# Application of AI and ML on Indian Banking Credit Rating System

**Dr. Mainak Chakraborty**
Assistant Professor, Department of Management,
School of Business and Economics, Adamas University
Phone - +91 86177 48110
Email – cmainack2013@gmail.com

**Sourav Kumar Das**
Assistant Professor, Department of Commerce and Economics
School of Business and Economics, Adamas University
Phone - +91 87776 25181
Email – skd84201@gmail.com

***Abstract***

AI and ML are the driving force of the modern financial industry today. From customer data analysis to investor requirement analysis, they are everywhere. The BlockChain system allows the data to be gathered and analyzed in a very systematic way. The way FinTech is evolving is giving a new dimension of credit analysis of customers. How AI is transforming financial wellbeing as well as financial analysis can be analyzed throughout the analysis. There are various ways to analyze the impact of AI and ML in the credit rating system. The credit rating in its sense analyzes the creditworthiness of the customers. These ratings are available through different credit rating agencies across different platforms. The analysis involves the application of Random Forest algorithms and Logistic Regression Analysis. The study also compares the effectiveness of these two algorithms on credit analysis. The study collects the observations from the Indian Banking sector to analyze the credit rating system. From the analysis, it has been proved that Random Forest Algorithm proves to be the best solution than the Logistic Regression Analysis.

***Keywords**:* Artificial Intelligence; Machine Learning; Credit Scoring; Random Forest Algorithm, Regression Analysis.

## 1. Introduction

Over the last 10 years, the use of Big Data, Business Analytics, Machine Learning, and BlockChain technologies changed the way the finance domain used to work. The advancement of information technology pushes the finance industry to use the latest technologies so that it can manage the growth of technology, FinTech emerged as a pool of opportunities to the investors and the users of the financial resources by using the recent tools.

Financial tech industry investment has already continued to rise worldwide in 2018, according to KPMG's "The Pulse of Fintech Report," with a roughly 12-fold rise from 2010 to 2018. With

2196 agreements, global Fintech investment totalled $111.8 billion. On the fintech ecosystem, several new areas are emerging, such as South America and Africa.

The advancement of FinTech calls for an improvement of the competitiveness among the firms that use FinTech as a tool to attract consumers. Some of the FinTech firms use Autobots to analyze the data automatically which aids in a quick and real-time data analysis. It also provides the clients of those companies to easily understand the market behaviour and to flow their investments in that sector. This also makes the services offered by those companies cheaper than those who do not use FinTechs. As a result, conventional financial service companies and banks have started to recognise the value of user experiences in building client loyalty and increasing market share, and some have sought to acquire or collaborate with Fintech enterprises.

FinTech firms' use of Artificial Intelligence and machine deep learning has aided in the restructuring of client interactions, such as customer contact, from conventional face-to-face contact to interactive engagement through online platforms with no human involvement. Starting from understanding and building Client Risk Profiles to informing the floor traders of the requirement of the trades, everything is handled by FinTech solutions. The client, as well as the trader both, can simultaneously see what is going on in the market.

Below are the project's contributions:

1. The bulk of extant research focuses on the influence of machine learning on a single Fintech service on a worldwide scale, independent of population.
2. However, the emphasis of this research is on evaluating the credit risk of Moroccan customers using a machine learning system.
3. The outcomes of the research offer clues for modifying Indian bank risk management methods by comparing algorithms.

## 2. The Way AI and ML and Shaping the Future –

First and first, it is necessary to distinguish between AI and machine learning, which is not always an easy distinction to make. Informally, AI and ML is not part that can be easily understood by businesses, thus they prefer an easier understanding of the subject matter. The most frequent definition of AI is intelligence displayed by computers, with intelligence defined in terms of what we consider intelligence to be in humans (Shieber, 2004). The mathematical algorithms fed in by human beings, which will help the system to understand the exact processes, to perform distinctive jobs, is known as machine learning. ML is a kind of AI system that helps to judge human behaviour on a predictive system, giving some input and data on how the human brain works. ML stands as a subset of AI which often helps to understand the mathematical aspects of the system. AI also includes much more updated computerized skills and techniques including advanced logical reasoning, coding systems, and other information technology solutions to build a robust system.

**2.1 Business and AI and ML –**

The term "artificial intelligence" is not a new one; it was first suggested more than 60 years ago. AI capabilities were only pushed out with the advancement of computers and storage. BCG Henderson Institute commented that the two biggest achievements of AI are, Natural language processing (NLP) and visual recognition of objects through defined mediums. The influence of AI on all businesses is undeniable, particularly in the financial sector. Artificial intelligence has lowered the amount of standardised and repetitive labour, changed the nature of certain employment, and improved efficiency. Simultaneously, new roles such as Data Scientist and Big Data Engineer Profile have been developed.

Previous to the advent of AI, businesses were more dependent on traditional tools and techniques like trend analysis, data mining, forecasting and scheduling, different statistical analysis, casualty analysis etc. With the advent of AI, the dependency on traditional tools decreased relatively. As a result, AI can help with risk management, marketing, and other areas by providing relevant analysis and intelligent decision-making. In other words, artificial intelligence tries to use advanced algorithms to better comprehend business operations, customer behaviour, and market trends by leveraging and processing the availability of enormous data sets. Furthermore, AI encourages educated decision-making, providing businesses with an advantage over their rivals. Credit scoring based on social networks, for example, optimising current scores or assessing individuals without a credit score are all possible using AI. The possibility of utilising improved machine learning algorithms in credit scoring is shown by Baidu, a Chinese corporation.

**2.2 Efficiency and Labor Cost Improvement**

They attract more customers and utilisation of new data sources, such as smartphones, tablets, and internet of things devices, that are characterised by geolocation, face, and voice recognition, have drastically altered how organisations access and utilise data. In reality, AI will have three major impacts on the commercial world: automated analysis, intelligent analysis and data processing, and, creation of new ideas and improved systems. As a result of intelligent analysis of the situational variables, through automation, new business models can easily be developed as it will also tell us the effectiveness of the models (David et al., 2018). AI is continuing to alter the finance sector by allowing for the automation of manufacturing processes as well as the automated generation of highly accurate calculations and reports. The fundamental benefit of automation is that it allows a company to boost overall productivity since robots, unlike humans, do not burn out or need rest periods (Hislop et al., 2017).

AI is influencing the type of employment and their efficiency as time goes on, by rendering certain jobs obsolete, generating new ones, and improving their efficiency. According to the BCG report, the banking sector would be the worst-hit sector where they would lose 1.06 million jobs by 2027, a decrease of 22% in terms of job cuts and efficiency gains. The other 78 percent of employees will experience a 42 percent efficiency boost, which equates to 2.4 hours per person per day saved in the same functional activity. As a result, AI is expected to lower working hours in the banking industry by 27% by 2027, according to a conservative estimate.

To summarise, artificial intelligence will dramatically enhance firms' capacity to generate future projections while lowering costs and increasing efficiency, resulting in better corporate production choices. Generally, AI is used for real-time analysis of data to be delivered at once (Partanen et al., 2017). Along with that, attainment of optimum operating efficiency will be the call for every firm as it aid to increase in competitive advantage. To further comprehend the AI mathematic component, we'll go through how machine learning plays a role in Fintech applications.

**2.3 Machine Learning Tools and Techniques –**

Machine learning, in particular, is gaining popularity as a result of digitised data, quicker machines/computers, and better algorithms for analysing acquired data.

When a computer or system learns from data, we refer to it as machine learning, but there are a few variances depending on the kind of machine learning utilised. In reality, machine learning may be divided into two types: supervised by a human being and unsupervised that is without the intervention of the human being. In the first mode, the AI reads a set of data to gather information and generate results on this. This is common to the traditional statistical data analysis model where the impact of independent variables on the dependent variables is seen. This is also known as Target Analysis.

Unsupervised learning is when you just have input data and wish to extract meaningful information or patterns from it while also learning more about the data structure, for as by grouping data by comparable qualities. Deep Leaning is another variant where the above two methods are combined to get a more efficient result.

Regression and Classification Algorithms are two types of supervised learning algorithms. Regression equations are used for the closely related categories, and it is a statistical determinant when the predictive variable is a continuous one. Predicting changes in a company's turnover based on the marketing campaign budget and other elements supplied as input is one example. Where there is a qualitative variable or target, the classification algorithm will be used. If we look into a problem of assessment of the credit or debt risk, the dependent variable would be the risk that the borrower failed to pay the debt. Here the risk can be sub-divided into classes – risky and zero risk. The risky variable will depend on various other dependent variables which will fix the degree of riskiness.

In the case of unsupervised machine learning techniques, Regression, classification, and clustering are the three machine learning tools that can be used. In reality, the clustering approach will be used when the user wants to fetch important information about the data structure for the segments with similar behaviour. It becomes easy to analyse each segment and use the best marketing approach for each speciality in this manner. Exploring data to discover customer preference or fraud, for example, without knowing which observations are fraudulent and which are legitimate, is an example.

**2.4 Finance and Machine Learning -**

The quantity of data gathered in financial institutions (FIs) has expanded dramatically in recent years as digitalization has progressed, and the share of unstructured data (pictures, videos, sound, and text...) has increased. This sort of data is often referred to as "big data." As a consequence, AI and ML have been evolved as a tool for analyzing vast information to get the extracts within a second on a real-time basis.

In today's world, the application of AI in the financial sector has evolved on a vast scale. Sectors like Credit Rating, Fraud Prevention, Automatic Trading, Financial Markets, ML has taken a way forward on how the sector should behave.

**2.5 Credit Score – the Introduction**

There has been a plethora of academic work on the application of machine learning approaches to estimating credit risk since the early 2000s. To assess a company's solvency, for example. Support Vector Machine (SVM) was employed by Auria and Moro (2008), who discovered that it produces more accurate predictions beyond the sample than other algorithms. Khandani et al. (2010) use a big dataset from a commercial bank to predict consumer credit risk using generalised classification and regression trees (CART). In analyzing Credit Risk for the customers (individual and corporate) ML has taken a lead over another system. The system involves real-time processing which includes the client's history, risk profiling, investment need, returns expected, loan repay analysis, etc.

The models like Linear, Logit, Probit are used to estimate the credit risk for financial institutions which also employ stress testing and internal risk management process which is employed to minimize the problems associated with the model interpretation. However, more complicated models, such as the neural network, Random Forest, or G-Boost, produced higher accuracy and lower prediction error. As a consequence, several financial institutions have lately begun experimenting with the use of machine learning approaches to enhance financial risk forecasts, owing to its capacity to grasp unstructured data semantically.

Furthermore, applicants for credit scores for a variety of goods, such as mortgages and car loans, have used a significant amount of credit history. The number of customers eligible for the credit is restricted by this technique, which is based on a small collection of historical data. To estimate the consumer's credit potential models are getting constructed with extensive and full information about the client's data like income, financial goals, education status, future employability, expected salary, future growth expectation, etc.

Consumers who may have been disqualified in the past may now acquire credit, according to the H2O.ai firm's research AI makes it possible to grow the number of clients with credit while offering a more accurate risk estimate who were previously missed from the band. The above-mentioned Baidu Chinese corporation is an example, but there are many more comparable companies functioning in this field.

Apart from the financial sectors, AI and ML also got an impact on other sectors like e-commerce. AI matches recommendations based on the client data gathered by the system. AI coupled with deep learning helps to make them a more effective suggestion to the clients and also helps the consumers to search products and to receive suggestions on the updated products. With actual examples, Yashoda (2018) examines AI-based inventory management technologies used in the e-commerce business in his paper.

**3. Literature Review –**

A literature review is one of the key elements of any research proposal or project. It identifies what happened there in the same field of evaluation as well as the present studies aiming at this particular phase.

According to Jennifer Ifft et al. (2018) the digital methods which are now the buzzword for credit scoring, have differentiated the use of the standard models used in econometrics and statistics with the use of the logistics regression model. According to the authors, there is five broad areas like Generalized line models, Bayesian Models, ensemble models, support vector machines, and nearest neighbour models.

According to Wijewardhana et al. (2018), there is much evidence that the logistic regression model can be successfully used to determine the classification problems not only on credit scoring, but also on debt recovery, and bad debts management.

According to Kandpal and Mehrotra (2019), the demonetization of 2016 helped the fintech entrepreneurs to look back to the underserved population, increasing the flow of debt, as well as rationalizing the credit scoring mechanisms.

As per Leo et al. (2019) the FinTech often faces difficulty in managing the four basic types of risk like credit, business, liquidity, and market risk. AI and ML helped a lot in managing this kind of risk in the firms.

According to Ozgur et al. (2020), there is a non–linear and non – parametric relationship with drivers of bank credit and outstanding bank loans at a single point in time.

According to O’Neill and Biallas (2020), AI and ML are an integral part of fintech analysis. As per their analysis, 85% of the banks use fintech solutions for the improvement of their efficiency and competitiveness.

Rafiei and Moradi (2020), studied the demographic parameters of the people who want to know their credit score for any reason. They have emphasized that there is hardly any study that helps to understand the effect of AI and ML on the credit rating system.

Linh et al (2019) in their study on Vietnamese farmers showed that the farmers have access to very little credit, and the financial institutions are not eager to analyse their needs or support their needs. The FIs of the opinion that they in general have a poor credit score so, loans can not be provided to them.

Bennouna and Tkiouat (2018) in their research paper said, that FIs use Fuzzy Logic – one of the ML-based credit rating systems which helps to identify the behaviour of the human being to predict the way the score may move.

**4. Credit Scoring in the using ML: Case of Indian Banking Industry**

Credit scores are used by banks to assess a customer's creditworthiness using a numerical score based on a variety of factors. Previously, Indian Banks were used outdated data modelling applications to rate their customers, based on statistical modelling. But with the advent of AI and ML, they have also joined the league to do more extensive research on credit rating tools available and to use them to rate their clients efficiently and error-free.

**4.1 Use of Single Statistical Model of Credit Scoring – Limitations**

Regression algorithms are the most well-known statistical model used in supervised learning to predict. Statistical models have been employed by businesses for a long time because of their common advantage: Because the variables are properly expressed as a mathematical equation, Statistical Models are easy to comprehend.

These statistical models, on the other hand, have grown less accurate and adaptable as new algorithms have arisen. The next sections outline the logistic regression method that we are interested in in this research, as well as the limitations of its use.

**4.1.1 Logistic Regression**

Logistic regression is a statistical procedure that is well-known for its simplicity. It entails categorizing data or output variables into binary or nominal extremes and separating them using a logarithmic line. David Cox created it and it is recognized as the most widely used statistical model in credit scoring. The model by Cox on the Logistic Regression assumes that the response variable belongs to a certain class of the continuous value and ranges from 0 to 1.

This is the chance of granting a loan or not in credit scoring. The goal of logistic regression is to build a mathematical model out of a collection of explanatory or predictor factors to predict a log transformation of the dependent variable.

Consider the standard example:

Assume,

A single independent/explanatory variable X,

Binary/Boolean dependent variable, Y.

Also assume that Y = 0 & 1.

Take 1 as Risky and 0 as no risk.

Let, p is the percentage of observations with Y = 1

Then 1-p is the probability of Y = 0.

The logistic regression equation is used in this scenario.

$$In\left(\frac{p}{1-p}\right) = \beta 0 + \beta 1\, X$$

β's are regression coefficients.

Here, the probability can be calculated using the inverse function that isknown as logistic.

There have been several studies that have compared Logistic Regression to other machine learning methods. Stephan and Lucila (2002), compared the use of AI neural networks and the logistic regression models for the categorization of biological data. Their analysis reveals that, depending on the sector of application and the goal sought, each algorithm has benefits and drawbacks. Elizabeth et al. (2016) analyzedML and logistic regression approaches for predicting hypertension, and discover that SVM outperforms logistic regression.

In actuality, Logistic Regression (LR), which is widely employed in Moroccan banks for credit scoring, has several drawbacks, including:

(1) LR requires assumptions such as variable normality, which is not usually confirmed in practice. (2) Overfitting is associated with a large number of parameters added in the model, resulting in a model that cannot generalize.

LR, on the other hand, is prone to underfitting, resulting in a simple model with a small number of parameters.

(3) Non-linear characteristics must be transformed in LR.

(4) It will have trouble handling a high number of categorical variables, which necessitates a transformation for each variable.

(5) LR isn't flexible enough to capture more complicated patterns or connections inherently.

### 4.2 The Proposed Non – Parametric Algorithm for Banking Credit Risk Scoring

Due to the lack of flexibility of logistic regression as a result of the arguments raised above, this research recommends using modern non-parametric algorithms like Random Forest, which are recognized for their accuracy and resilience, to score credit risk.

Machine learning models' credit scoring effectiveness has already been shown in several applications by international institutions such as Bank of America and Wells Fargo. It is better to compare with Random Forest Technique than to compare with other traditional statistical tools.Bootstrapping and Feature sampling are used in the model for data analysis, which is detailed below.

#### 4.2.1 Random Forest Algorithm – Introduction

The Random Forest is a kind of supervised ML algorithm that is widely used in regression and classification problems. It creates several decision trees. The trees are created using random data samples which are also known as Bootstrapping. It also uses a random set of variables or features which is also known as Feature Sampling. It is also known as a non – parametric model, as during the solution there is no assumption made about the frequency distribution. There is also a process which is known as bagging – it is a process of aggregating the bootstrap process, and this is also known as the first step of the random forest model. The process of bagging is set to pick "m" no, of the new dataset which is defined as Di, which is being extracted from the set D of "n" data size. With different split methods, multiple trees are being produced. Here each split contains different features and variables. A class is determined by each tree assigned to it. The voting process is initiated and the maximum voting determines the predicted class. The predicted class is determined by the trees with the highest percentage which is further assigned to one individual to a specified class. There is no limitation of the uses of several classes in the random forest model.

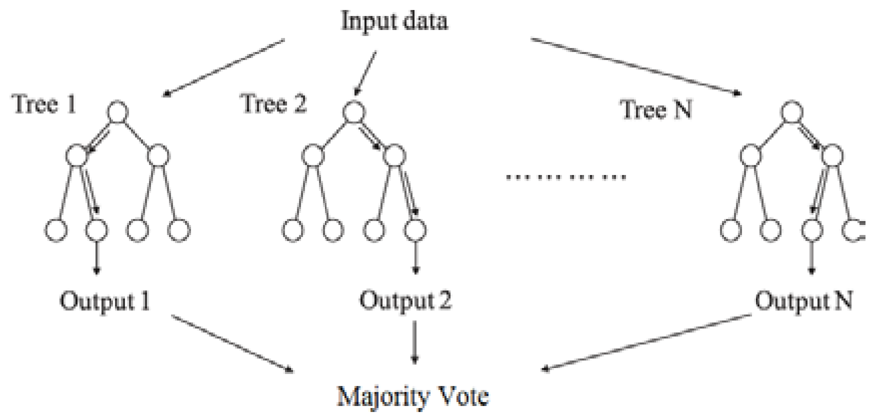


**Figure 1: Random Forest Model**

### 4.3 Credit Scoring Dataset and Its Application

Here, Random Forest and Logistic Regression algorithms are used on the same datasets with varying attributes at each stage to explain the extra value that machine learning delivers to financial sector applications. The purpose is to compare the accuracy of the results for each run to show that a nonparametric method outperforms a statistical model in Credit Scoring.

#### 4.3.1 The Dataset

Diverse datasets were gathered for this assignment to train and compare the models. We utilized genuine data from Moroccan bank clients who requested a personal loan. The dataset was not released to maintain data confidentiality. All of the data was organised in a certain way. On SPSS Software, data may be simply adjusted. The entire dataset, on the other hand, contains over 500,000 observations from 2015 to 2020, but due to performance issues when running the algorithm in the software, we were only able to use a subset of fewer than 5000 observations using Stratified Sampling, which is a method of reducing the sample size while maintaining the dataset's class balance. We gathered five separate datasets based on the demographic characteristics of the clients. Each dataset has a total of more than 7 randomly chosen characteristics. The target variable "Credit Risk," which is a Boolean variable that returns two values: 1 if the person is not solvent and credit will not be issued, and 0 if he is not risky, is present in all datasets. We used One-Hot-Encoding to encode all categorical variables during the Feature Engineering process to adjust Logistic Regression to provide appropriate results.

#### 4.3.2 Data Modelling

The IBM SPSS Modeler software platform, version 25.0, was used to carry out the full training and classification assignment. IBM SPSS Modeler is a piece of software that may be used for Feature Engineering and modelling. The Programme supports both logistic regression and random forests. The results of each algorithm were compared using IBM SPSS Statistics, which offered a large number of statistical tests to choose from. We employed a Partition to prevent overfitting by separating the data into two parts: 70% for training and 30% for testing. Every job to be completed has its node in the Classification stage, and the procedure consists of connecting nodes to produce a stream or flow chart, as shown in Figure 2.

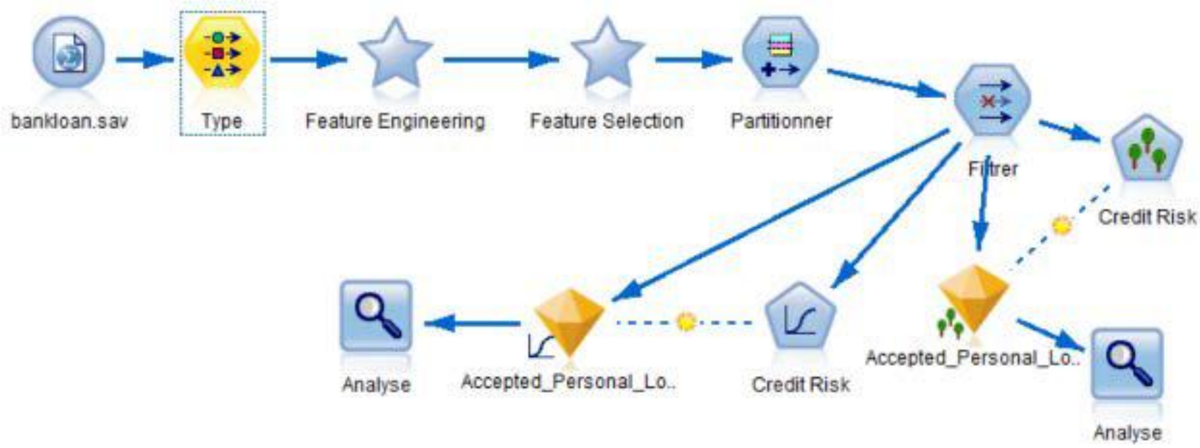


**Figure 2: SPSS Modeler**

### 4.3.3 Implementation and Analysis using Accuracy

Classification Accuracy is known as the percentage of the accurate prediction from the total number of observations of the event.To enable this accuracy, we have arranged each class with an equal amount of samples. Along with the classification accuracy, other factors like recall, the precision of data sets, F-Score, and Area Under ROC Curve, were also taken into consideration. Statistical tests were used to test the accuracy of the result obtained.

### 4.3.4 The Analysis of the Results

Two algorithms are tested using the five scenarios to be compared the result of which has been presented in Table 2.

We have the number of observations and predictor characteristics for each indexed dataset. 100 trees were used to train the Random Forest algorithms.

| Serials | A | B | C | D | E |
|---|---|---|---|---|---|
| No of Observations | 2500 | 3500 | 3800 | 1900 | 3900 |
| Random Features (No.) | 8 | 14 | 10 | 7 | 9 |
| Accuracy of Logistics Regressions | 82.78% | 81.42% | 78.39% | 87.10% | 88.91% |
| Accuracy of Random Forest | 94.88% | 89.77% | 85.74% | 88.52% | 90.21% |

**Table 2: Summary of data sets with accuracy based on Logistic Regression and Random Forest.**

The average performance of logistic regression is 83.72 percent across all data, and the average performance of random forest is 89.824 percent. Student's t-test, which determines if the difference in mean between the algorithms' accuracy is statistically significant or just coincidental has been used to compare the findings of the above two algorithms.

F = 0.997 with p = 0.351 is the test statistic for equal variance. We chose not to reject the null hypothesis of equal variance with this p-value.The equal variance has been ascertained by the test. We have calculated the t statistic where t = 2.642, and a matching p-value of 0.039. The significance level indicates how certain we are that the findings are significant rather than random. The resulting p-value is less than 0.05, which is the most frequent threshold of significance. As a result, the null hypothesis that the two performances are equal has been rejected.

From the above results, we can conclude that there isa difference in mean Random Forest and Logistic Regression performance and the difference is statistically significant. As a result, we may infer that in Credit Scoring, the Random Forest model outperforms the Logistic Regression model.

## 5. Summary and Conclusion

From the above discussions, it is evident that AI and ML have changed the way the financial sector behaves. It helps to tackle challenges like efficiency, performance, and increase in competitiveness of the system.

This project aims to compare the two methods – Logistics Regression and Random Forest – which are often used by the banks in India, to ascertain the credit scores. On analysis, it has been seen that the random forest helps to generate multiple decision trees, and it has a very low bias error, along with a large variance. The findings also demonstrate that the results based on Random Forest outperform the Logistic Regression. The difference in average accuracy between these two models is about 6.10% which is also statistically significant. So, in credit scoring – this particular model got a very positive impact on the data retrieved.

Finally, although having higher performance than linear or statistical models, the Random forest is still challenging to calibrate. In future research, we might evaluate the use of additional Machine Learning techniques that are recognized for their resilience, such as the Gradient-Boosting algorithm, which is used to discover non-linear feature interactions automatically.

## References:

1. Bennouna, Ghita, and Mohammad Tkiouat. 2018. Fuzzy Logic Approach Applied to Credit Scoring for Microfinance in Morocco. *Science Direct*,vol. 127: pp. 274-83.

2. Ifft, Jennifer, Ryan Kuhns, and Kevin Patrick. 2018. Can Machine Learning Improve Prediction – Application with Farm Survey Data. *International Food and Agri-Business Management Review,* vol. 21; pp. 1083 – 98.

3. Kandpal, Vinay, and Rajat Mehrotra. 2019. Financial Inclusion: The Role of Fintech and Digital Financial Services in India. *Indian Journal of Economics and Business.* vol. 19. pp. 85 – 93.
4. Leo, Martin, Sharma, Suneel, and Maddulety, K. 2019. Machine Learning in Banking Risk Management: A Literature Review. *Risks.* vol. 7. pp. 29 - 38.
5. Linh, Ta Nhat, Hoang Thanh Long, Le Van Choi, Le Thanh Tam, and Philippe Lebailly. 2019. Access to Rural Credit Markets in Developing Countries, the Case of Vietnam: A Literature Review. *Sustainability.* vol. 11. pp. 1468 - 1478.
6. O'Neill, Felicity, and Margarete Biallas. 2020. Artificial Intelligence Innovation in Financial Services. *International Finance Corporation, a member of the World Bank Group.* Vol. 85. pp. 1 – 8.
7. Ozgur, Onder, Erdal Tanas Karagol, and Faith Cemil Ozbugday. 2021. Machine Learning Approach to Drivers of Bank Lending: Evidence from an Emerging Economy. *Financial Innovation.* vol. 7. pp. 20 – 28.
8. Rafiei, Farimah Mokhatab, and Somayeh Moradi. 2019. A Dynamic Credit Risk Assessment Model with Data Mining Techniques: Evidence from Iranian Banks. *Financial Innovation.* vol. 5. pp. 15 – 23.
9. Wijewardhana, Udnai, Chinthaka Bandara, and Thesath Nanayakkara. 2018. A Mathematical Model for Predicting Debt Repayment: A Technical Note. *Australasian Accounting, Business, and Finance Journal.* vol. 12. pp. 127 – 135.
10. Shieber, S. M. (Ed.) 2004. The Turning Test: Verbal Behaviour as the Hallmark of Intelligence. *Cambridge: MIT Press.*
11. David He, Michael Guo, Jerry Zhou, and Venessa Guo. 2018. The Impact of Artificial Intelligence on Financial Job Market.
12. Auria. L., and R. Moro. 2008. Support Vector Machines as a Technique for Solvency Analysis. *Deutsches Institut fur Wirtschaftsforschung.* Discussion paper No. 811.
13. Khandani, E.A.J. Kimz, and A.W. Lox. 2010. Consumer Credit Risk Models via Machine Learning Algorithms. *Journal of Banking and Finance.* vol. 34. pp. 2767 – 2787.
14. Yashoda Kiran Lingam. 2018. The Role of Artificial Intelligence in Making Accurate Stock Decisions in E-Commerce Industry. *International Journal of Advanced Research, Ideas, and Innovations in Technology.* vol. 4. pp. 2281 – 2286.
15. Stephan Dreiseitl, Lucila Ohno-Machado. 2002. Logistic Regression and Artificial Neural Network Classification Models: A Methodology Review. *Journal of Biomedical Informatics.* vol. 35. pp. 352 – 359.
16. Held, Elizabeth et al. 2016. Comparing Machine Learning and Logistic Regression Methods for Predicting Hypertension Using a Combination of Gene Expression and Next-Generation Sequencing Data. *BMC Proceedings.* vol. 10. pp. 141 – 145.