House Price Prediction using Deep Learning

Dr.S.Russia[1], Varshini V S[2], Swathi K R[3], Sudhan P[4]
[1] Professor, Department of Computer Science and Engineering,
Velalar College of Engineering and Technology,
Erode, Tamil Nadu,
India[2,3,4]
Student, Department of Computer Science and Engineering,
Velalar College of Engineering and Technology,
Erode,
Tamil Nadu, India

ABSTRACT--Accurate prediction of future housing prices is essential for sellers and real estate agents to make informed decisions. By developing a housing price prediction model, it becomes possible to determine optimal selling prices and assist customers in timing their home purchases. The prediction of house prices is crucial for buyers, helping them make well- prepared financial decisions. The three key factors influencing house prices are physical conditions, concepts, and location. In this study, a deep learning approach utilizing LSTM and sentence embedding techniques is proposed for accurate house price prediction. The dataset includes features such as bedrooms, bathrooms, square footage, and location, with house price as the target variable. The findings highlight the effectiveness of LSTM and sentence embedding in providing valuable insights for real estate stakeholders and facilitating informed decision-making in the housing market. The sale of these houses, for those who bought, the prediction of the excellent house prices, better before they make one of the most important financial decisions in their lives, to expect what they prepare them for. Physical conditions, concepts, and where there are three factors that affect the price of the house to include. Housing prices is an important reflection of the economy, the scope of housing prices is a major concern for both the buyer and the seller. In this study, we propose a Deep learning approach for house price prediction using the LSTM and Sentence embedding. The dataset contains various features related to houses, such as the number of bedrooms, bathrooms, square footage, and location, with the target variable being the house price. The findings demonstrate the effectiveness of LSTM. Sentence embedding in accurately predicting house prices, providing valuable insights for real estate stakeholders and facilitating informed decision-making in the housing market.

Keywords--Housing prices prediction, Deep Learning ,LSTM, Sentence Embedding

I.INTRODUCTION

The primary focus of this article is the development of a robust real estate price prediction system utilizing machine learning algorithms for optimal accuracy. Situated within the realms of Machine Learning (ML) and Data Science, the project encompasses the creation of a comprehensive real estate price prediction model coupled with the construction of an intuitive website.The overarching objective is to deliver a sophisticated price prediction system complemented by a user-friendly front end. This interfaceaims to empower users, enabling them to select their desired destination and obtain accurate insights into prevailing price rates. The motivation behind this initiative lies in the critical importance of having precise information about property values before finalizing any real estate transactions.The analysis conducted in this paper relies on a dataset obtained from a reputable and trusted website. This dataset comprises an abundance of sample points, facilitating a more thorough and insightful analysis. Recognizing that an informed decision about real estate transactions hinges on the exact pricing of properties, the model takes into account a multitude of influential factors.Factors influencing the model include, but are not limited to, the area of the property, its location, local population dynamics, size, and the number of bedrooms and bathrooms. Additionally, considerations extend to amenities such as parking space, the presence of an elevator, the construction style, balcony space, the overall condition of the building, and the price per square foot. Each of these aspects contributes significantly to the holistic understanding of property values.In conclusion, this article endeavors to present a thorough exploration of real estate price prediction through machine learning. The fusion of accurate data analysis and the development of an accessible interface aims to empower users in making well-informed decisions regarding property transactions. The meticulous consideration of various factors in the model ensures a comprehensive understanding of the dynamic real estate market.

II. RELATED WORK

The effectiveness of machine learning techniques, such as decision trees, random forests, and support vector machines, in the field of housing price prediction has been amply proven by historical study. These algorithms have become extremely effective in forecasting and comprehending the intricate dynamics of real estate markets because of their robust performance and versatility across a wide range of datasets. By using decision trees, analysts can better understand the hierarchical structure of factors impacting house prices by having a clear and comprehensible framework for outlining the decision-making process.

1. HOUSING PRICE PREDICTION USING MACHINE LEARNING ALGORITHMS: THE CASE OF MELBOURNE CITY, AUSTRALIA

This literature endeavors to delve into the wealth of insights concealed within historical property market data, particularly focusing on the context of Australia. The study engages advanced machine learning techniques to meticulously analyze past real estate transactions, aiming to extract practical and valuable models that can serve as valuable tools for both buyers and sellers in the real estate market.A notable revelation arising from this analysis pertains to the stark contrast in home prices observed across different districts in Melbourne. The study sheds light on the significant disparity between the most expensive and least expensive areas, offering a comprehensive understanding of the diverse pricing dynamics within the real estate landscape.Moreover, the research emphasizes the utilization of the mean squared error measurement-based Stepwise and Support Vector Machine combination as a competitive and effective method. By employing these sophisticated techniques, the study aims to contribute not only to the theoretical understanding of the real estate market but also to the practical development of models that can enhance decision-making processes for those involved in real estate transactions. The incorporation of such advanced methodologies underscores the commitment to extracting nuanced and accurate insights from the intricate tapestry of historical property market data in the Australian context.

2. PREDICTION OF HOUSE PRICING USING MACHINE LEARNING WITH PYTHON

In this scholarly work by Jadhav Vaibhav et al., a comprehensive exploration is presented, outlining a novel methodology for predicting house prices through the application of diverse regression techniques facilitated by Python libraries. The research not only introduces an overarching perspective on the predictive modeling of house charges but also endeavors to enhance accuracy by delving into more nuanced aspects crucial for the calculation of residential property values.The proposed approach seeks to encapsulate the intricacies involved in determining house prices, aiming to provide more precise predictions than conventional methods. By incorporating various regression techniques, the research not only adds depth to the understanding of housing market dynamics but also offers practical insights into refining the prediction process.Moreover, the paper offers a succinct overview of graphical and numerical methods essential for the estimation of home prices. This multifaceted approach underscores the commitment to comprehensively address the various dimensions of house pricing, incorporating both visual and quantitative analyses to ensure a robust and accurate predictive model.A noteworthy aspect of this paper is itsintegration of machine learning techniques, elucidating how these methodologies contribute to the efficacy of the residence pricing model. The utilization of Python libraries further enhances the accessibility and replicability of the proposed approach.

3. HOUSE PRICE FORECASTING USING DATA MINING

In the research conducted by Nihar Bhagat et al., a novel approach is proposed that acknowledges the distinct spending patterns and investment strategies adopted by buyers of new homes. The study highlights a gap in existing methodologies, noting that current approaches often fail to account for the necessity of forecasting market changes and potential price fluctuations, which are crucial considerations in determining optimal house prices.The primary objective of this research is to develop a methodology that enables the prediction of the optimal house price tailored to the goals and financial circumstances of real estate clients. Unlike conventional methods, which often overlook the need for forecasting future market trends and price fluctuations, this approach seeks to leverage historical market data and upcoming changes to forecast future prices more accurately.Central to the implementation of this methodology is the development of a website interface that allows clients to input their requirements and preferences. Through the integration of data mining techniques, particularly the multiple linear regression method, the system generates forecasts of future house prices based on historical trends and forthcoming changes in the market.

4. REAL ESTATE VALUE PREDICTION USING LINEAR REGRESSION

The real estate market, deemed as one of the most price-centric and dynamic sectors within the economic system, emerges as an ideal domain for the application of machine learning principles to enhance and precisely predict property costs. This sector's inherent complexity, influenced by multifaceted factors, positions it as a fertile ground for the integration of advanced computational methodologies.The three pivotal elements identified to significantly impact housing prices are the property's location, its intrinsic characteristics or features, and its current physical state. The amalgamation of these factors creates a dynamic interplay that contributes to the fluidity and volatility observed in real estate valuations.In the current structure and prevalent methodologies, housing prices are estimated without incorporating crucial considerations such as market pricing trends and potential cost increases. This limitation underscores the necessity for innovative approaches that go beyond static assessments and embrace the dynamic nature of the real estate market.By recognizing the intricate relationship between location, property attributes, and physical condition,
machine learning principles can be harnessed to develop models that offer more nuanced and accurate predictions. This utilization of advanced computational techniques allows for a comprehensive understanding of the evolving real estate landscape, contributing to more informed decision-making processes for both buyers and sellers.

5. HOUSE PRICE PREDICTION USING OPTIMAL REGRESSION TECHNIQUES

The proposal by Vivek Singh Rana et al. centers around the profound consideration that individuals undertake when making investments, particularly in the context of housing. The pivotal aspect of this decision-making process, especially for middle-class and lower-class individuals, is the budget. Recognizing the significance of this financial constraint, the primary goal of their research is to develop a predictive model for housing costs tailored to the financial circumstances of individuals in these economic segments.This research contends that when individuals embark on the journey of acquiring a house, they take into account a myriad of factors, with budgetary considerations standing out as the most crucial. Understanding the unique challenges and financial constraints faced by middle-class and lower-class individuals, the research aims to provide a forecasting system that aligns with their specific circumstances.To achieve this goal, the study systematically evaluates and compares the outcomes of various regression techniques. The regression techniques under scrutiny include Support Vector Regression, XGBoost, Decision Tree Regression, and Random Forest Regression. This rigorous analysis aims to determine which regression technique proves to be the most effective in predicting house values accurately for individuals with diverse financial backgrounds.

III. PROPOSED SYSTEM

In the ambit of this comprehensive study, we present an innovative and resilient house price prediction system that harnesses the capabilities of the Long Short-Term Memory (LSTM) algorithm. This advanced system is meticulously crafted to serve as a valuable tool for real estate professionals and potential homebuyers, empowering them to make well- informed decisions through the provision of accurate and reliable house price estimates.The foundational step involves the meticulous collection of a comprehensive dataset enriched with pertinent features crucial for understanding the dynamics of house prices. Attributes such as the number of bedrooms, bathrooms, square footage, location, and other influential factors are meticulously curated to ensure the model captures the intricate nuances that contribute to house price variations.Subsequent to data acquisition, a rigorous preprocessing phase is initiated. This involves addressing missing values, converting categorical variables, and implementing standardized scaling techniques to ensure theuniformity and quality of the dataset. The careful execution of these preprocessing steps is pivotal in laying the groundwork for an accurate and robust prediction model.To optimize the model's performance and mitigate the risk of overfitting, feature selection techniques are employed. This strategic approach involves identifying the most significant predictors, thereby enhancing the model's efficiency and ensuring that it is streamlined to focus on the most impactful variables.

A.DATA PREPARATION

Data collection is the process of gathering and measuring information from countless different sources.Collecting data allows you to capture a record of past events so that we can use data analysis to find recurring patterns. The dataset has been collected from Kaggle and UCI repositories. Therefore the dataset includes LA Listings and London Listings. Fig 1. shows the overview of the LA dataset, and Fig 2. shows the overview of the London dataset.The data is raw and includes some missing and incorrect numbers. Preprocessing data is therefore an essential duty.In order to properly and cleanly present the data. Preprocessing and data purification are two aspects of preprocessing data. The dataset is categorized into several groups according to specific attributes of the data object.

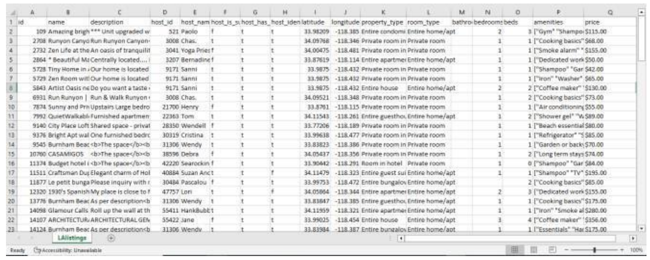

Fig 1. Overview of LS listings Dataset

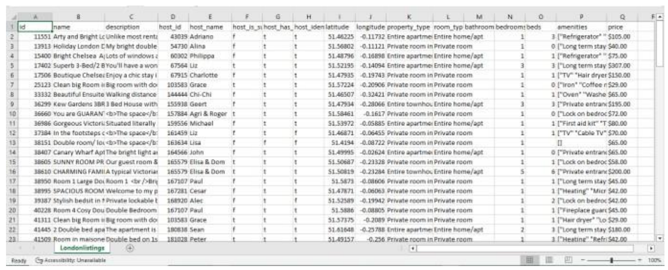

Fig 2. Overview of London Listings Dataset From those patterns, predictive models are built using machine learning algorithms that look for trends and predict future changes.

B.PRETRAINED SENTENCE EMBEDDING

In the innovative realm of house price prediction, our methodology incorporates a tandem of advanced techniques—pre-trained sentence embedding and Long Short-Term Memory (LSTM) networks. This dynamic synergy aims to elevate the accuracy and depth of understanding in predicting real estate prices. The utilization of pre-trained sentence embedding is a strategic choice, introducing a transformative approach to convert textual information related to real estate properties into numerical vectors. This embedding model excels in capturing the semantic nuances of the textual data, providing a compact yet information-rich representation that serves as a foundation for our predictive model.The pre-trained sentence embedding, acting as the linguistic backbone, plays a pivotal role in bridging the gap between raw textual information and numerical vectors. By distilling the intricate semantics embedded in real estate property descriptions, this embedding model ensures that the textual data's inherent meaning is preserved and efficiently communicated to subsequent stages of analysis. Its pre-trained nature signifies exposure to a diverse linguistic landscape, endowing it with the capability to comprehend a broad spectrum of meanings and contexts. This linguistic richness proves invaluable in enhancing the interpretability of textual data, a crucial aspect in the complex and nuanced realm of real estate descriptions. Complementing the pre-trained sentence embedding, our model incorporates LSTM networks, renowned for their proficiency in capturing temporal dependencies and intricate patterns within sequential data(Fig 3). The LSTM's role in our predictive system is to leverage its memory cell architecture to discern and retain essential information over extended sequences, thereby addressing the temporal dynamics inherent in real estate market trends. The synergy between pre-trained sentence embedding and LSTM networks creates a harmonious framework that not only preserves the semantic integrity of textual data but also harnesses the temporal context essential for making accurate predictions in the ever- evolving landscape of house prices. This integrated approach, marrying linguistic comprehension with temporal sensitivity, positions our house price prediction system as a cutting-edge solution in the realm of real estate analytics.

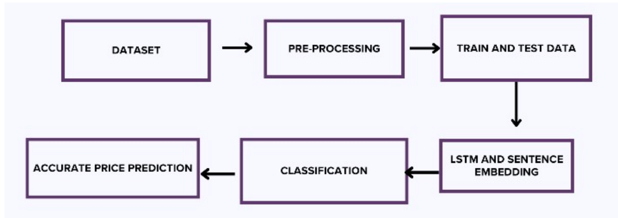

Fig 3.Training the modelC.PREDICTING

The prediction of house prices using LSTM (Long Short-Term Memory) and sentence embedding techniques represents a cutting-edge approach in real estate analytics. LSTM, renowned for its proficiency in handling sequential data, and sentence embedding, designed to capture semantic meaning in textual data, synergize to create a powerful predictive model.In this innovative methodology, historical data comprising various features such as property characteristics, location, and market trends serve as inputs to the LSTM model. By leveraging the sequential nature of time- series data, LSTM effectively captures temporal dependencies and complex patterns inherent in house price fluctuations over time.Fig 4 shows the prices of houses available. The model learns from past trends to make accurate predictions about future house prices.Complementing the LSTM model is the integration of sentence embedding techniques. Textual information associated with real estate properties, including descriptions, features, and amenities, are transformed into numerical vectors using pre-trained sentence embedding models. These vectors encapsulate the semantic essence of the textual data, enriching the feature set and enhancing the model's predictive capabilities.The combined LSTM and sentence embedding approach enables the model to analyze both numerical and textual data comprehensively. This holistic understanding of the underlying factors influencing house prices facilitates more accurate predictions and insights into market trends. Moreover, the model's ability to capture semantic context ensures that subtle nuances in property descriptions are not overlooked, further refining the predictive accuracy.

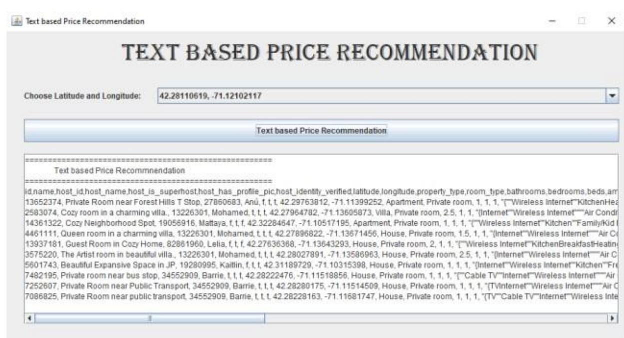

Fig 4.Predicting and listing the prices of houses in the given latitude and longitude

IV. CONCLUSION AND FUTURE WORK

In conclusion, this article has presented a comprehensive exploration into the prediction of real estate prices by leveraging machine learning algorithms, specifically utilizing Long Short-Term Memory (LSTM) and Sentence Embedding techniques. The overarching objective was to design an advanced house price prediction system within the realms of Machine Learning and Data Science, complemented by the development of a user-friendly interface.The primary aim of this model was to offer an accurate and efficient price prediction system, empowering users to make informed decisions when selecting a property. The user-friendly front end facilitates a seamless experience, allowing users to explore pricing information based on their preferences and desired destinations.The analysis conducted in this paper extensively utilized a dataset sourced from a trusted website, providing a robust foundation for the machine learning algorithms. The inclusion of various factors such as area, location, population, property size, bedrooms, bathrooms, parking space, elevator availability, construction style, balcony space, and building condition in the analysis contributes to a more nuanced understanding of housing prices.The incorporation of LSTM and Sentence Embedding techniques adds a layer of sophistication to the predictive model, enabling it to capture temporal dependencies and semantic information within the dataset. This enhances the accuracy of the predictions, making the model more adept at handling the dynamic nature of real estate markets.In the future, we plan to complete our ongoing platform for Dynamic Model Updating i.e. Implement a mechanism for real-time updates and continuous learning to ensure that the model adapts to changing market trends and dynamics.We also plan to develop a feedback loop where users can provide input on the accuracy of predictions, enabling continuous refinement of the model based on real-world user experiences.Morover we aim to strengthen the Sentence Embedding technique by incorporating advanced sentiment analysis to better capture qualitative aspects that may impact property prices.

REFERENCES

[1] Nihar Bhagat, Ankit Mohokar and Shreyash Mane. House Price Forecasting using Data Mining. International Journal of Computer Applications 152(2):23-26, October 2023.

[2] N. N. Ghosalkar and S. N. Dhage, "Real Estate Value Prediction Using Linear Regression," 2022 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, 2021, pp. 1-5, doi: 10.1109/ICCUBEA.2018.8697639.

[3] M. Jain, H. Rajput, N. Garg and P. Chawla, "Prediction of House Pricing using Machine Learning with Python," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2020, pp. 570-574, doi: 10.1109/ICESC48915.2020.9155839

[4] V. S. Rana, J. Mondal, A. Sharma and I. Kashyap, "House Price Prediction Using Optimal Regression Techniques,"2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), Greater Noida, India, 2020, pp. 203-208, doi: 10.1109/ICACCCN51052.2020.9362864

[5] Mu J, Wu F, Zhang A,Housing Value Forecasting Based on Machine Learning Methods,Abstract and Applied Analytics,2019,pp. 1-7 doi:10.1155/2020/648047

[6] C.Nagarajan and M.Madheswaran - 'Experimental verification and stability state space analysis of CLL-T Series Parallel Resonant Converter' - Journal of ELECTRICAL ENGINEERING, Vol.63 (6), pp.365-372, Dec.2012.

[7] C.Nagarajan and M.Madheswaran - 'Performance Analysis of LCL-T Resonant Converter with Fuzzy/PID Using State Space Analysis'- Springer, Electrical Engineering, Vol.93 (3), pp.167-178, September 2011.

[8] C.Nagarajan and M.Madheswaran - 'Stability Analysis of Series Parallel Resonant Converter with Fuzzy Logic Controller Using State Space Techniques'- Taylor & Francis, Electric Power Components and Systems, Vol.39 (8), pp.780-793, May 2011.

[9] C.Nagarajan and M.Madheswaran - 'Experimental Study and steady state stability analysis of CLL-T Series Parallel Resonant Converter with Fuzzy controller using State Space Analysis'- Iranian Journal of Electrical & Electronic Engineering, Vol.8 (3), pp.259-267, September 2012.

[10]Nagarajan C., Neelakrishnan G., Akila P., Fathima U., Sneha S. "Performance Analysis and Implementation of 89C51 Controller Based Solar Tracking System with Boost Converter" Journal of VLSI Design Tools & Technology. 2022; 12(2): 34–41p.

[11]C. Nagarajan, G.Neelakrishnan, R. Janani, S.Maithili, G. Ramya "Investigation on Fault Analysis for Power Transformers Using Adaptive Differential Relay" Asian Journal of Electrical Science, Vol.11 No.1, pp: 1-8, 2022.

[12]G.Neelakrishnan, K.Anandhakumar, A.Prathap, S.Prakash "Performance Estimation of cascaded h-bridge MLI for HEV using SVPWM" Suraj Punj Journal for Multidisciplinary Research, 2021, Volume 11, Issue 4, pp:750-756

[13]G.Neelakrishnan, S.N.Pruthika, P.T.Shalini, S.Soniya, "Perfromance Investigation of T-Source Inverter fed with Solar Cell" Suraj Punj Journal for Multidisciplinary Research, 2021, Volume 11, Issue 4, pp:744-749

[14]C.Nagarajan and M.Madheswaran, "Analysis and Simulation of LCL Series Resonant Full Bridge Converter Using PWM Technique with Load Independent Operation" has been presented in ICTES'08, a IEEE / IET International Conference organized by M.G.R.University, Chennai.Vol.no.1, pp.190-195, Dec.2007

[15]M Suganthi, N Ramesh, "Treatment of water using natural zeolite as membrane filter", Journal of Environmental Protection and Ecology, Volume 23, Issue 2, pp: 520-530,2022

[16]M Suganthi, N Ramesh, CT Sivakumar, K Vidhya, "Physiochemical Analysis of Ground Water used for Domestic needs in the Area of Perundurai in Erode District", International Research Journal of Multidisciplinary Technovation, pp: 630-635, 2019

[17]T. D. Phan, "Housing Price Prediction Using Machine Learning Algorithms: The Case of Melbourne City, Australia," 2019 International Conference on Machine Learning and Data Engineering (iCMLDE), Sydney, NSW, Australia, 2018, pp. 35-42, doi: 17.1109/iCMLDE.2018.00017.

[18]Li Li and Kai-Hsuan Chu, "Prediction of real estate price variation based on economic parameters", 2019 International Conference on Applied System Innovation (ICASI), 2019.

[19]Debanjan Banerjee and Suchibrota Dutta, "Predicting the housing price direction using machine learning techniques", 2019 IEEE International Conference on Power Control Signals and Instrumentation Engineering (ICPCSI)., 2019.

[20]Jang, H., Ahni, K., Kim, D., & Songs, Y. (2019, May). Detection and Prediction of House Price Bubbls: Evidencse from a New Cityes. In International Conferences on Computation Sciences (pp. 777-811). Springer, Cham.

[21]R. Manjula et al., "Real estate value prediction using multivariate regression models", Materials Science and Engineering Conference Series, vol. 263, no. 4, 2019.

[22]Fan C, Cui Z, Zhong X. House Prices Prediction with Machine Learning Algorithms. Proceedings of the 2019 10th International Conference on Machine Learning and Computing ICMLC 2018. doi:10.1145/3195106.3195133.

[23]N. N. Ghosalkar and S. N. Dhage, "Real Estate Value Prediction Using Linear Regression", 2019 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), pp. 1-5, 2019.

[24] R. A. Rahadii, S. K. Wirono, D. P. Koesrindartotor, and
[25]I.B. Syamwiil, ―Factors influencing the price for housing in Indonesia,‖ Int. J. Hous. Mark. Anal., vol. 10, no. 1, pp. 169– 188, 2018
[26]Lu S, Li Z, Qin Z, Yang X, Goh RSM. A hybrid regression technique for house prices prediction. 2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM) 2017. doi:10.1109/ieem.2017.8289904.
[27] Zhoui, J.., Zhang, H., Gud, Y., &Pantelouse, A. A. (2018). Affordable level of house price using fuzzys linear regression analysis: the case of Shanghasi. Soft Computers, 1-13.