NAVODAYA INSTITUTE OF TECHNOLOGY

MACHINE LEARNING LAB (BCSL606)

Program 4

4. For a given set of training data examples stored in a .CSV file, implement and demonstrate the Find-S algorithm to output a description of the set of all hypotheses consistent with the training examples.

PROGRAM:

csv file

Outlook	Temperat	Humidity	Windy	PlayTenni
Sunny	Hot	High	FALSE	No
Sunny	Hot	High	TRUE	No
Overcast	Hot	High	FALSE	Yes
Rain	Cold	High	FALSE	Yes
Rain	Cold	High	TRUE	No
Overcast	Hot	High	TRUE	Yes
Sunny	Hot	High	FALSE	No

PROGRAM:

```
import pandas as pd
def find_s_algorithm(file_path):
    data = pd.read_csv(file_path)
    print("Training data:")
    print(data)
    attributes = data.columns[:-1]
    class_label = data.columns[-1]
    hypothesis = ['?' for _ in attributes]
    for index, row in data.iterrows():
        if row[class_label] == 'Yes':
            for i, value in enumerate(row[attributes]):
                if hypothesis[i] == '?' or hypothesis[i] == value:
                    hypothesis[i] = value
                else:
                    hypothesis[i] = '?'
    return hypothesis
file_path = 'training_data.csv'
```

```
hypothesis = find_s_algorithm(file_path)

print("\nThe final hypothesis is:", hypothesis)
```

OUTPUT:

```
Training data:

     Outlook  Temperature  Humidity  Windy  PlayTennis
0      Sunny          Hot      High  False          No
1      Sunny          Hot      High   True          No
2   Overcast          Hot      High  False         Yes
3       Rain         Cold      High  False         Yes
4       Rain         Cold      High   True          No
5   Overcast          Hot      High   True         Yes
6      Sunny          Hot      High  False          No


The final hypothesis is: ['Overcast', 'Hot', 'High', '?']
```