Property Price Prediction Using Machine Learning

This presentation explores property price prediction through machine learning techniques, focusing on Linear Regression and Random Forest models. This work aims to provide data scientists and machine learning engineers with insights into practical modeling approaches for real estate valuation.

We will examine the methodology, datasets, model performances, and interpretations to understand how these algorithms can predict property prices accurately, combining statistical fundamentals with advanced non-linear models.

Members

Naman Jain

SV Prabhas

Anurag Pratap Singh

Shankar Singh

Property Price Prediction

Introduction to the Problem and Dataset

Challenge Overview

Predicting property prices is critical for buyers, sellers, and investors to make informed decisions.

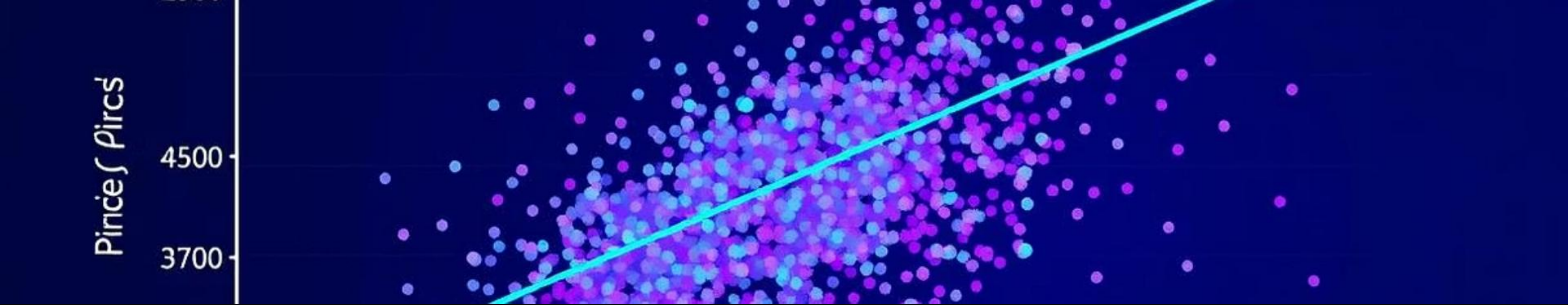
Data Description

Dataset includes property features like size, location, number of bedrooms, age, and amenities, alongside historical prices.

Data Preprocessing

Handling missing values, encoding categorical features, and scaling numerical attributes to prepare the data for modeling.

		Price	Property Featurg Fowtura	Square Feature	Location	Price
o sequared fetach	algition	29.00	$2,000	$81,000	$2,000	$2,0
uarer of bedrooom	alaition	1,0.00	$8,000	$81,000	$8,000	12,0
mber of recotemding	aloition	58.00	$2,000	$51,000	$2,000	$2,0
ratural	algiton	69.00	$6,000	$65,000	$9,000	$7,0
merrd of ter stadlow	alaition	13200	$8,000	$51,000	$0,000	15,0
aress Baty Sedelscal	alaition	$0.00	$6,000	$51,000	$3,000	17,0
are's derrplly danty	alution	19,04	$6,000	$72000	$6,000	$2,0



Linear Regression Model: Approach and Insights

Fundamental Concept

Linear Regression models the linear relationship between features and property price using a weighted sum.

Interpretability

Provides coefficients that quantify feature impacts, offering interpretable insights into price drivers.

Limitations

Assumes linearity and may underperform when relationships are complex or non-linear.

Random Forest Model: Approach and Benefits

Model Overview

A Random Forest ensembles multiple decision trees to capture complex patterns beyond linearity.

Advantages

- Handles non-linear feature interactions
- Robust against overfitting via bootstrapped aggregation
- Automatic feature importance estimation

Model Evaluation Metrics

Mean Absolute Error (MAE)

Measures average absolute difference between predicted and actual prices.

Root Mean Squared Error (RMSE)

Highlights larger errors by squaring differences before averaging and rooting.

R-Squared

Indicates proportion of variance explained by the model, higher is better.

Comparing Linear Regression and Random Forest

Linear Regression

- Simple and fast to train
- Good interpretability
- Limited to linear relationships

Random Forest

- Captures complex patterns
- More computationally intensive
- Less transparent but more accurate

Key Challenges and Practical Considerations

Data Quality

Inconsistent or incomplete data can bias models, necessitating careful cleaning and validation.

Feature Engineering

Creating meaningful features such as location scores or neighborhood crime rates improves predictions.

Model Interpretability

Balancing complexity and transparency is vital for trust and actionable insights in real estate markets.

Summary and Future Directions

1

2

3

Model Deployment

Integrating predictive models into real-time platforms for dynamic price estimation.

Enhanced Data

Incorporating additional data such as economic indicators and satellite imagery.

Advanced Techniques

Exploring deep learning and hybrid models to capture more subtle data patterns.

Property Price Prediction