**Roll no\_\_\_\_\_\_\_**

**School of Data Science and Forecasting**

**Devi Ahilya Vishwavidyalaya, Indore**

**End Semester Exam May’ 2022**

**Big Data Technologies**

**Time: 3 Hours Total Marks: 60**

**Important Instructions**

1. **All questions carry equal marks.**
2. **The student is required to submit a word document, json file and jupyter notebooks (pdf format).**
3. **Write the solution/output/summary of each question in the word document and submit the word file. Paste the screen shot of the final result of each question also in the word document.**
4. **It is important to submit the jupyter notebook (pdf format) of attempted questions. The pdf files should be numbered as rollno\_questionnumber (Example : 115\_2)**
5. **Paste the screen shot of Mongo Db in the word file.**
6. **Create your own different dataset wherever specified. No two students can have the same dataset.**
7. **Each notebook should have the student roll number in minimum 5 cells.**
8. **Marks will be awarded on the execution of the process and explanation of the results. The accuracy of the model is not the criterion for the grant of the marks.**

Q.I Considering left as dependent variable in HR dataset, split the dataset according to your last digit of roll no. (Example: if your roll no is ending with 0, the ratio will be 70, 30; if your roll no is ending with 1, the ratio will be 71, 29; if your roll no is ending with 2, the ratio will be 72, 28; if your roll no is ending with 3, the ratio will be 73, 27 etc.). Determine the accuracy of the model.

Q.II Considering the Item\_Outlet\_Sales as dependent variable in “Big Mart Sales” dataset, determine the accuracy of the model. Split the dataset according to your last digit of roll no. (Example: if your roll no is ending with 0, the ratio will be 70, 30; if your roll no is ending with 1, the ratio will be 71, 29; if your roll no is ending with 2, the ratio will be 72, 28; if your roll no is ending with 3, the ratio will be 73, 27 etc.).

III You are appointed as an Analyst at Hospital. You are required to manage enormous amount of data generated in the Hospital. Create a JSON file of your choice for representing the data that can be stored in the hospital. However, you should create only small number of observations considering the non-feasibility of generating a big data in this limited time.

Q.IV Import the data from JSON file created in Q.3 to MongoDB. Paste the screen shot of Mongo DB in word document.

1. Insert, delete, update and search a record in MongoDB.
2. Link Mongo DB with Python. Insert, delete, update and search a record using Python.

Q.V You are appointed as an analyst at Amazon. Create a RDD related to customer and execute any 10 functions that can be used for transformation and action. Discuss the results.

Q.VI Considering churn dataset, execute the following queries

Q1 What is the average Monthly Charges for customers having “DSL” Internet connection. Consider senior citizen , male and churned customers whose tenure is greater than 60.

Q2 What is the average amount for customers having “Mailed Check” as payment method and “One Year” as agreement. Consider the customers who have dependents and partner and have opted for call service.

Q3 What is the total Monthly Service Charges of customers having different billing method. Consider the male and senior citizen customers whose tenure is less than 20 and have multiple connections

Q4 How many male and female customers has dependents and no dependents. Consider those customers who have monthly service charges greater than 100.

Q5 How many number of customers have churned and not churned. Consider only female customers who have no dependents and has done call service and has preferred electronic check method.

Q6 How many male and female customers have no dependents and have multiple connections. Consider the customers who have call service and has preferred either electronic check method or mailed check method.

Q7 What is the average tenure of male and female customers who have no dependents and have partners. Consider the customers who have call service and has preferred either electronic check method or mailed check method.

Q8 What is the maximum monthly service charges of customers who have done payment by electronic check method? Consider only those customers who have agreement for on year or two years only.

Q9 What is the minimum total amount of male and female customers having one year or two year agreement type. Consider only those customers who have no internet connection, no online security no online backup and no device protection service.

Q10 what is the average total amount of male and female customers having partner and no partner