A robust machine learning approach for credit risk analysis of large loan level datasets using deep learning and extreme gradient boosting

9th biennial IFC Conference on "Are post-crisis statistical initiatives completed?"

Basel, 30-31 August 2018
The views expressed in this paper are those of the authors and not necessarily those of Bank of Greece.
Credit Risk Analysis Tool
In a nutshell

Novel Rating System for Greek banks business portfolios

Modeling technique
- Extreme Gradient Boosting
- Deep Neural Networks

Main Drivers
- Company Financial Ratios
- Macroeconomic factors

Implementation
- Corporate and SME loans of the Greek banking system, from the supervisory database of the Central Bank of Greece

Proactive monitoring
Economic assessment
Credit Risk Analysis
Machine and Deep learning techniques

- “Learn” without being explicitly programmed
- Unveiling new determinants and unexpected forms of dependencies among variables.
- Tackling non linear relationships.

- Use of ML and Deep Learning are favored by the technological advances and the availability of financial sector data.
- Supervisory authorities should keep up with the current developments.
Credit Risk Analysis
Bank of Greece – Regulatory Purpose

Exploit loan level regulatory data → Customer classification → Targeted regulatory measures

Early Warning System

Good Customer → Measure bank credit risk for addressing vulnerabilities

Bad Customer
AnaCredit will be a new dataset with detailed information on individual bank loans in the euro area.

The project was initiated in 2011 and data collection is scheduled to start in September 2018.
Multiple subsets are created from the original dataset, selecting observations with replacement and a base model (weak model) is created on each of these subsets.

- The models run in parallel and are independent of each other.
- The final predictions are determined by combining the predictions from all the models.
- Random Forests are common employed bagging techniques.

Credit Risk Analysis
Bagging – Different models vote for the result

source: Analytics Vidhya
Credit Risk Analysis
Boosting – Each model learns from the errors of the previous

- A base model is created based on a subset of the original dataset which is used to make predictions on the whole dataset.
- Errors are calculated and observations which are incorrectly predicted, are given higher weights (large plus signs).
- Another model is created which tries to correct the errors from the previous model.
- Similarly, multiple models are created, each correcting the errors of the previous model.
- The final model (strong learner) is the weighted mean of all the models (weak learners).

source: Analytics Vidhya
Credit Risk Analysis
Deep Neural Networks-Unlimited potential for Architectures

Deep neural network is simply a feedforward network with many hidden layers. It has the following advantages compared to one layer networks (“shallow”)

• A deep network needs less neurons than a shallow one
• A shallow network is more difficult to train with our current algorithms (e.g. it has more nasty local minima, or the convergence rate is slower)
Credit Risk Analysis
Deep Neural Networks-Unlimited potential for Architectures

This methodology provides the opportunity of creating a large combination of different structures based on:

- Number of layers,
- Selection of activation function
- Number of perceptrons
- Normalization layers
- Dropout adjustments

Which can be employed in the optimization process.
Credit Risk Analysis
Problem at hand

- We have collected loan level information on Corporate and SME loans of the Greek banking system, from the supervisory database of the Central Bank of Greece.
- A loan is flagged as delinquent if it is either 90 days past due or it gets rated as delinquent based on each bank’s internal rating rules.
- The forecast horizon for a default event is 1 year whereas the variables employed include macro data and company specific financial ratios.
Credit Risk Analysis
Many Predictor Candidates - Curse of dimensionality

- We employ **Boruta algorithm** for tackling the dimensionality issue. This is sequential Random Forest based algorithm which removes non relevant variables decreasing the dimensionality space.
Credit Risk Analysis
Many Predictor Candidates - Curse of dimensionality

Boruta Algorithm – steps:

- First, it adds randomness to the given dataset by creating shuffled copies of all features (shadow features).
- Then, it fits a Random Forest model (bagging model) on the extended dataset and evaluates the importance of each feature based on Z score.
- In every iteration, it checks whether a real feature has a higher importance than the best of its shadow features, and constantly removes features which are deemed unimportant.
Extreme Gradient Boosting
Variable Importance

Return on Equity before Taxes (lag1)
Economic sentiment indicator (lag1)
Available Working Capital (lag1)
Consumer Confidence Indicator (lag10)
Interest Expense Coverage (lag2)
Operating Profit Margin (lag1)
Total Capital to Capital Employed (lag1)
Stock Turnover Ratio (lag 3)
Creditors Ratio (lag 1)
Debtors Ratio (lag 5)
Extreme Gradient Boosting
Classification Accuracy

<table>
<thead>
<tr>
<th>Classification Accuracy Table 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model Comparison</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>KS</td>
</tr>
<tr>
<td>Logit 24%</td>
</tr>
<tr>
<td>LDA 23%</td>
</tr>
<tr>
<td>XGBoost 42%</td>
</tr>
<tr>
<td>MXNET 35%</td>
</tr>
</tbody>
</table>

Classification Accuracy Metrics: Kolmogorov - Smirnov (KS), Area Under ROC curve (AUROC).

**XGBoost** and **MXNET** algorithms provide better classification accuracy compared to traditional classification methods such as Logistic Regression and Linear Discriminant analysis.
Credit Risk Analysis
Calibrating a Rating system

Initial credit rating segmentation in 50 grades

Final credit rating segmentation in 9 grades
Deep Neural Networks
Rating System Performance

Rating System Calibration Metrics: Sum of Square Error (SSE), Brier's score (BRIER).

<table>
<thead>
<tr>
<th>Performance Metrics</th>
<th>Table 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Credit Rating System</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SSE</td>
</tr>
<tr>
<td>Logit</td>
<td>4.3%</td>
</tr>
<tr>
<td>LDA</td>
<td>4.8%</td>
</tr>
<tr>
<td>XGBoost</td>
<td>0.2%</td>
</tr>
<tr>
<td>MXNET</td>
<td>0.6%</td>
</tr>
</tbody>
</table>

Estimated and Actual default frequency metrics

<table>
<thead>
<tr>
<th>Estimated Probability of Default</th>
<th>Observed Default Rate (Out of sample)</th>
<th>Observed Default Rate (In sample)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logit</td>
<td>8.20%</td>
<td></td>
</tr>
<tr>
<td>LDA</td>
<td>7.80%</td>
<td>13.10%</td>
</tr>
<tr>
<td>XGBoost</td>
<td>13.50%</td>
<td></td>
</tr>
<tr>
<td>MXNET</td>
<td>15.00%</td>
<td></td>
</tr>
</tbody>
</table>

Estimated Probability of Default vs observed Default Rate in out-of-sample and in-sample population

- Based on SSE and Brier score the MXNET and XGBOOST rating systems perform better than Logistic Regression and Linear Discriminant analysis.

- The estimated PDs for MXNET and XGBOOST are closer to the observed default rates.
<table>
<thead>
<tr>
<th>Credit Risk Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Our Contribution</td>
</tr>
</tbody>
</table>

- Extensive exploration of advanced statistical techniques
- An automated algorithm for tackling dimensionality issues
- Application to a regulatory large size dataset
- Robust validation and Performance Measures
- Large potential for application in large datasets (Anacredit)
Credit Risk Analysis
Q&A

Questions

Answers
Thank you!