Revised Altman Model for Bulgarian SMEs Viability Assessment

Author: Dimitar Haralampiev Popov

Abstract: Due to their contribution to employment and the economy, small and medium enterprises (SMEs) are an integral part of EUs policies and national programs. Predicting their viability is an important aspect of management decision making. That is why, a modified model is created, based on the Altman's Z-score, a multivariate classification technique, to help determine the viability of Bulgarian SMEs by using financial ratios from their income statements and balance sheets. Since the data spans over thousands of SMEs, implementing new approaches from information technology is needed. This is the reason to use the cross industry standard process for data mining (CRISP-DM), which is a methodological framework for researching large amounts of data. The results of combining Altman's Z-score with CRISP-DM in a single SME viability research show that using logistic regression and naive Bayes classifiers can distinguish between viable and non-viable SMEs with a moderate predictive power.

Keywords: Altman Z-score, CRISP-DM, SMEs, viability

JEL: M42

1. INTRODUCTION

This article explores research into the use of a modified model based on Altman's Z-score to predict SMEs viability. Altman's original research focuses on big publicly traded corporations. Over the years, new modified models emerged having different research objects (Altman, 2000) (Altman, et al., 1995), methodology variations (Clapham, 2001) (Terdpaopong, 2011) and for different states (Agarwal & Taffler, 2005) (Nandi & Choudhary, 2011) (Armeanu & Cioaca, 2014). Such a model can be used by managers and analysts as an early warning indicator whether the enterprise's viability is in distress (ALTMAN, September 1984). Viability is predicted by using financial ratios as independent variables that are implemented to train a classification model. The large amount of SMEs' data is an opportunity to implement data mining techniques as CRISP-DM into the methodology.

2. LITERATURE REVIEW

2.1. SMEs Definition and Importance

Small businesses usually have limited financial and human resources. The main criteria that define them are the number of employees, the annual turnover, the value of the assets, the ownership and the interconnection between the enterprises. The Bulgarian Small and Medium Enterprises Act determines the size of the company according to these indicators. The criteria is shown in Table 1.

Table 1 SMEs Criteria Listing (Anon., n.d.).

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 [1]			

Source: Own data

Due to their size, SMEs are numerous and comprise a significant number of all the enterprises in Bulgaria. According to 2019 NSI data for big enterprises i.e., over 250 employees, their nominal share is less than 0.2% of all the country's enterprises. SMEs are 98.5% which means the vast majority of enterprises are under the EU's definition for SMEs (Anon., 2020). Bulgarian SMEs create 65% of value added and 75% of the employment in the country, which is about 10 percentage points more than the EU's average (Anon., 2019).

2.2. Altman's Z-score

A substantial foundational research about financial ratios as predictors of viability is done by Beaver (1966 et al.). 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 (Beaver, 1966). Altman's Z-score, developed by Edward Altman, predicts enterprise's viability using financial ratios via multiple discriminant analysis. It attempts to predict whether a publicly traded company will remain viable within a two-year period. Future viability/bankruptcy is predicted reasonably well using a company's Z-score as an outcome from the model implementation. In the original model, the lower the value of Z, the more likely the company is to go 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 (Altman, 2000). To meet the need for exploring businesses with other characteristics, Altman has developed several revised models. Altman Z' serves as a valuation of private manufacturing companies (Altman, 2000). 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, December 26, 2005). Terdpaopong prefers logistic regression because of its flexibility in the assumptions and the types of data that can be analyzed to mitigate

[1] According to BSME, 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.

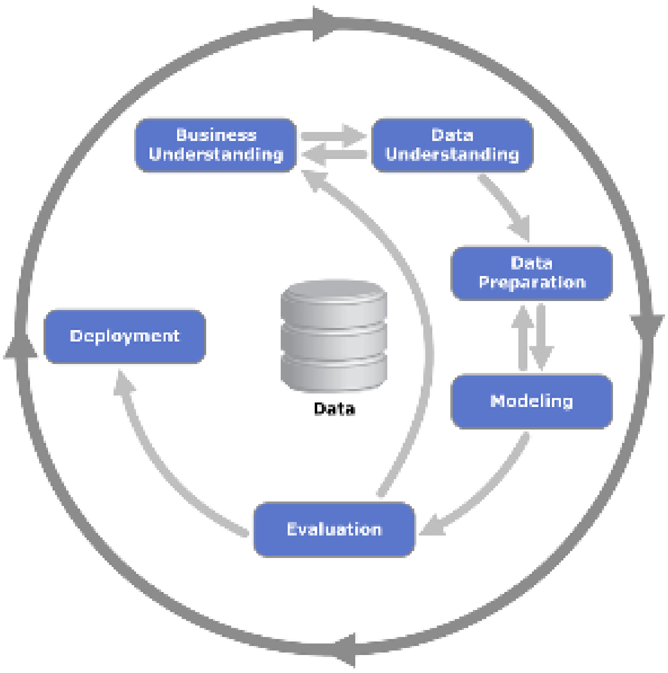
the lack of multivariate normality, value deviations, multicollinearity and singularity when developing a viability forecasting model (Terdpaopong, 2011). In order to avoid the turbulence of the global financial crisis, Lin samples companies before that period, which shows the importance of data temporal dimensions (Lin, 2015).

Altman's model shouldn't be taken as a universal instrument though. Different countries have their own specific economic characteristics. Differences in Bulgarian accounting legislation, although to some extent synchronized with European standards, the average size of companies surveyed and the conditions under which they find their viability jeopardized, must be taken into account.

3. RESEARCH METHODOLOGY

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 1 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.

Figure 1 Referent model CRISP-DM, phases. [1]



The methodology can be applied if we have to:

- Understand the business problem for SMEs. What specific goals are defined for the study?
- Select the available data for the issue. To describe and examine the available data.

[1] By Kenneth Jensen - Own work based on: ftp://public.dhe.ibm.com/software/analytics/spss/documentation/modeler/18.0/en/ModelerCRISPDM.pdf (Figure 1), CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=24930610

- Prepare the data, transforming it in a suitable for modeling format. Creating derivative and auxiliary variables.
- Create SMEs assessment model.
- 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 (Shearer, 2000).

4. RESULTS OVERVIEW

4.1. Empirical data

A sample compiled by the Bulgarian National Statistics Institute for SMEs with a total number of employees on an annual basis of not less than 10 and not more than 249, with an annual turnover not exceeding BGN 97.5 million for years 2014 through 2016. Balance of assets less than BGN 84 million and they are independent enterprises according to the Bulgarian SME Law. Micro-enterprises are excluded from the sample due to their high degree of individuality. They rely primarily on the personal abilities and skills of their creator / owner. The desired variables are from the Balance Sheet and Income Statement.

4.2. Data preparation

For usability, the business data are combined into a single dataset using their anonymous random individual code. The data arranged in this way can be used depending on the development of the enterprises in the period t to t + 2 for the purposes of the viability forecast. 85 data fields are received within three years. When preparing the data according to the literature discussed above, we need to create new fields derived from the already available data as financial ratios. Most of the data is complete and the missing values are imputed using A popular way to deal with this is to impute data. In our case, we will apply imputation through the R “mice” package, using random forest regression.

When performing analysis and modeling with multiple variables, it should be taken into account that the measured data scale is not always comparable between different measures. The purpose of normalization is to adjust the size and relative weight of individual variables. Given the enterprises’ scale difference, for example between 10 employees and 250 employees, it is quite possible to obtain a distortion of linear dependencies, despite the same unit of measurement. For the research purposes, data will be normalized according to the following formula:

$$n11 = (x/\mathrm{sqrt}(\textstyle\sum_{i=1}^{N} X_i^2)) \qquad (1)$$

where: n11 - quotient transformation

4.3. Multicollinearity check

In the presence of multicollinearity between the observed variables, the model variance coefficients increase, which causes a type II error. Variance inflation factor (VIF) will be used to perform the check, and by calculating the factor for each variable, the set of independent variables will be reduced to a set in which they are not collinear. Significant multicollinearity is observed at VIF values greater than 10. At a value between 5 and 10, moderate multicollinearity is observed. Using a lower value for VIF as a criterion will help reduce standard error. In this case, a value of 5 will be used. All features above 5 will be excluded from further use. Below we can see calculations and sorted values of all variables.

Table 2 Feature selection, variables excluded due to multicollinearity.

Variable	VIF value	Description
X2014_F15100_nm11_	531114.02	Net sales revenue
X2014_F10500_nm11_	464426.56	other expenses
X2014_LD.LDE_nm11_	222564.50	capitalization ratio
X2014_TL_nm11_	129275.90	total liabilities
X2014_F07002_nm11_	96653.74	long-term liabilities
X2014_TA_nm11_	66010.20	total assets
X2014_F07001_nm11_	49435.86	short-term liabilities
X2014_F05510_nm11_	5642.19	retained earnings
X2014_F05100_nm11_	4983.29	subscribed capital
X2014_F02300_nm11_	4891.32	long - term financial assets
X2014_WCEBITDA_nm11_	4406.98	working capital/ ebitda
X2014_F10200_nm11_	4089.15	cost of goods sold
X2014_F05400_nm11_	2882.31	reserves
X2014_F05520_nm11_	1729.08	retained loss
X2014_F03200_nm11_	1260.85	receivable
X2014_F05600_nm11_	1059.28	current profit/loss
X2014_F08000_nm11_	822.34	deferred revenue
X2014_F02200_nm11_	654.77	long-term material assets
X2014_F03100_nm11_	494.52	inventories
X2014_NT_nm11_	476.20	net sales revenue/average short-term assets
X2014_EBIT_nm11_	416.85	earnings before interest and taxes
X2014_DTA_nm11_	328.32	debt/total assets
X2014_F10300_nm11_	238.63	staff costs
X2014_F16300_nm11_	166.98	interest and other financial income
X2014_F15400_nm11_	152.01	other income
X2014_F11200_nm11_	145.88	interest and other financial income
X2014_F10400_nm11_	95.66	depreciation and impairment
X2014_lg_nm11_	94.83	short-term liabilities/short-term assets
X2014_F03400_nm11_	93.82	cash
X2014_F15200_nm11_	61.77	increase in work in process inventories
X2014_F05300_nm11_	51.73	revaluation reserve
X2014_F16100_nm11_	43.69	equity participation
X2014_F10100_nm11_	35.41	decrease in work in process inventories
X2014_F11100_nm11_	35.28	impairment loss in financial assets
X2014_CFD_nm11_	34.47	cash flow/debt
X2014_DTA.TAY_nm11_	32.94	depreciation and impairment costs/fixed assets
X2014_F06000_nm11_	26.75	provisions and similar liabilities
X2014_ITA_nm11_	21.44	intangible/total assets
X2014_CAPAS_nm11_	20.21	net income/sum of real assets
X2014_STDE_nm11_	18.71	short term debt/equity
X2014_F03300_nm11_	17.32	investments
X2014_F02100_nm11_	16.91	intangible
X2014_F02400_nm11_	13.41	deferred tax

Variable	VIF value	Description
X2014_TR_nm11_	12.10	total revenue
X2014_F17000_nm11_	8.32	extraordinary income
X2014_EBITDATA_nm11_	6.50	ebitda/total assets
X2014_CNS_nm11_	6.33	cash/net sales
X2014_F04000_nm11_	5.75	prepaid expenses
X2014_TOD_nm11_	5.73	duration of one turnover in days
X2014_F15300_nm11_	5.58	asset acquisition cost
X2014_LEV_nm11_	5.20	market value of equity/book value of total liabilities

4.4. Creating a training and test sample

It is necessary to divide the set of training and control sample. The ratio is 70/30 using the "createDataPartition" function from the R "caret" package. When specifying the dependent variable, the random sampling is performed by preserving the distribution of the variable in the data[1]. The number of non-viable enterprises reported in the dependent variable represents approximately 19% of the data in the population, test and training sample:

Table 3 Train and test sample summary:

	Percent „not_survive“	Total records
Population	19.7%	5839
Train sample	19.9%	4088
Test sample	19.1%	1751

Source: Own data

4.5. Model results

For predicting SME viability, the samples use dependent variable “not_survive” with Boolean values “1/0” where “0” is viability indication. Four popular classification models are to be trained – decision tree, random forest, logistic regression, naive Bayes. The models’ performance will be evaluated and their prediction results compared.

Evaluation of models according to ROC analysis

The abbreviation ROC comes from communication theory and means "operational characteristics of the receiver". The reason for using it is that in this way type I error and type II error are summarized in all possible indicators. The overall performance of the model is determined by the area under the ROC curve, for short AUC. The area under the curve should be as large as possible up to 1 and the curve should reach the upper left corner of the graph. As the AUC includes all possible indicators, it becomes an ideal indicator for comparison between different classification models.

Table 4 Classification Models Results.

Model	AUC	CA	F1	Precision	Recall
Tree	0.647	0.612	0.580	0.632	0.535
Random Forest	0.679	0.616	0.514	0.701	0.406
Naive Bayes	0.686	0.635	0.591	0.671	0.529
Logistic Regression	0.669	0.629	0.581	0.667	0.515

[1] http://topepo.github.io/caret/data-splitting.html

Model coefficients and variables for logistic regression are shown below:

Table 5 Variables coefficients for model:

Variable	Coefficient	Description
intercept	-0.06574	
X2014_RETA_nm11_	-24.2677	retained earnings / total assets.
X2014_V16140_nm11_	-16.6311	number of employees
X2014_ROI_nm11_	-1.07808	return on investment
X2014_NITA_nm11_	-27.1252	net income/total assets
X2014_F01000_nm11_	-7.40116	unpaid capital
X2014_ROE_nm11_	-4.15173	return on equity
X2014_ROS_nm11_	-3.15845	return on sales
X2014_EBITS_nm11_	37.3203	ebit/sales
X2014_COGSSALES_nm11_	13.4252	cogs/sales
X2014_TATL_nm11_	6.27271	total assets/total liabilities
X2014_DEB_nm11_	3.28092	liabilities/assets
X2014_ARL_nm11_	4.00216	account recievable/liabilities
X2014_TC_nm11_	8.36628	total cost
X2014_EBITDA_nm11_	5.42174	ebit, depreciation, and amortization
X2014_IOL_nm11_	-6.57876	investments/short-term liabilities
X2014_WCTA_nm11_	-48.8921	working capital/total assets

5. CONCLUSION

The results table shows that the models are performing differently depending on the emphasized indicators. The decision tree as a whole performs the weakest in almost all indicators, lagging behind the others in accuracy, precision and AUC. The only indicator where it is superior to others is the classification of true positives by a few percentage points. Although this is an important metric, this classifier shows the worst results compared to other metrics. The random forest shows the second lowest accuracy of the models despite the high AUC score. Another significant drawback is the lowest recall of all the models and the low percentage of true positives class. The lack of balance between indicators also plays a significant role in the model's poor performance. The logistic regression and naive Bayesian classifier have the best performance with balanced high result accuracy and can be used successfully in predicting SME viability.

References

1. Agarwal, V. & Taffler, R. J., 2005. *Twenty-five years of z-scores in the UK: do they really work?.* s.l.:Cranfield School of Management.

2. Altman, E., 2000. *PREDICTING FINANCIAL DISTRESS OF COMPANIES REVISITING THE Z-SCORE AND ZETA®MODELS.* New York: Journal of Banking & Finance.

3. Altman, E., Hartzell, J. & Peck, M., 1995. *Emerging Markets Corporate Bonds: A Scoring System.* New York: Salomon Brothers Inc.

4. ALTMAN, E. I., September 1984. *A Further Empirical Investigation of the Bankruptcy Cost Question.* s.l.:The Journal of the American Finance Association.

5. Altman, E. & Sabato, G., December 26, 2005. *Modeling Credit Risk for Smes: Evidence from the Us Market.* s.l.: SSRN, p.24.

6. Anon., 2019. *Republic of Bulgaria, Ministry of Economy and Industry, Current data on SMEs, https://www.mi.government.bg/bg/themes/aktualni-danni-za-msp-2103-285.html.* Sofia: s.n.
7. Anon., 2020. *National Statistical Institute.* Sofia, Bulgaria: s.n.
8. Anon., n.d. *Bulgarian Small and Medium Enterprises Act.* s.l.:s.n.
9. Armeanu, D. A. & Cioaca, S. I., 2014. *An assessment of the bankruptcy risk on the Romanian capital market.* s.l.:4th WORLD CONFERENCE ON EDUCATIONAL TECHNOLOGY RESEARCHES.
10. Beaver, W., 1966. *Financial Ratios As Predictors of Failure.* s.l.:Journal of Accounting Research Vol. 4 Empirical Research in Accounting: Selected Studies.
11. Clapham, P., 2001. *Cash is king [finance for SMEs].* s.l.:Engineering Management Journal, Volume 11, Issue 4.
12. Lin, H., 2015. *Default Prediction Model for SME's: Evidence from UK Market Using Financial Ratios.* s.l.:International Journal of Business and Management; Vol. 10, No. 2, p.102.
13. Nandi, J. K. & Choudhary, N. K., 2011. *Credit Risk Management of Loan Portfolios by Indian Banks: Some Empirical Evidence.* s.l.:IUP J. Bank Management.
14. Shearer, C., 2000. *The CRISP-DM model: the new blueprint for data mining.* s.l.:J Data Warehousing, Volume 5, Number 4, p.13.
15. Terdpaopong, K., 2011. *Identifying an SME's debt crisis potential by using logistic regression analysis.* s.l.:RANGSIT JOURNAL OF ARTS AND SCIENCES (RJAS) Volume 1, Number 1, 19.