

Iris Dataset- Clustering using K means

Abstract

The Iris dataset contains the data for 50 flowers from each of the 3 species - Setosa, Versicolor and Virginica. The data gives the measurements in centimeters of the variables sepal length and width and petal length and width for each of the flowers.

Goal of the study is to perform exploratory analysis on the data and build a K-means clustering model to cluster them into groups. Here we have assumed we do not have the species column to form clusters and then used it to check our model performance.