**JAIPURIA INSTITUTE OF MANAGEMENT, INDORE**

**PGDM**

**FOURTH TRIMESTER (Batch 2020-22)**

**END TERM IMPROVEMENT EXAMINATION, DEC-2021**

|  |  |  |  |
| --- | --- | --- | --- |
| Course Name | **Data Mining and Predictive Analytics** | Course Code | **BA401** |
| Max. Time | **2 hours** | Max. Marks | **40** |

**INSTRUCTIONS:**

|  |  |
| --- | --- |
| 1 | All questions are to be attempted. |
| 2 | In each worksheet question number and answer number are separately specified. |
| 3 | Answers are to be written in the provided End Term Examination Answer Sheet and screenshots with details are to be pasted in the specified worksheet. |
| 4 | All questions are to be answered with justification. (Business Inferences and Decision Making should be explained ) |
| 5 | After performing desired operations on the dataset provided required screenshots from RapidMiner ® Studio / MS- Excel are to be pasted at the appropriate worksheet/ location. |
| 6 | Question dataset is in question worksheet and answer worksheet contains question also. |
| 7 | MS- Excel and/or RapidMiner Studio are to be used as Software Tools. |
| 8 | Question Paper consist of **FOUR** questions. |
| 9 | Submit your Answer MS-Excel Sheet with file name in the format - **YOURROLLNUMBER\_DMPA\_202022\_IV\_Oct2021 along with End Term Answer Sheet.** |
| 10 | **Question paper is of 40 Marks.** |

**Question 1 (10 Marks)**

Discuss and implement process of cleaning textile dataset by handling missing values, outliers and non-numeric values and find Principal Components from the data set, also provide business inferences on the basis of PCA. Discuss importance of data cleaning process in data mining. (Use question 1 dataset)

**Question 2 (10 Marks)**

Apply suitable Predictive Analytics Technique/s on the data set and generate business insights for decision-making. (Use question 2 dataset)

**Question 3 (10 Marks)**

Apply and analyze suitable forecasting methods to forecast five values of Open, High , Low and Close of BSE 30. Also, explain why selected method is suitable for the provided data set? (Use question 3 dataset)

**Question 4 (10 Marks)**

A departmental super store chain wants to learn more about buyers purchase patterns. Specifically, they want to know what items(from the division) are purchased in conjunction with each other departments/division, for purposes of display, point of sale special offers, and to eventually implement a real time recommender system to cross-sell items at time of purchase. The data is in the form of customer transactions. On the basis of data analysis using suitable analytical technique, what actionable business inferences you generate for the organization/ store. Scenario - A random sample of customers is shown in the Data sheet. A "1" indicates a purchase has been made from a catalog in that division, a "0" indicates no purchase. (Use question 4 dataset)