**Project 3 – Association Rule Mining**

**CS548 / BCB503 Knowledge Discovery and Data Mining - Fall 2017**

**Prof. Carolina Ruiz**

**Students:** <replace this with your names ***in alphabetical order by last name***>

|  |  |
| --- | --- |
| **Dataset :*** Dataset Description
* Data Exploration
* Data Preprocessing – converting to bag of words, feature selection, …
 | /15/10/30 |
| **Code Description:** Association Rules | **Weka**  /20 | **Python** /20  |
| **Experiments:*** Guiding Questions
 | /15 |
| * Sufficient & coherent set of experiments
 | /15 | /15 |
| * Objectives, Parameters, Additional Pre/Post-processing
 | /15 | /15 |
| * Presentation of results
 | /15 | /15 |
| * Analysis of individual experiments’ results
 | /15 | /15 |
| Quantitative Analysis of Results and Discussion  | /20 |
| Qualitative Analysis of Results, Discussion, and Visualizations  | /20 |
| Advanced Topic | /30 |
| Total Written Report |  /300 = /100 |

**Dataset Description, Exploration, and Initial Preprocessing: (at most 1.3 page)**

**[15 points] Dataset Description: Include dataset domain, number of instances (= documents), range for the number of words in each document, where the dataset was collected from, and how it was collected (at most 1/2 page)**

**[10 points] Data Exploration: (e.g., comments on interesting or salient aspects of the dataset) (at most 1/4 page)**

**[30 points] Data preprocessing: Describe how the documents were converted into word vectors using Weka filters and/or Python libraries. Provide a detailed description including specific function names (at most 1/2 page, start here and continue on the next page)**

**[20 points] Weka Association Rules Code Description: Inputs, output, and process Weka’s code follows to construct the association rules (at most 1/3 page)**

**[20 points] Python Packages and Functions used for Association Rule mining. Describe inputs & outputs (at most 1/3 page)**

 **[15 points] Three Guiding Questions: (at most 1/3 page)**

1. **…**
2. **…**
3. **…**

|  |
| --- |
| **Summary of Association Rule Mining Experiments in Weka.** *At most 2/3 page.* **[60 points]** |
| **Pre-process** | **Parameters** | **Post-process** | **# of levels** | **# of rules** | **Interesting rules** | **Salient observations about experiment** |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

|  |
| --- |
| **Summary of Association Rule Mining Experiments in Python.** *At most 2/3 page.* **[60 points]** |
| **Pre-process** | **Parameters** | **Post-process** | **# of levels** | **# of rules** | **Interesting rules** | **Salient observations about experiment** |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

**[20 points] Quantitative Analysis of Weka and Python Results and Discussion (at most 1/3 page)**

**[20 points] Qualitative Analysis of Weka and Python Results and Visualizations (at most 1 page)**

(Remember also to analyze the results from the point of view of the dataset domain by searching the dataset’s domain (e.g., medical) literature, and discuss the answers that the experiments provided to your guiding questions.)

**Advanced Topic: Sequential Pattern Mining Using Association-Rules-like approach (at most 1 page)**

**[7 points] List of sources/books/papers used for this topic (include URLs if available). Most contain at least 2 references different from the textbook:**

* …
* …
* …

...

**[23 points] In your own words, provide an in-depth, yet concise, description of this topic. Make sure to cover all relevant data mining aspects of the topic beyond what it is covered in Section 7.4 of the textbook.**

**Authorship:** Although each student on the team is expected to be involved in every aspect of the project, describe in detail here the main contributions that each of the team members made to this project. This authorship description must accurately reflect the work done by each team member, and must be approved by all of the members of the team (at most 1/3 page)