

Tarea 02: Aplicación de Machine Learning en el Conjunto de Datos Iris

Objetivo:

El propósito de esta tarea es realizar un análisis y aplicación de Machine Learning en el conjunto de datos Iris, utilizando el algoritmo de k-Nearest Neighbors (KNN). Se espera que el participante descargue el conjunto de datos desde Kaggle, realice una exploración detallada de las características y, posteriormente, implemente y evalúe el modelo de Machine Learning.

Pasos a seguir:

Descarga del Conjunto de Datos:

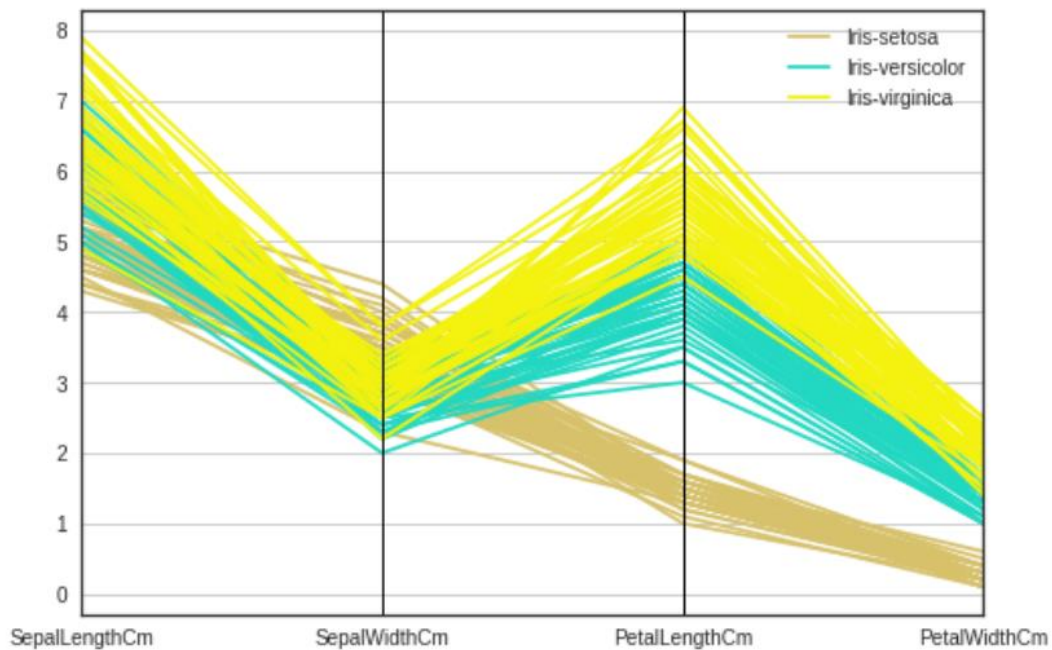
Descargar el conjunto de datos desde Kaggle mediante el siguiente enlace <https://www.kaggle.com/datasets/arshid/iris-flower-dataset>

Lectura del Dataset con Pandas:

Utilizar la biblioteca Pandas para leer el conjunto de datos desde el archivo CSV descargado. Explorar las filas del conjunto de datos y obtener información básica sobre las columnas.

Análisis de Características:

Realizar un análisis exploratorio de las características (columnas) del conjunto de datos. Utilizar visualizaciones como gráficos de dispersión, histogramas, o boxplots para identificar diferencias entre las distintas clases de flores Iris.



Implementación de KNN:

Implementar el algoritmo de k-Nearest Neighbors (KNN) visto en clase. Nota aclaratoria: No se permite utilizar bibliotecas para simplificar el algoritmo.

Seleccionar los parámetros apropiados para el algoritmo. Experimente con diferentes valores de N vecinos y otros parámetros según sea necesario (agréguelos a su informe).

Evaluación del Modelo:

Utilizar el 80% del conjunto de datos como conjunto de entrenamiento para realizar el aprendizaje del modelo y el 20% restante como conjunto de prueba.

Evaluar el modelo utilizando la métrica de precisión (Accuracy) y generar una matriz de confusión para analizar la clasificación de los ejemplos en el conjunto de datos de prueba.

Entrega:

Presentar los resultados obtenidos en un informe de Latex siguiendo la plantilla para artículos científicos de IEEE (usar Overleaf como recomendación) que incluya visualizaciones, snippets códigos utilizados (no es necesario incluir todo, solamente lo necesario) y el análisis de los hallazgos, coloque la tabla de rúbrica que se encuentra en este documento en su informe.

Adjuntar el código implementado en un archivo separado para su revisión. El código debe ser presentado en un Jupyter notebook.

Coloque un zip en el espacio del TecDigital, que contenga el código fuente del Latex de su informe, el propio PDF del informe y el código fuente de su trabajo.

Fecha de Entrega: 06-03-2024

Rúbrica:

Criterios	Puntuación máxima	Puntuación obtenida
Descarga y lectura del dataset	10	
Análisis de Características	20	
Implementación de KNN y experimentación de parámetros	30	
Evaluación del modelo con las métricas recomendadas	15	
Presentación de los resultados	15	
Estructura y claridad del informe	10	