdo con estas afirmaciones, (Robert P. Tett, 2003) identificaron una gama de factores que pueden indicar la expresión del neuroticismo, incluidos altos niveles de responsabilidad, falta de control, altos niveles de riesgo, conflictos interpersonales, reestructuración

organizacional, incertidumbre y amenazas externas.

2.3. Palabras más usadas según la personalidad

Las personas difieren al momento de expresar sus pensamientos, sentimientos y acciones, según estudios de (Jacob B. Hirsh, 2009) han identificado asociaciones sistemáticas entre la personalidad y el uso del lenguaje en una variedad de contextos diferentes, los resultados de tales estudios han confirmado el trabajo previo sobre la personalidad.

El estudio de (Yarkoni, 2010) analizó la relación entre la personalidad y el lenguaje utilizado en participantes cuya muestra de escritura es grande, de tópicos diversos y de fácil acceso, es decir bloggers. Los blogueros son libres de escribir sobre cualquier tema y al momento de escribir ellos no sabían que su escritura se analizaría con relación a la personalidad, estos datos proporcionan un enfoque natural donde no son influenciados por entes externos. (Yarkoni, 2010) en su estudio lo realizó con 700 blogs que contenían una media de 115,423 palabras cada uno, esto le permitió modelar confiablemente entre la personalidad y el uso de palabras no solo a nivel de categorías semánticas sino también a nivel de palabras individuales. Además, los participantes cumplían con todos los rasgos de los cinco grandes de la personalidad. El método fue mediante búsquedas aleatorias en el motor de blogs de Google y siguiendo los comentarios del autor del blog que quedaron en otros blogs. La comunicación con los participantes fue mediante correo, en la cual se invitaron a más de 5000 bloggers. Los que aceptaron la participación fueron dirigidos al sitio web del experimento, donde proporcionaron información demográfica básica y completaron un cuestionario de personalidad, los datos obtenidos de este estudio dieron como resultado un top de palabras según cada personalidad como se observa en la Tabla 1.

Tabla 1***Palabras más usadas según la personalidad***

Cinco grandes de la personalidad	Top palabras
Neuroticismo	terrible, aunque, perezoso, peor, deprimente, ironía, camino, terrible, sur, estresante, horrible, ordenar, visitó, molesto, avergonzado, suelo, prohibición, antiguo, invitado, completado.
Extraversión	Bar, otro, bebidas, restaurante, baile, restaurantes, gatos, abuelo, Miami, innumerable, bebida, disparo, computadora, niñas, gloriosa, menor, piscina, multitud, cantó, asado.
Apertura a la experiencia	Folk, humanos, de, poeta, arte, por, universo, poesía, narrativa, cultura, sorteo, siglo, sexual, películas, novela. Décadas, tinta, pasaje, literatura, blues.
Amabilidad	Maravilloso, juntos, visitar, mañana, primavera, porno, caminado, hermoso, quedar, sintió, costo, compartir, gris, alegría, tarde, día, momentos, abrazo, contento, joder.
Responsable	Completado, aventura, estúpido, aburrido, aventuras, desesperado, disfrutar, decir, Hawai, pronunciar, esos, extremo, baraja

Fuente: (Yarkoni, 2010)

2.4. Personalidad y Facebook

2.4.1. Extraversión y Facebook

Según (Wehrli, 2008) la extraversión se refiere a la medida en que las personas son sociables, alegres, optimistas, activas y comunicativas. Se espera que los individuos con alta extroversión participen en grandes cantidades de interacción social y se acerquen a otros más fácilmente. Los introvertidos ganan más al usar las redes sociales porque esta comunicación indirecta les permite compensar su falta de habilidades interpersonales, además les proporciona otra plataforma para poder comunicarse con sus amigos y contactos.

La extraversión y su uso en el Facebook según (Kelly Moore, 2012) están relacionados de modo que los usuarios más extrovertidos pasan más tiempo en Facebook, lo usan con mayor frecuencia, tienen más amigos de Facebook, publican más información en sus muros, publican más fotos y participan más en sí mismos

2.4.2. Amabilidad y Facebook

Los amables, según (Wehrli, 2008) son personas agradables con tendencias a ser comprensivas, corteses, flexibles, confiadas y clemente. Son personas con buena disposición para evitar el conflicto y que no rechazan una oferta de amistad, en términos generales la amabilidad influye en las interacciones sociales y su calidad percibida de manera favorable. (Kelly Moore, 2012) habla sobre la amabilidad y su uso en el Facebook de modo que los usuarios más amables usan el Facebook con menos frecuencia, hacen menos publicaciones en su muro sobre otros y expresan más angustia por su actividad en Facebook.

2.4.3. Responsabilidad y Facebook

La responsabilidad tiende a ser confiable, responsable, organizado auto disciplinado. Las personas responsables tienen motivación para lograr y mejorar su nivel de rendimiento laboral según (Wehrli, 2008). Si una persona es responsable no invertirá una gran cantidad de tiempo y recursos en Facebook porque prefieren mantenerse fieles a sus objetivos principales al evitar la distracción.

Según (Kelly Moore, 2012) las personas responsables están negativamente relacionadas con el uso del Facebook, de modo que las personas con mayor responsabilidad pasarán menos tiempo en Facebook, lo usarán con menos frecuencia, tendrán menos amigos, publicarán menos fotos, publicarán menos anuncios en sus muros y expresarán más arrepentimiento en sus publicaciones.

2.4.4. Neuroticismo y Facebook

Los de neuroticismo son individuos que muestran atributos negativos como desconfianza, tristeza, ansiedad, vergüenza y dificultad para manejar el estrés según (Wehrli, 2008). Los estudios de (Correa, 2010) encontraron que las personas con alto

neuroticismo eran usuarios de Internet más pesados que los extrovertidos, esto es porque los individuos con baja estabilidad emocional tienden a pasar más tiempo en las redes sociales donde intentan verse lo más atractivos posible.

Los de neuroticismo le gusta las salas de chat y los mensajes instantáneos, publican información privada pero menos fotos en su perfil de Facebook, esto tiene una explicación relacionada donde las personas neuróticas tienen más tiempo para contemplar lo que van a decir en lugar de comunicarse cara a cara, dado que las personas neuróticas son ansiosas y nerviosas por naturaleza, le es más probable que se enojen y lamenten por publicar algo de naturaleza cuestionable.

2.4.5. Apertura a la experiencia y Facebook

La apertura a la experiencia representa la curiosidad, la apertura mental y la disposición de los individuos a explorar nuevas ideas. Las personas con apertura a la experiencia y el uso del Facebook según (Kelly Moore, 2012) se involucran a una mayor sociabilidad en línea a través del Facebook. (Correa, 2010) encontró que las personas con altos puntajes en apertura son más abiertas a revelar información personal sobre ellos mismos en su perfil de Facebook, son propensos a usar y mantenerse al día en Facebook. Por lo tanto, las personas con mayor apertura a la experiencia están relacionadas positivamente con el uso de Facebook de modo que pasan más tiempo, lo usan con mayor frecuencia e inician publicaciones en su muro.

2.5. Inteligencia artificial

La inteligencia artificial es definida según la Encyclopedia of Artificial Intelligence de 2008:

La Inteligencia Artificial (IA) es un campo de la ciencia y la ingeniería que se ocupa de la comprensión, desde el punto de vista informático, de lo que se denomina comúnmente comportamiento inteligente. También se ocupa de la creación de artefactos que exhiben este comportamiento.

La IA se asocia a máquina con capacidades inteligentes, que desarrolla las

habilidades de aprender de una base de conocimiento, con el fin de solucionar problemas, además se menciona que simulan capacidades cognitivas humanas, acotando que no serán capaces de sustituir al ente humano.

2.5.1. Aprendizaje Automático

El aprendizaje automático o también conocido por su traducción en inglés como machine learning, es el proceso por el cual una computadora como primera instancia aprende a realizar una tarea al estudiar un conjunto de ejemplos de entrenamiento, lo cual significa crear su base de conocimiento. Para luego realizar la misma tarea con datos que no ha encontrado antes (Panos Louridas, Christof Ebert, 2016, pág. 110).

Consecuentemente de acuerdo con (Rouse Margaret, 2017) menciona en comparación entre el aprendizaje automático en contraste con la minería de datos se encuentran semejantes, debido a que ambos tienen el propósito de encontrar patrones en base a datos. Sin embargo, en lugar de extraer los datos para la comprensión humana –como es el caso de las aplicaciones de minería de datos– el aprendizaje automático utiliza esos datos para detectar patrones en los datos y ajustar las acciones del programa en consecuencia. La clasificación de los algoritmos de aprendizaje automático es supervisado o no supervisado. Los algoritmos supervisados pueden aplicar lo que se ha aprendido en el pasado a nuevos datos.

2.6. Api Graph

La API Graph es una herramienta de Facebook con el fin de introducir o extraer información en la plataforma de Facebook. Es una API de bajo nivel basada en HTTP que permite consultar datos, publicar historias nuevas, administrar anuncios, subir fotos y realizar varias tareas según la programación (Facebook, 2018).

2.6.1. Aspectos básicos

La API recibe su nombre a la idea de “grafica social”, que es una representación de la información disponible en Facebook, que comprende los siguientes elementos:

- Nodos: son básicamente objetos, estos pueden ser un usuario, una foto, una

página o un comentario.

- Perímetros: son las conexiones entre los objetos, como las fotos de una página o los comentarios de una foto.
- Campos: es la información sobre los objetos, como el cumpleaños de una persona o el nombre de una página.

La API se basa en HTTP, por lo que es compatible con cualquier lenguaje que cuente con una biblioteca de HTTP, además para el uso de la API es necesario el uso de identificadores de acceso como puede ser el inicio de sesión con Facebook.

2.6.2. Identificadores de acceso

Cuando alguien se conecta con una aplicación mediante el inicio de sesión con Facebook, la aplicación obtiene un identificador de acceso que proporciona acceso temporal y seguro a las API de Facebook. El identificador de acceso es una cadena opaca que identifica a un usuario, una aplicación o una página y que la aplicación puede utilizar para realizar llamadas a la API Graph, el identificador incluye información acerca de la caducidad y la aplicación que lo ha generado.

2.6.3. Permisos con el inicio de sesión con Facebook

Cuando una persona inicia sesión mediante el botón de sesión con Facebook, puedes acceder a un subconjunto de datos que tiene almacenados en su Facebook, estos permisos son la petición que realizan los usuarios para que permitan acceder a dichos datos. Se podrá acceder a una información u otra dependiendo de la configuración de la privacidad de la persona y de las peticiones que se realicen.

Como ejemplos a una solicitud de estos permisos puede ser el email que es la dirección de correo electrónico de una persona, `user_likes` es el acceso a la lista de intereses de una persona que ha indicado que le gustan. En este trabajo es necesario el permiso de `user_post`, este permiso proporciona acceso a las publicaciones de la bibliografía de una persona, incluida las suyas y las de otras personas.

2.7. El inventario de los cinco grandes

Para medir la personalidad de los cinco grandes existen diferentes test según (Elvira Salgado, 2016) son: Trait descriptive Adjectives, Neo Personality Inventor y el Big Five Inventory (BFI) de (Oliver P. John, 2008) , todos los test son de alta confiabilidad y el uso del uno con el otro dependen de los objetivos buscados para sus aplicaciones o el tiempo disponible de las personas para responder las preguntas. El BFI está compuesto por 41 preguntas de frases cortas y fáciles de entender, esto le permite a la persona responder con brevedad y simplicidad, evitando la ambigüedad o múltiples interpretaciones de las preguntas.

Las preguntas van anteceditas de la afirmación “Me veo a mí mismo como alguien que...”, los encuestados deben responder con base a la escala Likert de 5 puntos que va desde “No estoy de acuerdo” hasta “Totalmente de acuerdo”, estas escalas han mostrado una confiabilidad en el testeo, así como una convergencia fuerte con versiones más extensas para medir los cinco grandes de la personalidad.

2.8. Trabajos Relacionados

2.8.1. Apply Magic Sauce

Las aplicaciones que se dedican a la predicción de la personalidad con los perfiles de Facebook. Actualmente existe una aplicación nombrada Apply Magic Sauce, desarrollada en la Universidad de Cambridge, actualmente no existen otras aplicaciones similares a esta. La herramienta es desarrollada por (Cambridge, 2017) y es capaz de hacer un informe psicológico en base a los likes que realiza el usuario en Facebook, pero no en sus comentarios. Este desarrollo se basa en más de 6 millones de perfiles y los resultados coincidentes en las pruebas psicométricas, lo que hace la aplicación es dar acceso al perfil del usuario de Facebook bajo su permiso para que analice la actividad de los likes y lo compare con la información de más de seis millones de personas con los que se ha tenido un aprendizaje previo, de esta manera la aplicación predice la edad, el género, el estado civil, la religión, la afinidad política, la inteligencia, la amabilidad, la

responsabilidad, la timidez o extroversión (BIG 5) y lo feliz que es una persona según los datos extraídos de Facebook.

Lo que se pretende es tomar como base esta aplicación en lo visual, en la forma de mostrar los datos de manera intuitiva y rápida con unos cuantos clics, ya se puede predecir la personalidad, pero la diferencia es que Apply Magic tiene una base de datos de 6 millones de perfiles, la investigación propuesta se basa en el análisis de personalidad a partir de los likes, post y/o comentarios de los usuarios y un aprendizaje de texto que describe a cada uno de los grupos de personalidad a los que corresponde un usuario, se adiciona también cálculos matemáticos y algoritmos de aprendizaje automático que se utilizan para predecir la personalidad sin la necesidad de comparar entre millones de perfiles.

CAPÍTULO III

DESARROLLO

3.1. Introducción

En el presente capítulo se describe el proceso que se llevó a cabo para el desarrollo del aplicativo con las siguientes actividades que son la implementación de una aplicación Facebook que permita el acceso y recolección de datos usando la API Graph, luego de obtener los datos el usuario debe realizar un test de personalidad BFI, al finalizar el test el usuario puede visualizar sus resultados para su interpretación, al tener una gran cantidad de personas que han realizado el test se realiza el entrenamiento y clasificación de los datos obtenidos para finalmente realizar la predicción según los cinco grandes personalidades.

3.2. Metodología XP

Para el desarrollo del proyecto se utilizará las metodologías ágiles, según (Letelier, 2006) las metodologías ágiles permiten tener un control en el desarrollo del proyecto, tanto sobre los recursos humanos, como las herramientas a usar. Las metodologías ágiles buscan desarrollar el software más rápido sin tanta documentación, esto se debe a que el cliente o el experto deben estar en constante participación en el proyecto, por lo que el mismo puede sufrir varios cambios según los requerimientos del cliente.

El desarrollo de proyectos con metodologías ágiles permite una entrega frecuente de software al cliente, realizado por módulos, en vez de hacer todo el software este se lo divide en módulos o entregables, para ser revisados por el cliente y si existen cambios realizarlos inmediatamente, esto proporciona una retroalimentación del software a desarrollar y hacer los cambios al comienzo nos ahorra dinero y tiempo que realizarlos al final del desarrollo del software.

La metodología ágil para el presente proyecto será la de programación Extrema (XP), ya que permite avanzar el proyecto de manera rápida sin tanta documentación y con un enfoque más al desarrollo y al manejo de las herramientas según (Letelier, 2006). De esta manera permite ahorrar tiempo, haciendo pruebas a lo que se avanza el proyecto y cambios que se presenten, el cliente tiene un papel importante ya que es el que aprueba los cambios y verifica los módulos o entregables si están correctos si no es el caso se realiza los cambios.

3.3. Desarrollo aplicación Facebook

La aplicación Facebook permite la obtención de los datos del usuario de Facebook, en específico se necesita las publicaciones realizadas en el muro de cada usuario para su posterior análisis. Lo primero es crear una aplicación Facebook, para ello es necesario tener una cuenta Facebook e ingresar a la página de desarrolladores de Facebook, en la parte de aplicaciones se crea una nueva aplicación, lo primero que pide es ingresar el nombre de la aplicación y el correo de contacto del administrador del aplicativo.

Ya una vez creada la aplicación se debe tomar en cuenta ciertos requisitos por parte de Facebook que son:

- Tener un dominio.
- La página web debe tener un certificado.
- En el directorio de la página web debe tener las políticas de privacidad.

Estos requisitos son fundamentales para que Facebook permita el acceso a los datos del usuario y que el usuario a su vez se sienta en confianza en dar acceso a sus datos.

Al cumplir con los requisitos mencionados y tras obtener la aprobación de Facebook, es posible tener acceso a los siguientes datos del usuario como son: email, la lista de amigos, nombre, apellidos, foto de perfil, género y edad.

Pero si se necesita más información como las indicaciones de me gusta o las publicaciones en el muro del usuario es necesario realizar otra solicitud, en el caso de este proyecto es necesario el acceso a la información de las publicaciones en el muro del

usuario, para tener este permiso se necesita ir a la pestaña de revisión de la aplicación y gestionar un nuevo elemento para aprobación, donde Facebook revisa la propuesta que se envía, esta propuesta puede ser un video donde se explica dónde se va a utilizar los datos y en que parte el usuario se beneficia del uso de estos datos, si el usuario no puede visualizar el uso de sus datos, Facebook no dará permiso a la solicitud.

3.4. Desarrollo Página Web

Para el desarrollo de la Página web se consideró varias plataformas teniendo en consideración los requisitos de Facebook para el acceso a los datos del usuario, al final se optó por Heroku que es una plataforma como servicio según (Heroku, 2018), soporta varios lenguajes como son Java, Node.js, Scala, Clojure, Ruby, Python, PHP y Go. La base de su sistema operativo es Debian y está basado en un sistema de contenedor administrado, con servicios integrados. Los desarrolladores de Heroku se enfocan en la entrega de aplicaciones con la ayuda de las herramientas y flujos de trabajo que ofrece Heroku.

3.4.1. Plataforma Heroku

Heroku se basa en Dynos que son las piezas fundamentales del modelo de su arquitectura, que son las unidades que proveen la capacidad de cómputo dentro de la plataforma, estos contenedores están basados en Linux, cada uno de estos están aislados del resto por lo tanto los comandos y los archivos que se almacenan en ellos no afectan a los demás, además cada dyno tiene su propia arquitectura requerida según el desarrollador por la elección de su lenguaje de programación.

Heroku también facilita de un dominio o subdominio en el caso de este proyecto es <https://personalityprediction2018.herokuapp.com/>, también provee de un certificado para que todas las conexiones sean de forma segura y cuenta con bases de datos que son de ayuda para este proyecto como son Postgres con un límite de 20 conexiones simultaneas, esta base de datos contendrá la información extraída del usuario de Facebook.

3.4.2. Configuración de la plataforma

Para crear la aplicación en Heroku primero se debe crear una cuenta y posteriormente elegir en que lenguaje de programación se quiere crear el aplicativo, para este proyecto se eligió Node.js que es un lenguaje de programación basado en JavaScript y como la API Graph de Facebook funciona para este lenguaje le hace compatible para la realización del presente proyecto.

3.4.3. Configuración del Botón de Inicio de Facebook

Ya una vez creada la aplicación y que este en línea se procede a crear el botón de inicio de sesión Facebook, primero se registra la dirección web en la aplicación de Facebook, esto permite al aplicativo de Facebook corroborar con la dirección que se hacen las peticiones, si se cambia la dirección web también se debe cambiar en el aplicativo de Facebook sino el botón de inicio de sesión deja de funcionar.

Para realizar la conexión de la página web con el aplicativo de Facebook para el botón de inicio de sesión es necesario utilizar el SDK de Facebook que está escrito en JavaScript, sin la necesidad de descargar o instalar ningún archivo independiente, solo es necesario incluir el fragmento de código JavaScript en el código HTML, se cargara de forma asincrónica el SDK en la página web. Lo único que se debe cambiar para poder utilizarlo es poner el id del aplicativo de Facebook antes creado en el atributo 'appId' y que versión del API Graph se desea usar, para este proyecto se utiliza la versión 2.11.

El primero paso cuando se carga la página web es determinar si el usuario ya tiene sesión iniciada en el aplicativo, para determinar este proceso se comienza con una llamada a 'FB.getLoginStatus', esta función realiza una llamada a Facebook para obtener el estado de inicio de sesión, devolviendo un objeto response como respuesta que contiene varios campos que son:

- status: especifica el estado de inicio de sesión del usuario que usa la aplicación. El estado puede ser uno de los siguientes:
 - connected: el usuario ha iniciado sesión en Facebook y en tu aplicación.
 - not_authorized: el usuario ha iniciado sesión en Facebook, pero no el

- aplicativo.
- unknown: el usuario no ha iniciado sesión en Facebook y no se sabe si lo ha hecho en la aplicación.
- authResponse: se incluye este objeto si el estado es connected y contiene los siguientes elementos:
 - accessToken: contiene un identificador de acceso para el usuario que usa el aplicativo.
 - expiresIn: indica la hora en UNIX en que el identificador caduca.
 - signedRequest: es un parámetro firmado que contiene información sobre el usuario que usa el aplicativo.
 - userID: el identificador único del usuario que usa el aplicativo.

Ahora se incluye el botón de inicio de sesión, cuando el usuario haga clic en él, se abrirá un cuadro de diálogo que desencadena el proceso de inicio de sesión, el usuario que ya tenga iniciada sesión en el aplicativo no lo verá. En la ventana del cuadro de diálogo, el usuario tiene la posibilidad de elegir que permisos otorgar o denegar al aplicativo.

En el código de la creación del botón de Facebook se especifica que permisos son necesarios, en la sección “scope” se ingresa los permisos necesarios, para este proyecto es necesario dos permisos que son “public_profile” que devuelve los datos del usuario como son los nombres y la foto de perfil, el otro permiso es “user_posts” que permite el ingreso a los estados en el muro del usuario, estos pueden ser sus publicaciones, actualizaciones de perfil, fotos subidas a Facebook o estados compartidos.

3.4.4. Creación de la Base de datos.

El siguiente paso es extraer la información del usuario al momento de iniciar sesión en la página web y almacenar los datos en la base de datos Postgres, primero se configura la base de datos, se crea una nueva data en el panel de recursos de Heroku y se le asigna la base de datos al aplicativo creado en el caso de este proyecto se llama “personalityprediction2018”. Ya creada la base de datos en Postgres se crean las tablas con los atributos necesarios para este proyecto, para gestionar la base de datos se usó la

herramienta “pgAdmin” en la versión 4.2, para realizar la conexión primero debemos obtener las credenciales que otorga Heroku, cabe mencionar que las conexiones a la base de datos de Postgres se las realiza mediante protocolo SSL, esto le permite que los datos viajen seguros por la red de manera encriptada.

Con el pgAdmin de Postgres se crea las tablas necesarias para este proyecto, la primera tabla (ver Tabla 2) son los datos de usuario que se extraen de la Api de Facebook, a continuación, se describe la tabla con la descripción de cada una de las columnas de la tabla usuario.

Tabla 2
Tabla Usuario

Nombre	Descripción	Tipo
user_id	El identificador único que es extraído de Facebook y es la clave primaria de la tabla.	Texto
user_name	El nombre del usuario en su perfil de Facebook.	Texto
user_number_post	Es el número total de publicaciones realizadas en el muro del usuario.	Número
user_number_shares	Es el número de publicaciones que han sido compartidas por el usuario.	Número
user_number_photos	Es el número de veces que se publicó que se ha subido una foto, esto puede ser una o varias fotos por publicación.	Número



Continúa

user_number_update_perfil	Es el número de veces que el usuario actualizo su foto de perfil.	Número
user_posts	Es todo el texto que ha escrito en las publicaciones el usuario.	Texto
p_neuro	Es la sumatoria de todas las palabras encontradas en el texto de user_posts que corresponden a la personalidad neurótica.	Número
p_extro	Es la sumatoria de las palabras encontradas en el texto de user_posts que corresponden a la personalidad extrovertido.	Número
p_abierto	Es la sumatoria de las palabras encontradas en el texto de user_posts que corresponden a la personalidad abierto a experiencia.	Número
p_respo	Es la sumatoria de las palabras encontradas en el texto de user_posts que corresponden a la personalidad responsable.	Número
p_amab	Es la sumatoria de las palabras encontradas en el texto de user_posts que corresponden a la personalidad amabilidad.	Número
f1_e	Es la sumatoria de la respuestas correctas realizadas al test de personalidad que corresponden a la personalidad extrovertido.	Número
f2_n	Es la sumatoria de las respuestas correctas realizadas al test de personalidad que corresponden a la personalidad neurótico.	Número



Continúa

f3_ae	Es la sumatoria de la respuestas correctas realizadas al test de personalidad que corresponden a la personalidad abierto a la experiencia.	Número
f4_a	Es la sumatoria de la respuestas correctas realizadas al test de personalidad que corresponden a la personalidad amabilidad.	Número
f5_c	Es la sumatoria de la respuestas correctas realizadas al test de personalidad que corresponden a la personalidad responsabilidad.	Número
user_genero	Representa el género del usuario, f si es femenino o m si es masculino	Texto
user_personality	Es el resultado del test de personalidad realizado por el usuario según los factores antes descritos el de mayor sumatoria es la personalidad del usuario.	Texto

La siguiente Tabla 3, es de las palabras que definen a la personalidad, estas palabras fueron extraídas en las ya mencionadas palabras más usadas de cada personalidad según el estudio de (Yarkoni, 2010).

Tabla 3

Tabla de Palabras según su personalidad

Nombre	Descripción	Tipo
palabra	Es palabra que define a la personalidad y forma parte de la clave primaria de la tabla	Texto
personalidad	Permite identificar a qué tipo de personalidad pertenece la palabra, además forma parte de la clave primaria junto con la palabra, esto se debe a que existen palabras que pueden compartir diferentes personalidades con pesos diferentes.	Texto
peso	Es el valor de la palabra que le determina a la personalidad, si el valor es negativo significa que es al contrario de la personalidad.	Número

3.4.5. Extracción de datos del Usuario de Facebook

Ya creadas las tablas se procede a llenarlas para ello el usuario que se dirija a la dirección <https://personalityprediction2018.herokuapp.com/> y de clic en continuar tal como se muestra en la siguiente figura (ver Figura 5).

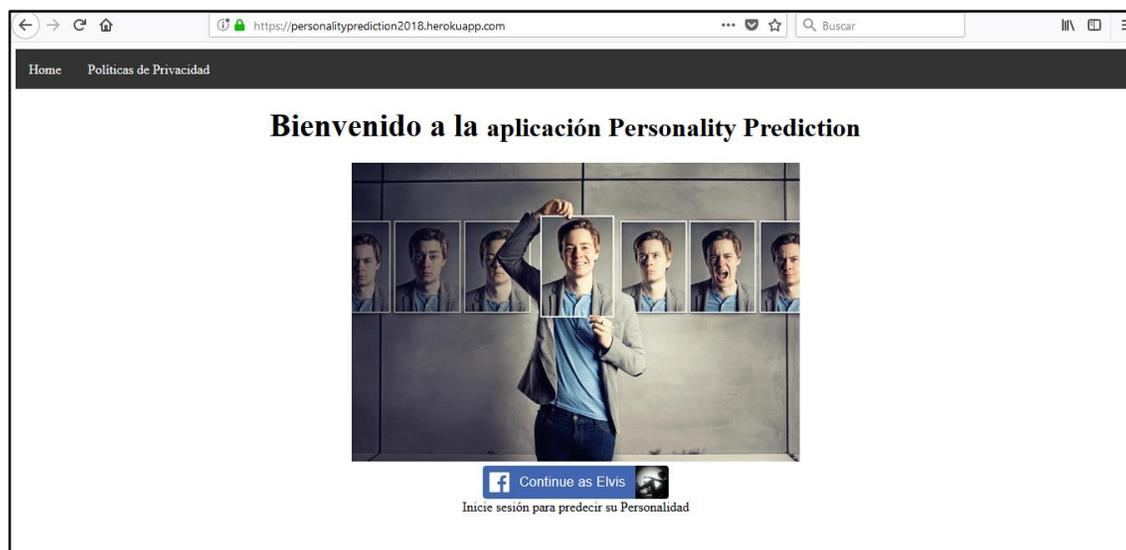


Figura 5 Inicio de sesión de la aplicación

Al dar clic se abrirá una ventana de dialogo creada por Facebook, en la cual contiene el nombre de la aplicación a la que se dará acceso y que datos se le proporcionara el acceso a la página web, es de total importancia que los dos campos estén seleccionados para la extracción de datos que son los datos del perfil público y las publicaciones de la biografía como se muestra en los siguientes gráficos (ver Figura 6 y 7).

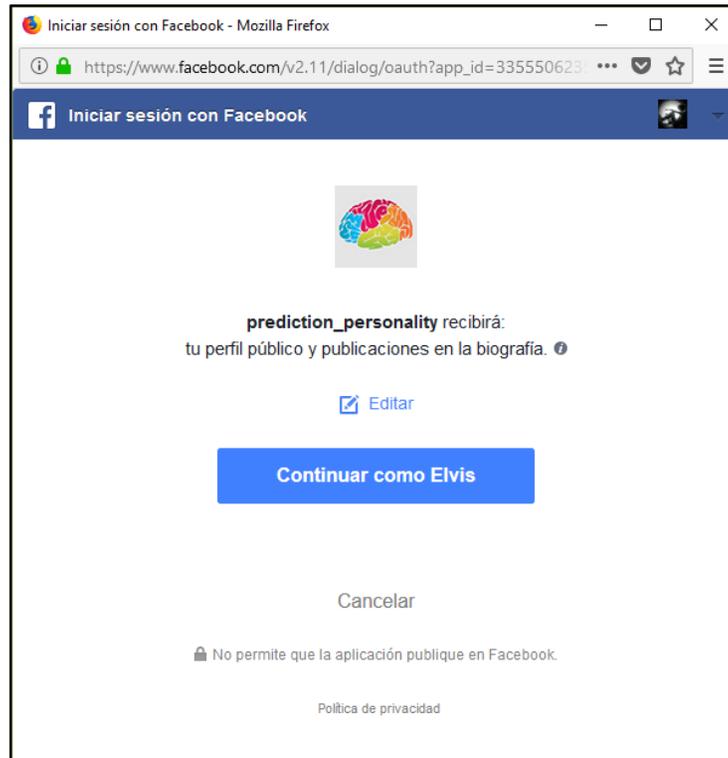


Figura 6 Ventana de diálogo de Facebook

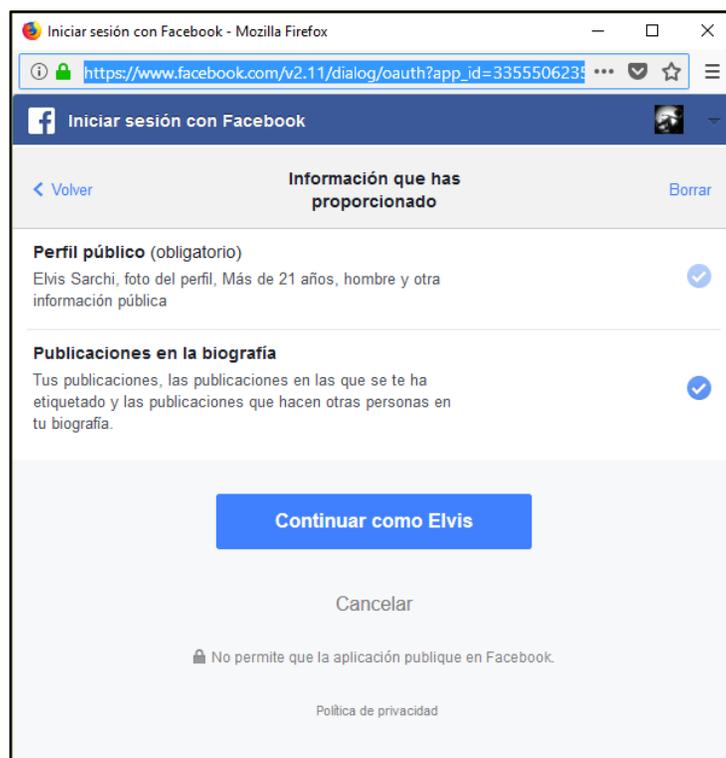


Figura 7 Ventana de información de Permisos

A continuación de dar clic en continuar la aplicación comienza con la extracción de los datos del usuario, esto puede demorar un poco dependiendo de la cantidad de publicaciones que el usuario tenga es su muro de Facebook. Se toma las siguientes consideraciones al momento de extraer las publicaciones:

- Las publicaciones son tomadas en los dos últimos años, esto permite tener datos más actualizados ya que las personas al pasar los años cambian sus ámbitos y su forma de representarse en las redes sociales publicando en su muro de diferente manera a la de años anteriores.
- Las publicaciones son entregadas por parte de Facebook en formato Json que contienen los siguientes datos:
 - id: es el identificador de la publicación.
 - message: es el texto de la publicación en el muro Facebook.
 - story: es el texto de la historia del usuario como pueden ser cuando dos amigos se hacen amigos, se sube una foto, se comparte una publicación

o se actualiza el perfil del usuario.

- created_time: hora en que se publicó inicialmente la publicación, esta será la fecha y hora del evento.
- Para definir qué publicaciones han sido compartidas se toma el texto de “story” y se busca la palabra reservada “shared”, si se encuentra esta palabra significa que la publicación ha sido compartida de otro lado.
- Para identificar el número de veces que se ha subido una o varias fotos se toma el texto de “story” y se busca la palabra reservada “new photo”, al encontrar esta cadena de texto significa que la publicación es relacionada a la publicación de fotos.
- Para definir el número de veces que el usuario actualizo su foto de perfil se toma el texto de “story” y se busca la cadena de texto “cover photo”, al encontrar esta cadena significa que la publicación se realizó por una actualización de foto de perfil.

3.4.6. Inserción de las palabras según su personalidad

Con los datos extraídos se realiza el análisis de la cadena de texto de todas las publicaciones realizadas por usuario que ahora están almacenados en la tabla usuario en la columna ‘user_posts’, primero se debe llenar la tabla de las palabras que corresponden a la personalidad con sus respectivos pesos como se muestra en la siguiente tabla, tiene un total de 288 palabras, estas palabras se obtuvo del trabajo de (Yarkoni, 2010) donde cada palabra que se obtuvo se le añadieron sus sinónimos, además cada palabra le corresponde un peso este puede ser positivo o negativo, si el peso es negativo significa que la palabra corresponde a la personalidad contraria a la que está definida, también se consideró la misma palabra con y sin tilde y en su forma singular, para usos prácticos a cada personalidad se le asigno su letra inicial como se explica a continuación:

- a: amabilidad.
- ae: abierto a la experiencia.
- e: extrovertido.

- n: neurótico.
- r: responsable.

En las siguientes tablas (ver Tablas 5, 6, 7, 8 y 9), se muestran las palabras pertenecientes a cada personalidad acompañada del peso correspondiente.

Tabla 4

Palabras de la personalidad amabilidad

Palabra	Personalidad	Peso
abrazo	a	0.22
alegría	a	0.22
alegría	a	0.22
auge	a	-0.25
bonito	a	0.23
calle	a	-0.22
caminado	a	0.23
caminar	a	0.23
ciudad	a	-0.21
compartir	a	0.23
contento	a	0.22
costo	a	-0.23
creciente	a	-0.25
día	a	0.22
disfrutar	a	0.22
divertido	a	0.22
enemigo	a	-0.24
esperar	a	0.23
estupendo	a	0.28
excelente	a	0.28
frescura	a	0.25
gay	a	-0.23
gris	a	0.22
hermoso	a	0.23
indecente	a	-0.25
insultar	a	-0.22
izquierda	a	0.23

Continúa

joder	a	-0.22
juntos	a	0.26
madrugada	a	0.26
maldito	a	-0.22
mañana	a	0.26
maravilloso	a	0.28
maricon	a	-0.23
mierda	a	-0.22
momentos	a	0.22
multitud	a	-0.21
pack	a	-0.25
pasear	a	0.23
pelucon	a	-0.2
permanecer	a	0.23
porno	a	-0.25
precio	a	-0.23
primavera	a	0.25
quedar	a	0.23
rechazado	a	-0.23
sentir	a	0.23
sexual	a	-0.24
sintio	a	0.23
sintió	a	0.23
tarde	a	0.22
tiempo	a	0.22
triste	a	0.22
valor	a	-0.23
visitar	a	0.26

Tabla 5
Palabras de la personalidad abierto a la experiencia

Palabra	Personalidad	Peso
amor	ae	-0.25
aniversario	ae	-0.25
arte	ae	0.29
belleza	ae	0.26
blues	ae	0.26
bonito	ae	0.25
cancion	ae	0.25
canción	ae	0.25
cerca	ae	0.28
cuidar	ae	-0.15
cultura	ae	0.28
decada	ae	0.27
década	ae	0.27
doctrina	ae	-0.16
emocionado	ae	-0.14
enfadado	ae	-0.13
enojado	ae	-0.13
familia	ae	-0.18
gente	ae	0.32
gracias	ae	-0.24
hermoso	ae	0.25
hogar	ae	-0.22
humanos	ae	0.31
iglesia	ae	-0.16
insultar	ae	-0.11
literatura	ae	0.27
maldecir	ae	-0.11
maldito	ae	-0.11
mierda	ae	-0.11
movimiento	ae	-0.18
musica	ae	0.26
narracion	ae	0.28
narración	ae	0.28
narrativa	ae	0.28



negacion	ae	-0.13
negación	ae	-0.13
novela	ae	0.27
optimismo	ae	-0.12
pasaje	ae	0.27
paso	ae	0.27
pelicula	ae	0.27
película	ae	0.27
personas	ae	0.31
poema	ae	0.29
poesia	ae	0.28
poesía	ae	0.28
poeta	ae	0.29
regalo	ae	-0.28
religión	ae	-0.16
sexual	ae	0.27
siglo	ae	0.28
sorteo	ae	-0.28
tiempo	ae	-0.21
tinta	ae	0.27
universo	ae	0.28

Tabla 6*Palabras de la personalidad extrovertido*

Palabra	Personalidad	Peso
abuelo	e	0.2
aceptar	e	-0.23
amable	e	-0.25
aprobar	e	-0.23
asado	e	0.18
atacames	e	0.2
autorizar	e	-0.23
bailar	e	0.2
bastante	e	-0.27
beber	e	0.19
bebidas	e	0.21


 Continúa

cantina	e	0.23
canto	e	0.18
cantó	e	0.18
casualidad	e	-0.13
causa	e	-0.13
chica	e	0.19
chupitos	e	0.19
colombia	e	0.18
computadora	e	-0.19
disparo	e	0.19
famoso	e	0.19
flojo	e	-0.26
fosh	e	0.23
gatos	e	-0.2
gentio	e	0.18
gentío	e	0.18
gloriosa	e	0.19
ilimitado	e	0.2
innumerable	e	0.2
interes	e	-0.22
interés	e	-0.22
jajaja	e	0.25
mascota	e	-0.2
menor	e	-0.19
mujer	e	0.19
multitud	e	0.18
ofrecer	e	-0.22
ordenador	e	-0.19
otro	e	-0.22
paginas	e	-0.25
páginas	e	-0.25
pájaro	e	-0.25
parrillada	e	0.18
patrón	e	-0.25
perezoso	e	-0.26
piscina	e	0.18
popular	e	0.19



Continúa

practico	e	-0.26
práctico	e	-0.26
razon	e	-0.13
razón	e	-0.13
restaurante	e	0.21
restaurantes	e	0.2
suficiente	e	-0.27
util	e	-0.26
útil	e	-0.26
vago	e	-0.26
zona	e	0.23

Tabla 7
Palabras de la personalidad neurótico

Palabra	Personalidad	Peso
abochornado	n	0.19
anticuado	n	-0.22
antiguo	n	-0.18
articulo	n	-0.16
artículo	n	-0.16
aunque	n	0.24
avergonzado	n	0.19
camino	n	-0.20
canabis	n	0.2
carretera	n	-0.20
censura	n	0.18
circunscripción	n	-0.23
completado	n	-0.18
condado	n	-0.23
cucho	n	-0.22
degradante	n	0.21
deprimente	n	0.21
distrito	n	-0.23
droga	n	0.2
especial	n	-0.15
estres	n	0.19

Continúa

estrés	n	0.19
estresante	n	0.19
estupefaciente	n	0.2
fastidioso	n	0.19
fatal	n	0.26
flojo	n	0.24
horrible	n	0.19
invitado	n	-0.18
invito	n	-0.18
invitó	n	-0.18
ironía	n	0.21
ironía	n	0.21
jurisdicción	n	-0.23
loco	n	-0.21
molesto	n	0.19
municipio	n	-0.23
numero	n	-0.09
número	n	-0.09
orden	n	0.19
ordenar	n	0.19
peor	n	0.21
perezoso	n	0.24
prohibicion	n	0.18
prohibición	n	0.18
salvaje	n	-0.21
sarcasmo	n	0.21
sureño	n	-0.2
suelo	n	-0.19
terminado	n	-0.18
terrible	n	0.2
territorio	n	-0.19
tremendo	n	0.26
vago	n	0.24
veterano	n	-0.22
viejo	n	-0.18
visitado	n	-0.19
visito	n	-0.19
visitó	n	-0.19
weed	n	0.2

Tabla 8
Palabras de la personalidad responsable

Palabra	Personalidad	Peso
aburrido	r	-0.22
ajuste	r	0.22
amigo	r	-0.15
angustiado	r	-0.2
artículo	r	-0.1
artículo	r	-0.1
aventura	r	0.22
aventuras	r	0.20
baraja	r	0.18
bobo	r	0.24
calle	r	-0.24
cansado	r	-0.22
ciudad	r	-0.25
completado	r	0.25
concluido	r	0.21
cubierta	r	0.18
decir	r	-0.2
desesperado	r	-0.2
discrepancia	r	-0.12
disfrutar	r	0.2
divergencia	r	-0.12
divertido	r	0.2
doctrina	r	-0.12
esmeraldas	r	0.19
esposa	r	-0.24
estupido	r	-0.22
expresar	r	-0.19
extremo	r	-0.19
facil	r	0.23
familia	r	-0.12
fastidioso	r	-0.22
fuego	r	-0.22
fumar	r	-0.26
futbol	r	-0.24



Continúa

fútbol	r	-0.24
gozar	r	0.2
hablar	r	-0.20
hazaña	r	0.22
hogar	r	0.12
iglesia	r	-0.12
limite	r	-0.19
límite	r	-0.19
listo	r	0.25
madre	r	0.23
pana	r	-0.15
pitada	r	-0.26
playa	r	0.19
preocupado	r	-0.2
pronunciar	r	-0.19
rara	r	-0.23
raro	r	-0.23
realista	r	0.22
religión	r	-0.12
superar	r	-0.23
superficie	r	0.18
tabaco	r	-0.26
terminado	r	0.25
tonto	r	-0.22
vencer	r	-0.23

3.4.7. Análisis del texto de las publicaciones

Ya con las palabras en la base de datos se realiza el análisis del texto de las publicaciones del usuario de Facebook, con el siguiente proceso:

- Primero se cuentan todas las palabras que hay en la cadena de texto, omitiendo los caracteres no alfabéticos e insertándolos en una lista.
- Cargar la lista de palabras que corresponden a la personalidad con sus respectivos pesos.
- Usando el método de stemming, el cual consiste en encontrar la palabra que contenga su raíz, por ejemplo, la palabra familia está contenida en familiares, con este método se compara si la palabra que pertenece a la personalidad está contenida en la lista de palabras que se determinó del texto de las publicaciones.

- Si la palabra está contenida o es igual, se verifica a qué tipo de personalidad le corresponde y se realiza la sumatoria dependiendo del peso de la palabra.
- Al terminar el proceso de la sumatoria de todas las palabras se guarda en la base de datos los pesos de cada personalidad.

3.5. Test de Personalidad

Para realizar la predicción de la personalidad del usuario primero se necesita de una base de conocimiento, con la ayuda del BFI que es un test de personalidad de (Oliver P. John, 2008) el cual consta de 41 preguntas, al realizar el usuario el test como resultado se le mostrara su personalidad y estos datos serán registrados en la base de datos en la tabla usuario, en la siguiente tabla (ver Tabla 9) se muestran las 41 preguntas del test BFI.

Tabla 9
Preguntas del Test de Personalidad

Número	Pregunta
1	Se enoja rápidamente con los demás.
2	Prefiere estar solo.
3	Está siempre en movimiento/realizando una actividad.
4	Puede convencer a otros para que hagan cosas.
5	Busca silencio.
6	Es asertivo y responsable de tareas.
7	Se retracta de expresar sus opiniones.
8	Disfruta de ser parte de un grupo.
9	Hace que las cosas avancen a su propio ritmo.
10	A menudo se siente triste, decaído emocionalmente.
11	No se molesta fácilmente.
12	Se estresa fácilmente.
13	Se siente abrumado por las emociones.
14	Está tranquilo, incluso en situaciones tensas.
15	Teme que haga lo incorrecto.
16	Se mantiene relajado.



Continúa

17	Hace cosas que más tarde lamenta.
18	No tiene una buena imaginación
19	Le encanta leer material desafiante (Leer cosas nuevas).
20	Está interesado en muchas cosas.
21	Intenta entenderme a sí mismo.
22	No está interesado en ideas abstractas.
23	Cree en la importancia del arte.
24	Prefiere quedarse con cosas que sabe, no ir más allá.
25	Tiende a votar por candidatos políticos conservadores.
26	Sospecha de motivos ocultos en otros.
27	Confía en otros.
28	Contradice a otros.
29	Valora la cooperación sobre la competencia.
30	Es fácil de satisfacer.
31	Piensa muy bien de sí mismo.
32	Está preocupado por los demás.
33	Pone a las personas bajo presión
34	Completa las tareas con éxito.
35	A menudo hace planes de último minuto.
36	Sobresale en lo que hace.
37	A menudo olvida volver a poner las cosas en su lugar.
38	Postpone las decisiones.
39	Trabaja mucho.
40	Paga sus cuentas a tiempo.
41	No ve las consecuencias de las cosas

Para la implementación del test se desarrolló la página web la cual contiene las 41 preguntas y como requisito para poder acceder al test, el usuario debe iniciar sesión en el aplicativo, una vez iniciada sesión el aplicativo le redirige a la página del test de personalidad tal como se muestra en la siguiente gráfico (ver Figura 8), se especifica la escala que va desde el uno que es no estoy de acuerdo al cinco que es totalmente de acuerdo, cada pregunta se la debe leer de tal manera que empieza con la frase “me veo como alguien que....” seguido de la pregunta, todas las preguntas deben ser contestadas para resolver el test de personalidad y mostrar el resultado al usuario.

No estoy de acuerdo	Parcialmente en desacuerdo	Ni de acuerdo ni en desacuerdo	Parcialmente de acuerdo	Totalmente de acuerdo
1	2	3	4	5

Me veo como alguien que:

Factor I

1. Se enoja rápidamente con los demás.*
 1 2 3 4 5

2. Prefiere estar solo.*
 1 2 3 4 5

3. Está siempre en movimiento/realizando una actividad.*
 1 2 3 4 5

4. Puede convencer a otros para que hagan cosas.*
 1 2 3 4 5

5. Busca silencio.*
 1 2 3 4 5

Figura 8 Test de Personalidad

3.5.1. Resultado Test de Personalidad

El resultado del test que se muestra al usuario está compuesto por los cinco factores de cada personalidad, el cual está determinado por el mayor puntaje que se obtuvo de la sumatoria de los cinco factores que le corresponde a cada personalidad.

En el siguiente gráfico (ver Figura 9) se muestra las cinco personalidades en porcentajes según las respuestas del test. Para definir el porcentaje primero se saca el máximo posible que puede obtener por ejemplo, para amabilidad, el máximo puntaje es de 40, se realiza una regla de tres y se resta 50 al valor, esto permite visualizar el resultado desde un punto medio, ahora si el resultado es 20, el porcentaje a mostrar es 0%, consecuentemente para esta personalidad el puntaje que sea menor a 20 tendrá resultados negativos y mayores a éste serán positivos, por ultimo a este valor se le multiplica por 2 para visualizar el resultado correctamente, si el puntaje es 40 el porcentaje a mostrar será del 100%. En la siguiente ilustración se observa un ejemplo de los resultados obtenidos de lo anteriormente explicado.

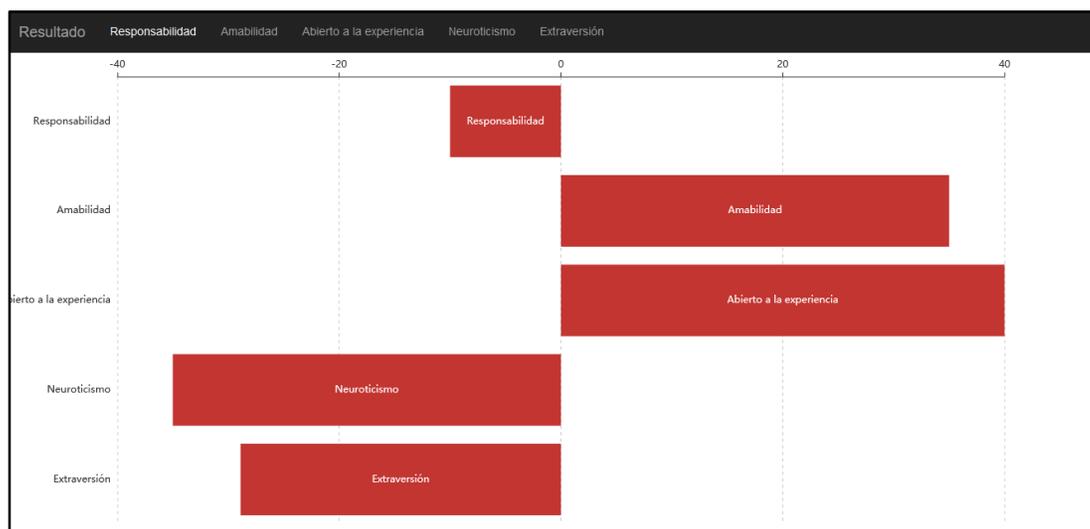


Figura 9 Resultado del Test de Personalidad

3.5.2. Aplicación del test de personalidad a Usuarios

Para la recolección de datos se necesita de personas que estén en búsqueda o futura búsqueda de trabajo, ya que es la población objetivo para este caso de estudio, los involucrados fueron los estudiantes próximos a egresar o de niveles superiores de la Universidad de las Fuerzas Armadas ESPE. Con el permiso del profesor a cargo de los estudiantes se les explicó el caso de estudio a realizar y como requisitos que deben tener una conexión a internet y una cuenta de Facebook, una vez ingresados en la página web <https://personalityprediction2018.herokuapp.com> y haber iniciado sesión en el aplicativo se les redirige a resolver el test de personalidad, tal como se muestra en las siguientes ilustraciones (ver Figuras 10, 11 y 12).



Figura 10 Escenario de Aplicación del Test de Personalidad

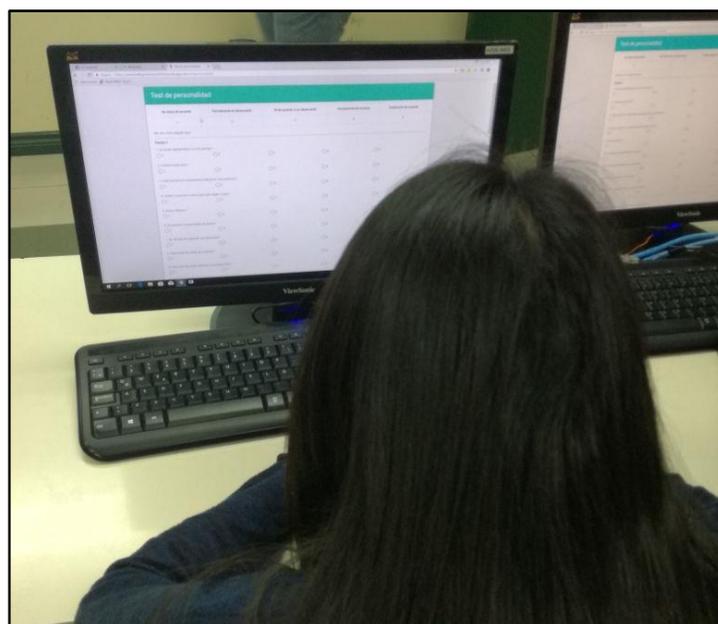


Figura 11 Usuario realizando el Test de Personalidad



Figura 12 Visualización de los resultados del Test de Personalidad

3.5.3. Resultados obtenidos del test de personalidad

Al finalizar los test que realizaron los estudiantes de la Universidad de las Fuerzas Armadas ESPE, brindó un total de 118 participantes, que arrojaron los siguientes resultados de personalidad:

En el siguiente gráfico (ver Figura 13) se muestra por porcentajes las personalidades, siendo el mayor con un 41% abierto a la experiencia, seguido de extraversión con un 37% y en tercer lugar a responsabilidad con 12%, las últimas dos personalidades poseen el 5% que son amabilidad y neuroticismo.

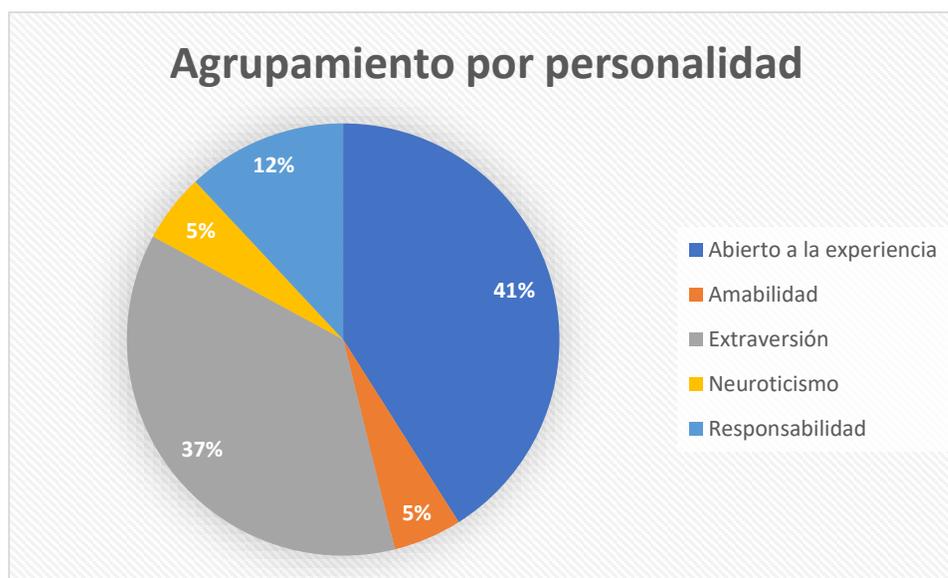


Figura 13 Agrupamiento por Personalidad

En el siguiente Gráfico (ver Figura 14) se observa la clasificación por género, donde las 118 personas que realizaron el test de personalidad, 74 de ellas fueron del género masculino mientras los 44 restantes del género femenino.

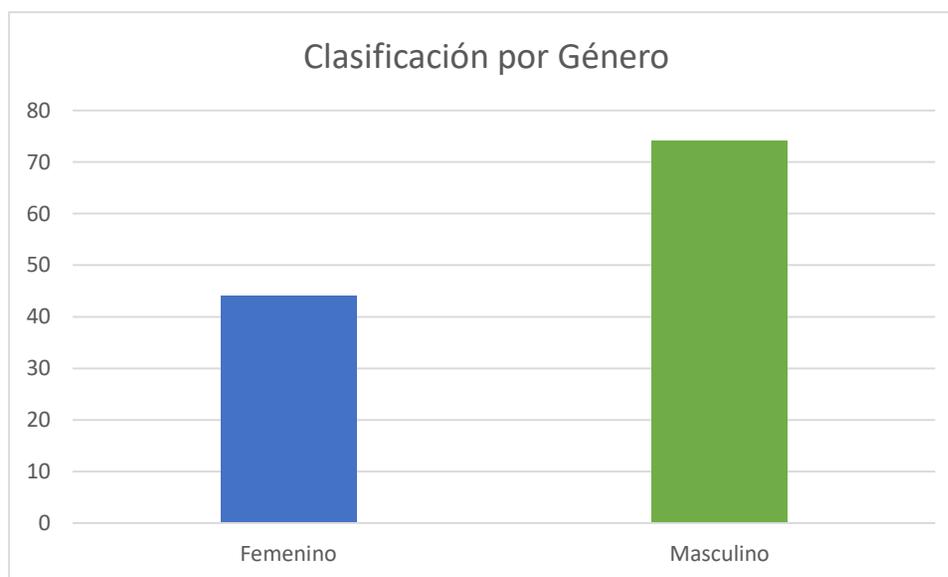


Figura 14 Clasificación por género

En el siguiente gráfico (ver Figura 15) se observa la clasificación por género y personalidad donde se obtuvo los siguientes resultados. Siendo el género masculino el de mayores resultados, para personalidad abierto a la experiencia se obtuvieron 33 por

parte masculina y 15 femenina, para la personalidad amabilidad se registraron 3 masculinos y 3 femeninos, de la personalidad extraversión se obtuvieron 27 masculinos y 17 femeninos, en la personalidad neuroticismo es el único caso donde el género femenino es mayor al masculino con 4 para femenino y 2 para masculino y por último la personalidad responsabilidad se obtuvieron 9 para masculino y 5 para el género femenino.

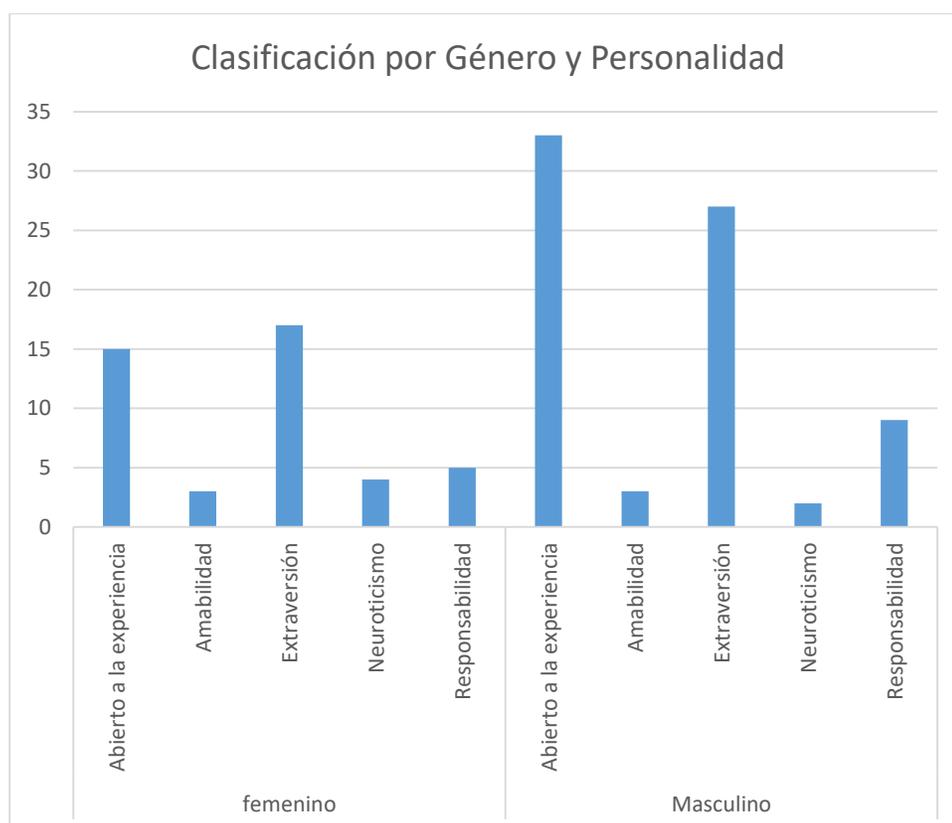


Figura 15 Clasificación por Género y Personalidad

Al final de este capítulo de desarrollo se obtuvo como resultado la base de conocimientos, que es esencial para el aprendizaje automático que se explica en el capítulo siguiente, de esta manera se describió cómo se obtuvieron los datos y todo el proceso de por medio para conseguir la base de conocimientos, también se explicó como fue el proceso y la realización del test de personalidad a los estudiantes de la Universidad de las Fuerzas Armadas ESPE y se finalizó mostrando los resultados de los test en gráficos.

CAPITULO IV

ANÁLISIS DE DATOS

4.1. Weka

Weka es una extensa colección de algoritmos de máquinas de conocimientos desarrollado por la universidad de Waikato en Nueva Zelanda, Weka esta implementado en Java y es de software de código abierto emitido bajo la licencia GNU, puede ser utilizado mediante la interface que ofrece o a su vez se puede implementar dentro de otras aplicaciones. Weka contiene las herramientas necesarias para realizar transformaciones sobre los datos, tareas de clasificación, regresión, clustering, asociación y visualización. Esta herramienta también tiene la posibilidad de añadir nuevas funcionalidades (Morate).

Weka en su interfaz de usuario se compone de las siguientes funcionalidades que son:

- Simple Cli: se trata de una consola que permite acceder a todas las opciones de Weka desde la línea de comandos.
- Explorer: dispone de varios paneles que dan accesos a los principales componentes del banco de trabajo como son:
 - Preproceso: este panel dispone de las opciones de importar datos de una base datos, de un fichero CSV, etc., y también para el filtrado de datos.
 - Clasificar: permite al usuario aplicar algoritmos de clasificación estadística y análisis de regresión, también permite estimar la exactitud del modelo predictivo.
 - Asociar: proporciona acceso a las reglas de asociación aprendidas que permiten identificar las interrelaciones importantes entre los atributos.
 - Cluster: permite el acceso a técnicas de clustering o agrupamiento.
 - Seleccionar atributos: permite ejecutar algoritmos para la selección de

atributos en sus datos para seleccionar los que son relevantes para la característica que se desea predecir.

- Visualizar: muestra una matriz de puntos dispersos(scatterplot) donde cada punto individual puede seleccionarse y agrandarse para ser analizados en detalle usando varios operadores de selección.
- Experimentar: permite la comparación sistemática de una ejecución de algoritmos predictivos de Weka sobre una colección de conjuntos de datos.

Weka es compatible con las tres plataformas principales: Windows, OS X y Linux, para poder ejecutar Weka es necesario tener instalado Java, la instalación viene incluido con un conjunto de datos experimentales estándar para ser aplicados y poder practicar.

Según lo menciona (Corso C, 2009), las técnicas de minería de datos provienen de la Inteligencia Artificial y de la estadística; las mismas no son más que algoritmos sofisticados aplicados a un conjunto de datos para obtener resultados.

Por otro lado, las técnicas de minería de datos se clasifican en: técnicas predictivas y técnicas descriptivas, parte esencial depende de la tarea que se va a realizar y el tipo de datos para seleccionar la técnica que se va a utilizar como se puede observar en el gráfico (ver Figura 16).

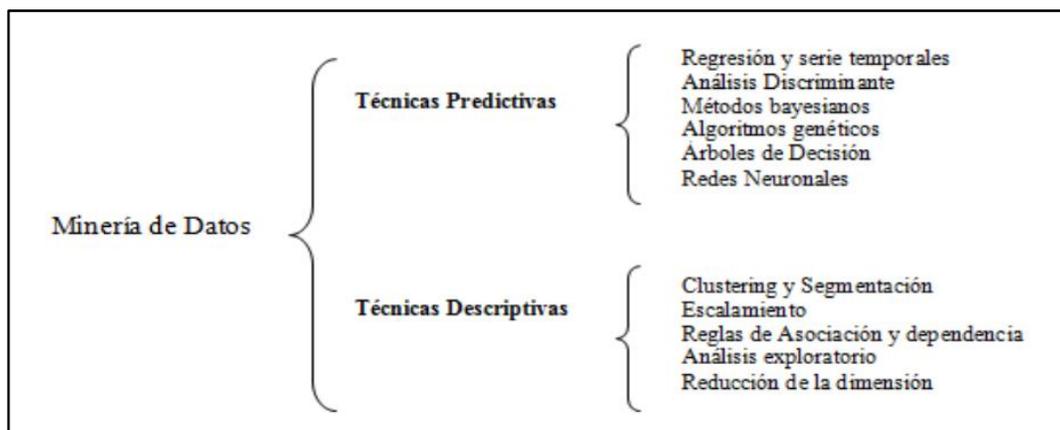


Figura 16 Técnicas de Minería de Datos

Fuente:(Corso C, 2009)

El estudio de Corso de 2009 define que el Árbol de decisión es:

Un árbol de decisión es un modelo de predicción utilizado en el ámbito de la inteligencia artificial, dada una base de datos se construyen estos diagramas de construcciones lógicas, muy similares a los sistemas de predicción basados en reglas, que sirven para representar y categorizar una serie de condiciones que suceden de forma sucesiva, para la resolución de un problema. Ejemplos: Algoritmo ID3, Algoritmo C4.5.

En el análisis de Díaz de 2015 define que los Algoritmos basados en Reglas:

Son una alternativa popular de los árboles de decisión. El antecedente o predicción de una regla es una serie de pruebas como las que se hacen en el nodo en árboles de decisión. El consecuente o conclusión da la clase o clases que aplica a instancias cubiertas por esa regla o tal vez da una probabilidad de distribución acerca de las clases. Una regla es generada por cada hoja. El antecedente de cada regla incluye la condición de cada nodo en el camino desde la raíz hasta la hoja y el consecuente de la regla es la clase asignada por la hoja.

4.2. Pre procesamiento de Datos

El pre procesamiento de datos se lo realiza mediante la obtención de un archivo plano, siendo que este archivo plano debe poseer las extensiones .csv o .arff.

Arff es una extensión del formato de archivo CSV donde se usa un encabezado que proporciona metadatos sobre los tipos de datos en las columnas, siendo el formato por default que recibe información el aplicativo Weka, su acrónimo es: Formato de Archivo de Relación Atributo.

Los campos o atributos son once los que se ha seleccionado para formar el archivo .cvs y son los siguientes (ver Tabla 10).

Tabla 10
Atributos del archivo .csv

Atributo	Descripción
user_number_post	Número de publicaciones por usuarios en Facebook.
user_number_shares	Número de enlaces compartidos en Facebook.
user_number_photos	Número de fotos compartidas en Facebook
user_number_update_perfil	Número de actualizaciones de perfil en Facebook.
p_neuro	Peso del número de palabras obtenidas en las publicaciones de los usuarios que se enmarcan bajo el perfil de neuroticismo.
p_extro	Peso del número de palabras obtenidas en las publicaciones de los usuarios que se enmarcan bajo el perfil de extrovertido.
p_respo	Peso del número de palabras obtenidas en las publicaciones de los usuarios que se enmarcan bajo el perfil de responsable.
p_amab	Peso del número de palabras obtenidas en las publicaciones de los usuarios que se enmarcan bajo el perfil de amabilidad.
user_genero	Género del usuario.
user_personality	Personalidad del usuario obtenida mediante el test de personalidad BFI.

Para el procesamiento se cuenta con 118 instancias y 11 atributos, de lo que se obtiene del grafico (ver Figura 18) que indica el numérico por cada atributo con relación a la personalidad de los usuarios. Estos atributos se representan en 5 colores que representan a los cinco tipos de personalidad (ver Figura 17) mencionados en el capítulo anterior.

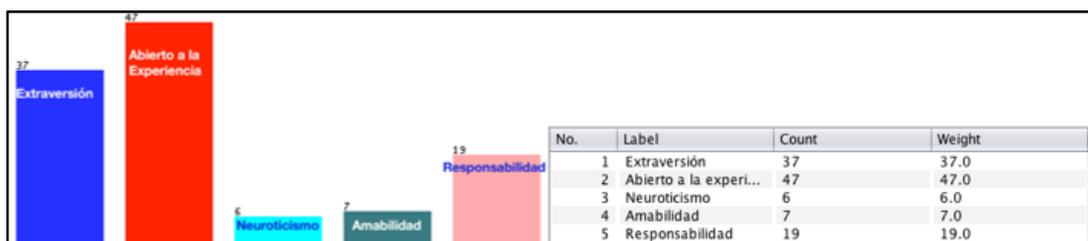


Figura 17 Numérico de los cinco tipos de personalidad

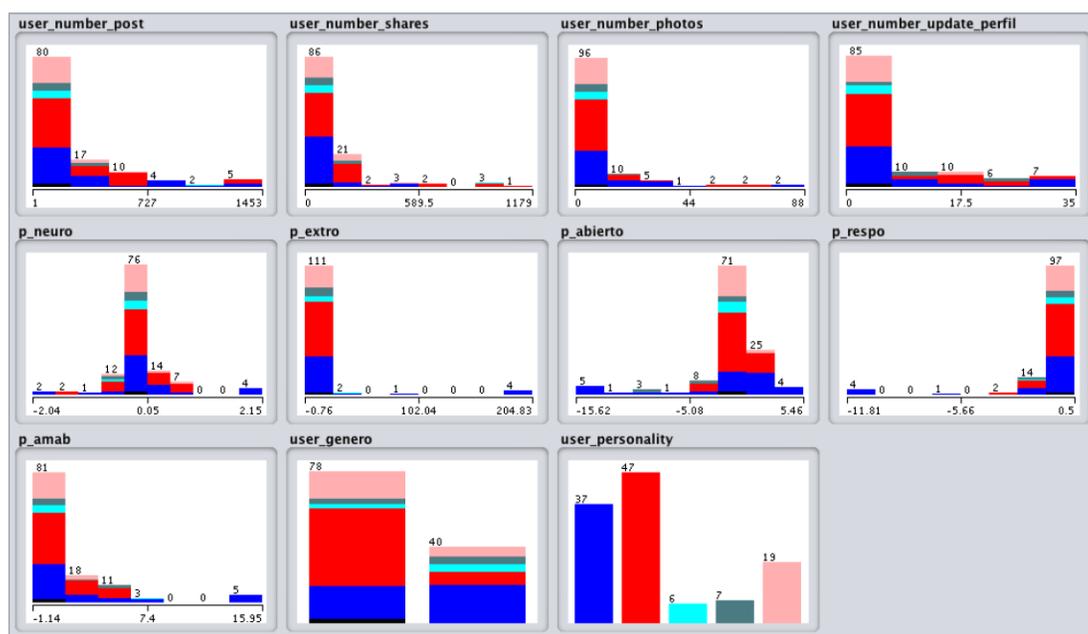


Figura 18: Atributo con relación a la personalidad de los usuarios.

4.3. Clasificación

En el presente apartado, se pretende construir un modelo que permita predecir la categoría de las instancias en función de una serie de atributos de entrada.

En ocasiones, el problema de clasificación se formula como un refinamiento en el análisis, una vez que se han aplicado algoritmos no supervisados de agrupamiento y asociación para describir relaciones de interés en los datos.

Para realizar la construcción del clasificador de datos, se procede a evaluar diferentes algoritmos de clasificación, esto se lo realizará en dos escenarios. Los dos

escenarios evalúan 12 algoritmos que se ubican bajo los diferentes clasificadores (ver Tabla 11) que posee el aplicativo Weka:

Tabla 11
Clasificación de los algoritmos

Tipo de Algoritmo	Algoritmo
Bayes	<ul style="list-style-type: none"> • NaiveBayes
Funciones	<ul style="list-style-type: none"> • Logistic • Multiplayer Perception
Lazy	<ul style="list-style-type: none"> • IBK • Kstar
Reglas	<ul style="list-style-type: none"> • ZeroR • Decision Table
Arboles	<ul style="list-style-type: none"> • J48 • Ramdom Forest • DecisionStump • Ramdom Tree • REPT Tree

Cada uno de los algoritmos expuestos en la Tabla 11, se encuentran especificados a continuación (ver Tabla 12).

Tabla 12
Funcionamiento de algoritmos de Weka

	ALGORITMO	FUNCIÓN
BAYES	NaiveBayes	Puede usar estimadores de densidad de kernel, lo que mejora el rendimiento si la suposición de normalidad es extremadamente incorrecta, también puede manejar atributos numéricos usando discretización supervisada
	Logistic	Construye modelos de regresión logística lineal.
FUNCIONES	Multiplayer	Red neuronal de propagación hacia atrás
	Perception	
LAZY	IBK	Es un clasificador vecino que usa la misma métrica de distancia. El número de vecinos más cercanos ($k = 1$ predeterminado) se puede especificar explícitamente en el editor de objetos o se puede determinar automáticamente usando la validación cruzada que deja salir uno, sujeto a un límite superior dado por el valor especificado. Las predicciones de más de un vecino se pueden ponderar según su distancia de la instancia de prueba, y se implementan dos fórmulas diferentes para convertir la distancia en un peso.
	KStar	Vecino más cercano con función de distancia generalizada.
REGLAS	ZeroR	Predice la clase mayoritaria (si es nominal) o el valor promedio (si es numérico).
	DecisionTable	Construye una tabla de decisión simple del clasificador mayoritario.
ÁRBOL	J48	Genera un árbol de decisión que utiliza el concepto de entropía de la información para la selección de variables que mejor clasifiquen a la variable (clase) estudiada.



Continúa

RandomForest	Construye árboles aleatorios. Son una combinación de predictores de árbol de modo que cada árbol depende de los valores de un vector aleatorio muestreado de forma independiente y con la misma distribución para todos los árboles en el bosque.
DecisionStump	Construye árboles de decisión de un nivel.
RandomTree	Construye un árbol que considera un número aleatorio de características dadas en cada nodo.
REPTree	Aprendizaje de árbol rápido que usa la poda en la reducción de errores.

Escenario 1: Usando data de entrenamiento

Para el escenario 1 se utilizará todas las 118 instancias para realizar el entrenamiento para construir el clasificador.

Escenario 1	ZeroR	Decision Table	J48	Ramdom Forest	DecisiónStump	Ramdom Tree	REPT Tree	NaiveBayes	Logistic	Multiplayer Perception	IBK	Kstar
Correctly Classified Instances	41.03%	42.74%	80.34%	100.00%	41.03%	100.00%	41.03%	29.91%	51.28%	62.39%	100.00%	100.00%
Incorrectly Classified Instances	58.97%	57.27%	19.66%	0.00%	58.97%	0.00%	58.97%	70.09%	48.72%	37.61%	0.00%	0.00%
Kappa statistic	0.000	0.055	0.704	1.000	0.000	0.000	0.000	0.167	0.228	0.405	1.000	1.000
Mean absolute error	0.273	0.273	0.108	0.101	0.265	0.000	0.271	0.290	0.229	0.206	0.013	0.009
Root mean squared error	0.368	0.367	0.232	0.141	0.364	0.000	0.368	0.468	0.339	0.314	0.016	0.055
Relative absolute error	100.00%	100.16%	39.59%	36.87%	97.20%	0.00%	99.26%	106.32%	84.02%	75.57%	4.81%	3.44%
Root relative squared error	100.00%	99.75%	63.14%	38.39%	98.94%	0.00%	99.98%	127.26%	92.05%	85.18%	4.45%	14.98%
Total Number of Instances	117	117	117	117	117	117	117	117	117	117	117	117
Ignored Class Unknown Instance	1	1	1	1	1	1	1	1	1	1	1	1

Figura 19 Ejecución de Algoritmos de 118 instancias aprendidas

En la figura 19, se observa que los algoritmos con mayor precisión en cuanto a la correcta clasificación son: Ramdom Tree, Ramdom Forest, IBK y Kstar, sin embargo si se toma en cuenta los errores se descarta Ramdom Forest y Kstar, debido a que lo que se pretende en el presente trabajo de investigación es obtener el menor error posible. El algoritmo Ramdom Tree presenta el 0% de error tras realizar el aprendizaje de 117 instancias. Es necesario mencionar que la instancia desconocida es la fila de la data que representa a cada campo.

A continuación, se presenta los resultados obtenidos del algoritmo Ramdom Tree.

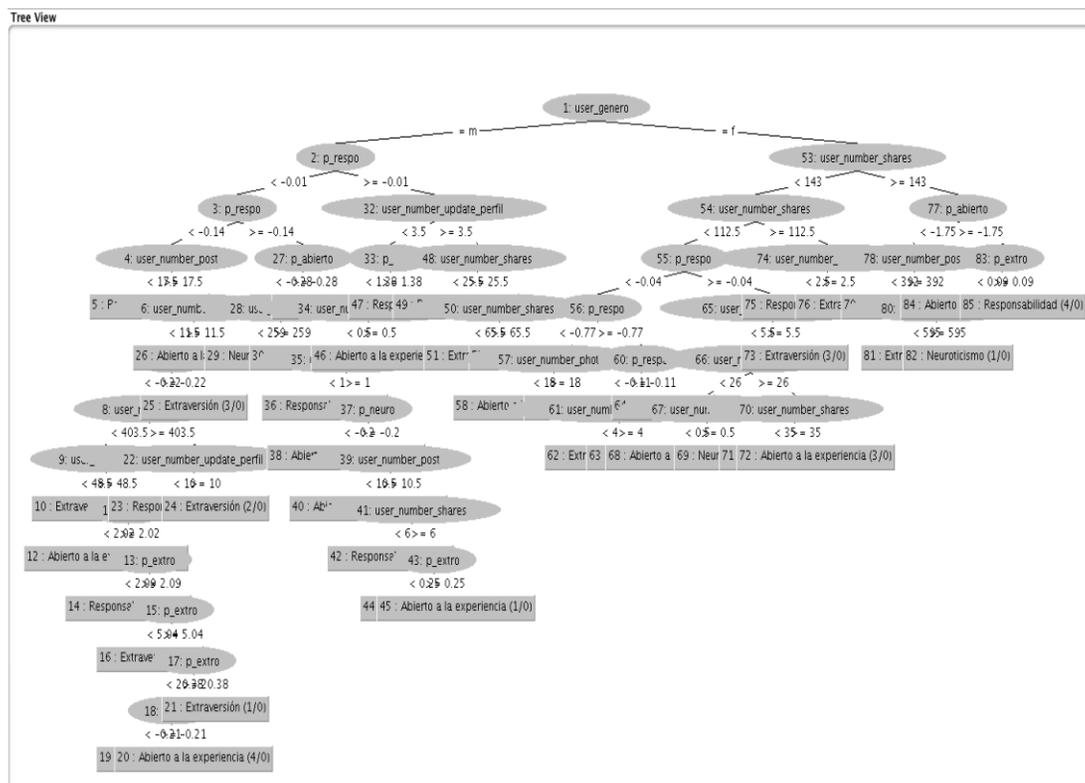


Figura 20 Árbol de decisiones.

En la Figura 20 se puede observar el árbol de decisiones del algoritmo RandomTree que se obtiene mediante el aplicativo Weka.

El algoritmo del árbol RandomForest es el siguiente:

```

user_genero = m
|
| p_respo < -0.01
| |
| | p_respo < -0.14
| | |
| | | user_number_post < 17.5 : Responsabilidad (3/0)
| | | user_number_post >= 17.5
| | | |
| | | | user_number_photos < 11.5
| | | | |
| | | | | p_respo < -0.22
| | | | | |
| | | | | | user_number_post < 403.5
| | | | | | |
| | | | | | | user_number_post < 48.5 : Extraversión (2/0)
| | | | | | | user_number_post >= 48.5
| | | | | | | |
| | | | | | | | p_extro < 2.02 : Abierto a la experiencia (12/0)
| | | | | | | | p_extro >= 2.02
| | | | | | | | |
| | | | | | | | | p_extro < 2.09 : Responsabilidad (1/0)
| | | | | | | | | p_extro >= 2.09
| | | | | | | | | |
| | | | | | | | | | p_extro < 5.04 : Extraversión (2/0)
| | | | | | | | | | p_extro >= 5.04
| | | | | | | | | | |
| | | | | | | | | | | p_extro < 20.38
| | | | | | | | | | | |
| | | | | | | | | | | | p_amab < -0.21 : Extraversión (1/0)
| | | | | | | | | | | | p_amab >= -0.21 : Abierto a la experiencia (4/0)
| | | | | | | | | | | | p_extro >= 20.38 : Extraversión (1/0)

```

```

| | | | | user_number_post >= 403.5
| | | | | user_number_update_perfil < 10 : Responsabilidad (1/0)
| | | | | user_number_update_perfil >= 10 : Extraversión (2/0)
| | | | | p_respo >= -0.22 : Extraversión (3/0)
| | | | | user_number_photos >= 11.5 : Abierto a la experiencia (12/0)
| | | | | p_respo >= -0.14
| | | | | p_abierto < -0.28
| | | | | user_number_post < 259 : Neuroticismo (2/0)
| | | | | user_number_post >= 259 : Extraversión (1/0)
| | | | | p_abierto >= -0.28 : Extraversión (4/0)
| | | | | p_respo >= -0.01
| | | | | user_number_update_perfil < 3.5
| | | | | p_extro < 1.38
| | | | | user_number_update_perfil < 0.5
| | | | | user_number_shares < 1 : Responsabilidad (2/0)
| | | | | user_number_shares >= 1
| | | | | p_neuro < -0.2 : Abierto a la experiencia (1/0)
| | | | | p_neuro >= -0.2
| | | | | user_number_post < 10.5 : Abierto a la experiencia (2/0)
| | | | | user_number_post >= 10.5
| | | | | user_number_shares < 6 : Responsabilidad (1/0)
| | | | | user_number_shares >= 6
| | | | | p_extro < 0.25 : Amabilidad (1/0)
| | | | | p_extro >= 0.25 : Abierto a la experiencia (1/0)
| | | | | user_number_update_perfil >= 0.5 : Abierto a la experiencia (8/0)
| | | | | p_extro >= 1.38 : Responsabilidad (4/0)
| | | | | user_number_update_perfil >= 3.5
| | | | | user_number_shares < 25.5 : Responsabilidad (2/0)
| | | | | user_number_shares >= 25.5
| | | | | user_number_shares < 65.5 : Extraversión (1/0)
| | | | | user_number_shares >= 65.5 : Amabilidad (2/0)
user_genero = f
| | | | | user_number_shares < 143
| | | | | user_number_shares < 112.5
| | | | | p_respo < -0.04
| | | | | p_respo < -0.77
| | | | | user_number_photos < 18 : Abierto a la experiencia (1/0)
| | | | | user_number_photos >= 18 : Amabilidad (1/0)
| | | | | p_respo >= -0.77
| | | | | p_respo < -0.11
| | | | | user_number_update_perfil < 4 : Extraversión (6/0)
| | | | | user_number_update_perfil >= 4 : Neuroticismo (1/0)
| | | | | p_respo >= -0.11 : Amabilidad (1/0)
| | | | | p_respo >= -0.04
| | | | | user_number_update_perfil < 5.5
| | | | | user_number_shares < 26
| | | | | user_number_photos < 0.5 : Abierto a la experiencia (2/0)
| | | | | user_number_photos >= 0.5 : Neuroticismo (2/0)
| | | | | user_number_shares >= 26
| | | | | user_number_shares < 35 : Extraversión (1/0)
| | | | | user_number_shares >= 35 : Abierto a la experiencia (3/0)
| | | | | user_number_update_perfil >= 5.5 : Extraversión (3/0)
| | | | | user_number_shares >= 112.5
| | | | | user_number_update_perfil < 2.5 : Responsabilidad (1/0)
| | | | | user_number_update_perfil >= 2.5 : Extraversión (9/0)
user_number_shares >= 143
| | | | | p_abierto < -1.75
| | | | | user_number_post < 392 : Amabilidad (2/0)
| | | | | user_number_post >= 392
| | | | | user_number_shares < 595 : Extraversión (1/0)
| | | | | user_number_shares >= 595 : Neuroticismo (1/0)
| | | | | p_abierto >= -1.75
| | | | | p_extro < 0.09 : Abierto a la experiencia (1/0)
| | | | | p_extro >= 0.09 : Responsabilidad (4/0)

```

El algoritmo RandomForest, presenta la matriz de confusión, en la que existen dos clasificaciones relevantes: i) Abierto a la experiencia y ii) Extraversión.

```

a b c d e <-- classified as
37 0 0 0 0 | a = Extraversión
0 47 0 0 0 | b = Abierto a la experiencia
0 0 6 0 0 | c = Neuroticismo
0 0 0 7 0 | d = Amabilidad
0 0 0 0 19 | e = Responsabilidad

```

Escenario 2: Utilizando un porcentaje de data de entrenamiento

Para ejecutar el escenario 2, se dividirá el conjunto de entrenamiento que conforman las 118 instancias, y se dividirá en dos partes: los primeros 70% de los datos se utilizará para construir el clasificador y el 30% finales, para realizar el test.

Escenario 2	ZeroR	Decision Table	J48	Random Forest	DecisiónStump	Random Tree	REPT Tree	NaiveBayes	Logistic	Multiplayer Perception	IBK	Kstar
Correctly Classified Instances	38,24%	44,12%	50,00%	70,59%	41,18%	67,65%	44,12%	47,06%	44,12%	44,12%	67,65%	76,47%
Incorrectly Classified Instances	61,76%	55,88%	50,00%	29,41%	58,82%	32,35%	55,88%	52,94%	55,88%	55,88%	32,35%	23,53%
Kappa statistic	0,0000	0,1151	0,2739	0,5658	0,0529	0,5621	0,1151	0,3045	0,1466	0,1545	0,5478	0,6596
Mean absolute error	0,2836	0,2733	0,2157	0,1698	0,2716	0,1294	0,2609	0,2143	0,2383	0,2429	0,1394	0,1235
Root mean squared error	0,3763	0,3764	0,3920	0,2833	0,3722	0,3597	0,3845	0,4394	0,3814	0,3666	0,3516	0,3115
Relative absolute error	100,00%	96,38%	76,07%	59,87%	95,77%	45,64%	92,02%	75,57%	84,04%	85,65%	49,17%	43,55%
Root relative squared error	100,00%	100,04%	104,18%	75,29%	98,92%	95,61%	102,20%	116,77%	101,37%	97,43%	93,45%	82,78%
Total Number of Instances	34	34	34	34	34	34	34	34	34	34	34	34
Ignored Class Unknown Instance	1	1	1	1	1	1	1	1	1	1	1	1

Figura 21 Escenario 2: Ejecución de Algoritmos de 35 instancias aprendidas

En la tabla (ver Figura 21), se observa que los algoritmos que mejores resultados brindaron tras el aprendizaje de 34 instancias que corresponde al 30% del total de la data, son los algoritmos: Random Forest y Kstar que están en un 70% de aciertos tras el aprendizaje obtenido del 30%.

4.3.1. Random Forest

El algoritmo Random Forest, tras realizar 100 iteraciones, presenta errores de clasificación del 29, 41%, motivos por el cual no es posible observar un árbol de decisiones, sin embargo, es posible observar en la figura (ver Figura 20) los errores obtenidos en la ejecución del aprendizaje del algoritmo:

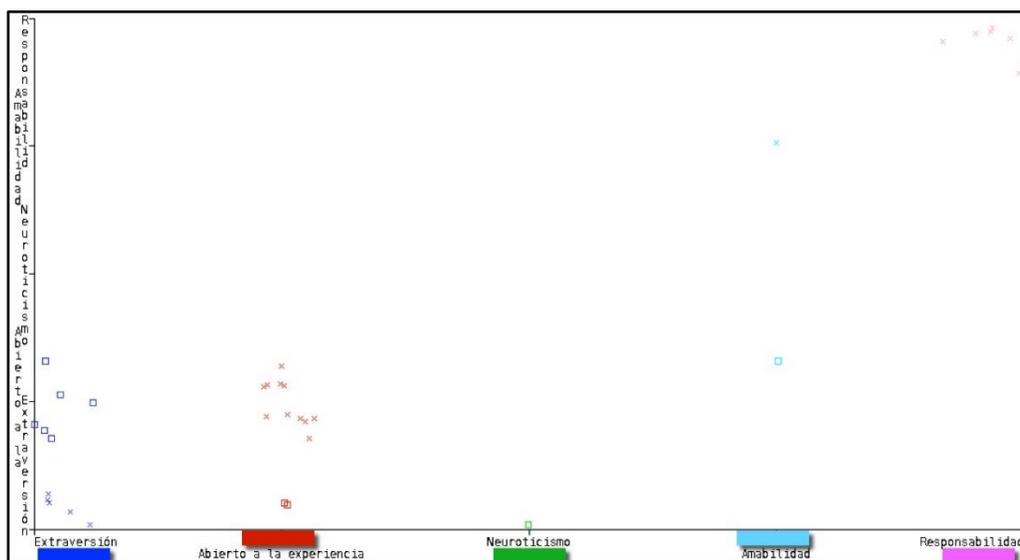


Figura 22 Errores generados de la ejecución de algoritmos Random

En la Figura 22 se puede observar los errores que generados, errores que de igual manera que en el algoritmo de RandomTree ejecutado en el escenario 1, se ha dividido en género: i) masculino () y femenino (x), de los cuales se puede observar que en la personalidad de extraversión presentan un balance en el error tanto hombres como mujeres. En la personalidad abierta a la experiencia, las mujeres presentan un mayor número de errores generados, estos errores son debidos a incompatibilidades de la data obtenida. En un escenario similar se encuentra la personalidad de responsabilidad, donde el error generado únicamente se encuentra en el género femenino.

El test de predicciones obtenido es el siguiente:

```
inst#,actual,predicted,error,prediction
1,1:Extraversión,1:Extraversión,,0.87
2,1:Extraversión,2:Abierto a la experiencia,+,0.62
3,2:'Abierto a la experiencia',2:Abierto a la experiencia,,0.8
4,5:Responsabilidad,5:Responsabilidad,,0.75
5,2:'Abierto a la experiencia',2:Abierto a la experiencia,,0.92
6,4:Amabilidad,2:Abierto a la experiencia,+,0.75
7,1:Extraversión,1:Extraversión,,0.64
8,1:Extraversión,1:Extraversión,,0.76
9,1:Extraversión,2:Abierto a la experiencia,+,0.69
10,5:Responsabilidad,5:Responsabilidad,,0.85
11,2:'Abierto a la experiencia',2:Abierto a la experiencia,,0.76
12,4:Amabilidad,4:Amabilidad,,0.66
13,2:'Abierto a la experiencia',2:Abierto a la experiencia,,0.93
14,2:'Abierto a la experiencia',2:Abierto a la experiencia,,0.77
15,2:'Abierto a la experiencia',1:Extraversión,+,0.37
16,2:'Abierto a la experiencia',2:Abierto a la experiencia,,0.8
```

```

17,5:Responsabilidad,5:Responsabilidad,,0.78
18,1:?,2:Abierto a la experiencia,,0.89
19,3:Neuroticismo,1:Extraversión,+,0.57
20,1:Extraversión,2:Abierto a la experiencia,+,0.51
21,2:'Abierto a la experiencia',2:Abierto a la experiencia,,0.77
22,1:Extraversión,2:Abierto a la experiencia,+,0.43
23,5:Responsabilidad,5:Responsabilidad,,0.71
24,1:Extraversión,1:Extraversión,,0.67
25,5:Responsabilidad,5:Responsabilidad,,0.75
26,2:'Abierto a la experiencia',2:Abierto a la experiencia,,0.82
27,1:Extraversión,1:Extraversión,,0.46
28,1:Extraversión,2:Abierto a la experiencia,+,0.45
29,5:Responsabilidad,5:Responsabilidad,,0.71
30,2:'Abierto a la experiencia',2:Abierto a la experiencia,,0.53
31,2:'Abierto a la experiencia',1:Extraversión,+,0.37
32,5:Responsabilidad,5:Responsabilidad,,0.75
33,2:'Abierto a la experiencia',2:Abierto a la experiencia,,0.53
34,1:Extraversión,2:Abierto a la experiencia,+,0.6
35,2:'Abierto a la experiencia',2:Abierto a la experiencia,,0.94

```

La matriz de confusión generada para el presente algoritmo, se expresa bajo la siguiente descripción:

```

a b c d e <-- classified as
5 6 0 0 0 | a = Extraversión
2 11 0 0 0 | b = Abierto a la experiencia
1 0 0 0 0 | c = Neuroticismo
0 1 0 1 0 | d = Amabilidad
0 0 0 0 7 | e = Responsabilidad

```

De la matriz de confusión se puede observar que sobresalen tres clasificaciones principales: i) Abierto a la experiencia, ii) Extraversión y iii) Responsable.

4.3.2. Kstar

El algoritmo Kstar, de 34 instancias aprendidas, 26 (76, 47%) instancias se han generado de manera correcta, mientras que 8(23,52%) instancias se han generado incorrectamente.

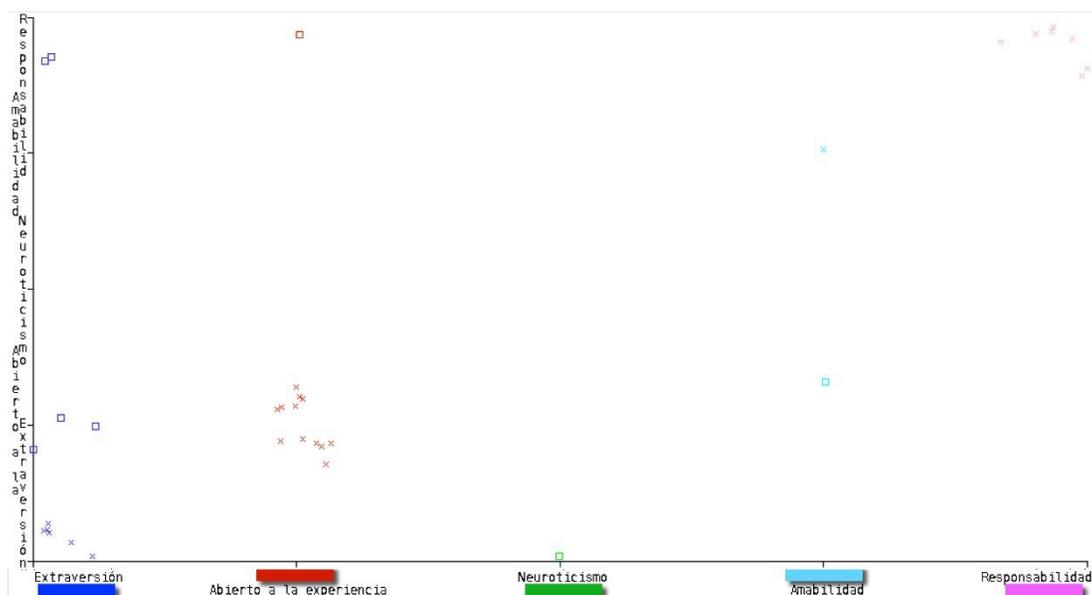


Figura 23 Errores generados de la ejecución del algoritmo Kstar

En la Figura 23 se observa que el mayor número de errores sucedidos se dan en la personalidad de abierto a la experiencia, seguido de la personalidad de extraversión y la personalidad de responsabilidad.

El test de predicción obtenido es el siguiente:

```

inst#,actual,predicted,error,prediction
1,1:Extraversión,1:Extraversión,,1
2,1:Extraversión,2:Abierto a la experiencia,+,0.997
3,2:'Abierto a la experiencia',2:Abierto a la experiencia,,1
4,5:Responsabilidad,5:Responsabilidad,,1
5,2:'Abierto a la experiencia',2:Abierto a la experiencia,,1
6,4:Amabilidad,2:Abierto a la experiencia,+,0.556
7,1:Extraversión,1:Extraversión,,1
8,1:Extraversión,1:Extraversión,,1
9,1:Extraversión,2:Abierto a la experiencia,+,1
10,5:Responsabilidad,5:Responsabilidad,,1
11,2:'Abierto a la experiencia',2:Abierto a la experiencia,,1
12,4:Amabilidad,4:Amabilidad,,1
13,2:'Abierto a la experiencia',2:Abierto a la experiencia,,1
14,2:'Abierto a la experiencia',2:Abierto a la experiencia,,1
15,2:'Abierto a la experiencia',2:Abierto a la experiencia,,0.496
16,2:'Abierto a la experiencia',2:Abierto a la experiencia,,1
17,5:Responsabilidad,5:Responsabilidad,,1
18,1:?,2:Abierto a la experiencia,,0.622
19,3:Neuroticismo,1:Extraversión,+,0.737
20,1:Extraversión,1:Extraversión,,0.383
21,2:'Abierto a la experiencia',2:Abierto a la experiencia,,1
22,1:Extraversión,2:Abierto a la experiencia,+,0.752
23,5:Responsabilidad,5:Responsabilidad,,0.523
24,1:Extraversión,1:Extraversión,,1

```

```

25,5:Responsabilidad,5:Responsabilidad,,1
26,2:'Abierto a la experiencia',2:Abierto a la experiencia,,0.492
27,1:Extraversión,1:Extraversión,,0.995
28,1:Extraversión,5:Responsabilidad,+,0.963
29,5:Responsabilidad,5:Responsabilidad,,1
30,2:'Abierto a la experiencia',2:Abierto a la experiencia,,0.93
31,2:'Abierto a la experiencia',2:Abierto a la experiencia,,0.496
32,5:Responsabilidad,5:Responsabilidad,,1
33,2:'Abierto a la experiencia',5:Responsabilidad,+,0.952
34,1:Extraversión,5:Responsabilidad,+,0.966
35,2:'Abierto a la experiencia',2:Abierto a la experiencia,,1

```

La matriz de confusión generada para el presente algoritmo, se expresa bajo la siguiente descripción:

```

a b c d e <-- classified as
6 3 0 0 2 | a = Extraversión
0 12 0 0 1 | b = Abierto a la experiencia
1 0 0 0 0 | c = Neuroticismo
0 1 0 1 0 | d = Amabilidad
0 0 0 0 7 | e = Responsabilidad

```

De la matriz de confusión se puede observar que sobresalen tres clasificaciones principales: i) Abierto a la experiencia, ii) Extraversión y iii) Responsable.

4.4. Clusterización

Para realizar la clusterización se lo ha realizado con los siguientes algoritmos de clusterización:

De los algoritmos de clusterización antes mencionados, se procede a realizar el desarrollo de cada uno de ellos, para cada algoritmo se ha tomado 500 iteraciones como un máximo.

4.4.1. Canopy

El algoritmo de agrupamiento canopy es un algoritmo no supervisado, algoritmo de agrupación introducido. Se utiliza a menudo como etapa de pre procesamiento para el algoritmo k-means. Está destinado a acelerar las operaciones de agrupamiento en grandes conjuntos de datos, donde el uso de otro algoritmo directamente puede ser poco práctico debido al tamaño del conjunto de datos.

El algoritmo procede de la siguiente manera:

- Comienza con el conjunto de puntos de datos que se agruparán.
- Retire un punto del conjunto, comenzando un nuevo “canopy”.
- Para cada punto que quede en el conjunto, asígnalo al nuevo canopy, si la distancia es menor que la distancia libre
- Si la distancia del punto es adicionalmente menor que la distancia, quítelo del conjunto original.
- Repita desde el paso 2 hasta que no haya más puntos de datos en el conjunto para agrupar.
- Estas marquesinas agrupadas relativamente baratas se pueden agrupar por debajo utilizando un algoritmo más caro pero preciso.

El algoritmo de Canopy provee cuatro clusters de información:

```
Cluster 0: 178.584615,110.692308,6.630769,6.046154,0.025385,3.364615,-0.637231,-
0.461231,1.359385,f,'Abierto a la experiencia',{65} <0,3,4>
Cluster 1: 837.333333,484,5,0.666667,-0.073333,20.96,-0.676667,-1.55,3.54,m,Neuroticismo,{3} <1,2,3>
Cluster 2: 224.189189,139.540541,5.297297,5.513514,-0.165405,2.953784,-0.974324,-
0.463243,1.119459,m,Extraversión,{37} <1,2,3,4>
Cluster 3: 850,545.4,38,14.4,0.096,8,4.03,-2.37,1.714,m,'Abierto a la experiencia',{5} <0,1,2,3>
Cluster 4: 452,227,34.5,8.5,0.32,23.075,-7.93,-3.495,6.715,f,Extraversión,{2} <0,2,4>
```

Los cuatro clusters generados, se ha tomado en cuenta la personalidad como atributo de la clase a evaluar en la clusterización, se presentan el siguiente porcentaje:

0	69 (58%)	Cluster 0 <-- Abierto a la experiencia
1	4 (3%)	Cluster 1 <-- No class
2	5 (4%)	Cluster 2 <-- No class
3	34 (29%)	Cluster 3 <-- Extraversión
4	4 (3%)	Cluster 4 <-- No class
5	2 (2%)	Cluster 5 <-- Neuroticismo

De los resultados obtenidos se expone que las personalidades de Abierto a la experiencia, Extraversión y Neuroticismo son los relevantes en el algoritmo de clusterización Canopy. Realizando una evaluación por género se obtiene la Figura 22, donde se observa de la evaluación de 118 usuarios, existen más hombres con personalidad de extraversión, abiertos a la experiencia y responsables que las mujeres. Adicional, se observa que existen hombres y mujeres igual con personalidad de Amabilidad.

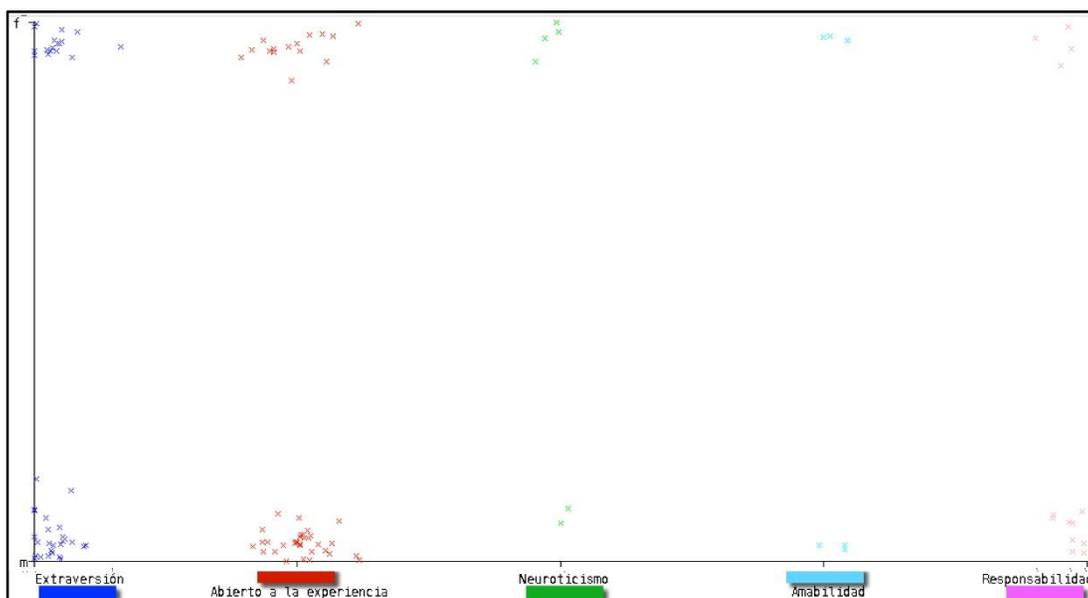


Figura 24 Evaluación del algoritmo Canopy mediante el género.

4.4.2. Filtered Clustered

Utiliza una clase para ejecutar un clúster arbitrario en datos que se han pasado a través de un filtro arbitrario. Al igual que el clúster, la estructura del filtro se basa exclusivamente en los datos de entrenamiento y las instancias de prueba serán procesadas por el filtro sin cambiar su estructura.

El algoritmo Filtered Clustered se ejecutó en 4 iteraciones, del cual los tipos de personalidades relevantes en las instancias son: i) Abierto a la experiencia (78%) y ii) Extraversión (22%).

Attribute	Full Data (118.0)	Cluster#	
		0 (26.0)	1 (92.0)
user_number_post	254.1271	726.7308	120.5652
user_number_shares	128.0678	326.5385	71.9783
user_number_photos	9.3983	29.6923	3.663
user_number_update_perfil	6.9153	18.5385	3.6304
p_neuro	-0.0208	-0.005	-0.0253
p_extro	11.9802	46.1219	2.3314
p_abierto	-1.0815	-4.4423	-0.1317
p_respo	-0.8719	-2.9438	-0.2863
p_amab	1.7791	5.8973	0.6152
user_genero	m	m	m

Figura 25 Resultado del Filtered Clustered

Las clases generadas en el clúster son los siguientes:

```

0 1 <-- assigned to cluster
10 27 | Extraversión
13 34 | Abierto a la experiencia
1 5 | Neuroticismo
2 5 | Amabilidad
0 19 | Responsabilidad

```

Al realizar un agrupamiento por género, en la Figura 26 se puede observar que utilizando el algoritmo Filtered Clustered, existen más hombres con una personalidad abierta a la experiencia.

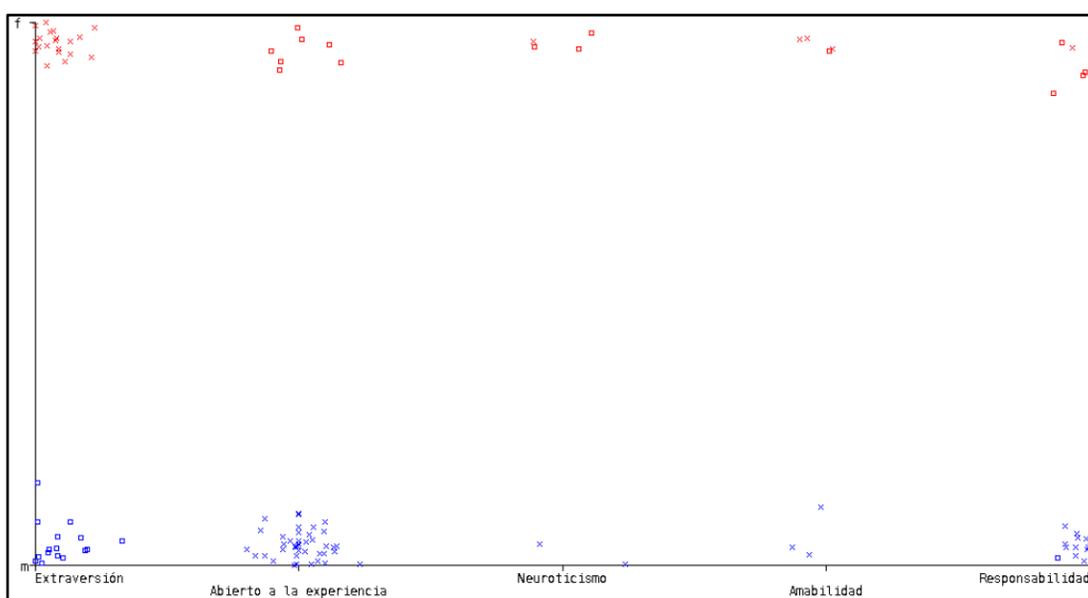


Figura 26 Filtered Clustered en agrupamiento por género

Los dos clústeres sobresalientes son Abierto a la experiencia y Extraversión, estos se representan numéricamente a continuación (ver Tabla 15).

Tabla 13
Resultados Filtered Clustered

Abierto a la experiencia	Extraversión	Género
60	18	Masculino
15	25	Femenino

4.4.3. Hierarchical Clusterer

El clúster jerárquico construye una jerarquía de clúster o, en otras palabras, un árbol de clústeres, también conocido como dendrograma.

Aglomerativo (de abajo hacia arriba)

1. Comience con 1 punto (singleton).
2. Agregue recursivamente dos o más clústeres apropiados.
3. Deténgase cuando se alcanza el número de k agrupaciones.

Divisivo (de arriba hacia abajo)

1. Comience con un gran grupo.
2. Divide recursivamente en grupos más pequeños.
3. Deténgase cuando se alcanza el número de k agrupaciones.

Las clases para el algoritmo Hierarchical Clusterer, se han dividido en las cinco personalidades, adicionando un clúster propio del algoritmo.

```
0 1 <-- assigned to cluster
33 4 | Extraversión
47 0 | Abierto a la experiencia
6 0 | Neuroticismo
7 0 | Amabilidad
19 0 | Responsabilidad
```

Sin embargo, el algoritmo ha definido dos clústeres relevantes:

```
Cluster 0 <-- Abierto a la experiencia 97% 114
Cluster 1 <-- Extraversión 3% 4
```

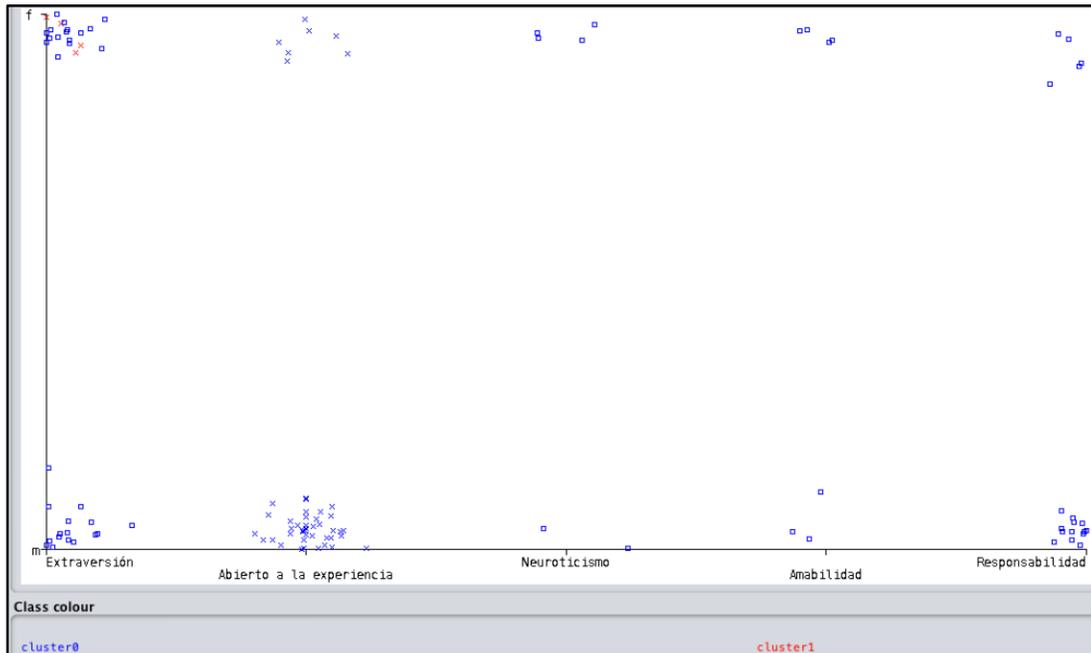


Figura 27 Clusterización Hierarchical por el género y la personalidad

En la Figura 27 se observa el algoritmo Hierarchical Clusterer en relación al género y a la personalidad, donde la personalidad de Abierto a la experiencia se presenta en un 97%, mientras que la personalidad de Extraversión obtiene un 4%.

4.4.4. Simple K means

K-means es un método de agrupamiento particionado ampliamente utilizado en las industrias. El algoritmo K-means es el algoritmo de clúster particionado más comúnmente usado porque se puede implementar fácilmente y es el más eficiente en términos del tiempo de ejecución. El algoritmo para la partición, donde el centro de cada grupo está representado por valor medio de los objetos en el clúster (Chaudhari B., 2012).

Método:

1. Elija arbitrariamente k objetos desde D como centro inicial del clúster.
2. Repita.
3. Reasignar cada objeto al clúster al cual el objeto es más similar usando Eq. 1, basado en el valor medio de los objetos en el clúster.

4. Actualice los medios del clúster, es decir, calcule el valor medio de los objetos para cada grupo.
5. Hasta que no haya cambios.

El algoritmo Simple K Means ha subdivido en dos clústeres tras ejecutar siete iteraciones:

```
Cluster 0: 622,594,1,3,0,-0.25,0,0,0,m,'Abierto a la experiencia'
Cluster 1: 9,1,1,0,-0.2,0.25,0,-0.3,0,m,Responsabilidad
```

La información obtenida es la siguiente:

Atributo	Full Data (118.0)	Cluster 0 (77.0)	Cluster 1 (41.0)
user_number_post	254.1271	193.6104	367.7805
user_number_shares	128.0678	102.8442	175.439
user_number_photos	9.3983	6.5714	14.7073
user_number_update_perfil	6.9153	5.3506	9.8537
p_neuro	-0.0208	-0.0718	0.0749
p_extro	11.9802	3.7639	27.4107
p_abierto	-1.0815	-0.2918	-2.5646
p_respo	-0.8719	-0.4468	-1.6702
p_amab	1.7791	0.8118	3.5956
user_genero	m	m	f
user_personality	Abierto a la experiencia	Abierto a la experiencia	Extroversión

Las instancias obtenidas, son las siguientes:

```
Cluster 0    77 ( 65%)  Abierto a la experiencia
Cluster 1    41 ( 35%)  Extroversión
```

En cuanto a la separación de género, se obtuvo que 60 hombres presentan una personalidad abierta a la experiencia y 18 hombres presentan una personalidad extravertida. Mientras que las mujeres, 15 presentan una personalidad abierta a la experiencia y 25 mujeres presentan una personalidad de extraversión.

Esto se puede observar en la Figura 28 se observa la división del algoritmo Simple K means considerando los dos clusters: i) Cluster 0: Abierto a la experiencia (azul) y ii) Cluster 1: Extraversión (rojo).

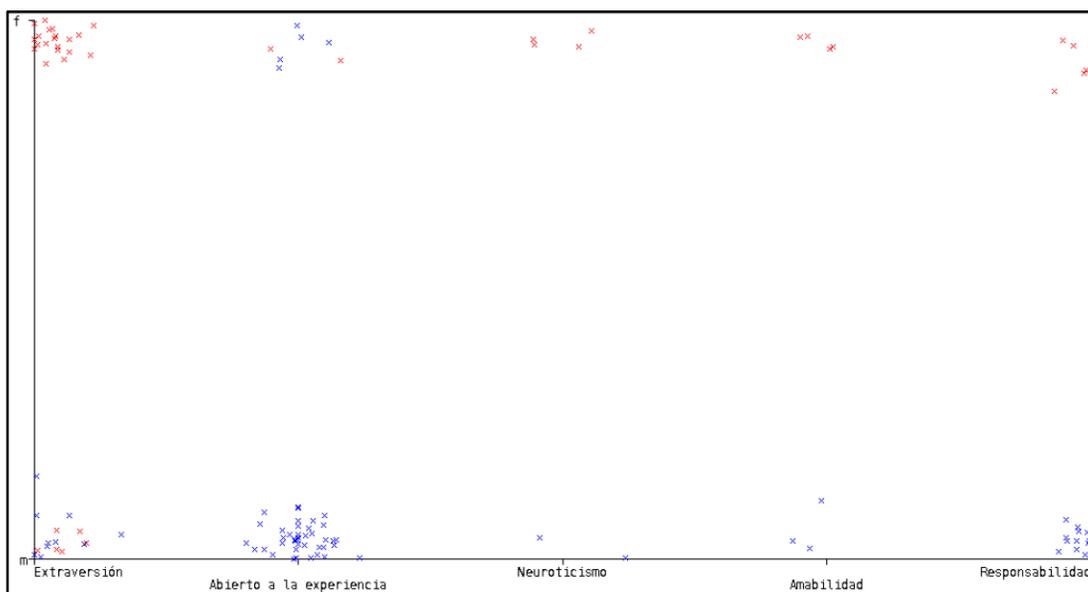


Figura 28: Simple Kmeans (Abierto a la experiencia y Extroversión)

4.5. Regla de Asociación

La regla de asociación para los algoritmos supervisados se ha tomado la regla FilteredAssociator que es una clase creada para ejecutar un asociador arbitrario en datos que se han pasado a través de un filtro arbitrario. Al igual que un asociador, la estructura del filtro se basa exclusivamente en los datos de capacitación y el filtro procesará las instancias de prueba sin cambiar su estructura.

A partir de las reglas de asociación se obtiene la visualización de los diferentes atributos que diferencian a cada personalidad.

Como se observa en la Figura 29, las personas extrovertidas pueden llegar a número de publicaciones de 1453, pero la mayoría de ellos se encuentran bajo las 727 publicaciones. Para la personalidad abierto a la experiencia tienen un número de publicaciones menor a los 727 y es fuera de lo normal que lleguen a los 1453. Para el resto de las personalidades se encuentran por debajo de las 300 publicaciones.

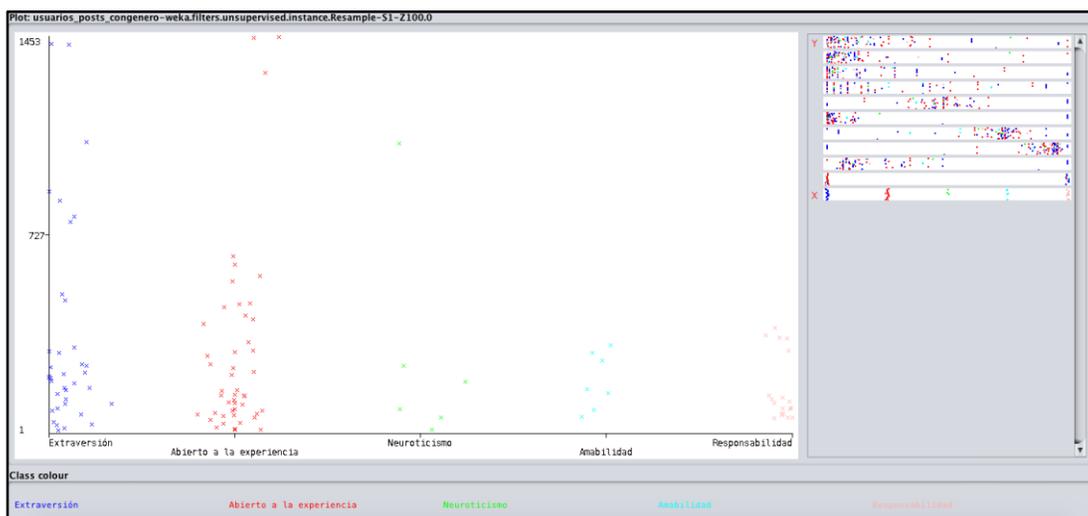


Figura 29 Personalidad con el número de post

En la Figura 30 las personas extrovertidas tienen como máximo las 589 publicaciones compartidas, los de abierto a la experiencia en su mayoría están por debajo de las 589, pero raramente pueden llegar hasta las 1179 publicaciones compartidas. Los neuróticos se encuentran en los niveles más bajos, siendo su tendencia a no compartir, mientras que amabilidad y responsabilidad están por debajo de las 250 publicaciones compartidas.

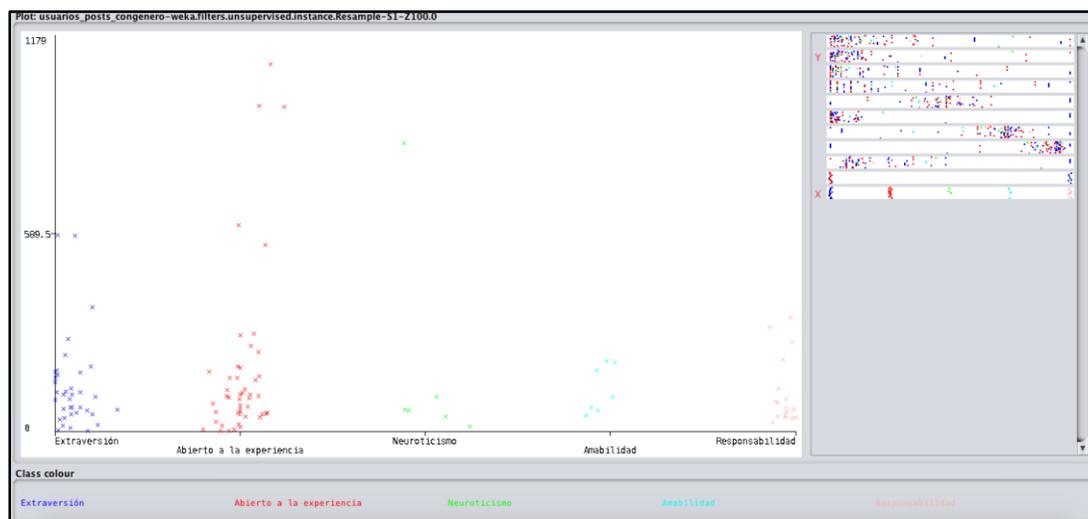


Figura 30 Personalidad con numero de shares

Tal como se puede observar en la Figura 31, las personas con la personalidad extraversión, llegan a tener unas 45 fotos subidas, pero en extraños casos llegan a las 88. Los de personalidad abierto a la experiencia llegan a unas 70 fotos, pero la mayoría se encuentra por debajo de los 30. Las personalidades neuroticismo, amabilidad y responsabilidad están por debajo de las 15 fotos subidas, siendo los de neuroticismo los que menos fotos suben.

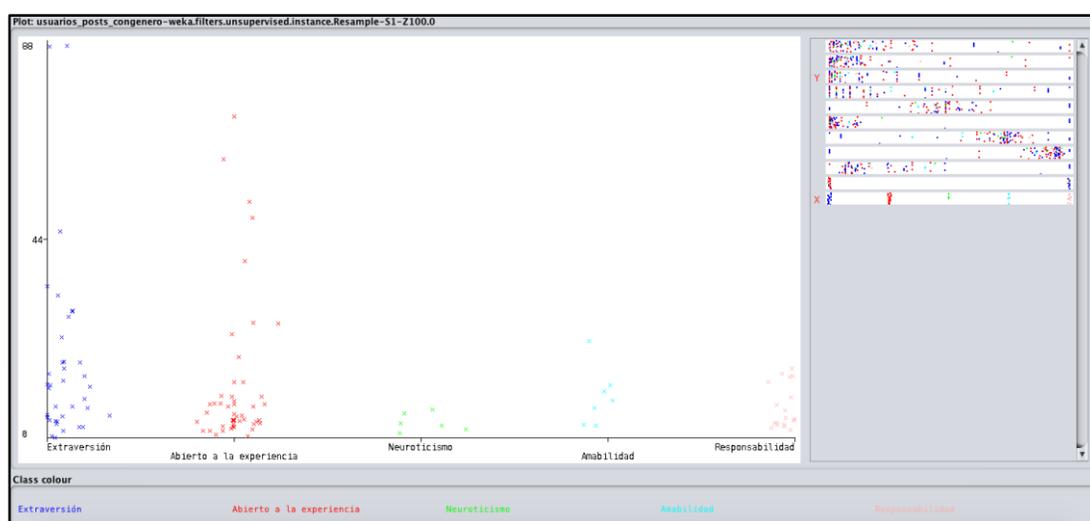


Figura 31 Personalidad con el número de fotos

En la Figura 32 se observa que las personalidades extroversión y abierto a la experiencia llegan a tener un máximo de 35 actualizaciones de foto de perfil. Los de personalidad responsabilidad llegan hasta las 18 actualizaciones y los de amabilidad menor a las 15 actualizaciones, pero raramente llegan a las 22 actualizaciones de la foto de perfil. Para este caso las personas neuróticas no tienen muchas actualizaciones de foto de perfil, manteniéndose por debajo de las 5 actualización de perfil

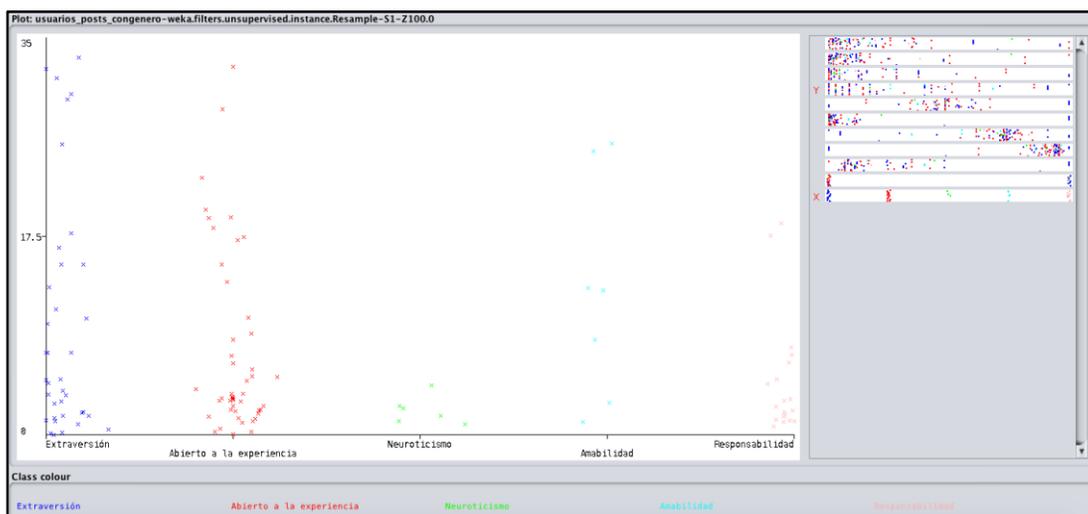


Figura 32 Personalidad con el número de actualizaciones

En la Figura 33 se observa como las cinco personalidades se están manteniendo en un rango entre 0 y 1 de las palabras enmarcadas en neurótico, pero la personalidad abierto a la experiencia tiene una tendencia a valores negativos, mientras que extraversión puede estar en rangos mayores de -2 a 2.

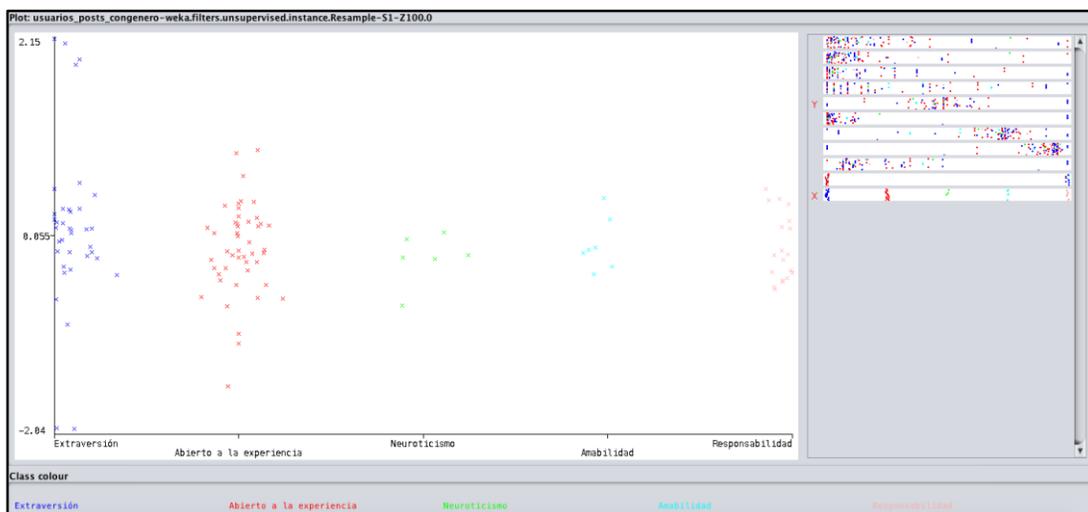


Figura 33 Personalidad con el peso de las palabras enmarcadas en neuróticas

Como se puede observar en la figura 34, las cinco personalidades están en un rango de -0.76 a 30 en el peso de las palabras enmarcadas en extrovertidas, pero la personalidad extraversión pueden llegar a tener un peso de 204.

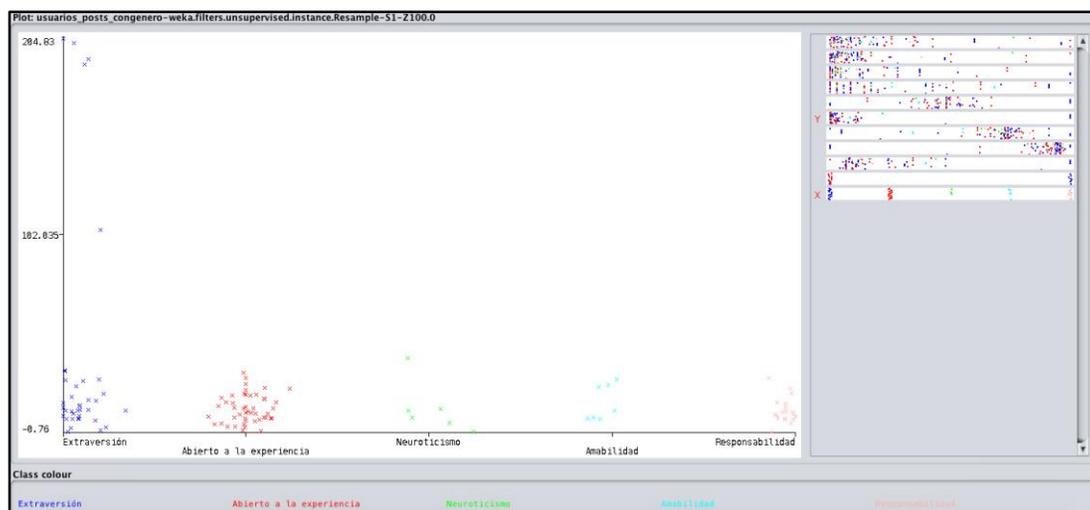


Figura 34 Personalidad con el peso de las palabras enmarcadas en extrovertidas

En la Figura 35 se muestra que todas las personalidades excepto extraversión se encuentran en un rango de -5.08 a 5.46 en el peso de palabras correspondiente en abiertos a la experiencia. La personalidad extraversión puede llegar a -15.62 en el peso, esto indica que una persona dice las palabras que van en contra de abierto a la experiencia le hacen una persona más extrovertida.

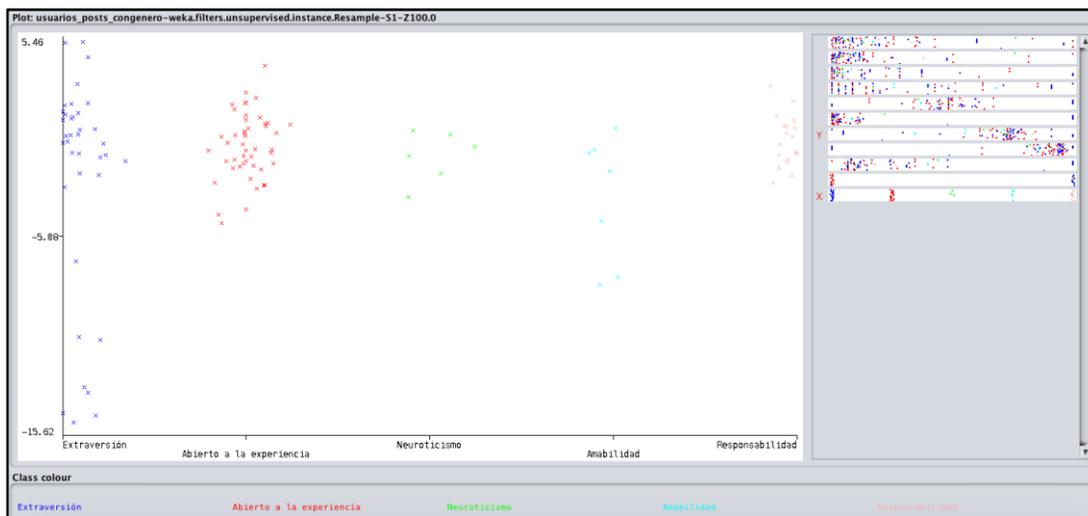


Figura 35 Personalidad con el peso en abiertos a la experiencia

En la figura 36 se observa que las personalidades están en un rango de 0.5 a -3 en el peso de las palabras correspondiente a responsabilidad, las personalidades amabilidad y neuroticismo tienden a cero. La personalidad extraversión en cambio puede llegar a -11.81 en los pesos, con el fin que entre mayor sea las palabras negativas en responsabilidad, la personalidad puede ser extraversión.

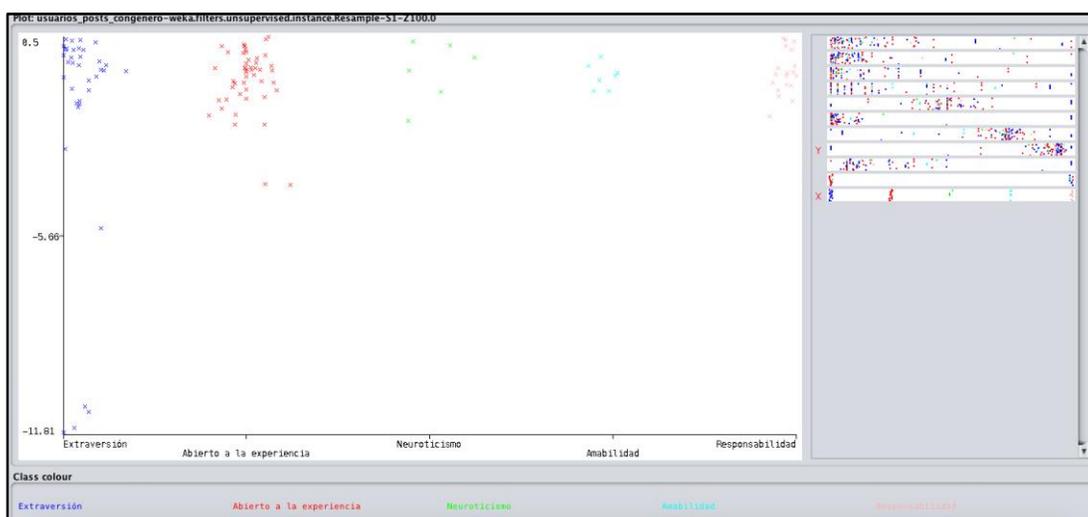


Figura 36 Personalidad con el peso en responsables

En la Figura 37 se puede determinar los siguiente:

- El género masculino es el que tiene mayor número de resultados.
- La personalidad neuroticismo tiene mayor frecuencia en el género femenino.
- El género masculino tiene mayor tendencia a ser abierto a la experiencia que la del género femenino.

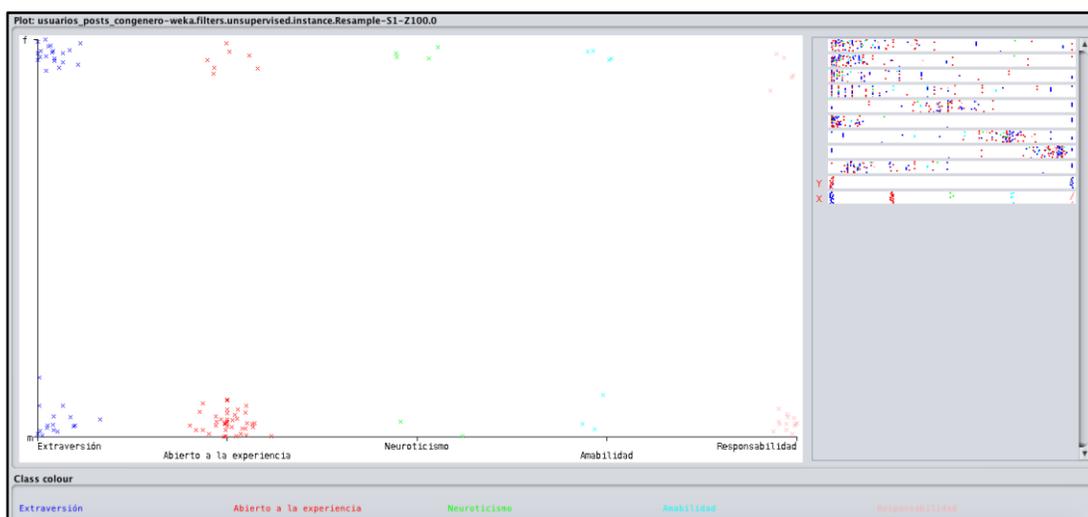


Figura 37 Personalidad según su género

4.6. Discusión

¿El enfoque del peso de las palabras según la personalidad pueden predecir la personalidad del usuario?

Con el análisis de datos que se obtuvieron de los 118 participantes, se demostró que las palabras no son de confiabilidad al momento de predecir la personalidad, ya que los resultados de las palabras para las personalidades en responsabilidad y abierto a la experiencia mientras más negativas eran, le correspondían a la personalidad extraversión. El único resultado favorable fue para las palabras que corresponden a la personalidad extraversión, mientras mayor sea su peso este le corresponde a su personalidad. Esto se debe a que las palabras ingresadas son sacadas de un trabajo

donde se involucró personas del idioma Ingles, mientras que las pruebas realizadas son a estudiantes del idioma español, este cambio de idiomas y de traducir las palabras al español, dio resultados no favorables al estudio de las palabras.

¿El aprendizaje automático en este caso de estudio puede predecir con exactitud la personalidad?

Con los datos obtenidos y los algoritmos utilizados si se puede predecir la personalidad, pero no en todos los casos fueron favorables, solo con el algoritmo "Random Tree" se obtuvo 0% de error, este algoritmo funciona de tal manera que se construye un árbol que considera un número aleatorio de características dadas en cada nodo, de esta manera la predicción se ajusta a las necesidades del algoritmo para dar el correcto resultado de la personalidad. También se demostró que el género femenino presenta características que le dificultan a los algoritmos en predecir su personalidad.

CAPÍTULO V

CONCLUSIONES Y FUTURAS LÍNEAS DE INVESTIGACIÓN

5.1. Conclusiones

En el Ecuador la red social más utilizada es Facebook, que cuenta con 11 millones de usuarios de los cuales 6 millones están activos mensualmente, lo cual para este proyecto le permite una extracción de datos con un alto porcentaje de confiabilidad.

La extracción de datos mediante la API de Facebook exige al desarrollador medidas de seguridad como son los protocolos SSH y demostrar que los datos extraídos son para mejorar la experiencia del usuario. Esto le concede al usuario una seguridad en que los datos otorgados a la aplicación serán para su beneficio.

El test de personalidad BFI permite conocer la personalidad del usuario, el cual es el sustento para llenar la base de conocimiento necesaria para el aprendizaje automático.

Las palabras que le corresponden a cada personalidad según los algoritmos utilizados no son de confiabilidad al momento de predecir la personalidad, esto se debe a la traducción de las palabras del inglés al español y que en el Ecuador se usa jergas diferentes a las usadas en el caso de estudio donde se obtuvieron las palabras.

Los resultados del aprendizaje supervisado en Weka, arrojaron datos en los cuales las mujeres tienen tendencias a ser de personalidad neurótica a comparación de los hombres. Estos datos también determinaron que las mujeres son más difíciles de predecir que los hombres con un mayor margen de error.

5.2. Recomendaciones

Se recomienda utilizar otras redes sociales tales como son Twitter o Instagram para la recolección de datos, de esta manera se puede tener más datos con los cuales se puede realizar una mejor predicción.

Se sugiere al desarrollador que utilice herramientas de desarrollo de mayor tendencia y que sean compatibles con la API Graph de Facebook para poder extraer los datos sin complicaciones.

Se recomienda usar expertos en el ámbito de psicología para realizar los test de personalidad que orienten a usar otro tipo de test que den mejores resultados, ya que esto permite tener un mayor grado de confiabilidad de los datos extraídos.

Se sugiere realizar un estudio de las palabras más usadas según la personalidad en la región del Ecuador, ya que se usan diferentes jergas a las de otras regiones, de esta manera se puede realizar una predicción de la personalidad con mayor confiabilidad.

Se recomienda realizar una base de conocimientos para el aprendizaje supervisado, donde exista un número equilibrado de hombres y mujeres de la muestra tomada y a su vez esta debe tener un mayor número de participantes que en el trabajo propuesto.

5.3. Futuras Líneas de Investigación

Para futuras líneas de investigación, se propone tomar los indicadores de me gusta del usuario de Facebook, este indicador para la personalidad del usuario refleja sus gustos, que tipo de páginas en Facebook le atraen al usuario, por ejemplo, un usuario puede tener mayores intereses en las artes, historia, ingeniería etc... este tipo de intereses pueden desarrollar un cierto tipo de personalidad.

La próxima propuesta es que se aumente la base de conocimientos, para este proyecto fue de 118 integrantes, entre más grande sea la base de conocimientos y los datos de en ella sean de mayor confiabilidad, tomando otras muestras de varias universidades o jóvenes que estén en búsqueda de empleo, la predicción de la personalidad será de mayor confiabilidad.

Se propone realizar un estudio a la foto de perfil del usuario de Facebook, en las redes sociales las fotos de cada persona son una forma de auto presentación ante los demás, se puede utilizar técnicas de transformación de imagen en un conjunto de datos

con el fin de obtener la personalidad del usuario.

Para un futuro estudio se propone aumentar diferentes rangos de edad de los participantes, con el fin de acercarse a predicciones más precisas y obtener mayor número de personalidades en los diferentes rasgos. Ya que este estudio está basado en un rango de edad de 20 a 25 años y la mayoría de estos tienen personalidades extrovertidas y abiertos a la experiencia.

Se propone realizar un estudio de las palabras más usadas según la personalidad en el Ecuador, porque el lenguaje usado de cada región es diferente y tienen léxicos propios que se pueden interpretar de diferente manera según donde se encuentre. De esta manera al realizar un nuevo estudio de las palabras, se puede llegar a obtener una mayor precisión en la predicción de la personalidad.

Finalmente, este trabajo fue realizado con el fin para dar soporte a recursos humanos al momento de elegir un nuevo personal. Para trabajos futuros se le propone realizar un estudio que dé como resultado el potencial de liderazgo de una persona con los datos extraídos del Facebook, ya que la mayoría de empresas buscan a potenciales líderes que ayuden al crecimiento de la empresa.

BIBLIOGRAFÍA

- Alexa Top Sites Service. (2017). *Alexa Top Sites Service*. Retrieved from Top Sites in Ecuador: <https://www.alexa.com/topsites/countries/EC>
- Bachrach Y. (2015, Octubre). Human Judgments In Hiring Decisions Based On Online Social Network Profiles. *IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, 1-10. doi:10.1109/DSAA.2015.7344842
- Bachrach, Y. (2015). Human Judgments In Hiring Decisions Based On Online Social Network Profiles. *IEEE*.
- Brady, E. a. (2006). Alarms sound over Facebook time;. *USA Today*, p. 1C.
- Cambridge. (2017). *Apply Magic Sauce*. Retrieved from The Psychometrics Centre: <https://applymagicsauce.com>
- Castells M. (2013). Autocomunicación de masas y movimientos sociales en la era de Internet. *Anuario del Conflicto Social*.
- Chaudhari B., P. M. (2012). A Comparative Study of clustering algorithms Using weka tools. *International Journal of Application or Innovation in Engineering & Management (IJAIEM)*, 154-158.
- Chiavenato, I. (2007). *Administración de los recursos humanos*. México: MC Graw Hill.
- Cooper, K. (2017, 07 08). *Keyword Research, Competitor Analysis, & Website Ranking | Alexa*. Retrieved from <http://www.alexa.com/topsites/countries/EC>
- Correa, T. (2010). Who interacts on the Web?: The intersection of users' personality and social media use. *Computers in Human Behavior*, 247-253.
- Corso C. (2009). Aplicación de algoritmos de clasificación supervisada usando Weka. *Córdoba: Universidad Tecnológica Nacional, Facultad Regional Córdoba*.

- Del Alcazar Juan. (2017, 02 22). *Formación Gerencial*. Retrieved from Ranking Redes Sociales, Sitios Web y Aplicaciones Móviles Ecuador 2017: <http://blog.formaciongerencial.com/ranking-redes-sociales-sitios-web-aplicaciones-moviles-ecuador-2017/>
- Elvira Salgado, E. V.-T. (2016). Use of the Big Five Inventory in a Colombian Sample. *Avances en Psicología Latinoamericana* , 365-382.
- Facebook. (2018). *Información general - API Graph - Documentación*. Retrieved from Facebook para desarrolladores: <https://developers.facebook.com/docs/graph-api/overview>
- García S., Luengo J., Herrera, F. (2015). *Data preprocessing in data mining*. New York: Springer.
- Heroku. (2018). *La plataforma Heroku como servicios y servicios de datos | Heroku*. Retrieved from <https://www.heroku.com/platform>
- Instituto Nacional de Estadística y Censos. (2017). *Instituto Nacional de Estadística y Censos*. Retrieved from Instituto Nacional de Estadística y Censos.: <http://www.ecuadorencifras.gob.ec/institucional/home/>
- Jacob B. Hirsh, J. B. (2009). Personality and language use in self-narratives. *Journal of Research in Personality*, 524-527.
- Kelly Moore, J. C. (2012). The influence of personality on Facebook usage, wall postings, and regret. *Computers in Human Behavior*, 267-274.
- Koch, K. (1957). *El test del árbol*. Buenos Aires: KAPELUSZ.
- Kornblum, J. a. (2006). What you say online could haunt you; Schools, employers scrutinize social websites such as MySpace and Facebook. *USA Today*.
- Letelier, P. (2006, 01 15). *Métodologías ágiles para el desarrollo de software: eXtreme Programming (XP)*. Retrieved from http://www.cyta.com.ar/ta0502/b_v5n2a1.htm

- Luke D Smillie, G. Y. (2006). Benefits of All Work and No Play: The Relationship Between Neuroticism and Performance as a Function of Resource Allocation. *American Psychological Association*, 131-155.
- Mababu Richard. (2016). Reclutamiento a través de las redes sociales: Reclutamiento 3.0. *Opción*, 32.
- Machover, K. (1976). *La Figura Humana: Test Proyectivo de Karen Machover*. Madrid.
- Mahasak Ketcham, T. G. (2016). The Forensic Algorithm On Facebook Using Natural Language Processing. *12th International Conference on Signal-Image Technology & Internet-Based Systems*, 624-627.
- Morate, D. G. (n.d.). *MANUAL DE WEKA*.
- Neal Andrew, Y. G. (2011). Predicting the form and direction of work role performance from the Big 5 model of personality traits. *Wiley Online Library*, 176-192.
- Oliver P. John, R. W. (2008). *Handbook of Personality, Third Edition: Theory and Research*. New York: The Guilford Press.
- Panos Louridas, Christof Ebert. (2016). Machine Learning. *IEEE Software*, 33, 110-115.
- Robert P. Tett, D. D. (2003). A Personality Trait-Based Interactionist Model of Job Performance. *American Psychological Association*, 500-517.
- Rouse Margaret. (2017, Enero). *Red de Techtarget*. Retrieved from Aprendizaje automático (machine learning): <http://searchdatacenter.techtarget.com/es/definicion/Aprendizaje-automatico-machine-learning>
- Samuel D. Gosling, S. G. (2007). Personality Impressions Based on Facebook Profiles. *ICWSM'2007 Boulder, Colorado, USA*.
- Sibel Adah, J. G. (2012). Predicting Personality with Social Behavior. *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*,

302-309.

Stuart C. Shapiro. (1987). *Encyclopedia of artificial intelligence*.

Thoresen, C. J. (2004). The Big Five Personality Traits and Individual Job Performance Growth Trajectories in Maintenance and Transitional Job Stages. *American Psychological Association* , 835-853.

Wehrli, S. (2008). Personality on Social Network Sites: An Application of the Five Factor Model. *Eth zurich sociology working papers, ETH Zurich. Predictors*, 33-54.

Witten et al. (2000). *Data Mining Practical Machine Learning Tools and Techniques* .

Yarkoni, T. (2010). Personality in 100,000 Words: A large-scale analysis of personality and word use among bloggers. *Journal of Research in Personality*, 363-373.