**Project 5 – Advanced Data Mining Applications**

**CS548 / BCB503 / CS583 Knowledge Discovery and Data Mining - Fall 2018**

**Prof. Carolina Ruiz and TA Huimin Ren**

**Student: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

|  |  |
| --- | --- |
| Description of the particular problem within the selected data mining topic to be addressed in this project | /15 |
| Description of the approach used in this project to tackle the above problem.  *All data mining techniques you use in this project for pre-processing, mining and evaluation must have been covered in class during this semester* | /25 |
| Description of the dataset selected | /15 |
| Appropriateness of the dataset selected with respect to this topic/problem | /10 |
| Guiding questions | /10 |
| Preprocessing | /10 |
| **Experiments:**   * Sufficient & coherent (most experiments must be performed in Python) | /25 |
| * Objectives, Data, Additional Pre/Post-processing | /20 |
| * Presentation of results | /20 |
| * Analysis of results | /30 |
| Overall discussion, comparisons, and conclusions | /20 |
| TOTAL | /200 |

Total Written Report: \_\_\_\_\_\_\_\_\_\_\_\_\_\_/200 = \_\_\_\_\_\_\_\_\_\_\_/100

Class Presentation: \_\_\_\_\_\_\_\_\_\_\_/100

Class participation during project presentation: \_\_\_\_\_\_\_\_\_\_\_/100

*Do not exceed the given page limits for this written report*

**Topic: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ <at most 1 page>**

1. **Description of the particular problem within the selected data mining topic to be addressed in this project:**
2. **Description of the approach used in this project to tackle the above problem:**
3. **Dataset Name:**
4. **Where found:**
5. **Dataset Description:**
6. **Initial data preprocessing, if any:**
7. **Three Guiding Questions about the dataset domain:**
8. …
9. …
10. …

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Summary of Experiments.** *At most 2 page.* | | | | | | | | | |
| **Tool** | **Pre-process** | **Mining**  **Technique** | **Results** | **…** | **Results** | **Time**  **taken** | **Evaluation** | **Observations about experiment**  **Observations about visualization**  **Interpretation results** | **You can add**  **other columns** |
| Weka?  Python? |  |  |  |  |  |  |  |  |  |
| … |  |  |  |  |  |  |  |  |  |
| … |  |  |  |  |  |  |  |  |  |
| … |  |  |  |  |  |  |  |  |  |
| … |  |  |  |  |  |  |  |  |  |
| … |  |  |  |  |  |  |  |  |  |

**<Continuation of summary of results>**

**Analysis of Results: (at most 1 page)** 1. Analyze the effect of varying parameters/experimental settings on the results. 2. Analyze the results from the point of view of the Domain, and discuss the answers that the experiments provided to your guiding questions. 3. Include and explain (some of) the best / most interesting results you obtained in your experiments. 4. Include visualizations.

**Summary of what you learned in this project:**