Routledge
Taylor & Francis Group

Economic Research-Ekonomska Istraživanja

ISSN: 1331-677X (Print) 1848-9664 (Online) Journal homepage: http://www.tandfonline.com/loi/rero20

Analysis of investors' preferences in the Montenegro stock market using data mining techniques

Ljiljana Kašćelan, Vladimir Kašćelan & Miomir Jovanović

To cite this article: Ljiljana Kašćelan, Vladimir Kašćelan & Miomir Jovanović (2014) Analysis of investors' preferences in the Montenegro stock market using data mining techniques, Economic Research-Ekonomska Istraživanja, 27:1, 463-482, DOI: 10.1080/1331677X.2014.970451

To link to this article: http://dx.doi.org/10.1080/1331677X.2014.970451

Published online: 30 Oct 2014.

Economic Research-Ekonomska Istraživanja, 2014
Vol. 27, No. 1, 463–482, http://dx.doi.org/10.1080/1331677X.2014.970451

Routledge
Taylor & Francis Group

Analysis of investors' preferences in the Montenegro stock market using data mining techniques

Ljiljana Kašćelan[a]*, Vladimir Kašćelan[a] and Miomir Jovanović[b]

[a]*Faculty of Economics, University of Montenegro, Jovana Tomaševića 37, 81000 Podgorica, Montenegro;* [b]*Biotechnical Faculty, University of Montenegro, Mihaila Lalića bb 37, 81000 Podgorica, Montenegro*

(*Received 10 June 2013; accepted 19 September 2014*)

This article analyses the preferences of different types of investors to stock characteristics in the Montenegrin stock market. The majority of papers deal with stock portfolio analysis of the institutional investors. Since the number of individual investors in the Montenegrin market is much higher, the analysis of their trading behaviour is also very significant. In this article, using data mining techniques, we tested trading behaviour with stocks for both types of investors. We prove that data mining techniques, such as logistic regression, clustering and decision trees, provide good results in this type of analysis. The analysis may be useful to the future investors, brokers and stock exchange.

Keywords: stock trading preferences; stock characteristics; investor features; logistic regression; decision trees

JEL classifications: G02, G11, C38, C53

1. Introduction

The process of mass voucher privatisation, which ended at the beginning of 2002, was a strong driving force for development of the Montenegrin capital market. Since that period, the stock market experienced intensive growth, especially until 2007. In this period, the number of trades in the stock market was growing each year compared to previous one, so that in 2007 it reached 221,000. Some of the development indicators for the Montenegrin market are given in Table 1. In the initial years of the market activities, there were a lot of anomalies and imperfections. Figure 1 shows the fluctuation of stock prices of one Montenegrin company during the period 2007 to 2011. A large number of companies had similar variations in stock prices in this period. It can be seen from the chart that in 2007, the stock prices of this company were high, between EUR 20 and 50, and in 2008 prices had a downward trend and went below EUR 5.

At the beginning of the crisis in 2008, the entire market experienced a significant downfall. In that year the number of stock trades, as well as the amount of traded stocks, was five times lower than in the previous year, in which the maximum was reached (Table 1). The downward trend continued and stability was reached in 2011. Because of that, for analysis in this article, we used the data from 2011, when the stock prices were more stable.

*Corresponding author. Email: ljiljak@ac.me

Table 1. Development indicators of the Montenegrin stock market.

	Issuers Companys (No.)	Outstanding Stocks (No.)	Accounts (No.)	Stock Trades (No.)	Traded value
2002	346	2,660,285,442	376,613	3,759	5,758,205.69
2003	379	2,814,700,322	378,850	20,414	18,937,651.79
2004	385	2,825,600,092	383,189	54,549	33,619,342.08
2005	387	2,865,498,762	387,174	101,108	159,064,807.79
2006	395	2,963,587,492	390,406	109,455	298,034,259.74
2007	421	2,987,735,146	395,942	221,086	699,199,800.17
2008	420	3,565,496,462	473,137	47,640	142,490,579.83
2009	419	3,876,686,095	474,441	35,136	163,487,672.32
2010	386	4,084,672,807	475,138	16,655	38,023,672.15
2011	371	4,186,011,356	475,542	12,375	50,129,059.70

Source: Central Depository Agency of Montenegro (CDA).

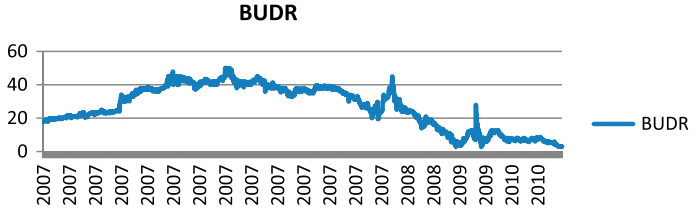

Figure 1. Fluctuation in prices of the Budva Riviera Company stocks from 2007 to 2011. Source: CDA.

In the Montenegro capital market, at the end of 2011, the number of issuers was 371, the number of outstanding stocks was over 4 billion, and the number of accounts of investors was over 475,000. So, the number of stocks offered to the investors in the market, grows from year to year, therefore making the decision in which stocks to invest more difficult. Imperfections in the Montenegro market and crisis conditions make this decision even more complex. The investors select stocks on the basis of easily available information, and that is quite scarce. The selection, of course, is also based on their capabilities, i.e. background. The investors with different backgrounds have different preferences in stock trading. So, for example, the investors with high trading frequency may prefer stocks with lower prices, or the wealthy investors may prefer stocks with higher dividends, etc. The analysis of the preferences, in such conditions, can be very useful to future investors.

In this article, we analysed the trading preferences of various types of investors in the Montenegrin stock market. In the analysis we used the data mining method based on the logistic regression, clustering and decision trees. The article is organised as follows. In the second section we give a review of related work, which deals with analysis of the trading behaviour of investors in stocks. In the third section we introduce the proposed method. In the fourth section, we define the initial data-set, we present the results obtained by the proposed method and discuss and analyse the obtained results. In the conclusion, we give our final considerations and possibilities for further research.

2. Related work

A large number of papers dealt with the preferences of investors in the stock markets. The majority of those papers analysed the preferences of institutional investors towards stocks, i.e. the companies with different characteristics.

In his paper, Elkinawy (2005) examined the influence of company characteristics, such as the financial condition, competition and management, on the stock portfolio of the Latin American investment funds, during the Asian–Russian financial crisis. The research showed that the investment funds preferred cross-listed companies, and they avoided companies which had main Russian exporters as the competition. Dahlquist and Robertsson (2001) considered the preferences of foreign investors in the Swedish stock market, and determined that they prefer big companies, low dividends and high cash positions on the balance sheets. Also, they determined that the visibility of the company on international market, in form of exports or stock listing on other Stock Exchanges, has significant influence on foreign investors. Same authors also researched the interests of the domestic investors, and have determined that the preferences towards big companies are a common point for all institutional investors. In their paper, Aggarwal, Klapper, and Wysocki (2005) analysed the selection of investments on emerging markets (markets in 30 countries were taken into consideration), by American investment funds. They examined the role of the country in which the investments were made, as well as the role of the characteristics of the company during the selection. At the national level, they discovered that the factors which had influence on selection of stocks were strong legal framework, higher rights of stockholders and transparent accounting. Concerning the companies, the funds had the tendency towards companies with American Depositary Receipt (ADRs; cross-listed companies) and with good accounting disclosure policy.

Very few studies considered individual investors in their analysis or examined how their features affected stock trading preferences.

For example, Grinblatt and Keloharju (2001) analysed the trades of individual and institutional investors in the Finnish stock market. The result shows that the trading behaviour of investors is affected by the company's past returns, the size of the holding period, capital gain or loss, and the company's historical price patterns. Ng and Wu (2006), studied both individual and fund investors, where they tested the preferences of Chinese individual investors depending on their level of wealth. They found that wealthier investors tend to invest in the stocks with high volatility and low book-to-market, while less-wealthy investors prefer to choose the stocks with low price and low earnings. Peress (2004), showed that wealthier individual investors were less risk averse and they had a higher demand for information. Chen, Kim, Nofsinger, and Rui (2004), obtained data from the accounts of brokers in China and showed that the accumulated experience of individual investors affected their behaviour during the stocks trading. Barber and Odean (2001), documented that men are more confident than women in accord with the behavioural finance model (men tend to trade more than women). Using US sample data, Jianakoplos and Bernasek (1998), found that women exhibit more risk aversion in financial decision-making than men. Barber, Lee, Liu, and Odean (2014), found that frequency of trade plays an important role in investor outcomes.

In recent years, individual characteristics are of first-order importance for portfolio choice. For example, Grinblatt and Keloharju (2009), and Grinblatt, Keloharju, and Linnainmaa (2011), showed that sensation seeking, overconfidence and IQ affect stock market participation, using data from the Finnish registry. Campbell, Ramadorai, and Ranish (2013), found that experience in good investment performance in the Indian

stock market pushes investors towards stock growth, large stocks and high-momentum stocks.

Barber and Odean (2011) provided an overview of research on the stock trading behaviour of individual investors in a larger view. These studies found that individual investors earn poorly, have a strong preferences for selling winners and holding losers (the disposition effect), repeat behaviours that previously coincided with pleasure (the reinforcement learning), consider only stocks that first catch their attention, tend to hold undiversified stock portfolios and buy stocks when most other investors sell and sell when they buy (contrarians).

Most of previous studies examined the stock market preferences of investors using the statistical method of regression analysis. The results of these studies give the impact of individual factors with statistical significance at the 1, 5 and 10% levels. However, a drawback of the regression models is a low interdependency of factors, because adjusted R^2 is generally less than 30%.

Because of these drawbacks to test the behaviour of stock market investors we proposed a data mining method based on logistic regression, clustering and decision trees. The advantage of this method, compared to the previous studies that used regression model, is the obtained interdependency of factors. Tsai, Lin, and Wang (2009), used neural networks and decision trees to identify trading preferences on the Taiwan stock market. In that study, they identified the stock portfolios and features of investors that affected their selections. The advantage of our analysis, compared to the mentioned paper, is that we used the data from entire stock market. In their paper, they used only data from one broker, and that can have the limiting influence on the results. Besides, we used decision tree algorithms obtained by component design which enables user to intelligently select in advance implemented components, best suited for specific data-set. Such defined algorithms can provide higher accuracy of classification, as well as the lower complexity of generated tree from the original decision tree algorithms.

Since the number of individual investors in the Montenegrin stock market is large, an analysis of their stock trading behaviour would be very useful. Thus, we tested behaviour of the both investors' types.

3. Method

In this article we tested the hypotheses that: (1) stock characteristics affect the high and low preferences of investors on the stock market; (2) stock characteristics affect the trading preferences of investors having a specific feature on the stock market; and (3) that the type of investor, i.e. a specific set of investor's features, affects the selection of the stock portfolio.

To test the hypothesis we considered three groups of factors which interactively influence the selection of stock portfolio: stock characteristics, balance sheet indicators of companies and investor features.

Motivation for the choice of the stock characteristics was found in the literature. There are many studies that examine the preferences for certain stock characteristics when choosing stock portfolio (Covrig, Lau, & Ng, 2006; Dahlquist & Robertsson, 2001; Ng & Wu, 2006; Tsai, Lin, & Wang, 2009). We used the stock characteristics (with acronyms in parentheses) as follows.

Age is the number of years since the stock was issued (Age). The stocks with newer dates can sometimes be more attractive to investors because they are current, people talk about them and they are publicly present. Dividend yield (DY) is the annual dividend

per stock divided with stock price. The stocks with high DY usually have stable prices and the investors who invest in them count on incomes from dividends. Stocks with low DY have big fluctuations in prices and mainly those who want to earn on fluctuation of prices invest in them. Earnings-per-share (EPS) is the net profit[1] divided by number of issued stocks at the end of period. EPS represents the earning of the investor per stock. The investors who want to achieve the good return invest in stocks with high EPS. Market-to-book (MB) is the average stock price divided by book value, i.e. ratio of stock capital and number of issued stocks. Stocks with high MB are growing stocks, i.e. stocks whose market value is high compared to book value. Investments in these stocks bear higher risks than investments in those with low market value compared to book value. Price is the average stock price in observed period (Price). Stocks with low prices are more attractive to the investors, because they require low expenses. Price/earnings ratio (PE) is the average stock price divided by EPS. Ratio PE is inversely proportional to investment return, meaning that the low PE usually means high Return on Investment (ROI). Return is the annual rate of stock return (the ratio of difference between the end and initial price increased by amount of the dividend, and initial price) (Return). Return tells us about the past performance of the stock, i.e. if the stock is the winner or loser at the end of the year. The winner means that the difference in stock price, at the beginning and at the end of year increased by dividend per stock, compared to the initial stock the price is high. Experienced investors will prefer the winning stocks. Size is the total market capitalisation (total amount of traded stocks) (Size). The investors usually trade with stocks whose size–trade value is high. The reason is that the companies with big market capitalisation are safer and more popular. However, that will primarily attract inexperienced investors, but not those who carefully analyse other characteristics of stocks in which they will invest. We also considered the company balance sheet indicators: return on assets (ROA) as the ratio of net profit and assets; return on equity (ROE) as the ratio of net profit and stock capital, and the debt to equity (DE) ratio.

Many studies (Barber & Odean, 2001; Barber & Odean, 2011; Chen et al., 2004; Peress, 2004) found that kind of investor (individual/organisation), wealth, region (foreign/domestic investors), trading frequency, experience and gender were the main factors that affect trading preferences in stock market. According to this, we used features of investors (with acronyms in parentheses) as follows: kind of investor, individual or institutional (IndOrg); gender of investor (Gender); wealth of investor as the total amount of stock ownership (Wealth); region of investor (Region); trading frequency, as the number of trades (Frequency) and experience, as the number of years from opening of the first account (Experience).

As the extent of the investor X preferences towards stock Y, we used the following measure:

$$\text{Pref} = \frac{\text{Number of trading toward stock Y by investor X during observed period}}{\text{Total number of trades by investor X during observed period}} \quad (1)$$

This measure is modified from the measure of stock preferences developed by Ng and Wu (2006). Their measure is based on the market value of the stocks. Such measures will underestimate less wealthy investors' preferences on stocks with high capitalisation. Therefore, we prefer a number of trades as the measure. Tsai, Lin, and Wang (2009) used similar measure based on the number of trades.

To test the above hypothesis we proposed three predictive classification data mining models. We decided to use the classification models because of the previously

mentioned drawbacks of regression to identify only the influence of certain factors, but not the interdependency all of them. Predictive classification models identify and predict interdependent influence of the predictors to the target variable (class label).

The model based on the first hypothesis identifies and predicts if and how the stock characteristics affect the high or low preferences of investors. The target variable in this model is the measure of investors' preferences defined by relation (1). In this classification model, we categorised the target variable using two values: L-low (lower preferences than average) and H-high (higher preferences than average). The predictors are the above stated stock characteristics. We categorised all of the predictors and we transformed them into dummy variables.

The model based on the second hypothesis identifies and predicts if and how the stock characteristics affect the investors with a specific feature. The target variables in this model are the different features of investors. The predictors are the above stated stock characteristics. All of the predictors are categorised and they are transformed into dummy variables.

For the model based on the third hypothesis, it was necessary to identify the types of investors (their features) that prefer certain stock portfolio. For this purpose the stocks were initially clustered based on the above stated stock characteristics. The produced clusters represent different group of investors who prefer certain stock portfolio. This model identifies and predicts the features of investors belonging to these groups. The target variable in this model is the cluster variable, the values of which identify the stock clusters. The predictors are the above stated features of the investors, which are categorised and transformed into dummy variables.

Data mining technique, which we proposed to develop the first two models, is logistic regression. This technique is used for prediction of binominal (0,1) or categorical (with limited set of categories) target variable on basis of predictor variables, where the data-set is classified in as many classes as the target variable has values. Logistic regression allows us to determine not only which predictor variable has the interconnected influence on the target variable, but also how big that influence is. For our models, this opportunity is essential and that is why we chose logistic regression.

Kernel logistic regression (KLR) is a nonlinear form of logistic regression, which is obtained by replication of data vector with help of kernel function. In this article we used KLR, which is based on fast dual algorithm (Keerthi, Duan, Shevade, & Poo, 2005). Ruping in 2003 implemented this algorithm in form of the programme MyKLR. In this article we also used the poly-nominal logistic regression (case when the target variable is categorical with multiple categories). This regression is implemented with help of binominal regression, with use of the method 'one in relation to all other'.

For development of the third model we proposed the data mining techniques clustering and decision trees. Clustering finds the similar groups within the data-set. We chose this method because it is necessary to divide investors into groups which have similar preferences in the selection of stock portfolio. Clustering method does not give explicit descriptions of the clusters. Because the decision tree can extract explicit rules from the clusters, we used this method to the result of clustering, in order to obtain more reliable conclusions. Thus the obtained rules identify features of investors who belong to the clusters.

This study uses the data mining technique k-mean clustering, which iteratively forms k clusters with help of functions for evaluation of distances and mean values of cluster (Hartigan, 1975).

Decision trees enable classification of data sets based on target variable, where the tree branches define classification rules expressed in terms of predictive attributes. There are several data mining algorithms for induction of decision tree. Majority of these algorithms implement decision tree induction by generating split for each attribute (tree node) which will be the best for prediction of target variable value.

One of the first decision tree algorithms is ID3 (Quinlan, 1986 a). This algorithm works only with categorical variables, it is based on 'multi-way' split and it uses 'information gain' as measure for split quality. This evaluation measure is biased towards choosing attributes with more categories. Breiman, Friedman, Stone, and Olshen (1984) proposed CART algorithm which works with both categorical and numerical variables, and for split evaluation it uses the 'Gini' measure. The algorithm supports only binary splits. Algorithm C4.5 (Quinlan, 1993 b), is an improvement of the ID3 algorithm which can work both, with categorical and numerical data. It uses a 'multi-way' split for categorical, and binary for numerical data. For split evaluation it uses a 'gain ratio' measure, which is not biased towards attributes with several categories. It also includes three pruning algorithms: reduced error pruning, pessimistic error pruning and error based pruning. CHAID algorithm was proposed by Kass, (1980). In this algorithm Chi-square test is used for evaluation of the split quality. QUEST algorithm (Loh & Shih, 1997), uses removal of insignificant attributes with chi-square test, for categorical, and ANOVA f-test, for numerical data.

In this article we used decision tree algorithms obtained by component design proposed by Delibasic, Jovanovic, Vukicevic, Suknovic, and Obradovic (2011). These algorithms are obtained by combining different components of original decision tree algorithms (split creation, split evaluation, stop criteria, etc.). Component design enables user to intelligently select in advance implemented components, the best suited for specific data-set. Such defined algorithms can provide higher accuracy of classification, as well as the lower complexity of generated tree than the original decision tree algorithms.

For the realisation of the previosuly defined models we used an open source data mining platform Rapid Miner (www.rapidminer.com), as well as the WhiBo (www.whibo.fon.bg.ac.rs), plug-in for Rapid Miner in this manner:

- We used the Rapid Miner X-Validation operator (tenfold, stratified sampling) to perform the cross-validation to estimate the statistical performance of the learning operator using an unseen data-set.
- For the classification performance evaluation, we used the operators Performance (Binominal Classification) and Performance (Classification).
- For the regression performance evaluation, we used the Performance (Regression) operator.
- For regression we used the Linear Regression (feature selection = 'M5 prime', with elimination of collinear features) and Polynomial Regression operators.
- For logistic regression we used the operators Logistic Regression (dot kernel and C = 1.0), which is based on a Java implementation of MyKLR software and Poly-nominal by Binominal Classification based on the method 'one in relation to all other'.
- For clustering, we used the operator k-Means, which performs clustering with the *k-means* algorithm.
- For component design of decision tree algorithms we used WhiBo plug-in and WhiBo Generic decision tree operator.

4. Empirical results

In this section we tested stated hypotheses on the data from the Montenegrin capital market, using the proposed data mining method. Empirical results verified the proposed method and they confirmed the hypotheses. When interpreting the obtained results, we identified the actual behavioural patterns of investors in the Montenegrin stock market, which may be useful for future trading strategies.

4.1. Data description

For the analysis we used the data from the Central Depository Agency of Montenegro on trades in the Montenegrin stock market in 2011. These data include basic information about the investors[2] such as the kind of investor (individual or institutional organisation), gender, region, date of opening of the first account (experience of the investor)[3] and wealth,[4] then data about the stocks, such as the unique trading symbol, initial, end and average price and number of outstanding stocks at the end of 2011, and in the end, data on trades, such as the date of trade, quantity and trading amount. In the initial data-set there were 12,375 records, i.e. stock trades. The stocks of 146 different companies were traded. Stock characteristics and balance sheet indicators are calculated on basis of information on companies, taken from the web site of the Securities Commission, such as the stock capital, net profit, amount of dividend, net assets and liabilities. Some of the smaller companies, which stocks were traded in 2011, were liquidated at the end of accounting year, and for some of them the necessary data did not exist. Trades with stocks of these companies are removed from the initial set. The total of 3584 records on stock trades remained. On this set we calculated two more derived values we needed for the analysis. Those are investor trading frequency (total number of trades within this data-set) and preferences of investor X for stock Y, as it is defined in Section 3. Thus we got the initial data-set for which the statistics are given in Table 2.

It can be noticed that the annual rate of stock return for majority of the companies is negative, and this is a consequence of downward trend in stock prices which were unrealistically high in previous periods (Figure 1). Also, it can be seen that the average value for attribute Size (amounts of traded stocks) and Pref (investor preferences for a stock) is high. This is result of the fact that large number of small companies, whose stocks are poorly traded, is removed from the initial data-set due to liquidation or lack of some balance sheet data. It is noticeable that the companies have low average EPS, very low average for DY, high average for MB and very small average value for ROA and ROE, as well as the high average DE. In total, stock characteristics and balance sheet indicators of the companies reflect quite bad conditions of Montenegrin companies, specific for the current crisis. It should be pointed out that the indicators would be even worse if we took into consideration all companies which stocks are traded in 2011, because we removed large number of small companies from the initial set. Wealth, frequency and experience are divided in groups according to criteria given in Table 3.

4.2. Data mining results

As in most of the previous studies, we firstly analysed influence of the stock characteristics on investor preferences with the linear regression. Target (dependent) variable is attribute Pref, which represents the preferences of stock investor as defined in Section 3,

Table 2. Initial data-set-metadata view.

Attribute Name*	Type	Statistics	Range
IndOrg	binominal	mode = I (2,677), least = O (907)	O (907), I (2,677)
Region	polynominal	mode = Central (2,440), least = South (264)	Central (2,440), Foreign (609), South (264), North (271)
Gender	polynominal	mode = M (1860), least = F (324)	NULL (1,400), M (1860), F (324)
Wealth	polynominal	mode = G3 (850), least = G5 (203)	G2 (794), G3 (850), G1 (610), G5 (203), G4 (220)
Frequency	polynominal	mode = G2 (2002), least = G3 (754)	G3 (754), G2 (2002), G1 (828)
Experience	polynominal	mode = G3 (1990), least = G2 (555)	G1 (1,039), G2 (555), G3 (1990)
MB	real	avg = 0.860 +/− 0.289	[0.152; 2.558]
Age	real	avg = 5.917 +/− 3.128	[1.486; 9.856]
DY	real	avg = 0.078 +/− 0.051	[0.011; 0.566]
EPS	real	avg = 0.432 +/− 0.759	[0.019; 12.076]
Price	real	avg = 7.455 +/− 85.383	[0.283; 1810.826]
PE	real	avg = 13.703 +/− 8.990	[1.767; 149.950]
Return	real	avg = −0.242 +/− 0.196	[−0.555; 0.183]
Size	real	avg = 2916934.949 +/− 1989983.789	[3864.044; 4853918.545]
DE	real	avg = 0.206 +/− 0.063	[0.000; 1.371]
ROA	real	avg = 0.057 +/− 0.033	[0.004; 0.418]
ROE	real	avg = 0.080 +/− 0.041	[0.006; 0.117]
Pref	real	avg = 0.765 +/− 0.289	[0.034; 1.000]

Note: *Acronyms for attribute names are explained in Section 3.
Source: Authors' calculation.

Table 3. Generalisation of attributes.

Wealth	EUR
G1*	x<5.000
G2	5.000<x<20.000
G3	20.000<x<100.000
G4	100.000<x<500.000
G5	x>500.000
Frequency	No.
G1**	x<5 (rare)
G2	5<x<50 (once a week)
G3	50<x<250 (every day)
Experience	Years
G1***	1<=x<=5
G2	5<x<10
G3	x>=10

Notes: *Group G1 includes investors with total stock ownership less than 5000 euros. Groups G2 to G5 are to be interpreted in a similar manner.
**Group G1 includes investors who traded less than five times in the observed period. Groups G2 and G3 are to be interpreted in a similar manner.
***Group G1 includes investors who are present in the stock market up to five years. Groups G2 and G3 are to be interpreted in a similar manner.
Source: Authors' calculation.

while the independent variables (regressors) are attributes representing the stock characteristics. All independent variables are categorised and for the needs of regression analysis they are transformed to dummy variables. The results of linear regression are presented in Table 4. The corrected coefficient of determination is very small, i.e. 0.032773, and only 26% of preferences have deviations lower than 20%. The results of the linear regression give the impact of individual factors with high statistical significance. However, a drawback of the model is a low interdependency of factors because adjusted R^2 is small. The small R^2 may indicate that the relationship between the dependent variable and the regressors is not linear. To test the non-linearity we applied the RESET test (Ramsey, 1969). For quadratic unrestricted functional form we got that $F_{(1;\ 3,553)} = 3{,}461.53$ which is greater then $F_{0.05\ (1;\ 3,553)} = 3.84$. Because the F test statistics is greater than the F critical value we reject the null hypothesis that the true specification is linear (which implies that the true specification is non-linear). However, after testing polynomial regression models (max degree = 5) we get small $R^2 < 0.040199$. This indicates low interdependency of factors in the non-linear regression, too.

Since the regression failed to provide good results in terms of interdependence, to develop our first model, we applied the logistic regression. For H as a positive class we obtained the model of logistic regression (the weighting coefficients of dot kernel model) in Figure 2. The model had an accuracy[5] of 61.56% (the logistic model performance is presented in Table A1 in the Appendix), and we showed that the stock characteristics affected the high and low preferences of investors in the Montenegrin stock market (first hypothesis).

With the logistic regression we also developed our second model. For this analysis the dependent variables were IndOrg, Region, Gender, Wealth, Frequency, Experience, respectively, while the independent variables were stock characteristics defined in Section 2. Since these dependent variables are categorical, with several categories, we applied poly-nominal logistic regression upon principle 'one against all'. In Table 5, we present some of the results obtained with this analysis. We only present results for which the class precision (in brackets) is higher than 30%. Table 5 shows the weighting coefficients of the dot kernel logistic regression that have an influence on the positive class, given in the header of the table. With this model we showed that the choice of stock characteristics is in relation with the individual investors' features (second hypothesis).

In order to realise the third model, it was necessary to divide investors into groups, with the same preferred stock portfolio. Because of that, we applied clustering by stock characteristics to the initial data-set (the clustering model performance is presented in Figure A1 in Appendix). In the first round we got four clusters. With further clustering we got nine clusters, out of which four clusters had small number of records, so we discarded them. The results of clustering according to the characteristics of stocks and companies are presented in Table 6.

Using the component-designed decision trees we generated the rules which express the features of the investors who prefer defined clusters. By using the WhiBo, we created 80 different decision tree algorithms combining the different components and then we tested their performance over initial data-set. For testing of performance we used WhiBo X-Validation operator (5X10-fold cross-validation test with stratified sampling) (differences in performance of these algorithms are presented in Table A2 in the Appendix). By testing we found the optimal algorithm for the data-set, with maximal accuracy and minimal complexity (the optimal decision tree algorithm performance is presented in Table A3 in the Appendix). Note that the optimal algorithm is not one of the original

Table 4. Results of the linear regression.

Attribute*	Coefficient	Std. Error	Std. Coeff.	Tolerance	t-Stat	p-value	Code
DY-High	−0.060083075	0.01774243	−0.125319923	0.490794	−3.38640638	7.61E+11	****
DY-Middle	−0.129943345	0.01126606	−0.149384685	0.754516	−11.5340583	0	****
DY-Low	0.192674982	0.00999518	0.244970872	0.995429	19.2767834	0	****
EPS-Middle	0.147494912	0.0114435	0.167646676	0.729094	12.8889651	0	****
EPS-Low	−0.144515216	0.00987406	−0.174532794	0.998108	−14.635847	0	****
MB-High	0.127589082	0.10458072	2.697912562	0.964049	1.22000573	0.307123	
MB-Middle	−0.129944263	0.01126606	−0.14938574	0.754516	−11.5341398	0	****
MB-Low	0.00402109	0.01139511	0.003514701	0.73661	0.35287866	0.727476	
Price-Middle	0.164009481	0.01147497	0.186206249	0.724893	14.2928038	0	****
PE-High 25+	−0.067149499	0.01267094	−0.143902882	0.995918	−5.29948648	1.26E+08	****
(Intercept)	0.746885436	Infinity	NaN	NaN	0	1	

Note: root_mean_squared_error: 0.281 +/− 0.000.
*Acronyms for attribute names are explained in Section 3.
The Code column in the table shows the statistical significance of the coefficients. High T-Stat values and p-value close to zero point to statistically significant coefficients which are marked with higher number of asterisks. Rapid Miner Linear Regression operator includes feature selection and for this reason the insignificant variables are excluded (collinear features are excluded, too).
Source: Authors' calculation.

Bias (offset): -0.289	w[Price-Middle] = -0.211	w[Age-Middle] = 0.087
w[DY-High] = 0.072	w[Price-Low] = 0.180	w[Age-Low] = 0.040
w[DY_Middle] = 0.028	w[P/E-Low 0-10] = -0.073	w[D/E- High] = -0.375
w[DY-Low] = -0.074	w[P/E-Middle 10-17] = 0.161	w[D/E-Middle] = 0.101
w[EPS-High] = 0.086	w[P/E-High 25+] = -0.000	w[D/E-Low] = 0.062
w[EPS-Middle] = -0.159	w[Return-Low] = -0.094	w[ROA-High] = 0.051
w[EPS-Low] = 0.124	w[Return-Negative] = 0.003	w[ROA-Low] = -0.073
w[M/B-High] = -0.308	w[Size-High] = 0.064	w[ROE-High] = 0.064
w[M/B-Middle] = 0.028	w[Size-Middle] = 0.040	w[ROE-Middle] = 0.602
w[M/B-Low] = -0.008	w[Size-Low] = -0.565	w[ROE-Low] = -0.225
w[Price-High] = 0.086	w[Age-High] = -0.256	

Figure 2. Model of logistic regression.
Note: Since that the class H is positive one, the factors with positive coefficients have an influence on the high preferences of investors, while the low preferences influence factors with negative coefficients. Coefficients with larger absolute value have a stronger impact in both cases. Thus, for example, from this model can be concluded that low total market capitalisation has the most influence on the low preferences of investors (factor Size-Low has a negative coefficient with the largest absolute value).
Source: Authors' calculation.

decision tree algorithms. This fact confirms that the component design provided algorithms with better performance that the original ones. With the optimal algorithms we get a decision tree model with 34 rules, out of which we selected only rules with index of confidence higher than 0.4 (Table 7). Using this model we showed that investors with a certain set of features prefer the same stock portfolio, i.e. that the type of investor has an influence on the stock portfolio selection (third hypothesis).

4.3. Discussion of the results

The results of the data mining analysis point out preferences of the investors with different features for characteristics of stocks and companies.

The model of logistic regression (Figure 2), gives the results for high and low preferences of the investors. According to this model, the preferences of investors in the Montenegro capital market are high for stocks with high DY, low EPS, middle MB, low prices, middle PE, negative return, higher size and low age. Regarding the company balance sheet indicators, the investors show high preferences for companies with low DE, high ROA and higher ROE. This shows that the investors in the Montenegrin stock market are mainly prone to low risk trades. However, high preferences for stocks with low EPS, as well as for stocks which are traded a lot and which are recent on the market, point to the existence of the investors – speculators, who are prone to risky trades. They count on earnings due to sudden change in stock price and they trade most current stocks.

Based on Table 5, it can be seen that individual investors are more prone to risky investments than institutional investors. The institutional investors select stocks with high earning per stock and high DY, with stable but high prices, good ROI, high trade value and higher age. The companies selected by the institutional investors have low DE, high ROA and low ROE. Individual investors prefer stocks with low earnings, middle MB ratio, low prices, middle and high PE ratio, middle trade value and those which

Table 5. Stock preferences of investors with different features.

	Ind (78,81%)	Org (44,96%)	Region-Central (72,41%)	Region-Foreign (33,33%)	Gender-Male (85,23%)	Wealth G2 (32,71%)	Wealth G3 (37,19%)	Freq G2 (66,71%)	Freq G3 (31,16%)	Experience G3 (42,06%)
DY-Middle	0.025	0.050		0.049	0.022	0.053	0.034			
DY-High			0.091	0.147						
DY-Low	0.003		0.001					0.039	0.131	0.061
EPS-Middle		0.005	0.127						0.093	0.159
EPS-High		0.093		0.176	0.081		0.045			
EPS-Low	0.054				0.218	0.085	0.100	0.033		
MB-Middle	0.025			0.049	0.022	0.053	0.034			
MB-Low		0.009	0.058					0.027	0.041	
MB-High		0.231							0.273	0.239
Price-Middle		0.125	0.140					0.011	0.079	0.079
Price-High		0.093		0.176	0.081		0.045			
Price-Low	0.182			0.057	0.124	0.115	0.005	0.009		0.023
PE-Low 0–10		0.028	0.092	0.052					0.003	0.033
PE-Middle 10–17	0.036			0.177	0.270	0.22	0.001	0.005	0.058	
PE-High 25+	0.030		0.009		0.036		0.156	0.051		
Return-Negative		0.001		0.003	0.010		0.009			
Return-Low	0.030		0.206			0.033		0.090	0.187	0.332
Size-High		0.006		0.106	0.059	0.006	0.049			
Size-Middle	0.039					0.102		0.075	0.069	0.136
Size-Low		0.122	0.573				0.147		0.439	0.012

(*Continued*)

Table 5. (*Continued*).

	Ind (78,81%)	Org (44,96%)	Region-Central (72,41%)	Region-Foreign (33,33%)	Gender-Male (85,23%)	Wealth G2 (32,71%)	Wealth G3 (37,19%)	Freq G2 (66,71%)	Freq G3 (31,16%)	Experience G3 (42,06%)
Age-Middle	0.027			0.036	0.211	0.025	0.204			
Age-High		0.127	0.347	0.053				0.008	0.144	0.111
Age-Low	0.039					0.102		0.075	0.069	0.136
DE-Middle	0.026			0.026	0.061	0.016	0.035	0.003		
DE-High	0.071		0.300	0.037					0.485	0.050
DE-Low		0.447			0.082	0.332		0.343		0.327
ROA-High		0,016		0.098	0.046		0.041			
ROA-Low	0.022		0.075			0.008		0.057	0.148	0.116
ROE-High		0.006		0.106	0.059	0.006	0.049			
ROE-Low		0.128	0.136					0.124	0.230	0.233
ROE-Middle	0.636			0.518	0.571	0.008	0.470			

Note: From the logistic regression model, shown in the first column of the table, it can be concluded that good ROE has the strongest influence to the individual investors. Also, looking at the fourth column, we can conclude that foreign investors choose stocks with high DY and EPS, medium MB, high prices, medium PE, high Size and Age, high DE and high ROA and ROE.
Source: Authors' calculation.

Table 6. Results of clustering (stock portfolios).

Cluster	2.2	2.1	1.1	0	3.1
DY	Middle	High	Low	Low	High
EPS	Middle	High	Low	Low	Low
MB	Middle	Low	Low	Low	Low
Price	Middle	High	Low	Low	Low
PE	Low 0–10	Low 0–10	Middle 10–17	High 25+	Middle 10–17
Return	Negative	Negative	Negative	Negative	Low
Size	High	High	Middle	Middle	Low
Age	Middle	High	Low	Low	Middle
DE	Middle	Middle	High	Middle	Low
ROA	High	High	Low	Low	Low
ROE	High	High	Low	Low	Middle

Source: Authors' calculation.

are more recent. Balance sheet indicators of the companies selected by these investors are bad. All of this points to the fact that in the Montenegrin capital market institutional investors carefully analyse the traded stocks, while the individual investors rely mainly on available information, their experience and luck. The domestic investors from central region and foreign investors prefer low risk investments with relatively certain return. The male investors behave in similar manner. Also, it can be noticed that the wealthier investors, with stock ownership between EUR 20 and 100 thousand, are more careful in selection of stocks than those with wealth between EUR 5 and 20 thousand. Those who are wealthy select high price and high stock age, stocks with better investment returns and good balance sheet indicators. Those less wealthy select the low prices, middle size and most recent stocks, as well as the companies with poor balance sheet indicators. The investors with once a week or daily trading frequency have preferences for stocks with low DY and earnings, low MB, low investment return, middle price and size, low age as well as for the unindebted companies with low ROA and ROE. The investors who trade every day prefer older and indebted companies with small market capitalisation. Therefore in the Montenegrin stock market the investors with high trading frequency are speculators, i.e. those who count on earnings from fluctuation in prices. The last column in Table 5 indicates that the experienced investors in Montenegrin capital market (those who are in the market for over 10 years) are conservative because they prefer middle values. They don't carefully analyse the companies whose stock they buy, because the balance sheet indicators of these companies are poor.

With clustering we have divided stocks into five clusters (Table 6). Cluster 2.1 represents the stocks and companies with good characteristics which can be considered as low risk. Cluster 2.2 represents stocks with middle values which present risk-free, i.e. conservative selection. The clusters 0, 1.1 and 3.1 are clusters with risky stocks which are different in only a few characteristics. For cluster 0 it is noticeable that those are stocks with very high PE. The investors who decided for such stocks are ready to take risk and to pay much higher amounts than the stock earnings. It is obvious that the selection of such stocks is characteristic for speculators in the market. Cluster 3.1 consists of stocks with high DY and middle PE, low price, middle age and medially good balance sheet indicators (low DE and middle ROE). This medially risky cluster is the selection of less experienced investors who intentionally take risks. At the end, cluster 1.1 represents stocks with poor balance sheet indicators of the companies (high indebtedness and poor profits), which can be the selection of either speculators or inexperienced investors.

Table 7. Rules generated with the decision tree model.

Rule	Cluster	IndOrg	Region	Gender	Wealth	Frequency	Experience	Ind Conf
1*	2.2 (57.95%)	I	Central	M	G1, G2	G1		0.49
2	2.2 (57.95%)				G1, G2	G3		0.63
3	2.2 (57.95%)	I	Central		G3	G2		0.61
4	2.1 (73.63%)	I		M		G3		0.9
5	2.1 (73.63%)	O	Foreign			G2		0.81
6	1.1 (42.9%)	I		M	G1,G2	G2	G3	0.42
7	1.1 (42.9%)	O	North			G2		0.87
8	1.1 (42.9%)	O				G1	G2,G3	0.47
9	0 (43.69%)	I	South, Central		G3,G4		G2,G3	0.84
10	0 (43.69%)	O	Central			G2	G1,G2	0.41
11	3.1 (23.08%)		Foreign		G4		G1	0.54

Note: *For example, rule 1, we can read: individual male investors from central region with a total stock ownership below EUR 20 000 who rarely trade, prefer stock portfolio from cluster 2.2 (Table 6).
Source: Authors' calculation.

The rules obtained by the decision tree model (Table 7) define the dependencies of investors' features in the Montenegrin stock market according to these clusters. It can be seen from the table that the individual investors from central region of Montenegro, male, with stock ownership in the amount of under EUR 20,000, who rarely trade, are conservative in selection of stocks (cluster 2.2). This is also the selection of the investors with stock ownership in the amount of under EUR 20,000 and with daily trading frequency, as well as the wealthy individual investors (with total stock ownership in the amount between 20,000 and 100,000) from central region with once a week trading frequency. Risk-free stocks with good characteristics (cluster 2.1) in the Montenegrin market are selected by the male investors who trade frequently (every day) and by foreign organisations who trade once a week. Risky stocks with poor characteristics and company balance sheet indicators (cluster 1.1) are mainly traded by institutional investors from northern region with trading frequency of once a week. Speculative cluster (cluster 0) is the selection of individual investors from central and southern region, with wealth up to EUR 500,000, who are present in the market for more than 10 years, as well as the organisations from central region with trading frequency of once a week and presence in the market of under 10 years. Medially risky stock cluster (cluster 3.1) is preferred by inexperienced foreign investors with stock ownership in value between EUR 100,000 and 500,000.

5. Conclusion

This article analyses the influence of the investors' features, stock characteristics and balance sheet company indicators to the selection of the stock portfolio.

In this regard, we addressed three crucial questions: (1) how do stock characteristics affect the high and low preferences of investors? (2) how do stock characteristics affect

the trading preferences of investors having a specific feature? and (3) how do the features of investors affect their stock portfolio selections?

In most of the previous studies, that addressed the preferences of stock market investors, are used regression analysis. In these studies the adjusted coefficient of determination is low, indicating a low interdependency of factors. Because of these drawbacks to test the behaviour of stock market investors we used a data mining method. In order to answer the questions, we proposed three predictive classification models based on logistic regression, clustering and decision trees. By applying our method to the Montenegrin stock market, we found that there is a high interdependency between stock characteristics and high and low preferences of investors. We also detected some correlations between stock characteristics and individual features of investors. According to the third hypothesis we found that features of investors, namely the type of investor affects their stock portfolio selections. Hence, the empirical results verified the proposed method. Using the models, we identified the actual behavioural patterns of investors in the Montenegrin stock market, which may be useful for future trading strategies.

In general, we determined that in the Montenegrin stock market, the investors have high preferences for safe and stable stocks and companies, but also for those which are risky. The institutional investors usually make good and careful (risk-free) selection of stocks, while some individual investors have high preferences for risky stocks (speculators). The investors from a central region, foreign investors and male investors prefer risk-free investments expecting return on them. Wealthy investors in the Montenegrin stock market are not prone to risk. The experienced investors are conservative in selection of stocks, but they do not analyse the balance sheets of the companies carefully. Individual investors from central region, who are male, less wealthy, and trade rarely, as well as some of the wealthy investors, who trade frequently, make conservative selection of stocks. The best selection of stocks (risk-free, stable stocks with guaranteed return on investment, stocks of companies with good balance sheet indicators) have male individual investors from central region who trade frequently, as well as the foreign organisations with weekly trading frequency. Risky stocks are traded by speculators from central and southern region, who are wealthy and experienced. It is interesting that some of the institutional investors from central region who trade once a week are speculators in the market. Due to lack of experience, even some of the institutional investors from northern region as well as some of the wealthy investors trade risky.

This analysis may be useful not only for future investors, but also for brokers, stock exchange, as well as for all other interested parties in the Montenegrin stock market.

In future research, it could be possible to test some other data mining techniques and determine if they provide better results (better classification performance) over the same data-set. Also, it would be interesting to make a similar analysis for 2007, when the maximum number of trades was reached and when the fluctuations were much higher. This analysis could be compared with the results obtained here.

Notes

1. Net profit is reduced by total amount of dividend of the preferred stocks. The companies analysed in this article did not have preferred stocks.
2. Personal information such as the registration number, family and first name or the name of institution, were not provided by the CDA, since this type of information is confidential.
3. This information, i.e. corresponding number of years from opening of the first stock account we have used to define the experience of the investor.

4. The wealth is the information on total amount on all stock accounts of an investor. Since this information is confidential, the CDA provided us only with information on investor affiliation to one of the five groups which represent different levels of wealth we defined.
5. Accuracy is a measure of evaluation of classification models. Accuracy alone is not sufficient to represent the quality of prediction because it will yield misleading results if the data-set is unbalanced (that is, when the number of samples in different classes vary greatly). For classification problems, there are a couple of measures (see Appendix).

References

Aggarwal, R., Klapper, L., & Wysocki, P. D. (2005). Portfolio preferences of foreign institutional investors. *Journal of Banking & Finance, 29*, 2919–2946.

Barber, B. M., & Odean, T. (2001). Boys will be boys: Gender, overconfidence, and common stock investment. *Quarterly Journal of Economics, 116*, 261–292.

Barber, B. M., & Odean, T. (2011). The behavior of individual investors. *Handbook of the Economics of Finance, 2*, 1533–1570.

Barber, B. M., Lee, Y. T., Liu, Y. J., & Odean, T. (2014). The cross-section of speculator skill: Evidence from day trading. *Journal of Financial Markets, 18*, 1–24.

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and regression trees*. London: Chapman & Hall/CRC Press.

Campbell, J. Y., Ramadorai, T., & Ranish, B. (2013). Getting better: Learning to invest in an emerging stock market. Available at SSRN, 2176222. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2176222

Chen, G. M., Kim, K. A., Nofsinger, J. R., & Rui, O. M. (2004, January). *Behavior and performance of emerging market investors: Evidence from China*. Unpublished Washington State University Working Paper.

Covrig, V., Lau, S. T., & Ng, L. (2006). Do domestic and foreign fund managers have similar preferences for stock characteristics? A cross-country analysis. *Journal of International Business Studies, 37*, 407–429.

Dahlquist, M., & Robertsson, G. (2001). Direct foreign ownership, institutional investors, and firm characteristics. *Journal of Financial Economics, 59*, 413–440.

Delibasic, B., Jovanovic, M., Vukicevic, M., Suknovic, M., & Obradovic, Z. (2011). Component-based decision trees for classification. *Intelligent Data Analysis, 15*, 671–693.

Elkinawy, S. (2005). Mutual fund preferences for Latin American equities surrounding financial crises. *Emerging Markets Review, 6*, 211–237.

Grinblatt, M., & Keloharju, M. (2001). What makes investors trade? *The Journal of Finance, 56*, 589–616.

Grinblatt, M., & Keloharju, M. (2009). Sensation seeking, overconfidence, and trading activity. *The Journal of Finance, 64*, 549–578.

Grinblatt, M., Keloharju, M., & Linnainmaa, J. (2011). IQ and stock market participation. *The Journal of Finance, 66*, 2121–2164.

Hartigan, J. A. (1975). *Clustering algorithms*. New York, NY: Wiley.

Jianakoplos, N. A., & Bernasek, A. (1998). Are women more risk averse? *Economic Inquiry, 36*, 620–630.

Kass, G. V. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied Statistics,* 119–127.

Keerthi, S. S., Duan, K. B., Shevade, S. K., & Poo, A. N. (2005). A fast dual algorithm for kernel logistic regression. *Machine Learning, 61*, 151–165.

Loh, W. Y., & Shih, Y. S. (1997). Split selection methods for classification trees. *Statistica Sinica, 7*, 815–840.

Ng, L., & Wu, F. (2006). Revealed stock preferences of individual investors: Evidence from Chinese equity markets. *Pacific-Basin Finance Journal, 14*, 175–192.

Peress, J. (2004). Wealth, information acquisition, and portfolio choice. *Review of Financial Studies, 17*, 879–914.

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning, 1*, 81–106.

Quinlan, J. R. (1993). *C4. 5: Programs for machine learning* (Vol. 1). San Francisco, CA: Morgan Kaufmann.

Ramsey, J. B. (1969). Tests for specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society, 1*, 350–371.
Rapid Miner User Manual. Retrieved from http://www.rapidminer.com
Tsai, C. F., Lin, Y. C., & Wang, Y. T. (2009). Discovering stock trading preferences by self-organizing maps and decision trees. *International Journal on Artificial Intelligence Tools, 18*, 603–611.
WhiBo User Manual. Retrieved from http://www.whibo.fon.bg.ac.rs

Appendix 1. Performance of classification and clustering models

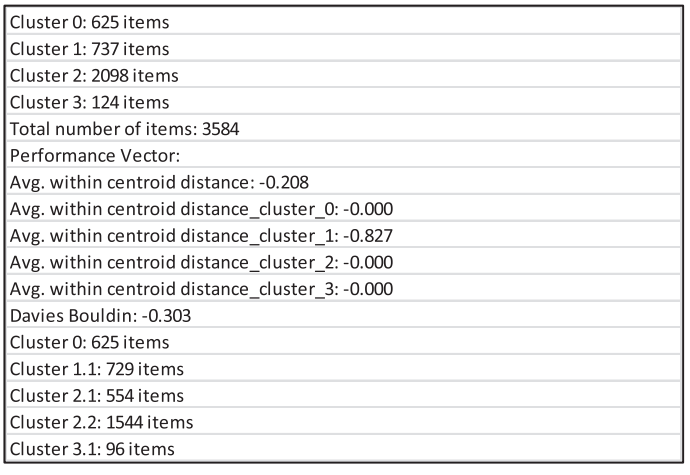

Figure A1. Cluster model and performance.
Note: Avg. within centroid distance is the average within cluster distance is calculated by averaging the distance between the centroid and all examples of a cluster (smaller value is preferred). The algorithms that produce clusters with low intra-cluster distances and high inter-cluster distances will have a low Davies–Bouldin index. The clustering algorithm that produces a collection of clusters with the smallest Davies–Bouldin index is considered the best algorithm based on this criterion.

Table A1. Performance of the model of logistic regression.

Accuracy: 61.56%			
	true H	true L	Class precision
pred. H	631	394	61.56%
pred. L	15	24	61.54%
Class recall	97.68%	5.74%	

Note: The table represents a Confusion matrix. Confusion matrix is a specific table layout that allows visualisation of the performance of a classification model. Each column of the matrix represents the instances in a predicted class, while each row represents the instances in an actual class. Accuracy is relative number of correctly classified examples or in other words percentage of correct predictions. Class precision is relative number of correctly as positive classified examples among all examples classified as positive. Class recall specifies the relative number of correctly as positive classified examples among all positive examples.
Source: Authors' calculation.

Table A2. Differences in performance of decision tree algorithms generated with component design.

Performance	Max	Min	Max-Min
Accuracy	53.38%	47.52%	5.86%
Max tree depth	5	5	0
WATD*	3.9	3.0	0.9
Total nodes	141	20	121
Total leaves	87	9	78
Execution time	00:02.45	00:00.37	00:02.08

Note: *Weighted average tree depth (WATD) represents the average length of the path in the tree necessary for classification of one example.
Source: Authors' calculation.

Table A3. Performance of the decision tree model generated with the optimal algorithm.

	true cluster_2.2	true cluster_2.1	true cluster_1.1	true cluster_0	true cluster_3.1	class precision
pred. cluster_2.2	594	154	157	98	22	57.95%
pred. cluster_2.1	16	67	1	7	0	73.63%
pred. cluster_1.1	95	17	151	77	12	42.90%
pred. cluster_0	64	39	51	128	11	43.69%
pred. cluster_3.1	3	0	5	2	3	23.08%
class recall	76.94%	24.19%	41.37%	41.03%	6.25%	Acc. 53.16%

Note: Performance of the tree complexity: Max tree depth = 5; WATD = 3.0; Total nodes = 57; Total leaves = 34.
Source: Authors' calculation.