The Proceedings Ekonbiz 2024
ISBN: 978-99955-45-44-4
UDK: 336.774.3:622.271 ...

Submitted: 19.04.2024.
Accepted: 30.04.2024.
Original Article

SOVEREIGN CREDIT RATING PREDICTION USING DATA MINING CLASSIFICATION TECHNIQUE

Rade Božić
University of East Sarajevo, Faculty of Business Economics Bijeljina, Bijeljina, Bosnia and Herzegovina
rade.bozic@fpe.ues.rs.ba
ORCID: 0000-0001-6956-8049

Predrag Katanić
University of East Sarajevo, Faculty of Business Economics Bijeljina, Bijeljina, Bosnia and Herzegovina
predrag.katanic@fpe.ues.rs.ba
ORCID: 0000-0002-6476-809X

***Abstract**: Before approving loans or buying securities, investors analyze the sovereign credit rating of a country that shows its ability to fulfill obligations. This information plays an important role for both, the debtor and the creditor. Calculation of this rating is performed by specialized agencies that provide their opinions based on appropriate information. It is expressed in the form of different categories and their calculation models are not publicly available. A country's credit rating shows how likely it is that the country will fulfill its obligations as a debtor on time. There are a lot of different opinions about the indicators that determine credit ratings and methods of their calculation. As data mining finds application in the economic sphere, the question is how successful are algorithms in determining country's credit rating. The aim of this paper is to use the data mining classification technique on selected data sets in order to predict sovereign credit rating. The methods used in this paper are Naive Bayes, k-nearest neighbours, decision tree and random forest. Evaluation measures of the models are presented and interpreted.*

***Key words**: data mining, classification, prediction, sovereign credit rating, credit rating agencies, financial markets*

JEL classification: E17, C53, B22

1. INTRODUCTION

In modern business conditions, one of the main goals of the investor is to reduce financial risk when creating an investment portfolio. This feature is also relevant in the area of international loans. In order to reduce the risk, it is necessary to gather as much information as possible about potential debtors. One of the important indicators is the sovereign credit rating, which is calculated and determined by specialized financial institutions. An appropriate category is assigned to a country as a result of a calculation based on different determinants. The determinants and methods of calculation are not publicly available. It is in the interest of every national economy to acquire the most attractive category of credit rating, which further provides an opportunity for more diverse sources of financing and economic activities. As data mining finds its application in the economic field, the question arises whether it is possible to determine the credit rating with the help of classification techniques and how successful are they in that task. To answer this question, two datasets with identical independent variables (attributes or determinants) were formed. Only difference between them is in the output variable (label) which represents the credit rating class of the company whose calculation is included in the data set. In this paper, 10 different indicators that are considered to have an impact on credit rating calculation were selected. Those are: Gross domestic product (GDP), GDP growth, GDP per capita, unemployment, inflation, public debt/GDP, corruption, fiscal balance, political stability and external balance. The first dataset shows the credit rating assigned by Standard & Poor's (S&P), while the second data set refers to Moody's company and its credit rating. Identical classification methods were applied to these two datasets (Naive Bayes, k-nearest neighbors, decision tree and random forest). Models were validated using accuracy, recall and false positive rate and the obtained results were presented and interpreted. A comparative analysis of the obtained results was also performed in order to determine which classification method had the greatest success.

2. SOVEREIGN CREDIT RATING AND CREDIT RATING AGENCIES

One of the options that countries use when obtaining financial funds are foreign loans. In this way, they provide an opportunity to pay current obligations or invest in the development and other projects. Recently, especially in the past three decades, government debt has been growing all over the world (Cecchetti, Mohanti, & Zampolli, 2011). As a negative stimulus, financial crisis has significantly contributed to the growth of this trend.

In order for countries to raise funds through borrowing, they need to access the international capital market. One of the main conditions for a successful performance in this market is the presentation of a credit rating, which shows the country's ability to settle its debt obligations on time. It can also be defined as "the ability and willingness of sovereign governments to repay existing and future commercial debt obligations on time and in full" (Fuchs & Gehring, 2013). For this reason, potential investors take this indicator into consideration when making a decision. It largely determines the conditions under which countries, especially developing ones, can raise funds on the international capital market. Credit rating agencies (CRAs) are independent and they estimate country's credit ratings by gathering information from a variety of sources. Based on them, the risk of non-fulfillment of obligations is determined, which is calculated in the form of a category marked with a letter (Kruck, 2011). The indicator itself is the basis of an important and often unpleasant relationship between countries that want to obtain funds in the international capital market and the agencies that determine the ability of countries to enter this market (Tennant & Tracei, 2016). However, this reduces the asymmetry of information between the creditor and the debtor. There are a large number of credit rating agencies in the world but the three most famous are US-based S&P, Moody's and Fitch. In addition to assigning credit ratings to the countries, their activity is to assess the risk of bonds and other debt instruments. This paper analyses credit ratings calculated by two companies, S&P and Moody's.

The S&P company was founded in 1860. by Henry Varnum Poor in order to help investors who wanted to invest in the railway industry by publishing a "History of Railroads and Canals in the United States". With further development, it became a credit rating agency bought by McGraw-Hill in 1966. With over 160 years of work history, the company deals with data processing and delivery of information that enables other companies, governments and individual investors to make the right business decisions. Headquarter is in New York (S&P, 2022).

Moody's Corporation is a global company for integrated risk assessment. John Moody founded Moody's in 1900. through the publication of the Moody's Manual of Industrial and Miscellaneous Securities. Moody's employs over 13,000 employees in more than 40 countries. They have been in business for over 100 years with the mission of providing reliable insights and standards that help decision makers do business with confidence (Moody's, 2022).

Table 1. Credit risk classification by S&P and Moody's - Investment-grade ratings

Interpret.	Moody's		Standard and Poor's	
	Long-term	Short-term	Long-term	Short-term
Highest credit quality	Aaa		AAA	
High credit quality	Aa1 Aa2 Aa3	Prime-1	AA+ AA AA-	A1+
Strong payment capacity	A1 A2 A3	Prime-2	A+ A A-	A1
Adequate payment capacity Last rating in investment-grade	Baa1 Baa2 Baa3	Prime-3	BBB+ BBB BBB-	A2 A3

Source: Elkhoury, 2008

Table 2. Credit risk classification by S&P and Moody's - Speculative-grade ratings

Interpret.	Moody's		Standard and Poor's	
	Long-term	Short-term	Long-term	Short-term
Speculative Credit risk developing, due to economic changes	Ba1 Ba2 Ba3		BB+ BB BB-	B
Highly speculative, credit risk present, with limited margin safety	B1 B2 B3	Not prime	B+ B B-	
High default risk, capacity depending on	Caa1 Caa2 Caa3		CCC+ CCC CCC-	C

Interpret.	Moody's		Standard and Poor's	
	Long-term	Short-term	Long-term	Short-term
sustained, favourable conditions			CC	
Default, Although prospect of partial recovery	Ca,C		C,D	D

Source: Elkhoury, 2008

Table 1. and 2. are shoing different classes of credit ratings that are calculated by the two companies for short-term and long-term loans. In this paper, long-term classes are analyzed. Moody's has 21 different classes expressed as a combination of uppercase and lowercase letters A, B and C with numerical designations 1, 2 and 3. S&P has 22 different categories with the letters A, B, C and D and the symbols "+" and "-" which show a tendency to move to a category above or below. In both agencies, class A is reserved for a high credit rating, B for an adequate one, and C for the one with the highest risk, with S&P also having a D mark for this category.

3. DATA MINING APPLICATIONS IN FINANCIAL MARKETS – LITERATURE REVIEW

Data mining is the process of collecting, processing, analyzing and obtaining usable insights from a data set. Observing different variants in the field of problem solving, practical application, formulation and presentation of data, this broad term seeks to explain different aspects of data processing. Data mining is gaining in importance in the modern age where automated systems provide a large amount of data for further processing (Aggarwal, 2015).

Data mining involves the use of statistical and machine learning methods that provide decision-making assistance, often in an automated way. As part of it, prediction is an important component. Instead of asking "what is the connection between advertising and sales?", there is often a greater interest in answering the question "which specific advertisement or product should be presented to the online consumer at the moment?". Also e.g., there is an interest in grouping customers in appropriate clusters which represent the basis of product offer differentiation. Data mining methods have the ability to automatically extract value from a large amount of data. Classification and prediction are performed through a large number of different methods, and each of them has advantages and disadvantages. They provide different results and their performance varies. For this reason, different methods are applied to solve the same problem in order to select the one that achieves the best results (Shmueli, Bruce, & Patel, 2016).

The variables in a data set are called attributes. In general, there are two types of data that are treated differently. The first type refers to a specially designated attribute where the goal is to predict its value for unseen instances. This type of data is called a label. Data mining with such data belongs to supervised learning. The two main tasks in this type of learning are called classification and regression. Classification is used for categorical attributes (such as very good, good, bad, etc.), while regression is used for numerically expressed attributes. For unlabelled attributes, data mining refers to unsupervised learning. Here, the goal is to extract as much useful information as possible from a given data set (Bramer, 2016).

The classification technique is used to determine the output label for unlabeled data in the test part of the data set, based on data from the another part related to algorithm training. It is also the most commonly used technique in data mining (Aggarwal, 2020).

Data mining finds application in various fields, especially in economics. As financial markets and business activities are driven by the movements of various factors, perfect time information provides an advantage in making business decisions at all levels. Data mining, which provides this type of information, plays a significant role in this process. It can find connections between features and create models that deal with predictions based on a wide range of data. Using historical data things like short-term exchange rates, interest rates and stock prices can be predicted (Hiľovská & Koncz, 2012).

Although not an easy task, predicting stock market prices and financial market movements has always attracted a large number of data science researchers. Their rise or fall are phenomena that investors seek to estimate when investing. One of the most widely used data mining techniques to solve this problem are neural networks with the basic assumption that similar input data after processing gives similar output data while ignoring daily fluctuations. Text mining also helps to predict stock price movements. Schumaker and Chen developed a model that downloaded S&P500 stock price news via the Yahoo finance platform to predict price movements 20 minutes into the future (2009). For solving this problem inference rules, statistical analysis, genetic algorithms, and data visualization were also used (Zhang & Zhou, 2004).

Data mining also provides assistance in portfolio management, i.e., in the selection of different types of securities and the distribution of available financial resources for their acquisition. The main goal is to choose their adequate combination that maximizes income while minimizing risk. To solve this problem, various data mining models based on neural networks, hidden Mark model, etc., have been developed. (Hariharan, 2018).

Globalization is forcing countries to liberalize their individual markets in order to attract foreign investment. Exchange rates are gaining importance as economic entities begin to operate multinationally. Forecasting the movement of exchange rates is used when making important business decisions because their fluctuations take place on a daily basis. Buying currencies at a lower price and selling them at a higher price is the main goal in the currency market. For this purpose, data mining models based on neural networks (especially multilayer perceptron) and various statistical techniques have been developed (Hariharan, 2018).

Zhang and Zhou wrote about the application of data mining in the analysis of borrowed capital risk. They forecasted the payment of debt, estimating the value of mortgages in the provision of services related to real estate and international currency trading (2004). In addition to the above examples, data mining can also be used in fraud detection, pattern recognition, anomaly detection, social media analysis and other activities that can provide assistance in business activities in financial markets.

When it comes to credit rating prediction, Ozturk, Namli and Erdan (2016) used classification and regression trees (CART), multilayer perceptron, support vector machines, Bayesian network and the Naive Bayes algorithm to determine Moody's credit rating for the period 1999-2010. The set consisted of 8 variables (financial balance, GDP balance, debt-to-GDP, GDP per capita, GDP changes, inflation, import-export ratio and government efficiency) and included 1,022 observations, or 92 countries. Compared to conventional statistical methods, their models provided over 90% accuracy with a tolerance of one or two notch deviations.

4. DATA SETS

Two data sets were used in this analysis. They contain identical variables for credit rating determination. The only difference between these two data sets is in the credit rating column itself, where the first set includes classes assigned by S&P and the second by Moody’s company. The same data mining methods will be applied on them so that the results can be compared and interpreted. It is important to note that the number of instances is not the same because companies provide credit ratings for only certain countries. All attributes were collected for the 2020. year, while the credit rating is from 2021. The models did not take historical credit rating data over a long period of time. The goal is to predict the credit rating in 2021. based on the data for the previous year. It is important to note that the global market was affected by the COVID-19 pandemic.

4.1. ATTRIBUTE SELECTION

As the calculation models and determinants for sovereign credit rating determination are unknown to the public, a large number of authors have addressed this issue such as Canton and Packer (1996), Tennant and Tracey (2016), Iyengar (2012), Sheng-Syan, Hsien -Yes. (2016), Wei Chee, Fan Fah and Nassir (2015) as well as many others. They categorized the determinants and formed their opinions based on the information given by the credit rating agencies. Based on the analysis of individual papers and written materials, the following most common attributes were selected for the data set.

Table 3. shows 11 attributes including the label. For each attribute, a link and source are provided through which the downloaded data can be accessed, as well as detailed information about what they actually represent. The last column shows the units in which attributes are expressed. All units are numeric except for the label, which is a combination of letters and numbers or symbols, depending on the rating agency. After eliminating countries with missing data, two data sets were formed. S&P data set has 95, while Moodi’s data set has 104 instances (different countries). Moody's data set contains all of the instances that are in the S&P set, and in addition it has 9 countries more. The S&P set has 19 different values (classes) of credit rating, while Moody’s has 21 different values.

Table 3. Data set attributes

No.	Attribute	Source	Link	Units

	name			
1.	GDP	Trading economics	https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?most_recent_year_desc=false	USD
2.	GDP growth	World bank	https://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG	%
3.	Unemployment	World bank	https://data.worldbank.org/indicator/SL.UEM.TOTL.ZS?most_recent_year_desc=false	%
4.	Inflation	World bank	https://data.worldbank.org/indicator/FP.CPI.TOTL.ZG	%
5.	GDPper capita	World bank	https://data.worldbank.org/indicator/NY.GDP.PCAP.CD	USD
6.	Public debt / GDP	Trading economics	https://tradingeconomics.com/country-list/government-debt-to-gdp	%
7.	Corruption	Trading economics	https://tradingeconomics.com/country-list/corruption-rank	Rank
8.	Fiscal balance	World bank	https://www.theglobaleconomy.com/rankings/fiscal_balance_percent_GDP/	%
9.	Political stability	The global economy	https://www.theglobaleconomy.com/rankings/wb_political_stability/	Index (from -2.5 to 2.5)
10.	External balance/ GDP	World bank	https://data.worldbank.org/indicator/NE.RSB.GNFS.CD	USD
11.	Credit rating	Trading economics	https://tradingeconomics.com/country-list/rating	Classes

Source: Collected data from different online sources and aggregated by author

5. DATA ANALYSIS

The open source WEKA data mining tool with version 3.8.6 was used in the analysis process. This tool enables the application of classification techniques and the validation of the obtained results provided by the models. The analysis was performed first on the Moody's, and then on the S&P data set. The k-fold cross-validation method was used to divide the set into a training part and a part for model testing. This method involves dividing a data set of N instances into k equal parts (subset). If the number N is not divisible without the remainder by the number k, then the last part has a smaller number of instances than the others. Each individual part is used for testing, while the others are used for training (Brammer, 2016). As S&P and Moody's data sets do not have a large number of instances, this method reduces bias in testing and training overall. In the analysis process, the data set is divided into 10 subsets, which means that testing in the WEKA tool is performed for each subset and once again for the whole data set (11 times in total).

In order to predict the credit rating class, the following classification methods were used in the analysis process: Naïve Bayes, decision tree, k-nearest neighbors (k-NN) and random forest.

Naïve Bayes uses probability theory for solving classification problems. It was named after Reverend Thomas Bayes (1702-1761) who was credited as the first mathematician to use probability in an inductive way. In classification, there is a set of alternative possible events, which are mutually exclusive and exhaustive, indicating that only one must always occur. Each of probabilities has to be between 1 and 0, and their sum has to be 1. Events are instances that have only one specific classification label. The training part of the data set contains a sample of trials that are used for predicting the classes of unseen instances. If set has k mutually exclusive classifications c_1, c_2,..., c_k, which have prior probabilities $P(c_1)$,$P(c_2)$,..., $P(c_k)$, and n attributes a_1,a_2,...,a_n which for a given instance have values v_1,v_2,...,v_n respectively, the posterior probability of class c_i can be calculated as:

$$P(c_i) \times P(a_1=v_1|c_1) \times P(a_1=v_1|c_1) \times \ldots \times P(a_n=v_n|c_n)$$

Product is calculated for each value of i from 1 to k and the classification which has the largest value is chosen (Brammer, 2016).

Decision tree is one of the most popular classification method in data mining. It starts from the root and consists of nodes and leaves. Every internal node has a splitting rule which divides data set based on attributes and sends data item to the node's children. This process is repeated until data reaches the leaf node which represents the predicted label class. Trees can also be pruned by getting shorter but maintaining similarclassification accuracy. They can use numerical or categorical attributes (Aggarwal, 2020).

Nearest Neighbour Classification can be used in both cases, when attribute values are continuous and categorical. If they are categorical, they need to be modified for this kind of use. The class of an unseen instance is determined by class of instances that are closest to it. In most cases, a small integer number of instances *k* is chosen to determine the class of an unseen one. The name of this method is *k-Nearest Neighbour* or *k-NN*. The basic k-NN algorithm finds *k* training instances with the smallest distance to the unseen instance and takes the most commonly occurring class of these *k* instances. The most popular methods used for the calculation of distance measures are Euclidian Distance, Manhattan Distance or City Block Distance and maximum dimension distance (Brammer, 2016).

Random Forest classifier is a combination of decision tree algorithms where each one is generated using a random vector and each tree votes for the class of the input vector. It consists of a large number of individual decision trees in the form of anensemble. Each tree predicts the class and a class with most votes represents the predicted label (Breinman, 2001). Random forest is described as a basic bagging method applied to decision trees involving training of each tree on a different part (subset) of the data set (Aggarwal, 2020).

Accuracy, recall and false positive rate were used to validate the models. They are calculated based on the confusion matrix as follows:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

$$Recall = \frac{TP}{TP+FN}$$

$$FPR = \frac{FP}{FP+TN}$$

When there are more than two classes in the selected data set, in the confusion matrix one class becomes positive while all the others are counted as negative. The confusion matrix consists of:

- TP (true positives) represents the number of positive instances that are classified as positive,
- TN (true negatives) the number of negative instances that are classified as negative,
- FP (false positives) number of negative instances that are classified as positive and
- FN (false negatives) number of positive instances that are classified as negative.

The accuracy of a classifier represents the percentage of instances that are correctly classified, recall of the model shows the percentage of correctly classified positives (TP) in the total number of positive instances, while FPR shows the percentage of negative instances that are incorrectly classified as positive (Brammer, 2016). Precision and F-score are not calculated due to the large number of different classes of the output attribute and the small number of samples in the set.

6. RESULT

The research results are presented in table form. The names of the methods are included together with the selected validation measures. First, the validation measures of the models applied to the Moody's data set are presented and interpreted. After that, the same procedure was applied to the S&P data set.

6.1. VALIDATION MEASURES OF METHODS APPLIED ON MOODY’S DATA SET

Table 4. shows the validation measures of methods applied to the Moody’s data set. Observing the obtained results of the applied methods at the Moody's set, a generally low level of accuracy can be noticed.

Table 4. Evaluation of the models applied on Moody’s data set

Method	Accuracy	Recall	FPR – false positive rate
Naive Bayes	16,35%	0,163	0,067
k-NN	15,38%	0,154	0,055
Decision tree	13,46%	0,135	0,061
Random forest	20,19%	0,202	0,058

Source: Authors’ calculation

The highest accuracy of 20.19% was achieved with the random forest method, while the lowest accuracy was achieved with the decision tree method (13.46%). Model recall reached its highest level again in the random forest method (20.2%), while the decision tree algorithm had the lowest level (13.5%). FPR was highest at Naive Bayes (6.7%), while it was lowest in the k-NN method (5.5%).

6.2. VALIDATION MEASURES OF METHODS APPLIED ON S&P DATA SET

Table 5. shows the validation measures of the S&P data set methods. As in the previous case, the applied methods on the S&P data set also gave a low level of accuracy.

Table 5. Evaluation of models applied on the S&P data set

Method	Accuracy	Recall	FPR – false positive rate
Naive Bayes	22,11%	0,221	0,061
k-NN	21,05%	0,211	0,055
Decision tree	16,84%	0,168	0,061
Random forest	25,26%	0,253	0,057

Source: Authors' calculation

The highest accuracy was provided by the random forest algorithm (25.26%), while the lowest accuracy was provided by the decision tree method (16.84%). Model recall was highest again in the random forest method (25.3%), while it was lowest in the decision tree algorithm (16.8%). The Naive Bayes and random forest methods have the same FPR (6.1%), while the k-NN method has the lowest (5.5%).

In both analyzed sets, the highest level of accuracy was achieved by the random forest method, while the lowest was achieved by the decision tree algorithm. The recall of the model was also the highest in the random forest method at both data sets, and the lowest with the decision tree method. FPR was lowest with the k-NN method in both data sets, while it was highest with the Naive Bayes method with the same level of FPR achieved in the S&P data set with the decision tree algorithm.

CONCLUSION

The highest accuracy (S&P 20.19% and Moody's 25.26%) and model recall in both data sets was achieved by the random forest method. All selected models achieved a generally low level of accuracy, FPR and model recall. This shows that they are not very successful in prediction of credit ratings, which further implies that the included variables do not sufficiently explain the assigned classes. This is also understandable because credit agencies include many more indicators compared to the number of attributes included in the models. In addition, one of the reasons is the short time period for which data was collected, only one year (2020), which implies that a period of one year is not enough to determine the credit rating class through the data mining classification technique. The large number of different credit rating values (classes) for a small number of instances affect the accuracy of the applied models.

Comparing the results obtained over the two selected data sets, methods were more successful in the S&P credit rating data base, meaning that the selected variables better explain the credit rating assigned by this company than by Moody's. One of the issues that arises and potentially explains the achieved results of the models is the potential bias of credit agencies in determining ratings.

One of the ways to increase the accuracy of the models is to reduce the classes to basic without subcategories, i.e. to A, B, C or A, B, C, D depending on the selected credit rating agency. This would significantly reduce the number of different label values. Future research may include observed variables over more time or add some new ones that deeply explain the sovereign credit rating, as well as new applied data mining techniques.

REFERENCES

[1] Aggarwal, C. (2020). Data Classification - Algorithms and Applications. Taylor & Francis Group.

[2] Bramer, M. (2016). Principles of Data Mining. London: Springer Verlag.

[3] Breinman, L. (2001). Random Forest. Machine Learning, 5-32

[4] Cantor, R., & Packer, F. (1996). Determinants and Impact of Sovereign Credit Ratings. Economic Policy Review.

[5] Cecchetti, S., Mohanty, M., & Zampolli, F. (2011). The Real Effects of Debt. BIS Working Papers. Retrieved from http://www.bis.org/publ/work352.pdf

[6] Elkhoury, M. (2008). Credit Rating Agencies and Their Potential Impact on Developing Coutries. Geneva: United Nations Conference on Trade and Development.

[7] Fuchs, A., & Gehring, K. (2013). The Home Bias in Sovereign Ratings. University of Heidelberg Department of Economics Discussion.

[8] Hariharan, N. (2018). Application of Data Mining in Finance. SSRN Electronic Journal.

[9] Hiľovská, K., & Koncz, P. (2012). Application of Artificial Intelligence and Data Mining Techniques to Financial Markets. ACTA VŠFS.

[10] Iyengar, S. (2012). The Credit Rating Agencies - Are They Reliable? A Study of Sovereign Ratings. Vikalpa: The Journal for Decision Makers, 69-82.

[11] Kruck, A. (2011). Private Ratings, Public Regulations: Credit Rating Agencies and Global Financial Governance. Basingstoke: Palgrave Macmillan.

[12] Moody's. (2022, 2 23). .moodys.io. Retrieved from Moodys: https://about.moodys.io/about-us

[13] Ozturk, H., Namli, E., & Erdan, H. (2016). Modelling sovereign credit ratings: The accuracy of models in a heterogeneous sample. Economic Modelling, 469-478.

[14] S&P, G. (2022, 2 23). Retrieved from spglobal: https://www.spglobal.com/en/who-we-are/our-history

[15] Schumaker, P., & Chen, H. (2009). A quantitative stock prediction system based on financial news. Information Processing & Management, 571-583.

[16] Sheng-Syan, C., Hsien-Yi, C., Chong-Chuo, C., & Shu-Ling, Y. (2016). The relation between sovereign credit rating revisions and economic growth. Journal of Banking and Finance, 90-100.

[17] Shmueli, G., Bruce, P., & Patel, N. (2016). Data Mining For Business Analytics - Concepts, Tecniques, and Applications with XLMiner. New Yersey: John Wiley & Sons.

[18] Tennant, F., & Tracey, R. (2016). Sovereign Debt and Credit Rating Bias. New York: Palgrave Macmillan.

[19] Wei Chee, S., Fan Fah, C., & Nassir, A. (2015). Macroeconomics Determinants of Sovereign Credit Ratings. International Business Research.

[20] Zhang, D., & Zhou, L. (2004). Discovering Golden Nuggets: Data Mining in Financial Application. IEEE Transactions on Systems, Man, and Cybernetics, 513-522.

[21] Aggarwal, C. (2015). Data Mining - The Textbook. Springer International Publishing Switzerland.