



Saturdays.IA

Predicción de Aprobación del Examen CENEVAL-EGEL

García-Dueñas, Pamela Zoe, Gatica-Martínez, Jessica, Lazcano-Calixto Ricardo
Neftali, Mejía-Blanco Erick & Sepúlveda-García Stephany

Eje Vertical: Educación

Lider del proyecto técnico: Hernandez-Mota, Rodrigo

Resumen

En el presente documento se presenta el modelo de Machine Learning, Random forest para determinar cuáles variables influyen en que un estudiante apruebe el Examen para el Egreso de la Licenciatura EGEL, así como la probabilidad el alumno apruebe el examen. Se encontró que las variables más importantes son: centro universitario, el tipo de examen y el promedio, la edad, cantidad de libros que tienen, la escolaridad de sus padres, cuántos cuartos y cuántas personas viven en su casa, hasta cual es el nivel de estudios al que los padres del estudiante deseaban que este llegara, además del conjunto de variables relacionadas con la vida laboral del estudiante. De los cuatro modelos probados Random Forest fue el más eficiente, ya que puede predecir correctamente el 72% de los casos, si un estudiante aprobará o no el examen con un 95% de confianza.

Introducción

En México sólo 1 de cada 2 estudiantes se titula (Martínez, 2015), lo que implica que el estudiante ya invirtió su tiempo y esfuerzo en terminar una licenciatura, pero no cierra el ciclo al no titularse, situación que es preocupante para las universidades, ya que invirtieron, tiempo y trabajo en un alumno que no concluye con su proceso.

Además de la titulación, la calidad educativa es uno de los temas cruciales para las instituciones de educación superior saber cómo es que sus egresados salen al mercado laboral, así es como surge CENEVAL, es una institución dedicada a evaluar estudiantes de educación media superior, es un organismo externo y autónomo que elabora exámenes con profesores de diversas universidades a través de dos consejos un técnico y un académico para cada examen. Actualmente CENEVAL elabora 39 exámenes para licenciaturas diferentes.

CENEVAL tomó relevancia al convertirse en una de las modalidades de titulación para los estudiantes de algunas universidades en México, ya que existen estudios como el de López, Salvo y García (1989) en el que se argumenta que una de las razones por las cuales los estudiantes no se titulaba era por lo complicado de los trámites administrativos, así como el tiempo necesario para poder realizar sus procesos de titulación, proporciona un método rápido para que los estudiantes se titulen a tiempo. En la búsqueda de aumentar el porcentaje de estudiantes que obtienen el grado.

Cada estudiante que realiza un el examen contesta una encuesta de 144 preguntas sobre su entorno socioeconómico, dichos datos pueden ayudar a identificar cuales son las variables socioeconómicas que determinan que un estudiante pase o no el Examen, ya que no todo tiene que ver con el desempeño académico, también influye las condiciones en las que vive.

Saber cuales son las variables socioeconómicas que afectan el desempeño de un estudiante, podría ayudar a las Instituciones de Educación superior a tomar medidas que reduzcan esas brechas y rezagos, que no necesariamente están ligadas con el desempeño de sus docentes o sus planes de estudios y elevar los resultados de sus estudiantes en el CENEVAL.

La propuesta es utilizar una técnica de Machine Learning que permita determinar cuales son las variables o factores socioeconómicos que influyen para que un estudiante pueda obtener un testimonio satisfactorio o sobresaliente y titularse.

Planteamiento del problema

Una de las dificultades que enfrentan las universidades es la titulación de los estudiantes, existe una brecha entre el periodo de egreso de los alumnos y la adquisición del grado de los mismos, ya que actualmente no es necesario contar con un título universitario para ejercer como profesionista. Rodríguez Betanzos (2014) señala que los universitarios titulados son personas que cuentan con mejores oportunidades para incursionar de manera más estable en el campo laboral. asimismo señala que la eficiencia terminal y el índice de titulación permiten observar los niveles de deficiencia que tienen las universidades. Además expone que la rigidez de los mecanismos de titulación provoca que los estudiantes se desincentivan a titularse.

De acuerdo con la Asociación Nacional de Universidades e Instituciones de Educación Superior (ANUIES), del total de jóvenes que ingresan a la educación superior en México, sólo el 50% logra titularse. El bajo índice de titulación es un tema de de interés para las universidades en México, pues omitir el proceso de titulación al egresar de un programa educativo no permite a las instituciones evaluar integralmente al egresado, ni a éste concluir con su formación académica (Martínez, 2015), (Vargas Pureko & Rivera Michelena 2006).

Otro de los factores a considerar es la calidad de la educación media superior, en esta incertidumbre es en la que surge CENEVAL (Gago, 2000), ya que es un organismo dedicado al diseño y aplicación de instrumentos de evaluación de conocimientos, habilidades y competencias de los estudiantes de las Instituciones de Educación Superior (IES) en México (CENEVAL, 2019).

La dificultad de los procesos de titulación aunado a un organismo dedicado a la evaluación de la educación externo a las universidades provocó que aprobar el examen se haya convertido en una de las opciones de titulación en los centros educativos de la Universidad de Guadalajara es la acreditación del examen CENEVAL-EGEL. Los alumnos que optan por esta modalidad tienen un 50% de probabilidad de acreditar el examen.

Sin embargo, CENEVAL mide únicamente los conocimientos, habilidades y capacidades de los estudiantes, no obstante, que estudiante se titule, no sólo depende de factores académicos, sino de factores intrínsecos al estudiante y su entorno, como su contexto social, familiar y económico. También se deben a otros factores que tienen que ver con la institución y todos los componentes del proceso educativo durante la trayectoria académica de alumno.

Encontrar las causas de este bajo rendimiento en la titulación nos podría ayudar a plantear estrategias que permitan el incremento en la acreditación del examen CENEVAL-EGEL

Problemática: Identificación de las variables de factores académicos, personales y socioeconómicos que influyen para que un alumno acredite el examen CENEVAL- EGEL mediante la aplicación de machine learning. Los resultados ayudarán a diseñar estrategias que incrementen la acreditación, por tanto, la titulación de los alumnos.

Descripción de la solución a la problemática detectada.

Se determinará el patrón de características que tiene un estudiante que aprueba el examen y se establecerá el grado de incidencia en que cada una de las variables independientes influyen y explican la acreditación de este.

Hipótesis

El contexto socioeconómico del alumno tiene relación con la acreditación del examen CENEVAL-EGEL.

Marco Teórico

La calidad de la educación es un tema primordial para cualquier institución educativa. Las universidades buscan acreditar sus programas académicos y que sus estudiantes alcancen estándares nacionales e internacionales de conocimientos. Existen diversas instituciones evaluadoras y acreditadoras en el país y una de ellas es el Centro Nacional de Evaluación para la Educación Superior (CENEVAL). CENEVAL es una asociación que realiza exámenes para evaluar los conocimientos de los egresados de todas las Universidades públicas del país, y de hecho otorga reconocimientos a las universidades en las cuales la mayoría de sus alumnos obtienen un resultado satisfactorio en el examen. Los estudiantes de licenciatura de México realizan el Examen General de Egreso de la Licenciatura EGEL, en la mayoría de los casos para obtener su titulación, debido a que este es un requisito para titularse o una forma para hacerlo.

Dichos exámenes son elaborados por profesores de todo el país y forman un excelente referente de los conocimientos de los estudiantes, aunque para poder tener una calificación aprobatoria en el mismo los conocimientos teóricos no son suficientes. Además, existen una serie de variables que influyen en que un alumno logre alcanzar un testimonio satisfactorio, como género, estado laboral, entre otras (Tinto, V., 1992).

En el año 2017 (Gonzalez-Marron, D., Enciso-Gonzalez, A., Hernandez-Gonzalez,..., A., 2017) realizaron un estudio “Evaluación de parámetros de encuesta de ingreso del CENEVAL para alumnos candidatos a ingresar al nivel superior, caso de estudio ITP”, en el cual evalúan las variables socioeconómicas del examen de Ceneval EXANI-II, analizaron los datos para los aspirantes a las licenciaturas de arquitectura, administración e ingenierías, estandarizaron los datos comunes en todas las encuestas, encontraron 23 atributos, los pasos que usaron el proyecto fue una reducción de dimensiones con PCA y como modelos de clasificación de datos nominales y regresión de datos numéricos, encontraron tres variables que determinan el desempeño de un estudiante; el promedio de bachillerato, el número de libros en casa, cantidad de horas de trabajo y la escolaridad del padre y de la madre.

En 2012 (Vera, C. M., Morales, C. R., & Soto, S. V., 2012) realizaron un estudio llamado “Predicción del Fracaso Escolar mediante Técnicas de Minería de Datos” en dicho estudio los autores pretenden encontrar los componentes socioeconómicos que influyen en el fracaso escolar de los estudiantes, suspendan o abandonen la escuela, realizaron una reducción de dimensiones con 10 algoritmos distintos, para quedarse con 15 atributos determinantes para su modelo, y utilizaron algoritmos de clasificación como árboles de decisión, encontraron que los árboles de decisión Prism permiten conocer a los alumnos que tienen mayor probabilidad de abandonar la escuela con un 99.8% de confianza.

Metodología

Descripción de los datos

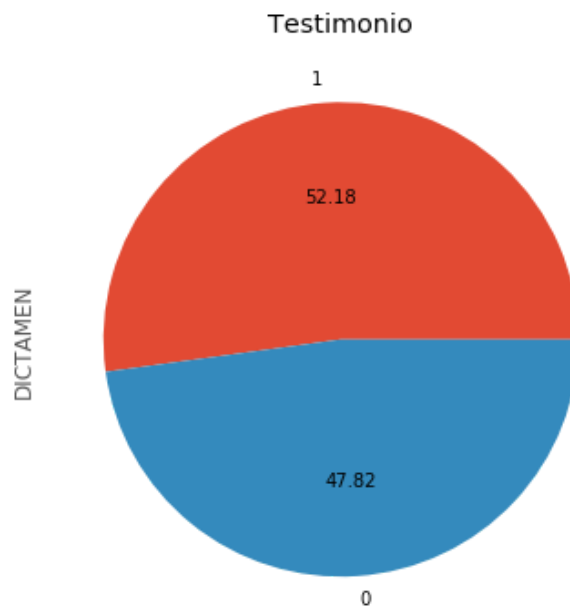
CENEVAL realiza un estudio socioeconómico a todos los estudiantes que realizan alguno de sus exámenes, ya sea de ingreso o de egreso, dicho estudio cuenta con 129 preguntas que describen el contexto social del estudiante (véase Anexo 1). Los datos de la encuesta fueron proporcionados por la Unidad de Evaluación y Acreditación de la Universidad de Guadalajara. Dicha base de datos contiene la información de 15,649 alumnos procedentes de todos los centros universitarios (40 licenciaturas) de la Universidad de Guadalajara que presentaron el examen CENEVAL-EGEL en el año 2017.

Todas las variables del examen fueron codificadas por términos de confidencialidad, no se puede identificar los Centros Universitarios, los estudiantes, ni los exámenes dentro de la base de datos.

La variable dependiente de esta estudio es ‘DICTAMEN’ que muestra si el alumno obtuvo una calificación satisfactoria o no del examen.

En la siguiente gráfica se observa que el 52.18% de los estudiantes que realizaron el examen obtuvieron un testimonio, mientras que el 47.82% no lo obtuvo.

Ilustración 1 Porcentaje de estudiantes con testimonio



Fuente: Elaboración propia

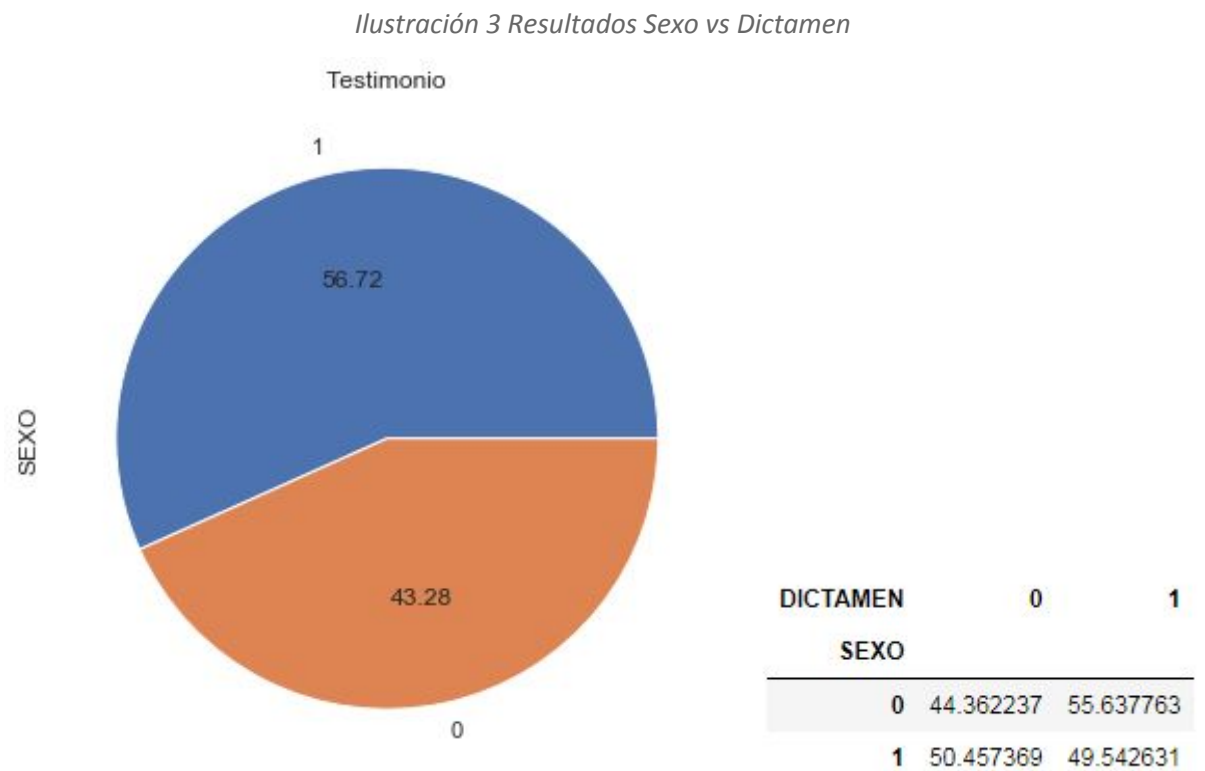
La Universidad aplica 35 de los 39 exámenes EGEL disponibles, dentro de los 16 centros de la Universidad de Guadalajara, se destaca que el centro universitario con mayor porcentaje de acreditados tuvo el 79% de alumnos aprobados, mientras que el centro que menos estudiantes aprobaron el examen obtuvo 19%.

Ilustración 2 Resultados Centro Universitario vs Dictamen

DICTAMEN	0	1
PTL_PROC		
1	58.3509514	41.6490486
2	38.5416667	61.4583333
3	29.9902153	70.0097847
4	33.1471136	66.8528864
5	40.010433	59.989567
6	20.2808112	79.7191888
7	80.9926082	19.0073918
8	58.9327146	41.0672854
9	70.9981168	29.0018832
10	40.948814	59.051186
11	73.7864078	26.2135922
12	66.6666667	33.3333333
13	61.2844037	38.7155963
14	77.3139746	22.6860254
15	54.1484716	45.8515284
16	56.8421053	43.1578947

Fuente: Elaboración propia

El 52.72% de los estudiantes que realizaron el examen fueron hombres, sin embargo, el 55.63% de las mujeres aprueba el examen, contra 49.5% de los hombres.



Fuente: Elaboración propia

El 62.68% de los estudiantes realizan el examen entre los 24 y 27 años, a su vez son los que tienen mayor probabilidad de aprobar el examen, se observa que el 63.60% de las personas con más de 40 años que realizan el examen no lo aprueban.

Ilustración 4 Resultados Sexo vs Dictamen

ANO_NAC		DICTAMEN		0	1
ANO_NAC		ANO_NAC			
20-23	0.90101604	20-23	45.3900709	54.6099291	
24-27	62.6877117	24-27	44.617737	55.382263	
28-29	21.1642916	28-29	51.9323671	48.0676329	
30-39	11.6301361	30-39	52.7472527	47.2527473	
40 y más	3.61684453	40 y más	63.6042403	36.3957597	

Fuente: Elaboración propia

El 60% de los estudiantes que realizan el examen tiene un promedio superior a 95, se observa que a mayor promedio, mayor probabilidad de aprobar el examen.

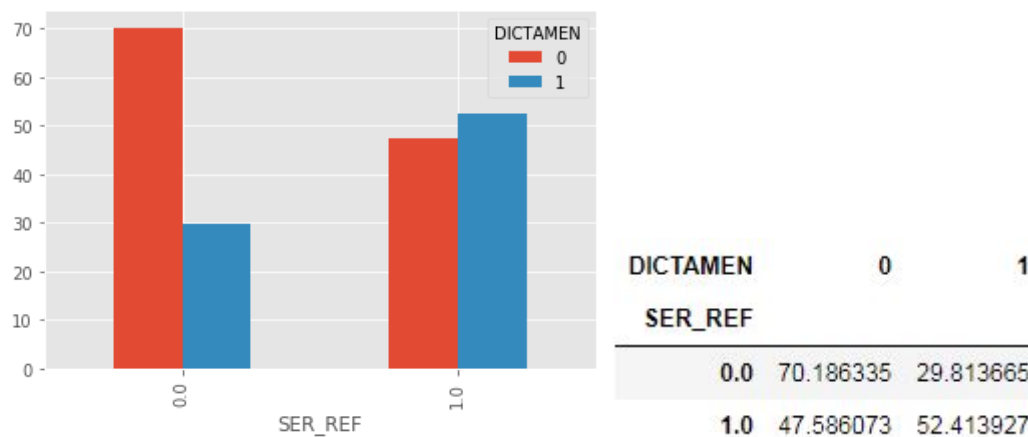
Ilustración 5 Resultados Promedio vs Dictamen

DICTAMEN	0	1
PRO_LIC		
6	77.272727	22.727273
7	79.503106	20.496894
8	61.466288	38.533712
9	44.810873	55.189127
10	18.510324	81.489676

Fuente: Elaboración propia

A continuación se muestran algunos resultados de variables destacables en el estudio descriptivo. La primera fue contar con un refrigerador, ya que solo el 29.81% de los alumnos que no tienen refrigerador aprueba el examen.

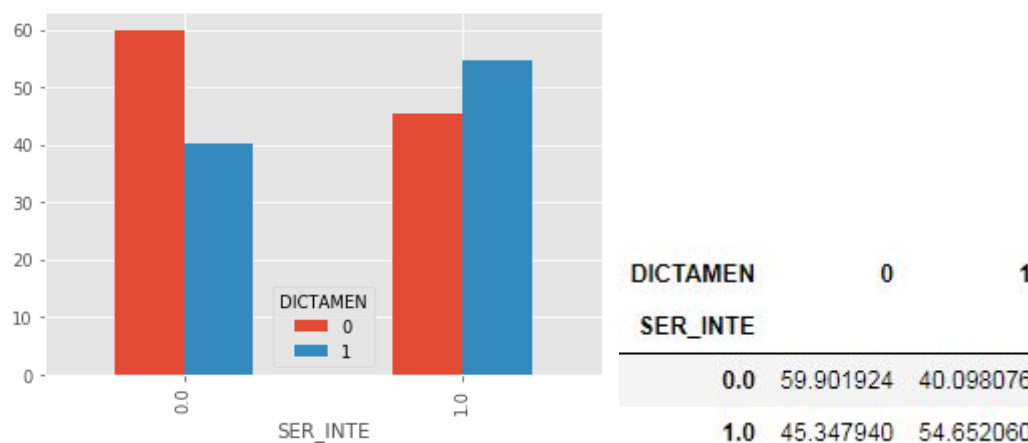
Ilustración 6 Resultados refrigerador vs Dictamen



Fuente: Elaboración propia

Contar con servicio de internet también fue una de las variables que representó mayor diferencia porcentual, ya que los estudiantes que tienen internet tienen 14.6% mayor probabilidad de aprobar el examen.

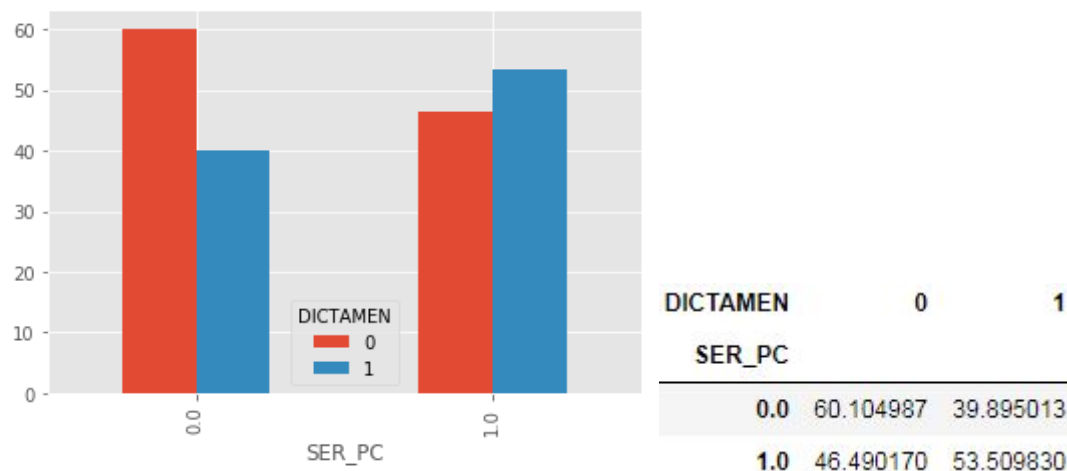
Ilustración 7 Resultados Servicio de internet vs Dictamen



Fuente: Elaboración propia

Con respecto a la computadora, tener una en casa resultó destacable en el análisis descriptivo, ya que el 60.10% de los estudiantes que no tienen computadora no aprueban el examen.

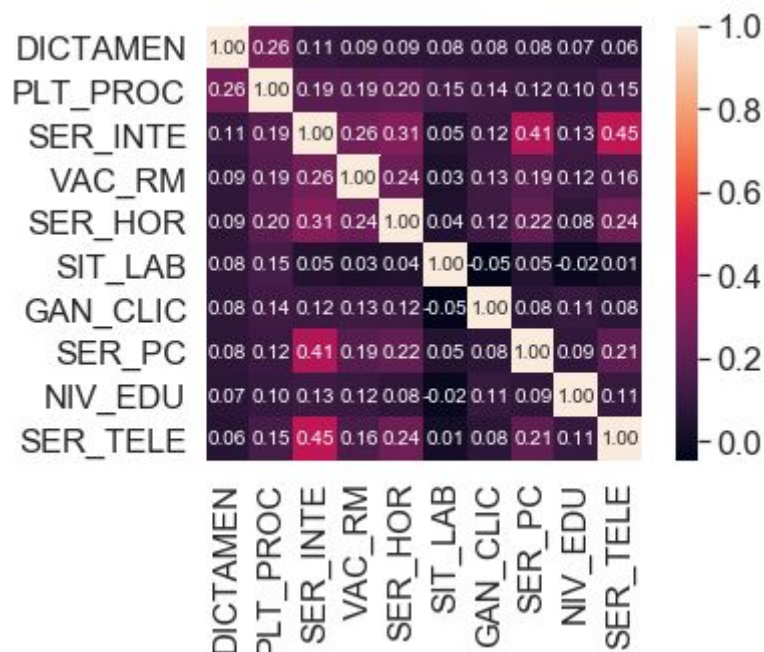
Ilustración 8 Resultados Servicio de PC vs Dictamen



Fuente: Elaboración propia

Las variables más destacadas detectadas a través de un análisis de correlación en orden de aparición fueron; centro universitario de procedencia, si cuenta con servicio de internet, si tomo vacaciones el último año, si tenía horno de microondas, cuál sería su ocupación laboral más viable al egresar, cuánto espera ganar cuando egrese, si tiene computadora en casa, el nivel educativo al que sus padres querían que llegara y por último si cuenta con televisión en su casa.

Ilustración 9 Correlación de las variables



Fuente: Elaboración propia

Se puede decir que las variables del contexto demográfico que explican cómo viven los estudiantes en casa son las más relevantes hasta el momento.

Descripción del modelo a utilizar

El caso de estudio corresponde a la predicción de clasificación, ya que la variable DICTAMEN tiene tres posibles resultados, al considerar los testimonios satisfactorio y sobresaliente como uno, la variable dependiente se convierte en binaria.

Los modelos de análisis de datos deben de ser comparados con otros modelos para poder elegir cual es el mejor, especialmente los modelos lineales o estadísticos simples, ya que, los nuevos modelos deben demostrar que es mejor que los modelos estadísticos básicos (Kitchenham & Mendes, 2009), debido a que los modelos algorítmicos en teoría requieren mayor capacidad de cómputo y son más complejos de interpretar, así que el presente estudio se compararon:

- **Regresión logística:** “es un modelo de regresión múltiple que busca explicar o predecir la probabilidad de que ocurra o no un evento” Berea-Baltierra, R., Rivas-Ruiz, R., Pérez-Rodríguez, M., Palacios-Cruz, L., Moreno, J., & Talaverab, J. O. (2014, p.2). Lo que convierte el modelo en un predictor de una variable dummie, variables en las cuales su valor es 0 o 1, por ejemplo la variable sexo, es hombre = 1 y mujer = 0, solo pueden tener dichos valores. Una característica relevante del modelo es que se pueden incluir variables de todo tipo, categóricas y continuas, siempre y cuando la variable dependiente sea categórica.
- **Árboles de decisión** los cuales son una forma gráfica y analítica de representar todos los eventos que pueden surgir a partir de una decisión asumida en cierto momento. Los árboles de decisión crean un modelo de clasificación basado en diagramas de flujo. Clasifican casos en grupos o pronostican valores de una variable dependiente (criterio) basada en valores de variables independientes (predictoras) (Berlanga-Silvente, V., Rubio-Hurtado, M. & Vilà-Baños, R., 2013).
- **Random Forest** Es una colección o conjunto de árboles de clasificación y regresión (CART) (Breiman et al., 1984) formados en los conjuntos de datos del mismo tamaño como conjunto de entrenamiento, llamados bootstraps, creado a partir de una muestra aleatoria remuestreo sobre el entrenamiento en sí mismo. RF sigue reglas específicas para el cultivo de árboles, la combinación de árboles, la autoevaluación y el procesamiento posterior, es robusto para sobreajustar y se considera más estable en presencia de valores atípicos y en espacios de parámetros dimensionales muy altos que otros algoritmos de aprendizaje automático (Sarica, A., Cerasa, A., & Quattrone, A., 2017)
- **Redes Neuronales** es el tercer modelo probado, este modelo de basa en la anatomía del cerebro humano cuenta con pequeñas unidades de procesamiento llamadas neuronas, lo que importa en este modelo es cómo interactúan unas con otras, la cual es ponderada por pesos y la transmisión de información a través de una función de activación, emplea un algoritmo de aprendizaje para minimizar una función de error en los datos y generar el mejor ajuste posible. (Salas, R., 2004).

La selección del mejor algoritmo está basada en el error del modelo o el accuracy de los modelos, todos los modelos fueron ejecutados en Python, con Jupyter.

Delimitaciones

Existen tres tipos de limitaciones de en el estudio:

1. **De datos.** Al contar únicamente con datos sociales no se puede hacer una predicción del nivel académico de los estudiantes, para identificar cuales son las ramas en las que estos necesitan reforzar sus habilidades, contar con dicha información haria un algoritmo más eficiente para predecir los resultados de los estudiantes.

Al ser realizado con datos de UdeG, se tendría que ajustar para el resto de las universidades para asegurar su funcionamiento adecuado, ya que se desconoce el comportamiento de sus estudiantes.

2. **Económicas** Al ser un proyecto social, la obtener recursos para operarlo es complicado debido a que no existen ganancias monetarias en la implementación del mismo, los beneficios son para la comunidad estudiantil.
3. **Temporales** Debido a que sólo tenemos lo datos del año 2017, el resultado está limitado a las circunstancias de esa generación. Lo ideal sería obtener datos de un período mayor.

Resultados

El objetivo de este proyecto es encontrar cuales son las variables socioeconómicas que influyen para que un estudiante obtenga un resultado favorable en el examen CENEVAL, para lo cual se hará una proyección para determinar si un estudiante pasará o no el examen.

El primer paso a realizar es estandarizar la base de datos y transformarla a valores normalizados para que los algoritmos alcancen la mayor precisión posible (véase en el anexo 2) de las 157 variables que contenía la base de datos, quitando las columnas repetidas quedaron 118, se eliminaron los registros vacíos, así que de 15,649 registros quedaron 15,613.

Con la base de datos procesada y limpia, se probaron cuatro modelos de machine learning, los cuales fueron medidos a través del accuracy en el grupo de prueba de cada modelo, se obtuvo lo siguiente.

Ilustración 10 Exactitud de la predicción de los modelos

Modelo	Accuracy
Regresión Logística	0.6686
Árboles de decisión	0.6834
Random Forest	0.7212
Redes Neuronales	0.6875

Fuente: Elaboración propia

Se observa que de los cuatro algoritmos el que tiene mayor exactitud es Random Forest, superando a redes neuronales, el modelo con el accuracy más bajo es regresión logística, sin embargo, destaca que la diferencia entre la predicción de los modelos no es elevada, sólo 5.26 puntos porcentuales, situación que puede estar dada por la naturaleza de los datos al ser socioeconomicos.

Las variables más determinantes en el modelo fueron el centro universitario de procedencia, el tipo de examen y el promedio de licenciatura, la edad de los muchachos, las condiciones de su casa, como la cantidad de libros que tienen, la escolaridad de sus padres, así como cuantos cuartos y cuántas personas viven en su casa, hasta cual es el nivel de estudios al que los padres del estudiante deseaban que llegara.

Resalta que 6 de las primeras 20 variables más importantes estaban relacionadas con la situación laboral de los estudiantes, tanto si trabajaban, como sus aspiraciones laborales al terminar la licenciatura.

De las variables relacionadas con el desempeño de los profesores, están el fomento a la investigación, si los profesores impartían clase, si incluir diversos factores al evaluar a sus estudiantes, así como si fomentaban la participación de los estudiantes.

Ilustración 11 Principales variables que influyen en el desempeño de un estudiante

Variable	Descripción	Indicador
PLT_PROC	Plantel de procedencia	0.086426
TIPO_EXA	Tipo de examen	0.053781
PRO_LIC	Promedio de licenciatura	0.033242
ANO_NAC	Año de nacimiento	0.032127
CUAN_LIB	¿cuántos libros tiene en su casa?	0.027744
ESC_PAD	Escolaridad del padre	0.026054
ESC_MAD	Escolaridad de la madre	0.025702
CPV_CASA	¿cuántas personas viven en su casa?	0.020308
SIT_LAB	¿cuál considera su opción laboral más viable al iniciar su ejercicio profesional?	0.018841
PRPF_MAET	Me ayudaron a encontrar trabajo	0.018535
GAN_CLIC	¿cuánto espera ganar al incorporarse al mercado laboral?	0.018432
TIEM_TRAB	Mientras realizó sus estudios de licenciatura, ¿cuánto tiempo trabajó?	0.017905
CUARTOS	Sin contar el baño, la cocina y los pasillos, ¿cuántos cuartos tiene su casa?	0.017836
SEMA_TRAB	Mientras realizó sus estudios de licenciatura, ¿cuántas horas a la semana le dedicaba a trabajar?	0.017467
HRS_TRAB	Actualmente, ¿cuántas horas a la semana le dedica a su trabajo?	0.015316
MAE_FIN	Fomentaban la investigación (científica y tecnológica)	0.014587
NIV_EDU	¿hasta qué nivel educativo querían sus padres que estudiara?	0.014539
MAE_FAC	Sus profesores faltaban a clases	0.014253
MAE_EIC	Incluían una evaluación integral al emitir una calificación	0.013492
MAE_PPDE	Promueven la participación activa de los estudiantes	0.013052

Fuente: Elaboración propia

Conclusiones finales

Se puede concluir que el contexto socioeconómico es vital en el desarrollo de un estudiante, el mejor modelo de los probados para este proyecto fue Random Forest, mediante el uso de este se pudo conseguir con una predicción correcta en el 72% de los casos, con un 95% de confianza.

Sorpresivamente las variables de rezago social no fueron las más representativas para el algoritmo, situación que puede estar determinada por la falta de muestra en esos casos.

La variable más importante para determinar si un estudiante pasa o no el examen es el Centro Universitario de procedencia, lo que revela que hay centros con mayor nivel educativo dentro de la red, la edad del estudiante es fundamental, así como las características de su casa y familia.

Otro conjunto de variables relevantes fue si el alumno trabajaba, ya que por el tipo de examen, si el estudiante tiene experiencia laboral su probabilidad de aprobar el examen aumenta. La investigación científica fue muy destacada, ya que si el alumno tiene deseos de ser investigador o si manifestó que los docentes impulsan la investigación, así como la participación su probabilidad de aumentar el examen es mayor.

La potencia de Random forest hizo que las variables más importantes en el análisis de correlación fueron totalmente distintas a las que resultaron con el procedimiento las variables que coinciden son el centro universitario, cuánto dinero esperan ganar al egresar y el sector en el que le gustaría trabajar al estudiante.

El estudio vislumbra que la universidad debe buscar medidas de estandarización de la calidad de la educación entre los centros universitarios, así como promover que sus estudiantes trabajen antes de egresar de sus licenciaturas, además de promover la investigación entre sus estudiantes, tal vez impulsando el trabajo con investigadores de la universidad para la publicación de artículos académicos o elaboración de tesis.

Trabajo futuro

- Se propone continuar con el trabajo y generar cadenas de colaboración para entrenar modelos que incluyan datos del desempeño académico de los estudiantes.
- Entrenar modelos con bases de datos de otras Instituciones de Educación Superior, tanto públicas y privadas para conocer el comportamiento de sus estudiantes.
- Programar modelos por cada centro universitario, hasta carrera universitaria, para poder promover medidas que fomenten un mejor desempeño de los estudiantes, basadas en sus características particulares.

Bibliografía

Berea-Baltierra, R., Rivas-Ruiz, R., Pérez-Rodríguez, M., Palacios-Cruz, L., Moreno, J., & Talaverab, J. O. (2014). *Del juicio clínico a la regresión logística múltiple*. *Rev Med Inst Mex Seguro Soc*, 52(2), 192-7.

Berlanga-Silvente, V., Rubio-Hurtado, M. & Vilà-Baños, R., (2013). *Cómo aplicar árboles de decisión en SPSS*. Universitat de Barcelona. Institut de Ciències de l'Educació

CENEVAL (2019) *Perfil institucional*. <http://www.ceneval.edu.mx/perfil-institucional>

Gago, A. (2000). El CENEVAL y la evaluación externa de la educación en México. *Revista Electrónica de Investigación Educativa*, 2 (2). Consultado el 20 de noviembre 2019 en: <http://redie.ens.uabc.mx/vol2no2/contenido-gago.html>

Gonzalez-Marron, D., Enciso-Gonzalez, A., Hernandez-Gonzalez, A. K., Gutierrez-Franco, D., Guizar-Barrera, B., & Marquez-Callejas, A. (2017). Evaluación de parámetros de encuesta de ingreso del CENEVAL para alumnos candidatos a ingresar al nivel superior, caso de estudio ITP. *Research in Computing Science*, 139, 135-147.

López, N. Salvo, B. y García, G. (1989). *Consideraciones en torno a la titulación en las instituciones de educación superior*. *Revista de la Educación Superior*, Vol. 18, No. 1 (69)

enero-marzo 1989.

Rodríguez-Betanzos, A., (2014) *Factors affecting students' graduation from a Mexican university*. Cuadernos de Investigación Educativa, Vol. 5, Nº 20, 2014, Montevideo (Uruguay), 117-127. ISSN 1510-2432 117 Universidad ORT Uruguay.

Salas, R. (2004). Redes neuronales artificiales. *Universidad de Valparaíso. Departamento de Computación*, 1.

Sarica, A., Cerasa, A., & Quattrone, A. (2017). Random forest algorithm for the classification of neuroimaging data in Alzheimer's disease: A systematic review. *Frontiers in aging neuroscience*, 9, 329.

Tinto, V., (1992). *El abandono de los estudios superiores: una nueva perspectiva de las causas del abandono y su tratamiento*, México: ANUIES-UNAM.

Vera, C. M., Morales, C. R., & Soto, S. V. (2012). Predicción del fracaso escolar mediante técnicas de minería de datos. *Revista Iberoamericana de Tecnologías del/da Aprendizaje/Aprendizagem*, 109.

Vera-Noriega, J., Ramos-Estrada, D., Sotelo-Castillo, M., Echeverría-Castro, S., Serrano-Encinas, D. y Vales-García, J. (2011), "Factores asociados al rezago en estudiantes de una institución de educación superior en México", en *Revista Iberoamericana de Educación Superior (ries)*, México, unam-iiue/Universia, Vol. III, núm. 7, pp. 41-56

Anexo 1

Descripción de las variables en la base de datos

Variable	Descripción
I. IDENTIFICACIÓN DE LA APLICACIÓN.	
TIPO_EXA	Tipo de examen
APLI	Número de aplicación
FECHA_APLI	Fecha de aplicación
II. IDENTIFICACIÓN DE LA INSTITUCIÓN SEDE	
IDENTIFICA	Identificación de la aplicación
CVE_SEDE	Clave de la institución en donde se llevó a cabo la aplicación
NOM_SEDE	Nombre de la sede de aplicación
CIU_SEDE	Ciudad o municipio donde se encuentra la sede de apl.
EDO_SEDE	Estado donde se encuentra la sede de aplicación.
III. IDENTIFICACIÓN DEL SUSTENTANTE	
FOLIO	Folio
MATRICULA	Clave que tiene la institución para la identificación del sustentante
IV. CUESTIONARIO DE CONTEXTO	
NOM_LIC	Nombre del examen
Sección 1. Datos generales	
NOMBRE	Nombre(s)
APE_PAT	Apellido paterno
APE_MAT	Apellido materno
Fecha de nacimiento. En los cuadros anote su fecha de nacimiento (día, mes y año) y rellene los óvalos correspondientes.	
DIA_NAC	Día de nacimiento
MES_NAC	Mes de nacimiento
ANO_NAC	Año de nacimiento
SEXO	Sexo
ENTI_NAC	¿En qué entidad nació?

¿Alguno de sus padres habla una lengua indígena?	
LENG_MA	Madre
LENG_PA	Padre
Sección 2. Datos escolares	
Identificación de la institución donde concluyó la licenciatura	
NOM_PROC	Nombre completo de la institución en la que concluyó la licenciatura
PLT_PROC	Campus o plantel de la escuela de procedencia
CIU_PROC	Nombre de la ciudad donde se ubica la institución en la que concluyó la licenciatura: si fue en el extranjero indique el país
EDO_PROC	Entidad de la escuela de procedencia
CVE_PROC	Clave de la institución en la que concluyó sus estudios de licenciatura
¿Cuál o cuáles son las razones por las que presenta el egel?	
RAZ_RAPT	Requisito administrativo para la titulación (sin importar el resultado obtenido)
RAZ_RACT	Requisito académico para la titulación (alcanzar determinado dictamen en el examen)
RAZ_OPT	Opción de titulación
RAZ_PCUM	Porcentaje de calificación de una materia
RAZ_CMNF	Conocer mi nivel de formación respecto a un estándar nacional
RAZ_OTR	Otras
REG_PROC	Indique el régimen de la escuela en la que cursó la mayor parte de sus estudios de licenciatura
MOD_LIC	¿En qué modalidad educativa cursó la mayor parte de sus estudios de licenciatura?
PRO_LIC	¿Cuál es su promedio general en la licenciatura?
¿RECIBIÓ BECA EN LA LICENCIATURA?	
BECA_DAC	Por desempeño académico
BECA_NEC	Por necesidad económica

LIC_POPC	La licenciatura que cursó, ¿fue su primera opción?
CAM_LIC	¿Se cambió de licenciatura?
MOT_CLIC	¿Cuál fue el principal motivo por el que cambió de licenciatura?
Sección 3. Características de la institución en la que cursó la licenciatura	
COM_PLIC	En comparación con otras instituciones educativas, ¿cómo considera que fue la calidad de su formación profesional en la licenciatura que cursó?
¿Cómo era la comunicación entre los siguientes actores en la institución donde cursó la licenciatura?	
COM_ESES	Estudiantes-estudiantes
COM_ESMA	Estudiantes-maestros
COM_ESAE	Estudiantes-autoridades educativas
Durante su formación profesional, ¿qué tan adecuados fueron los programas, instalaciones y servicios de la institución donde estudió?	
FOPR_ASAC	Asesoría académica
FOPR_TUTO	Tutoría
FOPR_APPS	Apoyo psicopedagógico
FOPR_PRMO	Programa de movilidad
FOPR_CLEX	Centro de lenguas extranjeras
FOPR_AULA	Aulas
FOPR_LATA	Laboratorios o talleres
FOPR_BIBL	Biblioteca
FOPR_SACO	Salas de cómputo
FOPR_ACIN	Acceso a internet
FOPR_ABRE	Acceso a bases de revistas electrónicas especializadas
FOPR_ACEL	Acervo electrónico (repositorio)
FOPR_BOTR	Bolsa de trabajo
FOPR_SDTE	Servicios de trámites escolares
FOPR_LHIN	Limpieza e higiene de las instalaciones

En qué medida la institución donde estudió la licenciatura fomentó lo siguiente:	
FOME_PECR	Pensamiento crítico
FOME_APAU	Aprendizaje autónomo
FOME_CARE	Capacidad reflexiva
FOME_FOPR	Formación práctica
FOME_FOTE	Formación teórica
FOME_CUGE	Cultura general
FOME_LIDE	Liderazgo
FOME_TRCO	Trabajo colaborativo
FOME_EXOR	Expresión oral
FOME_EXES	Expresión escrita
FOME_IDEX	Idiomas extranjeros
FOME_CMPP	Capacidad para poner en marcha proyectos propios
FOME_HBSI	Habilidad para la búsqueda y selección de información
FOME_TODE	Toma de decisiones
FOME_CANE	Capacidad de adaptación a nuevos escenarios
FOME_MDLT	Manejo de las tic
FOME_COET	Compromiso ético
FOME_RESO	Responsabilidad social
FOME_CCMA	Compromiso con el medio ambiente
¿Cuántos de sus profesores de la licenciatura tenían las siguientes características?	
MAE_DMAT	Dominaban la materia
MAE_UEDA	Utilizaban estrategias didácticas para un mejor aprendizaje de los temas
MAE_ASIS	Asistían a dar clase
MAE_CLAP	Iniciaban y terminaban las clases puntualmente
MAE_PPDE	Promovían la participación de los estudiantes
MAE_TAR	Dejaban tareas y trabajos adecuados para mi formación profesional

MAE_RET	Retroalimentaban las tareas y trabajos que dejaban
MAE_HEC	Hacían exámenes que correspondían con el programa del curso
PRA_PROF	En la institución donde cursó la licenciatura, ¿realizó prácticas profesionales?
¿Qué tan útiles fueron las prácticas profesionales en los siguientes aspectos?	
PRPF_AACA	Me ayudaron a aplicar los conocimientos adquiridos durante la licenciatura
PRPF_MPNC	Me proporcionaron nuevos conocimientos
PRPF_MAET	Me ayudaron a encontrar trabajo
¿Qué tan útil fue el servicio social en los siguientes aspectos?	
SER_MAAC	Me ayudó a aplicar los conocimientos adquiridos durante la licenciatura
SER_MPNC	Me proporcionó nuevos conocimientos
SER_MAET	Me ayudó a encontrar trabajo
Sección 4. Situación laboral. (En esta sección no deben tomarse en cuenta las prácticas profesionales como un trabajo)	
TIEM_TRAB	Mientras realizó sus estudios de licenciatura, ¿cuánto tiempo trabajó?
SEMA_TRAB	Mientras realizó sus estudios de licenciatura, ¿cuántas horas a la semana le dedicaba a trabajar?
INS_TRAB	Señale el tipo de institución en donde trabaja actualmente.
HRS_TRAB	Actualmente, ¿cuántas horas a la semana dedica a su trabajo?
REL_TRA	Indique la relación entre su trabajo actual y la preparación obtenida durante la licenciatura
Sección 5. Expectativas Laborales	
EVA_ETRA	Tomando en consideración el prestigio de la institución donde estudió, evalúe qué tan fácil o difícil le será encontrar un trabajo relacionado con su formación profesional.
GAN_CLIC	¿Cuánto espera ganar al incorporarse al mercado laboral?

SIT_LAB	¿Cuál considera su opción laboral más viable al iniciar su ejercicio profesional?
Sección 6. Entorno social	
ESC_MAD	¿Cuál es el máximo nivel de estudios alcanzado por su madre (aunque haya fallecido)?
ESC_PAD	¿Cuál es el máximo nivel de estudios alcanzado por su padre (aunque haya fallecido)?
CUAN_LIB	¿Cuántos libros hay en su casa? (no considere revistas, periódicos o libros de texto)
En su casa, ¿cuenta con los siguientes bienes y servicios?	
SER_LAV	Lavadora de ropa
SER_REF	Refrigerador
SER_HOR	Horno de microondas
SER_INTE	Internet
SER_TVP	Televisión de paga (izzi, sky, etc.)
Del siguiente listado de bienes, cuántos hay en su casa.	
SER_PC	Computadora
SER_TV	Televisor
SER_AUTO	Automóvil
SER_BANO	Baños completos
EDO_REP	¿Cuántos estados de la república mexicana ha visitado como turista?

Fuente: CENEVAL (2019)

Anexo 2

Original Name	Transformation
TIPO_EXA	Se transformó el tipo de examen al porcentaje de estudiantes que aprobaron, la clasificación quedó de 0 a 1
APLI	Se eliminó
FECHA_APLI	Se eliminó
NEMOTECNIA	Se eliminó
DESC_MNEMO	Se eliminó

IDENTIFICA	Se eliminó
DESC_IDENT	Se eliminó
CVE_SEDE	Se eliminó
NOM_SEDE	Se eliminó
CIU_SEDE	Se eliminó
EDO_SEDE	Se eliminó
TIPO_SEDE	Se eliminó
ID_INSTITU	Se eliminó
ID_CAMPUS	Se eliminó
ID_FACULTA	Se eliminó
FOLIO	Se eliminó
MATRICULA	Este atributo es el diferenciador de cada alumno, se eliminó
NOM_LIC	Los nombres de licenciatura son equivalentes al TIPO_EXA la variable se eliminó
PERFIL	Se eliminó
Duplicados	Se eliminó
DIA_NAC	Se eliminó
MES_NAC	Se eliminó
ANO_NAC	Se calculo la edad de los estudiantes y despues se transformo a cuartiles, al dato quedó 0.25, 0.50 y .075
SEXO	El 2 se transformó a 0
ENTI_NAC	Se eliminó
LENG_MA	El 2 se transformó a 0
LENG_PA	El 2 se transformó a 0
NOM_PROC	Se eliminó
PLT_PROC	Se transformó el tipo de examen al porcentaje de estudiantes que aprobaron, la clasificación quedó de 0 a 1
CIU_PROC	Se eliminó

EDO_PROC	Se eliminó
CVE_PROC	Se eliminó
RAZ_RAPT	El 2 se transformó a 0
RAZ_RACT	El 2 se transformó a 0
RAZ_OPT	El 2 se transformó a 0
RAZ_PCUM	El 2 se transformó a 0
RAZ_CMNF	El 2 se transformó a 0
REG_PROC	El 2=privado se transformó a 0
PRO_LIC	Se convirtieron a los porcentajes de los estudiantes que pasaron según cada promedio
BECA_DAC	El 2 se transformó a 0
BECA_NEC	El 2 se transformó a 0
OPI_NIVE	1 y 2 se transformaron a 1 3 y 4 se transformaron a 0
OPI_SLC	1 y 2 se transformaron a 1 3 y 4 se transformaron a 0
OPI_SERI	1 y 2 se transformaron a 1 3 y 4 se transformaron a 0
OPI_NIBA	1 y 2 se transformaron a 1 3 y 4 se transformaron a 0
OPI_ARM	1 y 2 se transformaron a 1 3 y 4 se transformaron a 0
COM_ESES	1 y 2 se transformaron a 1 3 y 4 se transformaron a 0
COM_ESMA	1 y 2 se transformaron a 1 3 y 4 se transformaron a 0
COM_ESAE	1 y 2 se transformaron a 1 3 y 4 se transformaron a 0
COM_EPAD	1 y 2 se transformaron a 1 3 y 4 se transformaron a 0
FOPR_ASAC	1 y 2 se transformaron a 0 3, 4 y 5 se transformaron a 1

FOPR_TUTO	1 y 2 se transformaron a 0 3, 4 y 5 se transformaron a 1
FOPR_APPS	1 y 2 se transformaron a 0 3, 4 y 5 se transformaron a 1
FOPR_PRMO	1 y 2 se transformaron a 0 3, 4 y 5 se transformaron a 1
FOPR_ELEX	1 y 2 se transformaron a 0 3, 4 y 5 se transformaron a 1
FOPR_CLEX	1 y 2 se transformaron a 0 3, 4 y 5 se transformaron a 1
FOPR_AULA	1 y 2 se transformaron a 0 3, 4 y 5 se transformaron a 1
FOPR_LATA	1 y 2 se transformaron a 0 3, 4 y 5 se transformaron a 1
FOPR_BIBL	1 y 2 se transformaron a 0 3, 4 y 5 se transformaron a 1
FOPR_SACO	1 y 2 se transformaron a 0 3, 4 y 5 se transformaron a 1
FOPR_CAFE	1 y 2 se transformaron a 0 3, 4 y 5 se transformaron a 1
FOPR_ESDE	1 y 2 se transformaron a 0 3, 4 y 5 se transformaron a 1
FOPR_SDTE	1 y 2 se transformaron a 0 3, 4 y 5 se transformaron a 1
FOPR_ACIN	1 y 2 se transformaron a 0 3, 4 y 5 se transformaron a 1
FOPR_ACEL	1 y 2 se transformaron a 0 3, 4 y 5 se transformaron a 1
FOPR_ABRE	1 y 2 se transformaron a 0 3, 4 y 5 se transformaron a 1
FOPR_BOTR	1 y 2 se transformaron a 0 3, 4 y 5 se transformaron a 1
FOPR_LIMP	1 y 2 se transformaron a 0 3, 4 y 5 se transformaron a 1

FOME_PECR	1 y 2 se transformaron a 0 3 y 4 se transformaron a 1
FOME_APAU	1 y 2 se transformaron a 0 3 y 4 se transformaron a 1
FOME_CARE	1 y 2 se transformaron a 0 3 y 4 se transformaron a 1
FOME_FOPR	1 y 2 se transformaron a 0 3 y 4 se transformaron a 1
FOME_FOTE	1 y 2 se transformaron a 0 3 y 4 se transformaron a 1
FOME_EXOR	1 y 2 se transformaron a 0 3 y 4 se transformaron a 1
FOME_EXES	1 y 2 se transformaron a 0 3 y 4 se transformaron a 1
FOME_IDEX	1 y 2 se transformaron a 0 3 y 4 se transformaron a 1
FOME_MDLT	1 y 2 se transformaron a 0 3 y 4 se transformaron a 1
FOME_HBSI	1 y 2 se transformaron a 0 3 y 4 se transformaron a 1
FOME_ABOR	1 y 2 se transformaron a 0 3 y 4 se transformaron a 1
FOME_ABRP	1 y 2 se transformaron a 0 3 y 4 se transformaron a 1
FOME_VCDE	1 y 2 se transformaron a 0 3 y 4 se transformaron a 1
FOME_LIDE	1 y 2 se transformaron a 0 3 y 4 se transformaron a 1
FOME_CMPP	1 y 2 se transformaron a 0 3 y 4 se transformaron a 1
FOME_TODE	1 y 2 se transformaron a 0 3 y 4 se transformaron a 1
FOME_TRCO	1 y 2 se transformaron a 0 3 y 4 se transformaron a 1

FOME_CUGE	1 y 2 se transformaron a 0 3 y 4 se transformaron a 1
FOME_REIN	1 y 2 se transformaron a 0 3 y 4 se transformaron a 1
FOME_EOAD	1 y 2 se transformaron a 0 3 y 4 se transformaron a 1
FOME_CANE	1 y 2 se transformaron a 0 3 y 4 se transformaron a 1
FOME_FTGP	1 y 2 se transformaron a 0 3 y 4 se transformaron a 1
FOME_RESO	1 y 2 se transformaron a 0 3 y 4 se transformaron a 1
FOME_COET	1 y 2 se transformaron a 0 3 y 4 se transformaron a 1
FOME_CECC	1 y 2 se transformaron a 0 3 y 4 se transformaron a 1
FOME_CCMA	1 y 2 se transformaron a 0 3 y 4 se transformaron a 1
MAE_DMAT	1 y 2 se transformaron a 0 3 y 4 se transformaron a 1
MAE_UEDA	1 y 2 se transformaron a 0 3 y 4 se transformaron a 1
MAE_ASIS	1 y 2 se transformaron a 0 3 y 4 se transformaron a 1
MAE_CLAP	1 y 2 se transformaron a 0 3 y 4 se transformaron a 1
MAE_PPDE	1 y 2 se transformaron a 0 3 y 4 se transformaron a 1
MAE_TAR	1 y 2 se transformaron a 0 3 y 4 se transformaron a 1
MAE_RET	1 y 2 se transformaron a 0 3 y 4 se transformaron a 1
MAE_HEC	1 y 2 se transformaron a 0 3 y 4 se transformaron a 1

MAE_FAC	1 y 2 se transformaron a 0 3 y 4 se transformaron a 1
MAE_ICC	1 y 2 se transformaron a 0 3 y 4 se transformaron a 1
MAE_VNC	1 y 2 se transformaron a 0 3 y 4 se transformaron a 1
MAE_DCO	1 y 2 se transformaron a 0 3 y 4 se transformaron a 1
MAE_ICI	1 y 2 se transformaron a 0 3 y 4 se transformaron a 1
MAE_RDE	1 y 2 se transformaron a 0 3 y 4 se transformaron a 1
MAE_EIC	1 y 2 se transformaron a 0 3 y 4 se transformaron a 1
MAE_FIN	1 y 2 se transformaron a 0 3 y 4 se transformaron a 1
PRA_PROF	1 y 2 se transformaron a 0 3 y 4 se transformaron a 1
PRPF_AACA	1 y 2 se transformaron a 0 3 y 4 se transformaron a 1
PRPF_MPNC	1 y 2 se transformaron a 0 3 y 4 se transformaron a 1
PRPF_MAET	1 y 2 se transformaron a 0 3 y 4 se transformaron a 1
SER_MAAC	1 y 2 se transformaron a 0 3, 4 y 5 se transformaron a 1
SER_MPNC	1 y 2 se transformaron a 0 3, 4 y 5 se transformaron a 1
SER_MAET	1 y 2 se transformaron a 0 3, 4 y 5 se transformaron a 1
TIEM_TRAB	1 y 2 se transformaron a 0 3 y 4 se transformaron a 1
SEMA_TRAB	1 y 2 se transformaron a 0 3 y 4 se transformaron a 1

INS_TRAB	El 1 se transformó en 0 el 2 y el 3 en 1
HRS_TRAB	1 y 2 se transformaron a 0 3 y 4 se transformaron a 1
REL_TRA	1 y 2 se transformaron a 0 3, 4 y 5 se transformaron a 1
EVA_ETRA	1 y 2 se transformaron a 0 3, 4 y 5 se transformaron a 1
GAN_CLIC	6 se transformó a 1 5 se transformó a 0.66 4 es 0.66 3 es 0.50 2 es 0.33 1 es 0
SIT_LAB	Se transformaron conforme a el porcentaje de alumnos con testimonio satisfactorio 4 es 0.50 3 es 0.30 2 es 0.17 1 es 0.03
ESC_MAD	Se transformaron conforme a el porcentaje de alumnos con testimonio satisfactorio 1 es 0.02 2 es 0.2 3 es 0.28 4 es 0.14 5 es 0.15 6 es 0.17 7 es 0.04 8 es 0.002
ESC_PAD	Se transformaron conforme a el porcentaje de alumnos con testimonio satisfactorio 1 es 0.02 2 es 0.19 3 es 0.22 4 es 0.15 5 es 0.08 6 es 0.24 7 es 0.06 8 es 0.02

CUAN_LIB	Se transformaron conforme a el porcentaje de alumnos con testimonio satisfactorio 1 es 0.005 2 es 0.13 3 es 0.22 4 es 0.23 5 es 0.19 6 es 0.12 7 es 0.07 8 es 0.04
NIV_EDU	Se transformaron conforme a el porcentaje de alumnos con testimonio satisfactorio 1 es 0.034 2 es 0.02 3 es 0.35 4 es 0.59
CPV_CASA	Se transformaron conforme a el porcentaje de alumnos con testimonio satisfactorio 1 es 0.015 2 es 0.08 3 es 0.19 4 es 0.28 5 es 0.26 6 es 0.19
SER_TELE	El 2=no se transformó a 0
SER_LAV	El 2=no se transformó a 0
SER_REF	El 2=no se transformó a 0
SER_HOR	El 2=no se transformó a 0
SER_INTE	El 2=no se transformó a 0
SER_TVP	El 2=no se transformó a 0
SER_TABL	El 2=no se transformó a 0
CUARTOS	Se transformaron conforme a el porcentaje de alumnos con testimonio satisfactorio 1 es 0.02 2 es 0.22 3 es 0.41 4 es 0.22 5 es 0.09 6 es 0.04

SER_PC	1 y 2 se transformaron a 0 3, 4 y 5 se transformaron a 1
SER_TV	1 y 2 se transformaron a 0 3, 4 y 5 se transformaron a 1
SER_AUTO	1 y 2 se transformaron a 0 3, 4 y 5 se transformaron a 1
SER_BANO	1 y 2 se transformaron a 0 3, 4 y 5 se transformaron a 1
VAC_RM	1 y 2 se transformaron a 0 3, 4 y 5 se transformaron a 1
P1	Se eliminó
A1	Se eliminó
P2	Se eliminó
A2	Se eliminó
P3	Se eliminó
A3	Se eliminó
P4	Se eliminó
A4	Se eliminó
P5	Se eliminó
A5	Se eliminó
P6	Se eliminó
A6	Se eliminó
DICTAMEN	Se transformó Aún no satisfactorio es 0 Satisfactorio y sobresaliente es 1
VECES_EXAM	Se calculó el número de veces que un estudiante realizó el examen

Fuente: Elaboración propia