

Sri Sivasubramaniya Nadar College of Engineering, Kalavakkam – 603 110
(An Autonomous Institution, Affiliated to Anna University, Chennai)

UCS2612 Machine Learning Laboratory

Academic Year: 2023-2024 Even

Faculty In-charges: Y.V. Lokeswari & Nilu R Salim

Batch: 2021-2025

VI Semester A & B

A. No. : 5 . **k-Nearest Neighbor algorithm**

Download the Online Shoppers Purchasing Intention Dataset dataset from the link given below:

<https://archive.ics.uci.edu/dataset/468/online+shoppers+purchasing+intention+dataset>

The dataset consists of 12,330 sessions, 84.5% (10,422) were negative class samples that did not end with shopping, and the rest (1908) were positive class samples ending with shopping.

Develop a python program to predict the Online Shoppers Purchasing Intention using K-Nearest Neighbour algorithm. Visualize the features from the dataset and interpret the results obtained by the model using Matplotlib library. **[CO1, K3]**

Use the following steps to do implementation:

1. Loading the dataset.
2. Pre-Processing the data (Handling missing values, Encoding, Normalization, Standardization).
3. Exploratory Data Analysis.
4. Feature Engineering techniques.
5. Split the data into training, testing and validation sets.
6. Provide test data.
7. Measure the performance of the model.
8. Represent the results in terms of ROC curves using graphs.

.....

Upload the code in GitHub and include the GitHub main branch link in the assignment PDF.

Hints to do the assignment:

Do the following:

1. Load the dataset.
2. Pre-Processing the data (Handling missing values, Encoding, Normalization, and Standardization).
3. Exploratory Data Analysis
4. Feature Engineering techniques.

Refer to

<https://machinelearningmastery.com/feature-selection-machine-learning-python/>
<https://www.analyticsvidhya.com/blog/2020/10/feature-selection-techniques-in-machine-learning/>
<https://www.datacamp.com/tutorial/feature-selection-python>

5. Split the data into training, testing and validation sets.
6. Apply K-Nearest Neighbor Classification algorithm on the test data and predict the class of test data.

Use Euclidean distance, Manhattan distance and Minkowski distance as distance metrics and compare the results.

<https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html>
<https://towardsdatascience.com/knn-using-scikit-learn-c6bed765be75>
<https://www.kdnuggets.com/2022/07/knearest-neighbors-scikitlearn.html>
<https://www.geeksforgeeks.org/ml-implementation-of-knn-classifier-using-sklearn/>
<https://www.datacamp.com/tutorial/k-nearest-neighbor-classification-scikit-learn>
<https://machinelearninggeek.com/knn-classification-using-scikit-learn/>

7. Implement K-Nearest Neighbor algorithm with user-defined functions and Euclidean distance metric.
8. Upload python project in GitHub and explore all git commands. Git Commands Tutorial : <https://git-scm.com/docs/gittutorial>

Upload IPython to GitHub

<https://reproducible-science-curriculum.github.io/sharing-RR-Jupyter/01-sharing-github/>

Additional Reference:

<https://www.youtube.com/watch?v=LlrKTV4-ftI>

.....

Upload the code in GitHub and include the GitHub main branch link in the assignment PDF.

Upload python project in GitHub and explore all git commands. Git Commands Tutorial : <https://git-scm.com/docs/gittutorial>