



FACTORES DE SALUD QUE INFLUYEN EN LA CALIDAD DEL SUEÑO

Introducción

El sueño saludable desempeña un papel importante, afectando de diferentes formas a nuestra calidad de vida, ya sea en relación con nuestra eficiencia laboral, en el desarrollo de nuestras tareas diarias, en la prevención de enfermedades o en el cuidado de nuestra salud mental. Dada la creciente prevalencia de los problemas de sueño en la actualidad, hemos decidido dedicar nuestro análisis a este tema.

¿Qué factores de salud son más relevantes para saber si alguien es más propenso a tener trastornos durante el sueño?

Para el análisis vamos a usar árboles de decisión y regresores logísticos.

Dataset

Fuentes: <https://www.kaggle.com/datasets/uom190346a/sleep-health-and-lifestyle-dataset>

 374 Filas
 13 Variables

Categorías:

BMI Category, Occupation, Blood pressure, Gender, Sleep Disorder.

Numéricas:

ID, Age, Sleep Duration, Sleep Quality, Physical Activity Level, Stress Level, Daily Steps, Heart Rate.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 374 entries, 0 to 373
Data columns (total 13 columns):
# Column Non-Null Count Dtype
---  -
0 Person ID 374 non-null int64
1 Gender 374 non-null object
2 Age 374 non-null int64
3 Occupation 374 non-null object
4 Sleep Duration 374 non-null float64
5 Quality of Sleep 374 non-null int64
6 Physical Activity Level 374 non-null int64
7 Stress Level 374 non-null int64
8 BMI Category 374 non-null object
9 Blood Pressure 374 non-null int64
```

Variables Categóricas

5 features categóricos:

Gender: ['Male', 'Female'],

Occupation: ['Software Engineer', 'Doctor', 'Sales Representative', 'Teacher', 'Nurse', 'Engineer', 'Accountant', 'Scientist', 'Lawyer', 'Salesperson', 'Manager']

BMI Category: ['Overweight', 'Normal', 'Obese', 'Normal Weight']

Blood Pressure: sistólica/diastólica. (Vamos a separarla en presión alta y presión baja (ej: '120/85' a 120 y 85))

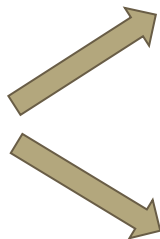
Sleep Disorder: ['None', 'Sleep Apnea', 'Insomnia']

Target

Problemática

Desorden de Sueño

Son dificultades para conciliar el sueño, o conductas anormales durante el sueño.



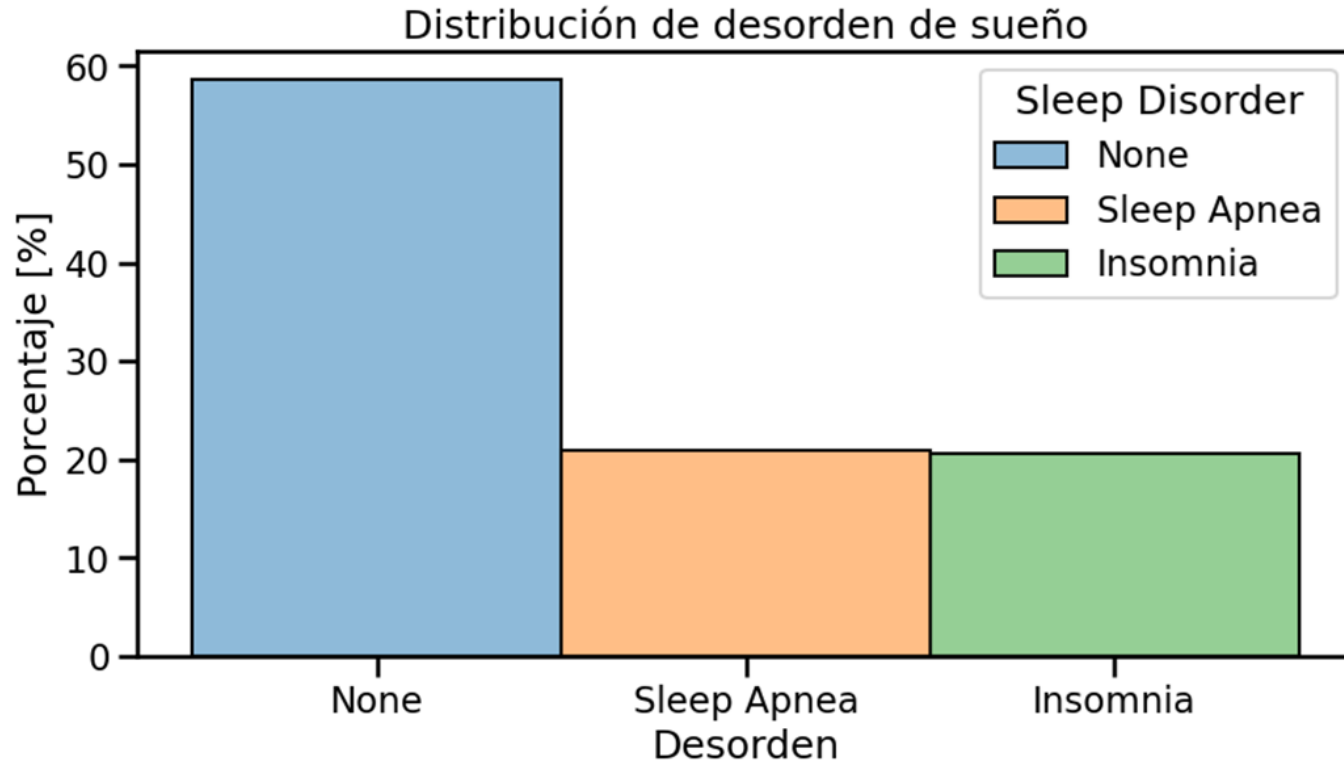
Insomnio

Es cuando se tiene dificultades para conciliar el sueño.

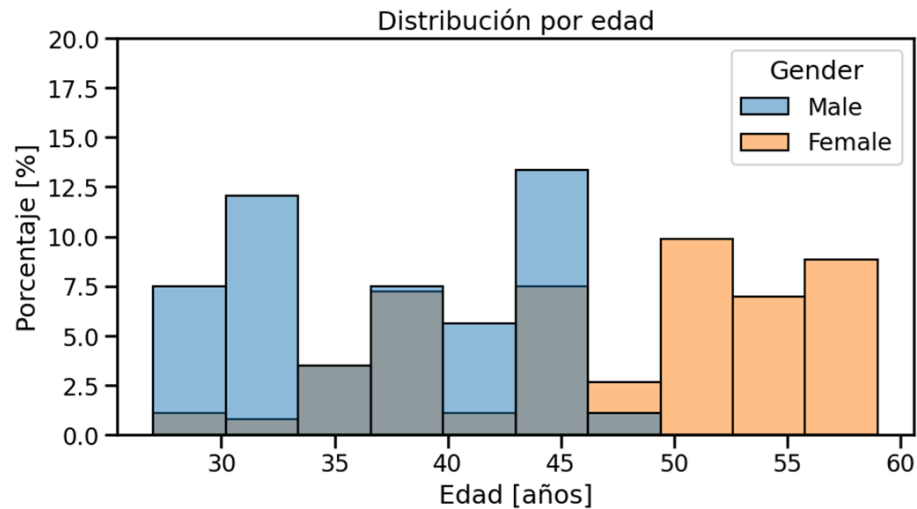
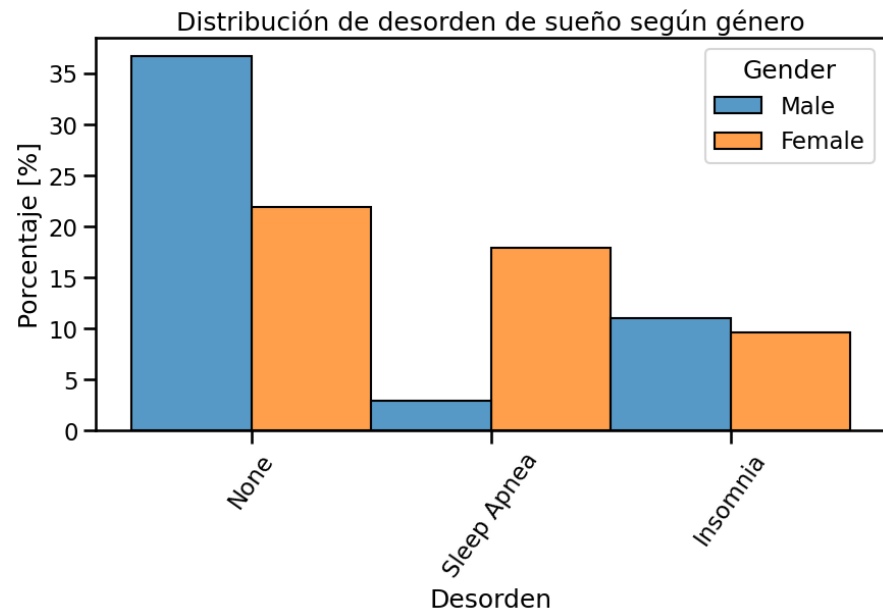
Apnea de Sueño

Es cuando la respiración se interrumpe o se hace muy superficial.

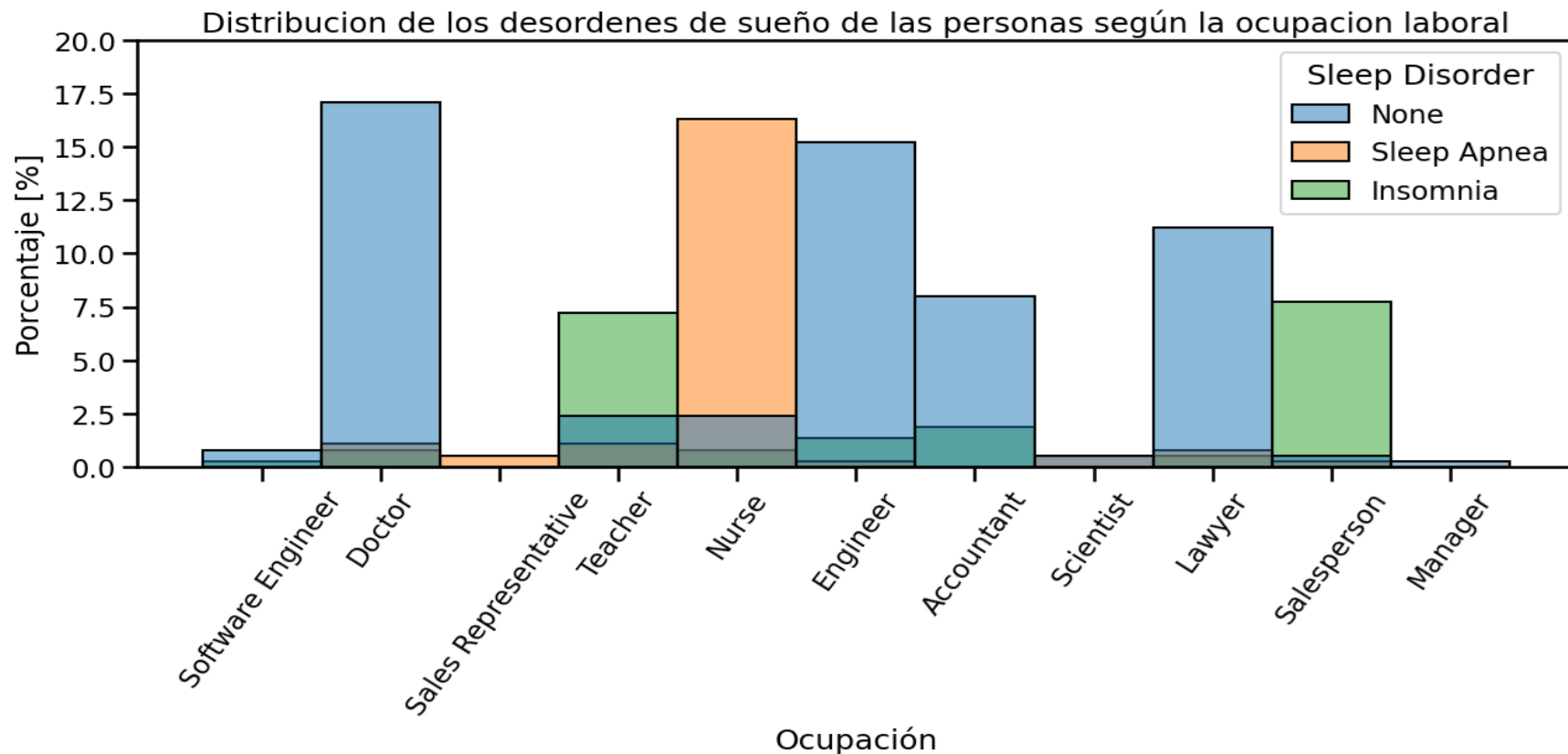
Análisis exploratorio



Análisis exploratorio

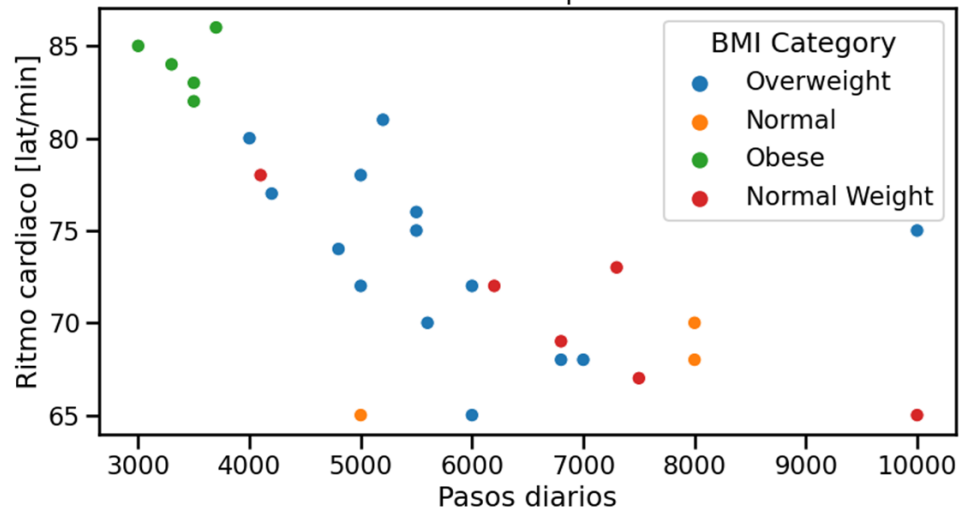


Análisis exploratorio

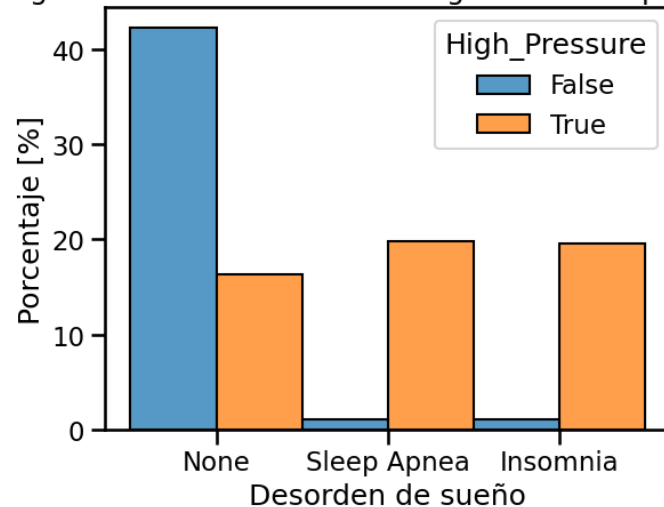


Análisis exploratorio

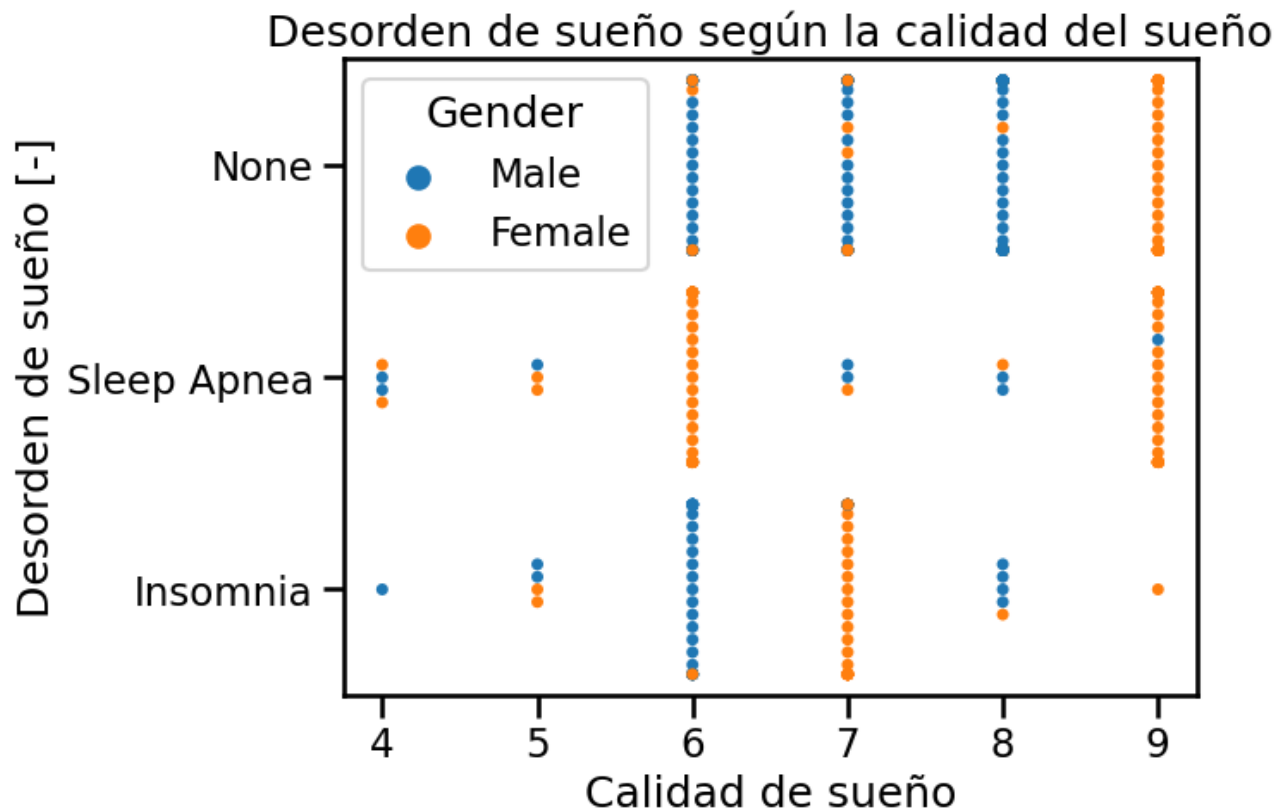
Ritmo cardiaco vs pasos diarios



Histograma desorden de sueño segun si tienen presion alta



Análisis exploratorio



Análisis exploratorio

Del análisis exploratorio junto a la preparación de datos decidimos que....

las variables descartadas son:

Categorías: Occupation.

Numéricas: ID y Sleep Quality.

Variables a utilizar son:

Categorías:

BMI Category, Blood Pressure, Gender, Sleep Disorder.

Numéricas:

Age, Sleep Duration, Physical Activity Level, Stress Level, Heart Rate, Daily Steps..

Preparación de los datos

Variables Categóricas: codificación por label encoding

Ocupación	→	Trabajador de salud	<ul style="list-style-type: none">• 1• 0
BMI Category	→	BMI	<ul style="list-style-type: none">• Normal = 0• Sobrepeso = 1• Obesidad = 2
Presión sanguínea	→	Presión Alta	<ul style="list-style-type: none">• 1• 0
Género	→	Mujer	<ul style="list-style-type: none">• 1• 0
Desorden del sueño	→	Desorden	<ul style="list-style-type: none">• None = 0• Apnea = 1• Insomnio = 2

Preparación de los datos

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 374 entries, 0 to 373
Data columns (total 10 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Age                                    374 non-null    int64
1   Sleep_Duration                        374 non-null    float64
2   Physical_Activity_Level               374 non-null    int64
3   Stress                                374 non-null    int64
4   Heart_rate                            374 non-null    int64
5   Daily_steps                           374 non-null    int64
6   BMI                                    374 non-null    int64
7   Desorden                              374 non-null    int64
8   Women                                 374 non-null    bool
9   High_Pressure                         374 non-null    bool
dtypes: bool(2), float64(1), int64(7)
memory usage: 24.2 KB
```

Entrenaminto: 300
Test: 74

k-fold entre 40 y 80

Selección del modelo de ajuste

Modelos



Métricas

- Accuracy Score para encontrar el mejor modelo de cada tipo
- Matriz de confusión comparar entre árbol y regresor
- Curva ROC comparar entre árbol y regresor

Determinación de hiper-parámetros a aplicar en el modelo

Árbol

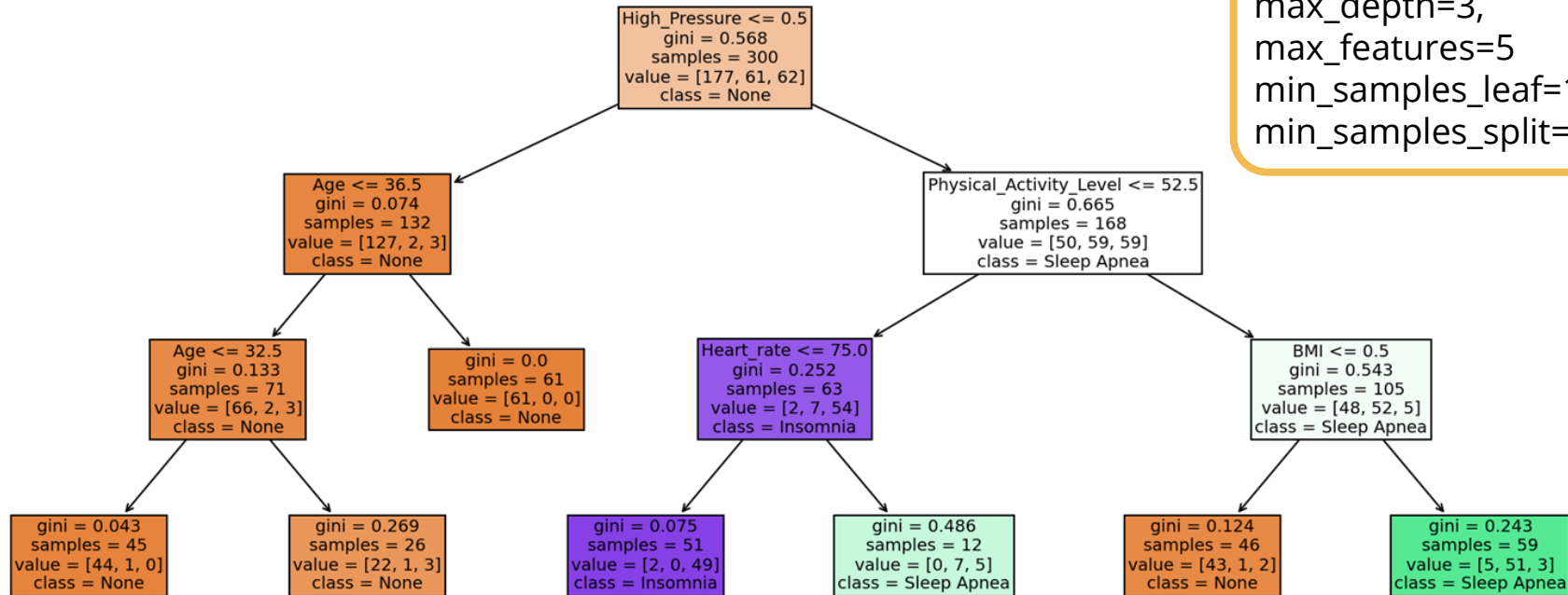
CV	Score	Max Depth	Max Feat	Min S leaf	Min S Split
80	0.905	5	4	10	40
80	0.901	9	5	10	20
80	0.909	9	6	10	20
80	0.898	5	3	20	20
80	0.905	5	3	20	40
80	0.909	9	5	10	20

Regresor

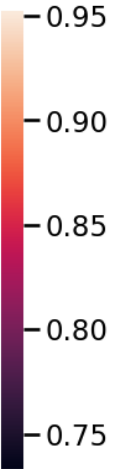
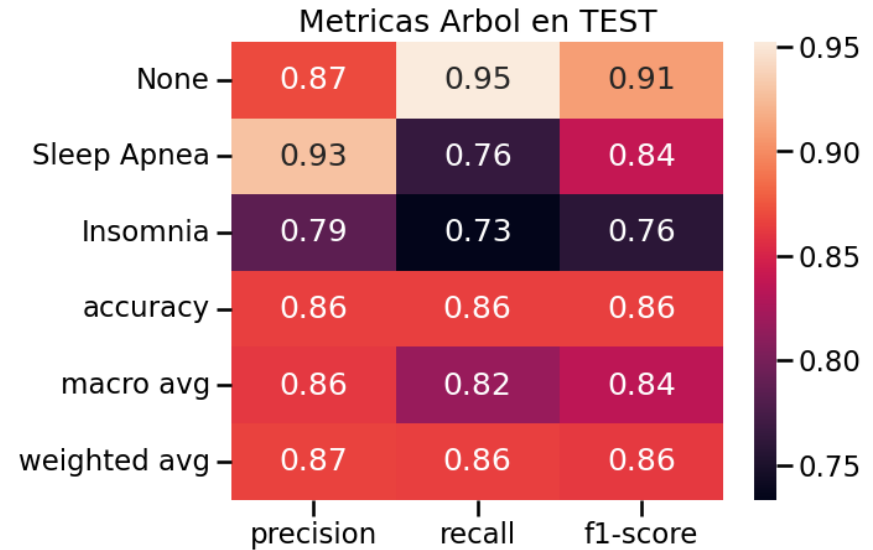
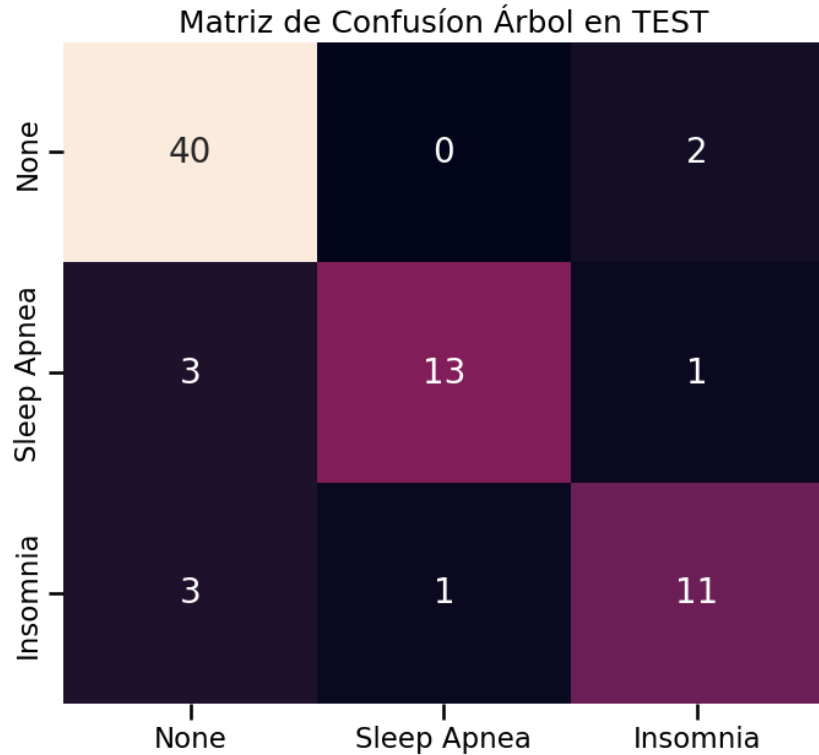
CV	Score	C	Class Weight	Max Iter	Penalty	Solver
10	0.869	10	none	500	l2	
10	0.890	1000	none	1000	l2	
20	0.897	10	none	1500	l2	
10	0.901	100		1500	l1	liblinear
10	0.923	100		1500	l1	liblinear
40	0.924	1		500	l1	liblinear

Árbol

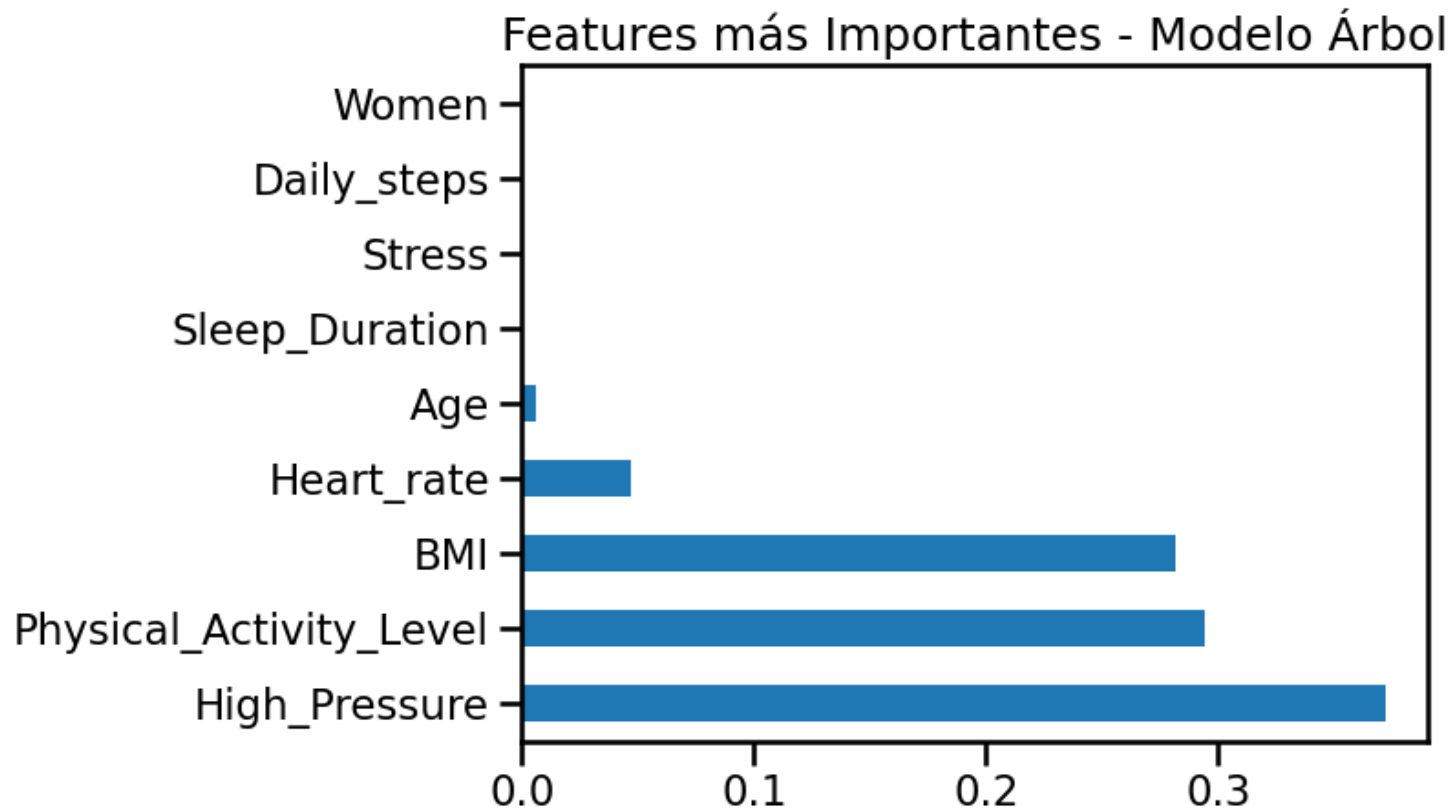
max_depth=3,
max_features=5
min_samples_leaf=10,
min_samples_split=20



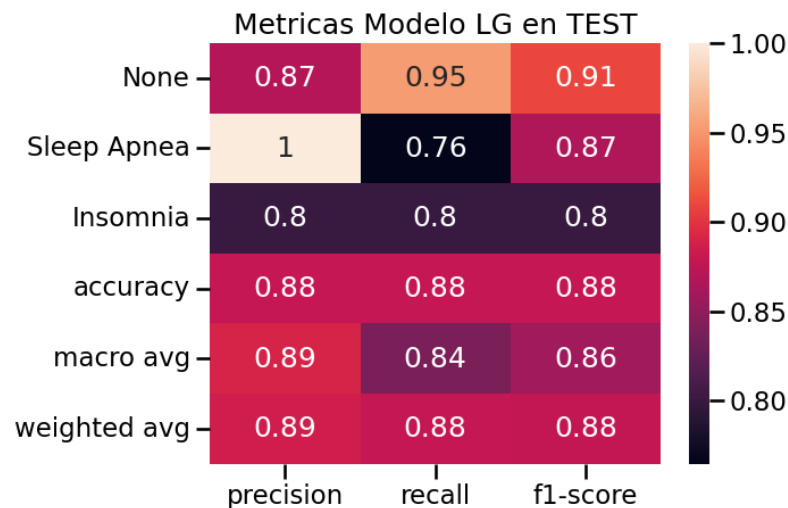
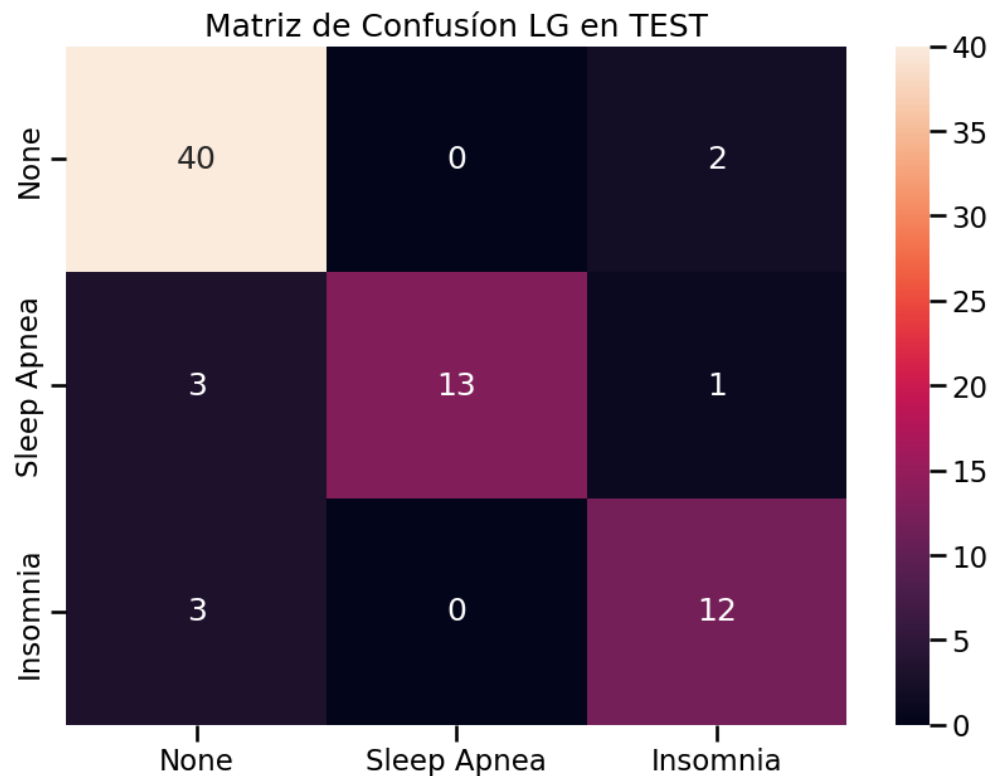
Métricas en Test - Modelo Árbol



Features más importantes - Modelo Árbol



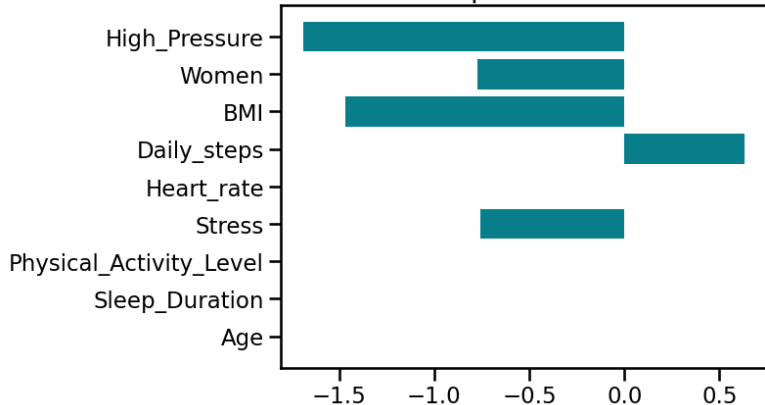
Modelo de Regresión Logística



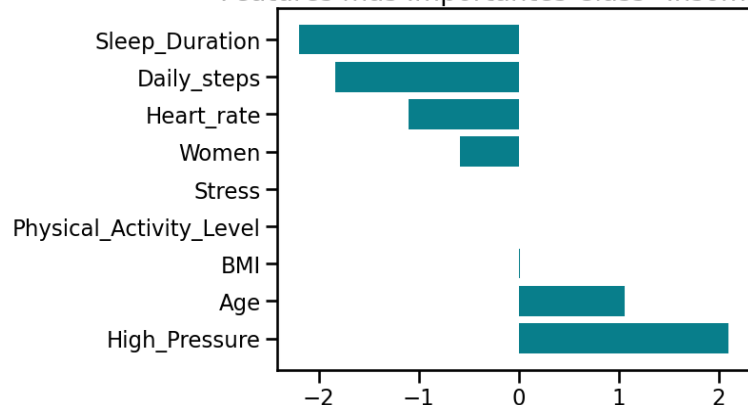
C=1
max_iter= 500
penalty= 'l1'
solver= 'liblinear'

Features más importantes - Modelo RL.

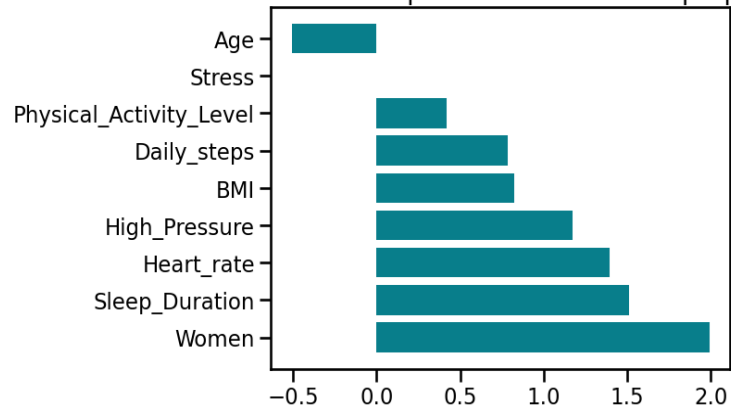
Features más Importantes Class "NONE"



Features más Importantes Class "Insomnia"

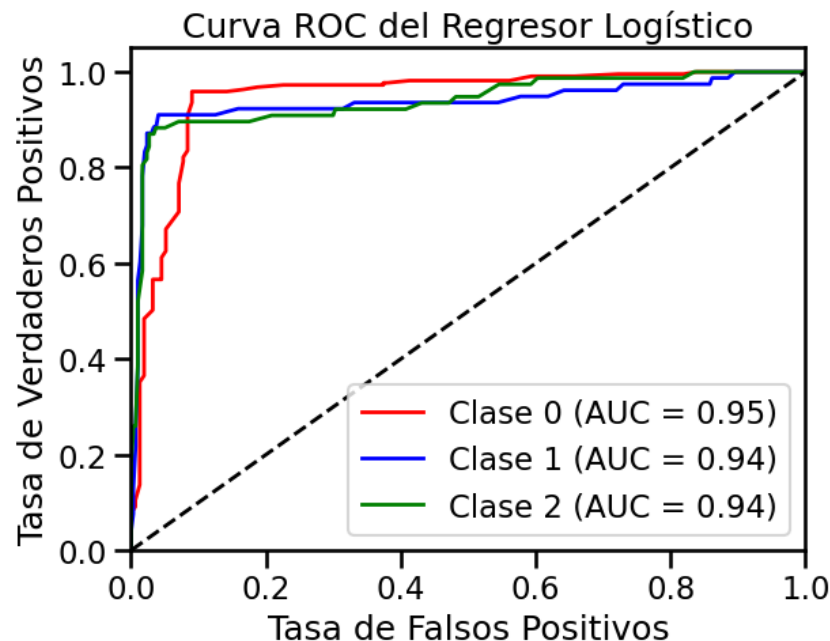
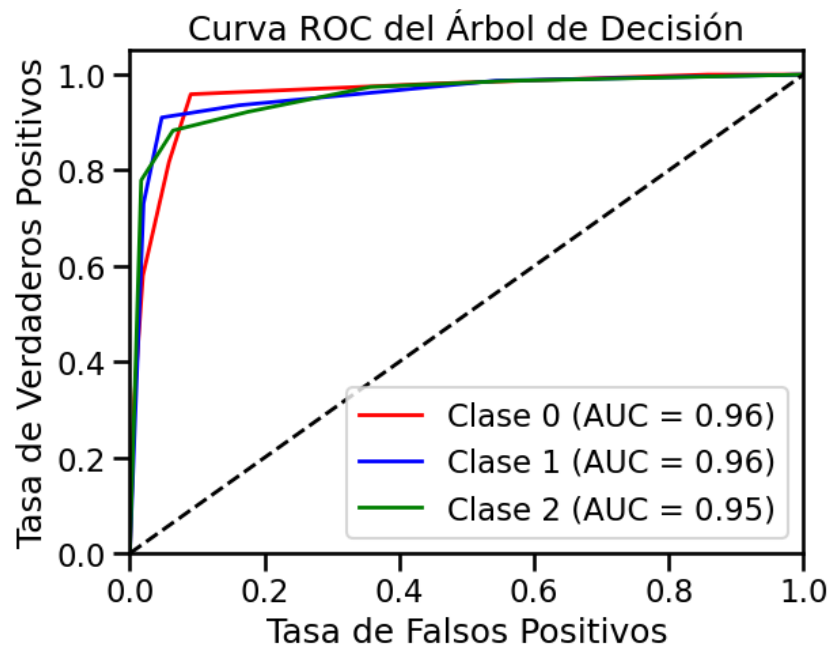


Features más Importantes Class "Sleep Apnea"



Evaluación y análisis de los modelos resultantes.

Los dos modelos son muy buenos en cuanto a métricas, incluso calculando la Curva ROC sin optimizarla da áreas en torno a 95 para todas las clases.



Conclusión

Los dos modelos tuvieron similares desempeños en cuanto a métricas, pero destacamos al RL que nos da información sobre cada clase de los features más importantes.

En cuanto a si los modelos responden la pregunta planteada, diremos que convergen al señalar algunos features como importantes: High Pressure, BMI, Physical_Activity_Level y Age, el RL pone mucho peso en el Género para la clase Sleep Apnea.

Trabajo a futuro...

Recopilar más datos para profundizar el estudio, poner foco en la ocupación y analizar si hay alguna relación de los trastornos de sueño con el ámbito médico. Y probar con Random Forest.



“GRACIAS POR TU ATENCION”