FORECASTING STOCK MARKET LIQUIDITY WITH MACHINE LEARNING: AN EMPIRICAL EVALUATION IN THE GERMAN MARKET

Bogdan Ionut ANGHEL, PhD Candidate*

Abstract

The study benchmarks four machine-learning algorithms—Random Forest, XGBoost, CatBoost and Long Short-Term Memory (LSTM) networks—for forecasting stock market liquidity in Germany's DAX equity market. Using data from January 2006 to May 2025, a Liquidity Score is constructed as a turnover-to-volatility ratio, designed to penalize wide intraday price swings while rewarding active trading behavior. This metric captures key microstructural aspects of liquidity and serves as the dependent variable throughout the analysis. It is paired with 41 independent variables that capture volatility, price ranges, return dynamics, technical indicators and cross-asset linkages. Empirical testing shows that the two gradient-boosting ensembles consistently outperform both Random Forest and the LSTM model, tracking sudden liquidity swings more accurately and delivering the tightest forecast errors. The evidence highlights (i) the practical superiority of tree-based boosting for high-frequency liquidity prediction, (ii) the value of rich, carefully engineered feature sets in modelling non-linear market micro-structure effects and (iii) the limitations of standard LSTM architectures when financial sequences are short and noisy. The findings offer actionable insights for traders, treasurers and regulators seeking real-time early-warning indicators of liquidity stress in European blue-chip equities.

Keywords: stock market, German equity market, liquidity, machine learning, time series forecasting

JEL Classification: G12, G15, C5, C8, C45, C53

* *Faculty of International Business and Economics, Bucharest University of Economic Studies, Bucharest, Romania. Corresponding author, e-mail: bogdan.anghel@rei.ase.ro*

1. Introduction

Sharp contractions in market liquidity can intensify price volatility, raise transaction costs, and transmit financial stress across asset classes, making short-term liquidity forecasting essential for traders, treasurers, and regulators. However, such forecasting remains inherently difficult due to the complex, non-linear, and regime-dependent interactions among order-flow variables, price dynamics, and cross-asset signals—features that traditional linear models are often ill-equipped to capture. Machine learning (ML) techniques offer a promising alternative by uncovering intricate patterns in high-dimensional, noisy datasets, yet their comparative effectiveness in forecasting liquidity in major European equity markets remains underexplored, with few studies employing standardized, real-time evaluation frameworks.

Against this backdrop the present study poses a single, guiding research question: Which of four widely used ML algorithms—Random Forest, XGBoost, CatBoost and Long Short-Term Memory network—provides the most accurate and robust day-ahead forecasts of DAX equity-market liquidity?

To answer this question the paper constructs a turnover-to-volatility Liquidity Score for the German blue-chip index and pairs it with forty-one predictors grouped into five conceptual blocks: volatility metrics, price-range measures, return dynamics, technical indicators and cross-asset signals. Each algorithm is tuned and assessed within an identical walk-forward framework that preserves the chronological order of observations and replicates the constraints of real-time deployment. The resulting head-to-head comparison allows the study to isolate the contribution of model architecture from that of feature engineering.

The article is structured as follows. Section 2 surveys the existing literature on liquidity forecasting and data-driven financial modelling, highlighting unresolved issues. Section 3 describes the data, defines the Liquidity Score and details the feature-engineering pipeline. Section 4 sets out the modelling framework, hyper-parameter tuning strategy and validation design. Section 5 reports empirical results, interprets comparative performance and discusses practical implications. Section 6 concludes, outlining limitations and suggesting directions for future research.

2. Literature review

The forecasting of liquidity in capital markets has become increasingly significant in the wake of heightened financial volatility and systemic risk. Over the past decade, the emergence of machine learning (ML) techniques has introduced sophisticated approaches capable of capturing the non-linear, dynamic patterns inherent in liquidity measures. This literature review synthesises findings from recent academic contributions on the application of ML to liquidity forecasting, highlighting trends, methodological advancements, and key challenges.

Traditional econometric models for liquidity forecasting have often relied on linear assumptions and limited feature sets. However, ML algorithms such as random forests, support vector machines, and deep learning architectures have demonstrated superior predictive capabilities, particularly in high-dimensional or noisy data environments (Guerra et al., 2022; Antony & Kumar, 2024).

Kirkby and Andrean (2024) applied supervised ML algorithms to forecast bid-ask spreads in foreign exchange markets, revealing that such models can effectively anticipate microstructural changes in market liquidity. Similarly, Cabrol et al. (2024) explored ML-based forecasting of bond illiquidity, noting that models incorporating non-linear interactions between fundamental variables outperformed benchmark methods.

From a regulatory standpoint, ML models offer promising tools for risk detection and early warning. Triepels et al. (2021) applied recurrent neural networks (RNNs) to monitor high-value payment flows in real time, identifying anomalous liquidity behaviours at the intraday level. Guerra et al. (2022) further argued that ML enables more nuanced supervisory modelling of liquidity risk, allowing regulators to identify stress scenarios not captured by conventional stress testing frameworks.

Furthermore, Pham et al. (2024) focused on exchange-traded funds (ETFs), demonstrating that ML can forecast ETF liquidity using trading activity, volatility, and underlying asset behaviour, offering practical insights for institutional investors.

Despite the growing success of ML in liquidity forecasting, several limitations persist. Issues such as model interpretability, overfitting in high-frequency datasets, and the limited availability of labelled liquidity events remain significant (Samitas et al., 2022; Yang et al., 2025). Additionally, data privacy concerns and the opacity of complex models, particularly deep learning, pose obstacles for regulatory adoption (Climent et al., 2019).

Nevertheless, ensemble models and explainable AI techniques are emerging as solutions to these challenges, offering balance between accuracy and transparency (Tavana et al., 2018; Zhu et al., 2020).

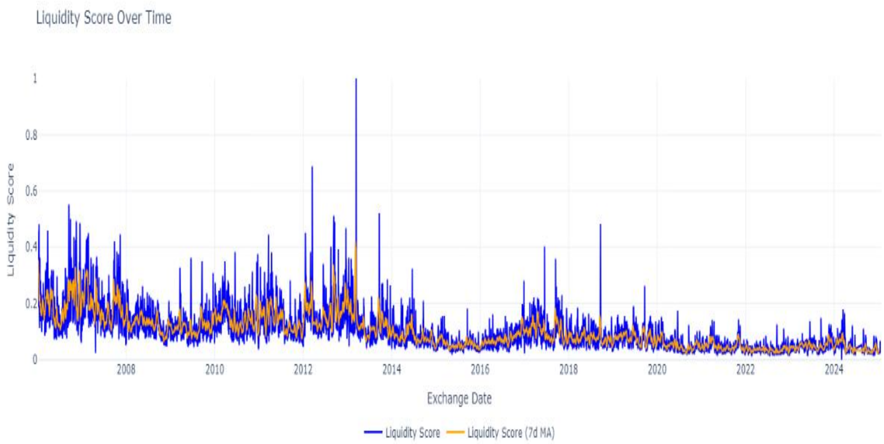
Overall, the literature confirms that machine learning offers substantial advantages over traditional methods in forecasting liquidity and managing liquidity risk. Through improved modelling of non-linearities, integration of textual data, and deployment in real-time supervisory systems, ML has emerged as a critical tool in modern financial analysis. Future research should focus on enhancing model explainability, integrating multi-modal data, and establishing standardised evaluation benchmarks for liquidity forecasting models.

3.Methodology

The dataset comprises daily observations from 1 January 2006 to 1 May 2025 for four financial markets: the German equity index (DAX), the German volatility index (VDAX), spot gold, and the EUR/USD exchange rate. For each trading day, the data include open, high, low, and close quotations, trading volume (where available), and bid/ask prices. All data employed in this study were obtained from Refinitiv Datastream.

Liquidity was proxied by a Liquidity Score, defined as a turnover-to-volatility ratio designed to penalise wide intraday price ranges while rewarding active trading behaviour. This metric captures core microstructural elements of liquidity and aligns with longstanding empirical approaches in financial market research (notably Kyle, 1985; Amihud, 2002). The Liquidity Score served as the dependent variable throughout the analysis. Missing values were handled using a backward fill method. To enhance interpretability and visualize prediction accuracy, seven-day moving averages of both observed and predicted Liquidity Scores were plotted (see Figure 1).

Figure 1
Comparative Predictions on Liquidity Score (7-Day Rolling Mean)



Source: Author's contribution

The construction of predictor variables was organised according to five conceptual blocks, reflecting volatility, price ranges, returns, technical indicators, and cross-asset interactions. Rolling standard deviations of closing prices were computed to capture asset-specific volatility dynamics. For the DAX index, 5-day and 10-day windows were employed. For VDAX and gold, 5-day rolling windows were used. High–low price ranges for DAX, gold, and EUR/USD were calculated, and bid–ask spreads were included where available to account for microstructural noise.

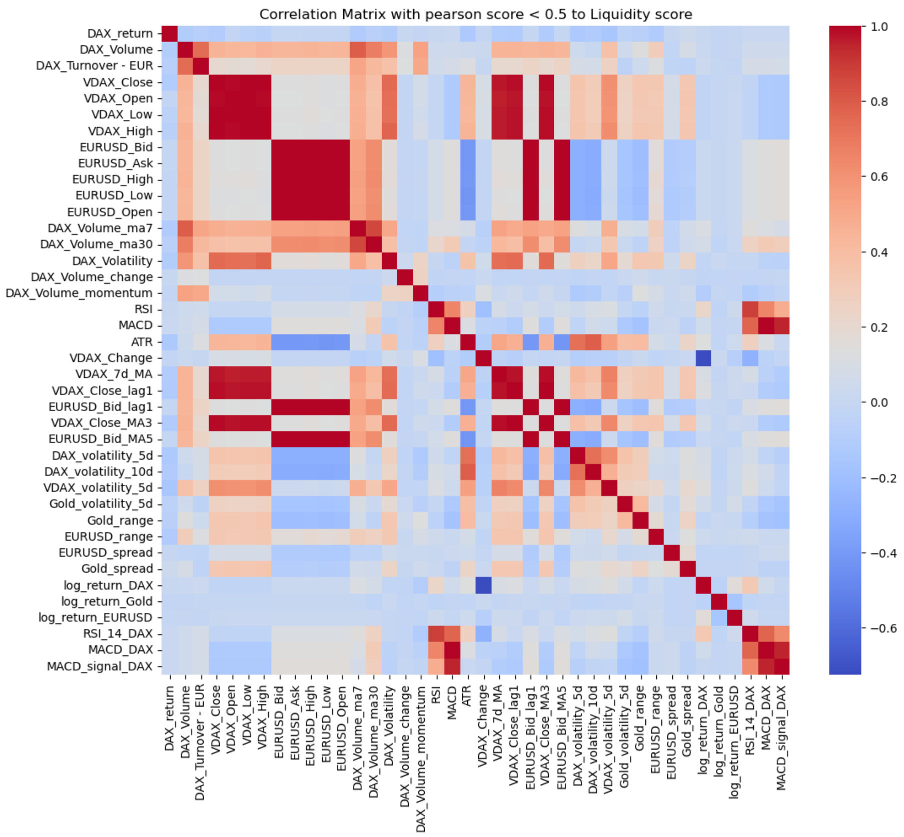
Logarithmic returns were computed as one-period, continuously compounded returns for DAX, gold, and EUR/USD, providing a standardised measure of price changes. Several technical indicators were incorporated to reflect momentum and trend-following dynamics. These included the 14-period Relative Strength Index (RSI), the Moving Average Convergence Divergence (MACD) along with its signal line, the 10-day and 50-day exponential moving averages (EMAs), the upper and lower bounds of Bollinger Bands, and the 14-day Average True Range (ATR). Additionally, one-day lags and short-term moving averages over 3-, 5-, and 7-day periods were included for

VDAX, gold, and EUR/USD to capture inter-asset linkages and short-run dynamics.

To mitigate the impact of multicollinearity while preserving explanatory strength, only predictors exhibiting an absolute Pearson correlation below 0.5 with the Liquidity Score were retained. This filtering procedure yielded a reduced yet robust subset of variables, forming the final feature matrix for model training. Figure 2 presents the matrix as a colour-graded heatmap, where warmer tones denote strong positive correlations, cooler tones indicate strong negative correlations, and the main diagonal reflects unit correlations by definition.

Figure 2

Correlation Matrix Heatmap



Source: Author's contribution

Figure 2 visualises the Pearson correlation matrix for the final set of retained predictors, all of which maintain an absolute correlation below the defined 0.5 threshold. The heatmap employs a diverging colour scale, with warm hues indicating positive correlations and cool tones representing negative relationships; the main diagonal, by construction, reflects unit correlation. The matrix reveals clusters of moderately interrelated features—such as lagged volatility measures and market microstructure indicators—which, although weakly correlated with the target individually, may contribute synergistically within non-linear modelling frameworks. This selective retention of weakly correlated variables reflects a modelling strategy that emphasises diversity, generalizability, and robustness over linear explanatory power, thereby reducing overfitting risks and enhancing model performance when employed in ensemble or deep learning architectures, as evaluated in Table 1 and depicted in Figure 3.

To account for the temporal autocorrelation structure characteristic of financial time series, the dataset was partitioned using a forward-chaining chronological split. The final 20% of the observations, corresponding to the most recent period in the time series, were reserved as an out-of-sample test set to ensure unbiased performance evaluation. The preceding 80% of the data was employed for model training and hyperparameter tuning under a time-consistent cross-validation scheme. An additive seasonal-trend decomposition using Loess (STL), parameterised with a 30-day trading seasonality window, was performed to isolate trend, seasonal, and residual components. The decomposition indicated the presence of weak but consistent cyclical behaviour in the liquidity time series. To assess the stochastic properties of the Liquidity Score, Augmented Dickey–Fuller (ADF) tests were conducted. The resulting p-values were consistently below the 0.05 threshold, leading to rejection of the null hypothesis of a unit root and supporting the stationarity assumption required for subsequent modelling.

Four supervised learning algorithms were selected for predictive modelling and comparative evaluation. These models were benchmarked using consistent performance metrics on the held-out test set to assess generalisation capability under realistic, temporally ordered conditions.

Random Forest (RF), introduced by Breiman (2001), aggregates an ensemble of decorrelated decision trees to reduce variance while preserving low bias. In a financial context, Liaw and

Wiener (2002) demonstrated RF's versatility for regression problems involving noisy, high-dimensional inputs. The method's robustness to multicollinearity and its built-in measure of variable importance make it an attractive baseline for liquidity forecasting. It was implemented with hyperparameters including the number of trees (ranging from 100 to 500), maximum tree depth (either unbounded or capped at 40), minimum samples per split (between 2 and 10), minimum samples per leaf (between 1 and 4), and feature selection strategy (auto, square root, or logarithm base 2). These parameters were optimised using a 50-draw randomised search combined with three-fold expanding-window cross-validation.

Chen and Guestrin (2016) proposed XGBoost as a highly efficient implementation of gradient-boosted decision trees, integrating regularisation, sparse-aware splitting and parallel computation. Its effectiveness in structured financial data has been showcased by Bentejac et al. (2021), who reported top-tier accuracy across a suite of time-series prediction tasks. The model was configured with similar optimisation procedures. The hyperparameter space included between 100 and 500 boosted trees, tree depths from 3 to 11, learning rates between 0.01 and 0.2, and subsample and column sample ratios ranging from 0.6 to 1.0.

The Long Short-Term Memory (LSTM) networks, originally formulated by Hochreiter and Schmidhuber (1997) to mitigate the vanishing-gradient problem, excel at capturing long-range temporal dependencies. Fischer and Krauss (2018) employed LSTMs to forecast stock returns and documented significant improvements over feed-forward networks and classical autoregressive models. Prior to training, Liquidity Score values were scaled to the [0,1] range and segmented into sequences of 30 days. The network architecture consisted of two LSTM layers: the first returned sequences, and the second produced a final state connected to a dense output layer. The model configuration involved a manual grid search across layer units, dropout rates, optimiser selection, number of epochs, and batch size. Early stopping with a patience threshold of five and a learning rate reduction on plateau were employed to prevent overfitting.

CatBoost, developed by Prokhorenkova et al. (2018), is a gradient boosting library that handles categorical features natively without the need for extensive preprocessing. By employing ordered boosting and a novel scheme for encoding categorical variables, CatBoost mitigates overfitting and achieves competitive performance on tabular datasets. In financial modeling, its efficiency in handling heterogeneous data types and robustness against overfitting make it particularly useful for tasks like credit scoring and market prediction, as evidenced in recent comparative studies (Hancock and Khoshgoftaar, 2020). Candidate settings involved 100 to 500 boosting iterations, depths from 3 to 11, learning rates between 0.01 and 0.2, L2-leaf regularisation values from 1 to 9, and border counts between 32 and 128. As with the other tree-based models, hyperparameter tuning was conducted via randomised search with cross-validation.

Predictive performance was evaluated on both training and test datasets using three standard metrics: mean absolute error (MAE), root mean squared error (RMSE), and the coefficient of determination (R^2).

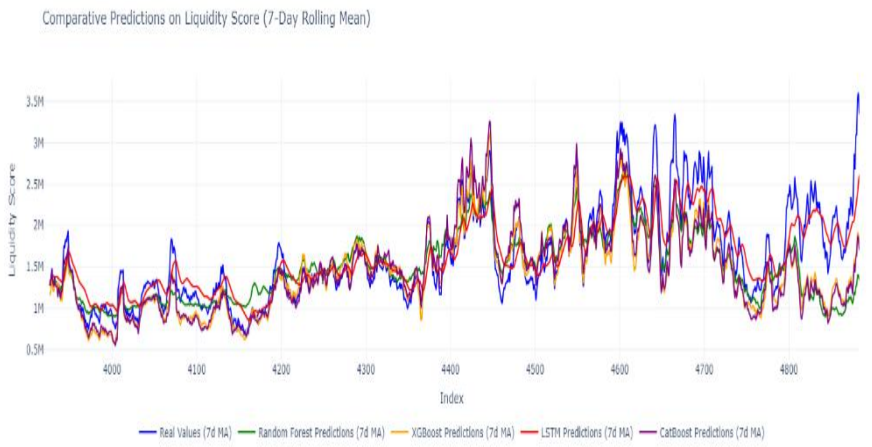
4. Results

Figure 3 presents a comparative time-series visualisation of the predicted and actual liquidity scores, smoothed using a 7-day rolling mean, for four distinct machine learning models: Random Forest, XGBoost, LSTM, and CatBoost. The application of the rolling mean serves to attenuate high-frequency fluctuations, thereby emphasising the underlying trends in liquidity dynamics.

This graphical representation substantiates the quantitative findings reported in Table 1, demonstrating that the gradient boosting models - CatBoost and XGBoost - exhibit superior performance in this forecasting task. Their effectiveness can be attributed to a favourable trade-off between bias and variance, as well as their capacity to model complex feature interactions, which enables more accurate and consistent liquidity predictions compared to both the Random Forest and LSTM models.

Figure 3

Predicted vs. Actual Liquidity Score (7-day rolling mean)



Source: Author's contribution

The real liquidity score is shown in blue, while the predicted values from the four models are shown in green (Random Forest), orange (XGBoost), red (LSTM), and purple (CatBoost). The visual comparison is directly contextualised by the predictive performance metrics summarised in Table 1, which reports the Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2 scores for both the training and test sets.

Table 1

Comparative Analysis of Models

Model	Set	MAE	RMSE	R^2
Random Forest	Train	97244.25	196117.72	0.9048
Random Forest	Test	377219.77	560692.67	0.514
XGBoost	Train	27862.92	35918.51	0.9968
XGBoost	Test	282033.17	418484.69	0.7293
LSTM	Train	301608.96	461820.94	0.474
LSTM	Test	476552.01	655444.26	0.3358
CatBoost	Train	40629.89	53966.24	0.9928
CatBoost	Test	281549.74	414466.17	0.7344

Source: Author's contribution

The XGBoost and CatBoost models clearly demonstrate superior generalisation capability, with test set R^2 values of 0.729 and 0.734 respectively, and the lowest RMSE and MAE scores among the models tested. Their forecast lines follow the actual liquidity values closely, capturing not only trend direction but also amplitude in periods of increased market turbulence — particularly around peak liquidity episodes.

In contrast, Random Forest exhibits significant performance degradation on the test set, dropping from an R^2 of 0.905 during training to just 0.514. This overfitting is also reflected visually: the green prediction line tends to lag behind actual liquidity changes and fails to replicate the higher-magnitude oscillations. The model is evidently not robust enough for forecasting extreme events.

The LSTM model, despite being a recurrent neural network capable of modelling sequential dependencies, underperforms both statistically and visually. With the lowest R^2 on the test set (0.336), it fails to capture the underlying signal effectively. The red curve appears smoothed and delayed, indicating that the model may have failed to learn the temporal structure inherent in the liquidity series, possibly due to insufficient tuning, model complexity, or lack of deeper layers.

5. Conclusion

The paper has examined the capacity of four leading machine-learning algorithms - Random Forest, XGBoost, CatBoost and Long Short-Term Memory network - to forecast daily liquidity in the German blue-chip equity market. Employing a Liquidity Score that balances turnover against intraday price dispersion and an extensive dataset, the study applied an identical walk-forward validation framework to ensure methodological comparability and to replicate real-time deployment conditions.

The empirical evidence indicates that gradient-boosting ensembles markedly surpass both Random Forests and the standard LSTM architecture. CatBoost, closely followed by XGBoost, consistently generated the lowest forecast errors and reproduced abrupt liquidity contractions more faithfully than its competitors. These results highlight the superior ability of boosting algorithms to capture complex, non-linear interactions among heterogeneous predictors while maintaining robustness against overfitting, and they underline the importance of a thoughtfully designed feature set that reflects the microstructural determinants of liquidity. By contrast, the LSTM's weaker performance suggests that recurrent networks may require longer or cleaner sequences—or more sophisticated tuning regimes—before they can rival tree-based methods in this context.

The study makes two substantive contributions to the literature. First, it introduces a parsimonious yet theoretically grounded liquidity metric that integrates both price impact and trading activity. Second, it proposes a systematic feature-engineering pipeline that balances informational breadth with parsimony by filtering for multicollinearity.

While the analysis focuses exclusively on the DAX index to ensure methodological depth and data consistency, this limits generalizability. Nonetheless, the modelling framework - especially the Liquidity Score and the structured feature set - is flexible and can be adapted to other developed equity markets. Future research could validate the findings by applying the same methodology to broader indices such as the EURO STOXX 50 or FTSE 100. Even a moderate cross-index comparison could strengthen external validity and reveal how model performance varies across market structures and liquidity regimes.

From a practical perspective, the study offers meaningful implications for multiple stakeholders. For traders and market-makers, the ability to anticipate short-term liquidity swings can inform order execution strategies, limit market impact, and optimise timing. Fund managers may integrate such forecasts into portfolio rebalancing processes or liquidity risk budgeting frameworks. Regulators and central banks can benefit from early-warning signals of liquidity stress, particularly in monitoring systemic risk in blue-chip segments of the market. The lightweight nature of the Liquidity Score and the real-time adaptability of boosting algorithms further enhance the feasibility of embedding such models into operational systems or supervisory dashboards.

Several limitations delimit the generalisability of the findings and motivate future research. The analysis is restricted to daily data for a single blue-chip index, leaving open the question of whether the observed performance hierarchy persists at intraday frequencies, in other asset classes or in less liquid markets. Furthermore, only four algorithms were evaluated, and issues of interpretability, computational efficiency and live implementation costs were beyond the present scope. Finally, the predictor set did not exploit textual sentiment, granular order-book information or alternative data sources that may refine short-horizon forecasts. Addressing these gaps, by extending the framework to higher-frequency horizons, incorporating richer feature spaces and deploying explainable, AI techniques represents a promising agenda for subsequent work.

Taken together, the results underscore the practical utility of gradient-boosting methods for short-horizon liquidity surveillance and provide a replicable methodological template for advancing research and practice in the forecasting of financial-market liquidity.

References

1. Amihud, Y. (2002). Illiquidity and stock returns: Cross-section and time-series effects. *Journal of Financial Markets*, 5(1). pp. 31–56.
2. Antony, T.M. and Kumar, B.S., 2024. Applying machine learning algorithms to predict liquidity risks. *Journal of System and Management Sciences*, [online] 14(1). pp. 21–30.
3. Bentejac, C., Lottier, N., & Henrot, R. (2021). A comparative review of gradient boosting algorithms. *Artificial Intelligence Review*, 54(3). pp. 1937–1967.
4. Box, G.E.P., & Jenkins, G.M. (1970). Time Series Analysis: Forecasting and Control. Holden-Day.
5. Breiman, L. (2001). Random forests. *Machine Learning*, 45(1). pp. 5–32.
6. Cabrol, A., Drobetz, W., Otto, T. and Puhan, T., 2024. Predicting corporate bond illiquidity via machine learning. *Financial Analysts Journal*, 80(2). pp. 223–240.
7. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. pp. 785–794.
8. Climent, F., Momparler, A. and Carmona, P. (2019). Anticipating bank distress in the Eurozone: An extreme gradient boosting approach. *Journal of Business Research*, 101. pp. 885–896.
9. Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, 270(2). pp. 654–669.
10. Guerra, P., Castelli, M. and Côrte-Real, N. (2022). Machine learning for liquidity risk modelling: A supervisory perspective. *Economic Analysis and Policy*, 73. pp. 1779–1790. https://doi.org/10.1016/j.eap.2022.01.005
11. Hancock, J. T., & Khoshgoftaar, T. M. (2020). CatBoost for big data: An interdisciplinary review. *Journal of Big Data*, 7(1). pp. 1–45.

12. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8). pp. 1735–1780.
13. Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: Principles and Practice (2nd ed.). OTexts.
14. Kirkby, J.L. and Andrean, V., 2024. Forecasting bid–ask spreads in foreign exchange: Analysis and machine learning prediction. *Journal of Financial Data Science*, 6(1). pp. 99–113.
15. Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica*, 53(6). pp. 1315–1335.
16. Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, 2(3). pp. 18–22.
17. Pham, S.D., Marshall, B.R., Nguyen, N.H. and Visaltanachoti, N. (2024). Predicting ETF liquidity. *Australian Journal of Management*, 49(1). pp. 34–53. https://doi.org/10.1177/03128962231114349
18. Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: unbiased boosting with categorical features. In: *Advances in Neural Information Processing Systems (NeurIPS)*, 31.
19. Samitas, A., Kampouris, E. and Kenourgios, D. (2022). Machine learning as an early warning system to predict financial crises. *Journal of International Financial Markets, Institutions and Money*, 78. 101565.
20. Tavana, M., Abtahi, A.R., Di Caprio, D. and Poortarighi, M. (2018). An artificial neural network and Bayesian network model for liquidity risk assessment in banking. *Expert Systems with Applications*, 94. pp. 70–81.
21. Triepels, R., Daniels, H. and Berndsen, R. (2021). Monitoring liquidity management of banks with recurrent neural networks. *Computational Economics*, 58(4). pp. 1151–1175. https://doi.org/10.1007/s10614-020-10059-y
22. Yang, J., Ren, R. and Kan, X. (2025). Improving the liquidity risk identification model of commercial banks using machine learning. *Academic Journal of Business & Management*, 7(2). pp. 315–325.
23. Zhu, Y., Lin, M. and Wang, H., 2020. Deep learning-based approach to liquidity risk management. *Systems Engineering*, 38(5). pp.1–11.