



datascience.pe



Expositor

Sergio Aarón Abanto León



978941882



sergio.abanto.13@gmail.com



<https://www.linkedin.com/in/sabantol/>

Tema

Análisis sobre la labor del PRONABEC respecto al programa Beca 18 y su impacto social, empleando técnicas de Procesamiento de Lenguaje Natural

**Tema: Análisis sobre la labor del PRONABEC respecto al programa Beca 18 y su impacto social,
empleando técnicas de Procesamiento de Lenguaje Natural.**

RESUMEN

En el país existe falta de oportunidades de educación, por ello el estado implementa programas sociales que permitan recortar las brechas de oportunidades en materia de educación que existan para la población que vive en condiciones de pobreza y pobreza extrema.

Para contribuir a combatir este problema social se creó el PRONABEC, la cual es un programa dependiente del Ministerio de Educación (MINEDU) que tiene como objetivo mejorar la igualdad de oportunidades en el acceso a la educación superior, buscando garantizar la permanencia y culminación de los estudios de las personas beneficiarias, quien en su totalidad son personas que viven en condiciones de pobreza y pobreza extrema.

La finalidad de este trabajo es presentar la implementación de un modelo de análisis de sentimientos basado en técnicas de NLP sobre la percepción que tienen los becarios y otras personas de la labor que realiza el PRONABEC respecto a “Beca 18” y el impacto social de dicho programa.

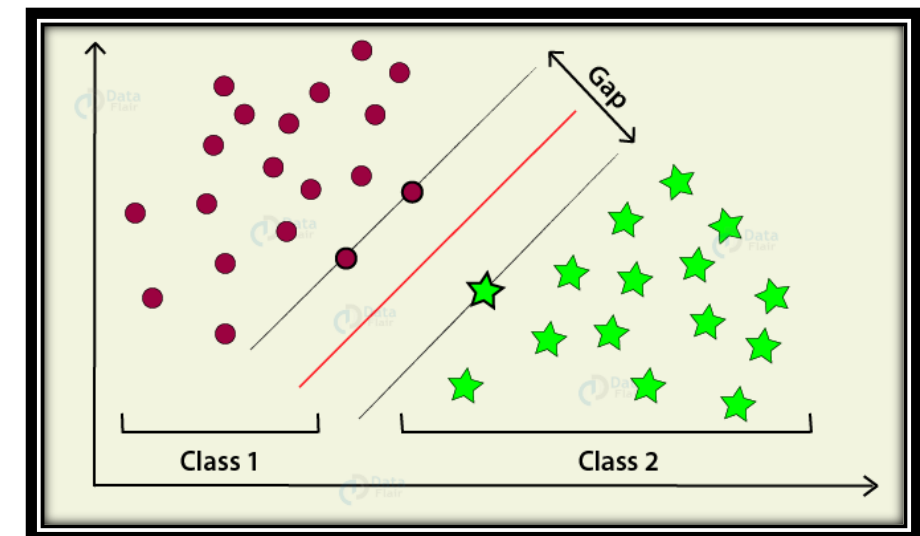
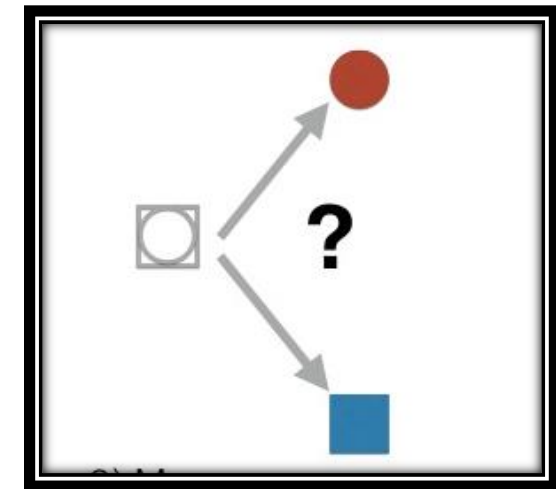


**Tema: Análisis sobre la labor del PRONABEC respecto al programa Beca 18 y su impacto social,
empleando técnicas de Procesamiento de Lenguaje Natural.**

INTRODUCCIÓN

Este proyecto tiene el propósito dar a conocer la percepción que tienen los becarios y las personas respecto a la gestión que tiene el PRONABEC sobre el programa “Beca 18” y el impacto social que genera; para ello, se realizará un análisis de sentimientos basado en técnicas de NLP con el fin de analizar las opiniones que tienen las personas respecto a “Beca 18”; luego de ello, se espera que esto sea de utilidad para notificar a las instituciones correspondientes sobre el nivel de aprobación que tiene dicho programa social, para que así, esto pueda contribuir al desarrollo de futuras políticas públicas que busquen mejorar el desempeño de dicho programa y su logro de objetivos.

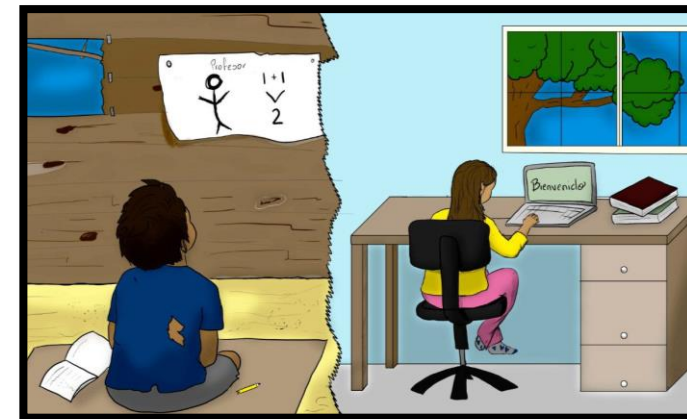
Al final del proyecto se implementará seis clasificadores binarios (Logística, Naive Bayes Bernoulli, Support Vector Machine, Nu-Support Vector Machine, Random Forest y Gradient Boosting) que permitan identificar tanto los mensajes de aprobación y desaprobación que existan alrededor de la gestión del PRONABEC respecto al programa “Beca 18”, cabe especificar, que se evaluará el rendimiento predictivo de dichos clasificadores binarios para identificar mensajes de aprobación y desaprobación.



Tema: Análisis sobre la labor del PRONABEC respecto al programa Beca 18 y su impacto social, empleando técnicas de Procesamiento de Lenguaje Natural.

SITUACIÓN PROBLEMÁTICA

En la actualidad en el Perú las desigualdades se han convertido en barreras invisibles que enfrentan las personas para poder lograr un mayor desarrollo personal, económico, familiar, entre otros. Este escenario, sin duda dificulta la convivencia y la paz en la sociedad, dado que, existen grupos de personas que se sienten excluidos de alcanzar todos los beneficios y oportunidades que brinda el sistema social, solo por el simple hecho de haber nacido en una realidad “diferente”. Según Vega (2017), la desigualdad en una sociedad se expresa en los accesos inequitativos a las diferentes oportunidades que pueden permitirnos un mayor desarrollo humano. Por lo tanto, es importante que la sociedad pueda “asegurar el acceso a la educación se considera como una vía para abatir la pobreza y la desigualdad social” (Márquez, 2016, p. 3).



Ante esta problemática, por parte del estado surgen iniciativas como el Programa Nacional de Becas y Crédito Educativo (PRONABEC) la cual es una entidad adscrita al Ministerio de Educación, con la finalidad de reducir la brecha de oportunidades de acceder a una educación superior de calidad que presentan las personas de bajos recursos económicos. De acuerdo, con la página web del PRONABEC, su función principal es:

“Otorgar becas y créditos educativos, a través de concursos públicos, a peruanos talentosos de escasos recursos económicos, brindándoles acceso y permanencia a una educación superior de calidad, hasta su culminación. Además, desarrollamos un modelo de acompañamiento integral durante todo el ciclo de vida del beneficiario, mediante una gestión moderna, eficiente, transparente y orientada al servicio” (PRONABEC, 2021).

PRONABEC

 **Data
Science**

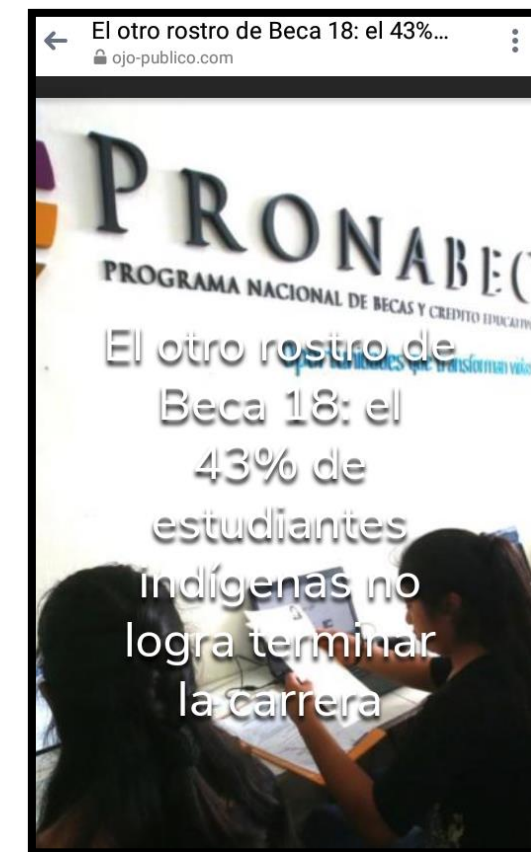
De acuerdo al PRONABEC, al año 2015 se habían otorgado aproximadamente 60'000 becas a nivel nacional en sus distintos componentes, promoviendo así, una mayor inclusión en materia de acceso a una educación superior de calidad para las personas que viven en condición de pobreza y pobreza extrema.

Sin embargo, a pesar de los buenos propósitos que tiene el PRONABEC, esta institución ha presentado inconvenientes al ejecutar sus diversos programas sociales, y en lo que respecta al programa “Beca 18” se puede identificar los siguientes problemas:

- ✓ Se ha evidenciado una alta deserción y falta de permanencia de becarios.
- ✓ Se deben mejorar los esfuerzos para ofrecer una información clara y detallada sobre el proceso de postulación y las obligaciones de los becarios, para evitar trámites innecesarios y confusos.
- ✓ Existen por parte de los becarios problemas de adaptación a nuevos estilos de vida, urbanos y modernos.

Es ante esta disyuntiva que nos cuestionamos aspectos como: ¿Cuál es la percepción de los becarios respecto a la labor que realiza el PRONABEC sobre el programa “Beca 18”? ¿La población considerará que “Beca 18” ha contribuido a reducir los problema de acceso a una educación superior de mayor calidad en el país? ¿Realmente se puede decir que “Beca 18” ha ayudado a que los jóvenes terminen sus estudios superiores de forma satisfactoria?

Así, esta problemática motiva a crear mecanismos que permita identificar el nivel de aprobación que existe sobre la labor del PRONABEC en el programa “Beca 18”, para poder luego con esta evidencia buscar implementar políticas públicas que ayuden a mejorar el desempeño de dicho programa y su logro de objetivos.



PROBLEMA GENERAL

¿Cuál es la percepción que tienen las personas en general sobre la labor que realiza el PRONABEC con respecto a “Beca 18” y el impacto social de dicho programa?

OBJETIVOS PRINCIPALES

Determinar la percepción de las personas en general sobre la gestión del PRONABEC respecto a Beca 18 y el impacto social que tiene dicho programa.

OBJETIVOS ESPECÍFICOS

- ✓ Recopilar datos consistentes sobre las opiniones que tienen las personas respecto a la labor del PRONABEC en la ejecución de Beca 18.
- ✓ Formatear y normalizar las opiniones de las personas respecto a la labor del PRONABEC en la ejecución de Beca 18.
- ✓ Dividir y codificar las opiniones recopiladas en datos de entrenamiento y prueba para entrenar y validar los modelos de clasificación binaria propuestos.
- ✓ Medir el rendimiento de los modelos obtenidos.
- ✓ Comparar el rendimiento de los modelos obtenidos
- ✓ Validar los resultados y evaluar los modelos.

ANTECEDENTES

1. Un estudio cualitativo de los becarios del programa beca 18.

AUTOR: Julio Cotler.

RESUMEN: Estudio cualitativo realizado entre agosto y octubre de 2015 sobre las estrategias que adoptó un pequeño grupo de jóvenes para acceder al Programa Beca 18 e insertarse en los centros de educación superior en Lima, así como los problemas que encontraron y los resultados que derivaron de dichas estrategias. El estudio permite sugerir algunas hipótesis generales relativas al desenvolvimiento del programa y ofrecer algunas recomendaciones destinadas a mejorar su ejecución.

CONCLUSIONES:

- ✓ Es necesario que el PRONABEC se esmere en brindar y difundir masivamente información clara y accesible sobre Beca 18 a los sectores pobres de la sociedad.
- ✓ Es crucial que el PRONABEC y los centros educativos que reciben becarios cuenten con un protocolo de acogida y recibimiento.
- ✓ Es importante asegurar el bienestar psicológico de los becarios, para lo cual el PRONABEC y los centros de estudios deben prestarles la ayuda necesaria mediante un sistema de tutorías y servicios psicopedagógicos.
- ✓ El programa Beca 18 tiene un impacto positivo en los sectores pobres que los jóvenes reconocen, a pesar de las muchas dificultades que enfrentan para gozar de sus beneficios.

ANTECEDENTES

2. Análisis de sentimientos: Aplicación sobre textos en redes sociales.

AUTOR: *Pedro Diego Leonardo.*

RESUMEN: La intención de este estudio es obtener tendencias de opinión pública sobre un producto de una empresa, mediante la implementación de técnicas de análisis de sentimientos sobre mensajes de clientes en redes sociales.

PREPROCESAMIENTO: Normalización, Tokenización y Stopwords.

CONCLUSIÓN: El impacto que genera el preprocesamiento de los datos es tanto o más importante que la elección de modelo a utilizar. Los resultados para los distintos modelos, sin ser idénticos, no distan significativamente entre sí; esto nos daría la pauta que es posible obtener resultados similares con más de un modelo.

RESULTADO: Accuracy Neural Network: 83%; Accuracy SVM: 86%; Accuracy Naive Bayes: 71%; Accuracy Random Forest: 73%

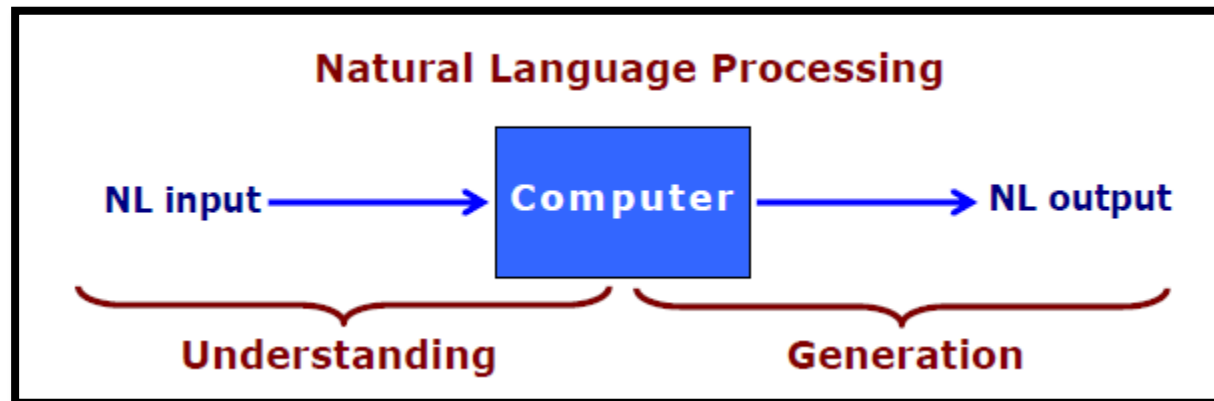
**Tema: Análisis sobre la labor del PRONABEC respecto al programa Beca 18 y su impacto social,
empleando técnicas de Procesamiento de Lenguaje Natural.**

MARCO TEÓRICO

EL PROCESAMIENTO DEL LENGUAJE NATURAL

El Procesamiento del Lenguaje Natural (PLN) es un subcampo de la inteligencia artificial y lingüística dedicado a hacer que las computadoras "entiendan" las declaraciones escritas en los lenguajes humanos (Kao y Poteet, 2007).

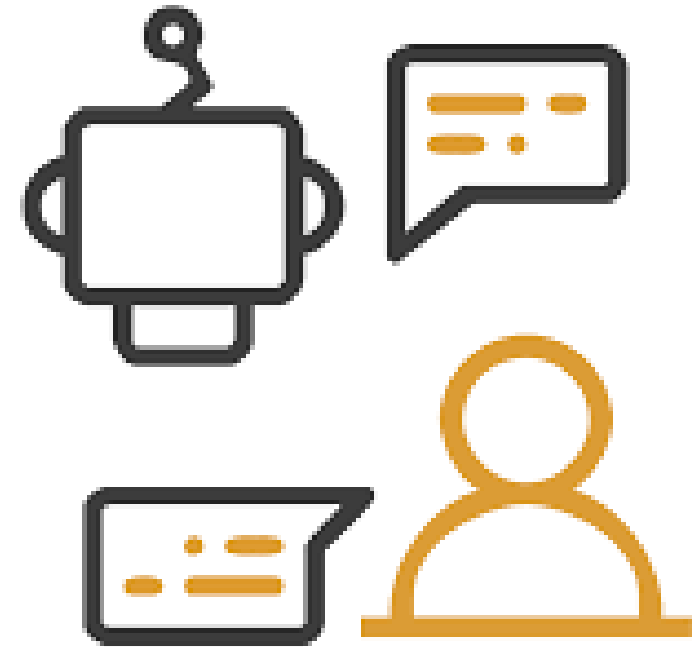
El Procesamiento del Lenguaje Natural es el campo de conocimiento de la Inteligencia Artificial que se ocupa de investigar la manera de comunicar las máquinas con las personas mediante el uso de lenguas naturales, como el español, el inglés o el chino (Moreno, 2018).



APLICACIONES DEL PROCESAMIENTO DEL LENGUAJE NATURAL

Las principales aplicaciones prácticas del Procesamiento de Lenguaje Natural (PLN) son las siguientes:

- ✓ Minería de datos.
- ✓ Sistemas de búsqueda de respuestas.
- ✓ Corrección ortográfica.
- ✓ Síntesis de voz.
- ✓ Sistemas de diálogo.
- ✓ Reconocimiento de voz.
- ✓ Análisis de sentimientos.
- ✓ Generación de resúmenes automáticos.
- ✓ Traducción automática.
- ✓ Recuperación y extracción de información.



ANÁLISIS DE SENTIMIENTOS

“El análisis de sentimiento es el proceso de determinar el tono emocional que hay detrás de una serie de palabras, y se utiliza para intentar entender las actitudes, opiniones y emociones expresadas en una mención online” (Bannister, 2015).

“El análisis de sentimiento se basa en la evaluación automatizada de los comentarios de los usuarios para determinar si un texto se ha escrito con una intención positiva o negativa. Para ello, se usa la minería de textos, es decir, el análisis automático de textos escritos en lenguaje natural” (IONOS, 2020).



PREPROCESAMIENTO DE LA INFORMACIÓN

El preprocesamiento es un paso necesario previo al entrenamiento de los modelos. El mismo permite limpiar lo que se conoce como 'Ruido del texto', estandarizar y convertir los datos de opiniones en estructurados, logrando así, un mejor entendiendo de las máquinas respecto a los datos, y por ende, una mejor clasificación.

Hacer un buen preprocesamiento asegurara el correcto desempeño de los clasificadores. Dentro del preprocesamiento tenemos variedad de posibles intervenciones de las cuales distinguimos:

- ✓ Normalización: Trata de unificar términos con mismo significado para darles el mismo tratamiento. Ej: Vehículo, Automóvil, Auto, Coche.

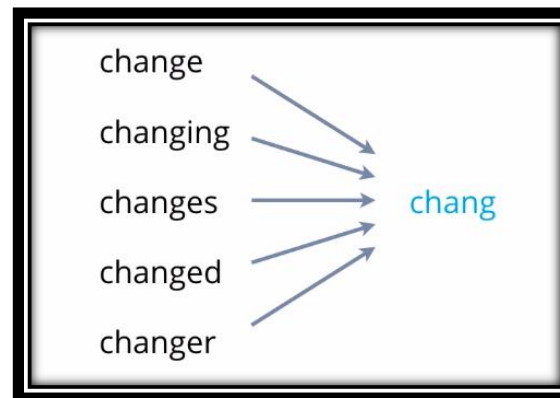
Ej: la eliminación de URLs, la eliminación de caracteres especiales como signos de puntuación, fechas, números, conversión de emoticones en términos mayúsculas, entre otros.

PREPROCESAMIENTO DE LA INFORMACIÓN

- ✓ Tokenización: Consiste en dividir las palabras de un texto en unidades léxicas para estructurar los datos. Cada uno de estos “Tokens” o unidades luego podrán ser tratados por separado.

Ej: “El vehículo se dirigía a toda velocidad” → [El] [vehículo] [se] [dirigía] [a] [toda] [velocidad].

- ✓ Stopwords: Las stopwords son palabras de elevada frecuencia, que no aportan significado a la oración. Usualmente artículos, preposiciones, contracciones y conectores forman parte de la lista. Al no aportar al significado son directamente quitados del texto.
- ✓ Stemming: Proceso por el cual se convierte cada palabra a su base léxica o raíz. No necesariamente cumple las leyes del idioma.



BAG OF WORDS

- ✓ Describe la ocurrencia de palabras dentro de un documento o una colección de documentos (corpus).
- ✓ Desarrolla un vocabulario de las palabras y una medida de su presencia.
- ✓ Ejemplo:

This is the best book ever. I loved the book and highly recommend it!!!

**{'This': 1, 'is': 1, 'the': 2 , 'best': 1 , 'book': 2,
'ever': 1, 'I':1 , 'loved':1 , 'and': 1 , 'highly': 1,
'recommend': 1 , 'it': 1 }**

BAG OF WORDS

La salida de un Bag of Words puede verse así:

	10	100	12	15	1984	20	30	40	451	50	...	wrong	wrote	year	years	yes	yet	you	young	your	yourself
0	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	1	0
1	0	0	0	0	0	0	0	0	0	0	...	0	0	0	1	0	0	1	0	0	0
2	0	0	0	0	0	0	0	0	0	0	...	0	0	0	1	0	0	2	0	0	0
3	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	...	0	1	0	0	0	0	3	0	1	0

TF-IDF: MÁS FORMAS DE TRANSFORMAR TEXTO

- ✓ Term frequency - inverse document frequency.
- ✓ Permite determinar las palabras más importantes en cada documento.
- ✓ **TF: term frequency:** Con qué frecuencia aparece una palabra determinada dentro de un documento en el corpus.
- ✓ **Inverse document frequency:** Relación logarítmica entre el número total de documentos y el número de documentos que contienen una palabra específica. Se usa para calcular el peso de las palabras que no ocurren con frecuencia.

TF-IDF: MÁS FORMAS DE TRANSFORMAR TEXTO

*TfIdf = term frequency * inverse document frequency*

$$w_{i,j} = tf_{i,j} * \log\left(\frac{N}{df_i}\right)$$

w_{i,j}: peso tf – idf para el token i en el documento j

tf_{i,j}: número de ocurrencias del token i en el documento j

df_i: número de documentos que contiene el token i

N: número total de documentos

- ✓ **Bag of Word (BOW) tradicional no tiene en cuenta la longitud de un documento, TfIdf si lo hace.**
- ✓ **Dado que penaliza las palabras frecuentes, hay menos necesidad de tratar explícitamente las palabras vacías.**
- ✓ **Muy útil en consultas de búsqueda y recuperación de información para clasificar la relevancia de los resultados devueltos.**

TF-IDF: MÁS FORMAS DE TRANSFORMAR TEXTO

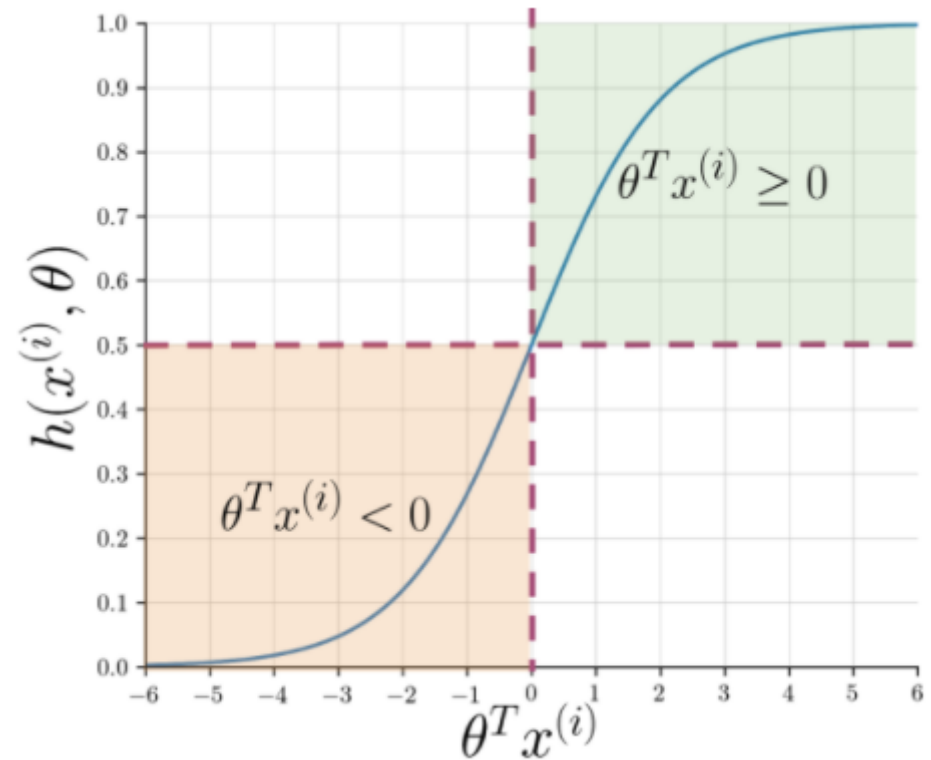
La salida de un Tf-Idf puede verse así:

	about	after	again	airline	all	am	americanair	amp	an	and	...	was	we	what	when	why	will	with	would	you	your
0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	...	0.0	0.0	0.668165	0.0	0.0	0.0	0.0	0.0	0.000000	0.000000
1	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	...	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.32904	0.000000
2	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	...	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	0.000000
3	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.431149	0.0	0.000000	...	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	0.332355
4	0.494872	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.279754	...	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	0.000000

REGRESIÓN LOGÍSTICA

- ✓ La Regresión Logística es una técnica de estadística multivariante que nos permite estimar la relación existente entre una variable dependiente no continua, en particular dicotómica y un conjunto de variables independientes continuas o discretas.
- ✓ La regresión logística hace uso de la función sigmoidea que genera una probabilidad entre 0 y 1. La función sigmoidea con algún parámetro de peso θ y algunos input $x^{(i)}$ se define de la siguiente manera:

$$h(x^{(i)}, \theta) = \frac{1}{1 + e^{-\theta^T x^{(i)}}}$$



MATRIZ DE CONFUSIÓN

Tabla cruzada que relaciona la variable objetivo real u observada con la variable objetivo pronosticada. Pueden calcularse diversas métricas como Verdaderos Positivos (Sensibilidad), Falsos Positivos, Verdaderos Negativos (Especificidad), Falsos Negativos, etc.

		Predicted Class	
		FALSE	TRUE
Observed Class	FALSE	TN	FP
	TRUE	FN	TP

TN: True Negative

FP: False Positive

FN: False Negative

TP: True Positive

$$Accuracy = \frac{TN + TP}{TN + FP + FN + TP}$$

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

**Tema: Análisis sobre la labor del PRONABEC respecto al programa Beca 18 y su impacto social,
empleando técnicas de Procesamiento de Lenguaje Natural.**

METODOLOGÍA

METODOLOGÍA

Para comprender la solución generada en este trabajo, debemos entender las etapas por las cuales se desarrolla. Para ello se utilizará la metodología KDD (Knowledge Discovery in Database), la cual consta de las siguientes fases:



NLTK

METODOLOGÍA



Algunas especificaciones:

- ✓ Lenguaje utilizado: Python.
- ✓ Entorno de desarrollo: Spyder.
- ✓ Librerías: NLTK, SKLEARN, PANDAS



PYTHON



FASES DE LA METODOLOGÍA

1. SELECCIÓN DE DATOS

Para el presente proyecto se utilizará una muestra de 160 opiniones de personas respecto al programa “Beca 18”. Estas opiniones fueron recopiladas de las siguientes trabajos y artículos:

- a. “Un estudio cualitativo de los becarios del programa beca 18”, realizado por Julio Cotler (2016).
- b. “Motivaciones de los postulantes seleccionados e ingresantes de Beca 18 que deciden no seguir la beca”, realizado por PRONABEC (2015).
- c. “Implementación de programas de inclusión social en territorios con población vulnerable: ¿Cómo está cambiando Beca 18 la vida de los y las jóvenes del valle de los ríos Apurímac, Ene y Mantaro (VRAEM)?”; realizado por los autores: Gabriela Guerrero, Vanessa Rojas, Santiago Cueto, Jimena Vargas y Sayuri Leandro (2019).
- d. “Beca 18 testimonios”, realizado por PRONABEC (2015).
- e. “El otro rostro de Beca 18: el 43% de estudiantes indígenas no logra terminar la carrera”, artículo web de Ojo Público redactado por Alicia Tovar y Gianfranco Huamán (2020).
- f. “Acceso a la educación superior de poblaciones vulnerables: el caso Beca 18, modalidad albergues del PRONABEC – Perú, 2012 – 2016”, tesis de Maestría realizada por Luz Espinoza.

Adicionalmente, se extrajo opiniones del fanpage de Facebook del PRONABEC.

FASES DE LA METODOLOGÍA

1. SELECCIÓN DE DATOS

Los opiniones de los becarios fueron recolectados en un archivos Excel, para su posterior importación al programa Python.

	A	B	C
64	Lamentablemente, el Estado tiene una incapacidad para salir de la caja y hacer una beca pensada con un enfoque intercultural. Para lograr una verdadera igualdad, tenemos que tener becas de acuerdo a las características de los estudiantes.	Negativo	0
65	La entidad necesita más presupuesto para cubrir los cursos que desaprueben los becarios. Ese ya es un factor de posible pérdida de la beca porque, si sus familias pertenecen a los últimos quintiles de pobreza, no tiene ningún sentido que se pretenda que puedan pagar esos créditos que les faltan para acabar la carrera.	Negativo	0
66	Cuando planifican que los estudiantes acaben en cinco años, eso es lo ideal. Pero ni los universitarios que salen de las familias más privilegiadas y con mejores oportunidades acaban en ese tiempo.	Negativo	0
67	No puede ser obstáculo para los becarios el sentirse solos, lejos de su comunidad. Son jóvenes que tenían una vida más colectiva, con diversas tareas, pero la vida en la ciudad es más individualista.	Negativo	0
68	Entiendo que dentro de la normativa actual debería ser posible que los chicos certifiquen su propia lengua. Como derechos lingüísticos de los pueblos indígenas, ahí hay algo que no está considerando el programa de Beca 18.	Negativo	0
69	El mes de abril decía que todavía no depositaban a los estudiantes y que van a depositar dentro de 15 días y ese mismo día teníamos que viajar a Lima porque las clases ya empezaban, entonces yo dije yo no tengo nada, mi papá está en el hospital y mi mama es la que trabaja y no gana nada, yo eso pensaba ¿Cómo hago? Mejor renuncio digo así.	Negativo	0
70	Este programa me abrió la puerta hacia el estudio, brindándome el apoyo que tanto necesitaba y la estimulación para culminarlo.	Positivo	1
71	Beca 18 es una de las mejores cosas que me pudo haber pasado. Me siento una de las personas más dichosas que existen en este planeta. Conocí a varias personas que también están beneficiadas con este apoyo del gobierno, sus historias son muy interesantes y cada uno tiene una trama particular. No podría decir que una historia es mejor que otra, lo único que puedo expresar es que gracias al Programa hemos logrado unirnos y conocernos más. Nosotros los becarios somos una gran familia, que a pesar de las diferencias tenemos una meta en común: "Construir un mejor Perú".	Positivo	1
72	El Programa Nacional Beca 18 me ha dado la oportunidad de acceder a una educación superior de calidad y así dar un gran paso para el cumplimiento de mi meta: ser un exitoso empresario. Así sacar a mi familia de la pobreza en la que se encuentra.	Positivo	1

FASES DE LA METODOLOGÍA

1. SELECCIÓN DE DATOS

POSITIVOS

A los jóvenes esta beca ya les cambió la vida totalmente, terminen o no terminen, ya hay un impacto positivo sobre sus condiciones de origen.
En el caso de ellos tanto la educación como la economía ya se está mejorando porque a mitad de carrera ellos ya pueden trabajar o practicar, y no va a ser un subempleo, va a ser un empleo. Entonces, yo creo que ya se les cambió la vida a todos ellos.
Yo creo que el principal impacto es personal. Es un su formación personal. Este es un modelo que vale la pena ser aplicado a otros becarios. Hay algunos chicos que sí vienen con la intención de iniciar su carrera y terminarla con éxito. Beca 18 les da la oportunidad de que estudien una carrera en las mejores universidades, en institutos reconocidos, cosa que no hubieran tenido si no hubieran salido de sus regiones.
Las enfermedades no eran problema porque los chicos tenían su seguro de MAPFRE y los atendían en las mejores clínicas con todas las comodidades.
Todo el camino andado no habría sido posible sin el apoyo del Pronabec del Ministerio de Educación. Antes de postular al concurso Beca 18 del Pronabec trabajaba. Gané la beca y me dediqué a la carrera y a realizar actividades extracurriculares de voluntariado social y ambiental.
La Beca 18 del Pronabec representa para mí y para mi familia una bendición.
Estudiar con Beca 18 era mi única forma de acceder a la PUCP. Ahora, que estoy a punto de concluir mi carrera, siento que aquí, en la universidad, ha continuado mi formación como ciudadana.
Ya en Tayabamba, orientado por mi profesor, postule a Beca 18 del Pronabec y, gracias a esta oportunidad académica que me otorgó el Estado, ahora soy un mecánico en construcciones metálicas luego de estudiar en el Senati.
Gracias a Beca 18 del Pronabec he podido trasladar todas mis ideas y hacerlas realidad. Cuando uno tiene metas y sueños, nada es imposible. Agradezco el apoyo que me brinda el Estado.
Si una persona quiere crecer, desarrollarse, solo debe dedicarse y estudiar mucho. Espero que muchos jóvenes sigan accediendo a las oportunidades que da el Pronabec con Beca 18.
Beca 18 ha cambiado bastante mi vida porque me ha dado muchas oportunidades, pe, de conocer, de conocer a gente. Yo allá en Pangoa en la selva conocía solamente a mis familiares, amigos, pero... no, no es como aquí en Huancayo, pues, aquí puedes conocer a más personas que te pueden dar más oportunidad de trabajo. Personas profesionales, aquí... que están ahí cerca, que te pueden aconsejar, que te pueden brindar información, que te pueden enseñar muchas cosas. Ese es la gran ventaja.

FASES DE LA METODOLOGÍA

1. SELECCIÓN DE DATOS

NEGATIVOS

Ellos, los tutores, te brindan asesorías en las que te preguntan "Oye, ¿cómo estás?" Pero se necesitan asesorías para cursos fuertes, un repaso o algo así.

Es chistoso, porque cuando tu firmas tu compromiso de la beca, te dicen te vamos a dar esto o esto, y así de la noche a la mañana publican "chicos, ya no van a recibir inglés, ustedes tienen que pagárselo en su misma universidad" ... ¿y ahora? Solo en mi universidad, el costo por ciclo es como mil y pico y el presupuesto que te da la beca es 250.

Lo que nos da el Pronabec es 250 para materiales y a veces eso no nos alcanza. En realidad no, porque tenemos que comprar telas. Entonces, lo que hace la mayoría, es tratar de ahorrar en algunas cosas para poder suplir ese gasto extra que tenemos en materiales, porque no alcanza.

La tecnología, en mi caso y a la mayoría, nos ha chocado porque ni siquiera la computadora sabía usar, ni siquiera Facebook conocía. Cuando en la postulación nos pedían nuestros correos electrónicos nosotros ni sabíamos que era eso. Qué será, decíamos. En verdad no sabía nadie que era ese correo, Facebook, nada.

Es cierto que los de pobreza o pobreza extrema son los que menos recursos tienen para acceder a la educación, pero tienen que darle información antes. Cualquiera dice: "ah ya, a esa universidad yo postulo", y al momento de entrar a esa universidad, quizás es demasiada exigencia académica y pierdes la beca. Y, o sea, si el chico realmente tiene las ganas, darle una asesoría paralela para que no pierda la beca. Un tutor que realmente le pueda apoyar. Porque de nada te sirve pagarle hasta el tercer ciclo para que al final pierda, y empiece de nuevo, quizás, en una nacional o en un instituto.

Es todo un mundo el problema de la vivienda, porque a veces un gestor o dos gestores para cuatrocientos o quinientos becarios que llegan al mismo tiempo es muy complicado.

Fue un coordinador que nos trajo de allá de mi ciudad. Y nos dejó a todos en el terminal. No nos llevó más allá. Nos dejó a todos ahí. Se fue con las mismas de regreso. Casi la mayoría no conocía a Lima

Teníamos que venir con nuestro propio dinero de Abancay hasta Lima. De la familia sus ahorros, por ejemplo. De mi papá, toda su plata nos hemos usado

FASES DE LA METODOLOGÍA

1. SELECCIÓN DE DATOS

Total de opiniones: **160**

Opiniones positivas: 80

Opiniones negativas: 80

FASES DE LA METODOLOGÍA

2. PREPROCESAMIENTO

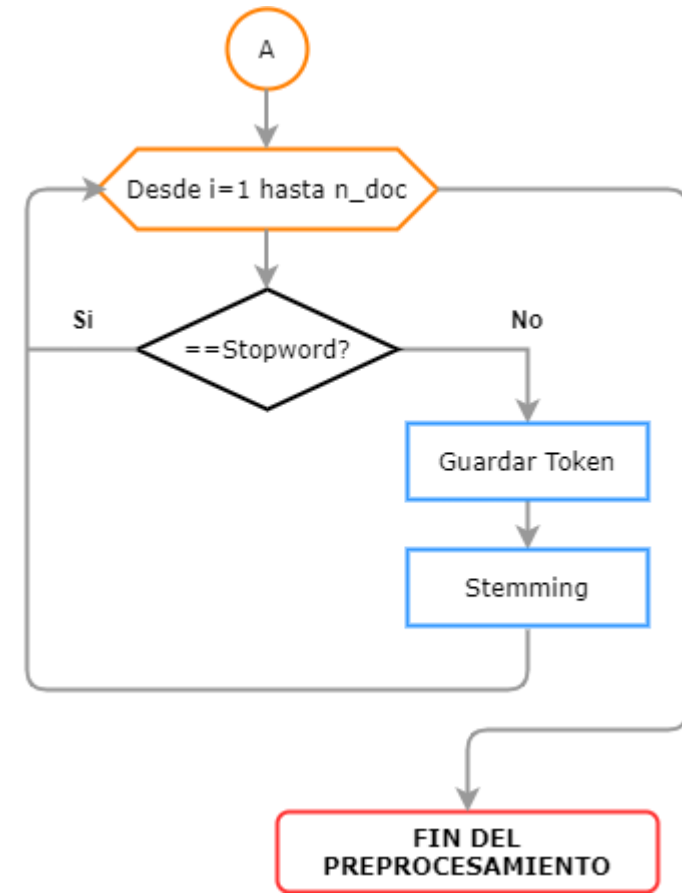
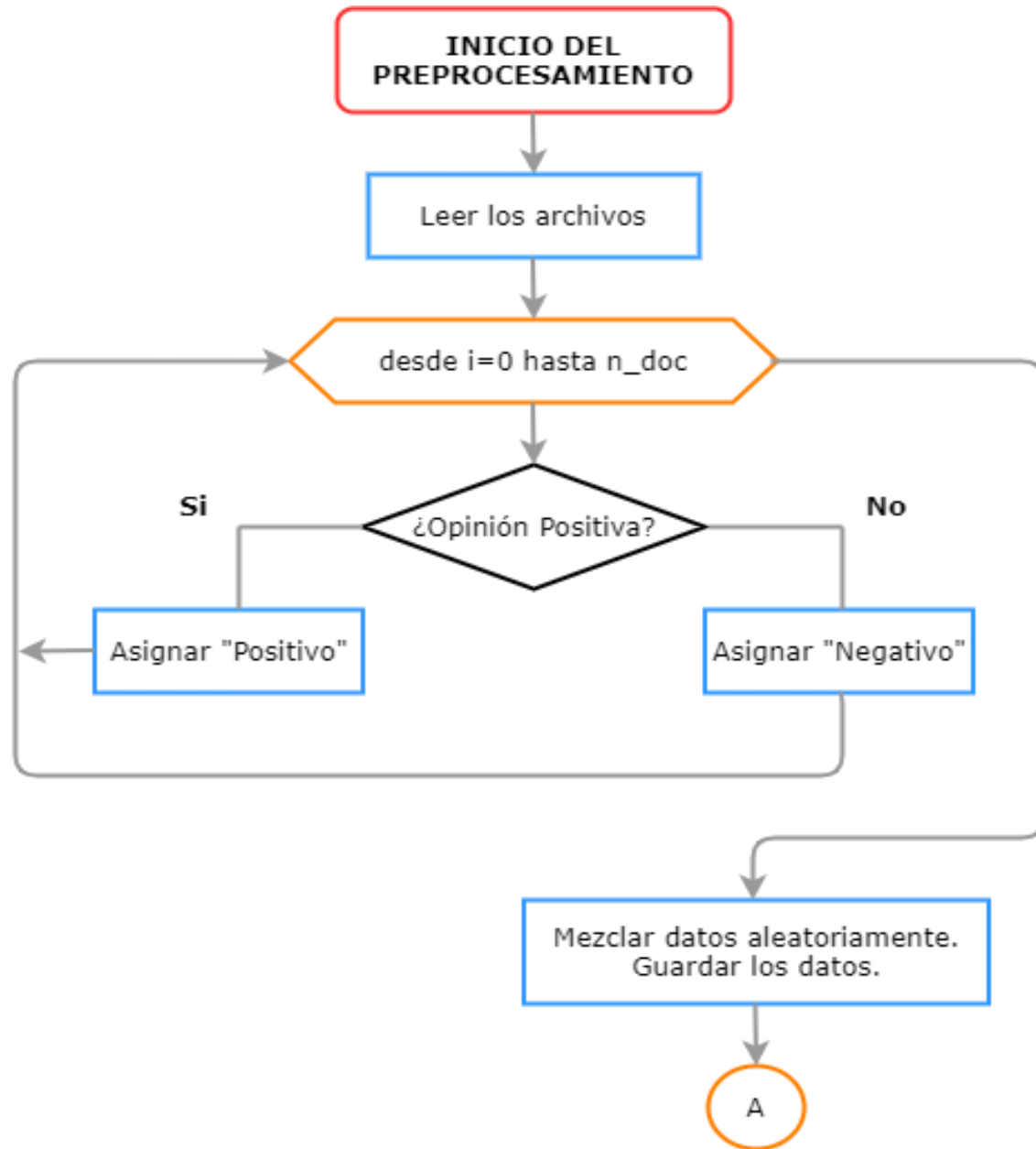
Corresponde a la limpieza, normalización y tokenización de los datos obtenidos, eliminación de palabras vacías. Para este proyecto se realizó tres tipos de preprocesamiento:

- ✓ Preprocesamiento inicial: Poner texto en minúscula eliminando dígitos y signos de puntuación.
- ✓ Preprocesamiento inicial eliminando stopwords.
- ✓ Preprocesamiento inicial eliminando stopwords con stemming.

Nube de palabras con el preprocesamiento inicial

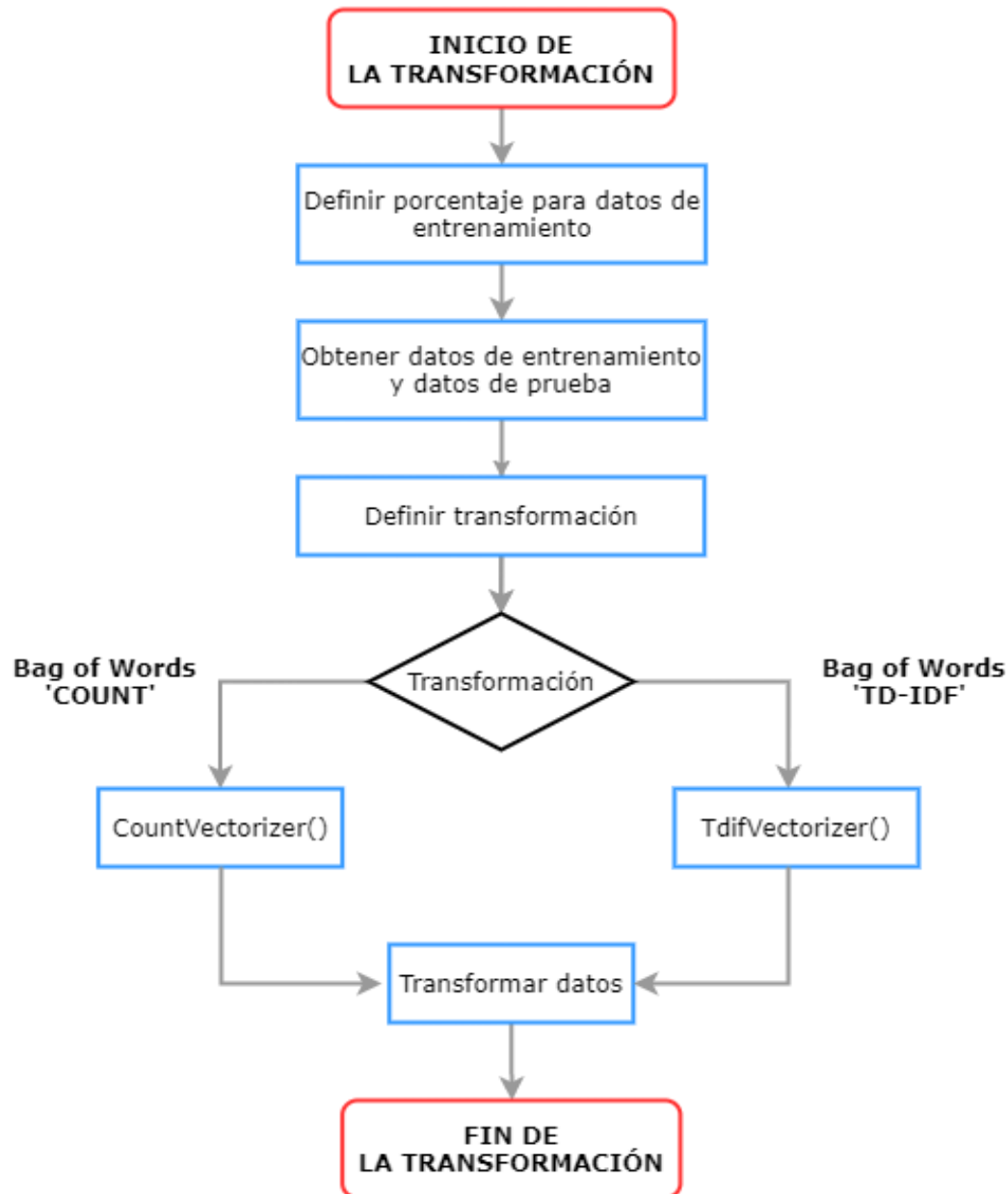


FASES DE LA METODOLOGÍA



2. PREPROCESAMIENTO

FASES DE LA METODOLOGÍA



3. TRANSFORMACIÓN

Para la transformación de la data primero hay que definir el porcentaje de datos que tendremos como prueba y como entrenamiento. Para este proyecto se consideró tener un **20%** de la data original como datos de prueba.

Luego hay que codificar la información para que los modelos de clasificación pueda interpretarla y hacer cálculos con ella.

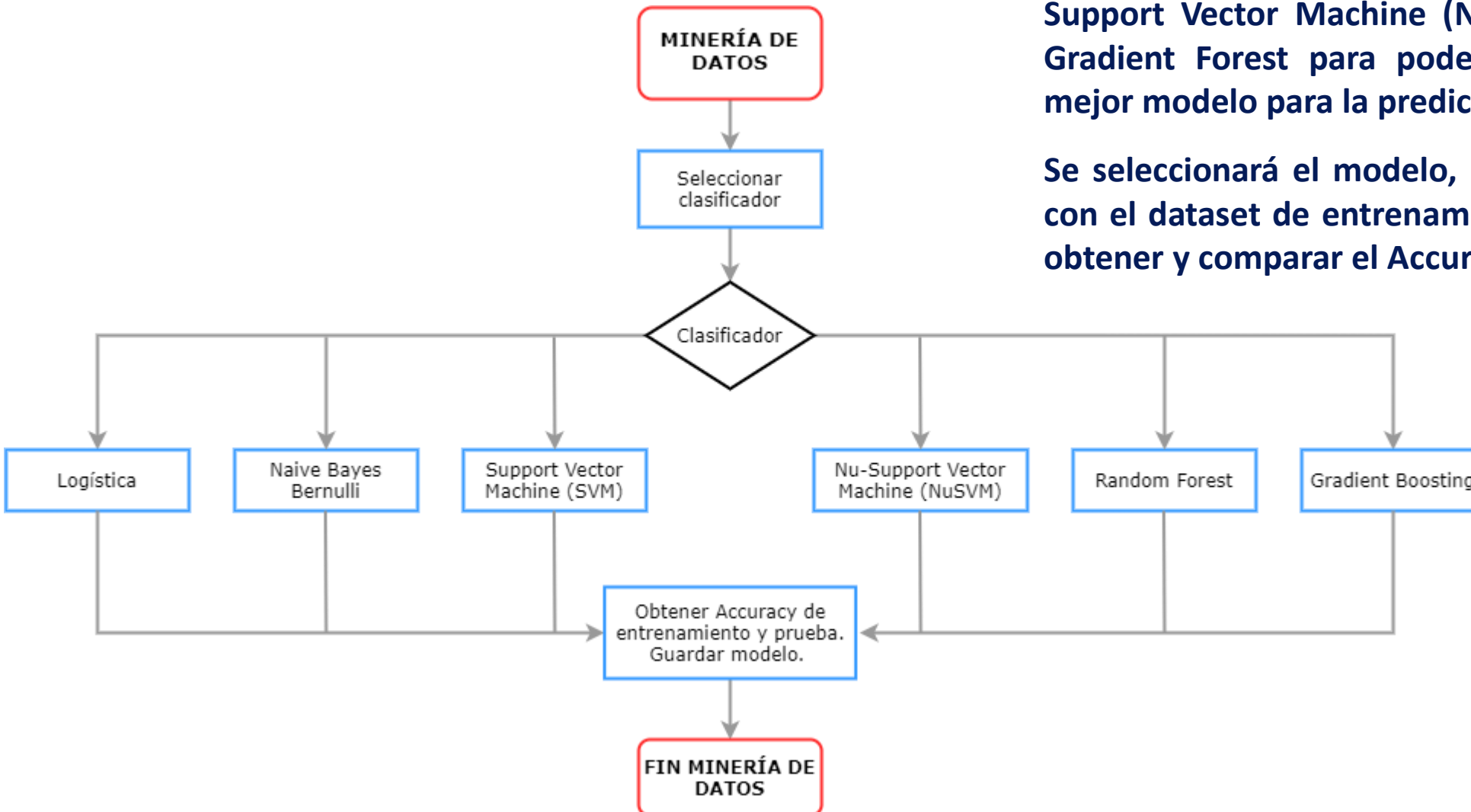
Las transformaciones consistirán en realizar un Bag of Words mediante las funciones `CountVectorizer()` y `TdifVectorizer()`.

FASES DE LA METODOLOGÍA

4. MINERÍA DE DATOS

Se implementará seis clasificadores: Logística, Naive Bayes Bernoulli, Support Vector Machine (SVM), Nu-Support Vector Machine (NuSVM), Random Forest y Gradient Forest para poder comparar y obtener el mejor modelo para la predicción.

Se seleccionará el modelo, se construirá y entrenará con el dataset de entrenamiento para posteriormente obtener y comparar el Accuracy.



FASES DE LA METODOLOGÍA

5. INTERPRETACIÓN Y EVALUACIÓN

Los datos obtenidos se verifican y se compara el Accuracy de los modelos.

Accuracy:

		Logística	NB Bernulli	SVM	Nu-SVM	Random Forest	Gradient Boosting
PreProc_Inicial	CountVectorizer()	0.84	0.88	0.78	0.88	0.88	0.84
	TdifVectorizer()	0.88	0.88	0.91	0.88	0.88	0.84
PreProc_Inicial sin StopWords	CountVectorizer()	0.88	0.88	0.81	0.81	0.91	0.81
	TdifVectorizer()	0.91	0.88	0.84	0.91	0.88	0.84
PreProc_Inicial sin StopWords + Stemming	CountVectorizer()	0.84	0.88	0.81	0.81	0.91	0.81
	TdifVectorizer()	0.88	0.88	0.91	0.91	0.91	0.88

**Tema: Análisis sobre la labor del PRONABEC respecto al programa Beca 18 y su impacto social,
empleando técnicas de Procesamiento de Lenguaje Natural.**

CONCLUSIONES

CONCLUSIONES

- ✓ Es posible desarrollar un sistema inteligente para clasificar eficazmente mensajes de aprobación y desaprobación alrededor de la labor que realiza el PRONABEC respecto a Beca 18.
- ✓ El impacto que genera el preprocesamiento de los datos es tanto o más importante que la elección del modelo a utilizar. Por ejemplo, el modelo de regresión logística presenta un mayor valor de Accuracy cuando se utiliza un preprocesamiento eliminando stopwords sin stemming.
- ✓ Para los distintos tipos de preprocesamiento en todos los modelos se presenta un mejor rendimiento cuando se utilizó la transformación TF-IDF para construir el Bag of Word; esto se explica dado que el TF-IDF penaliza las palabras frecuentes, por lo que, a diferencia de un Bag of Word tradicional por frecuencias, hay menos necesidad de tratar explícitamente las palabras vacías y la identificación de estas mismas puede ser mejor.
- ✓ Debido a la naturaleza del modelo, el Naive Bayes Bernoulli, obtiene un Accuracy aproximadamente igual en todas las estimaciones realizadas con los datos de prueba, esto se explica dado que se basa en probabilidades de aparición de las palabras más trascendentes y esto haría que no se viera modificado por los tipos de preprocesamiento considerados.

REFERENCIAS

- ✓ Bannister, K. (2015). *Entendiendo el análisis de sentimiento: qué es y para qué se usa*. Brandwatch. <https://www.brandwatch.com/es/blog/analisis-de-sentimiento/>
- ✓ Cotler, J. (2016). *Un estudio cualitativo de los becarios del programa beca 18*. PRONABEC. https://www.pronabec.gob.pe/modPublicaciones/descarga/serie7_educacionsuperior.pdf
- ✓ Diego, L. (2019). *Análisis de sentimientos: aplicación sobre textos en redes sociales* [Trabajo Final Integrador, Instituto Tecnológico de Buenos Aires]. Repositorio ITBA principal. <http://ri.itba.edu.ar/handle/123456789/1836>
- ✓ Digital Málaga (2015). *Procesamiento del Lenguaje Natural*. Digital Málaga. <http://www.digitalmalaga.com/fileBoletin/16/>
- ✓ Espinoza, L. (2018). *Acceso a la educación superior de poblaciones vulnerables: el caso Beca 18, modalidad albergues del PRONABEC – Perú, 2012 – 2016* [Tesis de Maestría, Pontificia Universidad Católica del Perú]. Repositorio Digital de Tesis y Trabajos de Investigación PUCP. <http://hdl.handle.net/20.500.12404/13481>
- ✓ Gil, V. (2018). Análisis de sentimientos sobre el impacto social de proyectos de vivienda en América Latina: el caso un TECHO para mi país (Colombia). *Espacios*, 39(44), 30.

REFERENCIAS

- ✓ Guerreo, G., Rojas, V., Cueto, S., Vargas, J. y Leandro, S. (2019). *Implementación de programas de inclusión social en territorios con población vulnerable: ¿Cómo está cambiando Beca 18 la vida de los y las jóvenes del valle de los ríos Apurímac, Ene y Mantaro (VRAEM)?* Grupo de Análisis para el Desarrollo [GRADE]. <https://www.grade.org.pe/wp-content/uploads/GRADEdi96.pdf>
- ✓ Kao, A. y Stephen, R. P. (2007). *“Natural Language Processing and Text Mining”*. Springer Science + Business Media. LLC.
- ✓ Márquez, A. (2016). La desigualdad social y las reformas educativas. *Perfiles Educativos*, 38(154), 3 – 18.
- ✓ Moreno, A. (2018). Procesamiento del lenguaje natural ¿qué es? Instituto de Ingeniería del Conocimiento. <https://www.iic.uam.es/inteligencia/que-es-procesamiento-del-lenguaje-natural/>
- ✓ PRONABEC (2015). *Beca 18 testimonios*. PRONABEC. <http://www.pronabec.gob.pe/inicio/publicaciones/documentos/beca18testimonios.pdf>
- ✓ PRONABEC (2015). *Motivaciones de los postulantes seleccionados e ingresantes de Beca 18 que deciden no seguir la beca*. PRONABEC. http://www.pronabec.gob.pe/inicio/publicaciones/documentos/motivaciones_postulantes_%20beca_18.pdf

REFERENCIAS

- ✓ *¿Qué es el análisis de sentimiento?* (2020). Digital Guide IONOS. <https://www.ionos.es/digitalguide/online-marketing/analisis-web/analisis-de-sentimiento/>
- ✓ Rodríguez, R., y Bundschus, M. (2016). Text mining patents for biomedical knowledge. *Drug Discovery Today*, 21(6), 997–1002.
- ✓ Tovar, A. y Huamán, G. (2020). *El otro rostro de Beca 18: el 43% de estudiantes indígenas no logra terminar la carrera*. Ojo público. <https://ojo-publico.com/2330/beca-18-el-43-de-indigenas-no-logra-concluir-estudios>
- ✓ Vega, P. (2017). La desigualdad invisible: el uso cotidiano de los espacios públicos en la Lima del siglo XXI. *Territorios* (36), 23-46. Doi: <http://dx.doi.org/10.12804/revistas.urosario.edu.co/territorios/a.5097>

Encuéntranos en

www.datascience.pe



-  Data Science Research
-  Data Science
-  Data Science
-  Data Science Research Perú
-  Data Science Research Perú

¡Muchas gracias!



#yomequedoencasa