**MCA 5161 MACHINE LEARNING LAB**

**Lab-00: Introduction to Python**

**Tutorial 1**

* Complete Anaconda Installation on laptop, following the instructions available online.
* Launch Jupyter notebook into your browser
* Use python 3 and write down the syntax and example for the following:

1. Print()
2. Comments, single line and multiline
3. Rules for Python variables:

* variable name must start with a letter or the underscore character
* variable name cannot start with a number
* variable name can only contain alpha-numeric characters and underscores (A-z, 0-9, and \_ )
* Variable names are case-sensitive (age, Age and AGE are three different variables)
* With regards to variable names, what is camel casing, pascal casing and snake case?

1. What are the data types supported in Python?
2. Type()
3. List the methods supported in python for casting variables from one type to another?
4. Create a variable to store your name and print it.
5. Define a variable name=”MICHAEL JACKSON”
6. What is the result of the following statements:
   1. name[::2]
   2. name[0:5:2]
   3. len(name)
   4. name1=name\*3

print(name1)

* 1. name2 = name1.lower()

print(name2)

* 1. name3=name.replace('MICHAEL','JANET')

print(name3)

1. What is the output of the following code?

fruits = ["apple", "banana", "cherry"]  
x, y, z = fruits  
print(x)  
print(y)  
print(z)

1. What is the output of the following code?

x = "Python is "  
y = "awesome"  
z =  x + y  
print(z)

1. What are lists in Python? Define a list to contain the following elements listA=[“hello”,10,1.2]
2. What are tuples in Python? Define a tuple to contain the following elements tupleA=(“hello”,10,1.2)
3. What is the difference between Lists and Tuples in Python?
4. Write down the syntax with example of the following statements in Python
   1. IF, IF ELSE, IF ELIF
   2. Logical operators – NOT , OR, AND
   3. FOR
   4. WHILE

**Lab-01: Statistical Analysis using Python**

* **Complete Tutorial 2 of Pandas through help reference**
* **Exer 1 – Data frame creation and manipulation**

1. Create a data frame with details of 5 students and columns as Roll Number, Name, Gender, Marks1, Marks2, Marks3.
2. Create a new column with total marks
3. Find the lowest marks in Marks1
4. Find the Highest marks in Marks2
5. Find the average marks in Marks3
6. Find student name with highest average
7. Find how many students failed in Marks2 (<40)

* **Exer 2 – Data Analysis using mtcars**
  1. Find the car with the best mpg
  2. Find the car with the worst mpg
  3. Find the car with the best horsepower
  4. Find 5 number summary of displacement
  5. Find median horse power
  6. What is average mpg for manual vs. automatic cars
  7. Draw a
  8. histogram of miles per gallon
  9. Boxplot of mpg for each cylinder type
  10. Create a crosstab displaying count of automatic vs. manual cars
  11. Create a crosstab displaying count of “am vs cyl”
  12. What is the correlation between the weight of the car and mpg

**Lab -02: Regression Techniques**

* **Exer 3 – Data Analysis using Titanic**
  1. Download the titanic data set, Train.csv and Test.csv, convert to data frames
  2. What is the survival rate in train dataset?
  3. What is the distribution in age among passengers?
  4. Visualise the distribution in fare.
  5. Tabulate the total number of Male vs Females
  6. Create a crosstab displaying Sex and Survival.
  7. Create a pivot table showing Passenger class and Survival
  8. Find the survival rates among passenger class.
  9. Using Chi Square test, find if there is a relationship between Passenger class and survived?
  10. Using Chi Square test, find if there is a relationship between Sex and Survived
* **Complete Tutorial 2 of NumPy**

**Lab -03: Classification Techniques, Performance Evaluation**

* 1. Write a user defined function ‘myLinReg(x,y) to perform Linear Regression given x and y. The function should return the coefficients of the straight line.
  2. Use mtcars to build a linear regression model to predict mpg given the weight of the car.
  3. Using myLinReg(x,y) , What is the mpg of a car, whose weight is 5.5?
  4. Use the boston.csv dataset and determine the best 5 features to predict ‘MEDV’.
  5. Using sklearn.linear\_model, find the multiple regression model for the boston.csv dataset using the best 5 features. (from sklearn.linear\_model import LinearRegression)
  6. Find the accuracy of the model using appropriate metrics.

**Lab -04: Classification Techniques, Performance Evaluation9**

1. Use the Train (Titanic) dataset and build a decision tree model (using sklearn) to predict class label survived using the attributes , Sex, Age and Pclass . Use the test data set and compute the performance of the classifer .
2. Use the IRIS data set to build a logistic regression model (using sklearn) and predict the class label ‘Species’. Divide the dataset into training and test set (70,30) using train\_test\_split method in sklearn. Use the test data set and compute the performance of the classifer .
3. Use the IRIS data and KNeighborsClassifier (using sklearn) and predict the class label ‘Species’. Divide the dataset into training and test set (70,30) using train\_test\_split method in sklearn. Use the test data set and compute the performance of the classifer .

**Lab -05: End Semester Examination**

**Lab -06: Clustering Techniques, Cluster Analysis**

* ***Assessment Test 2***

1. Use the sklearn package to perform clustering using the kmeans method . Use the Iris data set and 4 features of sepal length, sepal width, petal length, and petal width.
2. Write down the cluster centres for k=3 and k=5.
3. Visualize the group clusters for 3 species of flower using 2 features petal length and sepal width. Use matplotlib and pyplot method.