6-1-2022

Implementing artificial intelligence in forecasting the risk of personal bankruptcies in Poland and Taiwan

Tomasz Korol
Gdańsk University of Technology

Anestis K. Fotiadis
Zayed University

OECONOMIA COPERNICANA

VOLUME 13 ISSUE 2 JUNE 2022

p-ISSN 2083-1277, e-ISSN 2353-1827

www.oeconomia.pl

ORIGINAL ARTICLE

Citation: Korol, T., & Fotiadis, A. K. (2022). Implementing artificial intelligence in forecasting the risk of personal bankruptcies in Poland and Taiwan. *Oeconomia Copernicana, 13*(2), 407–438. doi: 10.24136/oc.2022.013

Contact to corresponding author: Tomasz Korol, tomasz.korol@zie.pg.gda.pl

Article history: Received: 13.09.2020; Accepted: 7.03.2021; Published online: 26.03.2022

Tomasz Korol
Gdańsk University of Technology, Poland
orcid.org/0000-0002-7623-3404

Anestis K. Fotiadis
Zayed University, United Arab Emirates
orcid.org/0000-0001-8806-7987

Implementing artificial intelligence in forecasting the risk of personal bankruptcies in Poland and Taiwan

JEL Classification: *G17; G51*

Keywords: *fuzzy logic; genetic algorithms; artificial neural networks; consumer bankruptcy; the financial crisis of households*

Abstract

Research background: The global financial crisis from 2007 to 2012, the COVID-19 pandemic, and the current war in Ukraine have dramatically increased the risk of consumer bankruptcies worldwide. All three crises negatively impact the financial situation of households due to increased interest rates, inflation rates, volatile exchange rates, and other significant macroeconomic factors. Financial difficulties may arise when the private person is unable to maintain a habitual standard of living. This means that anyone can become financially vulnerable regardless of wealth or education level. Therefore, forecasting consumer bankruptcy risk has received increasing scientific and public attention.
Purpose of the article: This study proposes artificial intelligence solutions to address the increased importance of the personal bankruptcy phenomenon and the growing need for reliable forecasting models. The objective of this paper is to develop six models for forecasting personal bankruptcies in Poland and Taiwan with the use of three soft-computing techniques.

Methods: Six models were developed to forecast the risk of insolvency: three for Polish households and three for Taiwanese consumers, using fuzzy sets, genetic algorithms, and artificial neural networks. This research relied on four samples. Two were learning samples (one for each country), and two were testing samples, also one for each country separately. Both testing samples contain 500 bankrupt and 500 nonbankrupt households, while each learning sample consists of 100 insolvent and 100 solvent natural persons.
Findings & value added: This study presents a solution for effective bankruptcy risk forecasting by implementing both highly effective and usable methods and proposes a new type of ratios that combine the evaluated consumers' financial and demographic characteristics. The usage of such ratios also improves the versatility of the presented models, as they are not denominated in monetary value or strictly in demographic units. This would be limited to use in only one country but can be widely used in other regions of the world.

Introduction

Household finance is the science devoted mainly to wealth, consumption level and patterns, financial literacy, portfolio allocation, and debt decisions (Gomes *et al*., 2021, pp. 919–1000; Guiso & Sodini, 2013, pp. 1397–1532). Hence, most studies are devoted to forecasting financial distress concerning enterprises, not households. Moreover, this sparse literature devoted to the forecasting of consumer bankruptcy in most cases concerns statistical analyzes concerning the identification of factors influencing the risk of household insolvency or the impact of macroeconomic events on the scale of bankruptcies in a country. Compared to studies devoted to forecasting the bankruptcy of enterprises, only a few of them are devoted to the development of a forecasting model. In contrast, the latest studies prove an increased risk of consumer insolvencies in the economies of countries worldwide (e.g., Ari *et al*., 2021; Li *et al*., 2020, pp. 472–500). The coronavirus pandemic's global character and the current Ukraine war affect most macroeconomic measures such as inflation, GDP growth, interest, and unemployment rates worldwide. The increasing inflation and interest rates directly influence households' financial standing. Moreover, it impacts the GDP growth rate by affecting consumers' consumption decisions. It can worsen the labor market's stability and reduce wealth, causing an even higher risk of bankruptcies among natural persons. The households use two basic types of resources to function: material assets, the general wealth accumulated from savings, and human capital, meaning work capabilities.

To fill this gap in the literature, the objective of this paper is to develop six models for forecasting personal bankruptcies in Poland and Taiwan with the use of three soft-computing techniques. The additional distinguishing feature of the paper is the implementation of the new type of ratios in the proposed models. The available models in the literature are based on a set of single demographic or financial information about households.

They lack any constructed ratios unlike in forecasting the bankruptcy risk of enterprises where financial ratios (e.g. liquidity, profitability, etc.) are used to evaluate a company's economic risk.

Bankruptcy forecasting models are particularly interesting for financial institutions due to the necessity to verify the economic risk of an entity requesting credit and monitor the loan repayment. However, household indebtedness is also receiving growing public attention due to the COVID-19 crisis and its consequences (i.e., increasing inflation and interest rates). Nowadays, the increasing number of private persons that have difficulties paying back their bank credits aroused public concern. Consumers are becoming more aware of problems with loan repayments and the risk of insolvency. It is worth adding that households can also use such models to evaluate their financial standing and verify how specific factors (financial and demographic) can affect their default risk.

This study proposes artificial intelligence solutions for two economic regions: Central Europe (on the example of Poland) and Far-East Asia (on the example of Taiwan), to address the increased importance of personal bankruptcies and the growing need for reliable forecasting models. The research contributes to the literature on forecasting the risk of personal bankruptcies in four folds: first, it evaluates the effectiveness of three innovative forecasting techniques–genetic algorithms, fuzzy logic, and artificial neural networks–in predicting this type of risk using the example of Polish and Taiwanese households separately. Second, it verifies the predictive properties of the new type of ratios that we constructed with a combination of demographic and financial data of households. Third, it discusses the advantages and disadvantages of the three techniques from the perspective of the economic region (Poland versus Taiwan) and the two types of errors generated. Fourth, it compares the artificial intelligence models' to the results obtained by the literature's most common statistical scoring models. While implementing the research objectives, the authors answered the following research questions:

– What types of ratios characterize the highest predictive properties? All households were described using six financial and five demographic variables. Based on those eleven variables, the authors of this research have created 12 ratios that simultaneously combine the financial and demographic information. It is worth noting that this is one of the first attempts in the literature worldwide to construct such unique entry variables that include different types of information to forecast the bankruptcy risk of natural persons.
– Are there any significant differences in forecasting the risk of household bankruptcy between Poland and Taiwan? On further investigation, we

found two aspects: first, the best forecasting method for each country; and second, the significant differences between these countries in terms of quality of forecast, error types, and the implementation of ratios used in the created models.
– Which model can obtain the lowest type-I errors? It is worth investigating because type I errors are considered costlier and more dangerous from the insolvency viewpoint of the bank than type II errors.

This paper is divided into five sections. In the first section, the authors justify the topic, research objectives, and innovations and contributions to the literature. Section 2 is devoted to a literature review covering the topic of credit scoring and forecasting bankruptcy risk. Section 3 introduces the assumptions of this study regarding the collected data, developed ratios, testing and learning samples, and prediction techniques implemented. Section 4 describes three prediction models for Polish households and three for Taiwanese households. The quality of the forecast based on the results obtained from the testing sample is also discussed. The last section concludes the paper.

Literature review

Credit scoring models are used to assess consumers' and enterprises' credit risk and estimate the probability of nonperforming loans. Forecasting the probability of credit delinquency has been the objective of credit scoring models for 50 years. However, forecasting the risk of bankruptcy was not included in the analysis of consumer loans (Zhang & Thomas, 2012, pp. 204–215). Tufano (2009, pp. 227–247) identifies several reasons why household finance and consumer bankruptcies predictions traditionally received limited attention from mainstream financial economists.

The objective of bankruptcy prediction is to estimate the risk of insolvency and declaration of bankruptcy due to the inability to repay credits, such as when the value of total credits exceeds the value of total assets. The main challenge in forecasting this is selecting the prediction technique and the set of variables with high predictive properties. Over the past five decades, many methods have been proposed for the bankruptcy prediction of enterprises. Starting from the development of the first model of multivariate discriminant analysis by Altman (1968, pp. 589–609), much research was carried out using a wide variety of statistical methods. A review of literature by Alaka *et al*. (2018, pp. 164–184) indicates that there are two popular and promising statistical models for corporate bankruptcy prediction — the multivariate discriminant analysis and logistic regression. The authors reviewed the articles published between 2010 and 2015. Similar conclusions were made by Barboza *et al*. (2017, pp. 405–417). The authors based on data from 1985 to 2013 on North American firms, compared the performance of estimated discriminant analysis, logistic regression, and neural networks models. In turn, Giannopoulos and Sigbjornsen (2019, pp. 1114–1128) investigated the practical application of common statistical models (e.g. z-score model of Altman) by re-estimating their coefficients using a recent sample period. Their study proved also that financial ratios are an efficient tool for forecasting the bankruptcy of companies in Greece. Another interesting recent study devoted to statistical models is research conducted by Mihalovic (2016, pp. 101–118), who estimated discriminant analysis and logistic regression models for enterprises operating in Slovakia. The results of the study suggest that the model based on a logit function outperforms the classification accuracy of the discriminant model (Mihalovic, 2016, pp. 101–118). The author identified also the most important financial ratios (return on assets and current liquidity) in forecasting the bankruptcy risk. Similar research was conducted by Delen *et al*. (2013, pp. 3970–3983) and Kieschnick *et al.* (2013, pp. 1827–1855). Delen *et al.* (2013) first identified the ratios with the highest forecasting properties among 72 different ratios and then estimated the decision tree model for Turkish enterprises. This study also proved that the methodology of C&RT is usable in predicting this type of risk. In the case of the study conducted by Kieschnick (2013), the relationship between working capital ratios, the wealth of the company, and the risk of bankruptcy was analyzed. The paper by Lukason and Hoffman (2014, pp. 80–91) examined also the relationship between the financial wealth of a company measured by bankruptcy models and the number and types of causes of firm failure. The authors proved that there is a relation between the results of logit models estimated for Estonian enterprises and the process of going bankrupt.

In the 2000s, the popularity of enterprise bankruptcy forecasting techniques has changed from statistical to soft-computing methods such as fuzzy sets, neural networks, vector support machines and genetic algorithms. In the last two decades, neural networks became a popular method for developing corporate bankruptcy models. Callejon *et al*. (2013, pp. 29–37) created such a model based on 500 European industry enterprises. Their model is characterized by the effectiveness of around 92%. Another interesting example of the usage of neural networks is the study of Hosaka (2019, pp. 287–299), who used the financial ratios as image inputs into convolution networks. In turn, Jardin (2018, pp. 64–77) used Kohonen neural networks to build the map of patterns of going bankrupt. This type of neural networks also proved to be an effective tool in evaluating the risk of bankruptcy for enterprises.

Genetic algorithms are the second most popular method among soft-computing techniques. We can point out three popular studies (Acosta-Gonzales & Fernandez-Rodriguez, 2014, pp. 133–157; Dong *et al*., 2018, pp. 204–220; Lin *et al*., 2014, pp. 2472–2483) as the examples of using genetic algorithms in the process of selection of the best variables and in estimating the forecasting models. All the models in those studies are characterized by the effectiveness above 80%.

The third popular technique is the vector support machines method. The latest example of such a model is the model created for Spanish enterprises by Garcia *et al*. (2019, pp. 1019–1031). In the study, it was proved that the support vector machine model is superior to discriminant analysis and logistic regression models. The second, latest example of using a support vector machines technique is a model created by Ptak-Chmielewska (2019, pp. 1–17) that was created based on 806 small enterprises in Poland. Ptak-Chmielewska proved that this technique performs as well as multilayer neural networks and decision trees.

In forecasting the bankruptcy risk of enterprises, the models use microeconomic information in the form of financial indicators (e.g., activity, profitability, liquidity, debt indicators). These ratios can be static and dynamic values (such as the dynamics of the equity return and the growth rate of current liquidity).

In the case of consumer credit scoring, the models use details about obligors that are static. Such models are used to determine whether an applicant should be granted credit based on data collected at the time of application and then remain fixed (Bellotti & Crook, 2013, pp. 563–574).

As mentioned before, most studies devoted to consumer credit scoring models focus on predicting the risk of the occurrence of nonperforming loans (NPL). This falls into two categories: macro and microeconomic research approaches.

The first kind of study (macroeconomic approach) quantifies the role of adverse, unforeseen shocks that may lead to consumer bankruptcies (Gross & Notiwigdo, 2011, pp. 767–778). An example of such a shock can be the likelihood of a recession, which may lead to an increase in interest rates and credit spreads, negatively affecting the financial situation of consumers (Luzzetti & Neumuller, 2016, pp. 22–39). Another example may be the situation of an increase in unemployment in the labor market, preventing borrowers from paying off their loans (e.g., Anastasiou *et al.*, 2016, pp. 116–119; Paskevicius & Jurgaityte, 2015, pp. 521–526; Barba & Pivetti, 2009, pp. 113–137). The third most frequently used macroeconomic factor is the inflation rate affecting the discussed interest rates. It influences the cost of living of natural persons. This results in worsening the financial strength of households. The inflation rate can also affect the unemployment rate, causing an even higher risk of insolvency among consumers (e.g., French & Vigne, 2019, pp. 150–156; Gross & Poblacion, 2017, pp. 510–528). Another important factor is the stability of income. Aristei and Gallo (2016, pp. 453–465) showed that growing unemployment, and thus the lack of stable income, is also a significant risk factor for consumers in Italy. Aller and Grant (2018, pp. 39–52) measured also the effect of the financial crisis on the default risk of the consumers in Spain. Similar research was conducted in Argentina. Zurawicki and Braidot (2005, pp. 1100–1109) verified the influence of the economic crisis on income, consumption, and insolvency risk for the middle class in Argentina. Diaz-Serrano (2005, pp. 153–177) identified the correlation between the monetary policy of banks and the personal bankruptcy rate in EU countries.

The second type of research (microeconomic approach) links consumer indebtedness and the economy, which is consumption (Kukk, 2016, pp. 764–785). In this approach, scholars have explored the microeconomic determinants of consumer bankruptcy such as income, credit card spending, mortgage expenditures, marital status, number of children, employment status, number of credit cards, the value of assets, substitution effect meaning buying cheaper brands, purchasing smaller packages, the increasing discount and neighborhood stores' popularity (e.g., Ghent & Kudlyak, 2011, pp. 3139–3186; Guiso *et at*., 2013, pp. 1473–1515; I-Cheng & Che-Hui, 2009, pp. 2473–2480; Patel *et al*., 2012, pp. 556–565; Worthington, 2006, pp. 2–15). This research approach also uses age, gender, and education as factors that may influence the financial vulnerability of households. Education positively influences the knowledge of finances. A proper understanding of credit terms can reduce consumers' risk of financial mistakes (Hira, 2021, pp. 502–507). Gender differences in financial risk behavior are posited by a few studies that show that women are more risk-averse when making financial decisions (e.g., Croson & Gneezy, 2009, pp. 448–474; Jianakoplos & Bernasek, 1998, pp. 620–630; Thorne, 2010, pp. 185–197).

Examples of the latest studies that implement the above-mentioned factors are the papers of Brygała (2022, pp. 1–13) and Nor *et al.* (2019, pp. 157–170).

Brygała examined the usefulness of logit regression in forecasting the consumer bankruptcy of households in the USA. The study proposed several forms of logit models depending on such factors as age, income level, number of children, marital status, and level of debts. The second example concerns forecasting the risk of bankruptcy for households in Malaysia.

The authors estimated the decision tree model with the use of such microeconomic factors as the level of loans, the number of dependants, and monthly income.

Research method

Statistical and software computing techniques are two broad categories of models that can predict bankruptcy in enterprises and households. In a previous study, the author estimated the prediction models for consumers using three statistical methods (Korol, 2021, pp. 1–14) — logistic regression models (LOG), multivariate discriminant analysis models (MDA), and classification and regression trees models (C&RT). The objective of this research is to develop six models to forecast the risk of insolvency: three for Polish households and three for Taiwanese consumers, using fuzzy sets (FL), artificial neural networks (ANN), and genetic algorithms used in the learning process of artificial neural networks (GA ANN). To develop the models, the authors used MatLab software.

The authors chose such three methods of artificial intelligence as they effectively deal with imprecisely defined problems, incomplete data, imprecision, and uncertainty. The phenomenon of forecasting the risk of personal bankruptcy has all of the above characteristics. In addition, these methods are suitable for use in systems that are designed to fit certain internal parameters to changing environmental conditions in a dynamic way.

Fuzzy logic was proposed by Zadeh (1965, pp. 338–353). It deals with inaccurate or incomplete knowledge. In the form of a mathematical system is modeling the imprecise information using linguistic terms. Opposite to binary logic, fuzzy sets use membership functions to deal with imprecise knowledge (Louzada *et.al*, 2016, pp. 117–134). The classical set theory uses the crisp boundary: an entity belongs to the set or is not a member of this set (true/false: 0,1). However, the human reasoning process works differently and less dichotomously. In fuzzy set theory, an entity can belong to more than one set to a certain degree. This degree is defined through membership functions (Akkoc, 2012, pp. 168–178). Membership functions are presented in any form and freely determined by the programmer. In the literature, the most common membership functions take one of three forms: triangular, trapezoidal, or Gaussian. The fuzzy set A in a certain nonempty set X ($A \subseteq X$) is specified as (Wu *et al*, 2010, pp. 774–787):

$$A = \{(x, \mu_A(x)) | \, x \in X\} \tag{1}$$

where μ_A: X → [0,1] is a membership function specifying the degree to which each element from X is assigned to fuzzy set A. The membership function $\mu_A(x)$: U⇒ [0,1] is defined as follows (Korol & Fotiadis, 2016, pp. 1451–1468):

$$\forall_{x \in U}\ \mu_A(x) = \begin{cases} f(x), x \in X \\ 0, x \notin X \end{cases} \tag{2}$$

where $\mu_A(x)$ is a function specifying the membership of x in set A, which is a subset of U, and f (x) is a function of the values in the range [0,1]. The values of this function are called degrees of membership. The membership function of each element x ∈ X assigns a degree of membership to fuzzy set A, in which we can distinguish three situations (Korol, 2018, pp. 165–188):

- $\mu_A(x) = 1$ indicates the entire membership of element x in fuzzy set A,
- $\mu_A(x) = 0$ indicates no membership of element x in fuzzy set A and
- $0 < \mu_A(x) < 1$ indicates fragmentary membership of element x in fuzzy set A.

The second forecasting technique used is the ANN. The principle of operation of a neural network is to simulate the working rules of the human brain (Xiao *et al*., 2012, pp. 196–206). ANN models learn from exemplary cases. An ANN consists of several processing elements that come together within the frame of particular rules, called neurons (Akkoc, 2012, pp. 168–178). The most common neural network used for predicting different financial phenomena is the feedforward multilayer neural network consisting of three layers of interconnected neurons (Sun *et al*., 2014, pp. 41–56). The first layer is called the input layer, in which the network uses entry data (e.g., financial ratios). All neurons in the entry layer transmit signals to the hidden layer, where the main processing and computation of the neural signals occurs. The exit layer generates the result of calculations outside of the network. In this study, the output neuron represents the forecast of the bankruptcy risk of consumers with neuron values from 0 to 1.

The last forecasting method employed in this study is the genetic algorithm (GA). The concept of GA is derived from the field of genetics. They are modelled on the theory of evolution, according to which the strongest and fittest individuals have the greatest chance of survival and subsequent reproduction (Tsai, 2014, pp. 46–58). This is related to the concept of the fitness function. These algorithms implement the following principles derived from biological laws (Mitchell, 1999, pp. 2–24):

- Search for a solution through the evolution processes of the population of solutions,
- inheritance of information through a single solution in successive generations of the solution population,
- changing the information in a single solution by crossing solutions with other solutions or mutation of the solution
- Select individual solutions based on matching a solution to the problem.

The basic operators of genetic algorithms are (Mitchell, 1999, pp. 2–24):

- selection and reproduction of better individuals,
- genetic crossover,
- random mutation of characteristics of entities.

The GA model takes the form (Louzada *et. al*, 2016, pp. 117–134):

$$AG = (N_{pop}, N_{gen}, N_{chr}, \Omega, f_{eval}, f_{sel}) \tag{3}$$

where N_{pop} is the number of population elements, N_{gen} is the number of generations, N_{chr} is the population size, f_{eval} is the adaptation function, f_{sel} is the reproduction selection rule. It is Ω□ the set of operators and their corresponding probabilities.

The distinguishing assumption of this study is the application of a new type of ratio that combines financial and demographic information. The standard economic predictors used in consumer credit scoring typically include education, occupation, marital status, number of children, age, income, and level of debt. Our study is one of the first to test the relationship between personality traits and behavioral inputs to financial variables in the form of ratios. Based on 11 variables, that is, annual income, monthly income, the total value of assets held, the value of all loans taken, credit cards debt, the value of monthly interest rates paid, age, level of education, number of children, marital status, and length of employment, we created 12 ratios that were used to create the forecasting models. Table 1 presents the ratios created with the given formulas and interpretations.

This study relied on four samples. Two were learning samples (one for each country), and the other two were testing samples (for each region). Both testing samples consist of 500 bankrupt and 500 nonbankrupt households, while each learning sample consists of 100 insolvent and 100 solvent natural persons. The authors calculated the values of the 12 proposed ratios for all 2400 consumers. It is worth noting that for each consumer the data was gathered and calculated individually as it was impossible to find a ready database of bankrupt and nonbankrupt households. The information was taken from the year 2015 to 2019 for both countries. Such a period helps to avoid the impact of a sudden, unforeseeable event such as the COVID-19 pandemic. Table 2 presents the example of the demographic characteristic for both testing samples. In both samples, males were the main group among bankrupt consumers, while for nonbankrupt consumers the distribution was close to 50/50 between females and males. Concerning the age of consumers, 71.7% of bankrupt women in Taiwan (99 cases) and 67.5% in Poland (108 cases) were from the group age of 27–50. In the case of males, the share of bankrupt consumers from the group age of 27–50 was lower — 59.4% in Poland (202 cases) and 58.3% in Taiwan (211 cases). A detailed demographic description of testing samples can be found in Korol (2021).

The following formulas were implemented to evaluate the quality of the programmed models (Korol, 2020, pp. 783–804):

- $S = \{1 - [(D1 + D2) / (BR + NBR)]\} \times 100\%$ (4)

- $E1 = D1 / BR \bullet 100\%$, (5)

- $E2 = D2 / NBR \bullet 100\%$, (6)

where: S is overall effectiveness, E1 is a type I error, E2 is a type II error, D1 is the number of bankrupted households classified by the model as nonbankrupt consumers, D2 is the number of nonbankrupt households classified as consumers at risk of bankruptcy, BR is the number of bankrupt households in the sample, and NBR is the number of nonbankrupt consumers in the sample.

A type I error shows the classification of a future insolvent natural person as an entity with good financial standing. Giving credit to a future bankrupt consumer will generate a loss to the bank due to the problem of recollecting the previously given loan. Type II error, on the other hand, presents the loss of potential income by the bank by rejecting the consumer loan, forecasting that this is a future insolvent household. Therefore, it is considered that type-I errors are more important to pay attention to than type II errors.

Results and discussion

The first two developed models are the FL models. The input ratios to these models were chosen with the use of the correlation matrix, selecting only those weakly correlated features and strongly correlated with the grouping variable, representing the status of the risk of bankruptcy or lack thereof for the given household. Such a research approach ensured the selection of variables that did not duplicate information provided by other developed ratios. Based on this, the following indicators were selected:

– for Polish consumers: X_2, X_3, X_5, X_{11},
– for Taiwanese entities: X_2, X_3, X_5, X_8, X_{11}.

The output of the model is a measure representing a prediction of bankruptcy risk for the analyzed consumer. Both FL models are based on the authors' rules in the state of "IF — THEN," in which the analysts' knowledge is defined. In the case of the FL model for Taiwanese consumers, there are five entry variables with three possible membership functions ("LOW," "AVG," and "HIGH"). The types and values of the membership functions are listed in Table 2. This resulted in a collection of 243 decision rules. In the case of the FL model for Polish households, there are four variables with the same three defined states of membership functions, resulting in a collection of 81 decision rules. Due to manuscript size constraints, only the top 20 decision rules are presented for each model.

For Polish consumers:

1. If X_2 is HIGH and X_3 is LOW and X_5 is HIGH and X_{11} is HIGH then output is LOW
2. If X_2 is HIGH and X_3 is LOW and X_5 is AVG and X_{11} is HIGH then output is LOW
3. If X_2 is HIGH and X_3 is LOW and X_5 is HIGH and X_{11} is AVG then output is LOW
4. If X_2 is HIGH and X_3 is HIGH and X_5 is HIGH and X_{11} is HIGH then output is LOW
5. If X_2 is HIGH and X_3 is AVG and X_5 is HIGH and X_{11} is HIGH then output is LOW
6. If X_2 is AVG and X_3 is LOW and X_5 is HIGH and X_{11} is HIGH then output is LOW
7. If X_2 is AVG and X_3 is LOW and X_5 is HIGH and X_{11} is AVG then output is LOW
8. If X_2 is AVG and X_3 is AVG and X_5 is HIGH and X_{11} is HIGH then output is LOW

9. If X_2 is LOW and X_3 is LOW and X_5 is HIGH and X_{11} is HIGH then output is LOW
10. If X_2 is HIGH and X_3 is LOW and X_5 is LOW and X_{11} is HIGH then output is LOW
11. If X_2 is LOW and X_3 is HIGH and X_5 is LOW and X_{11} is LOW then output is HIGH
12. If X_2 is AVG and X_3 is HIGH and X_5 is AVG and X_{11} is LOW then output is HIGH
13. If X_2 is LOW and X_3 is AVG and X_5 is LOW and X_{11} is LOW then output is HIGH
14. If X_2 is AVG and X_3 is AVG and X_5 is LOW and X_{11} is LOW then output is HIGH
15. If X_2 is LOW and X_3 is LOW and X_5 is LOW and X_{11} is LOW then output is HIGH
16. If X_2 is AVG and X_3 is HIGH and X_5 is LOW and X_{11} is LOW then output is HIGH
17. If X_2 is HIGH and X_3 is HIGH and X_5 is LOW and X_{11} is LOW then output is HIGH
18. If X_2 is LOW and X_3 is HIGH and X_5 is LOW and X_{11} is HIGH then output is HIGH
19. If X_2 is LOW and X_3 is HIGH and X_5 is AVG and X_{11} is AVG then output is HIGH
20. If X_2 is LOW and X_3 is AVG and X_5 is AVG and X_{11} is LOW then output is HIGH

For Taiwanese consumers:

1. If X_2 is HIGH and X_3 is LOW and X_5 is HIGH and X_8 is HIGH and X_{11} is HIGH then output is LOW
2. If X_2 is AVG and X_3 is AVG and X_5 is HIGH and X_8 is HIGH and X_{11} is HIGH then output is LOW
3. If X_2 is LOW and X_3 is LOW and X_5 is HIGH and X_8 is HIGH and X_{11} is HIGH then output is LOW
4. If X_2 is HIGH and X_3 is AVG and X_5 is AVG and X_8 is HIGH and X_{11} is HIGH then output is LOW
5. If X_2 is HIGH and X_3 is HIGH and X_5 is HIGH and X_8 is HIGH and X_{11} is HIGH then output is LOW
6. If X_2 is LOW and X_3 is AVG and X_5 is HIGH and X_8 is HIGH and X_{11} is HIGH then output is LOW
7. If X_2 is HIGH and X_3 is LOW and X_5 is HIGH and X_8 is AVG and X_{11} is AVG then output is LOW

8. If X_2 is AVG and X_3 is LOW and X_5 is AVG and X_8 is HIGH and X_{11} is HIGH then output is LOW
9. If X_2 is LOW and X_3 is LOW and X_5 is HIGH and X_8 is HIGH and X_{11} is LOW then output is LOW
10. If X_2 is HIGH and X_3 is LOW and X_5 is HIGH and X_8 is HIGH and X_{11} is LOW then output is LOW
11. If X_2 is LOW and X_3 is HIGH and X_5 is LOW and X_8 is LOW and X_{11} is LOW then output is HIGH
12. If X_2 is AVG and X_3 is HIGH and X_5 is LOW and X_8 is LOW and X_{11} is LOW then output is HIGH
13. If X_2 is HIGH and X_3 is HIGH and X_5 is LOW and X_8 is LOW and X_{11} is LOW then output is HIGH
14. If X_2 is LOW and X_3 is AVG and X_5 is LOW and X_8 is LOW and X_{11} is LOW then output is HIGH
15. If X_2 is LOW and X_3 is AVG and X_5 is AVG and X_8 is AVG and X_{11} is LOW then output is HIGH
16. If X_2 is AVG and X_3 is AVG and X_5 is LOW and X_8 is LOW and X_{11} is LOW then output is HIGH
17. If X_2 is LOW and X_3 is HIGH and X_5 is LOW and X_8 is LOW and X_{11} is AVG then output is HIGH
18. If X_2 is LOW and X_3 is HIGH and X_5 is LOW and X_8 is LOW and X_{11} is HIGH then output is HIGH
19. If X_2 is AVG and X_3 is HIGH and X_5 is LOW and X_8 is LOW and X_{11} is AVG then output is HIGH
20. If X_2 is LOW and X_3 is LOW and X_5 is LOW and X_8 is LOW and X_{11} is LOW then output is HIGH

Figures 1 and 2 present the ratio "X2" and "X11" in the model with defined membership functions. The ratio "X2" (annual income/total credit) represents the relationship between the annual salary of the consumer and the total credit value. In other words, this ratio informs us how long the consumer will pay back the credits. The higher the value, the shorter time it takes for the entity to pay it back, so the stronger the financial standing of the analyzed consumer and the smaller the risk of bankruptcy. Figure 1 shows that the threshold for the values of this ratio that are considered to influence positively or negatively the risk of household insolvency is fuzzified. Some values are fragmentary "LOW," partially "AVG," and fragmentary "HIGH." For example, the membership function "LOW" is a sigmoidal function with values below 0.5. However, results below 0.1 represent full membership (100%) to the fuzzy subset "LOW", while results from 0.1 to 0.5 represent both fuzzy subsets "LOW" and "AVG." Using the classical logic, we can only interpret the value of such ratio as 100% low, 100% average or 100% high. Table 3 specifies the exact range values of all three functions (“LOW”, “AVG”, and “HIGH”).

In Figure 2, we can see defined membership functions for ratio X11, which can be interpreted as the ability of consumers to pay back credits based on the evaluation of annual salary in relation to the education of the analyzed consumer. High results of this ratio represent the strong financial standing of a private person. Figure 2 and Table 3 specify exactly the range values for all three membership functions.

The second type of programmed models are artificial neural networks. While developing these models, we decided to maximize the amount of information supplied to the models. Therefore, at the input layer of each model, the values of all 12 developed ratios (Table 1) are given. In the case of the hidden layer, where the calculations are performed, there are six neurons in the model for Polish consumers and 12 neurons in the model for Taiwanese households. The need for a double hidden layer in the Asian model suggests that the forecasting process for this region is more complex than that for European consumers. The architectures of both models are presented in Figures 3 and 4.

The last two developed models were the GA-ANN. Both models are based on the multilayer perceptron structure that, in research samples with Taiwanese and Polish consumers, used a genetic algorithm with the following parameters: number of generations 200, size of the population 50. We used mutations and genetic crossover at a probability level of 0.9 for crossover and 0.1 for mutation. Both models had the same architecture of layers and neurons as the developed ANN models (Figures 3 and 4).

The output of all the above models (FL, ANN, and GA-ANN) in the learning process takes a value of 0 or 1. However, it should be noted that the output values generated by the six tested models using testing samples are not equal to the values specified in the learning sample but take values from the interval <0,1>. We adopted a threshold boundary of 0.5, meaning that households for which the model output adopts values below 0.5 are classified as at risk of bankruptcy. In contrast, models with output values above 0.5 indicate that these consumers are nonbankrupt.

After developing the six described forecasting models using two learning samples, we performed effectiveness analyses of these models on the testing samples. Table 4 presents the results.

From the obtained results, among the six models, the best effectiveness was gained by the two FL models. The FL model for Taiwanese households is characterized by 90.60 % correct classifications, while the FL model for European consumers achieved 3.60 percentage points better result (overall effectiveness of 93.90%). Another important feature of this study is that all six models generated lower type I errors (E1) than type II errors (E2). It is considered more important to pay attention to type-I errors because of the higher costs associated with this type of error. The FL model for European consumers is evaluated as the best, because of the highest overall effectiveness and the lowest type I error (with only 4.8% wrong classifications). It is also interesting to note that the ANN and GA ANN models obtained the same overall effectiveness in both consumer testing samples. In the case of Taiwanese households, it is 89.30%, and for Polish consumers, the result is better by 3.6 percentage points and equals 92.90%. Despite the same results for both models, we consider the ANN model superior to the GA-ANN model. Figures 5–7 (for the Taiwanese sample) and 8–10 (for the Polish sample) show detailed differences between the FL, ANN, and GA ANN models. All these Figures present the exact values of the forecast (from 0 to 1) against the real belonging of consumers to the solvent group (value 1) and the insolvent group (value 0). The forecast value represents the bankruptcy probability. For example, the value of 0.35 for the consumer can be interpreted as a 65% probability of going bankrupt; in other words, there is only a 35% probability that it is a future solvent consumer. The result 0 is considered a 100% bankrupt consumer, and 1 is treated as a 100% nonbankrupt household. Figures 5–10 prove that the FL and ANN models are better than the GA ANN models in predicting the bankruptcy risk of entities. Both GA ANN models noted the effect of "tightening" the results to full values (0 or 1), which resulted in the disappearance of risk classes. These models treat the analyzed consumers as 100% bankrupt or nonbankrupt. Such behavior is not consistent with real-life cases where households are often characterized by a higher or lower risk of going into bankruptcy processes. Using the FL and ANN models, we can identify the specific level of bankruptcy risk. Even if the ANN model achieved the same effectiveness as the GA ANN model, we can see that artificial neural networks were closer to the real situation with the consumers' risk classes. Therefore, both GA-ANN models are evaluated as the worst among the six proposed models from a practical perspective.

To enrich this research, we conducted comparative analyses of the effectiveness of the presented models with that of statistical models available in the literature (Korol, 2021, pp. 1–14). Table 5 shows the results of logistic regression models (LOG), classification and regression trees models (C&RT), and discriminant analysis (DA) models that were created using the same learning samples of Taiwanese and Polish households. It is important to note that a comparison of the effectiveness of developed artificial intelligence models to the effectiveness of statistical models that were estimated based on the same research samples guarantees high reliability of results, and conclusions as the same research conditions are kept.

Comparing the results in Tables 4 and 5, it is evident that the artificial intelligence models performed much better than the statistical models. First, four out of the six statistical models generated higher type I errors than type II errors. Only the DA and LOG models for Polish consumers obtained type I errors lower than type II errors. All six artificial intelligence models performed significantly better in this regard. Second, considering the overall effectiveness, only the LOG model for Taiwanese and Polish households achieved 90% or higher results. In the case of artificial intelligence models, four out of six gained effectiveness higher than 90%. It is also worth noting that the FL model for European entities outperforms the other 11 models, with effectiveness of 93.90%.

In the last stage of the study, we compared the effectiveness of developed models to the effectiveness of models from the literature. Contrary to the popular corporate bankruptcy forecasting models, it is difficult to conduct a comparative analysis of the effectiveness of household bankruptcy prediction models. As already mentioned, the literature on this subject is quite scarce, and the available articles focus mainly on the assessment of factors influencing such risk but do not propose the development of a prognostic model. In the latest papers, we found two studies. The first one concerns the consumers in Malaysia (Nor *et al.*, 2019, pp. 157–170) and the second one is devoted to households in the USA (Brygała, 2022, pp. 1–13).

The authors of both studies used two types of testing samples — balanced with an equal number of bankrupt and non-bankrupt consumers, and an imbalanced sample that consisted majority of non-bankrupt households. In Table 6, we can see that the C&RT model for Malaysian consumers (Nor *et al.*, 2019, pp. 157–170) generated better results for the imbalanced sample than for the balanced one (83.29 % vs 70.90%). Unfortunately, the authors of the C&RT model did not present the type I and II errors (E1, E2). Looking at the study by Brygała (2022) it is clear that although the imbalanced sample gives better results, in practice the model is completely unable to predict the risk for bankrupt consumers. The logit model estimated for American consumers is characterized by a very high type I error (99.71%). This means that in almost 100% of cases of future bankrupt households, the model made a mistake by classifying them as future non-bankrupt consumers. In the case of the balanced sample, the logit model of Brygała proves the good ability to forecast the future bankrupt and non-bankrupt consumers in the USA. The overall effectiveness of that model is at the level of 69.85%. This comparison proves that the artificial intelligence models have better abilities to forecast this phenomenon.

The above discussion and comparison of the results showed that it is necessary mentioning the common defects and limitations of bankruptcy forecasting models that rarely receive substantive discussion in the literature. The first defect has already been discussed — the structure of the sample. Using the imbalanced sample, we can receive higher overall effectiveness but very often it does not mean that the model has good predicting abilities. For example, if 90% of the consumers in the sample are non-bankrupt entities and only 10% of cases represent bankrupt ones, we cannot rely on overall effectiveness. It can happen that the model will generate a type I error at the level of 100%, so it has no ability to recognize the future risk of distress but overall performance will be at a high level due to the high proportion of nonbankrupt households in the sample. That is why we suggest using a balanced sample. First, it will enable the model to distinguish "good" and "bad" entities during the learning process. Second, during the testing stage, it will prove the equal abilities to recognize future bankrupt and non-bankrupt households.

The second controversy concerns the possibility of manipulating the threshold to maximize the classification results for statistical models such as logit, probit, and discriminant analysis models. Obviously, such manipulation will not increase the effectiveness of the model in the business practice (for example in the bank evaluating the real-life cases of consumers applying for the loans), but only in the given testing sample. In the case of artificial intelligence models (e.g. neural networks), it is not possible to manipulate the boundaries of classification rules.

The third inadequacy is expressed in the use of bivalent logic to describe and evaluate fuzzy, vague, and ambiguous phenomena. The financial situation of households is affected by many internal and external factors, which cannot be defined precisely. The traditional zero-one (good/bad) evaluation criteria of most statistical models have lost their relevance. In addition, a finding that a consumer is in a "good" or "bad" financial situation is imprecise because, in the current economic reality, analysts rarely have to deal with 100% "good" or 100% "bad" entities. Only by using the fuzzy sets, we can determine the precise degree of risk. Figures 5 and 9 show how usable is fuzzy logic in the estimation of this type of risk.

Conclusions

This study verifies the prediction abilities of three artificial intelligence techniques in forecasting the risk of household insolvency in two very different economic regions of the world. This proves that implementing FL guarantees high effectiveness and low-cost errors in evaluating consumer financial standing. Fuzzy sets models are superior to other artificial intelligence and statistical models in three aspects — overall effectiveness, lower type I errors, and, even more important — the use of explicit knowledge. Knowledge is a key element of such expert models. FL models require no assumptions regarding the learning procedure and are programmed with the use of analysts' knowledge and experience. Moreover, the ability to solve problems using different methods of inferences that are based on fuzzification, not on classical, sharp criteria, and in connection to the possibility to show the procedure of how the problem is solved gives economists a powerful tool with a wide spectrum of usability. The presented models can be quickly and easily updated in response to evolving economic or specific country characteristics. When changing any entry variable in other soft computing techniques or in the statistical models, we have to re-estimate a completely new model.

It is worth mentioning that this research approach uses a wide range of important information about households. The new types of ratios we proposed are not denominated in monetary value or strictly in demographic units that would be limited to only one country but can be widely used in other world regions increasing the versatility of the proposed models. It should be also noted that a single demographic indicator (for example — education level) or financial information (for example — the value of loans taken) does not support the forecasting model with a large amount of information crucial for the assessment and prediction of the financial standing of the analyzed consumer. Predicting household insolvency is an ambiguous and imprecise process. Going bankrupt is influenced by a wide variety of economic, demographic, and behavioral factors that we are not able to precisely define. This study presents a solution for effective forecasting through the implementation of a highly effective and usable method (fuzzy sets) and highly informative ratios that combine the financial and demographic properties of the evaluated consumers. The use of these ratios also improved the versatility of the presented models.

The authors are aware of the limitations of the conducted research. The main obstacle was the hard-to-collect reliable data. Moreover, it was a very time-consuming process as information about each consumer was collected individually. The authors will continue research towards the use of macroeconomic variables (such as exchange rates, GDP growth, unemployment rate, etc.) in developing the multifactor early warning system for predicting consumer bankruptcy risk. Such a system will forecast the factors negatively affecting the financial situation of households. It will be characterized by a cause-and-effect approach. For example, by forecasting the exchange rate of PLN / CHF that directly affects the deterioration of the economic situation of borrowers who have opened credit positions in Swiss Francs, we will be able to use it in evaluating the future bankruptcy risk for specific consumers. The system will consist of the “master” model and three "satellite" models. The “satellite” models will be responsible for the evaluation of macro variables and the “master” model taking into account the results of “satellites” and micro variables of consumers will generate the bankruptcy risk forecast.

References

Acosta-González, E., & Fernández-Rodríguez, F. (2014). Forecasting financial failure of firms via genetic algorithms. *Computational Economics*, *43*, 133–157. doi: 10.1007/s10614-013-9392-9.

Akkoc, S. (2012). An empirical comparison of conventional techniques, neural networks and the three stage hybrid Adaptive Neuro Fuzzy Inference System (ANFIS) model for credit scoring analysis: the case of Turkish credit card data. *European Journal of Operational Research*, *222*, 168–178. doi: 10.1016/j.ejor.2012.04.009.

Alaka, H. A., Oyedele, L. O., Owolabi, H. A., Kumar, V., Ajayi, S. O., Akinade, O. O., & Bilal, M. (2018). Systematic review of bankruptcy prediction models: towards a framework for tool selection. *Expert Systems with Applications*, *94*, 164–184. doi: 10.1016/j.eswa.2017.10.040.

Aller, C., & Grant, Ch. (2018). The effect of the financial crisis on default by Spanish households. *Journal of Financial Stability*, *36*, 39–52. doi: 10.1016/j.jfs.2018.02.006.

Altman, E. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *Journal of Finance*, *23*, 589–609. doi: 10.1111/j.1540-6261.1968.tb00843.x.

Anastasiou, D., Louri, H., & Tsionas, M. (2016). Determinants of non-performing loans: evidence from Euro-area countries. *Finance Research Letters*, *18,* 116–119. doi: 10.1016/j.frl.2016.04.008.

Ari, A., Chen, S., & Ratnovski, L. (2021). The dynamics of non-performing loans during banking crises: a new database with post-COVID-19 implications. *Journal of Banking and Finance*, *133*, 106–140. doi: 10.1016/j.jbankfin.2021.106 140.

Aristei, D., & Gallo, M. (2016). The determinants of households’ repayment difficulties on mortgage loans: evidence from Italian microdata. *International Journal of Consumer Studies*, *40*, 453–465. doi: 10.1111/ijcs.12271

Barba, A., & Pivetti, M. (2009). Rising household debt: its causes and macroeconomic implications—a long-period analysis. *Cambridge Journal of Economics*, *33*(1), 113–137. doi: 10.1093/cje/ben030.

Barboza, F., Kimura, H., & Altman, E. (2017). Machine learning models and bankruptcy prediction. *Expert Systems with Applications*, *83*, 405–417. doi: 10.1016 /j.eswa.2017.04.006.

Bellotti, T., & Crook, J. (2013). Forecasting and stress testing credit card default using dynamic models. *International Journal of Forecasting*, *29*, 563–574. doi: 10.1016/j.ijforecast.2013.04.003.

Brygała, M. (2022). Consumer bankruptcy prediction using balanced and imbalanced data. *Risks*, *10*(24), 1–13. doi:10.3390/risks10020024.

Callejon, A. M., Casado, A. M., Fernandez, M. A., & Pelaez, J. I. (2013). A system of insolvency prediction for industrial companies using a financial alternative model with neural networks. *International Journal of Computational Intelligence Systems*, *6*, 29–37. doi: 10.1080/18756891.2013.754167.

Croson, R., & Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature*, *47*(2), 448–474. doi: 10.1257/jel.47.2.448.

Delen, D., Kuzey, C., & Uyar. A. (2013). Measuring firm performance using financial ratios: a decision tree approach. *Expert Systems with Applications*, *40*, 3970–3983. doi: 10.1016/j.eswa.2013.01.012.

Diaz-Serrano, L. (2005). Income volatility and residential mortgage delinquency across the EU. *Journal of Housing Economics*, *14*, 153–177. doi: 10.1016/j.jh e.2005.07.003

Dong, M. C., Tian, S., & Chen, C. W. S. (2018). Predicting failure risk using financial ratios: quantile hazard model approach. *North American Journal of Economics and Finance*, *44*, 204–220. doi: 10.1016/j.najef.2018.01.005.

French, D., & Vigne, S. (2019). The causes and consequences of household financial strain: a systematic review. *International Review of Financial Analysis*, *62,* 150–156. doi:10.1016/j.irfa.2018.09.008.

Garcia, V., Marques, A. I., Sanchez, J. S., & Ochoa-Dominguez, H. (2019). Dissimilarity-based linear models for corporate bankruptcy prediction. *Computional Economics*, *53*, 1019–1031. doi: 10.1007/s10614-017-9783-4.

Ghent, A. C., & Kudlyak, M. (2011). Recourse and residential mortgage default: evidence from US states. *Review of Financial Studies*, *24*, 3139–3186. doi: 10.1093/rfs/hhr055.

Giannopoulos, G., & Sigbjornsen, S. (2019). Prediction of bankruptcy using financial ratios in the Greek market. *Theoretical Economics Letters, 9*, 1114–1128. doi: 10.4236/tel.2019.94072.

Gomes, F., Haliassos, M., & Ramadorai, T. (2021). Household finance. *Journal of Economic Literature*, *59*(3), 919–1000. doi: 10.1257/jel.20201461.

Gross, T., & Notowidigdo, M., J. (2011). Health insurance and the consumer bankruptcy decision: evidence from expansions of Medicaid. *Journal of Public Economics*, *95*, 767–778. doi: 10.1016/j.jpubeco.2011.01.012.

Gross, M., & Poblacion, H. (2017). Assessing the efficacy of borrower-based macroprudential policy using an integrated micro-macro model for European households. *Economic Modelling*, *61,* 510–528. doi: 10.1016/j.econmod.2016.1 2.029.

Guiso, L., Sapienza, P., & Zingales, L. (2013). The determinants of attitudes towards strategic default on mortgages. *Journal of Finance*, *68*, 1473–1515. doi: 10.1111/jofi.12044.

Guiso, L., & Sodini, P. (2013). Chapter 21 – household finance: an emerging field. In G. M. Constantinides, M. Harris & R. M. Stulz (Eds.). *Handbook of the economics of finance*, *2*(b). Elsevier, 1397–1532. doi: 10.1016/B978-0-44-459406-8.00021-4.

Hira, T. K. (2012). Promoting sustainable financial behaviour: implications for education and research. *International Journal of Consumer Studies*, *36*, 502–507. doi: 10.1111/j.1470-6431.2012.01115.x

Hosaka, T. (2019). Bankruptcy prediction using imaged financial ratios and convolutional neural networks. *Expert Systems with Applications*, *117*, 287–299. doi: 10.1016/j.eswa.2018.09.039.

I-Cheng, Y., & Che-Hui, L. (2009). The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients. *Expert Systems with Applications*, *36*(2), 2473–2480. doi: 10.1016/j.eswa.2007.12.020.

Jianakoplos, N. A., & Bernasek, A. (1998). Are women more risk averse? *Economic Inquiry*, *36*(4),620–630. doi: 10.1111/j.1465-7295.1998.tb01740.x

Jardin, P. (2018). Failure pattern-based ensembles applied to bankruptcy forecasting. *Decision Support Systems*, *107*, 64–77. doi: 10.1016/j.dss.2018.01.003.

Kieschnick, R., La Plante, M., & Moussawi, R. (2013). Working capital management and shareholders' wealth. *Review of Finance*, *17*, 1827–1852. doi: 10.109 3/rof/rfs043.

Korol, T. (2021). Examining statistical methods in forecasting financial energy of households in Poland and Taiwan. *Energies*, *14*, 1–14. doi: 10.3390/en14071 821.

Korol, T. (2020). Long-term risk class migrations of non-bankrupt and bankrupt enterprises. *Journal of Business Economics and Management*, *21*(3), 783–804. doi: 10.3846/jbem.2020.12224.

Korol, T. (2018). The implementation of Fuzzy Logic in forecasting financial ratios. *Contemporary Economics*, *12*(2), 165–187. doi: 10.5709/ce.1897-9254.270.

Korol, T., & Fotiadis, A. (2016). Applying Fuzzy Logic of expert knowledge for accurate predictive algorithms of customer traffic flows in theme parks. *International Journal of Information Technology & Decision Making*, *15*(6), 1451–1468. doi: 10.1142/S0219622016500425.

Kukk, M. (2016). How did household indebtedness hamper consumption during the recession? Evidence from micro data. *Journal of Comparative Economics*, *44*, 764–785. doi: 10.1016/j.jce.2015.07.004.

Li, L., Strahan, P., E., & Zhang, S. (2020). Banks as lenders of first resort: evidence from the COVID-19 crisis. *Review of Corporate Finance Studies*, *9*(3), 472–500. doi: 10.1093/rcfs/cfaa009.

Lin, F., Liang, D., Yeh, C. C., & Huang, J. C. (2014). Novel feature selection methods to financial distress prediction. *Expert Systems with Applications*, *41*, 2472–2483. doi: 10.1016/j.eswa.2013.09.047.

Louzada, F., Ara, A., & Fernandes, G. B. (2016). Classification methods applied to credit scoring: systematic review and overall comparison. *Surveys in Operations Research and Management Science*, *21*, 117–135. doi: 10.1016/j.so rms. 2016.10.001.

Lukason, O., & Hoffman, R.C. (2014). Firm bankruptcy probability and causes: an integrated study. *International Journal of Business and Management*, *9*, 80–91. doi: 10.5539/ijbm.v9n11p80.

Luzzetti, M., N., & Neumuller, S. (2016). Learning and the dynamics of consumer unsecured debt and bankruptcies. *Journal of Economic Dynamics & Control*, *67*, 22–39. doi: 10.1016/j.jedc.2016.03.007.

Mihalovic, M. (2016). Performance comparison of multiple discriminant analysis and Logit models in bankruptcy prediction. *Economics and Sociology*, *9*, 101–118. doi: 10.14254/2071-789X.2016/9-4/6.

Mitchell, M. (1999). *An introduction to genetic algorithms.* London: MIT Press.

Nor, S. H. S, Ismail, S., & Yap, B. W. (2019). Personal bankruptcy prediction using decision tree model. *Journal of Economics, Finance and Administrative Science*, *24*(47), 157–170. doi: 10.1108/JEFAS-08-2018-0076.

Paskevicius, A., & Jurgaityte, N. (2015). Bankruptcy of natural persons in Lithuania: reasons and problems. *Procedia - Social and Behavioral Sciences*, *213*, 521–526. doi: 10.1016/j.sbspro.2015.11.444.

Patel, A., Balmer, N. J., & Pleasence, P. (2012). Debt and disadvantage: the experience of unmanageable debt and financial difficulty in England and Wales. *International Journal of Consumer Studies*, *36*, 556–565. doi: 10.1111/j.1470-6431.2012.01121.x.

Ptak-Chmielewska, A. (2019). Predicting micro-enterprise failures using data mining techniques. *Journal of Risk and Financial Management*, *12*, 1–17. doi: 10.3390/jrfm12010030.

Sun, J., Li, H., Huang, Q., & He, K. (2014). Predicting financial distress and corporate failure—a review from the state-of-the-art definitions, modeling, sampling, and featuring approaches. *Knowledge-Based Systems*, *57*, 41–56. doi: 10.1016/j .knosys.2013.12.006.

Thorne, D. (2010). Extreme financial strain: emergent chores, gender inequality and emotional distress. *Journal of Family and Economic Issues*, *31*(2), 185–197. doi: 10.1007/s10834-010-9189-0.

Tsai, C. (2014). Combining cluster analysis with classifier ensembles to predict financial distress. *Information Fusion*, *16*, 46–58. doi: 10.1016/j.inffus.2011.12 .001.

Tufano, P. (2009). Consumer finance. *Annual Review of Financial Economics*, *1*, 227–247. doi: 10.1146/annurev.financial.050808.114457.

Worthington, A. C. (2006). Debt as a source of financial stress in Australian households. *International Journal of Consumer Studies*, *30*, 2–15. doi: 10.1111 /j.1470-6431.2005.00420.x.

Wu, D. D., Zhang, Y., & Olson, D. L. (2010). Fuzzy multi-objective programming for supplier selection and risk modeling: a possibility approach. *European Journal of Operational Research*, *200*, 774–787. doi: 10.1016/j.ejor.2009.01.026.

Xiao, Z., Yang, X., Pang, Y., & Dang, X. (2012). The prediction for listed companies' financial distress by using multiple prediction methods with rough set and Dempster-Shafer evidence theory. *Knowledge-Based Systems*, *26*, 196–206. doi: 10.1016/j.knosys.2011.08.001.

Zadeh, L. (1965). Fuzzy sets. *Information and Control*, *8*(3), 338–353. doi: 10.1016/S0019-9958(65)90241-X.

Zhang, J., & Thomas, L.C. (2012). Comparisons of linear regression and survival analysis using single and mixture distributions approaches in modelling LGD. *International Journal of Forecasting*, *28*, 204–215. doi: 10.1016/j.ijforecast.2010.06.002.

Zurawicki, L., & Braidot, N. (2005). Consumers during crisis: responses from the middle class in Argentina. *Journal of Business Research*, *58*, 1100–1109. doi: 10.1016/j.jbusres.2004.03.005.

Acknowledgments

This research has been prepared within the grant project No. 2017/25/B/HS4/00592, "Forecasting the risk of consumer bankruptcy in Poland." Research funded by the National Science Centre in Poland (Narodowe Centrum Nauki).

Annex

Table 1. The set of developed ratios used in programming the models

Symbol	Formula	Interpretation
X1	annual income / value of total assets	It represents the share of annual income in value of total assets owned by the consumer.
X2	annual income / total credits	It represents the relation of the annual salary of the consumer to the value of total credits. In other words, it can be used to evaluate how many years it will take to pay back all the credits. The higher the value, the stronger the financial standing of the analyzed consumer.
X3	interest rates paid / monthly income	It represents the burden of monthly salary with payments of interest rates of credits–the lower the value, the better the financial solvency of the analyzed household.
X4	credits cards / total credits	It represents the share of credit cards in the value of total credits taken by the consumer. The higher the value, the higher the insolvency risk of the analyzed consumer.
X5	(value of total assets – total credits) / total credits	It shows the net value of total assets to the value of total credits taken by the consumer. The higher the value, the stronger the financial standing of the analyzed consumer.
X6	value of total assets / total credits	It shows how many times greater the total assets owned by the consumer are in relation to the value of all credits taken by them. The higher the value, the lower the risk of bankruptcy of the analyzed consumer.
X7	monthly income/credits cards	It shows how many times greater the monthly salary is compared to the credit card debt. The higher the value, the lower the risk of insolvency of the analyzed household.
X8	education / age	The following values describe the variable 'education': 1 – elementary school, 2 – high skilled worker, 3 – bachelor's degree, 4 – master's or doctorate degree. The variable 'age' is described by the following values: 1 – from 27 to 50 years old, 2 – from 51 to 60 years old, 3 – younger than 26 or older than 60 years old. The higher the value of this relation, the lower the risk of bankruptcy of the analyzed consumer.
X9	education/number of children	The variable 'number of children' is described by the following values: 1 – from 0 to 2 children, 2 – from 3 to 4 children, 3 – more than 4 children. The higher the value of this relation, the lower the risk of insolvency of the analyzed consumer.
X10	marital status/length of employment	The variable 'marital status' is described by two values: 1 – married, 2 – single, widowed. The variable 'length of employment' is described by three values: 1 – up to 5 years of job experience, 2 – from 6 to 10 years of work experience, 3 – more than 11 years of work experience. The higher the value of this relation, the higher the risk of insolvency of the analyzed consumer.

Table 1. Continued

Symbol	Formula	Interpretation
X11	education / (total credits/annual income)	It represents the ability of the consumer to repay the credits taking into account the annual salary and the level of education of the consumer. The higher the value of this relation, the stronger the financial standing of the analyzed consumer.
X12	age / (total credits/ annual income)	It represents the ability of the consumer to repay the credits taking into account the annual salary and the age of the consumer. The higher the value of this relation, the lower the bankruptcy risk of the analyzed consumer.

Source: Korol (2021, pp. 1–14).

Table 2. Distribution of age and gender of evaluated consumers

Gender and age of consumers		Bankrupt	Nonbankrupt
		Poland	
Female	**<26**	17	37
	27-50	108	162
	51-60	32	36
	>60	3	25
	total	**160**	**260**
Male	**<26**	48	29
	27-50	202	154
	51-60	69	38
	>60	21	19
	total	**340**	**240**
		Taiwan	
Female	**<26**	11	32
	27-50	99	122
	51-60	24	47
	>60	4	25
	total	**138**	**226**
Male	**<26**	37	36
	27-50	211	169
	51-60	72	45
	>60	42	24
	total	**362**	**274**

Table 3. The defined membership functions of entry variables of FL models

Symbol	Description of variable	Type of membership functions and their range of values
X2	annual income / total credits	-"LOW" – the type of function – Sigmoidal. Results below 0.1 represent the fuzzy subset "LOW" with the full membership (100%), and results from 0.1 to 0,5 represent both fuzzy subsets "LOW" and "AVG"; -"AVG" – the type of function – Gaussian. Results from 0.1 to 0.5 represent two fuzzy subsets "LOW" and "AVG," and results from 0.5 to 1.0 represent two fuzzy subsets "AVG" and "HIGH"; -"HIGH" – the type of function – Sigmoidal. Results above 1.0 represent the fuzzy subset "HIGH" with the full membership (100%), and results from 0.5 to 1.0 represent two fuzzy subsets "AVG" and "HIGH."
X3	interest rates paid / monthly income	-"LOW" – the type of function – Sigmoidal. Results below 0.2 represent the fuzzy subset "LOW" with the full membership (100%), and results from 0.2 to 0.5 represent both fuzzy subsets "LOW" and "AVG"; -"AVG" – the type of function – Gaussian. Results from 0.2 to 0.5 represent two fuzzy subsets "LOW" and "AVG," and results from 0.5 to 0.8 represent two fuzzy subsets "AVG" and "HIGH"; -"HIGH" – the type of function – Sigmoidal. Results above 0.8 represent the fuzzy subset "HIGH" with the full membership (100%), and results from 0.5 to 0.8 represent fuzzy subsets "AVG" and "HIGH."
X5	(value of total assets – total credits) / total credits	-"LOW" – the type of function – Sigmoidal. Results below 0.5 represent the fuzzy subset "LOW" with the full membership (100%), and results from 0.5 to 4.0 represent both fuzzy subsets "LOW" and "AVG"; -"AVG" – the type of function – Gaussian. Results from 0.5 to 4.0 represent both fuzzy subsets "LOW" and "AVG," and results from 4.0 to 10.0 represent two fuzzy subsets "AVG" and "HIGH"; -"HIGH" – the type of function – Sigmoidal. Results above 10.0 represent the fuzzy subset "HIGH" with the full membership (100%), and results from 4.0 to 10.0 represent fuzzy subsets "AVG" and "HIGH."
X8	education / age	-"LOW" – the type of function – Sigmoidal. Results below 0.7 represent the fuzzy subset "LOW" with the full membership (100%), and results from 0.7 to 1.5 represent both fuzzy subsets "LOW" and "AVG"; -"AVG" – the type of function – Gaussian. Results from 0.7 to 1.5 represent two fuzzy subsets "LOW" and "AVG," and results from 1.5 to 2.0 represent fuzzy subsets "AVG" and "HIGH"; -"HIGH" – the type of function – Sigmoidal. Results above 2.0 represent the fuzzy subset "HIGH" with the full membership (100%), and results from 1.5 to 2.0 represent two fuzzy subsets "AVG" and "HIGH."
X11	education / (total credits/annual income)	-"LOW" – the type of function – Sigmoidal. Results below 0.5 represent the fuzzy subset "LOW" with the full membership (100%), and results from 0.5 to 1.0 represent both fuzzy subsets "LOW" and "AVG"; -"AVG" – the type of function – Gaussian. Results from 0.5 to 1.0 represent two fuzzy subsets "LOW" and "AVG," and results from 1.0 to 1.5 represent fuzzy subsets "AVG" and "HIGH"; -"HIGH" – the type of function – Sigmoidal. Results above 1.5 represent the fuzzy subset "HIGH" with the full membership (100%), and the results from 1.0 to 1.5 represent two fuzzy subsets "AVG" and "HIGH."

Table 4. The effectiveness of developed models based on the testing sample

Testing sample		Results		
		FL	**ANN**	**GA ANN**
Taiwanese consumers	E1	7.80% (39)	8.80% (44)	8.80% (44)
	E2	11.0% (55)	12.60% (63)	12.60% (63)
	S	**90.60%**	**89.30%**	**89.30%**
Polish consumers	E1	4.80% (24)	5.80% (29)	5.80% (29)
	E2	7.40% (37)	8.40% (42)	9.60% (48)
	S	**93.90%**	**92.90%**	**92.30%**

Table 5. The effectiveness of statistical models based on the testing sample

Testing sample		Results		
		DA	**LOG**	**C&RT**
Taiwanese households	E1	13.80% (69)	10.60% (53)	17.40% (87)
	E2	10.80% (54)	9.20% (46)	15.20% (76)
	S	**87.70%**	**90.10%**	**83.70%**
Polish households	E1	8.20% (41)	6.20% (31)	15.80% (79)
	E2	12.60% (63)	8.40% (42)	13.00% (65)
	S	**89.60%**	**92.70%**	**85.60%**

Source: Korol (2021, pp. 1–14).

Table 6. The effectiveness of exemplary models from the literature

Testing sample		**Results**	
C&RT model (Nor *et al.*, 2019)	**Imbalanced sample**		83.29 %
	Balanced sample		70.90 %
Logit model (Brygała, 2022)	**Imbalanced sample**	E1	99.71 %
		E2	0.00 %
		S	95.98 %
	Balanced sample	E1	29.41 %
		E2	30.88 %
		S	69.85 %

Source: Nor *et al.* (2019, pp. 157–170); Brygała (2022, pp. 1–13).

Figure 1. The defined membership functions for the ratio X2

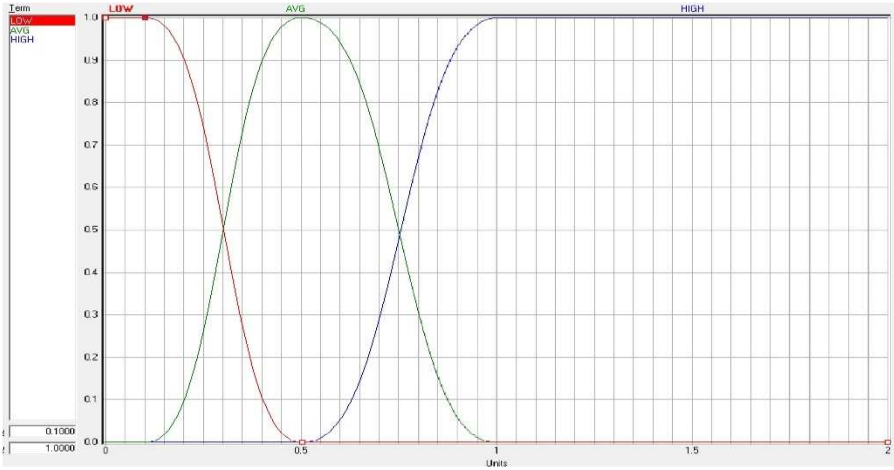

Figure 2. The defined membership functions for the ratio X11

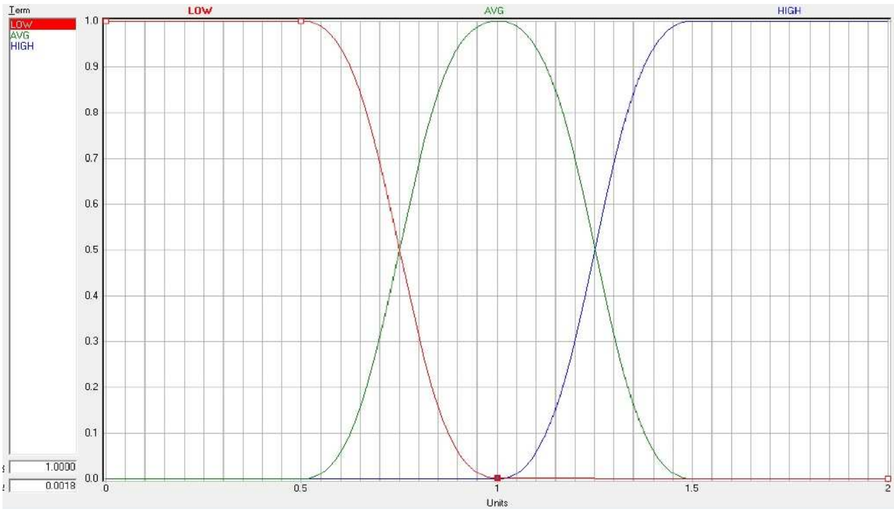

Figure 3. The structure of the ANN model for Polish consumers

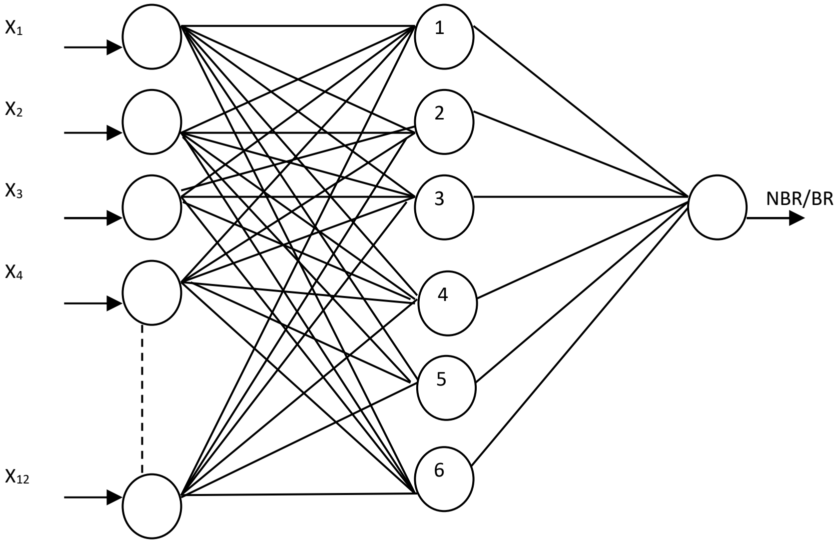

Figure 4. The structure of the ANN model for Taiwanese households

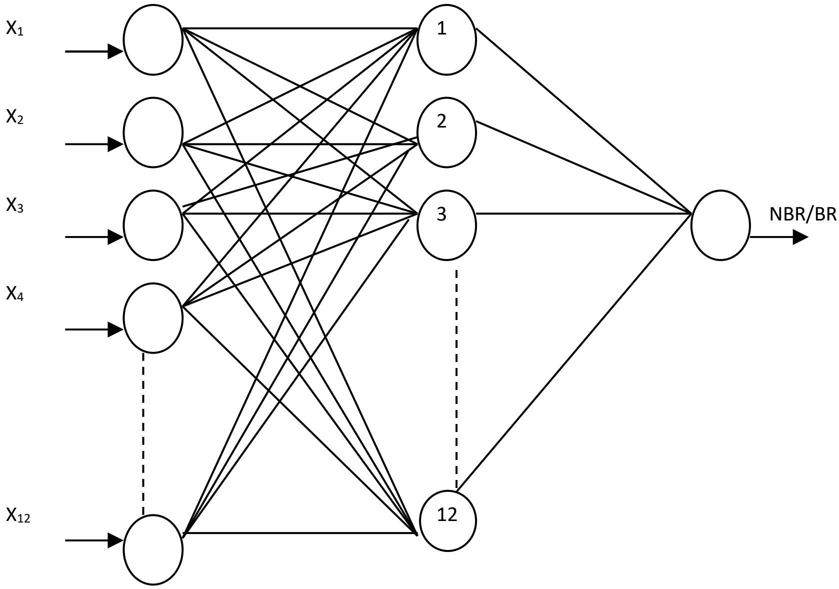

Figure 5. Accurate model classification results of FL model for Taiwanese consumers

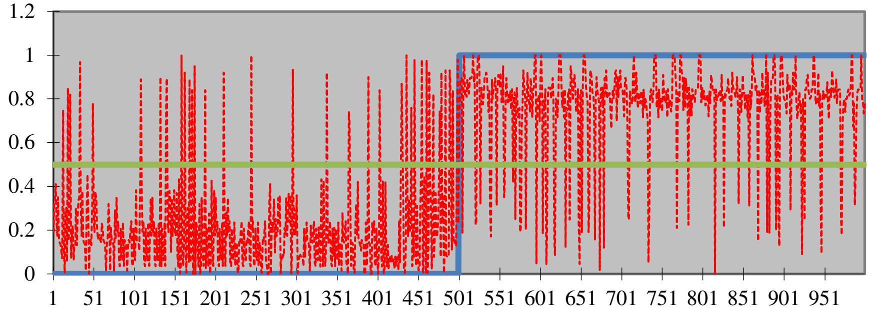

Figure 6. Accurate model classification results of GA ANN model for Taiwanese consumers

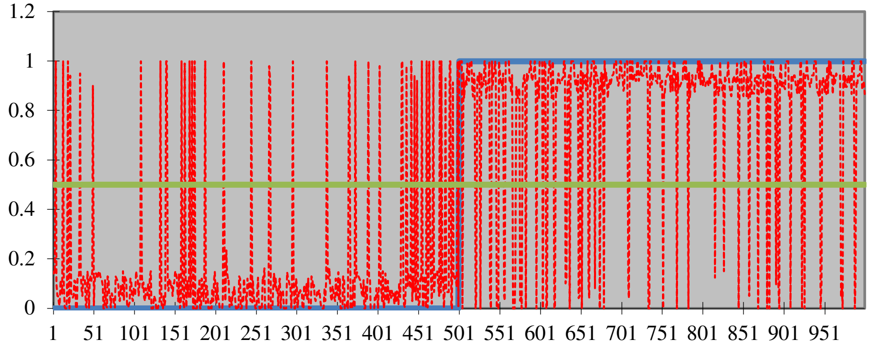

Figure 7. Accurate model classification results of ANN model for Taiwanese consumers

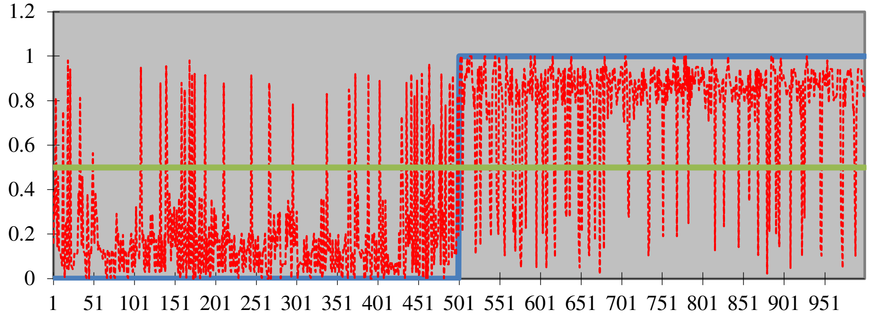

Figure 8. Accurate model classification results of GA ANN model for Polish consumers

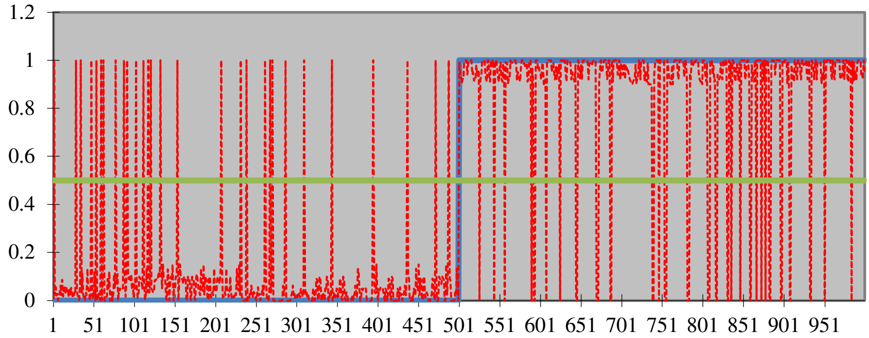

Figure 9. Accurate model classification results of FL model for Polish consumers

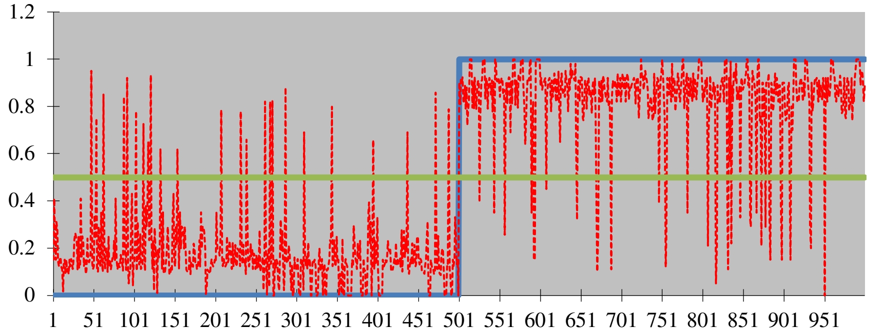

Figure 10. Accurate model classification results of ANN model for Polish consumers