USA House Price Prediction Project

Forecasting trends in American real estate market

Executive Summary

Project Overview and Performance

Project Objective

Predict house prices in the USA using machine learning to aid Real Estate Investment Trust decisions.

Model Development

RandomForestRegressor was chosen after testing multiple regression models for best predictive accuracy.

Hyperparameter Tuning

GridSearchCV optimized model parameters to max_depth 30 and n_estimators 300 for improved performance.

Deployment

Model deployed as interactive web app using Streamlit for real-time house price predictions.

Introduction

Business Problem and Project Goal

Business Problem Overview

The project addresses the need for precise house price predictions to aid real estate investment decisions.

Project Goal

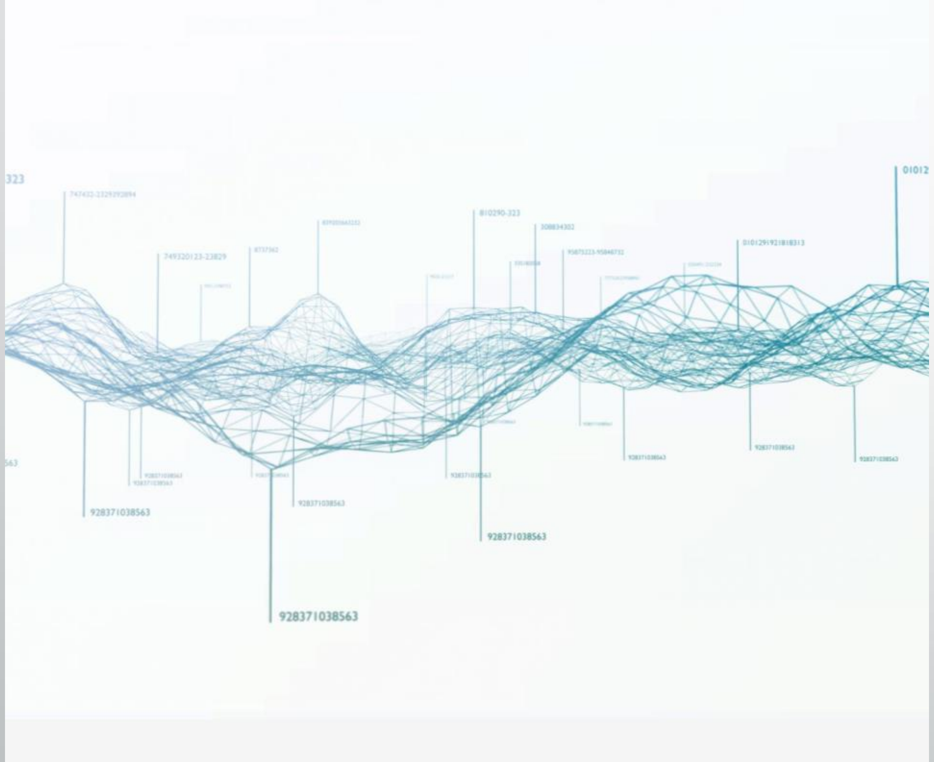
Aim to determine market house prices based on features to improve financial planning and investment.

Methodology

Utilize historical sales data and machine learning models to identify key factors influencing house prices.

Methodology

Data Processing and Modeling Approach



Data Cleaning and Wrangling

Irrelevant columns were dropped, missing values replaced with column means, and data types corrected for accuracy.

Exploratory Data Analysis

EDA was performed using Pandas, Seaborn, and Matplotlib to visualize feature distributions and correlations.

Regression Model Development

Multiple regression models including Linear, Ridge, Lasso, Decision Tree, Random Forest, and XGBoost were developed.

Model Optimization and Selection

GridSearchCV was used for hyperparameter tuning, and RandomForestRegressor was selected for its robustness and high R^2 score.

EDA Insights

Key Features Influencing House Prices

Influential Features

Square footage and overall grade are the most influential factors affecting house prices.

Waterfront Premium

Homes with waterfront views are significantly more expensive than others without such views.

Condition Impact

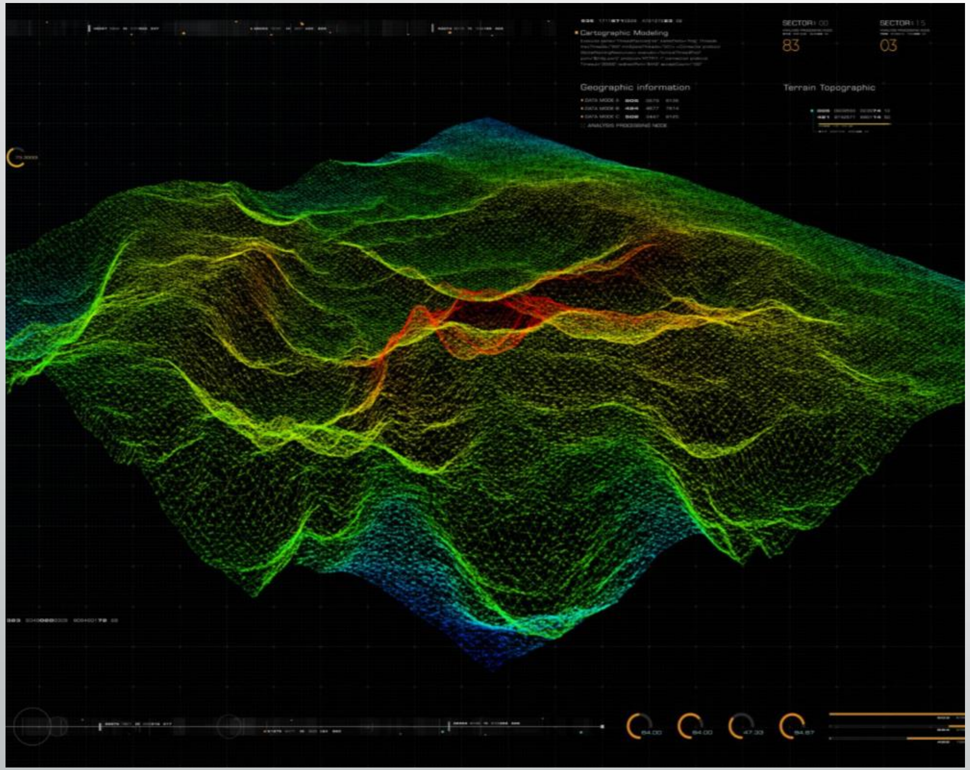
Condition of the house has less impact on price than initially expected.

Additional Square Footage

Extra square footage excluding basement positively correlates with higher house prices.

Model Development

Selected Model and Hyperparameter Tuning



Model Selection

RandomForestRegressor was chosen due to its superior performance on the regression task.

Hyperparameter Tuning

GridSearchCV identified optimal parameters: max_depth 30 and n_estimators 300 for best accuracy.

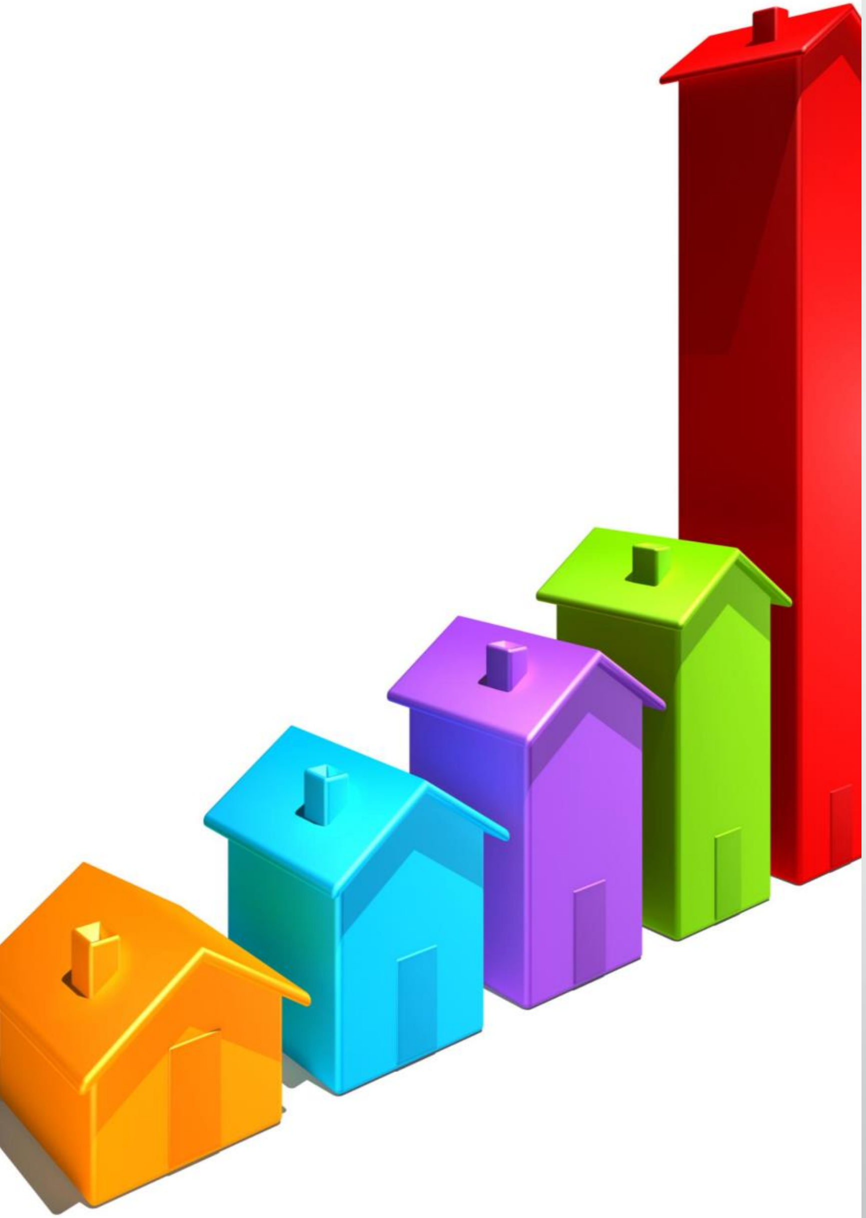
Model Performance

Tuned model achieved R^2 score of 0.78, demonstrating improved prediction accuracy and robustness.

Training and Testing Split

Model trained on 80% of data and tested on 20%, ensuring reliable evaluation of predictions.

Model Comparison



Performance of Regression Models

Baseline Regression Models

Linear, Ridge, and Lasso regression models provided baseline predictive performance for house prices.

Improved Model Performance

Decision Tree and XGBoost models showed moderate improvements over baseline regression models.

Top Performing Model

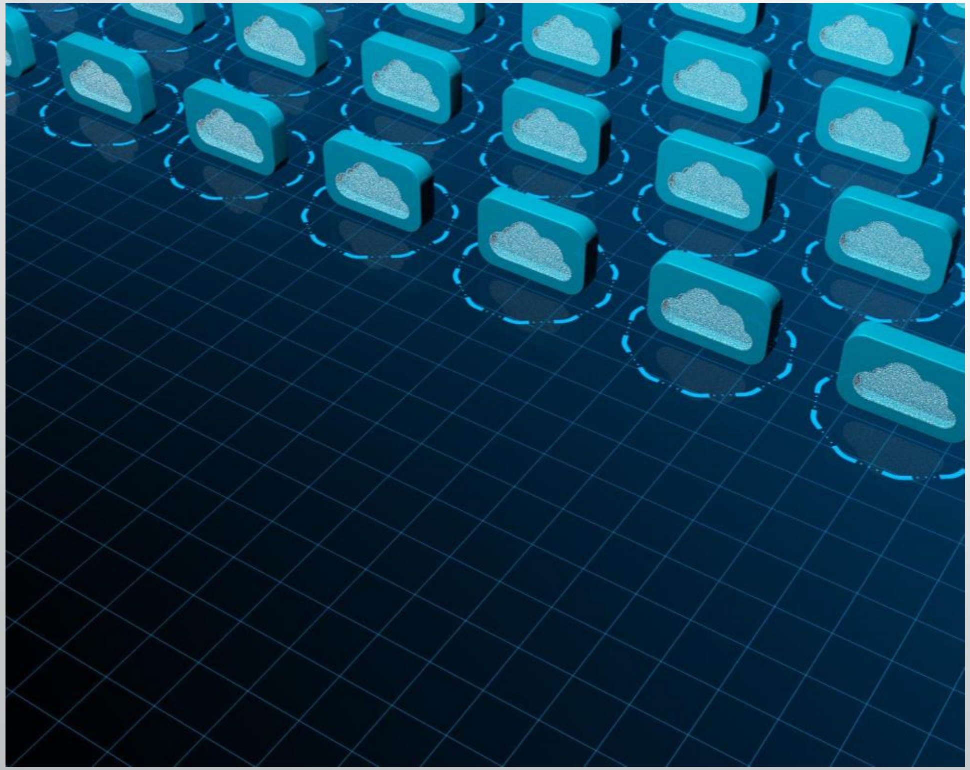
Random Forest achieved the highest R^2 score of 0.78, outperforming all other tested models.

Model Selection Importance

Selecting and tuning the right model is crucial for accurate house price prediction.

Deployment

Web Application and Hosting



Interactive Web Application

The model was deployed as a Streamlit web app allowing users to input data and get real-time predictions.

Model Storage and Access

The machine learning model is stored on Google Drive and accessed using Python libraries like requests and joblib.

Code Management and Hosting

Application code is maintained in GitHub and hosted publicly on Streamlit Cloud for accessibility and collaboration.

Stakeholder Accessibility

Hosting ensures stakeholders can easily access the app to make informed investment decisions using model outputs.

Key Findings

Insights for Real Estate Investment

Key Price Predictors

Square footage and housing grade are the strongest factors influencing property prices for investors.

Impact of Waterfront Views

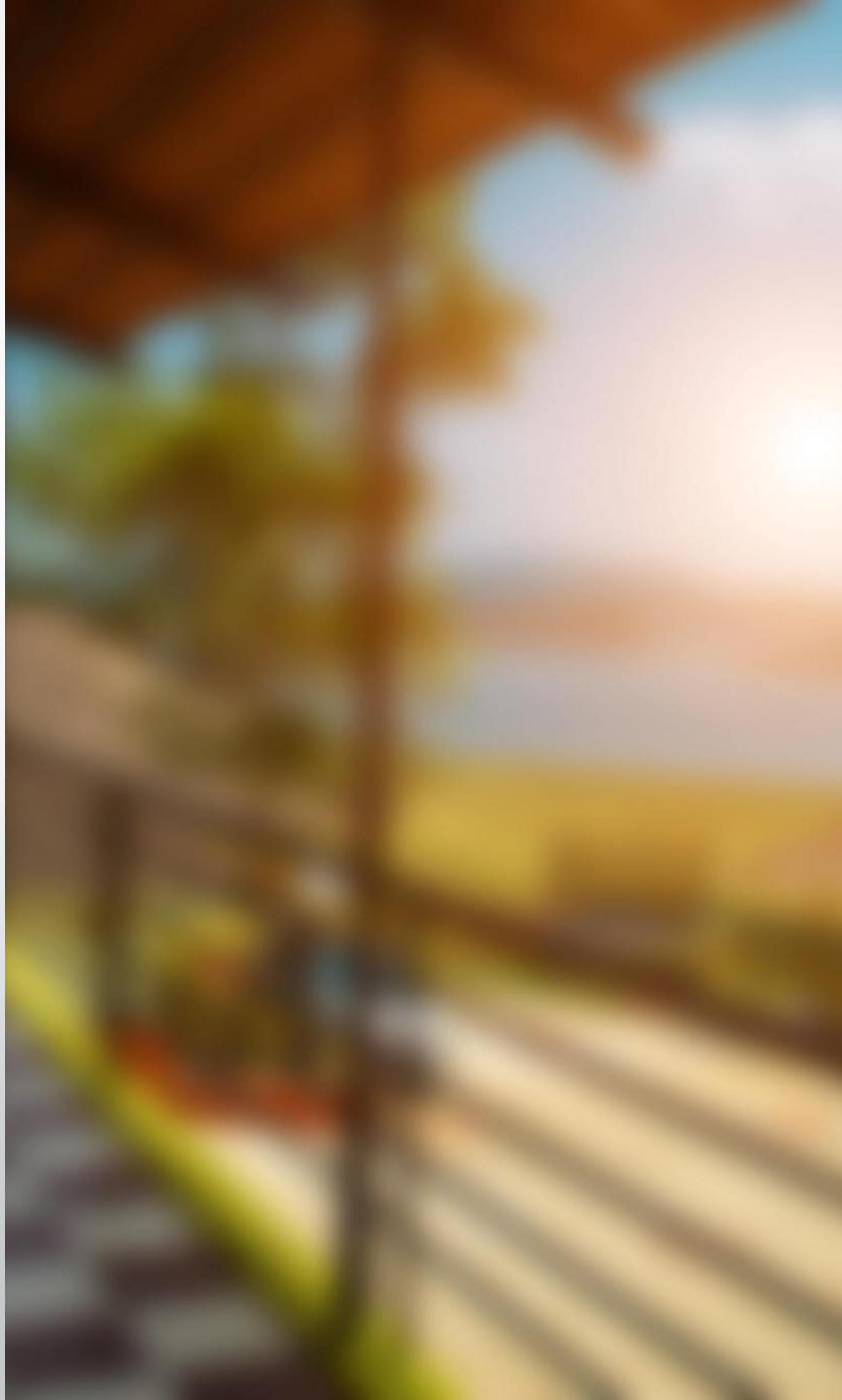
Properties with waterfront views command higher values, significantly boosting investment potential.

Minimal Effect of House Condition

Surprisingly, the house's condition has little impact on price according to the investment model.

Model for Investment Decisions

The analytical model aids investors in estimating prices and identifying valuable real estate opportunities.



Conclusion

Project Summary and Future Scope

Model Development and Performance

A RandomForestRegressor model was developed with an R^2 score of 0.78, offering reliable house price predictions.

Insights for Investment

Exploratory data analysis and model evaluation provided insights to support strategic investment decisions in housing market.

Future Enhancements

Future scope includes adding features, exploring advanced models like deep learning, and integrating real-time dynamic data.

Appendix

Technical Details and Tools Used

Dataset Overview

House sales data from USA between May 2014 and May 2015 includes features like square footage and waterfront view.

Data Preprocessing Steps

Data preprocessing involved managing missing values, encoding categories, and normalizing numerical features.

Machine Learning Models

Applied models include Linear Regression, Ridge, Lasso, Decision Tree, Random Forest, and XGBoost for prediction.

Tools and Evaluation Metrics

Used Python libraries, visualization tools, and evaluation metrics such as R^2 score and mean squared error.