**CS 612: Lab 1**

**Question 1: The following are Data Mining questions. Each can be “Predictive Data Mining” using “Classification”, predictive data mining using “Regression”, “Associative Data Mining”, “Clustering Data Mining” or “Anomaly Detection”. For each case, explain in 1 or at most 2 lines what each data mining case falls int?**

* 1. Predicting the future stock price of a company using historical record. We would attempt to create a model that can predict the continuous value of the stock price

**Predictive data mining using Regression because stock price is continuous value.**

* 1. An emergency room in a hospital measures 17 variables (ex: blood pressure, age, etc) of newly admitted patients. A decision must be taken whether to put the patient in an intensive-care unit. Due, to the high cost of ICU, those patients who may survive more than a month are given higher priority. The problem is to predict high risk patients and discriminate them from low risk patient.

**Predictive Data Mining using Classification because a patient may be high risk or not. So the result can be yes or no.**

* 1. Monitoring the heart rate of patient for abnormalities.

**Descriptive data mining , Anomaly Detection. Because it is the task of identifying the heart rate whose characteristics are significantly different from the rest of the data.**

* 1. Monitoring seismic waves of earthquake activities

**Descriptive data mining , Anomaly Detection. Because it is the task of identifying the seismic waves of eatrhquack whose characteristics are significantly different from the rest of the data.**

* 1. Working on the brain image of the monkeys in the zoo who show similar behavior.

**Descriptive Data Mining, Cluster Analysis. Because it is finding groups of similar behavior.**

* 1. A credit card company typically receives hundreds of thousands of applications for new cards. The application contains information regarding several different attributes, such as annual salary, any outstanding debts, age, etc. The problem is to categorize applications into those who have good credit, bad credit, or fall into gray area (thus require further human analysis).

**Descriptive Data Mining, Cluster Analysis. Because it is grouping set of related customer.**

* 1. People who watch CNN then most probably they don’t like to watch fox news.

**Descriptive Data Mining, Association Analysis. Because people who watch CCN are in different political views of people who watch Fox news. So people watching CNN most probably don’t like Fox news.They are associated values .**

* 1. Intrusion detection, example identifies strange patterns in the network traffic (that could signal a hack).

**Descriptive Data Mining, Anomaly Detection. . Because it is the task of identifying the strange pattern whose characteristics are significantly different from the rest of the network traffic.**

* 1. Astronomers have been cataloguing distant objects in the sky using long-exposure CCD images. The objects need to be labeled as star. Galaxy, nebula etc. The data is highly noisy, and the images are very faint. The cataloguing can take decades to complete. How can physicists automate the cataloguing process and improve its effectiveness?).

**Descriptive Data Mining, Cluster Analysis. Because it is grouping a set of related distant objects in the sky.**

* 1. Detection of fake news and misinformation in the Internet.

**Descriptive Data Mining, Anomaly Detection. Because it is the task of identifying news and information whose characteristics are significantly different from the rest of the information.**

**Question 2: Write a Python program that creates a matrix of 10 by 10 using the Data provided below.**

* 1. **Place the following data into a file and read the data into the matrix from a file called “Data.txt”**
  2. **Select three columns, columns 3, 1, and 9, sort them in ascending order of column 3 and make a matrix of 10 by 3 using these three columns**
  3. **Select another set of three columns, columns 5, 2, 7, sort them in descending order of column 5 and make a matrix of 10 by 3 using these three columns - Call it matrix2.**
  4. **Add the two sorted matrices to each other and put the result into another 10 by 3 matrix call it matrix3. No need to sort matrix 3.**
  5. **Add the content of each row of Matrix 3 and put it into a new 10 by 1 matrix call it Matrix 4.**
  6. **Sort Matrix4 in ascending order**

**Also, you are required to use all the libraries that are provided in numpy and Scikit-lean and any other libraries that can minimize your coding. This program MUST NOT BE WRIITTEN LIKE C++.**

**You are required to use the following 10 by 10 in a file to get started.**

**26 8 2 12 67 89 8 78 56 7**

**13 9 57 11 10 99 15 90 88 15**

**6 23 28 32 53 59 48 41 60 62**

**17 6 93 95 20 25 29 31 71 30**

**16 13 26 22 14 66 71 5 2 91**

**40 53 22 18 44 73 77 84 89 88**

**34 54 89 8 65 74 96 3 1 80**

**14 28 39 44 86 88 75 77 93 4**

**65 66 82 97 51 41 28 33 55 79**

**13 48 71 98 26 38 55 59 7 19**

**Your output should be as follows:**

**The original matrix is as follows:**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **26** | **8** | **2** | **12** | **67** | **89** | **8** | **78** | **56** | **7** |
| **13** | **9** | **57** | **11** | **10** | **99** | **15** | **90** | **88** | **15** |
| **6** | **23** | **28** | **32** | **53** | **59** | **48** | **41** | **60** | **62** |
| **17** | **6** | **93** | **95** | **20** | **25** | **29** | **31** | **71** | **30** |
| **16** | **13** | **26** | **22** | **14** | **66** | **71** | **5** | **2** | **91** |
| **40** | **53** | **22** | **18** | **44** | **73** | **77** | **84** | **89** | **88** |
| **34** | **54** | **89** | **8** | **65** | **74** | **96** | **3** | **1** | **80** |
| **14** | **28** | **39** | **44** | **86** | **88** | **75** | **77** | **93** | **4** |
| **65** | **66** | **82** | **97** | **51** | **41** | **28** | **33** | **55** | **79** |
| **13** | **48** | **71** | **98** | **26** | **38** | **55** | **59** | **7** | **19** |

**Suppose the next three random numbers are 3, 1, and 9. So you your out should prints:**

**The first three random numbers are 3, 1, and 9 which makes Matrix 1 as follows:**

|  |  |  |
| --- | --- | --- |
| **2** | **26** | **56** |
| **57** | **13** | **88** |
| **28** | **6** | **60** |
| **93** | **17** | **71** |
| **26** | **16** | **2** |
| **22** | **40** | **89** |
| **89** | **34** | **1** |
| **39** | **14** | **93** |
| **82** | **65** | **55** |
| **71** | **13** | **7** |

**The sorted version of Matrix1 in ascending order of its first column (column 3 of the original matrix) is:**

|  |  |  |
| --- | --- | --- |
| **2** | **26** | **56** |
| **22** | **40** | **89** |
| **26** | **16** | **2** |
| **28** | **6** | **60** |
| **39** | **14** | **93** |
| **57** | **13** | **88** |
| **71** | **13** | **7** |
| **82** | **65** | **55** |
| **89** | **34** | **1** |
| **93** | **17** | **71** |

**Suppose the next set of three random numbers are 5, 2, and 7. So you your output should prints:**

|  |  |  |
| --- | --- | --- |
| **67** | **8** | **8** |
| **10** | **9** | **15** |
| **53** | **23** | **48** |
| **20** | **6** | **29** |
| **14** | **13** | **71** |
| **44** | **53** | **77** |
| **65** | **54** | **96** |
| **86** | **28** | **75** |
| **51** | **66** | **28** |
| **26** | **48** | **55** |

**Sorting Matrix 2 in descending order of its first column (column 5 of the original matrix) we get:**

|  |  |  |
| --- | --- | --- |
| **86** | **28** | **75** |
| **67** | **8** | **8** |
| **65** | **54** | **96** |
| **53** | **23** | **48** |
| **51** | **66** | **28** |
| **44** | **53** | **77** |
| **26** | **48** | **55** |
| **20** | **6** | **29** |
| **14** | **13** | **71** |
| **10** | **9** | **15** |

**Adding the contents of the Matrix1 and Matrix2, and put the result in one column Matrix called Matrix 3**

**Matrix 1: Matrix 2 Matrix 3**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **2** | **26** | **56** |  | **86** | **28** | **75** |  | 88 | 54 | 131 |
| **57** | **13** | **88** |  | **67** | **8** | **8** |  | 124 | 21 | 96 |
| **28** | **6** | **60** |  | **65** | **54** | **96** |  | 93 | 60 | 156 |
| **93** | **17** | **71** |  | **53** | **23** | **48** |  | 146 | 40 | 119 |
| **26** | **16** | **2** |  | **51** | **66** | **28** |  | 77 | 82 | 30 |
| **22** | **40** | **89** |  | **44** | **53** | **77** |  | 66 | 93 | 166 |
| **89** | **34** | **1** |  | **26** | **48** | **55** |  | 115 | 82 | 56 |
| **39** | **14** | **93** |  | **20** | **6** | **29** |  | 59 | 20 | 122 |
| **82** | **65** | **55** |  | **14** | **13** | **71** |  | 96 | 78 | 126 |
| **71** | **13** | **7** |  | **10** | **9** | **15** |  | 81 | 22 | 22 |

**After adding the content of matrix 3 to get the matrix 4 we get:**

|  |
| --- |
| 273 |
| 241 |
| 309 |
| 305 |
| 189 |
| 325 |
| 253 |
| 201 |
| 300 |
| 125 |

**Sorting matrix 4 in ascending order:**

|  |
| --- |
| 125 |
| 189 |
| 201 |
| 241 |
| 253 |
| 273 |
| 300 |
| 305 |
| 309 |
| 325 |

**Test your program with a data set that is provided in this lab.**

* **Place the questions with the answers to question 1 in one file, call it “Lab1-Q1”**
* **the code for the program in another file call it “Lab1-Q2-Code”**
* **the snapshot of the output of the program in the third file., call it “Lab1-Q2-Output”**
* **Place all the files in a folder and call the folder based on your name and your partner’s name (ex:Jack-and-Nancy-Lab1)**
* **Zip the folder and only one person in the team is required to submit the work.**