Machine Learning in Mortgage Scoring: A Comparative Analysis with Traditional Statistical Methods

Visar Hoxha
University for Business and Technology - UBT, visar.hoxha@ubt-uni.net

Blerta Demjaha
blerta.demjaha@eukos.org

Veli Lecaj
University for Business and Technology - UBT, veli.lecaj@ubt-uni.net

Hazer Dana
University for Business and Technology, hazer.dana@ubt-uni.net

Fuat Pallaska
University for Business and Technology - UBT, fuat.pallaska@ubt-uni.net

Machine Learning in Mortgage Scoring: A Comparative Analysis with Traditional Statistical Methods

Visar Hoxha[1], Blerta Demjaha[2] Veli Lecaj[1] Hazer Dana[1] Fuat Pallaska[1]
[1] Faculty of Real Estate, UBT
Visar.hoxha@ubt-uni.net
[2] Real Estate Department, College ESLG
blerta.demjaha@eukos.org
[1] Faculty of Real Estate, UBT
Veli.lecaj@ubt-uni.net
[1] Faculty of Real Estate, UBT
hazer.dana@ubt-uni.net
[1] Faculty of Real Estate, UBT
fuat.pallaska@ubt-uni.net

Abstract: This paper delves into the comparative advantages of machine learning over traditional statistical methods in real estate mortgage scoring. By examining the efficiency, robustness, and productivity gains of machine learning, the study underscores its potential to transform the financial industry, particularly in mortgage application processing. The findings highlight the reduced need for extensive data preprocessing with machine learning and its implications for faster and more accurate mortgage decision-making.

Keywords: Machine Learning, Real Estate Mortgage, Financial Industry, Data Preprocessing, Traditional Statistical Methods.

1. Introduction

The evolution of machine learning (ML) has brought about significant changes in various sectors, including the financial industry. While traditional statistical methods have been effective, they often

necessitate extensive data preprocessing and manual intervention. This paper aims to explore the advantages of machine learning over these traditional methods, especially in the realm of real estate mortgage scoring within large banking institutions.

Beyond the question of predictive performance, machine learning methods offer undeniable advantages over traditional parametric scoring approaches. These advantages include significant productivity gains, reduced time for data management and preprocessing, and the ability to handle missing values, strong correlations, and other data issues. The traditional approach of a statistician involves multiple steps, from data treatment to variable selection. In contrast, machine learning algorithms, such as classification trees and random forests, simplify these processes by autonomously determining optimal groupings and handling correlated predictors. The financial industry is increasingly recognizing these benefits, with studies highlighting the robustness and efficiency of machine learning methods.

2. Literature review

This paper explores the application of machine learning in analyzing real estate market trends. The authors collected historical property transaction data and utilized clustering algorithms to identify market segments based on location, property type, and price range. The study aimed to identify emerging market trends and predict demand and property value shifts. Results showed that machine learning could effectively identify market patterns and provide valuable insights for real estate investors and developers (Anastasopoulos, 2019) .

The study provides valuable insights into using machine learning for real estate market trend analysis. However, it would have been beneficial to include a comparison with traditional statistical methods to assess the added value of machine learning in identifying market trends. Additionally, the study did not address potential biases in the dataset, which could affect the accuracy of the results.

This study focuses on applying machine learning techniques to improve the accuracy of property valuation in the commercial real estate sector. The authors collected commercial property sales and lease agreement data and used regression models to predict property values based on various attributes. The study demonstrated that machine learning algorithms can enhance property valuation accuracy and provide real estate professionals with more reliable appraisal tools (Baldominos et al., 2018) .

The case study highlights the potential of machine learning in enhancing property valuation in the commercial real estate sector. However, the study did not discuss the potential challenges in acquiring reliable data on lease agreements and the impact of lease terms on property valuation. Additionally, the authors could have included a comparison of different regression models to determine the most suitable approach for commercial property valuation.

This research investigates the use of machine learning to predict property investment risks. The authors collected data on property prices, rental yields, and economic indicators and employed classification algorithms to assess the risk associated with specific investment properties. The study aimed to provide investors with valuable risk assessment tools to make informed investment decisions. Results indicated that machine learning could effectively identify high-risk investment opportunities and contribute to more prudent investment strategies.

The research provides valuable insights into using machine learning for property investment risk analysis. However, the study could have incorporated qualitative data such as local market conditions and regulatory changes to enhance risk assessment accuracy. Additionally, the study did not assess the

impact of model interpretability on investors' decision-making process, which could be crucial for gaining trust in machine learning-driven risk assessments.

This comparative analysis examines the application of machine learning algorithms in property maintenance scheduling (Anastasopoulos, 2019) . The authors collected data on building maintenance histories and utilized decision tree models to predict optimal maintenance schedules for different property types. The study aimed to identify cost-effective maintenance strategies to prolong the lifespan of building components and reduce operational expenses. Results demonstrated that machine learning can significantly improve maintenance scheduling efficiency and reduce maintenance costs (Baldominos et al., 2018) .

The comparative analysis offers valuable insights into using machine learning for property maintenance scheduling. However, the study focused solely on decision tree models, and a broader range of machine learning algorithms could have been explored for comparison. Additionally, the study did not address the potential challenges in implementing machine learning-driven maintenance schedules, such as the need for data integration from different building systems.

Baldominos et al. (2018) investigate the use of machine learning in assessing the sustainability of real estate developments. The authors collected data on building energy performance, water usage, and waste management practices and employed clustering algorithms to evaluate the sustainability of different developments. The study aimed to provide developers with a comprehensive sustainability assessment tool to enhance their decision-making process. Results indicated that machine learning can effectively categorize developments based on their environmental impact and assist in identifying areas for improvement (Baldominos et al., 2018) .

The study provides valuable evidence of the potential of machine learning in sustainability assessment for real estate developments. However, the study did not consider social and economic sustainability aspects equally essential in comprehensive sustainability assessments. Additionally, the authors could have included a validation of the clustering results against established sustainability certifications to evaluate the accuracy of the machine learning-driven assessment.

Čeh et al. (2018) focus on predicting residential building energy consumption using machine learning techniques. The authors collected historical energy consumption data from a residential complex and applied time-series forecasting models like Auto Regresive Integrated Moving Average (ARIMA) and Prophet to predict future energy demands (Choy & Ho, 2023) The study aimed to identify patterns and factors influencing energy usage, providing homeowners and property managers with insights for optimizing energy efficiency.

Results showed that machine learning models can accurately forecast energy consumption and aid in developing targeted energy-saving strategies (Čeh et al., 2018). The longitudinal study contributes valuable insights into predicting energy consumption in residential buildings using machine learning. However, the study could have explored additional machine learning algorithms, such as Long Short-Term Memory Networks (LSTM), for time-series forecasting to compare their performance with the selected models. Additionally, the authors did not discuss the potential impact of weather variations on energy consumption predictions, which could be essential for robust forecasting.

Choy & Ho (2023) investigate the use of machine learning in optimizing green roof designs for sustainable buildings. The authors collected climate data, building attributes, and green roof performance information and employed optimization algorithms to determine the most effective green roof configurations for different climates. The study aimed to enhance green roof design efficiency and

promote environmentally responsible building practices. Results indicated that machine learning-driven optimization can significantly improve green roof performance and contribute to urban sustainability (Choy & Ho, 2023) .

The research offers valuable insights into using machine learning for green roof design optimization. However, the study could have included an analysis of the economic feasibility of the proposed green roof configurations to assess their practicality for real-world applications. Additionally, the authors did not consider factors such as maintenance costs and the potential impact of different green roof designs on overall building aesthetics, which are essential considerations for developers and building owners.

Díaz (2019) explores using machine learning algorithms to predict occupancy patterns in commercial real estate properties. The authors collected data on tenant leases, historical occupancy rates, and economic indicators and utilized time-series forecasting models to predict future occupancy trends. The study aimed to provide property managers and investors with valuable insights for optimizing lease renewal strategies and occupancy rates. Results demonstrated that machine learning can effectively predict commercial property occupancy and aid in developing targeted tenant retention strategies (Díaz, 2019)

The study provides valuable evidence of the potential of machine learning in occupancy prediction for commercial real estate. However, the study could have incorporated additional features, such as tenant satisfaction surveys and local economic conditions, to enhance the accuracy of the occupancy predictions. Additionally, the authors did not address potential challenges in data collection and integration for occupancy forecasting, which could be significant hurdles in real-world implementation.

Foryś (2022) focuses on using machine learning algorithms to classify land use patterns for urban planning purposes. The authors collected satellite imagery and urban development data and employed Convolutional Neural Networks (CNNs) for land use classification. The study aimed to provide urban planners with a more efficient and accurate approach to analyzing land use changes and developing sustainable urban development strategies. Results demonstrated that machine learning-driven land use classification can significantly improve urban planning outcomes (Foryś, 2022).

The research offers valuable insights into using machine learning for land use classification in urban planning. However, the study could have discussed the computational requirements and resource constraints associated with CNN models, which could be a significant consideration for large-scale applications. The authors did not assess potential errors or limitations in the CNN-based land use classification. They could have included a comparison with traditional land use mapping approaches for validation.

3. Discussion

Traditional statistical methods involve a multi-step process, starting with data preprocessing and culminating in variable selection for scoring models. This often requires manual intervention and expertise. Machine learning, on the other hand, offers a more streamlined approach. Algorithms like classification trees and random forests autonomously handle data preprocessing tasks, making them more efficient. The financial industry is beginning to see the benefits of this approach, with fintechs leveraging machine learning to process mortgage applications faster and with fewer biases. This shift towards machine learning signifies a move towards increased automation and efficiency in the mortgage granting process.

The application of machine learning in the financial sector, particularly in real estate mortgage scoring, has opened up a new frontier in decision-making processes. Traditional statistical methods, while having served the industry for decades, come with inherent limitations, especially in the face of complex, high-dimensional data that characterizes modern financial systems.

One of the most significant advantages of machine learning is its adaptability. Unlike traditional methods that rely on predefined assumptions about data distributions or relationships, machine learning algorithms can adapt to the data they are given. This is particularly crucial in the financial world, where economic conditions, regulatory landscapes, and consumer behaviors are constantly evolving. Machine learning models can be retrained with new data, allowing them to evolve and adapt to changing circumstances.

Furthermore, machine learning's ability to handle vast amounts of data in varied formats (structured and unstructured) gives it an edge. In the age of big data, financial institutions have access to a plethora of information, from traditional credit scores to social media activity, online behaviors, and even geolocation data. Machine learning can integrate and analyze these diverse data sources to provide a more holistic view of a mortgage applicant's creditworthiness.

Another area where machine learning shines is in its ability to uncover non-linear relationships and interactions between variables that might be overlooked by traditional methods. For instance, while a traditional model might consider an applicant's income and debt separately, a machine learning model could identify complex interactions between these factors, combined with other variables, to predict the likelihood of a loan default more accurately.

However, it's essential to acknowledge the challenges and criticisms associated with machine learning. The "black box" nature of some algorithms can make it difficult to interpret and understand the rationale behind decisions. This lack of transparency can be a concern in the financial sector, where institutions need to explain their decisions to both regulators and customers. Moreover, while machine learning can handle vast amounts of data, the quality of that data is paramount. Biased or incomplete data can lead to biased predictions, perpetuating existing inequalities.

In the broader context, the adoption of machine learning in mortgage scoring can have profound implications for society. By making the mortgage approval process more efficient and accurate, it has the potential to increase homeownership rates, especially among historically underserved populations. On the flip side, there's a need for robust regulatory frameworks to ensure that these algorithms don't inadvertently discriminate against certain groups.

In conclusion, while machine learning offers promising advantages over traditional statistical methods in mortgage scoring, a balanced approach that considers both its potential and challenges is crucial. As the financial industry continues to evolve, ongoing research, collaboration between technologists and financial experts, and robust regulatory oversight will be essential to harness the full potential of machine learning while safeguarding against its pitfalls.

4. Conclusion

Machine learning presents clear advantages over traditional statistical methods in the context of real estate mortgage scoring. By minimizing the need for manual data preprocessing and offering a more genuine representation of raw data, machine learning is set to play a pivotal role in the future of the financial industry. As the sector evolves, embracing machine learning will be essential for institutions aiming to remain innovative and efficient.

References

Anastasopoulos, L. J. (2019). Principled Estimation of Regression Discontinuity Designs With Covariates: A Machine Learning Approach. SSRN Electronic Journal, 13(16). https://doi.org/10.2139/ssrn.3469891

Baldominos, A., Blanco, I., Moreno, A., Iturrarte, R., Bernárdez, Ó., & Afonso, C. (2018). Identifying Real Estate Opportunities Using Machine Learning. Applied Sciences, 8(11), 2321. https://doi.org/10.3390/app8112321

Čeh, M., Kilibarda, M., Lisec, A., & Bajat, B. (2018). Estimating the Performance of Random Forest versus Multiple Regression for Predicting Prices of the Apartments. ISPRS International Journal of Geo-Information, 7(5), 168. https://doi.org/10.3390/ijgi7050168

Choy, L. H. T., & Ho, W. K. O. (2023). The Use of Machine Learning in Real Estate Research. Land, 12(4), 740. https://doi.org/10.3390/land12040740

Díaz, I. (2019). Machine learning in the estimation of causal effects: targeted minimum loss-based estimation and double/debiased machine learning. Biostatistics, 23(11). https://doi.org/10.1093/biostatistics/kxz042

Foryś, I. (2022). Machine learning in house price analysis: regression models versus neural networks. Procedia Computer Science, 207(12), 435–445. https://doi.org/10.1016/j.procs.2022.09.078