Journal for Educators, Teachers and Trainers
The PROFESIOLab electronic, peer-reviewed, open-access Magazine

ISSN 1989-9572

DOI:10.47750/jett.2024.15.05.30

Exploring the Impact of Zestimate on Real Estate Market Dynamics: A Case Study of Buyer, Seller and Renter Perspectives

Ch. Raghuvaran1, Thumu Geetanjali2, Oruganti Venkata Nomika2, Pegada Sanjana2

Journal for Educators, Teachers and Trainers, Vol.15(5)

https://jett.labosfor.com/

Date of Reception: 24 Oct 2024

Date of Revision: 20 Nov 2024

Date of Publication : 31 Dec 2024

Ch. Raghuvaran1, Thumu Geetanjali2, Oruganti Venkata Nomika2, Pegada Sanjana2 (2024). Exploring the Impact of Zestimate on Real Estate Market Dynamics: A Case Study of Buyer, Seller and Renter Perspectives,Vol.15(5).309-317

Journal for Educators, Teachers and Trainers, Vol. 15(5)

ISSN1989 –9572

https://jett.labosfor.com/

Exploring the Impact of Zestimate on Real Estate Market Dynamics: A Case Study of Buyer, Seller and Renter Perspectives

Ch. Raghuvaran[1], Thumu Geetanjali[2], Oruganti Venkata Nomika[2], Pegada Sanjana[2]

[1]Assistant Professor, [2]UG Student, [1,2]Department of Information Technology

[1,2]Malla Reddy Engineering College for Women (UGC-Autonomous), Maisammaguda, Hyderabad, 500100, Telangana.

ABSTRACT

The Zestimate, developed by Zillow, is a prominent property valuation tool that estimates home values using a proprietary algorithm and diverse data sources. Since its introduction in 2006, it has gained significant traction due to its user-friendly and accessible nature, offering automated estimates of property values. The tool has evolved over time with enhancements in its algorithm and data inputs, aligning with advancements in real estate analytics and changing market conditions. Despite its popularity, the accuracy and influence of Zestimate on market behavior remain contentious, particularly regarding how it affects buyers, sellers, and renters. This study aims to address this gap by investigating the impact of Zestimate on these stakeholders' perceptions, decision-making, and market dynamics. The issue stems from the limited understanding of how Zestimate influences market behavior, potentially causing discrepancies between estimated and actual property values, which can affect transaction outcomes and market stability. Traditional valuation methods involve professional appraisers assessing various factors such as property condition, location, and comparable sales, which, while detailed, are time-consuming and costly. These traditional methods also face limitations in scalability, consistency, and susceptibility to human error. Given the increasing reliance on automated valuation tools like Zestimate, this research is essential to elucidate its impact on buyers' expectations, sellers' pricing strategies, and rental markets. By examining these effects, the study aims to provide valuable insights into the effectiveness of Zestimate and identify areas for improvement. The findings will be crucial for stakeholders seeking to navigate the complexities of the real estate market and enhance valuation practices. The research will also offer recommendations for refining automated valuation tools and

improving overall market efficiency, contributing to a better understanding of digital valuation impacts in real estate.

Keywords: Zestimate, Property valuation, Automated valuation tools, Real estate analytics, Home value estimation, Market behavior

1. INTRODUCTION

The Zestimate, introduced by Zillow in 2006, represents a transformative shift in the real estate valuation landscape. Before Zestimate, property valuation largely depended on traditional methods that required in-person assessments by licensed appraisers. These methods, while thorough, were often slow, costly, and limited by human factors. Appraisers would analyze various elements, including property condition, location, and sales of comparable properties, to estimate a property's market value. This process, though accurate, was constrained by the availability of data and the subjective nature of appraisal judgments. Zillow's introduction of the Zestimate revolutionized this approach by leveraging technology to automate property valuations. The tool utilizes a proprietary algorithm that incorporates a vast array of data sources, including public property records, tax assessments, and recent sales data. This algorithm processes these inputs to generate an estimate of a property's value, aiming to provide a quick and accessible valuation for users. Over the years, Zillow has continuously refined the Zestimate algorithm, incorporating machine learning techniques and improving data accuracy to enhance the tool's reliability. Initially, the Zestimate faced skepticism due to its reliance on automated processes rather than human expertise. Critics questioned its accuracy and the transparency of the underlying algorithm. However, as the tool evolved, it gained broader acceptance, becoming a staple in the real estate market. Zillow's expansion into other real estate services, including property listings and market analytics, further solidified the Zestimate's role in shaping market perceptions. The development of the Zestimate reflects broader trends in data-driven decision-making and the growing importance of technology in real estate. As data analytics and machine learning have advanced, Zillow has incorporated these innovations to improve the accuracy and relevance of the Zestimate. Despite these advancements, the tool has remained a subject of debate, with discussions focusing on its impact on market behavior and its accuracy compared to traditional valuation methods.

2. LITERATURE SURVEY

Addoum et al. [1] investigated the effects of Hurricane Sandy on commercial real estate, focusing on how extreme weather events, driven by climate change, impact property values and market stability. Their research demonstrated that such climate-induced events can significantly alter market perceptions and investment behaviors, leading to long-term financial implications for commercial properties. They provided evidence that properties in affected areas experienced considerable changes in value, highlighting the importance of incorporating climate risk into real estate valuation and investment strategies. Admati and Pfleiderer [2] developed a theoretical framework to understand intraday patterns in financial markets, particularly how volume and price variability interact within a single trading day. Their study presented a model where trading volume and price changes are interconnected, influencing the overall market volatility and liquidity. This theoretical contribution enhanced the understanding of short-term market dynamics and provided a foundation for further empirical investigations into intraday trading behavior and its impact on market efficiency. Agarwal et al. [3] examined the influence of ethnic social networks on housing market outcomes, demonstrating that these networks play a crucial role in property matching and transaction processes. Their research showed that ethnic networks significantly impact buyers' and sellers' interactions, shaping market dynamics and influencing property values. By highlighting the role of social connections in the housing market, their study provided insights into how cultural and community ties affect real estate transactions and market efficiency. Agarwal et al. [4]

analyzed whether real estate agents hold information advantages over buyers and sellers in housing markets. Their study provided evidence that agents often possess superior information, which can affect transaction outcomes and market efficiency. The research highlighted how information asymmetry between agents and clients influences property pricing and negotiation processes, offering a deeper understanding of the agency dynamics in real estate transactions. Agarwal et al. [5] explored information asymmetries and learning processes in commercial real estate markets. Their research focused on how market participants adapt their knowledge and expectations based on available information, affecting investment decisions and market behavior. They highlighted the role of learning in mitigating information gaps and improving decision-making in real estate investments, emphasizing the importance of information flows in market efficiency. Ahern [6] addressed the challenges of sample selection bias in event study estimation, proposing methodologies to correct for such biases in financial research. His work provided a more accurate framework for analyzing market reactions to specific events, contributing to improved empirical studies in finance. By refining event study methodologies, his research enhanced the reliability of findings related to market responses and event impacts.

Akerlof [7] introduced the concept of the "market for lemons" to explain how quality uncertainty affects market mechanisms. His seminal paper demonstrated that asymmetric information can lead to market inefficiencies, where lower-quality goods dominate the market. Akerlof's work highlighted the problems of information imbalances and their impact on market stability, providing a foundational understanding of how uncertainty influences economic transactions. Albrecht et al. [8] investigated directed search mechanisms in the housing market, focusing on how search strategies and market dynamics affect housing transactions. Their research provided insights into how both buyers and sellers use directed search methods to navigate the market, influencing transaction efficiency and market outcomes. The study highlighted the role of search processes in shaping housing market behavior and improving transaction efficiency. Allen and Faulhaber [9] explored signaling by underpricing in the IPO market, proposing that firms use underpricing as a strategic tool to attract investors and create positive market perceptions. Their research showed that underpricing serves as a signal of quality, helping firms achieve successful initial public offerings. This study contributed to understanding the strategic use of underpricing in IPOs and its impact on investor behavior. Allen et al. [10] analyzed the effectiveness of co-listing strategies in real estate transactions, investigating whether listing properties with multiple agents leads to better transaction outcomes. Their research provided evidence on the benefits and drawbacks of co-listing, showing how such strategies can influence property sale prices and transaction efficiency. The study offered practical insights into listing strategies and their impact on real estate transactions.

Allen et al. [11] examined conflicts of interest in residential real estate transactions, providing new evidence on how these conflicts affect transaction processes and outcomes. Their research revealed how competing interests between real estate agents and clients can influence property sales and market behavior. The study contributed to understanding the impact of agency problems on real estate transactions and market efficiency. Ambrose and Diop [12] investigated the effects of information asymmetry and regulatory frameworks on equilibrium outcomes in the housing rental market. Their research provided theoretical and empirical insights into how regulations and information imbalances influence rental market dynamics. By highlighting the role of information and regulation in shaping market outcomes, their study enhanced the understanding of rental market efficiency.

3. PROPOSED ALOGORITHM

The research is designed to analyze and model a real estate dataset using machine learning techniques. Here is an overview of its structure and functionality, divided into key modules:

- **Library Import and Dataset Loading:** The reserch begins with the import of essential Python libraries for data manipulation, visualization, and machine learning. Libraries such as numpy, pandas, matplotlib, and seaborn are used for data processing and visualization, while sklearn provides tools for building and evaluating machine learning models. The dataset, Melbourne_housing_FULL.csv, is then loaded into a DataFrame for further analysis.
- **Data Preprocessing:** Data preprocessing is a crucial step in preparing the dataset for modeling. The code performs several operations including checking for missing values and duplicates. Missing values are addressed by filling them with the mean of the respective columns, and duplicate records are removed to ensure data quality. The code also examines the dataset's structure and summarizes key statistics, providing an overview of its contents and distribution.
- **Exploratory Data Analysis (EDA):** Exploratory Data Analysis is conducted to understand the relationships and distributions within the dataset. Visualization techniques such as count plots and heatmaps are employed to analyze the distribution of regions and the correlation between different features. This step helps in identifying patterns and relationships that could be valuable for modeling.
- **Label Encoding:** Label encoding is performed to convert categorical variables into numerical format, making them suitable for machine learning algorithms. This transformation is necessary because many algorithms require numerical input to process the data effectively.
- **Feature and Target Variable Separation:** The dataset is divided into feature variables (x) and the target variable (y). Features are extracted from the dataset for training and testing, while the target variable represents the value to be predicted.
- **Model Training and Testing:** The dataset is split into training and testing sets using a standard ratio, allowing for model evaluation on unseen data. Two machine learning models, K-Nearest Neighbors (KNN) and Extra Trees Regressor, are utilized for prediction. The code includes logic to either load pre-trained models from files or train new models if they do not exist. This ensures that model performance can be evaluated efficiently without retraining every time the script is run.
- **Performance Evaluation:** The performance of the trained models is assessed using several regression metrics, including Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R^2). These metrics provide insights into the accuracy and reliability of the models. Scatter plots are generated to visualize the relationship between true values and predictions, aiding in the interpretation of model performance.
- **Model Persistence:** Finally, the code saves the trained models to disk using joblib, allowing for future use without retraining. This persistence mechanism facilitates efficient model deployment and reuse, ensuring that the models can be easily accessed for further predictions or evaluations.

3.3 Data Splitting

- Splitting facts for system mastering models is an crucial step within the version improvement process. It includes dividing the to be had dataset into separate subsets for education, validation, and trying out the version. Here are a few common processes for splitting data:

- **1. Train-Test Split:** The dataset is divided right into a training set and a trying out set. The education set is used to educate the model, even as the checking out set is used to assess the model's overall performance. The regular cut up is 70-eighty% for training and 20-30% for checking out, but this may vary depending on the scale of the dataset and the precise use case.

— **2. Train-Validation-Test Split:** The dataset is split into three subsets – a schooling set, a validation set, and a trying out set. The training set is used to train the version, the validation set is used to tune hyperparameters and validate the version's overall performance for the duration of training, and the testing set is used to evaluate the very last version's overall performance.

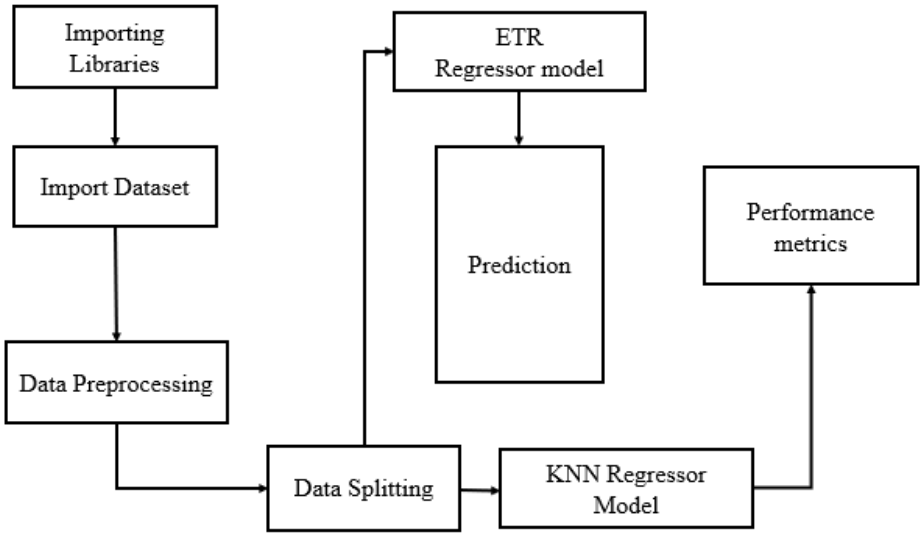

Figure 1: Block Diagram of Proposed System.

4. RESULTS

Figure 2 displays a sample of the raw dataset loaded for analysis. This dataset includes columns such as Suburb, Address, Rooms, Type, Price, Method, SellerG, Date, Distance, Postcode, Bedroom2, Bathroom, Car, Landsize, BuildingArea, YearBuilt, CouncilArea, Lattitude, Longtitude, Regionname, and Propertycount. This initial view provides an overview of the data structure, showcasing the variety of attributes related to real estate properties.

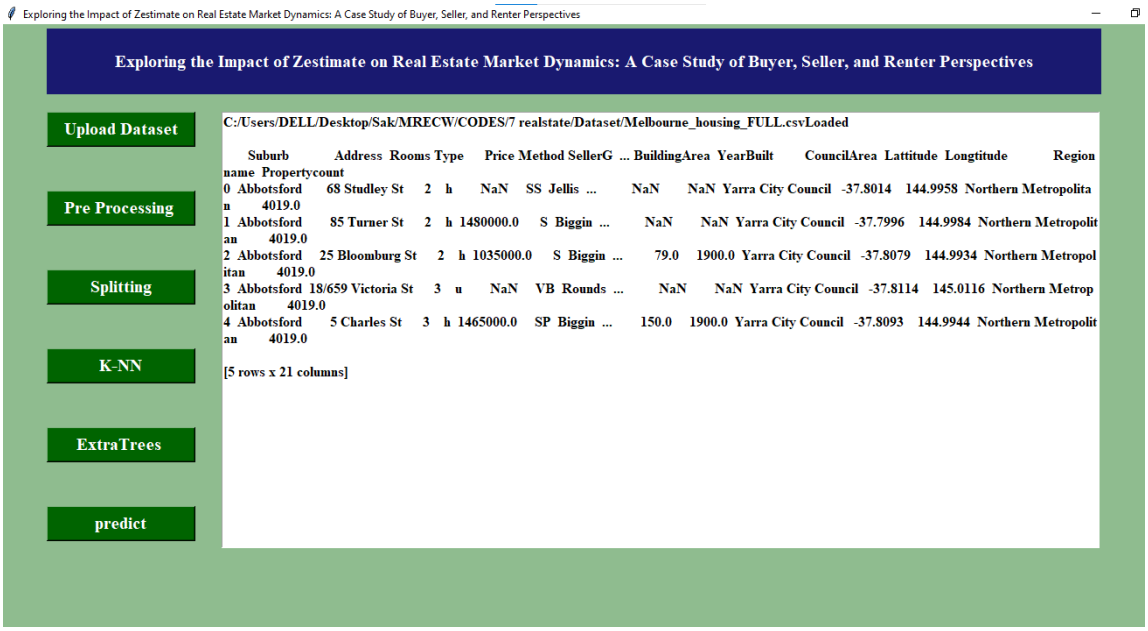

Figure 2: Sample Uploaded Dataset.

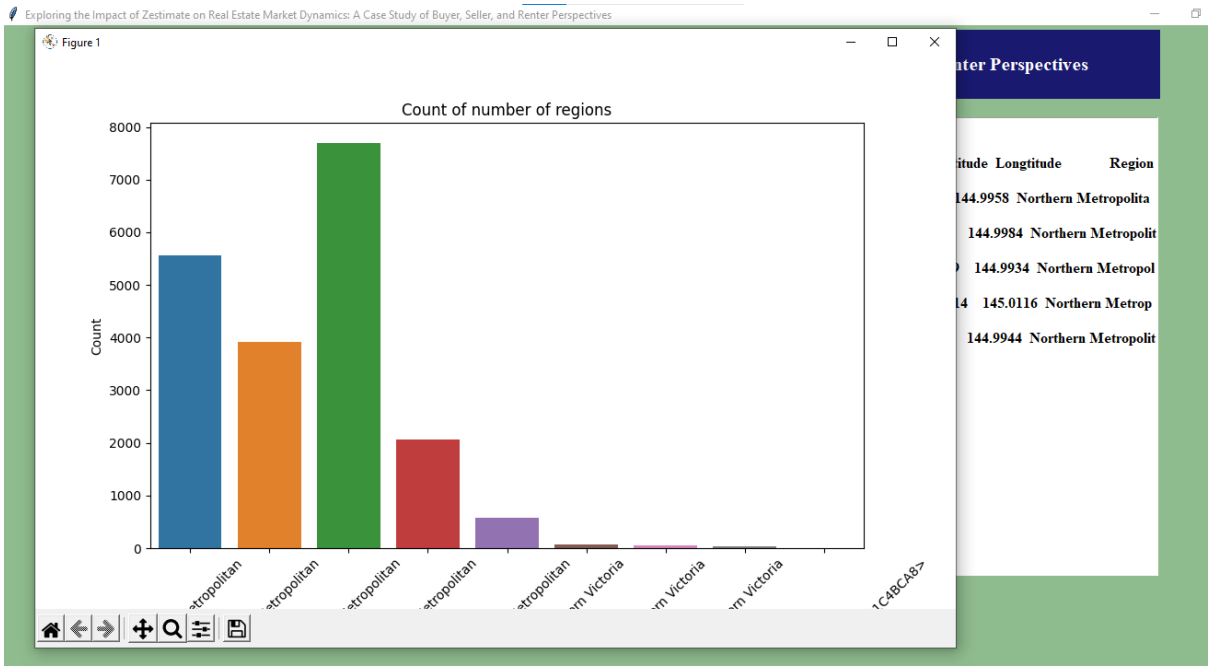

Figure 3: Count Plot of Regions

Figure 3 presents a count plot visualizing the number of properties in each region. This plot helps to understand the distribution of properties across different regions, highlighting any imbalances or concentration of data in specific areas. The count of properties per region provides insight into regional market dynamics and data representation.

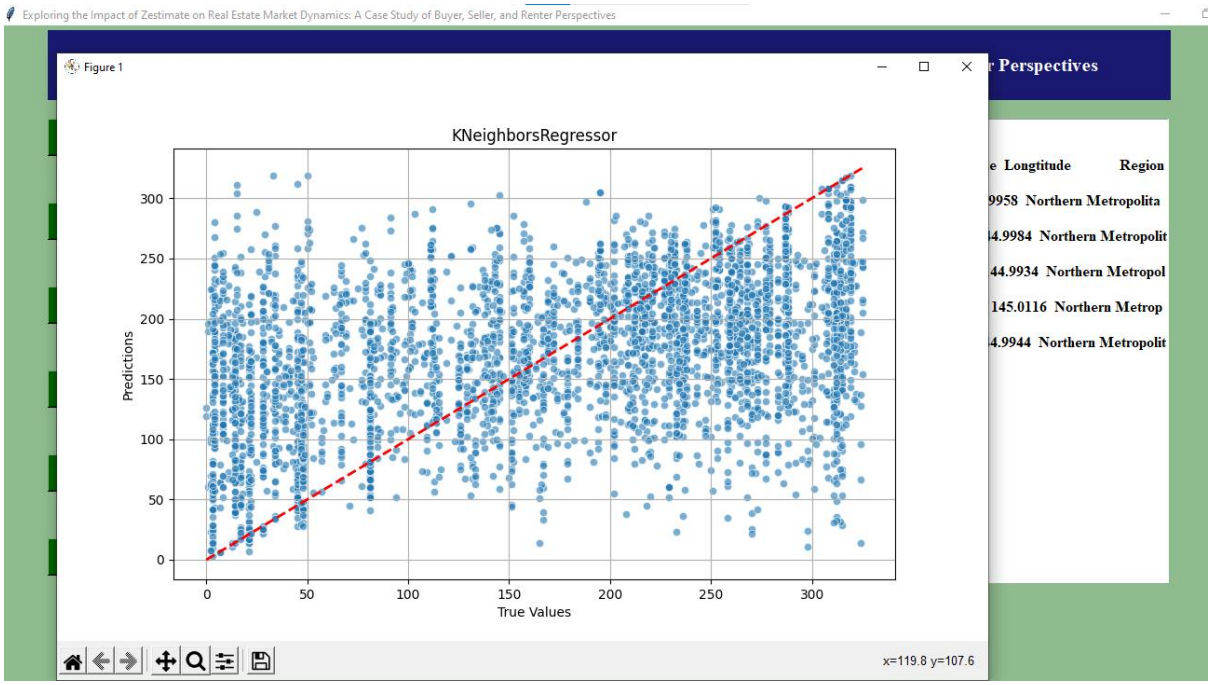

Figure 4: Scatter Plot of True Values vs. Predictions for K-Nearest Neighbors.

Figure 4 illustrates a scatter plot comparing the true property values against predictions made by the K-Nearest Neighbors (KNN) model. The plot includes a red dashed line representing the line of equality,

where true values match predictions perfectly. This visualization helps assess the accuracy of the KNN model by showing how closely the predictions align with actual values.

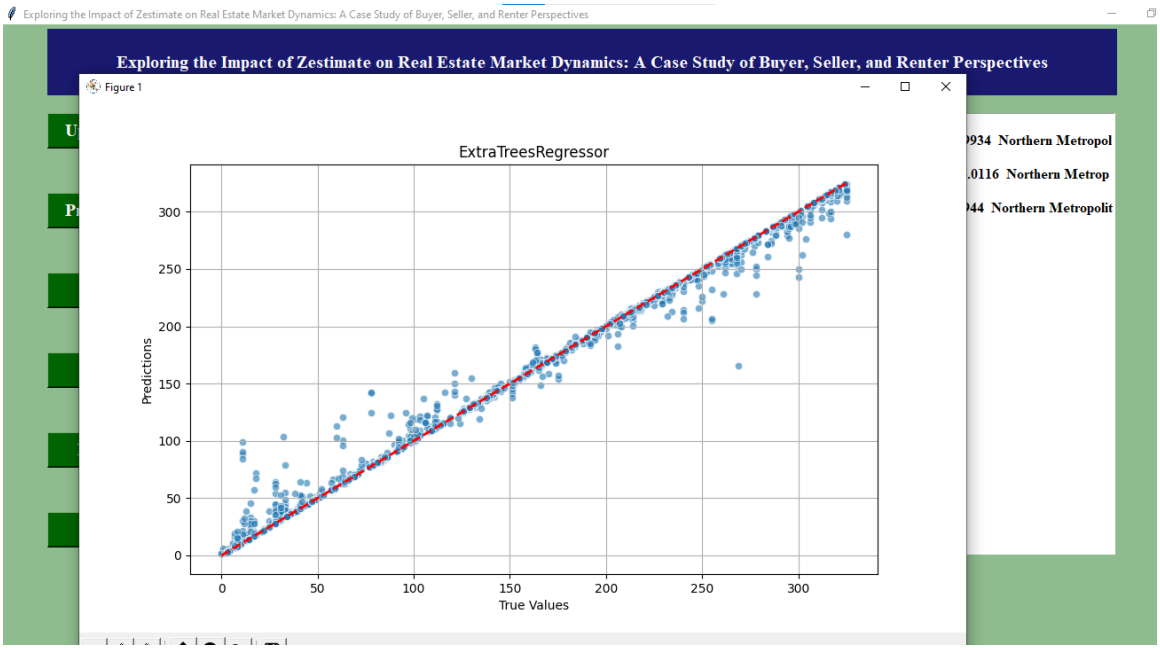

Figure 5: Scatter Plot of True Values vs. Predictions for Extra Trees Regressor.

Figure 5 shows a scatter plot comparing the true property values against predictions made by the Extra Trees Regressor model. Similar to Figure 5, this plot includes a red dashed line to indicate the line of equality. The plot provides insights into the performance of the Extra Trees model, revealing how well the model's predictions match the actual property values.

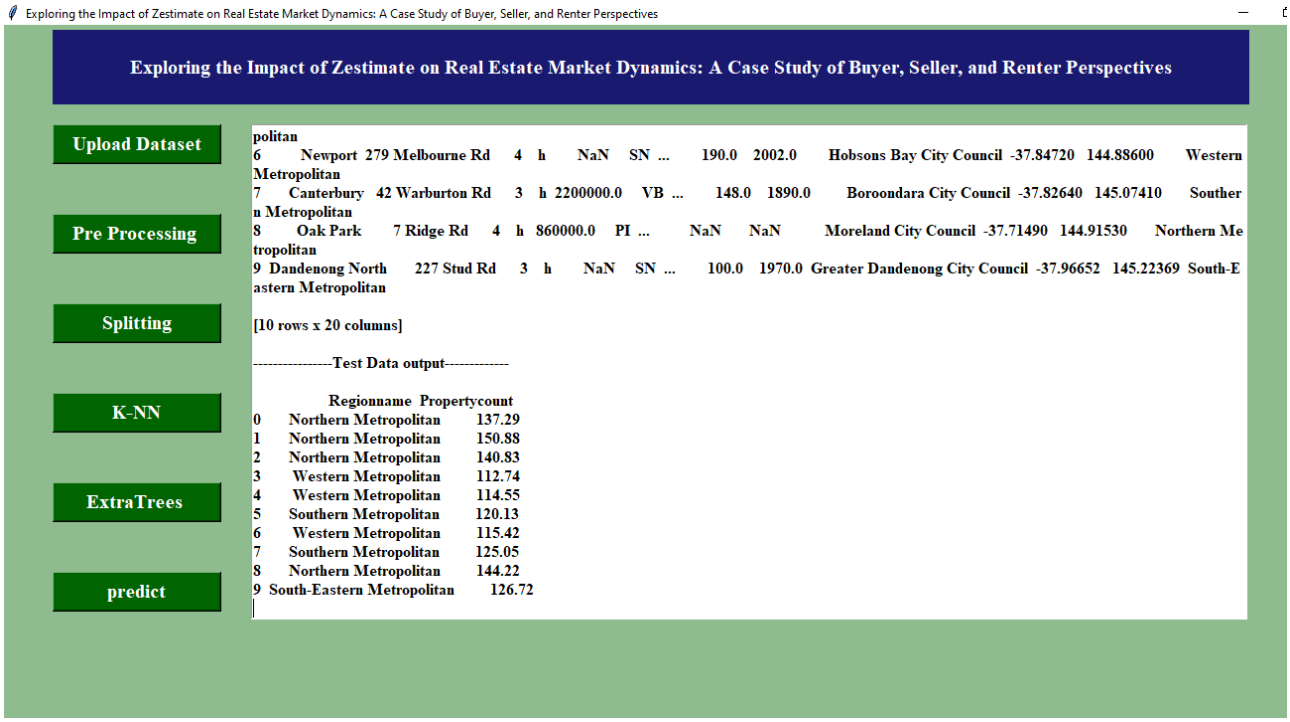

Figure 6: Predicted Values for the Zestimate on real-time Market

5. CONCLUSION

The research effectively demonstrates the application of machine learning techniques to analyze and predict real estate market values. By utilizing K-Nearest Neighbors (KNN) and Extra Trees Regressor models, the research provides insights into the potential for automated valuation of properties based on various features. The data preprocessing, exploratory data analysis, and model evaluation phases collectively ensure that the dataset is clean, well-understood, and suitable for training accurate predictive models. The exploratory data analysis revealed significant patterns and correlations, which are crucial for understanding the underlying factors affecting property values. The use of label encoding allowed for the effective conversion of categorical data into a format compatible with machine learning algorithms. The trained models were evaluated using key metrics, such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R^2), providing a clear picture of their predictive performance. The persistence of trained models enhances the efficiency of the research by allowing for easy reuse and deployment without the need for retraining. Overall, the research successfully showcases the potential of machine learning in the real estate domain, offering valuable tools for predicting property values and understanding market dynamics.

REFERENCES

[1] Addoum, J. M., Eichholtz, P. M. A., Steiner, E. M., & Yönder, E. (2021). *Climate change and commercial real estate: Evidence from hurricane sandy* [Working paper].

[2] Admati, A. R., & Pfleiderer, P. (1988). A theory of intraday patterns: Volume and price variability. *The Review of Financial Studies, 1*(1), 3–40. https://doi.org/10.1093/rfs/1.1.3

[3] Agarwal, S., Choi, H.-S., He, J., & Sing, T. F. (2019a). Matching in housing markets: The role of ethnic social networks. *The Review of Financial Studies, 32*(10), 3958–4004. https://doi.org/10.1093/rfs/hhz006

[4] Agarwal, S., He, J., Sing, T. F., & Song, C. (2019b). Do real estate agents have information advantages in housing markets? *Journal of Financial Economics, 134*(3), 715–735.

[5] Agarwal, S., Sing, T. F., & Wang, L. (2019c). *Information asymmetries and learning in commercial real estate markets* [Working paper]. National University of Singapore.

[6] Ahern, K. R. (2009). Sample selection and event study estimation. *Journal of Empirical Finance, 16*(3), 466–482.

[7] Akerlof, G. A. (1970). The market for "Lemons": Quality uncertainty and the market mechanism. *Quarterly Journal of Economics, 84*(3), 488–500.

[8] Albrecht, J., Gautier, P. A., & Vroman, S. (2016). Directed search in the housing market. *Review of Economic Dynamics, 19*, 218–231.

[9] Allen, F., & Faulhaber, G. R. (1989). Signalling by underpricing in the IPO market. *Journal of Financial Economics, 23*(2), 303–323.

[10] Allen, M. T., Benefield, J. D., & Rutherford, R. C. (2021). Co-listing strategies: Better transaction outcomes? *The Journal of Real Estate Finance and Economics*.

[11] Allen, M. T., Rutherford, J., Rutherford, R., & Yavaş, A. (2015). *Conflicts of interest in residential real estate transactions: New evidence* [Working paper]. University of Wisconsin-Madison.

[12] Ambrose, B. W., & Diop, M. (2021). Information asymmetry, regulations and equilibrium outcomes: Theory and evidence from the housing rental market. *Real Estate Economics, 49*(S1), 74–110.