Forecasting Bank License Revocation

Jacob Bortell
Jack Harding
Mike Giancola
Parmenion Patias
Acknowledgements

- Dr. Jaroslav Bologov
- Dr. Alexey Kozionov
- Dr. Oleg Pavlov
- Dr. Thomas Balistrieri
Outline

1. Background
2. Methodology
3. Results & Analysis
The Problem

- Bank licenses by Central Bank of Russia
- Difficult to forecast license revocation
- Affects general public
2013 CBR Reforms
- 2013 CBR Reforms
- 2014 Oil Crisis

Photo taken from: http://www.macrotrends.net/1369/crude-oil-price-history-chart
- 2013 CBR Reforms
- 2014 Oil Crisis
- Russian Wages
The Goal

Forecast the likelihood of license revocation
Central Bank of Russia (CBR)

- Banking Standards
  - N1 - Capital Adequacy Ratio
  - N2 - Instant Liquidity Ratio
  - N3 - Short-Term Liquidity Ratio
## Dataset Structure

<table>
<thead>
<tr>
<th>Lic Num</th>
<th>Date</th>
<th>N1</th>
<th>N2</th>
<th>N3</th>
<th>Quarters</th>
</tr>
</thead>
<tbody>
<tr>
<td>127</td>
<td>2015-10-1</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>3</td>
</tr>
<tr>
<td>2288</td>
<td>2014-3-1</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>&gt; 2 years</td>
</tr>
<tr>
<td>564</td>
<td>2015-7-1</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>Active</td>
</tr>
</tbody>
</table>
Building Models

● Build models which
  ○ Analyze financial data
  ○ Learn from data to make predictions

● Two Types:
  ○ Statistical: Logistic Regression
  ○ Machine Learning: Random Forest
- **Split data**
  - **Learn** from $\frac{2}{3}$ of data
    - | Lic. Num | Date     | N1 | N2 | N3 | --- |
      |-------|---------|---|---|---|-----|
      | 127   | 2016-1-1| 13| 0.4| 5 | ... |
      | 2288  | 2014-3-1| 11| 0.5| 2 | ... |
      |       |         |   |    |   | 3   |
      |       |         |   |    |   | > 2 years |
  - **Predict** from $\frac{1}{3}$ of data
    - | Lic. Num | Date     | N1 | N2 | N3 | --- |
      |-------|---------|---|---|---|-----|
      | 564   | 2015-7-1| 14| 0.6| 4 | ... |
      |       |         |   |    |   | ?   |
- Compare predictions to known values
Performance Metrics

- Relevant elements
- False negatives
- True negatives
- True positives
- False positives

Selected elements
Performance Metrics

Precision = \frac{5}{8} = \frac{5}{12}

Recall = \frac{5}{12}
Performance Metrics

F1 = \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 2

F1 = \frac{\frac{5}{8} \times \frac{5}{12}}{\frac{5}{8} + \frac{5}{12}} \times 2

F1 = \frac{1}{2} = 50\%
Results

- Dataset Composition
- Performance Based on F1
- Significance of Features
- Recommendations
- “Still Active” banks majority of dataset
- First 8 quarters only 10% of dataset
- Regression outperformed by guessing
- Random Forest outperforms Regression
Significant Factors
Significant Factors

- Loans to Businesses and Institutions (6 months to 1 year) 11%
- Loans to Businesses and Institutions (1 to 3 Years) 11%
- Loans to Businesses and Institutions (up to 6 Months) 10%
Recommendations

- Use Random Forest Model
- Research Clustering
- More Features
  - Risk Related
  - Socio-political
- Fewer Classes / More Granular Data

- Benefits
  - Banks
  - Investors
  - General Public