

El presente informe resume las conclusiones obtenidas a partir del análisis realizado en el conjunto de datos Sleep Health and Lifestyle Dataset. Se aplicaron diversos métodos de análisis y clasificación con el objetivo de comprender los factores relacionados con el sueño y los hábitos diarios, así como predecir la presencia de trastornos del sueño.

En primer lugar, se llevó a cabo un análisis exploratorio de los datos, seguido de la generación de una muestra aleatoria estratificada por "Sleep_Disorder". Posteriormente, se aplicaron métodos como el Análisis de Componentes Principales (PCA), el Análisis de Correspondencias Simple y Múltiple, el Análisis Discriminante, SVM, K-means y KNN.

El EDA reveló una alta correlación entre algunas variables, lo que indica que existen patrones comunes en la duración del sueño, la calidad del sueño, el nivel de estrés, la actividad física y las mediciones de presión arterial. Este análisis también evidencio, a través de gráficos boxplot, la existencia de outliers únicamente para la variable "Heart_Rate" cuando se observan las features de manera global y la existencia de outliers en prácticamente todas las variables cuando se observan las features discriminadas por "Sleep_Disorder".

El Análisis de Componentes Principales (PCA) identificó las variables más importantes que contribuyen a la variabilidad del conjunto de datos, siendo estas "Quality_of_Sleep", "Stress_Level", "Sleep_Duration" y "Heart_Rate".

El Análisis de Correspondencias permitió analizar la relación entre la categoría BMI, el trastorno del sueño y el género. Se observaron asociaciones significativas entre estas variables y esto se debe principalmente debido a la relación entre las categorías "Normal" (BMI), "None" (Sleep_Disorder) y "Hombre" (Gender), y la relación entre las categorías "Overweight" (BMI) y "Sleep Apnea" (Sleep_Disorder). Quedando expuesta la no independencia entre dichas variables.

Los modelos de Análisis Discriminante (LDA), Support Vector machine (SVM) y K-Nearest Neighbors Algorithm (KNN) fueron utilizados para clasificar los pacientes según la presencia de trastornos del sueño. Aunque los supuestos requeridos para el Análisis Discriminante no se cumplieron (Normalidad multivariada, Independencia de las observaciones y Homocedasticidad), se realizó el análisis y se obtuvieron resultados satisfactorios en términos de precisión de clasificación. Los modelos SVM y KNN mostraron un rendimiento similar y superaron al modelo de Análisis Discriminante en términos de precisión.

LDA

- Accuracy en train: 0.8862
- Accuracy en test: 0.8444

SVM

- Accuracy en train: 0.9176
- Accuracy en test: 0.8667

KNN

- Accuracy en train: 0.8706
- Accuracy en test: 0.8667

Por otro lado, los métodos no supervisados, como K-means, permitieron agrupar los datos en diferentes categorías o clusters. Mediante el Método de Elbow, se determinó el número óptimo de clusters (7), lo que proporcionó información sobre la estructura interna de los datos y permitió identificar patrones específicos en relación con las características de sueño y estilo de vida.

En resumen, los resultados obtenidos en este trabajo revelaron la importancia de las variables relacionadas con el sueño, el estilo de vida y la presencia de trastornos del sueño. Se destacaron variables como la duración del sueño, la calidad del sueño, el nivel de estrés, la actividad física y las mediciones de presión arterial como factores clave en el análisis y clasificación de los datos.

Los métodos supervisados, como SVM y KNN, demostraron un rendimiento similar y fueron superiores al modelo de Análisis Discriminante en términos de precisión de clasificación. Estos modelos podrían ser útiles para la predicción y detección temprana de trastornos del sueño en pacientes.

Es importante tener en cuenta que los supuestos requeridos para aplicar algunos métodos no se cumplieron plenamente. Por lo tanto, se recomienda validar estos supuestos en futuros análisis para obtener resultados más robustos y precisos.

En conclusión, este trabajo proporcionó conocimientos valiosos sobre la salud del sueño y los hábitos diarios a través del uso del análisis de datos. Los resultados obtenidos pueden contribuir a una mejor comprensión de los factores que influyen en los trastornos del sueño y apoyar la toma de decisiones en intervenciones personalizadas.