

Assignment-2

M.Sushanth-20CS10034

S.Bhanu Teja-20CS10059

1. Unsupervised Learning

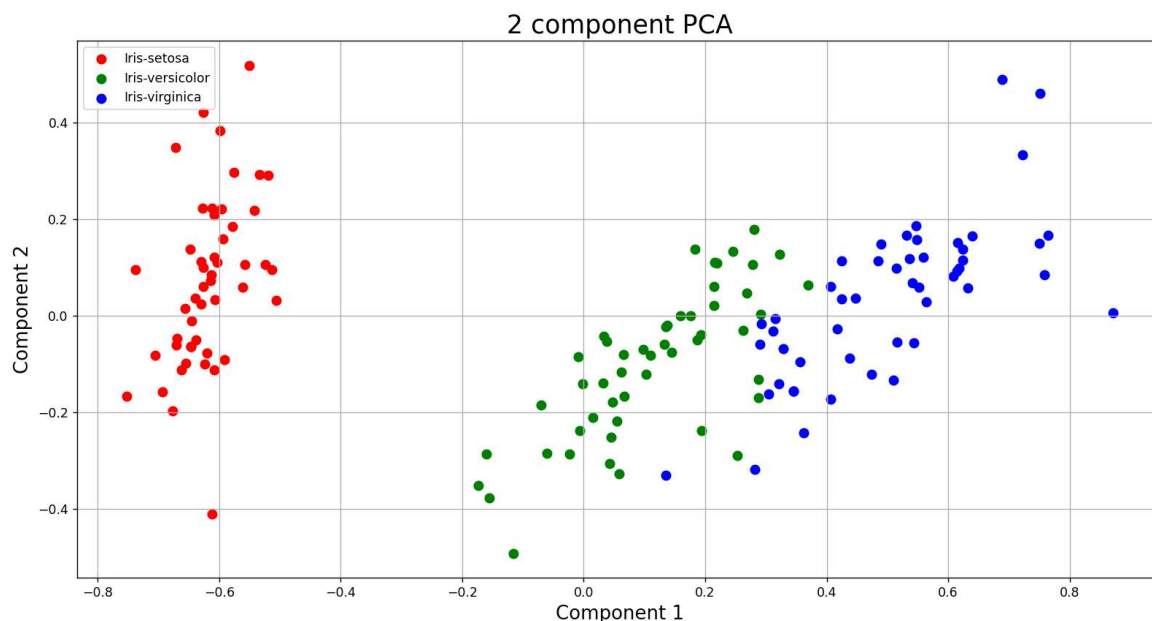
PCA:

Principal Component Analysis is an unsupervised learning algorithm that is used for dimensionality reduction in machine learning. It is a statistical process that converts the observations of correlated features into a set of linearly uncorrelated features with the help of orthogonal transformation. These new transformed features are called the Principal Components. It is one of the popular tools that is used for exploratory data analysis and predictive modelling. It is a technique to draw strong patterns from the given dataset by reducing the variances.

Total Variance with two components: **0.9587437487644829**

95% variance was preserved so number of components = 2.

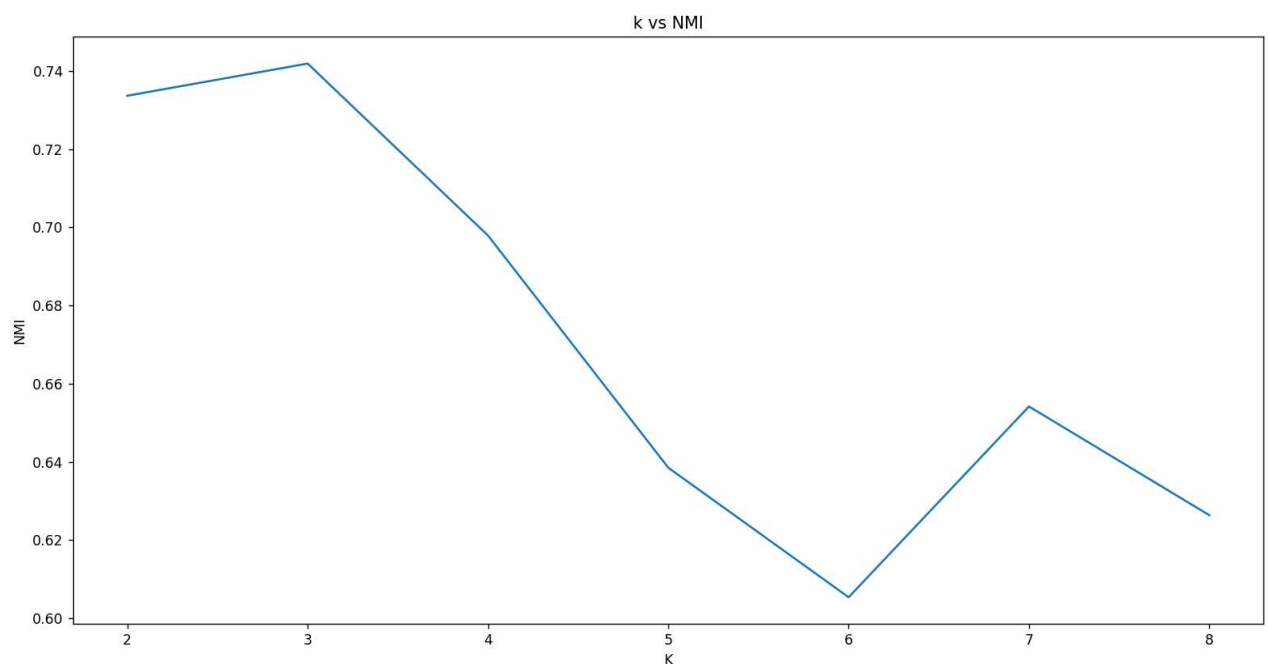
The Graph for PCA



K-Means Clustering:

K-means clustering is a simple unsupervised learning algorithm that is used to solve clustering problems. It follows a simple procedure of classifying a given data set into a number of clusters, defined by the letter "k," which is fixed beforehand. The clusters are then positioned as points and all observations or data points are associated with the nearest cluster, computed, adjusted and then the process starts overusing the new adjustments until a desired result is reached.

The graph of K vs normalised mutual information (NMI).



The value of K for which the NMI is maximum : **3**