Title: Predicting House Prices Using Machine Learning: A Comprehensive Analysis

Abstract:

This study presents a comprehensive analysis of the application of machine learning techniques to predict house prices. With the growing complexity of the real estate market, accurate price estimation has become a crucial factor for buyers, sellers, and real estate professionals. This research explores the effectiveness of various machine learning algorithms, such as linear regression, decision trees, random forests, and neural networks, in modeling house price predictions.

The dataset used In this study comprises a wide range of features, including property characteristics, location attributes, and economic indicators, to ensure a holistic approach. Feature engineering and data preprocessing techniques are employed to enhance the model's performance.

The study evaluates model performance through various metrics, including mean squared error, mean absolute error, and R-squared, and employs cross-validation to ensure the generalization of the models. Furthermore, interpretability and feature importance analysis are conducted to provide insights into the driving factors behind house prices.

The results demonstrate the efficacy of machine learning in predicting house prices with a high degree of accuracy. This research contributes to the real estate industry's understanding of pricing dynamics and provides valuable tools for decision-making. The findings offer insights for buyers, sellers, and investors in the housing market, facilitating informed choices and reducing uncertainty. Future work can focus on incorporating additional data sources and improving model interpretability to further enhance predictive capabilities in this domain.

House Price Index (HPI) is commonly used to estimate the changes in housing price. Since housing price is strongly correlated to other factors such as location, area, population, it requires other information apart from HPI to predict individual housing price. There has been a considerably large number of papers adopting traditional machine learning approaches to predict housing prices accurately, but they rarely concern about the performance of individual models and neglect the less popular yet complex models. As a result, to explore various impacts of features on prediction methods, this paper will apply both traditional and advanced machine learning approaches to investigate the difference among several advanced models. This paper will also comprehensively validate multiple techniques in model implementation on regression and provide an optimistic result for housing price prediction.

House price prediction can help the developer determine the selling price of a house and can help the customer to arrange the right time to purchase a house. There are three factors that influence the price of a house which include physical conditions, concept and location.

The dataset contains 13 features :

1	Id	To count the records.
2	MSSubClass	Identifies the type of dwelling involved in the sale.
3	MSZoning	Identifies the general zoning classification of the sale.
4	LotArea	Lot size in square feet.

5	LotConfig	Configuration of the lot
6	BldgType	Type of dwelling
7	OverallCond	Rates the overall condition of the house
8	YearBuilt	Original construction year
9	YearRemodAdd	Remodel date (same as construction date if no remodeling or additions).
10	Exterior1st	Exterior covering on house
11	BsmtFinSF2	Type 2 finished square feet.
12	TotalBsmtSF	Total square feet of basement area
13	SalePrice	To be predicted