

Teruel City

Döşoğlu, Neşe*

1 Introduction

Real estate valuation is a complex and challenging task with significant implications for various stakeholders, including homebuyers, sellers, lenders, real estate agents, and government agencies. Accurately predicting a property's value requires considering multiple factors, such as location, area, number of bedrooms, and whether it is remodeled or new. Traditionally, professional appraisers have been responsible for predicting property values, but their opinions may be influenced by the interests of different parties involved in a real estate transaction. The emergence of automated prediction systems has transformed the real estate industry, providing more objective and accurate property valuations. These systems use computer algorithms and machine learning techniques to analyze vast amounts of data and generate predictions based on historical sales data, market trends, and other relevant factors. As a result, automated prediction systems have become a valuable tool for homebuyers, sellers, lenders, and other stakeholders in the real estate market.

1.1 Literature Review

Numerous research studies have been conducted on house price prediction, where machine learning models have been frequently utilized. These models are well-suited for the task because they can analyze large datasets and detect patterns that may take time to be evident to humans.

Wang et al. (2019) uses house price prediction model based on deep learning is proposed in this paper, implemented on the TensorFlow framework. Park & Bae (2015) uses various regression techniques in this pathway, and our results are not the sole determination of one technique. Instead, it is the weighted mean of various techniques to give the most accurate results. Gebru et al. (2017) proposes a pipeline that uses a deep neural network model to automatically extract visual features from images to estimate house prices. Gokalani et al. (2022) focuses on applying different regression algorithms to find the sales price prediction of the house. Park & Bae (2015) concepts that housing prices are influenced by characteristics such as location, distance, and region is known as price prediction. It uses Linear Regression

*Student ID, [Github Repo](#)

(LR) and other Machine Learning algorithms to forecast the price of real estate. Konwar et al. (2021) develops algorithms, builds models from data, and uses them to predict new data. Also uses various algorithms explained below in various combinations, and each algorithm's result is given based on the accuracy percentage.

1.2 Dataset

We will analyze the Teruel City dataset, which includes the physical attributes and house prices in the center of Teruel city (Spain) area. This data set can be found on *Mendeley* (2016) website. This website is research data management platform that allows researchers to store, share, and discover data. It offers a secure and free cloud-based storage solution for researchers to store and manage their research data and provides tools to make research data discoverable and citable. The website also allows users to search for and access research data from other researchers and institutions. Teruel City dataset formed with the available house prices in the center of Teruel city (Spain) on December 30, 2016 from *Idealista* (2000). This website is a real estate platform that allows users to search for properties to rent or buy in Spain, Portugal, and Italy. The website features a large database of property listings, along with tools to help users find their ideal home, such as property alerts, maps, and detailed property descriptions. It also offers a mobile app for convenient access to its services on-the-go.

The Teruel City dataset contains 58 observations and 14 variables. Each variable will be described individually in the following sections. **Flat** : Type of the house is it apartment or individual. It is a categorical value with the values [1, 0]. **Duplex** : Is the property a duplex or not. It is a categorical variable with the values [1, 0]. **Attic** : Does the house contain an attic. It is a categorical variable with the values [1,0]. **Location** : Location of the house. It is a categorical variable with the values [San Julian 8, San Francisco 19, Parra 2, ..., san benito]. **Bedrooms** : The number of bedrooms in the house. It is a discrete variable with the values [1, 2, 3, 4, 5, 8]. **Area** : The area of the house. It is a continuous variable with the range of values [42, 259]. **Floor** : The floor number of the house. It is a discrete variable with the values [-1, 0, 1, 2, 3, 4, 8]. **Lift** : Does the house contain a lift. It is a categorical variable with the values [1,0]. **Garage** : Does the house contain a garage. It is a categorical variable with the values [1,0]. **GasHeating** : Does the house contain a gas heating system. It is a categorical variable with the values [1,0]. **RemodeledOrNew** : Is the house new or remodeled. It is a categorical variable with the values [1, 0]. **BoxRoom** : Does the house contain a box room. It is a categorical variable with the values [1,0]. **NumWC** : The number of bathrooms in the house. It is a discrete variable with the values [1, 2, 3]. **PriceThousandsEuros** : The price of the house in EUROS. It is a continuous variable with the range of values [33, 500].

2 References

- Gebreu, T., Krause, J., Wang, Y., Chen, D., Deng, J., Aiden, E. L., & Fei-Fei, L. (2017). Using deep learning and google street view to estimate the demographic makeup of neighborhoods across the united states. *Proceedings of the National Academy of Sciences*, 114(50), 13108–13113.
- Gokalani, L. B., Das, B., Ramnani, D. K., Kumar, M., & Shah, M. A. (2022). House price prediction of real time data (DHA defence) karachi using machine learning. *Sir Syed University Research Journal of Engineering & Technology*, 12(2), 75–80.
- Idealista*. (2000). Idealista, Inc. <https://www.idealista.com/en/>
- Konwar, R., Kakati, A., Das, B., Shah, B., & Muchahari, M. (2021). House price prediction using machine learning. *The Journal of Philosophy Psychology and Scientific Methods*, 9, 2455–6211.
- Mendeley*. (2016). Mendeley, Inc. <https://data.mendeley.com/>
- 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.
- Wang, F., Zou, Y., Zhang, H., & Shi, H. (2019). *House price prediction approach based on deep learning and ARIMA model*. 303–307.