“Forecasting models using Machine Learning (ML) techniques on banks’ credit rating.”

Grigorian Alexandros

SCHOOL OF ECONOMICS, BUSINESS ADMINISTRATION & LEGAL STUDIES

A thesis submitted for the degree of

Master of Science (MSc) in Banking and Finance

November 2017
Thessaloniki-Greece
“Forecasting models using Machine Learning (ML) techniques on banks’ credit rating.”

Student name: Grigorian Alexandros

SID: 1103160007

Supervisor: Prof. Gogas Periklis

I hereby declare that the work submitted is mine and that where I have made use of another’s work; I have attributed the source(s) according to the Regulations set in the Student’s Handbook.

November 2017

Thessaloniki-Greece

Grigorian Alexandros
Abstract

This dissertation was written as part of the MSc in Banking and Finance at the International Hellenic University. The main purpose of this study is to present an empirical model designed to forecast banks’ credit ratings using information from their financial statements. For this reason, we have used ratings provided by Fitch in 2012. The sample consists of 92 US banks with their financial statements from 2008 to 2011. Using the same financial data and for the same time frame as (Gogas, Papadimitriou and Agrapetidou 2014) in their study, we used machine learning (ML) models in order to examine whether they are more efficient than classical econometric techniques that were used on their research. For this purpose using the same data as (Gogas, Papadimitriou and Agrapetidou 2014) and following their classification we trained the data both with linear and non-linear support vector machines in order to examine whether the prediction accuracy is higher with support vector machines rather than with ordered probit model. According to the simulation results, in the optimal group of regressors, both linear and non-linear (RBF kernel) Support Vector Machines, can predict more accurately credit ratings with a 84.06 percent accuracy for the linear SVM which is slightly higher than the 83.70 percent accuracy achieved by the ordered probit model of (Gogas, Papadimitriou and Agrapetidou 2014). On the other hand, non-linear SVMs can predict much more accurately than ordered probit models with 99.64 percent prediction accuracy. In other words, using either Linear Support Vector Machine or non-linear RBF kernel we can predict credit ratings more accurately than Ordered Probit model.

Keywords: Credit ratings, support vector machines, kernel support vector machines ordered probit model, rating forecasting

Grigorian Alexandros

Monday, November 27, 2017
I would like to express my gratitude firstly and mostly, to my supervisor, Professor Gogas Periklis who was the main motivation to choose this particular study and the main contributor for the thesis to be completed properly and on time. Except from him, I would like to thank also his associates, for their valuable contribution for this thesis to be done.
“Forecasting models using Machine Learning (ML) techniques on banks’ credit rating.”

List of Contents

Financial crisis and the role of credit rating agencies .............................................. 6
Need for regulation ........................................................................................................ 8
Literature review ............................................................................................................. 9
Determinants of bank credit ratings .............................................................................. 11
Data ............................................................................................................................... 12
Basics of SVMs ............................................................................................................. 14
Linear Separable Case .................................................................................................... 14
Error Tolerant SVM ....................................................................................................... 15
Kernel Methods ............................................................................................................ 16
Empirical Results .......................................................................................................... 17
Conclusions ................................................................................................................... 26
Bibliography .................................................................................................................. 27
Appendix ....................................................................................................................... 32

List of Tables

Table 1: List of independent variables ........................................................................... Error! Bookmark not defined.
Table 2: Number of variables in each regressor group ................................................. Error! Bookmark not defined.
Table 3: R²-values for the optimum set of regressors for the six groups of variables ...... Error! Bookmark not defined.
Table 4: Variables selected from each method ............................................................. Error! Bookmark not defined.
Table 5: Rating forecasting accuracy for group 1 of regressors ..................................... Error! Bookmark not defined.
Table 6: Rating forecasting accuracy for group 2 of regressors ..................................... Error! Bookmark not defined.
Table 7: Rating forecasting accuracy for group 3 of regressors ..................................... Error! Bookmark not defined.
Table 8: Rating forecasting accuracy for group 4 of regressors ..................................... Error! Bookmark not defined.
Table 9: Rating forecasting accuracy for Group 5 of regressors ................................... Error! Bookmark not defined.
Table 10: Rating forecasting accuracy for Group 6 of regressors .................................. Error! Bookmark not defined.
Table 11: Rating prediction accuracy for all the group of regressors for each selection criterion .................................................................................................................. Error! Bookmark not defined.
Table 12: List of regressors in the optimal forecasting models and their sources .......... Error! Bookmark not defined.

Grigorian Alexandros
According to (Hull 2015) credit ratings provide all the necessary information needed by financial market participants in order to deal with credit risks. A credit rating is a way to measure the credit quality or assess the creditworthiness of a borrower or a debt instrument. A credit rating can be created for any entity that is seeking to borrow money (such as: individuals, corporations, governments) (Investopedia n.d.). According to (Gogas, Papadimitriou and Agrapetidou 2014) the three major credit rating agencies are Moody’s, S&P, and Fitch. The highest rating assigned by Moody’s is Aaa. Bonds, companies or governments with that rating are considered to have almost no chance of defaulting. Following the credit rating scale downwards, the next levels are: A, Baa, Ba, B, Caa, Ca, and C. For Moody’s credit ratings from Aaa to Baa are considered as “investment grade” and from Ba to C as “non-investment” grade or “junk” (Moody’s n.d.). The S&P credit rating scale AAA, AA, A, BBB, BB, B, CCC, CC, and C. In the same rational as in Moody’s the first four levels are considered as “investment grade” and the rest as “non-investment or speculative” (S&P n.d.). Fitch follows the same rating scale as S&P (Fitch n.d.). In our case, the credit ratings that have been used are assigned by Fitch. The financial crisis of 2008 unveiled important disadvantages of rating agencies and the way they operate. Some of them are analyzed in the next section.

The role of credit rating agencies in the financial crisis

The role of CRAs in the recent financial crisis has been a subject of intense criticism. According to (Darbellay and Partnoy 2012) the three major rating agencies rated with high grade eleven large financial institutions that where in fact problematic and later on bankrupted. They mention that two of the major CRA prediction failures were AIG and Lehman Brothers. AIG were rated with a double-A while Lehman Brothers was still in the investment-grade category even a few days prior to collapsing. Besides that, according to (Moloney 2008) and (McVea 2010) prior to the subprime mortgage crisis of 2007, the three major rating agencies retained the highest of credit rating (triple-A) on thousands of subprime-related instruments that were in fact worthless. (Angelides 2010) argue that Moody’s was in fact a triple-A ratings factory; from 2000 through 2007, Moody’s rated with a triple-A almost 42,625 residential mortgage-backed securities (RMBSs), while on 2006, 869 billion U.S. dollars worth of mortgage-related securities were rated triple-A by Moody’s. After only six months 83 percent of them were downgraded (Morgenson and Louise 2010). Fitch was also an important player on the ratings market but with a smaller share (Hill 2010). According to (Rom 2009) the main reasons that lead CRAs to underestimate the actual risks derived of the
subprime mortgage market that contributed the most on the financial crisis are economic incentives, CRAs’ ignorance, and that they became overwhelmed. Which are analyzed below:

Economic incentives: According to (Rom 2009) a main reason of the CRAs’ failures is due to their business model. (Diaz 2002) claim that there are a certain conflicts of interest, because the CRAs earn the main part of their income from issuing ratings. In other words, the companies that are seeking to have their securities rated pay the CRAs to rate them. This means that the CRAs have clear incentives to provide credible ratings in order to protect their reputation. However, another study conducted by (Covitz and Harrison 2003) found that the CRAs are primarily motivated by their credibility and not the conflict of interest, to inflate ratings. (Coffee 2008) argues that the CRAs had, in fact, further conflict of interest incentives related to timely changes of the ratings. He supports that the CRAs were paid to issue initial ratings; there were no policy for surveillance and any further upgrades or, more importantly, downgrades. He also supports that given the powerful negative impact of downgrades on issuers, downgrading may put the CRAs' relationships with issuers, investment banks and many institutional investors at risk. As a result, downgrades occur not as frequently as they should be and in general not on time. Besides that, credit ratings could also be out of date because of the fact that the changes in the way that the CRAs evaluate securities are not included in the previous rating methodology but only on new ratings (Rom 2009). The CRAs have no economic motivate to apply these new methods retrospectively, as they receive no additional pay for doing so, and doing so carries the risk of leading to downgrades.

Ignorance: According to (Rom 2009), the prediction accuracy of the credit ratings process is based mainly on two important elements: (1) the accuracy of the information that CRAs receive, and (2) the preexisting historical data on the credit risks of the securities that are already being rated and the way that this data is used. If the existing historical data is sufficient and the data provided on the assets are accurate, then CRAs’ rate can be reliable. However, if there is misinformation on the collateral, or the loan type performance is uncertain, then the ratings are likely to be worthless, or even harmful. Besides that, (Rom 2009) support that the rise of the subprime market, lead the CRAs to be much more ignorant. Some reasons underlined by (Rom 2009) that lead to CRAs’ ignorance were: Firstly, the CRAs’ low data on the historical performance of subprime loans. There was relatively little historical data to guide them. Second, the subprime loan market has changed. Conventional mortgages were only a small part of it. The new products had often little or no down payments, variable interest rates and were made without documenting the borrowers’ income. Loans were packaged in new innovative ways much more complicated, untested, and difficult to assess. As a result, the securities that CRAs were asked to rate were
complex instruments based on unknown quantities. Another reason supported by (Rom 2009) is that the CRAs increasingly rated new types of securities, without conducting any due diligence to ensure that the accuracy of the information provided.

Stress: Another issue reported by (Rom 2009) is that the CRAs were overwhelmed by the steep growth of credit products like RMBSs and CDOs in the 2000s. He reports that the number of ratings for RMBSs and CDOs, and the amount of revenue generated by these ratings, grew rapidly between 2002 and 2007 (SEC 2008). In fact over this period, RMBS revenue grew by more than 100 percent annually on average, while CDO revenue was almost tripled annually. Besides that, the number of ratings issued grew also tremendously, by an annual average of about 70 percent for RMBSs and some 250 percent for CDOs. However, staff selection grew in a much slower pace, by an annual average of about 70 percent for RMBSs and 50 percent for CDOs (SEC 2008). This means that large proportions of the rating issuance were conducted by new and inexperienced financial staff. The combination of this inexperience with the increasing complexity of the structured products, suggests that there was no enough of scrutiny as by more experienced raters to more conventional products. As a result the provided ratings had high probability of being inaccurate due to the low level of experienced staff and the high demand for credit ratings by the financial market.

Another study conducted by (Gupta, Mitta and Bhalla 2010) suggest that there was a clear conflict of interest for the CRAs and the collaborating investment banks that designed the structured products. In other words there was low probability for CRAs to assign a low rating for their own designed products. In fact CRAs had a dual role in the credit rating process both providing credit assessments of the underlying collateral and being involved in the design of structured products. In fact (Gupta, Mitta and Bhalla 2010) mention that CRAs where being significantly profitable from the boom in the structured products market. That kind of products had a great impact on their incremental earnings and that gave them a great incentive in the success of these products. (Brunnermeier 2008) Also underlines the possibility of the CRAs providing favorable ratings to structured products that were in fact highly profitable for them due to the high fees attached to them.

**Need for regulation**

According to (Gupta, Mitta and Bhalla 2010), the credit rating agencies (CRAs) are a main contributor on protecting the investors’ interest in terms of the ratings reliability and in the information asymmetry reduction between the issuers and investors. As a result, CRAs’ functions should be well-defined and regulated in order to avoid any investors’ confidence erosion. During the recent financial crisis, that was a major issue for debate regarding to the regulatory framework within which they should operate. IMF in its report on the Financial Stability Forum (IMF 2008) recommend some
measures to be taken regarding to the functions of the credit rating agencies. Some of these are:

- The CRAs’ should follow a new policy framework where the role of the rating agencies is redefined. It should be clear whether ratings are mere opinions or judgments.
- Besides that, the Review Committee (IMF 2008) has suggested that the dual role of CRAs should be banned. Neither an agency nor its subsidiary should be able to provide advisory services, either formal or informal, on the design of a structured financial instrument and at the same time to rate the product.
- Moreover, (IMF 2008) propose that better due diligence would provide a more structured role for the CRAs, and this can be achieved with a better documentation of procedures, better surveillance, greater transparency norms and a more competitive environment for the CRAs to operate.
- Another interesting suggestion by (IMF 2008) is that credit ratings should be mandatorily made from at least two rating agencies, and unaccepted ratings may also be disclosed. This would lead to greater transparency of the investor domain (Hunt 2009).

Furthermore (Hunt 2009) mentions that the competition among the three leading CRAs (Standard and Poor, Moody and Fitch) is limited. He argues that CRAs do not have to compete at all, as the market operates as effectively as a "partner monopoly" shared by Moody's, S&P and Fitch (SENATE COMMITTEE 2006). However, If CRAs are fully regulated this may lead to investors confidence violation and low quality of rating (Hunt 2009).

To conclude, it is debatable whether the regulatory measures of the CRAs should be taken under a regulated regime, or implemented under self-regulation of the industry. The recent financial crisis showed that self-regulation is not effective for the industry according to the report of (IMF 2008). On the other hand, the regulatory bodies should act pre-emptive and not functioning without taking corrective action. (IMF 2008) suggests that the right balance between self-regulation and legislation is crucial for markets to operate efficiently. Besides that, it also underlines that the CRAs should act more in a proactive way and less reactively. Lastly, another important addition on the regulatory framework might be included the conduction of stress tests on key model parameters, according to (IMF 2008). It is true that correct future rating could help investors optimize their portfolio and identify any particular market reaction on time.

**Literature review**
According to (Auria and Moro 2008) there are various statistical techniques, that are used for the credit rating assessment. According to them, some of the most commonly used techniques are traditional statistical techniques such as linear Discriminant Analysis (DA) and Logit or Probit Models and non-parametric statistical models like Neural Networks. According to (Huang, Chen and Wang 2007) Support Vector Machines (SVM) is a promising non-linear, non-parametric classification technique, which has already showed satisfying results in the medical diagnostics, optical character recognition, electric load forecasting. Besides that it was also applied to a number of financial problems like credit scoring (Huang, Chen and Wang 2007) (Martens, et al. 2007) (Schebesch and Stecking 2005). (Auria and Moro 2008) underlines that the common objective of all these classification techniques in solvency analysis, is to develop a function, which can accurately separate the space of solvent and insolvent companies, by benchmarking their score value. Besides that they mention that SVMs are related to and contain elements of non-parametric applied statistics, neural networks and machine learning while it is a technique suitable for binary classification tasks. SVMs also classify a company as solvent or insolvent based on its score value (through a function of selected financial ratios). However this function is neither linear nor parametric (Auria and Moro 2008).

(Bruder, Dao and Roncalli 2011) support that Support vector machine (SVM) is an important part of the Statistical Learning Theory. It was first introduced in the mid-90's (Vapnik and Bottou 1992) and has important applications in various domains. According to (Vapnik 1998) this technique can be useful in various ways such as classification, regression or density estimation. In the financial field it is developed in two main directions. In the first one SVM is used as non-linear estimator to forecast the market tendency or volatility. In this way, it is used as a regression technique with the possibility of extension to nonlinear case through the kernel approach. On the other way, SVM is used as a classification technique.

(Min and Lee 2005) in their study, suggest a bankruptcy prediction model using support vector machines (SVMs) technique. In order to find out the optimal parameter values of kernel function of SVM they use a grid-search technique of 5-fold cross-validation. They compare the prediction accuracy of SVMs with those of multiple discriminant analysis (MDA), logistic regression analysis (Logit), and three-layer fully connected back-propagation neural networks (BPNs). The simulation results show that SVM is the most accurate model.

(Schebesch and Stecking 2005) in their study examined the performance of SVMs in credit scoring using weighted classes and moderated outputs. The data-set consisted of 658 credit applicants out of 17158 credit applicants of the population of the German building and loan association. They grouped people into two categories: "defaulting" and "non- defaulting" based on their historical credit performance. They
also used logistic regression with cut off optimization in order to compare the results. The results of Logistic regression in credit scoring was comparable (or even slightly better) than linear discriminant analysis. They found that logistic regression with optimized cut off performed better than standard SVM but it was less efficient than non-standard SVM.

(Lee 2007) in his study investigated the performance of support vector machines (SVMs) in corporate credit rating. He used a data-set obtained by the Korea Information Service and it consists of 297 financial ratios and the corresponding bond rating of 3017 Korean companies rated from 1997 to 2002. In order to choose the optimal financial ratios, he applies multiple discriminant analysis (MDA) stepwise regression analysis. In cases of SVM, MDA, and CBR, each data set is split into two subsets: a training set of 80% (2413) and a holdout set of 20% (604) of the total data (3017) respectively. In order to find out the optimal parameter of RBF kernel function of SVM he used a grid-search technique of 5-fold cross-validation. He compares the prediction accuracy of SVMs with this of multiple discriminant analysis (MDA), case-based reasoning (CBR), and three-layer fully connected back-propagation neural networks (BPNs). The simulation results where: The overall classification accuracy of SVM’s holdout data was 67.22% and the prediction accuracy of the training data 77.62%. For the BPN, the classification accuracy of the holdout data was 59.93% and that of the training data 62.95%. MDA, showed an overall classification accuracy of the holdout data of 58.77%, and a prediction accuracy of the training data of 58.72%. For CBR, the overall classification accuracy of the holdout data was 63.41%. The researcher concludes that SVM outperforms BPN and MDA at 1% statistical significance level and outperforms CBR at 10% significance level. Besides that the classification accuracy among BPN, MDA, and CBR are not significantly different from each other.

**Credit ratings determinants**

According to the study of (Pasiouras, Gaganis and Zopounidis 2006) the most significant bank classification variables are: loan loss provisions, capitalization and region of operations. In the same rational (Pasiouras, Gaganis and Doumpos 2007) report that for credit rating forecasting the most important variables are net interest margin, short-term funding, return on average equity, the number of shareholders and subsidiaries and the region of operations. (Chen and Shih 2006) in their study report that banks in the investment grade category, are affected the most by their asset quality while the banks in the speculative grade depend mainly in capital adequacy. According to (Ioannidis, Pasiouras and Zopounidis 2010) country-specific variables play a significant role on classification accuracy. Another study conducted by (Bellotti, Matousek and Stewart 2011) show that banks are affected mainly by four...
Forecasting models using Machine Learning (ML) techniques on banks’ credit rating.

Factors which are: values of equity to total assets lagged, liquid assets to total assets, natural logarithm of total assets and net interest margin.

To conclude based on the above studies, financial variables play an important role in the assessment of banks’ credit ratings. Besides that the study of (Gogas, Papadimitriou and Agrapetidou 2014) show that CRAs depend on quantitative as well as qualitative information from many sources in their rating process.

**Data**

This study is based on 92 US banks’ long-term ratings freely available by Fitch. These rating were used in order to attempt a forecast of the ratings of 2012. In order to achieve that, we have collected, for each one of the 92 banks, 46 individual variables and ratios from their financial statements for the period of (2008-2011). In other words, we forecast the bank’s credit rating in 2012 using the set of regressors of each variable including its previous four years value. In total we have used 184 variables for each bank in order to forecast banks’ credit ratings. The financial data used is extracted by the database of the Federal Deposit Insurance Corporation (FDIC) . The dependent variable is ordinal and has six categories that are grouped in three categories. The banks get values from 0 to 2 according to the assigned rating. The three rating categories are: AA, A (2), BBB (1), BB, B, CCC (0) (in the parenthesis are the assigned score).

**Methodology**

The first step in the process was the identification of the factors that affected the most the bank ratings of the sample. In order to achieve that we followed a thorough variable selection procedure. In more detail we calculated the correlations $r_{i,R}$, (with i each individual variable and R the ratings dependent variable) for the total of 184 regressors. The next step is a prefiltering step, which include the creation of six groups of regressors. The first group includes all variables with $|r_{i,R}| \geq 0.5$ including their lags. This group consists of 20 variables. In the same rational, the second group includes all the variables with $|r_{i,R}| \geq 0.4$. This group consists of 44 variables. In the third group are included all the variables with $|r_{i,R}| \geq 0.4$ excluding their time instances. This group consists of 15 variables. Group four includes ten variables with the best five variables positively correlated and the best five variables negatively correlated. Group five consists of the best 30 highly correlated variables. The last
group includes all the variables of the sample. Table 2 presents the number of variables in each group.

<table>
<thead>
<tr>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
<th>Group 4</th>
<th>Group 5</th>
<th>Group 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>20 variables</td>
<td>44 variables</td>
<td>15 variables</td>
<td>10 variables</td>
<td>30 variables</td>
<td>184 variables</td>
</tr>
</tbody>
</table>

The next step in the procedure is to recognize and select from the groups described above, the most significant variables based on the ratings. In order to achieve that, we follow a combinatorial search procedure for every set of variables and we choose the one that has the highest $R^2$. In the same rational we select an augmented set of regressors including eight variables. In order to achieve that, we use stepwise forward method of least squares for the set of variables with p-value > 0.1.

Table 3 shows the results of the selection procedure described above. It shows the $R^2$ of the best selected variables achieved through regression. It is divided into six columns that represent the groups of regressors. The rows of the table describe the three section criterions. The first line shows the results of searching all possible combinations of the $R^2$ in order to select four regressors with the highest performance from each group. In the second line the same procedure is followed for eight regressors. The next line shows the selection results of using the stepwise-forward variable criterion to get the best $R^2$ of the variables. The last line shows the number of variables selected by the stepwise-forward method. The above selection procedure leads to all the optimal variables selected as best regressors through the three variable selection criterions and for each one of the six sets of regressors. In Panel A are included the best four variables derived from the combinatorial selection criterion for each group of regressors. Panel B consists of the best eight variables selected through the combinatorial methodology for each group of regressors. Finally, Panel C includes the variables selected through the stepwise – forward methodology. Case in total there was selected: three variables from group four, four variables from groups one and five, five variables from groups two and three and eleven variables from group six (it includes all 184 variables without any prefiltering).

Table 2: $R^2$-values for the optimum set of regressors for the six groups of variables

<table>
<thead>
<tr>
<th>Regressor selection</th>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
<th>Group 4</th>
<th>Group 5</th>
<th>Group 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Combinatorial 4</td>
<td>0,49</td>
<td>0,54</td>
<td>0,54</td>
<td>0,53</td>
<td>0,53</td>
<td>0,59</td>
</tr>
</tbody>
</table>

Grigorian Alexandros
“Forecasting models using Machine Learning (ML) techniques on banks’ credit rating.”

<table>
<thead>
<tr>
<th>Combinatorial 8</th>
<th>0.53</th>
<th>0.61</th>
<th>0.57</th>
<th>0.49</th>
<th>0.60</th>
<th>0.70</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stepwise-forward</td>
<td>0.40</td>
<td>0.55</td>
<td>0.55</td>
<td>0.46</td>
<td>0.53</td>
<td>0.71</td>
</tr>
<tr>
<td>variables selected by the stepwise-forward criterion</td>
<td>(4)</td>
<td>(5)</td>
<td>(5)</td>
<td>(3)</td>
<td>(4)</td>
<td>(11)</td>
</tr>
</tbody>
</table>

**Basics of SVMs**

According (Vapnik and Cortes 1995) Support Vector Machines (SVM) is a modeling methodology used for two-class data classification. Based on them, the main objective of SVMs is to select a small number of data points from the dataset (Support Vectors), defining a hyper plane and separating data points into two classes’. If the problem cannot be separated linearly, the SVM is combined with a non-linear Kernel mapping procedure, showing the data points in a higher dimensional space (feature space), in order to separate the classes linearly. The procedure is divided into two main steps: the training step and the testing step. In the first step, the largest part of the dataset is used for the estimation of the separating hyper plane. In the other step, there is an examination of the model’s performance in the small subset that wasn’t used in the first step in order to evaluate the generalization ability of the model. Typically, 80%-95% of the dataset is used for the training step and the rest 20%-5% for testing.

In the next sections follows a brief description the mathematical derivations of the SVM theory.

**Linear Separable Case**

Considering a dataset of vectors where \( x_i \in R^2 \ (i=1, 2, n) \), belonging into two classes: \( y_i \in \{-1, +1\} \), we suppose that the two classes are linearly separable. Then a separator is defined as:

\[
 f(x_i) = w^T x_i - b = 0 \quad (1)
\]

in such that \( y_i f(x_i) > 0 \ \forall i \), where \( w \) is the weight vector and \( b \) is the bias.

The optimal hyper plane is selected as the decision boundary that classifies each data vector to the correct class and has the maximum distance from both classes. This distance is often called “margin”. In Figure 1, the SV’s are represented with the pronounced contour, the margin lines (defining the distance of the hyper plane with each class) are represented with the continuous lines and the hyper plane is represented with the dotted line.

The problem of finding the hyper plane can be dealt through the Lagrange relaxation procedure on the following equation:

\[
 min_w, b, max_a \left\{ \frac{1}{2} \| w \|^2 - \sum_{i=1}^{N} a_i [y_i(w^T x_i - b) - 1] \right\} \quad (2)
\]
where \( \alpha = [a_1, \ldots, a_n] \) are the non negative Lagrange multipliers. Equation (2) is never used to estimate the solution. Instead we always solve the dual problem, defined as:

\[
\max_{\alpha} \{ \sum_{i=1}^{N} a_i - \sum_{j=1}^{N} \sum_{k=1}^{N} a_j a_k y_i y_k x_i^T x_k \}
\]

Subject to \( \sum_{i=1}^{N} a_i y_i = 0 \) and \( 0 \leq a_i \forall i \)

The solution of (3) gives the location of the hyper plane defined by:

\[
\hat{w} = \sum_{i=1}^{N} a_i y_i x_i
\]
\[
\hat{b} = \hat{w}^T x_i - y_i, \ i \in V \quad \text{(where } V = \{i: 0 < y_i\} \text{ is the set of the support vector indices).}
\]

Figure 1: Hyper plane selection and support vectors. The SV’s are represented with the pronounced red contour, the margin lines are represented with the continuous lines and the hyper plane is represented with the dotted line. (Source: Plakandaras et al., 2013)

**Error Tolerant SVM**

(Vapnik and Cortes 1995) introduced non-negative slack variables in order to allow a predefined level of error tolerance in the training procedure \( \xi_i \geq 0, \forall i \) and a parameter \( C \) describing the desired tolerance to classification errors. Equation (2) is now defined as:

\[
\min_{w,b,\xi} \max_{\alpha} \{ \frac{1}{2} ||w||^2 + C \sum_{i=1}^{N} \xi_i - \sum_{j=1}^{N} a_j [y_i (w^T x_j - b) - 1 + \xi_j] - \sum_{k=1}^{N} \mu_k \xi_k \}
\]

where \( \xi_i \) measures the distance of vector \( x_i \) from the hyper plane when classified erroneously.

The hyper plane is defined as:

\[
\hat{w} = \sum_{i=1}^{N} a_i y_i x_i
\]
\[
\hat{b} = \hat{w}^T x_i - y_i, \ i \in V
\]
where $V = \{ i : 0 < y_i < C \}$ is the set of the support vector indices.

**Kernel Methodology**

In the case that there is no linear separator in the two class dataset (Figure 2), then the SVM classification is combined with kernel methods.

Figure 2: The Data Space. The non-separable two class scenario. (Source: Plakandaras et al., 2013)
The dataset is projected through a kernel function into a richer space of higher dimensionality (feature space) where the dataset is linearly separable (Figure 3). The solution to the dual problem with projection of eq. (4) now transforms to:

\[
\max_a \sum_{i=1}^N a_i - \frac{1}{2} \sum_{j=1}^N \sum_{k=1}^N a_j a_k y_j y_k K(x_j, x_k)
\]  

(7)

Under the constraints \( \sum_{i=1}^N a_i y_i = 0 \) and \( 0 \leq a_i \leq C, \forall i \) where \( K(x_j, x_k) \) is the kernel function.

Although the SVM theory uses the structural risk minimization rule in order to select the hyper parameters, it always seeks for a globally optimized solution avoiding model over-fitting.

**Empirical Results**

As stated previously, the main purpose of this research is to investigate whether Machine Learning (ML) techniques are more efficient than classical econometrics techniques. In order to achieve that, we used the same data as (Gogas, Papadimitriou and Agrapetidou 2014) and we trained them in two different ways. Firstly using linear Support Vector Machines (SVM) and then using non-linear Gaussian Radial Basis Function (RBF). Due to the fact that these techniques are used in binary classification we dealt with that problem by following a one-to-all procedure comparing for
example the prediction accuracy of group 1 of regressors with the prediction accuracy of all regressors. We follow the same procedure step by step and one by one for all six groups of regressors. After training the data for all six groups of regressors we compared the results with the ones reached by (Gogas, Papadimitriou and Agrapetidou 2014) using Ordered Probit model in order to find the most efficient modeling technique in terms of prediction accuracy.

The next step after the thorough variable selection procedure that described previously, is the creation of three sets of regressors for each one of the groups of variables that lead to the creation of 36 models, in total, that were estimated and evaluated in terms of bank credit rating forecasting accuracy. Using the machine learning techniques described above, we forecast the banks’ credit ratings for the fiscal year of 2012. In the Tables 5-11 following and their corresponding graphs (1-5) shows the results of using the three modeling techniques in terms of forecasting prediction accuracy for each one of the 36 models tested. Each table corresponds to every set of the prefiltered regressors. Each column shows the results of the each model using the three selection criterions and each row corresponds to the prediction model used for each selection criterion.

Graph 1: Prediction accuracy of the three models for first set of regressors

Table 3: Rating forecasting accuracy for the Group 1 of regressors
The first group of regressors, includes 20 variables with a correlation (of each individual variable with its ratings dependent variable) higher than 0.5, including their lags. For the set of four variables (combinatorial 4) we report that the higher prediction accuracy is achieved by the non-linear RBF Kernel with a 88.77 percent accuracy. The second model, in terms of prediction accuracy, is the Linear Support Vector Machine technique which is almost ten percent less accurate than RBF Kernel. The least accurate model is Ordered Probit Model which can predict a combination of four variables with 69.57 percent accuracy. In other words, non-linear RBF Kernel can predict almost 10 percent more accurately than linear SVM and almost 20 percent more accurately than Ordered Probit Model. Moving further on, the third column of the Table 5, which represents the augmented regressor set with eight variables (combinatorial 8), shows more or less the same results as combinatorial 4. As a result of the fact that we increased the number of regressors we can report slightly higher accuracy for both Machine Learning techniques (RBF kernel and Linear SVM) compared to the augmented set of four regressors. The third column represents the set of variables that was selected using a stepwise forward method of least squares procedure. We can report that all three modeling techniques show the lowest prediction accuracy of the three selection procedures. Overall, for this group of regressors, as it is presented in Graph 1, for all three selection processes Machine Learning techniques show higher prediction accuracy. More specifically non-linear RBF Kernel has the higher prediction accuracy at every set of regressors and in augmented set of eight regressors it shows the higher accuracy (92.03 percent).
“Forecasting models using Machine Learning (ML) techniques on banks’ credit rating.”

The second group of regressors contains 44 variables with a correlation (of each individual variable with its ratings dependent variable) higher than 0.4. In the augmented set of four regressors (combinatorial 8) we report that linear SVM as well as non-linear RBF Kernel outperform by far Ordered Probit model. In more detail, linear SVM predicts ratings 81.52 percent accurately which is 10.87 percent more accurately that ordered probit and non-linear RBF kernel predicts ratings with 94.2 percent accuracy, in other words 23.55 percent more accurately that ordered probit model. In the augmented set of eight variables (combinatorial 8), the accuracy results are slightly higher than combinatorial 4, however the ranking of the models based on the prediction accuracy remain the same. In the third column, the set of regressors selected through stepwise forward method of least squares procedure, predicts ratings slightly less accurately than the other two set of regressors for Machine Learning Models while for Ordered Probit Model prediction accuracy is slightly higher. In total, as it is presented in Graph 2, the combinatorial 8 is the set of regressors that

Graph 2: Prediction accuracy of the three models for the second set of regressors

Table 4: Rating forecasting accuracy for Group 2 of regressors

<table>
<thead>
<tr>
<th>Group 2</th>
<th>Set of Regressors</th>
<th>Combinatorial 4</th>
<th>Combinatorial 8</th>
<th>Stepwise-forward</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ordered Probit</td>
<td>70.65%</td>
<td>75.00%</td>
<td>71.74%</td>
<td></td>
</tr>
<tr>
<td>Linear SVM</td>
<td>81.52%</td>
<td>83.70%</td>
<td>81.16%</td>
<td></td>
</tr>
<tr>
<td>Rbf Kernel</td>
<td>94.20%</td>
<td>95.29%</td>
<td>93.84%</td>
<td></td>
</tr>
</tbody>
</table>
produce the highest accuracy and non-linear Machine Learning Model (RBF kernel) is the model with the higher accuracy (95.29 percent).

![Graph 3: Prediction accuracy of the three models for the third set of regressors](image)

**Table 5: Rating forecasting accuracy for the Group 3 of regressors**

<table>
<thead>
<tr>
<th>Set of Regressors</th>
<th>Combinatorial 4 (1)</th>
<th>Combinatorial 8 (2)</th>
<th>Stepwise-forward (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ordered Probit</td>
<td>70.65%</td>
<td>76.09%</td>
<td>71.74%</td>
</tr>
<tr>
<td>Linear SVM</td>
<td>81.52%</td>
<td>81.16%</td>
<td>81.16%</td>
</tr>
<tr>
<td>Rbf Kernel</td>
<td><strong>94.20%</strong></td>
<td>92.03%</td>
<td>93.84%</td>
</tr>
</tbody>
</table>

The third set of regressors includes all variables with correlation higher than 0.4 without their corresponding time instances (15 variables). In the augmented set of four variables (combinatorial 4) Machine Learning techniques again outperform by far the classic econometric technique. More specifically, Linear SVM predicts almost 11 percent more accurately than ordered probit model while non-linear RBF Kernel predicts approximately 24 percent more accurately. In the augmented set of eight variables (combinatorial 8), ordered probit model predict more accurately than in combinatorial 4, prediction accuracy on Linear SVM is more or less the same and RBF kernel predicts slightly less accurately than in combinatorial 4. Lastly through stepwise
forward method of least squares procedure, prediction accuracy is more or less the same as in combinatorial 4. Overall, as it can be derived from Graph 3, in all selection procedures examined, both machine learning techniques predict ratings more efficiently than ordered probit model. The highest prediction accuracy in this set of regressors is achieved through the combinatorial 4 selection procedure and the model that provides that accuracy is RBF Kernel with 94.2 percent prediction accuracy.

Graph 4: Prediction accuracy of the three models for the fourth set of regressors

Table 6: Rating forecasting accuracy for the Group 4 of regressors

<table>
<thead>
<tr>
<th>Set of Regressors</th>
<th>Combinatorial 4 (1)</th>
<th>Combinatorial 8 (2)</th>
<th>Stepwise-forward (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ordered Probit</td>
<td>68,48%</td>
<td>71,74%</td>
<td>71,74%</td>
</tr>
<tr>
<td>Linear SVM</td>
<td>80,07%</td>
<td>79,71%</td>
<td>77,90%</td>
</tr>
<tr>
<td>Rbf Kernel</td>
<td>86,96%</td>
<td><strong>92,03%</strong></td>
<td>89,13%</td>
</tr>
</tbody>
</table>

The fourth group of regressors includes the five variables with the highest positive correlation and the five variables with the highest negative correlation for a total of ten variables. In the augmented set of four regressors (combinatorial 4) Machine Learning models again outperform the classical econometric technique (ordered probit model). If we increase the set of regressors into eight (combinatorial 8) we report a slightly higher accuracy both for the ordered probit model and for the RBF Kernel but lower prediction accuracy for linear SVM. In the set of regressors selected
“Forecasting models using Machine Learning (ML) techniques on banks’ credit rating.”

through stepwise-forward of least squares method, the prediction accuracy of Linear SVM is more or less the same as the other two sets of regressors. On the other hand, RBF kernel in this group of regressors predicts more accurately that combinatorial 4 but less accurately that combinatorial 8. Overall, as it can be easily derived from Graph 4, again both Machine Learning models provide more accurate prediction for all three set of regressors and the highest prediction accuracy is achieved using RBF kernel for combinatorial 8 set of regressors (92.03 percent).

Graph 5: Prediction accuracy of the three models for the fifth set of regressors

Table 7: Rating forecasting accuracy for the group 5 of regressors

<table>
<thead>
<tr>
<th>Set of Regressors</th>
<th>Combinatorial 4 (1)</th>
<th>Combinatorial 8 (2)</th>
<th>Stepwise-forward (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ordered Probit</td>
<td>68.48%</td>
<td>71.74%</td>
<td>68.48%</td>
</tr>
<tr>
<td>Linear SVM</td>
<td>80.07%</td>
<td>84.78%</td>
<td>80.07%</td>
</tr>
<tr>
<td>Rbf Kernel</td>
<td>86.96%</td>
<td>95.29%</td>
<td>86.96%</td>
</tr>
</tbody>
</table>

The fifth group of regressors includes the 30 variables with the highest correlation. In this set of regressors, we report identical prediction accuracy for combinatorial 4 and stepwise forward method of least squares procedure. For these selection procedures, both Linear SVM and non-linear RBF Kernel predict ratings more accurately than ordered probit model. Linear SVM predict by almost 11 percent more accurately than ordered probit while RBF Kernel prediction is by approximately 18 percent more
accurate. In the combinatorial 8 prediction accuracy for all three models is higher than the other two variable selection procedures. Overall, as it is shown in Graph 5, machine learning models predict more accurately than the classical econometric technique. More specifically in this set of regressors the most accurate model in prediction terms is RBF Kernel for combinatorial 8 selection procedure (95.29 percent).

![Graph 6: Prediction accuracy of the three models for the sixth set of regressors](image)

Table 8: Rating forecasting accuracy for the group 6 of regressors

<table>
<thead>
<tr>
<th>Set of Regressors</th>
<th>Combinatorial 4 (1)</th>
<th>Combinatorial 8 (2)</th>
<th>Stepwise-forward (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ordered Probit</td>
<td>71.74%</td>
<td>81.52%</td>
<td>81.52%</td>
</tr>
<tr>
<td>Linear SVM</td>
<td>82.25%</td>
<td>82.97%</td>
<td>84.06%</td>
</tr>
<tr>
<td>Rbf Kernel</td>
<td>91.30%</td>
<td>91.67%</td>
<td>99.64%</td>
</tr>
</tbody>
</table>

Last but not least, the sixth set of regressors includes all 184 explanatory variables of our sample. In the augmented set of four variables (combinatorial 4) as well as in the augmented set of eight variables (combinatorial 8) the most accurate predictions are achieved with Linear SVM and RBF Kernel. Both of these machine learning models perform better than ordered probit model. In more detail, RBF Kernel outperforms by approximately 20 percent both combinatorial 4 and combinatorial 8. Linear SVM
outperforms ordered probit model by 10 percent in combinatorial 4 while in combinatorial 8 the difference in accuracy is insignificant. As it is presented in Error! Reference source not found. above, through stepwise forward method, the prediction accuracy for all three models is higher than the other two. To conclude, for this set of regressors the highest prediction accuracy is achieved through non-linear RBF Kernel model and by stepwise forward selection procedure (99.64 percent).

Table 9: Rating prediction accuracy for all the group of regressors for each selection criterion

<table>
<thead>
<tr>
<th>Set of Regressors</th>
<th>Combinatorial 4</th>
<th>Combinatorial 8</th>
<th>Stepwise-forward</th>
</tr>
</thead>
<tbody>
<tr>
<td>GROUP 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ordered Probit</td>
<td>69,57%</td>
<td>68,48%</td>
<td>65,22%</td>
</tr>
<tr>
<td>Linear SVM</td>
<td>79,71%</td>
<td>80,43%</td>
<td>77,90%</td>
</tr>
<tr>
<td>Rbf Kernel</td>
<td>88,77%</td>
<td>92,03%</td>
<td>86,23%</td>
</tr>
<tr>
<td>GROUP 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ordered Probit</td>
<td>70,65%</td>
<td>75,00%</td>
<td>71,74%</td>
</tr>
<tr>
<td>Linear SVM</td>
<td>81,52%</td>
<td>83,70%</td>
<td>81,16%</td>
</tr>
<tr>
<td>Rbf Kernel</td>
<td>94,20%</td>
<td>95,29%</td>
<td>93,84%</td>
</tr>
<tr>
<td>GROUP 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ordered Probit</td>
<td>70,65%</td>
<td>76,09%</td>
<td>71,74%</td>
</tr>
<tr>
<td>Linear SVM</td>
<td>81,52%</td>
<td>81,16%</td>
<td>81,16%</td>
</tr>
<tr>
<td>Rbf Kernel</td>
<td>94,20%</td>
<td>92,03%</td>
<td>93,84%</td>
</tr>
<tr>
<td>GROUP 4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ordered Probit</td>
<td>68,48%</td>
<td>71,74%</td>
<td>71,74%</td>
</tr>
<tr>
<td>Linear SVM</td>
<td>80,07%</td>
<td>79,71%</td>
<td>77,90%</td>
</tr>
<tr>
<td>Rbf Kernel</td>
<td>86,96%</td>
<td>92,03%</td>
<td>89,13%</td>
</tr>
<tr>
<td>GROUP 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ordered Probit</td>
<td>68,48%</td>
<td>71,74%</td>
<td>68,48%</td>
</tr>
<tr>
<td>Linear SVM</td>
<td>80,07%</td>
<td>84,78%</td>
<td>80,07%</td>
</tr>
<tr>
<td>Rbf Kernel</td>
<td>86,96%</td>
<td>95,29%</td>
<td>86,96%</td>
</tr>
<tr>
<td>GROUP 6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ordered Probit</td>
<td>71,74%</td>
<td>81,52%</td>
<td>81,52%</td>
</tr>
<tr>
<td>Linear SVM</td>
<td>82,25%</td>
<td>82,97%</td>
<td>84,06%</td>
</tr>
<tr>
<td>Rbf Kernel</td>
<td>91,30%</td>
<td>91,67%</td>
<td>99,64%</td>
</tr>
</tbody>
</table>

If we concentrate the data of all the tables (5-10) we create an aggregate table of all groups of regressors (Error! Reference source not found.) for each selection criterion. For the combinatorial four criterion the best accuracy prediction is achieved with the regressors of groups 2 and 3 (94.20 percent) while for the combinatorial eight criterion the best accuracy prediction is achieved with the regressors of groups 2 and 5 (95.29 percent). In the case of Stepwise-forward criterion the best prediction accuracy is achieved by the group 6 that contains all of the variables (99.64 percent). According
to these results, the best accuracy is achieved through group six of regressors for all regressor selection criteria with 99.64 percent accuracy in bank rating forecast. Besides that, according to the results there is a strong correlation between them and historical data four years prior to the credit rating event. More specifically: the net operating income as a proportion of the total interest income lagged two years (NOI10), the size of the bank estimated by the total assets in the two years prior to the rating (TASSET11 and TASSET10), the proportion of interest expenses over interest income lagged one year (TIE11), the part of interest income that consists of securities gains (losses) 4 years prior to the rating (SEC8), goodwill and other intangibles as a part of the total assets lagged 4 years (GOI8), the contribution of employees on assets 2 years lagged (ASSPE10), the ratio of long-term assets over total assets four years before the rating (LTA8), the liabilities derived from trading as a part of the total assets, one year before the rating, the subordinated debt over the total assets one year lagged and the regulatory Tier 2 risk-based capital as a proportion of total assets four years prior to the rating. Another interesting conclusion can be reached if we examine the source, in other words the financial statement, of which we have extracted the regressors. The main source that leads to the optimal regressors is balance sheet statement as seven out of eleven regressors come from that statement. The remain four regressors are derived from Income Statement and as performance ratio (1 regressor). Another interesting thing is that the condition ratios (C/R) do not appear in any regressor.

**Conclusions**

As it was analyzed in detail previously, the CRAs have a great impact on the financial system. Their reports are a fundamental for both the debt issuer and debt holder as well as for the whole economy. Besides that, Investors ‘decisions as well as debt issuers’ decisions rely mostly on the credit ratings provided by these institutions. As a result, accurate future credit ratings forecasting based on historical financial information may be beneficial for both parties. The main goal and contribution of this work, following the suggestion of (Gogas, Papadimitriou and Agrapetidou 2014), was after taking into account the critical role of CRAs and their ratings, to achieve higher prediction accuracy using innovative the Machine Learning techniques that seem to be more efficient than models based classical econometrics. In order to achieve that, we replicated the relatively simple model that (Gogas, Papadimitriou and Agrapetidou 2014) used to forecast Fitch’s ratings. Using 184 financial variables of the US banks in the period from 2008 to 2011 we tried to fit a model that will accurately forecast the
next year’s ratings. In order to achieve that we reduced the great number possible regressors by prefiltering and selecting six groups of variables. For these groups of regressors we followed the three selection criterions analyzed in detail in the previous sections to extract the variables performed the best. This procedure lead to 18 sets of explanatory variables that are used to be trained in linear Support Vector Machine and non-linear Gaussian Radial Basis Function kernel. In total there were 36 models that used to forecast long-term ratings of the fiscal year of 2012 based on the ratings of Fitch. According to the results, the optimum model is achieved using eleven financial variables and has a 99.64 percent forecasting accuracy seven of these eleven forecasting regressors come from the banks’ balance sheets; three comes from the income statement and one from the performance ratios. While there are no condition ratios in this forecasting model. Moreover, based on the results, the assigned ratings have a strong correlation with the historical data four years prior to the credit rating event. As a result, the most important contributors to long-term ratings of banks are size, performance ratios and asset quality. This means that, during the recent financial crisis the problematic banks were downgraded rather late as the historical data was not used properly in order to avoid the domino effect.

Moving a step further, another controversial issue that derives from this research, is the role of regulation in CRAs. After the second major financial crisis the role of credit rating agencies where doubted. A widespread response has been the call for greater regulation of the rating agencies. However according to (White 2010), greater regulation will lead to barriers for new CRAs to entry as well as discourage innovation in the provision of bond creditworthiness information. As a result, these measures may have an adverse effect strengthening the centrality and the importance of the three significant agencies. (White 2010) Suggest that less regulation would be more efficacies but also a major revision in the prudential regulation of institutions’ would prevent future major financial crises.

Bibliography


Grigorian Alexandros


Grigorian Alexandros
"Forecasting models using Machine Learning (ML) techniques on banks' credit rating."


“Forecasting models using Machine Learning (ML) techniques on banks’ credit rating.”


“Forecasting models using Machine Learning (ML) techniques on banks’ credit rating.”


SENATE COMMITTEE. REPORT OF THE SENATE COMMITTEE ON BANKING, HOUSING, AND URBAN AFFAIRS TO ACCOMPANY. SENATE COMMITTEE, 2006.


"Forecasting models using Machine Learning (ML) techniques on banks' credit rating."


