

#SLEEP DISORDER PREDICTION Sarmiento Escobar Millaray

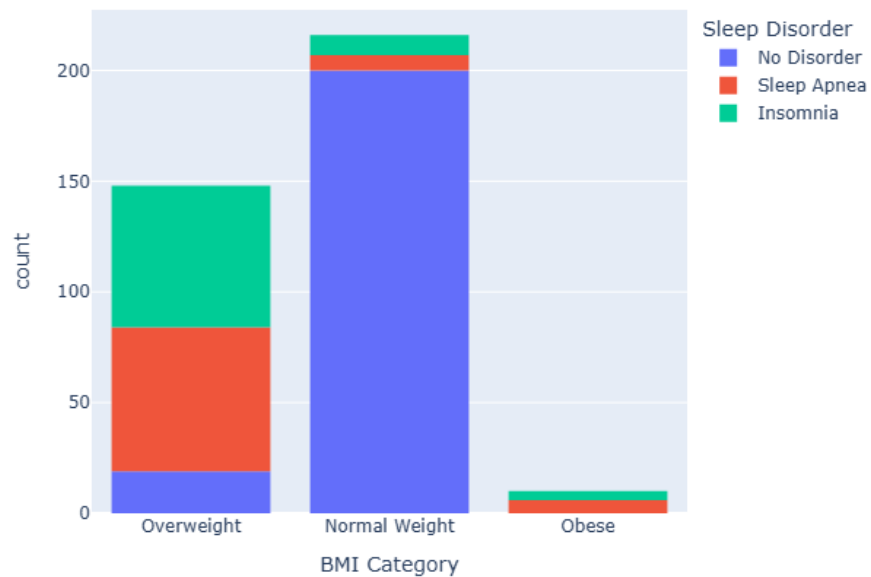
El trabajo se realizo en base al dataset de Sleep Disorder de kaggle, los datos pueden encontrarse en el siguiente link <https://www.kaggle.com/code/soumya g09/sleep-disorder>

#Objetivo

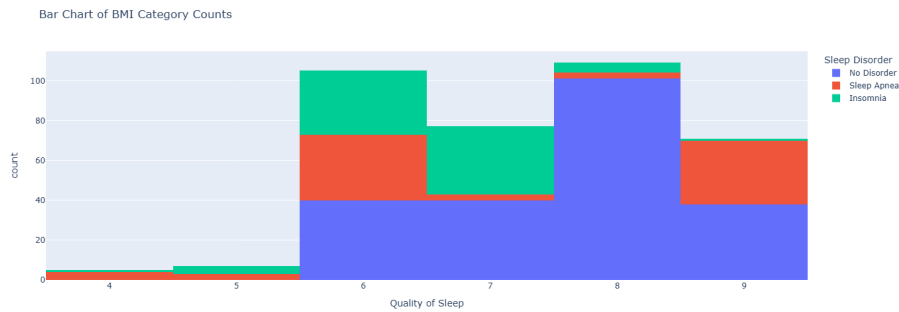
El trabajo se hizo con la finalidad de empezar a probar bases de dato de kaggle y para identificar factores de riesgo asociados al insomnio

#EDA

Bar Chart of BMI Category Counts

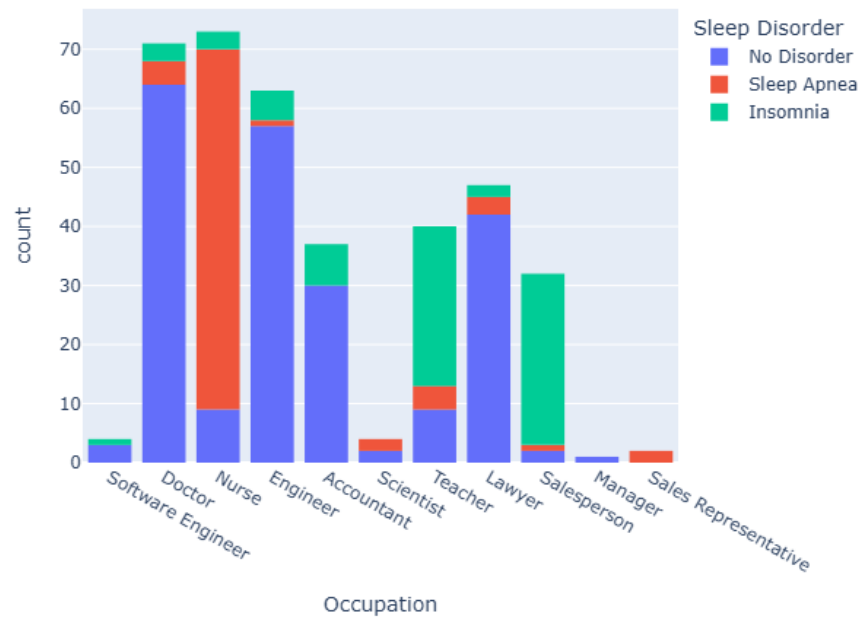


A partir de lo anterior y como menciona la literatura existe una relacion entre la obesidad y los trastornos del sueño



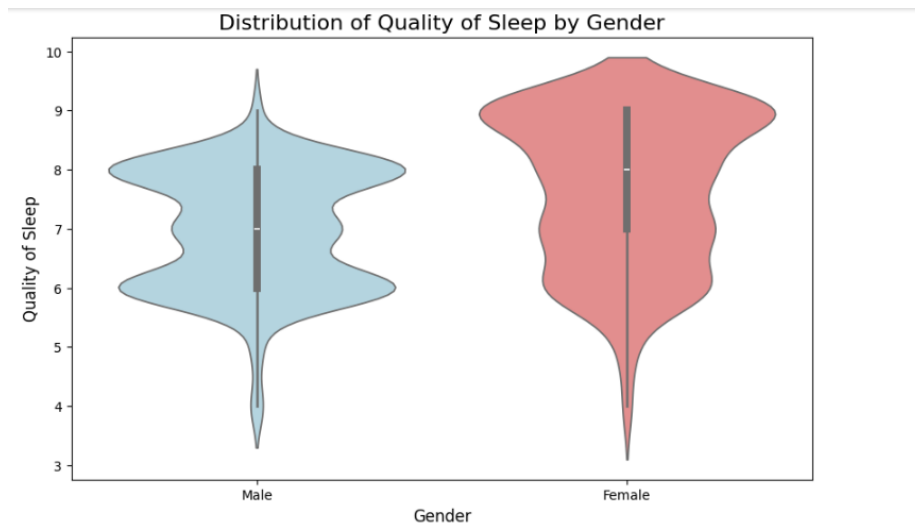
La calidad del sueño varia entre los trastornos, sin embargo se concentra mayormente entre 6 y 9

Bar Chart of BMI Category Counts



Los profesores, enfermeras y vendedores son los mas propensos a padecer trastornos del sueño

Esto es coherente, ya que en el sector salud estan expuestos a largas jornadas sin dormir, lo que puede alterar el ciclo circadiano "normal", con respecto a los maestros y vendedores un posible factor sea el estres en su vida diaria



En ambos generos encontramos valores inferiores a 5 lo que puede indicar una calidad baja del sueño, sin embargo en las mujeres encontramos una frecuencia mas altas, en los valores superiores de calidad de sueño



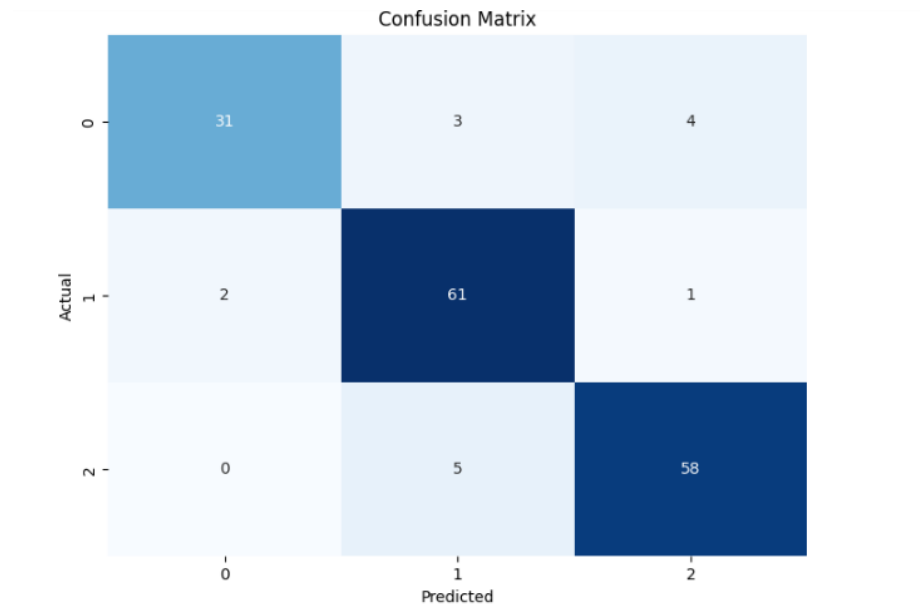
Encontramos una correlacion en donde a mayor nivel de estres, menor sera la calidad de sueño



- Se encuentran correlaciones positiva moderada con la presión sistólica (0.61) y diastólica (0.59), es decir que a mayor edad encontramos un aumento en la presión
- Encontramos una alta correlación entre la presión sistólica y diastólica (0.97)
- Hay correlación positiva entre nivel de actividad física con los pasos diarios (0.77)
- Encontramos una correlación negativa entre el ritmo cardíaco y la duración del sueño

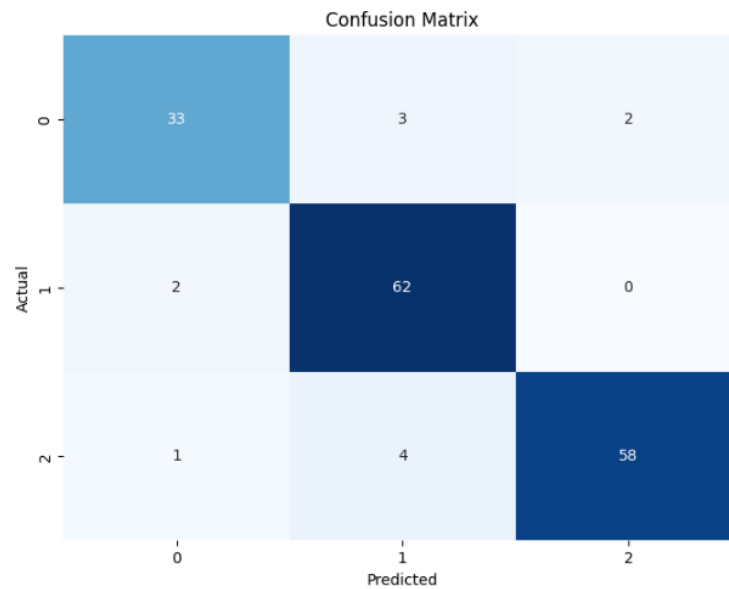
#MODELOS UTILIZADOS

Logistic Regression



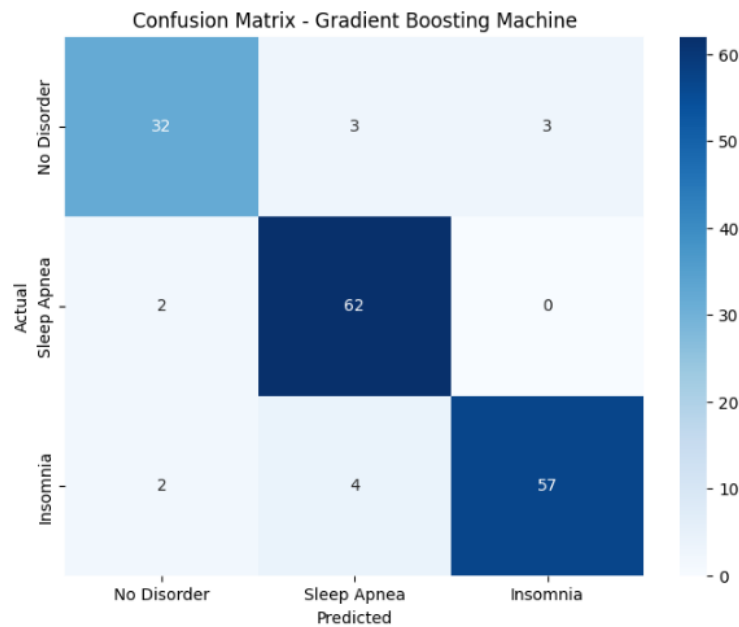
Logistic Regression es un modelo que se usa para calcular la probabilidad de que una instancia pertenezca a una clase, al ser un modelo bastante sencillo es fácil de interpretar sin embargo es sensible a outliers, tiene una precisión de .91 y una F1 score ponderado de .91, con un AUC de .94, es muy útil si ocupamos modelos simples y con datos lineales separables

XGBoost Classifier



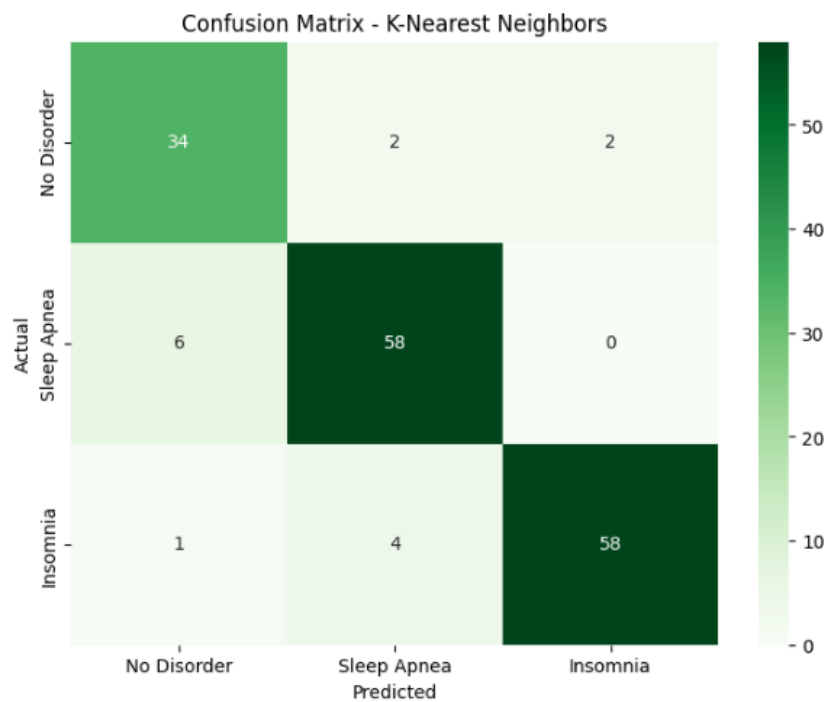
Con respecto a **XGBoost Classifier** este algoritmo construye es una version mas optimizada,y que trabaja con mas presicion que GBM ademas qde utilizar tecnicas para evitar el sobreajuste y manejar datos faltantes,cuenta con una presicion de .92 (mayor que LR),F1 score ponderado de .92 y un AUC de .95

Gradient Boosting Classifier



Gradient Boosting Machine(GBM) este modelo genera arboles en secuencia, con la finalidad de corregir el arbol anterior, por tanto es muy util,cuenta con una presicion de .92 y F1 score ponderado de .91 y un AUC de .95 este modelo es muy util con respecto a la interpretabilidad y el rendimiento

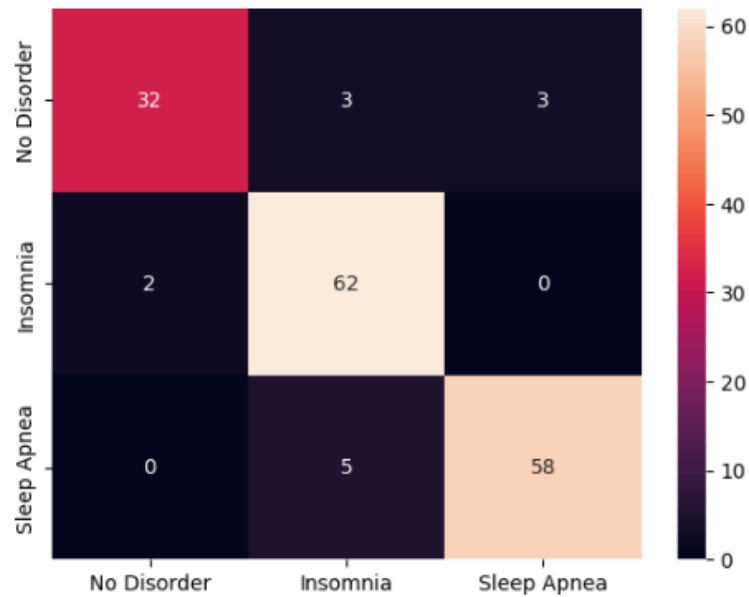
K Neighbors Classifier



K-vecinos, clasifica en base a los vecinos mas cercanos, es un diseño sencillo de entender e interpretar sin embargo su rendimiento es limitado en caso de presentar alta dimensionalidad, cuenta con una precision y F1-score de .91 y un AUC de .96

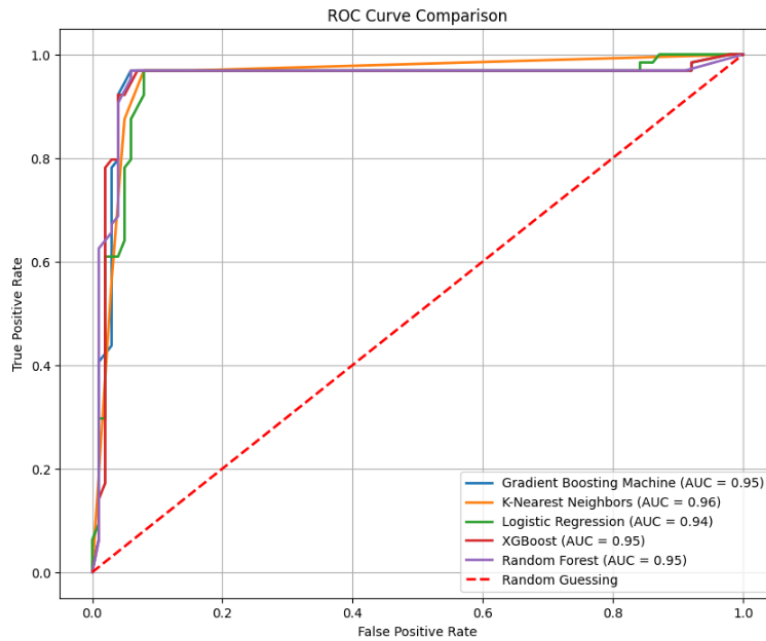
#Modelo implementado

Random Forest Classifier



Random Forest Classifier este es algoritmo que implemente, este funciona creando multiples arboles de desicion a partir de datos aleatorios de los datos y caracteristicas, toma en cuenta los votos mayoritarios de los arboles para tomar la desicion, este tipo es resistente al sobreajuste a la robustez, tiene una precision y F1 score de .92, y un AUC de .95

#COMPARACION ENTRE MODELOS



En terminos generales son modelos bastante buenos y que tiene mayor cantidad de verdaderos positivos que cualquier tipo de falso por lo que en base a lo que se requiera puede ser mas util un modelo que otro

#HALLAZGOS CLAVES

- Las personas con obesidad o sobrepeso tienen mayor probabilidad de sufrir insomnio o apnea del sueño.
- Los niveles altos de estrés están asociados con una menor calidad del sueño, lo que es coherente con la literatura neurocientífica sobre el cortisol (y por tanto niveles de estres)
- Las mujeres muestran mayor variabilidad en la calidad del sueño en comparación con los hombres
- Los modelos mostraron un rendimiento confiable para poder predecir los trastornos del sueño k- vecinos resalto notablemente por su AUC de .96 aunque tuvo una precision inferior a otros modelos