

## **Assignment 1: Introduction to machine learning**

CO1: To understand the need of machine learning for various problem solving.

Short questions (3-4 marks)

1. Define Machine Learning. Also, compare it with Human Learning.
2. Give the difference between supervised learning and unsupervised learning.
3. Explain the concept of penalty and rewards in reinforcement learning.
4. Write a short note on machine learning in finance and banking.
5. Identify main three features for following well-posed problem:
  - a. Fruit prediction Problem
  - b. Handwriting recognition Problem
6. Compare the different types of machine learning.
7. Explain types of machine learning in detail.
8. Define machine learning. Explain any two business applications of Machine Learning in detail.
9. Explain different tools and technology used in Machine Learning.
10. Describe basic concept of Machine Learning and its application.

## **Assignment 2: Preparing to Model**

CO2: Prepare machine learning model and learning the evaluation methods.

1. Give the difference between qualitative data and quantitative data.
2. Give the difference between histogram and box plot with example.
3. Find mean, median, mode and standard deviation for the following data:  
1, 1, 2, 4, 5, 5, 6, 7, 7, 7, 8, 9, 10
4. Define outliers. How can we take care of outliers in data?
5. State various strategies to handle missing values.
6. Explain PCA in brief.
7. Describe machine learning activities in detail.
8. Write a short note on feature subset selection.
9. Write a short note on dimensionality reduction.
10. Define following terms: Data pre-processing, Data remediation, Outliers, Imputation, Standard Deviation, Ratio Data, Ordinal Data

### Assignment 3: Modeling and Evaluation

CO2: Prepare machine learning model and learning the evaluation methods.

1. Define model. How can you train a model?
2. Give the difference between predictive model and descriptive model.
3. Define overfitting and underfitting. When does it happen?
4. Write a short note on bias-variance trade-off in context of model fitting.
5. State various ways to improve performance of a model.
6. Explain Holdout method in detail.
7. Describe k-fold cross validation in detail.
8. Explain bagging, boosting and stacking in detail.
9. Describe steps for model parameter tuning.
10. Consider the following confusion matrix of the win/loss prediction of cricket match. calculate the accuracy, error rate, sensitivity, specificity, precision, recall and F-measure of the model.

	Actual Win	Actual Loss
Predicted Win	82	7
Predicted Loss	3	8

## **Assignment 4: Supervised Learning - Classification and Regression**

CO3: Evaluate various supervised learning algorithms using appropriate dataset

Short questions(3-4 marks)

1. Explain classification model in brief.
2. Give the difference between classification and regression.
3. Draw the flowchart which shows the classification learning process.
4. Write and discuss k-NN Algorithm with advantages and disadvantages.
5. Discuss the SVM model in detail with its pros, cons and applications.
6. Explain logistic regression with advantage and disadvantage.
7. Write a short note on Single Linear Regression. Also, state applications of it.
8. Write a short note on Multiple Linear Regression. Also, state applications of it.
9. Explain any three applications of classification in detail.

## Assignment 5: Unsupervised Learning

CO4: Evaluate various unsupervised learning algorithms using appropriate dataset

1. Give the difference between supervised learning and unsupervised learning
2. Differentiate clustering with classification.
3. Define: Support, Confidence
4. State apriori property.
5. Write strength, weakness and applications of apriori clustering algorithm.
6. How unsupervised learning is useful in fraud detection?
7. Write and explain applications of unsupervised learning.
8. Write and explain apriori algorithm in detail.
9. Write and explain k-means clustering approach in detail.
10. You are given a set of one-dimensional data points: (5, 10, 15, 20, 25, 30, 35). Assume that  $k=2$  and first set of random centroid is selected as (15, 32) and then it is refined with (12, 30). Create two clusters with each set of centroid mentioned above following the k-means approach. 5. Generate frequent itemsets and generate association rules based on it using apriori algorithm. Minimum support is 50% and minimum confidence is 70%.

TID	Items
100	1,3,4
200	2,3,5
300	1,2,3,5
400	2,5

## **Assignment 6: Python libraries for Machine learning**

CO5: To understand the use of various existing machine learning libraries

Short questions (3-4 marks)

1. How to load dataset using Numpy? Explain.
2. How to load dataset using Panda? Explain.
3. How to plot a vertical line and a horizontal line using Matplotlib?
4. Explain features and applications of Pandas.
5. Explain features and applications of Numpy.
6. Explain features and applications of Matplotlib.
7. Explain features and applications of Scikit-Learn.
8. Write a Python program to load the iris data from a given csv file into a dataframe and print the shape of the data, type of the data and first 3 rows using Scikit-Learn.