

Real-Estate smart prediction system

Dr. Sherine Rady
Information Systems

Faculty of Computer Sciences Ain
Shams University
Cairo, Egypt
srady@cis.asu.edu.eg

Dr. Amal Moustafa
Information Systems

Faculty of Computer Sciences Ain
Shams University
Cairo, Egypt
aml.mostafa@cis.asu.edu.eg

Abeer Ahmed Sholkamy
Information Systems

Faculty of Computer Sciences Ain
Shams University
Cairo, Egypt
abeer20191700385@cis.asu.edu.eg

Osama Ahmed Ahmed Sayed
Information Systems

Faculty of Computer Sciences Ain
Shams University
Cairo, Egypt
osama20191700090@cis.asu.edu.eg

Shehab Adel Abdelgaleel
Information Systems

Faculty of Computer Sciences Ain
Shams University
Cairo, Egypt
shehab20191700316@cis.asu.edu.eg

Amany Adel Soliman
Information Systems

Faculty of Computer Sciences Ain
Shams University
Cairo, Egypt
amany20191700131@cis.asu.edu.eg

Abdelrahman Mohammed Ahmed
Information Systems

Faculty of Computer Sciences Ain
Shams University
Cairo, Egypt
abdelrahman20191700370@cis.asu.edu.eg

Esraa Mohammed Abdelhamid
Information Systems

Faculty of Computer Sciences Ain
Shams University
Cairo, Egypt
esraa20191700103@cis.asu.edu.eg

Abstract—A real estate management system is a software application that is designed to automate and streamline the various aspects of real estate transactions, including property listings, lead generation, marketing, contract management, and financial reporting. The system provides real estate professionals with a comprehensive set of tools and data to manage properties efficiently, improve market visibility, and increase transaction volume. The system can be used by various stakeholders in the real estate industry, including brokers, agents, property managers, investors, and buyers and sellers. Overall, a real estate management system aims to improve the efficiency, transparency, and security of real estate transactions, and to provide real estate professionals and clients with new opportunities and insights. This paper proposes a system that predicts house prices using a Xgboost machine learning algorithm[1], using Mutual Information Regression to measure the statistical dependence between two variables, the mutual information between the predictor variable(s) and the target variable is estimated, measure the performance by using Mean Squared Error and Mean Absolute Error.

I. INTRODUCTION

Every single organization in today's real estate business is operating fruitfully to achieve a competitive edge over alternative competitors. There is a need to simplify the process for a normal human being while providing the best results. This paper proposes a system that predicts house prices using a XGboost machine learning algorithm. In case you're going to sell a house, you have to recognize what sticker price to put on it[1].

The project holds immense importance and significance in today's digital age. With the rapid growth of the internet and technology, the real estate industry is evolving and becoming increasingly interconnected. An online management system revolutionizes the way properties are managed, offering a centralized platform for efficient data management, streamlined operations, and improved communication. By embracing this project, stakeholders in the real estate industry can harness the power of technology to enhance productivity, optimize financial management, provide a seamless tenant experience, and make data-driven decisions. It empowers the

industry to adapt to the changing needs of the market and leverage technology as a competitive advantage, ultimately leading to growth, profitability, and success in the dynamic world of real estate

II. RELATED WORKS

A. Propertybase

Propertybase is a real estate CRM and marketing platform that provides a comprehensive set of tools for managing real estate transactions. The platform includes features such as lead capture, property search, transaction management, and marketing automation.

B. Yardi

Yardi is a property management software that helps real estate professionals manage residential and commercial properties. The software includes features such as rent collection, maintenance tracking, tenant screening, and financial reporting.

C. CoStar

CoStar is a real estate information and analytics platform that provides access to a comprehensive database of commercial properties, sales and lease transactions, and market trends. The platform includes features such as property search, market analysis, and data visualization tools.

D. Buildout

Buildout is a marketing and deal management platform for commercial real estate professionals. The platform includes features such as property marketing, document management, and deal tracking.

E. RealPage

RealPage is a property management software that helps real estate professionals manage multifamily, single-family, and commercial properties. The software includes features such as leasing, maintenance, financial management, and marketing automation.

III. SYSTEM ARCHITECTURE

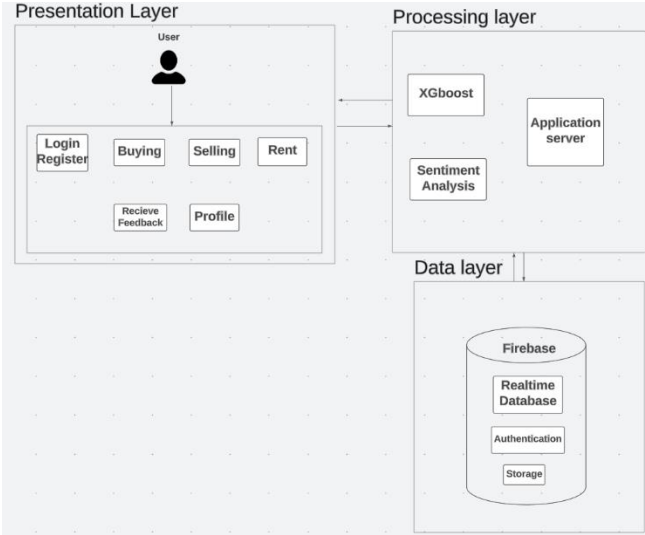


Fig. 1. System Architecture.

A. Presentation layer

It serves as the user interface or the front-end of the system, facilitating communication and interaction between users and the underlying system.

The main purpose of the presentation layer is to present data and information to users in a human-readable and meaningful format. It handles the transformation, formatting, and rendering of data received from the underlying layers, making it easily understandable and accessible to users. This layer is responsible for ensuring that the user interface is intuitive, visually appealing, and responsive.

Each user can login and register, can buy or sell estates, can either rent or offer his estate for rent, can edit his profile, can receive feedback.

B. Processing layer

the processing layer, also known as the business logic layer or application layer, is responsible for the core processing and functionality of the system. It serves as the intermediary between the presentation layer (user interface) and the data layer (data storage and retrieval).

The processing layer encapsulates the business rules, algorithms, computations, and decision-making logic that define how the system operates and processes data. It ensures that the system functions correctly, performs necessary operations, and enforces the business requirements.

Using XGboost to predict the price and using sentiment analysis to evaluate the rate of the user based on reviews.

C. Data layer

the data layer, also known as the data tier or data access layer, is responsible for managing the storage, retrieval, and manipulation of data within the system. It serves as the underlying infrastructure that handles data persistence and provides access to data for other layers of the system.

The data layer interacts directly with data sources, such as databases, file systems, or external data services, and abstracts the complexities of data storage and retrieval from the other layers.

Using Firebase: Real-time database, Authentication and Cloud storage.

IV. RESULTS

our study explored the relationship between property size and prices in a larger urban area. The findings revealed a significant positive association between the two variables, indicating that larger properties commanded higher prices. This suggests that buyers are willing to pay a premium for spacious homes, potentially due to the increased living space, greater privacy, and additional amenities that larger properties often offer. These findings have implications for homebuyers, real estate developers, and investors, emphasizing the value of larger properties in the housing market. It is important to note that other factors, such as location, condition, and neighborhood characteristics, may also influence property prices and should be considered in future research.

Using XGboost is better than ANN [4] to predict the price, it is high performance and speed, handling diverse data types, high interpretability, better to deal with missing data, high robustness against outliers, it avoids overfitting.

-measure the performance by using Mean Squared Error and Mean Absolute Error[5]

$$MSE = (1/n) * \sum (y_i - \bar{y})^2$$

where:

- MSE represents the Mean Squared Error.
- n is the total number of data points or observations.
- y_i represents the actual value of the i-th observation or data point.
- \bar{y} represents the predicted value of the i-th observation or data point.
- \sum denotes the summation symbol, indicating that you need to sum up the squared differences across all data points or observations.

$$MAE = (1/n) * \sum |y_i - \bar{y}|$$

where:

- MAE represents the Mean Absolute Error.
- n is the total number of data points or observations.
- y_i represents the actual value of the i-th observation or data point.
- \bar{y} represents the predicted value of the i-th observation or data point.
- $||$ denotes the absolute value function, which ensures that the differences between the actual and predicted values are positive.

Plot history using XGboost, where **X** axis represents epochs and **Y** axis represents mean squared error.

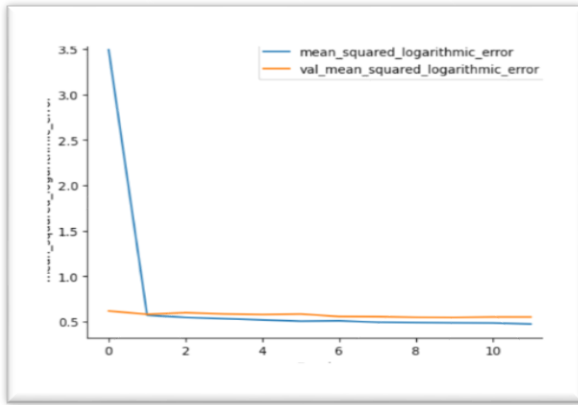


Fig. 2. Plot history.

TABLE I. COMPARISON BETWEEN ANN AND XGBOOST RESULTS

	ANN	XGboost
Mean absolute error	709162.66	1926584.31
Mean squared error	1334609082788.72	7823557174860.56
Median absolute error	428703.5	1337249.75
Explain variance score	0.56	0.69
R2 score	0.56	0.69

-using Mutual Information Regression to measure the statistical dependence between two variables, the mutual information between the predictor variable(s) and the target variable is estimated.

$$I(X;Y) = \sum \sum p(x,y) \log(p(x,y) / (p(x) * p(y)))$$

where:

- $p(x, y)$ is the joint probability distribution of X and Y (estimated from the data)
- $p(x)$ is the marginal probability distribution of X (estimated from the data)
- $p(y)$ is the marginal probability distribution of Y (estimated from the data)
- The summation is performed over all unique values of X and Y in the dataset.

TABLE II. IMPORTANCE OF FEATURE USING XGBOOST

Area	0.863977
Payment_Option	0.228052
Bathrooms	0.121363
Bedrooms	0.083963
Pool	0.050154
Elevator	0.030619
Pets_Allowed	0.029793

V. CONCLUSION

This paper proposes a system that predicts house prices using a Xgboost machine learning algorithm, using Mutual Information Regression to measure the statistical dependence between two variables, the mutual information between the predictor variable(s) and the target variable is estimated, measure the performance by using Mean Squared Error and Mean Absolute Error.

After completing the project with all the models achieving their goals, the main outcome resolved was that the area of a property is the most important factor in determining the price of the property.

The importance of this system comes from the market analysis, determining the pricing strategies and saving time for customers to find the appropriate property for them to buy or rent or even sell.

One of the drawbacks of this system is lack of the large datasets, large datasets help in making the predicting model more accurate and achieving that would help in making the customer experience better, this can be achieved by pipelining the stated that were sold/bought through the system to the dataset to increase model performance.

ACKNOWLEDGMENT

All praise and thanks to ALLAH, who provided me with the ability to complete this work. I hope to accept this work from me.

I am grateful of my parents and my family who are always providing help and support throughout the whole years of study. I hope I can give that back to them.

I also offer my sincerest gratitude to my supervisors, Prof. Dr. Sherine Rady and T.A. Amal Moustafa. who have supported me throughout my thesis with their patience, knowledge and experience.

Finally, I would like to thank my friends and all people who gave me support and encouragement..

REFERENCES

- [1] X. Ren, H. Guo, S. Li and S. Wang and J. Li, "A Novel Image Classification Method with CNN-XGBoost Model", International Publishing in Digital Forensics and Watermarking, pp. 378-390, 2017.
- [2] T. Chen and C. Guestrin, "XGBoost: "A Scalable Tree Boosting System"", Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, pp. 785-794, 2016.
- [3] Varma, A., Sarma, A., Doshi, S. et al.: House Price Prediction Using Machine Learning and Neural Networks. 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), 1936-1939 (2018).
- [4] Aswin Sivam Ravikumar, School of Computing , National College of Ireland, Real Estate Price Prediction Using Machine Learning , December 2017.
- [5] Dey, S. K., Urolagin, S.: Real Estate Price Prediction using Data Mining Techniques. 2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON), 1-4 (2021).

- [6] S.C. Dharmadhikari, Veerraju Gampala, Ch. Mallikarjuna Rao et al., "A smart grid incorporated with ML and IoT for a secure management system", *Microprocessors and Microsystems*, vol. 83, 2021.
- [7] Maclennan, D.: Some Thoughts on the Nature and Purpose of House Price Studies. *Urban Studies*, 14(1), 59–71 (1977).
- [8] Henry, C., Paul, D., Theo, D., Stephen A. J.: A spatio-temporal, Gaussian process regression, real-estate price predictor. In *Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (SIGSPACIAL '16)*. Association for Computing Machinery, New York, NY, USA, Article 68, 1–4 (2016).
- [9] Leo Breiman, Statistics Department, University of California, Berkeley, CA 94720, "RANDOM FORESTS", January 2001.
- [10] Debanjan Banerjee, Suchibrota Dutta, "Predicting the Housing Price Direction using Machine Learning Techniques", *IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI)*, 2017.