



**Universidad de San Carlos de Guatemala**

**Facultad de Ingeniería**

**Inteligencia Artificial “A”**

**Ingeniería en Ciencias y Sistemas**

**Manual Técnico**

**Nombre:** Pablo Josué Barahona Luncney

**Carné:** 202109715

**Fecha de presentación:** 2 de noviembre del 2024

## **Inicialización de Eventos y Variables**

En esta sección, se inicializan los eventos que permiten al usuario interactuar con la aplicación. Se configuran oyentes para la carga de archivos CSV y el entrenamiento de los modelos, asegurando que la aplicación responda adecuadamente a las acciones del usuario.

## **Manejo de Archivos CSV**

El manejo de archivos CSV es crucial para la entrada de datos en la aplicación. Al seleccionar un archivo, se utiliza un FileReader para leer su contenido. Una vez leído, se procesa el texto CSV, que se divide en filas y columnas, facilitando así el uso de los datos en los modelos de machine learning. Este procesamiento incluye la eliminación de espacios en blanco y la conversión de datos según sea necesario.

## **Selección de Columnas y Configuración**

Los usuarios pueden seleccionar qué columnas utilizarán como variables independientes y dependientes en sus modelos. Esto permite personalizar el entrenamiento de los modelos de acuerdo con los datos disponibles en los archivos CSV. Las funciones de configuración permiten actualizar las selecciones dinámicamente en función de las columnas disponibles en los datos.

## **Entrenamiento de Modelos**

La sección de entrenamiento es donde se implementan y ajustan los distintos modelos de machine learning. A continuación, se describen brevemente cada uno de los modelos utilizados:

### **Regresión Lineal**

Este modelo se utiliza para predecir un valor continuo. La regresión lineal encuentra la mejor línea que se ajusta a los datos, minimizando la distancia entre los puntos de datos y la línea de ajuste. Utiliza el método de mínimos cuadrados para calcular los coeficientes que definen la línea.

## **Regresión Polinómica**

La regresión polinómica se usa cuando la relación entre las variables es no lineal. En lugar de ajustar una línea recta, este modelo ajusta un polinomio a los datos, permitiendo capturar relaciones más complejas.

## **Árbol de Decisión**

Los árboles de decisión son modelos que dividen los datos en ramas para tomar decisiones. Cada nodo del árbol representa una prueba sobre una característica, y cada rama representa el resultado de esa prueba. Se entrenan utilizando técnicas como ID3 o CART, y son intuitivos y fáciles de interpretar.

## **Naive Bayes**

Este es un modelo de clasificación que se basa en el teorema de Bayes y en la suposición de independencia entre las características. Se utiliza comúnmente para clasificación de texto y es eficiente en términos de computación. La implementación incluye el cálculo de probabilidades para predecir la clase a la que pertenece un nuevo dato.

## **Red Neuronal**

Las redes neuronales son modelos inspirados en el funcionamiento del cerebro humano. Consisten en capas de neuronas (nodos) interconectadas, donde cada conexión tiene un peso. A través del entrenamiento, estos pesos se ajustan para minimizar el error en las predicciones. Las redes pueden ser profundas (DNN) o simples (perceptrón), dependiendo de la cantidad de capas.

## **Algoritmo K-means**

K-means es un algoritmo de clustering que agrupa los datos en k grupos basándose en la similitud. Funciona iterando sobre los puntos de datos y ajustando los centroides de los grupos hasta que se alcanza la convergencia. Es útil para identificar patrones y segmentar datos.

## **K-Nearest Neighbors (KNN)**

KNN es un método de clasificación basado en la distancia. Para clasificar un nuevo punto, el algoritmo busca los  $k$  vecinos más cercanos en el conjunto de datos y asigna la clase más común entre esos vecinos. Este método es simple pero efectivo para muchas aplicaciones de clasificación.

## **Visualización de Resultados**

La visualización es esencial para interpretar los resultados de los modelos de machine learning. Cada modelo tiene su propia función de visualización que representa gráficamente los datos y las predicciones, facilitando la comprensión de la precisión y la eficacia del modelo. Se utilizan bibliotecas como Chart.js para crear gráficos dinámicos que muestran las tendencias y patrones en los datos.

## **Cálculo de Métrica $R^2$**

La métrica  $R^2$  (coeficiente de determinación) es una medida estadística que se utiliza para evaluar qué tan bien se ajusta un modelo de regresión a los datos.  $R^2$  indica la proporción de la varianza en la variable dependiente que es predecible a partir de las variables independientes. Un valor de  $R^2$  más cercano a 1 indica un mejor ajuste.