REAL ESTATE INVESTMENT INFORMATION SYSTEM USING DATA MINING FROM WEB DATA

Suresh Nanayakkara[a]
[a*] *University of Sri Jayewardenepura, Sri Lanka*

Abstract

In present world, real estate is considered as a popular investment category. However, most of the investors in both large scale and small scale take real estate decisions based on traditional paper based or legacy computer applications mostly with outdated data. Outdated data and traditional or legacy applications may affect reliability, accuracy of the real estate investment decision which may cost hundreds of millions of rupees. In modern world, World Wide Web is used as a medium of publishing of real estate related advertisement, publication and news articles related to real estate and there for more updated and accurate set of data is available through the internet but is hard to gather all data as they are scatted across the web. Data mining is one of main technologies to collect and analyze those scatted data and same can be used to gather, store and analyze real estate and property related data to facilitate investors' by providing up to date and accurate information.

Keywords*:* Real Estate, Investment Decision, Data Mining

Introduction

This paper represents how information technology can be used as a supporting tool to the real estate investment decision making or as a decision supporting system for real estate investments. Land and property investment becomes a popular investment category in today's society in large scale as well as small scale not only with business community but with non-business communities as well. This paper reflects how information published on World Wide Web can be used to evaluate available investment opportunities in the market according to given criteria by an investor. Even though growing amount of digital data available on the internet and available most for free, the investors are not possible to go through all those paper or digital advertisements due to large amount of data to be processed and rapidly changing nature of those data and practically difficult to browse all web sites and gather and record only relevant data as they update on the web, due to limited time and resources. Further, traditional analysis of these data using paper based or spread sheet data collection and analyses is not possible or less effective due to this type of data changing rapidly over the time and dynamically changing their appearance in the web and in the market. Traditional data analysis methods mostly depend on manual data inputs and manual analysis of those data.

However the whole process can be automated more accurately using computer based programming algorithms and techniques to get even more reliable data collection, analysis and generate reports for decision making (Provost and Fawcett, 2013). Data mining techniques are

[*]*Corresponding Author: suresh@sjp.ac.lk*

a one of emerging concept in information Technology world (Bakırlı *et.al.,* 2012). It can be used to collect valuable data from very large set of data sources like data from web sites, information systems and social media sources. The same technology can be used in real estate investment decision making by collecting data form the published articles, advertisements, news from various web sites and then can be used to analyze, and visualize for increase the accuracy, responsiveness and reliability of the decisions made by real estate investors at any scale. The investor can pre-set their preferences like price, place and product type, and get more quick updates as they published on the web. There is lack of automated alert system for investors to get an idea about the market behavior.

The Objective of the Study

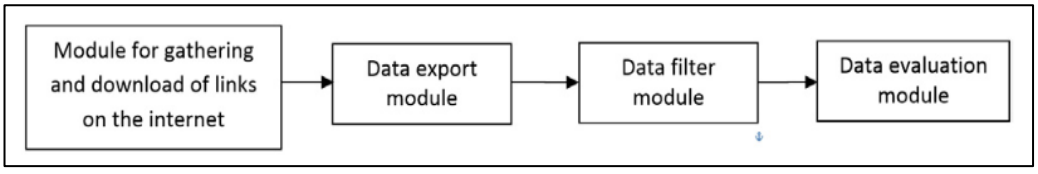
The aim of the research is to provide investors updated market data to take accurate investment decision using data mining methods. Even through real estate investment becomes a huge investment category, still investors' use the traditional paper based or an advertisement published on the internet. The research has been carried out to find the difficulties that they are facing while decision making and proposed a new level of decision-making system. In order to achieve the above goal, the few key objectives have been identified. These objectives range from critical review of existing approaches to real estate investment system to evaluation of the proposed approach. The data have been taken from the published web pages.

Literature Review

There are research who are interested in land and property investment has considered this area and they have proposed some systematic ways to fill this gap. Many of them analyzed the international market and observed is behavior. 'Filiz ERSOZ, Taner ERSOZ2 and Muhammet SOYDAN' has described an data mining methods which can be used for real state valuation using CHAID and C&RT algorithms (Ersoz *et.al*., 2018) for the analysis of data. Further, they have used advanced technologies such as machine learning, statistical modeling in real state valuation to determine the real value of the real estate. In their semi-automated approach, involve identifying real estate economic and constructional management through the data patterns and structure of data sources. In their paper they have clearly express how data classification, clustering and analysis can be done and how those knowledge base can be uses for decision makers.

'E. Hromada' examined the most common factors of real estate valuation and presented an innovative structure for real estate valuation named "Historical Market Price" (Hromada, E., (2016), which has designed using mathematics, statistical and database algorithms for valuation. The proposed method uses the true and exact information which has taken from the previous purchase and examine the technical and structural differences of given real estate and price changes for the given location such as city, region, street etc.

Figure 1: The Bacis Structure of the Software



Source: Hromada, E, (2016).

In the survey by 'Nick French, Simon French' describes that decision theory in mathematical models (French, 2001) help to identify the business decisions towards the investment. Gerhard Muggenhuber elaborated a decision making system for house hold and real estate economics in his paper on 'Geospatial Data Mining and Analytics for Real-Estate Applications' (Muggenhuber, 2018). The authors of the paper focused on mechanism of digital data acquisition, analyze and visualization of data from the housing market. This statistical model of forecasting real estate market based on web mining. These types of methods provide most relevant information for decision making for real estate market.

Methods

The investor decides the investment requirement first, such as the maximum price to pay, the area, district or province and distance to particular place. Further he can select the market size and availability of interested properties in selected market. While doing these in the frontend to the investor called "investor dashboard", the data collection and analyzing algorithms are running in the background using data mining techniques and web scrapping techniques using Scrapy. The web scrapping tools and their pre-defined algorithms continually look for newly available data from web sites, news articles, and advertisements capture them and bring those data to the backend processing system the valuable data mining algorithms.

Web scrapping tool used can automatically search in search engines to find the relevant advertisement published websites. The web scrapping tool visits websites in background and browse through web pages of classified and news items to collect relevant data. The algorithm sorts the captured real estate related data and stores to a MySql table in an indexed order created on a Linux server for high speed data retrieval. These data contain source, date stamp, location, and price at the time of the data storage. These steps continuously run in the background in pre-settable intervals to capture and store the newly posted advertisements and news posts.

Figure 2: High Level Overview of the System

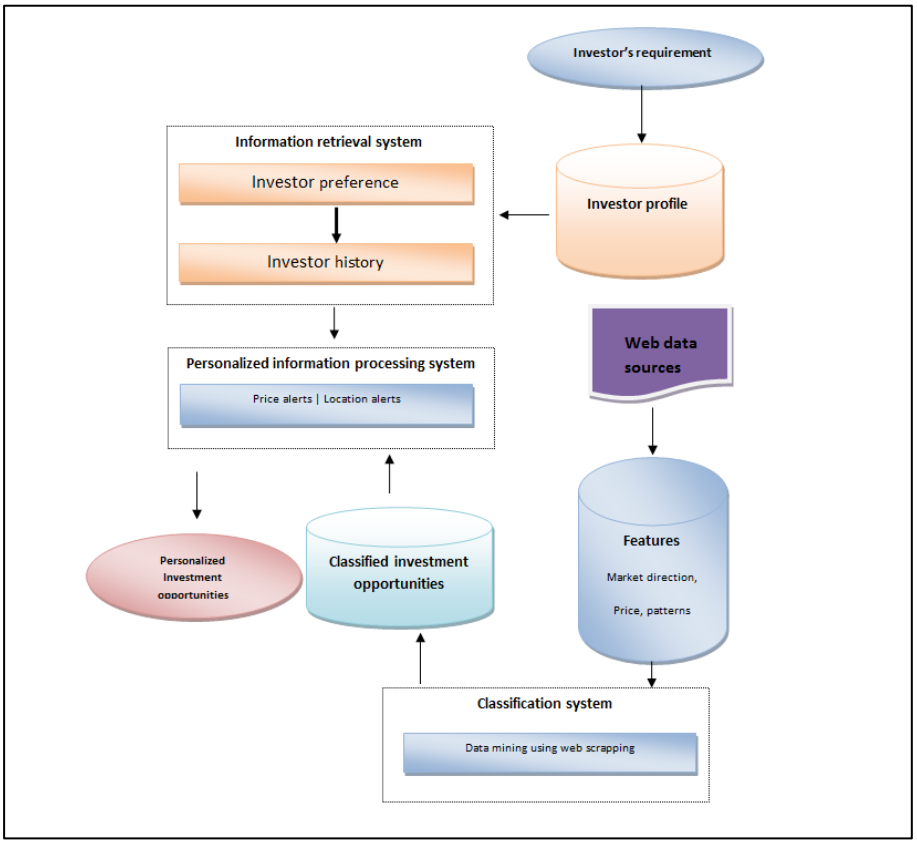
Source: Author, (2020).

Results and Discussion

The system started by data capturing process through web scrapping the system could capture data published on the internet, by browsing through different websites and pages then it could store relevant real estate and property related data to the database, according to the pre-

determined orderly manner. The web discovering interval was set to 30 minutes so that system visits internet every 30 minutes to browse and continuously collect new data, thus allowing the investor to receive alerts on matching investment opportunity within 30 minutes as they have published on the web. The investor interface which provide basic user interface for setting up the required filters and inputs was enabled to get investor inputs related to his investment need and preferences. The automated triggering system activated once system found relevant investment match for investor and send visual alert to his user interface, then the investor can click on it to see further details of the published advertisement or the news article and then he can make his investment decision accordingly. This helps investors to make decision fast and simple as details are in the hand. The system which captures data from web based sources and then facilitates investors to decision making based on those alerts generated by the system.

Figure 3: Investor Dashboard



Source: Author, (2020).

Figure 4: Web Scrapping Techniques Using Scrapy

```
import scrapy

class RedditSpider(scrapy.Spider):
```

Source: Author, (2020).

Conclusion

The mining technique used, could gather and store online data by browsing through web sites according to preset intervals, the data was sorted and categorized according to the characteristics they had and stored in a MySQL table. The investor could set his investment preferences via the front end and wait for results to appear. Once the system found a matching real estate or property from its data mining algorithm investor got alert about that. Then the investors do not have to find the real estate or property stressfully since he gets the alert with all the details.

References

Bakırlı, G., Birant, D., Mutlu, E., Kut, A., Denktaş, L. and Çetin, D., (2012). Data mining solutions for local municipalities. In Proceedings of the 12th European Conference on e-Government: ECEG (p. 74). Academic Conferences Limited.

Ersoz, F., Ersoz, T. and Soydan, M., (2018) Research on Factors Affecting Real Estate Values by Data Mining. Baltic Journal of Real Estate Economics and Construction Management, 6(1), pp.220-239.

Hromada, E., (2016). Real Estate Valuation Using Data Mining Software. Procedia Engineering, 164, pp.284-291.

French, N., (2001) Decision theory and real estate investment: an analysis of the decision-making processes of real estate investment fund managers. Managerial and Decision Economics, 22(7), pp.399-410.

Muggenhuber, (2018) Geospatial Data Mining and Analytics for Real-Estate Applications. Lecture Notes in Geoinformation and Cartography, pp.225-240.

Provost, F. and Fawcett, T., (2013) Data science and its relationship to big data and data-driven decision making. Big data, 1(1), pp.51-59.