

Chapter 15

Cloud Hosted Ensemble Learning–Based Rental Apartment Price Prediction Model Using Stacking Technique

Rajkumar S.

Vellore Institute of Technology, Vellore, India

Ramanathan L.

Vellore Institute of Technology, Vellore, India

Mary Nikitha K.

Vellore Institute of Technology, Vellore, India

Rajasekar Ramalingam

*University of Technology and Applied Sciences,
Sur, Oman*

Mudit Jantwal

Vellore Institute of Technology, Vellore, India

ABSTRACT

In this chapter, online rental listings of the city of Hyderabad are used as a data source for mapping house rent. Data points were scraped from one of the popular Indian rental websites www.nobroker.in. With the collected information, models of rental market dynamics were developed and evaluated using regression and boosting algorithms such as AdaBoost, CatBoost, LightGBM, XGBoost, KRR, ENet, and Lasso regression. An ensemble machine learning algorithm of the best combination of the aforementioned algorithms was also implemented using the stacking technique. The results of these algorithms were compared using several performance metrics such as coefficient of determination (R^2 score), mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), and accuracy in order to determine the most effective model. According to further examination of results, it is clear that the ensemble machine learning algorithm does outperform the others in terms of better accuracy and reduced errors.

DOI: 10.4018/978-1-6684-6001-6.ch015

INTRODUCTION

Rent is always a key variable when it comes to explaining urban phenomena, whether theoretically or empirically. Many real estate market participants, including investors, regulators, and policymakers, rely on accurate rental price forecasting. Rent predictions, for example, are critical in the property valuation used in discounted cash-flow models. The imputed rent is a significant component of the estimate of gross domestic income. For public housing policy, a thorough understanding of the structure and evolution of rents in local and national housing markets is also essential. Furthermore, a general idea of the features of a piece of real estate that were most likely to affect the price during the pandemic can play a major role in the real estate market (Grybauskas, 2021). The parties involved in the real estate industry often tend to depend on external valuation or various other internal approaches. Due to the large number of parameters involved, these metrics are unreliable and imperfect because they do not take into account all of the property's unique qualities during calculations (Viriato, 2019).

The real estate industry is growing by the minute and hence advanced and accurate predictions of prices of listings are the need of the hour. The market is changing all the time, and today many software juggernauts are turning to artificial intelligence for improved decisions and resolving certain complex real-world problems using large amounts of data. Machine learning has now become interdisciplinary and can be incorporated into a variety of industries including real estate so that it can aid not only investors to increase their business throughput but also give individuals access to accurate information about properties within their financial limits. As a result, the ML models have the potential to take into account a wide range of criteria while analyzing patterns in order to provide efficient results in terms of complexity and accuracy (Reddy, 2018).

The main difference between unsupervised learning and supervised learning is that in unsupervised learning, the learner has no prior information about the purpose of learning, and is left to make its conclusion just based on the information provided, while in supervised learning, there must be data present on the target, using which it makes further predictions. Regression, a type of supervised learning is employed to tackle this problem since our goal is to anticipate real estate pricing based on previously collected, relevant information (Dataken, 2018). The dataset is obtained by scraping a popular Indian rental listing website with 18 variables including the number of rooms, bathrooms, amenities available, and size among others which are factors that can influence the price of a house. Based on seven regression algorithms and boosting methods the data is preconditioned (pre-processed) and trained. The boosting algorithms include AdaBoost, CatBoost, LightGBM, XGBoost, KRR, ENet Lasso regression, and Stacked Regressor. And then the ensemble model is generated using the best combination of the aforementioned algorithms to increase the accuracy.

RELATED WORKS

A few studies have raised a query that what attributes or variables of apartments influence their price. Swarali M. Pathak et al (Swarali, 2021) explored the correlation between house price and a number of attributes and came to the conclusion that Location and size of the apartment had strong links with the house price. On the other hand, Andrius, G et al (Grybauskas, 2021) concluded that the TOM (Time on the market) attribute is the predominant and constant variable for price prediction.

8 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/cloud-hosted-ensemble-learning-based-rental-apartment-price-prediction-model-using-stacking-technique/314147

Related Content

Identifying Suggestions in Airline-User Tweets Using Natural Language Processing and Machine Learning

Rafael Jiménez, Vicente García, Karla Olmos-Sánchez, Alan Ponce and Jorge Rodas-Osollo (2021). *Handbook of Research on Natural Language Processing and Smart Service Systems* (pp. 481-498). www.irma-international.org/chapter/identifying-suggestions-in-airline-user-tweets-using-natural-language-processing-and-machine-learning/263117

Issues in the Syntactic Parsing of Queries for a Natural Language Interface to Databases

Alexander Gelbukh, José A. Martínez F., Andres Verastegui and Alberto Ochoa (2021). *Handbook of Research on Natural Language Processing and Smart Service Systems* (pp. 157-179). www.irma-international.org/chapter/issues-in-the-syntactic-parsing-of-queries-for-a-natural-language-interface-to-databases/263100

A Study of the State of the Art in Synthetic Emotional Intelligence in Affective Computing

Syeda Erfana Zohora, A. M. Khan, Arvind K. Srivastava, Nhu Gia Nguyen and Nilanjan Dey (2020). *Natural Language Processing: Concepts, Methodologies, Tools, and Applications* (pp. 1199-1212). www.irma-international.org/chapter/a-study-of-the-state-of-the-art-in-synthetic-emotional-intelligence-in-affective-computing/239986

An Extensive Text Mining Study for the Turkish Language: Author Recognition, Sentiment Analysis, and Text Classification

Durmu Özkan ahin and Erdal Klç (2021). *Natural Language Processing for Global and Local Business* (pp. 272-306). www.irma-international.org/chapter/an-extensive-text-mining-study-for-the-turkish-language/259794

Artificial Intelligence for the Identification of Endometrial Malignancies: Application of the Learning Vector Quantizer

Abraham Pouliakis, Niki Margari, Effrosyni Karakitsou, Evangelia Alamanou, Nikolaos Koureas, George Chrelias, Vasileios Sioulas, Asimakis Pappas, Charalambos Chrelias, Emmanouil G. Terzakis, Vasileia Damaskou, Ioannis G. Panayiotides and Petros Karakitsos (2020). *Natural Language Processing: Concepts, Methodologies, Tools, and Applications* (pp. 266-279). www.irma-international.org/chapter/artificial-intelligence-for-the-identification-of-endometrial-malignancies/239940