Center for
Economic Theories and Policies
Sofia University St. Kliment Ohridski
Faculty of Economics and Business Administration

ISSN: 2367-7082

Bulgarian Economic Papers

SME Viability Assessment Methodology: Combining Altman's Z-Score with Big Data

Dimitar Haralampiev Popov

BEP 04-2022
Publication: June 2022

SME Viability Assessment Methodology: Combining Altman's Z-Score with Big Data

Dimitar Haralampiev Popov[1]

Abstract: Due to their important place in an economy, small and medium enterprises (SMEs) viability is the focus of numerous scientific studies, European and national programs. One of the most widely used viability prediction model is Altman's Z-score. Altman's classical models are not suitable for all situations, though. SMEs' large nominal number in an economy presents another challenge to researchers. One possible solution to this issue is to use data mining tools that can lead to new knowledge discovery. Data mining is the result of a natural evolution of information technology. The cross industry standard process for data mining (CRISP-DM) is a methodological framework for researching large amounts of data. This paper aims to outline the characteristics of Altman's Z-score and CRISP-DM, and propose combining them into a methodology for predicting SMEs' viability.

Keywords: Altman Z-score, Data mining, CRISP-DM, SMEs, Bulgaria

JEL: M10, P12, C38, C55

[1] Dimitar Popov, University of National and World Economy, Department of Management, Sofia, Bulgaria.

1. Introduction

Small and medium-sized enterprises (SMEs) are an important source of economic growth, entrepreneurship and employment in the European economy (OECD, 2020). By accepting the Small Business Act for Europe(SBA) (European commission, 2008), the European Union (EU) provides a comprehensive framework for SME policies, incentivizes entrepreneurship and adopts the principle 'Think Small First' in the development of legislation and policies, in order to improve the SMEs' competitiveness. The National Strategy for 'Small and Medium-sized Enterprises 2021-2027' aims to increase capacity-building and support for the transition to sustainability and digitalization, reducing regulatory burden, improving market access and financing (European Commission, 2020), Altman's model that has put the beginning of solid methodological framework used by many researchers since. He uses financial ratios from the balance sheet and income statement of enterprises as viability predictors. The model is based on multiple discriminant analysis as a statistical technique for classifying a population. Altman's old model however, in addition to being outdated, is specifically focused for different business environment and other types of enterprises different from Bulgarian SMEs.

Another issue that arises looking at SMEs empirical data is the data volume that has to be explored. There are numerous financial ratios popular in the viability prediction literature. When combined with the sheer number of enterprises that SMEs comprise in an economy, it becomes clear that a new solution is to be looked up when researching viability. The rise of computational capacity and information rich environments leads to the development of data mining tools, which are a possible solution. This paper explores the aforementioned issues in detail and suggest a new methodological framework that combines Altman's Z-score as a viability predictor with CRISP-DM as a standard processing model.

2. SME Importance

SMEs traditionally have a very important role in the Bulgarian economy. According to the 2019 SBA Fact Sheet, they represent three quarters of the country's employment and two thirds of the total added value. In both cases, this is about 9 percentage points above the EU average. Through 2014-2018 the value added of Bulgarian non-financial SMEs increased by 50.7%. SME employment growth was moderately up by 8.6 percentage points. The European Commission expects to maintain the positive trend in SME value-added growth. For 2017-2018, SME value added continued to increase strongly by 15.0%, whereas SME employment grew by only 1.6% (European Commission, 2019).

SMEs in the Bulgarian SME Act

Small businesses are generally private companies, they have limited financial and human resources and a small amount of sales. Legislation that defines a small business may vary according to the criteria and size used. The main criteria are usually the number of employees, the annual turnover, the assets value, the ownership and the interconnection between the enterprises. The Bulgarian Small and Medium Enterprises Act determines the company's size according to these indicators.

Table 1: Company size

Criteria	Micro enterprises	Small enterprises	Medium enterprises
Staff headcount	Less than 10	Less than 50	Less than 250
Turnover	Less than BGN 3,9 mil.	Less than BGN 19,5 mil.	Less than BGN 97,5 mil.
Balance sheet total	Less than BGN 3,9 mil.	Less than BGN 19,5 mil.	Less than BGN 84 mil.
Autonomous enterprise within the meaning of Bulgarian SME Act[2]			

The Bulgarian legislation is harmonized with the EU and the criteria are identical with all other European countries (European Commission, 2008). This gives significant advantages in terms of research and empirical data comparability. From a national point of view, the harmonized criteria allows data accumulation on massive level by institutions such as the Bulgarian National Statistical Institute which in turn can be easily accessed and used for various scientific research such as SMEs' viability.

Table 2: SMEs basic figures

Class size	Number of enterprises			Number of persons employed			Value added		
	Bulgaria		EU-28	Bulgaria		EU-28	Bulgaria		EU-28
	Number	Share	Share	Number	Share	Share	Billion €	Share	Share
Micro	315,410	91.8%	93.0%	616,012	30.6%	29.7%	6.5	21.6%	20.8%
Small	23,471	6.8%	5.9%	477,693	23.7%	20.1%	7	23.2%	17.6%
Medium	4,248	1.2%	0.9%	432,689	21.5%	16.8%	6.1	20.5%	18.0%
SMEs	343,129	99.8%	99.8%	1,526,394	75.7%	66.6%	19.6	65.3%	56.4%
Large	623	0.2%	0.2%	489,587	24.3%	33.4%	10.4	34.7%	43.6%
Total	343,752	100.0%	100.0%	2,015,981	100%	100%	30	100%	100%

European Commission (2019)

[2] According to the Bulgarian Small and Medium Enterprises Act (2020), an independent enterprise is an enterprise in which no more than 25% of the capital or of the votes in the general meeting are controlled by another enterprise except for the specific cases mentioned.

Application for SMEs

Data mining for SMEs depends on the quantity of available data volume. Key SMEs characteristic is that their cumulative count in a country's economy is usually significant. As can be seen in Table 2 the 2020 National Statistical Institute (NSI) data for big enterprises i.e., over 250 employees, are less than 0.2% of all the country's enterprises. SMEs are 99.8% which means the vast majority of enterprises are under the EU definition for SMEs. Of course, this does not mean that the economic impact of SMEs surpasses that of the large enterprises, but it gives an indication that current Small and Medium Business issues will be important in the future. Table 3 provides an overview of the number of enterprises, employees and sectoral activities of SME in Bulgaria.

Table 3: Number of enterprises by groups and sectors relative to number of employees for 2020

Sectors by KID-2008	Total	Groups by employee number			
		0-9	10-49	50-249	250+
Agriculture, forestry and fishing	19126	17494	1513	106	13
Mining and quarrying	350	238	69	32	11
Manufacturing	30502	23881	4889	1474	258
Electricity, gas, steam and air conditioning	2137	2019	80	21	17
Water supply; sewerage, waste management	829	592	136	64	37
Construction	21893	18841	2553	462	37
Wholesale and retail trade; repair of motor	139376	132377	6147	770	82
Transportation and storage	22555	20608	1639	264	44
Accommodation and food service activities	25618	23469	1936	202	11
Information and communication	15632	14358	970	250	54
Real estate activities	25515	25032	447	..	..
Professional, scientific and technical activities	47648	46469	1042	119	18
Administrative and support service	11829	10688	843	242	56
Education	4211	4034	152	..	..
Human health and social work activities	14287	13353	654	208	72
Arts, entertainment and recreation	5809	5494	246	64	5
Other service activities	24247	24066	173	8	-
Total	411564	383013	23489	4286	715
Percentage of Enterprises	100.0%	93.1%	5.7%	1.0%	0.2%

".." Confidential data

"-" No cases

National Statistical Institute (2021)

3. Viability Assessment Methodology

The methodology has to give us the tools to handle two major characteristics of Bulgarian SME Viability Assessment. First, we need a conceptual model that can predict viability using SME's characteristics. Second, we need a standard process that will allow us to successfully examine a large amount of SMEs empirical data. For that purpose we will use the Altman Z-score and the CRISP-DM.

Altman Z-score

Company viability is studied using financial modeling, most notably the Altman Z-score. The model was developed in the 1960s by Edward Altman. It attempts to predict whether a public traded company will remain viable within a two-year period. The value obtained from the model is referred to as the company' Z-score, which is a reasonably accurate predictor for future viability/bankruptcy. The model is based on multiple discriminant analysis as a classification technique. In the original model, the lower the value of Z, the more likely the company is going bankrupt. As input variables, Altman uses a linear combination of four or five common business ratios, weighted by coefficients. The coefficients are assessed by identifying an aggregate of companies declared bankrupt and then collecting a cohesive sample of surviving firms with industry coincidence and approximate asset size. Reasons for creating a new model

There are several legitimate reasons why the viability classification model should be revised and improved. First, the original model is built for big publicly traded companies and is not suited for SMEs. Second, Altman shows that the new model should be up-to-date with regard to the temporal nature of the data. Third, previous models focus on a wide range of manufacturing enterprises or on specific industries. Altman believes that with appropriate analytical adjustments, particularly retail vulnerable banks, could be analyzed on an equal footing with manufacturers. An important feature of this type of study is that data of the financial statements need to be carefully analyzed to incorporate the latest changes in financial accounting standards and accepted good accounting practices. The purpose of these modifications is to make the model not only related to previous failures but also to data that will occur in the future. Finally, it is important to test and evaluate some of the latest achievements and still controversial aspects of discriminant analysis (Altman, 2000).

Considering the above reasons, it is recommended that the Altman model should not be taken as a universal instrument. Different countries have their own specific economic characteristics. Differences in local accounting legislation, although to some extent synchronized with European standards, the average size of companies surveyed and the conditions under which they find themselves in distress and bankruptcy, can be taken into account.

Financial ratios as predictors

A substantial foundational research about financial ratios as predictors of viability is done by Beaver (1966). He uses univariate analysis for viability prediction. Beaver discovers that certain ratios could be used to discern viable enterprises from defaulted ones, five years before the event occurs. His research creates an important framework for prediction using some enterprise characteristics. His research shows that best prognosticators are ratios for profit, liquidity, long-term solvency. The financial ratios eliminate part of the statistical effect of enterprise's size which increases their comparability.

Altman (2000) points out several shortcomings with the Beaver approach. The univariate analysis is focused on individual problems and misses the overall condition of multiple enterprises. Analyzing a single ratio can cause confusion and misinterpretation. For example, an enterprise can have small profit or even a loss, but having a long-term solvency, this could not be an issue. This is why Altman decides to include multiple meaningful variables in the analysis. The empirical study needs to answer three important questions:

- Which are the most important ratios predicting viability?
- What weight coefficients should be attached to them?
- In what objective way should these coefficients be calculated?

Altman's Z-score is a proven model for enterprise viability prediction. It is widely used in literature as a model with high predictive accuracy and foundational theoretical framework. In order to be used for Bulgarian enterprises, a new Altman based model should be trained. It uses financial ratios from local SMEs empirical data (Popov, 2018).

Table 4 Financial ratios by groups

GROUP III (DEBT TO TOTAL-ASSET RATIOS) 1. Current liabilities to total assets 2. Long-term liabilities to total assets 3. Current plus long-term liabilities to total assets 4. Current plus long-term plus preferred stock to total assets
GROUP IV (LIQUID-ASSET TO 1. Cash to total assets 2. Quick assets to total assets 3. Current assets to total assets 4. Working capital to total assets
GROUP V (LIQUID-ASSET TO RENT DEBT RATIOS) 1. Cash to current liabilities 2. Quick assets to current liabilities 3. Current ratio (current assets to current liabilities)
GROUP VI (TURNOVER RATIOS) 1. Cash flow to sales 2. Accounts receivable to sales 3. Inventory to sales 4. Quick assets to sales 5. Current assets to sales 6. Working capital to sales 7. Net worth to sales 8. Total assets to sales 9. Cash interval (cash to fund expenditures for operations) 10. Defensive interval (defensive assets to fund expenditures for operations) 11. No-credit interval (defensive assets minus current liabilities to fund expenditures for operations)

Beaver (1966)

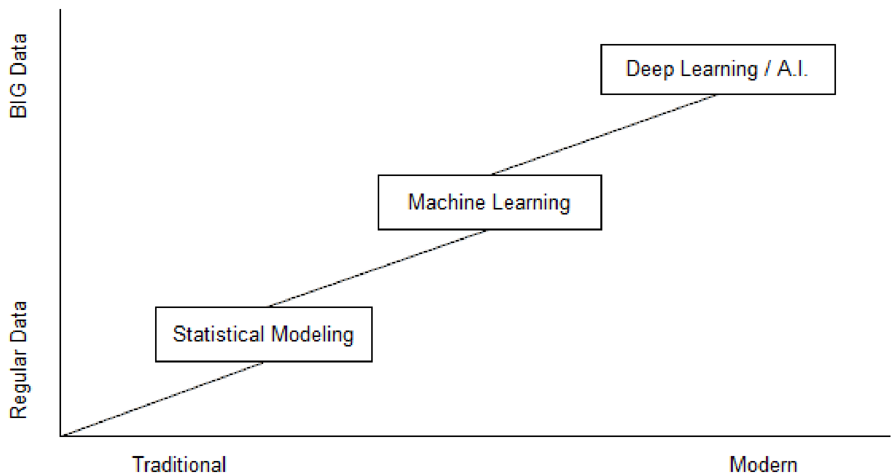
Data Mining Concept

Data mining is part of the natural evolution of information technology. The database management industry is evolving in the development of several critical functionalities. Since the 1960s, the databases and information technology are systemically developing from primitive file processing systems to complex and powerful database systems. The huge amount of data necessitates web-based applications. The integration and adaptation of such a big amount of data is still an issue that leads to the situation *data rich, but information poor*. The growing gap between data and information leads to the development of data mining instruments.

In the beginning, the popular term is *Practical Machine Learning*. The term *data mining* is introduced due to marketing reasons (Han et al., 2011). Data mining identifies information nuggets from massive datasets through various techniques. The information can be used for decision-making, forecasts, forecast evaluations. Except for the raw analysis, the technique includes aspects of database management, preliminary data processing, model considerations and conclusion, indicators of interest, complexity issues, open-structure processing, visualization and online actualization. Data mining is a step in the analytical process of knowledge discovery in databases (Fayyad et al, 1996).

For data mining, success comes with the combination of expert knowledge with advanced analytical techniques in which the computer identifies the main nods and functions in data. The data mining process generates models of historical data which are then used for forecasts, pattern discovery etc. This model generation technique is called machine learning or modeling.[3]

Figure 1: Machine learning relative to the data volume



Gavino & Anthony (2017), author's interpretation

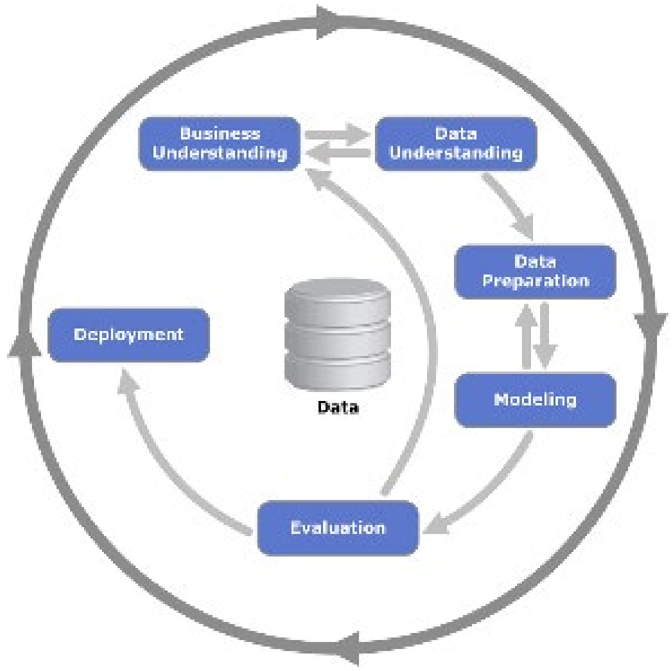
Cross-industry standard process for data mining (CRISP-DM)

Four leading companies in the data mining market create CRISP-DM in the late 1996. Daimler-Benz (now DaimlerChrysler), Integral Solutions Ltd. (ISL), NCR and OHRA. At that time, Daimler-Benz is leading most of the industrial commercial organizations in applied business operations data mining. A consortium is created a year later aiming at developing CRISP-DM. Because CRISP-DM is meant to be common industry standard with neutral application, the consortium gathers data from a wide range of organizations like data warehouse providers and management consultants who have interest in data mining.

[3] IBM® SPSS® Modeler 18.0 User's Guide, 23

To provide access to this information, a special interests group for CRISP-DM is created aiming at developing a standard processing model servicing the data mining community (Shearer, 2000).

Figure 2. Referent model CRISP-DM, phases [4]



CRISP-DM is a comprehensive data mining methodology and process model, which provides every novice expert a full plan for conducting a data mining project. CRISP-DM divides the data mining project life cycle into six phases: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation and Deployment.

Figure 3 shows the data mining process phases. The arrows point out the most important and frequent relations between the phases, while the outer circle symbolizes the cyclical character of data processing and illustrates that the lessons, learned during the data mining process and the deployed decision, can start new, often more focused questions. CRISP-DM provides a standardized approach for knowledge extraction from data for SMEs. The methodology can be applied if we have to:

- Understand the business problem for SMEs. Outline the specific goals that are defined for the study.
- Select the available data for the issue. To describe and examine the available data.
- Prepare the data, transforming it in a suitable for modeling format. Create derivative and auxiliary variables.
- Create SMEs assessment model.

[4] Graph by Jensen, K. (2012) https://commons.wikimedia.org/w/index.php?curid=24930610

- Evaluate the model performance. Deploy the model for the access and benefit of the stakeholders.

In each of the above phases, it is possible that the data scientist has to go back and reassess a previous methodology phase.

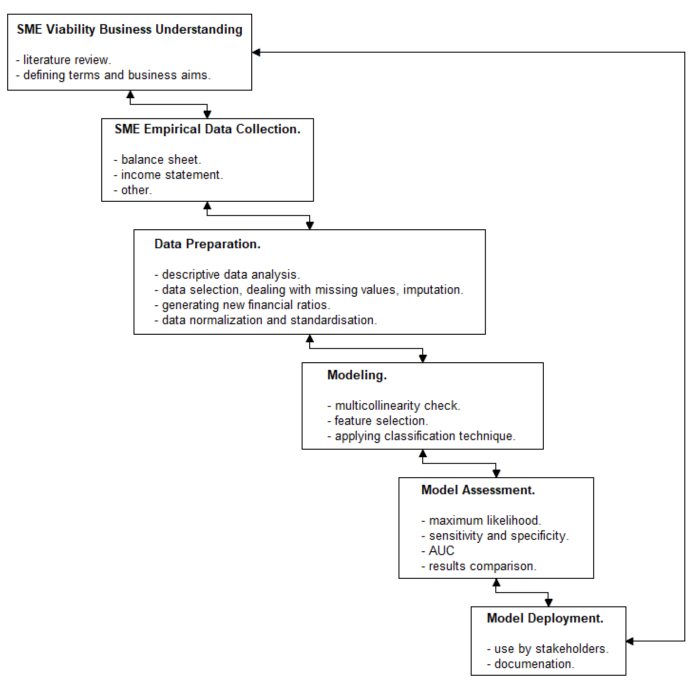
4. Combining Altman’s Z-score with Data Mining

In order to train an appropriate viability prediction model, that takes into account the massive amount of data, that SMEs generate, we will use CRISP-DM as an overarching general methodological framework. In each of the steps, we will built in the necessary concrete business understanding and characteristics specific to Altman’s and others research. The result will be an SMEs viability prediction model suitable for the Bulgarian economic landscape.

Research Process

Below is a block scheme showing the methodology workflow process. The Figure 3 goes through the methodology steps for training a new SME prediction model. Important note is that we can go back and forth between the steps as needed because with every step taken there is new information gained which clarifies the steps before. Ex: the descriptive analysis can contribute to Business understanding, Modeling can reveal new aspects for Data preparation and so forth. The methodology uses highly practical approach implementing *what works*, rather than sticking to the Altman old approach.

Figure 3: Methodology workflow



Data preparation

This section includes making a descriptive analysis that can lead to a better understanding of the gathered empirical data. Also the financial ratios used in previous research will be generated here. This step can double the predictor variables available for modeling. Also special attention has to be paid for solving the issue of missing values. Observations can be filled out using a form of imputation or outright skipped if too few data is available.

Classification models

After preparing the data, there needs to be a classification model that will distinguish between viable SMEs and the ones in distress. The below models will give us clear cut groups:

Logistic regression. Logistic regression is often preferred because it is more flexible in its assumptions and the types of data that can be analyzed to avoid the impact of lack of multivariate normality, value deviations, multicollinearity and singularity when developing a viability model forecasting (Terdpaopong, 2011). Altman and Sabato find that SMEs need a model specifically tailored to their needs and that prediction modeling is sensitive to the choice of model variables and modeling techniques (Altman & Sabato, 2005).

Decision tree. The decision tree classification method can be utilized in the context of data mining. Discrete and continuous values can be used as data for both independent and dependent variables. The process pruning not only significantly reduces the size of the Tree, but also improves the accuracy of classification (Kamiński et al., 2018).

Naive Bayes classifier. One of the classical algorithms in machine learning is the Naive Bayesian Classifier, which is based on Bayes' theorem for determining the a posteriori probability of an event. The application of the algorithm makes a "naive" assumption of strong independence between the pairs of variables. This means that each of the variables is assumed to contribute independently to determining the probability value of the independent variable (Hastie et al., 2009). Therefore, it is very important to do a proper feature selection and negate multicollinearity in data beforehand. Naive Bayes is highly scalable and usually has a high predictive accuracy.

Random forest. The method is the averaging of multiple decision tree algorithms to build a single model, as models can capture complex interactions. With a strong signal in the data, a non-random predictive result is obtained. A possible disadvantage is when the deviations between different trees are strongly correlated. In this case, they will be preserved in the random forest (Hastie et al., 2009).

Model Assessment

The results from training the classification models have to be properly assessed as to choose the most appropriate technique for viability prediction. The Area under the curve indicator combines all the available metrics but others too can be analyzed as part of the results. Another thing to consider is that results can lead to rework of some of the previous steps, for example, normalizing the data, implementing new imputation technique etc.

Overall, the research methodology allows for a very flexible approach towards the SME prediction process. Every step can lead back and forth when new information becomes available and additional techniques can be added if need be. Additional predictors can be added and iterated based on their predictive power and features can be excluded whenever they are correlated with each other. Furthermore, different models can be trained and compared based on their results.

5. Conclusion

SMEs are a very important factor for EUs economy, employment and general well-being. That is why SMEs' viability is of crucial importance. The sheer number of Bulgarian SMEs and the various data gathered in measuring their activity poses a challenge for future studies. Combining CRISP-DM with Altman's Z-score provides an easy to follow step by step practical approach to predictive modeling. The original Altman model used small samples of tens of enterprises. Backed by modern computation technology, the samples could rise to thousands of enterprises. Also, virtually all financial ratios suggested in the literature by other studies can be included, iterated through the methodology process and included in the final model training based on results. This practical approach of what works contributes to flexibility in choosing the appropriate techniques in the modeling process. Since the criteria is SME viability prediction, the model with highest predictive power can be chosen. That being said, the process should be limited as to the stated goals in the Business Understanding faze. In the end, a new methodology is created that serves to train a model suitable for the Bulgarian SMEs conditions and timeframe. Combining data mining with Altman's Z-score results in a complete methodology that solves both predicting viability and also handling big datasets that SMEs' sheer number produces.

References

Altman, E. (2000) Predicting Financial Distress of Companies: Revisiting the Z-Score and Zeta®Models. *Journal of Banking & Finance* 1, 8-36.

Altman, E.I.; Sabato, G. (2007) Modelling Credit Risk for SMEs: Evidence from the U.S. Market. *Abacus* 43, 332-357.

Beaver, W. (1966) Financial rations as predictor for failure. *Journal of Accounting Research* 4, 71-111.

Fayyad, U.; Piatetsky-Shapiro G.; Smyth, P. (1996) From Data Mining to Knowledge Discovery in Databases. *AI Magazine* 17(3), 37-54.

Gavino, D.; Anthony, A. (2017) Using Big Data to Enhance Library Services. *Philippine Association of Academic/Research Librarians*.

Han, J.; Kamber, M., Pei, J. (2011) Data mining: concepts and techniques. Morgan Kaufmann.

Hastie T.; Tibshirani R.; Friedman J. (2009) The Elements of Statistical Learning –Data Mining, Inference, and Prediction. Berlin: Springer.

Kamiński, B.; Jakubczyk, M.; Szufel, P. (2018) A framework for sensitivity analysis of decision trees. *Central European Journal of Operations Research* 26, 135-159.

Popov, D. (2018) Application of Altman Z-score for Viability of Small and Medium Enterprises. *Bulgarian Journal of Business Research* 2, 5-17.

Shearer C. (2000) The CRISP-DM model: the new blueprint for data mining. *Journal of Data Warehousing* 5(4), 13-22.

Terdpaopong, K. (2011) Identifying an SME's debt crisis potential by using logistic regression analysis. *Journal of Current Science and Technology* 1(1), 17-26.

National Statistical Institute (2021) Number of non-financial enterprises by size in terms of employed and economic activity groupings. *www.nsi.bg*. https://www.nsi.bg/en/content/8225/number-enterprises

Bulgarian Small and Medium Enterprises Act (2020)

European Commission (2020) Technical report including in-depth analysis of the SMEs in Bulgaria and preliminary recommendations for the directions of action of the new strategy. Brussels: Office for Official Publications of the European Communities 27.

The Organisation for Economic Co-operation and Development (2020) Financing SMEs and Entrepreneurs 2020: An OECD Scoreboard. Paris: The Organisation for Economic Co-operation and Development Publishing.

European Commission, Directorate-General for Enterprise and Industry (2019) SBA Fact Sheet. Brussels: EC Publications Office.

European Commission, Directorate-General for Enterprise and Industry (2008) The new SME definition: user guide and model declaration. Brussels: EC Publications Office.

European Commission, Directorate-General for Enterprise and Industry (2008) Communication from the Commission to the Council, the European Parliament, the European Economic and Social Committee and the Committee of the Regions: 'Think Small First': a 'Small Business Act' for Europe. Brussels: EC Publications Office.