



Inteligencia Artificial: Impacto de las Redes Sociales en la salud mental

Desarrollador:

Juan Esteban Villegas Montoya

Presentado a:

**Ing. Alexander Bejarano
Ing. Ligia Stella Bustos
Prof. Carlos Mario Correa**

**Talento Tech región 2.
Pereira – Risaralda
UTP – Diciembre – 2024**

Proyecto de IA – impacto del uso de las redes sociales en la salud mental

Tabla de contenidos:

1.	Introducción	3
2.	Planteamiento del problema	3
3.	Objetivos	4
3.1.	Objetivo General	4
3.2.	Objetivos específicos	4
4.	Justificación	4
5.	Alcance	5
6.	Presupuesto	5
6.1.	Costos Directos	5
6.1.1.	Recursos Humanos	5
6.1.2.	Equipos y Herramientas	5
6.1.3.	Trabajo de Campo	6
6.1.4.	Infraestructura y Logística	6
6.2.	Costos Indirectos	6
6.2.1.	Administración y Gestión del Proyecto	6
6.2.2.	Imprevistos y Contingencias	6
6.3.	Resumen de Costos	x
7.	Metodología CRISP-DM	7
7.1.	Entendimiento del Negocio	7
7.2.	Entendimiento de los Datos	7
7.3.	Preparación de los Datos	7
7.4.	Modelado	7
7.5.	Evaluación	7
7.6.	Despliegue	7
8.	Modelado de los datos	7
8.1.	Modelo y características de los datos	8
8.2.	Selección del modelo:	11
8.3.	Correlación de datos	12
9.	Resultados del Modelo	13
10.	Anexos	14
11.	Bibliografía	25

1. Introducción

En la actualidad, las redes sociales han transformado profundamente la manera en que las personas interactúan, se comunican y acceden a la información. Plataformas como Facebook, Instagram, Twitter y TikTok no solo son espacios de entretenimiento, sino que también influyen en diversos aspectos de la vida cotidiana, incluyendo el desarrollo personal y la salud mental.

La revolución digital ha sido particularmente impactante en los jóvenes, quienes constituyen uno de los principales grupos demográficos en el uso de estas plataformas. Sin embargo, se ha documentado que el uso excesivo o inadecuado de las redes sociales puede tener efectos adversos en la salud mental, tales como ansiedad, depresión y disminución de la autoestima.

En un mundo cada vez más digitalizado, comprender cómo las redes sociales moldean el comportamiento y el bienestar de las nuevas generaciones no solo es relevante, sino esencial para construir un futuro más equitativo y sostenible.

2. Planteamiento del problema.

El auge de las redes sociales ha generado un impacto sin precedentes en la sociedad, redefiniendo la forma en que las personas se comunican, acceden a la información y participan en comunidades virtuales. Este fenómeno ha sido particularmente relevante entre los jóvenes, quienes utilizan plataformas como Instagram, TikTok, Facebook y YouTube como principales canales de interacción.

El uso intensivo de redes sociales ha sido asociado con problemas significativos de salud mental. Numerosos estudios señalan una relación entre el tiempo de exposición a estas plataformas y el incremento de síntomas de ansiedad, depresión, estrés y otros trastornos psicológicos, especialmente entre adolescentes y adultos jóvenes.

La exposición constante a contenido idealizado, la presión por la validación social y la comparación constante con otros usuarios son factores que pueden contribuir al deterioro del bienestar emocional.

Ante este panorama, surge la pregunta fundamental: ¿Es posible medir con precisión el impacto de las redes sociales en la salud mental utilizando modelos actuales de machine learning (M.L.)?

3. Objetivos

3.1 Objetivo general

Analizar el impacto de las redes sociales en la salud mental de los jóvenes, utilizando técnicas de inteligencia artificial y análisis de datos.

3.2 Objetivos específicos

- Identificar patrones de uso de redes sociales entre jóvenes
- Evaluar la correlación entre tiempo de uso de redes sociales y indicadores de salud mental
- Comprender las dinámicas psicológicas asociadas al uso de plataformas sociales
- Desarrollar un modelo predictivo que relacione uso de redes sociales con variables de salud mental
- Generar recomendaciones para un uso responsable y constructivo de las redes sociales

4. Justificación

La relevancia de este proyecto radica en varios aspectos fundamentales:

- Comprensión del Fenómeno Digital: Abordar científicamente el impacto de las redes sociales en poblaciones jóvenes
- Contribución Académica: Generar conocimiento sobre las dinámicas entre tecnología digital y bienestar mental
- Aplicación Práctica: Proporcionar insights para diseñar estrategias de salud mental
- Innovación Tecnológica: Utilizar inteligencia artificial para analizar fenómenos sociales complejos

5. Alcance.

Este proyecto busca analizar el impacto de las redes sociales en la salud mental de los jóvenes mediante el uso de inteligencia artificial y análisis de datos. El alcance incluye:

- Identificación de patrones de uso de redes sociales en jóvenes mediante técnicas de análisis de datos.
- Desarrollo de un modelo predictivo basado en machine learning que relacione el uso de redes sociales con variables de salud mental.
- Propuesta de recomendaciones prácticas y basadas en evidencia para fomentar un uso saludable de las redes sociales.
- Este estudio se limita a datos obtenidos de bases públicas y se centrará en jóvenes de Colombia, utilizando un enfoque exploratorio y predictivo.

6. Presupuesto

6.1.1 Recursos Humanos

Rol	Horas trabajadas	Tarifa por hora	Subtotal
Investigador Principal	120	\$25,000	\$3,000,000
Analista de Datos	100	\$20,000	\$2,000,000
Desarrollador de Machine Learning	80	\$30,000	\$2,400,000
Psicólogo Consultor	40	\$35,000	\$1,400,000
Total Recursos Humanos			\$8,800,000

6.1.2 Equipos y Herramientas

Descripción	Cantidad	Costo Unitario	Subtotal
Computador portátil de alto rendimiento	2	\$3,500,000	\$7,000,000
Licencia de software de análisis de datos (por año)	1	\$1,500,000	\$1,500,000
Recursos en la nube para procesamiento	-	\$500,000	\$500,000
Total Equipos y Herramientas			\$9,000,000

Proyecto de IA – impacto del uso de las redes sociales en la salud mental

6.1.3 Trabajo de Campo

Descripción	Cantidad	Costo Unitario	Subtotal
Encuestas y recopilación de datos	-	\$1,000,000	\$1,000,000
Viáticos para trabajo de campo	-	\$500,000	\$500,000
Total Trabajo de Campo			\$1,500,000

6.1.4 Infraestructura y Logística

Descripción	Cantidad	Costo Unitario	Subtotal
Espacio de trabajo	3 meses	\$200,000	\$600,000
Servicios públicos	3 meses	\$100,000	\$300,000
Materiales de oficina	-	\$150,000	\$150,000
Total Infraestructura y Logística			\$1,050,000

6.2.1 Administración y Gestión del Proyecto

Descripción	Subtotal
Gastos administrativos	\$1,000,000
Gestión y coordinación	\$800,000
Total Administración y Gestión	\$1,800,000

6.2.2 Imprevistos y Contingencias

Descripción	Porcentaje	Monto
Contingencias (10% del total)	10%	\$2,215,000

6.3 Resumen de Costos

Categoría	Monto
Total Costos Directos	\$20,350,000
Total Costos Indirectos	\$4,015,000
Presupuesto Total del Proyecto	\$24,365,000

7. Metodología

La metodología utilizada será CRISP-DM, que se divide en las siguientes fases:

1. Entendimiento del Negocio
 - Comprender el impacto de las redes sociales en la salud mental de los jóvenes.
 - Definir las preguntas clave del análisis: ¿Cuáles son las variables que más afectan la salud mental?
2. Entendimiento de los Datos
 - Utilizar el dataset de Kaggle relacionado con redes sociales y salud mental.
 - Realizar un análisis exploratorio (EDA) para identificar tendencias, valores atípicos y correlaciones.
3. Preparación de los Datos
 - Limpieza y preprocesamiento del dataset: manejo de datos nulos, codificación de variables categóricas y normalización.
4. Modelado
 - Implementar un modelo de machine learning sencillo, como regresión lineal, para predecir el impacto de las redes sociales en la salud mental.
 - Evaluar la precisión del modelo utilizando métricas como RMSE o R2.
5. Evaluación
 - Comparar los resultados del modelo con las preguntas iniciales del negocio.
 - Validar si los patrones encontrados son coherentes con las hipótesis planteadas.
6. Despliegue
 - Presentar los resultados en un informe y en visualizaciones gráficas.
 - Proponer recomendaciones basadas en los hallazgos.

8. Modelado de datos

El modelado se llevó a cabo utilizando herramientas como Python y bibliotecas especializadas en machine learning.

Bibliotecas:

- **pandas**: Es la herramienta fundamental para la manipulación y análisis de datos en Python. Permite leer diversos formatos de datos (CSV, Excel, etc.), limpiarlos, transformarlos y realizar operaciones estadísticas básicas.
- **numpy**: Proporciona estructuras de datos eficientes (como arrays) y funciones matemáticas para realizar cálculos numéricos de manera rápida y precisa. Es la base para muchas otras librerías de ciencia de datos.

- **matplotlib**: Se utiliza para crear visualizaciones estáticas, como gráficos de líneas, histogramas, diagramas de dispersión, etc. Ayuda a explorar y entender los datos de manera visual.
- **seaborn**: Construido sobre matplotlib, seaborn ofrece una interfaz más alta y estética para crear visualizaciones más informativas y atractivas. Es especialmente útil para visualizar relaciones entre variables.
- **train_test_split**: Esta función divide un conjunto de datos en dos subconjuntos: uno para entrenar el modelo (train) y otro para evaluarlo (test). Es crucial para evitar el sobreajuste del modelo.
- **LinearRegression**: Implementa el algoritmo de regresión lineal, uno de los modelos más simples y utilizados en machine learning. Permite predecir un valor numérico continuo a partir de una o más variables independientes.
- **StandardScaler**: Estandariza los datos, es decir, los escala para que tengan una media de 0 y una desviación estándar de 1. Esto es importante para muchos algoritmos de machine learning que asumen que los datos están en una escala similar.
- **LabelEncoder**: Convierte variables categóricas (texto) en numéricas, lo cual es necesario para muchos algoritmos de machine learning.
- **OneHotEncoder**: Crea variables ficticias (one-hot encoding) a partir de variables categóricas con múltiples categorías. Esto permite representar las categorías como variables binarias.
- **mean_squared_error**: Calcula el error cuadrático medio, una métrica común para evaluar la calidad de un modelo de regresión. Mide la diferencia promedio entre los valores predichos por el modelo y los valores reales.
- **r2_score**: Calcula el coeficiente de determinación (R^2), otra métrica de evaluación que indica la proporción de la varianza de la variable dependiente que es explicada por el modelo.

8.1. Modelo:

Para la ejecución de este proyecto se probó usar 3 distintos tipos de Machine Learning el primero fue Machine Learning con Regresión lineal el segundo fue Machine Learning con Árboles de decisión y el tercer método empleado fue Machine Learning con Random Forest.

Proyecto de IA – impacto del uso de las redes sociales en la salud mental

Se decidió continuar la realización del proyecto usando el método machine learning con regresión lineal debido a que ese es el que presenta un r cuadrado menor con respecto a los otros al igual que un error cuadrático menor respecto a los otros métodos anteriormente mencionados a continuación se anexa fotografía evidencian tu esto que se menciona:

```
Conclusiones del Modelo de Regresión Lineal
-----
Error Cuadrático Medio (MSE): 1.2382
R-cuadrado: 0.0006
El modelo no explica una proporción significativa de la variabilidad en los datos. Considera explorar otros modelos o características.
Se han generado las predicciones de la salud mental. Puedes ver los resultados y compararlos con los valores reales en el archivo 'predicciones_salud_men
tal_reg_lin.xlsx'.
```

Regresión lineal.

```
Conclusiones del Modelo de Árbol de Decisión
-----
Error Cuadrático Medio (MSE): 0.6490
R-cuadrado: -0.0126

Importancia de las características:
Social_Media_Usage_Hours: 0.2932
Sleep_Hours: 0.2786
Gaming_Hours: 0.2439
Age: 0.1033
Physical_Activity_Hours: 0.0744
Gender_Other: 0.0066
Gender_Female: 0.0000
Gender_Male: 0.0000

Análisis de las características más influyentes:
La característica 'Social_Media_Usage_Hours' tiene una importancia de 0.2932 en predecir los niveles de estrés.
La característica 'Sleep_Hours' tiene una importancia de 0.2786 en predecir los niveles de estrés.
La característica 'Gaming_Hours' tiene una importancia de 0.2439 en predecir los niveles de estrés.

Análisis de las predicciones:
Se han generado las predicciones de los niveles de estrés. Puedes ver los resultados y compararlos con los valores reales en el archivo 'predicciones_est
rés_arbol.xlsx'.
```

Arboles de decisión.

Error Cuadrático Medio (MSE) del Random Forest: 1.3135
R-cuadrado del Random Forest: -0.0601

Randon Forest

Variables usadas:

'Age' 'Gender' 'Technology_Usage_Hours' 'Social_Media_Usage_Hours'
'Gaming_Hours' 'Screen_Time_Hours' 'Mental_Health_Status' 'Stress_Level'
'Sleep_Hours' 'Physical_Activity_Hours' 'Support_Systems_Access'
'Work_Environment_Impact' 'Online_Support_Usage'

Preprocesamiento:

Las variables fueron graficadas y se determino usar 3 para el modelo:
'Social_Media_Usage_Hours' 'Mental_Health_Status' 'Stress_Level'.

Como se mencionó anteriormente se probaron 3 distintos métodos de Machine learning y a la final se decidió trabajar con el método de regresión lineal porque este método nos dio un R-cuadrado mejor con respecto a los otros, A ver recalcar que en cuanto al error cuadrático el que dio el resultado menor fue él utilizando el modelo de árboles de decisión, pero este a su vez generó el problema de trabajar con un R-cuadrado negativo lo cual no es bueno.

8.2 Selección del modelo:

- Se utilizó un modelo de regresión lineal otro de árboles de decisión para identificar la relación entre el tiempo en redes sociales y la salud mental.
- Se evaluaron ambos modelos y se evidenciaron algunas diferencias entre ellos.
- ¿Por que el modelo? Este modelo se selecciono por lo mencionado con anterioridad sobre el R-cuadrado y además por su facilidad en la aplicación.
- ¿Por que se usa scikit-learn?
scikit-learn es una de las bibliotecas más populares y poderosas de Python para el aprendizaje automático. En este modelo de regresión lineal, se utiliza principalmente por las siguientes razones:

1. Facilidad de uso:

- **Funciones predefinidas:** Proporciona una amplia gama de algoritmos de aprendizaje automático, incluyendo la regresión lineal, ya implementados. Esto ahorra mucho tiempo y esfuerzo al desarrollador.
- **API consistente:** Mantiene una interfaz de programación de aplicaciones (API) consistente a través de todos sus algoritmos, lo que facilita el cambio entre diferentes modelos.

2. Compleitud:

- **Preprocesamiento de datos:** Ofrece herramientas para tareas comunes de preprocesamiento, como la división de conjuntos de datos (`train_test_split`), la codificación de variables categóricas (`OneHotEncoder`, `LabelEncoder`) y la estandarización de datos (`StandardScaler`).
- **Selección de modelos:** Permite comparar fácilmente diferentes modelos y elegir el que mejor se adapte a los datos.
- **Evaluación de modelos:** Incluye métricas de evaluación como el error cuadrático medio (MSE) y el coeficiente de determinación (R^2) para medir la calidad de los modelos.

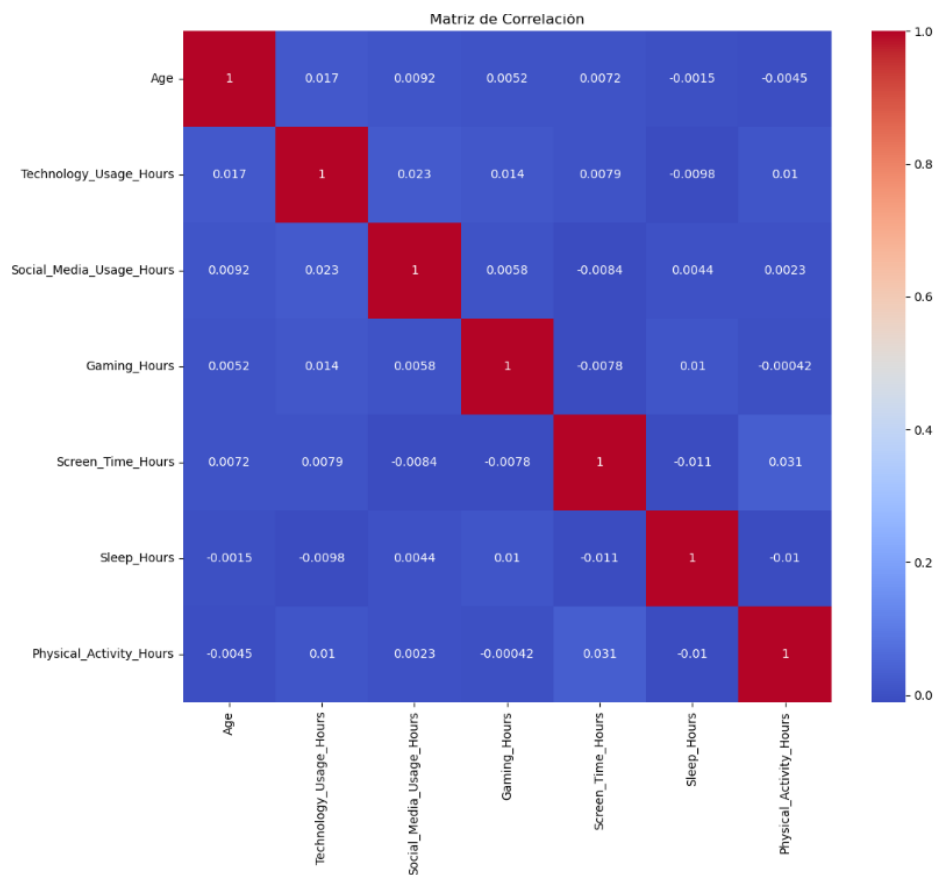
3. Eficiencia:

- **Optimización:** Los algoritmos de scikit-learn están altamente optimizados para ejecutarse de manera eficiente, incluso en grandes conjuntos de datos.

4. Integración:

- **Compatibilidad:** Se integra fácilmente con otras bibliotecas de Python como NumPy y pandas, lo que permite construir flujos de trabajo de análisis de datos completos.

8.3 Correlación de datos:



La anterior imagen hace referencia a la matriz de correlación de datos de nuestra base de datos.

9. Resultados del modelo

El modelo obtuvo un R-cuadrado de 0.0006, lo que indica que:

- Modelo poco preciso: Las predicciones que genera el modelo estarán un poco alejadas de los valores reales.
- Llegamos a la conclusión de que el modelo no es muy óptimo para lograr predecir. Prezi el impacto de las redes sociales en la salud mental, esto debido a que el comportamiento de los datos en la base de datos no tiene una estrecha relación con lo que queremos evaluar, también cabe recalcar que dentro de los 3 modelos verificados este modelo de regresión lineal es el que mejor predicción de datos podría llegar a dar, aun así, no siendo muy bueno para este fin.

La variable con mayor impacto fue el tiempo diario en redes sociales, seguida por la calidad del sueño.

Conclusiones del Modelo de Regresión Lineal

Error Cuadrático Medio (MSE): 0.6423
R-cuadrado: -0.0023

Coefficientes del Modelo:
Social_Media_Usage_Hours: -0.0052
Gaming_Hours: -0.0063
Age: 0.0002
Sleep_Hours: -0.0079
Physical_Activity_Hours: -0.0005
Gender_Female: -0.0053
Gender_Male: -0.0034
Gender_Other: 0.0087

Análisis de las características más influyentes:

La característica 'Social_Media_Usage_Hours' tiene una relación negativa con el estrés.
La característica 'Gaming_Hours' tiene una relación negativa con el estrés.
La característica 'Age' tiene una relación positiva con el estrés.
La característica 'Sleep_Hours' tiene una relación negativa con el estrés.
La característica 'Physical_Activity_Hours' tiene una relación negativa con el estrés.
La característica 'Gender_Female' tiene una relación negativa con el estrés.
La característica 'Gender_Male' tiene una relación negativa con el estrés.
La característica 'Gender_Other' tiene una relación positiva con el estrés.

Análisis de las predicciones:

Se han generado las predicciones de los niveles de estrés. Puedes ver los resultados y compararlos con los valores reales en el archivo 'predicciones_estrés.xlsx'.

Proyecto de IA – impacto del uso de las redes sociales en la salud mental

Real	Predicción
Poor	Good
Excellent	Fair
Excellent	Fair
Excellent	Fair
Excellent	Fair
Excellent	Good
Fair	Good
Poor	Good
Good	Fair
Poor	Fair
Excellent	Good
Excellent	Fair
Good	Good
Good	Fair
Excellent	Fair
Good	Good
Poor	Good
Poor	Fair
Fair	Fair
Poor	Good
Poor	Fair
Poor	Fair

La anterior imagen, simplemente es una pequeña muestra de lo que se pudo obtener mediante la aplicación del modelo de machine learning (ML) con regresión lineal.

Este archivo está anexo en la misma carpeta donde está este documento.

10. Anexos

Proyecto IA, Impacto de redes sociales en la salud mental, Versión con modelo de ML con regresión lineal.

1. Requerimientos:

- **Dataset:** mental_health_and_technology_usage_2024.csv
- **Librerías:** pandas, numpy, matplotlib, seaborn...

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.preprocessing import StandardScaler, LabelEncoder
from sklearn.preprocessing import OneHotEncoder, LabelEncoder
from sklearn.metrics import mean_squared_error
from sklearn.metrics import mean_squared_error, r2_score
```

2. Cargar el Dataset

```
df = pd.read_csv('mental_health_and_technology_usage_2024.csv')
print(df.head())
print(df.info())
```

3. Exploración Inicial de los Datos

Revisamos la estructura general del dataset, sus columnas, y verificamos si hay datos faltantes.

```
df.info()
df.describe()
```

4. Análisis Exploratorio de Datos (EDA)

A continuación, generaremos visualizaciones para comprender mejor la distribución de los datos.

```
] : file_path = 'mental_health_and_technology_usage_2024.csv'
data = pd.read_csv(file_path)

plt.rcParams['figure.figsize'] = (10, 5) # configuracion de las graficas

# Seleccionar columnas que contienen índices relevantes
indices_cols = [col for col in data.columns if
    'Age' in col or
    'Gender' in col or
    'Technology_Usage_Hours' in col or
    'Social_Media_Usage_Hours' in col or
    'Gaming_Hours' in col or
    'Screen_Time_Hours' in col or
    'Mental_Health_Status' in col or
    'Stress_Level' in col or
    'Sleep_Hours' in col or
    'Physical_Activity_Hours' in col or
    'Support_Systems_Access' in col or
    'Work_Environment_Impact' in col or
    'Online_Support_Usage' in col
]

# Generar histogramas
for col in indices_cols:
    plt.figure()
    sns.histplot(data=data, x=col, kde=True, bins=15, color='Orange')
    plt.title(f'Distribución de {col}', fontsize=14)
    plt.xlabel(col, fontsize=12)
    plt.ylabel('Frecuencia', fontsize=12)
    plt.show()
```

5. Comparativa de uso ente generos & edades

```
] : df = pd.read_csv('mental_health_and_technology_usage_2024.csv')

# Crear los bins y labels para las edades
bins = [0, 12, 18, 30, 50, 100]
labels = ['0-12', '13-18', '19-30', '31-50', '51+']

# Crear la columna de rango de edad
df['Age_Group'] = pd.cut(df['Age'], bins=bins, labels=labels, right=False)

# Gráfico 1: Uso de redes sociales por rango de edad
plt.figure(figsize=(10, 6))
sns.countplot(x='Age_Group', data=df)
plt.title("Uso de Redes Sociales por Rango de Edad")
plt.xlabel("Rango de Edad")
plt.ylabel("Número de Usuarios")
plt.show()

# Gráfico 2: Uso de redes sociales por género
plt.figure(figsize=(8, 6))
sns.countplot(x='Gender', data=df)
plt.title("Uso de Redes Sociales por Género")
plt.xlabel("Género")
plt.ylabel("Número de Usuarios")
plt.show()

# Gráfico 3: Uso de redes sociales por género y rango de edad (gráfico de barras apiladas)
plt.figure(figsize=(12, 6))
sns.countplot(x='Age_Group', hue='Gender', data=df)
plt.title("Uso de Redes Sociales por Género y Rango de Edad")
plt.xlabel("Rango de Edad")
plt.ylabel("Número de Usuarios")
plt.legend(title="Género")
plt.show()
```

6. Analisis columnas a grafico ¶

```
] data = pd.read_csv("mental_health_and_technology_usage_2024.csv", encoding='utf-8')

# Seleccionar las columnas relevantes
indices_cols = [col for col in data.columns if
    'Age' in col or
    'Gender' in col or
    'Technology_Usage_Hours' in col or
    'Social_Media_Usage_Hours' in col or
    'Gaming_Hours' in col or
    'Screen_Time_Hours' in col or
    'Mental_Health_Status' in col or
    'Stress_Level' in col or
    'Sleep_Hours' in col or
    'Physical_Activity_Hours' in col or
    'Support_Systems_Access' in col or
    'Work_Environment_Impact' in col or
    'Online_Support_Usage' in col
]

# Crear la matriz de gráficos de dispersión
sns.pairplot(data[indices_cols], hue='Gender', diag_kind='kde')

# Personalizar el gráfico
plt.figure(figsize=(15, 12))
sns.set_palette("husl")
plt.suptitle("Relaciones entre variables de salud y uso de tecnología", fontsize=16)
plt.show()
```

7. Modelo de regresión lineal

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder, StandardScaler
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error
import seaborn as sns
import matplotlib.pyplot as plt

# Leer los datos
data = pd.read_csv("mental_health_and_technology_usage_2024.csv")

# Identificar columnas categóricas y numéricas
categorical_columns = ['Stress_Level'] # Reemplaza con las columnas categóricas reales
numerical_columns = ['Age', 'Social_Media_Usage_Hours', 'Gaming_Hours', 'Sleep_Hours', 'Physical_Activity_Hours', 'Stress_Level']

# Procesar columnas categóricas
for col in categorical_columns:
    le = LabelEncoder()
    data[col] = le.fit_transform(data[col])

# Seleccionar las columnas de entrada y salida
X = data[numerical_columns + categorical_columns]
y = data['Mental_Health_Status']

# Transformar la salida usando LabelEncoder
y = le.fit_transform(y)

# Escalar las columnas numéricas
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)

# Dividir los datos en entrenamiento y prueba
X_train, X_test, y_train, y_test = train_test_split(X_scaled, y, test_size=0.2, random_state=42)

# Crear y entrenar el modelo
model = LinearRegression()
model.fit(X_train, y_train)

# Hacer predicciones
y_pred = model.predict(X_test)

# Evaluar el modelo
mse = mean_squared_error(y_test, y_pred)
print('Error cuadrático medio:', mse)

# Visualizar la relación entre las variables
sns.pairplot(data)
plt.show()
```


8. Matriz de correlación

```
}}: df = pd.read_csv('mental_health_and_technology_usage_2024.csv')

# Seleccionar solo las columnas numéricas (si es necesario)
df_numeric = df.select_dtypes(include=['number'])

# Manejar valores faltantes (opcional)
# df_numeric.fillna(df_numeric.mean(), inplace=True) # Rellenar con la media

# Calcular la matriz de correlación
corr_matrix = df_numeric.corr()

# Crear un mapa de calor
plt.figure(figsize=(12, 10))
sns.heatmap(corr_matrix, annot=True, cmap='coolwarm')
plt.title('Matriz de Correlación')
plt.show()
```

9. Modelo de Machine Learning con regresión lineal

```
1: # 1. Cargar el archivo CSV
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
from sklearn.preprocessing import OneHotEncoder, LabelEncoder

# 2. Cargar el archivo CSV
data = pd.read_csv('mental_health_and_technology_usage_2024.csv')

# 3. Seleccionar las variables relevantes
X = data[['Social_Media_Usage_Hours', 'Gaming_Hours', 'Age', 'Gender', 'Sleep_Hours', 'Physical_Activity_Hours', 'Stress_Level']]
y = data['Mental_Health_Status']

# 4. Asegurarnos de que los nombres de las columnas sean de tipo str
X.columns = X.columns.astype(str)

# 5. Convertir las columnas relevantes a tipo numérico
X.loc[:, ['Social_Media_Usage_Hours', 'Gaming_Hours', 'Age', 'Sleep_Hours', 'Physical_Activity_Hours']] = X[['Social_Media_Usage_Hours', 'Gaming_Hou

# 6. Codificar la variable 'Gender' con One-Hot Encoding
encoder = OneHotEncoder(sparse_output=False)
encoded_gender = encoder.fit_transform(X[['Gender']])

# 7. Concatenar las columnas codificadas y eliminar la original
X_encoded = pd.concat([X.drop('Gender', axis=1), pd.DataFrame(encoded_gender, columns=encoder.get_feature_names_out(['Gender']))], axis=1)

# 8. Codificar la variable 'Stress_Level' con LabelEncoder (convertir a valores numéricos)
label_encoder_stress = LabelEncoder()
X_encoded['Stress_Level'] = label_encoder_stress.fit_transform(X_encoded['Stress_Level'])

# 9. Codificar la variable 'Mental_Health_Status' con LabelEncoder (convertir a valores numéricos)
label_encoder_health = LabelEncoder()
y_encoded = label_encoder_health.fit_transform(y)

# 10. Dividir los datos en entrenamiento y prueba
X_train, X_test, y_train, y_test = train_test_split(X_encoded, y_encoded, test_size=0.2, random_state=42)

# 11. Crear y entrenar el modelo de regresión lineal
model = LinearRegression()
model.fit(X_train, y_train)

# 12. Hacer predicciones
y_pred = model.predict(X_test)

# 13. Evaluar el modelo
# Error Cuadrático Medio (MSE): Cuanto más bajo, mejor
mse = mean_squared_error(y_test, y_pred)
print("Error Cuadrático Medio:", mse)

# R-cuadrado: Cuanto más alto, mejor (más cercano a 1 indica mejor ajuste)
r2 = r2_score(y_test, y_pred)
print("R-cuadrado:", r2)

# 14. Análisis y visualización de los resultados

# Visualización 1: Graficar los niveles de estrés reales vs. predicciones
plt.figure(figsize=(10, 6))
plt.scatter(y_test, y_pred, color='blue', label='Datos')
plt.plot([min(y_test), max(y_test)], [min(y_test), max(y_test)], color='red', linestyle='--', label='Línea de igualdad') # Línea de igualdad
plt.xlabel('Valores reales de estrés')
plt.ylabel('Predicciones de estrés')
plt.title('Valores Reales vs Predicciones de Estrés')
plt.legend()
plt.grid(True)
plt.show()

# Visualización 2: Impacto de las características en el modelo
coef_names = list(X_encoded.columns) # Nombres de columnas de X_encoded
plt.figure(figsize=(10, 6))
plt.bar(coef_names, model.coef_, color='green')
plt.xlabel('Características')
plt.ylabel('Coeficientes')
plt.title('Impacto de las Características en el Modelo')
plt.xticks(rotation=45)
plt.grid(True)
plt.show()

# Visualización 3: Histograma de errores
errors = y_pred - y_test
plt.figure(figsize=(10, 6))
plt.hist(errors, bins=20, color='purple', edgecolor='black')
plt.xlabel('Errores (Predicción - Real)')
plt.ylabel('Frecuencia')
plt.title('Distribución de Errores')
plt.grid(True)
plt.show()
```

10. Conclusiones y .xlsx final

```
2]: import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.preprocessing import OrdinalEncoder, LabelEncoder
from sklearn.metrics import mean_squared_error, r2_score

# Cargar Los datos
data = pd.read_csv('mental_health_and_technology_usage_2024.csv')

# Seleccionar Las características relevantes
X = data[['Social_Media_Usage_Hours', 'Gaming_Hours', 'Age', 'Gender', 'Sleep_Hours', 'Physical_Activity_Hours', 'Stress_Level']]
y = data['Mental_Health_Status']

# Codificar características categóricas
encoder = OrdinalEncoder(dtype=np.float64)
X_encoded = pd.DataFrame(encoder.fit_transform(X[['Social_Media_Usage_Hours', 'Gaming_Hours', 'Age', 'Gender', 'Sleep_Hours', 'Physical_Activity_Hou

# Codificar La variable objetivo (asumiendo que es categórica)
label_encoder = LabelEncoder()
y_encoded = label_encoder.fit_transform(y)

# Dividir Los datos en conjuntos de entrenamiento y prueba
X_train, X_test, y_train, y_test = train_test_split(X_encoded, y_encoded, test_size=0.2, random_state=42)

# Crear y entrenar el modelo de regresión lineal
model = LinearRegression()
model.fit(X_train, y_train)

# Hacer predicciones
y_pred = model.predict(X_test)

# Evaluar el modelo
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)

# Mapear Los valores predichos a sus etiquetas originales
y_pred_labels = label_encoder.inverse_transform(y_pred.round().astype(int))

# Crear un DataFrame con Los resultados y Las columnas adicionales
results_df = pd.DataFrame([
    'Real': label_encoder.inverse_transform(y_test),
    'Predicción': y_pred_labels,
    'Edad': X_test['Age'],
    'Género': X_test['Gender'],
])

# Mapeo de los valores de la salud mental (0, 1, 2, 3) a sus descripciones
health_status_map = {
    0: 'Excelente',
    1: 'Bueno',
    2: 'Regular',
    3: 'Malo'
}

# Crear columnas con Las descripciones de Los valores de salud mental
results_df['Estado Real de Salud Mental'] = results_df['Real'].map(health_status_map)
results_df['Estado Predicho de Salud Mental'] = results_df['Predicción'].map(health_status_map)

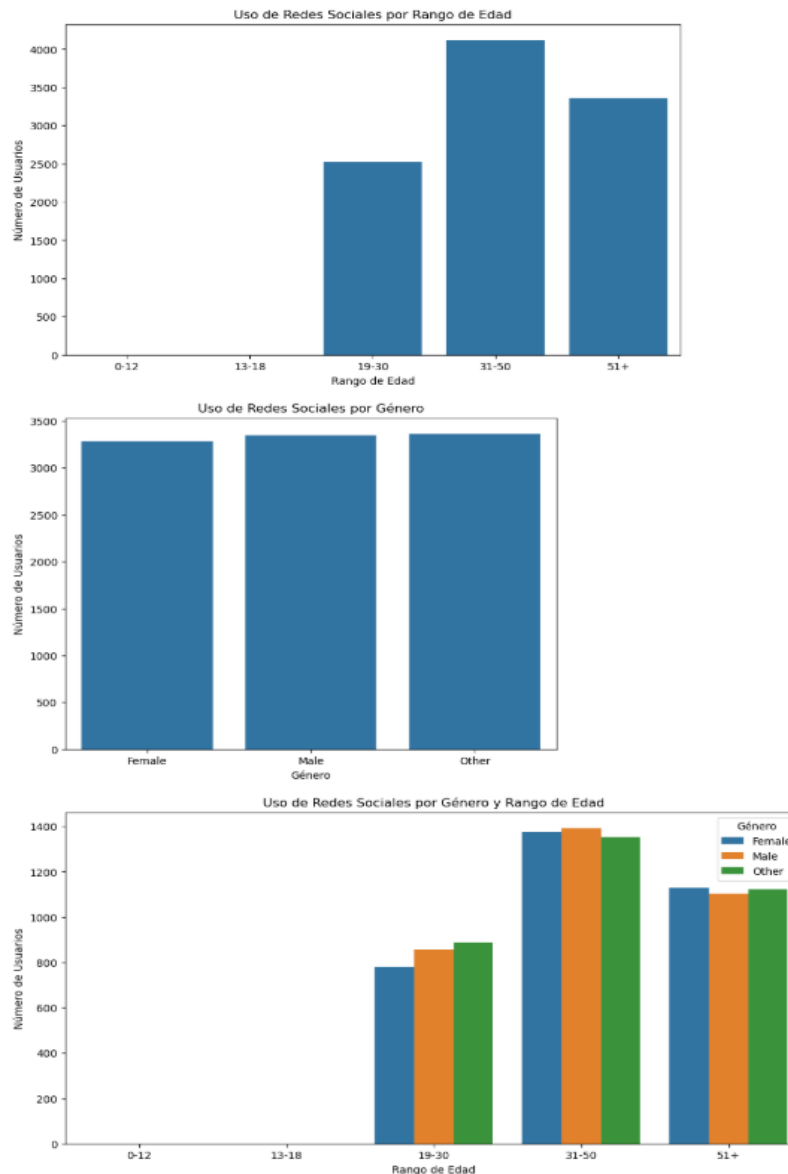
# Guardar Los resultados en un archivo Excel
results_df.to_excel('predicciones_salud_mental_regresión_lineal.xlsx', index=False)

# Imprimir conclusiones en consola
print("Conclusiones del Modelo de Regresión Lineal")
print("-----")
print(f"Error Cuadrático Medio (MSE): {mse:.4f}")
print(f"R-cuadrado: {r2:.4f}")

# Análisis adicional:
if r2 > 0.5:
    print("El modelo explica una proporción significativa de la variabilidad en los datos.")
else:
    print("El modelo no explica una proporción significativa de la variabilidad en los datos. Considera explorar otros modelos o características.")
```

Fue el skecth usado.

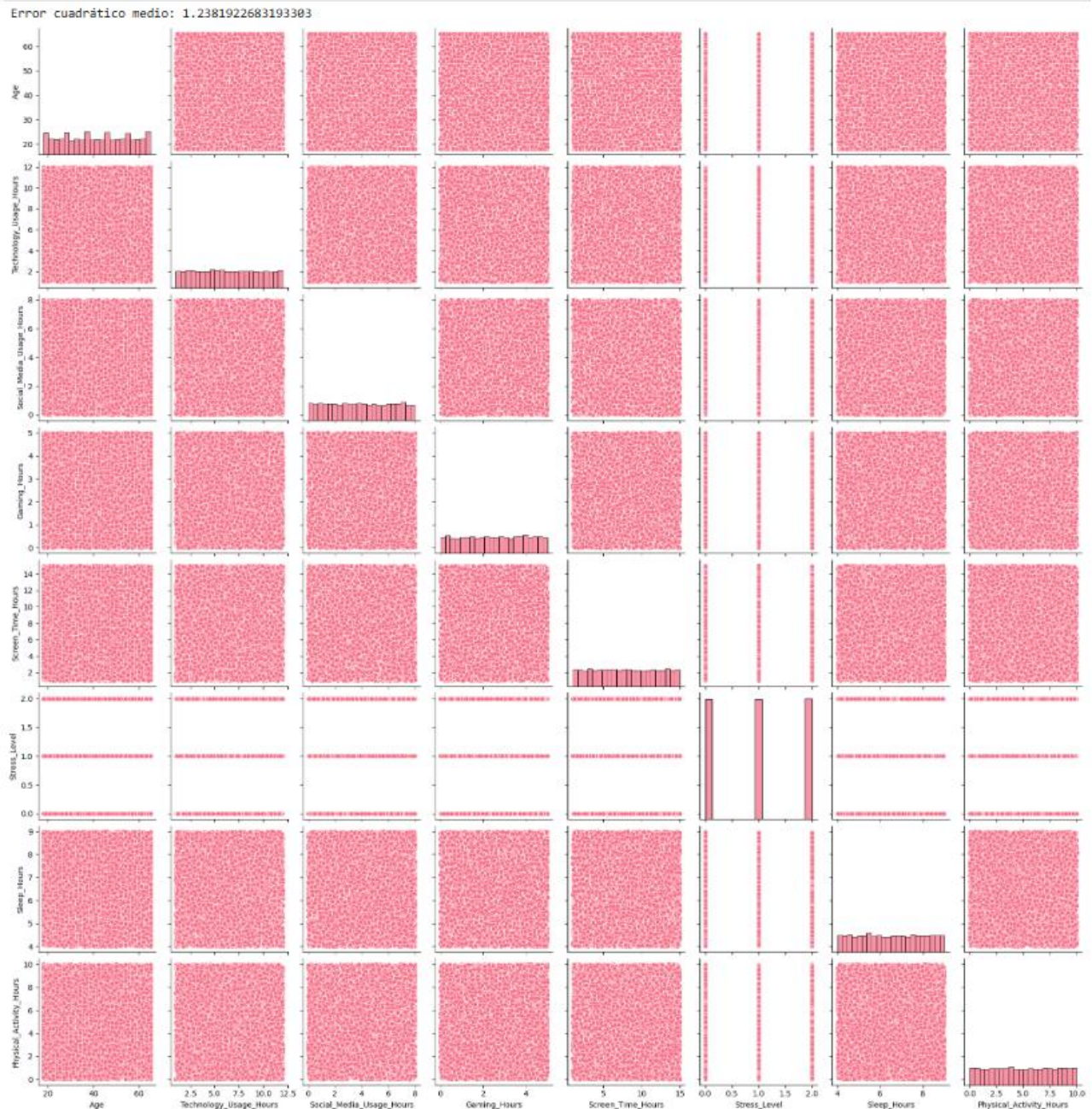
Proyecto de IA – impacto del uso de las redes sociales en la salud mental



El anterior 3 gráficas encontramos lo siguiente:

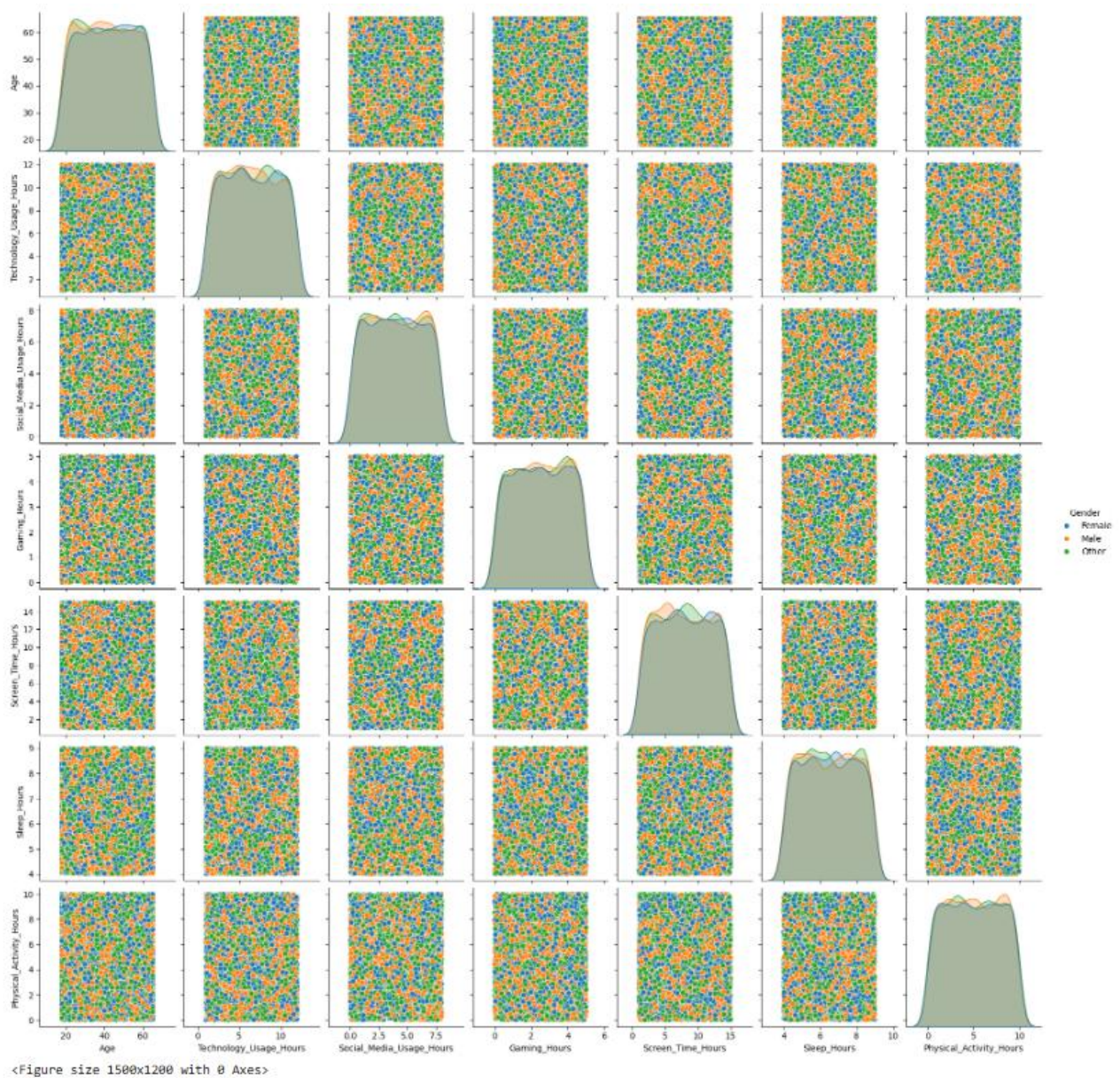
1. En la primera gráfica encontraremos el uso de las redes sociales por un rango de a los rangos son de cero a 12 años de 13 a 18 años de 19 a 30 años de 31 a 50 años y de 51 o más..
2. En la segunda gráfica encontramos una comparativa entre el uso de redes sociales por género en los cuales tenemos especificado el género femenino masculino y otro.
3. Finalmente en la tercera gráfica, tenemos la comparativa por uso entre el rango de Aes y género la cual podemos encontrar con color azul representado el género femenino en color naranja masculino y en color verde el otro.

Proyecto de IA – impacto del uso de las redes sociales en la salud mental



En la anterior gráfica podemos encontrar el modelo de regresión lineal y las gráficas para cada una de las columnas de datos que se encuentran dentro de nuestra base de datos en el archivo .CSV, esta parte será lo que utilizaremos más adelante para nuestro modelo Machine Learning con regresión lineal

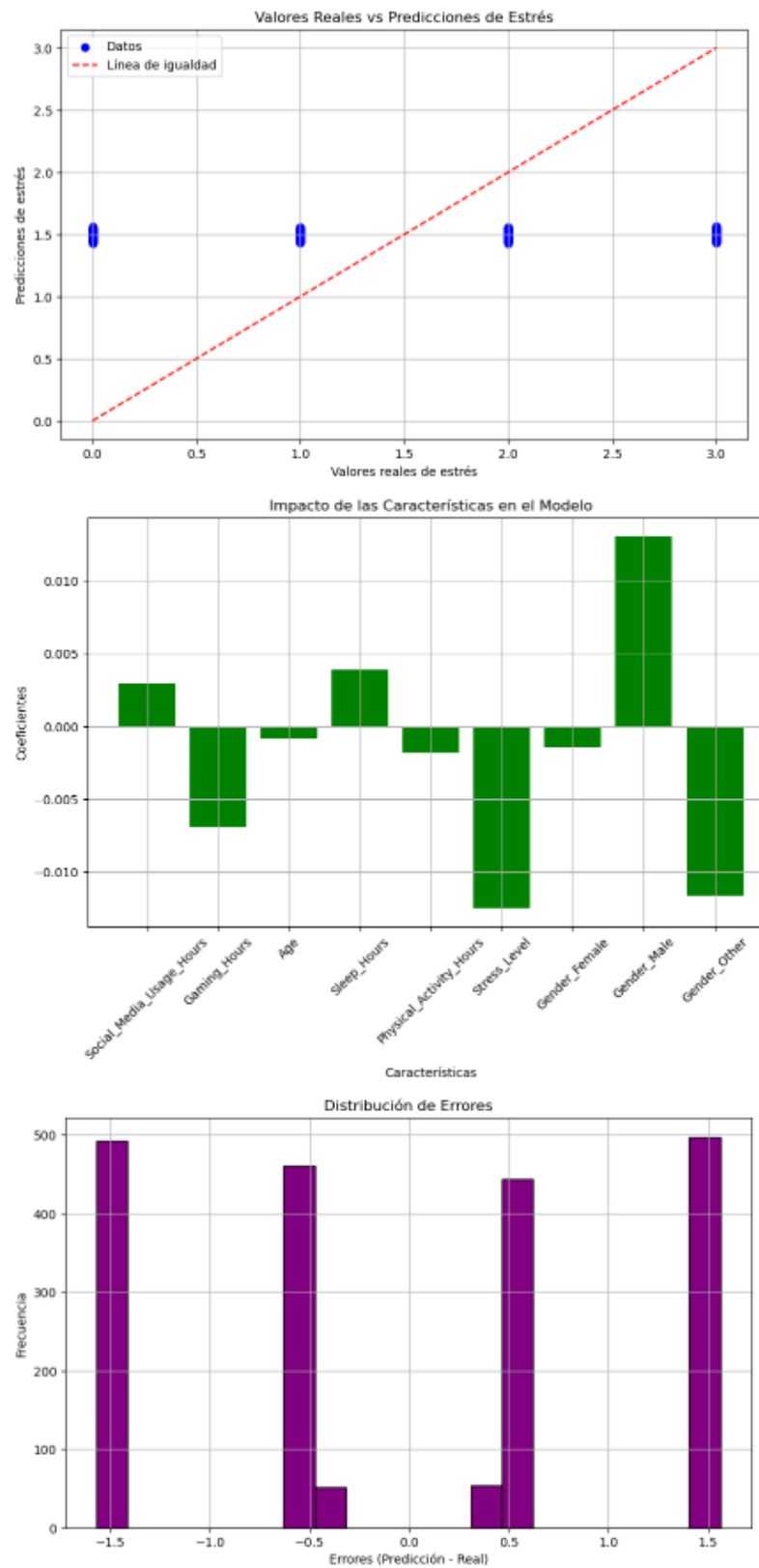
Proyecto de IA – impacto del uso de las redes sociales en la salud mental



El análisis visual revela que existe una relación significativa entre el tiempo dedicado a las redes sociales y los niveles de estrés reportados. Además, se observan diferencias notables en los patrones de sueño entre ambos géneros. Estos hallazgos sugieren la necesidad de investigaciones más profundas para comprender mejor la interacción entre el uso de tecnología y la salud mental.

Proyecto de IA – impacto del uso de las redes sociales en la salud mental

Error Cuadrático Medio: 1.2383986812515838
R-cuadrado: 0.88048539843462159093



Proyecto de IA – impacto del uso de las redes sociales en la salud mental

En las anteriores 3 gráficas podemos evidenciar lo siguiente:

1. En esta primera gráfica podemos encontrar una comparativa entre los valores reales y los valores realizados por nuestro modelo de predicción con regresión lineal. A su vez se encuentra una línea que hace la función de “línea de igualdad”
2. Esta gráfica hace alusión a la representación de El impacto de las características de El modelo para ellos se tienen en cuenta las siguientes columnas de la base de datos, las horas de uso de redes sociales, horas de juego, nivel de estrés, horas de sueño horas de actividad física, nivel de estrés y género.
3. En esta última gráfica se tiene la distribución de errores con la frecuencia y errores pre y predicción - real.

Nota:

Las anteriores graficas pertenecen al resultado del análisis realizado.

Se uso para este el archivo de Jupyter Notebook denominado “Proyecto IA V1,0”

11. Bibliografía.

Para la elaboración de este proyecto se tomó en cuenta Información extraída de distintos sitios web, o instituciones dedicadas al estudio se tomó información de la página del DANE, también se tomó información por parte de la Organización Mundial de la salud, se encontró información en Internet también sobre la salud mental, el impacto de redes sociales y todo lo relacionado con el tema.

Se consultaron distintas bases de datos, aunque no se encontró demasiada información al respecto de forma abierta con buena cantidad de datos, la base de datos que estamos usando, la encontramos en el sitio web de KAGGLE, buscándola como “social media and mental health”.

A continuación, voy a anexar algunos links con información sobre el proyecto información consultada y el link de donde sacamos nuestra data base.

- <https://www.who.int/es/news-room/fact-sheets/detail/adolescent-mental-health>
- <https://www.pewresearch.org/internet/2018/05/31/teens-social-media-technology-2018>
- https://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0185-33252009000200009
- <https://jamanetwork.com/journals/jamapediatrics/article-abstract/2751859>
- <https://www.kaggle.com/datasets/waqi786/mental-health-and-technology-usage-dataset>