Aplicación de Machine Learning para predecir la Felicidad y factores influyentes mediante los resultados del Saber-Pro de los estudiantes de la Universidad Distrital Francisco José de Caldas

(Abril de 2020)

GRUPO DE INVESTIGACIÓN THRISCUD

# Introducción

En el presente documento se plantea un modelo predictivo para conocer el estado emocional de un estudiante de acuerdo a si es Feliz o No y las razones del mismo, ,el modelo es desarrollado mediante el Machine Learning y la ciencia de datos lo cual brinda un respaldo matemático y estadístico junto a la aplicación de la ingeniería en campos computacionales, para esto se es utilizado 13 factores de prueba validados y estudiados más a profundidad en los siguientes documentos

“Análisis en SPSS sobre la influencia de diferentes factores en la felicidad de los estudiantes de la Universidad Distrital Francisco José de Caldas” y “Investigación Felicidad Universidad Distrital” donde se estudian a detalle las variables con las cuales se da pie a plantear un modelo de prueba para realizar el objetivo de estudio y se realiza un estudio práctico con los estudiantes que realizaron el saber Pro en el año 2018 para conocer el comportamiento del modelo en casos específicos..

# Metodología

## Diseño del Modelo

Los modelos predictivos de Machine Learning necesitan de dos tipos de variables:

* Variable Dependiente (Felicidad): Conocida como variable objetivo o a predecir, será aquella variable a conocer.

En nuestro caso es llamada Felicidad, para el entrenamiento del modelo se es necesario contar con datos en esta variable para poder entrenar y predecir en base a esta, por ende se utilizó el cuestionario de la Felicidad de Oxford el cual cuenta con 29 preguntas y del cual se realizó a 300 estudiantes de la universidad Distrital Francisco José de Caldas.

De los cuales se posee el resultado si es Feliz o Triste.

* Variables Independientes (13 Factores de Prueba): Variables predictivas o aquellas que tienen influencia en la variable dependiente.

Para esto planteamos 13 Preguntas con temas diferentes y la realizamos conjuntamente con el cuestionario de la universidad de la universidad de Oxford para obtener 300 muestras provenientes de los estudiantes de la Universidad Distrital.

Los 13 factores predictivos son los siguientes:

Edad

Género

Semestre

Estrato económico

Estado civil

Trabaja

¿El nivel de sus ingresos con respecto a sus gastos mensuales son?

Promedio Académico

Es creyente

¿Quién costea sus estudios?

Tiempo de desplazamiento casa-universidad

Tiempo promedio diario de uso de celular e internet

¿En qué tipo de casa vive?

* Algoritmo de Machine Learning: (Vectores de Soporte)

Debido al tipo de estudio y conformación de variables el algoritmo que ofrece mayor precisión es el conocido como vector de soporte, por lo cual será ajustado a los datos que se poseen.

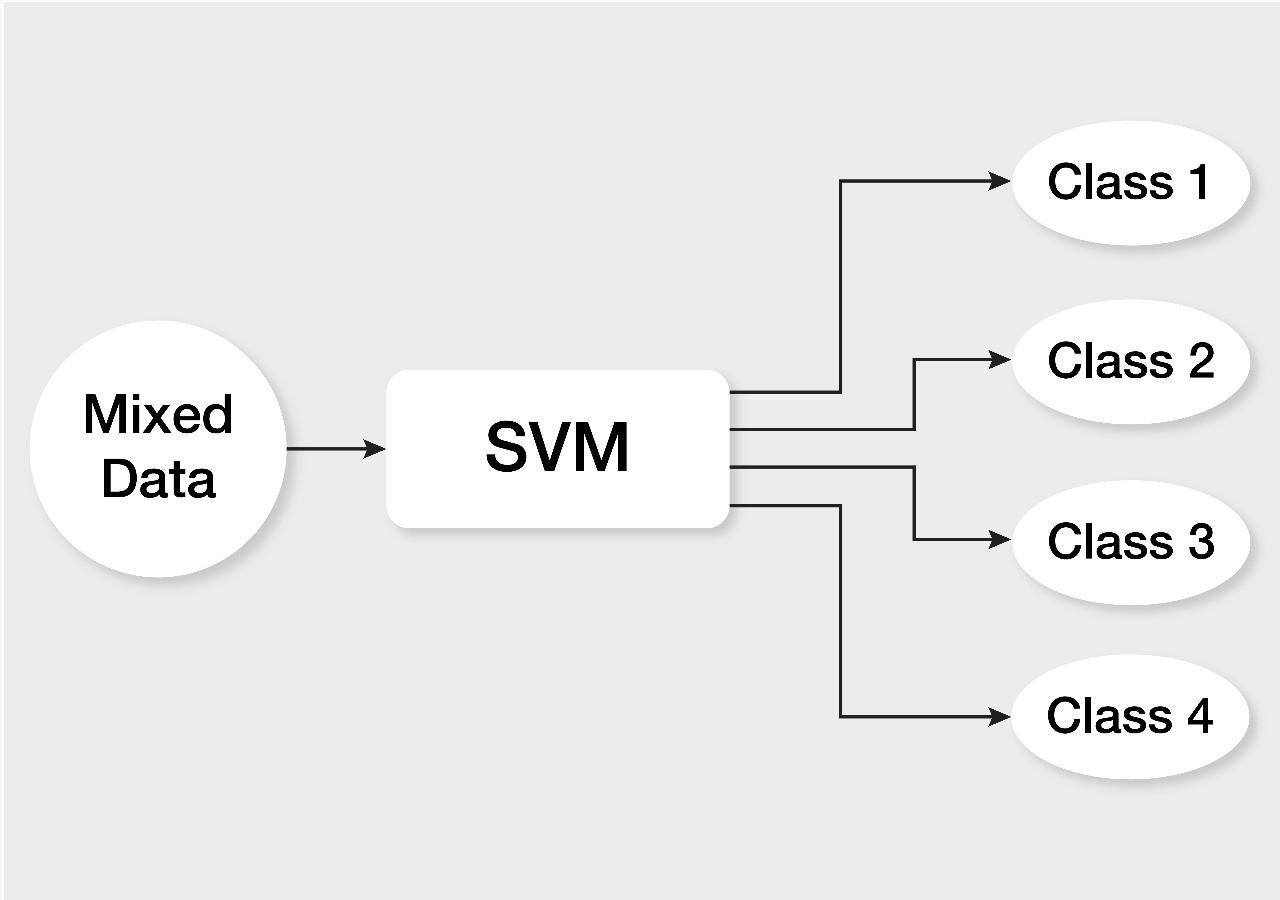


Imagen 1: Vector de Soporte

Caso de Estudio Específico: (Saber Pro 2018): Se aplica el modelo de Machine Learning mediante vectores de soporte a una muestra de datos de Estudiantes que realizaron el Saber Pro en su edición del primer semestre del año 2018.

* Variable Dependiente: Felicidad
* Variables Independientes (5 Factores en común ):

Género

Estrato

Trabaja

Estrato económico

Tiempo promedio diario de uso de celular e internet

## Muestra

Para este caso de estudio, se tomó una muestra de 302 estudiantes de pregrado de la Universidad Distrital pertenecientes a cualquiera de las facultades. La estrategia de muestreo fue aleatorio simple, donde el llenado de la encuesta fue totalmente voluntario.

En Cuanto a los resultados del Saber-Pro se utilizaron los 2000 registros respectivos a estudiantes pertenecientes a la universidad Distrital, siendo la misma el centro de estudio.

## Herramientas y Materiales

* Python 3.8 junto a las librerías Numpy,Pandas, Scikit-Learn.
* Base de Datos Saber-Pro 2018
* Resultados encuesta Felicidad junto a los 13 factores de los estudiantes de la Universidad Distrital.

## Procedimiento

1. Validación Encuesta Felicidad y 13 Factores

Primero se toma la encuesta de Felicidad junto a los 13 factores que serán tomados como variables , ya que se busca conocer si estas variables tienen incidencia dentro de la felicidad y de tenerla predecir con estas variables la felicidad de un estudiante sin necesitar de la aplicación del cuestionario.

Para esto se utiliza la variable dependiente u objetivo llamada Felicidad, la cual es el resultado del cuestionario de Oxford donde se tiene el registro de si un estudiante es feliz o no, y las variables dependientes siendo Edad,Género,Semestre, Estrato económico, Estado civil, Trabaja, ¿El nivel de sus ingresos con respecto a sus gastos mensuales son?,Promedio Académico ,Es creyente ,¿Quién costea sus estudios?

Tiempo de desplazamiento casa-universidad

Tiempo promedio diario de uso de celular

¿En qué tipo de casa vive?.

Una vez planteadas las entradas necesarias para aplicar los modelos de Machine Learning se utilizará el conocido como Vector de Soporte ya que es el más indicado para este tipo de casos.

Dependiendo del resultado del mismo conoceremos el SCORE, siendo esto la precisión resultante de las predicciones realizadas para conocer la felicidad de un individuo.

De obtenerse una precisión significante es decir entre más cercana a 100 % se puede validar o no la realización de un modelo a través de machine learning para predecir la Felicidad a través de los factores mencionados sin ser necesaria la aplicación del cuestionario a cada individuo,pero si debe utilizarse de este los resultados para una muestra significativa como variable dependiente que beneficiara al entrenamiento de la máquina.

1. Validación Factores de Felicidad conjunto a Datos Saber Pro 2018.

Una vez realizada la validación de los factores de Felicidad se podría realizar la predicción de la Felicidad a los estudiantes que realizaron el saber pro 2018 debido a que estos datos comparten varios de los factores de prueba utilizados.

Los cuales son:

* Género
* Estrato
* Trabaja
* Estrato económico
* Tiempo promedio diario de uso de celular e internet

Se realiza el procedimiento anterior con vectores de soporte esta vez con 2000 muestras del Saber Pro de las variables en común las cuales serán las variables independientes del modelo conjunto y la Felicidad como Variable dependiente.

En este caso al utilizarse sólo 5 variables dependientes se obtendrán varios resultados duplicados en la totalidad de los registros por ende se eliminan registros duplicados, y se realizará la predicción con todos los casos posibles que se presenten en las variables dependientes.

Finalmente el modelo es puesto a entrenar, se evidencia una desproporción en los datos debido a 300 muestras de felicidad contra 2000 muestras del saber-pro por lo cual el modelo predictivo no tendrá la precisión deseada-

# Resultados

1. La precisión del modelo mediante Vectores de Soporte para la predicción de Felicidad en cuanto a los 13 factores es de un 95 %, es decir se obtiene un modelo que al obtener los datos de los 13 factores analizados de una persona puede determinar si esta es Feliz o no con un grado alto de precisión

Imagen 2: Resultado Precision 95% Felicidad-13 Factores

1. .La precisión del modelo mediante Vectores de Soporte para la predicción de Felicidad en cuanto a los datos del Saber pro 2018 es de un 58 % , en otras palabras el modelo carece del entrenamiento suficiente para realizar una predicción de acuerdo a los 5 factores de un estudiante que realizó dicha prueba en este año, esto debido a la desproporción de datos entre felicidad y estudiantes del saber-pro.

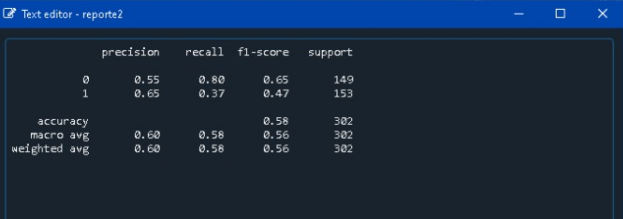


Imagen 3: Resultado Precision 60% Saber pro 2018

# Conclusión

De acuerdo a los resultados en el primer procedimiento es viable la realización de un modelo predictivo de la felicidad estudiantil a través de diversos factores, en este caso tan solo fueron 13 factores de prueba, pero de realizarse el cuestionario de la felicidad de Oxford con un tamaño de la muestra adecuado para obtener la variable dependiente Felicidad para el entrenamiento y de igual manera una agrupación de Factores como por ejemplo Factor Personal,Académico,Social y Económico que aportan variables con mayor impacto al estudiantado de la universidad podría obtenerse un modelo preciso y eficiente que pueda determinar si una persona es Feliz o No sino además determinar aquellas causas con mayor influencia y significancia en cada estudiante, lo cual posee una estrecha relación con la deserción universitaria.

En cuanto al segundo procedimiento se obtuvo un 60 % de predicción debido a la desproporción de datos sin embargo es un porcentaje que indica que de contar con un tamaño de la muestra satisfactorio el modelo propuesto es viable.

# Estudios futuros

Tomando en cuenta las conclusiones del presente documento, se plantea realizar un modelo predictivo de la Felicidad Estudiantil con variables más completas y con mayor impacto en la comunidad de estudiantes.

Esto debido al éxito del modelo con los 13 factores de prueba incidentes en la felicidad, para llevar esto a cabo se plantea agrupar factores por categorías:

Factor Personal,

Factor Académico

Factor Social

Factor Económico

Factor de Pasatiempos

Lo anterior permite analizar de acuerdo a cada factor determina no solo si la persona es feliz o no, sino que a su vez las causas del resultado obtenido referente a cada categoría, por lo cual con los datos de un estudiante de acuerdo a cada categoría se podrá conocer las variables que más fueron significantes y el resultado de si se considera una persona Feliz o Triste.

Se considera la propuesta de este modelo por categorías con una relación muy estrecha a la problemática de la Deserción Universitaria que le permitirá a la universidad conocer los estudiantes y las causas referentes a su bienestar en los aspectos propuestos.

Las categorías estarán basadas en las encontradas en las tesis de maestría y otros documentos académicos e investigativos relacionados a la Deserción universitaria específicamente en el alma mater, entre ellas “ MODELO REPRESENTATIVO DE DESERCIÓN ESTUDIANTIL VOLUNTARIA EN CARRERAS DE PREGRADO DE LA FACULTAD DE INGENIERÍA DE LA UNIVERSIDAD

DISTRITAL FRANCISCO JOSÉ DE CALDAS” realizada por

MÓNICA LIZETH SÁNCHEZ ARÉVALO .

VI. Anexos y Bibliografía

1. Análisis en SPSS sobre la influencia de diferentes factores en la felicidad de los estudiantes de la Universidad Distrital Francisco José de Caldas
2. Investigación Felicidad Universidad Distrital
3. Resultados Encuesta Felicidad y 13 Factores de prueba
4. Resultados Saber Pro 2018 <https://www.datos.gov.co/Educaci%C3%B3n/Resultados-Saber-Pro-Competencias-Genericas-2018-2/xqt7-7zji>
5. Imagen Vector de Soporte

<https://towardsdatascience.com/https-medium-com-pupalerushikesh-svm-f4b42800e989>