

HOUSE PRICE PREDICTION

A Course Project report submitted

In partial fulfillment of requirement for the award of degree

### BACHELOR OF TECHNOLOGY

in

### SPECIALIZATION

By

**SYED. UZMA YASMIN 2203A52055**

**YENAGANDULA.RAMYA 2203A52064**

**KOTHAPALLY.PAVITHRA 2203A52095**

Under the guidance of

### Dr. Soumik Podder

Assistant Professor, School of CS&AI.



**SR University**

Anantha Sagar, Warangal, Telagnana-506371



## CERTIFICATE

This is to certify that this project entitled “HOUSE PRICE PREDICTION”

is the bonafide work carried out by **SYED.UZMAYASMIN, YENAGANDULA.RAMYA,** **KOTHAPALLY.PAVITHRA,** as a Course Project for the partial fulfillment to award the degree **BACHELOR OF TECHNOLOGY** in **School of Computer Science and Artificial** **Intelligence** during the academic year **2023-2024** under our guidance and Supervision.

### Dr. Soumik Podder Prof. Sheshikala Martha

Assistant Professor, Professor & Head,

School of CS&AI, School of CS&AI,

SR University SR University

Anathasagar, Warangal Ananthasagar, Warangal.

### Reviewer-1 Reviewer-2

Name: Name:

Designation: Designation:

Signature: Signature:

# ACKNOWLEDGEMENT

We owe an enormous debt of gratitude to our Course project guide DR***.* SOUMIK PODDER, Assistant Professor** as well as Head of the School of CS&AI, **Prof. Sheshikala Martha,** for guiding us from the beginning through the end of the Course Project with their intellectual advices and insightful suggestions. We truly value their consistent feedback on our progress, which was always constructive and encouraging and ultimately drove us to the right direction.

We express our thanks to AI & ML course coordinator, Dr. Arpita Baronia, for her encouragement and support.

Finally, we express our gratitude and sincere thanks to all the teaching and non-teaching staff of the School of Computer Science & Artificial Intelligence, for their suggestions and timely support.

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| **S.no** | **Content** | **Page no** |
| 1 | Abstract | 4 |
| 2 | About the Organization | 5 |
| 3 | Introduction | 6 |
| 8 | Proposed Solution | 12 |
| 9 | System Architecture | 14 |
| 10 | Flow chart | 16 |
| 11 | Data flow diagram | 17 |
| 12 | Simulation Setup | 18 |
| 13 | Implementation | 20 |
| 14 | Result and Analysis | 21 |
| 15 | Learning Outcome | 23 |
| 16 | Conclusion with Challenges | 25 |
| 17 | Future Scope | 27 |
| 18 | References | 29 |
| 19 | Links of the projects | 30 |

**List of Figures**

|  |  |  |
| --- | --- | --- |
| **Fig no** | **Figure Title** | **Page no** |
| 1. | System architecture | 15 |
| 2. | Flowchart | 17 |
| 3. | Data flow Diagram | 18 |

**ABSTRACT**

## People look for a new house to be cozy and more conservative with their budgets and market price and strategies.

## This project aims to analyze many parameters like average income, analyze average area etc.

## This will help to predict the house price accordingly.

## To provide the best house without being cheated.

## To provide the best price without any problem causing for the customers.

## To enable and provide all the all needs for a customer before buying a new house.

## This will provide the information for buying a house.

## To enable the customer to search home as per the budget.

## This data set has the all information that would be helpful for a customer before buying a house.

## In this project the data set has the content like area, bedrooms, main road

## Guest room basement hot water heating, air conditioning, parking, preferred area.

## This will provide the accurate information to a customer before buying a house.

## We use linear regression, logistic regression and SVM for predicting the house and the price.

# ABOUT THE ORGANISATION

SR University is a private university located in Warangal, Telangana, India. It was established in 2018 under the Telangana State Private Universities (Establishment and Regulations) Act 2018. SR University is accredited with an 'A' grade by the National Assessment and Accreditation Council (NAAC).

SR University offers a variety of undergraduate and postgraduate programs in engineering, technology, management, commerce, and arts. The university has a strong focus on industry- relevant education and offers a variety of opportunities for students to gain hands-on experience through internships, projects, and workshops. SR University also has a strong incubation center that supports students in developing and launching their startups.

SR University has a well-equipped campus with state-of-the-art facilities, including classrooms, laboratories, libraries, sports facilities, and hostels. The university also has a strong commitment to research and has published several papers in reputed journals and conferences.

SR University has a good placement record. In 2023, the university achieved 90% placements for its engineering students. The university has a strong alumni network that includes several successful entrepreneurs and professionals.

Overall, SR University is a good choice for students who are looking for an industry-relevant education and a strong focus on innovation and entrepreneurship.

# INTRODUCTION

Machine learning is a subfield of artificial intelligence (AI) that uses algorithms and techniques to extract useful information from data. Machine learning **is** designed for big data because trying to process **big** data manually is impossible without the support of machines. In computer science, machine learning attempts to solve problems through algorithms rather than mathematics. **But** there are two types of machine learning: supervised learning and unsupervised learning. It is an unsupervised program that tries to find relationships and hidden patterns in data. Many types of machine learning algorithms are used to solve problems today. **But** as noted in **"There's** no **such** **thing** **as** **a** free lunch ,some may **be** better in **some** **cases.** Therefore, this article attempts to **compare** **the** **performances** **of** regression algorithms and artificial neural networks **(ANN)** in predicting **the** values **​​given** **to** **the** data. Performance **for** **house** **price** **prediction** will be evaluated **because** **in** **multivariate** **algorithms** prediction is based not only on specific features but also on **the** **behavior** of **unknowns,** **many** **of** **which** lead to price **predictions.** The price of **a** house depends on the **features** of **each** **house.** **The** **number** **of** **houses** **varies** **and** the **price** may **vary** depending on location. For example, if **the** house is located in a less desirable, affluent area, it will cost more than if it is **located** in a poor area. The data used in the experiment will be **subjected** **to** a series of **preprocessing** to **increase** prediction accuracy. **In** **addition,** some factors will be added to local data **and** **how** these factors **affect** sales **prices** **in** **Malmö** **will** **be** **examined.**

# PROBLEM STATEMENT

# Develop a machine learning model to accurately predict residential house prices based on features like location, size, bedrooms, and amenities. The model aims to assist buyers, sellers, and real estate agents in making informed decisions. Key objectives include data preprocessing, feature selection, model development, evaluation, and deployment for real-time predictions, with success measured by predictive accuracy and interpretability. Deliverables include a well-documented model, deployment in a production environment, and ongoing monitoring for performance maintenance.

# 

# REQUIREMENT ANALYSIS

Requirement analysis for a house price prediction system involves understanding the goals of the system, the data available for analysis, the features to be considered, and the methodology for predicting prices.

**1.Goal Definition:**

Define the purpose of the prediction (e.g., for real estate investment, pricing strategy, etc.).

Determine the scope of the analysis (e.g., geographical area, types of properties).

**2.Data Collection:**

Identify sources of data (e.g., real estate websites, public records).

Gather relevant data, including property features (size, location, amenities, etc.) and historical prices.

**3.Data Preprocessing:**

Clean the data (handle missing values, outliers, etc.).

Transform the data (normalize, encode categorical variables, etc.).

**4.Feature Selection:**

Identify relevant features that affect house prices (e.g., location, size, number of bedrooms/bathrooms, etc.).

Select features based on importance (e.g., using feature importance techniques).

**5.Model Selection:**

Choose a suitable machine learning model for prediction (e.g., linear regression, decision tree, neural network).

Consider factors like interpretability, performance, and scalability.

**6.Model Training:**

Split the data into training and validation sets.

Train the model using the training set.

**7.Model Evaluation:**

Evaluate the model using the validation set (e.g., using metrics like RMSE, MAE).

Fine-tune the model if necessary (e.g., hyperparameter tuning).

**8.Deployment:**

Integrate the model into a system (e.g., a web application, API).

Ensure scalability, reliability, and security.

**9.Monitoring and Maintenance:**

Monitor the model's performance over time.

Update the model periodically with new data.

**10.Documentation:**

Document the entire process, including data sources, preprocessing steps, model selection, and evaluation results.

# RISK ANALYSIS

### 1.Data Quality:

### Risk: Poor quality or insufficient data could lead to inaccurate predictions.

### Mitigation: Perform thorough data cleaning and validation. Use feature engineering to improve data quality.

### 2.Overfitting:

### Risk: The model may perform well on training data but poorly on new data.

### Mitigation: Use techniques like cross-validation, regularization, and feature selection to prevent overfitting.

### 3.Underfitting:

### Risk: The model is too simple to capture the underlying patterns in the data.

### Mitigation: Use more complex models or feature engineering to improve model performance.

### 4.Model Complexity:

### Risk: Overly complex models may be difficult to interpret and maintain.

### Mitigation: Balance model complexity with performance. Use simpler models when possible.

### 5.Feature Selection:

### Risk: Choosing irrelevant or redundant features can impact model performance.

### Mitigation: Use feature importance techniques and domain knowledge to select relevant features.

### 6.Data Leakage:

### Risk: Information from the validation or test set inadvertently leaks into the training set, leading to overly optimistic results.

### Mitigation: Split the data into training, validation, and test sets carefully. Use pipelines to ensure proper data handling.

### 7.Model Evaluation:

### Risk: Improper evaluation metrics or validation methods can lead to misleading results.

### Mitigation: Use appropriate evaluation metrics (e.g., RMSE, MAE) and validation techniques (e.g., cross-validation) for robust evaluation.

### 8.Deployment and Maintenance:

### Risk: Challenges in deploying the model into production or maintaining it over time.

### Mitigation: Involve DevOps and IT teams early in the process. Plan for regular updates and monitoring.

### 9.Ethical and Legal Issues:

### Risk: Biased predictions or misuse of data could lead to ethical and legal consequences.

### Mitigation: Ensure fairness and transparency in the model. Comply with data protection regulations (e.g., GDPR).

### 10.External Factors:

### Risk: Economic changes, regulatory changes, or other external factors could impact the housing market and affect model performance.

### Mitigation: Monitor external factors and update the model as needed to adapt to changes.

# PROPOSED SOLUTION

**1.Data Collection and Preparation**:

Gather data from sources like real estate websites, public records, and APIs.

Clean the data, handle missing values, and transform features as needed.

**2.Feature Engineering:**

Identify relevant features that could affect house prices (e.g., location, size, amenities).

Create new features or transformations to enhance the model's predictive power.

**3.Data Splitting:**

Split the data into training, validation, and test sets.

**4.Model Selection:**

Choose a regression model suitable for house price prediction (e.g., linear regression, decision tree regression, random forest regression).

Consider using ensemble methods for better performance.

**5.Model Training:**

Train the model using the training set.

Tune hyperparameters using cross-validation to improve performance.

**6.Model Evaluation:**

Evaluate the model using the validation set.

Use metrics like RMSE, MAE, and R-squared to assess performance.

**7.Model Deployment:**

Once satisfied with the model's performance, deploy it into a production environment.

Create an API or web interface for users to interact with the model.

**8.Monitoring and Maintenance**:

Monitor the model's performance in production.

Update the model periodically with new data and retrain if necessary.

**9.Documentation:**

Document the entire process, including data sources, preprocessing steps, model selection, and evaluation results.

Include instructions for deploying and maintaining the model.

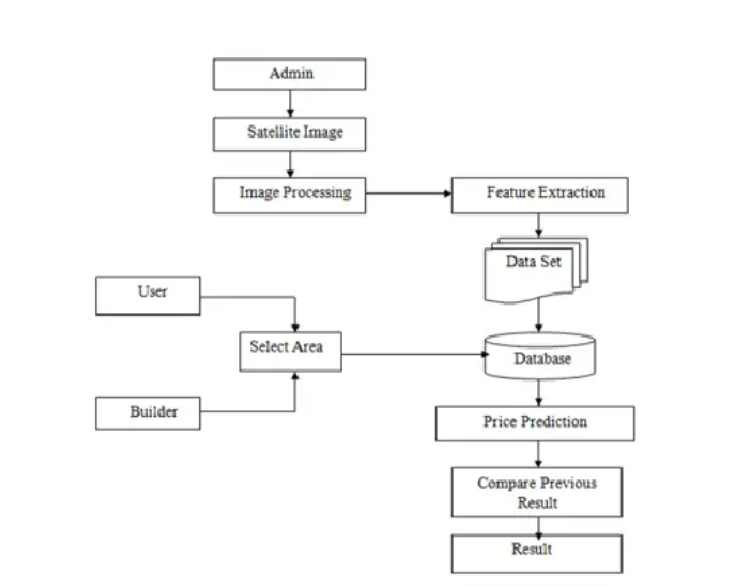
**10.Ethical Considerations:**

Ensure the model is fair and unbiased.

Handle sensitive data (e.g., personal information) according to data protection regulations.

By following these steps, you can build a robust house price prediction system using machine learning that provides accurate and reliable predictions.

# SYSTEM ARCHITECTURE



# A. Image Processing: - In order to get an enhanced image or to extract some useful information from it, some operations are required to perform on image. Image processing is the technique to extract useful information from image. In an image processing, input is an image and output may be image or characteristics or features associated with that image. We are using snapshots of satellite images to gain text-based information from image of particular areas.

# B. Feature Extraction: - In order to extract some significant characteristics of the object feature extraction is used. Feature is the function of measuring measurable property of an object. Feature extraction is related to dimensionality reduction. All the features of image can be classified into high-level feature and low-level feature extraction. Feature extraction may include color feature, shape feature and texture feature. Due to powerful parallel mechanism of computation neural networks are mainly used for feature extraction. In our system, we are using convolutional neural network for feature extraction

# C. Price Dataset: - In Machine Learning, the training data set is the actual dataset used to train the model for predicting house price.

# FLOWCHART

# A diagram of a data processing process Description automatically generated

# 

# DATA FLOW DIAGRAM

A diagram of a real estate

Description automatically generated

# SIMULATION SETUP

Setting up a simulation for house price prediction using machine learning involves creating a controlled environment to train and evaluate your model. Here's a general outline of how you can set up the simulation:

**1.Data Generation:**

Generate synthetic data that mimics the characteristics of real-world housing data.

Include features such as house size, number of bedrooms/bathrooms, location, and amenities.

Introduce noise to simulate real-world variability.

**2.Data Splitting:**

Split the generated data into training, validation, and test sets.

**3.Model Training:**

Select a machine learning model (e.g., linear regression, decision tree, random forest) for training.

Train the model using the training set.

**4.Hyperparameter Tuning:**

Perform hyperparameter tuning using the validation set to optimize the model's performance.

**5.Model Evaluation:**

Evaluate the model using the test set.

Use metrics like RMSE, MAE, and R-squared to assess the model's performance.

**6.Analysis:**

Analyze the results to understand how well the model predicts house prices.

Identify areas for improvement and potential biases.

**7.Repeat and Iterate:**

Repeat the simulation with different configurations (e.g., different models, feature sets) to compare performance.

Iterate on the model and data generation process to improve accuracy.

**8.Documentation:**

Document the simulation setup, including data generation methods, model selection, and evaluation results.

Include any insights or lessons learned from the simulation.

By setting up a simulation, you can experiment with different models and parameters in a controlled environment before applying them to real-world data, helping you build a more effective house price prediction model.

**IMPLEMENTATION**

House price prediction is a classic machine learning problem. To implement it, you can follow these general steps:

**Data Collection**: Obtain a dataset containing information about houses and their prices. This dataset should include features like the number of bedrooms, square footage, location, etc.

**Data Preprocessing**: Clean the data by handling missing values, encoding categorical variables, and scaling numerical features if necessary.

**Feature Selection/Engineering:** Select relevant features that might affect the house prices. You can also create new features based on existing ones if you think they could be valuable.

**Split the Data:** Divide the dataset into training and testing sets to evaluate the model's performance.

**Model Selection:** Choose a machine learning model to use for prediction. Common choices for regression tasks like house price prediction include linear regression, decision tree regression, and random forest regression.

**Model Training**: Train the selected model on the training dataset.

**Model Evaluation:** Evaluate the model's performance on the testing dataset using metrics like mean squared error (MSE), root mean squared error (RMSE), or R-squared.

**Hyperparameter Tuning:** Fine-tune the model by adjusting hyperparameters to improve its performance.

**Prediction:** Use the trained model to make predictions on new data.

# RESULT COMPARISON AND ANALYSIS

To compare and analyze the results of different machine learning models for house price prediction, you can follow these steps:

**1.Select Evaluation Metrics:** Choose appropriate metrics to evaluate the models, such as RMSE (Root Mean Squared Error), MAE (Mean Absolute Error), and R-squared (coefficient of determination).

**2.Evaluate Baseline Model:** Start by evaluating a simple baseline model (e.g., linear regression) to establish a benchmark for comparison.

Compare Different Models: Train and evaluate various machine learning models, such as linear regression, decision tree regression, random forest regression, and gradient boosting regression. Use the same evaluation metrics for consistency.

Hyperparameter Tuning: Perform hyperparameter tuning for each model using techniques like grid search or random search to optimize performance.

**3.Cross-Validation:** Use cross-validation to ensure the robustness of the results and mitigate overfitting.

**4.Result Analysis**: Analyze the results to identify the best-performing model based on the chosen evaluation metrics. Consider factors like model complexity, interpretability, and ease of deployment.

**5.Visualize Results**: Visualize the model performance using plots like actual vs. predicted prices to gain insights into the model's strengths and weaknesses.

**6.Statistical Significance:** Determine if the differences in performance between models are statistically significant using appropriate statistical tests.

Sensitivity Analysis: Perform sensitivity analysis to understand how changes in input variables affect the model's predictions.

**7.Final Model Selection:** Select the best-performing model as the final model for house price prediction based on the results of the comparison and analysis.

By following these steps, you can effectively compare and analyze the results of different machine learning models for house price prediction and select the most suitable model for your application.

# LEARNING OUTCOME

The learning outcomes for a house price prediction project using machine learning can include:

**1.Understanding of Machine Learning Concepts:** Gain a solid understanding of machine learning concepts such as regression, feature engineering, model selection, and evaluation metrics.

**2.Data Preprocessing Skills**: Learn how to clean, preprocess, and transform data to make it suitable for machine learning models.

**3.Feature Selection and Engineering:** Develop skills in selecting relevant features and creating new features to improve model performance.

**4.Model Selection and Evaluation:** Learn how to choose the right machine learning model for a given problem and evaluate its performance using appropriate metrics.

**5.Hyperparameter Tuning:** Gain experience in optimizing model performance by tuning hyperparameters using techniques like grid search or random search.

**6.Cross-Validation Techniques:** Understand the importance of cross-validation in assessing model performance and preventing overfitting.

Deployment Skills: Learn how to deploy a machine learning model into a production environment, including considerations for scalability and reliability.

**7.Data Visualization and Interpretation:** Develop skills in visualizing data and model results to gain insights and make informed decisions.

**8.Project Management Skills:** Learn how to manage a machine learning project from data collection to model deployment, including documentation and collaboration with stakeholders.

**9.Ethical Considerations**: Understand the ethical implications of machine learning, including bias and fairness, and learn how to address these issues in a real-world application.

Overall, a house price prediction project using machine learning can provide valuable hands-on experience and a deep understanding of various aspects of machine learning, which are applicable to a wide range of other machine learning projects

# CONCLUSION WITH CHALLENGES

**Conclusion:**

House price prediction using machine learning is a complex but rewarding task. By leveraging machine learning models, we can predict house prices with reasonable accuracy, providing valuable insights for buyers, sellers, and real estate professionals. Through careful data collection, preprocessing, model selection, and evaluation, we can develop robust models that capture the underlying patterns in housing market data.

**Challenges:**

**1.Data Quality:** Ensuring the quality of data is crucial, as inaccurate or incomplete data can lead to biased or inaccurate predictions.

**2.Feature Selection**: Identifying the most relevant features that affect house prices can be challenging, especially when dealing with a large number of potential features.

**3.Model Complexity:** Balancing model complexity with interpretability is a challenge, as more complex models may provide better accuracy but are harder to interpret and deploy.

**4.Overfitting and Underfitting:** Avoiding overfitting and underfitting is critical, as these can lead to poor generalization and inaccurate predictions.

**5.Deployment:** Deploying machine learning models into production systems can be challenging, requiring integration with existing infrastructure and ensuring scalability and reliability.

**6.Ethical Considerations**: Ensuring that the model is fair and unbiased is essential, as biased predictions can have negative consequences, such as perpetuating inequalities in housing markets.

Addressing these challenges requires a multidisciplinary approach, involving expertise in machine learning, data analysis, real estate, and ethics. By overcoming these challenges, we can develop more accurate and reliable models for house price prediction, benefiting both individuals and the real estate industry as a whole.

# FUTURE SCOPE

The future scope of house price prediction using machine learning is promising, with several potential advancements and applications:

**1.Improved Accuracy:** Continued research in machine learning algorithms and techniques will likely lead to models with higher accuracy in predicting house prices.

**2.Feature Engineering:** Further advancements in feature engineering, including the use of advanced data transformation techniques and the incorporation of new types of data (such as satellite imagery or social media data), could enhance model performance.

**3.Incorporating External Factors:** Integrating external factors such as economic indicators, neighborhood developments, and environmental factors could provide more holistic and accurate predictions.

**4.Deep Learning**: The use of deep learning models, such as neural networks, could lead to better feature representation and potentially higher prediction accuracy.

**5.Interpretability**: Improving the interpretability of machine learning models could increase their trustworthiness and adoption in real estate decision-making processes.

**6.Real-Time Prediction**: Developing models capable of real-time prediction could be valuable for stakeholders looking to make quick decisions in dynamic real estate markets.

**7.Personalized Predictions**: Tailoring predictions based on individual preferences and needs could provide more personalized and relevant insights for buyers, sellers, and investors.

**8.Market Trend Analysis**: Analyzing historical data to identify trends and predict future market conditions could help stakeholders make informed decisions about buying, selling, or investing in real estate.

**9.Automated Valuation Models (AVMs):** AVMs, powered by machine learning, could become more sophisticated and widely adopted for property valuation purposes.

**10.Ethical and Fairness Considerations**: Continued research into ethical and fairness considerations in machine learning models for house price prediction will be essential to ensure that these models do not perpetuate biases or discrimination.

Overall, the future of house price prediction using machine learning is likely to be characterized by increased accuracy, sophistication, and relevance to real-world applications, making it an exciting area for further research and development.

# REFERENCES

# Kaggle - House Price Prediction

# Git hub- House Price Prediction

# LINKS OF THE PROJECTS

1. Kaggle Link: https://www.kaggle.com/code/cematilgan05/house-price-prediction
2. GitHub Link: <https://github.com/topics/housing-price-prediction>