**Project 3 – Association Rule Mining**

**CS548 Knowledge Discovery and Data Mining - Spring 2015**

**Prof. Carolina Ruiz**

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

|  |  |  |
| --- | --- | --- |
| **Dataset :*** Dataset Description
* Data Exploration
* Initial Data Preprocessing (if any)
 | **Dataset 1: Mushroom** | **Dataset 2: Census-Income** |
| /05/10/05 | /05/10/05 |
| **Code Description:** Association Rules and Classification Association Rules | **Weka**/15 | **Python** /15 | **Weka**  | **Python**  |
| **Experiments:*** Guiding Questions
 | /10 | /10 |
| * Sufficient & coherent set of experiments
 | /05 | /05  | /15 | /15 |
| * Objectives, Parameters, Additional Pre/Post-processing
 | /05 | /05 | /15 | /15 |
| * Presentation of results
 | /05 | /05 | /15 | /15 |
| * Analysis of individual experiments’ results
 | /05 | /05 | /15 | /15 |
| Quantitative Analysis of Results and Discussion  | /10 | /20 |
| Qualitative Analysis of Results, Discussion, and Visualizations  | /10 | /20 |
| Advanced Topic | /30 |
| Total Written Report  | /340 |
| Total Written Report per student | /90 | /90 | /90 |
| Class Presentation per student | /07 | /07 | /07 |
| Class Participation per student | /03 | /03 | /03 |
| Total Project 3 per student | /100 | /100 | /100 |

**Dataset 1: Mushroom. Description, Exploration, and Initial Preprocessing: (at most 1 page)**

**[05 points] Dataset Description: (e.g., dataset domain, number of instances, number of attributes, distribution of target attribute, % missing values, …)**

**[10 points] Data Exploration: (e.g., comments on interesting or salient aspects of the dataset, visualizations, correlation, issues with the data, …)**

**[05 points] Initial data preprocessing, if any, based on data exploration findings: (e.g., removing IDs, strings, necessary dimensionality reduction, …)**

**Dataset 2: Census-Income. Description, Exploration, and Initial Preprocessing: (at most 1 page)**

**[05 points] Dataset Description: (e.g., dataset domain, number of instances, number of attributes, distribution of target attribute, % missing values, …)**

**[10 points] Data Exploration: (e.g., comments on interesting or salient aspects of the dataset, visualizations, correlation, issues with the data, …)**

**[05 points] Initial data preprocessing, if any, based on data exploration findings: (e.g., removing IDs, strings, necessary dimensionality reduction, …)**

**Weka Code Description: Inputs, output, and process followed by Weka’s code to construct the association rules (at most 2/3 page)**

**[10 points] Association Rules Code Description:**

**[10 points] Classification Association Rules (CARs) Code Description:**

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

**[10 points] Three Guiding Questions for the Dataset 1 Mushroom: (at most 1/3 page)**

1. **…**
2. **…**
3. **…**

|  |
| --- |
| **[20 points] Summary of Association Rule Mining & CAR Experiments in Weka with Dataset 1 Mushroom .** *At most 3/4 page.* |
| **Tech.** | **Pre-process** | **Parameters** | **Post-process** | **# of levels** | **# of rules** | **Interesting****rules** | **Salient observations about experiment** | **You can add other columns** |
| AR?CAR? |  |  |  |  |  |  |  |  |
| … |  |  |  |  |  |  |  |  |
| … |  |  |  |  |  |  |  |  |
| … |  |  |  |  |  |  |  |  |
| … |  |  |  |  |  |  |  |  |
| … |  |  |  |  |  |  |  |  |

|  |
| --- |
| **[20 points] Summary of Association Rule Mining & CAR Experiments in Python with Dataset 1 Mushroom .** *At most 3/4 page.* |
| **Tech.** | **Pre-process** | **Parameters** | **Post-process** | **# of levels** | **# of rules** | **Interesting****rules** | **Salient observations about experiment** | **You can add other columns** |
| AR?CAR? |  |  |  |  |  |  |  |  |
| … |  |  |  |  |  |  |  |  |
| … |  |  |  |  |  |  |  |  |
| … |  |  |  |  |  |  |  |  |
| … |  |  |  |  |  |  |  |  |
| … |  |  |  |  |  |  |  |  |

**[10 points] Quantitative Analysis of Weka and Python Results on Dataset 1 and Discussion (at most 1/2 page)**

**[10 points] Qualitative Analysis of Weka and Python Results on Dataset 1 and Visualizations (at most 1/2 page)**

(Remember also to analyze the results from the point of view of the dataset domain, and discuss the answers that the experiments provided to your guiding questions.)

**[10 points] Three Guiding Questions for the Dataset 2 Census-Income: (at most 1/3 page)**

1. **…**
2. **…**
3. **…**

|  |
| --- |
| **[20 points] Summary of Association Rule Mining & CAR Experiments in Weka with Dataset 2 Census-Income .** *At most 3/4 page.* |
| **Tech.** | **Pre-process** | **Parameters** | **Post-process** | **# of levels** | **# of rules** | **Interesting****rules** | **Salient observations about experiment** | **You can add other columns** |
| AR?CAR? |  |  |  |  |  |  |  |  |
| … |  |  |  |  |  |  |  |  |
| … |  |  |  |  |  |  |  |  |
| … |  |  |  |  |  |  |  |  |
| … |  |  |  |  |  |  |  |  |
| … |  |  |  |  |  |  |  |  |

|  |
| --- |
| **[20 points] Summary of Association Rule Mining & CAR Experiments in Python with Dataset 2 Census-Income.** *At most 3/4 page.* |
| **Tech.** | **Pre-process** | **Parameters** | **Post-process** | **# of levels** | **# of rules** | **Interesting****rules** | **Salient observations about experiment** | **You can add other columns** |
| AR?CAR? |  |  |  |  |  |  |  |  |
| … |  |  |  |  |  |  |  |  |
| … |  |  |  |  |  |  |  |  |
| … |  |  |  |  |  |  |  |  |
| … |  |  |  |  |  |  |  |  |
| … |  |  |  |  |  |  |  |  |

**[10 points] Quantitative Analysis of Weka and Python Results on Dataset 2 and Discussion (at most 1/2 page)**

**[10 points] Qualitative Analysis of Weka and Python Results on Dataset 2 and Visualizations (at most 1/2 page)**

(Remember also to analyze the results from the point of view of the dataset domain, and discuss the answers that the experiments provided to your guiding questions.)

**Advanced Topic: <include name of the topic here>**

**[7 points] List of sources/books/papers used for this topic (include URLs if available):**

* …
* …
* …

...

**[20 points] In your own words, provide an in-depth, yet concise, description of your chosen topic. Make sure to cover all relevant data mining aspects of your topic.**

**[3 points] How does this topic relate to association rule mining?**

**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)