



ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL

Facultad de Ingeniería en Electricidad y Computación

“DESARROLLO DE UNA APLICACIÓN DE MINERÍA DE DATOS QUE PERMITA DETECTAR PREFERENCIAS Y PATRONES EN TEXTOS Y TÓPICOS OBTENIDOS DE TWITTER”

INFORME DE MATERIA INTEGRADORA

Previo a la obtención del título de:

LICENCIADA EN SISTEMAS DE INFORMACIÓN

JESSICA ALEXANDRA NAVARRETE BERMEO

GUAYAQUIL – ECUADOR

AÑO: 2017

AGRADECIMIENTOS

Agradecida con Dios enormemente por quien en vida fue María Guadalupe Bermeo Proaño, mi madre de crianza quien me dio todo su amor como si yo hubiese gestado en su vientre; solo deposito en mi corazón palabras de amor y sabios consejos que me permitieron convertirme en la mujer que soy ahora. Además agradezco todo el apoyo incondicional que me ha brindado mi hermana Roxana Oliveros Bermeo a lo largo de toda mi carrera.

Jessica Alexandra Navarrete Bermeo

DEDICATORIA

A mi añorada madre, María Guadalupe Bermeo Proaño, me dio el más grande y atesorado consejo: todo lo que te propongas hazlo con infinito amor y dedicado siempre a Dios.

Jessica Alexandra Navarrete Bermeo

TRIBUNAL DE EVALUACIÓN

MSc. Ronald Alfredo Barriga Díaz

PROFESOR EVALUADOR

MSc. Ronny Enrique Santana Estrella

PROFESOR EVALUADOR

DECLARACIÓN EXPRESA

"La responsabilidad y la autoría del contenido de este Trabajo de Titulación, me corresponde exclusivamente; y doy mi consentimiento para que la ESPOLE realice la comunicación pública de la obra por cualquier medio con el fin de promover la consulta, difusión y uso público de la producción intelectual"

Jessica Alexandra Navarrete Bermeo

RESUMEN

En la actualidad existe una gran cantidad de información subjetiva en internet y esta crece exponencialmente con el paso del tiempo. A esto se suma la aparición de redes sociales donde miles de usuarios a cada segundo generan contenido, lo que convierte al internet en una valiosa fuente de información.

Desde el punto de vista del procesamiento del lenguaje natural se ha despertado un gran interés en el estudio de opiniones (Minería de Datos) y sentimientos (Análisis de Sentimientos) que las personas hacen públicas en la red.

En este proyecto se propone desarrollar una herramienta que permita extraer dicha información, clasificarla y detectar en ella preferencias y patrones de comportamientos. La información será extraída de Twitter, específicamente de los seguidores de una cuenta registrada y el ingreso de tópicos o más conocidos como hashtag.

ÍNDICE GENERAL

AGRADECIMIENTOS	ii
DEDICATORIA	iii
TRIBUNAL DE EVALUACIÓN	iv
DECLARACIÓN EXPRESA	v
RESUMEN	vi
ÍNDICE GENERAL.....	vii
CAPÍTULO 1.....	1
1 GENERALIDADES	1
1.1 Antecedentes	1
1.2 Objetivo General	2
1.3 Objetivos Específicos	2
1.4 Justificación.....	2
1.5 Soluciones Similares	3
1.6 Descripción del proyecto	5
CAPÍTULO 2.....	6
2 SOLUCIÓN PROPUESTA.....	6
2.1 Metodología utilizada	6
2.1.1 Lista priorizada	6
2.1.2 Historias de Usuarios.....	7
2.1.3 Blackboard SCRUM.....	11
2.2 Desarrollo de la aplicación	12
2.3 Informe de pruebas	13
2.3.1 Métodos para corrección	14
CAPÍTULO 3.....	16
3 SOLUCIÓN TECNOLÓGICA IMPLEMENTADA.....	16
3.1 Instalación e Inicio Aplicación MineTweetEC	16
3.2 Inicio de sesión y autorización a MineTweetEC	16
3.3 Proceso por selección de Cuentas	17
3.4 Proceso por Ingreso de Tópico	18
3.5 Visualización del Grafico de Barras.....	18

3.6 Descarga Reporte PDF	19
CONCLUSIONES Y RECOMENDACIONES	20
BIBLIOGRAFÍA.....	21
ANEXOS.....	22

CAPÍTULO 1

1 GENERALIDADES

Este capítulo se centra en la descripción del proyecto, sus objetivos y el alcance que se espera obtener. Además se exponen trabajos similares con redes sociales y la herramienta que se ha utilizado para la minería de datos.

1.1 Antecedentes

Hoy en día las redes sociales son una fuente valiosa de información; diariamente y sin pausas aglutinan gustos, preferencias y opiniones de usuarios alrededor de todo el mundo. Obtener esta información subjetiva es de gran interés y valor para muchas empresas y para el poder político.

Hace no más de una década la única manera de que las empresas tengan una idea de que opinan los clientes sobre sus productos y servicios era a través de encuestas o analizando las ventas por periodo; pero ahora gracias al internet y las redes sociales esta información está al alcance de todos ya que los usuarios expresan sus opiniones de forma pública e inclusive se dejan influenciar de estas a la hora de tomar decisiones de compra o inclinaciones a un partido político.

La minería de datos es un proceso dentro del campo de la estadística y las ciencias computacionales, que intenta descubrir patrones en grandes volúmenes de conjuntos de datos [1]. Extrae información subjetiva oculta en un texto mediante algoritmos y técnicas de análisis clasificándolo como positivo o negativo. Dicho proceso puede ser aplicado en texto cortos, como los más de 500 millones de Tweets por día que escriben cerca de 100 millones de usuarios activos en Twitter [2].

La disponibilidad de muchas herramientas de minería de datos, como RapidMiner, R, Python, WEKA entre otros; abre una gama de opciones

para ser aplicadas en el desarrollo de aplicaciones usadas para segmentar clientes, para el marketing directo y el marketing interactivo, además del análisis de tendencias.

1.2 Objetivo General

Crear una aplicación de minería de datos que permita detectar preferencias y patrones de comportamiento en texto obtenido de Twitter.

1.3 Objetivos Específicos

- Convertir los mensajes escritos (Tweets) en la red social Twitter en un conjunto de datos que pueda ser clasificado y categorizado.
- Identificar la polaridad (positivo, negativo) del conjunto de datos obtenido (Tweets) en cuentas de la red social Twitter.
- Analizar el conjunto de datos obtenido (Tweets) y construir un modelo de clasificación y agrupación que permita evaluar características relevantes en los resultados como patrones y similitudes.
- Resumir y mostrar la clasificación de los Tweets de modo que de ellos surjan nuevas hipótesis.

1.4 Justificación

La minería de datos en redes sociales es actualmente uno de los campos de investigación más estudiados. Tiene como objetivo analizar la opinión de los usuarios hacia temas de interés como la política, experiencias con productos o servicios, entre otros.

Esta investigación se ha extendido fuera de la informática a las ciencias de gestión y ciencias sociales debido a su importancia para las empresas y la sociedad en su conjunto. La importancia creciente hacia la minería de datos coincide con el aumento en el uso de redes sociales en internet, como por ejemplo Twitter.

Twitter actualmente se ha convertido en una fuente inagotable de información que a diario sus usuarios ingresan. El desarrollo de una aplicación que permita realizar el análisis de opinión sobre temas de interés puede ser muy útil para la comercialización y posicionamiento de productos o servicios, e incluso medir la intención de voto en el ámbito de la política, ya que se expone la inclinación de los usuarios en diferentes dimensiones, tales como tiempo, ubicación geográfica y experiencia.

El seguimiento de cómo las opiniones o discusiones evolucionan con el tiempo puede ayudarnos a identificar tendencias y patrones interesantes y comprender mejor las formas en que la información se propaga en Internet.

1.5 Soluciones Similares

Data Mining.Remote

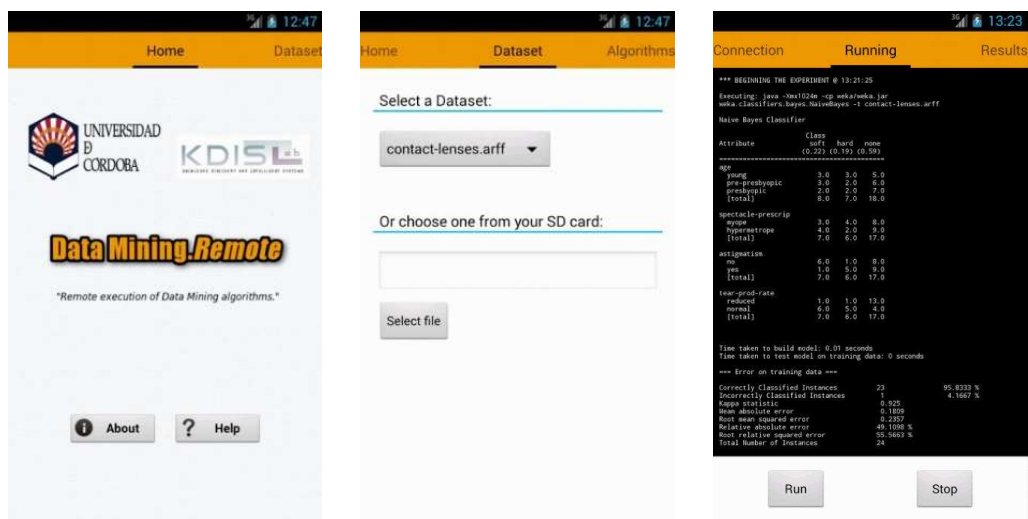


Figura 1.1 Capturas de Pantallas Data Mining.Remote.

Esta aplicación hace una ejecución remota de algoritmos de Minería de Datos. Está basada en los algoritmos de la aplicación WEKA. El usuario selecciona un conjunto de datos ya clasificado o sube un archivo

compatible con la herramienta WEKA (.arff) y muestra los resultados en ejecución y permite la descarga del mismo [3], ver Figura 1.1. Cuenta con estadísticas aceptables con respecto a las opiniones de sus usuarios [4], ver Figura 1.2.

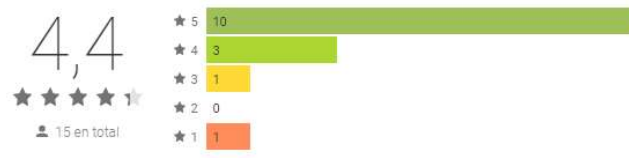


Figura 1.2 Estadística Opiniones Data Mining.Remote.

Knowledge Analysis Tool

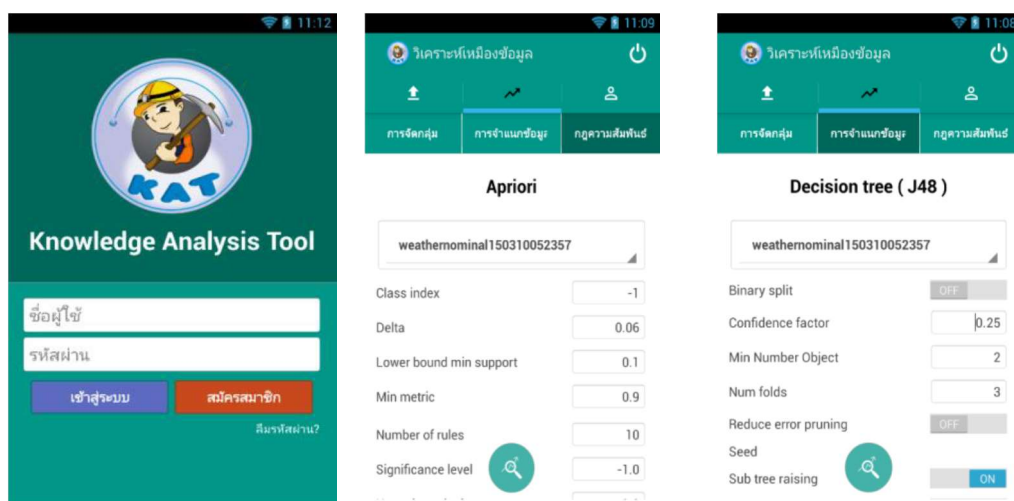


Figura 1.3 Capturas de Pantallas Knowledge Analysis Tool.

Esta aplicación fue desarrollada por el Departamento de Matemáticas Estadísticas e Informática de la Facultad de Ciencias en la Universidad Ubon Ratchathani, Tailandia. Tiene como objetivo determinar una herramienta de minería que alcance un nivel aceptable de rendimiento. El resultado fue una aplicación donde el usuario envía un conjunto de datos y selecciona la técnica de minería de datos, ver Figura 1.3. Los conjuntos de datos (Atributo-Relación Archivo Formato ARFF o Comma-Separated Values CSV) se pueden cargar a través del teléfono

inteligente en el servidor para posteriormente ser procesados con WEKA [5].

1.6 Descripción del proyecto

La aplicación móvil funcionará en dispositivos con sistema operativo Android, la misma que le permitirá al usuario:

- Iniciar sesión con su cuenta de Twitter.
- Acceder a una lista de las cuentas que sigue en Twitter entre otros datos de su perfil.
- Seleccionar de que cuentas de Twitter desea obtener los Tweets para el proceso de minería de datos.
- Ingresar un texto o tópico para la recolección de Tweets a ser clasificados.
- Visualizar el resultado del proceso de minería de datos aplicado a los Tweets que se hayan recolectado según las cuentas seleccionadas o el texto/tópico ingresado.

CAPÍTULO 2

2 SOLUCIÓN PROPUESTA

Después de plantear, justificar y contrastar lo que se desea alcanzar; se detalla a continuación todas las actividades para llevar a cabo el cometido de este proyecto.

2.1 Metodología utilizada

La metodología que se usará es SCRUM ya que aplica de manera regular un conjunto de buenas prácticas para trabajar colaborativamente, en equipo, y obtener el mejor resultado posible de un proyecto [6].

2.1.1 Lista priorizada

La lista priorizada es la planificación de las actividades para completar los objetivos/requisitos seleccionados para la iteración [6], ver Tabla 1.

ID	Actividad	Usuario	Descripción	Prioridad	Estimación	Sprint
1	Inicio sesión con Twitter	Usuario	El usuario podrá registrarse en la aplicación mediante su cuenta de Twitter.	10	7 días	1
2	Menú Flotante	Usuario	El usuario podrá visualizar un menú flotante con las opciones de la aplicación.	8	3 días	1
3	Lista de Seguidores	Usuario	El usuario, desde su perfil obtenido de su cuenta enlazada, podrá visualizar una lista de sus seguidores.	10	6 días	2
4	Selección de Cuentas	Usuario	El usuario podrá elegir de qué cuentas desea recolectar Tweets para poder analizar la información.	10	4 días	3
5	Ingreso de Tópicos	Usuario	El usuario podrá ingresar tópicos para recolectar Tweets y analizar la información contenida.	10	3 días	3
6	Proceso Minería	-	Internamente, luego de seleccionar las cuentas de seguidores del usuario que inicio sesión o de ingresar un tópico; el usuario podrá ejecutar el proceso de minería con los Tweets recolectados.	10	7 días	4
7	Mostrar Resumen Tweets recolectados	Usuario	El usuario podrá ver datos nominales y numéricos de los Tweets recolectados en base a las cuentas o tópico ingresado.	8	3 días	4

8	Mostrar Resultado Proceso Minería	Usuario	El usuario podrá ver gráficos de barra por tópico o texto digitado de la polaridad de los Tweets recolectados (Positivo, Negativo, Ninguno y Neutro).	8	3 días	5
9	Descarga de Reportes	Usuario	El usuario podrá descargar los datos de una lista variada de reportes prediseñados.	10	4 días	5

Tabla 1: Lista Priorizada

2.1.2 Historias de Usuarios

A continuación, ver desde la Tabla 2 hasta la Tabla 13, el detalle de las funcionalidades que deben ser incorporadas y validadas en la aplicación mediante las historias de usuarios:

Código	001				
Nombre	Inicio Sesión con Twitter				
Actor	Usuario				
Descripción	Como usuario debo iniciar sesión con mi cuenta de Twitter ingresando usuario y contraseña.				
H.U. Relacionada(s)	Código		Nombre		
Modulo					
Criterios de Aceptación	Condición			Resultado	
	El usuario debe contar con una cuenta de Twitter.			El usuario inicia sesión con sus credenciales de Twitter.	

Tabla 2: HU 001 – Inicio Sesión con Twitter

Código	002				
Nombre	Autorizar a MineTweetEC				
Actor	Usuario				
Descripción	Como usuario debo autorizar a MineTweetEC para que acceda a la información de mi cuenta de Twitter.				
H.U. Relacionada(s)	Código	001	Nombre	Inicio Sesión con Twitter	
Modulo					
Criterios de Aceptación	Condición			Resultado	
	El usuario debe iniciar sesión con su cuenta de Twitter.			Una vez que inicio sesión con su cuenta de Twitter puede autorizar a MineTweetEC para acceder a los datos de su cuenta.	

Tabla 3: HU 002 – Autorizar a MineTweetEC

Código	003		
Nombre	Perfil del Usuario		
Actor	Usuario		
Descripción	Como usuario quiero ver el nombre de mi cuenta de Twitter.		
H.U. Relacionada(s)	Código	001	Nombre
		002	
			Inicio Sesión con Twitter
			Autorizar a MineTweetEC
Modulo			
Criterios de Aceptación	Condición		Resultado
	<p>El usuario debe iniciar sesión con su cuenta de Twitter.</p> <p>El usuario debe autorizar a MineTweetEC para acceder a los datos de su cuenta de Twitter.</p>		Una vez que haya iniciado sesión con sus credenciales de Twitter y autorizado a MineTweetEC para acceder a sus datos la aplicación puede mostrar algunos de los datos de su perfil como el nombre de usuario.

Tabla 4: HU 003 – Perfil del Usuario

Código	004		
Nombre	Menú Flotante		
Actor	Usuario		
Descripción	Como usuario quiero tener acceso a las opciones de la aplicación por medio de un menú flotante.		
H.U. Relacionada(s)	Código	001	Nombre
		002	
			Inicio Sesión con Twitter
			Autorizar a MineTweetEC
Modulo			
Criterios de Aceptación	Condición		Resultado
	<p>El usuario debe iniciar sesión con su cuenta de Twitter.</p> <p>El usuario debe autorizar a MineTweetEC para acceder a los datos de su cuenta de Twitter.</p>		Una vez que haya iniciado sesión con sus credenciales de Twitter y autorizado a MineTweetEC para acceder a sus datos, podrá visualizar el menú flotante con las opciones disponibles.

Tabla 5: HU 004 – Menú Flotante

Código	005		
Nombre	Listar Seguidores de Twitter		
Actor	Usuario		
Descripción	Como usuario quiero visualizar una lista de los seguidores de mi cuenta de Twitter.		
H.U. Relacionada(s)	Código	001	Nombre
		002	
			Inicio Sesión con Twitter
			Autorizar a MineTweetEC
Modulo			
Criterios de Aceptación	Condición		Resultado
	El usuario debe iniciar		Una vez que haya iniciado

	sesión con su cuenta de Twitter. El usuario debe autorizar a MineTweetEC para acceder a los datos de su cuenta.	sesión con sus credenciales de Twitter y autorizado a MineTweetEC para acceder a sus datos, se carga una lista con los seguidores del usuario.
--	--	--

Tabla 6: HU 005 – Listar Seguidores de Twitter

Código	006		
Nombre	Tipo de Recolección de Tweets		
Actor	Usuario		
Descripción	Como usuario quiero seleccionar el tipo de recolección de Tweets a ser procesados.		
H.U. Relacionada(s)	Código	005	Nombre Listar Seguidores de Twitter
Modulo			
Criterios de Aceptación	Condición		Resultado
	El usuario debe poder visualizar la lista de seguidores.		Una vez que el usuario pueda acceder a la pantalla de lista de seguidores podrá visualizar los tipos de recolección de Tweets: Seguidores/Tópico.

Tabla 7: HU 006 – Tipo de Recolección de Tweets

Código	007		
Nombre	Seleccionar Cuentas de Twitter		
Actor	Usuario		
Descripción	Como usuario quiero seleccionar uno o más seguidores de mi cuenta de Twitter para recolectar Tweets.		
H.U. Relacionada(s)	Código	005	Nombre Listar Seguidores de Twitter
Modulo			
Criterios de Aceptación	Condición		Resultado
	El usuario debe poder visualizar la lista de seguidores.		Una vez que el usuario tenga acceso a la lista de seguidores podrá seleccionar una o varias cuentas de Twitter para recolectar los Tweets que serán purgados de contenido innecesario previo al proceso de minería de datos.

Tabla 8: HU 007 – Seleccionar Cuentas de Twitter

Código	008
Nombre	Ingreso de Tópico
Actor	Usuario
Descripción	Como usuario quiero ingresar un texto como tópico para la

	recolección de Tweets.			
H.U. Relacionada(s)	Código	005	Nombre	Listar Seguidores de Twitter
Modulo				
Criterios de Aceptación	Condición		Resultado	
	El usuario debe poder visualizar la lista de seguidores. El usuario debe presionar el botón que le permite cambiar el tipo de recolección de Tweets.		Una vez que el usuario tenga acceso a la lista de seguidores podrá visualizar el botón para cambiar el tipo de recolección de Tweets a tópico, donde ingresara un texto para el proceso de minería.	

Tabla 9: HU 008 – Ingreso de Tópico

Código	009			
Nombre	Ejecutar Proceso de Minería			
Actor	Usuario			
Descripción	Como usuario debo ejecutar el proceso de minería.			
H.U. Relacionada(s)	Código	007	Nombre	Seleccionar Cuentas de Twitter
		008		Ingreso de Tópico
Modulo				
Criterios de Aceptación	Condición		Resultado	
	El usuario debe seleccionar uno o varios seguidores de la lista o ingresar un tópico para la recolección de Tweets.		Una vez seleccionado el tipo de recolección de Tweets (Selección de seguidores o ingreso de tópico), el usuario puede enviar a ejecutar proceso de minería de datos.	

Tabla 10: HU 009 – Ejecutar Proceso de Minería

Código	010			
Nombre	Mostrar Resumen Tweets Recolectados			
Actor	Usuario			
Descripción	Como usuario quiero visualizar un resumen de los Tweets recolectados.			
H.U. Relacionada(s)	Código	009	Nombre	Ejecutar Proceso de Minería
Modulo				
Criterios de Aceptación	Condición		Resultado	
	El usuario debe ejecutar el proceso de minería de datos.		Una vez que el usuario ejecute el proceso de minería, al finalizar este; muestra un resumen de los Tweets recolectados ya listos para ser presentados en el reporte.	

Tabla 11: HU 010 – Mostrar Resumen Tweets Recolectados

Código	011		
Nombre	Descargar Reporte		
Actor	Usuario		
Descripción	Como usuario quiero poder descargar un archivo PDF con el reporte.		
H.U. Relacionada(s)	Código	009	Nombre
Modulo	Ejecutar Proceso de Minería		
Criterios de Aceptación	Condición		Resultado
	El usuario debe ejecutar el proceso de minería de datos.		Una vez que el usuario ejecute el proceso de minería, al finalizar este; el usuario podrá visualizar las opciones de reportes para seleccionar.

Tabla 12: HU 012 – Descargar Reporte

Código	012		
Nombre	Reporte		
Actor	Usuario		
Descripción	Como usuario quiero poder visualizar los datos detallados de Tweets en un archivo de PDF.		
H.U. Relacionada(s)	Código	009	Nombre
Modulo	Ejecutar Proceso de Minería		
Criterios de Aceptación	Condición		Resultado
	El usuario debe ejecutar el proceso de minería de datos.		Una vez que el usuario ejecute el proceso de minería y haya seleccionado el tipo de reporte podrá descargar el mismo en su dispositivo móvil.

Tabla 13: HU 012 – Reporte.

2.1.3 Blackboard SCRUM

Para gestionar la lista de objetivos a completar en la interacción, se utilizó el tablero de tareas que ayudo a promover la comunicación y transparencia en el desarrollo del proyecto, ver Figura 2.1.

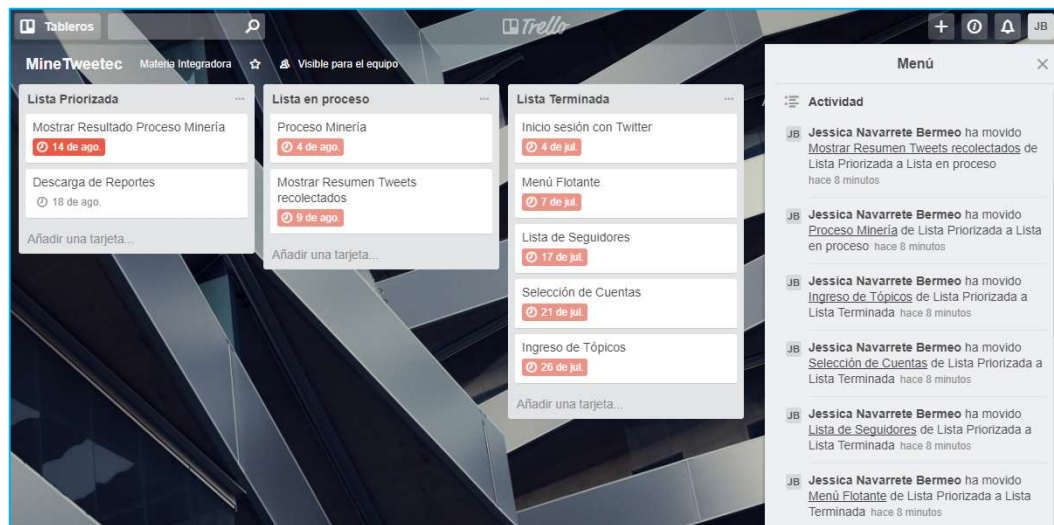


Figura 2.1 Blackboard SCRUM

2.2 Desarrollo de la aplicación

La aplicación ha sido desarrollada bajo una plataforma móvil, ver Tabla 14. Para tener un detalle del acceso al servicio de Twitter, además de la creación de la solución, ver en Anexos literal A.

Hardware	
Procesador	Intel(R) Core(TM) i3-3227U CPU 1.90GHz 64bits
Memoria RAM	5.90 GB
Disco Duro	500 GB
Software	
Sistema Operativo	Windows 10 pro
IDE	Android Studio 2.3.3
SDK	Android API 16-4.1-Jelly Bean
Librerías Referenciadas	<ul style="list-style-type: none"> ▪ SDK Twitter 2.2 ▪ wekaSTRIPPED.jar ▪ google GSON 2.7 ▪ retrofit2 2.1.0 ▪ retrofit2 GSON 2.1.0
Herramienta de Minería	WEKA 3.8 - Waikato Environment for Knowledge Analysis.
Base de Datos	SQLite 3.7.11

Tabla 14: Plataforma de Desarrollo

2.3 Informe de pruebas

Se definió una matriz de pruebas internas en la que se validó el correcto funcionamiento de cada una de las pantallas de la aplicación MineTweetEC. Para tener un detalle de las capturas de pantalla según las historias de usuario, ver en Anexos literal B.

A continuación se detalla dicha matriz de pruebas con su respectivo resultado, ver Tabla 15:

Pantalla	Prueba	Descripción	Tiempo Respuesta (en segundos)	Resultado
Inicio de Sesión	Inicio de Sesión con una cuenta activa de Twitter.	Se validó que un usuario pueda iniciar sesión con una cuenta activa de Twitter.	2	Exitoso
	Autorización a MineTweetEC.	Se verificó que se autorice a la aplicación MineTweetEC para acceder a los datos del usuario.	El tiempo varía de 2 a 3 segundos según la conexión a internet.	
Selección de Cuentas	Lista de seguidores.	Se verificó que la lista de seguidores de quien inicia sesión se cargue correctamente	0 Carga automáticamente luego del inicio de sesión.	Exitoso
	Proceso de Minería de Datos.	Se validó que el proceso de recolección y minería de los Tweets se realice correctamente.	2 El tiempo de respuesta varía según la conexión a internet, la capacidad de procesamiento y del volumen de información a procesar.	Exitoso
Gráfico de Barras	Resumen gráfico de barras.	Se verificó que se muestre correctamente en el gráfico barras, con el resumen de la minería de datos que clasifica la polaridad de los Tweets recolectados.	0 El gráfico se genera automáticamente una vez terminado el proceso de minería de Tweets.	Exitoso
	Descargar Data	Se validó que se descargue el archivo PDF con los datos detallados de la polaridad de los Tweets recolectados.	0 Depende del volumen de la información que fue recolectada.	
Proceso por Tópico	Ingreso de Tópico	Se verificó que el usuario pueda digitar cualquier texto como tópico de referencia para la recolección de Tweets a procesar.	0 El tiempo que tome depende del usuario.	Exitoso
	Proceso de Minería de Datos.	Se validó que el proceso de recolección y minería de los Tweets se realice correctamente.	2-3 Depende de la conexión a internet, la capacidad de procesamiento y del	

			volumen de información a procesar.	
Perfil Twitter	Perfil de Twitter.	Se verificó que el usuario pueda visualizar sus datos en la pantalla.	0 Los datos se cargan automáticamente luego del inicio de sesión.	Exitoso
	Publicar Tweet.	Se validó que el usuario pueda publicar un tweet desde la aplicación MineTweetEC.	1 El tiempo varía de 2 a 3 segundos según la conexión a internet del dispositivo.	Bueno No verificado

Tabla 15: Matriz de prueba.

2.3.1 Métodos para corrección

Para cada Sprint se realizó una reunión con el líder del proyecto para revisar el porcentaje de avance según actividades definidas; en dichas reuniones se dio seguimiento al desarrollo del proyecto, validando el esfuerzo y estado de los mismos.

En las siguientes tablas se detallan las pilas de los Sprint que sirvieron como método de corrección a lo largo del desarrollo de la aplicación, ver desde la Tabla 16 hasta la Tabla 20.

SPRINT 1				Avance de tareas:		15%
Inicio:	19-Jun-2017	Fin:	30-Jun-2017	Duración:	30 horas	
PILAS DEL SPRINT						
ID	Actividad	Tipo	Estado	Responsable	Esfuerzo	
1	Inicio sesión con Twitter	Desarrollo	Entregado	Jessica Navarrete	8	
2	Menú Flotante	Desarrollo	Entregado	Jessica Navarrete	8	

Tabla 16: Métodos Corrección – SPRINT 1.

SPRINT 2				Avance de tareas:		15%
Inicio:	03-Jul-2017	Fin:	10-Jul-2017	Duración:	18 horas	
PILAS DEL SPRINT						
ID	Actividad	Tipo	Estado	Responsable	Esfuerzo	
3	Lista de Seguidores	Desarrollo	Entregado	Jessica Navarrete	7	

Tabla 17: Métodos Corrección – SPRINT 2.

SPRINT 3				Avance de tareas:	
Inicio:	11-Jul-2017	Fin:	24-Jul-2017	Duración:	21 horas
PILAS DEL SPRINT					
ID	Actividad	Tipo	Estado	Responsable	Esfuerzo
4	Selección de Cuentas	Desarrollo	Entregado	Jessica Navarrete	10
5	Ingreso de Tópicos	Desarrollo	Entregado	Jessica Navarrete	8

Tabla 18: Métodos Corrección – SPRINT 3.

SPRINT 4				Avance de tareas:	
Inicio:	25-Jul-2017	Fin:	07-Ago-2017	Duración:	30 horas
PILAS DEL SPRINT					
ID	Actividad	Tipo	Estado	Responsable	Esfuerzo
6	Proceso de Minería	Desarrollo	Entregado	Jessica Navarrete	10
7	Mostrar Resumen Tweets recolectados.	Desarrollo	Entregado	Jessica Navarrete	7

Tabla 19: Métodos Corrección – SPRINT 4.

SPRINT 5				Avance de tareas:	
Inicio:	08-Ago-2017	Fin:	21-Ago-2017	Duración:	21 horas
PILAS DEL SPRINT					
ID	Actividad	Tipo	Estado	Responsable	Esfuerzo
8	Mostrar Resultado Proceso Minería	Desarrollo	Entregado	Jessica Navarrete	6
9	Descarga de Reportes	Desarrollo	Entregado	Jessica Navarrete	7

Tabla 20: Métodos Corrección – SPRINT 5.

CAPÍTULO 3

3 SOLUCIÓN TECNOLÓGICA IMPLEMENTADA

Finalmente, en este capítulo se describe la implementación de la solución en un dispositivo móvil que tenga las características requeridas para la plataforma bajo la cual ha sido desarrollada.

3.1 Instalación e Inicio Aplicación MineTweetEC

Una vez instalada la aplicación MineTweetEC el usuario podrá acceder con su cuenta de Twitter ingresando su correo y contraseña luego de presionar el botón “*Inicia sesión con Twitter*”, ver Figura 3.1.

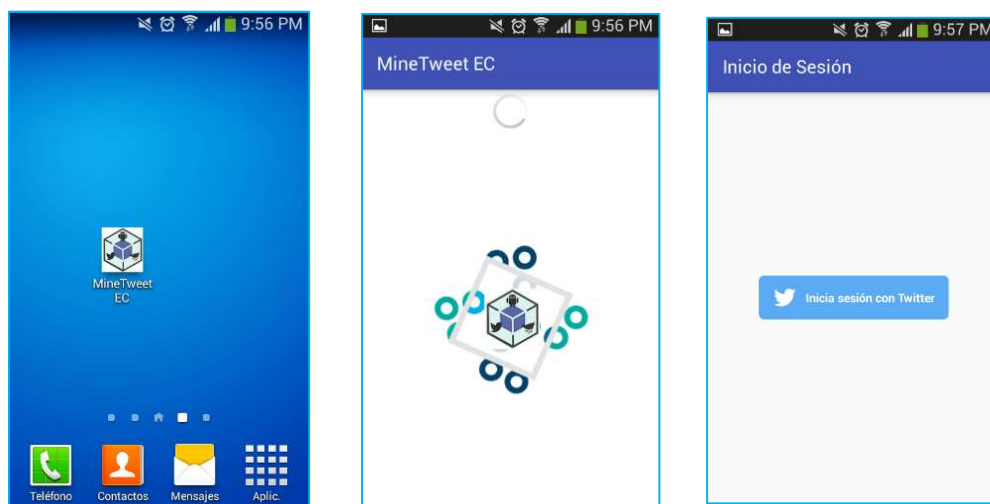


Figura 3.1. Instalación e Inicio Aplicación MineTweetEC.

3.2 Inicio de sesión y autorización a MineTweetEC

Para iniciar sesión el usuario debe digitar su correo y contraseña de su cuenta de Twitter y luego presionar el botón “*Autorizar la aplicación*”. Se muestra una pantalla informativa para re direccionarte a MineTweetEC, ver Figura 3.2.

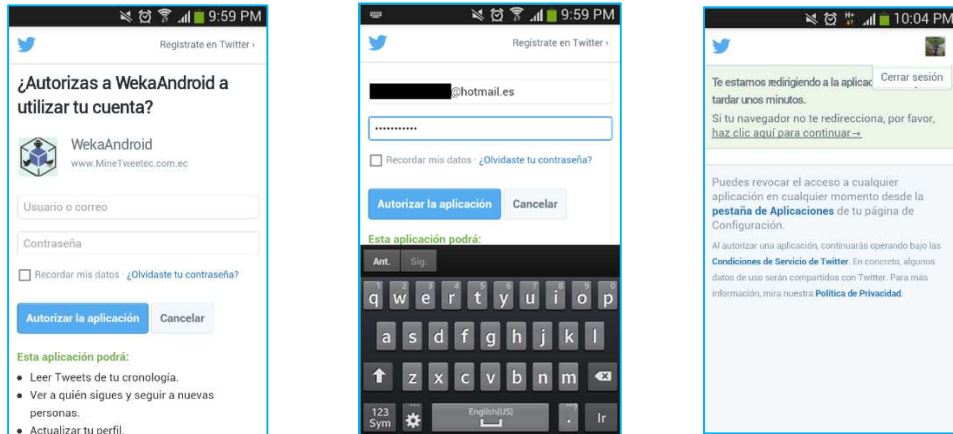


Figura 3.2. Inicio de sesión y autorización a MineTweetEC.

3.3 Proceso por selección de Cuentas

Una vez iniciado sesión y que el usuario haya autorizado a MineTweetEC, podrá visualizar una lista de sus seguidores. El usuario podrá seleccionar una o varias cuentas para proceder al proceso de recolección y minería de Tweets presionando el botón “*MINETWEETS AHORA*”. Luego de este evento se muestra una lista de los Tweets recolectados, mismos que fueron analizados y clasificados según su polaridad; es decir si el Tweets es positivo, negativo o neutro, ver Figura 3.3.

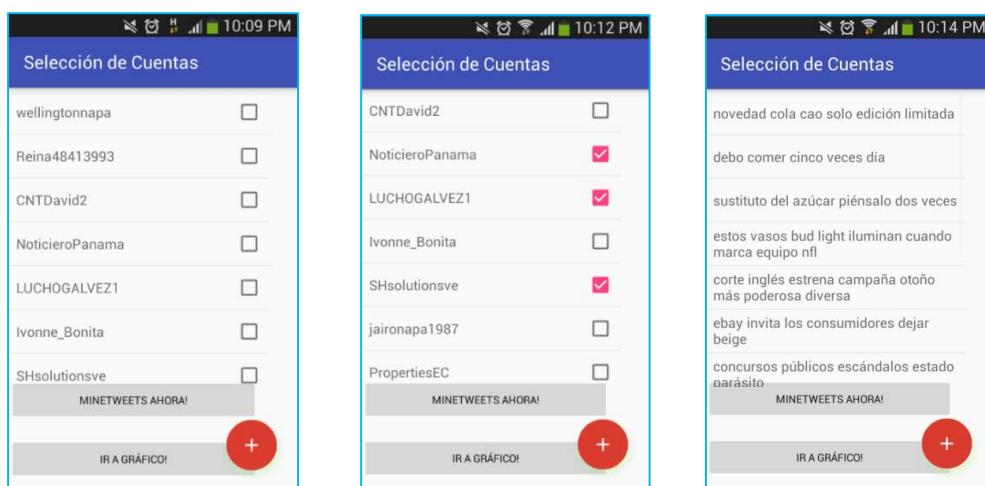


Figura 3.3. Proceso por selección de Cuentas.

3.4 Proceso por Ingreso de Tópico

El usuario también podrá ejecutar el proceso de recolección y minería por medio del ingreso de Tópicos o más conocidos como hashtags.

Una vez que el usuario digite el texto puede presionar el botón “MINETWEETS AHORA” y automáticamente se carga una lista de los Tweets recolectados, mismos que fueron analizados y clasificados según su polaridad; es decir si el Tweets es positivo, negativo o neutro, ver Figura 3.4.

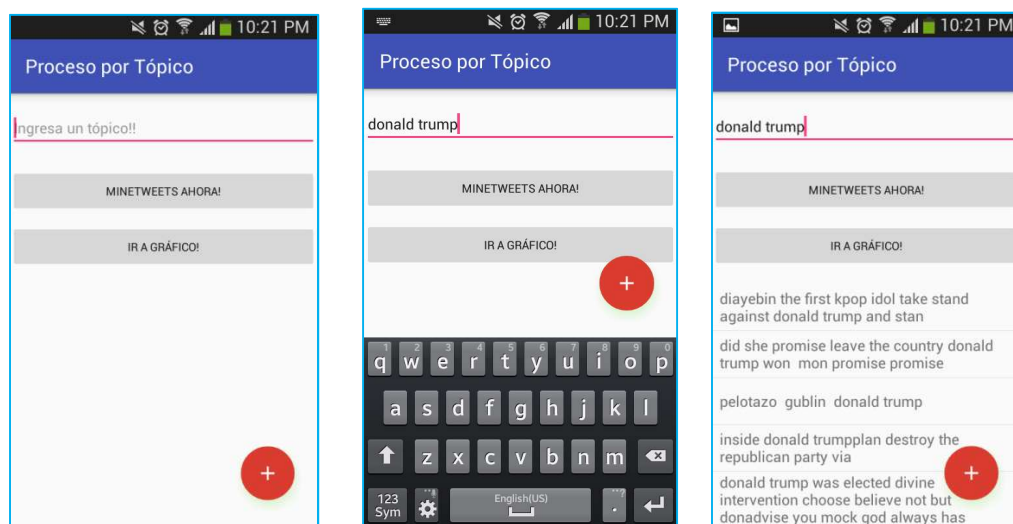


Figura 3.4. Proceso por Ingreso de Tópico.

3.5 Visualización del Grafico de Barras

Luego de ejecutar el proceso de minería por medio de cualquiera de los dos tipos de recolección de Tweets, el usuario podrá visualizar un gráfico de barras que le muestra el resumen de la clasificación de polaridad resultante de la minería aplicada, ver Figura 3.5.

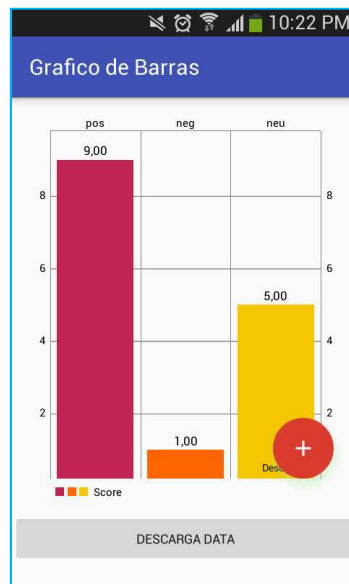


Figura 3.5. Visualización del Gráfico de Barras.

3.6 Descarga Reporte PDF

Finalmente el usuario puede descargar un archivo PDF con el detalle de los Tweets recolectados y clasificados con su respectiva polaridad asignada por el proceso de minería, ver Figura 3.6.

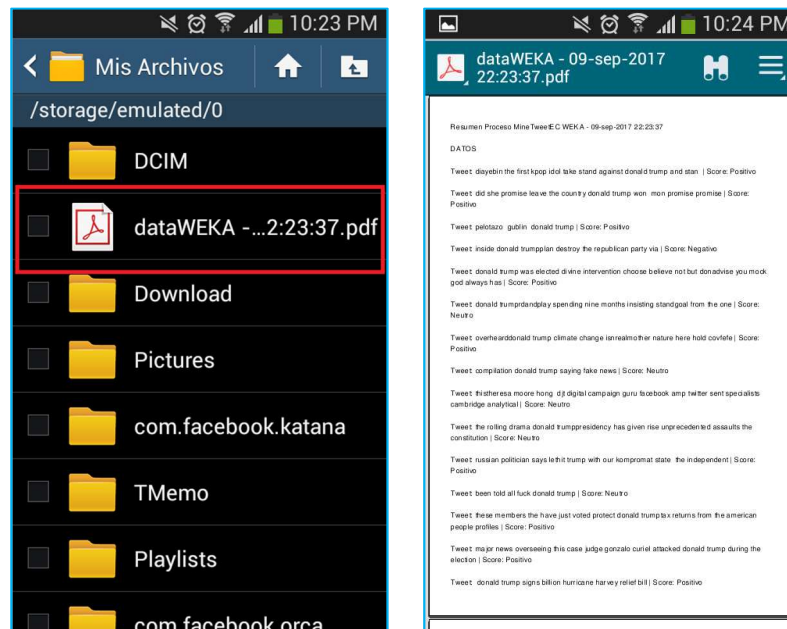


Figura 3.6. Descarga Reporte PDF.

CONCLUSIONES Y RECOMENDACIONES

El objetivo de este proyecto es desarrollar una aplicación de minería de datos capaz de analizar y clasificar la información de modo que permita detectar preferencias y patrones de comportamiento en los tweets recolectados de cuentas o tópicos ingresados por una persona. Este objetivo se ha cumplido según las condiciones que se fueron dando a medida que se fue desarrollando el proyecto ya que el servicio de Twitter para obtener la información tiene limitantes.

Se han detectado errores cuando se valida la información en el reporte en cuanto a la asignación de polaridad de cada Tweets pero esto es debido a varios factores como el idioma y otros elementos en el Tweet que hacen imposible detectar si este es negativo o positivo y termina siendo clasificado como neutro. Pero en su mayoría los Tweets son clasificados de forma correcta logrando así un resultado bastante aceptable con respecto al fin deseado.

Este proyecto fue pensado y desarrollado para aprovechar el aumento de procesamiento y capacidad en los dispositivos móviles que se lanzan al mercado; de modo que la minería de datos sea accesible para todo tipo de persona, pequeñas o grandes empresas y por qué no, para corporativos y la política.

En conclusión el trabajo realizado solo es el inicio del camino hacia una meta a largo plazo para desarrollar una aplicación con todas las bondades que ofrece la herramienta WEKA y poder acercarse más a una verdadera experiencia en predicción para la toma de decisiones. Este es el primer paso para futuros proyectos de Minería de Datos desarrollados en plataformas para móviles.

BIBLIOGRAFÍA

- [1] O. Maimon and L. Rokach, "Data Mining and Knowledge Discovery Handbook", Springer, New York, 2010.
- [2] "Twitter by the Numbers (2017): Stats, Demographics & Fun Facts", Omnicoreagency.com, 2017. [Online]. Disponible en:
<https://www.omnicoreagency.com/twitter-statistics/>.
- [3] M. Pérez, "Data Mining.Remote", AppCrawlr, 2014. [Online]. Disponible en:
<http://appcrawlr.com/android/data-mining-remote#authors-description>.
- [4] "Data Mining Remote", En Google Play, 2014. [Online]. Disponible en:
https://play.google.com/store/apps/details?id=com.dataminingws_android.
- [5] Ubon Ratchathani University, "Development of a Data Mining Tool on Android Smartphones", Ubon Ratchathani, Tailandia, 2015.
- [6] "SCRUM", Proyectos Ágiles, Qué es SCRUM. [Online]. Disponible en:
<https://proyectosagiles.org>.

ANEXOS

A. INSTALACIÓN DE LA PLATAFORMA DE DESARROLLO

Para el desarrollo se instaló el IDE Android Studio en su versión 2.3.3, el SDK utilizado es Jelly Bean (Android API 16) para que la aplicación MineTweetEC pueda instalarse en dispositivos móviles Android a partir de la versión 4.1.

Se requiere la creación de una Aplicación de Twitter, para ello es necesario contar con una cuenta activa de Twitter.

Creación App Twitter:

- 1) Acceder al link <https://apps.twitter.com> para crear una nueva Aplicación de Twitter, ver Figura A.1.

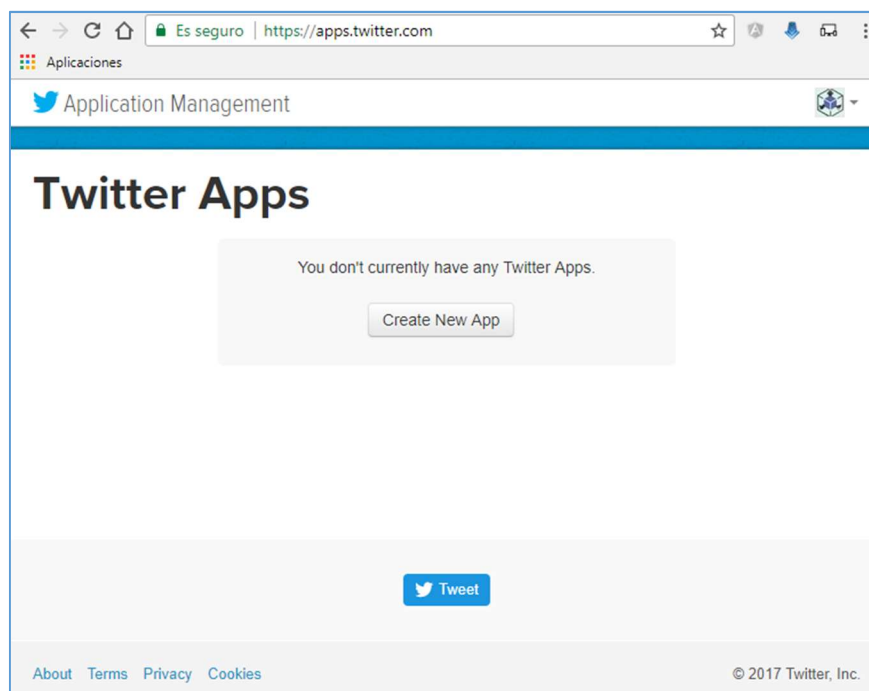


Figura A.1. Nueva Aplicación de Twitter

- 2) Ingresar los datos pertinentes para identificar la aplicación, ver Figura A.2.

← → ↻ 🏠 Es seguro | https://apps.twitter.com/app/new ☆ 🔒 📄

Aplicaciones

Create an application

Application Details

Name *

WekaAndroid

Your application name. This is used to attribute the source of a tweet and in user-facing authorization screens. 32 characters max.

Description *

Minería de Datos Twitter con WEKA desde Android

Your application description, which will be shown in user-facing authorization screens. Between 10 and 200 characters max.

Website *

http://www.MineTweetec.com.ec

Your application's publicly accessible home page, where users can go to download, make use of, or find out more information about your application. This URL will be used for source attribution for tweets created by your application and will be shown in user-facing authorization screens. (If you don't have a URL yet, just put a placeholder here but remember to change it later.)

Figura A.2. Ingreso Datos Aplicación de Twitter.

- 3) Una vez creada la aplicación tendrá acceso a las claves y a los Tokens de acceso, ver Figura A.3.

← → ↻ 🏠 Es seguro | https://apps.twitter.com/app/[redacted] ☆ 🔒 📄

Aplicaciones


Application Management

Your application has been created. Please take a moment to review and adjust your application's settings.

WekaAndroid

Test OAuth

Details Settings Keys and Access Tokens Permissions

 Minería de Datos Twitter con WEKA desde Android
http://www.MineTweetec.com.ec

Organization

Information about the organization or company associated with your application. This information is optional.

Organization None

Organization website None

Figura A.3. Acceso a claves y Tokens de acceso.

- 4) Se requiere la generación de las claves para poder consumir el API de Twitter, ver Figura A.4.

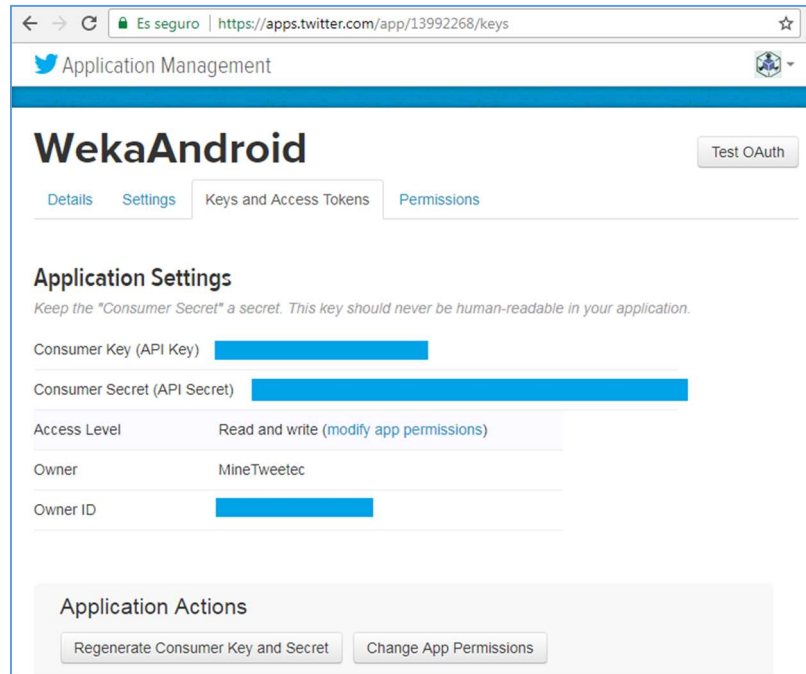


Figura A.4. Claves consumo API Twitter

Creación Proyecto Android Studio:

- 1) Creación de un nuevo proyecto Android Studio, ver Figura A.5.

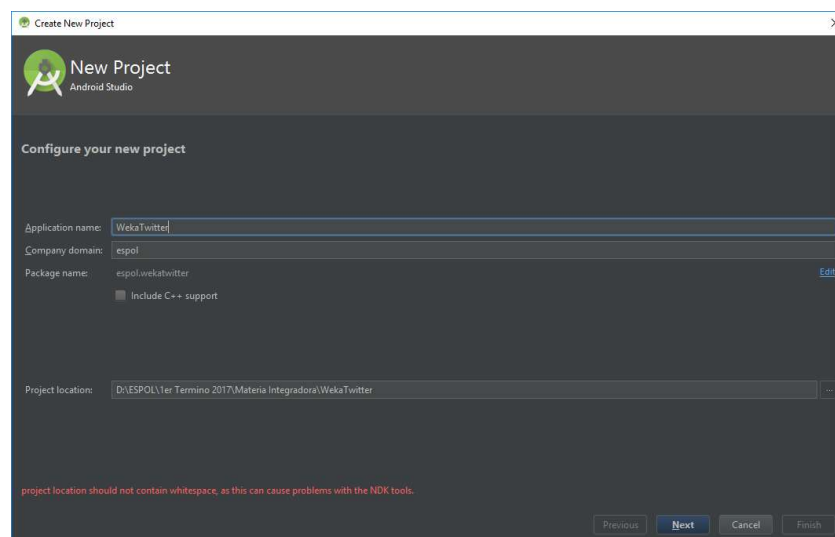


Figura A.5. Proyecto Android Studio.

2) Lista de SDK disponibles para crear el proyecto, ver Figura A.6.

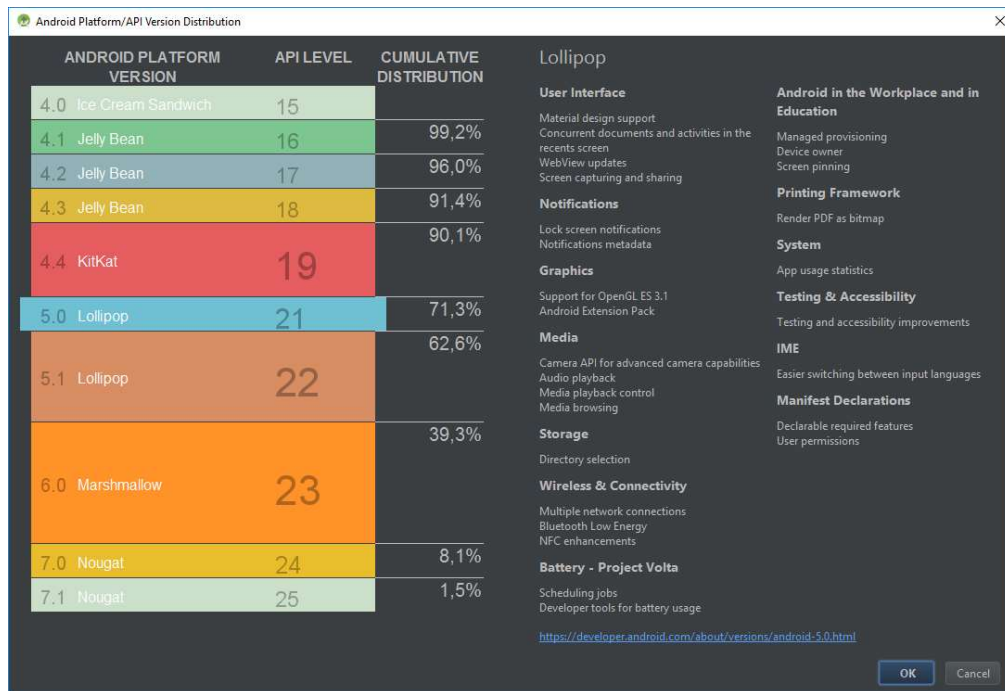


Figura A.6. SDK Disponibles.

3) Selección del SDK mínimo para correr MineTweetEC, ver Figura A.7.

Create New Project

Target Android Devices

Select the form factors your app will run on

Different platforms may require separate SDKs

Phone and Tablet

Minimum SDK: [Dropdown menu]

Lower API levels target more devices, but have fewer features available.
By targeting API 21 and later, your app will run on approximately 71.3% of the devices that are active on the Google Play Store.
[Help me choose](#)

Wear

Minimum SDK: [Dropdown menu]

TV

Minimum SDK: [Dropdown menu]

Android Auto

Previous Next Cancel Finish

Figura A.7. SDK Mínimo para MineTweetEC

- 4) Para empezar con el desarrollo se debe crear una actividad vacía, ver Figura A.8.

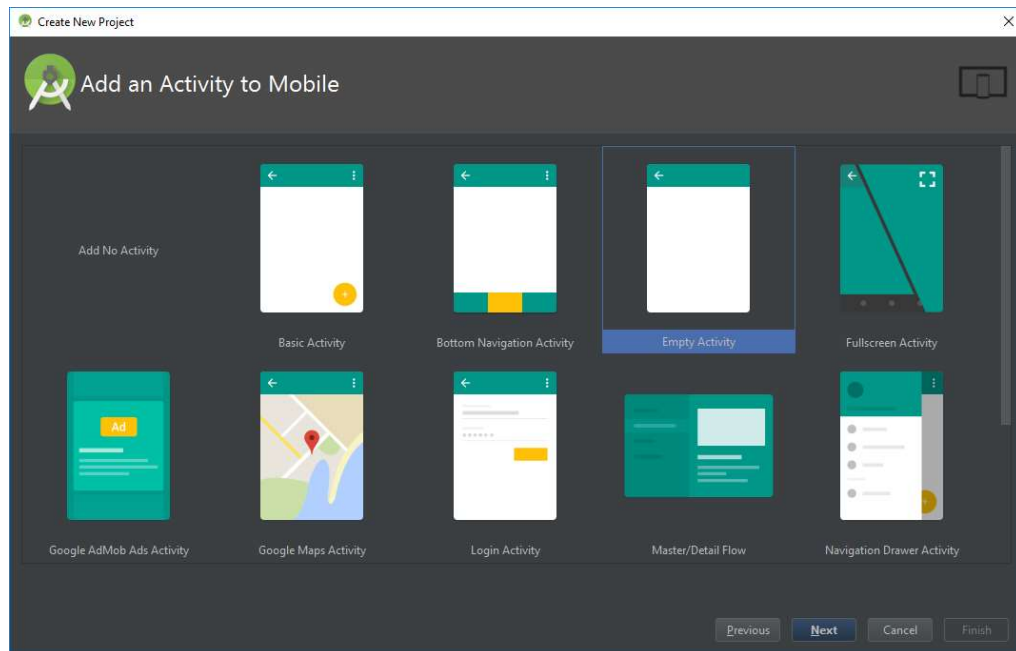


Figura A.8. Creación actividad

B. CRITERIO DE ACEPTACIÓN

En esta sección se demuestra la efectividad de los criterios de aceptación de cada una de las historias de usuarios con la matriz de las pantallas capturadas, ver desde la Tabla 21 hasta la Tabla 32.



Tabla 21: Pantalla Inicio Sesión con Twitter



Tabla 22: Pantalla Autorizar a MineTweetEC

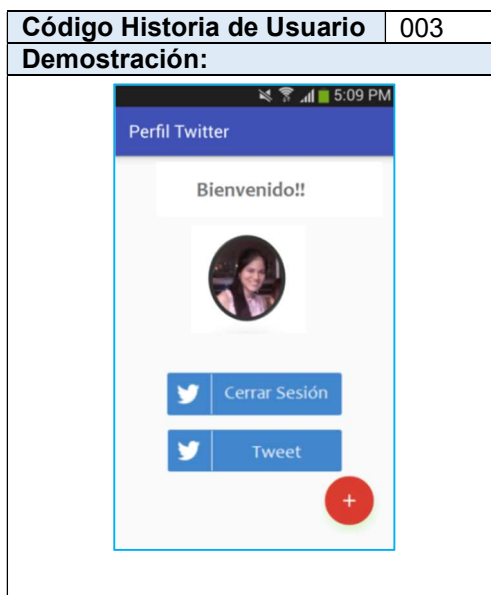


Tabla 23: Pantalla Perfil del Usuario



Tabla 24: Pantalla Menú Flotante

Código Historia de Usuario	005
Demostración:	

Tabla 25: Pantalla Listar Seguidores de Twitter

Código Historia de Usuario	006
Demostración:	

Tabla 26: Pantalla Tipo de Recolección de Tweets

Código Historia de Usuario	007
Demostración:	

Tabla 27: Pantalla Seleccionar Cuentas de Twitter

Código Historia de Usuario	008
Demostración:	

Tabla 28: Pantalla Ingreso de Tópico

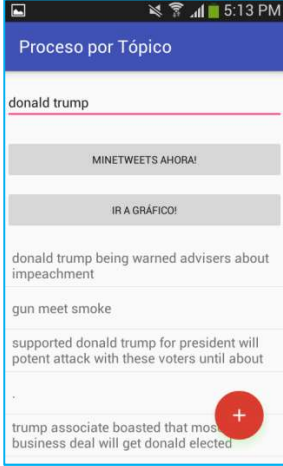
Código Historia de Usuario	009
Demostración:	
	

Tabla 29: Pantalla Ejecutar Proceso de Minería

Código Historia de Usuario	010
Demostración:	
	

Tabla 30: Pantalla Mostrar Resumen Tweets Recolectados

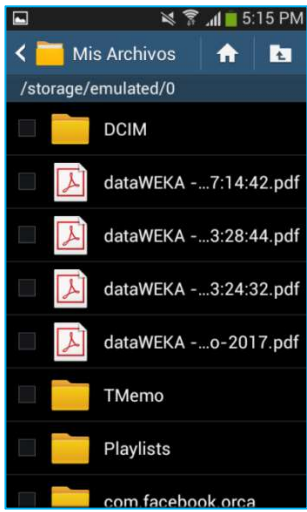
Código Historia de Usuario	011
Demostración:	
	

Tabla 31: Pantalla Descargar Reporte

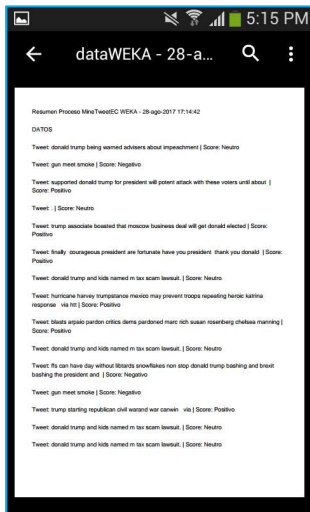
Código Historia de Usuario	012
Demostración:	
	

Tabla 32: Pantalla Reporte