

Análisis-Trastorno del sueño, Tarea1

Alumna Stephani Montserrat Vázquez Juárez
No.Cuenta 31524683-4
Correo stephani_farreraciencias.unam.mx

Trastornos del sueño

El sueño es un proceso biológico fundamental, un estado transitorio del sistema nervioso que permite la regulación de diversos mecanismos relacionados con el bienestar físico, mental y emocional de los seres humanos. Los trastornos del sueño afectan a millones de personas a nivel mundial, deteriorando su calidad de vida y aumentando el riesgo de enfermedades crónicas[1, 5, 6].

De acuerdo con la Clasificación Internacional de los Trastornos del Sueño (ICSD, por sus siglas en inglés) de la Academia Estadounidense de Medicina del Sueño, en su tercera edición, los trastornos del sueño se clasifican en[5, 6]:

1. Insomnio
2. Trastornos respiratorios asociados con el sueño
3. Trastornos centrales de la hipersomnolencia
4. Trastornos del ritmo circadiano del sueño
5. Parasomnias
6. Trastornos del movimiento relacionados con el sueño

Los trastornos del sueño afectan tanto a infantes como a adultos; sin embargo, presentan una mayor incidencia y prevalencia en la población adulta. Los trastornos del sueño pueden provocar fatiga, problemas de memoria, cambios de humor e incluso asociarse a enfermedades cardiovasculares, diabetes y obesidad [1, 5, 6].

El insomnio y la apnea del sueño son los trastornos del sueño más frecuentes en la población adulta. El insomnio se caracteriza por dificultades para iniciar el sueño (latencia del sueño prolongada) y mantener el sueño. Por otro lado, la apnea se manifiesta a través de interrupciones breves y repetidas de la respiración durante el sueño[5].

Diagnóstico y tratamiento

El diagnóstico de los trastornos del sueño requiere un análisis del historial clínico del paciente, estudio de polisomnografía, cuestionarios clínicos(Escala de Somnolencia de Epworth, índice de calidad del sueño de Pittsburgh) y estudios de laboratorios tales como: gasometría arterial, test de función tiroidea y exámenes de toxicidad(drogas y alcohol).

Su tratamiento depende del tipo y cronicidad del trastorno; por lo que puede ser farmacológico, conductual o mixto[6].

Casos de uso con Machine Learning

- Clasificación del tipo de trastorno que se padece considerando variables como el estrés, salud física, consumo de alcohol, calidad y duración del sueño
- Clustering de pacientes según patrones/similitudes
- Predicción de horas de sueño considerando variables fisiológicas y variables relacionadas con el estilo de vida del paciente

0.1 Hipótesis

- La edad y el estilo de vida influyen en la modificación el ciclo normal del sueño
- Factores como la edad, la ocupación y el nivel de estrés impactan significativamente en la calidad del sueño.
- La población adulta masculina tiene mayor predisposición a desarrollar apnea del sueño.

Mejoras al análisis

- Visualizar el balance de los datos respecto al sexo
- Visualizar si existen outliers
- Realizar gráficos de distribución de la calidad y horas de sueño en función de la edad, el sexo, ocupación y movilidad
- Transformar las variables categóricas a variables numéricas con el objetivo de realizar un mapa de correlación entre todas las variables
- Modificación del conjunto de variables que tienen correlación con la variable a predecir
- Realizar una validación cruzada
- Justificar la elección de las variables predictivas

Comparación de modelos usados en el código base

En el dataset presentado se implementaron los siguientes modelos de aprendizaje supervisado para predecir los trastornos del sueño:

- Regresión Logística (ordinal): Modelo de clasificación y análisis predictivo que permite hallar relación entre una variable dependiente y una o más variables independientes. Por lo general, se emplea para realizar predicciones sobre una variable categórica frente a variables continuas. Su funcionamiento se basa en el cálculo de probabilidades asociadas a la ocurrencia de un evento [4].

Desventajas:

1. Es propensa al sobre-ajuste cuando existe una gran cantidad de variables predictivas
 2. La predicción se ve afectada cuando existe multicolinealidad entre las variables predictivas
- Gradient Boosting: Método de aprendizaje supervisado que emplea árboles de decisión de manera secuencial para minimizar errores y mejorar la precisión de las predicciones . Puede emplearse en problemas de regresión y clasificación [2].

Desventajas:

1. Es propensa al sobre-ajuste
 2. La velocidad de entrenamiento es lenta, ya que se da de manera secuencial
 3. Interpretabilidad
- XGBoost: Optimización del modelo Gradient Boosting [2].

Ventajas:

1. Alto rendimiento
 2. Menos propenso al sobre-ajuste
 3. Manejo automático de valores faltantes
- K-Nearest Neighbors: Algoritmo de aprendizaje supervisado inferencial empleado en problemas de regresión y clasificación. El objetivo de este algoritmo es identificar los puntos cercanos a un punto específico, para poder asignarle una etiqueta de clase. Emplea la distancia como métrica de asignación [3].

Desventajas:

1. Las predicciones dependen del número de vecinos cercanos al punto de interés (K) y la distancia usada
2. Velocidad de predicción baja y sensibilidad a outliers

Los datos del DataSet se dividieron en un conjunto de entrenamiento y un conjunto de prueba empleando Train/Test Split con una relación 75%-25% de separación de datos. Luego, se realizó un sobremuestreo para balancear las clases en cada conjunto. Para evaluar el rendimiento de cada modelo, se incluyeron las siguientes métricas: Accuracy, Precision, Recall, y F1-score. En la Tabla 1 se muestran los valores de las métricas obtenidos en los modelos de predicción del código base.

Modelo	Accuracy	F1-score	Precision	Recall
Regresión Logística	0.909	0.908	0.910	0.909
KNN	0.921	0.921	0.923	0.921
XGBoost	0.927	0.927	0.928	0.927
Gradient Boosting	0.915	0.914	0.916	0.915

Table 1: Métricas de desempeño de los modelos de predicción empleados en el código original. De acuerdo a las métricas Accuracy y F1-score, El algoritmo XGBoost presenta mejor desempeño al clasificar los trastornos del sueño

Luego de optimizar el análisis exploratorio, la matriz de correlación entre todas las variables del conjunto de datos permitió identificar que se pueden descartar algunas variables debido a su baja correlación con la variable de interés(trastorno del sueño). Por lo tanto, conservaremos únicamente las siguientes:

- Edad
- Ocupación
- Presión diastólica
- Presión sistólica
- Género
- Pasos diarios
- Nivel de actividad física
- Frecuencia Cardíaca

En la Tabla 2, se muestran los valores de las métricas de evaluación para cada modelo tras la optimización del código base.

Modelo	Accuracy	F1-score	Precision	Recall
Regresión Logística	0.915	0.915	0.915	0.915
KNN	0.945	0.945	0.947	0.945
XGBoost	0.9636	0.9637	0.9647	0.9636
Gradient Boosting	0.9636	0.9637	0.9646	0.9636

Table 2: Métricas de desempeño de los modelos de predicción empleados en el código optimizado. Se observa que todos los modelos presentan mejoras en todas las métricas, sin embargo, de acuerdo a las métricas Accuracy y F1score, XGBoost y Gradient Bossting presenta el mejor desempeño al clasificar los trastornos del sueño

EL incremento en el desempeño de cada modelo también se visualiza en un análisis comparativo entre las matrices de confusión asociadas a cada modelo de predicción.

En todos los modelos se observa un incremento en la clasificación de personas que no padecen trastornos del sueño. en el número de personas clasificación de personas que no padecen trastornos del sueño.

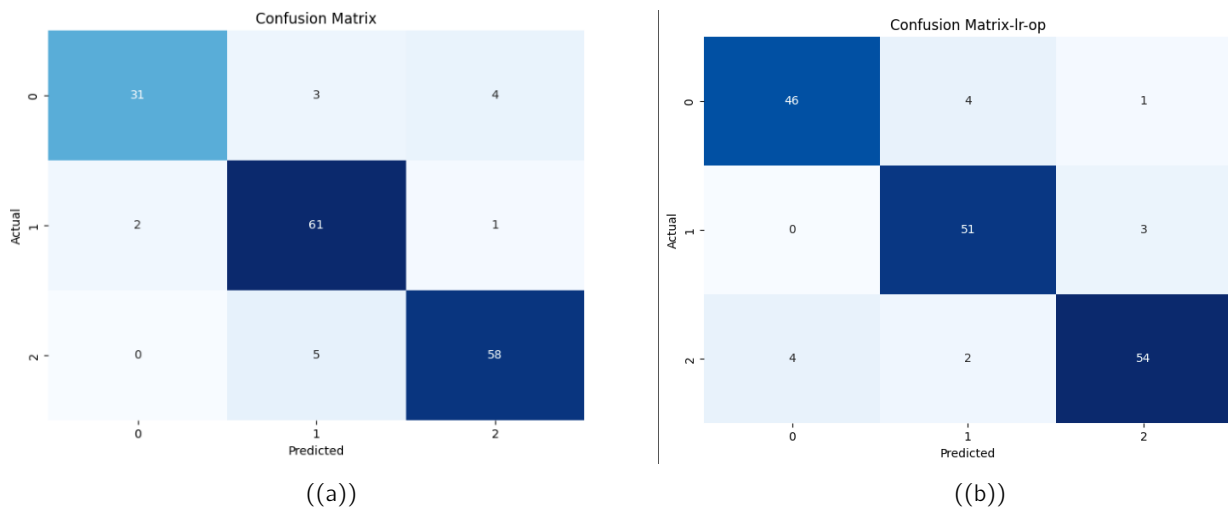


Figure 1: Matrices de confusión obtenidas en el modelo de regresión logística. a) Matriz de confusión obtenida en el código original. b) Matriz de confusión obtenida tras la optimización del análisis exploratorio y del pre-procesamiento de datos

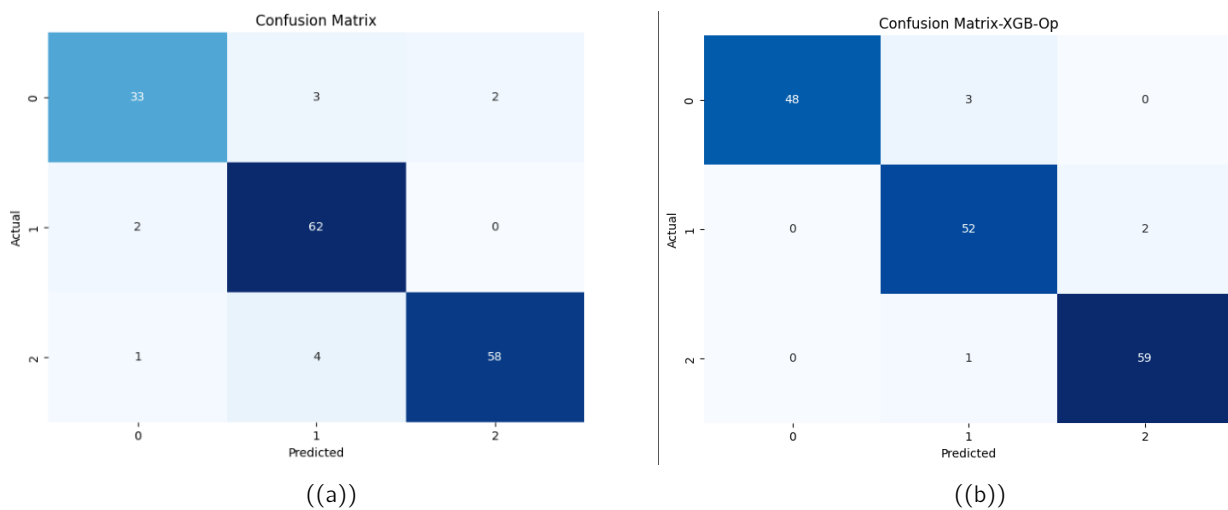


Figure 2: Matrices de confusión obtenidas en el modelo de XGBoost a) Matriz de confusión obtenida en el código original. b) Matriz de confusión obtenida tras la optimización del análisis exploratorio y del pre-procesamiento de datos

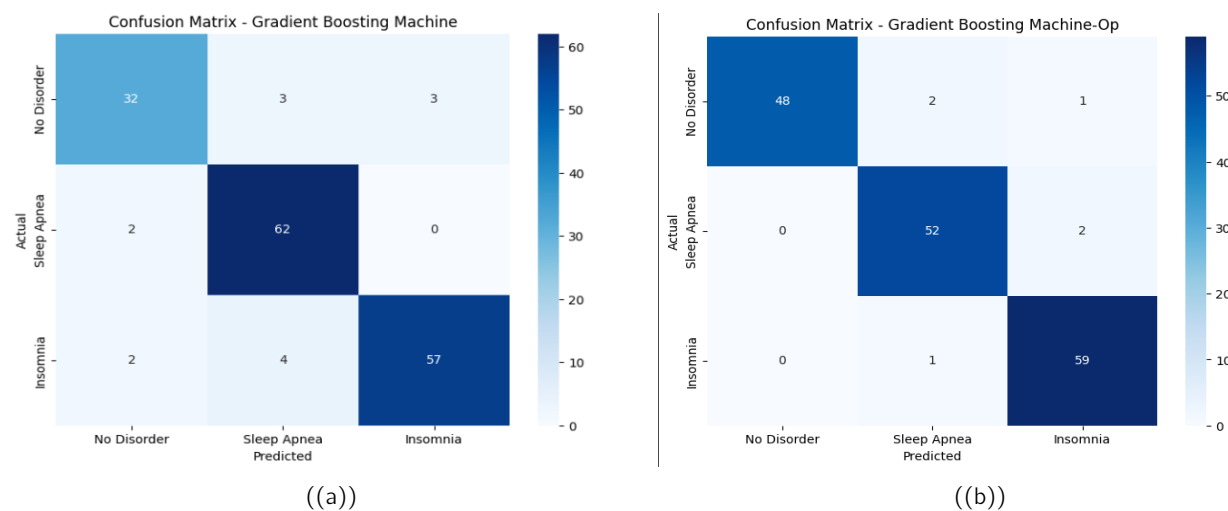


Figure 3: Matrices de confusión obtenidas en el modelo de Gradient Boosting a) Matriz de confusión obtenida en el código original. b) Matriz de confusión obtenida tras la optimización del análisis exploratorio y del pre-procesamiento de datos

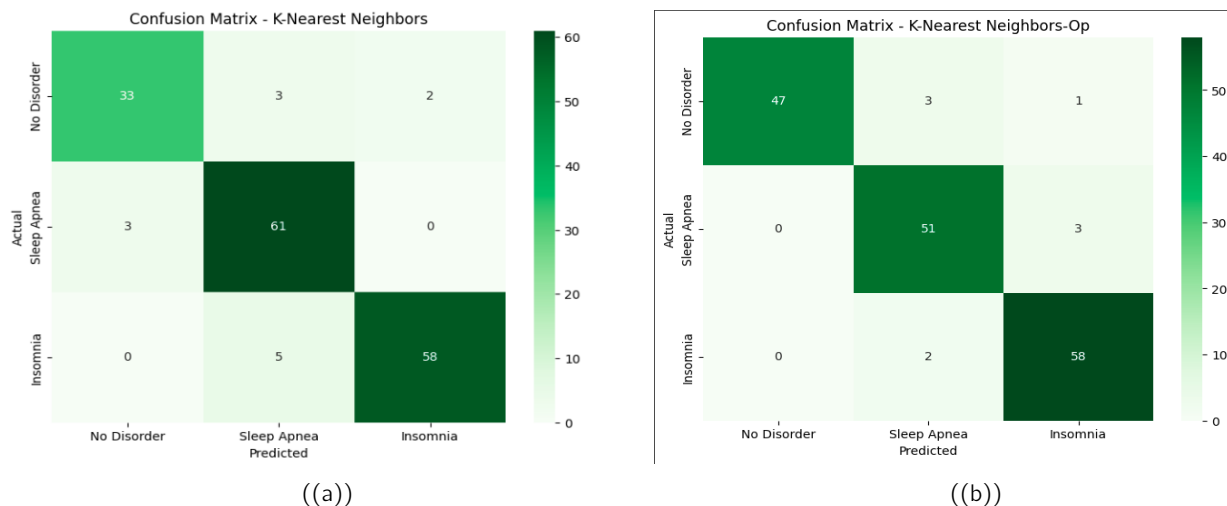


Figure 4: Matrices de confusión obtenidas en el modelo de KNN a) Matriz de confusión obtenida en el código original. b) Matriz de confusión obtenida tras la optimización del análisis exploratorio y del pre-procesamiento de datos

Conclusiones

Del análisis del DataSet se concluye que:

- Es posible predecir cuando una persona padece trastornos del sueño utilizando variables asociadas a su estilo de vida.
- Los trastornos del sueño afectan la calidad del descanso; en particular, las personas con insomnio suelen experimentar una calidad de sueño significativamente reducida
- Quienes trabajan en enfermería, docencia, ventas o representación comercial son más propensos a experimentar trastornos del sueño
- La apnea del sueño es más frecuente en personas de género femenino, mientras que el insomnio se presenta con mayor frecuencia en personas de género masculino
- Personas con buen estado físico reportan mejor calidad de sueño.
- Los niveles de estrés influyen en la calidad del sueño. Niveles altos de estrés disminuyen la calidad del sueño

1 Repositorio GitHub

Todos los recursos del proyecto están disponibles en el siguiente repositorio <https://github.com/StephaniVJ/SleepDisorders-SMJ-T1>:

Incluye:

Código replicado y comentado

Documento PDF de este informe

Archivo README con instrucciones

References

- [1] Jaqua, E. E., Hanna, M., Labib, W., Moore, C., & Matossian, V. (2023). Common Sleep Disorders Affecting Older Adults. The Permanente journal, 27(1), 122–132. <https://doi.org/10.7812/TPP/22.114>
- [2] Clark, Bryan, and Fangfang Lee. "What Is Gradient Boosting?" Ibm.com, 7 Apr. 2025, www.ibm.com/topics/gradient-boosting. Accessed 20 Apr. 2025.
- [3] IBM. "KNN." Ibm.com, 4 Oct. 2021, [www.ibm.com/think/topics/knn?mhsrc=ibmsearch_a\\$&\\$mhq=knn\\$%\\$20](http://www.ibm.com/think/topics/knn?mhsrc=ibmsearch_a$&$mhq=knn$%$20). Accessed 22 Apr. 2025.

- [4] —. “Logistic Regression.” Ibm.com, 16 Aug. 2021, www.ibm.com/think/topics/logistic-regression?mhsrc=ibmsearch_a&mhq=logistic%20regression%20. Accessed 22 Apr. 2025.
- [5] Karna, Bibek, and Vikas Gupta. “Sleep Disorder.” PubMed, StatPearls Publishing, 2020, www.ncbi.nlm.nih.gov/books/NBK560720/. Accessed 20 Apr. 2025.
- [6] Szentkirályi, András, et al. “Sleep Disorders: Impact on Daytime Functioning and Quality of Life.” Expert Review of Pharmacoeconomics & Outcomes Research, vol. 9, no. 1, Feb. 2009, pp. 49–64, <https://doi.org/10.1586/14737167.9.1.49>. Accessed 22 Apr. 2025.