

CIND 820 Data Analytics final project Abstract

Name: Rabia Danish

Student Number: 501210698

Title: Prediction of housing prices of Boston

Supervisor's Name: Not assigned yet

Date of submission: 15th May 2023

Abstract

In today's ever challenging world of real estate, where housing prices are constantly on the rise, it has become increasingly important for individuals to make well-informed decisions when it comes to property investments. Our research uses available "Housing Prices Dataset" for Boston. By harnessing the power of this extensive dataset, our goal is to unravel the intricate factors that influence housing prices and provide individuals with the knowledge necessary to navigate the market with confidence and foresight. This will allow to determine future prices based on the available features which will enable people to make reliable decisions. Our dataset consists of various features related to the properties, including price, area, number of bedrooms, bathrooms, stories, and other factors that may influence housing prices. The objective of this project is to develop precise prediction models that can assist in estimating housing prices.

We aim in conducting exploratory data analysis to gain valuable insights into the dataset, identifying patterns, and understanding the relationships between the features and the target variable (price). Data preprocessing techniques such as handling missing values, feature selection, and encoding categorical variables are applied to ensure the data is in a suitable format for analysis.

In this study, we will employ several machine learning techniques, including linear regression, decision tree regression, and random forest regression, to develop predictive models for housing price estimation using cross Validation technique. These models will consider input features such as the number of bedrooms, bathrooms, stories, and the area of the house, aiming to capture the underlying linear relationship with the target variable, price. To assess the performance of these models, we will utilize commonly used evaluation metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) and R2 coefficient.

By comparing the performance of these models using these evaluation metrics, we will gain insights into their accuracy and reliability in predicting housing prices.

Additionally, correlation analysis is conducted to examine the relationships among the different attributes of houses using python. By determining the degree of correlation between variables, we can identify highly correlated attributes and assess their impact on housing prices. This analysis helps address the issue of multicollinearity, where highly correlated attributes can introduce redundancy and distort the accuracy of the prediction models. By identifying the most influential attributes, we can refine the models and focus on the features that have a significant impact on housing prices.

Overall, this data analytics capstone project aims to use the provided dataset to develop accurate prediction models for housing prices by considering different features which may impact those prices, therefore, enabling stakeholders to make informed decisions regarding real estate investments, pricing policies, and market analysis.

Data set:

The Project aims to use data to achieve the goal:

<https://www.kaggle.com/datasets/yasserh/housing-prices-dataset>