

Duke Forest House Price Prediction

A.A.Barakat, Hamza Aslan*

1 Introduction

When individuals are in the market for a new home, they typically seek out properties that are reasonably priced and possess the desired features they are looking for. A variety of factors contribute to a property's value, including its location, number of bedrooms, appearance, and size. While professional appraisers have traditionally been responsible for predicting property values, their opinions may be influenced by the interests of the lender, mortgage broker, buyer, or seller. As a result, an automated prediction system serves as an unbiased third-party source of information.

Access to accurate property value predictions is crucial for central banks, financial supervision authorities, investors, and homeowners. A house price prediction system can aid in their decision-making process by helping them determine whether the property is worth the asking price. For those looking to sell their homes, a price prediction system can assist them in identifying which features they should add to their property to increase its value.

1.1 Literature Review

Given the significance of the housing market in the global economy, there has been extensive research on the development of house price prediction models, with a particular focus on using machine learning and statistical techniques to accurately forecast housing prices.

H. Ahmed & N. Moustafa (2016) propose using visual and textual features to predict the house price. Then, these extracted features are passed through a neural network model (NN). Through experiments the visual features increased the performance. Different 11 machine learning regression models are used by Imran et al. Park & Bae (2015) using and evaluated to determine the best model for a better housing price prediction of the capital Islamabad. Nur et al. Alfiyatin et al. (2017) modeling rely on three factors: physical conditions, concept, and location to predict the price of a house. Different tests, such as linear regression and particle swarm optimization (PSO) methods are leveraged to predict the house price.

Chen et al. Yang et al. (2018) build an array of machine learning models to predict the price of a product given its image. Additionally, this work visualizes the features that result

*21080637, [Github Repo](#)

in higher or lower price predictions. For price regression, they use linear regression on histogram of oriented gradients, convolutional neural network (CNN) features. Nissan et al. Mysore et al. Kumar et al. (2021) prediction compare different algorithms, such as linear regression, support vector regressor, XGBoost regressor, CatBoost regressor, and others to predict the house prices. CatBoost Regressor and SVR have the highest accuracy by 90%. Recently, Kumar et al. Mysore et al. (2022) prediction use the Decision Tree machine learning algorithm to predict house prices in Mumbai city, where additional features, such as air quality and crime rate are used to help in predicting the prices.

1.2 Dataset

We will use the “Sale prices of houses in Duke Forest” dataset which is publicly available at the OpenIntro website. This website provides free and open-source statistics and data science educational materials, including textbooks, videos, and online courses. The organization aims to make high-quality educational resources accessible to all students and promote statistical transparency. Their materials are widely used in academic courses and by self-learners.

The dataset contains the data on houses that were recently sold in the Duke Forest neighborhood of Durham, NC in November 2020. It is collected from Zillow.com which is a popular online real estate marketplace that provides information on homes for sale, rent, and recently sold properties in the United States. It offers users various tools to search for and estimate the value of a home, as well as resources for home buyers and sellers. The website is widely used by those interested in buying, selling, or investing in real estate.

The “Sale prices of houses in Duke Forest” dataset consists of 98 observations and 13 variables. In the following, we will describe each of the variables separately.

- **address** Address of house.
- **price** Sale price, in USD.
- **bed** Number of bedrooms.
- **bath** Number of bathrooms.
- **area** Area of home, in square feet.
- **type** Type of home (all are Single Family).
- **year__built** Year the home was built.
- **heating** Heating sytem.
- **cooling** Cooling system (other or central).
- **parking** Type of parking available and number of parking spaces.

- **lot** Area of the entire property, in acres.
- **hoa** If the home belongs to an Home Owners Association, the associted fee (NA otherwise).
- **url** URL of the listing.

2 References

- Alfiyatin, A. N., Febrita, R. E., Taufiq, H., & Mahmudy, W. F. (2017). Modeling house price prediction using regression analysis and particle swarm optimization case study: Malang, east java, indonesia. *International Journal of Advanced Computer Science and Applications*, 8(10).
- H. Ahmed, E., & N. Moustafa, M. (2016). House price estimation from visual and textual features. *8th International Conference on Neural Computation Theory and Applications*, 62–68.
- Kumar, G. K., Rani, D. M., Koppula, N., & Ashraf, S. (2021). Prediction of house price using machine learning algorithms. *2021 5th International Conference on Trends in Electronics and Informatics (ICOEI)*, 1268–1271.
- Mysore, S., Muthineni, A., Nandikandi, V., & Behera, S. (2022). Prediction of house prices using machine learning. *Int. J. Res. Appl. Sci. Eng. Technol*, 10(6), 1780–1785.
- Park, B., & Bae, J. K. (2015). Using machine learning algorithms for housing price prediction: The case of fairfax county, virginia housing data. *Expert Systems with Applications*, 42(6), 2928–2934.
- Yang, R. R., Chen, S., & Chou, E. (2018). AI blue book: Vehicle price prediction using visual features. *arXiv Preprint arXiv:1803.11227*.