Proyecto de Inteligencia Artificial*

1st Diana Sofia Reyes

Departamento de Ciencias Naturales e Ingeniería Universidad Jorge Tadeo Lozano Bogotá, Colombia dianas.reyesl@utadeo.edu.co

3rd Adrián Esteban Villalobos

Departamento de Ciencias Naturales e Ingeniería Universidad Jorge Tadeo Lozano Bogotá, Colombia adriane.villalbao@utadeo.edu.co

Abstract—En este documento presentaremos un problema generado por las aplicaciones existentes en la Play Store de Android que consta de el sentimiento producido por 10 mil aplicaciones utilizadas por los usuarios con el fin de conocer qué sentimiento lídera estas críticas y así analizar el mercado de Android. Por ende, obtuvimos las críticas de estos usuarios como información importante para analizar, procesar y asegurar el éxito de estas aplicaciones. Finalmente, procesamos esta información y visualizamos los datos importantes mediante diferentes tipos de gráficas.

Index Terms—Análisis de datos, Metodología CRISP, Machine Learning.

I. INTRODUCTION

La página web scraping obtuvo datos de 10 mil aplicaciones de la App Store con el fin de analizar y profundizar el mercado de Android. Aunque varios sitios web como Kaggle han proporcionado datos sobre la App store de Apple, no hay tanta información a diferencia de Play Store. Todos los datos obtenidos son a causa de Google Play Store, y de acuerdo a esta es posible la recopilación de esta información.

A partir de esta información, se analizarán las críticas generadas por los usuarios de la App Store acerca de estas 10 mil aplicaciones con el fin de conocer qué sentimiento les genera a estos usuarios al utilizar estas aplicaciones. Todo lo anterior con el propósito de asegurar el éxito de las mismas.

II. METODOLOGÍA CRISP

La metodología CRISP es un modelo de procesamiento de datos que proporciona un enfoque estructurado para planificar un proyecto de minería de datos y es un modelo que se lleva a cabo en secuencias de eventos.

En el caso de los proyectos de implementación de minería de datos esta metodología ha representado un grado alto de utilización en desarrollo de estos proyectos.

2nd Francisco Javier Vasquez

Departamento de Ciencias Naturales e Ingeniería

Universidad Jorge Tadeo Lozano

Bogotá, Colombia

franciscoja.vasquezg@utadeo.edu.co

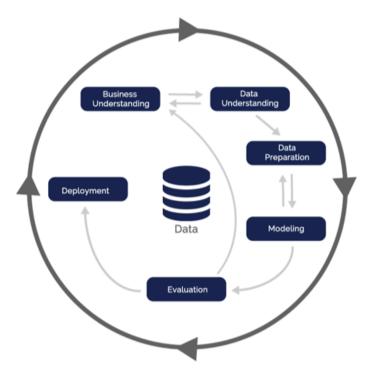


Fig. 1. Metodología CRISP

A. Fases

La metodología incluye un modelo estructurado en seis fases:

 Compresión del negocio: Es muy importante convertir la información adquirida del negocio en un problema de análisis de datos y que, de igual manera, permita alcanzar los objetivos de negocio propuestos.
 Lo que se quiere lograr es asegurar el éxito que tienen algunas de las aplicaciones existentes en Play Store según el sentimiento que les genera a los usuarios a

partir de sus críticas.

- 2) Comprensión de los datos: Las reviews y comentarios que hacen los usuarios en Play Store pueden ser una herramienta de feedback muy poderosa si se interpreta de la manera adecuada y a partir de esa interpretación se toman decisiones que lleven el producto a alcanzar los objetivos esperados.
- 3) Preparación de los datos: Para lograr el análisis de éxito de las aplicaciones, se le otorgó importancia al sentimiento que genera cada aplicación para los usuarios (positivo o negativo) y el valor de polaridad y subjetividad de cada crítica.
- 4) Modelado: La técnica de modelado a utilizar en este caso ha sido elegida en función de los siguientes criterios:
 - Ser adecuada para el problema Disponer de los datos necesarios para el problema Cumplir los requisitos del problema Tiempo necesario para adecuar el modelo
 - Las variables empleadas en la generación del modelo dependen de las características de los datos y la precisión que se quiere lograr con este proyecto.
- 5) Evaluación: Se evalúa el el modelo teniendo en cuenta los criterios de éxito del problema y conocer si este es útil para las necesidades del negocio.
- 6) Despliegue: Si al evaluar el modelo este comple con las necesidades del negocio y produce los resultados esperados, el modelo ya puede llevarse a cabo para su implementación.

III. ANÁLISIS Y VISUALIZACIÓN DE DATOS

Inicialmente, es necesario comprender todos los datos a analizar y el comportamiento de los mismos. Por ende, en la siguiente gráfica logramos representar la forma en que se comporta la polaridad de las reviews respecto a cada sentimiento:

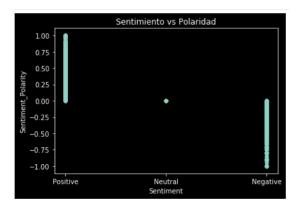


Fig. 2. Sentimiento y Polaridad

Evidentemente, el valor de la polaridad tiende a representar lógicamente el sentimiento generado en los usuarios. Es decir:

- Si el valor de polaridad es menor que 0 o negativo el sentimiento es Negativo.
- Si el valor de polaridad es igual a 0 el sentimiento es Neutro
- Si es valor de polaridad es mayor que 0 o positivo el sentimiento es Positivo.

Para lograr comprender los datos desde la pregunta más lógica: ¿Cuántas reviews positivas y negativas se presentan? Es importante representar los datos a partir de las siguientes gráficas:

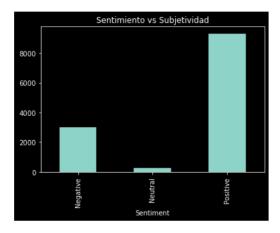


Fig. 3. Gráfico de barras Sentimiento y Subjetividad

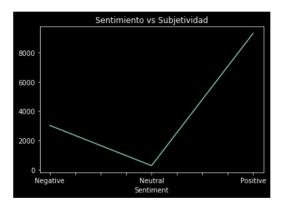


Fig. 4. Gráfico líneal Sentimiento y Subjetividad

- Como podemos ver, la cantidad de reviews positivas es mayor que las reviews negativas y/o neutrales.
- Las review neutrales no serán muy útiles ya que representan menos del 5% de las reviews totales
- Las review negativas se deben tener en cuenta, pero, probablemente no aporten demasiado
- La subjetividad del sentimiento de valor positivo debería ser revisado en parte de manera manual.

- Será necesario equilibrar los datos obtenidos para el modelo.
- Un tema importante que merece ser tenido en cuenta es la alta concentración de polarización.

IV. CONCLUSIONES

Los datos de las aplicaciones de Play Store tienen un enorme potencial para impulsar al éxito a las empresas de creación de aplicaciones. Se pueden extraer conocimientos prácticos para que los desarrolladores trabajen y capturen el mercado de Android.

La metodología de CRISP para proyectos de minería de datos es muy útil para comprender esta tecnología o extraer ideas para diseñar o revisar métodos de trabajo para proyectos con esas características.

REFERENCES

- [1] Galan Cortina, V.(2015). Aplicación de metodología Crisp-DM a un priyecto de minería de datos en el entorno universitario.
- [2] [Rodríguez, 2010] Dr. Oldemar Rodríguez Rojas. Metodología para el Desarrollo de Proyectos en Minería de Datos CRISP-DM, 2010.
- [3] ESPINOSA-ZUNIGA, Javier Jesús. Aplicación de metodología CRISP-DM para segmentación geográfica de una base de datos pública. Ing. invest. y tecnol. [online]. 2020, vol.21, n.1, e00008. Epub 03-Ago-2020. ISSN 1405-7743. https://doi.org/10.22201/fi.25940732e.2020.21n1.008.

URL GitHub:https://github.com/franciscoJavierV/inteligencia $_artificial_2021$ URLVideo:https://youtu.be/C8zc68Gf7Fk