

Course Code: CS481	Course Name: Data Science
Instructor Name: Dr Muhammad Atif Tahir	
Student Roll No:	Section No:

Instructions:

- Create a folder name _____ at Desktop
- You are going to use Spyder to run python code. There should be no files except question1.py and question2.py during the whole exam.
- Return the question paper.
- Read each question completely before answering it. There are **2 questions and 2 pages**
- In case of any ambiguity, you may make assumption. But your assumption should not contradict any statement in the question paper.
- All programs should be properly commented. Write your **student id** (not student name) in both programs

Time: 60 minutes.

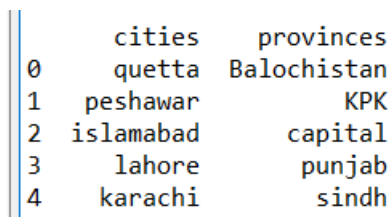
Max Marks: 10 points

Question 1: Complete the following program. Save this program as question1.py [3 Points]

```

import pandas as pd
data = {'cities': ['lahore','karachi'], 'provinces': ['punjab','sindh']}
frame1 = ..... # create DataFrame object for data
..... # print frame1
frame2 = ..... # read data.json file
frame3 = ..... # combine frame1 and frame2
frame3 = ..... # remove duplicates rows
frame3 = ..... # sort based on provinces
frame3 = ..... # re-arrange all indexes
..... # print frame3

```



	cities	provinces
0	quetta	Balochistan
1	peshawar	KPK
2	islamabad	capital
3	lahore	punjab
4	karachi	sindh

Figure1: Screen shot of Final Output for Question 1.

Question 2: Implement the following system using Python. Use german_train.csv as training file and german_test.csv as test file. Save this program as question2.py. You need to calculate classification accuracy from each model and ensemble of models. Final output is shown in Figure 2. [7 Points]

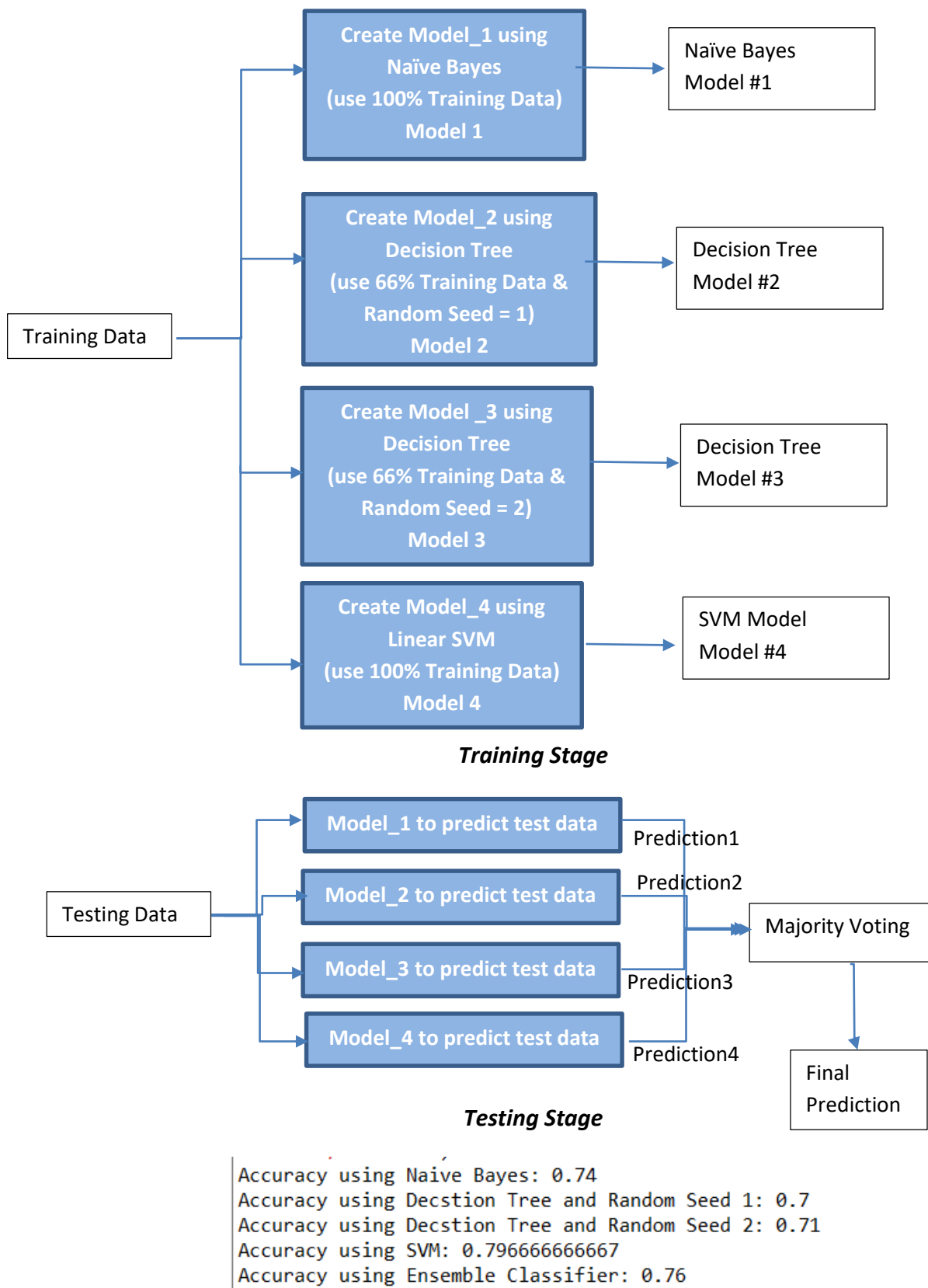


Figure2: Screen shot of Final Output for Question 2. The accuracies can differ in your implementation.

BEST OF LUCK!