Oct 28th, 8:00 AM - Oct 29th, 6:00 PM

Revolutionizing Real Estate Mortgage Scoring: The Superiority of Machine Learning Over 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

Recommended Citation

Hoxha, Visar; Demjaha, Blerta; Lecaj, Veli; Dana, Hazer; and Pallaska, Fuat, "Revolutionizing Real Estate Mortgage Scoring: The Superiority of Machine Learning Over Traditional Statistical Methods" (2023). *UBT International Conference*. 6.
https://knowledgecenter.ubt-uni.net/conference/IC/LAW/6

Revolutionizing Real Estate Mortgage Scoring: The Superiority of Machine Learning Over 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 explores the advantages of machine learning over traditional statistical methods in the context of real estate mortgage scoring. While traditional methods require extensive data preprocessing, machine learning offers a more streamlined and efficient approach. The financial industry is recognizing these benefits, with machine learning enabling faster mortgage application processing and reduced modeling biases. The findings underscore the potential of machine learning to revolutionize the financial sector.

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

1. Introduction

The evolution of machine learning (ML) has revolutionized various sectors, including the financial industry. Traditional statistical methods, while effective, often require extensive data preprocessing and manual intervention. Machine learning, with its ability to autonomously process and analyze data, offers a promising alternative. This paper delves into the advantages of machine learning over traditional statistical methods, particularly in the context of real estate mortgage scoring within large banks.

2. Literature review

Beyond the question of predictive performance, machine learning methods have an undeniable advantage over the usual parametric scoring approaches since they allow significant productivity gains. In particular, machine learning algorithms make it possible to reduce the time devoted to the data management and preprocessing stages before the modeling stage in a strict sense (Milunovich, 2019) [1]. Of course, this does not mean that machine learning makes it possible to dispense with the work of construction and data quality control, which remains necessary.

To fully understand this point, let's return to the traditional approach of a statistician in charge of building a scoring model real estate mortgage within the risk department of a large bank. The first step of his work is to apply different treatments to the training data. Among these is the processing of missing or outlying values, which requires the implementation of detection, imputation, and exclusion procedures. The other treatments generally concern categorizing categories of the discrete explanatory variables and discretizing the continuous variables. For each of the qualitative variables, the modalities are grouped in such a way as to reduce the number of classes and maximize the discriminating power of the variable. All the continuous explanatory variables are discretized (Milunovich, 2019) [2].

On the one hand, this is to capture potential non-linear effects and, on the other hand, reduce the influence of extreme values or uncorrected outliers. The number of classes and the discretization thresholds are determined by iterative algorithms built to maximize a measurement of Cramer's V type association or chi-square statistic, between the target variable (the default) and the explanatory variable. The second step consists in analyzing the correlations between the predictors to verify that these variables are not too correlated with each other. Depending on these correlations, the expert then removes certain redundant variables according to the principle of parsimony. The third step is the selection of the explanatory variables of the score model (Milunovich, 2019) [3]. Under a given scoring model (e.g., logistic regression), we select from among all the reprocessed variables the best predict the default. Depending on the number of variables available, this selection can be made manually or using automatic approaches such as stepwise. The automatic selection is often complemented by business expertise and a finer analysis of the model (marginal effects, odds ratios).

Conversely, using a classification tree or tree-based algorithms, such as random forests, renders discretizing continuous variables and grouping categories obsolete. These techniques autonomously determine the optimal discretization and groupings of modalities (Stang et al., 2022) [4]. Analyzing correlations between predictors is less crucial because most machine learning algorithms can integrate strongly correlated predictors. Penalized regression methods such as the Lasso or the Ridge precisely make it possible to select the relevant variables and overcome multicollinearity. More generally, the advantage of machine learning algorithms is precisely to use the data to determine the optimal functional form of the model in the sense of a certain criterion. This, therefore, renders the step of selecting the explanatory variables of the score model obsolete.

[1] Milunovich, 2019

[2] Ibid

[3] Ibid

[4] Stang et al., 2022

These productivity gains associated with machine learning are now highlighted in the financial industry (Stang et al., 2022) [5]. Grennepois et al. (2018) [6] highlight the fact that the predictive performance of random forests is generally robust to the non-imputation of missing values, to the presence of strong correlations between certain explanatory variables, to the non-grouping of the modalities of the discrete variables, and to the non-discretization of the continuous variables. This robustness, therefore, potentially makes it possible to limit the preprocessing steps on the data.

Beyond productivity gains, limiting pre-processing of the data can also reduce any modeling biases since, in the end, machine learning lets the raw data do the talking. Machine learning thus allows increased automation of real estate mortgage granting processes, including in the construction phase and revision of risk models. Considering data on the processing time of mortgage applications in the United States, Fusteret al. (2018a) [7] show that Fintechs process mortgage applications around 20% faster than other lenders, and this is without a noticeable deterioration in the quality of mortgage selection.

3. Discussion

Traditional statistical approaches involve multiple stages of data preprocessing. A statistician would first treat training data, handling missing or outlying values, categorizing discrete explanatory variables, and discretizing continuous ones. The aim is to maximize the discriminating power of each variable. Subsequent steps involve analyzing correlations between predictors, removing redundant variables, and selecting the most relevant explanatory variables for the scoring model.

In contrast, machine learning offers a more streamlined approach. Algorithms like classification trees or random forests eliminate the need for discretizing continuous variables or grouping categories. They determine the optimal groupings autonomously. Moreover, machine learning algorithms can handle strongly correlated predictors, making the analysis of correlations less critical. Penalized regression methods, such as Lasso or Ridge, address multicollinearity and select relevant variables. The inherent advantage of ML is its ability to determine the optimal functional form of the model using the data, rendering the step of selecting explanatory variables obsolete.

The financial industry is beginning to recognize these productivity gains. Machine learning's robustness to various data issues, such as missing values or strong correlations between variables, reduces the need for extensive preprocessing. This not only results in productivity gains but also minimizes modeling biases, allowing for a more genuine representation of raw data. The automation capabilities of ML have also been observed in the real estate mortgage granting processes. Fintechs, leveraging ML, have been shown to process mortgage applications faster without compromising the quality of mortgage selection.

4. Conclusion

[5] Ibid

[6] Grennepois et al., 2018

[7] Fusteret al., 2018a

Machine learning offers undeniable advantages over traditional statistical methods, especially in the realm of real estate mortgage scoring. By reducing the need for extensive data preprocessing and offering a more genuine representation of raw data, ML is poised to revolutionize the financial industry. As the sector continues to evolve, embracing machine learning will be crucial for institutions aiming to stay at the forefront of innovation and efficiency.

References:

Fuster A., Goldsmith -Pinkham P., Ramadorai T. and Walther A. (2018b), "Predictably Unequal? The Effects of Machine Learning on Credit Markets? », SSRN, Working Paper

Grennepois N. and Robin E. (2019), "Explain Artificial Intelligence for Credit Risk Management", Deloitte Risk Advisory , July.

Grennepois N., Alvirescu M. A. and Bombail M. (2018), "Using Random Forest for Credit Risk Models, Deloitte Risk Advisory , September.

Milunovich, G. (2019). Forecasting Australian Real House Price Index: A Comparison of Time Series and Machine Learning Methods. SSRN Electronic Journal, 6(56). https://doi.org/10.2139/ssrn.3417527

Stang, M., Krämer, B., Nagl, C., & Schäfers, W. (2022). From human business to machine learning—methods for automating real estate appraisals and their practical implications. Zeitschrift Für Immobilienökonomie, 13(21). https://doi.org/10.1365/s41056-022-00063-1