Real Estate Analysis

Predicting House Prices with Supervised Learning

An engaging exploration of how supervised learning techniques can effectively forecast house prices, presented by Sujal Gupta, Shivam Chaudhary, and Shubhranshu, under the guidance of Mrs. Komal Salgotra.

Sujal Gupta, Shivam Chaudhary, Shubhranshu.

Predicting House Prices with Regression

A comprehensive approach to modeling house prices

- **Project Overview**

  Develop a predictive model for house prices using linear regression based on various property features.

- **Dataset Features**

  Includes crucial features like bedrooms, bathrooms, living area, and more that influence house pricing.

- **Data Preparation**

  Loading and cleaning data, handling missing values by using column means for numerical columns.

- **Feature Selection**

  Choosing relevant features that significantly influence the price for effective model training.

- **Data Splitting**

  The dataset is split into training (80%) and testing (20%) sets to ensure model validation.

- **Training the Model**

  A Linear Regression model is trained on the training set to learn the relationship between features and price.

- **Making Predictions**

  After training, predictions are made on the test set to evaluate the model's performance.

- **Performance Evaluation**

  Model performance is assessed using RMSE, which penalizes larger errors in predictions.

- **Real-World Application**

  A sample prediction is showcased for a hypothetical house to demonstrate practical use.

Comprehensive Methodology Overview

Overview of Steps in House Price Prediction

Data Acquisition and Handling

Load dataset from CSV with error handling. Assumes residential property features with price as target.

Data Cleaning Techniques

Impute missing numerical values with column means to maintain data integrity and avoid bias.

Feature Selection Criteria

Select significant features like bedrooms and sqft to influence price predictions. Construct input matrix X.

Evaluation Metrics Used

Measure model performance via Root Mean Squared Error (RMSE) to evaluate prediction accuracy.

Training and Testing Splits

Split dataset into training (80%) and testing (20%) sets for model evaluation and generalization.

Model Training Process

Utilize a Linear Regression model to fit the relationship between selected features and house prices.

Visualization of Results

Use visualizations like scatter plots and histograms to interpret model outcomes and errors.

Predicting House Prices Using ML

A step-by-step guide to predicting house prices

Data Loading and Preprocessing

The dataset is loaded and missing values in numerical columns are replaced with their mean.

Feature Selection

Important features selected include bedrooms, bathrooms, and sqft_living to predict house prices.

Data Splitting

Data is split into training and testing sets (80% train, 20% test) to evaluate model performance.

Model Training

A Linear Regression model is trained using the training dataset to learn the relationship between features and price.

Price Prediction

The trained model predicts house prices based on provided features for new houses.

Model Evaluation

Root Mean Squared Error (RMSE) is calculated to assess the model's prediction accuracy.

In-Depth Visualization Insights

Exploring machine learning in real estate

- **Scatter Plots for Feature Analysis**

  Scatter plots reveal relationships between features and house prices, enhancing understanding of data patterns.

- **Actual vs. Predicted Prices**

  A comparison plot of actual versus predicted prices evaluates model performance and accuracy in predictions.

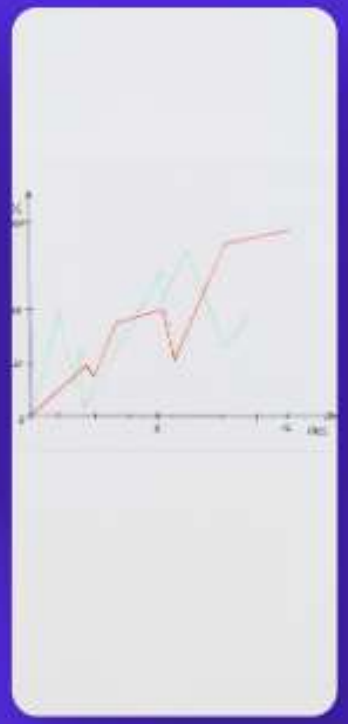
- **Residuals Distribution Analysis**

  Residuals distribution plot assesses prediction errors, ensuring randomness without unexplained patterns.

Evaluating Output and Performance

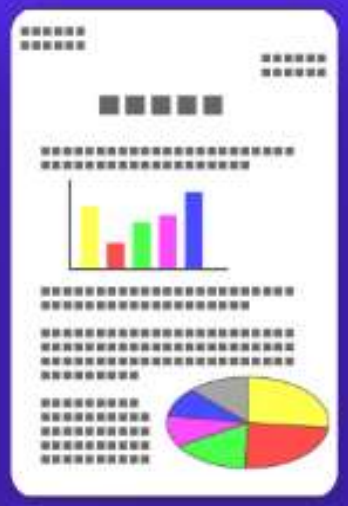
Feature vs Price Analysis

Visualizes relationships between features and house prices using scatter plots.



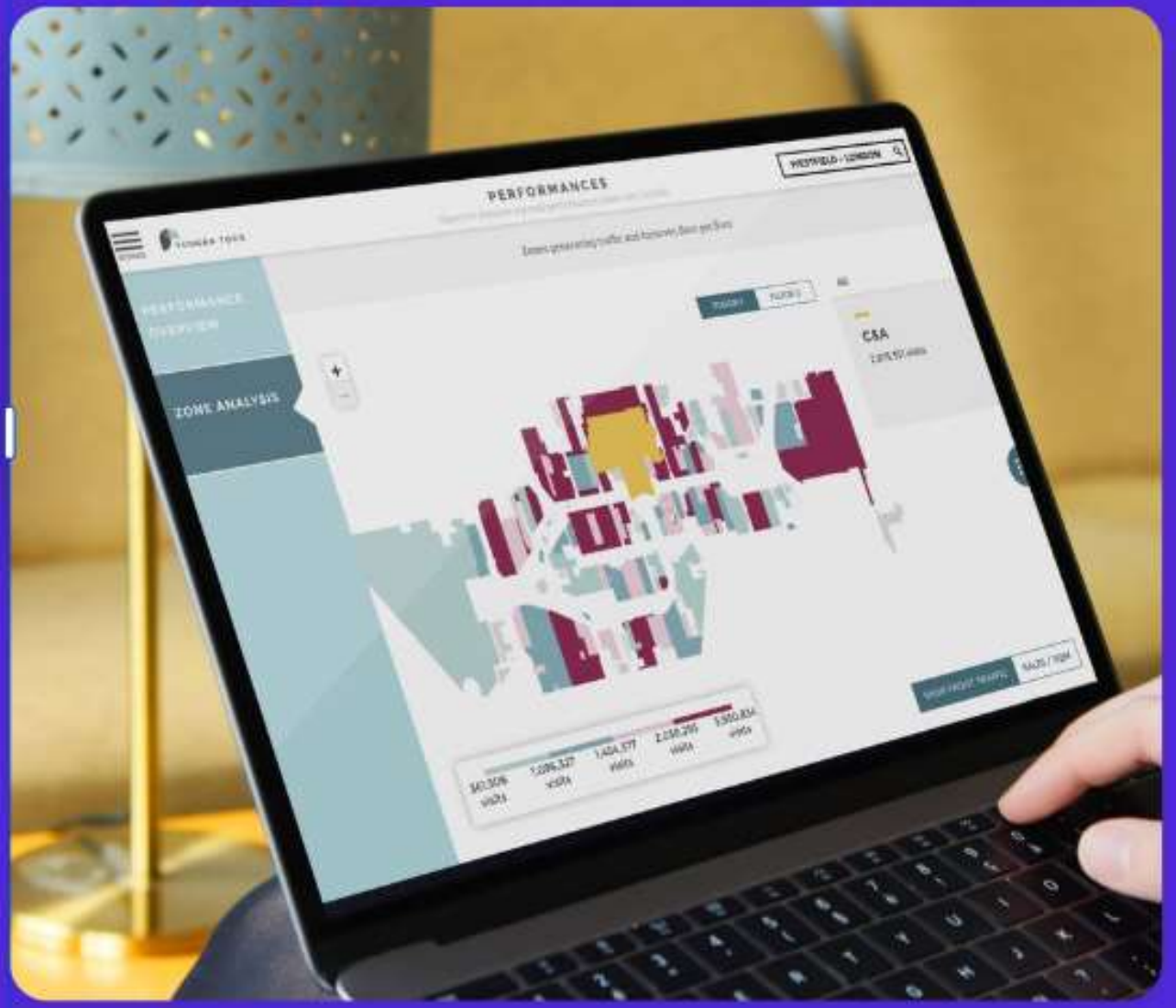
Actual vs Predicted Prices

Compares actual house prices against predicted values to assess model



Residuals Distribution

Displays distribution of prediction errors to evaluate model performance.



Meet Our Team Members for MCA

Shivam Chaudhary

Student ID: 202410116100199, passionate about technology and learning.

Sujal Gupta

Student ID: 202410116100214, a dedicated member of our team.

Shubhranshu

Student ID: 202410116100205, innovative thinker and problem solver.

Engage with Us: Let's Discuss Your Ideas

Thank you for your attention *We invite you to ask questions and explore this project further. Contact us to start the conversation.*