Journal of Artificial Intelligence and Data Mining (JAIDM), Vol. 11, No. 1, 2023, 131-148.

Shahrood University of Technology

Journal of Artificial Intelligence and Data Mining (JAIDM)
Journal homepage: http://jad.shahroodut.ac.ir

Research paper

Investigating Revenue Smoothing Thresholds That Affect Bank Credit Scoring Models: An Iranian Bank Case Study

Seyed Mahdi Sadatrasoul*, Omid Mahdi Ebadati and Amir Amirzadeh Irani

Information Technology Management department, Management school, Kharazmi, Tehran, Iran.

Article Info

Article History:
Received 05 June 2022
Revised 07 December 2022
Accepted 11 January 2023

DOI:10.22044/jadm.2023.11982.2346

Keywords:
Data Mining, Credit Scoring, Credit Models, Revenue Smoothing, Stacking.

**Corresponding author: msadatrasoul@khu.ac.ir (S.M. Sadatrasoul).*

Abstract

Companies have different motivations for using and implementation of revenue smoothing based on the various possible effects that it has on income, costs, and expenses, as well as profitability, which can be seen in terms of changes in their annual financial statements. Revenue smoothing can affect credit scoring models reliability. It can cause to provide/not provide facilities to nonworthy/worthy organizations orderly, which are both known as decision errors, and are reported as "type I" and "type II" errors. This work investigates this issue for the first time in credit scoring studies on the authors knowledge and searches. The data of companies associated with a major Asian Bank are first applied using logistic regression, and different smoothing scenarios are tested. The study results indicates that traditional credit scoring models have significant errors when revenue smoothing procedures have more than 20% fluctuation in financial statement parameters using the Wilcoxon statistical tests.

1. Introduction

Companies with various objectives and motivations in their financial statements tend to make legal adjustments and changes based on their particular type of objective, which can ultimately lead to positive/negative changes and effects in the company's revenue as well as their profitability, which is called revenue/income smoothing. Furthermore, it is essential to know that revenue smoothing is applied in many organizations and companies; besides, it can be implemented both realistically as well as fraudulently, depending on the company's circumstances and objectives. Additionally, one of the main reasons for its implementation is to prevent major fluctuations in the overall performance, income, and profit of the company, specifically for those public companies that are listed on the stock exchange, for it can possibly have significant impact on the shareholder's perspectives. Moreover, the mangers typically engage on revenue smoothing, due to the fact that shareholders are always most interested in investing in companies that tend to have stability in their growth as well as minimum-levels of fluctuations in the overall revenue. It is important to know that there are also other motivations and incentive for the implementation of revenue smoothing, which include firm size and the identification of its precise share in the entire market [1]. There is another research work conducted by Holthausen [2], on the administrative costs, employee salaries, and service compensations claims and refunds, as well as the study and analysis on the variety of bonuses of CEOs, staff members, and board of directors. The Kanagaretnam, Lobo, and Mathieu research [3] study mainly focuses on the ownership control and job security of the managers. The Magicson, Jordan-Wagner, and Wooton [4] study is conducted mainly on the taxation on revenue, deviations in operational activities, income variability, and changes in sales volume, as well as the use of financial facilities at preferential rates that require special conditions depending on the industry. Studies on the effect of revenue smoothing on credit scoring models are not known on the authors searches.

It is extremely important to know that banks and financial institutions do, in fact, use these previously smoothed financial statements in order to determine whether that particular company is qualified enough to be granted financial facilities and guarantees. Additionally, such revenue smoothed financial statements tend to mainly represent a manipulated and false image of companies, and can very well lead to inappropriate and wrong decision-makings by banks. Moreover, since one of the most important and essential tasks of any bank or financial institution is to provide financial services in order to support the manufacturing and service industries, therefore, they are always looking to find for solutions and improve the current method in order to efficiently reduce the possible risks of providing financial facilities to the customers. It is also essential to know that due to the high risk of such activities, banks and financial institutions tend to have credit scoring models and risk measurement systems in place in order to increase the efficiency to allocate financial resources of their potential customers; besides, the methods or credit models used in this particular case are usually mathematical, computer-based or statistical credit scoring models.

In the current study, the authors have investigated the possible effects of revenue smoothing changes on statistical credit scoring models, and it is for this purpose that the research background is examined in the second part of the research; besides, the third section includes mostly discussions about the research methodology in particular, in which the researched credit models are often obtained by implementing logistic regression statistical algorithms on training data, in addition to their potential upgrades through bagging and boosting techniques. Furthermore, in the fourth section of the research, the data used from one of the country's banks, which includes the financial statements of a thousand different companies, is first introduced. Additionally, using this data, the companies that have requested the financial facilities are divided into two groups in terms of repaying the borrowed amount, which are essentially creditworthy and non-repaying or unworthy applicants. Moreover, in order to appropriately conduct the smoothing procedure, the accounting variables affecting revenue smoothing must first be identified; then the obtained data is divided into the two categories of test and train, and the testing data must slightly be changed. In addition, in order to change the effect of each step, the overall performance evaluation indices of the classification algorithms are evaluated. At the end, the table of these indicators is shown and fully examined through Wilcoxon and Kruskal-Wallis statistical test. Finally, in the fifth and final part of the research work, an ultimate conclusion is reached.

2. Research Background and Theoretical Foundations

Credit scoring or accreditation is essentially a set of decision-making models and techniques used in order to assist the creditors in providing services and accrediting. Additionally, in a much broader definition Louzada, Ara, and Fernandes [5] claimed that descriptive credit scoring was essentially a number achieved based on the analysis of the level of competency of customers and their creditworthiness, which can be used as an extremely useful tool and method for the assessment as well as total prevention of potential accreditation risks.

It is important to know that there are many credit scoring techniques used in order to form a credit scorecard; in addition, the logistic regression credit model seems to be the most commonly used amongst the banking industry due to it extremely desirable features and characteristics such as robustness and transparency.Although new techniques such as support vector machines have recently been applied and shown great and accurate credit scoring results, the obtained results are often extremely complex and their interpretation may not be so simple. As a result, according to Dong, Lai, and Yen [6], these particular credit models have not been widely used in practice, due to the competition growth and increased pressure to generate higher revenues that have led credit bureaus and other financial institutions to seek for more efficient ways to attract trusted customers while reducing potential risks simultaneously. Furthermore, the relentless efforts of marketers have led to greater risks for potential customers; in addition, the need for a quick and effective method of processing these risks has caused major upgrade and growth of software solutions, which can aid in credit scoring as well as guarantee limitations, and therefore, can ease the decision-making process subsequently. Moreover, according to Siddiqi [7], the current challenge for the managers to overcome is to decide a method or model that will not only assess the customers reliability successfully, but it must also be a credit model that has cost efficient processing power for each of these potential risks at the same time. Besides, customer service upgrades are essential and necessary in order to minimize the automated process of discrediting trusted and creditworthy customers, while maximizing the rejection of non-creditworthy who have the potential for default. Additionally, in the current state, the risk-based scorecard introduces a new powerful experimental method that is essential for many businesses as well as organizations. It is also important to know that the risk-based scorecard method has widely been used in many various industries such as non-repayment estimation and bankruptcy, as well as fraud, guarantee refund requests, and recovering the amounts owed on accounts. According to Thomas [8], this particular credit scoring method can very well provide a purposeful and efficient way of potential risk assessment, as well as being a consistent approach to minimize the possibility of system failures as the same time.

In another research conducted by Blochlinger and Leippold [9], one of the main tasks of banks as a financial service provider is discovered to be the ability to provide financial facilities and capital, in addition to the identification and classification of potential credit risks. Additionally, one of the major issues in the overall management of commercial banks is to measure and find out the actual creditworthiness of customers who refer to the bank for financial facilities. Furthermore, the appropriate design and implementation of credit scoring measurement models in the banking system can very well play an effective role in the rise of the productivity levels of the country's banks as well as improving efficiency and accuracy in the allocation of resources. Take note that, providing financial facilities to the right customer is, in other words, both art and science simultaneously. Moreover, the overall success of the banks may very much depend on various factors that include bank's total knowledge on their customers and the techniques they use, as well as the loan application provided by the customers and accurate credit scoring of potential customers. It is also important to know that in the recent years, banks have rapidly increased their use on credit scoring techniques in order to properly evaluate the financial facilities requests made by the customers. However, providing such financial services by any bank or financial institution is considered the first step in creating potential credit risks for them. In addition, in the traditional credit scoring methods previously used, financial facilities were offered to those applicants who had a positive net present value, and the rest with the negative net present value were mostly rejected; but with the rapid growth and development of statistical credit scoring techniques, the banks and financial institutions are finally able to use such credit scoring models in order to calculate the precise default risk.

According to Zhao [10], due to the current intense competition and rapid increase in the growth of the customer credit market, credit scoring models are strongly suggested and frequently used to evaluate precise credit allocations and worthiness. Take note that credit scoring is essentially a form of potential credit risk model used for the limitation of credit requests. Additionally, financial institutions and research developers use credit scoring models in order to properly address the issues that may rise during the evaluation process. Previously, this evaluation was done by an analyst using the rules and regulations that he himself had created for credit scoring or, in other words, accreditation. Furthermore, with the rapid increase of credit applicants, it was no longer neither economically nor workforce-wise possible to evaluate each customer this way. Therefore, according to Ince and

Aktan [11], as a result of these shortages, many brand-new methods were created and introduced in order to efficiently aid in credit scoring decision-makings. Moreover, at first, credit scoring models were created for mainly one purpose, and that was to classify and categorize the customers based on their various characteristics into two separate groups: acceptable or rejected customers. Besides, the core purpose of credit scoring models remains the same, which is essentially the classification of customers into two groups with slightly different names in comparison with the past including creditworthy (that refers to the customers who are likely to payback the borrowed amount), and uncreditworthy customers (who are mostly rejected due to their credibility default).

The term revenue smoothing refers to the company's management using their authority in the organization's financial statements to intentionally adjust the occurred fluctuations to their liking. Additionally, a number of research works have been conducted in relation with revenue smoothing including Beidleman [12], which showed proof as well as evidence that companies have, in fact, attempted to use revenue smoothing often. In addition, the conducted questionnaires by Graham, Harvey, and Rajgopal [13] also indicated that financial managers tend to be specifically interested in the shortcuts that revenue smoothing can possibly provide. Furthermore, according to Chen [14], the implementation and use of revenue smoothing can potentially have strong impacts on various factors such as cost of equity, earnings informativeness, and liquidity as well as bond ratings. Besides, in another research work conducted by LaFond, Lang, and Skaife [15] that focused on the effect of revenue smoothing on corporate stock liquidity risks, came to the conclusion that the implementation of revenue smoothing could, in fact, have an massive impact on the transparency of accounting data, and may very well lead the investors to be more willing towards investing in that particular investment; however, it is extremely significant to know that a reduction in transparency rate can possibly lead to less liquidity in the market. Moreover, revenue smoothing is usually used by the management in order to reduce or hide revenue fluctuation rates using accounting tools and techniques, which has recently become an extremely significant issue in the fields of accounting and finance to deal with. As previously mentioned, revenue smoothing is the use of accounting techniques to help shift the income and revenues from one period of time to another mainly for the financial statements to seem fixed and normal. It is also important to know that companies and organizations may go to great lengths in order to conduct revenue smoothing, because they are aware that the investors are much more likely to invest in organizations with fixed and predictable revenue streams rather than companies with high rate of fluctuations in their income and revenue. Additionally, keep in mind that the investors are constantly looking to increase their capital and profit margins, and since the forecast of future earnings is of great value to any investor, therefore, revenues and income with higher prediction rate and accuracy can be extremely effective and influential towards investors decision-making on whether to hold or sell their current shares. In addition, reported revenues has always held a special significance for the investors and their financial decision-making; besides, the general users of financial statements have always considered income and revenue as one of the more important factors in their reviews as well as their overall judgment on the organization, meaning any factor that can potentially affect the revenue is also of great importance, because there are economic consequences for it. As a result, revenue smoothing is considered one of the most important factors that can possibly affect the overall revenue as well as its report. Take note that it is possible for the banks to make the wrong decisions in their reviews, when it comes to lending finance to companies and organizations. However, this error or mistake is often caused by companies showing false financial statements in the structure of their expenses and overall costs as well as their revenues, which may seemingly be profitable (in other words, showing a false positive credit score in order to gain the trust of investors or can even show an adjusted loss and negativity in their financial statements (in order to reduce the rates or possibly evade the taxation required from them) based on their needs at the time, whether it is to show false positivity and stability in order to gain the trust of investors and banks or perhaps providing a negative review when needed in order to potentially evade taxes. Additionally, this article specifically tends to examine the effect of revenue smoothing on customer credit scoring models using the data obtained from one of the banks of Iran.

According to what was previously discussed on the topics of research backgrounds and theoretical foundations and their reviews, it was obtained that there have been many studies conducted on various credit scoring models and techniques; however, every single of the research works provided their final obtained results and outcome mostly on the basis of performance appraisal indicators, which can be misleading. Additionally, there have been many discussions on revenue smoothing in the fields of accounting and financial science, in addition to a number of researches that have been conducted on the factors affecting the revenue smoothing process in general, which most importantly include motivations, reasons, advantages, and drawbacks of revenue smoothing implementation. However, there has never been a major relationship found between the implementation of revenue smoothing and credit scoring in any of the previously conducted research works. Furthermore, for this particular reason, the authors of this article decided to examine the effects of revenue smoothing on the credit scoring models used in the banks and financial institutions. Finally, the rest of the article includes a step-by-step description of the research work, and then at the end, the credit model procedures used in the research work are shown.

3. Research Methodology

The purpose of this article is to evaluate the revenue smoothing effect on credit scoring problem; therefore, in the first step of this section, the research methodology steps is discussed, and in the second section, the hypothesis development is presented.

3.1. Research methodology steps

This particular article tends to use bagging and stacking techniques, due to the frequent discovery of unbalanced data in the credit scoring rates. The bagging estimation technique developed by Breiman [16] is essentially a method used for the production of multiple versions of a predictor classifier, using all these various versions in order to arrive at an optimal integrated predictive credit model. Additionally, when the dependent variable (predictable variable) has a numerical output, the credit model attempts to find out the mean, and at other times when the dependent variable is considered class-by-class, it tends to multiply the plurality vote. In addition, Skurichina and Duin [17] used the bagging technique in order to improve the current poor-quality of classifiers. It is also important to know that these techniques are based on the integration as well as combination of such classifiers. Moreover, a simple or perhaps weighty plurality vote is often used as the combination of rules for such techniques; however, other combinations can also be used including averaging, multiplication, and mediation. The term stacking is essentially a technique in which the estimations of a set of classifiers must first be implemented as the input variables to the second-stage learning algorithm (a step that includes the results of such predictions right after the output-generation step). Furthermore, a second-stage algorithm is used in order to learn how to optimally combine credit model estimations, and form one final outcome or prediction. Besides, many experts in the fields of algorithm and machine learning including Sill *et al*. [18] used stacking and a combination of related techniques in order to efficiently increase the accuracy of their estimations, instead of using each model individually. The remaining stages in the establishment of data mining, using the (CRISP) data mining cycle specifically, is displayed on the next page in five detailed steps as Figure 1.

In the first step, data cleansing (purification) needs to be applied (which includes the removal or separation of illegible factors from the obtained data); additionally, what follows after involves the identification and replacement of missing data, as well as the selection of variables and measurement variables with the use of least discriminant analysis (LDA), in hopes to identify the most important variables, in addition to possibly provide more weight and value to the variables that are currently in use.

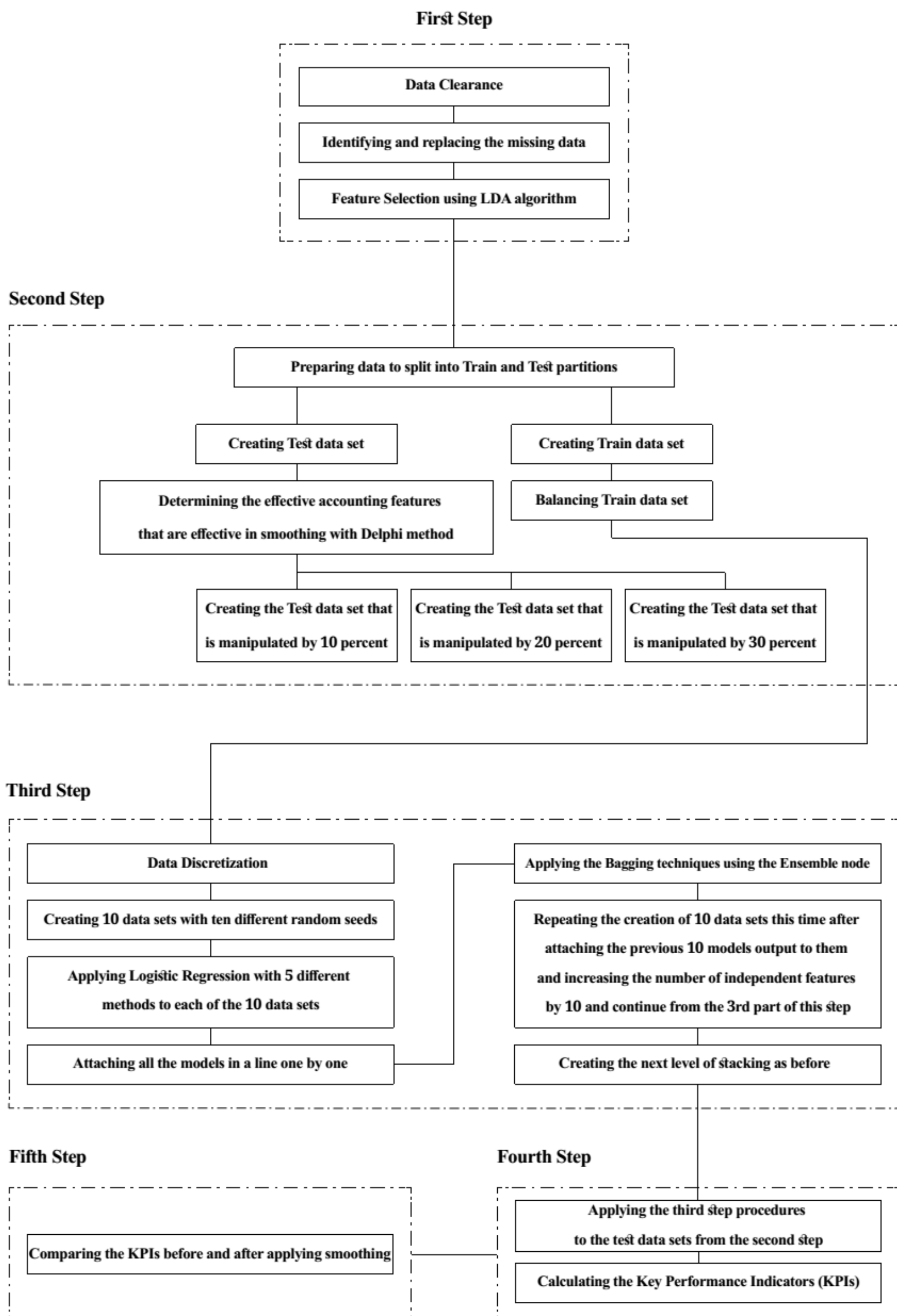

Figure 1. Research methodology process.

During the second step, what occurs is the formation of test and train database, in addition to the maintaining of balance amongst the creditworthy/unworthy applicants in the training database; besides, determining the accounting variables that can potentially affect the revenue smoothing through Delphi method, which is often used by accounting experts in order to identify and find out those variables that are frequently false or perhaps being manipulated in the organization's financial statements more often. Furthermore, the test database is established in order to assist in the revenue smoothing process procedure by smoothing these revenues from 10%, 20%, to upwards of 30% (the rise and fall found in the percentages may vary depending on the effectiveness of variables used).

In the third step, data discretization takes place, which, in other words, is the sort and organization of data in various classes and hereby creating a scorecard, in addition to the implementation of logistic regression algorithm with five different methods and quality control techniques, as well as using stacking techniques, due to their ability in providing the most accurate and capable credit scoring models, and finally, the measurement of the overall performance and quality of evaluation indicators.

In the next step, the obtained models are first applied to the test database, and then the performance evaluation indicators are measured. Additionally, in the fifth and final step, the indicators obtained in the previous part (step 4) are compared with each other using the hypothetical tests.

3.2. Hypotheses development

Hypotheses for this research work are organized with the objective of identifying the threshold of changes (conversions) of variables that are necessary in smoothing process; these changes may start from 10% and can possibly go up to a maximum of 30% increase or decrease in variable quantity. Fifteen hypotheses from (1) to (15) are organized in order to handle the issue of recognizing the threshold of changes in variables that finally affect the credit scoring model of the companies.

- **Main hypothesis**
  - The implementation of revenue smoothing does have a direct and significant effect on the credit scoring models of banks as well as financial institutions.

- **Hypothesis (1)**
  - A 10% adjustment in accounting variables that can potentially affect the revenue smoothing process has no significance or major effect on the credit scoring models.
  - A 10% adjustment in accounting variables that can potentially affect the revenue smoothing process has a major significance and effect on the credit scoring models.

- **Hypothesis (2)**
  - A 20% adjustment in accounting variable that can potentially affect the revenue smoothing process has no major or significant effect on the credit scoring models.
  - A 20% adjustment in accounting variable that can potentially affect the revenue smoothing process can possibly have a major impact on the credit scoring models.

- **Hypothesis (3)**
  - A 30% adjustment in accounting variable that can potentially affect the revenue smoothing process has no major or significant effect on the credit scoring models.
  - A 30% adjustment in accounting variable that can potentially affect the revenue smoothing process can possibly have major impact on the credit scoring models.
- **Sub-hypothesis**
  - The separation of the type of regression algorithm from the type of bagging method used in addition to stacking stages, as well as the conduct of 10, 20, and 30% adjustment/no-adjustments for every single available test database.

- **Hypothesis (4)**
  - The implementation of stacking stages to test databases with a 10% adjustment, practically has no major or significant effect.
  - The implementation of stacking stages to test databases with a 10% adjustment can very well have major or significant effects as a result.

- **Hypothesis (5)**
  - The implementation of stacking stages to test databases with a 20% adjustment, practically has no major or significant effects.
  - The implementation of stacking stages to test databases with a 20% adjustment can very well have major or significant effects.

- **Hypothesis (6)**
  - The implementation of stacking stages to test databases with a 20% adjustment, practically has no major or significant effects.

- The implementation of stacking stages to test databases with a 20% adjustment or change can very well have major or significant effects.

- **Hypothesis (7)**
  - The conduct of bagging technique on the test database, using voting method with a 10% adjustment, has practically no major effects whatsoever.
  - The conduct of bagging technique on the test database, using voting method with a 10% adjustment, can possibly have a significant impact in the aftermath.

- **Hypothesis (8)**
  - The conduct of bagging technique on the test database, using confidence method with a 10% adjustment, has practically no major effect at all as a result.
  - The conduct of bagging technique on the test database, using confidence method with a 10% adjustment, can possibly have significant effects.

- **Hypothesis (9)**
  - The conduct of bagging technique on the test database, using highest confidence method with a 10% adjustment, has practically no major effect at all.
  - The conduct of bagging technique on the test database, using highest confidence method with a 10% adjustment, can possibly have a significant impact.

- **Hypothesis (10)**
  - The conduct of bagging technique on the test database, using voting method with a 20% adjustment, has practically no major effect whatsoever.
  - The conduct of bagging technique on the test database, using voting method with a 20% adjustment, can possibly have a significant impact in the aftermath.

- **Hypothesis (11)**
  - The conduct of bagging technique on the test database, using confidence method with a 20% adjustment, has practically no major effect at all.
  - The conduct of bagging technique on the test database, using confidence method with a 20% adjustment, can possibly have significant effects as a result.

- **Hypothesis (12)**
  - The conduct of bagging technique on the test database, using highest confidence method with a 20% adjustment, has practically no major effect at all.
  - The conduct of bagging technique on the test database, using highest confidence method with a 20% adjustment, can possibly have a significant impact.

- **Hypothesis (13)**
  - The conduct of bagging technique on the test database, using voting method with a 30% adjustment, has no major effect whatsoever.
  - The conduct of bagging technique on the test database, using voting method with a 30% adjustment, can possibly have a significant impact as a result.
- **Hypothesis (14)**
  - The conduct of bagging technique on the test database, using confidence method with a 30% adjustment, has practically no major effect at all.
  - The conduct of bagging technique on the test database, using confidence method with a 30% adjustment, can possibly have significant effects.

- **Hypothesis (15)**
  - The conduct of bagging technique on the test database, using highest confidence method with a 30% adjustment, has practically no major effect at all.
  - The conduct of bagging technique on the test database, using highest confidence method with a 30% adjustment, can possibly have significant or major effects as a result.

4. Presentation of Research Results and Findings

It is important to know that the samples used in this particular research work have been obtained through the database of one of the official banks of the country, during the four previous years (1989-1993) that included the financial data of various companies that had previously received financial facilities from that specific bank (involving 1000 companies, in addition to 41 default independent predictor variables). Additionally, such variables are often obtained through either the balance sheet or the profit and loss statements of the organization. Besides, according to the data mining methods used in this research work, through computing techniques, we are hereby enabled to figure out the precise number of cases of companies with defaulted facilities which amount to 135 cases, which is (13.5%) of the total, in addition to the companies and organizations that have completely repaid their provided facilities that amounts to approximately 865 cases or in other words (86.5%) of all the provided facilities. Moreover, the SPSS modeler software is used in order to establish an appropriate credit scoring model, as well as taking care of the process procedure stages, in addition to the SPSS statistics software, which is also used to efficiently analyze the obtained data statistical tests. The following paragraphs explains the credit model implementation steps and stages in detail.

4.1 First step

At first, we sample the obtained training dataset ten times with various seed for cross-validation (sampling with different and random numbers), then as a result, ten different databases are obtained and formed. Secondly, each database with logistic regression algorithm, which is essentially an independent variable of refund/non-refund that must be applied in different ways; additionally, the obtained credit model outputs must be established and put together, as shown in Figure 2., In addition, take note that the regression nugget must be connected and attached to each other ten times repeatedly. Furthermore, ensembles are also used after the implementation of the last credit model in order for the bagging theory to take place. Moreover, various methods used for logistic regression algorithm including backward step-wise, backward, forward, stepwise, enter and bagging, which are conducted respectively (from left to right) in accordance with other methods used such as highest confidence wins voting as well as confidence-weighted voting.

As a result, each credit model and its related hexagon form and add a new column to our current database. Additionally, the number of the net variables used is consequently increased from 41 to 51 due to the addition of new predictor variables outcome obtained from the previous credit model results.

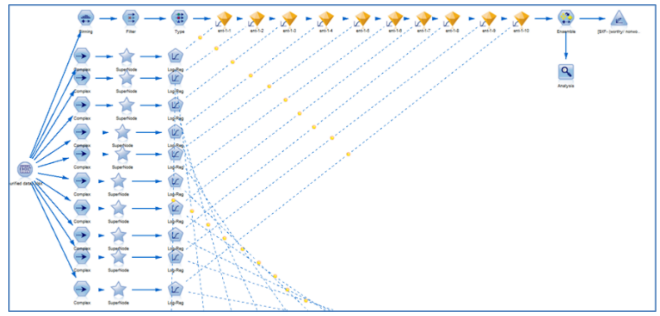

Figure 2. Addition of output of regression credit models to the original data as an independent variable (SPSS modeler scheme).

4.2 Second step

We continue the second stage by adding an ensemble to the other ten sample selections, and then apply the obtained credit model to each selected sample precisely, as we did in the previous bagging stage. In the next step, the credit models are implemented to the previous data that has been updated, and ten new variables have been added to them simultaneously. Furthermore, after the regression nuggets for the second stage of credit models are obtained; they can then be added as a set of series to the previous series of data. In conclusion, the obtained results from the variables sums around 61 cases. Moreover, this process is shown in great details on the next page as Figure 3.

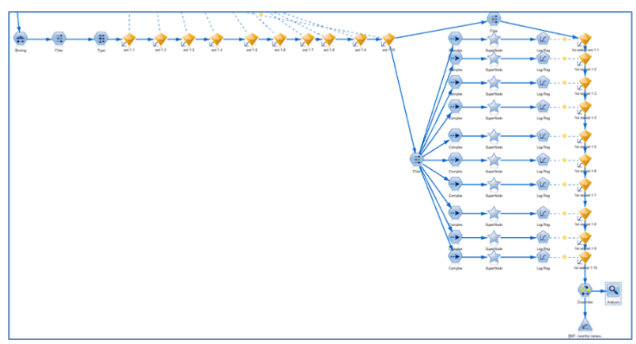

Figure 3. Addition of second stage data to the original series of data (SPSS modeler scheme).

4.3 Third step

At this stage, we conduct the exact same procedure as the previous second step, and convert 61 variables into 71. Additionally, in order to fully understand the differences in each stage, an ensemble is placed at the end of that particular stage, along with two analysis outputs as well as the area located below the curve. Therefore, with the use of such ensembles, we are enabled to obtain the

desired performance indicators, in addition to finding out the possible effects of each step of bagging and stacking.
The precise number of effective obtained outcome can be found in the following list. The obtained outcome includes five logistic regression algorithm methods, as well as three methods used for ensembles, as well as three operational stages without the conduct of stacking, in addition to first and second stages of stacking, that at the end form the establishment of ensemble, furthermore, five different data types are used, ($3 \times 3 \times 5 \times 5$) that results in (225) various outputs, which shall be described and displayed in the following. Additionally, these obtained outputs are displayed as 15 numbers, and each number has 15 separated outputs as a result.

4.4. Fourth step

In this particular step, the models are run, and key performance indicators of each test data bases are reported in Table 3 in the appendix 2carefully.

4.5. The Fifth Step

It is essential to use Wilcoxon statistical test of performance evaluation indicators in order to fully test the initial hypotheses of the research obtained from the table, in addition to conduct of a two-by-two comparison of each index in the test database without any changes against test databases that seem to have adjustments level of 10, 20 & 30%, the obtained results can be seen in Table 1 below:

Table 1. Various results on the Acceptance or Rejection of the hypotheses presented in this research.

Hypothesis No	Status	Hypothesis No	Status	Hypothesis No	Status
Hypothesis 1	H_0 is Accepted	Hypothesis 6	H_0 is Accepted	Hypothesis 11	H_0 is Accepted
Hypothesis 2	H_0 is Rejected	Hypothesis 7	H_0 is Accepted	Hypothesis 12	H_0 is Rejected
Hypothesis 3	H_0 is Rejected	Hypothesis 8	H_0 is Accepted	Hypothesis 13	H_0 is Accepted
Hypothesis 4	H_0 is Accepted	Hypothesis 9	H_0 is Rejected	Hypothesis 14	H_0 is Accepted
Hypothesis 5	H_0 is Accepted	Hypothesis 10	H_0 is Accepted	Hypothesis 15	H_0 is Rejected

5. Discussions & Conclusions

The main objective of this research, was to investigate the effects of smoothing in the overall performance of credit scoring models of banks and financial institutions. The research importance comes from the fact that smoothing is effectively being practiced in many companies nowadays, from which most of them apply for bank loans, and therefore it can potentially have major impact on the acceptance/rejection of the bank loan applicants and could exposed banks to higher levels type I and type II errors and of course higher credit risk. In order to investigate the issue, Logistic regression is used to build credit scorecards which are practically used by banks. 10, 20 & 30% adjustment was tested on the test data, revealing that from 20% adjustment and higher percentages the credit scoring models shows significant errors and expose the banks to higher and undesired level of credit risk. There could be different solution in order to handle the issue, one solution could be using a two staged credit scoring models. In this situation, the companies with more than 20% adjustments came from worthy to non-worthy as well as the ones came from non-worthy to worthy could be listed in a watch list and other features and factor that could not be adjusted in revenue smoothing could be used to finally evaluate the creditworthiness of them. Based on the practical aspect of the current research, logistic regression algorithm was used as the optimal scorecard building method for the creation of credit scoring models. Additionally, bagging & stacking techniques were also introduced and used in this particular research, in order to improve the overall quality of the credit models. However, it is also possible to use decision trees and mathematical programming in order to build appropriate scorecards. This research could be used for banks in which their active customer or at list special customer segments of them use high levels of smoothing in their financial statements.

References

[1] S. Watts, L. Ross and J. Zimmerman, “Towards a positive theory of the determination of accounting standards”, *Accounting review*, vol. 53, pp.34-112, Jan.1978.

[2] R. W. Holthausen, “Evidence on the effect of bond covenants and management compensation contracts on the choice of accounting techniques: The case of the depreciation switch-back”, *Journal ofAccounting and Economics*, vol. 3, pp. 73-109, June 2003.

[3] K. Kanagaretnam, G. J. Lobo, and R. Mathieu. “Managerial incentives for income smoothing through bank loan loss provisions”, *Review of Quantitative Finance and Accounting*, vol. 20, pp, 63-80., 2003.

[4] S. E. Michelson, J. Wagner, and C. W. Wootton. “The relationship between the smoothing of reported income and risk-adjusted returns”, *Journal of Economics and Finance*, vol. 24, pp 59-141, 2000.

[5] S. M. Sadatrasoul, M. Gholamian, M. Siami, and Z. Hajimohammadi, "Credit scoring in banks and financial institutions via data mining techniques: A literature review." *Journal of AI and Data Mining* 1, vol. 2 ,pp. 119-129, 2013.

[6] F. Louzada, A. Ara and B. F. Guilherme, “Classification methods applied to credit scoring: Systematic review and overall comparison”, *Surveys in Operations Research and Management Science*, vol. 21, pp34-117., 2016.

[7] S. M. Sadatrasoul, "Matrix Sequential Hybrid Credit Scorecard Based on Logistic Regression and Clustering" *Iranian Journal of Management Studies*, vol. 11 , pp: 91-111, 2018.

[8] G. Dong, L. Kin Keung and J. Yen, “Credit scorecard based on logistic regression with random coefficients”, *Procedia Computer Science*, vol. 1, pp.2463-2468, 2010.

[9] N. Siddiqi, “*Credit risk scorecards: developing and implementing intelligent credit scoring”,* John Wiley & Sons, ISBN: 978-1-119-20173-1, 2012.

[10] L. C.Thomas, “*Consumer Credit Models: Pricing , Profit and Portfolios: Pricing, Profit and Portfolios*”, OUP Oxford, 2009.

[11] A. Blöchlinger and L. Markus, “Economic benefit of powerful credit scoring”, *Journal of Banking & Finance*, vol. 30, pp. 873-851., 2006.

[12] H. Zhao, “A multi-objective genetic programming approach to developing Pareto optimal decision trees”, *Decision Support Systems*, vol. 43, pp. 809-826., 2007.

[13] H. Ince and A. Bora, “A comparison of data mining techniques for credit scoring in banking: A managerial perspective”, *Journal of Business Economics and Management*, vol. 10, pp. 40-233., 2009.

[14] C. Beidleman, "Income Smoothing: The Role of Management ,'"*The Accounting Review*, vol.48, no. 4, pp. 653-667., 1973.

[15] J. R. Graham, R. H. Campbell and S. Rajgopal, "The economic implications of corporate financial reporting", *Journal of accounting and economics*, vol. 40, pp 3-73., 2005.

[16] R. LaFond, M. H. Lang and H. A. Skaife, "Earnings smoothing, governance and liquidity: international evidence", 2007.

[17] M. Skurichina, R. P. W. Duin, "The Role of Combining Rules in Bagging and Boosting." *Advances in Pattern Recognition*, Vol. 1876. pp. 631-640 10, 2000.

[18] J. Sill, G. Takács, L. Mackey and D. Lin, "Feature-weighted linear stacking.", *arXiv preprint arXiv:0911.0460*, 2009.

[19] M.-W. Li, D.-Y. Xu, J. Geng and W.-C. Hong, "A ship motion forecasting approach based on empirical mode decomposition method hybrid deep learning network and quantum butterfly optimization algorithm" *Nonlinear Dynamics* vol. 107 Issue 3 , pp. 2447-2467, 2022.

Appendix 1

Variables included in credit dataset, and their types are sorted alphabetically and shown in Table 2.

Table 2. List of features in bank credit dataset.

Feature	Type
Accounts Receivable	Continuous
Accumulated Gains or Losses	Continuous
Active in Internal Market	Categorical
Audit Report	Categorical
Average Exports Over the Past Three Years	Continuous
Capital	Continuous
Company Background (Number of Years)	Continuous
Current Account Weighted Average	Continuous
Current Accounts Creditor Turn Over	Continuous
Current Assets	Continuous
Current Liabilities	Continuous
Current Period Assets	Continuous
Current Period Sales	Continuous
Current Period Shareholder Equity	Continuous
Experience With Bank(Number Of Years In 5 Categories)	Continuous
Export Price Index	Continuous
Financial Costs	Continuous
Gross Profit	Continuous
Inflation Rate	Continuous
Inventory Cash	Continuous
Last Three Years Average Imports	Continuous
Long-Term Financial Liabilities	Continuous
Mangers History	Continuous
Net Profit	Continuous
Non-Current Assets	Continuous
Non-Current Liabilities	Continuous
Number of Countries That The Company Export to	Continuous
Other Accounts Receivable	Continuous
Prior Period Assets	Continuous
Prior Period Sales	Continuous
Prior Period Shareholder Equity	Continuous
Sale	Continuous
Seasonal Factors	Categorical
Shareholder Equity	Continuous
Short-Term Financial Liabilities	Continuous
Stock	Continuous
Target Market Risk (From 1 To 5)	Continuous
Tehran Stock Exchange Index	Continuous
Three Prior Year Foreign Exchange Rate	Continuous
Top Mangers History	Categorical
Total Assets	Continuous
Total Liabilities	Continuous
Two-Prior Period Assets	Continuous
Two-Prior Period Sales	Continuous
Two-Prior Period Shareholder Equity	Continuous
Type of Book: Accredited Auditor (=1,Other=0)	Categorical
Type of Book: Audit Organization (=1,Other=0)	Categorical
Type of Book: Tax Declaration(=1,Other=0)	Categorical
Type of Company: Cooperative (=1, Other =0)	Categorical
Type of Company: Limited And Others (=1, Other =0)	Categorical
Type of Company: PJS (=1, Other =0)	Categorical
Type of Company: Stock Exchange (=1, Other =0)	Categorical
Type of Company: Stock Exchange(LLP) (=1, Other =0)	Categorical
Type of Industry: Agricultural (=1, Other =0)	Categorical
Type of Industry: Chemical (=1, Other =0)	Categorical
Type of Industry: Industry And Mine (=1, Other =0)	Categorical
Type of Industry: Infrastructure and Service (=1, Other =0)	Categorical
Type of Industry: Oil and Petrochemical (=1, Other =0)	Categorical
Year of Financial Ratio	Categorical
Basel: Creditworthy (=1, Other =0)	Categorical

Appendix 2

Results of experiments for different algorithms and adjustments (horizontal axis) versus different performance indicators of credit scoring model (Vertical axis).

Table 3. Various Categories of Performance Indicators in Different Conditions

Regression method	Bagging method	Stacking level	DB type	TP	FP	TN	FN	Accuracy	Miss classification Rate	TPR (Sensitivity)	FPR	Specificity	Precision	Prevalence	AUC	Gini
										Performance indicator						
Enter	Voting	Without stacking	Train	527	60	545	77	0.887	0.113	0.873	0.099	0.901	0.898	0.887	0.96	0.92
			Test	23	80	153	14	0.652	0.348	0.622	0.343	0.657	0.223	0.652	0.673	0.346
			Test 10%	24	80	153	13	0.656	0.344	0.649	0.343	0.657	0.231	0.656	0.676	0.353
			Test 20%	26	83	150	11	0.652	0.348	0.703	0.356	0.644	0.239	0.652	0.662	0.324
			Test 30%	26	82	151	11	0.656	0.344	0.703	0.352	0.648	0.241	0.656	0.661	0.321
		One level	Train	597	19	586	7	0.978	0.022	0.988	0.031	0.969	0.969	0.978	0.998	0.997
			Test	23	69	164	14	0.693	0.307	0.622	0.296	0.704	0.250	0.693	0.656	0.311
			Test 10%	23	74	159	14	0.674	0.326	0.622	0.318	0.682	0.237	0.674	0.661	0.323
			Test 20%	22	75	158	15	0.667	0.333	0.595	0.322	0.678	0.227	0.667	0.639	0.278
			Test 30%	22	76	157	15	0.663	0.337	0.595	0.326	0.674	0.224	0.663	0.639	0.279
		Two level stacking	Train	597	15	590	7	0.982	0.018	0.988	0.025	0.975	0.975	0.982	0.999	0.998
			Test	21	72	161	16	0.674	0.326	0.568	0.309	0.691	0.226	0.674	0.646	0.292
			Test 10%	22	76	157	15	0.663	0.337	0.595	0.326	0.674	0.224	0.663	0.65	0.299
			Test 20%	21	82	151	16	0.637	0.363	0.568	0.352	0.648	0.204	0.637	0.627	0.254
			Test 30%	22	79	154	15	0.652	0.348	0.595	0.339	0.661	0.218	0.652	0.628	0.256
	Weighed confidence voting	Without stacking	Train	534	48	557	70	0.902	0.098	0.884	0.079	0.921	0.918	0.902	0.964	0.928
			Test	23	67	166	14	0.700	0.300	0.622	0.288	0.712	0.256	0.700	0.67	0.34
			Test 10%	24	68	165	13	0.700	0.300	0.649	0.292	0.708	0.261	0.700	0.671	0.342
			Test 20%	22	71	162	15	0.681	0.319	0.595	0.305	0.695	0.237	0.681	0.653	0.307
			Test 30%	24	70	163	13	0.693	0.307	0.649	0.300	0.700	0.255	0.693	0.655	0.31
		One level stacking	Train	528	8	553	0	0.993	0.007	1.000	0.014	0.986	0.985	0.993	0.932	0.864
			Test	20	71	147	16	0.657	0.343	0.556	0.326	0.674	0.220	0.657	0.587	0.174
			Test 10%	20	76	142	16	0.638	0.362	0.556	0.349	0.651	0.208	0.638	0.597	0.194
			Test 20%	20	77	141	16	0.634	0.366	0.556	0.353	0.647	0.206	0.634	0.565	0.13
			Test 30%	20	75	143	16	0.642	0.358	0.556	0.344	0.656	0.211	0.642	0.568	0.136
		Two level stacking	Train	528	12	549	0	0.989	0.011	1.000	0.021	0.979	0.978	0.989	0.932	0.864
			Test	18	74	144	18	0.638	0.362	0.500	0.339	0.661	0.196	0.638	0.583	0.165
			Test 10%	19	75	143	17	0.638	0.362	0.528	0.344	0.656	0.202	0.638	0.589	0.179
			Test 20%	16	78	140	20	0.614	0.386	0.444	0.358	0.642	0.170	0.614	0.555	0.111
			Test 30%	16	76	142	20	0.622	0.378	0.444	0.349	0.651	0.174	0.622	0.559	0.117
		Without stacking	Train	458	94	511	146	0.801	0.199	0.758	0.155	0.845	0.830	0.801	0.806	0.611
			Test	23	80	153	14	0.652	0.348	0.622	0.343	0.657	0.223	0.652	0.62	0.241
			Test 10%	21	85	148	16	0.626	0.374	0.568	0.365	0.635	0.198	0.626	0.593	0.186
			Test 20%	21	86	147	16	0.622	0.378	0.568	0.369	0.631	0.196	0.622	0.578	0.157
			Test 30%	21	84	149	16	0.630	0.370	0.568	0.361	0.639	0.200	0.630	0.582	0.165
		One level stacking	Train	528	69	492	0	0.937	0.063	1.000	0.123	0.877	0.884	0.937	0.84	0.679
			Test	19	78	140	17	0.626	0.374	0.528	0.358	0.642	0.196	0.626	0.554	0.108
			Test 10%	18	88	130	18	0.583	0.417	0.500	0.404	0.596	0.170	0.583	0.532	0.063
			Test 20%	18	83	135	18	0.602	0.398	0.500	0.381	0.619	0.178	0.602	0.536	0.071

Regression method	Bagging method	Stacking level	DB type	Performance indicator												
				TP	**FP**	**TN**	**FN**	**Accuracy**	**Miss classification Rate**	**TPR (Sensitivity)**	**FPR**	**Specificity**	**Precision**	**Prevalence**	**AUC**	**Gini**
Two step Entrance			Test 30%	19	79	139	18	0.620	0.380	0.514	0.362	0.638	0.194	0.620	0.542	0.084
		Two level	Train	528	58	503	0	0.947	0.053	1.000	0.103	0.897	0.901	0.947	0.89	0.78
			Test	17	72	146	19	0.642	0.358	0.472	0.330	0.670	0.191	0.642	0.523	0.046
			Test 10%	13	73	145	23	0.622	0.378	0.361	0.335	0.665	0.151	0.622	0.492	-0.017
			Test 20%	12	77	141	24	0.602	0.398	0.333	0.353	0.647	0.135	0.602	0.426	-0.077
			Test 30%	12	77	141	24	0.602	0.398	0.333	0.353	0.647	0.135	0.602	0.426	-0.077
	Voting	**Without stacking**	Train	517	36	569	87	0.898	0.102	0.856	0.060	0.940	0.935	0.898	0.97	0.94
			Test	25	65	184	12	0.731	0.269	0.676	0.261	0.739	0.278	0.731	0.73	0.46
			Test 10%	24	63	185	13	0.733	0.267	0.649	0.254	0.746	0.276	0.733	0.719	0.439
			Test 20%	23	63	186	14	0.731	0.269	0.622	0.253	0.747	0.267	0.731	0.723	0.446
			Test 30%	23	61	188	14	0.738	0.262	0.622	0.245	0.755	0.274	0.738	0.723	0.447
		One level stacking	Train	591	14	591	13	0.978	0.022	0.978	0.023	0.977	0.977	0.978	0.992	0.985
			Test	22	60	189	15	0.738	0.262	0.595	0.241	0.759	0.268	0.738	0.725	0.45
			Test 10%	21	55	193	16	0.751	0.249	0.568	0.222	0.778	0.276	0.751	0.713	0.427
			Test 20%	21	59	190	16	0.738	0.262	0.568	0.237	0.763	0.263	0.738	0.724	0.448
			Test 30%	20	56	193	17	0.745	0.255	0.541	0.225	0.775	0.263	0.745	0.718	0.436
		Two level stacking	Train	598	14	591	6	0.983	0.017	0.990	0.023	0.977	0.977	0.983	0.999	0.998
			Test	21	61	188	16	0.731	0.269	0.568	0.245	0.755	0.256	0.731	0.719	0.438
			Test 10%	22	54	194	15	0.758	0.242	0.595	0.218	0.782	0.289	0.758	0.708	0.416
			Test 20%	20	57	192	17	0.741	0.259	0.541	0.229	0.771	0.260	0.741	0.72	0.44
			Test 30%	19	56	193	18	0.741	0.259	0.514	0.225	0.775	0.253	0.741	0.712	0.425
	Weighed confidence voting	**Without stacking**	Train	504	39	566	100	0.885	0.115	0.834	0.064	0.936	0.928	0.885	0.968	0.935
			Test	25	67	182	12	0.724	0.276	0.676	0.269	0.731	0.272	0.724	0.73	0.459
			Test 10%	22	64	184	15	0.723	0.277	0.595	0.258	0.742	0.256	0.723	0.717	0.435
			Test 20%	21	61	188	16	0.731	0.269	0.568	0.245	0.755	0.256	0.731	0.722	0.445
			Test 30%	23	61	188	14	0.738	0.262	0.622	0.245	0.755	0.274	0.738	0.726	0.452
		One level stacking	Train	591	13	592	13	0.978	0.022	0.978	0.021	0.979	0.978	0.978	0.991	0.982
			Test	20	53	165	16	0.728	0.272	0.556	0.243	0.757	0.274	0.728	0.634	0.268
			Test 10%	20	45	173	16	0.760	0.240	0.556	0.206	0.794	0.308	0.760	0.641	0.283
			Test 20%	21	48	170	15	0.752	0.248	0.583	0.220	0.780	0.304	0.752	0.634	0.267
			Test 30%	20	48	170	16	0.748	0.252	0.556	0.220	0.780	0.294	0.748	0.625	0.25
		Two level stacking	Train	591	13	592	13	0.978	0.022	0.978	0.021	0.979	0.978	0.978	0.991	0.982
			Test	21	55	163	15	0.724	0.276	0.583	0.252	0.748	0.276	0.724	0.635	0.27
			Test 10%	20	46	172	16	0.756	0.244	0.556	0.211	0.789	0.303	0.756	0.641	0.283
			Test 20%	21	50	168	15	0.744	0.256	0.583	0.229	0.771	0.296	0.744	0.633	0.267
			Test 30%	20	50	168	16	0.740	0.260	0.556	0.229	0.771	0.286	0.740	0.625	0.25
	Highest Confidence wins	**Without stacking**	Train	425	86	519	179	0.781	0.219	0.704	0.142	0.858	0.832	0.781	0.781	0.561
			Test	21	77	172	16	0.675	0.325	0.568	0.309	0.691	0.214	0.675	0.63	0.26
			Test 10%	21	79	169	16	0.667	0.333	0.568	0.319	0.681	0.210	0.667	0.618	0.236
			Test 20%	21	79	170	16	0.668	0.332	0.568	0.317	0.683	0.210	0.668	0.626	0.252
			Test 30%	21	78	171	16	0.671	0.329	0.568	0.313	0.687	0.212	0.671	0.628	0.256
		One level	Train	591	37	568	13	0.959	0.041	0.978	0.061	0.939	0.941	0.959	0.958	0.916
			Test	21	58	160	15	0.713	0.287	0.583	0.266	0.734	0.266	0.713	0.62	0.24
			Test 10%	19	61	157	17	0.693	0.307	0.528	0.280	0.720	0.238	0.693	0.596	0.192
			Test 20%	19	64	154	17	0.681	0.319	0.528	0.294	0.706	0.229	0.681	0.584	0.167
			Test 30%	19	62	156	17	0.689	0.311	0.528	0.284	0.716	0.235	0.689	0.587	0.175
		T	Train	591	37	568	13	0.959	0.041	0.978	0.061	0.939	0.941	0.959	0.958	0.916
			Test	21	58	160	15	0.713	0.287	0.583	0.266	0.734	0.266	0.713	0.62	0.4

Regression method	Bagging method	Stacking level	DB type	TP	FP	TN	FN	Accuracy	Miss classification Rate	TPR (Sensitivity)	FPR	Specificity	Precision	Prevalence	AUC	Gini
				Performance indicator												
Forward	Voting		Test 10%	19	61	157	17	0.693	0.307	0.528	0.280	0.720	0.238	0.693	0.596	0.192
			Test 20%	19	64	154	17	0.681	0.319	0.528	0.294	0.706	0.229	0.681	0.584	0.167
			Test 30%	19	62	156	17	0.689	0.311	0.528	0.284	0.716	0.235	0.689	0.587	0.175
		Without stacking	Train	522	36	569	82	0.902	0.098	0.864	0.060	0.940	0.935	0.902	0.971	0.943
			Test	24	60	189	13	0.745	0.255	0.649	0.241	0.759	0.286	0.745	0.726	0.452
			Test 10%	24	67	181	13	0.719	0.281	0.649	0.270	0.730	0.264	0.719	0.731	0.463
			Test 20%	24	62	187	13	0.738	0.262	0.649	0.249	0.751	0.279	0.738	0.734	0.468
			Test 30%	24	63	186	13	0.734	0.266	0.649	0.253	0.747	0.276	0.734	0.729	0.458
		One level stacking	Train	598	19	586	6	0.979	0.021	0.990	0.031	0.969	0.969	0.979	0.998	0.997
			Test	24	55	194	13	0.762	0.238	0.649	0.221	0.779	0.304	0.762	0.722	0.444
			Test 10%	23	52	196	14	0.768	0.232	0.622	0.210	0.790	0.307	0.768	0.719	0.437
			Test 20%	21	54	195	11	0.769	0.231	0.656	0.217	0.783	0.280	0.769	0.723	0.447
			Test 30%	21	54	195	16	0.755	0.245	0.568	0.217	0.783	0.280	0.755	0.712	0.425
		Two level stacking	Train	592	16	589	12	0.977	0.023	0.980	0.026	0.974	0.974	0.977	0.999	0.998
			Test	24	56	193	11	0.764	0.236	0.686	0.225	0.775	0.300	0.764	0.719	0.438
			Test 10%	23	50	198	11	0.784	0.216	0.676	0.202	0.798	0.315	0.784	0.713	0.426
			Test 20%	21	54	195	16	0.755	0.245	0.568	0.217	0.783	0.280	0.755	0.718	0.437
			Test 30%	21	53	196	16	0.759	0.241	0.568	0.213	0.787	0.284	0.759	0.707	0.414
	Weighed confidence voting	**Without stacking**	Train	521	39	566	83	0.899	0.101	0.863	0.064	0.936	0.930	0.899	0.97	0.94
			Test	25	64	185	12	0.734	0.266	0.676	0.257	0.743	0.281	0.734	0.726	0.453
			Test 10%	23	62	186	14	0.733	0.267	0.622	0.250	0.750	0.271	0.733	0.732	0.464
			Test 20%	22	64	185	15	0.724	0.276	0.595	0.257	0.743	0.256	0.724	0.732	0.465
			Test 30%	23	63	186	14	0.731	0.269	0.622	0.253	0.747	0.267	0.731	0.731	0.462
		One level stacking	Train	574	11	594	30	0.966	0.034	0.950	0.018	0.982	0.981	0.966	0.989	0.979
			Test	22	52	166	14	0.740	0.260	0.611	0.239	0.761	0.297	0.740	0.63	0.26
			Test 10%	20	44	174	16	0.764	0.236	0.556	0.202	0.798	0.313	0.764	0.622	0.244
			Test 20%	19	51	167	17	0.732	0.268	0.528	0.234	0.766	0.271	0.732	0.607	0.214
			Test 30%	20	52	166	16	0.732	0.268	0.556	0.239	0.761	0.278	0.732	0.597	0.193
		Two level stacking	Train	578	11	594	26	0.969	0.031	0.957	0.018	0.982	0.981	0.969	0.989	0.979
			Test	22	52	166	14	0.740	0.260	0.611	0.239	0.761	0.297	0.740	0.63	0.26
			Test 10%	20	44	174	16	0.764	0.236	0.556	0.202	0.798	0.313	0.764	0.622	0.244
			Test 20%	19	51	167	17	0.732	0.268	0.528	0.234	0.766	0.271	0.732	0.607	0.214
			Test 30%	20	52	166	16	0.732	0.268	0.556	0.239	0.761	0.278	0.732	0.597	0.193
	Highest Confidence	**Without stacking**	Train	425	86	519	179	0.781	0.219	0.704	0.142	0.858	0.832	0.781	0.781	0.561
			Test	21	78	171	16	0.671	0.329	0.568	0.313	0.687	0.212	0.671	0.63	0.261
			Test 10%	21	78	170	16	0.670	0.330	0.568	0.315	0.685	0.212	0.670	0.62	0.239
			Test 20%	21	78	171	16	0.671	0.329	0.568	0.313	0.687	0.212	0.671	0.629	0.258
			Test 30%	21	77	172	16	0.675	0.325	0.568	0.309	0.691	0.214	0.675	0.631	0.262
		O n	Train	579	42	563	25	0.945	0.055	0.959	0.069	0.931	0.932	0.945	0.944	0.888
			Test	22	51	167	14	0.744	0.256	0.611	0.234	0.766	0.301	0.744	0.646	0.292

1

Regression method	Bagging method	Stacking level	DB type	Performance indicator													
				TP	FP	TN	FN	Accuracy	Miss classification Rate	TPR (Sensitivity)	FPR	Specificity	Precision	Prevalence	AUC	Gini
Backward			Test 10%	19	50	168	17	0.736	0.264	0.528	0.229	0.771	0.275	0.736	0.617	0.234
			Test 20%	19	53	165	19	0.719	0.281	0.500	0.243	0.757	0.264	0.719	0.592	0.185
			Test 30%	16	51	167	20	0.720	0.280	0.444	0.234	0.766	0.239	0.720	0.571	0.142
		Two level stacking	Train	579	42	563	25	0.945	0.055	0.959	0.069	0.931	0.932	0.945	0.944	0.888
			Test	22	51	167	14	0.744	0.256	0.611	0.234	0.766	0.301	0.744	0.646	0.292
			Test 10%	19	50	168	17	0.736	0.264	0.528	0.229	0.771	0.275	0.736	0.617	0.234
			Test 20%	18	53	18	165	0.142	0.858	0.098	0.746	0.254	0.254	0.142	0.592	0.185
			Test 30%	16	51	167	20	0.720	0.280	0.444	0.234	0.766	0.239	0.720	0.571	0.142
	Voting	Without stacking	Train	527	35	570	77	0.907	0.093	0.873	0.058	0.942	0.938	0.907	0.972	0.943
			Test	24	62	187	13	0.738	0.262	0.649	0.249	0.751	0.279	0.738	0.726	0.452
			Test 10%	23	62	186	14	0.733	0.267	0.622	0.250	0.750	0.271	0.733	0.731	0.462
			Test 20%	24	62	187	13	0.738	0.262	0.649	0.249	0.751	0.279	0.738	0.734	0.468
			Test 30%	24	63	186	13	0.734	0.266	0.649	0.253	0.747	0.276	0.734	0.729	0.458
		One level stacking	Train	598	18	587	6	0.980	0.020	0.990	0.030	0.970	0.971	0.980	0.998	0.997
			Test	24	61	188	13	0.741	0.259	0.649	0.245	0.755	0.282	0.741	0.722	0.444
			Test 10%	23	54	194	14	0.761	0.239	0.622	0.218	0.782	0.299	0.761	0.719	0.437
			Test 20%	21	54	195	16	0.755	0.245	0.568	0.217	0.783	0.280	0.755	0.723	0.447
			Test 30%	21	54	195	16	0.755	0.245	0.568	0.217	0.783	0.280	0.755	0.712	0.425
		Two level stacking	Train	592	14	591	12	0.978	0.022	0.980	0.023	0.977	0.977	0.978	0.999	0.998
			Test	24	57	192	13	0.755	0.245	0.649	0.229	0.771	0.296	0.755	0.719	0.438
			Test 10%	23	52	196	14	0.768	0.232	0.622	0.210	0.790	0.307	0.768	0.713	0.426
			Test 20%	21	58	191	16	0.741	0.259	0.568	0.233	0.767	0.266	0.741	0.718	0.437
			Test 30%	21	55	194	16	0.752	0.248	0.568	0.221	0.779	0.276	0.752	0.707	0.414
	Weighed confidence voting	Without stacking	Train	521	40	565	83	0.898	0.102	0.863	0.066	0.934	0.929	0.898	0.97	0.94
			Test	25	62	187	12	0.741	0.259	0.676	0.249	0.751	0.287	0.741	0.726	0.453
			Test 10%	24	66	182	13	0.723	0.277	0.649	0.266	0.734	0.267	0.723	0.733	0.465
			Test 20%	23	62	187	14	0.734	0.266	0.622	0.249	0.751	0.271	0.734	0.734	0.468
			Test 30%	23	61	188	14	0.738	0.262	0.622	0.245	0.755	0.274	0.738	0.731	0.462
		One level stacking	Train	579	13	592	25	0.969	0.031	0.959	0.021	0.979	0.978	0.969	0.989	0.979
			Test	22	52	166	14	0.740	0.260	0.611	0.239	0.761	0.297	0.740	0.63	0.26
			Test 10%	20	44	174	16	0.764	0.236	0.556	0.202	0.798	0.313	0.764	0.622	0.244
			Test 20%	19	51	167	17	0.732	0.268	0.528	0.234	0.766	0.271	0.732	0.607	0.214
			Test 30%	20	52	166	16	0.732	0.268	0.556	0.239	0.761	0.278	0.732	0.597	0.193
		Two level stacking	Train	578	11	594	26	0.969	0.031	0.957	0.018	0.982	0.981	0.969	0.989	0.979
			Test	22	52	166	14	0.740	0.260	0.611	0.239	0.761	0.297	0.740	0.63	0.26
			Test 10%	20	44	174	16	0.764	0.236	0.556	0.202	0.798	0.313	0.764	0.622	0.244
			Test 20%	19	51	167	17	0.732	0.268	0.528	0.234	0.766	0.271	0.732	0.607	0.214
			Test 30%	20	52	166	16	0.732	0.268	0.556	0.239	0.761	0.278	0.732	0.597	0.193
	Highest Confidence wins	Without stacking	Train	519	47	558	85	0.891	0.109	0.859	0.078	0.922	0.917	0.891	0.961	0.921
			Test	21	78	171	16	0.671	0.329	0.568	0.313	0.687	0.212	0.671	0.63	0.261
			Test 10%	21	78	170	16	0.670	0.330	0.568	0.315	0.685	0.212	0.670	0.62	0.239
			Test 20%	21	78	171	16	0.671	0.329	0.568	0.313	0.687	0.212	0.671	0.629	0.258
			Test 30%	21	77	172	16	0.675	0.325	0.568	0.309	0.691	0.214	0.675	0.631	0.262
		One level stacking	Train	597	21	584	7	0.977	0.023	0.988	0.035	0.965	0.966	0.977	0.998	0.997
			Test	22	51	167	14	0.744	0.256	0.611	0.234	0.766	0.301	0.744	0.646	0.292
			Test 10%	19	50	168	17	0.736	0.264	0.528	0.229	0.771	0.275	0.736	0.617	0.234
			Test 20%	18	53	165	18	0.720	0.280	0.500	0.243	0.757	0.254	0.720	0.592	0.185
			Test 30%	16	51	167	20	0.720	0.280	0.444	0.234	0.766	0.239	0.720	0.571	0.142

Regression method	Bagging method	Stacking level	DB type	TP	FP	TN	FN	Accuracy	Miss classification Rate	TPR (Sensitivity)	FPR	Specificity	Precision	Prevalence	AUC	Gini
				Performance indicator												
Backward Stepwise		Two level	Train	579	42	563	25	0.945	0.055	0.959	0.069	0.931	0.932	0.945	0.944	0.888
			Test	22	51	167	14	0.744	0.256	0.611	0.234	0.766	0.301	0.744	0.646	0.292
			Test 10%	19	50	168	17	0.736	0.264	0.528	0.229	0.771	0.275	0.736	0.617	0.234
			Test 20%	18	53	165	18	0.720	0.280	0.500	0.243	0.757	0.254	0.720	0.592	0.185
			Test 30%	16	51	167	20	0.720	0.280	0.444	0.234	0.766	0.239	0.720	0.571	0.142
	Voting	Without stacking	Train	522	51	554	82	0.890	0.110	0.864	0.084	0.916	0.911	0.890	0.96	0.921
			Test	22	65	168	15	0.704	0.296	0.595	0.279	0.721	0.253	0.704	0.674	0.348
			Test 10%	22	70	163	15	0.685	0.315	0.595	0.300	0.700	0.239	0.685	0.676	0.352
			Test 20%	23	72	161	14	0.681	0.319	0.622	0.309	0.691	0.242	0.681	0.662	0.325
			Test 30%	22	74	159	15	0.670	0.330	0.595	0.318	0.682	0.229	0.670	0.659	0.318
		One level stacking	Train	597	20	585	7	0.978	0.022	0.988	0.033	0.967	0.968	0.978	0.998	0.997
			Test	23	70	163	14	0.689	0.311	0.622	0.300	0.700	0.247	0.689	0.655	0.311
			Test 10%	22	75	158	15	0.667	0.333	0.595	0.322	0.678	0.227	0.667	0.661	0.322
			Test 20%	22	76	157	15	0.663	0.337	0.595	0.326	0.674	0.224	0.663	0.639	0.277
			Test 30%	22	72	161	15	0.678	0.322	0.595	0.309	0.691	0.234	0.678	0.64	0.279
		Two level stacking	Train	597	16	590	7	0.981	0.019	0.988	0.026	0.974	0.974	0.981	0.999	0.998
			Test	22	72	161	15	0.678	0.322	0.595	0.309	0.691	0.234	0.678	0.646	0.293
			Test 10%	22	77	156	15	0.659	0.341	0.595	0.330	0.670	0.222	0.659	0.65	0.299
			Test 20%	21	81	152	16	0.641	0.359	0.568	0.348	0.652	0.206	0.641	0.627	0.254
			Test 30%	21	77	156	16	0.656	0.344	0.568	0.330	0.670	0.214	0.656	0.628	0.256
	Weighed confidence voting	Without stacking	Train	534	48	557	70	0.902	0.098	0.884	0.079	0.921	0.918	0.902	0.964	0.928
			Test	23	67	166	14	0.700	0.300	0.622	0.288	0.712	0.256	0.700	0.67	0.34
			Test 10%	24	68	165	13	0.700	0.300	0.649	0.292	0.708	0.261	0.700	0.671	0.342
			Test 20%	22	71	162	15	0.681	0.319	0.595	0.305	0.695	0.237	0.681	0.653	0.307
			Test 30%	24	70	163	13	0.693	0.307	0.649	0.300	0.700	0.255	0.693	0.655	0.31
		One level stacking	Train	528	8	553	0	0.993	0.007	1.000	0.014	0.986	0.985	0.993	0.932	0.864
			Test	20	71	147	16	0.657	0.343	0.556	0.326	0.674	0.220	0.657	0.587	0.174
			Test 10%	20	76	142	16	0.638	0.362	0.556	0.349	0.651	0.208	0.638	0.597	0.194
			Test 20%	20	77	141	16	0.634	0.366	0.556	0.353	0.647	0.206	0.634	0.565	0.13
			Test 30%	20	75	143	16	0.642	0.358	0.556	0.344	0.656	0.211	0.642	0.568	0.136
		Two level stacking	Train	528	8	553	0	0.993	0.007	1.000	0.014	0.986	0.985	0.993	0.932	0.864
			Test	20	71	147	16	0.657	0.343	0.556	0.326	0.674	0.220	0.657	0.587	0.174
			Test 10%	20	76	142	16	0.638	0.362	0.556	0.349	0.651	0.208	0.638	0.597	0.194
			Test 20%	20	77	141	16	0.634	0.366	0.556	0.353	0.647	0.206	0.634	0.565	0.13
			Test 30%	20	75	143	16	0.642	0.358	0.556	0.344	0.656	0.211	0.642	0.568	0.136
	Highest Confidence wins	Without stacking	Train	458	94	511	146	0.801	0.199	0.758	0.155	0.845	0.830	0.801	0.806	0.611
			Test	23	80	153	14	0.652	0.348	0.622	0.343	0.657	0.223	0.652	0.62	0.241
			Test 10%	21	85	148	16	0.626	0.374	0.568	0.365	0.635	0.198	0.626	0.593	0.186
			Test 20%	21	16	86	147	0.396	0.604	0.125	0.157	0.843	0.568	0.396	0.578	0.157
			Test 30%	21	84	149	16	0.630	0.370	0.568	0.361	0.639	0.200	0.630	0.582	0.165
		One level stacking	Train	528	36	492	0	0.966	0.034	1.000	0.068	0.932	0.936	0.966	0.84	0.679
			Test	19	78	140	17	0.626	0.374	0.528	0.358	0.642	0.196	0.626	0.554	0.108
			Test 10%	18	88	130	18	0.583	0.417	0.500	0.404	0.596	0.170	0.583	0.532	0.063
			Test 20%	18	83	135	18	0.602	0.398	0.500	0.381	0.619	0.178	0.602	0.536	0.071
			Test 30%	18	79	139	18	0.618	0.382	0.500	0.362	0.638	0.186	0.618	0.542	0.084
		Two level stack	Train	528	69	492	0	0.937	0.063	1.000	0.123	0.877	0.884	0.937	0.84	0.679
			Test	19	78	140	17	0.626	0.374	0.528	0.358	0.642	0.196	0.626	0.554	0.108
			Test 10%	18	88	130	18	0.583	0.417	0.500	0.404	0.596	0.170	0.583	0.532	0.063

Regression method	Bagging method	Stacking level	DB type	Performance indicator													
				TP	**FP**	**TN**	**FN**	**Accuracy**	**Miss classification Rate**	**TPR (Sensitivity)**	**FPR**	**Specificity**	**Precision**	**Prevalence**	**AUC**	**Gini**
			Test 20%	18	83	135	18	0.602	0.398	0.500	0.381	0.619	0.178	0.602	0.536	0.071
			Test 30%	18	79	139	18	0.618	0.382	0.500	0.362	0.638	0.186	0.618	0.542	0.084

بررسی میزان اثر هموارسازی درآمد بر مدل‌های اعتبار سنجی بانک‌ها و مؤسسات مالی با استفاده از داده‌کاوی: مورد یک بانک ایرانی

سید مهدی سادات رسول*، امید مهدی عبادتی و امیرامیرزاده ایرانی

گروه مدیریت فناوری اطلاعات و عملیات، دانشگاه خوارزمی، تهران، ایران.

ارسال ۲۰۲۲/۰۶/۰۵؛ بازنگری ۲۰۲۲/۱۲/۰۷؛ پذیرش ۲۰۲۳/۰۱/۱۱

چکیده:

شرکت‌ها با اهداف و انگیزه‌های گوناگونی در صورت‌های مالی خود تنظیماتی را انجام می‌دهند که در نهایت منجر به تغییر در مبلغ متغییرهای صورت مالی و ترازنامه خصوصا درآمد، هزینه و سود و شرکت می‌گردد، که اصطلاحا آن را هموارسازی درآمد می‌گویند. یکی ازعلل شایع این موضوع جلوگیری از نمایش نوسان عملکرد شرکت‌ها، بخصوص شرکت‌های سهامی عام که در بورس هستند، در نظر سهامداران می‌باشد؛ البته دلایل دیگری نیز برای این تنظیمات می‌تواند وجود داشته باشد. موسسات مالی و اعتباری عمدتا از صورت‌های مالی و ترازنامه شرکت‌ها به عنوان مبنایی برای اعتبارسنجی وتخصیص تسهیلات و ضمانت‌نامه به آنها استفاده می‌نمایند و این تنظیمات هموارسازی باعث تغییراتی در ورودی این مدل‌ها و درنتیجه اعتبار نتیجه مدل‌های ساخته شده می‌شود. این مقاله درصدد است تا اثر هموارسازی درآمدها را بر مدل‌های اعتبارسنجی بررسی نماید. داده‌های ۱۰۰۰ شرکت مختلف از یک بانک ایرانی جمع آوری شده و با انجام تغییرات مختلف روی این داده‌ها اثر هموارسازی درآمد در مدل‌های اعتبارسنجی از نوع کارت امتیاز اعتباری که توسط رگرسیون‌های لجستیک مختلف ساخته شده است بررسی شده است. بررسی فرضیه‌ها با استفاده از آزمون آماری ویلکاکسون نشان می‌دهد که تغییرات هموار سازی که بیش از ۲۰ درصد منجر به تغییر داده‌ها شوند می‌توانند تاثیر قابل توجهی بر مدل های هموارسازی داشته باشند. تغییر در میزان برخی از مبالغ متغییرهای صورت های مالی و ترازنامه می تواند منجر به ایجاد خطا در مدل های اعتبارسنجی گردد.

کلمات کلیدی: داده کاوی، رتبه بندی اعتباری، مدل های اعتبارسنجی، هموار سازی درآمد.