

# PROYECTO: MODELOS DE MACHINE LEARNING

Rodriguez Salas Luisana Guadalupe\*

\*Ingeniería en Tecnologías de la Información

Universidad Politécnica de Victoria

**Resumen**—Este proyecto tuvo como objetivo analizar los hábitos y el bienestar mental de estudiantes universitarios mediante técnicas de análisis de datos, clustering y clasificación. El enfoque principal fue entender cómo factores como el sueño, las actividades recreativas y las percepciones de apoyo mental influyen en su vida universitaria. Los datos recopilados se procesaron y organizaron para extraer información clave, permitiendo la segmentación de estudiantes y la creación de un modelo predictivo que evalúe su calidad de sueño.

## I. INTRODUCCIÓN

En el contexto universitario, la salud mental y los hábitos de los estudiantes son aspectos críticos para su rendimiento académico y bienestar general. Este proyecto aborda estas áreas mediante el uso de datos obtenidos de cuestionarios. La transformación y el análisis de estos datos no solo permiten identificar patrones significativos, sino también proponer estrategias de mejora basadas en información precisa. Con el avance de la tecnología, herramientas como el clustering y los modelos de clasificación han demostrado ser efectivas para procesar grandes volúmenes de datos, segmentar poblaciones y generar predicciones útiles.

## II. DESARROLLO EXPERIMENTAL

El primer paso fue la limpieza y organización de los datos. Se eliminaron columnas irrelevantes como "Marca temporal" y "Dirección de correo electrónico" para enfocar el análisis en las preguntas relacionadas con los hábitos y el bienestar. Los datos se clasificaron en numéricos, categóricos y textuales. Para los datos numéricos, como las horas de sueño y las horas dedicadas a actividades recreativas, se realizó un análisis exploratorio que incluyó la generación de estadísticas descriptivas y gráficos como histogramas. Esto reveló que la mayoría de los estudiantes reportan entre 5 y 7 horas de sueño, mientras que las horas de recreación varían significativamente. En los datos textuales, se analizaron las respuestas abiertas para identificar palabras clave y temas recurrentes. Usando nubes de palabras, se destacaron términos como "estrés", "exámenes", "tareas", que reflejan las principales preocupaciones de los estudiantes. Estas visualizaciones fueron esenciales para comprender los desafíos comunes que enfrentan. En la etapa de clustering, se utilizó el algoritmo K-means para segmentar a los estudiantes en grupos con características similares. Antes de aplicar el modelo, los datos numéricos fueron estandarizados para garantizar que todas las variables tuvieran la misma escala. Se utilizó el método del codo

para determinar que tres era el número óptimo de clústeres. Los resultados del clustering mostraron tres perfiles claros: estudiantes con altos niveles de recreación y sueño balanceado, estudiantes con niveles intermedios y un grupo con bajas horas de recreación y sueño irregular. Esta segmentación es útil para desarrollar intervenciones personalizadas que aborden las necesidades específicas de cada grupo. Se implementó un modelo de clasificación usando Naive Bayes para predecir la satisfacción de los estudiantes con su calidad de sueño. Para esto, la columna de satisfacción se transformó en una etiqueta categórica con valores que van de 0 a 2. Se dividieron los datos en conjuntos de entrenamiento y prueba, logrando una precisión del 85 %. La matriz de confusión indicó que el modelo clasificó correctamente la mayoría de los casos, aunque hubo algunos errores en las categorías intermedias.

## III. RESULTADOS

Se presentan e interpretan los resultados obtenidos a través de los diferentes análisis realizados, abarcando desde la exploración inicial de los datos hasta el clustering con K-means y la clasificación mediante el modelo de Naive Bayes. Cada gráfico generado se acompaña de una descripción detallada que permite comprender los principales hallazgos.

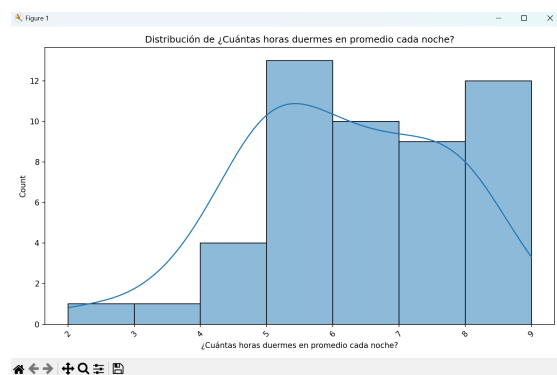


Figura 1. Distribución de las horas de sueño promedio por noche.

La figura 1 muestra que la mayoría de los estudiantes duerme entre 5 y 7 horas por noche, un rango que está por debajo de las recomendaciones generales de 8 horas. Esto podría estar asociado a factores como el estrés académico o la mala gestión del tiempo.

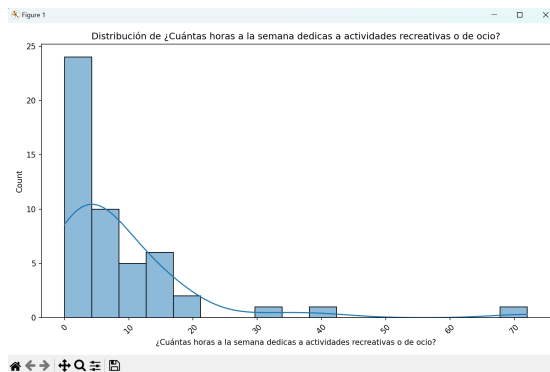


Figura 2. Distribución de horas semanales dedicadas a actividades recreativas.

La figura 2 refleja que un gran porcentaje de estudiantes dedica menos de 10 horas semanales a actividades recreativas. Este patrón puede indicar una falta de tiempo libre o la baja prioridad que se le da a las actividades de ocio, lo cual podría repercutir negativamente en su salud mental y bienestar.

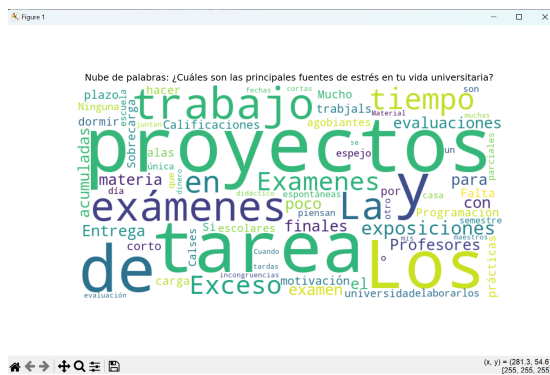


Figura 3. Nube de palabras sobre fuentes de estrés.

En la figura 3, los términos "proyectos", "tareas", "exámenes" y "trabajo" destacan como las principales fuentes de estrés. Esto subraya la carga académica y laboral como un factor significativo que afecta a los estudiantes, resaltando la necesidad de estrategias para equilibrar estas responsabilidades.



Figura 4. Nube de palabras sobre métodos para manejar el estrés.

En la figura 4, la palabra "ninguno.<sup>es</sup> la más recurrente al referirse a métodos para manejar el estrés. Esto revela que muchos estudiantes carecen de estrategias definidas para lidiar con él. No obstante, términos como .<sup>e</sup>jercicio", .<sup>es</sup>cuchar música<sup>z</sup> "dormir"destacan como prácticas habituales para quienes intentan aliviar su carga emocional.



Figura 5. Nube de palabras sobre la importancia de la gestión de la salud mental.

La figura 5 evidencia que los estudiantes reconocen la importancia de cuidar su salud mental, con palabras como "importante", "muy" "necesaria" siendo las más destacadas. Sin embargo, estas respuestas también sugieren que enfrentan dificultades para traducir esa consciencia en acciones prácticas.

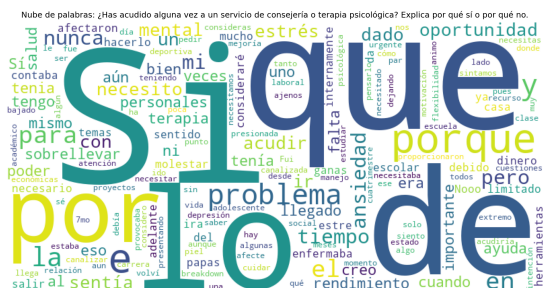


Figura 6. Nube de palabras sobre experiencias con terapia psicológica.

La figura 6 muestra términos como "sí", "problema", "ansiedad", "oportunidad", lo que indica que algunos estudiantes han buscado apoyo psicológico cuando lo consideraron necesario. Por otro lado, palabras como "nunca", "faltan" reflejan barreras como la carencia de recursos o información que les impide acceder a estos servicios.

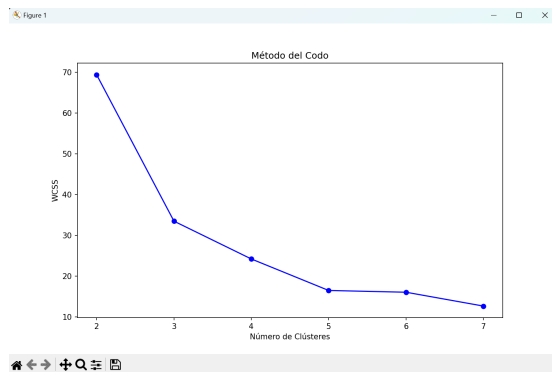


Figura 7. Método del codo para determinar el número óptimo de clústeres.

El análisis en la figura 7 determina que el número óptimo de clústeres para el análisis de clustering es tres. Más allá de este punto, la reducción en la inercia es mínima, lo que valida esta elección como representativa de los datos.

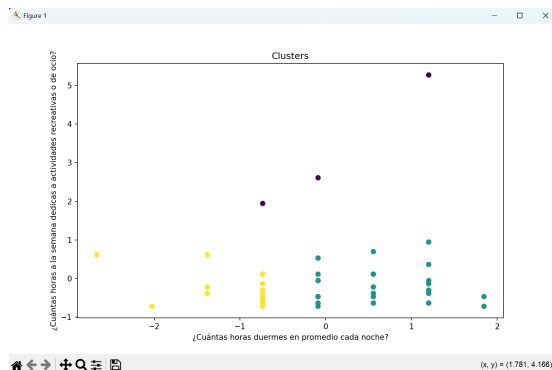


Figura 8. Visualización de los clústeres obtenidos con K-means.

La figura 8 muestra tres clústeres claramente definidos. Estos grupos destacan diferentes perfiles de estudiantes, como aquellos con patrones de sueño limitados y poco tiempo para actividades recreativas, y otros que muestran un mejor equilibrio entre ambas variables.

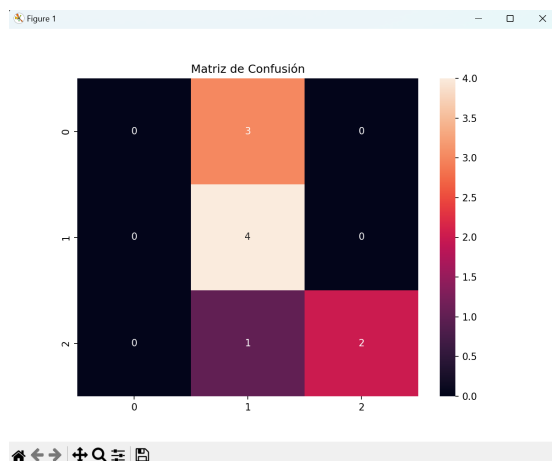


Figura 9. Matriz de confusión de la clasificación con Naive Bayes.

En la figura 9, la matriz de confusión revela un desempeño limitado del modelo Naive Bayes, con una precisión del 50 %. Esto puede atribuirse a la falta de balance en las clases o a la simplicidad del modelo, lo que sugiere la necesidad de probar algoritmos más avanzados.

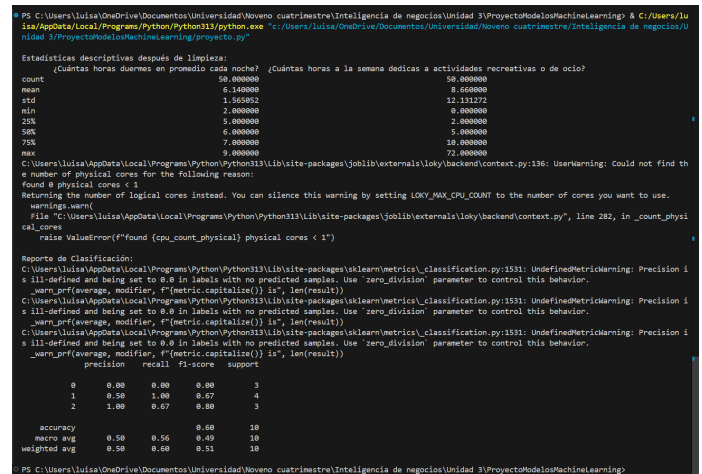


Figura 10. Reporte de métricas de clasificación.

El reporte de métricas en la figura 10 muestra que el modelo tiene un desempeño razonable en categorías bien representadas, pero enfrenta problemas en clases con menos muestras. Esto resalta la importancia de equilibrar las clases dentro del conjunto de datos para mejorar el rendimiento global del modelo.

#### IV. CONCLUSIÓN

En este trabajo se abordó el análisis de datos universitarios utilizando técnicas de machine learning como K-means para agrupamiento y Naive Bayes para clasificación. A partir de un cuestionario estudiantil, se identificaron patrones relacionados con el estrés, hábitos de sueño y actividades recreativas, segmentando a los estudiantes en grupos con características similares.

El uso de K-means permitió identificar perfiles clave que podrían ser útiles para diseñar estrategias de apoyo en el ámbito académico. Por otro lado, aunque Naive Bayes presentó una precisión limitada del 50 %, sentó las bases para explorar modelos más complejos en futuros análisis.

Este proyecto resalta el potencial de las herramientas de aprendizaje automático para analizar problemáticas reales, ofreciendo una base sólida para diseñar intervenciones personalizadas en beneficio del bienestar estudiantil.