Analyzing factors affecting land prices in urbanized areas using machine learning: A basis for future 3D property valuations

Peyman Jafary [1,2,*], Davood Shojaei [2], Abbas Rajabifard [2], Tuan Ngo [1,3]

[1] Building 4.0 CRC, Caulfield East, Victoria, Australia

[2] Centre for Spatial Data Infrastructures and Land Administration, Department of Infrastructure Engineering, The University of Melbourne, Melbourne, VIC, Australia

[3] Department of Infrastructure Engineering, The University of Melbourne, Melbourne, VIC, Australia

Abstract: Accurate land valuation promotes fairness and efficiency in the property market, aiding property development and land use planning. Land price is also a key component for property valuation using cost approach, particularly with 3D models like Building Information Modeling (BIM). Identifying and creating a database on value-related features is essential to establishing a robust land valuation system. Advanced Machine Learning techniques can handle non-linear relationships between land value and the driving factors. This study analyzes factors affecting land prices in Melbourne Metropolitan, considering a wide range of features. The importance of each factor is calculated through an ensemble method based on four techniques: Random Forest, XGBoost, Recursive Feature Elimination (RFE) and Mutual Information (MI). The results show that land area, longitude, land use, latitude, distance to the Central Business District (CBD), elevation, mortgage rate and primary school zone have the highest impact on the land value in the study area.

Keywords: Land price; Machine learning; Feature importance analysis; 3D property valuation using BIM; Ensemble learning

Introduction

Urbanization is a significant social and economic phenomenon that has been occurred throughout the world. In modern countries, land availability and land price deeply influence the process of horizontal urban growth and sprawl, as well as the vertical development trend. In general, areas with lower land prices tend to have urban development characterized by lower-density and shorter buildings, with a preference for horizontal expansion. Conversely, areas with higher land prices tend to have urban development with taller buildings and higher population density, with a preference for vertical expansion. Hence, land value has a remarkable impact on the urban development process (Deng and Huang, 2004; Hu et al., 2013; Silveira and Dentinho, 2018). Furthermore, one of the main approaches for the valuation of houses is the cost approach, which is based on land value plus reconstruction cost deducted by the depreciation cost. There is a growing literature on the integration of this approach with 3D models such as Building Information Modeling (BIM) and CityGML. Such 3D models can provide precise information to estimate the reconstruction cost and the depreciation cost (Arcuri et al., 2020a; Jafary et al., 2022a; Khodabakhshian and Toosi, 2021). However, it is still required to develop advanced land valuation methods to provide the land prices as the other main component of the cost approach.

To automatically estimate the price of lands at a large scale using big data, a mathematical model is used to estimate the value based on a variety of complex and interdependent factors that influence it (Hong et al., 2020; Yilmazer and Kocaman, 2020). These factors may not have a simple cause-and-effect relationship with the value and may interact with each other in complex ways, leading to non-linear relationships. Additionally, due to the spatial nature of many of these factors, they may not have a consistent impact on the value across all real estate assets or locations, leading to further non-linearity in the relationships (L. Krause and Bitter, 2012; Sampathkumar et al., 2015; Tsutsumi et al., 2011). Hence, as the first step for the automatic land valuation process, it is essential to extract the valuation factors and then analyze the importance of them in relation to the land value using methods that can mitigate the non-linearity problem. Machine Learning (ML) algorithms such as Random Forests (RF), Gradient Boosting (GB) and Artificial Neural Networks (ANN) can model such non-linear relationships by using multiple layers of nodes or decision rules to capture complex patterns in the data. These models can automatically learn to compute the importance of each feature and identify non-linear interactions between them, allowing the methods to better capture the complexity of real-world relationships. Additionally, ML methods can often handle a larger number of features than traditional regression models, making them more suitable for datasets with many potential predictors (Baur et al., 2023; Jafary et al., 2022b; Ma et al., 2020; Zhang et al., 2021).

Although some previous studies have attempted to analyze the driving factors influencing urban land value, most of them have employed statistical methods such as Hedonic Price Models (HPMs) (Kim et al., 2007), least-squares methods (Wen and

* Corresponding author.
E-mail address: pjafary@student.unimelb.edu.au

Goodman, 2013) and path analysis (Mirkatouli et al., 2018). Such methods, however, have limitations in fully handling the non-linear relationships between the land value and the affecting factors. Ma et al. have conducted a study to cover more influential factors and find the relationships between them and the land value using advanced methods in ML (Ma et al., 2020). Nevertheless, their study does not cover legal and environmental factors. Jalali et al. have emphasized that planning and zoning regulations affect the housing and land markets (Jalali et al., 2022). In addition, the impact of environmental factors such as temperature, rainfall, vegetation cover and air quality have almost been neglected in studies on land valuation and the value-related factors. Such environmental characteristics can differ across various parts of the huge urban areas, so this paper hypothesize that they could affect land value. In general, making linear assumptions, experiencing multicollinearity within the applied research models, limited number of considered features and a lack of feature selection and feature importance analysis procedures are among common limitations in the relevant studies in the literature (Ma et al., 2020). Accordingly, it is still necessary to analyze the impact of a comprehensive range of driving factors of land value using the advanced methods in ML.

Land value as a basis for 3D property valuation using BIM

For accurate valuation of constructed properties, it is important to take advantage of utilizing Three-Dimensional (3D) models since built properties have 3D origins. The 3D characteristics of a property can have a significant impact on its value, both in terms of its structural integrity and its visual appeal. Therefore, it is recommended to use 3D modeling and 3D Geographic Information Systems (GIS) to gather data and information on various 3D features that contribute to a property's value (El Yamani et al., 2019; Jafary et al., 2022b). One of the widely used approaches for the valuation of constructed properties is the cost approach, which has a very high compatibility with 3D models, especially BIM. The cost approach is based on the summation of land value and reconstruction cost deducted by depreciation cost, and the second and third components can be estimated using BIM (Jafary et al., 2022a). Some studies have attempted to develop land valuation systems based on cost approach benefitting from 3D information provided by BIM (Arcuri et al., 2020b; Couto et al., 2021; Khodabakhshian and Toosi, 2021). However, these studies have mainly focused on the estimation of reconstruction cost and depreciation cost based on BIM and have neglected a detailed methodology for land valuation as the third component of the cost approach. This could be a source of uncertainty in their valuations. Indeed, a precise valuation based on cost approach requires an accurate land valuation to be a basis for further addition of reconstruction cost and deduction of depreciation cost. Accordingly, this paper can highly contribute to the research in property valuation using 3D models such as BIM by providing a backbone for accurate land valuation through the identification of driving factors of land price.

Study area and data on land value

The study area covers 19 out of 31 Local Government Areas (LGAs) of Melbourne Metropolitan, Victoria, Australia (Figure 1). The study area spans about 2,400 square kilometers (km), including nearly 900,000 million land polygons. In order to analyze the affecting factors on land price, the land values of 26,700 properties, called Site Value (SV), received from Valuer-General Victoria (VGV), Australia, were used as input data to the ML-based feature importance analysis methods.

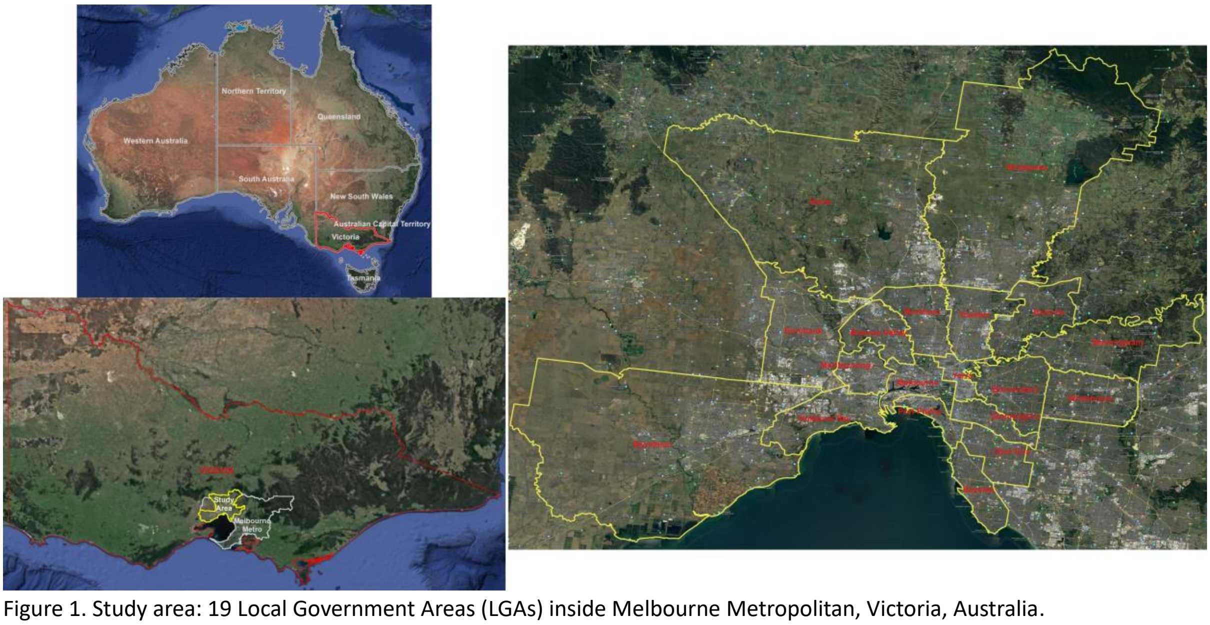

Figure 1. Study area: 19 Local Government Areas (LGAs) inside Melbourne Metropolitan, Victoria, Australia.

Land valuation factors

In order to identify a list of potential factors that can affect land value, the literature was reviewed to discover features that have previously been considered by different scholars. Most of the considered factors in the literature are based on geographic and location-based characteristics (Carranza et al., 2022; Derdouri and Murayama, 2020; Zhang et al., 2021). Some other studies have also focused on factors from socio-economic (Sampathkumar et al., 2015) or land use and planning aspects (Kim et al., 2021). Using the previous experiences, this study establishes a feature set including a broad range of affecting factors from various domains. Table 1 presents the 37 factors considered in this study, which will be further explained in the below sections. All the factors are created using the different functions provided in Esri ArcMap 10.8.1 and Quantum Geographic Information System (QGIS) 3.22.16.

Table 1. List of potential features affecting land value considered in this study.

Category	Factor	Number
Physical	Area, shape (convexity), elevation and slope	4
Geographical	Latitude, longitude, LGA, locality, road type, distance to Central Business District (CBD), distances to the different amenities (sustenance, shopping centers, higher level education centers, health centers, childcares and kindergartens, public transportation, natural landscapes and water bodies) and walkability	15
Socio-economic	Population density, employment rate, Gross Regional Product (GRP), mortgage rate, ethnicity, job opportunities, income, crash, criminal activities and offensive incidents	10
Environmental	Temperature, rainfall, green vegetation and air quality	4
Legal	Zoning, land use, primary school zone and secondary school zone	4

Physical factors

In this study, four physical features are utilized, including land area (in squared meter), shape as convexity and topography-related variables of elevation (in meter above sea level) and slope (in percentage). For elevation and slope, the average values in each land polygon were computed using the Digital Elevation Model (DEM) of Victoria with a spatial resolution of 10 meters received from the Data Share Portal of the Department of Environment, Land, Water and Planning (DELWP), Victoria (DELWP, 2023).

Geographical factors

Geographical factors are mainly based on the location of the lands, as well as their distances from different Points of Interest (PoIs). In this study, 14 geographical factors were produced for the land valuation. The locations of the PoIs were extracted from the Open Street Map (OSM) data (OSM, 2023). For the access to public transportation, the layer related to the points of train, tram and bus stations of Public Transport Victoria (PTV) was received from DELWP's Data Share Portal. For the walkability, the walk scores of the different suburbs were extracted from the Walk Score portal (Walk-Score, 2023).

Socio-economic factors

This study covers different socio-economic factors. For the population, mortgage, ethnicity, job and income factors, the total population data, the mortgage repayment rate ($/month), ancestries of the respondents, total jobs and the median total personal income ($/weekly) data were received from the Australian Bureau of Statistics (ABS) based on Census 2021 (ABS, 2023). Data on employment rate was received from the Australian Urban Research Infrastructure Network (AURIN) database (AURIN, 2023). GRP data of the different LGAs for the year 2021 was also received from the REMPLAN through web scraping (REMPLAN, 2023). Points of crashes during the past five years in Victoria were received from Victoria's Department of Transport (Department-of-Transport, 2023). Subsequently, the average distance of each land polygon from the closest ten crashes was calculated. Finally, the latest statistics (year ending September 2022) in relation to the criminal incident rate and the offense rate per 100,000 population in each LGA were received from the Crime Statistics Agency (CSA) (CSA, 2023).

Environmental factors

For the climatic factors of rainfall and temperature, mean annual rainfall and mean annual temperature maps of Australia were downloaded from the Australian Bureau of Meteorology (Bureau-of-Meteorology, 2023). For the green vegetation, the maximum annual Normalized Difference Vegetation Index (NDVI) image of the study area was produced using Sentinel-2 images from 1 January 2021 to 31 December 2021 through the Google Earth Engine (GEE) platform. For air quality, The Environment Protection Authority Victoria (EPA) Air Watch yearly data of hourly averages of air quality in the different sites in Victoria (EPA, 2023) was downloaded from the DELWP Data Share Portal for the reference year of 2021.

Legal factors

For the land use factor, the layer of Victorian land use information system 2016-2017, for the zoning, the planning scheme layer of Victoria, and the primary and secondary school zones were downloaded from the DELWP database to prepare the legal factors in this study.

Feature importance analysis using Machine Learning (ML)

Feature importance analysis is a technique used in ML to identify the most important features that contribute to the predictive accuracy of a model. It involves evaluating the contribution of each feature in the model's output and identifying the features that have the greatest impact on the target variable. Feature importance analysis is a powerful tool for understanding the most important variables that contribute to the predictive accuracy of an ML model. By identifying these key features, it is possible to optimize the model's performance and gain deeper insights into the underlying factors that drive the target variable (Salcedo-Sanz et al., 2018). There are several methods for performing feature importance analysis in ML, including coefficient values, tree-based methods, Recursive Feature Elimination (RFE), Mutual Information (MI), Genetic Algorithm (GA), Fisher Score (F-Score) and Principal Component Analysis (Dhal and Azad, 2022).

In this study, an ensemble feature importance analysis is carried out based on four robust methods of RF, XGBoost, RFE and MI. After preparation of the spatial database based on the considered 37 features, some pre-processing steps are carried out before running the ML-based feature analysis algorithms, including data cleaning, encoding and normalization. Using each of these algorithms, the importance of each feature is computed. Subsequently, the relevant importance value for each variable is normalized. The final importance ranking is then achieved by calculating the average value based on the four considered feature importance analysis methods (Solano et al., 2022). Figure 2 shows the methodology framework applied in this research to analyze driving factors of land value in the study area. Analysis is conducted using R (R-Core-Team, 2023) and RStudio (RStudio-Team, 2023).

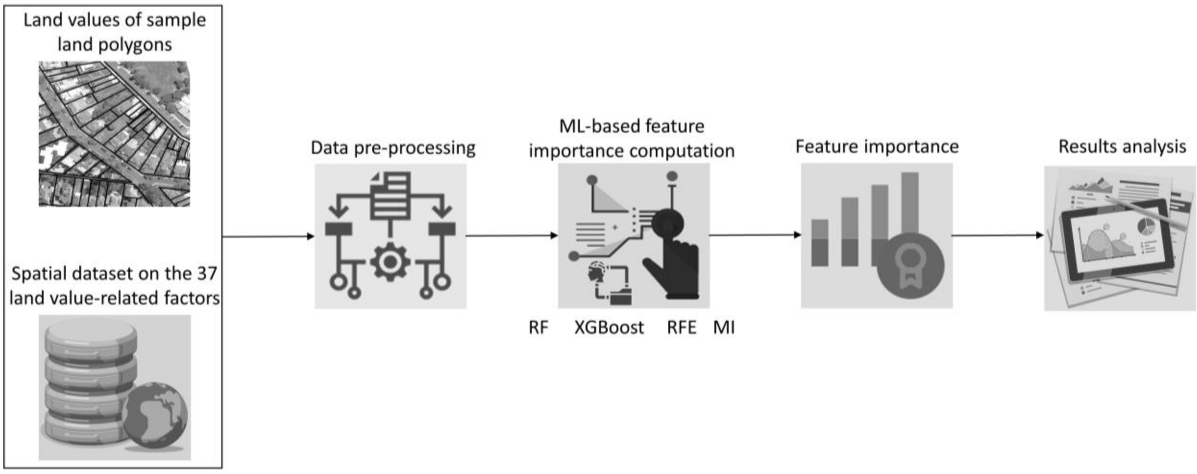

Figure 2. Research framework applied in this study.

Random Forest (RF)

Tree-based methods allow for model-specific analysis, detection of non-linear relationships and feature interactions and are robust to outliers and missing data. These advantages make tree-based methods a popular and powerful approach for understanding the most important variables that contribute to the predictive accuracy of an ML model (Clark and Pregibon, 2017). Among different tree-based methods, RF is an ensemble learning method that creates multiple decision trees and combines their results to make predictions. In addition to predicting the target variable, it can also provide valuable insights into which features are most important for making accurate predictions. To perform feature importance analysis using RF, after model training, the feature importance can be calculated using the Mean Decrease Gini (MDG) or Mean Decrease Accuracy (MDA) metrics. The MDG is based on the decrease in Gini impurity when a particular feature is used for splitting the data at a node in the RF algorithm. The MDA is based on the decrease in accuracy when a particular feature is permuted randomly. Both metrics provide a score for each feature, which can be used to rank the features in order of importance (Hong et al., 2016; Ma et al., 2020). In this paper, the average value of MDG and MDA output for each feature is considered as its importance based on the RF algorithm.

XGBoost

Extreme Gradient Boosting, known as XGBoost, is another tree-based ML algorithm that has the ability to identify the importance of different features in predicting the target variable. XGBoost, as an ensemble method, combines multiple weak learners to create a strong learner that can make accurate predictions. Similar to Rf, XGBoost provides a built-in method for calculating feature importance based on the contribution of each feature to the decision trees in the ensemble. It uses a metric called "gain" to measure the importance of each feature. The gain represents the improvement in the model's loss function that is achieved by splitting the data based on a particular feature. Features with higher gain values are considered more important (Chen et al., 2020; Li et al., 2022).

Recursive Feature Elimination (RFE)

RFE algorithm works by recursively removing features from a dataset and re-fitting the model until the desired number of features is reached. RFE is a general feature selection method and can be used with different types of models, such as linear regression, logistic regression, and Support Vector Machine (SVM). The basic idea behind RFE is to fit a model on the entire dataset and rank the importance of each feature based on their contribution. In this study, the SVM-RFE method is used as an RFE-based method benefiting from the SVM capability to deal with non-linear data (Cai et al., 2018; Dhal and Azad, 2022).

Mutual Information (MI)

MI is a statistical method used to quantify the relationship between two variables. In the context of feature importance analysis, it can be used to determine how much information about the target variable can be gained from a particular feature. Features with high MI values are considered to be more important, as they provide more information about the target variable. In the case of MI as a filter-based method, the feature ranking is based on the strength of the relationship between each feature and the target variable, measured using the MI score (Wang et al., 2019).

Results and discussions

At the first step of driving factors analysis, the correlation between the different features was mapped as shown in Figure 3. According to this map:

- Accessibility-related features are highly correlated, which means that the lands that are closer to some amenities are usually close to the other amenities as well. On the other hand, lands that are closer to natural landscapes have a negative correlation with accessibility-related factors.
- There is a high correlation between topographic and environmental features and the location of the lands. To be more specific, latitude and elevation are highly correlated, which is rational since lower latitudes are closer to sea level in Melbourne Metropolitan. Also, rainfall and slope are correlated with longitude.

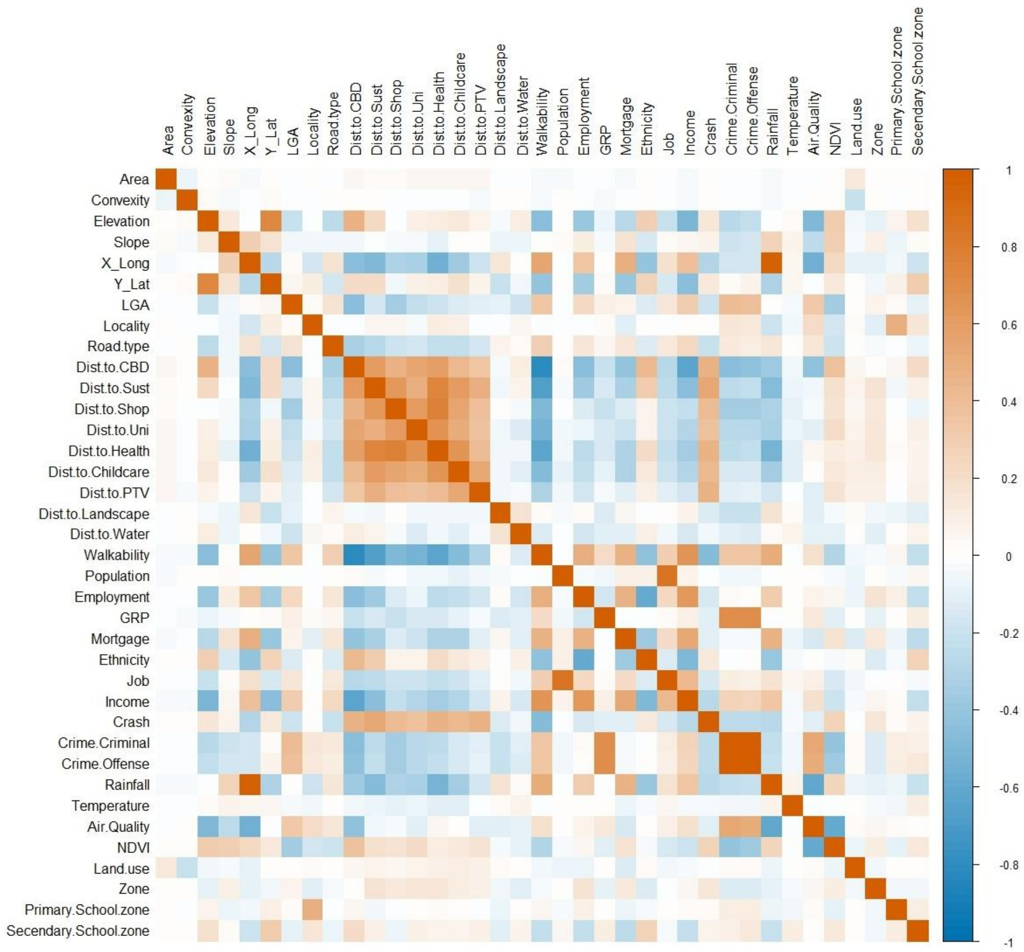

Figure 3. Correlation map between the different 37 features.

- There is a positive correlation between the school zones and the localities and LGAs that demonstrates the importance of schools in the different suburbs.
- Walkability has a high negative correlation with distance to CBD, which means that the lands closer to the city center have a lower walk score.
- Two crime-related features are highly correlated, which means that areas with higher criminal activities are associated with higher offense incidents.

Next, the importance of the different features was calculated using the four algorithms of RF, XGBoost, RFE and MI. Figure 4 presents the comparison between the normalized importance values of the different features based on the four methods. The ranking in this figure is based on alphabetical letters.

- In terms of RF, land area, distance to natural landscapes, distance to public transportation (PTV), distance to water bodies and crashes are the main factors affecting land value in the Melbourne Metropolitan. Conversely, ethnicity, GRP, crime features and LGA have the less importance in terms of land value.
- XGBoost ranks the most important features as land area, longitude, primary school zone, latitude and walkability, respectively. Contrarily, it highlights the low importance of ethnicity, crime features, distance to PTV and job.
- According to RFE, land use, latitude, distance to CBD, land area, and air quality are respectively the most important driving factors of land prices, and ethnicity, distance to shopping centers, health centers and water bodies and population have a very low effect on land value in the study area.
- MI also ranks the importance of the factors by identifying land use, LGA, latitude, criminal incidents and zone at the top, and temperature, crash, slope, distance to PTV and GRP at the bottom of the list.

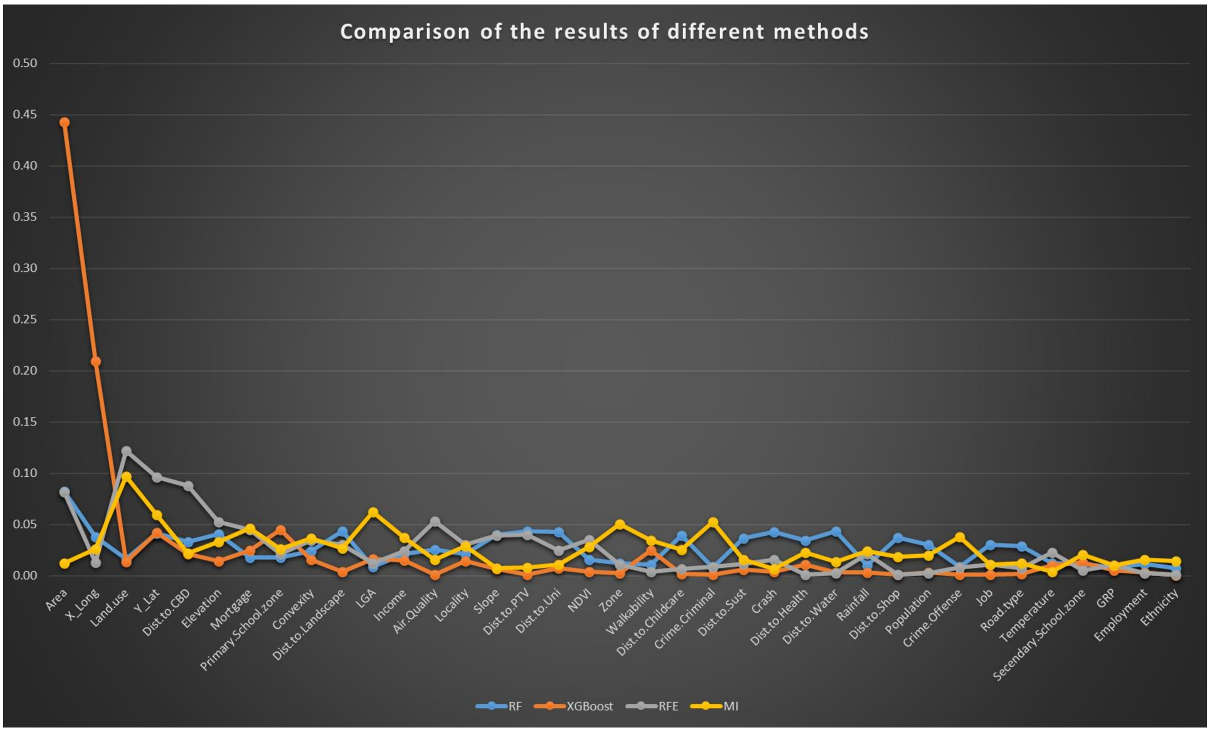

Figure 4. Importance values of the 37 features based on the different algorithms.

Based on the results of the different methods, it is clear that each method has its own specific procedures for the feature importance determination. There are many differences regarding the ranking of the most important and the least important features among the four methods. These differences are so significant that, in some cases, their results are completely opposite. Accordingly, it could be discussed that integrating the different algorithms through an ensemble method is inevitable to find an optimized feature importance ranking. The values related to the importance of each feature using each method were based on different metrics. Accordingly, in order to ensemble all the four importance values for each factor based on each method, it was required to normalize the importance values of the features based on each method. Then, through averaging, one final importance value was assigned to each factor. The final importance of each factor based on the applied ensemble method is presented in Figure *5*.

According to the results of the applied ensemble method, land area, longitude, land use, latitude, distance to Central Business District (CBD), elevation, mortgage rate and primary school zone have the highest impact on the land value in the study area.

On the other hand, ethnicity, employment, GRP, secondary school zone and temperature are associated with the least importance values in terms of land price in the Melbourne Metropolitan. With a close look at the final ranking, some key points can be discussed:

- The land area has the highest impact on the land price in Melbourne, Australia. This is reasonable since the valuation of land in the study area is based on considering the whole package of each land polygon (not square meter).
- Three factors related to location are deeply influential in land valuation, and this proves the theory of many experts on the high importance of location-based features in the valuation of real estate assets.
- Two of the considered environmental factors, including air quality and NDVI, are among the first 20 important features. This can confirm our hypothesis about the importance of such factors. It seems that previous studies' failure to take these features into account has caused uncertainties in their carried-out automatic land valuations.

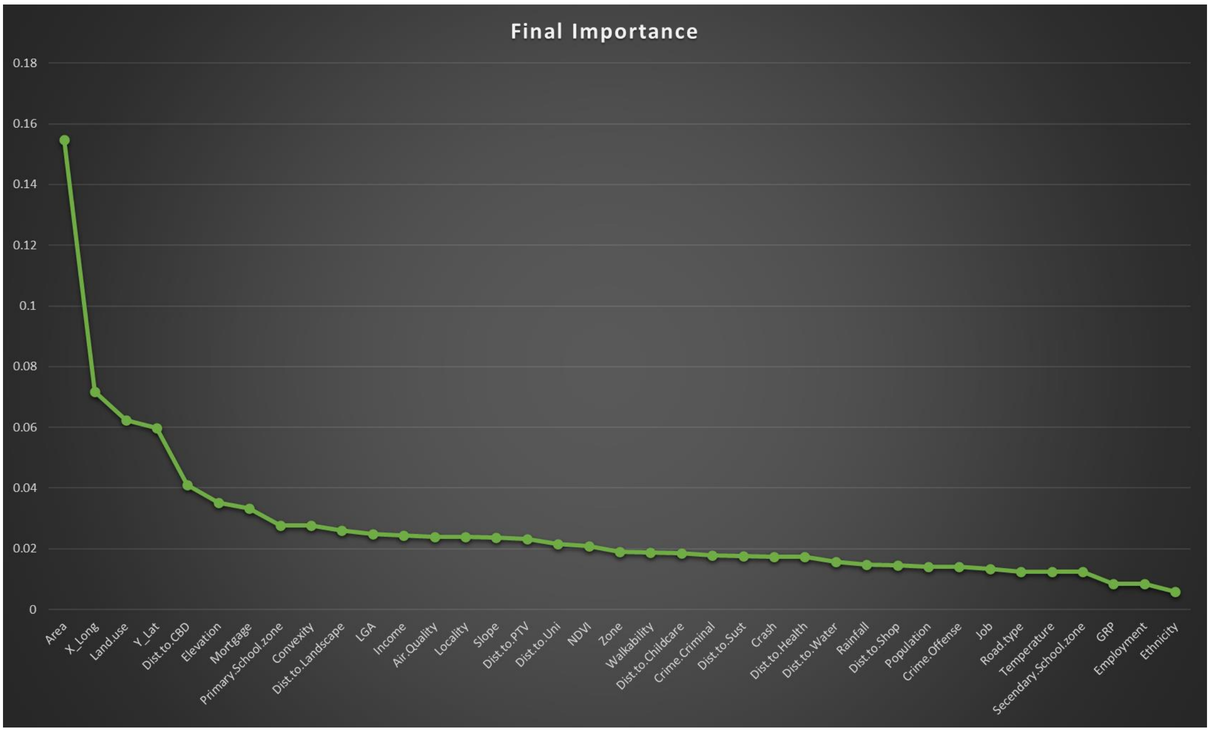

Figure 5. Final importance values of the different features.

Conclusions

This study adopts a wide range of driving factors of land value from physical, geographical, socio-economic, environmental and legal aspects to identify the most important features affecting land price in Melbourne, Australia. Considering the capabilities of ML techniques in handling non-linear relationships between different value-related features, the application of these techniques for feature importance analysis was considered. However, due to the differences in nature and structure of the different ML-based models, as well as the specific characteristics of the input dataset, the rankings of the features through different models enormously vary. To tackle this problem, an ensemble method of different ML-based techniques of RF, XGBoost, RFE and MI was applied to achieve a more robust feature importance ranking model.

Feature importance analysis in this study acknowledges the key role of location and geographical features in land valuation, as previously highlighted by different scholars. Moreover, this study recommends consideration of environmental features such as air quality and green vegetation for mass valuation of lands in large areas like Melbourne Metropolitan, where such features vary in different areas. The analyzed factors in this study and the results could be later used for improving the accuracy of the automatic land valuation models based on advanced techniques in ML. Furthermore, this paper contributes to research on 3D property valuations using cost approach and BIM through the identification of driving factors affecting land value as an important component in the cost approach. Future work in the field can also focus on adding visual features that can be extracted from satellite images through computer vision and deep learning to the current set of features and analyze their possible impacts on the land value.

Acknowledgements

This research is supported by Building 4.0 CRC. The support of the Commonwealth of Australia through the Cooperative Research Centre Programme is acknowledged.

Resources

ABS, 2023. Australian Bureau of Statistics (ABS).
Arcuri, N., De Ruggiero, M., Salvo, F., Zinno, R., 2020a. Automated Valuation Methods through the Cost Approach in a BIM and GIS Integration Framework for Smart City Appraisals, Sustainability, p. 7546.
Arcuri, N., Ruggiero, M., Salvo, F., Zinno, R., 2020b. Automated Valuation Methods through the Cost Approach in a BIM and GIS Integration Framework for Smart City Appraisals. Sustainability 12, 7546.
AURIN, 2023. Australian Urban Research Infrastructure Network (AURIN).
Baur, K., Rosenfelder, M., Lutz, B., 2023. Automated real estate valuation with machine learning models using property descriptions. Expert Systems with Applications 213, 119147.
Bureau-of-Meteorology, 2023.
Cai, J., Luo, J., Wang, S., Yang, S., 2018. Feature selection in machine learning: A new perspective. Neurocomputing 300, 70-79.
Carranza, J.P., Piumetto, M.A., Lucca, C.M., Da Silva, E., 2022. Mass appraisal as affordable public policy: Open data and machine learning for mapping urban land values. Land Use Policy 119, 106211.
Chen, C., Zhang, Q., Yu, B., Yu, Z., Lawrence, P.J., Ma, Q., Zhang, Y., 2020. Improving protein-protein interactions prediction accuracy using XGBoost feature selection and stacked ensemble classifier. Computers in Biology and Medicine 123, 103899.
Clark, L.A., Pregibon, D., 2017. Tree-based models, Statistical models in S. Routledge, pp. 377-419.
Couto, P., Silva, M., Salvado, A., 2021. Real-Estate Valuation Based on BIM Methodology, pp. 15-19.
CSA, 2023. Crime Statistics Agency (CSA) - Crime statistics.
DELWP, 2023. DataShare - A search and discovery tool that enables assessment of DELWP's Spatial Data resources. Department of Environment, Land, Water and Planning (DELWP) - The Victorian Government
Deng, F., Huang, Y., 2004. Uneven Land Reform and Urban Sprawl in China: The Case of Beijing. Progress in Planning 61, 211-236.
Department-of-Transport, 2023. Road Crashes for five Years Victoria.
Derdouri, A., Murayama, Y., 2020. A comparative study of land price estimation and mapping using regression kriging and machine learning algorithms across Fukushima prefecture, Japan. Journal of Geographical Sciences 30, 794-822.
Dhal, P., Azad, C., 2022. A comprehensive survey on feature selection in the various fields of machine learning. Applied Intelligence 52, 4543-4581.
El Yamani, S., Ettarid, M., Hajji, R., 2019. Building Information Modeling Potential for an Enhanced Real Estate Valuation Approach based on the Hedonic Method. WIT Transactions on the Built Environment 192, 305-316.
EPA, 2023. EPA AirWatch.
Hong, H., Xiaoling, G., Hua, Y., 2016. Variable selection using Mean Decrease Accuracy and Mean Decrease Gini based on Random Forest, 2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS), pp. 219-224.
Hong, J., Choi, H.H., 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, 140-152.
Hu, S., Cheng, Q., Wang, L., Xu, D., 2013. Modeling land price distribution using multifractal IDW interpolation and fractal filtering method. Landscape and Urban Planning 110, 25-35.
Jafary, P., Shojaei, D., Rajabifard, A., Ngo, T., 2022a. BIM and real estate valuation: challenges, potentials and lessons for future directions. Engineering, Construction and Architectural Management ahead-of-print.
Jafary, P., Shojaei, D., Rajabifard, A., Ngo, T., 2022b. A FRAMEWORK TO INTEGRATE BIM WITH ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING-BASED PROPERTY VALUATION METHODS. ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci. X-4/W2-2022, 129-136.
Jalali, D., MacDonald, H., Fini, A.A.F., Shi, S., 2022. Effects of planning regulations on housing and land markets: A system dynamics modeling approach. Cities 126, 103670.
Khodabakhshian, A., Toosi, H., 2021. Residential Real Estate Valuation Framework Based on Life Cycle Cost by Building Information Modeling. Journal of Architectural Engineering 27, 04021020.
Kim, J., Won, J., Kim, H., Heo, J., 2021. Machine-Learning-Based Prediction of Land Prices in Seoul, South Korea, Sustainability, p. 13088.
Kim, K.S., Park, S.J., Kweon, Y.-J., 2007. Highway traffic noise effects on land price in an urban area. Transportation Research Part D: Transport and Environment 12, 275-280.
L. Krause, A., Bitter, C., 2012. Spatial econometrics, land values and sustainability: Trends in real estate valuation research. Cities 29, S19-S25.
Li, H., Shi, L., Gao, W., Zhang, Z., Zhang, L., Zhao, Y., Wang, G., 2022. dPromoter-XGBoost: Detecting promoters and strength by combining multiple descriptors and feature selection using XGBoost. Methods 204, 215-222.
Ma, J., Cheng, J.C.P., Jiang, F., Chen, W., Zhang, J., 2020. Analyzing driving factors of land values in urban scale based on big data and non-linear machine learning techniques. Land Use Policy 94, 104537.

Mirkatouli, J., Samadi, R., Hosseini, A., 2018. Evaluating and analysis of socio-economic variables on land and housing prices in Mashhad, Iran. Sustainable Cities and Society 41, 695-705.
OSM, 2023. Open Street Map (OSM)
R-Core-Team, 2023. R: A language and environment for statistical computing., R Foundation for Statistical Computing, Vienna, Austria.
REMPLAN, 2023. Economy, Jobs and Business Insights.
RStudio-Team, 2023. RStudio: Integrated Development Environment for R., RStudio, PBC, Boston, MA.
Salcedo-Sanz, S., Cornejo-Bueno, L., Prieto, L., Paredes, D., García-Herrera, R., 2018. Feature selection in machine learning prediction systems for renewable energy applications. Renewable and Sustainable Energy Reviews 90, 728-741.
Sampathkumar, V., Santhi, M.H., Vanjinathan, J., 2015. Forecasting the Land Price Using Statistical and Neural Network Software. Procedia Computer Science 57, 112-121.
Silveira, P., Dentinho, T.P., 2018. A spatial interaction model with land use and land value. Cities 78, 60-66.
Solano, E.S., Dehghanian, P., Affonso, C.M., 2022. Solar Radiation Forecasting Using Machine Learning and Ensemble Feature Selection, Energies.
Tsutsumi, M., Shimada, A., Murakami, D., 2011. Land price maps of Tokyo Metropolitan Area. Procedia - Social and Behavioral Sciences 21, 193-202.
Walk-Score, 2023. Walk Score - Cities in Victoria.
Wang, X., Guo, B., Shen, Y., Zhou, C., Duan, X., 2019. Input Feature Selection Method Based on Feature Set Equivalence and Mutual Information Gain Maximization. IEEE Access 7, 151525-151538.
Wen, H., Goodman, A.C., 2013. Relationship between urban land price and housing price: Evidence from 21 provincial capitals in China. Habitat International 40, 9-17.
Yilmazer, S., Kocaman, S., 2020. A mass appraisal assessment study using machine learning based on multiple regression and random forest. Land Use Policy 99, 104889.
Zhang, P., Hu, S., Li, W., Zhang, C., Yang, S., Qu, S., 2021. Modeling fine-scale residential land price distribution: An experimental study using open data and machine learning. Applied Geography 129, 102442.