Business Analytics Applications for Consumer Credits

Claudia ANTAL-VAIDA
The Bucharest University of Economic Studies, Romania

The fast-paced and dynamic economical background determines all the industries to quickly adapt to change and adopt emerging technologies to remain competitive on the market. This tendency led to high volumes of data generated each second and to a decreasing ability of the manpower to analyze it and use if for beneficial purposes. This paper reviews the impact of Digital Transformation on the Banking area and how financial institutions leverage the advantage created by this trend, especially in the credit risk management field. Multiple papers on consumer credit scoring models written after the financial crisis from 2007 were reviewed and their results were summarized in this article, to increase the accuracy of future analysis by leveraging the already known results.
Keywords: *Business Analytics, Machine Learning, Banking, Credit Risk Assessment, Scoring Models, Consumer Credits*

1 Introduction

Considering the current economic background, public and private companies are facing lots of challenges and difficulties when setting up their growth strategy due to the highly competitive environment where they operate. In order to overcome them, these entities require changes in their approach when making decisions, relying more on historical facts. Therefore, these companies should change focus – instead of intuition, they should aim to become data-driven companies, but for this approach to be successful, they need to invest in data-processing and methods of generating insights to get valuable information.

Change is all around us and if we analyze how fast technology, economy, medicine and other fields are evolving, we will notice that data and data-analysis are key players in it. As most of the industries are highly developed nowadays, any organizational decision should be taken based on results generated by business analytics to reach an ascending trend and a sustainable growth. Fortunately, companies understood the importance of data and the business value add they can bring when used correctly and consistent.

Gartner defines the Analytics area as a bunch of methods and techniques used for building analytical models and simulations to understand reality and predict the future state of a system. [1] It is a must-have area due to its impact on financial growth and profitability by improving company's competitive advantage and minimizing the error when making a decision. Examining the growing interest in this area and the resources invested in research by multiple companies, we can deduce that the popularity of this field has exponentially increased and companies acknowledged the impact of more sophisticated analytical decision-making tools for creating new opportunities, choosing the timing and enhancing the know-how.

The Banking sector is facing similar changes and needs to invest in business analytics as well. Its focus should be on understanding customer's behavior, how to improve interactions with customers and what motivates him to carry through their obligations. [2]

Even though this industry seems to be complicated, its activity is quite straightforward: the banking sector handles cash, credits and secures funds aiming to make more money out of them. Its products mainly consist of credit and debit accounts, loans and mortgages, helping people and companies to invest in their future, especially when they do not have enough liquidities to do that.

Players in each industry have a key focus on improving profitability, looking into cost reduction, revenue increase and process optimization and the banking field does not make an exception. Having said that, Business Analytics can play a key role in achieving those goals and there are already examples on the market which exceeded expectations. Some applications worth mentioning would be the algorithms of predicting risks, that analyze transactions and identify uncommon behavior of a customer or the credit-scoring algorithms that predict whether a customer will pay his debts, or he will miss them.

2. Algorithms used in the Banking Area

The following part of this paper explains why and how the concept of big data appeared in the banking field and how companies decided to leverage the advantages created. All the data created each second increased the influence of Business Analytics in business applications, with lots of solutions already implemented and many others being explored.

2.1. Digital Transformation of Banking

The Digital Era is characterized by new technologies which increase the speed of knowledge turnover [3]. The emerging development of analytics, cloud, social media and mobility technologies caused overall disruption across industries, Banking being one of the most impacted fields by this trend. [4]

Companies understood the importance and impact of technology in their activity, therefore they started to invest in research and development labs that use not only social media analytics, machine learning algorithms and big data, but also research on possible innovative scenarios that leverage artificial intelligence, robotics, automation, advanced data visualization and not only.

American Express, one of the giants in the Banking and Financial services from US, has set up a new tech lab to focus on big data, cloud computing, analytics and mobile technologies, as well as on futuristic ones. Thru this initiative, they aim to analyze customer's behavior on the market and quickly respond to it with customized products. [5]

Another example of company acknowledging the importance of new technologies and investing in its development would be the Fidelity Investments, one of the largest asset management companies in the world. In 2014, they announced the opening of Financial Labs, a research unit that will partner with The Massachusetts Institute of Technology (MIT) and Stanford University to get the "outside view" and develop innovative applications. [6]

These were only few examples of Financial companies investing into the tech area. The trend in the industry is to digitalize as much as possible and behave almost like tech companies. Banks must be quick in converting an idea into a service in order to survive on the market but also stay relevant in the years to come. Moreover, technology offers new opportunities to address untouched markets, by simplifying the communication and removing the geographical barriers. [7]

From previous examples, we draw the conclusion that companies have become aware of the importance of technology and the role they play in business performance management. Moreover, they started to massively invest in developing new technical capabilities to optimize their processes and improve relationships with customers.

What's the result of these investments? How digitalization translates in the real world? It results in high volumes of unstructured data that are harder and harder to analyze manually and that's the perfect scenario for Big Data and Analytics to come into the picture. As data is not meaningful enough, companies had to identify ways of converting data into information and insights to monetize digitalization.

2.2. Machine learning at a glance

Business Analytics represents the use of data, technologies, statistics, mathematics and computer-based models to help management understand the business, solve issues and make fact-based decisions. [8] This area has four stages that have different business impacts, depending on their complexity and level of knowledge required [9]:

- **Descriptive Analytics** is answering the question "What happened?", by analyzing and displaying historical data in reports and dashboards to simplify the decision-making process;
- **Diagnostic Analytics** researches the causes and effects of a certain event in order to avoid it or increase its frequency in the future, depending on the impact. It usually answers the question "Why did it happen?";
- **Predictive Analytics** provides an answer to the question "When can it happen?" and implies statistical methods and Machine Learning techniques;
- **Prescriptive analytics** in the area that recommends decisions based on simulations and process optimizations, trying to answer the question "How can it happen again?".

Going forward, this paper will focus more On Machine Learning and its applicability. **Machine Learning** is defined as a tool or mechanism that uses statistical models to facilitate the solutioning of a problem, by studying the past behavior, identifying patterns and constantly improving itself based on the data analyzed. Its main purpose is to develop an adaptable algorithm to solve an issue, despite the external variables that might influence it or its complexity. [10] There are different types of algorithms used by machine learning listed below [11]:

- **Supervised algorithms** know the outcome from the beginning and its learning is guided by human observations and feedback thru tags and labels inputted from the beginning;
- **Unsupervised algorithms** rely exclusively on clustering separate data based on similar features and modifies the calculation process to respond to initial inputs; this type does not involve any external feedback, nor tags to be considered for data processing;
- **Reinforced algorithms** are about taking the most appropriate action to maximize the output, regardless the situation. If in supervised learning the expected outcome is known from the beginning, this type of algorithm has the liberty to decide what is best to do to perform a task and tends to learn from its own experience.

Even though Machine Learning sounds evolutionary and promises to revolutionize the way things work, having a positive impact on the areas where applied, it has both advantages and disadvantages that should be considered when deciding to use them for a real use case.

Thus, few of the advantages worth mentioning of Machine Learning are: [12]

- **Easily identified trends and patterns**. When given a large dataset to analyze, the Machine Learning algorithms can quickly identify specific trends that might not even be obvious to human beings.
- **Constant Improvement**. While exposed to computation of data, the ML algorithms gain experience and improve efficiency and accuracy, leading to more reliable decisions and results.
- **Various applications.** No matter the area you work on, you will find for sure an application that would involve Machine Learning algorithms and would be beneficial for your area.

On the other hand, Machine Learning has some limitations that should be known before deciding to use and invest in them:

- **Requires high volumes of data**. For excellent results, Machine Learning algorithms require high volumes of data to train on. Besides volumes, data quality of the train dataset plays an important role as well, outputs being highly dependent on inputs.
- **Time and Resources**. Depending on the complexity of analysis you want to perform, Machine Learning algorithms

may require time to learn and massive computer resources.

- **May produce biased results**. Machine Learning algorithms are highly susceptible to error, but this aspect mostly depends on the diversity of the dataset it trains on. If the train set is small and not inclusive enough, the results might be biased, leading to irrelevant interpretation.
- **Interpretability**. Unfortunately, it is hard to understand the reasoning behind a decision taken by an algorithm, reason why these are considered “black-boxes”.
- **Scalability**. Once a model is proven to be efficient, companies implementing it need to overcome the challenge of scaling it. This can become expensive due to the resources required, the need of further optimizing it and integrating it with other systems.

All these challenges can be overcome if the company is willing to invest and few of them might not even appear, it always depends on the use case. Therefore, Machine Learning usage can surface the potential and value of unstructured data in each company. Its ability to form adaptive behavior in the process, without being programmed for this, makes them incredibly powerful when used for the right processes and analysis.

To sum up, the main benefits of the ML algorithms in this area is given by the complexity of the analysis performed (due to various parameters considered), reduction of approval time, less human resources involved, thus avoiding the human error, fraud of subjectivism

2.3. Algorithms for Risk Management

Risk management has become more important in the banking fields since the global financial crisis took place, moment since when banks started to research on how risks can be detected, measured and managed . [13]

There are different types of risks in the financial area that can be addressed by Machine Learning, but we will analyze the applicability only for three of them: Operational risk, market risk and credit risk.

Operational risk management assumes that a firm wants to foresee the direct or indirect risk of financial loss due to a host of potential operational breakdown. [14] The risk can be triggered by internal factors (people, system, deficient processes) or by external ones (global economic background, frauds, operational errors). Considering the increasing variety and complexity of risks, especially for financial institutions, machine learning and artificial intelligence applicability increased consistently and started to play a key role in predicting these events, assessing their impact and minimizing their effects. [15]

Banks pursuit the evaluation of the best ways to protect their data, systems and clients and machine learning can support that. Process automation can increase the execution of routine tasks, minimize human error, analyze data to outline the relevant content and increases the ability to evaluate risky clients and networks. Machine Learning can also generate and prioritize alerts for uncommon activities and asses the risk involved.

Another risks that is worth investigating is the **Market Risk** to which the banks (and not only) are exposed due to investing, trading and playing on the financial markets. Machine Learning is mainly proper for identifying inadvertent risk in trading behavior, for understanding the impact of the firms that trade on their market price, for establishing new patterns and connections between assets and how they influence each other or even for creating bots to constantly monitor the financial indicators and send alerts once a trade would be profitable.

Finally**, Credit risk** is one of the highest risks faced by banks and usually the one requiring the most capital, therefore its management is of high interest for the financial institutions.

The objective of credit risk management is to optimize the credit portfolio and reduce the risk of customers not meeting their obligations. The high and extensive complexity of credit risk assessment made this area proper for machine learning applications.

3. Machine Learning for Consumer Credit

Utilizing Machine Learning techniques is not a new trend, but it is a growing one. Back in the '90s, a comparative analysis between traditional statistical models of distress and bankruptcy prediction and an alternative neural network algorithm proved to be an effective combinations, with a significant increase in accuracy. [16] And the research in this area just started at that point. Over years, there were multiple implementations of machine learning techniques supporting risk management, which proved to be very efficient, making the most optimal decision.

The following part of this paper mainly focuses on the applications that were developed and deployed for consumer credits in the risk management area after the financial crisis.

3.1. Scoring Models Overview

One of the tools most used in the credit risk management are the credit scoring models, defined as statistical methods that consider financial indicators to predict the default risk of individuals or companies. These indicators are given a relative importance and are considered when predicting the creditworthiness, pointing out the probability of default of the borrower. [13]
In Table 3.1. are listed in a chronological order all the articles considered in this paper with the utilized algorithm(s). They are all tackling Credit Scoring models for Consumer Credit, reason why they were considered:

Article	**Author(s)**	**Year**	**Algorithm**
Credit scoring with a data mining approach based on support vector machines	Huang, Chen, Wang	2007	Hybrid Support Vector Machines
Support vector machines for credit scoring and discovery of significant features	Bellotti, Crook	2009	Support Vector Machines
Consumer Credit Risk Models via Machine-Learning Algorithms	Khandani, Kim, Lo	2010	Classification and Regression Trees (CART), Linear Regression
A Proposed Classification of Data Mining Techniques in Credit Scoring	Keramati, Yousefi	2011	Artificial Neural Networks; Bayesian classifier; Discriminant Analysis; Logistic regression; K-Nearest Neighbor; Decision Tree; Survival Analysis; Fuzzy rule-based system; Support Vector Machine; Hybrid Models
Loan Default Prediction on Large Imbalanced Data Using Random Forests	Zhou, Wang	2012	Random forest
Quantitative credit risk assessment using support vector machines: Broad versus Narrow default definitions	Harris	2013	Support Vector Machines
Large Unbalanced Credit Scoring Using Lasso-Logistic Regression Ensemble	Wang, Xu, Zhou	2015	Lasso logic regression
Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research	Lessmann, Baesens, Seow, Thomas	2015	Artificial Neural Networks, Support Vector Machine, Ensemble Classifier, Selective Ensemble Classifier, Threshold metric, Area under receiver operating characteristics curve,

			H-measure, Statistical Hypothesis Testing.
redit scoring with a feature selection approach based deep learning	Van-Sang, Nguyen	2016	Deep Learning
A novel multistage deep belief network based extreme learning machine ensemble learning paradigm for credit risk assessment	Yu, Yang, Tang	2016	Deep briefing Network, Extreme Machine Learning
Ensemble Learning or Deep Learning? Application to Default Risk Analysis	Hamori, Kawai, Kume, Murakami, Watanabe	2018	Bagging, Random Forest, Boosting

Table 3.1. Articles list on Scoring Models for Consumer Credit

3.2. Machine Learning Algorithms for Consumer Credit Models

First reference found is the study presented by Huang, Chen and Wang in 2007 that proposed a hybrid SVM-based credit scoring model that analyzes the credit applications based on the information provided by applicants. [17] This application would help the banks decide whether to grant credit to consumers who require it. The authors compared the SVM classifiers with neural networks, genetic programming and decision tree classifiers, and even though the number of imputed features was low, the result was comparable with the other techniques' outputs. Moreover, the SVM algorithm was proven to be very effective in the classification area, successfully grouping credit applications into accepted or not, hence minimizing the money lose due to underperforming credits. Lastly, the researched revealed that if the SVM classifiers are combined with genetic algorithms, they can serve multiple purposes: performing feature selection tasks but also optimize the model parameters.

Even though most of the conclusions were positive, the authors outlined some negative sides of the hybrid Support Vector Machine – Genetic Algorithms models too. One of their downsides would be the long time required for training and the high computational complexity demanded for a good classification accuracy. Another inconvenience would be the "black-box" nature of SVMs, but this could be overcome with the use of SVM extraction techniques or the use in combination with other interpretable models.

In 2009, another article focusing on SVMs for credit scoring was published by Belloti and Crook. They compared three traditional techniques (logistic regression, discriminant analysis and K-Nearest neighbors) against SVMs using a dataset of around 25000 customers. The results reinforced the outcomes of previous researches (that SVMs are successful in classifying credit card customers), but also revealed that they can be well applied selecting the features that have the highest impact on likelihood of default. They also discovered that a very important indicator in this analysis is the type of credit card, as this could influence the other variables to be examined in the model.

Though, one major disadvantage acknowledged is the high number of support vectors required for the best performance, mainly due to the broad indicative nature of credit data. [18]

In 2010, Khandani, Kim and Lo [19] published a study in which they were using machine-learning techniques for forecasting models of consumer credit risks. The framework used consisted of generalized classification and regression trees (CART)

By combining credit bureau data with customer transactions and applying linear regression R^2 with a delinquency rate of 85%, they reached a better accuracy of classification rates for the default of an application. They also analyzed the patterns of the time-series of delinquency rates and concluded that aggregated consumer credit-risk analytics may have a high influence in forecasting systemic risks. Moreover, by applying machine learning forecast models on the decision to cut credit lines, they estimated a cost saving between 6% to 25%.

On the same topic, Keramati and Yousefi presented a study in 2011 in which they acknowledge the importance of analyzing the huge amount of data generated by credit scoring in a fast-growing credit industry, but also the human impossibility of manually reviewing and interpreting it, hence the need of data mining techniques to support this effort. In their paper, they analyzed ten different data mining approaches and their results outlined the following [20]:

- K-Nearest Neighbor (K-NN) proved to obtain the best results for the credit scoring purposes;
- The Employ Multi-Group Hierarchical Discrimination (M.H.DIS) resulted to have better classification abilities than the traditional models;
- Support Vector Machine – MARS (SVM MARS), logistic regressions and neural network are very good for classification, but LDA and CART are easy to use in building such a model,
- Integration of Self Organization Maps (SOMs) with supervised classification methods proved to bring more advantages.
- Kernel Based RBF neural network was the best choice in identifying the true positive.
- The comparison between discriminant analysis, logistic regression, neural network and regression trees for predictions and classification tasks outlined that CART and neural network are the best to apply for best results.
- When evaulating the accuracy of K-NN, SVMs and neural networks, it resulted that the integration of all these methods with effective feature selection improved the accuracy of the classification.

One of the main conclusions they draw was that calculating the probability of default for an applicant if more meaningful that classifying them into the binary classes.

One year later, in 2012, Zhou and Wang proposed a study in which they applied improved random forest algorithms in the binary classification field, by attributing weights to the decision trees in the forest. These weights were calculated based on the previous performance, namely errors in training. Their approach outperformed expectations, beating the result of benchmark algorithms like traditional random forest, SVMs, KNNs in terms of accuracy and proved that parallel random forests can considerably reduce the learning time [21].

In 2013, Terry Harris published an article on Support Vector Machines (SVMs) applied for credit-scoring models from two perspectives: a broad one, considering the credits that are less than 90 days past due, and a narrow perspective, analyzing the credits that are more than 90 days past due, reaching the conclusion that the last produced more accurate, mainly for severe cases of default. The main explanation for this conclusion could be the greater number of cases fed to the model, leading to a better learning of the pattern for un-creditworthiness. [22]

Wang, Xu and Zhou published a new article in 2015, in which they outlined a new mix of algorithms for credit scoring that exceeded expectations and previously known results. Their approach consisted of applying clustering and bagging algorithms to generate balanced training data and diversify data, applying Lasso-logistic regression ensemble to evaluate credit risks. [23]

During the same year, Lessman, Baesens, Seow and Thomas updated the study started by Baesens et al., including new classification algorithms used in the credit scoring area. They considered 41 classifiers for 8 credit scoring data sets and their results proved that there are more performant classifiers than the standard logistic regression. More than that, they outlined the

business value add of improving the prediction models, variable selection and data quality and suggested that focus should change into these areas. [24]
One year later, in 2016, Van-Sang and Nguyen published a study on Deep Learning, a powerful classification tool that provides training stability, generalization and scalability with big data. This method surpassed results previously obtained with baseline methods and showed competitive performance with other feature selection models extensively used in credit scoring area. The study also outlined that fewer features considerer for the evaluation procedure allow for collecting essential variables, hence reducing the resources allocated on performing the research. Moreover, parallel processing proved to decrease processing time, whilst obtaining the same results. [25]
During the same year, Yu, Yang and Tang proposed a novel multistage deep belief network based extreme learning machine (DBN – based ELM) ensemble learning methodology as a promising mix for credit scoring problems. These 2 techniques were already known for the time-saving characteristics and for the high-learning capacity thru hidden layers. The structure of multistage ensemble learning model, working in three stages, conducted to better results than typical single classifiers. The steps followed for this analysis are the following: in the first stage, the bagging sampling algorithms are applied to generate multiple and diverse training subsets of data; during the second stage, the ELM is utilized as classifier and different ensemble components can be properly defined with right subsets and different initial conditions. The last stage merges the individual results to form the final classification output thru the DBN model, which can effectively outline the relevant information hidden in ensemble members. [26]
In 2018, Hamori, Kawai, Kume, Murakami and Watanabe published an article in which they assessed payment data and compared the prediction accuracy and the classification ability of three ensemble-learning methods with neural-network methods. The three methods assessed were bagging, random forest and boosting. The study outlined that the boosting method has a superior classifying ability, even when compared to neural networks. The performance of the lastly mentioned proven to be highly dependent on the choice of activation function, dropout inclusion and number of hidden layers. [27]

3.3. Discussions

All these articles are approaching the same topic from different angles and thru different methods. The observations and conclusions obtained by the authors can be further leveraged in analysis, hence improving efficiency and accuracy by using the already known results.
Overall, the main idea that was outlined by each article is that Business Analytics plays a key role in the evolution of the financial institutions and in the optimization of their processes. When applied, it minimizes costs by reducing the number of human resources involved and increases the accuracy of the decisions.
Credit scoring algorithms assigns numerical values to the client outlining whether the entity is likely to default or not. Most of the studies were focused on this area, treating it as a classification problem in order to facilitate the credit decision, but also minimize the credit risk exposure.
Machine Learning algorithms performed better than the traditional techniques in classification steps and obtained increased prediction accuracy. The SVMs were widely tested and proved to be very effective in the classification area and in the feature selection process, especially when combined with genetic algorithms. Another algorithm that exceeded expectations was the Random Forest applied in the binary classification area, which showed outstanding results and a reduced learning time. Deep learning was outlined as one of the tools that provides training stability, generalization and scalability.

Beside all these, another valid point that should be considered is the dataset used for researches: it should be varied, integer, divers, to cover as many scenarios as possible and reduce biased results. Having said that, if the availability of real data would increase, if would encourage more researches on evaluating all the problems encountered in credit risk management, and not only.

4. Conclusions

Companies acknowledged the importance of adopting new technologies and the business value it creates and it is expected that financial institutions will increase the machine learning applications in the risk management fields to enhance their capabilities.

Even though these applications have some known limitations, a major one being the inability of understanding the mechanism used to reach a decision (mentioned in the literature as “operating like a black box”), the business value it creates it significantly higher, main benefit worth outlining being: high complexity of analysis performed, limited number of human resources involved, minimal error and reduction of performing time.

Machine Learning proved to be evolutionary and promises to revolutionize the way things work, hence it has the potential to transform the risk management area and enables the discovery of complex, nonlinear patterns in broad datasets.

This paper introduced an assessment of the researches around credit scoring algorithms for consumer credit within the banking industry, mostly because credit risks is considered the highest risk for a financial institution. However, the advantages and disadvantages of various machine learning tools for credit scoring can be further studied to refine them, improve results and maximize their values.

In conclusion, even though there have been different studies performed in this area, there’s still room for research and improvement to extend the beneficial applicability and impact of machine learning in the financial field and not only.

References

[1] "Gartner Glossary," Gartner, Available: https://www.gartner.com/en/information-technology/glossary/business-analytics.

[2] M. Dwight, A framework for Applying Analytics in Healthcare – What can be Learned from the Best Practices in Retail, Banking, Politics and Sports, Pearson Education Inc, 2013.

[3] G. Doukidis, N. Mylonopoulos and N. Pouloudi, Social and Economic Transformation in the Digital Era, IGI Global, 2004.

[4] B. Raghynathan and R. V. Maiya, SMACing the Bank - How to Use Social Media, Mobility, Analytics, and Cloud Technologies to Transform the Business Processes of Banks and the banking Experience, CRC Press, 2018.

[5] T. Groenfeldt, "Forbes," 24 Dec 2014. [Online]. Available: https://www.forbes.com/sites/tomgroenfeldt/2014/12/24/american-express-opens-tech-lab-in-palo-alto/.

[6] I. Schmerken, "WallStreet and Technology," 10 Feb 2014. [Online]. Available: http://wallstreetandtech.com/asset-management/fidelity-labs-takes-innovation-to-the-next-level/d/d-id/1316288d41d.html?.

[7] R. Browne, "CNBC," 18 Nov 2019. Available: https://www.cnbc.com/2019/11/18/banks-must-behave-like-tech-companies-to-survive-amid-fintech-threat.html.

[8] J. R. Evans, Business Analytics - Methods, Models and Decisions, Second Edition, Pearson Education, Inc., 2016.

[9] H. Chahal, J. Jyoti and J. Wirtz, Understanding the Role of Business Analytics, Spinger, 2019.

[10] G. Sunila, Practical Machine Learning, Packt, 2016.

[11] O. Theobald, Machine Learning from Absolute Beginners, Scatterplot Press, 2017.

[12] D. TEAM, "Data Flair," 1 Jan 2019. [Online]. Available: https://data-

flair.training/blogs/advantages-and-disadvantages-of-machine-learning/.
[13] L. Martin, S. Suneel and K. Maddulety, "Machine Learning in Banking Risk Management: A Literature Review," Risks, 2019.
[14] I. A. Moosa, Operational Risk Management, Palgrave MACMILLAN, 2007.
[15] T.-M. Choi, H. K. Chan and X. Yue, "Recent Development in Big Data Analytics for Business Operations and Risk Management," IEEE Transactions on Cybernetics, 2007.
[16] E. I. Altman, M. Giancarlo and F. Varetto, "Corporate distress diagnosis: Comparison using linear discriminant analysis and neural networks (the Italian experience)," Journal of Banking & Finance, 1994.
[17] C. L. Huang, M. C. Chen and C. J. Wang, "Credit scoring with a data mining approach beased on support vector machines," ScienceDirect - Expert Systems with Applications, vol. 33, pp. 847-856, 2007.
[18] T. Bellotti and J. Crook, "Support vector machines for credit scoring and discovery of significant features," ScienceDirect - Expert Systems with Applications, vol. 36, pp. 3302-3308, 2009.
[19] A. E. Khandani, A. J. Kim and A. W. Lo, "Consumer credit-risk models via machine leraning algorithms," Journal of Banking & Finance, vol. 34, pp. 2767-2787, 2010.
[20] A. Keramati and N. Yousefi, "A Proposed Classification of Data Mining Techniques in Credit Scoring," in International Conference on Industrial Engineering and Operations Management, Kuala Lumbur, Malaysia, 2011.
[21] L. Zhou and H. Wang, "Loan Default Prediction on Large Imbalanced Data Using Random Forests," TELKOMNIKA Indonesian Journal of Electrical Engineering, vol. 10, pp. 1519-1525, 2012.
[22] T. Harris, "Quantitative credit risk assessment using support vector nachines: Broad versus Narrow deafult definitions," Elsevier - Expert Systems with Applications, vol. 40, 2013.
[23] H. Wang, Q. Xu and L. Zhou, "Large Unbalanced Credit Scoring Using Lasso-Logistic Regression Ensemble," PloS Ones, 2015.
[24] S. Lessman, B. Baesens, H.-V. Seow and L. C. Thomas, "Benchmarking state-of-the-art classiÞcation algorithms for credit scoring: An update of research," European Journal of Operational Research, 2015.
[25] V.-S. Ha and H.-N. Nguyen, "Credit scoring with a feature selection approach based deep learning," MATEC Web of Conference, vol. 54, 2016.
[26] L. Yu, Z. Yang and L. Tang, "A Novel Multistage Deep Belief Network Based Extreme Learning Machine Ensemble Learning Paradigms for Credit Risk Assessment," Flexible Services and Manufacturing Journal, vol. 28, 2016.
[27] S. Hamori, M. Kawai, T. Kume, Y. Murakami and C. Watanabe, "Ensemble Learning or Deep Learning? Application to Default Risk Analysis," Journal of Risk and Financial Management, 2019.

Claudia ANTAL-VAIDA is a graduate of the Faculty of Economic Cybernetics, Statistics and Information at the Bucharest Academy of Economic Studies, bachelor’s degree in Cybernetics and master’s degree in Business Analysis and Enterprise Performance Control. She is currently a PhD Student at the same University, mostly interested in business analytics, machine learning and data structures.