Applied Machine Learning Workshop (CSE 3193)

ASSIGNMENT-5: INTRODUCTION MACHINE LEARNING AND PROBLEM FORMULATION

1. Write a python program to create a default dict which contains the list of interest of each user.
   interests = [(”Amit”, ”Hadoop”), (”Amit”, ”Big Data”), (”Amit”, ”Java”), (”Amit”, ”Spark”), (”Raj”, ”NoSQL”), (”Raj” , ”MongoDB”), (”Raj”, ”HBase”), (”Raj”, ”Postgres”), (”Akash”, ”Python”), (”Akash”, ”scikit-learn”), (”Akash”, ”scipy”), (”Akash”, ”numpy”), (”Akash”, ”pandas”), (”Sheetal”, ”R”), (”Sheetal”, ”Python”), (”Sheetal”, ”machine learning”), (”Sheetal”, ”Spark”), (”Shyam”, ”Python”), (”Shyam”, ”R”), (”Shyam”, ”Java”), (”Shyam”, ”C++”), (”Rahul”, ”MongoDB”), (”Rahul”, ”Hadoop”), (”Rahul”, ”Java”), (”Rahul”, ”Python”), (”Miley”, ”scikit-learn”), (”Miley”, ”pandas”)]

2. Write a Python program to print the list of users with interest in “Python” for the above dataset.

3. Write a Python program to compute the average salary for each year of experience using the default dictionary for the given dataset. Also, print the average salary for maximum and minimum years of experience.
   salaries_and_exp = [(83000, 8.7),(82000, 8.7),(88000, 8.7),(48000, 0.7),(45000, 0.7),(76000, 6),(69000, 6), (76000, 7.5),(77000, 7.5),(60000, 2.5), (83000, 10), (48000, 1.9), (63000, 4.2)]

4. Write a Python program that creates a counter of the vowels in the word ”Python Exercises.”

5. Write a Python program that creates a ’Counter’ for a list of items and converts it to a list of unique items with their counts.
   items = [”Red”, ”Green”, ”Black”, ”Black”, ”Red”, ”red”, ”Orange”, ”Pink”, ”Pink”, ”Red”, ”White”]

6. Write a Python program to define a function Person and store and print the value associated with the function like name of the person, age, phone_Number, and city using keyword argument.

7. What is the output of the following code:

```
def func(a,b,*args,**kwargs):
    print(a, b)
    print(args)
    print(kwargs)
func(1, 3,4,5, Name = 'Tim', Salary = 60000)
```

8. Write Python code from scratch to implement Precision, Recall (Sensitivity), F1 Score and accuracy.

9. A classifier is built which will predict the class for the new input sample. The model is tested with 165 samples, and the following results are obtained: True Positive is 100, True Negative is 50, False Positive is 10, and False Negative is 5. Write and false positive rate.

10. For the generated values from a classifier for spam and not spam, write a Python program to compute and plot a confusion matrix and compute the metrics like accuracy, precision, recall, and f1 score based on the confusion matrix data.
    Actual=[1,1,1,0,1,0,1,1,0,0,1,0,1,0,1]
    Predicted =[1,0,1,0,1,1,1,1,0,0,0,1,0,1,1]