

Predicting Problem Gambling: An Analysis of Ladbrokes and Coral Self-Excluders

Alexander Szoke - MSc Business Analytics, University College London
 Academic Supervisor: Deyu Ming
 Entain Supervisor: Disa Ramadinha



Introduction

Background

As the popularity of online gambling has grown, so too has the public health concern for gambling addiction. As a result, gambling operators have been required to implement Responsible Gambling (RG) tools to provide better player protection and increase safer gambling awareness.

One of the most extreme RG tools is self-exclusion, a programme where gamblers who recognise that they have a gambling problem can ask the operator to deactivate their account for a pre-set period, typically 6 months.

The ability to predict self-exclusion is beneficial for gambling operators because:

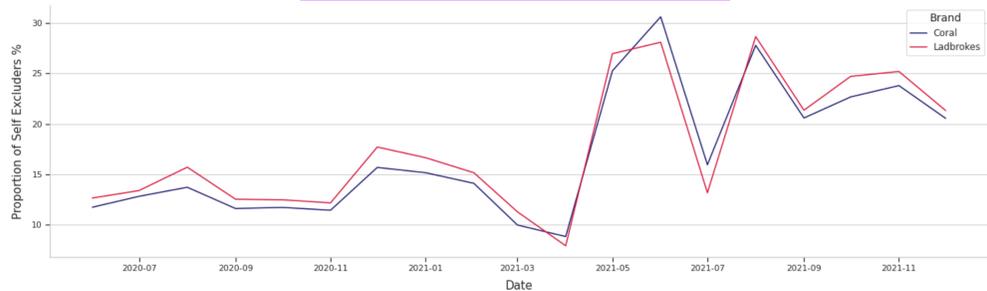
1. High-risk players can be pre-emptively managed and protected
2. Players will interact with Entain's products in a more financially sustainable way, stabilising long-term revenues.

Objective

Using the gambling activity of a sample of Ladbrokes and Coral Sports bettors, the objectives of this study were to:

1. Explore the difference in the gambling behaviour between Self-Excluders and control gamblers (i.e., those who did not self-exclude)
2. Develop binary classification models, capable of predicting self-exclusion
3. Explore how the performance of these models varied based on the length of the observational period on which the models were trained

Monthly Self-Excluders as a Proportion of new Coral and Ladbrokes customers (2021 – 2021)



Data Overview

Data Origin

Almost all data used was sourced from Entain's database. The final data set summarised the 2021 sports betting activity of 60,000 UK-based Ladbrokes and Coral players who played sportsbook for the first time in 2021 with real money at least once.

Variable Overview

The variables created could be grouped into 5 categories:

1. **Player characteristics:** Included age, gender, brand and geospatial features
2. **Bet Behaviour:** Included summary statistics for all bets placed e.g., number of bets, number of distinct days played, Std. of bet value (£)
3. **Bet Type:** Summarised the risk level of bets placed e.g., average odds, percentage of accumulator bets
4. **Transaction Behaviour:** Included summary statistics for all deposits and withdrawals made
5. **Trajectory variables:** Linear regression coefficients capturing the trajectory of bet behaviour over time

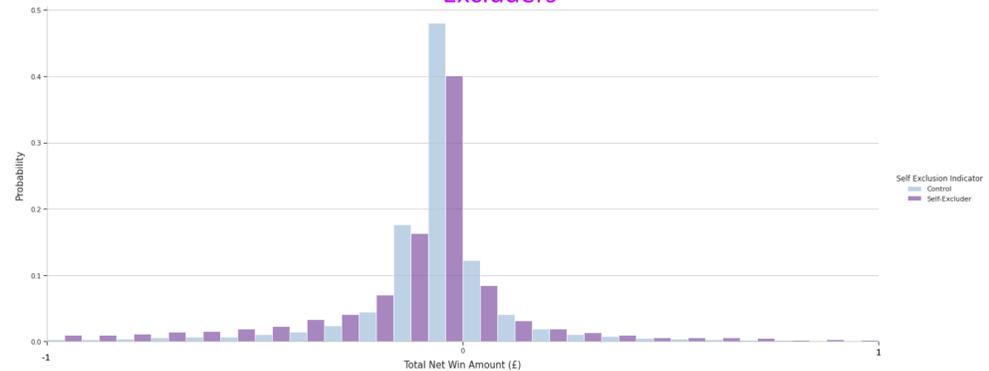
Mean Variable Value for Controls vs. Self-Excluders

Variable	Controls (n=46,875)	Self-Excluders (n=14,844)
Age	38.64	35.79
Tenure	2.68	3.24
Number of Entain-affiliated accounts	1.81	2.43
Number of days played	9.39	10.26
Frequency of days played	0.63	0.67
Total Net Win %	-0.30	-0.33
Average Deposit : Average Withdrawal (£) ratio	0.53	0.35
Deposit Frequency	0.73	0.84
Live bet Frequency	0.18	0.36
Accumulator bet Frequency	0.29	0.53

Exploratory Data Analysis

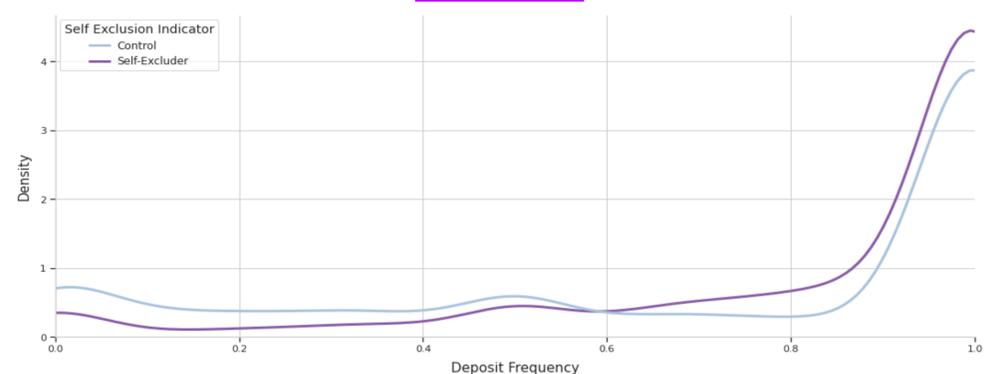
Both classes of players tended to be losers over the observation period. However, Self-Excluders tended to be the bigger losers.

Probability Distribution of Net Win Amount (£) for Controls and Self-Excluders

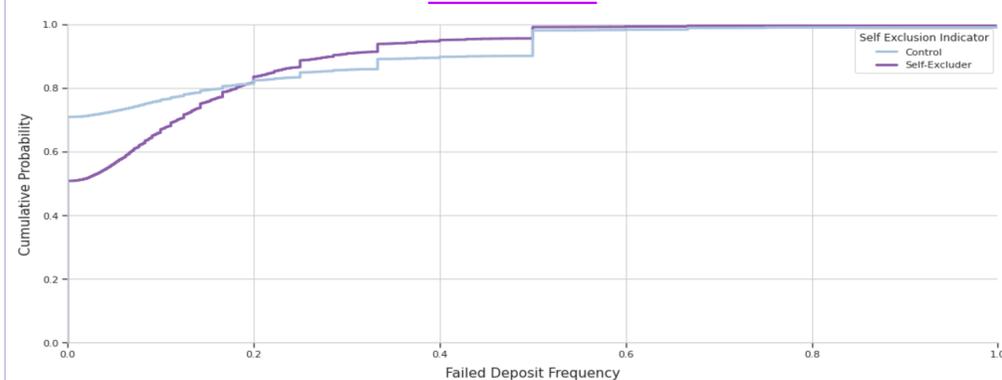


Self-Excluders deposited more often but also failed their deposits more often (due to lack of funds etc.)

Kernel Density Estimate of Deposit Frequency for Control Gamblers and Self-Excluders



Cumulative Probability of Failed Deposit Frequency for Control Gamblers and Self-Excluders



Modelling and Results

Modeling

A set of 12 different supervised learning algorithms were developed to determine which was most effective in predicting self-exclusion. After hyperparameter tuning, the optimal model was deemed to be an ensemble Voting Classifier consisting of a KNN classifier, a Gradient Boost classifier, a Random Forest classifier and a Neural Network.

The final model was able to correctly classify 73% of players using only a two-week observation period of data, surpassing any test accuracy seen in prior literature.

Performance of Final Voting Classifier on Varying Observation Period

