Journal of Business, Economics and Finance (JBEF), ISSN: 2146 – 7943, http://www.pressacademia.org/journals/jbef

PressAcademia

Journal of Business, Economics and Finance

Year: 2018 Volume: 7 Issue: 1

PREDICTION OF RESIDENTIAL GROSS YIELDS BY USING A DEEP LEARNING METHOD ON LARGE SCALE DATA PROCESSING FRAMEWORK

DOI: 10.17261/Pressacademia.2018.801
JBEF- V.7-ISS.1-2018(12)-p.124-129

Semra Erpolat Tasabat[1], Olgun Aydin[2], Ali Hepsen[3]
[1]Mimar Sinan University, Department of Statistics, Istanbul, Turkey,
semra.erpolat@msgsu.edu.tr, ORCID: 0000-0001-6845-8278
[2]Mimar Sinan University, Department of Statistics, Istanbul, Turkey,
olgun.aydin@ogr.msgsu.edu.tr, ORCID: 0000-0002-7090-0931
[3]Istanbul University, Faculty of Business Administration, Department of Finance, Istanbul, Turkey,
alihepsen@yahoo.com, ORCID: 0000-0002-3379-7090

To cite this document
Tasabat, S. E., Aydin, A., Hepsen, A.. (2018). Prediction of residential gross yield rates by using a deep learning method on large scale data processing framework. Journal of Business, Economics and Finance (JBEF), V.7(1), p.124-129.
Permemant link to this document: http://doi.org/10.17261/Pressacademia.2018.801

ABSTRACT

Purpose- Households, investors and companies who want to make an investment on residential properties are interested in sales prices and rental values that vary depending on regional factors, location and attributes of residential units. It is the preference of investors to buy a new house with higher rental income. Real estate developers and real estate consultants as well as the real estate investors are also interested in investigating relationship between gross yield rate and location, regional factors, attributes of residential units. The purpose of this study is to examine the relationship between attributes of the residential units, location of the units and the gross yield rate.
Methodology - In this study, the prediction model of residential gross yield rates with the help of city, county, district, residential attributes information, was created by using LSTM, which is a deep learning method, on big data platform Spark.
Findings- According to test results, it has been proven that gross yield rates could be estimated with high accurate model by the aid of Long short term memories. With this model, researchers can predict gross yield rate of any specific flat.
Conclusion- The LSTM network has been built in this study shows that the residential gross yield rate could be estimated using city, county, district, number of rooms, number of bathrooms, floor number, total floor attributes. This study also shows that the Spark framework can be used to deal with the growing size of data in real estate and to develop deep learning applications on distributed data processing platforms.

Keywords: LSTM, real estate, residential gross yield rates, big data
JEL Codes: R31, R10, R30, C13, C82

1. INTRODUCTION

Real estate industry is based on balance of supply-demand. Due to the reasons such as the rapidly increasing population of Turkey, urbanization, increasing income levels and improving life standards, there is an increasing demand for the properties on a constant basis. Even though political and economic unrest in the MENA and Turkey, Turkey succeed to remain politically and economically stable and the average of GDP growth rate is 7% in the last 15 years. However, inflation rate reached to %11 in October 2017. There were 1.03 million house sales transactions were recorded in the first 9 months of 2017 which is 10% more than same period in 2016. Turkey's population has just crossed 80 million and around 35 million people are between the productive age group of 20-55 years and 49% of the population is under 30 years. Based on the population stats, there are around 400 thousand marriages take place. This is the direct effect on real estate demand. According to the REIDIN Turkey September 2017 report, residential sales prices increased 11.5 percent in the last one year and 44 percent in the last 3 years. On the other hand, rental values increased 4.89 percent in the last one year and 32 percent in the last three years. Average gross rental yield for a standard house is 5.5 percent as of September 2017. Additional to 400 thousand yearly new supply, there is an urban regeneration in the Turkey currently and around 7.5 million units needs to be rebuilt in the next 20 years which will have market value around USD 1.5 trillion. Turkish real estate also attracts foreign investors especially from GCC region. Investment from foreign investors could accelerate the growth in the upcoming years. (Reidin Turkey, September 2017 Report)

Real estate is also defined as a land being a physical asset and structures built on this land by humans. Housing is thought as both a sheltering instrument and an investment instrument in Turkey. When housing is thought as an investment instrument, investors both expect a capital income from the difference between purchase and sale and plan to have a regular rental income. At this point, it becomes important to invest on residential unit with efficient characteristics. In this case estimating real estate prices and analysing gross yield rates become crucial.

Gross yield rate is the return of the income from the investment that is calculated without including any costs. Interest rates are also excluded when gross returns are calculated. Gross yield rate in residential investment is the ratio calculated using a residential unit's sale price and gross rental income of this residential unit.

Gross yield rate is an excellent indicator of the overall outlook of the rental income of a residential unit compared to its value. Gross yield rate of a residential unit helps to quickly understand the characteristics of different units. For example, it is obvious that which of the residential units with a gross yield rate of 4% and gross yield rate 10% brings more cash flow.

It could be asserted that optimization in residential investments is overemphasized for people who not only content with capital gain but also prefer residential investments to obtain a regular rental income in long-run period.

2. LITERATURE REVIEW

There are some studies in the literature about estimating real estate prices and analysing gross rental. Ratchatakulpat el al. (2009) noted that the factors that prospective buyers consider when purchasing residential property in Queensland, Australia. A drop-off survey was used, with 376 property buyers and a response rate of 62.7 percent. Affordability, maintenance and interior design, and a good neighbourhood were considered as most important. Of least importance are the affluence and quality of the area, water, views and roads, and features, such as a pool or air-conditioning. Therefore, location is important in the sense of neighbourhood and community, rather than prestige. In another study, Lee (2001) presented an elegant and simple approach to the decomposition of property type and regional influences on property returns, and thus provided a quantitative framework for analysing the relative impact of these two diversification categories to real estate portfolio selection. Using data on retail, office and industrial properties spread across 326 real estate locations in the United Kingdom, over the period 1981 to 1995, the results showed that the performance of real estate was largely property type-driven, a result in line with previous work. This implied that the property type composition of the real estate fund should be the first level of analysis in constructing and managing the real estate portfolio. Consequently, real estate fund managers needed to pay more attention to the property type allocation of their portfolios than to the regional spread. Jackson (2002) conducted an alternative classification study to examine the development of the return on investment in rent for the regional retail market. Baker studied about the gross yield rate of real estate in Australia (Baker, 2001) and Goetzmann et. al. (2001) applied clustering algorithm to effective rents for twenty-one metropolitan U.S. office markets, and to twenty-two metropolitan markets using vacancy data. Unlike other clustering studies, they found strong evidence of bi-coastal city associations among cities such as Boston and Los Angeles. They presented a bootstrapping methodology for investigating the robustness of the clustering algorithm and developed a means for testing the significance of city associations. While the analysis was limited to aggregate rent and vacancy data, the results provided a guideline for the further application of cluster analysis to other types of real estate and economic information.

3. DATA AND METHODOLOGY

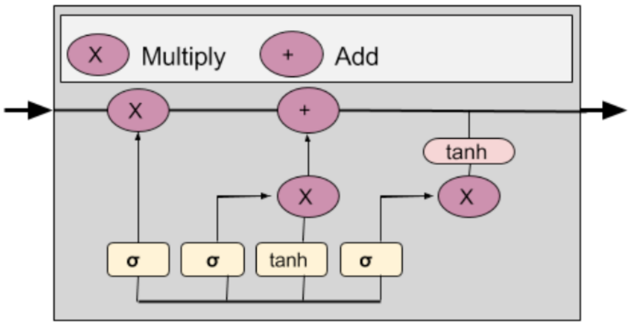
Neural Nets (NNs) can estimate almost any linear or non-linear function in an efficient and stable manner, when the underlying data relationships. The NN model is a nonlinear, adaptive modelling approach which relies on the observed data rather than on an analytical model. The architecture and the parameters of the NN are determined solely by the dataset (Feng, 2005). NNs has been evolving for years.

Nowadays Deep Neural Nets (DNN) are one of the trending topics in this domain. With DNNs, researchers are able to build NN model in a more efficient way. Deep Learning is rising star of Machine Learning and Artificial Intelligence domains. Until 2006, many researchers had attempted to build deep neural networks, but most of them failed. In 2006, it was proven that deep neural networks are one of the most crucial inventions for the 21th century. Nowadays, deep neural networks are being used as a key technology for many different domains: self-driven vehicles, smart cities, security, automated machines

In this paper, we applied Long Short Term Memories (LSTM) which is one of DNNs on big data platform Spark. We trained and compared our LSTM models at various numbers of parameters and configurations. LSTM models converges quickly and gave effective prediction performance.

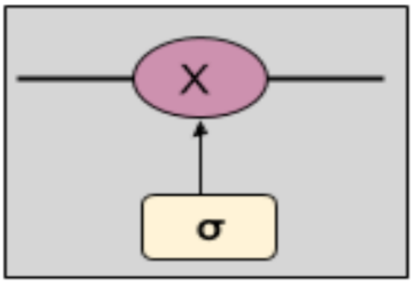
Sepp Hochreiter and Bergen Schmidhuber proposed LSTM in the mid 90s for vanishing gradient problem. LSTMs have a chain of repeating modules of neural network as standard Recurent Neural Networks (RNN) (Mahammad et.al, 2010). Repeating modules in standard neural networks have the simple structures like tanh and sigmoid layers; however, LSTMs have different repeating modules comparing other type of neural networks. Instead of having a single neural network layer, LSTMs have four interacting special layers as seen on Figure 1.

Figure 1: LSTM Modules Including Four Layers



Each layer carries an entire vector from the output of layer to inputs of the next layer. LSTMs have an ability to add information or remove information while going through gates that are a way to decide how much information to carry among the layers Figure 2 shows the point wise multiplication operation which carries the information among the cells. (Greff et. al, 2016).

Figure 2: LSTM Module Including Add Operation



In this study, the relationship between the attributes of the residential units, location of the units and the gross yield rate was examined. This relationship was modelled by LSTM, one of the deep learning methods. All model development processes were built using big data platform Spark. Spark, one of the distributed data processing architects, was used to increase data processing power and speed. Apache Spark has been started some research group's discussions with Hadoop users. The main advantage in Apache Spark is resilient distributed dataset (RDD), Users can explicitly cache an RDD and it is available to reuse in multiple MapReduce-like parallel (Zaharia et. al, 2010).

Residential unit listing data obtained from the online listing portals serving in Turkey were used. The dataset covers 150081 flat listings for sale from 2981 districts, 308 counties and 74 cities. In this context, the final dataset including the city, county, district, number of rooms, number of bathrooms, number of floors, total floor numbers, building age, rental value and sales price information was used. The residential gross yield rate for each listing was calculated by using the sales price and rental value of each listings.

For all these operations, Keras, which is a high-level neural networks library, written in Python and capable of running on top of TensorFlow, was used (https://keras.io/). Another Python library, Elephas, was used to enable the model developed with Keras to run on Spark (https://github.com/maxpumperla/elephas).

4. FINDINGS AND DISCUSSIONS

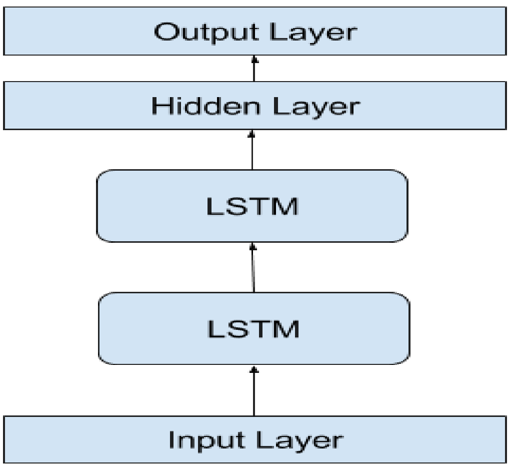
In the generated model for estimating residential gross yield rates; city, county, district, number of rooms, number of bathrooms, floor number, total floor number, building age were used as independent variables and residential gross yield rate were used as dependent variable. The dataset was divided two parts as train set and test set. 70% of the data set was used as train set and the remaining parts of the dataset was used as test set. According to Table 1 which shows descriptive statistics of all the dataset, Gross yield ratio is in the range between 1.34% and 17.23%.

Table 1: Descriptive Statistics of Dataset

	Noofrooms	**NoofBaths**	**Size (sqm)**	**Tot. Floor**	**NoofFloor**	**Yield**
Min	1	1	30	3	1	1.34%
Max	6	3	650	35	35	17.23%
Mean	2.8129	1.3926	154.17	5.8612	3.5983	5.21%
Variance	1.1055	0.7016	10.6553	8.6789	9.8724	0.02%

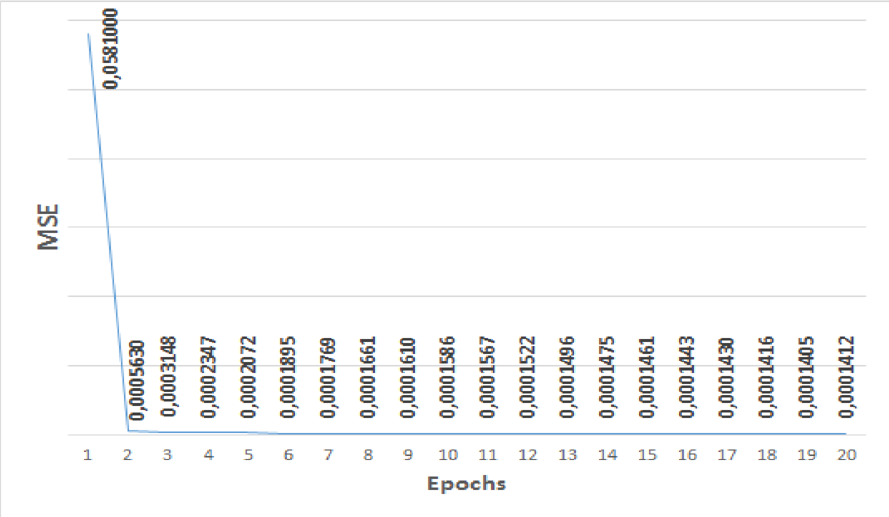
To train the LSTM network, the number of epoch was set as 20 and the batch size set as 30. The mean square error was used as the loss function and the Adadelta algorithm was used as the optimization algorithm. The LSTM network has been built shown in Figure 3.

Figure 3: LSTM Network has been built



According to results shown in Figure 4, at the end of final epoch MSE was measured as 0.00014122. Rooted Mean Square Error (RMSE) could be calculated as 0.01192, by using this information.

Figure 4: Training Process



Finally, to investigate performance of model, we have calculated biases for test set and descriptive statistics of them. It could be clearly said that mean bias is 0.6502 and standard deviation of bias is 0.2009 by the aid of Table 2.

Table 2: Descriptive Statistics of Bias

Min	Max	Mean	Variance	Std. Deviation
0.0801	0.6374	0.6502	0.0004	0.2009

According to results shown in Table 3, the residential gross yield rates for the 210 sqm sized unit at the 6th floor of 12 floored building has 3 rooms, 2 bathrooms and with the age of 8 in Adana, Cukurova, Belediye Evleri could be calculated as 3.22% and for the 80 sqm sized apartment unit at the 4th floor of 4 floored and has 1 room, 1 bathroom and with the age of 35 in Istanbul, Besiktas, Abbasaga could be calculated as 5.2%.

Table 3: Some Estimated Gross Yield Rates

Location	N.of Room	N.of Bathroom	Size	Floor Number	Total Floor	Estimated Gross Yield Rate	Observed Gross Yield Rate
Adana, Cukurova, Belediye E.	3	2	210	6	12	3.22%	3.87%
Istanbul, Besiktas, Abbasaga	1	1	80	4	4	5.2%	5.28%

5. CONCLUSION

The shining star of the real estate sector, which has an important place in the economies of emerging countries, is the housing sector. It is very important to know how much a house will have a rate of return and to make investment according to this knowledge. With this information, investors have the chance to estimate in advance which property in which region has the highest return rate. Thus, investors can buy houses with a high rate of return.

The LSTM network generated in this study shows that the residential gross yield rate could be estimated using city, county, district, number of rooms, number of bathrooms, floor number, total floor information. This study also shows that the Spark framework can be used to cope with the growing volume of data in the real estate sector and to develop deep learning applications on distributed systems.

REFERENCES

Akerlof, G.A. (1970). The Market for "Lemons": Quality Uncertainty and the Market Mechanism, Quarterly Journal of Economics, vol. 84, no. 3, pp. 488-500.

Alessandri, T., Cerrato, D. and Depperu, D. (2014). Organizational slack, experience, and acquisition behavior across varying economic environments, Management Decision, vol. 56, no. 5, pp. 967-982.

Baker, B. (2001). Residential Rental Real Estate: An Investment in Need of a Theory. Pacific Rim Real Estate Society Conference, Christchurch, New Zealand, 20-23 January 2001.

Elephas: Distributed Deep Learning with Keras & Spark, https://github.com/maxpumperla/elephas

Goetzmann, W. N., Wachter, S. M. (1995). Clustering methods for real estate portfolios. Real Estate Economics, 23, 271-310.

Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., and Schmidhuber (2016). LSTM: A search space odyssey. IEEE transactions on neural networks and learning systems. 99, 1-11.

Jackson, C. (2002). Classifying local retail property markets on the basis of rental growth rates. Urban Studies, 39, 1417-1438.

Keras: Deep Learning Library for Theano and TensorFlow, https://keras.io/

Lee, S. (2001). The relative importance of property type and regional factors in real estate returns., Journal of Real Estate Portfolio Management, 7(2), 159-167.

Mahamad, A. K., Sharifah S., Takashi H. (2010). Predicting remaining useful life of rotating machinery based artificial neural network. Computers & Mathematics with Applications. 60, 1078-1087.

Ratchatakulpat T., Miller P., Marchant T. (2009). Residential real estate purchase decisions in Australia: is it more than location?, International Real Estate Review, 12, 273-294.

REIDIN, September 2017 Report, http://blog.reidin.com/reidin-turkey-real-estate-indices-september-2017-results/

TurkStat, Press Releases, http://www.turkstat.gov.tr/PreTabloArama.do?metod=search&araType=hb_x

Zaharia, M., Chowdhury, M., Franklin, M. J., Shenker, S., and Stoica, I. (2010). Spark: cluster computing with working sets. HotCloud, 10, 10-10.