Journal of Financial Management Strategy
Alzahra University
Vol. 11, No. 42, Fall 2023
pp. 1-4

Research Paper

Comparison of Statistical and Machine Learning Models for Predicting Cash Holdings and Providing the Optimal Model[1]

Sajjad Mirzaei[2], Mehdi Mohammadi[3], Gholamreza Mansourfar[4]

Received: 2023/02/16 Accepted: 2023/09/24

INTRODUCTION

The significance of cash holdings cannot be overstated, as it plays a pivotal role in both personal and business finances. In personal finance, cash assets serve as a safety net during emergency situations. For businesses, maintaining cash is essential for effective cash flow management and operational financing. Positive cash flow is indispensable for the survival of any business, ensuring that the company has ample funds to cover daily expenses, pay employees, suppliers, and creditors. Additionally, liquid assets empower companies to seize opportunities,

1. DOI: 10.22051/JFM.2023.42943.2789

2. M.Sc. Department of Financial Management, Faculty of Economics and Management, Urmia University, Urmia, Iran. Email: st_s.mirzaei@urmia.ac.ir.

3. M.Sc. Department of Management Accounting, Faculty of Economics and Management, Urmia University, Urmia, Iran. Email: mehdim0719@gmail.com.

4. Associate Professor, Department of Accounting, Faculty of Economic and Management, Urmia University, Urmia, Iran. Coresponding Author. Email: g.mansourfar@urmia.ac.ir.

such as investing in new equipment or acquiring other businesses. In essence, liquidity forecasting aids managers in determining how cash can be utilized to generate more profits and safeguard the company from financial challenges. Hence, it is imperative to measure and predict the optimal amount of cash. Two common methods for cash retention forecasting are statistical models and machine learning. Statistical models, including univariate and multivariate approaches, are commonly used for modeling. While statistical models are simple to implement, they come with certain drawbacks, including numerous assumptions and limitations. Conversely, recent advancements in machine learning have resulted in the successful application of machine learning algorithms for forecasting in various domains beyond finance and accounting. Therefore, the aim of this paper is to compare the performance of machine learning models and statistical models in predicting cash retention.

MATERIALS AND METHODS

The statistical population for this research comprises all companies listed on the Tehran Stock Exchange during the period 2010-2021. The systematic elimination method was employed to select the statistical sample, resulting in the analysis of 174 companies as the research sample. The research follows a two-stage approach for cash holding forecasting. In the first stage, modeling is conducted with the entire set of research variables. In the second stage, the impact of the feature selection approach on prediction results is investigated. Additionally, efforts are made to mitigate potential collinearity issues by utilizing the variable set suggested by Lasso regression. Two statistical models, namely multivariate linear regression and generalized linear models, are employed, alongside 10 machine learning models. The machine learning models include deep learning methods, decision trees, random forests, tree gradient boosting, XGBoost, support vector regression, KNN, symbolic regression, MARS regression, and neural networks. Data analysis is carried out using the SPM algorithm, RapidMiner, Eureqa data mining software, and Stata econometric and statistical software.

RESULTS AND DISCUSSION

The results of fitting the models without using the Lasso regression feature selection approach revealed that the highest accuracy coefficients were associated with symbolic regression models using genetic algorithms, MARS regression, Support Vector Regression, tree gradient boosting, neural network, and XGBoost, in that order. Other models exhibited accuracy coefficients below 50%, indicating poor performance and lower rankings. On the other hand, the results of estimating the models using the Lasso regression feature selection approach demonstrated that the highest accuracy coefficients were achieved by symbolic regression, MARS regression, and tree gradient boosting models.

CONCLUSION

The results of fitting the models indicate that the symbolic regression machine learning model, utilizing the genetic algorithm with an accuracy coefficient of 70%, exhibits the best performance among both statistical and machine learning models. This superiority may be attributed to the model's ability to explore a wide space of probabilistic models, increasing the likelihood of finding a model that fits the data well. The MARS regression machine learning algorithm ranks second, possibly due to its high flexibility in handling input data. The newer reinforcement algorithms of machine learning, such as gradient boosting, demonstrated higher accuracy, as expected, given their reinforcement processes and the research context. However, their high complexity did not result in significantly higher accuracy compared to models with medium complexity, which fall in the middle range of models investigated in this study. Their placement below symbolic regression and MARS may be influenced by the limited amount of data in this study. Among the two statistical models examined, the generalized linear model outperformed the linear regression model, possibly due to its greater suitability with non-normally distributed data. While the majority of machine learning models exhibited higher performance than statistical models, some machine learning algorithms were less accurate than their statistical counterparts. The simplicity of the structure in statistical forecasting models may allow these parametric methods to

achieve suitable and higher forecasting accuracy compared to certain machine learning models.

Keywords: Lasso Regression, Machine Learning, Predict Cash Holdings.

JEL Classification: C52, C53, G32.