**CS548 Knowledge Discovery and Data Mining. Spring 2014. Project 2: Trees. Student’s Name: <write your name here>**

**[2 points] Dataset Description and initial data preprocessing if any (at most ½ page):**

**[3 points] Three Guiding Questions (at most ½ page):**

1. **…**
2. **…**
3. **…**

|  |
| --- |
| **[8 points] Summary of Classification Experiments.** *At most 1 page.* |
| **Tech.** | **Tool** | **Guiding****questions** | **Train/****Test** | **Pre-process** | **Parameters** | **Post-process** | **Confusion matrix****& Accuracy** | **Time to build model** | **Size of model** | **Interesting patterns in the model** | **Salient observations about experiment** | **You can add other columns** |
| ZeroR?OneR?J4.8?…? | Weka?Matlab?RapidM? | 1? 2? 3? | 10-fold cross-val? |  |  |  |  |  |  |  |  |  |
| … |  |  |  |  |  |  |  |  |  |  |  |  |
| … |  |  |  |  |  |  |  |  |  |  |  |  |
| … |  |  |  |  |  |  |  |  |  |  |  |  |
| … |  |  |  |  |  |  |  |  |  |  |  |  |
| … |  |  |  |  |  |  |  |  |  |  |  |  |

|  |
| --- |
| **[8 points] Summary of Regression Experiments.** *At most 1 page.* |
| **Tech.** | **Tool** | **Guiding****questions** | **Train/****Test** | **Pre-process** | **Parameters** | **Post-process** | **Error** | **Time to build model** | **Size of model** | **Interesting patterns in the model** | **Salient observations about experiment** | **You can add other columns** |
| ZeroR?OneR?Linearregr?Regr.trees?Modeltrees? | Weka?Matlab?RapidM?SPSS? | 1? 2? 3? | 10-fold cross-val? |  |  |  | Specifywhat metric(s)you use here |  |  |  |  |  |
| … |  |  |  |  |  |  |  |  |  |  |  |  |
| … |  |  |  |  |  |  |  |  |  |  |  |  |
| …. |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
| … |  |  |  |  |  |  |  |  |  |  |  |  |

**[7 points]** **Analysis of Results: include Analysis of the Classification Experiments’ Results, and Analysis of the Regression Experiments’ Results.** 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 (a part of) the best classification model, and (a part of) the best regression model obtained in your experiments.

**[7 points] Advanced Topic: <include name of the topic here>**

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

* …
* …
* …

...

**Summary of what you learned:**

**How does this topic relate to trees and the material covered in this course?**