Data Mining Application for Real Estate Valuation in the city of Skopje

Zoran Gacovski, Josip Kolic, Rosica Dukova, Marko Markovski

FON University, Bul. Vojvodina, bb, Skopje, Macedonia
```
{zoran.gacovski,josip.kolic, rosica.dukova,
        marko.markovski}@fon.mk
```

Abstract. In this paper we present an application of data mining techniques in order to make price prediction of the real estate properties in the city of Skopje. The current research on the real estate data is insufficient, resulting in miss-understandings between the key players - local government, construction companies, real estate agencies and the potential clients. A dataset from over 1000 transactions in the past three years was used. Five variables (attributes) for each apartment were taken into consideration. We have used SQL Server database and Microsoft Business Intelligence tool (three different algorithms – decision trees, neural networks and logistic regression) in order to perform the price prediction. Also, it can be useful for the prediction of future trends in the urban development of city of Skopje.

Keywords: Data mining techniques, Real estate property value, Microsoft Business Intelligence.

1   Introduction

This paper examines the factors that determine housing prices in a sample of over 1000 home sales in Skopje's region during the period of 2009-2011. Our analysis can be used in real-estate Agencies, and can help the potential buyers to estimate whether the property price is in accordance with the existing market trends.

This paper is organized as follows: in Section 2 – a problem of price prediction is described and data mining algorithms (neural networks, decision trees and logistic regression) are presented. In Section 3- an overview of the Business intelligence software is given, and in Section 4 – simulation results are presented.

2   Problem description and algorithms

The problem we have investigated is the real estate market in Macedonia, particularly - apartment sales in Skopje. According to available dataset for about 1000 transactions from the previous three years (2009-2011), and by applying the Business Intelligence and data mining, the task is to predict the price of an apartment with known attributes (characteristics).

For population of apartment sales database (training set) we have used data from Macedonian real-estate agencies for transactions from the past 3 years. For each apartment we have used the following data: suburb (settlement) of the apartment, quadrature/ surface area (in m^2), floor of the apartment, number of rooms, heating

(central, electricity, fossils, etc.) and apartment price (in Euros). Finally, our database consisted of 1200 apartments, which was the training set for our predictions.

We have used Business Intelligence Studio from Visual Studio 2008. For solution of our problem (price prediction) we have applied 3 different techniques: Decision Trees, Neural Network and Logistic Regression.

After creation of the data mining model, it can be evaluated in Business Intelligence Studio, in order to make the predictions for the price (for each apartment within the table). We have shown the predicted prices graphically.

3   Overview of the Business Intelligence tool

The Business Intelligence software offers useful tools which can be applied in the process of strategic planning and management in the companies. This software enables the companies to discover their critical operations via different reporting and analyzing tools. BI deliverables can incorporate different components, like tabular reports, shared lists, diagrams and graphs. Although traditional BI systems were developed via host terminals and printed reports, the current development of BI applications is performed via Web (Internet). It is also possible to develop interactive BI applications, which are optimized for mobile devices, smart-phones and e-mail usage.

The BI environment is well integrated in Microsoft Visual Studio.NET, in order to enable faster development of BI applications. The data mining project developed in a BI environment is known as a solution.

4   Simulation Results

In our research – we have investigated and simulated the problem of price prediction – with the three algorithms offered in Microsoft Business Intelligence. Bellow, the diagrams for comparison of the apartment selling prices with the predicted prices obtained by these algorithms, are given.
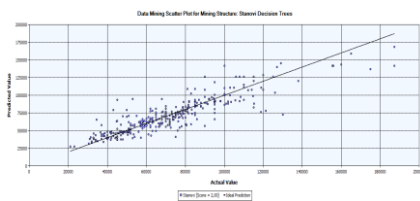

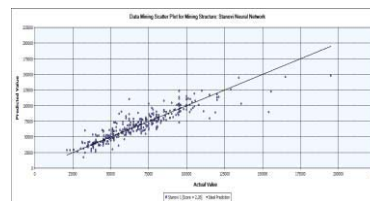Fig. 1. Diagram of predicted prices by Decision Trees algorithm.

Fig. 2. Diagram of selling/predicted prices obtained by Neural Network.

5   Conclusion

In this paper we have used dataset from over 1000 transactions in the past three years. Five variables (attributes) for each apartment were taken into consideration. We have used SQL Server database and Microsoft Business Intelligence tool (three different algorithms – decision trees, neural networks and logistic regression) in order to perform the price prediction. Our analysis shows that logistic regression algorithm gives better (closer) prediction of the home prices than other two algorithms.