



UNIVERSIDAD ESAN
FACULTAD DE INGENIERÍA
INGENIERÍA DE TECNOLOGÍAS DE INFORMACIÓN Y SISTEMAS

Predicción del estado de financiamiento de proyectos de tecnología en web de crowdfunding
Kickstarter mediante modelo(s) de Aprendizaje Automático

Tesis para optar el Título de Ingeniero de Tecnologías de Información y Sistemas que
presenta:

Alonso Augusto Puente Ríos
Asesor: Marks Arturo Calderón Niquin

Lima, 23 de junio de 2020

Esta tesis denominada:

PREDICCIÓN DEL ESTADO DE FINANCIAMIENTO DE PROYECTOS DE
TECNOLOGÍA EN WEB DE CROWDFUNDING KICKSTARTER MEDIANTE
MODELO(S) DE APRENDIZAJE AUTOMÁTICO

ha sido aprobada.

.....
(Jurado Presidente)

.....
(Jurado)

.....
(Jurado)

Universidad ESAN
2020

PREDICCIÓN DEL ESTADO DE FINANCIAMIENTO DE PROYECTOS DE
TECNOLOGÍA EN WEB DE CROWDFUNDING KICKSTARTER MEDIANTE
MODELO(S) DE APRENDIZAJE AUTOMÁTICO

Agradecimiento y dedicatoria

Durante la inducción en la empresa en la cual realicé mis segundas prácticas profesionales, se realizaron varias actividades, entre ellas, una que me marcó positivamente. Esta consistía en comparar los tiempos de llegada de un punto a otro de una persona corriendo. Se caracterizó porque quien asumió el reto tuvo presente en su mente las personas y las razones por las cuales todos los días lucha y son su principal fuente de motivación.

Por ello, quiero dedicar este gran esfuerzo personal de trabajo de tesis a quienes siempre han estado a mi lado en los mejores y peores momentos, aquellos críticos en que definen el destino. Mi amada hermana Clarisabel, mis queridos padres Augusto e Isabel, mi familia en especial mis abuelos; y mis pocos, pero verdaderos y leales amigos de la universidad, colegio y trabajo. Todos ellos han sido y son cada uno, piedra fundamental en el desarrollo de mi ser como persona y profesional, así como también seres con los cuales siempre comparto gratos momentos. Su presencia en mi vida no ha sido una suerte más sino parte de mi destino. Asimismo, luchar por mis sueños y mi país, y pensar cada día en solidificar su planificación me motivan emocionalmente hasta en aquellos momentos en que parece haber imposibles.

Quiero concluir esta sección, muy especial para mí, agradeciendo también a mi alma máter, la Universidad Esan, y al Programa Nacional de Becas (Pronabec) por hacer que estos 5 años entre el 2015 y 2019 sean mágicos y muy fructíferos. Tuve la oportunidad no solo de incrementar y potenciar mis conocimientos en distintas áreas académicas sino también de aprender de excelentes profesionales como mis profesores, conocer grandes amigos dentro y fuera de su campus (desde el primer ciclo como cachimbo hasta el último ciclo, en el CADE Universitario 2019, estudiantes de diferentes universidades y otras partes del Perú), ponerme a prueba en el exterior (en el II Congreso Internacional de Investigación en Colombia) y formar parte de la gran familia UE.

Por todos ellos, simplemente gracias.

Índice general

Índice de Figuras	7
Índice de Tablas	8
1. PLANTEAMIENTO DEL PROBLEMA	11
1.1. Descripción de la Realidad Problemática	11
1.2. Formulación del Problema	13
1.2.1. Problema General	14
1.2.2. Problemas Específicos	14
1.3. Objetivos de la Investigación	15
1.3.1. Objetivo General	15
1.3.2. Objetivos Específicos	15
1.4. Justificación de la Investigación	15
1.4.1. Teórica	15
1.4.2. Práctica	16
1.4.3. Metodológica	16
1.5. Delimitación del Estudio	16
1.5.1. Espacial	16
1.5.2. Temporal	16
1.5.3. Conceptual	16

1.6. Hipótesis	17
1.6.1. Hipótesis General	17
1.6.2. Hipótesis Específicas	17
1.6.3. Matriz de Consistencia	17
2. MARCO TEÓRICO	18
2.1. Antecedentes de la investigación	18
2.1.1. Supervised Learning Model For Kickstarter Campaigns With R Mining (Kamath & Kamat, 2018)	18
2.2. Bases Teóricas	19
2.2.1. Machine Learning	19
2.2.2. Natural Language Processing (NLP)	20
2.3. Marco Conceptual	20
3. METODOLOGÍA DE LA INVESTIGACIÓN	21
3.1. Diseño de la investigación	21
3.1.1. Enfoque de la investigación	21
3.1.2. Alcance de la investigación	21
3.1.3. Tipo de la investigación	22
3.1.4. Descripción del prototipo de investigación	22
3.2. Población y muestra	22
3.2.1. Población	22
3.2.2. Muestra	22
3.2.3. Unidad de análisis	23
3.3. Operacionalización de Variables	24
3.4. Instrumentos de medida	24
3.5. Técnicas de recolección de datos	24

3.6. Técnicas para el procesamiento y análisis de la información	25
3.7. Cronograma de actividades y presupuesto	25
4. DESARROLLO DEL EXPERIMENTO	27
4.1. X	27
4.2. Y	27
4.3. Z	28
5. ANÁLISIS Y DISCUSIÓN DE RESULTADOS	29
5.1. X	29
5.2. Y	29
5.3. Z	30
6. CONCLUSIONES Y RECOMENDACIONES	31
6.1. Conclusiones	31
6.2. Recomendaciones	31
Anexos	32
A. Anexo I: Matriz de Consistencia	33
B. Anexo II: Resumen de Papers investigados	35
BIBLIOGRAFÍA	37

Índice de Figuras

1.1. Resultados y ratios obtenidos en la encuesta por GEM y ESAN. Fuente: Redacción Gestión, 2018	12
1.2. Ratio de éxito de proyectos en Kickstarter desde 2009 hasta 2019 (Febrero). Fuente: The Hustle, 2019	14
3.1. Prueba de Figura	23

Índice de Tablas

3.1. An example table.	26
4.1. An example table.	27
5.1. An example table.	29
A.1. Matriz de consistencia. Fuente: Elaboración propia	34
B.1. Cuadro Resumen de Papers investigados. Fuente: Elaboración propia	36

Resumen

Conocer el destino del financiamiento de proyectos siempre ha sido el principal deseo de todos los emprendedores que los promocionan en Internet, en especial, de la categoría de tecnología por ser los que presentan las ratios más bajas de éxito debido a sus altas metas que buscan alcanzar. El presente trabajo de investigación se basó en construir un modelo predictivo cuyo objetivo es la de estimar el éxito o fracaso de financiamiento de proyectos tecnológicos en la plataforma de crowdfunding Kickstarter durante la duración de su campaña a partir de su metainformación, imagen y/o descripción. Para ello, se crearon modelos de Aprendizaje Automático (SVM, MLP y CNN) para cada una de estas partes. Luego de analizar todos los modelos con las mismas métricas mencionadas en el décimo antecedente, se concluyó que solo los de metainformación tuvieron niveles excelentes de acuerdo a sus puntajes AUC (0.8377 para SVM y 0.7043 para MLP), mientras que los modelos de descripciones tuvieron un rendimiento regular (0.6746 para SVM con TF-IDF y 0.6709 para SVM con BoW) y finalmente el modelo de imágenes no logró clasificar las clases correctamente. Como trabajo a futuro, estos últimos modelos de descripción e imagen serán mejorados para construir, junto con los de metainformación, un modelo ensamblado basado en considerables variables del proyecto.

Palabras claves: metainformación, descripción, imagen del proyecto, Máquina de Vectores de Soporte (SVM), Perceptrón Multicapa (MLP), Red Neuronal Convolutiva (CNN).

Knowing funding projects destiny has always been the main desire of all entrepreneurs who promote them on the Internet, especially in the technology category because they have the lowest success rates due to their high goals. The present research work was based on building a predictive model whose objective is to estimate the success or failure of technological projects funding in the Kickstarter crowdfunding platform during the duration of its campaign based on its metadata, image and/or description. For this, Machine Learning models (SVM, MLP and CNN) were created for each of these parts. After analyzing all the models with the same metrics mentioned in the tenth antecedent, it was concluded that only metadata models had excellent levels according to their AUC scores (0.8377 for SVM and 0.7043 for MLP), while the description models had a regular performance (0.6746 for SVM with TF-IDF and 0.6709 for SVM with BoW) and finally the image model failed to classify the classes correctly. As future work, these latest models of description and image will be improved to build, together with metadata ones, an assembled model based on considerable project variables.

Palabras claves: project metadata, project description, project image, Support Vector Machine (SVM), Multilayer Perceptron (MLP), Convolutional Neural Network (CNN).

Introducción

Por muchos años, en especial en las dos últimas décadas, diversos proyectos emprendedores han sido lanzados en distintas plataformas web, buscando un objetivo compartido por todos: ser financiados en un determinado plazo para hacer realidad estas ideas. Entre fracasos y éxitos, han surgido nuevas tendencias, así como nuevos enfoques de estudios de estos casos para encontrar la clave que descifre las variables de éxito.

El presente trabajo de investigación se basó formular un modelo ensamblado robusto que determine el estado final de un proyecto, agregando un nuevo enfoque: basarse solamente en proyectos de tecnología, la segunda categoría con más baja probabilidad de éxito al final de una campaña. En estudios previos, los modelos planteados resultaron bastante aceptables debido a que el resto de categorías de proyectos balancearon la inequidad existente en las dos clases del estado.

El reto principal fue el de construir modelos predictivos que consideren las tres características más importantes de un proyecto: la primera basada en la metainformación, la segunda, en la imagen principal del proyecto y la tercera, en la descripción del mismo, para ser ensamblados más adelante en uno solo.

Para ello, se recolectó un total de 27,251 proyectos tecnológicos en Kickstarter entre los periodos 2009-2019, de los cuales 27,035 registros finalmente fueron usados para cada uno de los tres modelos. Algunos proyectos provenientes de países fuera del territorio de los Estados Unidos y en distintos idiomas fueron considerados dentro de esta cantidad ya que no afectó al rendimiento general como en casos particulares de algunos estudios previos.

La principal motivación de este trabajo fue la de aportar una herramienta de ayuda para los emprendedores que les permita estimar el estado final del financiamiento de su proyecto de tecnología, es decir, éxito o fracaso, con un nivel confiable de probabilidad de éxito del mismo durante el transcurso de su campaña, permitiendo además servir de soporte en la toma de decisiones de cara a lograr su principal objetivo.

Capítulo 1

PLANTEAMIENTO DEL PROBLEMA

1.1. Descripción de la Realidad Problemática

El emprendimiento hoy en día es una realidad en todo el mundo. Desde crear productos nuevos hasta crear nuevas formas de hacer las cosas, todo gracias a ideas nacidas a partir de querer satisfacer nuestras propias necesidades.

Nuestro país no es ajeno a ello. El 50.6 % de la población entre 18 y 64 años tiene la expectativa de iniciar un emprendimiento dentro de los tres próximos años de acuerdo al último reporte de Global Entrepreneurship Monitor (GEM) 2014. El 62.3 % de la población entre ese rango de edad, además, tiende a ser más optimista en su percepción de oportunidades. Asimismo, según informa la Cámara de Comercio de Lima, la iniciativa emprendedora responde más a la identificación de una oportunidad de negocio que a una falta de oportunidad de empleo ([Redacción Gestión, 2015](#)). Sin embargo, en un estudio más reciente basado en una encuesta realizada a residentes peruanos entre junio y julio del 2017 desarrollada por el equipo GEM Perú y ESAN a 2080 personas entre el mismo rango de edad, el 24.6 % de emprendimientos se encuentra en fase temprana, es decir, representa una dificultad para el emprendedor peruano llegar a etapas más avanzadas como un emprendimiento establecido (negocios con más de 3.5 años, que representan solo el 7.4 % para Perú), ubicando así a nuestro país en la posición 25 de 54 economías a nivel mundial ([Redacción Gestión, 2018](#)). En la Figura 1.1 se aprecian algunos ratios del estudio.

Estos resultados desfavorables tienen como base el ecosistema poco beneficioso para los emprendimientos que permitan su establecimiento en el entorno nacional, con condiciones asociadas al acceso de financiamiento, políticas gubernamentales que alienten la implementación de Innovación y Desarrollo en las empresas, acceso a infraestructura física y asesoría

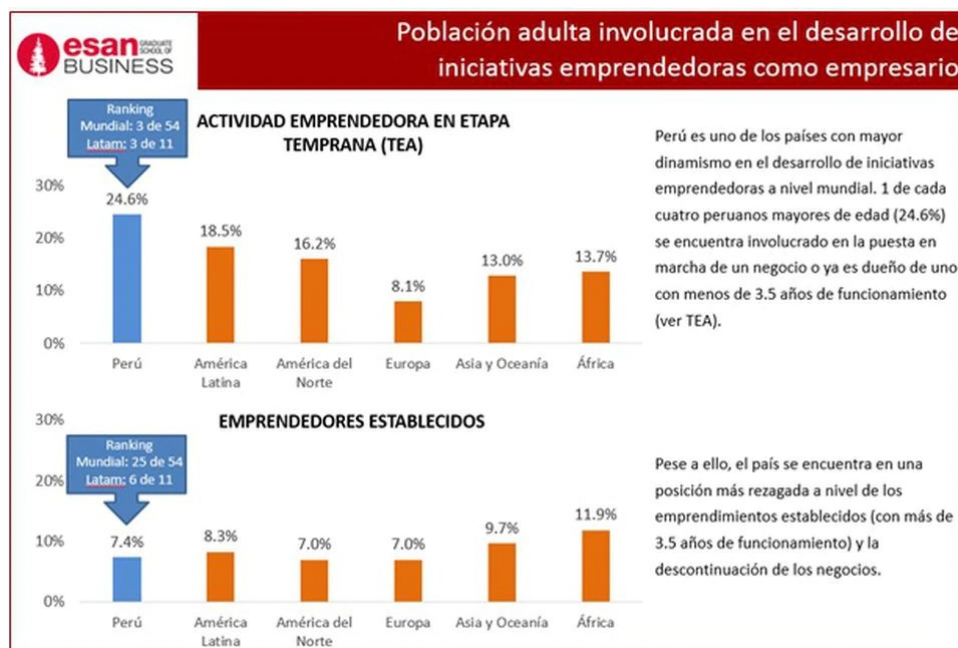


Figura 1.1: Resultados y ratios obtenidos en la encuesta por GEM y ESAN. Fuente: [Redacción Gestión, 2018](#)

a nivel comercial y profesional, como sostiene el investigador del equipo GEM Perú Carlos Guerrero ([Redacción Gestión, 2018](#)). La Asociación de Emprendedores de Perú (ASEP) afirma, asimismo, que en la región solo se invierte el 1.5 % del PIB en actividades de ciencia, tecnología e innovación, y las limitaciones son dadas por barreras burocráticas ejercidas por el Gobierno y el sector privado ([Asociación de Emprendedores de Perú, 2018](#)). En adición a esto, otras razones que representan barreras para emprender son la falta de conocimientos en la iniciación de un negocio, su tramitación, la fuente de financiamiento del proyecto o búsqueda de inversionistas, la cultura, la falta de fomento de emprendimiento y la falta de una red de contactos ([Sandoval, s.f.](#)).

Ante estas limitaciones, en la actualidad muchos emprendedores se ven forzados a mostrar sus proyectos al público en la Internet con el fin de captar personas interesadas en ayudarlos en el financiamiento de estos. Por ello, se han creado plataformas web con el fin de permitir la interacción entre los proyectos publicados en un determinado tiempo, el cual puede variar entre 30 y 120 días, y la comunidad en general que desee colaborar con una cantidad de dinero para su financiamiento. El sitio web solo servirá para mostrar los proyectos presentados a detalle por los creadores y la promoción de estos al público. La idea es que, al término de este plazo de tiempo, el proyecto sea financiado y se logre convertir en una realidad. A esta práctica se le conoce como crowdfunding ([Universo Crowdfunding, s.f.](#)).

En Latinoamérica, son muy pocos los países los que se incorporan en el crowdfunding,

tales como Chile, México, Argentina y Brasil. Sin embargo, el modelo funciona distinto a países de Norteamérica y Europa debido a la cultura diferente y resistencia a su implementación por la poca confianza en el éxito de los proyectos. En los últimos años se decidió seguir una manera muy similar a los modelos de Estados Unidos, basados en la creación de campañas de un emprendedor para obtener fondos para sus ideas con la moneda norteamericana pero limitados a las leyes económicas de cada país ([Solidaridad Latina, s.f.](#)).

Entre los sitios web más conocidos de crowdfunding están Kickstarter e Indiegogo. Kickstarter, desde su inicio en 2009, es una plataforma de financiamiento de proyectos creativos de todo tipo, los cuales incluyen películas, juegos, música, arte, diseño y tecnología. Actualmente, se han registrado más de 162 mil proyectos realizados, 16 millones de contribuyentes y 4,3 miles de millones de dólares fondeados ([Kickstarter, s.f.-a](#)). La plataforma utiliza un modelo de financiamiento llamado “todo o nada”, el cual consiste en que si un proyecto no alcanza su meta de financiamiento en un determinado plazo de tiempo, no se realiza ninguna transacción de fondos ([Kickstarter, s.f.-b](#)). Si bien los patrocinadores apoyan estos proyectos por motivos personales y distintos para hacerlos realidad, ellos no obtienen la propiedad o los ingresos de los proyectos que financian, sino que los creadores conservan la totalidad de su trabajo ([Kickstarter, s.f.-c](#)).

Para los proyectos tecnológicos, en contraste, la ratio de éxito es uno de los más bajos de las categorías existentes (20 %) solo por delante de Artesanía y Periodismo, como se aprecia en la Figura 1.2.

Ya existen estudios previos para predecir la probabilidad de éxito de financiamiento para este tipo de proyectos utilizando técnicas de Aprendizaje Automático. Sin embargo, la mayoría de los modelos predictivos propuestos no arrojan resultados con exactitud muy alta ya que su rango varía entre 60 y 70 %. Esto conlleva a generar imprecisión para pronosticar confiablemente el éxito de financiamiento de estos proyectos de tecnología. Para el presente trabajo de tesis, se creó un modelo predictivo alimentado de datos históricos de la plataforma para estimar el estado final de financiamiento de un proyecto aleatorio, así como su probabilidad de éxito.

1.2. Formulación del Problema

Para la formulación de los problemas de la presente investigación, se elaboró un «árbol de problemas» (véase Anexo 1).

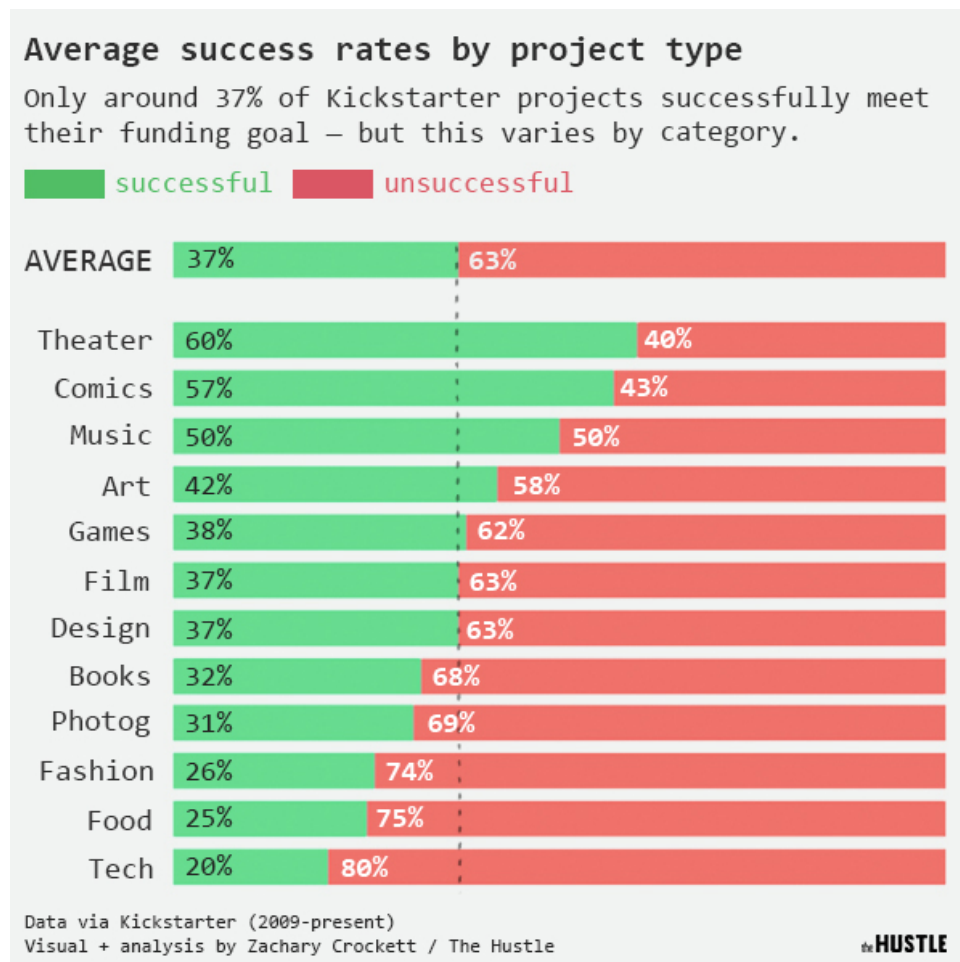


Figura 1.2: Ratio de éxito de proyectos en Kickstarter desde 2009 hasta 2019 (Febrero). Fuente: [The Hustle, 2019](#)

1.2.1. Problema General

Bajos niveles de precisión de modelos entrenados de Aprendizaje Automático para cualquier categoría para predecir estado de financiamiento de proyectos de tecnología.

1.2.2. Problemas Específicos

- Variables de proyectos no normalizadas y varianzas altas.
- Datos faltantes o incompletos de proyectos.
- Parámetros de modelos no ajustados.
- Sobreajuste de aprendizaje de modelos y clasificación incorrecta de las dos clases del estado final de financiamiento (exitoso o fracasado).

- Predicción incorrecta de estado de financiamiento de un proyecto tecnológico.

1.3. Objetivos de la Investigación

Para la formulación de los objetivos de la presente investigación, se elaboró un «árbol de objetivos» (véase Anexo 2)

1.3.1. Objetivo General

Construir modelo(s) de Aprendizaje Automático entrenado(s) para predecir correctamente proyectos de tecnología con nivel de precisión aceptable.

1.3.2. Objetivos Específicos

- Normalizar variables de proyectos y reducir niveles altos de varianza.
- Eliminar datos faltantes o incompletos de proyectos.
- Ajustar parámetros de modelos.
- Evitar sobreajuste de aprendizaje de modelos.
- Predecir correctamente el estado final de financiamiento de cualquier proyecto tecnológico (éxito o fracaso).

1.4. Justificación de la Investigación

1.4.1. Teórica

Esta investigación se basa en crear un modelo de Aprendizaje Automático que sea aplicable a proyectos de tecnología de la plataforma Kickstarter por presentar bajas performances en antecedentes.

1.4.2. Práctica

Al culminar la investigación, se ofrecerá un modelo predictivo confiable que ayude a los emprendedores en la toma de decisiones respecto a sus proyectos a partir del insight obtenido de los resultados que deriven a la manipulación de los datos de entrada.

1.4.3. Metodológica

Se creará un modelo predictivo a partir de las variables finales seleccionadas, previa limpieza de datos. Luego, será entrenado y evaluado por las métricas correspondientes. Finalmente, se lanzará una versión de prueba que reciba datos de entrada para predecir la viabilidad de un proyecto de tecnología.

1.5. Delimitación del Estudio

1.5.1. Espacial

Para la presente investigación, se considerará el territorio de los Estados Unidos ya que tanto la campaña del proyecto a servir para la investigación como los datos fuentes de proyectos relacionados financiados previamente, que servirán para la elaboración del modelo predictivo, se encuentran en dicho país.

1.5.2. Temporal

El periodo de tiempo abarcará desde el año 2009, fecha en el cual se tiene registrado los primeros conjuntos de datos de proyectos en Kickstarter hasta el mes de agosto del año 2019, últimos registros descargados hasta el inicio del presente trabajo.

1.5.3. Conceptual

La presente investigación consistirá en la implementación de un modelo predictivo del estado de financiamiento de un proyecto tecnológico en Kickstarter basado en técnicas y conceptos de Aprendizaje Automático, previamente evaluando cuál de todas las existentes genera un mejor desempeño para su uso y análisis de resultados.

1.6. Hipótesis

1.6.1. Hipótesis General

El modelo entrenado de Aprendizaje Automático logrará predecir correctamente proyectos de tecnología con nivel de precisión aceptable.

1.6.2. Hipótesis Específicas

- Las variables de los proyectos descargados se normalizarán y se reducirán los niveles altos de varianza.
- Los datos faltantes o incompletos de los proyectos serán eliminados.
- Los parámetros de los modelos usados serán ajustados.
- Se evitará el sobreajuste de aprendizaje de modelos para clasificar correctamente las dos clases del estado final de financiamiento.
- El estado final de financiamiento de cualquier proyecto tecnológico será predicho correctamente.

1.6.3. Matriz de Consistencia

A continuación se presenta la matriz de consistencia elaborada para la presente investigación (véase Anexo [A.1](#)).

Capítulo 2

MARCO TEÓRICO

2.1. Antecedentes de la investigación

En esta sección se presentarán diversos trabajos de investigación basados en la predicción de éxito o fracaso de campañas en Kickstarter o plataformas similares y el análisis de estas utilizando conjunto de datos de la propia plataforma o almacenadas en otros repositorios, con sus variables respectivas. En la mayoría de casos consideraron variables básicas que se obtienen del repositorio de las plataformas de crowdfunding, en otros casos consideraron variables cualitativas basadas en texto y descripción de proyectos, y en otros antecedentes usaron nuevas técnicas poco convencionales como el Aprendizaje Profundo e híbridos de modelos para obtener los mejores resultados posibles. Asimismo, a continuación se presenta un cuadro resumen (véase Anexo [B.1](#)) de lo que se presenta en esta sección.

2.1.1. Supervised Learning Model For Kickstarter Campaigns With R Mining ([Kamath & Kamat, 2018](#))

Kamath y Kamat realizaron un artículo de investigación el cual fue publicado en la revista «Resources Policy» en el año 2018. Este fue titulado «Copper price estimation using bat algorithm» la cual traducida al español significa «Estimación del precio del cobre utilizando el algoritmo bat».

2.1.1.1. Planteamiento del Problema y objetivo

hhhhj

2.1.1.2. Técnicas empleadas por los autores

Los autores plantearon emplear una combinación entre la función de series de tiempo y el aljhkk.

2.1.1.3. Metodología empleada por los autores

gfhhhh

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (O_i - T_i)^2}{N}} \quad (\text{Ecuación 2.1})$$

gfghf tal forma mejorar aún más la precisión de la predicción del precio del cobre.

2.1.1.4. Resultados obtenidos

Las funciones de serie de tiempo más importantes se usaron para estimar los cambios en el precio del cobre. Entre ellos, la serie BMMR con una media de RMSE de 0.449 presentó la mejor estimación. El algoritmo Bat se usó para modificar la función de tiempo BMMR debido a su alta capacidad para estimar los cambios en el precio del metal. Se obtuvo un RMSE de 0.132 de la ecuación modificada con BA. Los resultados obtenidos tienen una precisión mucho mayor y, a diferencia del BMMR, están más cerca de la realidad.

2.2. Bases Teóricas

2.2.1. Machine Learning

Es un subcampo de l]ecutar dificultosos procesos aprendiendo de datos, en lugar de seguir reglas preprogramadas ([Royal Society Working Group, 2017](#)).

es importante mencionar que existen también cinco tipos de problemas de aprendizaje que se pueden enfrentar: regresión, clasificación, simulación, optimización y clusterización ([Gollapudi, 2016](#)). Por otro lado, el aprendizaje automático también posee una división por subcampos que se puede observar en la Figura 14.

2.2.2. Natural Language Processing (NLP)

Naturalmano ([Goyal y col., 2018](#)). Otra definición para este término implica que es un campo especializado de la informática que es

De acuerdo con [Goyal y col. \(2018\)](#), e

2.3. Marco Conceptual

Para de

Capítulo 3

METODOLOGÍA DE LA INVESTIGACIÓN

3.1. Diseño de la investigación

En esta sección del documento se explicará cual es el diseño, el tipo y el enfoque del trabajo de investigación, así como también la población y la muestra.

3.1.1. Enfoque de la investigación

El presente trabajo tendrá un enfoque cuantitativo ya que se busca diseñar y desarrollar instrumentos, en este caso modelos predictivos, para responder al problema estudiado a partir de medición de datos históricos en la plataforma Kickstarter con herramientas basadas en la estadística y matemáticas que puedan ser interpretadas por cualquier investigador.

3.1.2. Alcance de la investigación

El alcance del presente trabajo será descriptivo ya que se recolectarán datos en un determinado rango de tiempo (desde 2009 hasta el presente año 2019) para describir el comportamiento de las campañas de proyectos tecnológicos en Kickstarter a partir de las características de sus variables y con ello, pronosticar su posible éxito o fracaso antes de finalizar la campaña con un nivel óptimo de precisión.

3.1.3. Tipo de la investigación

Para determinar el tipo de la investigación, primero es necesario definir el actual trabajo como Diseño Experimental ya que las variables que se tienen serán controladas, es decir, serán agregadas o quitadas en el o los modelos construidos en el experimento para analizar el impacto que este o estos tendrán en los resultados obtenidos. Dentro de esta categoría se clasifica como Diseño Experimental Puro ya que se busca medir la variable dependiente, en este caso Status (el estado actual del proyecto en Kickstarter) a partir de la manipulación de las demás variables independientes agregando o desagregándolas para comparar los rendimientos obtenidos de los instrumentos de medición y determinar cuáles de ellas finalmente serán tomadas en cuenta.

3.1.4. Descripción del prototipo de investigación

Teniendo como referencia y base principal el décimo antecedente explicado en el Capítulo II, la idea del prototipo final consistió en ensamblar las tres partes básicas de un proyecto: la primera consiste en el tratamiento de la metainformación (en la cual se realizarán, asimismo, tres experimentos independientes), el segundo, en el contenido visual y el último, el contenido textual respectivamente pero con el valor diferenciado de adaptar el modelo general de acuerdo a las variables y conjuntos de datos disponibles para el presente trabajo. Para ello, se representa cada una de las tres partes agrupadas en el marco de trabajo de la Figura 1.1.

3.2. Población y muestra

3.2.1. Población

La población que será considerada para el presente trabajo será de 27,251 proyectos en Kickstarter de la categoría tecnología de todas las subcategorías entre los periodos 2009-2019, en su mayoría del territorio de los Estados Unidos de América.

3.2.2. Muestra

Debido a que se 214 imágenes del contenido visual no pudieron re-dimensionarse, así como 2 proyectos no contaban con descripciones en el contenido textual, se procedió a remover los 216 proyectos incompletos tanto en la metainformación como en las otras bases de datos, resultando finalmente en 27,035 registros en cada una de los tres conjuntos de datos.

Sin embargo, la división en subconjuntos fue distinta en los tres casos y se dio de la siguiente manera:

- Para la metainformación y el contenido textual, el conjunto de datos total de cada uno fue dividido en un subconjunto de entrenamiento (80 %) y uno de prueba (20 %) siguiendo las proporciones dadas en el octavo antecedente.
- Para el contenido visual, el conjunto de datos total fue dividido en tres subconjuntos: entrenamiento (80 %), validación (10 %) y prueba (10 %) siguiendo las proporciones dadas en el décimo antecedente.

Nisi porta lorem mollis aliquam ut porttitor leo. Aenean pharetra magna ac placerat vestibulum. Est placerat in egestas erat imperdiet sed euismod. Velit euismod in pellentesque massa placerat. Enim praesent elementum facilisis leo vel fringilla. Ante in nibh mauris cursus mattis molestie a iaculis. Erat pellentesque adipiscing commodo elit at imperdiet dui accumsan sit. Porttitor lacus luctus accumsan tortor posuere ac ut. Tortor at auctor urna nunc id. A iaculis at erat pellentesque adipiscing commodo elit. La Figura 3.1 y el Cuadro 3.1

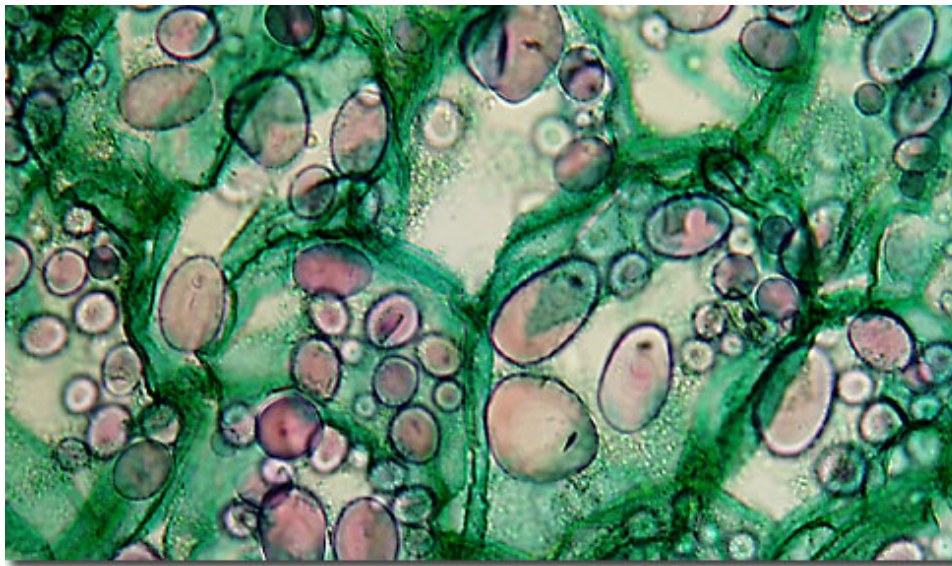


Figura 3.1: Prueba de Figura

3.2.3. Unidad de análisis

La unidad de análisis para el presente trabajo será un proyecto en Kickstarter de la categoría tecnología de cualquier subcategoría entre los periodos 2009-2019 dentro del territorio de los Estados Unidos de América.

3.3. Operacionalización de Variables

Nisi porta lorem mollis aliquam ut porttitor leo. Aenean pharetra magna ac placerat vestibulum. Est placerat in egestas erat imperdiet sed euismod. Velit euismod in pellentesque massa placerat. Enim praesent elementum facilisis leo vel fringilla. Ante in nibh mauris cursus mattis molestie a iaculis. Erat pellentesque adipiscing commodo elit at imperdiet dui accumsan sit. Porttitor lacus luctus accumsan tortor posuere ac ut. Tortor at auctor urna nunc id. A iaculis at erat pellentesque adipiscing commodo elit.

3.4. Instrumentos de medida

Nisi porta lorem mollis aliquam ut porttitor leo. Aenean pharetra magna ac placerat

- muscle and fat cells remove glucose from the blood,
- cells breakdown glucose via glycolysis and the citrate cycle, storing its energy in the form of ATP,
- liver and muscle store glucose as glycogen as a short-term energy reserve,
- adipose tissue stores glucose as fat for long-term energy reserve, and
- cells use glucose for protein synthesis.

3.5. Técnicas de recolección de datos

Nisi porta lorem mollis aliquam ut porttitor leo. Aenean pharetra magna ac placerat vestibulum. Est placerat in egestas erat imperdiet sed euismod. Velit euismod in pellentesque massa placerat. Enim praesent elementum facilisis leo vel fringilla. Ante in nibh mauris cursus mattis molestie a iaculis. Erat pellentesque adipiscing commodo elit at imperdiet dui accumsan sit. Porttitor lacus luctus accumsan tortor posuere ac ut. Tortor at auctor urna nunc id. A iaculis at erat pellentesque adipiscing commodo elit.

L^AT_EX is great at typesetting mathematics. Let X_1, X_2, \dots, X_n be a sequence of independent and identically distributed random variables with

$$S_n = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{1}{n} \sum_i^n X_i \quad (\text{Ecuación 3.1})$$

La Ecuación [Ecuación 3.1](#) denote their mean. Then as n approaches infinity, the random variables

$$\sqrt{n}(S_n - \mu)$$

converge in distribution to a normal $\mathcal{N}(0, \sigma^2)$.

3.6. Técnicas para el procesamiento y análisis de la información

Nisi porta lorem mollis aliquam ut porttitor leo. Aenean pharetra magna ac placerat vestibulum. Est placerat in egestas erat imperdiet sed euismod. Velit euismod in pellentesque massa placerat. Enim praesent elementum facilisis leo vel fringilla. Ante in nibh mauris cursus mattis molestie a iaculis. Erat pellentesque adipiscing commodo elit at imperdiet dui accumsan sit. Porttitor lacus luctus accumsan tortor posuere ac ut. Tortor at auctor urna nunc id. A iaculis at erat pellentesque adipiscing commodo elit.

You can make lists with automatic numbering ...

1. Like this,
2. and like this.

... or bullet points ...

- Like this,
- and like this.

3.7. Cronograma de actividades y presupuesto

Nisi porta lorem mollis aliquam ut porttitor leo. Aenean pharetra magna ac placerat vestibulum. Est placerat in egestas erat imperdiet sed euismod. Velit euismod in pellentesque massa placerat. Enim praesent elementum facilisis leo vel fringilla. Ante in nibh mauris cursus mattis molestie a iaculis. Erat pellentesque adipiscing commodo elit at imperdiet dui accumsan sit. Porttitor lacus luctus accumsan tortor posuere ac ut. Tortor at auctor urna nunc id. A iaculis at erat pellentesque adipiscing commodo elit.

Item	Quantity
Widgets	42
Gadgets	13

Tabla 3.1: An example table.

Capítulo 4

DESARROLLO DEL EXPERIMENTO

4.1. X

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn? Kjift ”not at all!...

4.2. Y

Nisi porta lorem mollis aliquam ut porttitor leo. Aenean pharetra magna ac placerat vestibulum. Est placerat in egestas erat imperdiet sed euismod. Velit euismod in pellentesque massa placerat. Enim praesent elementum facilisis leo vel fringilla. Ante in nibh mauris cursus mattis molestie a iaculis. Erat pellentesque adipiscing commodo elit at imperdiet dui accumsan sit. Porttitor lacus luctus accumsan tortor posuere ac ut. Tortor at auctor urna nunc id. A iaculis at erat pellentesque adipiscing commodo elit.

Item	Quantity
Widgets	42
Gadgets	13

Tabla 4.1: An example table.

4.3. Z

Nisi porta lorem mollis aliquam ut porttitor leo. Aenean pharetra magna ac placerat vestibulum. Est placerat in egestas erat imperdiet sed euismod. Velit euismod in pellentesque massa placerat. Enim praesent elementum facilisis leo vel fringilla. Ante in nibh mauris cursus mattis molestie a iaculis. Erat pellentesque adipiscing commodo elit at imperdiet dui accumsan sit. Porttitor lacus luctus accumsan tortor posuere ac ut. Tortor at auctor urna nunc id. A iaculis at erat pellentesque adipiscing commodo elit.

El paper es citado y el otro paper .

Capítulo 5

ANÁLISIS Y DISCUSIÓN DE RESULTADOS

5.1. X

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn? Kjift ”not at all!...

5.2. Y

Nisi porta lorem mollis aliquam ut porttitor leo. Aenean pharetra magna ac placerat vestibulum. Est placerat in egestas erat imperdiet sed euismod. Velit euismod in pellentesque massa placerat. Enim praesent elementum facilisis leo vel fringilla. Ante in nibh mauris cursus mattis molestie a iaculis. Erat pellentesque adipiscing commodo elit at imperdiet dui accumsan sit. Porttitor lacus luctus accumsan tortor posuere ac ut. Tortor at auctor urna nunc id. A iaculis at erat pellentesque adipiscing commodo elit.

Item	Quantity
Widgets	42
Gadgets	13

Tabla 5.1: An example table.

5.3. Z

Nisi porta lorem mollis aliquam ut porttitor leo. Aenean pharetra magna ac placerat vestibulum. Est placerat in egestas erat imperdiet sed euismod. Velit euismod in pellentesque massa placerat. Enim praesent elementum facilisis leo vel fringilla. Ante in nibh mauris cursus mattis molestie a iaculis. Erat pellentesque adipiscing commodo elit at imperdiet dui accumsan sit. Porttitor lacus luctus accumsan tortor posuere ac ut. Tortor at auctor urna nunc id. A iaculis at erat pellentesque adipiscing commodo elit.

Capítulo 6

CONCLUSIONES Y RECOMENDACIONES

6.1. Conclusiones

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn? Kjift ”not at all!...

6.2. Recomendaciones

Nisi porta lorem mollis aliquam ut porttitor leo. Aenean pharetra magna ac placerat vestibulum. Est placerat in egestas erat imperdiet sed euismod. Velit euismod in pellentesque massa placerat. Enim praesent elementum facilisis leo vel fringilla. Ante in nibh mauris cursus mattis molestie a iaculis. Erat pellentesque adipiscing commodo elit at imperdiet dui accumsan sit. Porttitor lacus luctus accumsan tortor posuere ac ut. Tortor at auctor urna nunc id. A iaculis at erat pellentesque adipiscing commodo elit.

Anexos

Anexos A

Anexo I: Matriz de Consistencia

PROBLEMAS	OBJETIVOS	HIPÓTESIS
Problema General	Objetivo General	Hipótesis General
Bajos niveles de precisión de modelos entrenados de Aprendizaje Automático para cualquier categoría para predecir estado de financiamiento de proyectos de tecnología.	Construir modelo(s) de Aprendizaje Automático entrenado(s) para predecir correctamente proyectos de tecnología con nivel de precisión aceptable.	El modelo entrenado de Aprendizaje Automático logrará predecir correctamente proyectos de tecnología con nivel de precisión aceptable.
Problemas Específicos	Objetivos Específicos	Hipótesis Específicas
Variables de proyectos no normalizadas y varianzas altas.	Normalizar variables de proyectos y reducir niveles altos de varianza.	Las variables de los proyectos descargados se normalizarán y se reducirán los niveles altos de varianza.
Datos faltantes o incompletos de proyectos.	Eliminar datos faltantes o incompletos de proyectos.	Los datos faltantes o incompletos de los proyectos serán eliminados.
Parámetros de modelos no ajustados.	Ajustar parámetros de modelos.	Los parámetros de los modelos usados serán ajustados.
Sobreajuste de aprendizaje de modelos y clasificación incorrecta de las dos clases del estado final de financiamiento (éxito o fracasado).	Evitar sobreajuste de aprendizaje de modelos.	Se evitará el sobreajuste de aprendizaje de modelos para clasificar correctamente las dos clases del estado final de financiamiento.
Predicción incorrecta de estado de financiamiento de un proyecto tecnológico.	Predecir correctamente el estado final de financiamiento de cualquier proyecto tecnológico (éxito o fracaso).	El estado final de financiamiento de cualquier proyecto tecnológico será predicho correctamente.

Tabla A.1: Matriz de consistencia. Fuente: Elaboración propia

Anexos B

Anexo II: Resumen de Papers investigados

Tipo	N°	Título	Autor	Año	País	Fuente
Problema	1	Copper price estimation using bat algorithm	Dehghani Bogdanovic	2018	United Kingdom	Resources Policy
	2	Alternative techniques for forecasting mineral commodity prices	Cortez, Saydam, Coulton, Sammut	2018	Netherlands	International Journal of Mining Science and Technology
Propuesta	3	Prediction of the crude oil price thanks to natural language processing applied to newspapers	Trastour, Genin, Morlot	2016	USA	Standfort University ML repository
	4	Stock Price Prediction Using Deep Learning	Tipirisetty	2018	USA	Master's Theses San Jose State University
	5	Deep Learning for Stock Prediction Using Numerical and Textual Information	Akita, R., Yoshihara, A., Matsubara, T., Uehara, K.	2016	USA	2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS)
Técnica	6	Stock Prices Prediction using the Title of Newspaper Articles with Korean Natural Language Processing	Yun, Sim, Seok	2019	Japan	2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC)
	7	A Method of Optimizing LDA Result Purity Based on Semantic Similarity	Jingrui, Z., Qinglin, W., Yu, L., Yuan, L.	2017	China	2017 32nd Youth Academic Annual Conference of Chinese Association of Automation (YAC)
	8	Qualitative Stock Market Predicting with Common Knowledge Based Nature Language Processing: A Unified View and Procedure	Rao, D., Deng, F., Jiang, Z., Zhao, G.	2015	USA	2015 7th International Conference on Intelligent Human-Machine Systems and Cybernetics
	9	Fuzzy Bag-of-Words Model for Document Representation	Zhao, R., Mao, K.	2018	USA	IEEE Transactions on Fuzzy Systems (Volume: 26 , Issue: 2 , April 2018)

Tabla B.1: Cuadro Resumen de Papers investigados. Fuente: Elaboración propia

BIBLIOGRAFÍA

- Asociación de Emprendedores de Perú. (2018). Avances y limitaciones del emprendimiento peruano. <https://asep.pe/index.php/avances-limitaciones-emprendimiento-peruano/>
- Dehghani, H. & Bogdanovic, D. (2018). Copper price estimation using bat algorithm. *Resources Policy*, 55, 55-61.
- Gollapudi, S. (2016). *Practical machine learning*. Packt Publishing Ltd.
- Goyal, P., Pandey, S. & Jain, K. (2018). Deep learning for natural language processing. *Deep Learning for Natural Language Processing: Creating Neural Networks with Python [Berkeley, CA]: Apress*, 138-143.
- Kamath, R. S. & Kamat, R. K. (2018). Supervised Learning Model For Kickstarter Campaigns With R Mining. *International Journal of Information Technology, Modeling and Computing (IJITMC)*, 4(1). <https://doi.org/10.5281/zenodo.1228716>
- Kickstarter. (s.f.-a). *Acerca de nosotros: Kickstarter*. <https://www.kickstarter.com/about?ref=global-footer>
- Kickstarter. (s.f.-b). *Financiamiento: Kickstarter*. <https://www.kickstarter.com/help/handbook/funding?lang=es>
- Kickstarter. (s.f.-c). *Prensa: Kickstarter*. <https://www.kickstarter.com/press?ref=hello>
- Redacción Gestión. (2015). Emprendimiento en el Perú se origina más por oportunidades de negocio que por desempleo. *Diario Gestión*. <https://gestion.pe/economia/emprendimiento-peru-origina-oportunidad-negocio-desempleo-80578>
- Redacción Gestión. (2018). Perú es el tercer país con mayor cantidad de emprendimientos en fase temprana a nivel mundial. *Diario Gestión*. <https://gestion.pe/economia/peru-tercer-pais-mayor-cantidad-emprendimientos-fase-temprana-nivel-mundial-240264>
- Royal Society Working Group. (2017). *Machine learning: the power and promise of computers that learn by example* (inf. téc.). Technical report.
- Sandoval, L. (s.f.). Barreras del Emprendedor ¿Por qué cuesta tanto hacerlo? <https://www.emprender-facil.com/es/barreras-del-emprendedor/>
- Solidaridad Latina. (s.f.). ¿Cómo funciona el crowdfunding en Latinoamérica? <https://solidaridadlatina.com/actualizacion/como-functiona-crowdfunding-latinoamerica/>

The Hustle. (2019). What are your chances of successfully raising money on Kickstarter? <https://thehustle.co/crowdfunding-success-rate>

Universo Crowdfunding. (s.f.). ¿Qué es el crowdfunding? <https://www.universocrowdfunding.com/que-es-el-crowdfunding/>