Applying Data Mining Technique to Sales Forecast

1Erkin Guler, 2Taner Ersoz and 1 Filiz Ersoz

1Karabuk University, Department of Industrial Engineering, Karabuk, Turkey
erkn.guler@yahoo.com, fersoz@karabuk.edu.tr
2Karabuk University, Department of Actuarial and Risk Management, Karabuk, Turkey
tanerersoz@karabuk.edu.tr

Abstract

This paper addresses the issues and technique for Agricultural Machinery and sale prices of the machines using data mining techniques. Data mining means the efficient discovery of previously unknown patterns in large databases. It is an interactive information discovery process that includes data acquisition, preprocessing data, data exploration and model building, interpretation and evaluation. In this study, method of data mining have been applied to the sales data of agricultural machinery products that were obtained from CANSA company in between 2011-2013. Data mining techniques to the data obtained from the CHAID algorithm was applied. Classification technique based analyses were used while data mining and decision model about sale amounts and variables affecting sales were found by this method. According to the analysis results; the R&D spending increase, the amount of agricultural machinery sales is increasing.

Keywords: Agricultural Machinery, Sales, Data Mining, Decision Trees, Chaid Algorithm

1 Introduction

Turkey is one of the few countries which has the most fertile land in the world. In Anatolia many great civilizations were founded; agriculture has been the source of livelihood in these soils for inhabitants for centuries allowing them to develop their culture.

Agriculture is the process of obtaining any product with soil agricultural activities. It is still used as a first-class source of income in many countries helping indirectly development that also produce their own food in the country, out in the increase of exports in a country. However, the importance of agriculture is decreased and has been replaced by mechanization after the industrial revolution in Europe. With the begin of industrial mass production the need for manpower has been reduced.

The mechanization of agriculture continues that industrialization contrary to reduce manpower in the fields, the mechanization of agriculture became a new source of employment for people in factories.

Agricultural Machinery Industry is becoming important to achieve more efficient products, in less time, in terms of operations from year to year, together with the development of agriculture in the world. This sector holds an important place in Turkey's exports. The agricultural machinery sector in Turkey is quite lively. Agricultural machinery manufacturing industry is a sub-sector of the manufacturing industry that producing investment goods.

The main objective is to obtain high yields for per unit area, improvement of living and business conditions in agricultural production. However, as a result of intensive agricultural activities, especially in developing countries soil erosion, salinization, soil compaction, drought are emerging.1

The purpose of this studying, is to determine a company's operating in the sector of agricultural machinery how its sales amount changes depending on which various variable.
2 Material and Method

This research material constitutes sales data between 2011-2013 from operating a company in the agricultural machinery industry in Adana, Turkey. In the study, taking the amount of sales as dependent variable price, temperature (Seasonality), monthly rainfall, monthly inflation, promotional expenses, R&D (Research and Development) expenditures, competitive pricing, advertising with the help of the relations between the CHAID analysis were analyzed. Classification technique based analyses were used while data mining and decision model about sale amounts and variables affecting sales were found by this method.

Data mining is an interdisciplinary study that uses statistics, database technology, machine learning, artificial intelligence and of visualizing. A data mining process generally includes the following four steps: Data acquisition, preprocessing data, data exploration and model building, interpretation and evaluation.

Classification technique based analyses were used while data mining and decision model about sale amounts and variables affecting sales were found by this method. The goal of classification is to develop a model that maps a data item into one of several predefined classes. Once developed, the model is used to classify a new instance into one of the classes. Decision trees are part of the Induction class of DM techniques. An empirical tree represents a segmentation of the data that is created by applying a series of simple rules. Each rule assigns an observation to a segment based on the value of one input. One rule is applied after another, resulting in a hierarchy of segments within segments. The hierarchy is called a tree, and each segment is called a node. The original segment contains the entire data set and is called the root node of the tree. A node with all its successors forms a branch of the node that created it. The final nodes are called leaves. For each leaf, a decision is made and applied to all observations in the leaf. The type of decision depends on the context. In predictive modeling, the decision is simply the predicted value.

Specific decision tree method include the count or Chi-squared Automatic Interaction Detection (CHAID) algorithm. CHAID is decision tree techniques used to classify a data set. The following discussion provides a brief description of the CHAID algorithm for building decision trees. For CHAID, the inputs are either nominal or ordinal. The IBM SPSS Modeler packages accept interval inputs and automatically group the values into ranges before growing the tree. For nodes with many observations, the algorithm uses a sample for the split search, for computing the worth and for observing the limit on the minimum size of a branch. The samples in different nodes are taken independently. For binary splits on binary or interval targets, the optimal split is always found. For other situations, the data is first consolidated.

3 Application

The purpose of this studying, is to determine a company's operating in the sector of agricultural machinery how its sales amount changes depending on which various variable. The price of the product, which is in the same industry to do research on competing firms to increase the sale price of their machines, the amount of the monthly promotion spending, monthly the amount of research and development activities as an argument made and used. In addition to the model, between the years 2011-2013 of monthly precipitation amounts, monthly temperature values, monthly inflation rates in Turkey and the Turkish Liras has been included in the foreign
markets, too. For pre-processing activities in the modeling study data cleansing, merging and conversion process has been done through IBM SPSS Statistics program. Work on the data in IBM SPSS Modeler 11.1, Decision Tree algorithm was applied CHAID algorithms.

In this study, the sales amount’s minimum and maximum value have found minimum 42 and maximum 1462 units. The average of monthly sales amounted to 358 units.

3.1 Chaid Analysis Results of Data Mining

Chaid analysis applied with the algorithm as a result of the dependent variable is the most important variable affecting the amount of agricultural machinery sales, R&D has been seen. Other variables, in order of importance; temperature, the amount of the monthly rainfall promotion and inflation.

In the model, the estimation of the amount of sales as the dependent variable selected results are shown below. Variables that most influence the amount of sales in R&D expenditure was found. There are 37 nodes in the analysis. Decision tree all notes are as Figure 3-1.

Fig 3-1 Decision Tree All Nodes
The detailed results are given to nodes of 13, 28 and 35. The amount of sales forecasts has emerged as the average value, the smallest and largest. Decision tree sales amounts (Minimum) node of 13 is as Figure 3.2.

**Fig. 0-2** Decision Tree Node of the 13 and Estimated Sales Amount of 13.
In the case of being monthly temperature ≤ 23.3 °C and > 27.3 °C, Research & Development ≤ 2.500 TL and > 2.750 TL, promotion ≤ 2.000 TL and Research & Development ≤ 2.750 TL; estimated sales amount will be 45 unit as calculated.

Decision tree sales amounts (Average) node of 28 is as Figure 3-3.

![Decision Tree Node of the 28 and Estimated Sales Amount of 28.](image)

In the case of being monthly rainfall > 63 mm, Research & Development ≤ 27.500 TL and > 41.250 TL; estimated sales amount will be 462 unit as calculated.

Decision tree sales amounts (Maximum) node of 35 is as Figure 3-4.

![Decision Tree Node of the 35 and Estimated Sales Amount of 35.](image)

In the case of being monthly temperature ≤ 13.6 °C, Research & Development > 63.250 TL; estimated sales amount will be 1462 unit as calculated.
4 Conclusion and Evaluation

The study focused on classification using data mining methods and the amount of sales on a monthly basis in 2011-2013, which were obtained using the classifier model. In this study, the amount of sales in Agricultural Machinery variables affecting the sector, data mining methods have been tried to predict.

The data used in this study is taken from CANSA Agricultural Machinery Co. These coefficients are multiplied by a certain coefficient given due to confidentiality requirements modeling was conducted. This information is evaluated in various ways, the connection between the independent variables as the dependent variable were performed to determine their effect.

In this study, the sales amount’s minimum and maximum value have found minimum 42 and maximum 1462 units. The average of monthly sales amounted to 358 units.

The main purpose of the development of the models, the coefficient of the independent variables in the model by looking at how many sales they can occur is to create a model that can predict. In CHAID analysis, in order of importance; R&D spending, promotional expenses, temperature, monthly rainfall and inflation rate were significant and included in the model.

According to the analysis results, we may state: the R&D spending increases, the amount of agricultural machinery sales is increasing.

There are some obstacles in front of the agricultural machines and data mining applications. Primarily; in Turkey, companies operating in the agricultural machinery sector can not be fully institutionalized. Because of inability of the company to be institutionalized, difficulties in providing data are experienced.

In further analyses conducted in the future, the authors hope to continue studies that could lead to further researches and, at the end, strengthened sustainable growth and improved living conditions, in Turkey and other emerging and developing countries.

References