

Development of Bankruptcy Prediction Model for Latvian Companies

Arnis Stasko^{1*}, Ilze Birzniece¹, and Girts Kebers²

¹ Riga Technical University, Riga, Latvia

² Lursoft, Riga, Latvia

{arnis.stasko, ilze.birzniece}@rtu.lv;
kebers@lursoft.lv

Abstract. This article addresses the financial performance prediction for Latvian companies. It is of critical importance to be able to provide timely warnings to management, investors, employees, stakeholders and other interested parties who wish to reduce their losses. There are literature review structures that previously made research into company performance prediction. Estimating the risk of bankruptcy of Latvian companies has been carried out by applying two commonly used approaches: Altman's Z-score estimation and an experience-based machine learning approach using C4.5 Decision Tree. The results show that Altman's Z-score method predicts bankruptcy for a massive number of companies, while the ML method predicts bankruptcy for only a few. Each of these approaches has its drawbacks. We propose an extended company performance prediction model that considers other factors that influence distress risk, e.g., changes in regulation and other environmental factors. Expert opinion is of great value in estimating a company's future performance; therefore, an automated solution supporting experts in their decision-making is presented.

Keywords: Insolvency, Bankruptcy, Prediction, Altman's Z-score, Machine Learning.

1 Introduction

The Directive of The European Parliament and the Council on preventive restructuring frameworks on the discharge of debt and disqualifications (further on – Directive) for each member state requires implementing the company warning system. Estimating the risk of bankruptcy of a specific company substantiates economic impact to its owners, investors and stakeholders [1], [2]. Such an alert system would allow these companies to make timely adjustments to their operations in order to survive or show better financial results. Two main components of developing an early warning system are (1) selection of financial ratios and (2) creation of classifier design [3].

Available warning systems are based on analyzing the financial ratios most commonly using the

* Corresponding author

© 2021 Arnis Stasko, Ilze Birzniece, and Girts Kebers. This is an open access article licensed under the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>).

Reference: A. Stasko, I. Birzniece, and G. Kebers, "Development of Bankruptcy Prediction Model for Latvian Companies," Complex Systems Informatics and Modeling Quarterly, CSIMQ, no. 27, pp. 45–59, 2021. Available: <https://doi.org/10.7250/csimq.2021-27.02>

Additional information. Author's ORCID iD: I. Birzniece – <https://orcid.org/0000-0002-5775-6138>. PII S225599222100155X. Received: 7 June 2021. Accepted: 19 July 2021. Available online: 30 July 2021.

Altman Z-score algorithm. There is an assumption underlying bankruptcy prediction that leading macro-economic indicators (e.g., inflation, interest rates) and company characteristics are represented in financial reports [1]. Some solutions learn to predict insolvency using machine learning methods from the financial data of insolvent or liquidated companies.

According to the Directive, a warning system should be implemented in Latvia. The aim of this research is to introduce a performance prediction concept for Latvian companies which may serve as a core for developing a full-scale warning system. We start by implementing two contrasting prediction systems based on (1) Altman's Z-score and (2) the machine learning approach. We compare the results of the two methods and conclude on their performance on Latvian companies' data. To improve the results, the concept of the extended method is presented. We hypothesize that global or country-specific events should be taken into account in addition to financial data from annual reports. More precisely, we offer to predict insolvency or liquidation of companies by analyzing the country's legislation changes.

Our proposed approach is based on the idea that, in addition to financial data, it is possible to add historical information about significant changes in legislation in a given year to learn from and report on which industries and which companies in those industries are most affected by such changes. The possibility that a similar change in the law will be repeated in the country after several years is unlikely. Therefore, the opportunities to learn from our experience are minimal. At the same time, the chance to learn from a legislation change in another country is very likely. If it is planned to adopt a law or tax change in one country that has already been similarly adopted in the past in another country, it is possible to learn from the experience of that other country.

The article is structured as follows. Section 2 discusses related work. Section 3 demonstrates the application of traditional approaches to bankruptcy prediction for Latvian companies. This includes Altman's Z-score and a machine learning approach with the decision tree classification solution. After analyzing the results, we introduce the extended approach for company performance prediction in Section 4. Finally, conclusions and future work are discussed in Section 5.

2 Related Work

Techniques for the prediction of bankruptcy or insolvency have been researched both by academics and practitioners. For instance, structural models are applied that use an explicit function based on a theory of companies and insolvency [4] and can be associated with traditional statistical techniques. In contrast, data-driven empirical models (machine learning models) are built and assessed using predictive performance as the criterion. Machine learning (ML) is the ability of a computer program to improve its own performance, based on past experience [5]. Thus techniques for the prediction of bankruptcy can be characterized in two main groups [6], [7]:

- Traditional statistics techniques, e.g., Altman's Z-score (a variation of the traditional z-score), Discriminant Analysis, Logistic Regression, Generalized Linear Models;
- Machine learning (ML) models, e.g., Support Vector Machines (SVM), Bagging, Boosting, and Random Forest (RF), Artificial Neural Networks (ANN), Decision Trees (DT), k-Nearest Neighbour (kNN), Ensemble Techniques, Rough Sets, Evolutionary Programming.

2.1 Literature Review

The literature review is performed to structure available research into company performance prediction that uses different data sets and applies machine learning techniques. We included research works that were addressing *bankruptcy*, *insolvency* and *financial distress* and these were retrieved through ACM digital library from the last 15 years. From the retrieved search results we selected those papers which described a clear application of company performance prediction by

applying machine learning techniques. This literature analysis aimed to identify the following aspects for each selected research paper:

- The goal of the research, including the application scope;
- Characteristics of factors used for analysis;
- Analytical approaches, including data set characteristics;
- Machine learning techniques that were applied, including evaluation metrics.

The summary of the literature review is given in Table 1.

Table 1. Literature analysis on bankruptcy prediction

Res. paper	The goal of analysis, scope	Factors used	Analytical approaches	ML techniques, metrics
[7]	Improvement in prediction accuracy using machine learning techniques using Altman's Z-score variables and six complimentary financial indicators. <i>Applied to USA companies.</i>	Altman's ratios extended by operating margin, change in return-on-equity, change in price-to-book, and growth measures related to assets, sales, and the number of employees (as suggested by Carton and Hofer in 2006).	Balanced data set, consecutive years. Training from years 1985–2005, testing for the years 2006–2014.	SVM, Bagging, Boosting, RF. <i>Metrics: ROC, AUC, Accuracy</i>
[6]	Bankruptcy prediction using Extreme Gradient Boosting for learning an ensemble of decision trees. Introduction of synthetic features, where each synthetic feature can be seen as a single regression model. <i>Applied to Polish companies.</i>	Synthetic features – generated by random selection of two existing features (64 economic features) and random selection of arithmetical operations to be performed on them.	Imbalanced data set, consisting of companies five years before bankruptcy in the period of 2007–2013 and 2000–2012 for still operating companies. Five trials to the bankruptcy prediction in the 1st, 2nd, 3rd, 4th, and 5th year.	Boosted Trees trained with the Extreme Gradient Boosting method. <i>Metrics: AUC</i>
[8]	Bankruptcy prediction using SVM and RF. <i>Applied to Polish companies.</i>	64 economic features	Comparison of balanced and imbalanced data set, showing better results for balanced. Five years before bankruptcy in the period of 2007–2013 and 2000–2012 for still operating companies. Five trials to the bankruptcy prediction in the 1st, 2nd, 3rd, 4th, and 5th year. Randomly split into train and test sets.	SVM and RF <i>Metrics: Error Rates.</i>
[4]	Using standard credit scoring methodology on a new data source, demonstrate that insolvency risk can be estimated using income tax return data. <i>Applied to Australian companies.</i>	Altman's ratios and eight financial variables as defined by Ohlson in 1980, including Financial Distress Indicator Variables, Age of the company and industry classification using ABS industry codes.	Balanced data set. The corporate insolvency risk is defined as the probability that a company will become insolvent in the next year. Years up to 2005 are used for training to predict the year 2006, testing with predicting the year 2007.	Logistic Regression, Ada Boost, and RF <i>Metrics: AUC</i>

Table 1. Continued

Res. paper	The goal of analysis, scope	Factors used	Analytical approaches	ML techniques, metrics
[9]	Bankruptcy prediction using genetic algorithm-based two-step classification model that can be adapted to new situations, both in terms of the model architecture and the features used. <i>Applied to Russian companies.</i>	55 features (financial ratios and macro/micro business environment factors).	Balanced data set, normalized, split into train and test. 912 observations. Extend financial factors with environmental ones. Evaluation of feature importance. Comparison with traditional bankruptcy prediction models.	Hybrid classifier system, based on ensembles of kNN, Logistic Regression, SVM, Naive Bayes, DT. <i>Metrics: Accuracy, Precision, Recall</i>
[3]	Analyze financial ratio selection in bankruptcy prediction, compare different predictive methods. <i>Applied to Taiwan companies (can be implied, not stated directly).</i>	64 features, including nonfinancial ratio Taiwan Corporate Credit Risk Index	600 companies obtained from the financial statements of the Taiwan Economic Journal database. The ratio of bankrupt to nonbankrupt companies used 1:3.	The hybrid approach, Genetic Algorithm, Fuzzy Clustering. <i>Metrics: Precision, Error Rates</i>

2.2 Discussion

The results indicate that Altman’s Z-Score ratios are often included in the factors to predict a company’s performance. However, they are mainly extended by other financial factors. A limited number of studies investigated non-financial aspects to predict a company’s insolvency. We propose to consider changes in legislation as one of the characteristics describing the environment. Another inference from the research examined is the localization of implemented systems or experiments. None of the studies covers more than one national market, which led us to conclude that predicting company wellbeing is prone to individual factors dependent on national peculiarities. To the best of our knowledge, no research for the Baltic States or Latvia, in particular, was found. The variety of applied ML methods is wide. None of them dominates. However, SVM is among the most common ones.

Evaluation metrics provide a systematic way of evaluating different methods and settings. For classification problems, it is natural to measure a classifier’s performance in terms of the error rate [10]. However, different metrics exist and it is a typical challenge for machine learning researchers to compare the performance of different ML solutions objectively. As can be seen from the comparison of existing applications in Table 1, varied metrics have been used to evaluate ML algorithms in bankruptcy and insolvency prediction. All of them (Accuracy, Precision, Recall, ROC, AUC, Type I and Type II Error Rates) fall under classical ML evaluation techniques and are based on a confusion matrix.

Having other researchers’ experiences in mind, we define the following scenario to introduce performance prediction for Latvian companies. First, we apply the traditional approach of evaluating Altman’s Z-score to create a baseline for prediction performance in the Latvian case. Second, we develop an ML-based solution and compare it with Altman’s Z-score. Third, we conclude on results and examine potential ways of improvement.

3 Predicting Bankruptcy for Latvian Companies

In this section, we predict bankruptcy using Altman’s Z-score and an ML-based solution. Then we compare the results between the two methods.

3.1 Altman's Z-score

The most discussed method to predict insolvency/bankruptcy is Altman's Z-score. Altman's Z-score calculation formula has three common variations [11], [12], [13].

For public enterprises:

$$\text{Altman Z-Score} = 1.2 \times A + 1.4 \times B + 3.3 \times C + 0.6 \times D + 1.0 \times E$$

(bankruptcy threshold = 1.81, safety threshold = 2.99)

For private companies:

$$\text{Altman Z-Score} = 0.717 \times A + 0.847 \times B + 3.107 \times C + 0.420 \times D' + 0.998 \times E$$

(bankruptcy threshold = 1.23, safety threshold = 2.99)

For private (non-manufacturing) companies:

$$\text{Altman Z-Score} = 6.56 \times A + 3.26 \times B + 6.72 \times C + 1.05 \times D'$$

(bankruptcy threshold = 1.1, safety threshold = 2.6)

Where:

- A = working capital / total assets;
- B = retained earnings / total assets;
- C = earnings before interest and tax / total assets;
- D = market value of equity / total liabilities;
- D' = book value of shares / total liabilities;
- E = sales / total assets.

A score above the safety threshold predicts that a company has an insignificant probability of facing bankruptcy (Safe). Having a score between the bankruptcy threshold and safety threshold, a company has a moderate chance of bankruptcy (Grey). At the same time, a Z-score below the bankruptcy threshold predicts that a company has a very high probability of bankruptcy (Bankruptcy).

We calculated Altman's Z-score for each company in Latvia based on the annual reports from 2010 till 2014 (Table 2). According to the 2010 reports, the table shows that the Z-score predicted a safe future for 18 387 companies, a moderate chance of bankruptcy for 8 610 companies, and a very high probability of bankruptcy for 50 502 companies. Similarly, predictions from the data for 2011–2014 annual reports are presented.

Table 2. Bankruptcy prediction based on Altman's Z-score

Prediction	2010	2011	2012	2013	2014
Safe	18 387	22 285	27 306	30 341	32 993
Grey *	8 610	13 102	17 406	22 253	27 772
Bankruptcy	50 502	51 466	55 136	58 823	61 173
Total	77 499	86 853	99 848	111 417	121 938

* Included companies which Altman's Z-score has not been calculated due to missing data

To assess the Z-score predictions, we examined whether the companies were the subject of insolvency proceedings or liquidation in the next five years (Table 3). We combined the Safe and Gray zones to compare the results with the machine learning approach discussed in the next section. The Safe / Gray zone companies are considered to be predicted "Correct" if they were not the subject of insolvency proceedings or liquidation in the next five years after the analyzed year, otherwise "Incorrect". The companies in the Bankruptcy zone are considered to be predicted "Correct" if they were the subject of insolvency proceedings or liquidation in the next five years after the analyzed year, otherwise "Incorrect".

Table 3. Evaluation of Bankruptcy prediction based on Altman's Z-score

Year	Prediction	Companies	Correct	Incorrect	Precision	Recall	F-Measure
2010	Safe/Grey	26 997	24 139	2 858	89.4%	36.5%	51.9%
	Bankruptcy	50 502	8 577	41 925	17.0%	75.0%	27.7%
	Total/Avg	77 499	32 716	44 783	78.7%	42.2%	48.3%
2011	Safe/Grey	35 387	29 950	5 437	84.6%	41.6%	55.7%
	Bankruptcy	51 466	9 341	42 125	18.1%	63.2%	28.2%
	Total/Avg	86 853	39 291	47 562	73.3%	45.2%	51.1%
2012	Safe/Grey	44 712	35 066	9 646	78.4%	44.2%	56.5%
	Bankruptcy	55 136	10 849	44 287	19.7%	52.9%	28.7%
	Total/Avg	99 848	45 915	53 933	66.4%	46.0%	50.8%
2013	Safe/Grey	52 594	37 064	15 530	70.5%	44.5%	54.5%
	Bankruptcy	58 823	12 552	46 271	21.3%	44.7%	28.9%
	Total/Avg	111 417	49 616	61 801	58.1%	44.5%	48.1%
2014	Safe/Grey	60 765	37 329	23 436	61.4%	44.3%	51.5%
	Bankruptcy	61 173	14 243	46 930	23.3%	37.8%	28.8%
	Total/Avg	121 938	51 572	70 366	49.6%	42.3%	44.5%

The results show that Altman's Z-score predicts bankruptcy for a significantly larger number of companies than actually will be the case in the next five years. Looking at individual companies, we see that in Latvia, at a low Z-score value, many companies are not liquidated but continue to exist, even with negative financial indicators, over a long period.

From the perspective of evaluating potential business partners, Z-score appears to be very skeptical or pessimistic – possibly resulting in a decision not to cooperate with a large number of companies because they are unlikely to be able to cover their liabilities and may go bankrupt. Guided only by such criteria, it would be challenging to build a business.

3.2 Machine Learning Approach

As a machine learning approach (further – ML-approach), we use the decision tree learning method, specifically the C4.5 algorithm. It is one of the most popular classification algorithms [14], creating the human-interpretable knowledge representation form – a decision tree. It is widely used for many real-life tasks. To distinguish from other machine learning paradigms, e.g., artificial neural networks, support vector machines, k-nearest neighbors, which lack explanatory power, inductive learning methods in the form of decision trees are highly regarded due to their interpretability [15]. In order to build a C4.5 decision tree from training data, we applied its J48 implementation by a Weka tool [16] and used it for prediction tasks.

Five training sets are presented – one for each year (2010 – 2014). Each training set is based on a single year's financial ratios from the annual reports. The following attributes characterize each sample or company in the training set:

- Industry;
- Company age;
- 109 attributes from balance sheets of the company's annual reports (according to Legal Act of The Republic of Latvia [17]);
- 62 attributes from profit-loss statements of the company's annual reports (according to Legal Act of The Republic of Latvia [17]).

Training set instances are divided into the following classes:

- Safe/Grey – companies have not been liquidated and have not started insolvency procedures five years from the analyzed year;

- Bankruptcy - companies have been liquidated or have started insolvency procedures five years from the analyzed year.

The number and distribution of examples of training sets in the classes are given in Table 4. The table shows that in the #1 training set, 66 735 companies have not been liquidated and did not start insolvency procedures during 2011–2015; 10 764 companies have been liquidated or have started insolvency procedures during 2011–2015. Similarly, other training sets are constructed.

Table 4. Training sets for Bankruptcy prediction using ML-approach

Training set:	#1	#2	#3	#4	#5
Sample classification	2010	2011	2012	2013	2014
Safe/Grey	66 735	72 590	79 822	83 773	84 689
Bankruptcy	10 764	14 263	20 026	27 644	37 249
Total	77 499	86 853	99 848	111 417	121 938

For each training set described in Table 4, using the Weka J48 algorithm, a decision tree was built. Then, all training examples were classified using the appropriate decision tree for each training set. Classification distribution is given in Table 5. To assess the ML-approach predictions, we compared classification results against actual results – whether the companies were the subject of insolvency proceedings or liquidation in the next five years (Table 6).

Table 5. Bankruptcy prediction based on ML-approach

Training set:	#1	#2	#3	#4	#5
Prediction	2010	2011	2012	2013	2014
Safe/Grey	74 455	84 550	96 921	108 460	120 627
Bankruptcy	3 044	2 303	2 927	2 957	1 311
Total	77 499	86 853	99 848	111 417	121 938

The Safe / Gray zone companies are considered to be predicted “Correct” if they were not the subject of insolvency proceedings or liquidation in the next five years from the analyzed year, otherwise prediction is “Incorrect”. The companies in the Bankruptcy zone are considered to be predicted “Correct” if they were the subject of insolvency proceedings or liquidation in the next five years from the analyzed year, otherwise – “Incorrect”.

Table 6. Evaluation of Bankruptcy prediction based on ML-approach

Year	Prediction	Count	Correct	Incorrect	Precision	Recall	F-Measure
2010	Safe/Grey	74 455	66 058	8 397	88.7%	99.0%	93.6%
	Bankruptcy	3 044	2 367	677	77.8%	22.0%	34.3%
	Total/Avg	77 499	68 425	9 074	87.2%	88.3%	85.3%
2011	Safe/Grey	84 550	72 134	12 416	85.3%	99.4%	91.8%
	Bankruptcy	2 303	1 847	456	80.2%	12.9%	22.3%
	Total/Avg	86 853	73 981	12 872	84.5%	85.2%	80.4%
2012	Safe/Grey	96 921	79 063	17 858	81.6%	99.0%	89.5%
	Bankruptcy	2 927	2 168	759	74.1%	10.8%	18.9%
	Total/Avg	99 848	81 231	18 617	80.1%	81.4%	75.3%
2013	Safe/Grey	108 460	83 043	25 417	76.6%	99.1%	86.4%
	Bankruptcy	2 957	2 227	730	75.3%	8.1%	14.6%
	Total/Avg	111 417	85 270	26 147	76.3%	76.5%	68.6%
2014	Safe/Grey	120 627	84 389	36 238	70.0%	99.6%	82.2%
	Bankruptcy	1 311	1 011	300	77.1%	2.7%	5.2%
	Total/Avg	121 938	85 400	36 538	72.1%	70.0%	58.7%

The results show that the described ML-approach predicts bankruptcy for a significantly smaller number of companies than actually occurs in the next five years. Examining the examples of companies individually, we see that in many cases the training set contains samples with identical financial ratios. In contrast, only a few of them have been liquidated or have started insolvency procedures. As a result, the ML-approach follows the majority opinion, and many examples of liquidation are not taken into account in the training set.

From the perspective of evaluating potential business partners, the ML-approach offers to be very optimistic – likely resulting in a decision to cooperate with almost every company, except just a few, because almost all are likely to cover their liabilities. Guided only by such criteria, it would be risky to build a business. Cooperation with a potentially bankrupt company is presumed to result in losses.

3.3 Summary

The methods considered give very different results. If we compare the methods technically using precision, recall, and F-measure ratios, the ML-approach overscores the Z-score. The precision measures the percentage of cases in which the classification algorithm correctly predicts the result. It is compared in Table 7 and Figure 1. The ML-approach, with an average precision of 80%, outperforms Z-score, with an average precision of 65%, by 15%. In other words, if the ML-approach predicts the future for 100 companies, it is correct in 85 cases. In comparison, Z-score is correct only for 65 companies.

Table 7. Precision: Altman’s Z-score vs ML-approach

Method	2010	2011	2012	2013	2014	Avg
Altman’s Z-score	78.7%	73.3%	66.4%	58.1%	49.6%	65%
ML-approach	87.2%	84.5%	80.1%	76.3%	72.1%	80%

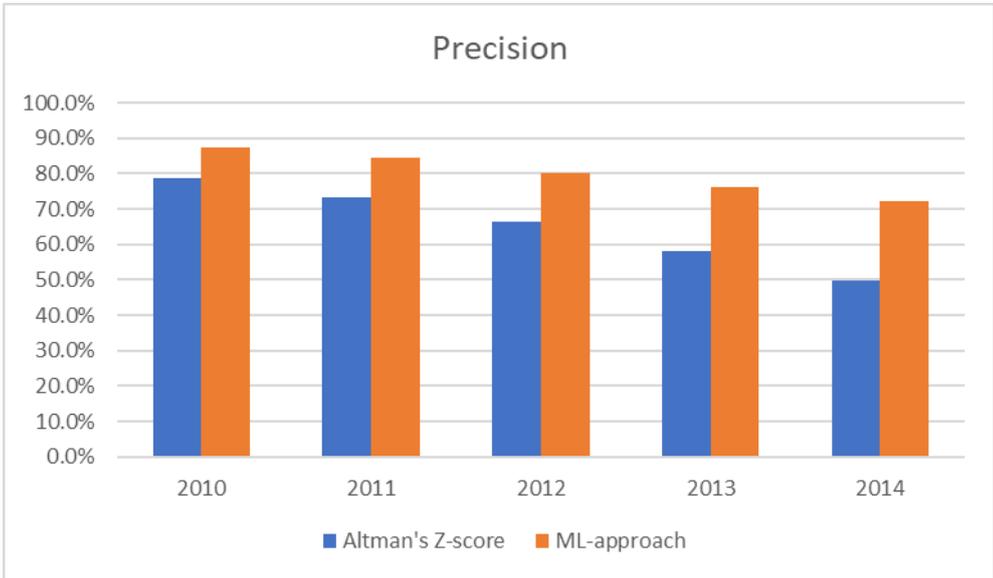


Figure 1. Precision: Altman’s Z-score vs ML-approach

The recall is compared in Table 8 and Figure 2. The ML approach, with an average recall of 80%, outperforms Z-score, with an average recall of 44%, by 36%. Here the recall shows the average percentage of bunkruptcy cases that the method misses. Altman’s Z-score predicts bunkruptcy for many companies, so it misses the actual Safe/Grey class. In contrast, the ML-approach misses many bunkruptcy cases predicting a safe future for them.

Table 8. Recall: Altman's Z-score VS ML-approach

Method	2010	2011	2012	2013	2014	Avg
Altman's Z-score	42.2%	45.2%	46.0%	44.5%	42.3%	44%
ML-approach	88.3%	85.2%	81.4%	76.5%	70.0%	80%

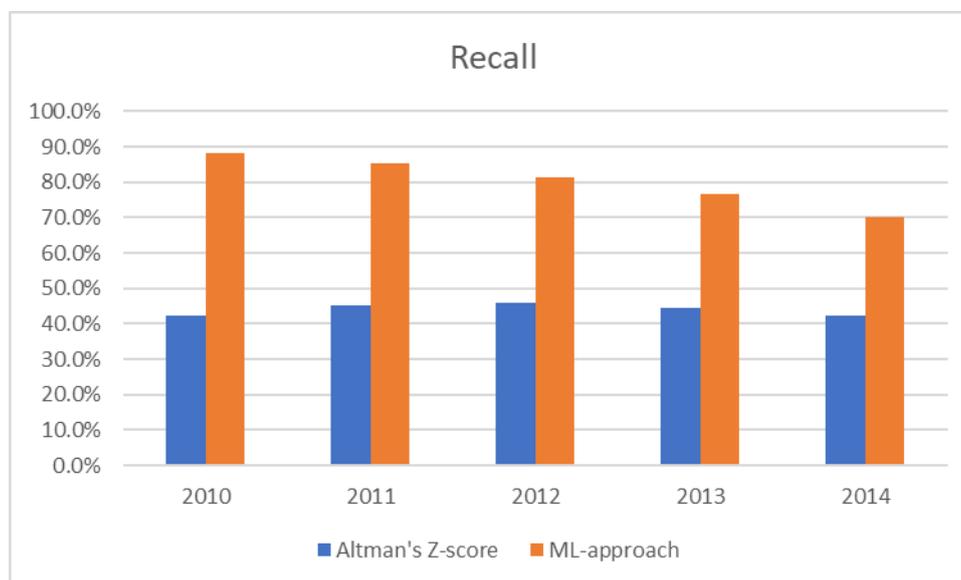


Figure 2. Recall: Altman's Z-score VS ML-approach

The F-measure shows the harmonic mean of precision and recall. Thus, F-Measure shows the overall performance of each method. It is compared in Table 9 and Figure 3. The ML approach, with an average F-Measure of 74%, outperforms Z-score, with an average F-Measure of 49%, by 25%.

Table 9. F-Measure: Altman's Z-score VS ML-approach

Method	2010	2011	2012	2013	2014	Avg
Altman's Z-score	48.3%	51.1%	50.8%	48.1%	44.5%	49%
ML-approach	85.3%	80.4%	75.3%	68.6%	58.7%	74%

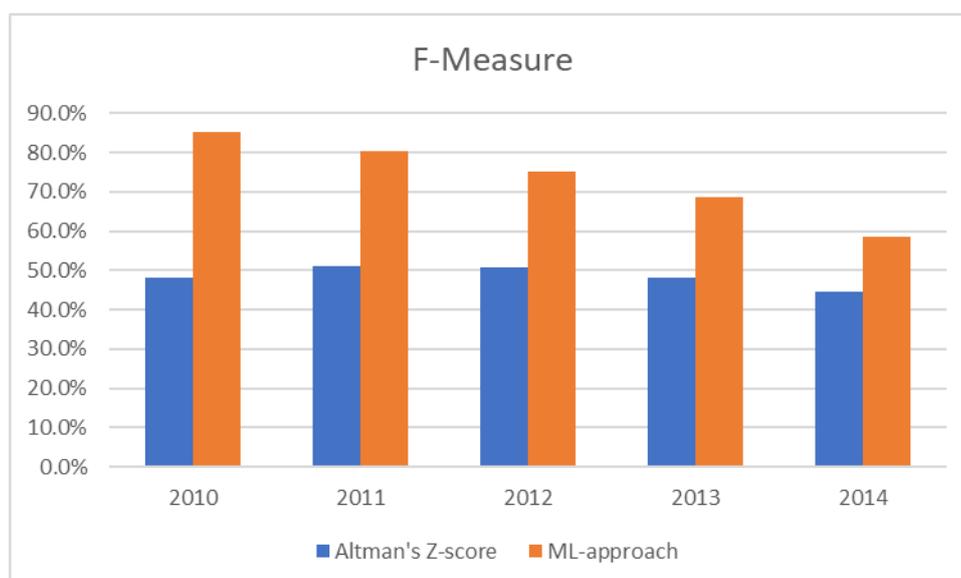


Figure 3. F-Measure: Altman's Z-score VS ML-approach

While the Z-score method predicts bankruptcy for a massive number of companies, the ML method predicts bankruptcy for only a few. If we look more closely at both methods' capabilities of capturing bankruptcy cases, the Altman's Z-score's bankruptcy prediction precision average is 19.9%, and the ML-approaches recall average is 11.3%. In other words, Altman's Z-score prediction of bankruptcy is coming true only for 2 out of 10 companies, while ML-approach misses 9 out of 10 bankruptcy cases. Each of the approaches has its drawbacks. If knowledge extraction from the data does not perform well enough with given methods, we can infer that the prediction model is unable to capture the relationships between previous and future performance. A search for other factors affecting the solvency of companies is thus required.

4 Towards Extended Approach for Company Performance Prediction

To improve the results of company financial performance prediction we analyze the potential of including more refined attributes to characterize the company's inner and outer factors.

4.1 Inclusion of Non-Financial Factors

Basically, in classical models, only financial factors are used for predicting a company's insolvency or bankruptcy. However, there are some attempts to include other factors. Already in [18] it is mentioned that essential qualitative information can be extracted from annual reports and other text-based documents. In [19] the use of qualitative information from the annual reports to forecast a company's operating performance is applied. For more general corporate financial forecasts, formalized non-financial indicators such as the company's size and corporate governance are also included [20]. In [9] indicators that reflect the economic situation and the policy of the Central Bank, and factors that describe the firm's non-financial characteristics are added: the presence of government control, the presence of economic sanctions, market share, the inclusion of the firm in the public list of unreliable suppliers, etc. Authors of [9] conclude that the financial indicators are the most important factors for predicting bankruptcy. However, out of the 10 top significant features, according to predictive models built, six were environmental factors. They mention that the research was carried out with Russian companies and might be relevant especially to developing economies [9].

The most commonly used method to predict insolvency is Altman's Z-score [12]. Multiple variations exist of what coefficients to apply to calculate Altman's Z-score. In the meantime, none of them has proved to be perfect. Our solution assumes that the essence of the problem lies elsewhere. Namely, particular global events affect companies in different countries every year. Global events can take the form of changes in national legislation, changes in international relations, natural disasters, global crises and more.

Thus, we put forward the assumption that other factors influence insolvency risk as well, e.g., changes in legislation and other environmental factors.

4.2 Development of the Extended Prediction Model

Knowledge of bankruptcy or insolvency risks can be obtained from historical data. Given that the experience of one company does not provide convincing arguments for the success of other companies in the industry, such knowledge should be sought in the overall indicators of industries. If financial data of companies for several years are available, then it is possible to calculate the average financial indicators of the sectors and visualize the changes of these financial indicators in graphs. Looking at the graphs, it is possible for the expert to find the years in which one of the industries suffered significant losses, as a result of which several companies were financially distressed and went bankrupt in subsequent years.

Such events, when the industry is facing difficulties, are related to external factors – *force majeure* events, changes in international relations, changes in national legislation and other global

events. The high complexity is introduced by the fact that several influential events can occur simultaneously in one year, each of which worsens or improves the situation of the industry with a certain weight. It is hard to handle this complexity, because the information is distributed over several sources, and it is difficult to model such effects and explain them precisely. At the same time, if the expert has identified the deterioration of the industry in one of the years and is convinced that this deterioration is largely due to a specific change in legislation, then these local cases can be analyzed with machine learning methods to assess all companies in that industry. Most likely, in this situation, some companies will experience insolvency, while others will survive. For the machine learning task, we get a training set in which companies are divided into successful and unsuccessful. Accordingly, machine learning methods can extract the knowledge and explain the situation.

In the broadest sense, the task could be defined as *predicting the performance of a particular company in the future*. It is based on the following factors, described in Table 10.

Table 10. Extended criteria for company performance prediction

Factor	Related issues
Financial ratios	How to choose and combine financial ratios
Environmental factors	Human expert evaluation is required, however, could be assisted by automated solutions
Industry	A company belongs to a certain industry that could be specifically affected by environmental factors
Company specifics	Within the same industry, companies could have different factors affecting their performance
Definition of performance	Could be defined as bankruptcy/non-bankruptcy or have a more complex division, e.g., including gray area.

While every business sector can be affected significantly by almost any internal or external factor, to limit the scope of the research, we will focus only on legislation change analysis in the remainder of this article. Environmental factors could be characterized by regulatory changes in the country and beyond, and other non-determined factors which could be associated with the year. Thus, the proposed solution will complement the Z-score assessment of a company’s insolvency by adding knowledge about certain legislation change impacts to specific business sectors. Such knowledge will be extracted from historical data using expert advice and machine learning methods.

The proposed solution consists of two parts:

- I. Find cases in history where changes in regulatory enactments have affected the solvency of the industry (the expert oriented method);
- II. Extract knowledge of which companies are affected and how (the machine learning method).

Figure 4 represents Part I. At the stage where the expert analyzes the impact of external factors, his decision is supported by two automated techniques: (1) mathematical analysis, summary and graphical representation of the selected financial parameters (points 5, 1, and 2 in Figure 4); (2) a selected list of changes in the regulations, which is used as an external service (point 6). The result of this stage is the hypothesis put forward by the expert that *changes in a particular regulation in a particular year cause difficulties for companies in a particular industry*.

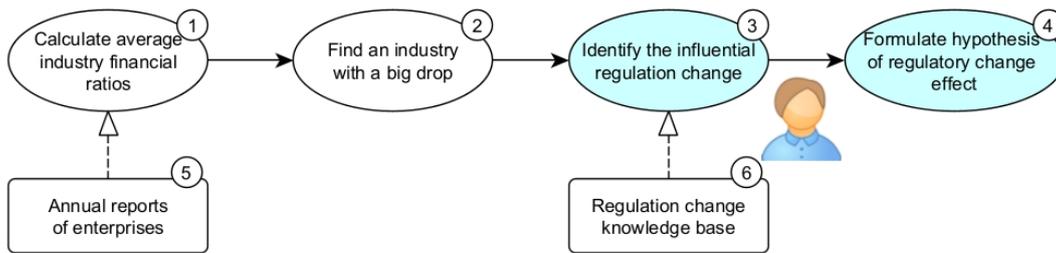


Figure 4. Extended model (Part I) – the expert method

Figure 5 represents Part II. To address the issue of choosing financial ratios, the company’s financial data (point 1 in Figure 5) are taken from annual reports (point 12). Each implementation of the prediction system can differ in the complexity of classification (points 2 and 13). Two or more class learning models can be defined, based on insolvency, bankruptcy or more refined factors derived from financial ratios. These decisions along with data preparation lead to a company data set (point 3), where we propose to (1) respect the industry of a company; (2) split data into years, representing each company in each year; (3) include expert’s hypothesis of regulatory change effect as a factor for particular year and industry (point 4 in Figure 5, corresponding to point 4 of the expert’s method in Figure 4). With this type of multi-year company knowledge base, we can model the situation in three steps.

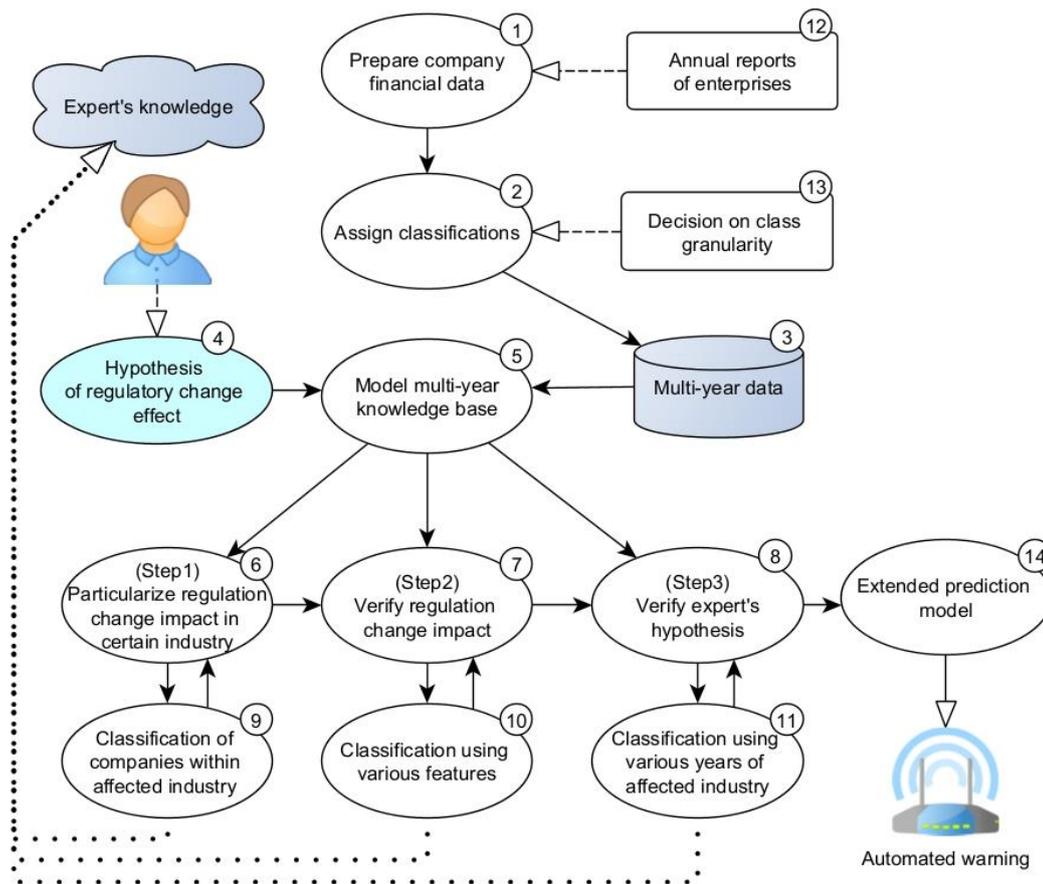


Figure 5. Extended model (Part II) – the machine learning method

Step 1 helps to evaluate why some companies in the same industry are affected by regulatory change more than others.

Step 2 clarifies which features are essential for explaining variation between companies (e.g., particular legislation change, a year, etc.).

Step 3 finalizes the verification of the expert's hypothesis by comparing the results between the years of the same companies.

The findings of these steps not only could help to develop an extended prediction model for a company's performance and automated early warning system but also would help to improve the expert's knowledge with data-driven insights.

5 Conclusions and Future Work

This article is devoted to the research on company bankruptcy prediction solutions, experiments on bankruptcy prediction for Latvian companies, and the development of a more intelligent company bankruptcy prediction model that can serve as an early warning system. Traditionally, only financial ratios are used to estimate company financial distress, insolvency, or bankruptcy. Evaluating the circumstances, we hypothesize that changes in national legislation, changes in international relations, natural disasters, global crises and more can have a major impact on business performance. Thus, even with the same financial ratios but at different times and places, the results of companies may differ.

The literature review shows that current implementations for predicting bankruptcy or insolvency are local and relate to particular national scope. Attributes that are used for prediction cover different financial factors. Applied ML methods and their evaluation metrics vary.

Considering the findings of the literature review, the authors experimented with the Altman Z-score approach and the ML decision tree algorithm to predict bankruptcy of Latvian companies. Comparing the overall results, for Latvian company bankruptcy prediction, the ML-approach, with an average F-Measure of 74%, outperformed Z-score, with an average F-Measure of 49%, by 25%. However, if we consider the ability to capture the bankrupt class, the results are less satisfying. The Altman Z-score approach predicts bankruptcy for too many companies while the ML-approach is quite optimistic and predicts bankruptcy for only a few companies. These results are especially useful as a basis to offer improvements and create a better early bankruptcy warning system.

Literature review and initial experiments with predicting bankruptcy for Latvian companies assure us that there is room for improvement. The authors believe that the basic idea of Altman's Z-score is in the right direction, as well the ML-approach. However, both methods can be improved. Altman's Z-score approach can be refined for Latvian companies by adjusting the bankruptcy thresholds based on historical data. The ML-approach can be supplemented with additional attributes such as tax arrears, calculated attributes such as debt/equity, and more. By taking the best parts of each approach, the results of both methods can be combined to provide a synergistic assessment of the company's future. In addition, the impact of global circumstances can be considered to correctly interpret the financial indicators resulting from the global situation in each fiscal year. Accordingly, an extended prediction model is proposed to improve the current results.

Development of an extended prediction model includes an expert method and a ML method to combine the background knowledge of the effect of legislation changes, from both expert and machine learning capabilities, in order to find relationships in a large amount of data.

Our research has both theoretical and practical implications. Previous research on company performance prediction is reviewed and analyzed. The bankruptcy prediction solution for Latvian companies has been introduced and experiments with two different approaches are carried out and discussed. As a result, the proposal is made for an extended prediction model which takes into account both financial and non-financial factors.

The research is currently limited to Latvian companies but, because the proposed model does not include any country-level specifics, as a theoretical framework it may serve for wider use.

Future work is to implement the proposed extended company performance prediction model. Some parts of the model represented in Figure 4 are already developed, and experts are working to define the hypothesis for specific dropped ratio cases in the Latvian setting. The questions to be

answered are: How to find regulatory changes which impacted company insolvency? Is that even possible? Is it possible to learn from such situations, and what knowledge can we extract from the data using machine learning methods?

Another future development segment is to introduce comparison with other countries' experience, e.g., the UK, where publicly available company financial data are available.

Acknowledgments

The research leading to these results has received funding from the project "Competence Centre of Information and Communication Technologies" of EU Structural funds, contract No. 1.2.1.1/18/A/003 signed between IT Competence Centre and Central Finance and Contracting Agency, Research No. 1.19 "Comparative analysis of regulatory and financial data of companies from different countries for forecasting business results".

References

- [1] T. Van Gestel et al., "Bankruptcy prediction with least squares support vector machine classifiers," *IEEE/IAFE Conference on Computational Intelligence for Financial Engineering, Proceedings (CIFER)*, 2003, Janua, pp. 1–8, 2003. Available: <https://doi.org/10.1109/CIFER.2003.1196234>
- [2] H. T. Kesgin et al., "Bankruptcy Risk Forecast Based on Company Balance Sheet Data Using Machine Learning," *4th International Conference on Computer Science and Engineering*, pp. 195–200, 2019. Available: <https://doi.org/10.1109/UBMK.2019.8907029>
- [3] C. H. Chou, S. C. Hsieh, and C. J. Qiu, "Hybrid genetic algorithm and fuzzy clustering for bankruptcy prediction," *Applied Soft Computing Journal*, vol. 56, pp. 298–316, 2017. Available: <https://doi.org/10.1016/j.asoc.2017.03.014>
- [4] R. A. Baxter, M. Gawler, and R. Ang, "Predictive model of insolvency risk for australian corporations," *Conferences in Research and Practice in Information Technology Series*, vol. 70, pp. 21–27, 2007.
- [5] K. J. Cios and L. A. Kurgan, "Hybrid Inductive Machine Learning: An Overview of CLIP Algorithms," *New Learning Paradigms in Soft Computing*, vol. 84, Springer, pp. 276–322, 2002. Available: https://doi.org/10.1007/978-3-7908-1803-1_10
- [6] M. Zięba, S. K. Tomczak, and J. M. Tomczak, "Ensemble boosted trees with synthetic features generation in application to bankruptcy prediction," *Expert Systems with Applications*, vol. 58, pp. 93–101, 2016. Available: <https://doi.org/10.1016/j.eswa.2016.04.001>
- [7] F. Barboza, H. Kimura, and E. Altman, "Machine learning models and bankruptcy prediction," *Expert Systems with Applications*, vol. 83, pp. 405–417, 2017. Available: <https://doi.org/10.1016/j.eswa.2017.04.006>
- [8] H. Chang, "The application of machine learning models in company bankruptcy prediction," *Proceedings of the 2019 3rd International Conference on Software and e-Business*, vol. 122, pp. 199–203, 2019. Available: <https://doi.org/10.1145/3374549.3374550>
- [9] Y. Zelenkov, E. Fedorova, and D. Chekrizov, "Two-step classification method based on genetic algorithm for bankruptcy forecasting," *Expert Systems with Applications*, vol. 88, pp. 393–401, 2017. Available: <https://doi.org/10.1016/j.eswa.2017.07.025>
- [10] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*. 2nd ed., The Morgan Kaufmann Series in Data Management Systems. Elsevier, 2005.
- [11] E. I. Altman, "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy," *The Journal of Finance*, vol. 23, no. 4, pp. 589–609, 1968. Available: <https://doi.org/10.1111/j.1540-6261.1968.tb00843.x>
- [12] E. I. Altman, *Corporate Financial Distress: A Complete Guide to Predicting, Avoiding, and Dealing with Bankruptcy*. 1st ed., New York: Wiley, 1983.
- [13] E. I. Altman, "Predicting financial distress of companies: revisiting the Z-Score and ZETA® models," *Handbook of Research Methods and Applications in Empirical Finance*, Edward Elgar Publishing, pp. 428–456, 2013. Available: <https://doi.org/10.4337/9780857936097.00027>
- [14] X. Wu et al., "Top 10 algorithms in data mining," *Knowledge and Information Systems*, Springer, vol. 14, pp. 1–37, 2008. Available: <https://doi.org/10.1007/s10115-007-0114-2>
- [15] M. S. Aksoy, "Dynamic system modelling using RULES3 induction algorithm," *Information Sciences*, vol. 10, no. 1, pp. 121–132, 2005. Available: <https://doi.org/10.3390/mca10010121>
- [16] I. Witten et al., *Data Mining: Practical Machine Learning Tools and Techniques*. 2016.

- [17] Law on the Annual Financial Statements and Consolidated Financial Statements. Latvijas Vēstnesis, no. 222, 12.10.2015. Available: <https://likumi.lv/ta/en/en/id/277779-law-on-the-annual-financial-statements-and-consolidated-financial-statements>
- [18] T. J. Ward and B. P. Foster, "A Note on Selecting a Response Measure for Financial Distress," *Journal of Business Finance & Accounting*, vol. 24, no. 6, pp. 869–879, 1997. Available: <https://doi.org/10.1111/1468-5957.00138>
- [19] P. Chou et al., "Integrated annual report readability and dominance-based rough set theory to forecast corporate's operating performance," *2019 IEEE 8th Global Conference on Consumer Electronics, GCCE 2019*, pp. 602–604, 2019. Available: <https://doi.org/10.1109/GCCE46687.2019.9015506>
- [20] Y. J. Chen et al., "Financial Forecasting with Multivariate Adaptive Regression Splines and Queen Genetic Algorithm-Support Vector Regression," *IEEE Access*, vol. 7, pp. 112931–112938, 2019. Available: <https://doi.org/10.1109/ACCESS.2019.2927277>