

## 1. Analizador de Sentimientos en Reseñas de Productos

**Propósito:** Ayudar a las empresas a identificar si las reseñas de productos son positivas o negativas.

**Objetivo:** Construir un modelo supervisado que clasifique texto utilizando técnicas básicas de procesamiento de lenguaje natural (PLN).

**Relación con el curso:** Aplica limpieza de texto, tokenización, y modelos simples como Naive Bayes o Regresión Logística.

- **Justificación:** Este proyecto es accesible porque el preprocesamiento y los algoritmos son básicos, y los datos textuales son fáciles de obtener.
- **Temas clave:** Limpieza de texto, tokenización, técnicas de PLN, clasificación supervisada.
- **Datos:** Reseñas simuladas de productos disponibles en [Kaggle](#).
- **Salida esperada:**
  - Gráficos de distribución de sentimientos.
  - Un clasificador con métricas como precisión y recall.

## 2. Predicción del Precio de Casas

**Propósito:** Permitir a compradores o vendedores estimar el precio de viviendas según características clave como ubicación, tamaño y antigüedad.

**Objetivo:** Entrenar un modelo de regresión para predecir precios a partir de datos estructurados.

**Relación con el curso:** Refuerza la comprensión de regresión lineal, manejo de valores faltantes y visualización de datos.

- **Justificación:** Este proyecto utiliza datos estructurados, lo que simplifica el preprocesamiento, y permite a los estudiantes trabajar con técnicas ampliamente aplicables.
- **Temas clave:** Preprocesamiento, visualización, regresión lineal.
- **Datos:** [Boston Housing Dataset](#).
- **Salida esperada:**
  - Gráficos que muestren la correlación entre variables.
  - Predicciones con un modelo ajustado y evaluado.

## 3. Clasificación de Manuscritos Digitalizados

**Propósito:** Crear un sistema que reconozca dígitos manuscritos, útil en la digitalización de formularios y documentos.

**Objetivo:** Implementar una red neuronal básica para clasificar imágenes del dataset MNIST.

**Relación con el curso:** Introduce redes neuronales simples con TensorFlow/Keras y el preprocesamiento de imágenes.

- **Simplificación:** Los estudiantes usarán un dataset preprocesado (MNIST) y se enfocarán en ajustar parámetros básicos como el número de épocas.
- **Justificación:** Este proyecto permite una introducción sencilla a redes neuronales sin enfrentar complejidades asociadas a datos reales.
- **Temas clave:** Redes neuronales, entrenamiento y evaluación de modelos.
- **Datos:** [MNIST Dataset](#).
- **Salida esperada:**
  - Precisión del modelo > 90%.
  - Visualización de ejemplos mal clasificados para análisis.

## 4. Predicción de Accidentes de Tránsito

**Propósito:** Identificar patrones y predecir la probabilidad de accidentes según variables como clima, hora y ubicación.

**Objetivo:** Entrenar un modelo de clasificación supervisada que prediga la gravedad de accidentes.

**Relación con el curso:** Refuerza el uso de algoritmos supervisados (e.g., árboles de decisión), manejo de datos categóricos y visualización de patrones.

- **Justificación:** Este proyecto trabaja con datos reales, introduciendo a los estudiantes en la limpieza y análisis de datos complejos, pero sin requerir técnicas avanzadas.
- **Temas clave:** Limpieza de datos, clasificación supervisada.
- **Datos:** [US Accidents Dataset](#).
- **Salida esperada:**
  - Mapas de calor con puntos críticos de accidentes.
  - Modelo predictivo con métricas claras.

## 5. Resumen Automático de Artículos

**Propósito:** Generar resúmenes concisos de textos largos, como artículos o noticias, para ahorrar tiempo al lector.

**Objetivo:** Implementar un sistema de resumen extractivo basado en conteo de palabras clave.

**Relación con el curso:** Introduce técnicas de procesamiento de texto simples como tokenización y conteo de frecuencias.

- **Simplificación:** Se evitarán métricas avanzadas como ROUGE y se empleará un método básico que seleccione oraciones más frecuentes o significativas.
- **Justificación:** Este proyecto es accesible al usar textos preprocesados y métodos intuitivos, permitiendo un aprendizaje significativo en PLN.
- **Temas clave:** Tokenización, procesamiento de texto, resumen extractivo.
- **Datos:** Artículos en inglés de [Kaggle](#).
- **Salida esperada:**
  - Comparación del resumen generado con uno creado manualmente.
  - Evaluación utilizando la métrica ROUGE.

## 6. Predicción de Demanda de Ventas

**Propósito:** Ayudar a las empresas a gestionar inventarios mediante la predicción de ventas futuras basándose en datos históricos.

**Objetivo:** Crear un modelo de regresión para analizar y predecir tendencias de ventas mensuales.

**Relación con el curso:** Refuerza conceptos de regresión, análisis de series temporales y visualización de datos.

- **Justificación:** Este proyecto es adecuado porque los datos estructurados son intuitivos y el modelo puede implementarse usando herramientas básicas como Scikit-learn.
- **Temas clave:** Series temporales, visualización, regresión.
- **Datos:** [Retail Sales Dataset](#).
- **Salida esperada:**
  - Gráficos de tendencias y predicciones.
  - Modelo con métricas de rendimiento como MAE.