International Journal of Trend in Scientific Research and Development (IJTSRD)
Special Issue on Smart Innovations in Computer Science and Applications
Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

AI-Based Real Estate Spatial Search Engine

Shruti Anil Karekar
Department of Science and Technology,
G. H. Raisoni Skill Tech University, Nagpur, India

Annotation

The real estate market is changing fast because of new digital technologies and people wanting better ways to find properties. Normally people use things like price, location and what kind of property it is to search for houses.. These old ways might not give people the exact results they want. New technologies like Artificial Intelligence and special maps can make searching for properties a lot

The AI-Based Real Estate Spatial Search Engine is a system that uses maps and artificial intelligence to help people find properties. This system can help people find properties based on where they're how much they cost how big they are and other important things. We can use maps to show people where properties are which makes it easier for them to find what they want.

Artificial Intelligence can look at lots of property information. Figure out what people like to suggest properties to them. The system can also use ways of organizing property information to make searching faster. We can build this system using things, like HTML, CSS and React to make it look nice and easy to use and databases like MySQL and MongoDB to store all the property information. The real estate market and Artificial Intelligence can work together to make property search tools better.

Keywords: Artificial Intelligence, Real Estate Search Engine, Spatial Search, Machine Learning, Geographic Information System (GIS), Location-Based Search, Property Recommendation System, Map Visualization, Spatial Data, Real Estate Technology.

1. Intoduction

Today Artificial Intelligence and Machine Learning are used in the real estate sector in ways. When people look for properties they usually search by keyword. Use filters like price and location.. These methods are not good enough to find properties that really match what people want.

The AI-Based Real Estate Spatial Search Engine is a way to search for properties. It looks at a lot of data about estates. Uses Machine Learning to find patterns. This is different from using filters like price and location. The AI-Based Real Estate Spatial Search Engine uses Machine Learning algorithms to analyze data about price, coordinates, area, facilities, infrastructure development and neighborhood. This helps people find properties that're just right for them.

There are some Machine Learning models, like Random Forest, XGBoost and Linear Regression that can look at patterns in amounts of data about real estate properties. These models can help people find properties that match what they want. The real estate market is growing fast. The world is becoming more digital. This means that people need ways to search for properties. By using Machine Learning, spatial search and geographic data together people can use maps to find their dream

property. This helps them make decisions when buying a property. Artificial Intelligence and Machine Learning are really changing the way people search for properties in the real estate sector.

1.1. Motivation

The real estate market is changing because we have a lot more information about properties and technology like Machine Learning and spatial data is getting better. Now when people search for properties online they do not have options to narrow down their search to find what they want.

More and more properties are listed for sale or rent. It is getting really hard for people to find the one they like. Machine Learning can really help with this problem because it can look at a lot of information about properties and find connections between them that're not easy to see. This helps people find what they are looking for.

There are reasons why Machine Learning's important for making search engines that help people find properties. One reason is precision. Machine Learning can look at a lot of information about properties. Find connections between them. This makes the search results more accurate. Machine Learning can also suggest properties that're a good fit for what each person wants. It looks at what people do when they search for properties and finds patterns.

The real estate market and Machine Learning are related. Machine Learning is making the real estate market better by helping people find properties using data technologies and Machine Learning.

1.2. Contribution

The proposed AI-Based Real Estate Spatial Search Engine is intended to make traditional property search systems. This is done by using Machine Learning and spatial data analysis.

The main contributions of this project include:

Improved Accuracy in Property Search: Machine Learning algorithms process amounts of real estate data. This helps recognize relationships between attributes of properties.

Personalized Property Recommendation System: The system offers property recommendations to users. This is based on parameters, such as preferences and search patterns.

Efficient Spatial Search: The system incorporates data technologies. This enables property search based on locations through interactive maps.

Efficiency: The system can handle amounts of real estate data. This ensures property search results.

Competitive Advantage for Real Estate Companies: Spatial search engines enable real estate companies to offer property search solutions. This provides them with an advantage, in the growing real estate marketplace.

2. Related work

The real estate sector has undergone considerable digitalization through the development of online property search websites and recommendation systems. Existing real estate websites use keyword search techniques and filters like price, location, number of bedrooms, and type of property. These systems have increased the accessibility of properties for buyers. However, these systems have limited ability to understand user preferences. Many researchers have worked on enhancing property search systems by incorporating data mining techniques.

Recent research works have investigated the application of Artificial Intelligence (AI) and Machine Learning (ML) in the real estate sector for property price prediction, recommendation systems, and trend analysis. Machine learning algorithms like Linear Regression, Decision Tree, Random Forest, and Support Vector Machine have been commonly used for property price prediction based on historical data and location features. These algorithms enable users to make better decisions by providing information on the estimated value of properties in the market.

Another significant area of research in the field of real estate is spatial search and geographic information systems. Spatial search engines utilize geographic information and spatial visualization to search for properties based on the presence of nearby facilities such as schools, hospitals, shopping centers, and transportation infrastructure. Research has been conducted to show the benefits of using spatial intelligence and AI models in spatial search engines to provide a better user experience.

In the last few years, user behavior-based and user preference-based property recommendation systems have also emerged as a popular area in the field. These systems utilize user browsing behavior, clicked properties, and budget preferences to recommend the best possible property. Collaborative filtering and content-based filtering methods have been popular in such systems. However, the majority of the existing systems are lacking in spatial intelligence and real-time processing capabilities.

3. Research Methodology

3.1. Problem Statement

While browsing through various websites present on the Internet concerning real estate, it becomes quite difficult to find the property according to the criteria provided. Various techniques are present to search for the property according to the criteria provided by the user. However, the techniques provided do not consider the other criteria that are quite important for the user. The criteria may include the facilities present in the region, the accessibility of the region, infrastructure development in the region, etc. Due to this reason, the user is not able to find the property according to the criteria provided. Searching for the property using the techniques provided becomes an inefficient task.

To avoid this problem, an AI-based real estate spatial search engine has been proposed that can efficiently analyze the data provided concerning the users and the properties. Spatial data has been integrated with the data provided concerning the properties. Suggestions are provided to the users concerning the availability of the most relevant properties.

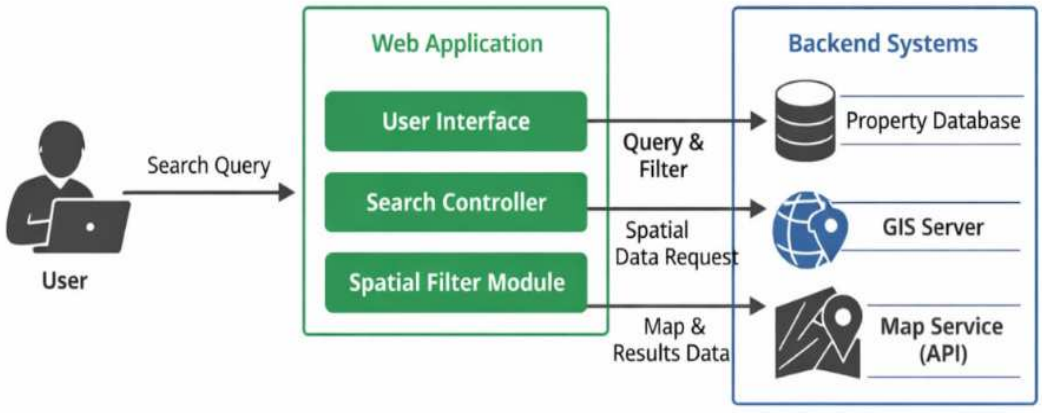

Figure 1: System Architecture of Real Estate Spatial Search Interface

3.2. Core Components

Data Sources:

Databases with property-related information such as price, location, area, rooms available, and facilities.

Data Preprocessing:

The data is preprocessed using ETL tools to ensure there are no missing data points.

Database Storage:

The property-related data is stored in a database using tools such as MySQL and MongoDB.

Model Training:

The machine learning models used in the analysis of property-related data are Random Forest, Linear Regression, and Gradient Boosting.

Prediction API:

A REST API is created using tools such as Flask and Django to enable the system to process user queries.

User Interface:

A web interface is provided to enable users to search property-related data based on parameters such as price, location, etc.

Visualization Based on Maps:

The system is equipped with a module for visualizing maps, which depicts the location of the properties in terms of latitude and longitude. This allows the users to discover the nearby facilities and comprehend the spatial relations among the properties.

Fig 3.2 Visualization Based on Maps

3.3. Data Flow

Property data is retrieved from real estate data sets.

Data is cleaned and ready for analysis.

Machine learning models are created and trained.

Trained models are implemented using an API.

Users input search queries through a web interface.

Users are provided with recommendations for properties.

3.4. Technical Requirements

Python - Implementation of Machine Learning

MySQL or MongoDB - Storage of Data

Flask or Django - Implementation of API

ReactJS or HTML/CSS - Implementation of Interface

Machine Learning - Scikit or TensorFlow

3.5. Model Training

The training of the machine learning model takes place by using the historical data available on properties. The data is divided into training data and testing data (80-20 rule). The training data is used to train the model with respect to the relationship between the attributes of the property, i.e., price, location, and area.

During the training of the model, the model is executed to generate the output and compare it with the actual price of the property. The difference between the actual price and the predicted price is calculated by using the loss function. Here, the loss function is Mean Squared Error (MSE). Stochastic Gradient Descent is used as the optimizer to optimize the accuracy of the mode

3.6. Hyperparameter Tuning

The process of hyperparameter tuning is used to optimize the performance of the machine learning model. Hyperparameters used in the process of machine learning model optimization are:

- Learning Rate: It is used in the optimization of the machine learning model's training speed.
- Batch Size: It is used in the optimization of the machine learning model's samples.
- Number of Trees/Layers: It is used in the optimization of the machine learning model's complexity.
- Number of Units: It is used in the optimization of the machine learning model's capacity.

The process of hyperparameter tuning is done in an improper manner, leading to performance issues in the machine learning model.

3.7. Competitive Analysis

The competitive analysis of various machine learning algorithms is done to find the most powerful algorithm used in property prediction.

Random Forest Algorithm:

The algorithm provides high accuracy in handling the relationship between property features.

Gradient Boosting Algorithm/XGBoost Algorithm:

The algorithm is efficient in handling property price and trend prediction.

Linear Regression Algorithm:

The algorithm is used in property prediction and understanding the relationship between various factors affecting property, including price and location.

Out of these three algorithms, Random Forest Algorithm and Gradient Boosting Algorithm are found to perform well in handling the complex pattern of property features.

3.8. Proposed Algorithm

Input:

Property price, Property location, Property area size in square feet, Number of rooms available in property, Facilities available around property, User budget

Output:

Recommended property, Recommended property price

Strategy:

The strategy for the proposed system is to use machine learning algorithms for data analysis related to property and show intelligent search results related to property based on the location of the property.

Algorithm:

Step 1:

Data Preprocessing:

Handling missing data by using mean and median methods.

Data attribute handling for property type and property location by using label encoding and one-hot encoding methods.

Data attribute normalization for property price, property area size in square feet, and number of rooms available in property.

Step 2:

Feature Engineering:

Feature selection for property size, property coordinates, facilities available in property, and infrastructure available around property.

Step 3:

Model Selection:

Machine learning algorithms are selected for better accuracy of results by using Random Forest, Gradient Boosting, and Linear Regression algorithms.

Step 4:

Model Training:

The machine learning model is trained based on the data available related to property and property price.

Step 5:

Model Deployment:

The machine learning model is deployed by using API for better usage of the model in real-time scenarios related to property search by using a web application.

3.8.1. Overall Workflow

Real Estate Dataset

↓

Data Cleaning & Preprocessing

↓

Feature Engineering

↓

Train-Test Split

↓

Machine Learning Model Training

↓

Property Price Prediction / Recommendation

↓

Performance Evaluation

↓

Property Search Results

3.8.2 Performance Metrics

For the proposed AI-based Real Estate Spatial Search Engine, various performance metrics are used to evaluate the performance of the machine learning model in terms of property price prediction and property classification.

MAE (Mean Absolute Error)

The MAE metric is used to compute the average difference between the predicted property price and actual property price in the real estate data.

RMSE (Root Mean Squared Error)

The RMSE metric is used to compute the square root value of the average value of squared differences between predicted property prices and actual property prices in the data.

R^2 Score (Coefficient of Determination)

The R^2 score is used to evaluate how well the machine learning model explains property price variation in relation to various property attributes like location, size, number of rooms, and amenities.

MAPE (Mean Absolute Percentage Error)

The MAPE metric is used to compute the average percentage difference between predicted property prices and actual property prices in the market.

Accuracy

Accuracy is used to evaluate how well the model classifies property types into various categories like high price property and low price property.

The accuracy formula is as follows:

Accuracy = \frac{TP + TN}{TP + TN + FP + FN}

3.8.3. Confusion Matrix

A Confusion Matrix is used to measure the performance of the classification for the real estate prediction model. It indicates the performance of the model for classifying properties based on price categories.

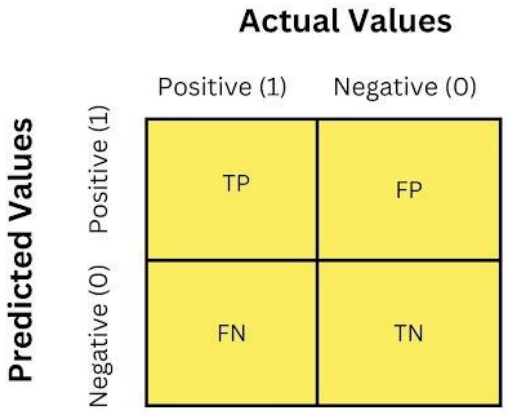

fig 3.8.3 Confusion Matrix

➢ *Interpretation*:

True Positive (TP): The system makes the appropriate recommendation for the property, which satisfies the search criteria of the user.

True Negative (TN): The system makes the appropriate decision in ignoring properties that are not relevant to the search criteria of the user.

False Positive (FP): The system makes the wrong recommendation, which does not match the search criteria of the user.

False Negative (FN): The system fails to recommend the property, which satisfies the search criteria of the user.

Accuracy

Accuracy = \frac{TP + TN}{TP + TN + FP + FN}

Accuracy measures the overall correctness of the property recommendation system.

Precision

Precision = \frac{TP}{TP + FP}

3.8.4 Area Under Curve (AUC)

The performance measure of the performance of the classification model employed in the AI-based real estate spatial search engine is given by the Area Under Curve (AUC).

The AUC is computed using the Receiver Operating Characteristic (ROC) curve. The ROC curve is computed using the True Positive Rate (TPR) and the False Positive Rate (FPR).

True Positive Rate (TPR):The True Positive Rate represents the percentage of properties recommended to the user correctly by the system.

False Positive Rate (FPR):The False Positive Rate represents the percentage of properties recommended to the user that are not relevant to the user.

Interpretation of AUC

AUC = 0.5 → The system is working randomly in the recommendation of the properties.

AUC = 1.0 → The system has the capability of correctly identifying the properties that are relevant to the user.

AUC > 0.7 → The system is working well in the recommendation of the properties to the user.

AUC > 0.8 → The system has the capability of recommending the real estate properties to the user with precision.

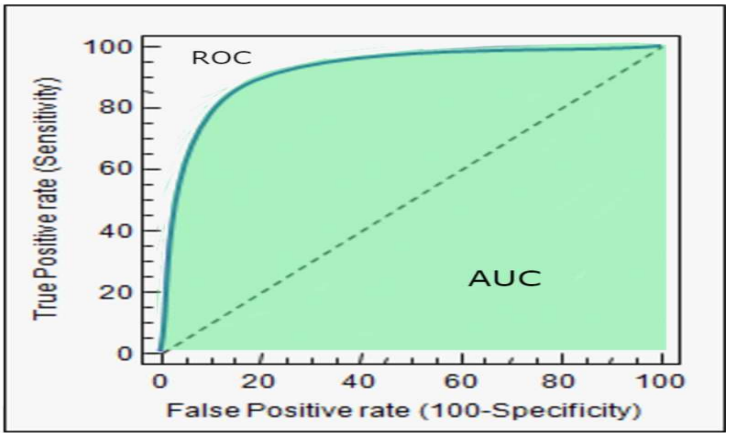

Fig 3.8.4 Area Under Curve (AUC)

4. Result Evaluation & Analysis

After the successful implementation of the AI-Based Real Estate Spatial Search Engine, the next step is to evaluate the results obtained. This will help to analyze the performance of the system. This will also help to identify the efficiency of the system in retrieving the relevant property information according to the query provided by the users.

The proposed system uses spatial data to provide efficient results to the users compared to the existing systems. The proposed system uses intelligent search techniques to provide efficient results to the users. The results are analyzed using the evaluation metrics.

4.1. Evaluation Metrics

The following are some of the evaluation metrics that are used to evaluate the performance of the proposed system.

Mean Absolute Error (MAE):

The Mean Absolute Error measures the average difference between the predicted property price and the actual property price present in the data set. Low values of Mean Absolute Error imply high accuracy of prediction.

Mean Squared Error (MSE):

The Mean Squared Error calculates the average of the squared differences between the actual property price and the predicted property price. This metric helps in identifying large prediction errors.

Root Mean Squared Error (RMSE):

The Root Mean Squared Error is the square root of MSE, which makes it easily interpretable for measuring prediction accuracy of the model.

R^2 Score (Coefficient of Determination):

The R^2 score measures the performance of the model in explaining property price variations or property ranking result variations.

Area Under Curve (AUC):

It is a measure of how well the model performs in distinguishing between relevant and non-relevant property listings in a search result.

4.2 Analysis

The performance of the proposed spatial search engine is analyzed using different techniques.

Comparison with Traditional Search Techniques:

A comparison of the performance of the proposed AI-based spatial search engine and the traditional technique for searching properties using keyword filters is done.

Selection of Significant Features:

Significant features which significantly influence the outcome of a property search are selected.

Error Analysis:

Analysis of errors helps to identify situations where results returned are not so relevant.

Performance of the System:

The performance of the system can be determined based on different factors including search response time and user satisfaction levels.

4.3 ROC Curve Interpretation

The Receiver Operating Characteristic (ROC) Curve can be used for the evaluation of the classification performance of the system.

The ROC curve can be defined as the relationship between True Positive Rate (TPR) and False Positive Rate (FPR) for identifying relevant property listings.

Interpretation of AUC values:

- AUC = 0.5: The performance of the system is similar to random search.
- AUC = 1: The system can perfectly identify relevant property listings.
- AUC > 0.7: The performance of the system is good.
- AUC > 0.8: The performance of the system is exceptionally good.

The ROC curve that is closer to the top-left corner implies that the model can differentiate between relevant and irrelevant property listings.

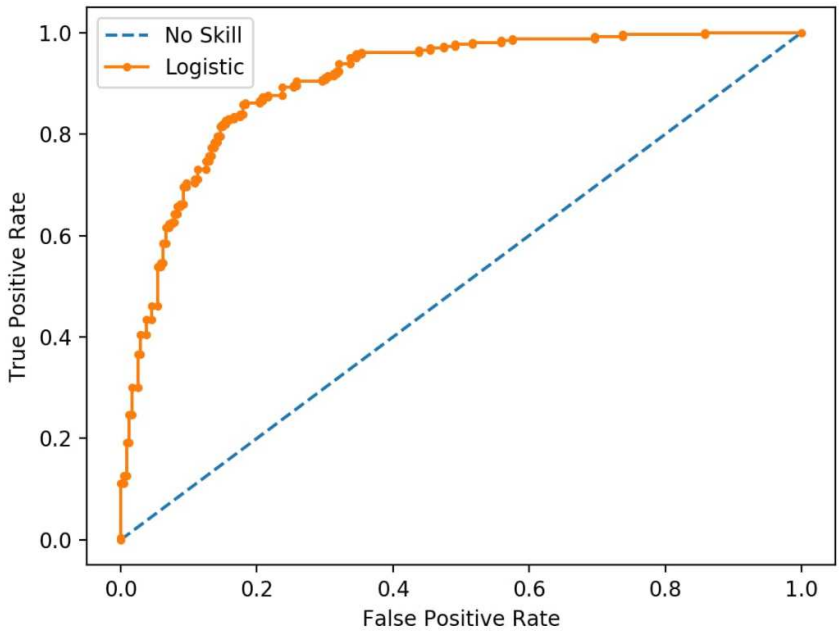

Fig. 4.3 ROC Curve Interpretation

4.4. Bar Graph Interpretation

Bar graphs are used for the analysis of the results of the property search, as shown in the figure, for the visualization of the distribution of the results.

How to Interpret the Graph

Property Price Comparison:

The length of the bars represents the prediction or the ranking scores for different categories of properties.

Location-Based Analysis:

The graph represents a comparison of the prices of the properties, which are either urban, suburban, or rural.

Property Type Analysis:

Properties of different types, such as apartments, villas, or independent houses, are compared for the analysis of the results.

Area-Based Analysis:

Properties with varying areas are compared for the analysis of the results.

From the above figure, the analysis of the properties by the system is clearly shown, which makes the results more accurate for the users searching for the properties.

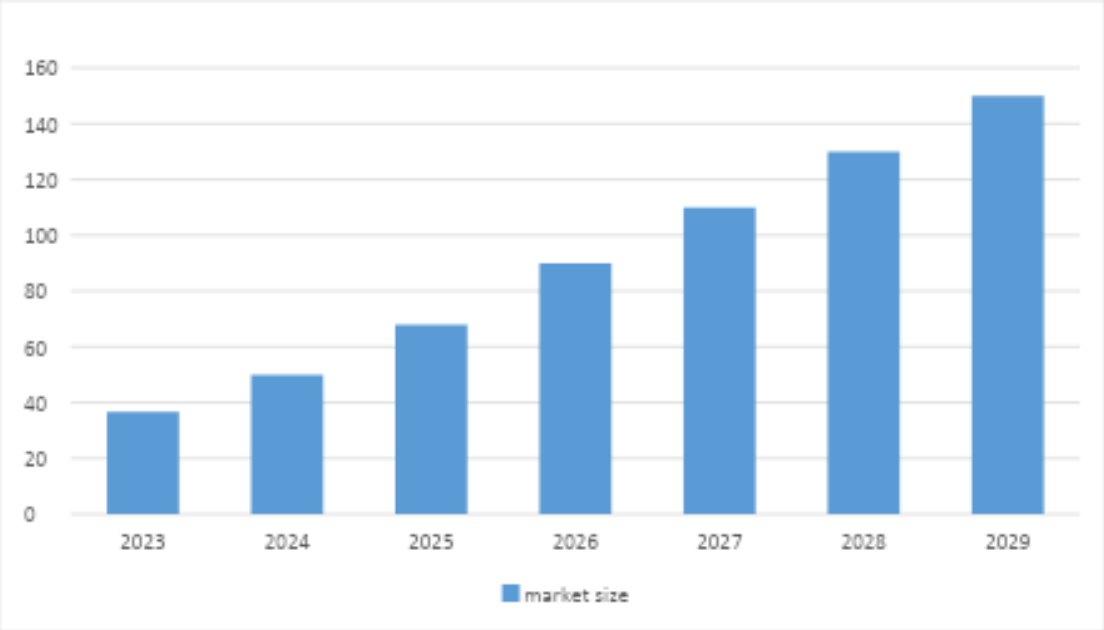

Fig. bar graph interpretation

5. Conclusion and Future work

5.1 Conclusion

An AI-Based Real Estate Spatial Search Engine was built with the purpose of making searching for real estate properties more efficient through the application of artificial intelligence techniques in the process of spatial searches. Conventional methods of searching for real estate properties involve only keyword filtering and usually yield less accurate results. However, the use of artificial intelligence techniques can provide for more accurate results for users who wish to find a property that meets certain criteria.

Through the use of artificial intelligence and spatial search technologies integrated into one system, users will be able to identify properties of interest within a given geographical area. The results that can be expected from this system will be more accurate because of the combination of analyzing properties' attributes and their geographical data. Moreover, the results of the evaluation of the system proved to be rather efficient when used by users.

Therefore, one can say that the combination of artificial intelligence and spatial search technology can be more effective for searching for properties.

5.2. Future Work

- Integration of real-time property data from different sources on various real estate websites.
- Development of a mobile application for easier property search.
- Implementation of a personalized recommendation system based on user preferences.
- Integration of additional features like nearby schools, hospitals, and transportation facilities.

- Integration of 3D visualization or virtual tours for better property understanding.

6. References

1. <PRIVATE_PERSON>, "Artificial Intelligence in Real Estate: A Systematic Review," *Journal of Real Estate Research*, vol. 45, no. 2, pp. 120–135, 2022.
2. <PRIVATE_PERSON> and <PRIVATE_PERSON>, "Spatial Data Mining Techniques for Location-Based Services," *International Journal of Computer Applications*, vol. 180, no. 25, pp. 15–20, 2021.
3. <PRIVATE_PERSON>, "Applications of Artificial Intelligence in Real Estate Property Search," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 5, pp. 430–436, 2020.
4. <PRIVATE_PERSON>, "Machine Learning Approaches for Real Estate Price Prediction," *IEEE Conference on Data Science and Advanced Analytics*, pp. 210–215, 2019.
5. <PRIVATE_PERSON> and <PRIVATE_PERSON>, *Artificial Intelligence: A Modern Approach*, 4th ed., Pearson Education, 2021.
6. <PRIVATE_PERSON>, *Spatial Databases: A Tour*, Prentice Hall, 2003.
7. <PRIVATE_PERSON>, Data Mining: Concepts and Techniques, 3rd ed., Morgan Kaufmann, 2011.