**CSE 446 Final Project Milestone**

**Damir M. ZhaksilikovJason Frazier**

Department of Computer Science Department of Computer Science

University of Washington University of Washington

Seattle, WA 98105 Seattle, WA 98105

*damir@cs.washington.edu* *jasonf56@cs.washington.edu*

**Abstract**

The aim of this experiment is to develop a model based on Logistic Regression to successfully predict the outcome of the 2017 NCAA March Madness tournament. As a part of Kaggle’s March Machine Learning Mania 2017 challenge, we were provided with initial data for regular and tournament game scores [1]. Our course of action is to first confidently predict regular season games using logistic regression coupled with feature selection techniques. Next, we plan to experiment with additional datasets to move towards predicting tournament outcomes using regular season data. These datasets would supplement our isolated features to gear our model towards predicting tournament results rather than regular season games. After these adjustments, our model will be tested against the 2014-16 March Madness Tournament results.

**1 Project Outline (Damir)**

**2 Data Overview (Damir)**

**1.1 Regular Season Games**

**1.2 March Madness Games**

**1.2.1 Seeding**

**1.2.2 Las Vegas Point Spread**

**1.3 Feature selection results**

**3 Model Selection (Jason)**

**4 Tournament Adjustments (Damir)**

**5 Results (Jason)**

**6 Conclusions (Both)**