UNIVERSITY OF CALIFORNIA

Los Angeles

Beating the Book: A Machine Learning Approach to NBA Win Probabilities in Search of an Edge Over the Betting Odds

A thesis submitted in partial satisfaction of the requirements for the degree Master of Applied Statistics

by

Guy Dotan

2020

© Copyright by Guy Dotan 2020

ABSTRACT OF THE THESIS

Beating the Book: A Machine Learning Approach to NBA Win Probabilities in Search of an Edge Over the Betting Odds

by

Guy Dotan  
Master of Applied Statistics in

University of California, Los Angeles, 2020

Professor Frederic R. Paik Schoenberg, Chair

In this study, we set out to better understand the voter behavior of different demographic partitions of Los Angeles County (LAC). We developed a methodology to enrich and analyze LAC voter data by joining it with US census data and modelling vote by mail rates and turnout rates using general additive models. In doing so, we were able to forecast where in LAC we expect to see the highest rates of turnout by vote by mail and turnout in person for the 2020 presidential election. These findings are timely: In 2020, LAC will roll out a new and novel voting system called Voting Solutions for All People. The findings of the current study provide guidance on how to best allocate resources to high-turnout populations and target outreach to low-turnout populations in preparation for the first election under the new system. This paper presents a combination of visualizations, predictive models, and summary statistics that can inform LAC of where to focus outreach to expand the electorate and build a more representative electorate while also improving preparation in areas where turnout has been historically high.

Chargeback fraud is a massive problem for e-commerce businesses. Using historical ticket order data, several machine learning models are trained and tested to predict which trans- actions are high risk for chargeback. The results of this thesis show that many fraudulent transactions can be successfully identified and stopped before they are processed. Using these types of models could significantly reduce chargebacks, saving companies time and money.

The thesis of Guy Dotan is approved.

Vivian Lew

Zili Liu

Frederic R. Paik Schoenberg, Committee Chair

University of California, Los Angeles

2020

Table of Contents

1. Introduction
2. Exploratory Analysis
3. Random Forest
4. Neural Network
5. Betting Results
6. Kelly Criterion
7. Conclusions

List of Figures

List of Tables

Acknowledgements

Chapter 1

Introduction