

# Exploring Machine Learning Prediction for Soccer Outcomes

Ryan Baker  
University of Utah  
u0724394@utah.edu

Nathan Wilkinson  
University of Utah  
u0295388@utah.edu

## ABSTRACT

### 1 INTRODUCTION

### 2 RELATED WORK AND BACKGROUND

Sports betting is a common activity among avid sports fans in all sports, including soccer, basketball, football, baseball, and many other sports. To facilitate betting, many organizations publish public betting lines that describes which team is favored to win a game, and how much they are favored to win a game. However, many, or perhaps even most, of these models are strictly statistical and probabilistic. We seek to apply machine learning. Our goal is to focus on soccer, particularly in the United States, and to explore how different algorithms may produce better results, and if different train/test splits on data have an impact on learning results (e.g. explore if it is more useful to split data randomly or linearly within the course of a single season).

We have explored several different studies of machine learning with sports betting. These studies explore data from NCAA (college) basketball in the United States [3], NFL (professional) football in the United States [2], and EPL (professional) soccer in England [1]. These reports have utilized Bayesian methods [1, 3], decision trees and random forests [3], and Gaussian processes [2]. An important common aspect of all of these studies is that they recognize that using simple, raw statistics is not nearly as useful in machine learning algorithms. That is, they recognize the need for alternative features to help represent matches. This includes factors like home field advantage, winning- and losing-streaks, and player injuries. We will continue to explore these conclusions as we seek to form our feature representations and identify our desired algorithms.

### 3 DESCRIPTION OF DATA

### 4 METHODS

### 5 EVALUATIONS

### 6 FUTURE WORK

### 7 CONCLUSION

## ACKNOWLEDGEMENTS

This report was completed for the course project in CS 5350: Machine Learning at the University of Utah in the Fall 2017 semester. We would like to acknowledge the help of Prof. Vivek Srikumar who taught us the principles and tools necessary to complete this project. Much of the implementation of different algorithms for this project was completed as part of assignments throughout the semester, and was modified to fit our needs for this project.

## REFERENCES

- [1] Anthony C Constantinou, Norman E Fenton, and Martin Neil. 2012. pi-football: A Bayesian network model for forecasting Association Football match outcomes. *Knowledge-Based Systems* 36 (2012), 322–339.
- [2] Jim Warner. 2010. *Predicting margin of victory in nfl games: Machine learning vs. the las vegas line*. Technical Report. Technical Report.
- [3] Albrecht Zimmermann, Sruthi Moorthy, and Zifan Shi. 2013. Predicting college basketball match outcomes using machine learning techniques: some results and lessons learned. *arXiv preprint arXiv:1310.3607* (2013).