Role of AI and ML in real estate market.

Chaitanya Dilip Shinde[1], Dr. Harshali Patil[2]

[1]Department of Information Technology and Computer Science, 31031424037, SK Somaiya College, Somaiya Vidhyavihar University, Mumbai, India. chaitanya.s@somaiya.edu

[2]Department of Information Technology and Computer Science, S K Somaiya College, Somaiya Vidhyavihar University, Mumbai, India harshali.yp@somaiya.edu

ABSTRACT :

The relevance of modern technologies like Artificial Intelligence and Machine Learning is rising so the scholars, industries and researchers are working in these fields. This study aims to explore the potential uses of artificial intelligence (AI) and machine learning (ML) in analyzing real estate markets, predicting trends as well as assess how those analyses would influence decision-making processes within the industry. Through the lens of this example, which is such an essential subject like property valuation, it will be investigated how these technologies can further improve accuracy and efficiency in real estate markets. This study will serve as a systematic review of these technologies, aiming to provide an educated background for the creation of more accurate and efficient instruments regarding market analysis use cases in real estate market.

Keywords: Real Estate, Artificial Intelligence, Machine learning, analysis, prediction, regression, linear regression.

Introduction :

The real estate market in India has undergone significant transformations in recent years, with the emergence of new technologies playing a crucial role in shaping its evolution. One such technology that has had a profound impact is Artificial Intelligence & Machine learning. As the Indian economy continues to grow and urbanization gains momentum, the real estate sector has become increasingly important, driving the need for innovative solutions to address the complexities and challenges faced by this industry. (Llp, n.d.) This paper explores the transformative impact of Artificial Intelligence & Machine learning on the real estate sector, highlighting their crucial role in shaping the future of the industry, and also explores how using AI & ML in the real estate sector can help professionals identify new opportunities and mitigate risks in their market decisions. This study questions and demonstrates how AI & ML (Artificial Intelligence & Machine learning) technologies are enabling more informed, efficient, and strategic decision making actions in the real estate market.

Review of Literature :

Housing is one of basic human needs. Today the real estate market is a huge field. Real estate market analysis traditionally relied on economic indicators, location-based factors, and expert judgment. The real estate market has evolved to incorporate big data and has implemented technologies such as Artificial Intelligence & Machine learning. With the advent of AI & ML, the real estate market has seen a significant shift towards more data-driven practices, accurate and scalable methods of analysis.

Traditionally, value of assets was calculated using manual evaluation methods. (Pagourtzi et al., 2003) Now, many real estate stakeholders such as house owners, buyers, investors and agents make use of AI & ML tool such as House price prediction, Property valuation, Trend prediction, Improving customer experience, and Property management. A similar picture emerges for the application of AI & ML in the real estate industry. By treating a real estate asset as the sum of its individual characteristics, the hedonic price regression has established itself as the main approach for price and rent estimation. ML models have proven helpful in real estate hedonic modelling for predictive purposes. Nevertheless, hedonic models lack transparency and do not reveal the underlying theoretical relationships, says S. Mullainathan and J. Spiess (Mullainathan & Spiess, 2017)

Various machine learning models can be used to predict and analyse market prices, Luca Rampini & Fulvio Re Cecconi (2022) examines the application of machine learning (ML) techniques to predict house prices in Italy. His study specifically focuses on two cities, Brescia and Varese, utilizing an extensive dataset to test three popular machine learning models: ElasticNet, XGBoost, and Artificial Neural Network (ANN) (Rampini & Re Cecconi, 2022). His study suggests that ANN is the most effective model for predicting house prices in the Italian market, especially for medium to low-priced properties. However, more expensive properties may require larger and more diverse datasets for better accuracy.

In a study by (Chen, 2024) synthesizes and builds upon existing research to implement and enhance house price prediction methods, evaluating the performance of various techniques, including Convolutional Neural Networks (CNNs), Decision Trees, and K-Nearest Neighbors (KNN), with a particular emphasis on the Most Correlated Features (KNN-MCF) approach. Despite the advancements and refinements of these techniques over the years, they continue to encounter challenges by integrating available resources, their study aims to provide objective guidance for addressing these common challenges in house price prediction.

In a study title: "An Algorithm to Predict Real Estate Price using Machine Learning" (Vyas & Sharma, 2023) the primary focus of the author is to develop a method for predicting property prices accurately and efficiently, thereby alleviating the challenges faced by buyers in determining the fair market value of properties. Their paper presents several important conclusions regarding the use of machine learning in real estate price prediction.

In another study Winky K.O. Ho, Bo-Sin Tang & Siu Wai Wong (2021) discusses the application of various machine learning algorithms—specifically support vector machine (SVM), random forest (RF), and gradient boosting machine (GBM) in predicting property prices (Ho et al., 2021). They find that RF and GBM generally provide better predictive performance compared to SVM, evidenced by lower error metrics. Nonetheless, SVM is noted for its ability to deliver accurate predictions quickly.

The article (Dhar & P, 2023) "A Literature Review on Using Machine Learning Algorithm to Predict House Prices" by Tanmoy Dhar and Dr. Manikandan P offers an in-depth examination of various machine learning methods utilized for predicting residential property values. The authors highlight the importance of factors such as size, location, building type, age, number of bedrooms, and garages in influencing house prices. They investigate multiple machine learning algorithms, including decision trees, linear regression, support vector regression, and Lasso regression, to create predictive models. The research assesses these models using performance metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and overall accuracy. Their findings reveal that neural networks and gradient boosting algorithms generally achieve the highest accuracy in predicting house prices, with random forests and support vector machines following closely behind. The authors assert that choosing appropriate features is essential for improving the accuracy of the predictive models. They highlight the promise of machine learning as a significant asset in real estate valuation and propose that additional research could enhance these models and their practical application in real-world contexts.

In "Advanced Machine Learning Techniques for Predictive Modeling of Property Prices" (Mathotaarachchi et al., 2024), the author reviews the use of machine learning (ML) in real estate price prediction, emphasizing its growing role in improving accuracy and supporting decision-making. The study highlights gaps in the literature, such as limited exploration of hybrid ML-econometric models and challenges in interpreting ML predictions.

In 2018, Neelam Shinde and Kiran Gawande published a study titled "The valuation of house prices using predictive techniques". (Shinde & Gawande, n.d.) This work employs four algorithms: Logistic Regression, Support133 (Dhar and Manikandan, P2023, Vol. 05, Issue 05). Svector regression, Lasso regression, and Decision Tree.   The Decision tree has the highest accuracy among the four algorithms, at 84.64%.   Others are logistic regression, support vector regression, and lasso regression, with 72.81%, 67.81%, and 60.32%, respectively, and they have some parameters such as accuracy, R-squared value, RMSE (Root Mean Square Error), MAE (Mean Absolute Error), and MSE (Mean Square Error) for comparing those models.

In this paper we will discuss some of these various techniques and methods used in the real estate market, and talk about their advantages, disadvantages and the challenges faced.

Design :

Many research papers have been published that present the usage of the AI & ML applied to real estate market analysis, where they claim accurate result with few steps of analysis. Nevertheless, the real estate industry, much like the overall tech sector, confronts constant new challenges that require innovation and agility.

The real estate market needs to learn to fly alongside these upcoming new technologies in order to offer more accurate and actionable predictions. Without this advancement, not only might these tools find their predictive models and algorithms aged, but their potential for analyzing market trends, consumer behavior and investment opportunities will wane as well. The more advanced AI gets, the more important it becomes to be current in market analysis and making appropriate decisions in order to avoid being left behind.

As stated in the introduction, this paper will focus on how AI & ML algorithms play an important part in real estate market. The goal is to investigate how insights from AI & ML assist investors in making informed choices by forecasting future market trends, evaluating investment risks, and enhancing property portfolios. This study will delve into particular AI tools that are shaping real estate investment strategies. The aim of this study is to examine how AI & ML methods, especially deep learning models, are applied to anticipate future market demand

I will be using the following methods for researching about this study:

   A. Literature Review: To conduct a comprehensive review of existing literature on AI & ML in real estate market analysis to understand the current state of research in this area. Study recent developments and applications of Artificial Intelligence & Machine learning in this field to identify research gaps. I will also utilize tools such as Connectedpapers.com, Scispace.com, Chatpdf.com, Mendeley.
   B. Empirical Research: Conducting experiments and studies to demonstrate the effectiveness of AI & ML in real estate.
   C. Case Studies: Analyse and present case studies of real-world applications of AI & ML in the real estate sector to showcase practical relevance and impact. Highlight successful implementations and their outcomes.

Discussion :

Before diving into why Artificial Intelligence & Machine learning play such an important role in the real estate market, I want to write about Artificial Intelligence & Machine learning in general. What Artificial Intelligence & Machine learning really are and why do these two terms matter so much in today's 2024 world?

Artificial intelligence (AI) and machine learning (ML) are related technologies that use algorithms and models to perform tasks. AI is a broad field that refers to the use of technology to mimic human intelligence, while ML is a subset of AI that focuses on how machines can learn and improve from experience. These systems leverage machine learning, natural language processing, and computer vision to analyse data, make predictions, and automate processes. It is impossible to overestimate the significance of AI as it improves decision-making, increases productivity, and opens up new avenues for innovation ('Artificial Intelligence(AI) in Tech', 2024) .

These days, AI has become a key component in many different fields of work, providing important advantages such as:

1. Automation
2. Enhanced data analysis (EDA)
3. Improving user experiences
4. Provides creative solutions

On the other hand, machine learning enables computers to learn without explicitly programmed by hand. The name itself makes it clear that it allows the computer to learn, which makes it more like a human.

Nowadays, machine learning is being actively used to enhance a variety of tasks in a variety of professional domains. The following are some characteristics of machine learning:

1. Because machine learning relies on data, it needs an absurdly high volume of data to function accurately and precisely. Every day, organizations generate vast amounts of data. Organizations can thus help obtain significant insights from data analysis and, as a result, make better decisions by establishing noteworthy relationships in the data.
2. The machine can automatically improve its decision-making by learning from historical data.
3. It finds different patterns in the data from the provided dataset, which is
4. Because it works with vast amounts of data, it necessitates extensive data mining.

Certain techniques involve informing the machine in advance about the independent (input) and dependent (output) variables. By examining a collection of data known as the "training dataset," the machine discovers the connection between these two categories of variables. Several steps are taken to pre-process the data before training a model.

Traditional Methods for Real Estate Pricing:

As reAlpha shares in their blog (*Traditional Methods and Advanced Techniques for Pricing Real Estate*, 2024), Lets understand what types of methods and techniques were used in the past and are used today. Following are some traditional methods which are used for appraising real estate properties:

1. Sales comparison
2. Cost approach
3. Income approach
4. Comparative Market Analysis (CMA)
5. Broker Price Opinion (BPO)

Advanced Methods in Real Estate Pricing:

Real estate pricing estimation is a complex process that requires a deep understanding of the market, the property, and the buyer. While traditional methods like the sales comparison approach, cost approach, and income approach are still widely used, advanced methods are increasingly being used to provide more accurate and insightful valuations. Following are some advanced methods which are used for appraising real estate properties:

1. Automated Valuation Models (AVMs):

Automated Valuation Models (AVMs) are computer algorithms that analyse multiple data points and use statistical models that analyse data and take actions to generate a valuation for a property. AVMs play a significant role in real estate valuation because they provide automated and data-driven estimates of a property's value.

AVMs contribute to real estate valuation as shown in the following ways:

- Efficiency and Speed
- Data Analysis
- Comparable Sales
- Mathematical Modelling
- Market Trends and Conditions
- Portfolio Analysis
- Initial Screening

2. Machine Learning Models:

Machine learning models play a significant role in analysing large datasets to uncover patterns and trends. In the real estate sector, these models can evaluate various features such as location, size, age, and sales data to estimate property values. Additionally, they can forecast future trends and highlight potential risks.

Machine learning models are utilized in various ways for real estate pricing. Here are some common applications of machine learning in real estate pricing:

- Property Valuation
- Feature Selection
- Comparable Sales Analysis
- Demand Prediction
- Price Optimization
- Market Segmentation
- Risk Assessment

3. Predictive Analytics:

Predictive analytics is that domain of data analysis, or the use of statistical algorithms and machine learning methodologies in order to detect patterns, trends, or even a mixture of both within a dataset. In the context of the real estate business, therefore, technology can be applied for historical study of sales data as well as dynamics of market capture, and enable further predictions of price movements in future states of development. Consequently, investors and developers can make strategic decisions regarding the best location for investment or property development activities.

Predictive analytics has immense importance in real estate pricing strategies as it also provides useful information along with data-driven guidance. The following examples depict how predictive analytics impacts real estate pricing strategies:

- Accurate Property Valuation
- Market Demand Forecasting
- Pricing Elasticity Analysis
- Competitive Analysis
- Time-On-Market Prediction
- Optimized Pricing Strategies
- Investment Analysis

4. Spatial Analysis:

Spatial analysis is plays an important role because it incorporates the spatial relationships and geographic contexts associated with properties. The process involves the scrutinizing and interpretation of data within a geographical framework to determine the degree to which location affects property values. The following are essential aspects of the importance of spatial analysis in valuing real estate:

- Location Importance
- Neighbourhood Effects
- Comparable Sales Analysis
- Market Segmentation
- Geographic Constraints and Opportunities
- Trend Identification
- Geographic Visualization

The most important step in making a ML model is actually acquiring relevant data. Most of the collected data is not satisfactory, so part of it needs to be polished, and from this polished part, extraneous data needs to be eliminated to present appropriate data sets for our machine learning model. Calculating real estate prices is a complex procedure that would require deep knowledge of the market, the property in question, and the buyer. Chief sources of data used to determine prices for real estate are drawn from public records, electronic real estate listings, MLS, websites for real estate companies, county assessor's offices, real estate research firms, government agencies, trade associations, mortgage brokers, banks, and real estate value websites. Let's take a look were we can source this data about real estate pricing from various sources:

- Public Records and Real Estate Values
- Multiple listing services (MLS) Data and Real Estate Pricing
- Property Appraiser Records and Real Estate Valuation
- Tax Assessments and Real Estate Pricing
- Real Estate Market Reports and Pricing Strategies
- Satellite Imagery and Real Estate Pricing

Creating a Model for Housing Price Prediction using Linear Regression.

Linear regression is an excellent starting point for building a house price prediction model, providing a solid foundation for understanding how different features influence prices.

I will be using the Mumbai Dataset from Housing Prices in Metropolitan Areas of India dataset by (Bhatia, 2020), which contains various features and prices of new and resale houses. The following steps will be used to create and predict house prices:

1. Data Preparation
    a. Import
    b. Explore
    c. Split
2. Building a Model
    a. Baseline
    b. Iterate

Results communication :
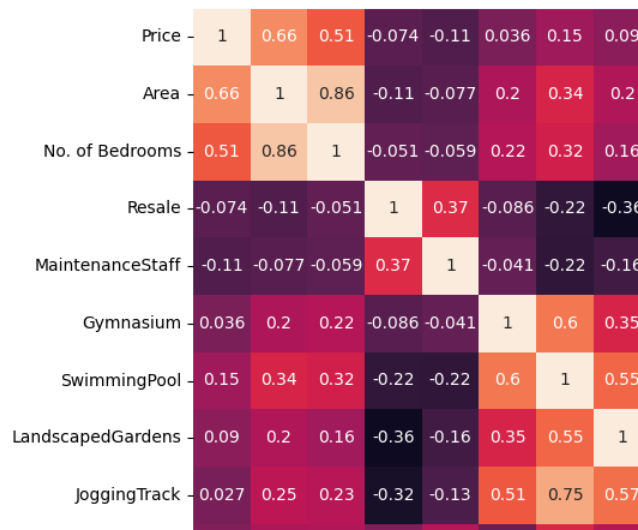
Fig 1.1 (Zoomed in heatmap of the mumbai.csv dataset) (Credit: Original)

From the zoomed in heat map plot in fig 1.1, it can be seen that the column 'Area' is strongly related column to the column 'Price'. Now we know that 'Area' has the highest correlation with 'Price'. (Prices are represented in multiples of one crore rupees, while Area is represented in Sq.ft)
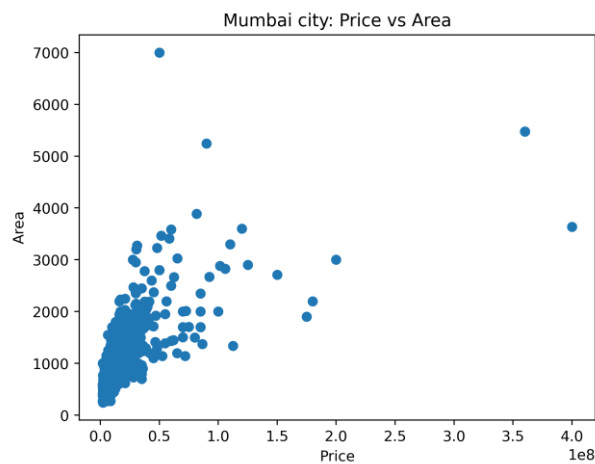

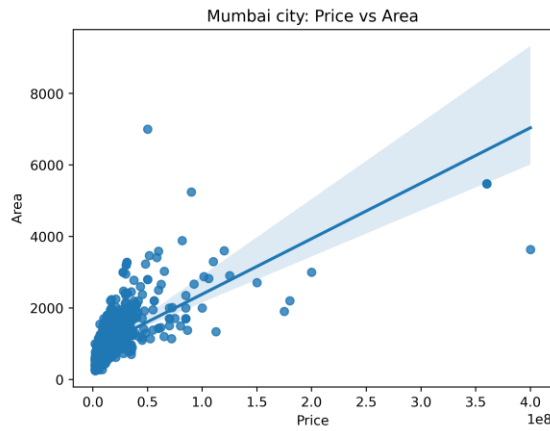**Fig 1.2 (Mumbai city: Area vs Price Scatterplot) (Credit: Original)**

Fig 1.3 (Mumbai city: Area vs Price Regplot) (Credit: Original)

Here I have successfully identified and filtered the outliers and below is how the final filtered data looks like
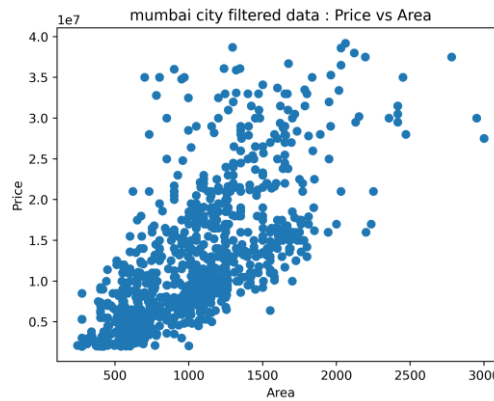

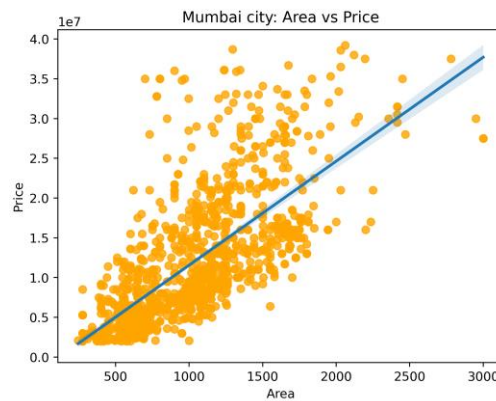**Fig 1.4 (Filtered data) (Credit: Original)**

Fig 1.5 (Filtered data regplot) (Credit: Original)

As I said earlier we are going to use a Linear Regression model to create our model to predict housing prices. I will fit the training set onto the model to train the model and then I will use the Test set to test the model and generate results. To check how accurate our results are, I am going to use Mean Absolute Error, Mean Squared Error, Root Mean Square Error values.
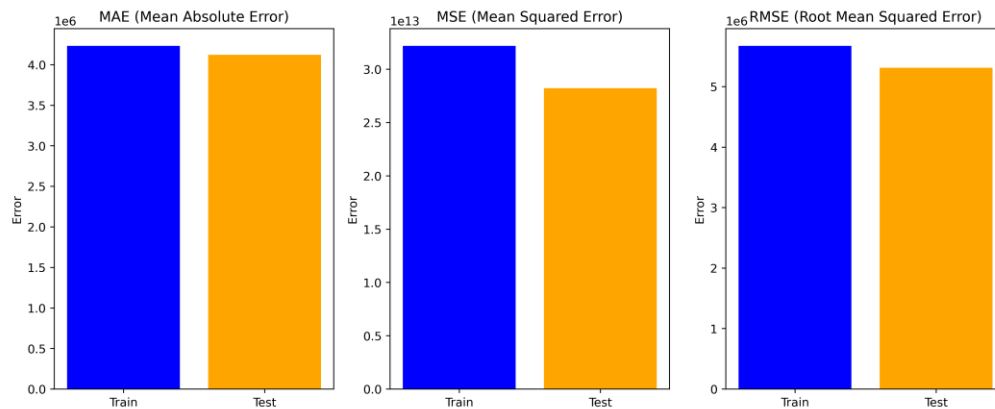
Fig 1.6 (Error comparison) (Credit: Original)

The article "Predicting Property Prices with Machine Learning" (Ho et al., 2021) compares the performance of support vector machine (SVM), random forest (RF), and gradient boosting machine (GBM) in predicting property prices using a dataset of 40,000 housing transactions in Hong Kong. RF and GBM outperformed SVM, though SVM delivered quicker predictions. The study highlights the potential of machine learning in property valuation.

In "Developing a Forecasting Model for Real Estate Auction Prices Using AI" (Kang et al., 2020), genetic algorithms (GA) performed best in predicting auction prices, followed by artificial neural networks (ANN) and regression analysis. The study emphasizes the importance of regional segmentation to improve prediction accuracy.

In another study titled "Artificial intelligence algorithms to predict Italian real estate market prices" by Luca Rampini and Fulvio Re Cecconi (Rampini & Re Cecconi, 2022), the authors utilized three popular machine learning techniques to predict house prices: ElasticNet, XGBoost, and Artificial Neural Network. They found out that the performance of the Artificial Neural Network (ANN) was superior to the other models, achieving a Mean Absolute Error (MAE) that was 5% lower than the second-best model, which was XGBoost. The ANN provided accurate predictions for medium to low-priced houses and had the lowest error for the most expensive ones, despite all models showing weaknesses in predicting the prices of the most expensive houses due to limited data

Results :

As you can see from Fig 1.6, the error between the training and testing set is very low, and the error of the testing set is lesser than the training set. This means that our model is performing good. But the R2 score of 0.53 says different and it means that our model's accuracy is moderate. (Error figures are represented multiples of 10 Lakh rupees, while MSE is in 100 crore rupees)

This mean now we can put out model to use to predict house prices, for that we create a formula as follows:

The formula: house_price = -1401539.58 + 12964.32 * surface_area_covered

(Where -1401539.58 is the model intercept and 12964.32 is the model coefficient.)
Red line in the figure 1.7 below represents the best fit of our linear model. Now we can create a function which takes Area as a parameter and outputs the predicted price and use this function wherever to generate predicted values and communicate our results

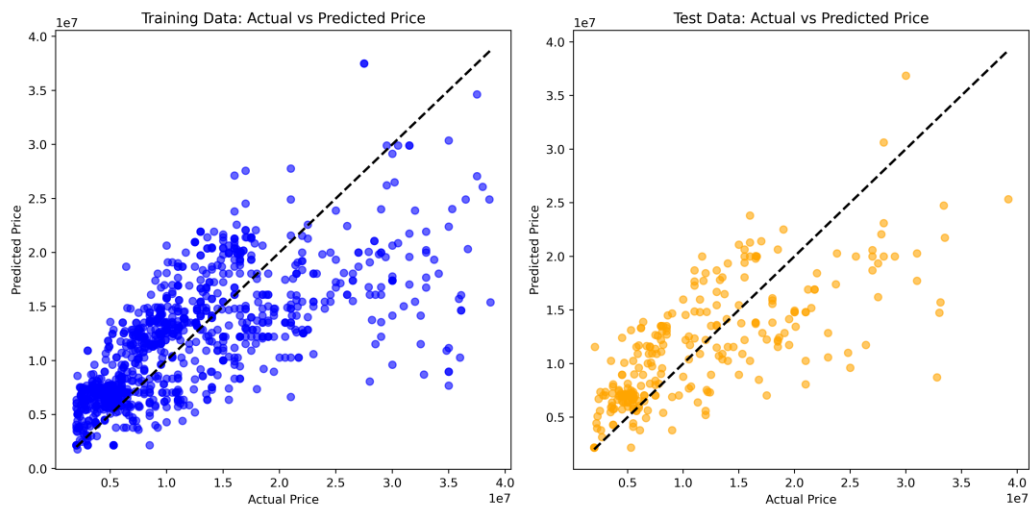**Fig 1.7 (Linear plot) (Credit: Original)**
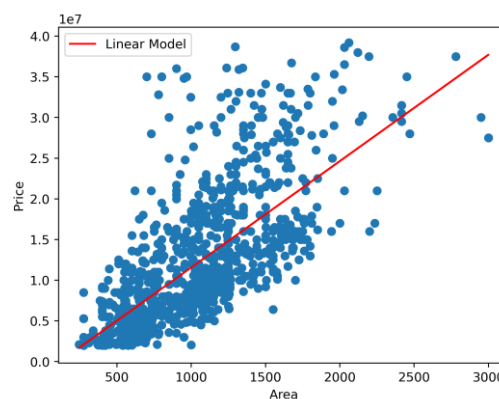
Fig 1.8 (Training vs Testing) (Credit: Original)

This paper attempted to illustrate how machine learning can provide more accurate price predictions than traditional non automated statistical techniques used in the real estate market. Generating accurate and precise results in real estate property market while taking various features into consideration by ML algorithms is an important task in itself that can be done easily with the help of AI & ML automated processes. The more data accumulated, the larger the valid size of the data required. This might cause some concerns over how the data privacy of the subject matters is maintained. In addition to this, there is always some likelihood that by mistake, private information is leaked or misused. Algorithms from machine learning can immensely affect aspects of others' lives, such as loan or employment approval rates. Any future research who is interested in this dataset can apply a better model to it, in order to gain better results.

Conclusion :

One can use these different machine learning techniques and apply them in different ways to the real world. In a paper published by (Ub & Saxena, 2023) the author has created a property price estimator model and has developed a model and applied to the real world by creating a mobile application for predicting the property prices in Bangalore. Following are the challenges and limitation which AI & ML need to tackle in the real estate industry:

- Challenges associated with property price estimation
- Data Accuracy and Real Estate Valuation
- Subjective Elements in Real Estate Pricing
- Market Volatility and Pricing Strategies
- Lack of Comparable Properties
- External Factors Affecting Property Decisions
- Legal and Regulatory Compliance
- Dependency on Technology

While real estate pricing techniques assist in valuing properties, they are indeed replete with challenges. From inaccurate data and subjective factors to market volatility and outside influencing factors, there are many types of obstacles that challenge the reliability and effectiveness of real estate pricing methodologies. Such obstacles have to be appreciated by buyers, sellers, and investors within the marketplace. This research sheds light on how AI & ML are being currently applied in real estate.

Even though this research reports some of the most recent applications for AI & ML technologies in real estate, it does not purport to report future applications or potential future changes to existing technologies. This research forms a good founding knowledge of what the present and the past functionality and applications are in place. It pays attention to any gaps or emerging areas that should further be considered by future researchers.

REFERENCES :

1. Artificial Intelligence(AI) in Tech. (2024, July 7). [Educational]. Artificial Intelligence(AI) in Tech. https://www.geeksforgeeks.org/ai-in-tech/?ref=ml_lbp
2. Bhatia, R. (2020). Housing Prices in Metropolitan Areas of India [Csv; Csv]. kaggle. https://www.kaggle.com/datasets/ruchi798/housing-prices-in-metropolitan-areas-of-india

3. Chen, Z. (2024). Integrating Machine Learning Techniques for Real Estate Analysis. Highlights in Science, Engineering and Technology, 92, 252–256. https://doi.org/10.54097/cvn9mr84

4. Dhar, T., & P, M. (2023). A Literature Review on Using Machine Learning Algorithm to Predict House Prices. International Research Journal on Advanced Science Hub, 5(Issue 05S), 132–137. https://doi.org/10.47392/irjash.2023.S017

5. Ho, W. K. O., Tang, B.-S., & Wong, S. W. (2021). Predicting property prices with machine learning algorithms. Journal of Property Research, 38(1), 48–70. https://doi.org/10.1080/09599916.2020.1832558

6. Kang, J., Lee, H. J., Jeong, S. H., Lee, H. S., & Oh, K. J. (2020). Developing a Forecasting Model for Real Estate Auction Prices Using Artificial Intelligence. Sustainability, 12(7), 2899. https://doi.org/10.3390/su12072899

7. Llp, G. T. B. (n.d.). Realising the potential of real estate.

8. Mathotaarachchi, K. V., Hasan, R., & Mahmood, S. (2024). Advanced Machine Learning Techniques for Predictive Modeling of Property Prices. Information, 15(6), 295. https://doi.org/10.3390/info15060295

9. Mullainathan, S., & Spiess, J. (2017). Machine Learning: An Applied Econometric Approach. Journal of Economic Perspectives, 31(2), 87–106. https://doi.org/10.1257/jep.31.2.87

10. Pagourtzi, E., Assimakopoulos, V., Hatzichristos, T., & French, N. (2003). Real estate appraisal: A review of valuation methods. Journal of Property Investment & Finance, 21(4), 383–401. https://doi.org/10.1108/14635780310483656

11. Rampini, L., & Re Cecconi, F. (2022). Artificial intelligence algorithms to predict Italian real estate market prices. Journal of Property Investment & Finance, 40(6), 588–611. https://doi.org/10.1108/JPIF-08-2021-0073

12. Shinde, N., & Gawande, K. (n.d.). VALUATION OF HOUSE PRICES USING PREDICTIVE TECHNIQUES. 5(6).

13. Ub, D., & Saxena, S. (2023). Real Estate Property Price Estimator Using Machine Learning. 2023 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES), 895–900. https://doi.org/10.1109/CISES58720.2023.10183626

14. Vyas, R., & Sharma, J. (2023). An Algorithm to Predict Real Estate Price using Machine Learning. Asian Journal of Computer Science and Technology, 12(1), 31–34. https://doi.org/10.51983/ajcst-2023.12.1.3592

15. What are the traditional methods and advanced techniques for pricing real estate? (2024, April 23). https://www.realpha.com/blog/real-estate-pricing-methods-and-techniques