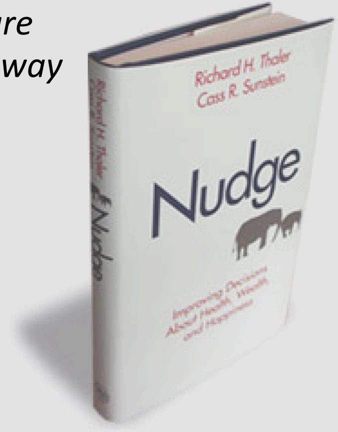
A Nudge

Evaluation Insights for Betfair

"A nudge is any aspect of the choice architecture that alters people's behaviour in a predictable way without forbidding any options."

-Richard Thaler
Professor of Behavioural Economics
University of Chicago



Big picture

The World’s Biggest Betting Community

- Betfair is an online ‘marketplace’

- Match odds, bet sizes are unknown *a priori*

- ‘Nudging’ user behaviour can be useful

Big picture

betfair

The World's Biggest Betting Community

- Betfair is an online 'marketplace'

- Match odds, bet sizes are unknown *a priori*

- 'Nudging' user behaviour can be useful

Examples of 'nudges'

- Informing (passive) users that their bets are likely to lapse
- Giving options to (active) users who want to cancel their bets
- Using a user skill database to incentivize selected (skilled) users
- All of the above need predictive capabilities

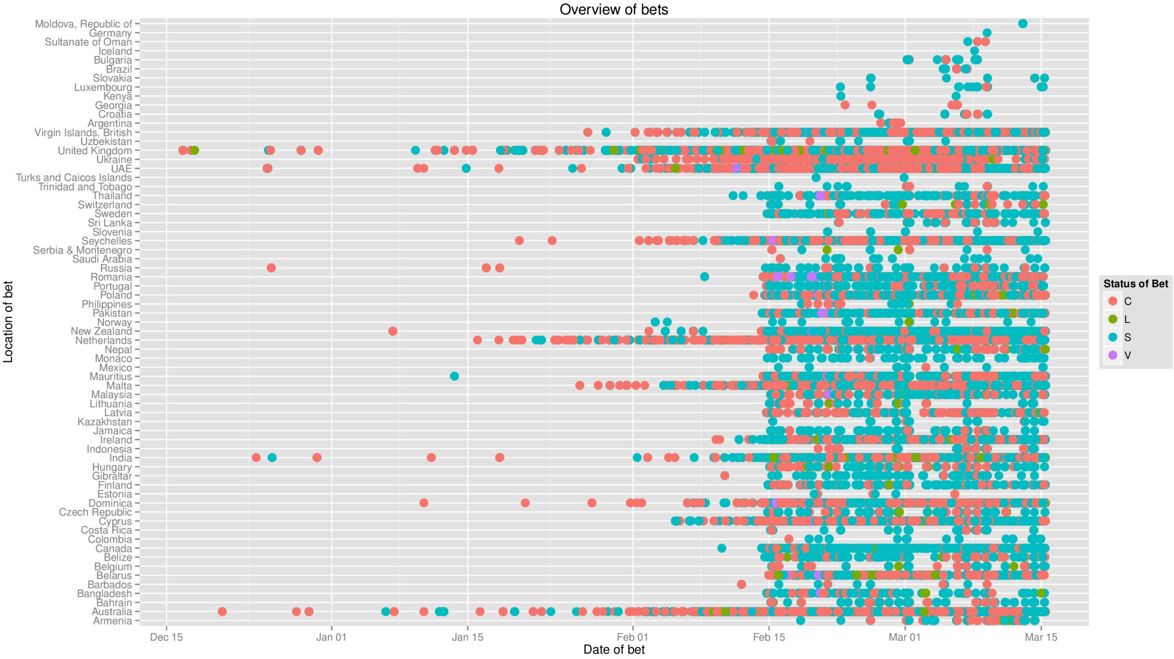
Two key questions:

- What is the probability of any given bet being cancelled or lapsed?
- What is the probability of an individual user winning a transaction?

The goal of our analysis is to develop models to answer these questions.

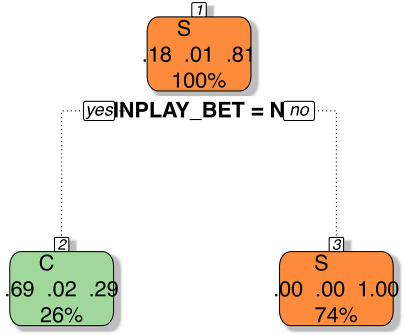
Data Analysis

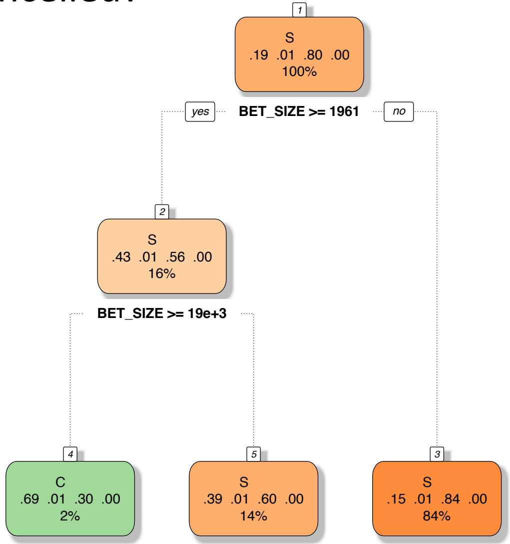
Problem: Way too much data!



Overview of settled bets with bet size and profit/loss
Location of bet
Moldova, Republic of
Germany
Sultanate of Oman
Iceland
Bulgaria
Brazil
Slovakia
Luxembourg
Kenya
Georgia
Croatia
Argentina
Virgin Islands, British
Uzbekistan
United Kingdom
Ukraine
UAE
Turks and Caicos Islands
Trinidad and Tobago
Thailand
Switzerland
Sweden
Sri Lanka
Slovenia
Seychelles
Serbia & Montenegro
Saudi Arabia
Russia
Romania
Portugal
Poland
Philippines
Pakistan
Norway
New Zealand
Netherlands
Nepal
Monaco
Mexico
Mauritius
Malta
Malaysia
Lithuania
Latvia
Kazakhstan
Jamaica
Ireland
Indonesia
India
Hungary
Gibraltar
Finland
Estonia
Dominica
Czech Republic
Cyprus
Costa Rica
Colombia
Canada
Belize
Belgium
Belarus
Barbados
Bangladesh
Bahrain
Australia
Armenia
Dec 15
Jan 01
Jan 15
Feb 01
Feb 15
Mar 01
Mar 15
Date of bet
Winnings
<−50000
−50000−0
0−50000
50000>
Size of bet
0−100
100−1000
1000−10000
10000+
Status of Bet
C
L
S

Which bets are likely to be cancelled?





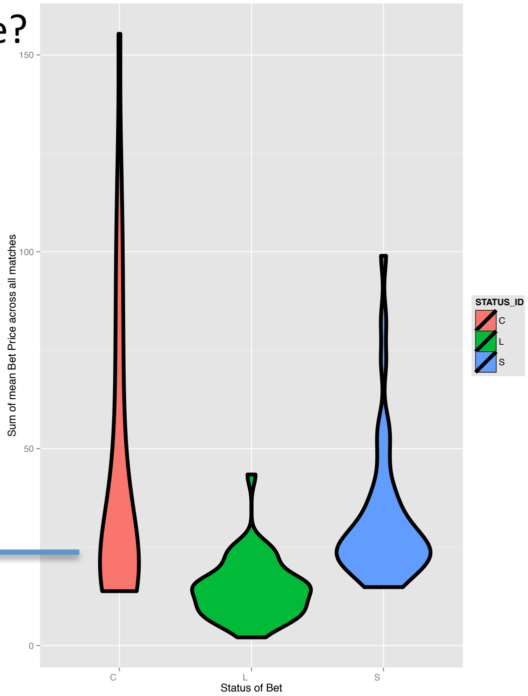
C = Cancelled L = Lapsed S = Settled V = Void

Which bets are likely to lapse?

<u>Aggregation Approach</u>

- Mean bet price/match
- Summed over all matches

Thickness indicates frequency



- Cancelled bets correlate with bet size and whether they are made ‘in-play’.
- Lapsed bets correlate with mean bet price.
- Possible confounders?

- Cancelled bets correlate with bet size and whether they are made ‘in-play’.

- Lapsed bets correlate with mean bet price.

- Possible confounders?

<u>Second approach – look at users</u>

- Separated users into ‘Regular’ (Total bets < 1 million AUD), ‘Power’ (Total bets < 10 million AUD) and ‘Super’ (Total bets > 10 million AUD) categories.

- ~18000 Regulars (R), ~1650 Power (P) and ~350 Super (S) users identified.

- Cancelled bets correlate with bet size and whether they are made ‘in-play’.

- Lapsed bets correlate with mean bet price.

- Possible confounders?

<u>Second approach – look at users</u>

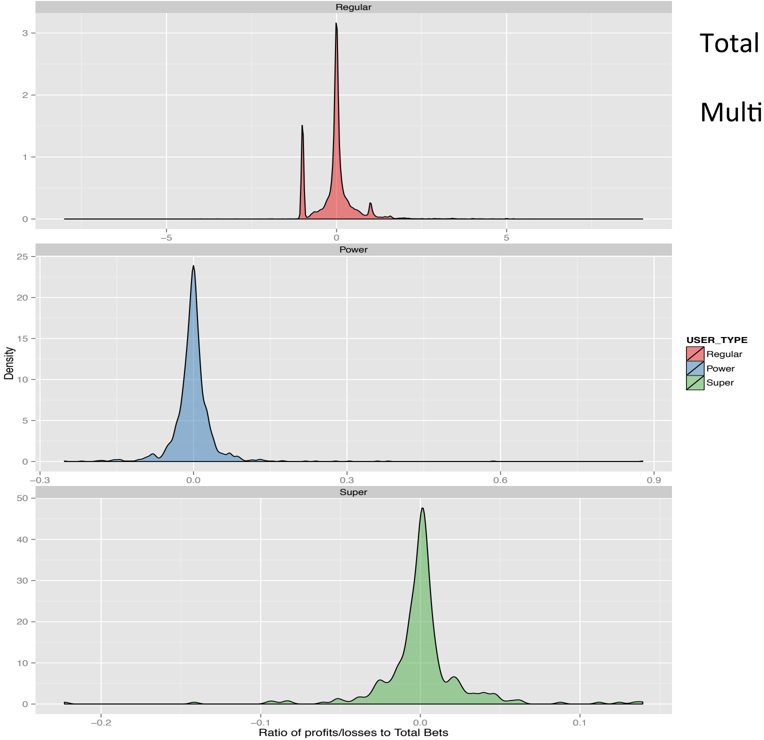
- Separated users into ‘Regular’ (Total bets < 1 million AUD), ‘Power’ (Total bets < 10 million AUD) and ‘Super’ (Total bets > 10 million AUD) categories.

- ~18000 Regulars (R), ~1650 Power (P) and ~350 Super (S) users identified.

- Key findings - Mean Bet sizes per user– ~370 (R), ~1250 (P) and ~2700 (S) (all AUD)

- Mean transactions per user – ~50 (R), ~500 (P) and ~4450 (S) !

- Big players bet far more regularly – but are they better?



Total Profits/losses divided by total bets

Multiple modes = Lost everything

'Power' users do better than 'regular' ones

'Super' users do the best
- Right-shifted curve

Trickle up effect?


Building Classifiers

- # Model 1 – Lapsed/cancelled bet classifier

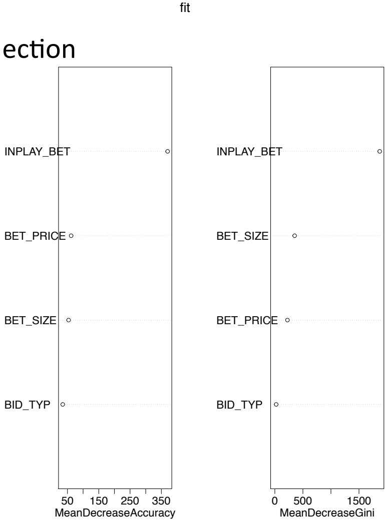
Methodology:

- Separated first 40 match (train) and QF data (test)
- Used Random Forest algorithm along with feature selection

• Model 1 – Lapsed/cancelled bet classifier

Methodology:

- Separated first 40 match (train) and QF data (test)
- Used Random Forest algorithm along with feature selection



Results:

- 92.47 % accuracy (CI: 91.94 – 92.98)
- High sensitivity and specificity for cancelled bets.

Limitations

- Poor performance with predicting lapsed bets
- Feature selection can be improved

- Model 2 – Individual user win/lose classifier

<u>Methodology</u>:

- Separated data into blocks, employed time series approach to each user
- Comprehensive feature engineering introduced including cumulative wins at sample, prior average bet size, prior profit/loss, number of prior bets etc
- Used a logistic regression classifier to predict win/loss

- Model 2 – Individual user win/lose classifier

<u>Methodology</u>:

- Separated data into blocks, employed time series approach to each user
- Comprehensive feature engineering introduced including cumulative wins at sample, prior average bet size, prior profit/loss, number of prior bets etc
- Used a logistic regression classifier to predict win/loss

<u>Results</u>:

- It performs reasonably well.

<u>Limitations</u>:

- Future plans to try alternative approaches including random forest and neural networks to improve accuracy

Conclusions and Future Work

Conclusions

- Suggested mechanism - ‘nudging’ customers
- Requires predictive analytics
- Feature design for classification was informed by extensive data analysis
- Two classifier models were developed for bet status and individual user win/loss
- The approaches show promise