MACHINE LEARNING LAB MANUAL (LPCCS-110)

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENT

FOR THE AWARD OF THE DEGREE.

BACHELOR OF TECHNOLOGY

(COMPUTER SCIENCE AND ENGINEERING)

SUBMITTED BY:

Name: Navkirat Singh

CRN: 2215122

URN: 2203512

SUBMITTED TO:

Prof. Kamaldeep Kaur

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

GURU NANAK DEV ENGINEERING COLLEGE, LUDHIANA

INDEX:

S No.	Practical Title	Date	Remarks
1.	Introduction to Python, Variables, Loops, Control Structures, Libraries and File Handling		
2.	Write a program to demonstrate FIND-S algorithm for finding the most specific hypothesis based on a given set of training data samples. Read the training data from a .CSV file.		
3.	Write a program for Candidate Elimination algorithm for finding the consistent version space based on a given set of training data samples. The training data is read from a .CSV file.		
4.	Build an Artificial Neural Network by implementing the Back propagation algorithm and test the same using appropriate data sets.		
5.	Write a program to implement the naïve Bayesian classifier for a sample training data set stored as a .CSV file. Compute the accuracy of the classifier, considering few test data sets.		
6.	Assuming a set of documents that need to be classified, use the Naive Bayesian Classifier model to perform this task. Calculate the accuracy, precision, and recall for your data set.		
7.	Write a program to construct a Bayesian network considering medical data. Use this model to demonstrate the diagnosis of heart patients using standard Heart Disease Data Set.		
8.	Apply EM algorithm to cluster a set of data stored in a .CSV file. Use the same data set for clustering using k-Means algorithm. Compare the results of these two algorithms and comment on the quality of clustering.		
9.	Write a program to implement k-Nearest Neighbour algorithm to classify the iris data set. Print both correct and wrong predictions.		
10.	Implement the non-parametric Locally Weighted Regression algorithm in order to fit data points. Select appropriate data set for your experiment and draw graphs.		
11.	Write a program to predict high risk patients based on variables (e.g. blood pressure, age etc.) and discriminate them from low risk patients.		
12.	Develop a genetic algorithm for optimization of hyper parameters in machine learning.		

Practical 1

AIM: Introduction to Python, Variables, Loops, Control Structures, Libraries and File Handling

Introduction to Python:

Python is a high-level, interpreted programming language known for its **simplicity** and **readability**. It supports multiple programming paradigms, including **procedural, object-oriented, and functional programming**. Python language is used in fields of programming including Development, AI/ML etc.

Why Python?

- Easy to learn and read
- Extensive standard library and third-party packages
- Used in Web Development, Data Science, AI/ML, Automation, and more
- Large community and strong support

Variables:

- Variables are containers for storing data values.
- A variable is a named location in memory used to store data.
- Python variables do not require explicit type declaration
- Primary Data Types in Python include String, Integer, Float and Boolean etc.
- Other Data Types in Python include List, Tuples, Sets, Dictionaries etc.

Loops:

- Loops allow repeated execution of a block of code.
- The main types of loops are:
  For loops (counting through items) and While loops (based on conditions)

Control Structures:

- Control structures allow decision-making.
- The if-elif-else statements let us execute different blocks of code based on conditions.

Python Libraries:

A Python library is a collection of related modules. It contains bundles of code that can be used repeatedly in different programs. It makes Python Programming simpler and convenient for the programmer.

The Python Standard Library contains the exact syntax, semantics, and tokens of Python. It contains built-in modules that provide access to basic system functionality like I/O and some other core modules.

Top Python libraries include NumPy, Pandas, Matplotlib, TensorFlow, PyTorch, Scikit-learn etc.

File Handling:

File handling refers to the process of performing operations on a file such as creating, opening, reading, writing and closing it, through a programming interface like Python. It involves managing the data flow between the program and the file system on the storage device, ensuring that data is handled safely and efficiently.

Code:

```
# Import Libraries

import random

import math


# Variables

name = "Alice"  # String variable

age = 25        # Integer variable

height = 5.4    # Float variable


print("Name: Navkirat Singh, URN: 2203512")

print(f"Name: {name}, Age: {age}, Height: {height}")


# List of Random Numbers using Random Library

numbers = [random.randint(1, 10) for _ in range(5)]

# Square Root List using Math Library
```

```
sqrt_numbers = [round(math.sqrt(num),2) for num in numbers]

# Control Structures

if age >= 18:

    print(f"{name} is an adult.")

else:

    print(f"{name} is a minor.")


# Loops: While and For

index = 0

sum_numbers = 0

product_numbers = 1

while index<len(numbers):

    sum_numbers += numbers[index]

    product_numbers *= numbers[index]

    index+=1


# for num in numbers:

#     sum_numbers += num

#     product_numbers *= num


print(f"Numbers: {numbers}")

print(f"Square Root Numbers: {sqrt_numbers}")


print(f"Sum: {sum_numbers}")

print(f"Product: {product_numbers}")
```

```
# File Handling

# Writing to a file

with open("example.txt", "w") as file:

    file.write("Hello, Python!\nThis is file handling.")

print("Data written to File")


# Reading from a file

with open("example.txt", "r") as file:

    content = file.read()

    print("File Content:\n", content)
```

Output:

```
Name: Navkirat Singh, URN: 2203512
Name: Alice, Age: 25, Height: 5.4
Alice is an adult.
Numbers: [10, 5, 7, 7, 8]
Square Root Numbers: [3.16, 2.24, 2.65, 2.65, 2.83]
Sum: 37
Product: 19600
Data written to File
File Content:
 Hello, Python!
This is file handling.
```

Practical 2

AIM: Write a program to demonstrate FIND-S algorithm for finding the most specific hypothesis based on a given set of training data samples. Read the training data from a .CSV file.

Code:

```
import csv

import pandas as pd

def findS():

    h = ["^", "^", "^", "^", "^", "^"]

    with open('./SampleData.csv', 'r') as file:

        data = csv.reader(file)

        print(pd.read_csv('./SampleData.csv'))

        for row in data:

            if(row[len(row)-1]=="Yes"):

                for i in range(len(row)-1):

                    if(h[i]!=row[i]):

                        if(h[i]=="?"):

                                continue

                        elif(h[i]=="^"):

                            h[i]=row[i]

                        else:

                            h[i]="?"

    return h

print("Name: Navkirat Singh, URN: 2203512")

final_h = findS()

print("final hypothesis: ", final_h)
```

Output:

```
Name: Navkirat Singh, URN: 2203512
     Sky  Temp Humidity    Wind   Water Forecast EnjoySport
0  Sunny  Warm   Normal  Strong    Same      Yes        Yes
1  Sunny  Warm     High  Strong    Same      Yes        Yes
2  Rainy  Cold     High  Strong  Change       No         No
3  Sunny  Warm     High  Strong    Same      Yes        Yes
final hypothesis:  ['Sunny', 'Warm', '?', 'Strong', 'Same', 'Yes']
```