

UNIVERSIDAD PERUANA UNIÓN

FACULTAD DE INGENIERIA Y ARQUITECTURA

Escuela Profesional de Ingeniería Sistemas



Una Institución Adventista

Predicción del nivel de estrés en estudiantes universitarios utilizando técnicas de machine learning

Por:

Hernan Vilca Masco

Asesor:

Ing. Angel Rosendo Condori Coaquira

Juliaca, 02 de Diciembre 2019

DECLARACIÓN JURADA DE AUTORÍA DEL TRABAJO DE INVESTIGACIÓN

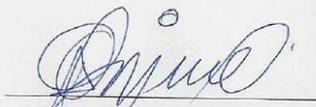
Angel Rosendo Condori Coaquira, de la Facultad de Ingeniería y Arquitectura, Escuela Profesional Ingeniería de Sistemas, de la Universidad Peruana Unión.

DECLARO:

Que el presente trabajo de investigación titulado: "PREDICCIÓN DEL NIVEL DE ESTRÉS EN ESTUDIANTES UNIVERSITARIOS UTILIZANDO TECNICAS DE MACHINE LEARNING" constituye la memoria que presenta el estudiante Hernan Vilca Masco para aspirar al grado académico de bachiller en Ingeniería de Sistemas, cuyo trabajo de investigación ha sido realizada en la Universidad Peruana Unión bajo mi dirección.

Las opiniones y declaraciones en este trabajo de investigación son de entera responsabilidad del autor(es), sin comprometer a la institución.

Y estando de acuerdo, firmo la presente declaración en Juliaca, a los 2 días del mes de diciembre del año 2019.



Ing. Angel Rosendo Condori
Coaquira

ACTA DE SUSTENTACIÓN DE TRABAJO DE INVESTIGACIÓN



En Puno, Juliaca, Villa Chullunquiani, a 2 día(s) del mes de diciembre del año 2019 siendo las 9:00 horas se reunieron los miembros del jurado en la Universidad Peruana Unión campus Juliaca, bajo la dirección del (de la) presidente(a) : Mtro. Lennin Henry Centurión Tulca el (la) secretario(a) : Mg. Abel Angel Sullen Macalupu y los demás miembros : Ing. Eder Gutierrez Quispe y el (la) asesor(a)

..... con el propósito de administrar el acto académico de sustentación del trabajo de investigación titulado: Reducción del nivel de estrés en estudiantes universitarios utilizando técnicas de machine learning

..... de los (las) egresados (as): a) Hernan Vilca Masco b)

..... conducente a la obtención del grado académico de Bachiller en Ingeniería de Sistemas (Denominación del Grado Académico de Bachiller)

El Presidente inició el acto académico de sustentación invitando candidato(a)/s hacer uso del tiempo determinado para su exposición. Concluida la exposición, el Presidente invitó a los demás miembros del jurado a efectuar las preguntas, y aclaraciones pertinentes, las cuales fueron absueltas por candidato(a)/s. Luego, se produjo un receso para las deliberaciones y la emisión del dictamen del jurado.

Posteriormente, el jurado procedió a dejar constancia escrita sobre la evaluación en la presente acta, con el dictamen siguiente:

Candidato/a (a): Hernan Vilca Masco

CALIFICACIÓN	ESCALAS			Mérito
	Vigesimal	Literal	Cualitativa	
<u>Aprobado</u>	<u>14</u>	<u>C</u>	<u>acceptable</u>	<u>Bueno</u>

Candidato/a (b):

CALIFICACIÓN	ESCALAS			Mérito
	Vigesimal	Literal	Cualitativa	

(*) Ver parte posterior

Finalmente, el Presidente del jurado invitó candidato(a)/s a ponerse de pie, para recibir la evaluación final y concluir el acto académico de sustentación procediéndose a registrar las firmas respectivas.

[Firma]
Presidente/a

[Firma]
Miembro

[Firma]
Secretario/a

Asesor/a

Miembro

[Firma]
Candidato/a (a)

Predicción del nivel de estrés en estudiantes universitarios utilizando técnicas de machine learning

Vilca Masco Hernan¹, Ing. Condori Coaquira Angel Rosendo²*

^aEP Ingeniería de Sistemas, Facultad Ingeniería y Arquitectura, Universidad Peruana Unión

Resumen:

El diagnóstico en el sector salud es de suma importancia para el tratamiento de enfermedades de salud mental como el estrés, las tecnologías de inteligencia artificial específicamente en machine learning están cambiando el mundo en el sector de la salud mental el tratamiento de los datos con técnicas de aprendizaje de máquina son herramientas que apoyan a la toma de decisiones con una mayor confiabilidad de los profesionales del área sector salud. El uso de librerías y lenguajes de programación libres de licencias permiten tener un mejor control y desarrollar herramientas de software con bajos costos que permitan su uso por profesionales de la salud y personas libre de pagos de licencias a terceros. La metodología se basó en la propuesta de IBM Knowledge denominada CRISP-DM que consta de las siguientes fases: Comprensión del negocio, Comprensión de los datos, Preparación de los datos, Modelado, Evaluación y distribución, todo esto con el objetivo de entrenar un modelo computacional basado en Aprendizaje de Máquina para predecir del nivel de estrés en estudiantes universitarios.

Palabras claves: Machine Learning, Limpieza de datos, Estrés, Predicción, Instrumento del estrés.

Abstract:

The diagnosis in the health sector is of utmost importance for the treatment of mental health diseases such as stress, artificial intelligence technologies specifically in machine learning are changing the world in the mental health sector the treatment of data with techniques of Machine learning are tools that support decision-making with greater reliability for

* Autor de correspondencia: Vilca Masco Hernan
Km. 6 Carretera Arequipa. Villa Chullunquiani.
Teléfono 947014960
E-mail: hernan.vilca@upeu.edu.pe

professionals in the health sector. The use of license-free libraries and programming languages allow better control and development of low-cost software tools that allow their use by health professionals and individuals free of license fees to third parties. The methodology was based on the IBM Knowledge proposal called CRISP-DM that consists of the following phases: Business understanding, Data understanding, Data preparation, Modeling, Evaluation and distribution, all with the objective of training a model Computational based on Machine Learning to predict stress level in college students.

Keywords: Machine Learning, Data cleaning, Stress, Prediction, Stress instrument.

1. Introducción

Machine Learning (ML) es un campo de estudio de la rama de evolución de algoritmos que brinda a las computadoras la capacidad de aprender sin ser programado explícitamente y emular la inteligencia humana aprendiendo del entorno de los datos (Naqa & Murphy, 2015).

El estrés académico es todo aquello que es producido por la demanda que se exige en el ámbito educativo (Caldera Montes & Pulido Castro, 2007) así mismo (Nathaly Berrío García; Rodrigo Mazo Zea, 2011) define que “El estrés académico es generada por las demandas que imponen el ámbito educativo. En consecuencia, tanto alumnos como docentes, pueden verse afectados por este problema” Por ende el estrés es estudiado desde múltiples disciplinas como psicología, biología, tecnologías avanzadas, etc.

Según los datos de la Organización Mundial de la Salud (OMS), una de cada cuatro personas, 450 millones en el mundo, sufren de al menos un trastorno mental. El 50% de peruanos con problemas de salud mental no considera el estrés o la depresión u otras enfermedades relacionadas como problemas de salud (Hermoza, 2015) Así mismo (Redaccionlr, 2018) menciona que “En un estudio del Instituto de Opinión Pública de la Pontificia Universidad Católica del Perú (PUCP) señala que el 80% de peruanos sufren o han sufrido estrés”.

En el trabajo realizado por (Dueñas Zuñiga, 2017) denominado “Estrés académico y estrategias de afrontamiento en estudiantes de la universidad nacional del altiplano puno”, El

método que utilizo dicha investigación se realizó con una encuesta utilizando instrumentos psicológicos que cumplan la validez y la fiabilidad en donde tomo 377 muestras de los cuales el resultado fue que el 75.3% de los estudiantes se encuentran con algún nivel de estrés.

En el trabajo realizado por (Barraza, 2008) su tipo de estudio fue descriptivo, el universo estuvo constituido por los alumnos de educación media superior que asisten a las instituciones con bachillerato general, los resultados fueron que el 85% de los alumnos reconocen la presencia de estrés académico, las evaluaciones de los profesores 92%, los estresores tienen que ver con la competencia con compañeros del grupo 91%, el tipo de trabajo que piden los maestros 91%, sobrecarga de tareas y trabajos escolares y exceso de responsabilidades por cumplir las obligaciones escolares con 90%, los síntomas que se presentan con mayor intensidad en los alumnos son los psicológicos.

El objetivo de esta investigación utilizar los datos psicológicos para analizar y pronosticar el nivel de estrés y así ayudar a mejorar el tratamiento a los profesionales de la salud. por lo que (Aragón, 2017) correlaciona que el estrés es la enfermedad del ciclo XXI, para ello es de suma importancia detectar sus síntomas, tratarlos y evitar repercusiones más graves, es por ello que se propone utilizar aprendizaje maquina ya que nos ayuda a descubrir patrones de la enfermedad de salud mental, además el estrés que es una enfermedad de salud mental está muy ligada con enfermedades de ansiedad y depresión lo cual es de suma importancia de descubrir los patrones y así poder predecir los niveles de estrés.

2. Materiales y Métodos

La presente investigación se desarrolló en los alumnos de una Universidad en la ciudad de Juliaca, en donde investigador envió un test (Instrumento) mediante las redes sociales y correos para así llegar a los estudiantes de dicha Universidad, la encuesta se realizó en el formulario de google para la recopilación de los datos que en total respondieron 50 estudiantes. Para ello se pretende predecir enfermedades mentales del estrés utilizando técnicas de machine learning para la respectiva predicción del nivel de estrés académico.

2.1 Software y Librerías

Arquitectura: en esta investigación realizada se propuso una arquitectura de solución *Figura 1* donde se utilizó una data de validación y data de entrenamiento para luego usar los algoritmos de árboles de decisión y mostrar los resultados y la predicción del modelo.

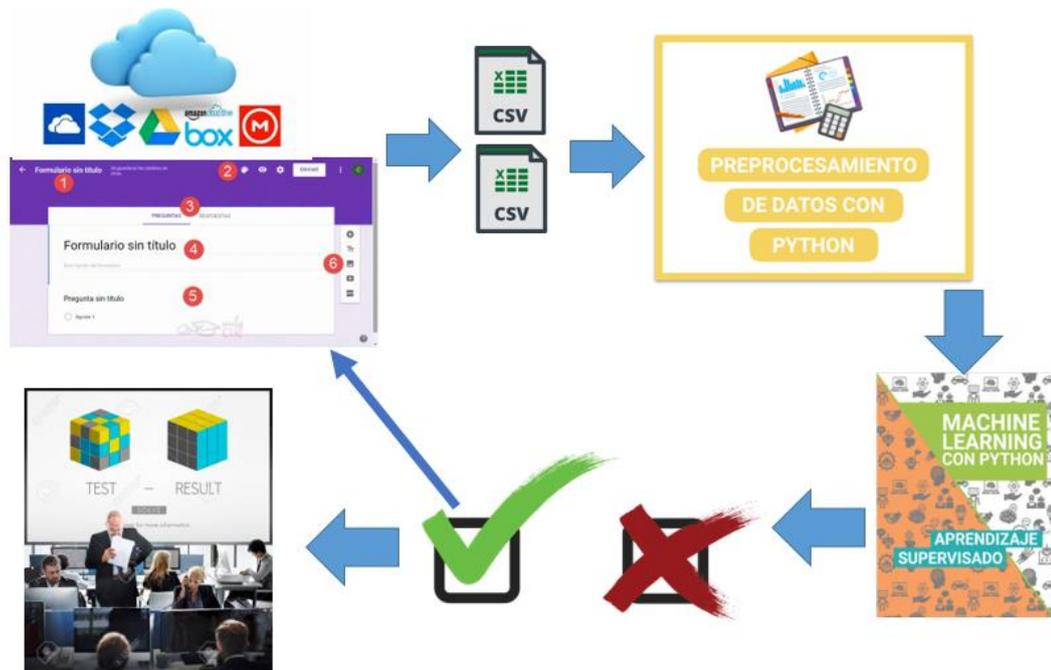


Figura 1. arquitectura de solución para realizar la investigación.

Materiales utilizados para el desarrollo de la investigación

Tabla 1

Software y librerías

Materiales	Versión		Descripción
Python	3.7	Libre	Lenguaje de programación orientado a la ciencia de datos
Azure Notebooks	4.3	Libre	Servicio gratuito en la nube para desarrollar y ejecutar en lenguajes de Python, R.
Pandas	0.25	Libre	Leer y escribir datos en diferentes formatos.
Sklearn	0.21	Libre	Algoritmos populares de machine

Matplotlib	3.1	Libre	Genera diagramas, histogramas, espectros de potencia, gráficos de barras, diagramas de error, diagramas de dispersión, etc.
Numpy	1.17	Libre	Garantiza cálculos eficientes con matrices.
Graphviz	2.38	Libre	Archivo para generar DOT.

Fuente: Elaboración Propia

2.2 Entender el Negocio

Es importante que se tenga en cuenta que para diseñar el modelo computacional es importante conocer los parámetros de que con llevan al estrés los cuales están definidos en estudios anteriores, se seleccionó el instrumento de medición de estrés de MSI, y se usó una escala de 6 niveles que son las siguientes: 1 nunca, 2 casi nunca, 3 pocas veces, 4 algunas veces, 5 relativamente frecuente, 6 muy frecuente.

2.2.1 Instrumento

Permite al investigador entender las preguntas y medir el nivel de estrés de las personas que responden dicha encuesta.

	1 Nunca	2 Casi nunca	3 Pocas veces	4 Algunas veces	5 Relativamente frecuente	6 Muy frecuente
Imposibilidad de conciliar el sueño.	1	2	3	4	5	6
Jaquecas y dolores de cabeza.	1	2	3	4	5	6
Indigestiones o molestias gastrointestinales.	1	2	3	4	5	6
Sensación de cansancio extremo o agotamiento.	1	2	3	4	5	6
Tendencia de comer, beber o fumar más de lo habitual.	1	2	3	4	5	6
Disminución del interés sexual.	1	2	3	4	5	6
Respiración entrecortada o sensación de ahogo.	1	2	3	4	5	6
Disminución del apetito.	1	2	3	4	5	6
Temblores musculares (por ejemplo tics nerviosos o parpadeos).	1	2	3	4	5	6
Pinchazos o sensaciones dolorosas en distintas partes del cuerpo.	1	2	3	4	5	6
Tentaciones fuertes de no levantarse por la mañana.	1	2	3	4	5	6
Tendencias a sudar o palpitaciones.	1	2	3	4	5	6

Figura 2. Preguntas que se hicieron en google encuestas.

2.2.2 Formulario Google

El formulario de google nos permite realizar encuestas, es muy importante en la recopilación de los datos de una manera eficaz y rápida. Lo que facilita al investigador a recolectar datos de las personas que respondan la encuesta y esto se almacena en una hoja de cálculo para luego realizar los pasos limpieza de datos.

2.2.3 Azure notebooks

Para ya ejecutar o importar los datos se utilizó la herramienta de azure notebooks que nos facilita poder trabajar en la nube (Internet) sin la necesidad estar preocupado por un ordenador, ya que tiene todas las herramientas para realizar machine learning o inteligencia artificial.

2.3 Entender y preparar la Data

Para realizar la limpieza de datos se ejecutaron los siguientes pasos:

- Descargar la data del formulario de google tal como está.
- Se creó una cuenta en azure notebooks para posteriormente realizar la limpieza de datos con el lenguaje de programación de Python.
- Importación de librerías como numpy, pandas, sklearn, matplotlib.
- Importación de los datos en csv.
- Visualización de datos.
- Función para convertir de "1 Nunca", "2 Casi Nunca", "3 Pocas Veces", "4 Algunas Veces", "5 Relativamente Frecuente", "6 Muy Frecuente" a un formato numérico 1,2,3,4,5,6.
- Función para convertir de "SinEstres", "Leve", "Medio", "Alto", "Grave" a un formato numérico 1,2,3,4,5.
- Visualización de los datos en formato numérico.
- Función para convertir los números sumados a letras.
- Exportar data ya limpiada a formato csv o seguir con el procedimiento.

Marca temporal	Imposibilidad de conciliar el sueño.	Jaquecas y dolores de cabeza.	Indigestiones o molestias gastrointestinales.	Sensación de cansancio extremo o agotamiento	Tendencia de comer, beber o fumar más de lo habitual.	Disminución del interés sexual.	Respiración entrecortada o sensación de ahogo.	Disminución del apetito.
0 2019/10/11 12:43:53 p.m. GMT-5	2 Casi Nunca	2 Casi Nunca	2 Casi Nunca	4 Algunas Veces	4 Algunas Veces	3 Pocas Veces	2 Casi Nunca	1 Nunca
1 2019/10/11 1:44:59 p.m. GMT-5	4 Algunas Veces	6 Muy Frecuente	4 Algunas Veces	6 Muy Frecuente	1 Nunca	1 Nunca	2 Casi Nunca	1 Nunca
2 2019/10/11 1:50:11 p.m. GMT-5	3 Pocas Veces	5 Relativamente Frecuente	4 Algunas Veces	5 Relativamente Frecuente	2 Casi Nunca	1 Nunca	3 Pocas Veces	4 Algunas Veces
3 2019/10/11 1:54:38 p.m. GMT-5	3 Pocas Veces	3 Pocas Veces	3 Pocas Veces	4 Algunas Veces	4 Algunas Veces	2 Casi Nunca	3 Pocas Veces	2 Casi Nunca

Figura 3. Features(Características) data.
Fuente: Propia

2.4 Preparar la Data

El pre procesamiento se desarrolló con el lenguaje de Python.

	pre1	pre2	pre3	pre4	pre5	pre6	pre7	pre8	pre9	pre10	pre11	pre12	nivel	Puntaje
0	2	2	2	4	4	3	2	1	2	2	3	1	Leve	28
1	4	6	4	6	1	1	2	1	1	5	6	3	Medio	40
2	3	5	4	5	2	1	3	4	2	3	5	3	Medio	40
3	3	3	3	4	4	2	3	2	3	2	4	3	Medio	36
4	1	1	1	2	1	2	1	1	1	1	5	1	SinEstres	18

Figura 4. Data pre procesados con Python limpia.
Fuente: Propia

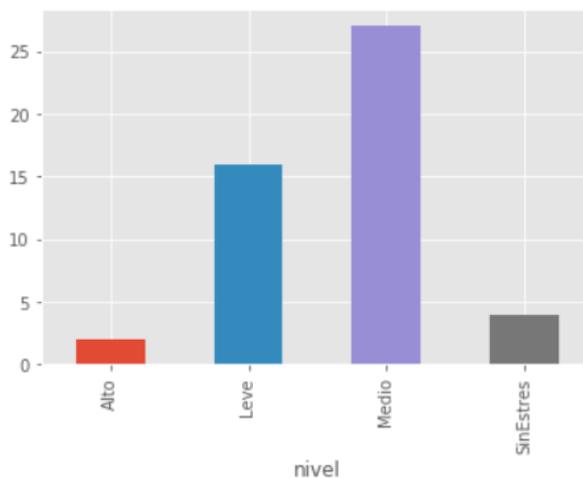


Figura 5. Visualización del nivel de estrés según a los datos.
Fuente: Propia

2.5 Modelo Algoritmo

Para realizar el entrenamiento del algoritmo se utilizó algoritmo supervisado de árboles de decisión ya que es el más adecuado de acuerdo a los datos.

- Se indicó las variables predictores las preguntas correspondientes del 1 al 12 que se abrevio como: "pre1", "pre2", "pre3", "pre4", "pre5", "pre6", "pre7", "pre8", "pre9", "pre10", "pre11", "pre12".
- Se indicó la variable del target que en este caso los niveles de estrés: "SinEstres", "Leve", "Medio", "Alto", "Grave".
- Se creó las variables para el entrenamiento del árbol y para las pruebas: X_entrenamiento, X_test, y_entrenamiento, y_tes.
- Se Construyó el árbol con los datos de entrenamiento.
- Se verificó la probabilidad de predicción del árbol con los datos de prueba.
- Se verifico la probabilidad de predicción del árbol con los datos de entrenamiento.

Tabla 2
Data targets y predictores.

Predictores	Targets
Pre1 = Imposibilidad de conciliar el sueño	Sin estrés, leve, medio, alto, grave
Pre2 = Jaquecas y dolores de cabeza.	Sin estrés, leve, medio, alto, grave
Pre3 = Indigestiones o molestias gastrointestinales.	Sin estrés, leve, medio, alto, grave
Pre4 = Sensación de cansancio extremo o agotamiento.	Sin estrés, leve, medio, alto, grave
Pre5 = Tendencia de comer, beber o fumar más de lo habitual.	Sin estrés, leve, medio, alto, grave
Pre6 = Disminución del interés sexual.	Sin estrés, leve, medio, alto, grave
Pre7 = Respiración entrecortada o sensación de ahogo	Sin estrés, leve, medio, alto, grave
Pre8 = Disminución del apetito	Sin estrés, leve, medio, alto, grave
Pre9 = Temblores musculares (por ejemplo tics nerviosos o parpadeos).	Sin estrés, leve, medio, alto, grave
Pre10 = Pinchazos o sensaciones dolorosas en distintas partes del cuerpo.	Sin estrés, leve, medio, alto, grave

Pre11 = **Tentaciones fuertes de no levantarse por la mañana.** Sin estrés, leve, medio, alto, grave

Pre12 = **Tendencias a sudar o palpitaciones.** Sin estrés, leve, medio, alto, grave

Fuente: Propia.

2.6 Validación

Para realizar la validación del árbol se utilizó k-fold lo cual nos ayuda a poder a cuanta profundidad el árbol es adecuado.

Tabla 4

Según k-fold

Profundidad	Accuracy
1	49%
2	61%
3	61%
4	61%
5	61%

Fuente: Propia

3. Resultados

Para el desarrollo del algoritmo de machine learning se trabajó con la profundidad de 4 porque el algoritmo de validación cruzada nos indicó que a partir de la profundidad 2 es el adecuado ya que vale un 61%. Así mismo el árbol llegó a una precisión del 97% lo cual nos indica que es perfecto para este tipo de datos lo cual servirá de mucha ayuda para la toma de decisiones.

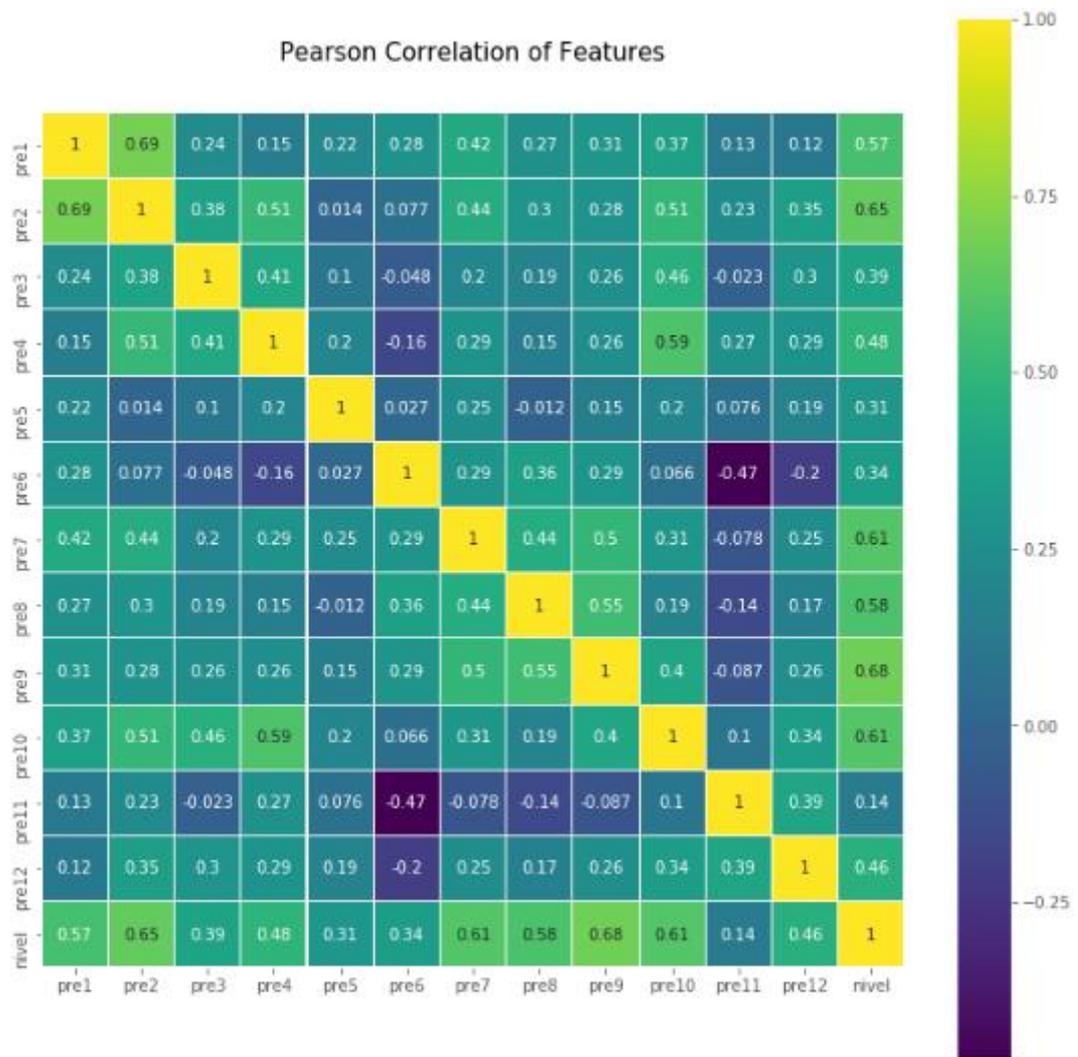


Figura 6. Correlación de Pearson.
Fuente: Propia

Tabla 5

modelo algoritmo.

Profundidad	Modelo	Accuracy
4	DecisionTreeClassifier	97%

Fuente: Propia

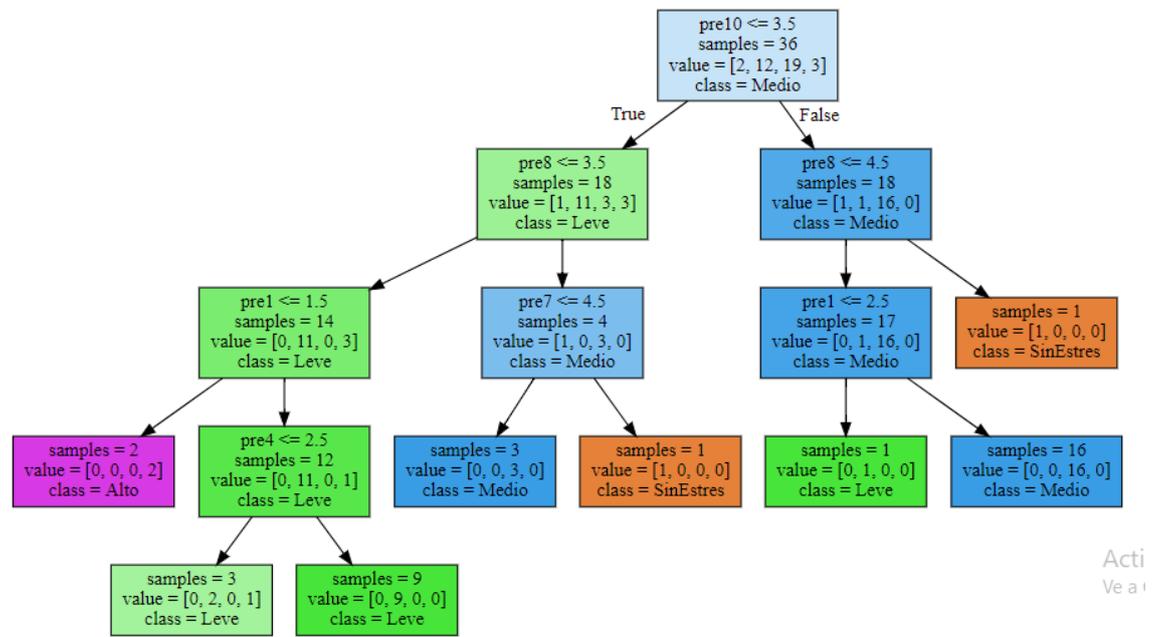


Figura 7. Visualización del Árbol.
Fuente: Propia

4. Conclusiones y discusiones

Según los resultados encontrados se concluye que los estudiantes de la Universidad donde se probó con técnicas de Machine Learning usando el modelo de predicción árbol de decisión la cual arroja una precisión del 97% lo cual demuestra que se puede aplicar estas técnicas de Aprendizaje de Maquina para aplicar a este tipo de problemas como el estrés, así mismo se puede alimentar el modelo con una mayor cantidad de datos con los cual este pueda aprender y de esa manera pueda ayudar en la toma de decisiones, por ende en un futuro se propone diseñar e implementar una interfaz de usuario que permita la facilidad de uso mediante una interfaz web o móvil que interactúe con las personas de una forma sencilla y clara. Finalmente, las aplicaciones de machine learning en el ámbito de la Psicología están siendo herramientas de gran ayuda para el trabajo de profesionales en el tratamiento de enfermedades de salud.

5. Referencias

- Aragón, D. del A. (2017). Estrés, enfermedad del siglo XXI. *Diario Del AltoAragón*, 15. Retrieved from <https://www.cofhuesca.com/archivos/noticia/3158/4236-171126-estres-enfermedad-del-siglo-xxi.pdf>
- Caldera Montes, J. F., & Pulido Castro, B. E. (2007). Niveles de estrés y rendimiento académico en estudiantes de la carrera de Psicología del Centro Universitario de Los Altos. *Revista De Educacion Y Desarrollo*, 77–82.
- Dueñas Zuñiga, H. F. (2017). Estrés Académico Y Estrategias De Afrontamiento En Estudiantes De La Universidad Nacional Del Altiplano Puno. *Repositorio Institucional UNA-PUNO*, 1–113.
- Hermoza, D. (2015). El estrés, aliado de la pobreza de Latinoamérica. Retrieved October 28, 2019, from <https://www.bancomundial.org/es/news/feature/2015/12/28/el-estres-aliado-de-la-pobreza-de-latinoamerica>
- Macías, A. B. (2008). Academic stress in master's students and its modulatory variables: A between-groups design. *Avances En Psicología Latinoamericana*, 26(2), 270–289.
- Naqa, I. El, & Murphy, M. J. (2015). Machine Learning in Radiation Oncology. *Machine Learning in Radiation Oncology*, 3–11. <https://doi.org/10.1007/978-3-319-18305-3>
- Nathaly Berrío García;Rodrigo Mazo Zea. (2011). *Estrés académico*. Retrieved from http://www.ghbook.ir/index.php?name=فرهنگ و رسانه های نوین&option=com_dbook&task=readonline&book_id=13650&page=73&chckhashk=ED9C9491B4&Itemid=218&lang=fa&tmpl=component
- Redaccionlr. (2018). El estrés, ese enemigo silencioso que afecta al 80% de peruanos | Sociedad - La República. Retrieved October 28, 2019, from <https://larepublica.pe/sociedad/1218645-el-estres-ese-enemigo-silencioso-que-afecta-al-80-de-peruanos/>