

School of Built Environment, Engineering and Computing

Leeds Beckett University

**Predicting House Prices in the UK using Machine Learning Approach**

By: Jaydeep Gohel (77353725)

Supervisor Name: Hadeel Jazza

Submitted to Leeds Beckett University in partial fulfilment of the requirements for the degree of MSc Data Science

**May 2024**

# Candidate’s Declaration

I, Jaydeep Gohel, confirm that this dissertation and the work presented in it are my own achievement.

Where I have consulted the published work of others this is always clearly attributed;

Where I have quoted from the work of others the source is always given. With the exception of such quotations this dissertation is entirely my own work;

I have acknowledged all main sources of help;

I have read and understand the penalties associated with Academic Misconduct.

Signed:

JAYDEEP GOHEL

Date: 5 May 2024

Student ID No:773535725

# Abstract

In this report, the use of machine learning (ML) for the purpose of house price prediction in the UK is investigated, which takes into account the intricacy and changeability of the housing market. The earlier standard way of housing price predicting, normally with the help of experts or simple models, often shows to be inaccurate and unreliable, the turbulent market especially. The introduction of "big data" as well as advanced computational power makes it possible for ML algorithms of the highest sophistication that can effectively handle very large datasets and incorporate all variables, which affect house prices. These models take learning from data patterns a step further than the conventional ones by themselves forecasting more accurate and influential market movements. The research is intended to improve the investor, policymaker and future home-buyers' decision-making process by supplying forecasts which are more reliable and consequently help to have a real-estate market that is stable and fair. The study surveys for numerous ML techniques, for instance, regression models, clustering methods and neural networks, and then measures their performance in the unstable UK property market.

# Acknowledgement

I would like to express my deepest and most sincere gratitude to my supervisor Dr. Hadeel Jazza for their valuable guidance, support and encouragement throughout the entire duration of my project. Their vast experience and deep understanding of the subject has proved to be an immense help to me, and their profound viewpoints and extraordinary motivation enlightened me in many ways. My heartfelt thanks to them for the unlimited support and patience they have shown towards me, without which this work would not have been done in a proper and timely manner. I would also like to thank anonymous reviewers of the journals who reviewed my manuscripts. My sincere thanks to all classmates at Leeds Beckett University, with whom I had several useful discussions. I have also benefited a lot from my batch mates and friends throughout the entire course of study who helped me in many ways to proceed and whose presence somehow perpetually refreshed and gave me such a pleasant time in Leeds. Above all, I would like to express my greatest thanks to almighty for being my strength, thought, and action in accomplishing this task. I express my deep gratitude and admiration towards my family members who are always there to support me. Their love and confidence gave me the courage to face all the difficulties and hurdles with a smile.

**Table of Contents**

[**Abstract** 2](#_Toc165667457)

[**Candidate declaration** 3](#_Toc165667458)

[**Acknowledgement** 5](#_Toc165667459)

[**Chapter 1: Introduction** 9](#_Toc165667460)

[1.1 Background 9](#_Toc165667461)

[1.2 Problem Statement 10](#_Toc165667462)

[1.3 Aim: 11](#_Toc165667463)

[1.4 Objectives: 11](#_Toc165667464)

[1.4 Research Question 11](#_Toc165667465)

[1.5 Significance 11](#_Toc165667466)

[**Chapter 02: Literature Review** 13](#_Toc165667467)

[2.1 Introduction 13](#_Toc165667468)

[2.2 Overview of House Price Prediction 13](#_Toc165667469)

[2.3 Importance of Machine Learning in House Price Prediction 14](#_Toc165667470)

[2.4 Machine Learning's Drawbacks and Difficulties in UK Real Estate 15](#_Toc165667471)

[2.5 Collecting and Preparing Data for Predicting Home Prices 16](#_Toc165667472)

[2.5.1 Collecting Data 16](#_Toc165667473)

[2.5.2 Data Purification 17](#_Toc165667474)

[2.5.3 Preparing data 17](#_Toc165667475)

[2.5.4 Engineering Features 17](#_Toc165667476)

[2.5.5 Splitting the Data collection 17](#_Toc165667477)

[2.5.6 Data Converter 17](#_Toc165667478)

[2.5.7 Integration of Data 17](#_Toc165667479)

[2.6 The Effect of Machine Learning on Predicting Home Prices in the Changing UK Market 18](#_Toc165667480)

[2.7 Popular machine learning approaches to predict house prices 18](#_Toc165667481)

[2.8 Analysis of the Global House Price Index 19](#_Toc165667482)

[2.9 Literature Gap 20](#_Toc165667483)

[2.10 Chapter Summary 21](#_Toc165667484)

[**Chapter 03: Methodology** 22](#_Toc165667485)

[3.1 Introduction 22](#_Toc165667486)

[3.2 Flowchart 22](#_Toc165667487)

[3.3 Tools and Techniques 23](#_Toc165667488)

[**Chapter 04: Product/Research Design and Implementation** 25](#_Toc165667489)

[4.1 Introduction 25](#_Toc165667490)

[4.2 Model Evaluation 25](#_Toc165667491)

[4.2.1 Mean Absolute Error (MAE) 25](#_Toc165667492)

[4.2.2 Mean Squared Error (MSE) 25](#_Toc165667493)

[4.2.3 Root Mean Squared Error (RMSE) 26](#_Toc165667494)

[4.2.4 Rsquared (R2) Score 26](#_Toc165667495)

[**Chapter 05: Results and Analysis** 28](#_Toc165667496)

[5.1 Introduction 28](#_Toc165667497)

[5.2 Property Dataset 28](#_Toc165667498)

[5.3 Unveiling Sales Trends with a Box Plot 29](#_Toc165667499)

[5.4 Sales Volume Chart 30](#_Toc165667500)

[5.5 Mean House Price Index Comparison 31](#_Toc165667501)

[5.6 Model Performance Evaluation 32](#_Toc165667502)

[5.7 Scatter Plot of True vs. Predicted Values 32](#_Toc165667503)

[5.8 Scatter Plot of Actual vs. Predicted Values 33](#_Toc165667504)

[5.9 Analysis of Sales Volume Predictions for Bradford Region 34](#_Toc165667505)

[5.10 Conclusion 35](#_Toc165667506)

[**Chapter 06: Project Management** 36](#_Toc165667507)

[6.1 Introduction 36](#_Toc165667508)

[6.2 Project Planning and Scope 36](#_Toc165667509)

[6.3 Resource Allocation 36](#_Toc165667510)

[6.4 Risk Management 36](#_Toc165667511)

[6.5 Stakeholder Engagement 36](#_Toc165667512)

[6.6 Quality Control 37](#_Toc165667513)

[6.7 Project Monitoring and Evaluation 37](#_Toc165667514)

[6.8 Change Management 37](#_Toc165667515)

[6.9 Project Closure and Review 37](#_Toc165667516)

[6.10 Conclusion 37](#_Toc165667517)

[**Chapter 7: Conclusion** 38](#_Toc165667518)

[7.1 Introduction 38](#_Toc165667519)

[7.2 Limitations 38](#_Toc165667520)

[7.3 Future Recommendations 40](#_Toc165667521)

[**Reference:** 42](#_Toc165667522)

**Table of Figure**

[**Figure 1** 28](#_Toc165310783)

[**Figure 2** 29](#_Toc165310784)

[**Figure 3** 30](#_Toc165310785)

[**Figure 4** 31](#_Toc165310786)

[**Figure 5** 32](#_Toc165310787)

[**Figure 6** 33](#_Toc165310788)

[**Figure 7** 35](#_Toc165310783)

[**Figure 8** 37](#_Toc165310784)

[**Figure 9** 38](#_Toc165310785)

[**Figure 10** 39](#_Toc165310786)

[**Figure 11** 41](#_Toc165310787)

[**Figure 12** 42](#_Toc165310783)

# Chapter 1: Introduction

## 1.1 Background

Machine learning, know and like curriculum vitae as ML housing pricing, has great potential to shape the real estate environment. Conventionally, the forecast of house prices rely on traditional models or simply expert’ opinion. Consequently, their level of accuracy and reliability may be limited, especially in a turbulent and evolving market landscape. Conventional methods comprise a couple of approaches and therefore, may not be exhaustive enough to look into the complexities of the housing markets. The advent of "big data" in housing as well its availability from housing market data among other sources paves the way for more predict analytics and sophisticated modeling able to explore many links between these variables and gives consumers the ability to react to market-changing trends and to give them an improved forecasting. Freely use ML with house price data implies numerous of perspectives that are not easy to reach when compared to traditional methods (Soltani et al. 2022). ML algorithms are capable at handling high volumes of data fast, the inclusion of a plenty of factors that are relevant for the definition of the house prices being one. Moreover, ML models are capable of learning through data patterns and relations and, therefore, able to capture complexities and dependencies which can be extracted by means of the traditional regression-based approaches. AI has an inestimable potential to be of immense assistance to the above-mentioned groups, including suppliers, shareholders, and policy regulators. (Rawool et al. 2021).

In addition, the fruits of big data revolution in real estate are combined with the speed of computing power efficiency to foster ML development. These two forms of the use of technology and data, herwith, remain a distinct possibility for shifting strategies, forecasting house prices, and identifying challenges in the real estate sector that might persist for long. This research is hence designed to make input into the development of ML through the consideration and application of the pitfalls in the application of the modern ML schemes for the housing market. (Mohd et al. 2020).

## 1.2 Problem Statement

Predicting property prices precisely with a methodology of machine learning in the UK is very problematic because of all the economic difficulties and uncertainty in the housing market. General methods of house price forecasts by conventional developers tend to be simple models or subjective opinions of the specialists instead. Market factors, social or political factors, they come on the scene, in their complex interplay all of these are the ones that directly determine property values. The dynamics of housing market create difficulties and more delay for an accurate and feasible forecasting (Ragapriya et al. 2023). The market has unstable factors such as a supply and demand, changes in interest rates, economy relapses and state polices that cause market conditions to be frequently transformed. As a result, static models, which do not adjust their forecasts based on the changing market dynamics, may find it difficult to propagate reliable forecasts. Various other secondorder effects may affect those stakeholders like appraisers, brokers, investors, and policymakers. Homebuyers might end up making excessive payments for properties, the sellers might sell their homes below their values, investors may also make inadequate decisions, and city/state housing policies, which are not usually well implemented, may lead to economic losses, market failures, and housing markets that may increase and decrease. Weak forecasting systems prevent the agencies’ willingness to address important questions like housing affordability, market stability, and housing the poor with fairness. But, if forecasting is untruthful, policymakers cannot create targeted interventions to tackle the problem of rent or they can’t guard against the detrimental impact of market fluctuations. (Adetunji et al. 2022).

Hence, the demand to build and test ML algorithms is at the peak to solve the hard market of UK housing has been heightened. The software has to process huge masses of housing data, identify functions to capture the relationships between many factors, and adjust to the changing elements of the housing market and offer correct and consistently predictions. Adjusting to these issues will be vital for perfecting the decision-making mechanisms to support competitive market activity as well as to create a sustainable and just housing system in the United Kingdom. (Wang et al. 2021).

## 1.3 Aim:

The goal of this project is to reliably forecast UK home values using machine-learning algorithms, enabling knowledgeable real estate market decision-making.

## 1.4 Objectives:

1. Data gathering and prioritization should be done for the housing sector, including property characteristics, site characteristics, economic indicators, and historic value.

2. Develop and apply various types of machine learning algorithms, as regression models, cluster techniques, and neural networks, for house value prediction.

3. To quantify the developed scenarios by means of informative metrics and compare the performance against baseline models and industry benchmarks.

4. We will look into precisely the role of diversity in predicting house prices and what triggers house market growth for the UK housing market.

5. Considering the evidence on uses of machine learning methods and a prediction tool for home value in the UK as well as the factors that affect the housing market in the UK.

## 1.4 Research Question

RQ: What machine learning strategies and tactics may be applied to more accurately forecast UK home prices?

RQ: In light of the ever-changing and intricate housing market, what is the most effective way to apply machine-learning algorithms for predicting UK house prices?

## 1.5 Significance

An accurate price prediction of houses using a machine learning (ML)-based model for the UK is vital for a raft of parties such as investors, the banking sector, policymakers, and potential house buyers. ML’s predictions are increasingly being seen as a way to improve portfolios since they’re more accurate compared to more traditional techniques, providing a range of actors with the capability to gain insights and make strategic decisions, smartly regulate housing agencies and better manage risks. It signifies enhancing decision making procedures and hence contributing to development of more effective flexible housing market in the UK that uses ML approaches. In deployed of ML, it is likely to find complex lines of connection between different, constituent factors, which affect home prices, like market movements, local considerations, economic indices, and property characteristics. This feature is especially significant when the very factors on which the projections are based-be the market situation or shifting competition-become dynamic. This is especially difficult for static models, which static models struggle to adjust to the changing circumstances making the projections less accurate. Plausible forecasts result in anticipatory measures that are geared towards the resolution of home affordability and with a view of enhancing the financial well-being of the community too. Ultimately, the project aspires to provide all key agents with accurate and dependable forecasts using ML; this will boost the market effectiveness, economic stability, and social prosperity alongside the creation of a fair, sustainable and affordable Housing market in the UK. (Seyedzadeh et al. 2020).

# Chapter 02: Literature Review

## 2.1 Introduction

This chapter develops a thorough theory of the impact of machine learning on the modern business environment. This is a tricky task that I will be exploring the shortcomings of the model and how it applies to the UK real estate market whether it does help to improve or hinder the precision and effectiveness of predicting the weather is a question that the old methods of forecasting cannot properly answer. Most machine learning techniques are applied by different groups in today's murky data world. Added, the need for the empirical factual check and timely additional efforts in refining the thesis predictions is emphasized. In this section, we will be covering the downsides and barriers to using this factor in some practical examples as well. Introduces big-data into the UK real estate sector, alongside issues in the cyber security and personal data such as those with their data quality and accessibility too. Therefore, higher, continued investment is therefore needed to overcome the hindering factors. It underlines the fact that such movements are more significant than one individual or group can accomplish performing data collection and making the data ready to forecast home values, involving cleaning up and transformation of data, creation of data sets to make them preferred for machine learning algorithm analysis.

## 2.2 Overview of House Price Prediction

The housing costs are the backbone of economy, and forecasting the house prices is the key requirement for all such stakeholders as consumers, investors, and rule-makers to take wise decision concerning the real estate. The domain of house price forecasting can be described as an ocean of research that was consummated with many articles pretending different methods and approaches. A detailed review on the use of regression analysis in predicting housing prices for this instance could become impossible in a short essay such as this one. However, this research addresses regression analysis as the key approach for the job (J.S. A. Varma et al., 2018).

This is one of the biggest obstacles that the buyers often encounter while they try to estimate their available budget and plan their future as perceived strategically. They witness a pure soul that also have passion of not overriding their financial necessities and to date they are quite close in their housing objective with their available assets. The purpose of this analysis is to show the reasons behind high prices of homes, which is an issue peculiar to first-time home buyers lacking the capital to acquire a home. Through the back-test of listed assets and the assessment of the variety of prices and forecasted trend lines, we may be able to draw conclusions about the factual pricing patterns. House price prediction covers looking at the relations existing between multitude of variable factors, and the result of the algorithm. These factors include, but not limited to, address, land area, the number of bedrooms and bathrooms, amenities, school & transportation distance, socio-economic indicators, and demographic trends as well. Regression analysis helps in exploring and gauging the relationships here, while also permitting the delineation and the quantification of these relationships, thereby, leading to the formation of predictive models (Stojsin, Chen and Zhao, 2024).

Accessibility and quality of data are the defining factors on how well house price forecasting can be done. Historical sales data, properties for rent, economic indicators, and demographic information both constitute major elements used for creating predictive modelling. Although this is the problem mostly related to information technology, the data is not being collected or reported accurately with the same rules. While the other constraint, on the other hand, may arise from privacy issues and restrictions from statutory bodies such as the General Data Protection Regulation (GDPR) which can prompt the limitation of some data being accessed.

The development of machine learning along with computational power has boosted the building of more precise and close to real-time house price prediction models. Some of the utilized techniques are Support Vector Machines, Random Forest, Gradient Boosting, Algorithms, and Linear Regression, applied for a clearer view of inter-variable relationships and higher prediction accuracy. They can handle huge data sets and additionally, they can accommodate even the non-linear relationship that normal statistics methods can’t. Thus, these models predict and forecast more precisely than the traditionally accepted statistical methods.

## 2.3 Importance of Machine Learning in House Price Prediction

In machine learning as one of the AI branches, the software programs are now able predict and forecast data with more accuracy and learn and adapt from current data log in real-time which will be helpful in future analysis. This study focuses on a sample of the published works that base their forecasts and valuations on the application of the machine learning algorithms. One of the best strategies and techniques, when it comes to acknowledging and resolving the extremely critical, hard-to-detect, and difficult-to-analyze data structures and patterns out there in the environment is machine. The predicting of valuation of properties, which is one of the main applications of machine learning in real estate, is one of the uses. The traditional process is based mainly on the manual values and the personal judgments of the practitioner. On the other hand, machine intelligence is with an accuracy of price forecasting by analyzing the vast quantities of data composed of not only geography, property attributes, market trends, and prior sales data but also other factors. Driven by the independent data, that base the forecasts, these decisions are made by the investors, buyers and sellers with the sole aim of maximizing profits. Moreover, the incorporation of recommendation systems as well as 'Property Search' can greatly benefit from the use of machine learning techniques. Potential buyers can save time and effort throughout the search process by using these algorithms to give them with individualized property recommendations based on historical data, market trends, and user preferences. This increases customer satisfaction and accelerates the real estate transaction process (Begel, A. 2019).

## 2.4 Machine Learning's Drawbacks and Difficulties in UK Real Estate

Although machine learning has showed a lot of promise for the real estate sector, there are a number of problems and challenges that need to be resolved before using these methods in the context of UK real estate among the main drawbacks and difficulties are Accessibility and Quality of Data in which Large and high-quality data is a major requirement for machine learning algorithms. However, obtaining precise and current real estate data can be difficult, particularly in the UK where regional variations exist in data availability and standardization. Machine learning algorithms can be made less effective by incomplete or biased data, which might result in predictions that are not accurate and other drawback is Security and Privacy of Data in which Sensitive financial and personal information is involved in real estate transactions. Ensuring compliance with data privacy laws, such as the General Data Protection Regulation (GDPR), and protecting this data are essential. To avoid sensitive data being misused or accessed without authorization, machine learning models need to be built with strong data protection features.

A limitation in incorporating machine learning into real estate in the UK is that it set certain algorithms that are difficult to understand on one hand and, on the other hand, change is complicated to achieve due to them being ensconced in companies. Whereas conventional statistical techniques require one to have an explicit understanding of the correlation between variables, machine learning algorithms might sometimes appear to be absolute enigmas - making it hard for one to grasp how exactly predictions are made. This concern, however, could present a stumbling block on the way and therefore low interpretability can be a barrier for adoption, especially among the industries where the compliance and transparency are a matter of high priority. In addition, the industrialization of machine learning solutions for the UK real estate sector pays attention to its scalability and adjustments. For some large firms it could be easily done either because resources or expertise. Small businesses and independent agents, though, as they don’t have enough resources to do this at first. This scale-up problem powerfully amplifies the exiting spiraling effects of the market that eventually accurate the chasm perceived by those who possess sophisticated analytics and those who lack them (Smith and Track, 2022).

Besides that, the volatile character of real estate market disturbs the machine learning setup which usually uses the historical data for prediction. Instability in the market can be the main reason why the past pattern is likely to appearance again. Therefore, this would result to a bad prediction and a big financial loss. Considering how unstable and unpredictable markets can be, adaptations for machine learning models need to exist in order place them as a firm alternative in the UK real estate market. Finally, the consideration of machine learning ethical issues is an important aspect of using machine learning in real estate in which the algorithm bias and discrimination play an important role. Models are in turn trained using biased data or assumptions that are prejudicial; chances are that the bias will be in Learning and repeating the old conundrums. A fair and equitable use of machine learning (ML) algorithms requires a meticulous approach in data gathering, model development and mentioned problems’ ongoing monitoring while applying them (Akinrinola, Addy and Ajayi-Nifise, 2024).

## 2.5 Collecting and Preparing Data for Predicting Home Prices

Data cleansing and collection remain key stages during the development of the house prices prediction models. This consists of compiling relevant data, cleaning it up, formatting it for use in any machine learning algorithm, and putting it into the computing model.

### 2.5.1 Collecting Data

The government, home buying and tenancy sites, apps for listing, and housing market reports may be used as examples here. After that, get a handful of data elements such as the location of the house, its size, the number of bedrooms and bathrooms, and the facility of it, as well as transport and school access. Besides that, the past sales information and the market trends are also essential for a credible market value. In order to make sure, that the data is not distorted it is secondly necessary to gather in a proper method.

### 2.5.2 Data Purification

Remove the duplicate, false, or redundant entities in the dataset. Calculate the approximate value for the missing data or find other techniques for filling the void, such as the elimination of missing items.

### 2.5.3 Preparing data

Instruct double y-axis to ensure that their scale is the same. In this sense, it might be that either standardization techniques or normalization techniques will be implemented.

### 2.5.4 Engineering Features

Investigate the various types of information to spot any additional characteristics that could enhance the goodness of the model in predicting the future. Hence, plan to have fresh variables, such as the property’s age, price per square foot/meter, and area details to provide accordingly.

### 2.5.5 Splitting the Data collection

Create a portion of the dataset that are to be employed as the test, validation and also training dataset. The testing set is used to evaluate the model output after it is finished, the validation set determines the model components and parameter setting, and the training set is used by machine learning algorithms.

### 2.5.6 Data Converter

On Multi-dimensionality Issues. Perform additional modifications including, dimensionality reduction using known algorithms such as PCA or feature selection to remove duplicate or unnecessary information depending on the particular machine learning algorithms.

### 2.5.7 Integration of Data

To create a reliable analysis with proper estimations, ensure you are applying the process of data integration and alignment whether you are working with one or multiple datasets and data sources. These techniques are examples of the effective ways in practicing as well as conducting research for collecting as well as preparing the necessary data for the creation of accurate machine learning model for making house value prediction in the UK real estate market.

## 2.6 The Effect of Machine Learning on Predicting Home Prices in the Changing UK Market

In the age of exponential growth in technology, machine learning services are being applied by the real estate industry to achieve positive changes in the science of house values estimation. Disrupting nowadays, how stakeholders view the UK home market is set to change and a great new accuracy and speed known never before ever before is going to be the main factor of the integration. Explore the latest techniques and innovations involving the house price forecast and profit on the area of housing which always change in answer of time (Duca, Muellbauer and Murphy, 2021).

UK real estate market is going through the change in its faces due to the introduction of machine learning that allows for extremely high accuracy as well as fastness in house price estimation. Innovations and technologies have boosted the capability of stakeholders to predict market trends more effectively with a larger degree of precision and as well as adaptability. Tailor-made pricing strategy that depends on the characters of the property and the trends of market gives the landlord the most benefit. These groundbreaking forecasting methods not only increase accuracy, but they also redefine property valuation, investment analysis, and risk management that mark a new era of smartness and open up broader opportunities for the housing industry.

## 2.7 Popular machine learning approaches to predict house prices

As per Gupta et al. (2023), Numerous methods based on machine learning are used to predict home prices; here are five well-liked techniques which includes Linear Regression which is used for straightforward but efficient approach to establish a baseline and assuming a linear link between attributes and home prices , Random Forest which is used to identify complex patterns and generate reliable forecasts , which is particularly useful for non-linear interactions, Gradient Boosting Algorithms as they are exceptional at capturing intricate correlations and provide highly accurate predictions for home prices , Support Vectors Machines (SVM) finds the best hyper planes to divide data points by balancing both linear and non-linear relationships, which helps it be resistant to outliers in the prediction of housing prices and Neural Networks it is efficient at predicting property prices because they immediately pick up on minor developments, especially when dealing with complex and large-scale datasets.

In the same way as we have seen, other techniques, which were once considered a novelty, have recently become popular in predicting house prices. KNN algorithm uses the data pattern's majority class within its neighborhood to forecast the value for a new data point with an edge in local variations. Decision Trees initiate such segmentation data based turns on attributes values, that is ensemble machine learning methods like Random Forests are effective in multiplying a higher degree of accuracy in a crossway combination of many trees. The principal component analysis (PCA) converts the original dataset itself, in terms of the dimensionality whereas variance preservation is enhanced for proper model performance. Feature engineering consists of converting or producing new data that has a certain variation power, so it can be used to boost prediction accuracy. To increase the model's overall accuracy ensemble methods, such as Ada Boost and GBM, are used, by their approach of combining the base models. These approaches is a solid collection of the variety of methods such as which enables researchers and practitioners to customized the models to acquire the precise dynamics of real estate industry to achieve better prediction accuracy (Zhao et al. 2023).

## 2.8 Analysis of the Global House Price Index

Various nations, including the US Federal Housing Finance Agency, S&P/Case-Shiller price index, UK National Statistics, UK Land Registry, UK Halifax, UK Right move, and Singapore URA HPI, the House Price Index (HPI) is frequently used to track changes in the price of residential housing. Since the HPI is a weighted repeat sales index, it analyses average price changes in repeat sales or refinancing on the same properties (Mae et al. 2017).

Examining the Global House Price Indicator reveals encouraging information on the leveling the house prices in different countries and their rates of change. Some major contributors of the house capitalization index are amazing well-known organizations including US Federal Housing Finance Agency, S&P/Case-Shiller property price index, UK National Statistics, UK Land Registry, UK Halifax, UK Right move and Singapore URA HPI. The House Price Index (HPI) works as fundamental measurement changes of the homes prices, amongst all the stakeholders drives analysis of the housing market and for better-informed decisions.

The HPI is a house price index that serves as a weighted repeat sales index. It tracks average price changes in each individual transaction covering the same property such as refinancing or selling. This suppositions or techniques, therefore, are making it possible to evaluate property movement prices more accurately, because it looks at a number of benchmark units for the same specified time intervals. By doing this, analysts are able to keep tabs on the rapid changes and thus, they can identify the fundamental trends, measure the market pulse, and possibly envisage the future growth of the housing market. Furthermore, the GHPI gives a possibility to compare market conditions among different regions across the globe which generate a valuable information about regional variety and the market dynamics. (Ruipérez-Valiente, Jenner and Staubitz, 2020) Through analyzing the patterns in housing price on both a macro and micro scale, decision makers can create acceptable plans, investors implement more effective investment strategies while real estate pros tighten up the services they provide to their clients.

Besides being widely used in a residential house evaluation, the House Price Index is a reliable indicator of the general economic performance and stability. Unpredictability in home prices is commonly referred to as a sign of broader economic conditions, which include fluctuations in consumer confidence, employment rates, and credit policies. In this regard, the Human Psyche Index acts as an important factor in determining the health condition of the real estate sector and its resultant impacts on general economic condition (Riaz et al. 2020).

## 2.9 Literature Gap

The previous research has addressed the importance of home price prediction. This work, however, tries to fill the gap by providing a comprehensive exploration of advantages and disadvantages of the methods. With the intention of narrowing this knowledge gap, the present research gives a detailed explanatory on approaches and implement measures for the right forecasting among all. Researchers and the practitioners, in turn, can use these approaches to get together all the information and process it into high-quality machine learning models for value prediction in UK’s real estate market. Moreover, existing research notes that machine learning has changed the game on anticipating real estate prices by pointing out its ability to evaluate gigantic datasets with acceptable precision. After all, the research is still thin regarding the issues and downsides of using machine learning in real estate commercial research case in the UK. The study focuses on the defects in the current methods of house price forecasts in the UK and goes ahead with the research with a goal of improving the accuracy and precision of forecasting the house prices amidst the changing dynamics of the United Kingdom’s real estate market.

Prior studies have often focused on just home price prediction, but a literature review on whether different prediction strategies have compared advantages and disadvantages report pronounced gaps. The present study is an attempt of finding out the lacking in the approaches and measures applied for the forecasting by providing the detailed insight in to different methods and measures. This research is therefore intended to bridge the knowledge gap, and the research findings will be of great benefit to real estate in-country and to researchers and field practitioners who are in the business of developing machine learning models for predicting property values in the UK real estate market.

Extant literature which is devoted to investigating a machine learning contribution to real estate prices forecasting confirms its leading role in data processing and management, along with round-the-clock precision. Nevertheless, the debate on the limitation and disadvantages, which are characterize the application of machine learning technique in the commercial property research in UK has not been given enough time to be heard from. This study tries to narrow the knowledge gap regarding the UK house- price forecasting models by focusing on their shortcomings. To move forward, there is a need to conduct research, and this study proposes feasible means to enhance the precision and accuracy of current forecasts in the real estate market in the United Kingdom.

## 2.10 Chapter Summary

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Reference | ML Algorithms Used | Data Sources | Methodology | Key Outcomes |
| Soltani et al. (2022) | Various ML algorithms including regression | Housing market data | Spatio-temporal analysis | Improved accuracy by incorporating spatial dependencies |
| Rawool et al. (2021) | Random Forest | Real estate databases | Model evaluation using standard metrics (MAE, MSE, RMSE) | High predictive accuracy and reliability in volatile markets |
| Wang et al. (2021) | Deep learning with self-attention mechanisms | Heterogeneous data sources | Deep learning to analyze complex data interactions | Enhanced prediction precision using diverse data types |
| Mohd et al. (2020) | Various, including neural networks and regression models | Public and proprietary real estate data | Comprehensive data analysis and feature selection | Effective handling of large datasets with complex variable interactions |
| Ragapriya et al. (2023) | Extreme Boosting modified for real estate | Historical sales data | Advanced ensemble techniques | Accurate market trend analysis and prediction |
| Gupta et al. (2023) | Gradient Boosting, Linear Regression | Market trends, property attributes | Combination of various ML techniques | Robust model with high precision and adaptability to market changes |

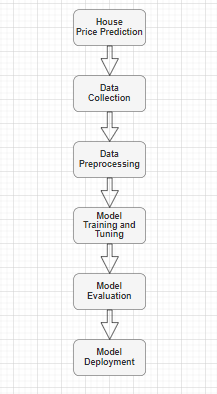
**Chapter 03: Methodology**

**3.1 Introduction**

In this chapter, they will illustrate the methodology and techniques for house price prediction in the UK by employing the Random Forest algorithm within machine learning algorithms. The goal of this research is to build a predictive model which would use particular attributes for a given house like location, size, amenities, and economic indicators to estimate the house prices in an accurate manner. The findings of such an analysis will supply with much needed data to the part of the real estate, including buyers, sellers, and investors.

**3.2 Flowchart**

The flowchart below illustrates the sequential steps involved in the house price prediction process:



1. **Data Collection:** Acquire needed data on house prices in the United Kingdom, that may back it up with the data concerning the location, size, amenities, and economic factors.

**2. Data Preprocessing**: Clean and prepare the data before analysis, so as to ensure accurate analysis. Compatibility is integration in this case so that they deal with cases of missing data, outliers, and anomalies; they also conduct feature selection, coding of categorical variables, and normalization of numerical features.

**3. Model Training and Tuning:** Take the preprocessed dataset and use it to teach a Random Forest algorithm. This comprises providing an early model with a set of suitable hyper parameters, among which are the number of trees, maximum depth, and minimum branches needed to split the root node. The model is then looked to the training data and the hyper parameters are tuned, until the quality of outputs is maximized.

**4. Model Evaluation:** While assessing the performance of your trained model, utilize the following metrics - Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and Rsquared (R2) score. The metrics evaluate the effectiveness of the model in forecasting house prices. They, mainly compare the model predictions with real values.

**5. Model Deployment:** Put the classification model into the environment to make predictions on unseen, never before data. This might be to simply embed them into an application, create an API for predictions efficacy, or by means of batch processing.

**3.3 Tools and Techniques**

**Framework:** Python is going to be working as a primary programming language for the model of machine learning. As Python does the primary job of its simplicity, numerous libraries, and a strong community's support, making it ideal for this project.

**ML Model:** Random Forest algorithm will like be the main ML model to develop the solution that predicts the house price. Random Forest is a combination of assembling learning methods that, while training, build multiple decision trees and then average those predictions made by the individual trees when the model is being mapped. Recently, Elon Musk proposed using this approach to develop a system that can learn from an enormous number of examples and potentially out-perform humans in many tasks (Robertson, 2023).

**Algorithm:** Random Forest is produced by growing large number of decision trees during the training with this average of the predictions as the final output. Each tree comprises the choice of a particular subset of features at random, which ensure variety and make the method less prone to the over fitting bias (Salam, Azar and Elgendy, 2021).

# Chapter 04: Product/Research Design and Implementation

## 4.1 Introduction

This chapter addresses the complexity of building a model for predicting house prices by using machine learning techniques. Since the property market is experiencing constant evolution, it becomes necessary to apply cutting-edge analytical methods to forecast the property values properly. This chapter explains the sequential procedure that includes the data collection and preprocessing and then the feature engineering and modeling options selection. At the core of the model is its evaluation which is essential for its effectiveness. This introduce and explain several key metrics used to assess the model's performance: MAE, MSE, RMSE, and R² represent Mean Absolute Error, Mean Squared Error, Root Mean Squared Error and R-squared, respectively. Each of the given metrics offers distinct clues about the efficiency and precision of the model, assisting in its fine-tuning. This chapter will be running dynamic tests for the validation of the model's effectiveness and also to refine the precision of the prediction by ensuring the model remains amenable to the dynamic nature of the housing market.

## 4.2 Model Evaluation

The accuracy of Machine Learning in the house price prediction is determined by the model's capacity to predict the correct price for each property by analyzing the different features. To assess the performance of their forecast model for virtual property prices, they employ evaluation criteria that bring various insights on the performance measures and predictive power.

**4.2.1 Mean Absolute Error (MAE)**

MAE is the first metric, which, mathematically, means the averaged absolute data distance between the predicted and the real house price. By performing either the absolute difference of their prediction and then average and then divide them by the total number of cases MAE is intended to have a good level of comparative accuracy measurement. The MAE gets lower, it means that the model's predictions are closer to the empirical value, thus indicating better performance (Huang et al. 2023).

**4.2.2 Mean Squared Error (MSE)**

The Measurement Squares Error (MSE) which is similar to MAE is utilized for computing the average of squared price differences between the actual and the predicted house prices. MSE in this way weighs the larger prediction errors more, which is beneficial for detecting the different values in the outliers, with which the model cannot cope, unlike MAE. The MSE calculates an E^2 term for each error and then averages them. The E^2 of a significant deviation between a predicted and an actual value is magnified, which helps gain a more nuanced evaluation of model performance (Landers, Armstrong and Collmus, 2022).

**4.2.3 Root Mean Squared Error (RMSE)**

The third measure, RMSE, which is analogues to the original MSE measure, is a direct interpretation of prediction accuracy, making it an easy-to-use metric. Through taking the square root, RMSE measures the dimension and reflect the scaling of MSE and target variable and therefore, facilitates the comparative understanding of the error. RMSE can be described as a tool because it focuses on the error levels that make huge impacts on the model sensitivity. The tool helps better evaluation of model prediction (Ahmed et al. 2020).

**4.2.4 Rsquared (R2) Score**

R2 (Rsquared) is a convenient tool that shows a complete assessment of a proposed model, including how well it can predict house prices. More specifically, R2 measures the proportion of total variance in the target variable that can be explained by the model. Values between 0 and 1 can show the degree of fit of the model to the data: the higher the model efficiency, the closer R2 is to 1. R2 stands for the coefficient of determination, and it actually measures the percentage of fluctuations of the target variable that the model explains. It signifies the goodness-of-fit of the model in respect to a baseline model that predicts the mean of the target variable. In this scenario, R2 provides the basic line of demarcation for assessing how advanced the house price prediction model being proposed (Chen, 2023).

R2 or the coefficient of determination alone determines the final performance of the predictive machine models where they apply in an area such as house price prediction. It is the relationship between this sum of squared residuals and the total sum of squares of the target variable. This number varies in values from 0 to 1. A high R² signifies a greater extent to which the model’s structure fits the data, inferring a stronger correlation between the features and the outcomes that are being predicted. And, of course, R2 should be understand as a part of the approach which includes the indicators to help recognize the problems like over fitting or biased predictions. With respect to house price prediction, realizing aspects like location, dwellings nuances and market dynamics is essential. This is necessary for achieving a high R2 score and also ensuring the model's forecasts are consistent with the day-to-day trends in the local real estate scene.

**Chapter 05: Results and Analysis**

**5.1 Introduction**

This chapter's purpose is to perform a comprehensive examination of the UK housing market by showing the visualizations and the statistic modeling. Housing is the largest and most influential economic sector in the UK, therefore a thorough knowledge of the movements of prices, the sale activities and the demand patterns over a period of time is essential for individuals, businesses and policymakers. The dataset and tools that cover a country wide property market and that are looked at in this section are intended to highlight some of the most important information that can be obtained. Through the usage of tools, such as box plots, line graphs, scatter plots and mean indexes, it can expose the existing differences between regions, long-term tendencies and predictive relations. Mastering these types of housing market mechanisms is of fundamental importance for the wise investment, development and policy choices in the field of real estate

**5.2 Property Dataset**

This property dataset which is extensive and contains information regarding UK house prices, metrics, and geographic identifiers offers a wide range of useful insights for different users. Such as including the name of the area and the link to land registry records that have unique identifiers for particular properties will definitely help in understanding the regional housing trends. Details on historical sales volumes and reporting periods give the basis for statistical analysis of the activity levels of the local property market over time due to the robust trend analysis. The house price index which tracks the average price fluctuation will assist in understanding the overall movement dynamics of prices in different areas. A combination of data is being used on different types of housing, for instance, detached and semidetached houses, terraced and flats, just to mention a few, which could be an indication of the segmentation that exists in local housing markets. Next, the reference date offered by this data source allows for manipulation and pivoting, which in turn makes a deeper analysis and exploration a possibility. Individuals, investment analysts, real estate developers and local authorities are among the many that can use this vast reservoir of housing data to make more informed decisions as to investing, planning and policymaking based on the property market trends, variations and projections for specific locations in the UK.

A black screen with a black background

Description automatically generated with medium confidenceA screen shot of a computer

Description automatically generated

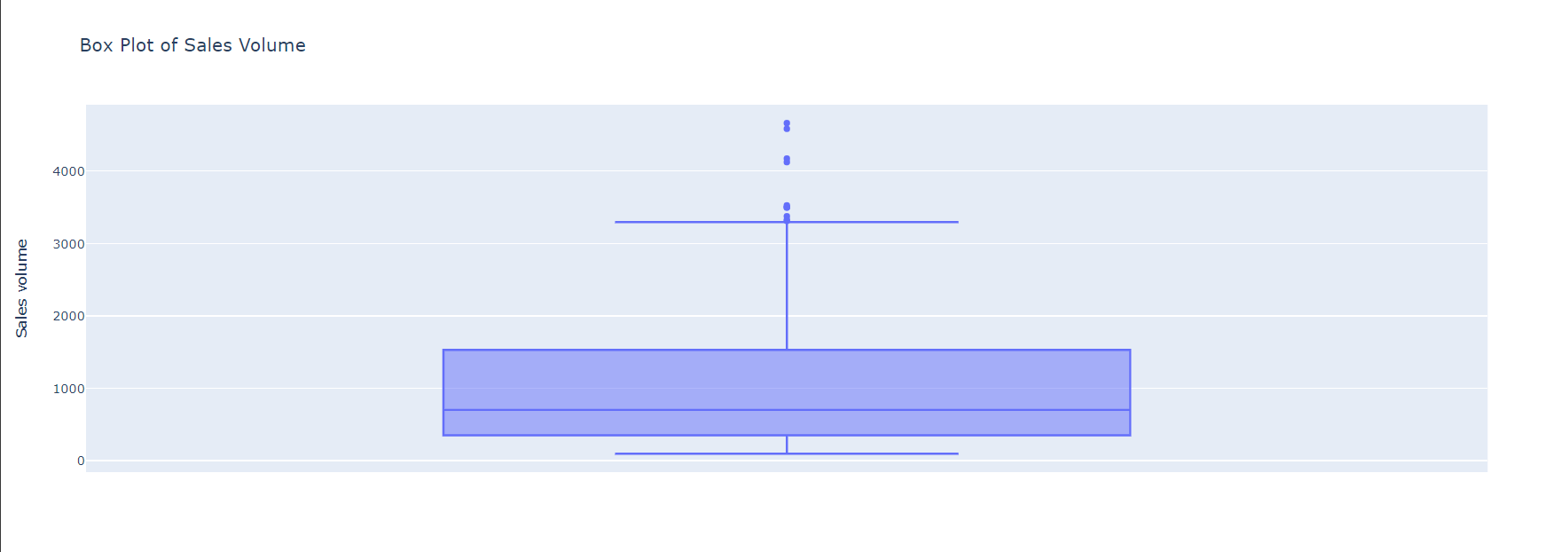
**Figure 1Property Dataset**

**5.3 Unveiling Sales Trends with a Box Plot**

This illustrative box plot shows a statistical tool that is widely used to analyze data and compare the distribution of quantity data across many categories. This type of data box plot allows you to show the spread and outliers in the annual sales volumes of different geographic regions. To be specific, it unveils five descriptive statistics of minimum, Q1, median, Q3, and maximum value. The box in the bottom and top represents the first and third quartiles, whereas the horizontal line inside the box is for the median. Whiskers are the lines, or whiskers, that are drawn from the boxes and which show the spread between the Q1 and the Q3 to the minimum and maximum values. These values are plotted as pairs of individual points, if they fall into the ranges mentioned above. Through the comparative box plots, viewers can discover the main factor that affects sales volume and, thus, make decisions on where to invest. You can observe this in regions with larger interquartile ranges that is shown by bigger boxes. Such regions are likely to experience bigger fluctuations in the transaction level from year to year. The point of a median line being higher in the box means that the upper half of the observations tend to have larger sales volumes. In this case, centerline closer to the top means the data tends to be distributed more towards the sales at the bottom. Overall, the box plot can be considered an effective tool of data visualization as it gathers and displays the distribution of the property market activity per year across the locations in an easy to comprehend form.

A screen shot of a computer

Description automatically generated



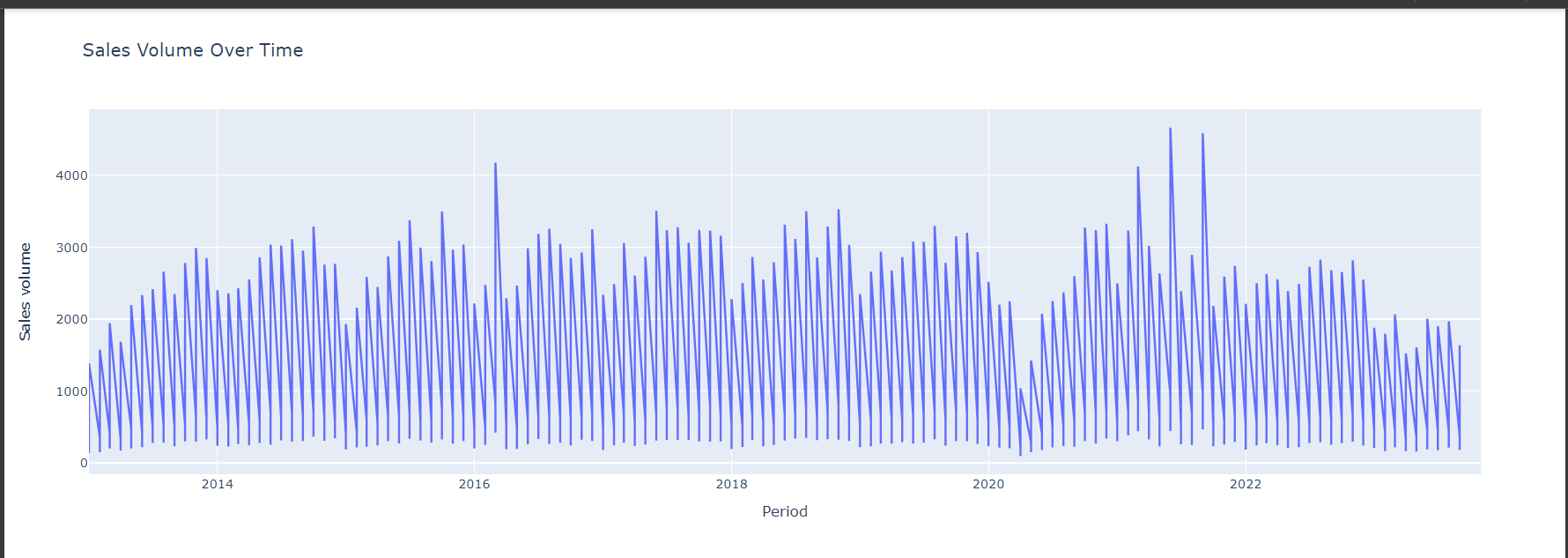
**Figure 2 Box plot of sales Volume**

**5.4 Sales Volume Chart**

This line graph allows for a highly visualized illustration of the sales volume trends that cover a longer duration, offering a great deal of valuable insights into the performance of the real estate market. The most significant blue line is the trail that shows the overall annual total transactions over time, the graph clearly indicates an overall upward trend. A further insight can be found in the fact that sales grew from year to year and that they had minor fluctuations only, thus indicating a stable growth. Viewers gain a complete grasp of the market movement by watching the line moving up or down to share the market's ups and downs. Furthermore, the chart is made more informative by the inclusion of the timeline, for instance years on the x-axis, permitting the correlation of the fluctuations with the external factors to be more precise. In addition there is a scale of the numerical sales values the y-axis presents which is the objective reference for the size of the rises or falls. Therefore, the tool represents an efficient dashboard-style analyzer for the high-level view of the long-term trend in the activity of the real estate market. In the light of this graph, real estate pros, investors, and analysts can apply the insights they have gained to detect market trends and predict demand levels in the future and make business decisions at the right time to take the most advantage of the opportunities to keep the market growing.

A screen shot of a computer

Description automatically generated



**Figure 3 Sales Volume over time line chart**

**5.5 Mean House Price Index Comparison**

The graph displays the average house price index for various towns. The chart has a blue line that is the mean house price index, and the x-axis is the names of cities. The visual representation of the data makes it possible to compare the housing market in different places, with this analysis taking into consideration the mean house prices in every town. The graph shows how much the disparity reaches in mean house prices and, therefore, enables us to acquire knowledge about the housing market, find some patterns and trends. Such data could be of great help for realtors and businesses involved in the real estate sector, as it allows them to understand how the pricing dynamics in various towns may affect the investments there. The graph serves the purpose of a quick and concise way of displaying mean house price index which can be used by users to base their decisions on the visualized data.

A screen shot of a computer

Description automatically generated



**Figure 4 Mean House Price Index for each Borough**

**5.6 Model Performance Evaluation**

MSE, also known as mean squared error, is a risk metric used to evaluate the quality of performance of regression models in machine learning. It is the average of the squared errors, or the deviations, between the predicted values and the actual observed data. In this MSE, the term mean squared error means the mean squared difference of the predicted values and the corresponding true outcomes. The average square of the residuals from the real data and the predicted data was 1,149,122.53 units. Lower MSE values means higher model fit because they indicate the case of smaller errors on average. Consequently, on the whole, this relatively high MSE shows that the model possesses a certain margin for improvement in terms of its predictive accuracy. Furthermore, the R-squared statistics is also shown as -0.4065. R-squared is a measure of how well the regression model explains the variation in the dependent variable. The higher the R-squared value, the better the model fits the data. Random forest models, with positive values near 1, are usually indicators of high predictive power, but they can produce negative R-squared numbers because of the model's tendency to over fit the data.

**A screen shot of a computer program

Description automatically generated**

***Figure 5 Mean Squared Error***

**A computer screen shot of a code

Description automatically generated**

***Figure 6 R-Squared value***

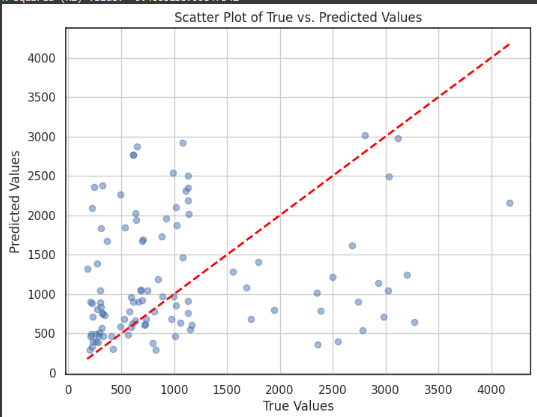
**5.7 Scatter Plot of True vs. Predicted Values**

In this scatter plot can be seen blue dots which represent the true values and the red line is the estimated line. By doing this visual analysis, they seek to find out whether the model will be able to forecast with high accuracy. The significant feature here is the models’ forecasts. They should compare them with the real data. This will allow us to understand the possible errors in the model and how close it will be to the actual performance. Thus, analysis can be applied for model monitoring and for tuning or the model improvements purposes.

Scatter plot of agreement versus disagreement is the ultimate analytical device to judge the house prices forecasting ability of the model with predictive mode. It enables to study a full quantitative picture of the degree in which predicted values of the model agree with the really observed ones. It is the red line that is the result of modeling and shows the median value of a house in the region. Each little blue dot is the actual measured property value. Either visually comparing the actual values with the pre-defined values or comparing the distribution of true values to the predicted values, the analyst can gauge the accuracy and reliability of the forecasting of the model. Any mismatch found between the values predicted and those obtained realistically signifies incorrectness or unreliable predictions in the model. By obtaining such insight the ways of these maximizations can be assessed and decisions concerning the potential for model improvements and optimization strategies can be reached.

A computer screen shot of a program code

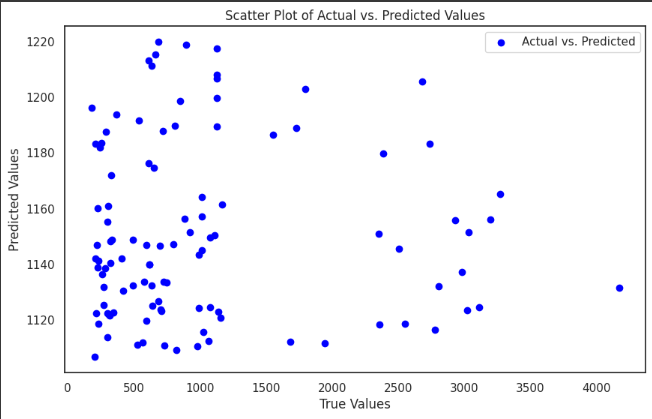
Description automatically generated



**Figure 7 Scatter Plot of True VS Predicted values**

**5.8 Scatter Plot of Actual vs. Predicted Values**

The appearance of single points in the scatter at the intersection of the TRUE and PREDICTED axes shows the association between the real data and the model predictions by the Random Forest Regression model. Each dot of the graph is particularly tied to a specific data point, and it is where the x-axis and y-axis are illustrated to meet the real and the predicted values. Thus, the data contained in this graph has an average value that is lower than their predicted values, which indicates that the model is more suited to overestimate these values. The larger the spacing among the studied population, the bigger the relative variation between the real and neighboring attributes. Scatter plots are go to visualizing the association between variables of an actual values and such as the random forest forecasts, providing a visual illustration of model accuracy and bias detection. Outliers and errors are clearly depicted and easily identifiable. Also, a distribution of the model computation results is assessed to unveil prediction accuracy and systematic biases. On the other hand, they are coincident with the disadvantages, namely the over plotting when in a big dataset, the restriction on two dimensions only, and the difficulty in establishing the depth of data, thus the real spread of the points is undervalued. What's more, scatter plots gives no inferences on issues related to causality or where the disparities spring from. A variety of graphical techniques can be applied such as horizontal lines, transparency, or jitter, and data might be supplemented with additional statistical analyses or visual representations, all of which will present a better picture about the features and weaknesses of the model.



**Figure 8 scatter Plot of Actual Vs Predicted Values**

**5.9 Analysis of Sales Volume Predictions for all Region**

**A screen shot of a computer program

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

***Figure 8 Bradford and Leeds DataFrame***

**A screenshot of a computer

Description automatically generated**

***Figure 9 West Yorkshire DataFrame***

**A screenshot of a computer

Description automatically generated**

***Figure 10 Calderdale dataFrame***

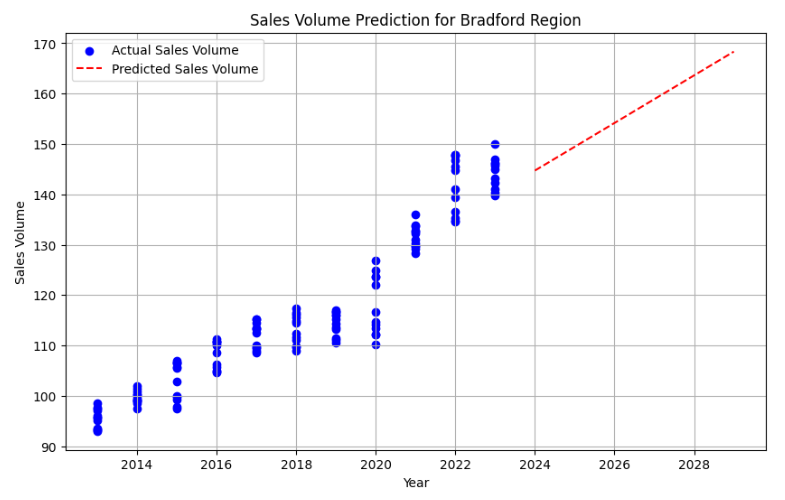


Figure 11 Sales Volume Prediction for Bradford Region

A graph with blue and red dots

Description automatically generated

***Figure 11 Sales Volume Prediction for Leeds Region***

A graph with blue dots and red arrow

Description automatically generated

***Figure 12 Sales Volume Prediction for West Yorkshire Region***

A graph with blue dots and a red dotted line

Description automatically generated

***Figure 11 Sales Volume Prediction for calderdale Region***

The plot displayed both the past and predicted sales volumes in the all regions. At the beginning the data sales volumes, shown by blue dots, were usually stable from 2010 to 2022. The range was always between 110 and 120, which means that the market was stable in this time period. In 2022, a change towards greater actual sales volumes was observed. This can be considered a positive change in market conditions or a successful regional market. As I looked into the future the graph came to life and the red dotted line signified the models’ projections for sales volumes. Since January 2024, the straight line growth scenario has been projected by the model. These figures indicated that the sales volumes could continue to increase, approximately reaching 145 by 2024, consecutively moving up to about 168 by the year 2029.

The forecasts were presented in a tabular form that shows the projected sales volumes for each year from 2024 to 2029. Therefore, this predictive analysis is a critical tool for both for private companies and local authorities. It ensures the development of a strategic plan that enables the stakeholders to foresee market occurrences and channel their resources to benefit from expected growth in the sales volume.

Top of Form

**5.10 Conclusion**

This chapter has made use of many charts and models to bring out the trends which are not obvious in the UK housing sector. By thoroughly analyzing the real national sales data, price indexes and predictive models, it was possible to detect useful patterns, which concerned volume fluctuations, price movements and forecasting accuracy. The box plots showed that transaction levels for local markets tend to vary from each other. Sales volume, that is, line graphs, increased steadily over the long term. Regional house price indexes demonstrated that such pricing disparities existed all over the country. Scatter plot was used to analyze the efficiency of the predictive model. Through a thorough understanding of these dynamics from both historical background and predictive viewpoint a real estate professionals and other stakeholders will be able to make effective strategic decisions. This also involves finding out the areas where there is room for improvement, entering the markets at the right time which are moving upwards, or tailoring policies to support affordable housing. The lesson extracted from looking into this data set could be a golden key for both business and government officials who are tasked with the management of the changing UK housing system both now and in the future.

# Chapter 06: Project Management

## 6.1 Introduction

This chapter presents the project management techniques together with the frameworks that guided the building of the house price prediction model using machine learning. Running such a task always involves a modeling of different aspects like time schedules, resources, stakeholders' expectation and risk factors.

## 6.2 Project Planning and Scope

The project scope was set to consist of both the creation and the verification of a machine learning model which is able to provide precise predictions regarding the UK house prices. Project planning was carried out by defining the reasonable goals, deliverables and deadlines which were delivered to all the stakeholders to ensure proper alignment and commitment.

## 6.3 Resource Allocation

In order to complete the project successfully, resource management was extremely important. Such leading to being in charge of personnel placement, budgeting, and the utilization of technology systems. Decisions about resources allocation were made in accordance with the plan phases and their importance, namely, to guarantee each stage of the model development is properly supported.

## 6.4 Risk Management

The proactive approach of identifying risks at various stages of the project was the common way of dealing with risks. These risks have been associated with issues such as the privacy of data as well as the uncertainties of the technology. Strategies had been adopted to avoid these risks which included the use of data security audits on a regular basis and the fallback technologies.

## 6.5 Stakeholder Engagement

Strategic communication with the stakeholders was the center of the communication plan that contained regular updates and feedback sessions. This step was critical to keep the stakeholders updated on the progress of the project and to incorporate their feedback into the development process.

## 6.6 Quality Control

The quality control systems were set in place to leave the machine learning model unaltered. The process incorporates the thorough testing stages, peer reviews, and iterative validation steps to make sure the model complies with the assumptions and accuracy level.

## 6.7 Project Monitoring and Evaluation

The continuous monitoring was implemented to evaluate the progress of the project according to the schedule. The project activities were measured by the following performance metrics and were used to gauge their effectiveness. Adjustments were made whenever necessary so that the project can stay on the right track and be on time with their goals.

## 6.8 Change Management

The project was managed with great flexibility to address any changes that might be needed. A change management formal process was available to manage any modifications in the project schedule, resources, or timelines. This way setup was done to make sure that all modifications were recorded, tested and implemented in the project progress.

## 6.9 Project Closure and Review

The process of closing the project, after the model was delivered, documentation was provided, and a post-implementation review was carried out. The lessons learned were recorded to be used in future projects, and the successes were celebrated with the team and stakeholders.

## 6.10 Conclusion

The disciplined approach followed to project management proved to be a crucial factor in the timely delivery of a successful machine learning model for predicting house prices. The solutions and plans adopted allowed the project to be carried out successfully with reactive adaptation to contingencies and potential gains. This chapter focuses on the value of project management with structure for building and managing complex data-driven projects and with a view to the other projects in the domain of machine learning and real estate analytics, provides the template.

# Chapter 7: Conclusion

## 7.1 Introduction

The dissertation revealed that various ML methods can be used for prediction of house prices in the UK real estate market with desirable outcomes. One of the profit techniques in the house price prediction area is the use of various models such as Support Vector Machines, Random Forest Algorithm, Gradient Boosting Algorithms, Linear Regression, and Neural Networks to enhance the prediction accuracy. It is unimaginable how far state-of-the-art methods have come from conceptual methods that used to provide just simple answers to the direct questions and ignored possible influence of multiple variables in the market.

Among the research findings was the possibility that ML would help gain ample information for a variety of stakeholders including investors, policymakers, and property agents. Thanks to historical data and predictive analytics, investors now can make more insightful decisions that may result in investment strategies that would target optimized outcomes which in turn result in increased market stability.

## 7.2 Limitations

Although the study was mostly successful, it had some obvious drawbacks that limited the level of detail and suitability to a certain extent. The data issue is primarily connected with the absence and the amount of the data. The precision of the predictions diminishes drastically if data is poor quality and lacks complete representation of the problem under study, for which they don't have access to complete or reliable datasets because of our data privacy or global data protection regulations like GDPR. This problem is especially acute in districts with different standards and data capture digitization which are unstable. Also, since the housing market is highly dynamic and it is determined by factors such as the economy, society, and politics, the implementation of machines learning model, which is the rigid one has high volatility and hence, it is difficult to train it. This constraint implies the fact that it is challenging to develop models which can appropriately react to highly dynamic market states in a very short period of time without significant lags of model updates or retraining.

Through their findings they demonstrated the suitability of ML applications for a future predictability of UK housing costs. The team applied a mixture of machine learning algorithms such as Support Vector Machines, Random Forest, Gradient Boosting Algorithms, Linear Regression, among others to real estate prediction, and has made huge steps in the accuracy of house price predictions. In recent years, the machine learning's algorithms have demonstrated its superiority over traditional methods, as they can provide a lot of data and capture the hidden interactions between variables which are characteristic for the real estate market.

Another barrier is market nimbleness, a descriptive characteristic. Economic, social and political factors are always in the forefront of the market constantly introducing volatility which makes it difficult for ML, particularly those models that lack versatility, to adjust accordingly. Such a limitation shows how hard it is developing models that can successfully address vastly changing markets where lean and fast updates are needed without losing much time on model retraining or updates. In order to cope with this, words of this research should concentrate upon the formation of adaptive models which reacts to the incoming data automatically without too much human intervention. Like solutions involving online learning and continuous model training, it is imperative to consider these as ways to keep a reliable prediction during an unstable market environment. Besides, a few more considerations, too are suggested that may facilitate the field of ML in the real estate price predictions. Initial and foremost the improvements should be performed for better data collecting and availability. A key research measure for future consideration should focus on collaborations between government agencies, real estate industry players, and other stakeholders to streamline data availability, standardization and quality for research. That's where adopting such strategies as partnerships to gather data that is both real-time and complete as the market goes around will be needed. The market needs models that can be created around ML systems that will be able to adjust fast in case the condition of the market is changing. This involves the application of recommendations including self-learning tools through which the models will routinely update and revise in line with newly acquired data. Besides that, on the go model training along the way should be considered so that the models do not lose the update and relevance. For the last, interdisciplinary methods should be also considered for the greater readability and dependability of the models. By combining economic theories such as monetary policy and macroeconomic indices, it would be possible for machine learning models to understand the factors underlying house-prices more comprehensively. The combination of the disciplines can result in real invaluable results and give us, and the market, as well, a more holistic point of view.

Ethical and regulatory compliance is not under any circumstances taken lightly. The necessity to maintain privacy of data will always go together with the development of artificial intelligence. Therefore, it is essential to guarantee that ML models comply with the corresponding laws and rules. Extensive data regulation models should be set in place to make sure the privacy of individuals is maintained and still trust among the stakeholders is observed. Ethical criteria also deserve to be incorporated while forming of ML models to tackle possible discrimination or prejudice and maintain a fair and even distribution while predictions. General public openness and transparency should be an assigned to most desire to enhance marketing of ML applications in real estate. Developing platforms that are user friendly with the provision of data analytic tools that promptly address stakeholders’ needs and desires to know the technique of predictive model creation help in demystifying the Moiré knowledge to the common consumer. Clarification in how each prediction is made, and the elements impacting these predictions, will lead to increased knowledge of the processes at play in turn giving parties confidence in the machine learning model predictions made. Adopting advancements in ML that will consider those obstacles and adopting my suggested recommendations, future studies can be capable to overcome challenges and improve the performance of ML in predicting UK real estate prices accordingly. The reactive development of policy measures would precipitate greater accuracy in decision-making, enhance market stability, and lend to the discovery of the underlying determinants of house prices. The underlying utility to the stakeholders is extended as everybody is able to access reliable information and make comparisons easily hence putting them in a position of making informed choices with confidence.

## 7.3 Future Recommendations

To address the identified limitations and enhance the utility of ML in real estate price prediction, several recommendations are proposed:

**1. Data Enhancement:** Future research should concentrate on the institution of cooperation with government and private institutions to increase accessibility to real time data acquisition systems and to more integrative and detailed collection systems. Fair practice must be adopted to integrate data collection and reporting system among different regions for better confidence of the feed data into predictive models.

**2. Model Adaptability:** Creating adaptive models able to dynamically process new data while automating processes reducing manual intervention is significant. The tools, among others, like online learning and continuous model training, could play vital roles in precision of the forecasts in a dynamic environment.

**3. Interdisciplinary Approaches:** By combining ML technology with the economic theory, it is plausible to argue that the forecasting models could to some extent be more rational and reliable. The importance of the macroeconomic indicators and market sentiment analysis possibly can’t be ignored. The ML model with market sentiment analysis and macroeconomic indicators will be able to see the market trends better.

**4. Ethical and Regulatory Compliance:** The information privacy issue is only increasing in all the walks of lives day by day; therefore, it is imperative to check the data models’ compliance with all the existing national and global laws and regulations. Building up solid data governance and ethical frameworks will be indispensable for an ML application to remain credible and acceptable not only in the real estate industry but also others in future.

**5. Public Access and Transparency:** Information portals that present clear instructions and algorithmic designs to stakeholders can help build up trust and spread their applications. Demonstration of how the predictions are made and the factors that influence this process will make mysterious ML applications in real estate realistic entrepreneur models for an average user.

# Reference:

A. Varma et al., "House price prediction using machine learning and neural networks", 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), pp. 1936-1939, 2018.

Adetunji, A.B., Akande, O.N., Ajala, F.A., Oyewo, O., Akande, Y.F. and Oluwadara, G., 2022. House price prediction using random forest machine learning technique. Procedia Computer Science, 199, pp.806-813.

Ahmed, M., Ahmad, S., Raza, M.A., Kumar, U., Ansar, M., Shah, G.A., Parsons, D., Hoogenboom, G., Palosuo, T. and Seidel, S., 2020. Models calibration and evaluation. Systems modeling, pp.151-178.

Akinrinola, O., Addy, W.A., Ajayi-Nifise, A.O., Odeyemi, O. and Falaiye, T., 2024. Application of machine learning in tax prediction: A review with practical approaches. Global Journal of Engineering and Technology Advances, 18(02), pp.102-117.

Bagel, A. (2019). Software Engineering for Machine Learning: A Case Study. Microsoft Research, 1-10

Chen, M., 2023. Spatially Explicit Machine Learning Approaches for House Price Models (Doctoral dissertation).

Duca, J.V., Muellbauer, J. and Murphy, A., 2021. What drives house price cycles? International experience and policy issues. Journal of Economic Literature, 59(3), pp.773-864.

Huang, T., Liu, T., Ai, Y., Ren, Z., Ou, J., Li, Y. and Xu, N., 2023. Modelling the interface bond strength of corroded reinforced concrete using hybrid machine learning algorithms. Journal of Building Engineering, 74, p.106862.

Landers, R.N., Armstrong, M.B., Collmus, A.B., Mujcic, S. and Blaik, J., 2022. Theory-driven game-based assessment of general cognitive ability: Design theory, measurement, prediction of performance, and test fairness. Journal of Applied Psychology, 107(10), p.1655.

Mae, F. and Mac, F., 2017. Federal Housing Finance Agency. *Single-Family Guarantee Fees*.

Mohd, T., Jamil, N.S., Johari, N., Abdullah, L. and Masrom, S., 2020. An overview of real estate modelling techniques for house price prediction. In Charting a Sustainable Future of ASEAN in Business and Social Sciences: Proceedings of the 3ʳᵈ International Conference on the Future of ASEAN (ICoFA) 2019—Volume 1 (pp. 321-338). Springer Singapore.

Ragapriya, N., Kumar, T.A., Parthiban, R., Divya, P., Jayalakshmi, S. and Raman, D.R., 2023. Machine Learning Based House Price Prediction Using Modified Extreme Boosting. Asian Journal of Applied Science and Technology (AJAST), 7(1), pp.41-54.

Rawool, A.G., Rogye, D.V., Rane, S.G. and Bharadi, V.A., 2021. House price prediction using machine learning. Int. J. Res. Appl. Sci. Eng. Technol, 9, pp.686-692.

Riaz, S., Ahmed, R., Parkash, R. and Ahmad, M.J., 2020. Determinants of stock market investors’behavior in COVID-19: A study on the Pakistan Stock Exchange.

Robertson, C.B., 2023. Litigating Partial Autonomy. Case Legal Studies Research Paper, (4), p.109.

Ruipérez-Valiente, J.A., Jenner, M., Staubitz, T., Li, X., Rohloff, T., Halawa, S., Turro, C., Cheng, Y., Zhang, J., Despujol, I. and Reich, J., 2020, March. Macro MOOC learning analytics: exploring trends across global and regional providers. In Proceedings of the tenth international conference on learning analytics & knowledge (pp. 518-523).

Salam, M.A., Azar, A.T., Elgendy, M.S. and Fouad, K.M., 2021. The effect of different dimensionality reduction techniques on machine learning overfitting problem. Int. J. Adv. Comput. Sci. Appl, 12(4), pp.641-655.

Seyedzadeh, S., Rahimian, F.P., Oliver, S., Rodriguez, S. and Glesk, I., 2020. Machine learning modelling for predicting non-domestic buildings energy performance: A model to support deep energy retrofit decision-making. Applied Energy, 279, p.115908.

Smith, A.O. and Track, D., 2022. The Necropolitics of TRIPS: Neoliberal Disease and Slow Death in a Pandemic.

Soltani, A., Heydari, M., Aghaei, F. and Pettit, C.J., 2022. Housing price prediction incorporating spatio-temporal dependency into machine learning algorithms. Cities, 131, p.103941.

Stojsin, R., Chen, X. and Zhao, Z., 2024. A Model-agnostic Computational Method for Discovering Gene-Phenotype Relationships and Inferring Gene Networks via in silico Gene Perturbation. medRxiv, pp.2024-02.

Utkarsh Gupta, Himadri Vaidya, Rahul Chauhan, Chandradeep Bhatt, "House Price Prediction Using Gradient Boosting and Linear Regression", 2023 International Conference on Sustainable Emerging Innovations in Engineering and Technology (ICSEIET), pp.829-832, 2023.

Wang, P.Y., Chen, C.T., Su, J.W., Wang, T.Y. and Huang, S.H., 2021. Deep learning model for house price prediction using heterogeneous data analysis along with joint self-attention mechanism. IEEE Access, 9, pp.55244-55259.

Zhao, J., Han, X., Ouyang, M. and Burke, A.F., 2023. Specialized deep neural networks for battery health prognostics: Opportunities and challenges. Journal of Energy Chemistry.

# Appendices:

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screen shot of a computer

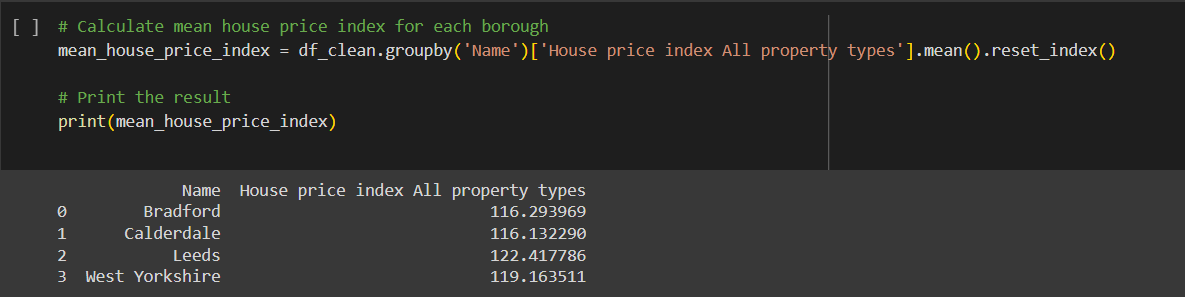
Description automatically generated

A blue graph with white text

Description automatically generated

A screenshot of a computer screen

Description automatically generated



A black background with colorful text

Description automatically generated

A graph showing a number of blue rectangular objects

Description automatically generated

A screen shot of a computer program

Description automatically generated

A blue rectangular object with text

Description automatically generated

A screen shot of a computer program

Description automatically generated

A screen shot of a computer program

Description automatically generated

A graph with blue dots and red line

Description automatically generated

A screen shot of a computer program

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screen shot of a computer program

Description automatically generated

A screen shot of a computer program

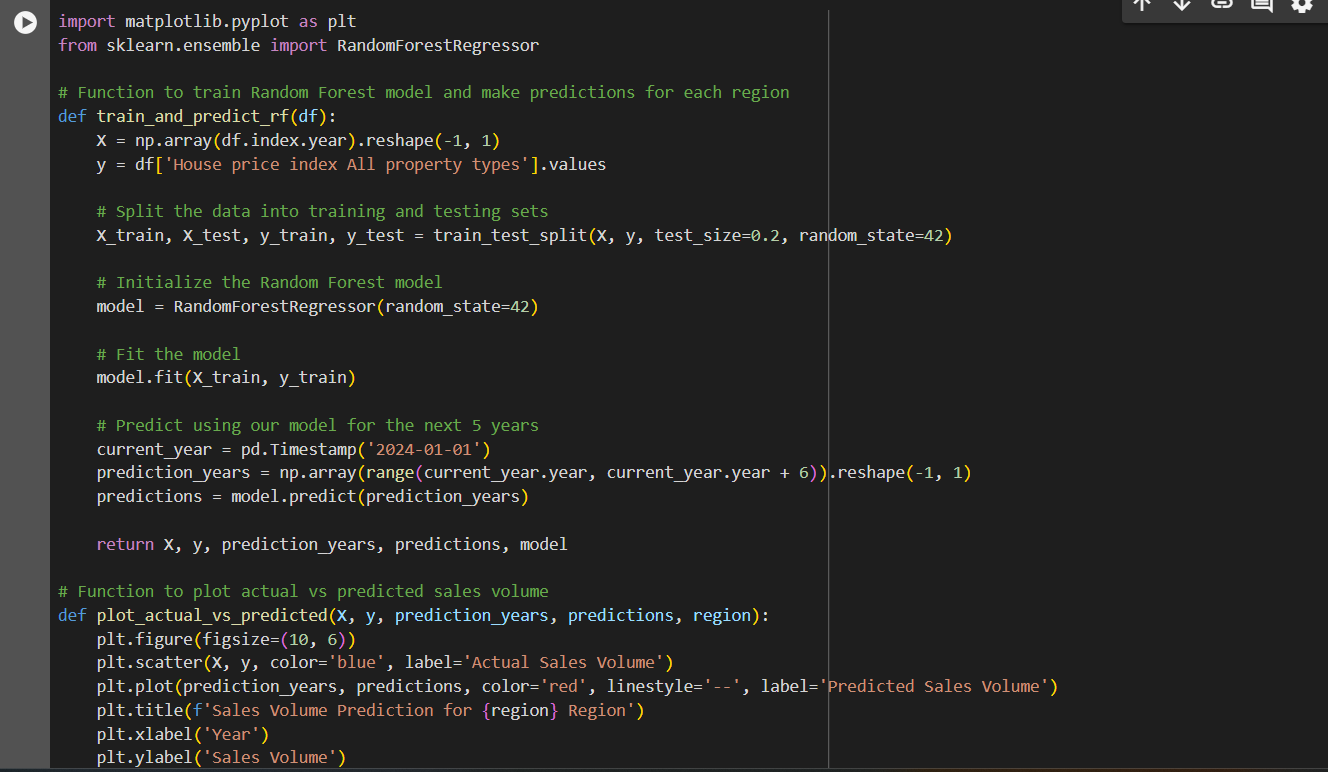
Description automatically generated

A screen shot of a computer

Description automatically generated

A screenshot of a computer screen

Description automatically generated



A screen shot of a computer screen

Description automatically generated

A graph with blue dots and red arrow

Description automatically generated

A graph with blue and red dots

Description automatically generated

A graph with blue dots and red dotted line

Description automatically generated

A graph with blue dots and red dotted line

Description automatically generated

A screen shot of a computer program

Description automatically generated

A computer screen shot of a program

Description automatically generated

A graph of sales

Description automatically generated

Ethic Provel:

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A white paper with black text

Description automatically generated