

ANÁLISIS DE SALUD EMOCIONAL UTILIZANDO INTELIGENCIA ARTIFICIAL

Emotional Health Analysis using Artificial Intelligence

Erika Morales Portillo	erika.moralesp@alumno.buap.mx
María del Carmen Santiago Díaz	marycarmen.santiago@correo.buap.mx
Ana Claudia Zenteno Vázquez	ana.zenteno@correo.buap.mx
Judith Pérez Marcial	judith.perez@correo.buap.mx
Abraham Sánchez López	abraham.sanchez@correo.buap.mx
Gustavo Trinidad Rubín Linares	gustavo.rubin@correo.buap.mx

Facultad de Ciencias de la Computación, Benemérita Universidad Autónoma de Puebla, México.

RESUMEN

La ansiedad es una respuesta natural al estrés, pero los trastornos de ansiedad son condiciones donde esta respuesta se vuelve excesiva y persistente, afectando la vida diaria. Tipos comunes incluyen trastorno de ansiedad generalizada, trastorno de pánico, trastorno obsesivo-compulsivo y fobias. Los factores genéticos, ambientales y biológicos pueden contribuir a su desarrollo. Los síntomas abarcan inquietud, fatiga, problemas de concentración y sueño. El presente trabajo presenta los resultados obtenidos y tratados mediante herramientas y técnicas de inteligencia artificial para la identificación automática del estado de ansiedad y su correlación a los niveles de estrés y salud emocional humana.

Palabras Clave: Trastorno de Ansiedad, Trastorno Obsesivo-Compulsivo, Trastorno de Estrés Postraumático, Inteligencia Artificial.

development. Symptoms include restlessness, fatigue, concentration, and sleep problems. This paper presents the results obtained and treated using artificial intelligence tools and techniques for the automatic identification of the state of anxiety and its correlation to stress levels and human emotional health.

Keywords: Anxiety Disorder, Agoraphobia, Social Phobia, Generalized Anxiety Disorder, Specific Phobia, Obsessive-Compulsive Disorder, Post-Traumatic Stress Disorder, Artificial Intelligence.

► I. Introducción

La ansiedad es una parte integral de la existencia humana; todos experimentan un grado moderado de ella, actuando como una respuesta adaptativa. Desde su tardía inclusión en la Psicología en 1920, la ansiedad ha sido un tema de gran interés, abordado desde diversas perspectivas. Según el Diccionario de la Real Academia Española, el término ansiedad proviene del latín "anxietas", refiriéndose a un estado de agitación, inquietud o zozobra del ánimo, siendo una de las emociones más frecuentes y complicadas del ser humano. Esta emoción displacentera se manifiesta a través de una tensión emocional acompañada de

ABSTRACT

Anxiety is a natural stress response, but anxiety disorders are conditions where this response becomes excessive and persistent, affecting daily life. Common types include generalized anxiety disorder, panic disorder, obsessive-compulsive disorder, and phobias. Genetic, environmental, and biological factors may contribute to its

síntomas físicos como palpitaciones, sudoración y temblores. [1]

Aunque la ansiedad está relacionada con el miedo, se diferencia en que el miedo es una respuesta a un peligro inmediato y real, mientras que la ansiedad se relaciona con la preocupación por posibles peligros futuros e inciertos. [2] La intensidad de la ansiedad varía entre las personas debido a factores biológicos y experiencias de vida. Algunas personas ven muchas situaciones como amenazantes y reaccionan con ansiedad, mientras que otras no se preocupan tanto.

En los últimos años, la ansiedad se ha convertido en un problema de salud mental significativo entre los estudiantes, afectando su rendimiento académico, bienestar emocional y calidad de vida. La identificación temprana y el manejo adecuado de la ansiedad en el contexto educativo son cruciales para mejorar tanto los resultados académicos como la salud mental de los alumnos. [3] En este sentido, diversas herramientas como el General Anxiety Disorder 7 (GAD-7), el Social Phobia Inventory (SPIN), y la Satisfaction With Life Scale (SWL) se han utilizado ampliamente para medir distintos aspectos de la ansiedad y la satisfacción con la vida en los estudiantes. El GAD-7 es una herramienta que mide la gravedad del trastorno de ansiedad generalizada; el SPIN evalúa la fobia social, y el SWL mide la satisfacción con la vida, proporcionando una visión integral de la salud mental de los estudiantes. [4]

Este documento se enfoca en la aplicación de técnicas de minería de datos para analizar un dataset que contiene información relacionada con los niveles de ansiedad y satisfacción de vida de los estudiantes, utilizando las métricas GAD-7, SPIN y SWL. La minería de datos es una herramienta poderosa para descubrir patrones y relaciones significativas en grandes volúmenes de datos, permitiendo una comprensión más profunda de los factores que contribuyen a la ansiedad estudiantil. [5] A través del análisis de estos datos, es posible identificar tendencias, perfiles de riesgo, y potenciales intervenciones que podrían ser implementadas para apoyar a

los estudiantes que experimentan altos niveles de ansiedad o insatisfacción con la vida. [6]

El objetivo de este trabajo es presentar la metodología empleada para realizar un análisis exhaustivo del dataset mediante técnicas de minería de datos. Esto incluirá la limpieza de datos, selección de variables, aplicación de algoritmos de clasificación y agrupamiento, y la interpretación de los resultados obtenidos. A partir de estos análisis, se espera proporcionar recomendaciones basadas en datos que puedan ser utilizadas por instituciones educativas para mejorar el apoyo psicológico y el bienestar general de los estudiantes.

II. Metodología

La Selección es la primera etapa crucial del proceso de preparación de datos. En esta fase, el objetivo es identificar y extraer las variables más relevantes del dataset que aportan valor para el análisis y los objetivos planteados. Esta selección se basa en el conocimiento previo del dominio del problema y en el entendimiento de los datos disponibles. Se eliminan los atributos que son irrelevantes, redundantes o que no tienen relación con el problema a resolver. Al reducir la dimensionalidad del dataset, se mejora la eficiencia y efectividad de los modelos de minería de datos, asegurando que solo se utilicen los datos necesarios para generar resultados significativos y precisos.

Primero, se importa las librerías necesarias y se carga el dataset en un DataFrame de pandas. Como se muestra en la figura 1.

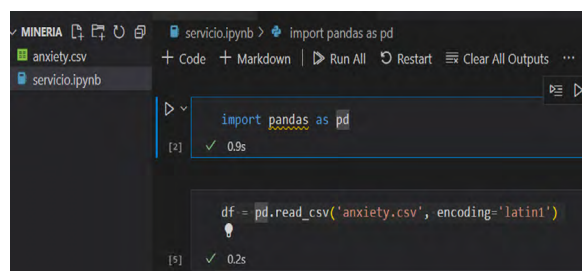


Fig. 1. *Importación del dataset*

Después se explora el dataset para entender mejor los datos con los que se está trabajando. En este

caso el dataset cuenta con 13464 respuestas de estudiantes, y contando con 53 columnas. En donde 22 columnas contienen datos con float, 16 columnas con el tipo de datos int y 15 columnas con el tipo de dato object.

Para la selección de variables, es importante considerar las columnas que serán útiles para el análisis de la ansiedad de los estudiantes. En este caso, dado que queremos analizar la ansiedad utilizando las métricas GAD, SPIN, y SWL, seleccionaremos las siguientes columnas:

Variables principales: GAD_T, SWL_T, SPIN_T (totales de cada instrumento).

Variables de análisis adicionales: gender, age, hours, whyplay.

Eliminaremos las columnas que no son relevantes para el análisis actual, como, 'timestamp', 'playstyle', 'earnings', 'degree', 'birthplace', 'residence', 'work' y 'league'.

Antes de proceder con el análisis, es crucial verificar si existen datos faltantes en las columnas, esto se logra hacer con la línea de código `print(df_selected.isnull().sum())`. Esto muestra que las columnas Hours y SPIN_T tienen datos faltantes, como se muestra en la figura 2.

```
S. No.      0
Gender      0
Age         0
Game        0
Hours       30
whyplay     0
GAD_T       0
SWL_T       0
SPIN_T      650
dtype: int64
```

Fig. 2. Datos faltantes en la columna del dataset.

Para solucionar esto, se eliminarán las filas que tienen estos datos faltantes con la línea de código `df_selected = df_selected.dropna()`. El resultado de esto se muestra en la figura 3.

```
S. No.      0
Gender      0
Age         0
Game        0
Hours       0
whyplay     0
GAD_T       0
SWL_T       0
SPIN_T      0
dtype: int64
```

Fig. 3. Eliminación de filas con datos faltantes.

A. Preprocesado

El preprocesado incluye la limpieza de datos, la transformación de variables, y la preparación de los datos para ser utilizados en modelos de minería de datos.

Para empezar, dado que la variable gender es categórica, será necesario convertirla en un formato que pueda ser interpretado por los algoritmos de minería de datos, ya que estos generalmente requieren valores numéricos. Esto se logra con la línea de código `df_encoded = pd.get_dummies(df_selected, columns=['Gender'])`.

```
Gender_Female  Gender_Male  Gender_Other
False         True         False
False         True         False
True          False        False
False         True         False
False         True         False
```

Fig. 4. Variable categórica a valor booleano.

Pero como resultado tendremos que las columnas se imprimieron con 'true' y 'false' como se observa en la figura 4. Y aunque es correcto y no afectará ya que internamente true y false son equivalentes a 1 y 0, respectivamente en Python. Se recomienda verlos explícitamente como 0 y 1 para mayor claridad o compatibilidad con ciertas herramientas. Esto se consigue con la línea de código:

```
bool_columns = df_encoded.select_dtypes(include=[bool]).columns
```

```
df_encoded[bool_columns] = df_encoded[bool_columns].astype(int)
```

Con esto podemos visualizar correctamente la variable 'gender' en binario en la figura 5.

Gender_Female	Gender_Male	Gender_Other
0	1	0
0	1	0
1	0	0
0	1	0
0	1	0

Fig. 5. Variable 'gender' actualizada a binaria.

Por último, haremos una Matriz de Correlación, para ver si alguna variable tiene una alta correlación (mayor a 0.9) con otra, si esto es verdadero, puede ser redundante y se tendría que eliminar.

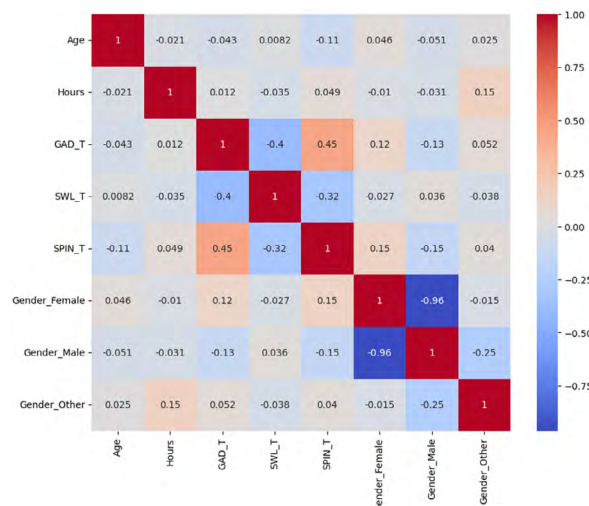


Fig. 6. Matriz de Correlación

Las columnas Gender_Female, Gender_Male, y Gender_Other contienen la misma información que la columna original Gender, pero en un formato que los algoritmos de minería de datos pueden entender. Si bien existe redundancia (ya que saber el valor de dos columnas te da la tercera), la mayoría de los algoritmos de minería de datos manejan esta situación sin problemas.

B. Selección de Características

La Selección de Características es la siguiente etapa en el proceso de minería de datos. En esta

etapa, el objetivo es seleccionar las características más relevantes del dataset para mejorar la eficiencia y el rendimiento de los modelos de minería de datos. Este proceso ayuda a eliminar el "ruido" o la información irrelevante que puede confundir a los modelos.

Aquí se utilizará el modelo de clasificación RandomForestClassifier para entrenar un modelo y obtener la importancia de las características.

El random forest es un algoritmo de machine learning de uso común registrado por Leo Breiman y Adele Cutler, que combina la salida de múltiples árboles de decisión para alcanzar un solo resultado. Su facilidad de uso y flexibilidad han impulsado su adopción, ya que maneja problemas de clasificación y regresión. [7]

Para lograr esto en python primero se importan las librerías necesarias que son RandomForestClassifier para el modelo, matplotlib.pyplot para la visualización y pandas para la manipulación de datos. Después se define 'x' que contiene las variables predictoras (todas las columnas excepto SPIN_T). Y la variable objetivo 'y' (SPIN_T).

En seguida se crea el modelo de Random Forest con 100 arboles y se entrena con los datos ('x' y 'y').

Para finalizar se extrae la importancia de cada característica del modelo entrenado. Se crea un DataFrame con las características y su importancia de mayor a menor y por último se visualiza la importancia de las características mediante un gráfico de barras horizontales de la figura 7.

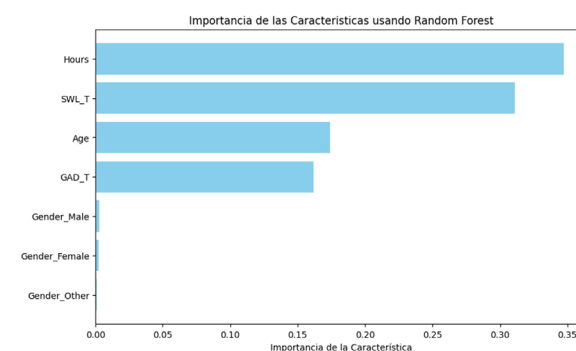


Fig. 7. Importancia de las características usando Random Forest.

► III. Resultados

A. Extracción de conocimiento

En la etapa de Extracción de Conocimiento, el objetivo es interpretar los resultados del análisis y obtener información útil y comprensible a partir de los datos procesados. Aquí se busca traducir los hallazgos del modelo en conocimientos aplicables que puedan ayudar a tomar decisiones informadas.

Primero se examina cuáles características tienen mayor impacto en la predicción del modelo. En este caso, se observa qué características (Hours, SWL_T, Age, GAD_T) son las más relevantes en la figura 8.

Características más importantes:		
	Feature	Importance
1	Hours	0.347347
3	SWL_T	0.310823
0	Age	0.173920
2	GAD_T	0.161866
5	Gender_Male	0.002545
4	Gender_Female	0.002503
6	Gender_Other	0.000996

Fig. 8. Características más importantes.

En esta etapa es importante crear gráficos y tablas que ayuden a comunicar los hallazgos de manera clara. Por ejemplo, los gráficos de barras que muestran la importancia de las características como en la figura 7.

En este gráfico podemos decir que la variable 'Horas de Juego' tiene la mayor importancia, lo que sugiere que la cantidad de tiempo dedicado a jugar videojuegos es un predictor crucial para la ansiedad. Esto podría indicar que un mayor tiempo de juego está asociado con niveles más altos de ansiedad. También La 'Satisfacción con la Vida' muestra una importancia considerable. Esto sugiere que una menor satisfacción con la vida puede estar relacionada con mayores niveles de ansiedad. Por último, la edad y las puntuaciones en el GAD también contribuyen a la predicción de la ansiedad, aunque con menor peso comparativo en relación con las horas de juego y la satisfacción con la vida.

B. Interpretación y Evaluación

En esta etapa final del proceso de minería de datos, se interpretan los resultados obtenidos del modelo Random Forest y se evalúa su desempeño. La evaluación se realiza utilizando métricas como la matriz de confusión y el análisis de la importancia de las características.

La Matriz de Confusión nos mostrara como se distribuyen las predicciones del modelo en comparación con las etiquetas reales de la figura 9. Una matriz de confusión es una herramienta utilizada en clasificación para evaluar el rendimiento de un modelo. Muestra el conteo de las predicciones correctas e incorrectas del modelo en comparación con las etiquetas verdaderas. Es particularmente útil para entender cómo se comporta el modelo en relación con cada clase y para identificar patrones de errores.

Características más importantes:		
	Feature	Importance
1	Hours	0.347347
3	SWL_T	0.310823
0	Age	0.173920
2	GAD_T	0.161866
5	Gender_Male	0.002545
4	Gender_Female	0.002503
6	Gender_Other	0.000996

Fig. 9. Matriz de confusión

C. No Ansioso

Verdadero Negativo (TN): El valor real es negativo y la prueba predijo también que el resultado era negativo. 2684 casos fueron correctamente predichos como "No Ansioso".

Falso Positivo (FP): El valor real es negativo, y la prueba predijo que el resultado es positivo. No hay predicciones incorrectas en esta categoría.

D. Ansioso

Falso Negativo (FN): El valor real es positivo, y la prueba predijo que el resultado es negativo. Hay un caso que fue clasificado incorrectamente como

“No Ansioso” cuando realmente es “Ansioso”.

Verdadero Positivo (TP): El valor real es positivo y la prueba predijo también que era positivo. Ocho casos fueron correctamente predichos como “Ansioso”.

E. Importancia de las Características

El análisis de la importancia de las características nos proporciona información crucial sobre cuáles de las variables del conjunto de datos tienen un mayor impacto en la predicción de la ansiedad de los estudiantes.

Las horas de juego resultaron ser la característica más importante según el modelo Random Forest. Este resultado sugiere una fuerte relación entre el tiempo dedicado a los videojuegos y los niveles de ansiedad de los estudiantes.

Un número elevado de horas dedicadas a los videojuegos podría estar asociado con un mayor aislamiento social y menos tiempo dedicado a actividades físicas o sociales que promuevan una mejor salud mental. El modelo sugiere que, a medida que aumentan las horas de juego, también lo hace la probabilidad de tener altos niveles de ansiedad social, medidos a través del SPIN_T.

Este hallazgo sugiere que las intervenciones dirigidas a reducir las horas de juego excesivo podrían ser efectivas para disminuir los niveles de ansiedad en los estudiantes. Por ejemplo, programas de concienciación sobre el tiempo de pantalla y la promoción de hábitos de vida saludables que equilibren el tiempo de juego con otras actividades podrían ser beneficiosos.

La satisfacción con la vida (medida a través de SWL_T) es otra característica clave que el modelo considera significativa para predecir la ansiedad.

» IV. Conclusiones

Los estudiantes con bajos niveles de satisfacción con su vida suelen mostrar niveles más altos de ansiedad social. Esto sugiere que la percepción de bienestar personal, felicidad y cumplimiento

juega un papel importante en la salud mental de los estudiantes. Un bajo nivel de satisfacción con la vida puede estar relacionado con factores como el estrés académico, problemas familiares, falta de objetivos claros, o malas relaciones sociales, todos ellos son factores de riesgo para desarrollar ansiedad.

Mejorar la satisfacción con la vida de los estudiantes a través de programas de bienestar, apoyo emocional, asesoramiento, y actividades extracurriculares que promuevan un sentido de logro y pertenencia podría reducir significativamente los niveles de ansiedad.

La puntuación total de General Anxiety Disorder (GAD_T), que mide la ansiedad general, es también una característica importante en el modelo.

La puntuación de GAD_T indica que los estudiantes que ya experimentan ansiedad generalizada son más propensos a presentar también ansiedad social. La ansiedad general puede manifestarse en múltiples contextos y, al superponerse con la ansiedad social, puede amplificar sus efectos. Por lo tanto, aquellos con puntuaciones altas en GAD_T podrían tener niveles más altos de SPIN_T, que mide la ansiedad social.

» V Referencias

- [1] Kaplan H. I, Sadock B. J. Sinopsis de psiquiatría. 8a ed. Ed. Lippincott Williams & Wilkins-Panamericana. 1998. Pág.324 y 683.
- [2] Kolb L. C., Psiquiatría clínica moderna, 6a ed. PRENSA MÉDICA MEXICANA S.A. DE C.V. MEXICO 1992. 562 páginas.
- [3] Jenike M. A., et al. Trastornos obsesivo-compulsivos. Manejo práctico. 3a ed. 2001. Ed.Harcourt. 127 páginas.
- [4] Patiño Rojas J. L. Psiquiatría Clínica. 2a edición. Salvat. México, 1990. 413 páginas.
- [5] Kandel E. R., Schwartz J. H. & Jessell T. M. Principios de Neurociencias, 4a Edición. McGraw-Hill Interamericana, 2001, 1395 páginas.
- [6] Coolhuntermx, R. (2021, 24 agosto). 4 apps para la salud mental y emocional. Coolhuntermx. <https://coolhuntermx.com/apps-de-inteligencia->

- artificial-para-la-salud-mental-y-emocional/
- [7] Doonamis. (2024, 5 febrero). El poder de las Apps para mejorar la salud mental - Doonamis. Doonamis. <https://www.doonamis.com/poder-apps-mejorar-salud-mental/>
- [8] J. Díaz-Verdejo, "Ejemplo de bibliografía", En Actas de las XI Jornadas de Ingeniería Telemática, vol. 1, n. 1, pp. 1-5, 2013.