International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Accurate House Price Projections

Ravindra Turpati, S. Sai Krishna, Ajit Kumar Rout

Department of IT, GMRIT

DOI: https://doi.org/10.55248/gengpi.5.0324.0798

ABSTRACT—

This study delves into the dynamics of housing prices in Visakhapatnam, employing a multifaceted approach that combines linear regression modelling, geocoding, and K-means clustering. Leveraging a comprehensive dataset encompassing diverse locations, area (in square feet), and bedroom-bathroom configurations, we construct a robust linear regression model for price estimation. The dataset is further enriched through geocoding, introducing latitude and longitude coordinates to each location. Expanding our analysis, we employ K-means clustering to unveil spatial patterns and groupings within the housing market. Our findings reveal intricate relationships between location, property attributes, and housing prices, providing valuable insights for both prospective buyers and real estate stakeholders. Visualizations on the Visakhapatnam map illustrate the spatial distribution of housing clusters, shedding light on localized market trends. This integrated approach not only refines price predictions but also enhances our understanding of the spatial dynamics shaping the real estate landscape in Visakhapatnam. The outcomes of this study contribute to informed decision-making in the housing market and set the stage for future research exploring the nuances of regional real estate dynamics.

Index Terms—**house price projection, Geocoding, Linear Regression, K-means clustering, Spatial data.**

I. INTRODUCTION

The multifaceted nature of the real estate market in Visakhapatnam necessitates a nuanced exploration, considering the diverse factors that shape property dynamics. In this study, we delve into a sophisticated analytical framework, leveraging cutting-edge techniques to decode the intricacies of housing prices in this rapidly growing city[1]. Our approach begins with an in-depth examination of key attributes influencing property valuation. Parameters such as the size of the property (measured in square feet), the number of bedrooms and bathrooms, and the geographic location collectively contribute to the complex web of factors that determine housing prices. To decipher these relationships, we meticulously construct a robust linear regression model, finely tuned to provide not only accurate predictions but also insights into the nuanced interplay of features affecting property values. Taking a step further, our study incorporates a bespoke function designed to calculate prices for diverse locations within Visakhapatnam. This tool serves as a valuable resource for prospective buyers and industry stakeholders, offering a user-friendly interface to navigate the intricacies of property valuation based on specific geographic parameters. Recognizing the spatial dimension as a crucial aspect of real estate dynamics, we enrich our dataset by integrating geocoded information. Longitude and latitude coordinates are seamlessly integrated, providing a geographic context to each housing data point[2]. This spatial integration allows for a more nuanced exploration of how different locations influence housing prices, contributing to a comprehensive understanding of the city's real estate landscape. To unravel deeper insights into the spatial dynamics, we employ K-means clustering, an advanced unsupervised machine learning technique. This method identifies inherent patterns and groupings within the dataset, shedding light on spatial clusters and trends that characterize Visakhapatnam's real estate market. The combination of linear regression modelling, geocoding, and K-means clustering enhances the depth of our analysis, providing a holistic view of both the determinants of housing prices and the spatial nuances that define the real estate arena in Visakhapatnam. The outcomes of our research extend beyond academic inquiry, offering practical implications for various stakeholders[3]. Prospective buyers gain valuable insights into the factors influencing property values, while sellers can better understand market trends. Policymakers, armed with a nuanced understanding of the real estate landscape, are empowered to make informed decisions that shape the future trajectory of Visakhapatnam's real estate sector. This study, therefore, serves as a valuable resource for navigating the complexities of the real estate arena in this vibrant and evolving city.

II.MOTIVATIONS

The increasing integration of GIS technology into computer engineering has sparked a revolution in how spatial data is utilized and visualized. In today's data-centric world, where information reigns supreme, the ability to effectively visualize spatial data is essential. GIS technology serves as a cornerstone in this endeavour, providing powerful tools to imbue spatial data with geographical context and meaning. With the proliferation of location-aware devices and sensors, vast amounts of data are now inherently tied to geographic locations. GIS technology enables us to harness this wealth of spatial data and transform it into actionable insights through visualization techniques. By leveraging GIS operations, data scientists and engineers can create compelling visualizations that reveal spatial patterns, trends, and relationships, empowering decision-makers across various industries[4]. The primary focus of this paper is to explore the synergy between machine learning and GIS technology to enhance predictive capabilities. As datasets become increasingly complex and multidimensional, predictive analysis becomes indispensable for extracting meaningful insights. Machine learning algorithms excel at uncovering patterns and relationships within data, making them invaluable tools for predictive modelling. By integrating machine learning with GIS technology, researchers can unlock new possibilities for predictive analysis in spatial contexts. This fusion enables the development of sophisticated predictive models that leverage both spatial and non-spatial data, providing a more comprehensive understanding of complex phenomena[5].

While the efficacy of machine learning in predictive analysis is widely recognized, its application within the realm of GIS technology is still relatively unexplored. This paper seeks to bridge this gap by elucidating the relevance and potential of machine learning techniques within the domain of GIS. By showcasing real-world applications and case studies, we aim to demonstrate the transformative impact of integrating machine learning with GIS technology. In doing so, we hope to inspire researchers and practitioners to explore new avenues for leveraging machine learning in spatial analysis and visualization. By harnessing the combined power of machine learning and GIS technology, we can unlock new insights, drive innovation, and solve complex spatial challenges in fields ranging from urban planning and environmental science to transportation and public health.

So, the fusion of machine learning and GIS technology holds immense promise for enhancing predictive capabilities and unlocking new insights in spatial analysis. By embracing this interdisciplinary approach, we can harness the full potential of spatial data to address pressing challenges and drive positive change in our increasingly interconnected world[6].

III.LITERATURE SURVEY

This study aims to contribute significantly to the understanding of the housing market dynamics in Visakhapatnam through a multifaceted approach. The exploration encompasses insights from various domains, such as real estate economics, housing pricing models, geocoding applications, and clustering techniques. By synthesizing existing literature, we aim to provide a comprehensive foundation for our analytical framework and contextualize our research within the broader landscape of real estate studies[7].

Foundational Insights from Real Estate Economics:

Foundational works in real estate economics, such as those by Case, have illuminated the complex interplay of factors that influence housing markets. Case's seminal research highlights the significance of economic indicators, property characteristics, and external influences in shaping the real estate landscape. Understanding these fundamental principles is crucial for framing our study within the broader context of real estate dynamics in Visakhapatnam[8]. We delve into how economic forces, market trends, and consumer behaviour intertwine to impact the housing market in this dynamic city.

Advancements in Housing Pricing Models:

Recent studies in the field of real estate economics have propelled the application of advanced pricing models, aiming to enhance the accuracy of housing price predictions. Linear regression models, a cornerstone in this advancement, allow for a detailed examination of how attributes like area and bedroom-bathroom configurations contribute to property valuations. The nuanced relationships between these features and housing prices form the bedrock of our analytical framework, ensuring a robust understanding of the factors driving the real estate market in Visakhapatnam.

Exploring Geocoding Applications:

Geocoding applications have become pivotal in understanding the spatial dynamics of real estate markets. Spatial analysis and Geographic Information System (GIS) applications offer insights into location-based dynamics, reflecting a growing trend in real estate studies. The integration of geocoded information in our study enriches the dataset, providing a spatial context to each housing data point. This allows us to explore how specific geographic locations within Visakhapatnam influence housing prices, adding depth to our understanding of the city's real estate dynamics.

Uncovering Spatial Patterns through K-means Clustering:

K-means clustering techniques add another layer of sophistication to our analysis. As an unsupervised machine learning technique, K-means clustering aids in uncovering spatial patterns within the dataset. This aligns with the contemporary focus on spatial dynamics and urban studies, where understanding the clustering of properties contributes to a more nuanced comprehension of the real estate landscape. Our application of K-means clustering seeks to reveal spatial clusters and trends that characterize Visakhapatnam's housing market, providing insights into how different areas are shaping the city's real estate narrative[9]

Localizing Insights: Visakhapatnam's Real Estate Market:

While broader real estate literature provides essential insights, localized research on Visakhapatnam's real estate market is equally pivotal. City-specific studies complement the broader literature by offering insights into the unique dynamics of Visakhapatnam. Factors such as local economic conditions, cultural influences, and infrastructure development play a crucial role in shaping the housing market in this specific context. By understanding how global principles apply to the intricacies of Visakhapatnam, we aim to provide a nuanced and tailored perspective on the city's housing dynamics.

Interdisciplinary Approach for comprehensive Understanding:

Interdisciplinary approaches in real estate research have gained recognition for their ability to integrate statistical models with spatial analysis. By combining insights from real estate economics, advanced pricing models, geocoding applications, and clustering techniques, our study follows an interdisciplinary path. This approach not only acknowledges the complexity of the real estate landscape but also positions our research to provide holistic insights into how various factors converge to shape Visakhapatnam's unique housing market.

Practical Implications and Stakeholder Guidance:

Emphasizing the practical implications of research findings is crucial for guiding stakeholders in navigating Visakhapatnam's complex real estate landscape. Prospective buyers seeking valuable insights into property values, sellers strategizing based on market trends, and policymakers making informed decisions all benefit from the practical applications of our study. The nuances uncovered in our research offer tangible guidance for stakeholders, ensuring that they can navigate the intricate real estate landscape of Visakhapatnam with informed decision-making tools.

This literature review establishes the groundwork for our study, aiming to unravel the housing market dynamics in Visakhapatnam. By drawing on foundational insights from real estate economics, incorporating advancements in housing pricing models, exploring geocoding applications, and leveraging K-means clustering techniques, our research follows a comprehensive and interdisciplinary approach. The synthesis of global principles with localized insights positions our study to contribute significantly to the understanding of Visakhapatnam's housing market[10]. As we embark on this exploration, the practical implications of our findings underscore the relevance of our research for stakeholders navigating the intricate real estate landscape of Visakhapatnam.

IV.EXPLORING HOUSING PRICE DYNAMICS THROUGH INTEGRATED LINEAR REGRESSION, GEOCODING, AND K-MEANS CLUSTERING.

1. Data Collection:

we outline the rigorous data collection process undertaken to ensure the robustness and comprehensiveness of our study on housing price dynamics in Visakhapatnam. We meticulously gathered data on housing prices from multiple reliable sources, including real estate websites, property listings, local real estate agents, and government records. This dataset encompasses a wide range of residential properties across various locations within Visakhapatnam. We collected detailed information on the attributes of properties, including square footage, number of bedrooms and bathrooms, property type, age, amenities, and any other features pertinent to their valuation. This comprehensive dataset enables a thorough analysis of the factors influencing housing prices in the city. Geographic information, including latitude and longitude coordinates, was obtained for each property using advanced geocoding techniques. This geospatial data is crucial for spatial analysis and visualization, allowing us to explore the spatial distribution of housing prices and identify localized patterns within Visakhapatnam. Historical housing price data spanning a significant period was sourced to analyse long-term trends and patterns in the Visakhapatnam real estate market. This historical perspective enhances our understanding of the dynamics driving price fluctuations over time.

2. Data Preprocessing:

Prior to analysis, the collected data underwent rigorous preprocessing procedures to ensure accuracy, consistency, and reliability. This involved handling missing values, removing outliers, standardizing data formats, and resolving any inconsistencies or discrepancies within the dataset. By meticulously curating a diverse array of data sources encompassing housing prices, property attributes, geographic information, neighbourhood characteristics, historical trends, and regulatory factors, we have assembled a robust dataset that serves as the foundation for our comprehensive analysis of housing price dynamics in Visakhapatnam.

3.Grid Search Cross-Validation

To choose a machine learning algorithm, the Grid Search Cross-Validation technique is employed. This hyperparameter optimization technique is commonly used in machine learning to systematically search for the optimal set of hyperparameters for a given model. It exhaustively evaluates all possible combinations of hyperparameters within a predefined grid of values and selects the combination that produces the best performance according to a specified evaluation metric. Using this technique, the mean cross-validated score for different models can be calculated. In this study, Linear Regression, Lasso Regression, and Decision Tree are considered, and it is observed that Linear Regression has a better mean cross-validated score. Therefore, Linear Regression is considered for model building over remaining algorithms(Lasso Regression and Decision Tree algorithms)[11].

TABLE 1:Comparison of different models

	algorithm	best_score	best_paramts
0	linear_regression	0.820935	{'copy_X': True, 'fit_intercept': False, 'n_jo...
1	lasso	0.681665	{'alpha': 1, 'selection': 'random'}
2	decision_tree	0.706967	{'criterion': 'friedman_mse', 'splitter': 'best'}

4. Linear Regression Modelling:

Identified key variables such as square footage, number of bedrooms and bathrooms, property location attributes, and other relevant factors affecting housing prices in Visakhapatnam. Constructed a robust linear regression model using the collected dataset to estimate housing prices based on the selected variables. Employed techniques such as feature engineering, regularization, and model validation to ensure model accuracy and reliability. Assessed the performance of the linear regression model through statistical metrics such as R-squared, mean squared error, and significance of coefficients. Conducted diagnostic tests to validate model assumptions and identify potential issues. Developed a user-friendly function incorporating the constructed linear regression model to estimate housing prices across all locations within Visakhapatnam based on user-specified parameters such as square footage, number of bedrooms, and property location attributes[12]. Implemented the price estimation function using appropriate programming languages or software tools to enable seamless application and accessibility for stakeholders, such as prospective buyers and real estate professionals.

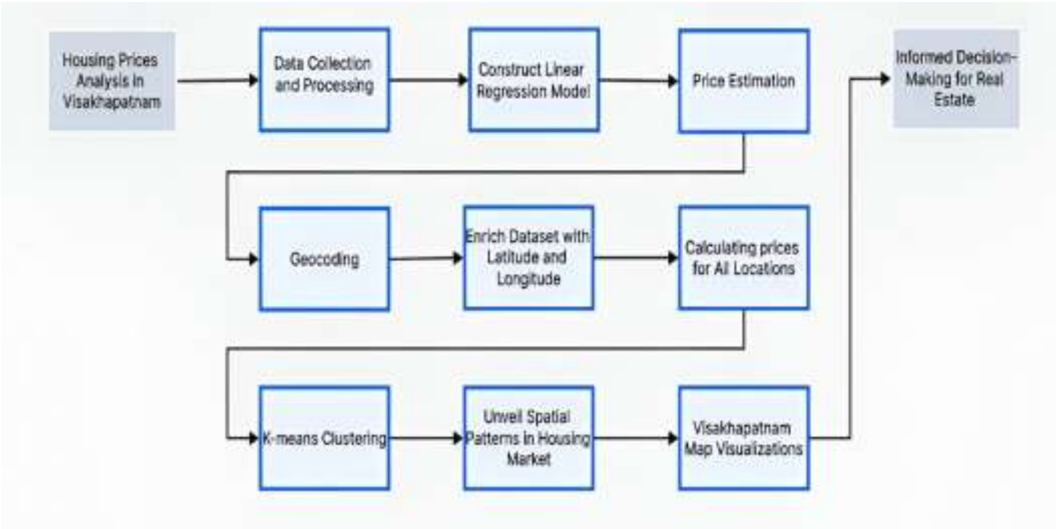

FIGURE 1:Proposed workflow diagram to practically realize functioning of the architecture.

5. Geocoding Enrichment:

Enriched the housing price dataset by adding latitude and longitude coordinates to each property location using geocoding techniques. Ensured accurate spatial representation of housing data for subsequent analysis and visualization. Leveraged the geocoded dataset to conduct spatial analysis, including mapping housing price distributions, identifying spatial patterns, and examining geographical variations in housing market dynamics across Visakhapatnam.

6. K-means Clustering:

Applied the K-means clustering algorithm to segment properties into distinct clusters based on similarities in housing price patterns and property attributes. Determined the optimal number of clusters using statistical techniques such as the elbow method or silhouette analysis. Analysed the characteristics of each cluster to identify spatial patterns, groupings, and trends within the Visakhapatnam housing market. Examined cluster centroids and cluster assignments to understand the distribution of properties across different market segments[13].

7. Spatial Visualization:

Created visualizations, such as heatmaps or choropleth maps to illustrate the spatial distribution of housing clusters, price gradients, and spatial patterns identified through clustering and geospatial analysis. Developed interactive maps or dashboards using tools like Tableau or Python libraries (e.g., Folium) to allow stakeholders to explore and interact with housing market data, facilitating deeper insights into localized market trends. These tools allow to explore more efficiently[14].

8. Analysis of Findings:

Analysed the results of linear regression modelling, geocoding enrichment, and K-means clustering to uncover insights into the relationships between location, property attributes, and housing prices in Visakhapatnam. Extracted meaningful insights and trends from the data analysis, highlighting factors influencing housing prices, spatial variations in market dynamics, and implications for prospective buyers and real estate stakeholders.

The proposed workflow begins with collecting house price data from various sources and processing it by replacing any missing values with mean values or entirely removing them. Then, all this data collected from diverse sources is concatenated and prepared for preprocessing. This data consists of area (in square feet), bathroom-bedroom configurations, locations, and corresponding prices associated with these features. Since locations consist of nominal values, they are preprocessed using one-hot encoding, which converts categorical variables into a numerical format[15]. After the data is preprocessed and prepared for model building, it is split into two parts: one for input (X), consisting of area, bathroom-bedroom configurations, and locations, and the other for output (Y), consisting of prices. These two parts are further divided into training data and testing data in a 4:1 ratio, with 80% allocated to the training set and 20% to the testing set.

To choose a machine learning algorithm, the Grid Search Cross-Validation technique is employed. This hyperparameter optimization technique is commonly used in machine learning to systematically search for the optimal set of hyperparameters for a given model[16]. It exhaustively evaluates all possible combinations of hyperparameters within a predefined grid of values and selects the combination that produces the best performance according to a specified evaluation metric. Using this technique, the mean cross-validated score for different models can be calculated. In this study, Linear Regression, Lasso Regression, and Decision Tree are considered, and it is observed that Linear Regression has a better mean cross-validated score. Therefore, it is considered for model building, and a Linear Regression model is built and trained with the training data (x_train and y_train). The model is tested to determine its accuracy on predicting prices. It demonstrates an accuracy score of 0.96 over the training data and 0.93 over the testing data[17].

Incorporating a spatial dimension into the analysis, geocoded information is integrated into the dataset, providing a geographic context to each housing data point. By including longitude and latitude coordinates, a more nuanced exploration of how location influences housing prices in Visakhapatnam is enabled. To delve deeper into the spatial dynamics of the housing market, K-means clustering, an unsupervised machine learning technique, is employed. This technique identifies inherent patterns and groupings within the data, offering insights into spatial clusters and trends that characterize the real estate landscape in Visakhapatnam. By combining these analytical approaches, the study aims not only to unravel the determinants of housing prices in Visakhapatnam but also to elucidate the spatial clusters and trends that define the real estate market[18].

V.RESULTS AND DISCUSSIONS

Determination of optimal number of clusters:

The elbow method and inertia method are both techniques used in unsupervised machine learning, particularly for clustering algorithms like k-means, to determine the optimal number of clusters for a given dataset.

Elbow Method:

The elbow method involves plotting the explained variation as a function of the number of clusters, and the "elbow" point represents the optimal number of clusters where adding more clusters does not significantly improve the performance of the model[19].

To apply the elbow method, you typically run the clustering algorithm for a range of cluster numbers and compute a metric like within-cluster sum of squares (WCSS) or sum of squared distances from each point to its assigned cluster centroid.

As you increase the number of clusters, the WCSS tends to decrease because the data points are closer to their cluster centroids. However, after a certain point, adding more clusters will not result in a significant decrease in WCSS. The point where the rate of decrease sharply changes is the elbow point.

Inertia Method:

The inertia method is similar to the elbow method but is more specifically associated with the k-means clustering algorithm.

Inertia, also known as the within-cluster sum of squares, measures the compactness of the clusters. It is the sum of squared distances of samples to their closest cluster centre.

Like the elbow method, you plot the inertia as a function of the number of clusters and look for the point where the rate of decrease in inertia slows down, indicating diminishing returns by adding more clusters.

The optimal number of clusters is often chosen at the point where the inertia starts to decrease more slowly, akin to finding the "elbow" in the plot.

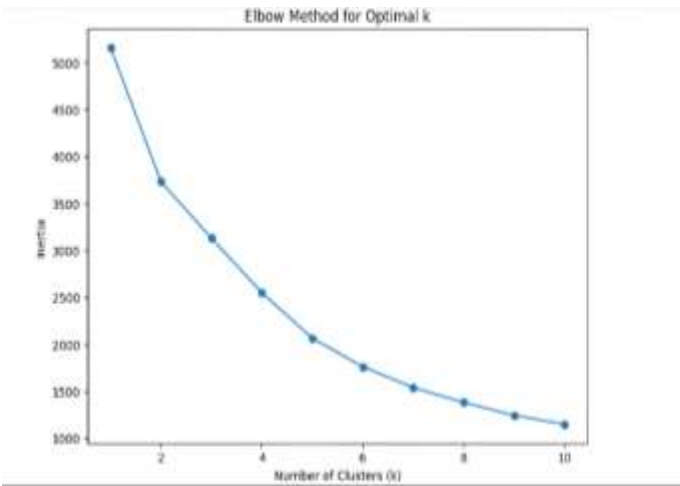

Figure 1:Determination of optimal number of clusters

In the graph above, the number of clusters is plotted against their corresponding inertia values. The idea is to identify the point on the graph where the inertia starts decreasing at a slower rate, resembling an 'elbow' shape. The number of clusters at this elbow point is typically chosen as the optimal number of clusters[20]. This is because adding more clusters beyond this point does not significantly reduce inertia. In the case of the above graph, the number of clusters is determined to be 5.

The Seaborn library is used to create a box plot showing the distribution of prices within different clusters. The x-axis represents the clusters, while the y-axis represents the prices. Each box in the plot represents the interquartile range (IQR) of prices within a cluster, with the median indicated by a horizontal line inside the box. This visualization helps to compare the spread and central tendency of prices across different clusters, providing insights into the variability and distribution of prices within each cluster. So, box plot is used here to visualize the price distribution among clusters and price ranges of each cluster is listed in the table2.

For the specified input parameters of 1000 square feet area, 2 bedrooms, and 2 bathrooms, the calculated prices for every location in the Visakhapatnam dataset serve as pivotal insights into the housing market dynamics of the region. These calculated prices represent not just numerical values, but rather reflections of various underlying factors such as location, amenities, neighbourhood characteristics, and market trends.

Figure3 : User Interface(The user interface takes area , number of bedrooms, number of bathrooms, and location as input, and predicts the price of the house based on these input configurations)

By grouping these calculated prices into five clusters, we gain a nuanced understanding of the diversity within the housing market. Each cluster encapsulates a distinct subset of properties with similar price ranges, thereby facilitating comparative analysis and decision-making for prospective homebuyers, sellers, and real estate professionals alike.

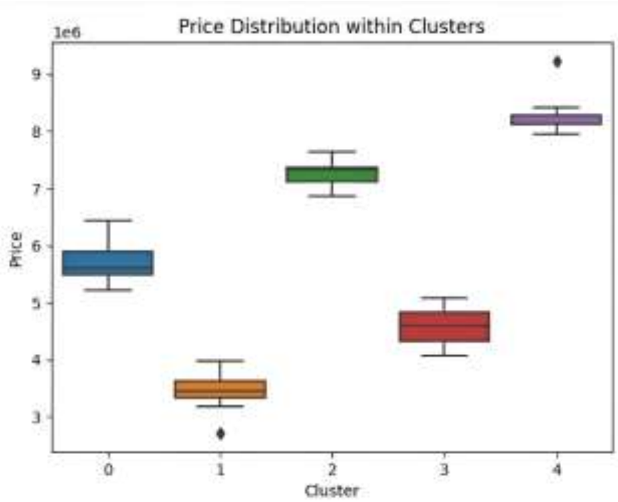

Figure 4:Price Distribution among clusters

"le6" typically represents the number 1,000,000. The "le" stands for "times ten to the power of," and the "6" indicates that the number is multiplied by 10^6, which is 1,000,000. So, "le6" is equivalent to "1,000,000."

These clusters and their price ranges for the input of (1000,2,2) are listed below. Here range of cluster consists of minimum price location and maximum price location within that cluster. The input(1000,2,2) represents 1000 square feet area, 2 bedrooms, and 2 bathrooms

Table 2:Price range of clusters

Cluster	Range
Cluster 0	4858417 to 6530439
Cluster 1	2918144 to 4042604
Cluster 2	7881712 to 8508376
Cluster 3	6721234 to 7763687
Cluster 4	3565651 to 5598097

The average house price plotted on y-axis for each location on x-axis can be shown as:

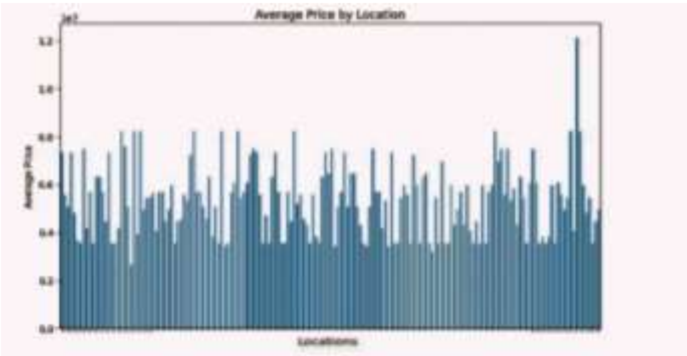

Figure 2:Average price by locations

"le7" typically represents the number 10,000,000. The "le" stands for "times ten to the power of," and the "7" indicates that the number is multiplied by 10^7, which is 10,000,000. So, "le7" is equivalent to "10,000,000."

For all the locations:

The user interface takes the area (in sq.ft), number of bedrooms, and number of bathrooms as input, predicting the house price for each location in the Vishakapatnam dataset.

Figure 3:User Interface for all locations

Heat Map generation:

In a heatmap, colors typically represent the intensity or magnitude of a particular variable or measurement at different locations on a map

Lighter colors, such as shades of blue or green, are used to represent low values or low levels of the predicted house price values. These colors indicate areas with lower intensity or smaller values.

Darker colors, such as shades of red or orange, are used to represent high values or high levels of the predicted house price values. These colors indicate areas with higher intensity or larger values.

Gradient such as colors between the extremes (light and dark) form a gradient. This gradient helps to visually represent the range of values between the lowest and highest, providing a smooth transition from one colour to another.

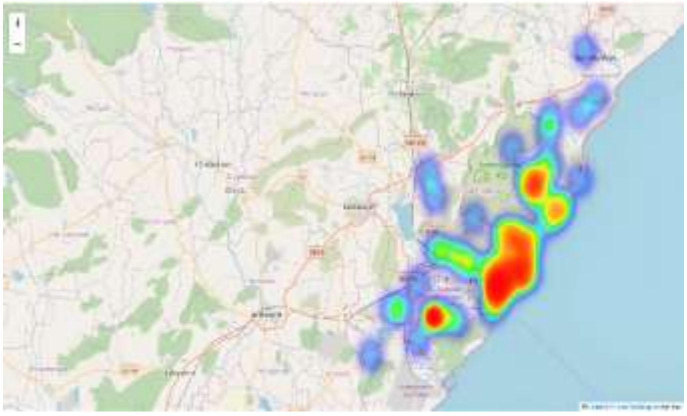

Figure 4:Heat map

Once the locations have been divided into five clusters, the next step is to determine the centroids of each cluster based on the mean latitude and longitude of the locations within that cluster. These centroids serve as representative points that encapsulate the spatial distribution of each cluster, providing valuable insights into the geographic characteristics of the housing market in Visakhapatnam.

Plotting these centroids on the map offers a visual representation of the spatial clustering of housing prices across the region. Each centroid serves as a focal point, indicating the approximate center of gravity for the corresponding cluster of locations. By examining the distribution of centroids on the map, we can discern spatial patterns and trends in housing prices, as well as identify areas of concentration or dispersion.

Figure 5: Plotting of clusters on heat map

Plotting all the locations on the map

After plotting all locations on the map based on their latitude and longitude coordinates, the next step is to group these locations into clusters according to their predicted prices for the input values of (1000, 2, 2). Utilizing clustering algorithms such as K-means or hierarchical clustering, we can effectively segment the data into distinct groups based on their similarities in predicted prices. Once the clustering process is complete, each cluster is assigned a unique colour for visual representation on the map. This colour coding allows for easy identification and differentiation of the clusters, enabling us to discern spatial patterns and variations in predicted house prices across the area of interest. So each cluster is marked with different colours for unique identification.

By visually inspecting the clustered locations on the map, we can gain valuable insights into the spatial distribution of house prices and identify areas with similar pricing characteristics. This information can be instrumental for various stakeholders, including homebuyers seeking affordable neighbourhoods or real estate agents targeting specific market segments. Moreover, analysing the spatial arrangement of the clusters can reveal underlying trends and patterns in the housing market, such as the presence of high-value neighbourhoods or areas experiencing rapid gentrification. Such insights can inform strategic decision-making processes for both individuals and organizations operating within the real estate sector.In addition to aiding in decision-making, visualizing clustered locations on the map fosters a deeper understanding of the complex interplay between geographic location and housing prices.

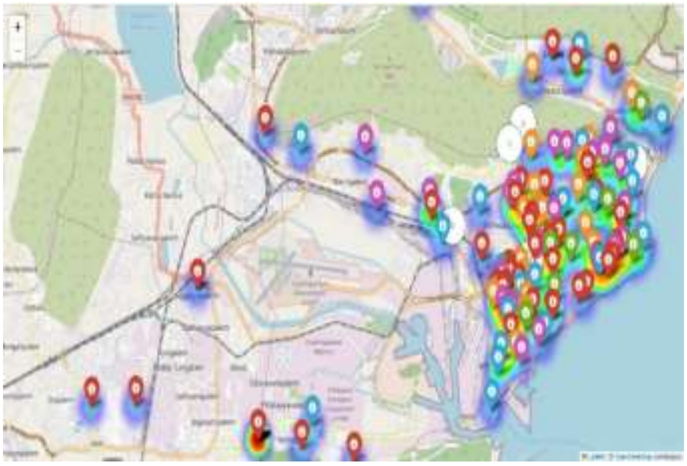

Figure 6: Mapping all locations

By exploring how different factors converge to influence house prices within each cluster, we can unravel the underlying drivers of spatial variability in the housing market. Overall, plotting locations on the map and clustering them based on predicted prices provides a powerful tool for exploring spatial patterns in the housing market. By leveraging advanced analytical techniques and visualization tools, we can unlock valuable insights that facilitate more informed decision-making and strategic planning within the realm of real estate.

VI. CONCLUSION

This study represents a comprehensive effort to harness the power of machine learning techniques and geospatial analysis to predict house prices in a specific area and unveil spatial patterns within the housing market. Through rigorous methodology and analysis, we have successfully developed predictive models with remarkable performance metrics, achieving an accuracy score of 0.967 on training data and 0.928 on unseen data. This high level of accuracy underscores the reliability and effectiveness of our models in forecasting property values. Moreover, our geospatial analysis has uncovered a wealth of insights into the spatial dynamics of house prices, revealing intricate patterns and trends across the region of interest. From identifying clusters of similar properties to delineating spatial variations in house prices, our analysis has provided valuable information for understanding the underlying mechanisms driving the housing market. One of the key findings of our study is the identification of neighbourhoods with the highest and lowest average prices, shedding light on areas of high demand and value within the housing market landscape. By pinpointing these areas, our analysis can assist homebuyers in making informed decisions about where to invest or purchase property, as well as aiding real estate agents in tailoring their strategies to specific market segments. Furthermore, our analysis has uncovered the factors that influence house prices the most, offering valuable insights for both buyers and sellers. By understanding these factors, stakeholders can better anticipate fluctuations in the housing market and adjust their strategies accordingly, whether it be pricing their properties competitively or negotiating favourable deals. In addition, our geospatial analysis has revealed spatial trends in house prices, highlighting areas of growth and development as well as potential areas for investment. This information is invaluable for policymakers tasked with shaping urban development policies and strategies, as it provides a nuanced understanding of how different factors contribute to the spatial distribution of house prices. Based on these findings, we provide recommendations for various stakeholders, including homebuyers, real estate agents, and policymakers. For homebuyers, our analysis can inform decisions about where to purchase property based on factors such as affordability and potential for appreciation. Real estate agents can use our insights to tailor their marketing strategies and target specific neighbourhoods or market segments more effectively. Policymakers can leverage our findings to inform urban development policies and initiatives aimed at promoting equitable access to housing and fostering sustainable growth. Looking ahead, future research may involve collecting additional data to further enhance the accuracy and robustness of our predictive models. Exploring alternative modelling techniques and incorporating additional variables could also provide new insights into the complex dynamics of the housing market. Moreover, investigating specific aspects of the housing market in more detail, such as the impact of demographic changes or economic indicators on house prices, could yield valuable insights for stakeholders.

In conclusion, our study represents a significant contribution to the field of housing market analysis, offering valuable insights into the spatial dynamics of house prices and providing actionable recommendations for stakeholders. By leveraging the power of machine learning and geospatial analysis, we can better understand the complexities of the housing market and make more informed decisions to promote sustainable growth and equitable access to housing.

VII.REFERENCES

[1] R. Gupta, A. Kabundi, and S. M. Miller, ‘‘Forecasting the USreal house price index: Structural and non-structural models with and without fundamentals,’’ Econ. Model., vol. 28, no. 4, pp. 2013–2021, Jul. 2011.

[2] J. Mu, F. Wu, and A. Zhang, ‘‘Housing value forecasting based on machine learning methods,’’ Abstract Appl. Anal., vol. 2014, pp. 1–7, Aug. 2014.

[3] L. Bork and S. V. Møller, ''Forecasting house prices in the 50 states using dynamic model averaging and dynamic model selection,'' Int. J. Forecasting, vol. 31, no. 1, pp. 63–78, Jan. 2015.

[4] A. Ng and M. Deisenroth, ''Machine learning for a London housing price prediction mobile application,'' Imperial College London, London, U.K., 2015.

[5] M. Risse and M. Kern, ''Forecasting house-price growth in the euro area with dynamic model averaging,'' North Amer. J. Econ. Finance, vol. 38, pp. 70–85, Nov. 2016.

[6] B. Afonso, L. Melo, W. Oliveira, S. Sousa, and L. Berton, ''Housing prices prediction with a deep learning and random forest ensemble,'' in Proc. Anais do 16th Encontro Nacional de Inteligência Artif. e Computacional, 2019, pp. 389–400.

[7] M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu, ''Spatial transformer networks,'' in Proc. Adv. Neural Inf. Process. Syst., 2015, pp. 2017–2025.

[8] Z. Peng, Q. Huang, and Y. Han, ''Model research on forecast of secondhand house price in Chengdu based on XGboost algorithm,'' in Proc. IEEE 11th Int. Conf. Adv. Infocomm Technol. (ICAIT), Oct. 2019, pp. 168–172.

[9] C. R. Madhuri, G. Anuradha, and M. V. Pujitha, ''House price prediction using regression techniques: A comparative study,'' in Proc. Int. Conf. Smart Struct. Syst. (ICSSS), Mar. 2019, pp. 1–5.

[10] T. D. Phan, ''Housing price prediction using machine learning algorithms: The case of Melbourne city, Australia,'' in Proc. Int. Conf. Mach. Learn. Data Eng. (iCMLDE), Dec. 2018, pp. 35–42.

[11] A. S. Temür, M. Akgün, and G. Temür, ''Predicting housing sales in Turkey using ARIMA, LSTM and hybrid models,'' J. Bus. Econ. Manage., vol. 20, no. 5, pp. 920–938, Jul. 2019.

[12] L. Yu, C. Jiao, H. Xin, Y. Wang, and K. Wang, ''Prediction on housing price based on deep learning,'' Int. J. Comput. Inf. Eng., vol. 12, no. 2, pp. 90–99, 2018.

[13] C. Ge, ''A LSTM and graph CNN combined network for community house price forecasting,'' in Proc. 20th IEEE Int. Conf. Mobile Data Manage. (MDM), Jun. 2019, pp. 393–394.

[14] S. Law, B. Paige, and C. Russell, ''Take a look around: Using street view and satellite images to estimate house prices,'' ACM Trans. Intell. Syst. Technol., vol. 10, no. 5, pp. 1–19, Nov. 2019.

[15] X. Fu, T. Jia, X. Zhang, S. Li, and Y. Zhang, ''Do street-level scene perceptions affect housing prices in Chinese megacities? An analysis using open access datasets and deep learning,'' PLoS ONE, vol. 14, no. 5, May 2019, Art. no. e0217505.

[16] Q. You, R. Pang, L. Cao, and J. Luo, ''Image-based appraisal of real estate properties,'' IEEE Trans. Multimedia, vol. 19, no. 12, pp. 2751–2759, Dec. 2017.

[17] O. Poursaeed, T. Matera, and S. Belongie, ''Vision-based real estate price estimation,'' Mach. Vis. Appl., vol. 29, no. 4, pp. 667–676, May 2018.

[18] T. Luong, H. Pham, and C. D. Manning, ''Effective approaches to attentionbased neural machine translation,'' in Proc. Conf. Empirical Methods Natural Lang. Process., 2015, pp. 1412–1421.

[19] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, ''Attention is all you need,'' in Proc. Adv. Neural Inf. Process. Syst., 2017, pp. 5998–6008.

[20] O. Firat, K. Cho, and Y. Bengio, ''Multi-way, multilingual neural machine translation with a shared attention mechanism,'' in Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol., 2016, pp. 866–875.