MVP Samaj's
K.T.H.M. College Nasik
Department of Statistics
MSc.(Statistics)-II (sem_III)
STS-615-MJP: Practical based on Machine Learning

INDEX

STS-615-MJP: Practical based on Machine Learning

Title : 1] Supervised and unsupervised learning method.

1. Use the Iris dataset to classify flowers into species (setosa, Versicolor, or Virginia) based on petal and sepal measurements. [Load the Iris dataset using sklearn.datasets]
2. Predict car fuel efficiency (mpg) using features such as horsepower, weight, and displacement from the Auto MPG dataset.

   Dataset: Use the Auto MPG dataset from seaborn.
3. Perform customer segmentation by grouping customers into clusters based on their purchasing behavior using the Wholesale Customers dataset

   .Dataset: Use the Wholesale Customers dataset from UCI Machine Learning Repository (download manually or load using a CSV).
4. Reduce the dimensionality of the Wine Quality dataset and visualize it in 2D.

   Dataset: Use the Wine Quality dataset from UCI Machine Learning Repository (download manually or load using a CSV)

STS-615-MJP: Practical based on Machine Learning

Title : 2] Feature selection and extraction

1] Feature Selection Using Random Forest: Wine Quality Prediction

Problem: Identify the most important features for predicting wine quality using the Wine Quality dataset.

Dataset: Wine Quality Dataset

archive.ics.uci.edu/ml/machine-learning-databases/wine-quality/winequality-red.csv

2] Feature Extraction with PCA: Breast Cancer Classification

Problem: Reduce the dimensionality of the Breast Cancer dataset using PCA and build a classifier.

Dataset: Available from sklearn.datasets

3] Dimensionality Reduction with PCA: Customer Segmentation

Problem: Use PCA to reduce the dimensionality of the Wholesale Customers dataset and perform customer segmentation with K-Means clustering.

Dataset: Wholesale Customers Dataset

archive.ics.uci.edu/ml/machine-learning-databases/00292/Wholesale customers data.csv

STS-615-MJP: Practical based on Machine Learning

Title : 3] Classification of data using k-Nearest Neighbor (k-NN).

1] Classifying Handwritten Digits Using k-NN

Dataset: Available from sklearn.datasets (load_digits function)

Problem: Classify handwritten digits (0-9) using k-NN.

2] Classifying Diabetes Diagnosis Using k-NN

Dataset: Pima Indians Diabetes Dataset

Problem: Predict whether a patient has diabetes based on medical examination data

https://raw.githubusercontent.com/jbrownlee/Datasets/master/pima-indians-diabetes.data.csv

3] Classifying Banknote Authentication Using k-NN

Dataset: Banknote Authentication dataset

Problem: Classify whether a banknote is authentic or fake based on image processing measures.

archive.ics.uci.edu/ml/machine-learning-databases/00267/data_banknote_authentication.txt

STS-615-MJP: Practical based on Machine Learning

Title :4] Classification of data using Naïve Bayes classifier.

1] Classifying Breast Cancer Using Naïve Bayes

Dataset: Breast Cancer Wisconsin (Diagnostic) Dataset

Problem: Predict whether a tumor is malignant or benign based on features computed from digitized images.

https://archive.ics.uci.edu/dataset/17/breast+cancer+wisconsin+diagnostic

2] Classifying Titanic Survival Using Naïve Bayes

Dataset: Titanic Dataset

Problem: Predict whether a passenger survived the Titanic disaster based on features like age, sex, and class.

https://raw.githubusercontent.com/datasciencedojo/datasets/master/titanic.csv

3] Classifying Diabetes Diagnosis Using Naïve Bayes

Dataset: Pima Indians Diabetes Dataset

Problem: Predict whether a patient has diabetes based on medical examination data.

https://raw.githubusercontent.com/jbrownlee/Datasets/master/pima-indians-diabetes.data.csv

STS-615-MJP: Practical based on Machine Learning

Title : 5] Support Vector Machine.

1] **Classifying Handwritten Digits Using SVM**

Dataset: Available from sklearn.datasets (load_digits function)

Problem: Classify handwritten digits (0-9) using SVM.

2] **Classifying Iris Flower Species Using SVM**

Dataset: Iris dataset

Problem: Classify iris flowers into three species (Setosa, Versicolor, Virginica) based on features such as sepal length, sepal width, petal length, and petal width.

archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data

3] **Classifying Breast Cancer Using SVM**

Dataset: Breast Cancer Wisconsin (Diagnostic) Dataset

Problem: Predict whether a tumor is malignant or benign based on features derived from a digitized image of a breast mass.

https://archive.ics.uci.edu/dataset/17/breast+cancer+wisconsin+diagnostic

STS-615-MJP: Practical based on Machine Learning

Title : 6] Practical based on Artificial Neural Network

1] Predicting House Prices Using ANN

Dataset: California Housing Prices Dataset

Question: How to predict house prices based on features like median income, house age, and latitude/longitude?

https://raw.githubusercontent.com/ageron/handson-ml/master/datasets/housing/housing.csv

2] Classifying Fashion Items Using ANN

Dataset: Fashion MNIST Dataset

Question: How to classify images of fashion items (e.g., shirts, shoes, etc.)?

https://github.com/zalandoresearch/fashion-mnist

3] Predicting Diabetes Using ANN

Dataset: Pima Indians Diabetes Dataset

Question: How to predict whether a person has diabetes based on health metrics?

https://raw.githubusercontent.com/jbrownlee/Datasets/master/pima-indians-diabetes.data.csv

STS-615-MJP: Practical based on Machine Learning

Title : 7] Computation of impurity measure and construction of Decision Tree.

1] Predicting Titanic Survival Using Decision Trees

Dataset: Titanic Dataset

Question: How can a decision tree be used to predict the survival of passengers based on features like age, sex, and class?

2] Predicting Heart Disease Using Decision Trees

Dataset: Heart Disease UCI Dataset

Question: Can decision trees predict whether a patient is likely to have heart disease based on medical data?

https://archive.ics.uci.edu/ml/machine-learning-databases/heart-disease/processed.cleveland.data

3] Predicting Diabetes Using Decision Trees

Dataset: Pima Indians Diabetes Dataset

Question: Can decision trees predict diabetes based on health measurements such as glucose, BMI, and age?

https://raw.githubusercontent.com/jbrownlee/Datasets/master/pima-indians-diabetes.data.csv

STS-615-MJP: Practical based on Machine Learning

Title : 8] Agglomerative hierarchical clustering.

1] Clustering Cancer Data Using Agglomerative Clustering

Dataset: Breast Cancer Wisconsin Dataset

Question: How can Agglomerative Clustering group tumor samples based on cell characteristics?

https://archive.ics.uci.edu/dataset/17/breast+cancer+wisconsin+diagnostic

2] Clustering Wine Varieties Using Agglomerative Clustering

Dataset: Wine Dataset

Question: How can wine varieties be clustered based on physicochemical properties?

https://archive.ics.uci.edu/dataset/109/wine

3] Clustering Iris Species Using Agglomerative Clustering

Dataset: Iris Dataset

Question: How can Agglomerative Clustering cluster different species of iris based on petal and sepal measurements?

https://archive.ics.uci.edu/ml/datasets/iris

M. V. P. Samaj's
K. T. H. M. College, Nashik

Department of Statistics

As per NEP Syllabus

M.Sc. II **Date:**

Paper: STS-615-MJP **Practical No: 09**

Title of Practical: Practical based on Association Rule.

Q.1) For the following transaction data set, generate rules using apriori algorithm. Consider support=50% and confidence=75%

Transaction ID	Items Bought
1	Milk, Bread, Butter
2	Milk, Bread
3	Milk, Butter
4	Bread, Butter
5	Milk, Bread, Butter, Cheese
6	Milk, Cheese

Q.2) For the following transaction data set, generate rules using apriori algorithm. Consider support=50% and confidence=60%

Transaction ID	Items Bought
1	Phone, Headphones, Charger
2	Phone, Case
3	Phone, Headphones
4	Laptop, Mouse
5	Laptop, Charger, Headphones
6	Phone, Charger, Case