**How to not Blow a 3-1 Lead in this Year’s Bracket Predictions**

**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 (Damir)**

As of recent, the NCAA March Madness tournament has piqued the interests of the computing and machine learning community. The tournament’s seemingly unpredictable nature questions whether tournament brackets and the associated upsets can be accurately predicted. In this document, we will explore the existence of upset indicators, success of various models in predicting brackets, and impact of mixing raw team statistics with various measures of team rankings. First, we describe the results of regular season data manipulation and model performance of L2-Normalized Logistic Regression, Random Forest, Multilayer Perceptron, and K-Nearest Neighbors models, of which Random Forest consistently outperformed others. Next, we outline and model improvement to cater towards the nature of March Madness Tournament play. Our model’s success in prediction is then quantified – we estimate that it has an accuracy of about 70 and 60 percent accuracy in predicting regular and tournament games respectively. Our techniques in tuning our model provides a 25% increase in correctness from predictions made by non-expert humans [1].

**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 (Jason)**

**References**

[1] <http://stephenpettigrew.kinja.com/11-million-brackets-vs-espn-cbs-and-fox-experts-who-1561354312/1562354592>