

# Medición del éxito de un videojuego en Google Play antes de su lanzamiento

Danny Sebastián Díaz Padilla  
Facultad de ingeniería en sistemas  
Escuela Politécnica Nacional  
Andalucía y av. Ladrón de Guevara, 170525, Quito, Ecuador  
danny.diaz@epn.edu.ec

**Resumen-** Este documento presenta la definición formal del proyecto junto a las estrategias de minería de datos que se utilizarán que permitan identificar patrones en bases de datos gigantes para la posterior generación de conocimiento y sabiduría con el fin de ayudar al resto de personas que deseen iniciar un negocio por medio de un videojuego publicado en la plataforma Google Play. Se analiza también los indicadores claves de trabajos anteriores y relacionados con esta temática.

**Palabras Clave:** Videojuego, Minería de datos, Google Play, Web Scrapping.

## I. INTRODUCCIÓN

### A. Antecedentes

Debido a la adquisición de teléfonos inteligentes, desde el 2008 el mercado de aplicaciones en tiendas como App Store y Google Play ha incrementado [1]. Al momento de diseñar una aplicación existe una incertidumbre de si va a ser exitosa o no.

Tal y como lo expresa Nayebi et al. [2], la mitad de los desarrolladores tienen una clara estrategia de lanzamiento para tener una buena retroalimentación de los usuarios, sin embargo, esta estrategia se basa en sus recursos cognitivos, una lista del top 10 de aplicaciones más descargadas y su instinto. Así que este proyecto es un complemento a la estrategia de lanzamiento que tenga un desarrollador o un emprendedor, independientemente de la madurez que tenga la estrategia.

### B. Descripción de la problemática

La creación de un videojuego para móviles supone un gran costo sobre todo cuando al ser publicado no genera los beneficios esperados [3].

Los emprendedores ante grandes dudas sobre la idea que será más rentable invertir, necesitan de una referencia para optimizar su toma de decisiones.

Además, el uso de las aplicaciones en general, se ve reducido a medida que transcurre el tiempo, incluso se estima que el 65% de los usuarios dejan de utilizar una app 3 meses después de instalarla [1]. De forma que es importante analizar también cuáles son los videojuegos más “adictivos” (de gran enganche) de forma que el emprendimiento sea sostenible a largo plazo.

### C. Propósito del proyecto

El proyecto busca crear un modelo para predecir el éxito de un videojuego en base a las siguientes características: calificación, vistas, categoría y relación número de instalaciones-revisiones de los usuarios.

Tiene un propósito netamente comercial dirigido hacia emprendedores y programadores independientes que comprenden muy poco o nada acerca del mercado de videojuegos en teléfonos inteligentes.

### D. Importancia del proyecto

El proyecto ayuda a la toma de decisiones. Con el fin de elegir la mejor alternativa para desarrollar una aplicación dependiendo del estado del mercado en un tiempo predeterminado o en tiempo real.

La importancia radica en la evasión de proyectos sin un futuro claro y por ende una pérdida de dinero/tiempo considerable para el inversor.

Aumentará las probabilidades de que el producto a crear sea exitoso, mas no dará una recomendación de un producto infalible, porque el diseño y la creatividad de cada persona en el desarrollo de los mismos difiere ya sea en los que producirán el producto o el nicho de mercado que vaya a consumirlo.

## II. TRABAJOS PREVIOS Y RELACIONADOS

Los siguientes trabajos se enmarcan en el análisis de aplicaciones en las distintas tiendas. No existe un trabajo serio y fiable acerca del nicho específico de los videojuegos, sin embargo, estos trabajos aportan conclusiones muy interesantes y determinan en general indicadores claves que pueden reutilizarse en este trabajo.

**A. Bachelor of Science Incomputer Science and Engineering Exploratory Data Analysis and Success Prediction of Google Play Store Apps, Abdul Mueez et al. [4]**

En este trabajo se analiza la inmensa competencia dentro del mercado de las aplicaciones de Google Play para informar a un desarrollador para que sepa si está avanzando en la dirección correcta.

Se destaca sus conclusiones: los indicadores principales son la cantidad de instalaciones y las calificaciones de los usuarios que ha recibido a lo largo de su vida útil en lugar de los ingresos que generó. Descubrió características específicas como, por ejemplo, el número de palabras en el nombre de una aplicación afecta las instalaciones, y usarlos para averiguar qué aplicaciones tienen más probabilidades de tener éxito. Además del análisis del número de aplicaciones que se han hecho muchas veces acorde a una categoría como lo muestra en la Fig. 1.

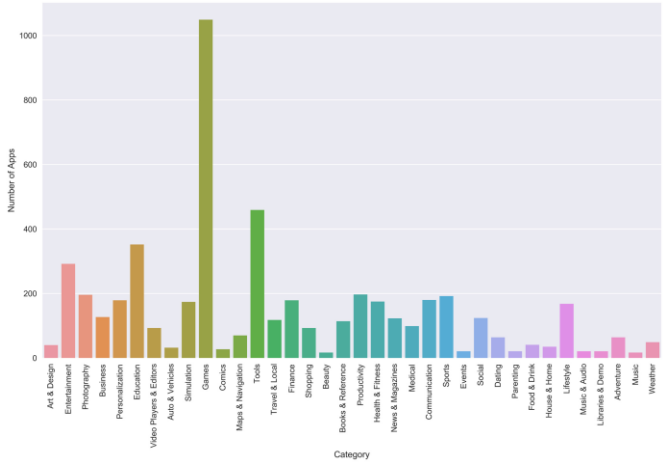


Fig. 1 Gráfica de barras de categorías contra número de aplicaciones

Y su análisis de correlaciones entre instalaciones y características de la aplicación como se muestre en un extracto en la Tabla 1.

Tabla 1

Tabla de Correlación entre número de instalaciones y otras características

Correlation of installs versus	Correlation value
Installs	1.000000
No_of_Ratings	0.141699
Size	0.060767
Rating	0.058889
Subjectivity_Mean	0.032624
Sentiment_Mean	-0.042139
Recent_Rating_Mean	-0.045481

### B. Causal Impact Analysis for App Releases in Google Play William Martin et al [3]

En este trabajo se analiza el impacto de los lanzamientos de software propios y de competidores. Utilizando 38,858 aplicaciones populares de Google Play, durante un período de 12 meses.

El trabajo encontró que el 33% de estos lanzamientos causaron un cambio estadísticamente significativo en las calificaciones de los usuarios.

Evaluaron a 56 empresas y el 78% estuvo de acuerdo con la evaluación causal, de las cuales el 33% afirmó que su compañía consideraría cambiar su estrategia de lanzamiento de la aplicación como resultado. de estos hallazgos.

En sus conclusiones se tiene que la frecuencia general de lanzamiento no está correlacionada con el éxito posterior de la aplicación, pero que hay evidencia de que el precio y el tamaño del texto de lanzamiento y el contenido juegan un papel importante. Es más probable que los lanzamientos con precios más altos sean significativos y, quizás sorprendentemente, que tengan una calificación de efecto positivo. Todo esto resumido en la Fig. 2.

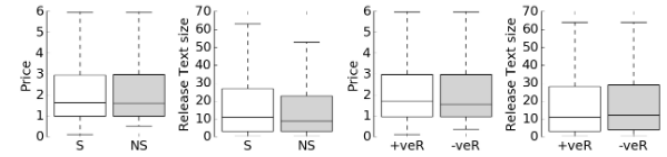


Fig. 2, Diagrama de cajas de precio y tamaño del texto de lanzamiento, comparando (S) emisiones significativas y (NS) no significativas, y emisiones que aumentaron la calificación (+ veR) y disminuyeron la calificación (-veR).

Se demuestra finalmente que el análisis causal puede ser una herramienta útil para los desarrolladores.

### C. Trabajo de fin de grado de C. E. Carazo de la Universidad de la Rioja: Citylok, una app de eventos global análisis de éxito. [5]

En este trabajo se identifica claves para crear una app teniendo en cuenta la oportunidad en el sector de las apps y las dificultades de rentabilidad.

Se analiza solamente un caso a profundidad, Citylok que es una app que reúne la oferta cultural, gastronómica y de ocio y la presenta de una forma geolocalizada, calendarizada y categorizada en función del tipo de evento. El caso se analiza desde el punto de vista de la oferta y la demanda.

Se concluye que el punto fuerte de Citylok es su idea de negocio que se valida como relevante tanto desde el punto de vista de sus creadores como de sus usuarios ya que las 161 personas encuestadas en ese trabajo así lo han declarado. Por medio de encuestas determinan también si la idea de negocio de la aplicación es rentable y útil como se muestra en la Fig. 3.

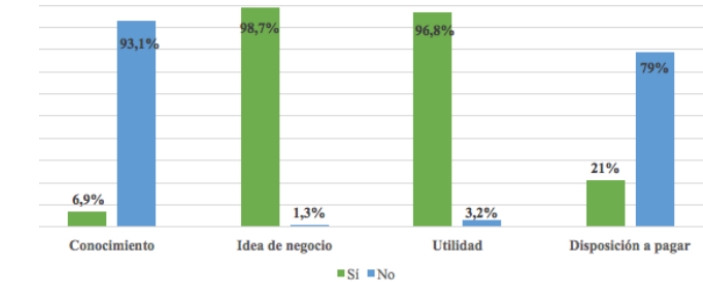


Fig. 3. Variables conocimiento, idea de negocio, utilidad y disposición a pagar

Y recomiendan para todos los emprendedores en nuevas tecnologías que utilicen las herramientas que les brinda el marketing móvil y de contenidos para crecer, consolidarse consiguiendo una comunidad del tamaño suficiente que les permita ser rentables.

#### D. A Survey of App Store Analysis for Software Engineering William Martín et al [6]

App Store Analysis estudia información sobre aplicaciones extraídas de tiendas de aplicaciones. Incorpora esta información no técnica con información técnica para conocer tendencias y comportamientos dentro de estas formas de repositorios de software. Esta encuesta describe y compara las áreas de investigación que se han explorado hasta ahora, extrayendo nuevas direcciones que la investigación futura debe tomar para abordar problemas y desafíos abiertos.

Muestran también las distribuciones de campos analizados en todos los estudios recopilados de las tiendas de aplicaciones

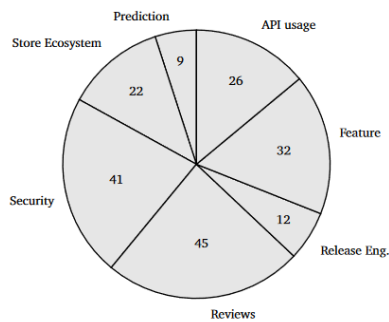


Fig. 4 Gráfico circular que muestra las distribuciones generales de los subcampos que muestran el período comprendido entre 2010 y el 27 de noviembre de 2015.

Como conclusiones rescatan que para la ingeniería de software se han identificado los subcampos clave de App Store Analysis hasta la fecha: análisis de API, análisis de características, ingeniería de versiones, análisis de revisión, análisis de seguridad, comparación de ecosistemas de tiendas y análisis de predicción de tamaño y esfuerzo.

Han observado la aparición de nuevas áreas de Análisis de App Store, y la progresión de ideas conceptuales a estudios empíricos prácticos que aplican y los refinan.

### III. DESCRIPCIÓN DE LA PROPUESTA DEL PROYECTO

#### A. Objetivo general y específicos

##### Objetivo general

Crear una guía para determinar las variables de éxito de un videojuego por medio de los procesos de Rapid Miner para un desarrollador de aplicaciones o emprendedor de forma que pueda determinar la tasa de éxito de su producto.

#### Objetivos específicos

- Crear un “dashboard” con los indicadores clave más importante del éxito de un videojuego en Google Play.
- Entrenar varios modelos matemáticos para la predicción de una variable de éxito la cual es dependiente de otras variables: puntuación de usuarios, descargas respecto revisiones de usuarios, revisiones de usuarios y la categoría.
- Seleccionar el mejor modelo matemático e incluir 50 ejemplos de prueba en el simulador.

#### B. Detalle

Para el proyecto se utilizará una cantidad de 267mil aplicaciones, de las cuales se filtrará únicamente los que tengan categorías relacionados a los videojuegos.

Para la selección de esa cantidad de aplicaciones se ha indagado en más de 15 bases de datos y se han filtrado aquellas bases de datos que serán útiles para el análisis.

Además, como un extra se pretende recolectar con Web Scraping información de los videojuegos más exitosos dado que no existe una base de datos confiable que reúna características de este entorno en específico

Posteriormente se realizará una limpieza de datos utilizando una herramienta open source llamada Open Refine y usando un lenguaje de programación (Python) se obtendrá un conjunto de datos de calidad con desviación estándar aceptable e información confiable.

Con la base de datos limpia se la incluirá en la herramienta Rapid Miner y se procederá a realizar filtros especiales y gráficas para encontrar patrones.

Se utilizará la herramienta de auto modelamiento para seleccionar el mejor algoritmo de predicción en base a las características que se determinen más importantes del análisis anterior. Y utilizando el simulador se incluirá 50 ejemplos de prueba para ver que factores tienen más “peso” al momento de crear un videojuego.

Este mismo proceso se realizará para la recolección de los datos utilizando Web Scraping. Se determinará por dos vías diferentes los mejores factores, los mejores algoritmos y finalmente se recomendará un claro conjunto de campos para que el desarrollador y el emprendedor realicen una elección correcta sobre su próxima idea de negocios.

#### IV. HERRAMIENTAS QUE SE PRETENDE USAR

##### A. *Open Refine (De Google)*

Es una herramienta originalmente creada por Google para el manejo de bases de datos. Permite limpiar bases de datos, exportarlas en diferentes formatos, y arreglar y manejar las bases para un mejor uso. Actualmente el proyecto ya no es financiado por Google y se encuentra como proyecto abierto.[7]

##### B. *Rapid Miner*

Rapidminer es una herramienta perfecta para crear modelos y a posterior la realización de análisis predictivos de grandes volúmenes de datos.

Es una solución que facilita el autoservicio de análisis predictivo permitiendo una avanzada analítica empleando solamente drag and drop y opcionalmente la generación de código.

Se utiliza para realizar análisis de minería de datos (Data Mining) en aplicaciones empresariales, gobierno y academias. [8]

##### C. *PyCharm (Como IDE para Python)*

Aunque RapidMiner ofrece una forma de limpiar los datos, no es tan completa como la de OpenRefine (excepto en la formación de terceras columnas con nueva data de forma sencilla).

Python es el lenguaje estándar usado en Open Refine para aplicar funciones en toda una columna de datos de forma que se pueda formatear las columnas y asignarles un tipo de valor en el proceso de limpieza.

PyCharm es uno de los entornos de desarrollo más completos para Python. Es parte de la suite de herramientas de programación ofrecidas por JetBrains, que cuenta con entornos para construir código en distintos idiomas como PHP y Ruby. Sin embargo, es muy cómodo de utilizar. [9]

##### D. *Web Scrapper*

Es una extensión de navegador que permite la extracción de datos de una página de forma intuitiva y fácil mediante un mapeo del sitio y automatización del proceso de interacción usuario -> página web. [10]

Se concluye que este proyecto al abarcar un segmento específico de las aplicaciones otorgará un nuevo aporte o al menos una nueva perspectiva en el desarrollo y emprendimiento.

De trabajos anteriores y relacionados se rescata que los indicadores principales son: la cantidad de instalaciones, las calificaciones de los usuarios que ha recibido a lo largo de su vida útil, número de palabras en el nombre de la aplicación, el precio, el contenido (la idea de negocio y su utilidad).

Además, la herramienta RapidMiner ofrece un auto modelamiento de inteligencia artificial y limpieza de datos, pero con menos características que la limpieza de datos que ofrece open refine.

#### AGRADECIMIENTOS

Un agradecimiento a Gautham Prakash<sup>1</sup>, el cual ha recolectado la principal base de datos a utilizarse en el proyecto usando Web Scrapping junto con Python.

#### REFERENCIAS

[1] I. Group, "Factores que marcan el éxito o el fracaso de una app", *Digitaltransformation.ituser.es*, 2016. [Online]. Available:

<https://digitaltransformation.ituser.es/noticias/2016/09/factores-que-marcen-el-exito-o-el-fracaso-de-una-app>. [Accessed: 19- Nov- 2019].

[2] M. Nayebi, B. Adams, and G. Ruhe. Mobile app releases – a survey research on developers and users perception. In IEEE 23rd International Conference on Software Analysis, Evolution and Reengineering (SANER'16). IEEE, 2016.

[3] Martin, W., Sarro, F. and Harman, M. (n.d.). *Causal Impact Analysis for App Releases in Google Play*. [online] ucl. Available at: [http://www0.cs.ucl.ac.uk/staff/f.sarro/resource/papers/Martin\\_FSE\\_Causal\\_PrePrint.pdf](http://www0.cs.ucl.ac.uk/staff/f.sarro/resource/papers/Martin_FSE_Causal_PrePrint.pdf) [Accessed 19 Nov. 2019].

[4] A. Mueez, K. Ahmed, T. Islam and W. Iqba, "Exploratory Data Analysis and Success Prediction of Google Play Store Apps", *Dspace.bracu.ac.bd*, 2018. [Online]. Available: [http://dspace.bracu.ac.bd/xmlui/bitstream/handle/10361/11407/15101108%2C15101020%2C15101109%2C15141002\\_CS\\_E.pdf?sequence=1&isAllowed=y](http://dspace.bracu.ac.bd/xmlui/bitstream/handle/10361/11407/15101108%2C15101020%2C15101109%2C15141002_CS_E.pdf?sequence=1&isAllowed=y). [Accessed: 19- Nov- 2019].

[5] C. Escudero Carazo, "Citylok, una app de eventos global", *Biblioteca.unirioja.es*, 2018. [Online]. Available: [https://biblioteca.unirioja.es/tfe\\_e/TFE003059.pdf](https://biblioteca.unirioja.es/tfe_e/TFE003059.pdf). [Accessed: 19- Nov- 2019].

#### V. CONCLUSIONES

---

<sup>1</sup> Su perfil <https://www.kaggle.com/gauthamp10> y la base de datos que recolectó <https://www.kaggle.com/gauthamp10/google-playstore-apps>

[6] W. Martin, F. Sarro, Y. JiaYuanyuan Zhang and M. Harman, "A Survey of App Store Analysis for Software Engineering", *Cs.ucl.ac.uk*, 2016. [Online]. Available: [http://www.cs.ucl.ac.uk/fileadmin/UCL-CS/research/Research\\_Notes/RN\\_16\\_02.pdf](http://www.cs.ucl.ac.uk/fileadmin/UCL-CS/research/Research_Notes/RN_16_02.pdf). [Accessed: 19-Nov- 2019].

[7] "Open Refine", *https://es.schoolofdata.org/2014/06/30/openrefine/*, 2014. [Online]. Available: <https://es.schoolofdata.org/2014/06/30/openrefine/>. [Accessed: 19- Nov- 2019].

[8] B. Fernández García, "Rapidminer: software data mining | Clarcat", *Clarcat*. [Online]. Available: <https://www.clarcat.com/rapidminer/>. [Accessed: 19- Nov- 2019].

[9] "PyCharm: uno de los mejores IDE para Python — Escuela de Python", *Escuela de Python*, 2019. [Online]. Available: <https://www.escuelapython.com/pycharm-uno-de-los-mejores-ide-para-python/>. [Accessed: 19- Nov- 2019].

[10]"Web Scraper - The #1 web scraping extension", *Webscraper.io*. [Online]. Available: <https://webscraper.io/>. [Accessed: 27- Nov- 2019].