# Lab Report: K-Nearest Neighbors from Scratch

**By Aniruddha Sahay Varma 23/CS/054**

**Aim:** To implement K-Nearest Neighbours algorithm from scratch using Numpy and Pandas

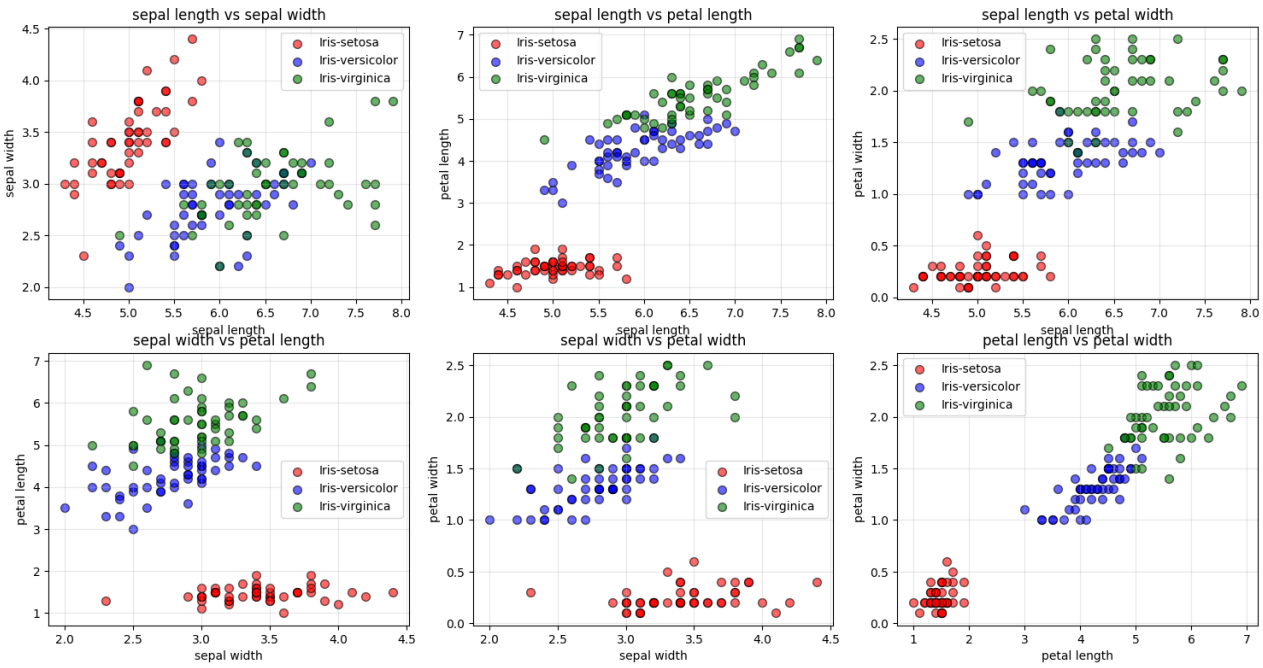
**Theory:**

K-Nearest Neighbors (KNN) is a non-parametric, instance-based supervised machine learning algorithm employed for both classification and regression tasks. Its core mechanism operates on the principle of proximity, predicting the class or value of a new data point by examining its 'K' closest neighbors within the training dataset, as determined by a specified distance metric (e.g., Euclidean). For classification, the new point is assigned to the class most frequently occurring among its K neighbors (majority vote); for regression, its value is typically predicted as the average or median of the values of its K neighbors. As a "lazy learner," KNN defers computation until prediction time, storing the entire training dataset rather than constructing an explicit model during a distinct training phase.

**Observations:**

**Datasets Used**: Iris, Wine

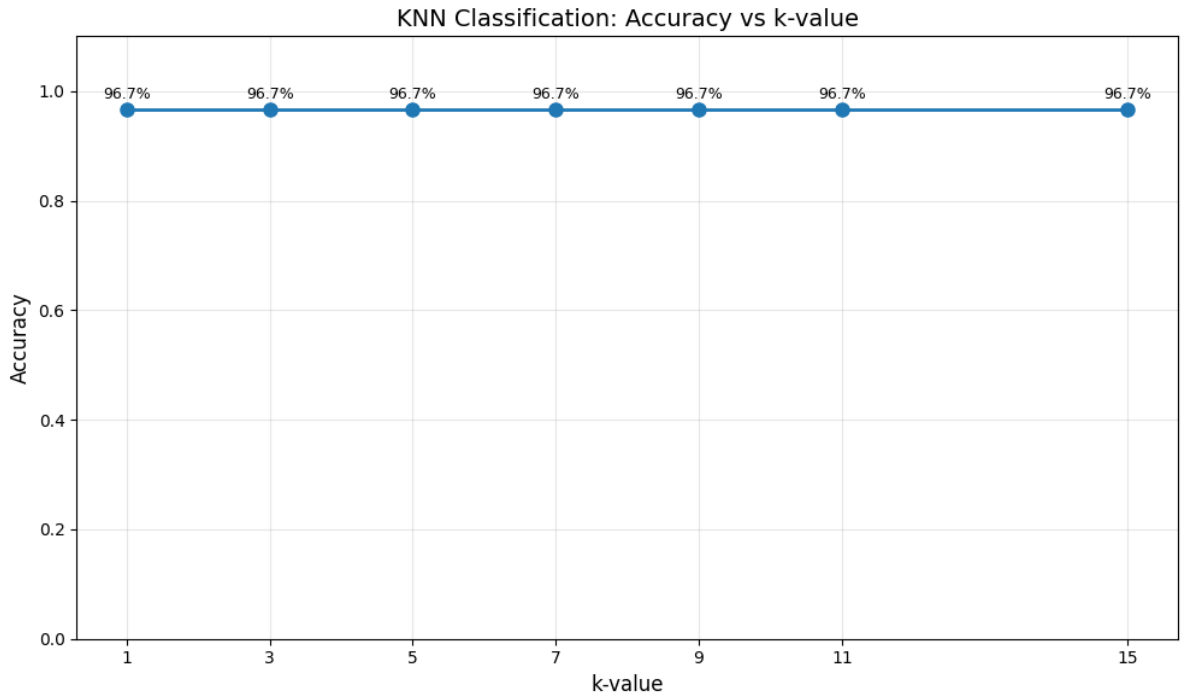
EDA for Iris Dataset:



**Analysis:**

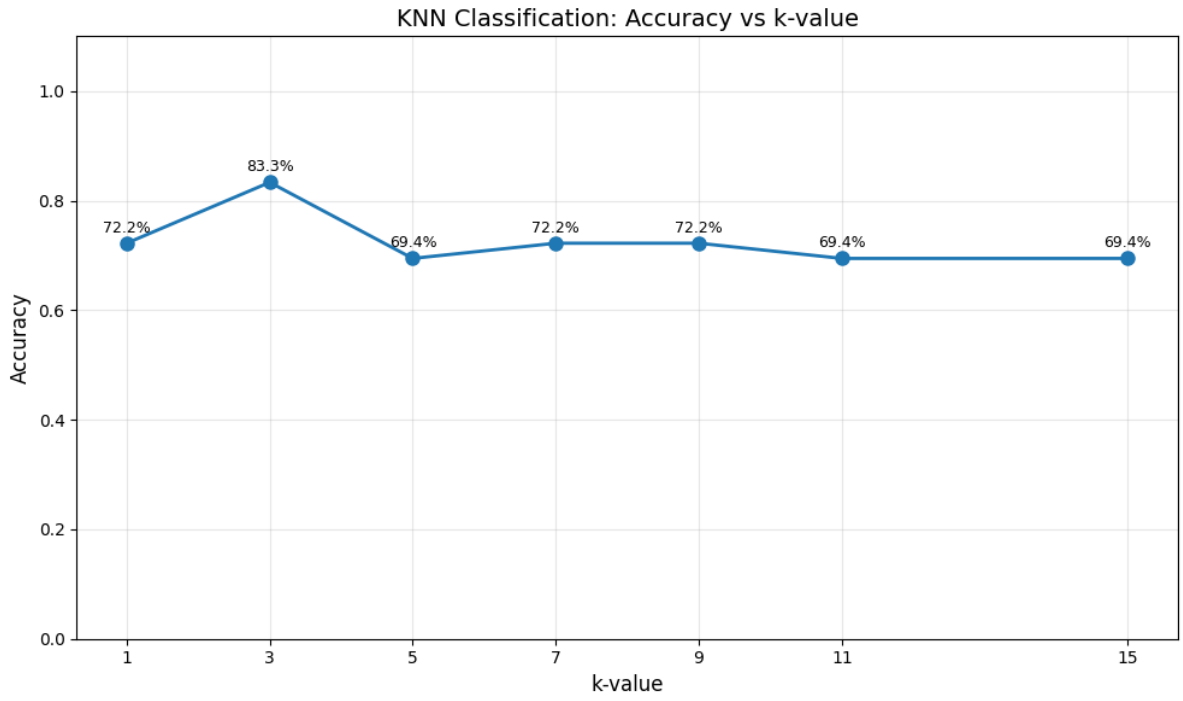
- feature pair Petal length vs Petal Width provides the best and cleanest separation between the three classes.
- The setosa class is inherently easier to distinguish, as it forms a distinct, linearly separable cluster in most of the plots, especially those involving petal dimensions.

**Performance on IRIS dataset**




For all K values from 1-15 the model performs the same due to the inherent ease in the problem in the dataset

**Performance on Wine dataset**




**Best Performing K value: 3**

**Conclusion**: Knn algorithm was successfully implemented using numpy and pandas, K values vs accuracy was plotted and best values were found for IRIS and Wine dataset