Using Text Mining & ML to Enhance the Credit Risk Assessment Process

Bruce Brenkus
Chief Risk Officer in Underwriting, Risk & Recoveries
11th of October
What is Spotcap?

An innovative on-balance sheet lender

Offers versatile and accessible finance to SMEs

Has developed a unique, sophisticated, proprietary credit algorithm

Assesses real-time business performance next to traditional credit scores
Traditional credit assessment practices are disconnected from the modern world

**Traditional credit assessments**

- Based on credit scores & backward looking information
- Often use outdated financial data
- Can take up to several weeks

**Today’s balance sheet lenders must:**

- Have an in-depth understanding of customers to remain competitive
- Meet complex regulatory requirements
- Deal with a wealth of data about their customers
- Make decisions faster than ever
- Safeguard themselves against fraud & defaults
Modern lenders need to balance business needs with speed.
Spotcap has turned tradition on its head

MAIN COMPONENTS OF RISK ASSESSMENT (PROBABILITY OF DEFAULT MODELS) FOCUS ON:

- Borrower profile
- Bank account behaviour
- Business profile + financials

MODELS LEVERAGE ML & DECISION SCIENCE TO DELIVER SMARTER RISK ASSESSMENT

- Rating Models (SpoRT, incl. Advanced Bank Analytica Score)
- Limit Determination (ALCo)
- Decision Algorithm (DAL)

Spotcap Risk Technology - Rating Model
ALCo Auto Limit Computation Algorithm
DAL Decision & Approval Logic
Real-time, high-quality data sits at the center of our credit assessment process

**QUANTITATIVE**
- Fact based
- Fixed, measurable
- Numeric in design
- Credit reports
- Financial statements
- Accounting software
- Tax Data/Interims
- Bank account
- Shipments data

**QUALITATIVE**
- Understanding based
- Human behavior
- Text, written word
- Photo, sound
- Annual report written sections
- Adverse media
- Bank account transaction level text
- Questionnaires (In App)

Unstructured to Structured
As does finding the right balance between automation & that human touch

**AUTOMATION**
- Data depth and breadth
- Sufficient to make sound risk decision
- Size of transaction
- Quality of business
- Size/quality Interaction
- Unsecured loans

**HUMAN INTERACTION**
- More involved transactions
- Size of transaction
- Company type
- Market/economic attributes
- Human expertise
- Security interest
- Machine based pre-decisions
- Machine based conditions to clear
- Override controls
- Enable your team
Data understanding and preparation is key

PROMOTE BEST PRACTICES
Promotion of R and data science within the company through developing ‘spotcapR’ - an internal R package to foster DS knowledge, as well as optimize and speed-up data science project workflow.

FOCUS ON REPRODUCIBILITY
Every project and analysis is designed to be easily refreshed and repeated at any point in time.

SHARING KNOWLEDGE
Insights and ideas are shared with respective team members in an interactive and appealing way through RMarkdown reports.
Understanding the data science lifecycle

Business Idea

Need for a swift proof-of-concept and feasibility assessment

Classic Data Science Project Lifecycle

Data Understanding
- Data sources
- Data frequency & environment
- Data structure

Data Preparation
- Data modelling
- Data exploration
- Feature engineering

Deployment
- Productionizing selected model
- Monitoring and evaluation

Evaluation
- CV, holdout tests
- Selecting strongest models
- Accuracy vs. speed

Modelling
- Testing many model types
- Optimal parameters search
Spotcap has made the data science project lifecycle more efficient

**Business Idea**

Need for a swift proof-of-concept and feasibility assessment

**Data Understanding**
- Data sources
- Data frequency & environment
- Data structure

**Data Preparation**
- Data modelling
- Data exploration
- Feature engineering

**Deployment**
- Productionizing selected model
- Monitoring and evaluation

**Evaluation**
- CV, holdout tests
- Selecting strongest models
- Accuracy vs. speed

**Modelling**
- Testing many model types
- Optimal parameters search

Platform enables Spotcap to significantly reduce modelling, evaluation and deployment times, and allows us to focus on what’s most important - building data knowledge and generating ideas.
ML models can add value, but only when you know your data

- **KYC** - first and foremost
- Where do you start (Evolution)?
  - **Expert system**
  - Rules engine
  - Quantitative models
- Balancing **energy and results**
- Build a simple data laboratory
  - **Know the data** (Expertise)
  - Exploratory data analysis
  - Simple models to determine value
- Move the **value to production**
- Explore the remaining
- Cut and run
- **Results, Results, Results**
  - Monitor the portfolio
  - Track the models
  - Score & attribute migration
Our risk assessment utilizes hundreds of data points

Data processed and applied:

<table>
<thead>
<tr>
<th>CREDIT BUREAU</th>
<th>TRADITIONAL FINANCIALS</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Not part of the main performance model except for general borrower characteristics</td>
<td>• App Lite (&lt;50k amounts): no requirements to submit documents</td>
</tr>
<tr>
<td>• Credit Score and adverse information used for warning signals / hard declines</td>
<td>• Standard Application (&gt;50k amounts): annual accounts and VAT returns to derive standard financial ratios and trends</td>
</tr>
<tr>
<td>• Primary use: verification</td>
<td>• Relative operating surplus used to determine eventual credit amount (combined with BA)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TRANSACTIONS (BANK ACCOUNT)</th>
<th>USER / WEB</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Main source for probability of default model</td>
<td>• Financial and non-financial information asked to verify data from 3rd parties</td>
</tr>
<tr>
<td>• Transactional data used from bank account connections (proprietary) or parsed from bank statements provided</td>
<td>• Web search (for larger tickets) for more holistic assessment also using web automation tools (RPA)</td>
</tr>
<tr>
<td>• 200+ attributes used for algorithm; variety of semantic checks for warning signals and profiling</td>
<td></td>
</tr>
</tbody>
</table>
The Bank Account 1st Gen model offers a highly predictive picture of default probability

The Bank Account 1st Generation Model is specified from over 200 numerical variables that can be organized into three different groups that constitute together a comprehensive and highly predictive picture of default probability in the next 12 months.

Bank account general characteristics
Set of variables that provide essential information about the number of transactions, their volume and frequency.

Relative bank account activity indicators
Using relative figures can improve model performance significantly.

Bank account balance development & trends
Changes in the bank account activity or the volume of transactions could be vital for client’s ability to repay.
2nd Gen Bank Account model significantly improved credit performance …

KEY OBSERVATIONS

- Historical loan cohorts with slightly higher credit losses especially around D credit grade

- Simulation and application of 1st generation bank account (BA) model for 2017 cohorts onwards

- H1 2017 cohort credit performance already reveals significant improvements while maintaining conducive approval rate levels

Credit losses before/after BA model application
Percent of loan volume issues, Q1 15 – Q1 17 cohorts*

<table>
<thead>
<tr>
<th></th>
<th>Projected</th>
<th>Actual</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>0.9</td>
<td>1.7</td>
</tr>
<tr>
<td>C</td>
<td>3.4</td>
<td>1.5</td>
</tr>
<tr>
<td>D</td>
<td>4.2</td>
<td>2.5</td>
</tr>
<tr>
<td>E</td>
<td>6.8</td>
<td>6.8</td>
</tr>
</tbody>
</table>

SPoRT Credit Model
Credit Loss: 2.6%

SPoRT Credit & BA Model 2.0
Expected Credit Loss: 1.4%

* Percent of loan volume issues, Q1 15 – Q1 17 cohorts
Our models have and will continue to evolve

**Start Model**
- Q3 2014
  - Base
  - Multi-variables model
  - Expert System

**SpoRT 1st Generation**
- Q2 2015
  - Basel IRB model structure
    + Bank account algorithm
    + Merchant account profile algorithm
  - Expert calibrated

**SpoRT 2nd Generation**
- Q1 2017
  - Logistic Regression Models
    + Advanced Analytics & Machine Learning
    + Semantic Analysis (Text Mining)

**SUCCESS**
- Q4 2017
  - Robust Performance History
  - Profitability & CLV Algorithms
  - Data Attributes Expansion (Text Mining)
  - Expanded Auto Rejections & Approvals
That evolution has had a significant business impact

**SpoRT** models are regularly monitored and further developed in order to maximize their predictive capabilities, as well as reflect Spotcap’s current risk appetite and application processes.

- **SpoRT 1.0**
- **SpoRT 2.0**
- **SpoRT 2.0 & Bank Model**

SpoRT 2.0 & Bank Model outperforms other models in lowest and highest deciles, which shows that it has the highest “good-bad” differentiation potential.

Identification of clients in this area will further improve in SpoRT next generations.
Thank you.

bruce.brenkus@spotcap.com
@spotcapglobal
spotcap-global
spotcap-global-services

© Spotcap, 2018