**Lab Week 5**

**Supervised Learning**

**Part A: Classification**

Topic 1: Decision Tree

1. Import libraries such as numpy and pandas.
2. Load file: “Job dataset.csv”.
3. View the data frame.
4. Set input and target value.
5. View the input and target value.
6. Transform text input into numerical input:
   1. Import LabelEncoder library: from sklearn.preprocessing import LabelEncoder
   2. For each input column, create a transformation model using LabelEncoder.
   3. For each input column, create new columns with only numerical values.
   4. Remove the text columns.
7. Import Decision Tree library: from sklearn import tree
8. Create a decision tree model.
9. Fit the model with input and target values.
10. Check the accuracy score of the model.
11. What is the prediction based on the following input:
    1. Google, Sales Executives, Master
    2. Facebook, Programmer, Degree
    3. ABC Pharma, Business Manager, Master

Topic 2: K Nearest Neighbor

1. Import libraries:
   1. numpy
   2. pandas
   3. seaborn
   4. matplotlib.pyplot
   5. iris dataset: load\_iris from sklearn.datasets
2. Load iris dataset
3. View the feature names from iris dataset (input variables)
4. View the species names from iris dataset (target variable)
5. Create a data frame based on input and target variables.
6. View the data frame.
7. Add a column “species name” into the data frame. It shows the name based on code. i.e. 0 = setosa, 1 = versicolor, 2 = virginica.
8. Split the data frame into three sub data frame categorized by the target group.
9. Perform data visualization on based on:
   1. Setosa and Vesicolor (x-axis: Sepal length, y-axis: Sepal width)
   2. Setosa and Vesicolor (x-axis: Petal length, y-axis: Petal width)
10. Assign the input as X and target as y values.
11. Import train\_test\_split library.
12. Split the dataset into 80% training and 20% testing.
13. Import KNN library.
14. Create a KNN model with three neigborhood.
15. Fit the model with training and testing data.
16. Check the accuracy score of the model.
17. Import confusion matrix library.
18. Use the KNN model to predict X\_test.
19. Compare your model prediction result with y\_test using confusion matrix.
20. Display the confusion matrix using heatmap from seaborn.

**Part B: Regression**

1. Import pandas library.
2. Load “homeprices.csv” dataset.
3. View data frame and check the descriptions.
4. Filter missing values (if any).
5. Use linear regression model,
   1. predict the price if the house is 3,000 square ft area with three bedrooms and age of 40 years.
   2. predict the price if the house is 2,500 square ft area with four bedrooms and age of five years.