

Mathematical Statistics Stockholm University Bachelor Thesis **2021:15** http://www.math.su.se

Predicting the Final Result of a Football Match

Alexander Crompton*

June 2021

Abstract

In this paper we study how different variables effect the possible outcomes of a football match played in the British Premier League. The goal is to find the best trained model to forecast the data correctly as frequently as possible. We use machine learning techniques and logistic regression for predictions. The study is on three different outcome variables, with the first one being a binary outcome variable for a home team win and then two multinomial variables with three and five possible outcomes. The data contains information on all matches played between the 2014/15 season to the 2018/19 season. We find that the logistic regression and the SVM classifiers have the highest predictability and perform the most consistently for all three response models. The classification trees provide decent results predicting but do not perform quite at the same rate. The results regarding the importance of the match variables is that the expected goals is the most important for result predictions while the shots on goal, red cards, clearance and ranking variables also show importance in the predictions and improve the predictions significantly.

^{*}Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: alexcrompton@hotmail.se. Supervisor: Dmitry Otryakhin.