Data Mining in Real Estate Business

Basavaraj Jakkali
Associate Professor, Department of Computer Science, BMSCE, Bangalore

***Abstract*- Searching for a house/land at a reasonable rate in metropolitan city has become a biggest issue these days. Different house/land rates depend on different attributes like number of rooms in the house, facilities available, location of the property etc. It is important to providing one stop information to customers about the variation of prices with respect to different attributes and also help customer to choose the best suitable property within their budget.**

Index Terms- Data Importing, Data preprocessing, Data Mining, Visualization.

I. INTRODUCTION

Data is everywhere and it is expanding exponentially, Data is being generated from multiple sources and in multiple formats. Data is generated from different sectors like health care sector, financial sector, social media, and telecom sector. Data can be in structured, unstructured and semi-structured formats. Data mining is the computing process of discovering patterns from large datasets involving methods at the intersection of machine learning, statistics and database system. Discovering knowledge from database involves Data selection, Data pre-processing i.e. understanding structure of data, Data mining i.e. applying intelligent operations such as clustering classification, regression to extract pattern, pattern evaluation i.e. obtained pattern is correct useful and new, knowledge representation i.e. represent data in understandable format like graphs.

Data mining in real estate business is interdisciplinary in nature, it involves combination of Machine Learning, Big Data and Statistics analysis. Data about all the house in a metropolitan city constitutes for a very large amount of data i.e., Big data. Extracting useful information from this larger sets of data is Data Mining and making System understand about this data is Machine learning, all these disciplines goes hand in hand. Initially all data about a property is collected and is refined later on this data is used to plot graphs for prices of the property against different attributes, testing data is given to the used to find the accuracy of the system along with error rate

A. PROBLEM STATEMENT

To analyze and predict prices of different house/land with respect to different attributes and testing the new sample data against existing database to know the difference in the price.

II.LITERATURE SURVEY

The analyses of previous market trends and price ranges, to predict future prices have been observed. It brings together the latest research on prediction markets to further the utilization by economic forecasters. It provides a description of prediction markets, and also the current markets which are useful in understanding the market which helps in making useful predictions. Thus, there is a need to predict the efficient house pricing for real estate customers with respect to their budgets and priorities. The database of property rates contains attributes like quarter, upper, average and lower, where each year is divided into 4 quarters. The column upper consists of the average values of the houses that are high in prices, likewise average and lower column consists of average values of middle range and low range house. To use linear regression the quarter attribute is assigned on x-axis and the values of rates on y-axis. Analyzes previous market trends and price ranges, and also upcoming developments future prices will be predicted [1].

The real estate market in Macedonia, particularly apartment sales in Skopje has been investigated. Based on the dataset from the previous three years, and by applying the Business Intelligence and data mining, the task is to predict the price of an apartment with known attributes (characteristics).

The Business Intelligence software enables the companies to discover their critical operations via different reporting and analyzing tools. [2].

The methodology is a case study of current literature on Big Data applications. The findings show that Big Data is having an increasing impact on the evaluation and understanding of real estate markets. We create a typology of real estate data including Core, Static Spatial and Peripheral. While many of the initial applications of Big Data in the property sphere have focused on valuation and residential property sales, real estate investment and development are primed to benefit (as well as suffer disruption) from the explosion of data and the growth of advanced analytical techniques [3].

The Microsoft Decision Trees algorithm is a classification and regression algorithm provided by Microsoft SQL Server Analysis Services for use in predictive modelling of both discrete and continuous attributes. The Microsoft Neural Network algorithm creates a network that is composed of up to three layers of neurons. Decision Trees and Neural Networks models have been built to using the actual residential property transactions in King County. These prediction values were saved in SQL Server database to calculate the prediction errors from the decision trees and neural networks, respectively [4].

The first phase of this paper is all about differentiating the residential areas available in each of the zones of Delhi and ranking them. For this, we listed different attributes that affects the pricing of real estate in an area and then, identified the most influencing attribute for ranking. The city of Delhi, in total has 5 zones, namely, South Delhi, East Delhi, West Delhi, North Delhi and Central Delhi. Then it involves identifying amenities between structure and unstructured data [5].

III.EXISTING SYSTEM

A. Places that have housing complexes or multi-storey apartments which were located in commercial areas were not included and included only those apartments that were constructed by the Development Authority.
B. Any economic conditions external to the properties were not factored in the analysis.

IV.METHODOLOGY

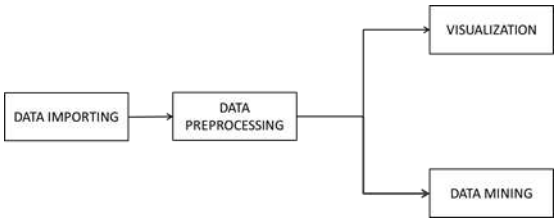

Fig 1.High level system design

A. Data Importing
Data Importing involves collecting all real estate data and save it in a single file for processing
B. Data Preprocessing
Data pre-processing involves removal of anomaly, null values and unwanted attributes
C. Data Mining
In data mining we extract useful information about prices of the houses and also predict prices based on different attributes
D. Visualization
Data visualization in done through simple graphs using linear regression Training Data Set

V.SYSTEM DESIGN

A.Data Mining diagram

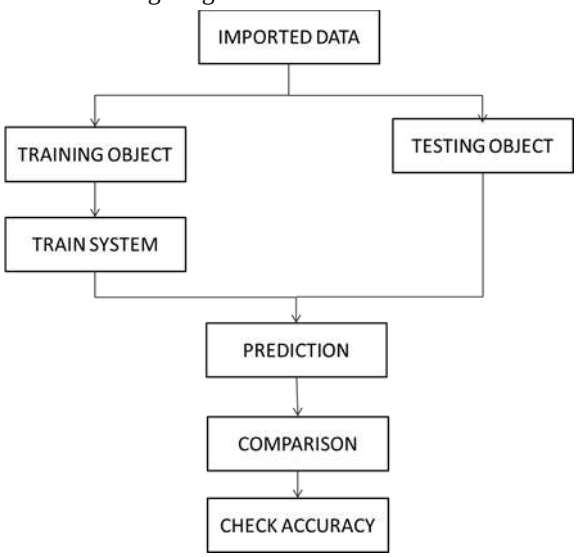

Fig 2. Data Mining diagram

In data mining, imported data is split into training object and testing object. Training object is used to train the system and create model, testing data is used to predict price of the house based on the model built. Predicted values are compared against actual values, error in prediction and accuracy is calculated to check the performance of model being built. Different models can be built on same training and

testing data, model with least error and more accuracy in prediction is chosen as the best model for prediction.

VI. CONCLUSION

In today's highly competitive real estate business, it has become difficult to store such huge data and extract them for one's own requirement. Also, the extraction should be useful which is another complicated step; otherwise, there would be no use of such data. Implementing the concept of data mining applications in such a field has definitely proven to be useful, time saving and advantageous to the customers. One of the major future scopes is adding real estate database of more cities which will provide the user to explore more estates and to reach an accurate decision. More factors that effects the price of house shall be added, this will help the system to run on a larger level.

REFERENCES

[1] Nihar Bhagat, Ankit Mohokar, Shreyash Mane, "House Price Forecasting using Data Mining", International Journal of Computer Applications (0975 – 8887)Volume 152 – No.2, October 2016

[2] Zoran Gacovski, Josip Kolic, Rosica Dukova, Marko Markovski, "Data Mining Application for Real Estate Valuation in the city of Skopje", ICT Innovations 2012 Web Proceedings - Poster Session ISSN 1857-7288

[3] Kimberly Winson-Geideman and Andy Krause," Transformation in real estate research: The Big Data revolution", 22nd Annual Pacific-Rim Real estate society conference, sunshine coast, Queensland, Australia 17-20 January

[4] Victor Gan, Vaishali Agarwal, " Data mining analysis and prediction of real estate prices", Issues in Information SystemsVolume 16, Issue IV, pp. 30-36, 2015

[5] Vishal Venkat Raman, Swapnil Vijay, Sharmila Banu K , "Identifying Customer Interest in Real Estate Using Data Mining Techniques", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (3) , 2016

[6] https://www.r-bloggers.com/installing-r-packages/

[7] https://www.rdocumentation.org/packages/base/versions/3.5.0/topics/library