**Nearest Neighbors**

**Describing the Nearest Neighbors method:**

Nearest neighbors is an supervised machine learning and classification algorithm. The guideline behind closest neighbor is to discover a predefined number of preparing tests nearest in distance to the new point. It figures the distance of new data point to all other training data points. The quantity of tests can be user defined. The distance can be any measurement, but the method Euclidean distance is the most widely recognized utilized. It very well may be summed up in the accompanying manner: -

* Load the data information.
* Measure the euclidean distance from the new data to all other data that has been classified.
* Get the K smaller distances.
* Check for the number of classes that had smaller distance and appeared the most.
* Classify the new data with the correct class.

**Explaining what the criteria was for selecting the Three attributes:**

The measures utilized for choosing the Three attributes was **correlation using Pearson**. Connection alludes to the mutual relationship or relationship between two amounts. It figures pairwise relationship between columns excluding NA values. The connection of each quality with the objective is taken and the characteristics having most noteworthy qualities are taken. The attributes that had the most values were picked

Correlation matrix with heatmap shows all attributes. There is a positive correlation (increase in one value of three features increasing with the target variable).

Chart

Description automatically generated

From the figure we can observe that the observed three attributes pclass, fare, sex have better correlation with target variable(survived)

**Training Accuracy vs Test Accuracy:**

**Chart, line chart

Description automatically generated**

From above graph it is clear that the best k value is 5 followed by 13 when compared or checked to accuracies.

1. **Interpreting and comparing the results.**

The dataset used here is Titanic dataset where we have taken first 3 attributes as **pclass, fare, sex**. We can use confusion matrix to evaluate the accuracy of the classification with different values of K.

**Confusion Matrix for k=5:**

The Accuracy is: 74.6268656716418

**Confusion matrix:**

Chart, treemap chart

Description automatically generated

Table

Description automatically generated

**Confusion Matrix for k=1:**

The Accuracy is: 77.23880597014924

**Confusion matrix:**

**Chart, treemap chart

Description automatically generated**

Table

Description automatically generated

**Confusion Matrix for k=11:**

The Accuracy is: 72.76119402985076

**Confusion matrix:**

**Chart

Description automatically generated**

Table

Description automatically generated

Sources

<https://www.datacamp.com/community/tutorials/k-nearest-neighbor-classification-scikit-learn>

<https://scikit-learn.org/stable/modules/neighbors.html>

<https://machinelearningmastery.com/k-nearest-neighbors-for-machine-learning/>