Using data and predictive analytics in the online gaming industry

Catherine Wilks, PwC UK
Agenda

• The building blocks of predictive analytics
• Proof of concept at UK gaming client
• Using industry held data to minimise harm in remote gambling
The analytics maturity curve
Is predictive analytics suitable for all organisations?

- **Level 1 limited**: Data inconsistency, little centralisation, spreadsheet based
- **Level 2 evolving**: Departmental approach, local governance, little ownership
- **Level 3 functional excellence**: Data management, defined data strategy, central approach
- **Level 4 integrated excellence**: MI strategy driving business performance, single view of data
- **Level 5 information premium**: Timely analysis using core and external datasets, data-driven innovation

**Evolution**

**Low maturity** → **Premium maturity**
Does it align with your strategy?

**Strategy and Vision** – what do we want to achieve and how will we measure success? Defining a clear set of goals and setting the vision for Business Intelligence is key for a consistent and improved use of data across the organisation.

**Operating Model** – how should we structure ourselves to achieve success? As part of the strategy, we provide a high level target operating model. This uses our operating model framework which encompasses nine key components such as scope, policies and tools.

**People and Culture** – how do our people and culture underpin the strategy? How do you grow a team of data professionals with the range of skills and experience needed for the future - the strategy will drive training and recruitment needs.

**Processes, Policies and Standards** – how should we manage and use data? As part of the data strategy we will help define a high level framework for processes, policies and standards for managing and using data, and therefore a clear methodology for these for the future.

**Technology and Tools** – what tools do we need to achieve the best results from our data? Recognising the existing technology landscape has a bearing on how effectively data can be used within the organisation. As part of the data strategy, we map out the target technology architecture and integration with the existing landscape. We can then go on to support the selection of the right tools for the business.
Do you have a specific use case and data to support?

- If the use case is not clearly defined and linked to a key business driver then you could be doing analytics for analytics sake.
- Do you fully understand the data and is it of sufficient quality?
- The use case needs to be specific and have a recognised, agreed output and intervention.
- Defining the use case and obtaining the data to support is an iterative process – would you consider a less critical use case which has better quality data to support?
What will you do with the output?

- Predictive analytics is not a cheap investment. Will the benefits outweigh the costs?
- Do you have appropriate technical and business resources to deliver?
- Will you act on the output and stage appropriate intervention? Have you factored the costs of this into your cost benefit analysis?
- Consider a well defined proof of concept before rolling out further.
Predictive analytics fundamentals
Our client is an online gaming client. Can a Proof of Concept prove the value of predictive analytics?
Establishing the use case

**Risk Factoring**
Predict the risk associated with a customer based on their expected profit and loss and begin to drive commercial effectiveness by anticipating winnings given the operator’s tools to manage their exposure.
*Output: a predicted profit and loss figure for each customer*

**Cross Sales Penetration**
Understand the probability around a customer’s potential to purchase other gaming products, as well as the ideal time to cross sell to maximise profitability.
*Output: probability of likelihood to purchase product*

**Customer Churn**
Predicting the likelihood of a customer closing their account or suspending their services, prompting action of a targeted marketing campaign to minimise churn.
*Output: probability of customer closing account*

**Bonus Effectiveness**
Predicting, on an individual customer basis, the optimum time to send relevant marketing messages, specifically around bonuses or promotions, to maximise uptake based on campaigns.
*Output: percentage or probability of uptake on promotional offer*
Establishing the use case

**Lifetime Value**
Understanding, early on in a customer’s lifecycle, how much a customer can contribute to the organisation over all, enabling a ‘cost and benefit’ analysis of acquisition costs.
*Output: profit and loss figure from organisation’s perspective*

**Responsible Gambling**
Being able to predict customers who have the potential to become dependant on gambling before they actually do is crucial in being a responsible gaming organisation. Once identified, steps can be taken to protect the most vulnerable customers.
*Output: probability percentage of customer’s likelihood of problem gambling*

**Fraud Detection**
Understanding the potential that a customer will commit fraud in the future, for example through fraudulent use of credit cards.
*Output: probability percentage of customer’s likelihood of fraudulent activity*
**Identifying the data sources**

**Risk Factoring**
- Historical customer transaction listing
- Customer profile information
- Betting markets history information
- Customer registration data
- Marketing campaign data

**Cross Sales Penetration**
- Historical customer transaction listing
- Customer activity across books
- Product information

**External Sources**
- Events data
- Weather data

**Customer Churn**
- Historical customer transaction listing
- Customer data on closed accounts

**Bonus Effectiveness**
- Historical customer transaction listing
- Historical bonus transactions
- Customer profile information

**Lifetime Value**
- Historical customer transaction listing
- Customer registration data
- Customer profile information
# Define the approach

## Predictive analytics engagement

<table>
<thead>
<tr>
<th>Step</th>
<th>Detail</th>
</tr>
</thead>
<tbody>
<tr>
<td>Requirements gathering</td>
<td>Define the exact requirements for the engagement which all parties sign up to.</td>
</tr>
<tr>
<td>Data definition</td>
<td>Define exactly what fields are to be extracted from source systems to feed into the predictive model.</td>
</tr>
<tr>
<td>Create data set and load</td>
<td>Create reusable extraction and transformation scripts to transform data ready for import into the predictive models.</td>
</tr>
<tr>
<td>Build predictive models</td>
<td>Build predictive models based on business question. Optimise and train models.</td>
</tr>
<tr>
<td>Run analysis</td>
<td>Execute analysis based on predictive model results. Where relevant, define key features of the model which impact the results.</td>
</tr>
<tr>
<td>Analyse and disseminate results</td>
<td>Socialise analysis with the organisation. Work with organisation to analyse the findings and agree next steps.</td>
</tr>
<tr>
<td>Embed strategy</td>
<td>Define strategies to address business question. Work with the organisation to integrate predictive analytics in the ‘Business as Usual’ processes.</td>
</tr>
</tbody>
</table>
The proof of concept

- Risk factoring use case.
- Active trading team who monitor bets and apply limits to events and customers.
- Current model identifies users based on known unusual behaviour:
  - Account address which has been previously excluded
  - Initial bet over a threshold
  - Early bettors
  - Consistently beating the odds
- The identification took up to 4 weeks.
Technical solution

Data from client systems → Neuro-Bayes neural network → Predictive model outputs

**Pre-Processing**
- Data loaded into database
- Data quality is assessed and assured
- Initial linear modelling takes place

**Network Training**
- Non-linear correlations learnt
- Black Swan events understood and eradicated
- Over fitting avoided

**Post-Processing**
- The individual probability density functions are defined
- Results qualified for accuracy and appropriateness

Results can be fed back in real time to client systems
Data from client systems

• All online customers who had placed sports bets in 2013 and 2014.
• Sources were:
  • Customer transactions, including deposits and withdrawals
  • Customer data
  • Sports events data
  • Accounts top ups
Neuro-Bayes neural network

- Data cleansing required on customer data due to self service.
- Reformatting required on all data sources.
- 2013 data used as a training set for 2014 predictions.
- The model took two weeks to develop and once productionised, could run a years’ worth of data in a few hours.
**Predictive model outputs**

- The output was the predicted Profit and Loss of a single customer:
  - 6 months from registration (i.e. with no bets)
  - 6 months from 1 day after first bet (i.e. with 1 bet)
  - 6 months from 7 days after first bet (i.e. with > 1 bet)
- Model predicted a 30% uplift in revenue from a 10% exclusion of customers, or 19% uplift in revenue from a 4% exclusion.
- There was both an intersection and difference between the results from the client’s existing model and our predictive model.
Using industry held data to minimise harm in remote gambling

- Project commissioned by Gamble Aware, led by Andy MacGilp and Nigel Issa of PwC working alongside the Responsible Gambling Council of Canada
- Inform practical applications of harm minimisation for remote gambling operators serving British customers. The emphasis is on how harmful and risky behaviour can be mitigated online, not just if it can be identified and mitigated online.
- A particular priority of the research is to explore the usefulness of industry-held data and behavioural analytics in delivering these research aims.
- Phase II results will be published imminently and presented at upcoming industry conferences.
Data analytics performed

Operators provided two years’ of data (bets, deposits, withdrawals, account details) for all online gaming excluding poker.

1. Data collection and cleansing
2. Clustering
3. Establish markers

- 200 data attributes across 10,600 customers giving 1.5bn data points
- 9 clusters of customers with similar betting behaviours, for example high volume gaming or high stakes sports

Demographic (age, gender, job status) and Behavioural (bet volume, deposit frequency) identified through regression. Daily triggers (in the moment reactions to wins and losses) through pattern recognition

Risk scoring model
**Key learnings**

One size doesn’t fit all! To hone in on the most influential markers, ensure the data is clustered prior to analysis.

The resulting risk thresholds need to be set appropriately, and in line with the business strategy.

<table>
<thead>
<tr>
<th>e.g.</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision</td>
<td>63%</td>
<td>71%</td>
<td>86%</td>
</tr>
<tr>
<td>Hit rate</td>
<td>81%</td>
<td>67%</td>
<td>26%</td>
</tr>
</tbody>
</table>
Thank you

Cathy Wilks
Tel: 07730147787
Email: catherine.j.wilks@pwc.com