ASSESSMENT OF FACTORS AFFECTING FINANCIAL PERFORMANCE OF TOURISM COMPANIES IN *BIST* BY MEANS OF DATA MINING ALGORITHMS IN FINANCIAL RATIOS

Duygu Arslantürk Çöllü, Assistant Professor[1]
Ayaz Yusuf Altin, Assistant Professor[2]
Leyla Akgün, Assistant Professor[3]
Ecevit Eyduran, Professor[4]

DOI: https://doi.org/10.31410/tmt.2018.425

Abstract: *The present study was conducted on seven tourism companies in BIST Tourism Index in order to describe continuous financial factors which affect net profit margin (NPM) as a continuous response variable through CART (Classification and Regression Tree), CHAID (Chi-Square Automatic Interaction Detector), Exhaustive CHAID and MARS (Multivariate Adaptive Regression Splines) algorithms. In the present study, the data of these companies from the period 2011-2017 were evaluated. Predictive performances of CART, CHAID, Exhaustive CHAID and MARS in predicting NPM were measured based on model goodness of fit criteria, viz. r (Pearson correlation coefficient between actual and predicted values in NPM), coefficient of determination (R^2), adjusted coefficient of determination (Adj.R^2), standard deviation ratio (SD_{RATIO}), root of mean square error (RMSE), global relative approximation error (RAE), mean absolute deviation (MAD), Akaike's information criterion (AIC) and the corrected Akaike's information criterion (AICc). In the study, financial factors used in the prediction of NPM were current ratio (CR), acid-test ratio (ACTR), asset turnover ratio (ASTR), accounts receivable turnover ratio (ACRTR), equity turnover ratio (EQTR), short term liabilities to total assets ratio (SHTLTAR), long term liabilities to total assets ratio (LOTLTAR), total assets to equity ratio (TOAER), long term liabilities to equity ratio (LOLER) and total debt to total assets ratio (TODTAR) as predictors. In the prediction of the NPM and the description of the influential financial factors influencing the NPM, the highest predictive accuracy was obtained by MARS algorithm (r=0.980) and the statistically significant order was found as MARS (r=0.980) > Exhaustive CHAID (r=0.915) = CART (r=0.873) = CHAID (r=0.868) algorithms.*

In conclusion, the achieved results indicated that, i) the regression tree diagram constructed by Exhaustive CHAID algorithm displayed that tourism companies with LOTLTAR < 0.3715 and EQTR < 0.0311 had the highest average NPM of 2.778, ii) CART tree-based algorithm showed that the companies with EQTR > - 0.2125 and ASTR < 0.0246 had the highest average NPM of 4.226, iii) the diagram of CHAID tree-based algorithm revealed that the companies with TODTAR < 0.6145 and EQTR < 0.0311 had the highest NPM with the average of 2.778. It is recommendable that data mining algorithms capture optimal cut-off values of influential factors, which may ensure the highest NPM values.

Keywords: *Tourism companies, financial ratios, data mining algorithms, MARS, BIST*

[1] Iğdır University, Faculty of Economics and Administrative Sciences, Department of Business Administration, Iğdır, Turkey

[2] Iğdır University, Faculty of Economics and Administrative Sciences, Department of Business Administration, Iğdır, Turkey

[3] Iğdır University, Faculty of Economics and Administrative Sciences, Department of Business Administration, Iğdır, Turkey

[4] Iğdır University, Faculty of Economics and Administrative Sciences, Department of Business Administration, Iğdır, Turkey

1. INTRODUCTION

Tourism is perceived as a very important mean of economic development in terms of GDP, employment ratio, financing of current budget deficits, international foreign exchange revenues both in developing and developed countries worldwide [1]- [2]. The tourism sector encompasses many areas such as accommodation, food and beverage services, logistics, promotion and marketing. At the same time, tourism is one of the most rapidly developing and expanding sectors in the world economy which constitutes the movement of people and capital between countries [3]- [4]. Since the 1980s, Turkish tourism has begun to show the development and increased its growth rate by adapting to rapidly changing competitive environment from year to year. Tourism revenue has increased from 326 million dollars to 26,3 billion dollars within the last 37 years in Turkey, a tourism center thanks to its geological location and natural resources according to data of World Tourism Organization (UNWTO). Tourism as an important investment area is expected to be among the locomotive sectors of the changing world in future [5]. This expectation makes it necessary for tourism companies to increase their financial performance for having more competitive power in the worldwide.

Financial performance means a monetary assessment of the results of the strategies and operations determined by a company, and the results are expected to be reflected in the returns from the investments and assets of a company. This allows financial performance to be regarded as a subjective criterion of how well firms can use their assets; thereby to generate revenue, to provide a general measure of a firm's overall financial health over a given period of time, and to compare their competitive advantages with similar companies across the same industry or to compare sectors in a cluster [6]. Financial performance analysis, also referred to as financial structure analysis, is used to assess the stability and profitability of a business. The tourism companies operate in a sector where uncertainty, exchange rate risk and fixed capital investments are high and having difficulties in finding funds. Determining the most influential factors that affect the financial performance becomes a necessity for tourism companies in order to utilize their resources in the most effective way and to provide maximum returns with their available resources [7]. Therefore, the present study is conducted to ascertain the factors affecting the financial performance of the tourism companies included in the BIST Tourism Index. Determining the factors which adversely affect their financial performance allows tourism companies to take necessary precautions with the objective to sustain growing their profitability.

There are many methods i.e. vertical percentage analysis, ratio analysis, comparative table analysis and trend percentile analysis in order to measure the financial performance [8]. Among them, the ratio analysis is the most commonly used technique. The ratio analysis, which aims to reveal the relationship between the two values in the financial statement tables, provides easier interpretations of the financial data and gives an idea to the managers on the financial health of their companies [9]. The ratio analysis can be divided into a number of sub-headings such as liquidity ratios, activity ratios, capital structure ratios, profitability ratios, market rates, etc. and among them, the most frequently used groups are profitability ratios which are considered to present more detailed and healthy information about financial structures of the companies.

In the direction of the aim of this study, Net Profit Margin (NPM) as a type of profitability ratios is used as a continuous response variable in the data mining applications, because NPM reflects both all the losses of the company and the final outcomes of an activity while evaluating the effectiveness of any activity [10]. In the agreement with the present study, [11]- [12]- [13]- [14] employed NPM as the output (response) variable within the scope of the data envelopment analysis (DEA). Some selected liquidity, activity and capital structure rates were included as predictors as also described in material and methods section of the study.

Financial analysis, carried out with the aid of financial statements of firms, requires the evaluation of many values in these tables. However, overabundance of financial statement data sets prevents making accurate interpretations and a large number of interactions between them cause the complexity of performing financial performance for managers and supervisors of long-established companies. In this regard, use of data mining algorithms is adopted in this context to analyze data obtained from large databases and to capture the influential predictors affecting a response variable in order to make successful management decisions [15]. Data mining algorithms, which have many types of classification, regression, clustering, visualization, decision trees, association rules, neural networks, support vector machine, etc., are used in the study of decision tree analysis. It is aimed to predict the behavior of the response variable by using algorithms. Among those, tree-based algorithms i.e. CART and CHAID allow researchers to easily make interpretations of the obtained visual results and to obtain cut-off values of the significant predictors, which permit researchers to obtain optimum results of the response variable. MARS is a powerful data mining algorithm that reveals the complex high dimensional relationships between sets of responses and predictors, without needing distributional assumptions of the handled variables. To our best knowledge, there is no information about using CART, CHAID, Exhaustive CHAID and especially MARS algorithms to measure financial performances in tourism companies. In this study where quantitative analysis methods are applied, the tree-based algorithms (CART, CHAID, Exhaustive CHAID) along with MARS algorithm are therefore used to determine the effect of liquidity, activity and capital structure rates on financial performance and also the methods are compared with each other. This study was conducted by using the financial statements of the seven companies registered in BIST TOURISM index for the years 2009-2017.

This study provides significant contributions to the literature in two ways such as methodology and application. Within a new approach to reveal financial structure of tourism companies, the study aims to find the best one among the data mining algorithms used for statistically analyzing the data of tourism companies. The study consists of four parts. At the first part, a literature review is conducted to determine the ratios which have possible impact on the NPM and to examine the algorithms used to measure financial performance of tourism companies. Then at the methodology part, the data structure, analysis methods and algorithms are mentioned. The CHAID, Exhaustive CHAID, CART and MARS algorithms are used in the study. After performing the analysis, comparison of the algorithms in their predictive accuracy are demonstrated for tourism companies. At the end, the suggestions are made in the direction of the obtained results.

2. LITERATURE REVIEW

There are many studies that have been conducted in finance literature to evaluate the financial performance of tourism companies. However, there is no information about using data mining algorithms such as CART, CHAID, Exhaustive CHAID and MARS in order to reveal the financial performance of the tourism companies in literature. In this context, [16] determined that operating and profitability rates are the most important ratios in their studies to determine the financial ratios commonly used in the hospitality sector and to determine importance of these financial ratios at financial decision making. [9] examined the club sector which is one of the branches of the tourism sector by the help of ratio analysis and they concluded that the percentage of personnel costs, the percentage of the costs of food and beverages sold, the current ratio and the debt-equity ratios are the most commonly used rates that affect financial performance. [17] measured financial performance of the hotel sector in their study and they performed a ratio analysis for the accommodation, restaurant, airline and entertainment sectors that form the branches of the hotel sector in the period of 1997-2001. It was found that the profitability rates of net profit margin, return on equity and return on assets were lower for companies in the entertainment sector compared to other sectors. It has been determined that this sector is the best in terms of profitability among the other sectors.

[18] analyzed financial performances of hotels and restaurants in the tourism sector on the basis of 15 financial ratios. The results of the study conducted to compare the hotels and restaurants in terms of financial performance indicated that the hotels had higher performance than restaurants in terms of liquidity and activity rates. When they are evaluated in terms of payment power ratios, it was found that restaurants were more successful than hotels according to paying their long-term debts. No significant difference in profitability rates was found between hotels and restaurants.

[19] investigated usage of financial ratios in terms of the frequency of use and importance by 191 five-star hotel financial/accounting managers in the Mediterranean region. The results of the study which covers the February-July period of 2012, showed that the most important commonly used ratios are activity ratios. In addition, cash ratio, borrowing ratio, receivables turnover rate, operating profitability and net profit margin ratios are the most important ratios.

[12] evaluated the financial performances of DEA in 2010 by using the financial ratios of 9 tourism companies traded on the BIST. In the study, current ratio, total debt total assets, tangible fixed assets/continuous capital are used as input variables and return on equity, asset profitability, net profit margin, operating expenses + sold goods cost/sales ratio are considered as output variables. As a result, the performances of AYCES, FVORI, MAALT, NTTUR, PTKENT and TEKTU companies were found high while the financial performances of MARTI, METUR and UTYP enterprises were found to be insufficient.

[20] measured the financial performance of nine companies listed in Borsa İstanbul with the help of financial ratios for the period 2008-2012. In the study, 17 financial ratios were used to evaluate the financial performance in the scope of liquidity, leverage, profitability and activity indicators through the gray relational analysis (GRA) method. Findings revealed that the most effective ratio group in financial performance measurement was leverage ratio, followed by profitability, activity and liquidity ratios, respectively.

[21] evaluated the financial data of the two major hotel chains in Pakistan for the period 2011-2012. The findings showed that the firms had difficulties in terms of both the current ratio and the acid test rate and they could not provide sufficient assets in terms of debt. However, it was found that other liquidity ratios were acceptable. It was also found that the profitability ratios determined by the net profit margin and return on assets were satisfactory.

[13] used the DEA and the Malmquist Total Factor Productivity (MTFV) Index methods in their study in order to investigate the financial activities of the 9 tourism enterprises in the BIST 100 for the period 2009-2013. Current ratio, financial leverage ratio (total debt/total asset), tangible fixed assets/equity ratio were considered as predictors (input variables), while equity, profitabil-

ity, net profit margin and (operating expenses + cost of sales)/sales ratios were taken into consideration as response (output) variables in the previous study. As a result, full efficacy scores of AVTUR, MAALT and NTTUR companies were available for the period 2009-2013, while UTPYA and ULAS companies were not able to show full efficacy scores during this period.

[22] evaluated the financial performances of 10 tourism companies listed in Borsa İstanbul in 2015 by using the GRA method. They aimed to evaluate the financial performance of the companies available in the tourism sector with the help of financial ratios. The results of the earlier study illustrated that the leverage ratio (69.75%) was the most important one among financial ratios used to measure the financial performance of tourism companies.

[23] analyzed the financial performance of 13 tourism companies included in Borsa İstanbul through TOPSIS method in the 2011-2015 period. In the former study, 12 ratios were employed under the headings of liquidity ratios, activity rates, profitability ratios and financial structure ratios. Companies have to give importance to solvency ratios and activity ratios and their financial structures should be spread to an appropriate debt equity base.

3.METHODOLOGY

Company managers have difficulties in performing financial performance and making successful decisions because of availability of many ratios related to each other as a part of financial performance. The growth of the databases complicates finding, extracting and analyzing data for decision makers who try to generate only required data for their companies. These difficulties are required to be revealed different methodologies for finding, extracting and analyzing financial data. Data mining algorithms are developed in this context to find invaluable information from large databases and the main goal of data mining techniques is to extract knowledge in order to make successful management decisions [15].

Data Structure

In the present study, there are only seven tourism firms listed in BIST Tourism Index. These are:

- AVTUR: Avrasya Petrol ve Turistik Tslr Ytrmr AŞ
- MAALT: Marmaris Altinyunus Turistik Tesisler AŞ
- MARTI: Marti Otel Isletmeleri AŞ
- METUR: Metemtur Otelcilik ve Turizm Islet. AŞ
- TEKTU: Tek Art Insaat Ticaret Tur San ve Yat AŞ
- ULAS: Ulaslar Turizm Yatirimlari AŞ
- UTPYA: Utopya Turizm Insaat Isletmecilik AŞ

The data were collected from the site kap.org.tr. for the period of 2009-2017. Rows including missing values were excluded from the data set.

Variables in the study

Financial performance gives detailed information about business organization's revenues, expenses and net income. In other words, it shows the current financial status of a company. Financial performance is important to understanding if a business is profitable and, if not, where to make needed changes. Financial performance allows a company's top management to identify the influential factors affecting net profit margin, return on investment, return on assets, value added, etc. The most frequently used technique in the applications of financial performance analysis is the ratio analysis.

While performing the performance analysis of the companies, the profitability ratios are the most emphasized ratios because it is accepted that the profitability ratios are more informative about the financial structures of the companies among other ratio analysis. Therefore, NPM as a type of profitability ratios is the response variable in the current analysis, because NPM considers both all the losses of the company and the final outcomes of an activity while evaluating the effectiveness of any activity [10]. Indeed, in literature [11]- [12]- [13]- [14] used NPM as the response variable in their DEA.

There are 10 independent variables categorized into three group such as liquidity ratios, activity ratios and capital structure ratios in the models. The groups of independent variables are listed as follows:

Liquidity Ratios:
- Current Ratio (CR) = Current assets/ Current liabilities
- Acid-Test Ratio (ACTR) = (Cash equivalents + marketable securities + accounts receivables)/ current liabilities

Activity Ratios:
- Asset Turnover Ratio (ASTR) = Net Sales / Total Assets
- Accounts Receivable Turnover Ratio (ACRTR)= Net Credit Sales/ Accounts Receivable
- Equity Turnover (EQTR) = Net Sales /Equity

Capital Structure Ratios:
- Short Term Liabilities to Total Asset Ratio (SHTLTAR)
- Long Term Liabilities to Total Asset Ratio (LOTLTAR)
- Total Asset to Equity Ratio (TOAER)
- Long Term Liabilities to Equity Ratio (LOLER)
- Total Debt to Total Asset Ratio (TODTAR)

Method of Analyses

To predict NPM from some financial ratios, tree-based data mining algorithms (i.e. CART, CHAID and Exhaustive CHAID) along with MARS algorithm were examined comparatively in the present work. Among those, MARS algorithm allows analysts to obtain a prediction equation for NPM as a response variable.

CART is a tree-based data mining algorithm that recursively divides a subset into 2 small subsets by starting a root node at the top of the regression tree until homogenous subsets are obtained as soon as possible in the regression tree structure with the aim of ensuring minimum error variance. However, CHAID algorithms create a regression tree that generates recursively multi-splits by starting a root node, until reaching up to maximum variance among subsets in the tree structure [24]- [25].

CHAID algorithm just handles nominal or ordinal categorical predictors. Due to this reason, continuous predictors are transformed into ordinal predictors before employing the following algorithm. For a given set of break points $a_1,\ a_2,...,a_{K-1}$ (in ascending order), a recognized x is mapped into category $C(x)$ herein below:

$$C(x)=\begin{cases}1 & x\le a_1\\ k+1 & a_k<\chi\le a_{k+1},\quad k=1,...,K-2\\ K & a_{K-1}<\chi\end{cases}$$

When K is the preferred number of bins, for the computation of the break points x_i frequency weights are unified in calculating the ranks. In the case of being tied, the average rank is employed. The rank and the matching values in the ascending order can be stated as $\left\{r_{(i)},x_{(i)}\right\}_{i=1}^{n}$.

For k = 0 to (K–1), set $I_k=\left\{i:\left[r_{(i)}\dfrac{K}{N_f+1}\right]=k\right\}$ where *(x)* shows the floor integer of x.

If I_k is not empty, $i_k=\max\left\{i:i\in I_k\right\}$. Adjustments of the break points are made by becoming equal to the x values corresponding to the i_k, excluding the largest [26].

CHAID and Exhaustive CHAID algorithms permit multiple splits of a node and have three stages i.e. splitting, merging and stopping. Splitting and stopping stages are the same in both CHAID algorithms. Merging stage employs an exhaustive procedure for merging any similar pair, until only a single pair obtains. Bonferroni adjustment was made for both CHAID algorithms in order to calculate Adjusted P values of F values. In the IBM SPSS 23 software, the tree based CHAID algorithms, with an automatically pruning process to exclude needless nodes from the regression tree structure, use F significance test for a continuous response variable.

Multivariate Adaptive Regression Splines Algorithm

MARS algorithm proposed by [27] is a non-parametric regression analysis technique used to explain the high dimensional functional relationship between sets of predictors and responses, as well as it is a good alternative to response surface analysis. It is a nonparametric statistical method where the training data sets are split into separate piecewise linear segments (splines) of various gradients (slopes). The splines are connected smoothly to each other. The basis functions describing piecewise curves powerfully reveal linear, non-linear and interaction effects associated with the influential predictors. The point connections of the pieces that derive in the algorithm are knots. The probable knots are located randomly inside the range limit of each predictor. As part of a stepwise procedure, the own basis functions are generated by allowing for all probable knots and interactions among the influential predictors. In order to achieve the pairs of the basis functions, the forward procedure captures the probable knots at random within the range boundary of each predictor. The MARS model generated at every stage takes for the knots and their pairs of basis functions with the smallest GCV error. Until the complicated MARS model is derived, the involvement of the basis functions into the model goes on. The non-significant terms are excluded from the MARS prediction model by specifying the backward procedure in the MARS algorithm [28].

The MARS model can be described as follows:

$$\hat{y} = \beta_0 + \sum_{m=1}^{M} \beta_m \prod_{k=1}^{K_m} h_{km} \left(X_{v(k,m)}\right) \quad (1)$$

Where

$\hat{y}$ is the predicted score of the response variable (NPM), β_0 is an intercept, β_m are coefficients of basis functions, $h_{km}\left(X_{v(k,m)}\right)$ is the basis function, where $v(k,m)$ is an index of the predictor for the m^{th} component of the kth product, K_m is the parameter limiting the order of interaction [29].

After building the most complex MARS model, the basis functions that did not provide the contribution to the model fitting performance are excluded from the model in the pruning process on the basis of the generalized cross-validation error (GCV) [30]:

$$GCV(\lambda) = \frac{\sum_{i=1}^{n} \left(y_i - y_p\right)^2}{\left[1 - \frac{M(\lambda)}{n}\right]^2} \quad (2)$$

Where:

n is the number of training cases, y_i is the observed value of a response variable (NPM), y_{ip} is the predicted value of a response variable (NPM), $M(\lambda)$ is a penalty function for the complexity of the model with λ terms.

Only in the case of penalty=-1, the GCV is equal to RSS/n (according to the manual for the R "earth" package and the Friedman's paper on MARS - Friedman, Jerome H. „Multivariate adaptive regression splines." - The annals of statistics (1991): 1-67). But this penalty is necessary for preventing over-training (or overfitting) of the MARS model to the training data. If you set the penalty in the earth package to -1, the backward procedure of pruning (which is necessary to find the MARS model that is not too flexible and has good generalization abilities) is practically turned off. In this regard, it must be specified as penalty = 2, 3 and 4. However, a good solution by significantly increasing nk (number of maximum terms) and specifying a small nprune (number of the terms desired by a researcher) together with penalty= -1 can be also obtained [31].

Goodness of Fit Criteria

Goodness of fit criteria for comparing predictive performances of the examined algorithms are indicated below:

1. Pearson correlation coefficient (r) between predicted and observed values in the NPM,
2. Akaike Information Criterion (AIC) calculated as:

$$AIC = n \ln\left[\frac{1}{n}\sum_{i=1}^{n}(y_i - y_p)^2\right] + 2k \text{, if n/k>40} \quad (3)$$

or:

$$AIC_c = n \ln\left[\frac{1}{n}\sum_{i=1}^{n}(y_i - y_p)^2\right] + 2k + \frac{2k(k+1)}{n-k-1} \text{, otherwise} \quad (4)$$

3. Root-mean-square error (RMSE) expressed by the following formula:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - y_{ip}\right)^2} \tag{5}$$

4. Mean error (ME) presented by the following equation:

$$ME = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - y_{ip}\right) \tag{6}$$

5. Mean absolute deviation (MAD):

$$MAD = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - y_{ip}\right| \tag{7}$$

6. Standard deviation ratio (SD_{ratio}):

$$SD_{ratio} = s_m / s_d \tag{8}$$

7. Global relative approximation error (RAE):

$$RAE = \sqrt{\frac{\sum_{i=1}^{n}\left(y_i - y_{ip}\right)^2}{\sum_{i=1}^{n} y_i^2}} \tag{9}$$

8. Mean absolute percentage error (MAPE):

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - y_{ip}}{y_i}\right| \cdot 100 \tag{10}$$

where: n – the number of cases in a set, k – the number of model parameters (number of the selected terms), y_i – the observed NPM value, y_{ip} – the predicted value of a response variable (NPM), s_m – the standard deviation of model errors, s_d – the standard deviation of the response variable, NPM.

It is an essential aim to capture the ideal MARS predictive model that can provide the smallest GCV error. Smaller is better in goodness of fit criteria numbered 2-8, but greater is better in Pearson's correlation coefficient between observed and predicted values in NPM, a continuous response variable. Greater is better in square of the Pearson's correlation coefficient which is defined as coefficient of determination (R^2). Adjusted R^2 is described as follows

$$\text{Adjusted } R^2 = (1-((1- R^2)(n-1) / (n-k-1)) \tag{11}$$

Where n is sample size and k is number of the selected terms in the MARS model whereas, in the tree-based algorithms, k is number of significant variables entered into the tree structure. Analysis of the tree-based algorithms was performed using IBM SPSS 23 software [32]. Also, the earth package developed by [33] was used for statistical analysis of MARS algorithm (with interaction effect) in R software [34]. All the algorithms were analyzed based on ten cross-validation. See notes of [31] for more detailed information on the earth package.

Results

CART Tree-Based Algorithm

The regression tree structure constructed by CART data mining algorithm for predicting NPM is depicted in Figure 1. Very strongly Pearson's correlation coefficient between actual NPM and predicted NPM values was found as 0.873 (P<0.01). Only two financial ratios i.e. EQTR and ASTR were significant in explaining total variability in NPM. As seen from Figure 1, overall NPM average was -0.330. Node 0 at the top of the tree structure is called root node, and it was split into the smaller subgroups named Node 1 (the subgroup of the companies with EQTR < -0.2125) and Node 2 (the subgroup of the companies with EQTR > -0.2125) on the basis of EQTR as a significant predictor (-8.352 vs. -0.002 in NPM). Splitting process for Node 1 was stopped at the first depth of the generated tree structure. In this regard, Node 1 is called a terminal node.

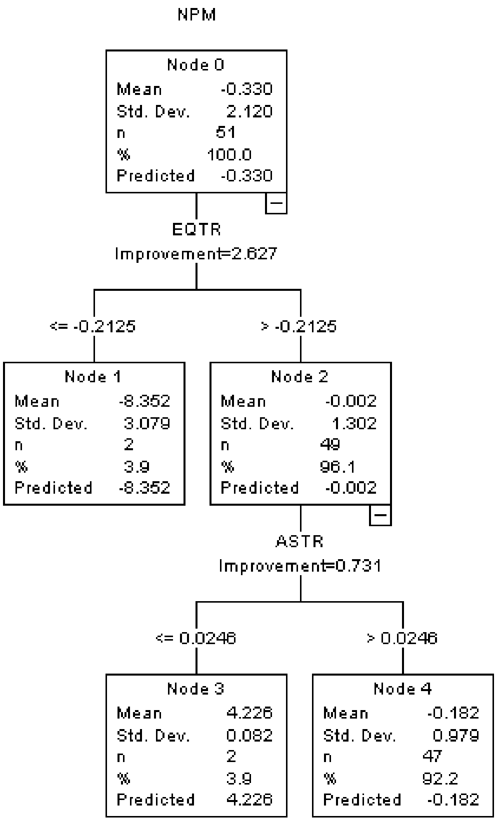

Figure 1: Regression Tree Structure of CART Algorithm

Node 2 was divided by means of ASTR predictor into the smaller subgroups i.e. Node 3 (the subgroup of the companies with EQTR > -0.2125 and ASTR < 0.0246) and Node 4 (the subgroup of the companies with EQTR > -0.2125 and ASTR > 0.0246) (4.226 vs. -0.182 in NPM). The best ideal result was obtained by Node 3. As also reported above, two cut-off values (-0.2125 EQTR and 0.0246 ASTR) could provide remarkable clues for financial researchers.

CHAID Tree-Based Algorithm

The visual structure of CHAID tree-based algorithm used for the prediction of NPM as a response variable is depicted in Figure 2. The predicted NPM values was very strongly correlated with the actual NPM values within the scope of the BIST Tourism companies (r=0.868, P<0.01).

Node 0, at the top of the constructed tree structure, gave overall average of -0.330 in NPM. Influential financial predictors i.e. TODTAR, EQTR and TOAER were determined (Figure 2) for the CHAID tree-based algorithm. In the first depth of the tree structure, Node 0 was divided into two smaller subgroups i.e. Node 1 (the subgroup of the companies with TODTAR < 0.6145) and Node 2 (the subgroup of the companies with TODTAR > 0.6145) according to TODTAR (0.291 vs. -1.819 in NPM). In other words, Node 1 was a higher value in NPM compared with Node 2. Node 1 was split into Node 3 (the subgroup of the companies with TODTAR < 0.6145 and EQTR < 0.0311) and Node 4 (the subgroup of the companies with TODTAR < 0.6145 and EQTR > 0.0311) according to EQTR (2.778 vs. 0.065 in NPM). As also seen from Figure 2, the highest NPM average was achieved by Node 3.

Node 4 was partitioned into the smaller subgroups named Node 7 (the subgroup of the companies with TODTAR < 0.1462 and EQTR > 0.0311) and Node 8 (the subgroup of the companies with 0.1462 < TODTAR < 0.6145 and EQTR > 0.0311) according to TODTAR again (0.443 vs. -0.213 in NPM).

Node 2 was branched into Node 5 (the subgroup of the companies with TODTAR > 0.6145 and TOAER < 1.0424) and Node 6 (the subgroup of the companies with TODTAR > 0.6145 and TOAER > 1.0424), at the second tree depth, according to TOAER (-8.352 vs. -0.814 in NPM). Node 5 produced the lowest NPM average (-8.352).

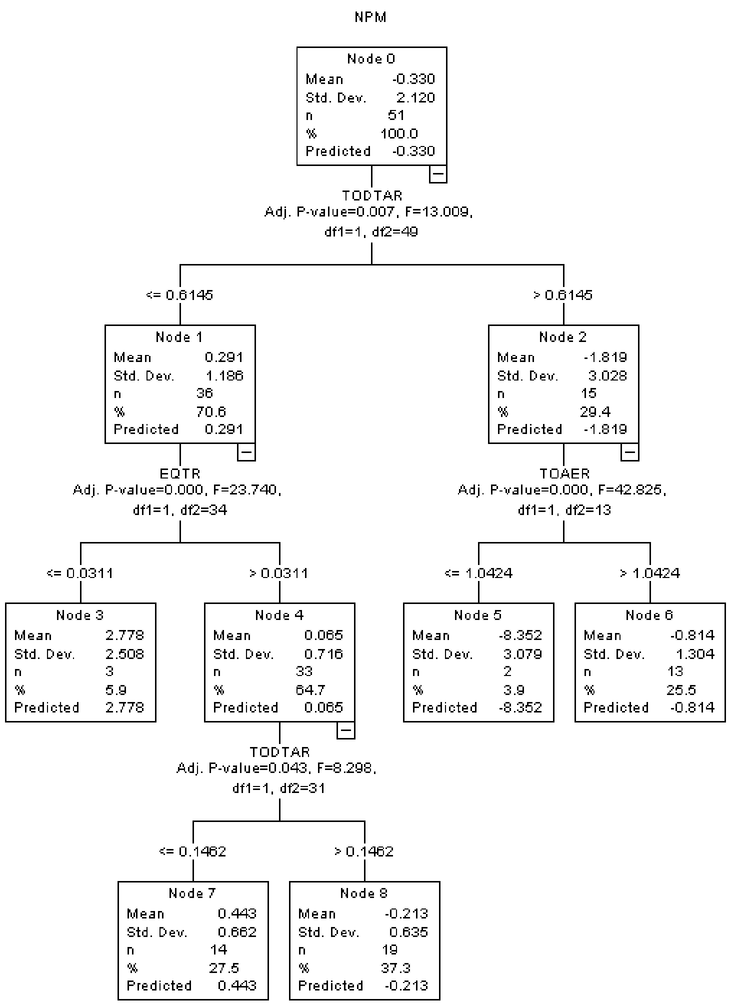

Figure 2: Regression Tree constructed by CHAID Algorithm

Exhaustive CHAID Tree-Based Algorithm

In the prediction of NPM as a response variable, the Exhaustive CHAID tree diagram is depicted in Figure 3. Very strongly Pearson correlation coefficient between the actual NPM values and the NPM values predicted by the tree-based algorithm was 0.915 (P<0.01). Among the tree-based algorithms, the Exhaustive CHAID had a bit better predictive performance; however, MARS outperformed them. Node 0 was split into the smaller subgroups i.e. Node 1 (the subgroup of the companies with LOTLTAR < 0.3715) and Node 2 (the subgroup of the companies with LOTLTAR > 0.3715) at the first tree depth according to LOTLTAR (0.212 vs. -2.549 in NPM). Most of the nodes derived from Node 1 had a positive NPM average, but all the nodes obtained from Node 2 had negative NPM averages. A cut-off value of 0.3715 for LOTLTAR might be important for next studies.

Node 1 was divided by EQTR into two smaller subgroups i.e. Node 3 (the subgroup of the companies with LOTLTAR < 0.3715 and EQTR < 0.0311) and Node 4 (the subgroup of the companies with LOTLTAR < 0.3715 and EQTR > 0.0311) at the second tree depth (2.778 vs. 0.009 in NPM). The greatest NPM average was obtained from Node 3 as a terminal node. However, Node 4 was re-branched into three smaller subgroups i.e. Node 7 (the subgroup of the companies with LOTLTAR < 0.3715, EQTR > 0.0311 and CR < 0.7708), Node 8 (the subgroup of the companies with LOTLTAR < 0.3715, EQTR > 0.0311 and 0.7708 < CR < 10.1824) and Node 9 (the subgroup of the companies with LOTLTAR < 0.3715, EQTR > 0.0311 and CR > 10.1824) by means of CR at the third tree depth. The corresponding NPM values were -0.518, 0.202 and 0.887.

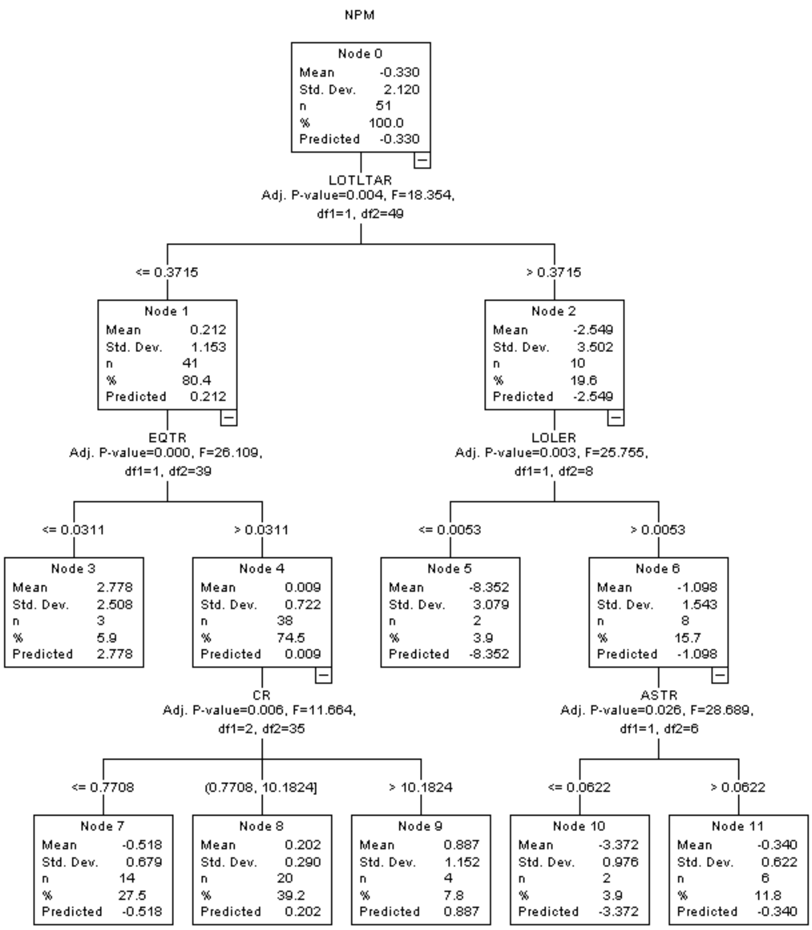

Figure 3: Regression tree constructed by Exhaustive CHAID Algorithm

Node 2 was split by means of LOLER into two smaller subgroups i.e. Node 5 (the subgroup of the companies with LOTLTAR > 0.3715 and LOLER < 0.0053) and Node 6 (the subgroup of the companies with LOTLTAR > 0.3715 and LOLER > 0.0053) at the second tree depth (-8.352 vs. -1.098 in NPM). As a terminal node, Node 5 gave the lowest NPM average (-8.352) when compared with other Nodes from Exhaustive CHAID algorithm. Afterwards, Node 6 was divided into two smaller subgroups i.e. Node 10 (the subgroup of the companies with LOTLTAR > 0.3715, LOLER > 0.0053 and ASTR <0.0622) and Node 11 (the subgroup of the companies with LOTLTAR > 0.3715, LOLER > 0.0053 and ASTR >0.0622) at the third tree depth according to ASTR (-3.372 vs. -0.340 in NPM).

MARS Data Mining Algorithm

As also explained above, MARS algorithm produced the best predictive accuracy and captured eight predictors i.e. ACTR, ASTR, EQTR, SHTLTAR, LOTLTAR, LOLER, TODTAR financial ratios and company factor. The MARS predictive model with the selected 15 terms which ensured the lowest GCV error yielded the ideal predictive power. All the coefficients in the MARS were found significantly (P<0.01). The correlation coefficient between the real and the predicted values in NPM was estimated for MARS algorithm as 0.980 (P<0.01).

Basis Functions	**Coefficients**
Intercept	0.72561
max (0, 2.088-ACTR)	1.0467
max (0, 0.057-ASTR)	94.0780
max (0, ASTR-0.057)	95.2610
max (0, ASTR-0.069)	-91.6650
max (0, 0.23-EQTR)	5.5018
max (0, SHTLTAR-0.063)	-53.7200
max (0, 0.138- SHTLTAR)	-19.1210
max (0, SHTLTAR-0.276)	16.1820
max (0, LOTLTAR-0.081)	-43.8640
max (0, 0.034-LOLER)	-0.90699
max (0, TODTAR-0.12)	33.1900
max (0, TODTAR-0.43)	9.4499
COMPANY-A* max (0, 0.23-EQTR)	-25.6760
max (0, 2.088-ACTR) * max (0, 0.186-SHTLTAR)	-20.285

Table 1: Results of MARS algorithm in predicting NPM

No effect of the terms 2 and 15 on the NPM was found for the companies with ACTR > 2.088. In this case, the change in SHTLTAR in term 15 had no effect on the NPM. The effect of ACTR on the NPM could be mentioned to be based on SHTLTAR.

When ASTR < 0.057 was considered, no effect of ASTR on the NPM was found for the terms 4 and 5 of the MARS; however, there was a positive effect in increasing the NPM for the term 3. For ASTR = 0.057, no effect of ASTR on the NPM was noted for the terms 3, 4 and 5.

When EQTR > 0.23 for the terms 6 and 14 were taken into consideration, the mentioned terms had no effect on the NPM. However, there was an increasing effect of the term 6 for EQTR < 0.23 while for the effect of the term 14 on the NPM was based on the company factor for EQTR < 0.23.

4. CONCLUSION AND SUGGESTIONS

In conclusion, the obtained results indicated that, i) the regression tree diagram constructed by Exhaustive CHAID algorithm displayed that tourism companies with LOTLTAR < 0.3715 and EQTR < 0.0311 had the highest average NPM of 2.778, ii) CART tree-based algorithm showed that the companies with EQTR > - 0.2125 and ASTR < 0.0246 had the highest average NPM of 4.226, iii) the diagram of CHAID tree-based algorithm revealed that the companies with TODTAR < 0.6145 and EQTR < 0.0311 had the highest NPM with the average of 2.778. It is recommendable that data mining algorithms capture optimal cut-off values of influential factors, which may ensure the highest NPM values.

Considering the model fit statistics, the best algorithm tested here in predictive accuracy was MARS, followed by the tree-based algorithms such as Exhaustive CHAID, CHAID and CART.

	CHAID	**Exhaustive CHAID**	**CART**	**MARS**	**Average**
r	0.868[b]	0.915[b]	0.873[b]	0.980[a]	0.909
SDratio	0.497	0.403	0.488	0.197	0.396
RMSE	1.043	0.846	1.025	0.414	0.832
ME	0.000	0.000	0.000	0.000	0.000
RAE	0.491	0.398	0.482	0.195	0.392
MAD	0.675	0.532	0.628	0.295	0.533
R^2	0.753	0.838	0.762	0.961	0.829
R^2_{ADJ}	0.737	0.820	0.752	0.944	0.813
AIC	10.249	-7.048	6.477	-60	-12.58
AIC_C	10.759	-5.715	6.727	-46	-8.557

Table 2: Model Fit Statistics For The Studied Algorithms

Also, model fit statistics for the studied algorithms are given in Table 1. Considering the model fit statistics, the best algorithm tested here in predictive accuracy was MARS, followed by the tree-based algorithms such as Exhaustive CHAID, CHAID and CART. The lowest SD ratio, RMSE, ME, RAE, MAD, AIC and AICc values as well as the greatest r, R^2 and R^2_{ADJ} values were produced for MARS data mining algorithm. Nearly all of the variability in NPM was explained by the selected 15 terms in the MARS interactive modeling. When regression tree algorithms i.e. CHART, CHAID and Exhaustive CHAID are compared, it is seen that the best predictive performance is provided by Exhaustive CHAID, followed by CHAID and CART algorithm. Firstly, the Exhaustive CHAID algorithm shows that LOTLTAR (long term liabilities to total asset ratio) should be lower than 0.3715 to reach the highest positive NPM. These results indicate that a maximum of 37.2% of the assets should be covered by long-term debts, indicating that 62.8% of the assets should be provided by using equity and short-term debt. In order to reach the highest NPM, CHAID revealed that TODTAR should be lower than 0.6145. These results emphasize that a maximum of 61.5% of total assets should be covered by long-term or short-term debts and the remaining 38.5% of the assets should be provided by using equity. In addition, both CHAID and Exhaustive CHAID algorithms show that the NPM value will be positive in case of EQTR ≤0.0311. This means that the EQTR (Net Sales to Equity Ratio) should be decreased. This situation can be interpreted as tourism companies, which have difficulties in finding long-term funds should increase their equity capitals by merging through strategic alliances with other companies or issuing new shares with the decision of the General Assembly after the payment of the shares corresponding to the basic capital of a company. On the other hand, it can be said that MARS is the most appropriate method to evaluate the data of tourism companies, because it produces the most meaningful results among the four algorithms used. The most important results of MARS show that in order to achieve a positive NPM value, the coefficients of term 3 and 4 tried to be positive while term 5 tried to be neutralized. In this context, either ASTR should be <0.057 or a value between 0.057 <ASTR <0.069. If ASTR is greater than 0.069, it is not desirable, because this time the largest negative coefficient has to be used. Another important finding is that TODTAR value should be greater than 0.43. This indicates that at least 43% of active investments should be made in debt, and it encourages tourism companies to use debt to up to 51%. Eventually, tourism companies should pay attention to their capital structures and determine the most appropriate debt-equity composition for them. When companies have a healthy capital structure, they can provide their sustainability even in the absence of long-term funds for their investments or activities.

REFERENCES

[1] Aziri, B., Nedelea, A. (2013) Business Strategies in Tourism, Ecoforum, 2(1-2), pp.5-11. Available: http://www.ecoforumjournal.ro/index.php/eco/article/view/21/12.

[2] Bahar, O., Bozkurt, K. (2010) Gelişmekte Olan Ülkelerde Turizm-Ekonomik Büyüme İlişkisi: Dinamik Panel Veri Analizi, *Anatolia: Turizm Araştırmaları Dergisi*, 21 (2), pp.1-11.

[3] Gai, L., Ielasi, F. and Ventisette, E. (2017) *Overcoming Borders: a Financial Analysis of Tourism Sector in Italy*, Journal of Economics and Development Studies, 5(4), pp. 17-30.

[4] Çımat, A., Bahar, O., (2003) Turizm Sektörünün Türkiye Ekonomisi İçindeki Yeri ve Önemi Üzerine Bir Değerlendirme, *Akdeniz Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi,* 3 (6), 1-18.

[5] Karataş,M., Babür, S. (2013) Gelişen Dünya'da Turizm Sektörünün Yeri, *KMÜ Sosyal ve Ekonomik Araştırmalar Dergisi,* 15 (25), pp. 15-24.

[6] Ravichandran, M., Subramanian, M. V. (2016) A Study on Financial Performance Analysis of Force Motors Limited. *International Journal for Innovative Research in Science & Technology*, 2 (11), pp. 662-666.

[7] Karadeniz, E., Beyazgül, M., Dalak, S. and Günay, F. (2017) Türk Turizm Sektörünün Finansal Performansının Dikey Analiz Yönteniyle İncelenmesi: BİST Turizm Şirketleri ve TCMB Sektör Bilançoları Üzerinde Bir Araştırma, *Sosyoekonomi Dergisi*, 25(32), pp. 105-119.

[8] Dağlı, H. (2013) *Finansal Yönetim,* Derya Kitabevi, Trabzon, pp.61.

[9] Scmidgall, S. R., Defranco, L. A. (2004) Ratio Analysis: Financial Benchmarks For The Club İndustry, *Journal of Hospitality Financial Management*, 12 (1), pp. 1-16.

[10] Subačiene, R., Senkus, K. (2013) Net Profitability Analysis of Travel Organizers and Travel Agencies, *Ekonomika*, 92(1), pp. 156-165.

[11] Soba, M., Akcanlı, F. and Erem, I. (2012) İMKB'ye Kayıtlı Seçilmiş İşletmelere Yönelik Etkinlik Ölçümü ve Performans Değerlendirmesi: Veri Zarflama Analizi ve Topsis Uygulaması, *Selçuk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 27, pp. 229-243.

[12] Gülcü, A., Cenger, H. (2013) İMKB'de İşlem Gören Turizm İşletmelerin Veri Zarflama Analizi Yöntemiyle Mali Performanslarının Ölçümü ve Benchmarking Uygulaması, *International Journal of Social Science*, 6(8), pp. 853-870.

[13] Yakut, E., Harbalıoğlu, M. and Pekkan, N. Ü. (2015) Turizm Sektöründe BIST'a Kayıtlı İşletmelerin Veri Zarflama Analizi ve Toplam Faktör Verimliliği ile Finansal Performanslarının İncelenmesi, *Journal of Business Research Turk*, 7(2), pp. 235-257.

[14] Tarım, Ş., Dener, H. I. and Tarım, A. (2000) Efficiency Measurement in The Hotel Industry: Output Factor Constrained Dea Application, *Anatolia: An International Journal Of Tourism And Hospitality Research*, 11(2), pp. 111–123.

[15] Bach, M. P., Zoroja, J., Jakovic, B. and Sarlija, N. (2017) Selection of Variables For Credit Risk Data Mining Models: Preliminary Research, 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), 22-26 May 2017.
[16] Singh, A. J., Schmidgall, R. S. (2002) Analysis Of Financial Ratios Commonly Used, Us Lodging Financial Executives', *Journal Of Leisure Property*, 2(3), pp. 201-213.
[17] Kim, W., B. Ayoun (2005) Ratio Analysis For The Hospitality Industry: A Cross Sector Comparation of Financial Trends in The Lodging, Restaurant, Airline and Amusement Sectors, *Journal of Hospitality Financial Management*, 13(1), pp.1-33.
[18] Kim, D. J. (2006) A Comparative Study of Financial Ratios between Hotels and Restaurants, *International Journal of Tourism Sciences*, 6(1), pp. 95-106.
[19] Karadeniz, E., Kahiloğulları, S. (2013) Beş Yıldızlı Otel İşletmelerinde Finansal Oranların Kullanımı: Akdeniz Bölgesi'nde Bir Araştırma. *Seyahat ve Otel İşletmeciliği Dergisi*, *10*(3), pp. 84-106.
[20] Ecer, F., Günay, F. (2014) Borsa İstanbul'da İşlem Gören Turizm Şirketlerinin Finansal Performanslarının Gri İlişkisel Analiz Yöntemiyle Ölçülmesi, *Anatolia: Turizm Araştırmaları Dergisi,* 25(1), pp. 35-48.
[21] Asdullah, M. A., Rehman, Z. (2015) A Comparative Study on Financial Performance of Hotel Industry in Pakistan (Sarena Hotel & Marriott Hotel), Journal *of Tourism, Hospitality and Sports*, 8, pp. 42-54.
[22] Karkacıer, O., Yazgan, A. E. (2017) Turizm Sektöründe Gri İlişkisel Analzi (GİA) Yöntemiyle Finansal Performans Değerlemesi, *Selçuk Üniversitesi Sosyal Bilimler Entitüsü Dergisi,*(37), pp.154-162.
[23] Erdoğan, M., Yamaltdınova, A. (2018) Borsa İstanbul'a Kayıtlı Turizm Şirketlerinin 2011-2015 Dönemi Finansal Performanslarının TOPSİS ile Analizi, *Optimum Journal of Economics and Management Sciences,* 5(1), pp.19-36.
[24] Ali, M., Eyduran, E., Tariq, M.M., Tirink, C., Abbas, F., Bajwa, M.A., Baloch, M.H., Nizamani, A.H., Waheed, A., Awan, M.A., Shah, S.H., Ahmad, Z. and Jan, S., 2015. Comparison of artificial neural network and decision tree algorithms used for predicting live weight at post weaning period from some biometrical characteristics in Harnai sheep. Pakistan J. Zool., 47: 1579-1585.
[25] Akin, M., Hand, C., Eyduran, E. and Reed, B.M., 2018. Predicting minor nutrient requirements of hazelnut shoot cultures using regression trees. Pl. Cell Tiss. Cult., 132: 545:559.[25]
[26] Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.I., 1984. Classification and regression trees. Wadsworth, Belmont, California, USA.
[27] Friedman, J.H. 1991 Multivariate adaptive regression splines, The *Annals of Statistics*, 19 (1991), pp. 1-141.
[28] Zhang, W., Goh, A.T.C., 2016 Multivariate Adaptive Regression Splines and Neural Network Models For Prediction of Pile Drivability, Geoscience Frontiers, 7, pp. 45-52.
[29] Grzesiak, W., Zaborski, D., 2012 *Examples of The Use of Data Mining Methods in Animal Breeding. in: Data Mining Applications in Engineering and Medicine* (ed. A. Karahoca), In-Tech, Rijeka, Croatia, pp. 303-324. https://doi.org/10.5772/50893
[30] Kornacki, J. and Ćwik, J., 2005 Statistical Learning Systems, WNT, Warsaw, Poland.
[31] Milborrow, S. 2018 Notes on the earth package. http://www.milbo.org/doc/earth-notes.pdf pp:1-64.
[32] IBM Corp. Released 2015 IBM *SPSS Statistics for Windows*, Version 23.0. Armonk, NY: IBM Corp.
[33] Milborrow, S 2011 Derived from mda:mars by T. Hastie and R. Tibshirani. earth: Multivariate Adaptive Regression Splines, 2011, R package.
[34] R Core Team., 2014 R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, 2014. http://www.R-project.org.