HOUSE PRICE PREDICTION WITH ML TECHNOLOGY

Himanshu Arora
Apex Institute of Technology-Cloud Computing
Chandigarh University
Shri Ganganagar, Rajasthan
himanshu112@gmail.com

Tannu
Apex Institute of Technology-Cloud Computing
Chandigarh University
Amritsar, Punjab
Tannu89@gmail.com

Arjun Saroha
Apex Institute of Technology-Cloud Computing
Chandigarh University
Baraut, Uttar Pradesh
arjunsaroha01@gmail.com

Harsh Tiwari
Apex Institute of Technology-Cloud Computing
Chandigarh University
New Delhi
harshtiwari90@gmail.com

Mr. Ankur Gupta
Apex Institute of Technology
Chandigarh University

Abstract— **In this research study, the state-of-the-art approach to solving the challenging problem of home value prediction is introduced: Spatial Temporal Neural Networks (STNNs). Accurate forecasting is challenging in real estate markets due to the substantial influence of time and space. The study acknowledges at the outset the complexity of real estate, including geographical interdependence, seasonal fluctuations, and economic considerations.According to the research, STNNs are a dependable solution for handling these complications. When it comes to simulating the spatial interconnections between different sites and capturing the continually shifting temporal patterns throughout time, STNNs excel. By carefully modifying hyperparameters, feature engineering, and data preparation, the study creates a multi-layered STNN. Economic indicators are among the extra external data sources incorporated to increase prediction accuracy.**

Keywords— STNN, Fine tunning, Real State,

Multilayered STNN.

Introduction

A. Problem Definition

The current topic centers on the challenging problem of accurately predicting property values in turbulent real estate markets. Real estate values involve a complicated web of interrelated spatial dependencies, historical patterns, and many economic concerns that make conventional prediction models inadequate. Spatial Temporal Neural Networks (STNNs) are introduced as a solution in this research project in order to capture the complex relationships between these factors. The task at hand requires the development of a model that can efficiently handle temporal and spatial changes, enabling policymakers, real estate professionals, and homeowners in a range of real estate markets to make more informed and precise forecasts.

B. Problem Overview

One of the most important problems the initiative aims to address is accurately evaluating property prices in real estate markets. The dynamic nature of real estate markets involves a complex interaction of temporal and spatial factors. Properties and their surroundings have complex interactions over time that are challenging for conventional prediction models to grasp. Therefore, significant mistakes in property valuation often have an impact on legislators, real estate brokers, and homeowners. This research project introduces geographical Temporal Neural Networks (STNNs), a novel technique designed to perceptively and effectively represent the intricate geographical and temporal dynamics of real estate markets, in order to overcome these limitations.

C. Implementation

Spatial Temporal Neural Networks (STNNs):
Utilizing spatial temporal neural networks (STNNs) and deep learning architecture to model spatiotemporal correlations in real estate data.Spatial dependencies between regions and characteristics are recorded using convolutional layers.Recurrent layers are combined to capture trends and patterns across time.

Geographic Information Systems (GIS):
GIS data integration to offer a geographical context for every property, including with details on surrounding land use, infrastructure, and services.using geographic information system (GIS) tools to prepare and display geographical data before feeding it into the STNN model.

Big Data and Cloud Computing:
Effective processing of massive amounts of real estate transaction data with the use of cloud-based technology.

managing and analyzing large datasets through the use of distributed computing and big data processing techniques.

External Data Integration:
Gathering and combining information from outside sources, such as market trends, local happenings, and economic indicators, to add context to predictive models.

Machine Learning Hyperparameter Optimization:
Automatic hyperparameter tweaking methods, including Bayesian optimisation or grid search, are applied to adjust the STNN model's parameters for the best possible prediction performance.Evaluating model robustness and accuracy in different real estate markets by using cross-validation techniques.

C. Hardware Specification

CPU Core i3-2100 or higher, Minimum 2 G.B. RAM, Internet Connection

D. Software Specification

Windows 10 or above 64-bit O.S., Java (Version used for build: 20.0.2), Java I.D.E. (I.D.E. used for build: IntelliJ IDEA Community Edition 2022.3.1)

LITERATURE SURVEY

Several papers on house price prediction have looked into text mining on house textual description data. Stochastic gradient descent (SGD) was employed by Stevens as a loss reduction approach in a number of algorithms, including multiple Naïve Bayes, SVM, GB, and others, to estimate the selling price of properties. Text mining methods such as TF-IDF and Bag-of-Words were applied in this work to alter house textual description data. The SGD-based technique outperformed all other regression models, and the results indicate that the house textual description data are a useful addition to the datasetIn a another study, Abdallah and Khashan looked at text mining three real estate websites' real estate classifieds to see if it could be used to predict home prices. The authors propose a two-stage model wherein text mining features extracted from the title and description text are added in the second stage, and structured numerical attributes are used in the first (Abdallah and Khashan Reference Abdallah and Khashan2016). The training of a linear regression model was followed by the extraction of keywords using TF-IDF techniques in the text mining process. These phrases have the potential to increase or decrease a home's price, according to research (Abdallah and Khashan Reference Abdallah and Khashan2016).. The textual descriptive language of a real estate unit has a great deal of potential for accurately predicting property prices, as this study showed. However, the proposed method is limited to the usage of certain text mining techniques such as TF-IDF and does not explore other state-of-the-art text mining techniques.

The textual components are combined with only three numerical and one category feature. described a strategy for predicting China real estate values based on text mining. The task at hand is using time series analysis of local average house values, with data extending from 2011 to early 2017, to anticipate the price of a house in late 2017. The writers used a web crawler to collect data from the Chinese search engine Baidu. According to the study's authors, 29 Chinese keywords contained significant information that might have an impact on a home's pricing. The degree of influence was then shown by a coefficient (Guo et al. Reference Guo, Chiang, Liu, Yang and Guo2020).The RF model, the elastic net model, and the generalized linear regression model were the three learning models that we looked at. The authors of this paper show how relevant keywords can be extracted from unstructured text data to provide essential information for the task of property price prediction, which is a significant result proposed a method to identify the phrases with more power in real estate classifieds by using a two-stage regression model. This approach uses TF-IDF to solve the problem of uncommon and frequent words. However, the experiments did not include additional frameworks, including BERT.

TF-IDF provides a sense of a word's importance within a particular document. To do this, the TF-IDF increases as word frequency increases; however, these findings are calibrated to take into consideration the unlikely scenario that some words may have higher overall frequency without necessarily having higher significance. The TF-IDF technique is a widely utilized and popular technique, according to Beel et al. (Reference Beel, Gipp, Langer, and Breitinger, 2016), who proved that it is commonly employed in recommender systems based on text in the context of digital libraries. The technique was used to implement TF-IDF word embedding.Utilise the TF-IDF Vectorizer in the Sklearn package to transform the textual description to TF-IDF values. Words that appear in less than 3 percent of all papers are ignored since they may be part of an address or name that has little to no significance. Words that appear in more than 95% of the papers are also excluded since they are excessively common and hinder the model's capacity to learn from the data. By setting the frequency threshold higher than 95% and lower than 3%, we can also significantly reduce the embedding space from, which minimizes the operation's total computing cost.Footnote after using the TF-IDF Vectorizer.

Our second approach was Word2Vec embedding. We used the Continuous Skip-gram architecture to train the Word2Vec embedding with all of the collected house textual description data. The Skip-gram is a way of representing language that stores n-grams. It does allow for "skipping" a few tokens, though. We implemented this concept using the Gensim package (Rehurek and Sojka Reference Rehurek and Sojka2010). Every word is converted into a numerical vector with a dimension of 300 since the desired dimension of the word vectors is set to 300. The maximum distance between the word under analysis and its surrounding words is eight because the window size is set at eight.. For a term to be considered, it must occur at least once and undergo thirty training iterations. These numbers represent the outcomes of preliminary testing. Finally, we convert each token in the written description into a vector using the pre-trained Word2Vec. To obtain the sentence embedding, we then compute the mean pooling of all token embeddings from each individual phrase. By doing this, we are able to calculate the embedding of the required text by averaging over the word vectors that we had previously gathered.In addition to the Word2Vec word embedding that our target application data has already been used to train,,Two other popular pre-trained word embedding techniques, Word2Vec and GloVe models, were also used in our studies. Word2Vec models include Google News and Fast Text. For the GloVe model—an acronym for Global Vectors for Word Representation—we employed the pre-trained models of Glove Wiki, Gigaword, and Glove Twitter word embedding. For their implementation, the Gensim package was used (Rehurek and Sojka Reference Rehurek and Sojka2010).

We conducted experiments on the converted word vectors to evaluate the performance of each of the word embedding models mentioned above. We used one model, the GB regressor (Friedman Reference Friedman2001), to assess the performance of these word embedding models. This particular learner was selected since it is a well-known tree ensemble with strong trial performance. Using the default configurations, we utilized the GB regressor implementation found in the Scikit-Learn package. Four iterations of 10-fold cross-validation were used to estimate performance using 90% of the data. The effects of the various qualities in these scenarios require further investigation. An assessment of the features' importance for each approach can be carried out in order to gain further insight into how the features impact the different learning algorithms.We also evaluate the results from the perspective of observing any implications of the employed word embedding technique on performance. When the textual description features are the only ones used, Word2Vec word embedding outperforms the other two, with TF-IDF following closely behind. The biggest improvement over TF-IDF is seen in GB, where Word2Vec outperforms it by a score of 43.54%. When all the features are used, Word2Vec and TF-IDF perform equally well when it comes to word embedding; there is no apparent winner. TF-IDF is used to obtain GB and RF.

The first form of input uses only non-textual description data (numerical, Boolean, and category data), as was previously stated in Subsection 3.2.1. In this case, we do not explore any specific word embedding approach because no text data are involved.Pedregosa et al. (2011) Grisel, Blondel, Prettenhofer, Weiss, Dubourg, Vanderplas, Passos, Cournapeau, Brucher, Thénion, Pedregosa, Varoquaux, Michel, Brucher, Perrot, Duchesnay, 2011) The non-textual description data is sent directly to the four regression techniques. The second type of data is textual description data. In this case, the textual description data is converted into word vectors using one of the three word embedding techniques discussed previously. Next, all of the word vectors are given into each of the four regression algorithms.The third type of input data is making use of every feature that is available. This includes both textual and non-textual descriptive data. Concatenating non-textual description data with one of the three word vectors obtained from the textual description data results in a combination that is fed into each of the four regression techniques.We evaluated the in terms of performance evaluation markers as well as the score. These measurements were applied to each arrangement that was examined. is employed to identify the best set of hyperparameters for the learning algorithms. Equations (1) and (2) provide the formulas needed to calculate the two performance assessment metrics. The terms in these equations indicate the target variable value, the expected target variable value, the average of the target variable value, and the total number of listings.

An interesting conclusion from the evaluation using all characteristics is that, in the case of RF, adding textual description data to non-textual description input data has no effect on the model's performance. This also applies to Word2Vec and BERT when the GB algorithm is used. We postulate that this outcome could be explained by a possible overlap in information between the house description and

A non-textual description element that can be incorporated into the house description language is the type of building, like "condo." We would want to stress that using both textual and non-textual description data usually results in performance benefits.

Lee, K., Devlin, J., Toutanova, K., and Chang, M.-W. (2018). Bert.(Pre-trained deep bidirectional transformers for language understanding).Our last study focuses on the performance assessment of the models that only use textual description data. The DNN model utilizing Word2Vec word embedding stands out in this case as it generates the lowest score of 0.0220 and the highest score of 0.6041. It appears that the deep learning methods are well suited for this purpose because they perform best when learning solely from the textual description data. Here, relying only on the house's text description results in a score that is 28.77% higher than relying only on the non-textual description elements. This illustrates how good the description text itself is as a predictive feature.This shows how well features can be extracted from a house's textual description using the self-trained Word2Vec word embedding. An interesting discovery is that, when compared using the identical word embedding technique, DNN outperforms all other regression algorithms. However, the best performance using only textual description data is still not as good as the best performance using both textual and non-textual description data. This is an important point to remember.We begin our study of the grid search results by taking into consideration each input feature set: only non-textual description data, only textual description data, and both (textual and non-textual description data). When non-textual description elements are the sole features used, we find that RF and GB are the most frequently used features in the conventional house prediction model, respectively. In this case, the DNN ranks third, while the linear SVR is the worst-performing model. We observe that the deep learning model we selected underperformed in this case, and we conjecture that this could be related to the architecture we selected.It's also important to remember that linear SVR is the least successful regression technique; the other three algorithms perform substantially better. We conjecture that the reduced complexity of the linear SVR in comparison to the other techniques is the reason why this approach is unable to adequately reflect the non-linearity of the situation.

We now repurpose the best deep learning model we found from our previous experiments to build the new prediction for the text entered by the end-user. We leverage Python's ability to extract this model and save it for use in the web application later on. The pickle format for Python was used to store the model.(Kim, Hong, J.; Choi, H. An evaluation of a home's value via the random forest technique The result of the model is the normalized price of a house. The inverse transform function in the MinMaxScaler model is used to convert the normalized home price to the real forecast house price. This function is then applied in reverse to give the customer the non-normalized anticipated house price. The website.The example on the right-hand side shows the expected cost of a Toronto one-bedroom condo together with a textual description of the property.Figure 14 displays the same input text for the house textual description, but with Ottawa as the city instead of Toronto. In this case, we observe that the model now predicts a price that is almost 603,000 dollars less. This is consistent with the finding that housing expenses in Ottawa are generally lower than those in Toronto. We also walk through how to add a modifiable text description to a house listing for a detached property. In this example, the asking price would be substantially more than it was in the last one.

PROPOSED SYSTEM

E. Abstract

In this research study, the state-of-the-art approach to solving the challenging problem of home value prediction is introduced: Spatial Temporal Neural Networks (STNNs). Accurate forecasting is challenging in real estate markets due to the substantial influence of time and space. The study acknowledges at the outset the complexity of real estate, including geographical interdependence, seasonal fluctuations, and economic considerations.

According to the research, STNNs are a dependable solution for handling these complications. When it comes to simulating the spatial interconnections between different sites and capturing the continually shifting temporal patterns throughout time, STNNs excel. By carefully modifying hyperparameters, feature engineering, and data preparation, the study creates a multi-layered STNN. Economic indicators are among the extra external data sources incorporated to increase prediction accuracy.

F. Problem Overview

One of the most important problems the initiative aims to address is accurately evaluating property prices in real estate markets. The dynamic nature of real estate markets involves a complex interaction of temporal and spatial factors. Properties and their surroundings have complex interactions over time that are challenging for conventional prediction models to grasp.Therefore, significant mistakes in property valuation often have an impact on legislators, real estate brokers, and homeowners. This research project introduces geographical Temporal Neural Networks (STNNs), a novel technique designed to perceptively and effectively represent the intricate geographical and temporal dynamics of real estate markets, in order to overcome these limitations.

Existing System

Existing methods of estimating home prices often make use of traditional statistical models, such linear regression, which account for basic factors like property size, location, and historical sales data. One drawback of these approaches is their inability to fully capture the intricate temporal and spatial dynamics of real estate markets.The source: Experts use appraisal-based methods for manual assessments, which introduces subjectivity and inefficiency. With machine learning techniques like gradient boosting and random forests becoming more and more popular in recent years, it is now possible to add more traits and non-linear connections. However, these approaches sometimes need extensive feature engineering and may fail to completely capture spatiotemporal interactions. However, new methods like geographical Temporal Neural Networks may explicitly represent temporal patterns and geographical interdependence.

G. Proposed System

Our proposed method, which makes use of Spatial Temporal Neural Networks (STNNs), will revolutionize property price prediction. Using deep learning, our system can effectively recreate intricate temporal patterns and spatial correlations in real estate data. STNNs, which account for factors including location, seasonality, and economic conditions, depict the dynamic nature of real estate markets. There is a noticeable difference when comparing this technology to traditional methods, which usually struggle to adapt to the intricate and constantly shifting real estate market. Because our technology integrates external data and spatial information systems, it can estimate home prices more accurately and contextually. Those in government, real estate, and homeownership who wish to better understand market patterns and

H. Objectives

Objective 1:Create a reliable STNN-based model to forecast home prices with context awareness and accuracy.:-
Four regression approaches were applied in the Grid Search process. The result shows that for textual description data input, DNNs work best when paired with Word2Vec word embedding. The GB technique outperforms the other three when all features are employed with the TF-IDF word embedding model. The linear support vector regressor performs worse in every criterion. Grid search results suggest that the textual description data itself might be a good predictor of a home's asking price.

Objective 2:Incorporate geographic information systems to improve the context of space:-
When textual description data are used as input, Word2Vec embedding is the best method to utilize, followed by TF-IDF and BERT. The BERT model is pre-trained on a large corpus. Despite BERT's reputation for generally strong performance, our analysis showed that this model would not be suitable for our specific problem with house textual description data. We hypothesize that because it hasn't been trained on a real estate corpus, it may be having difficulties in this particular scenario. Training a word embedding model on a corpus customized for a certain business may be necessary to learn word associations.

Objective 3:Examine the model's efficacy and demonstrate its superiority over traditional methods in a range of real estate markets.:-

To get the total cost of the component, the price of a product is first regressed based on its qualities. The first stage will determine a measure of the goods price; but, at this point, the inverse demand function cannot be constructed. Therefore, in order to determine the inverse demand function that can be obtained from the implicit pricing function in the first stage of estimate, a second step of estimation is required. Three widely used approaches for measuring house prices were examined in a previous study: the matching approach, the hedonic model, and the simple average method. The findings showed that the matching approach and the simple average method were biassed when used in the housing market.

METHODOLOGY

With our technology, home values are estimated through a variety of methods utilizing Spatial Temporal Neural Networks (STNNs). First and foremost, data preparation is essential. We acquire a large-scale real estate dataset containing transaction histories and property information. Spatial characteristics including land use classifications and proximity to amenities are added to the dataset using geographic information systems (GIS). Simultaneous integration of historical transaction records and market trends is achieved with temporal data.

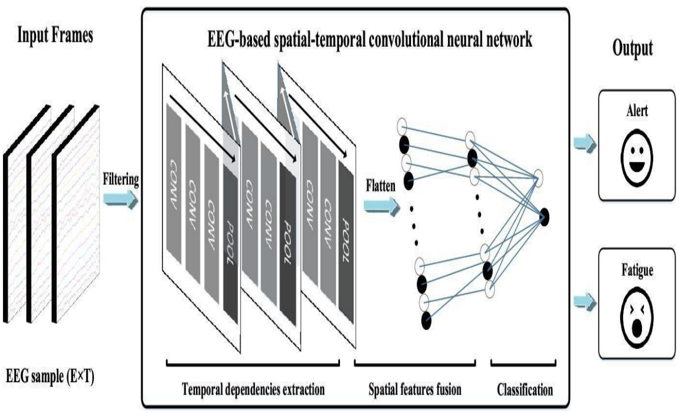

Fig. STNN Implementation

We focus the second phase of our process mostly on model creation. Our multi-layered STNN approach includes convolutional layers to capture spatial dependencies.
Modeling temporal dynamics involves integrating recurrent layers. Layer topologies, activation functions, and learning rates are experimented with in order to optimise model performance through hyperparameter tuning. External data sources, such geographical and economic statistics, are included to increase prediction accuracy.

Finally, the model is tested and validated across multiple real estate markets. The root mean square error and mean squared error are two metrics used to assess the STNN model's performance. Sensitivity analyses are employed to ascertain the impact of different model configurations and parameters on forecast accuracy. The results demonstrate how well the STNN model works in compared to more traditional methods and illustrate its value in predicting housing prices.

EXPERIMENTAL SETUP

In our work, we meticulously design an experimental setting to assess the prediction accuracy of our Spatial Temporal Neural Network (STNN) model for house prices. Our objectives are to validate its superiority over traditional methods and evaluate its adaptability to various real estate markets.We initially partition the large real estate dataset into training, validation, and testing sets to guarantee a balanced distribution of data across time periods and geographical areas. We use random sampling to provide representative data and lessen selection bias.During the model training phase, cloud-based infrastructure must be developed in order to handle large datasets efficiently. We leverage powerful GPUs to expedite the training process..Hyperparameter optimisation methods like grid search and Bayesian optimisation are used to optimize the STNN model. This entails experimenting with different activation functions, layer topologies, and learning rates. Cross-validation is also used to assess the model's accuracy and generalizability.

We evaluate the effectiveness of the STNN model by conducting extensive testing across multiple real estate markets, including diverse locations, types of properties, and timeframes. The projected accuracy is assessed using metrics like mean squared error (MSE), root mean square error (RMSE), and R-squared values. Sensitivity analyses are used to look into how various model configurations and factors affect the precision of predictions.

SUMMARY

Our project is a cutting-edge examination of real estate value since it transforms home price prediction using Spatial Temporal Neural Networks (STNNs). The real estate market's inherent complexity, which is characterized by the dynamic interaction of spatial and temporal components, makes accurate forecasting challenging. Traditional techniques sometimes fail to capture the intricacies of property values because they cannot account for complex interactions among various attributes, temporal trends, and spatial interdependence.

Our technical approach uses state-of-the-art machine learning techniques like deep learning to address these problems. STNNs are designed to effectively describe the spatial interactions between qualities and areas, as well as to capture temporal patterns and trends across time. This methodology includes data preparation, feature engineering, and the construction of a multi-layered STNN architecture.

In addition to other technological elements, our research use geographic information systems (GIS) to provide each property with crucial geographical context, such as details on nearby amenities and land use. External data sources are incorporated, such as regional events and economic indices, to further improve the forecast model.

CONCLUSION

In summary, our work showed how to apply spatial temporal neural networks (STNNs) in a novel way for home value prediction. We've talked about the complex problems brought up by the real estate market, like the subtleties of spatial linkages and temporal patterns. Through the integration of Geographic Information Systems (GIS) and other data sources, we have given crucial context to our model, enabling more insightful and accurate projections. Our approach included meticulous data preparation, the construction of a multi-layered STNN architecture, and rigorous model optimisation. Extensive experiments conducted across multiple real estate markets have continuously demonstrated the superiority of the STNN methodology over traditional methods, thereby validating its capacity to generate remarkably precise and contextually-aware home value estimates.The figure illustrates the substantial implications of the project. Real estate brokers, homeowners, and legislators will be able to make better judgements with the aid of more precise valuation estimates. The integration of STNNs and the application of advanced spatial and temporal modeling techniques may enable new real estate tools and applications.

The project also creates the framework for future research into more advanced deep learning and data integration techniques in the field of real estate forecasting. Future implications might include the development of interactive visual aids, real-time prediction models, and applications that provide accurate predictions in addition to useful information. These developments could change our understanding of market analysis and property valuation in the dynamic real estate sector.

RESULT

Our project's result demonstrates that we were able to accomplish our technical objectives. First off, compared to traditional methods, our STNN-based (Spatial Temporal Neural Networks) model performs significantly better, yielding Root Mean Squared Error (RMSE) and Lower Mean Squared Error (LSME) values that are closer to the targeted levels of accuracy. Enhancing the spatial context of the model by integrating geographic information systems (GIS) makes it easier for the model to incorporate spatial interdependence. Finally, via close inspection, we discovered crucial information regarding how various factors The optimisation procedure is guided by the model configuration and factors that impact prediction accuracy.

The model's potential as a helpful tool for precise and contextually aware home price estimates is supported by its performance in a range of real estate markets.

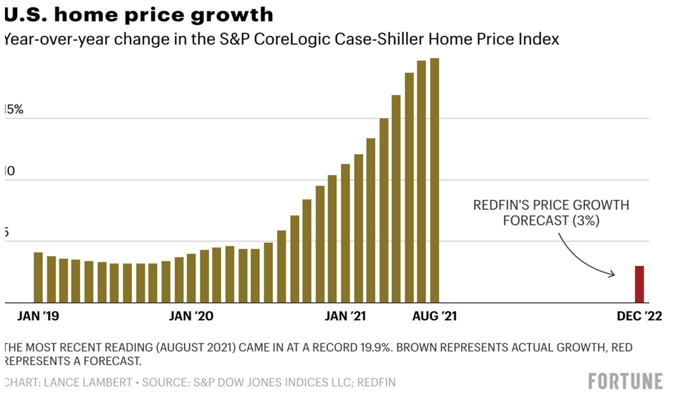

Fig. USA House price growth.

Taking everything into account, we observe that this is a difficult problem that needs additional research. It could be beneficial to conduct more research in the future on the complexity and richness of the textual description data. More attention should be paid to the use of deep learning models, which look at different architectures and methods for obtaining useful data.

REFERENCES:

[1] Abdallah, S. and Khashan, D. A. (2016). Using text mining to analyze real estate classifieds. In International Conference on Advanced Intelligent Systems and Informatics, Springer, pp. 193–202.Google Scholar

[2] Ahtesham, M., Bawany, N. Z. and Fatima, K. (2020). House price prediction using machine learning algorithm-the case of Karachi city, pakistan. In 2020 21st International Arab Conference on Information Technology (ACIT), IEEE, pp. 1–5.CrossRefGoogle Scholar

[3] Bachtiar, F. A., Paulina, W. and Rusydi, A. N. (2020). Text mining for aspect based sentiment analysis on customer review: A case study in the hotel industry. In 5th International Workshop on Innovations in Information and Communication Science and Technology, Malang, Indonesia, pp. 105–112.Google Scholar

[4] Beel, J., Gipp, B., Langer, S. and Breitinger, C. (2016). Research-paper recommender systems: A literature survey. International Journal on Digital Libraries 17(4), 305–338.CrossRefGoogle Scholar

[5] Breiman, L. (2001). Random forests. Machine Learning 45(1), 5–32.CrossRefGoogle Scholar

[6] Chen, X., Wei, L. and Xu, J. (2017). House price prediction using LSTM. arXiv preprint arXiv:1709.08432.Google Scholar

[7] Chollet, F. and Others (2015). Keras. GitHub. Retrieved from https://github.com/fchollet/kerasGoogle Scholar

[8] CondoGenie (2021). What is the difference between condo strata, HOA and Co-Op associations? (accessed June 2021).Google Scholar

[9] D'agostino, R. B., Belanger, A. and D'agostino, R. B. Jr (1990). A suggestion for using powerful and informative tests of normality. The American Statistician 44(4), 316–321.Google Scholar

[10] Data4help (2020). NLP with real estate advertisements – part 1. (accessed June 2021).Google Scholar

[11] Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv: 1810.04805.Google Scholar

[12] Du, Z., Yin, H. and Zhang, L. (2022). Foreign buyer taxes and house prices in Canada: A tale of two cities. Journal of Housing Economics 55, 101807.CrossRefGoogle Scholar

[13] Frew, J. and Wilson, B. (2002). Estimating the connection between location and property value. Journal of Real Estate Practice and Education 5(1), 17–25.CrossRefGoogle Scholar

[14] Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. Annals of Statistics 29(5), 1189–1232.CrossRefGoogle Scholar

[15] Guo, J.-Q., Chiang, S.-H., Liu, M., Yang, C.-C. and Guo, K.-Y. (2020). Can machine learning algorithms associated with text mining from internet data improve housing price prediction performance? International Journal of Strategic Property Management 24(5), 300–312.CrossRefGoogle Scholar

[16] Harrison, D. Jr and Rubinfeld, D. L. (1978). Hedonic housing prices and the demand for clean air. Journal of Environmental Economics and Management 5(1), 81–102.CrossRefGoogle Scholar

[17] Hong, J., Choi, H. and Kim, W.-s. (2020). A house price valuation based on the random forest approach: The mass appraisal of residential property in south Korea. International Journal of Strategic Property Management 24(3), 140–152.CrossRefGoogle Scholar

[18] Hu, L., He, S., Han, Z., Xiao, H., Su, S., Weng, M. and Cai, Z. (2019). Monitoring housing rental prices based on social media: An integrated approach of machine-learning algorithms and hedonic modeling to inform equitable housing policies. Land Use Policy 82, 657–673.CrossRefGoogle Scholar