Portfolio Cash Liquidity Forecasting with Machine Learning

Project Overview

This project was carried out at a major European asset management firm. Its goal was to enhance cash liquidity forecasting for mutual fund portfolios regulated under the UCITS (Undertakings for Collective Investment in Transferable Securities) directive. Effective forecasting helps funds meet daily redemption requests, comply with regulation. To address these challenges, the project developed and tested a range of machine learning models to predict next-day (T+1) liquidity for equity, fixed income, and multi-asset portfolios.

Motivation and Context

Modern finance increasingly relies on data-driven insights. In recent years, machine learning (ML) has shown strong potential to improve forecasting in asset management. Research has emphasized that while ML can boost predictive accuracy, transparency and interpretability are equally important, especially in regulated industries.

Reflecting this, the project adopted a combination of advanced ML models like Random Forest, XGBoost, paired with SHapley Additive exPlanations (SHAP) to make model decisions more transparent. The aim was not only to forecast liquidity accurately, but also to provide clear explanations for model outputs so portfolio managers and risk teams could trust and act on the results.

Analytical Approach

Data Preparation

The analysis used anonymized daily data from 44 UCITS portfolios for the year 2024. Key variables included:

- Cash positions
- Investor flow dynamics
- Liquidity profiles across time horizons
- Regulatory risk measures

Additional macroeconomic data, such as market volatility, interest rates, and economic trends were sourced from the Federal Reserve and European Central Bank. Data cleaning, standardization, and anonymization were applied throughout to ensure quality and confidentiality.

Model Development

Three main techniques were explored:

- ARIMA: A classical statistical approach, serving as a baseline
- Random Forest & XGBoost: Modern ML models capable of capturing more complex patterns and relationships

Key steps included feature engineering (adding lagged and rolling variables), strict rolling-window out-of-sample validation, and the use of SHAP values for interpretability. To improve transparency further, simplified "distilled" versions of the machine learning models were also built.

Key Outcomes

Predictive Performance

- Machine learning models (Random Forest and XGBoost) clearly outperformed the traditional ARIMA baseline, with the best and most consistent results in equity portfolios.
- Random Forest achieved the lowest error rates and the highest explanatory power, especially for equity funds.
- Distilled linear models generally matched the performance of their more complex parent models, while making results easier to interpret.

Transparency and Business Insights

- Liquidity features, particularly current and recent cash positions, were the strongest drivers of forecasts.
- Macroeconomic variables, such as market volatility and interest rates, improved predictions during periods of stress.

- SHAP analyses helped to understand the models, supporting better decisions making.

Benefits to the Firm

The developed models provided the firm with:

- Supports to understand how to have more effective cash allocations
- Possible to develeop better compliance and model transparency
- It can facilitate earlier identification of potential liquidity pressures

Learning Experience and Lessons Learned

Technical and Operational Lessons

This project demonstrated that machine learning models, while powerful, must be explainable to gain acceptance. Performance varied across asset classes, underlining the importance of richer and more granular data for fixed income and multi-asset funds. The process also highlighted the importance of reproducibility and data privacy.

Personal and Professional Growth

Through this project, I advanced my skills in Python, machine learning, data cleaning, and model interpretability. I learned to balance technical complexity and I deepened my understanding of how to deliver solutions to business stakeholders.

Conclusion

In summary, this project advances the use of machine learning for cash liquidity forecasting in UCITS portfolios, providing a transparent, scalable foundation for future enhancements. The best results were seen in equity portfolios, but overall accuracy across asset classes remains moderate, reflecting the complexity of liquidity dynamics.

Future improvements should include additional explanatory variables, such as bond maturities and applications to new asset classes and longer-term forecasting. Most importantly, these models should be seen as decision-support tools, complementing professional judgment rather than replacing it. This approach allows asset managers to implement, refine, and validate ML-driven liquidity forecasting in a robust, controlled way.

Copyright Notice

AUTHOR'S DECLARATION

I, the undersigned, Károly Takács, candidate for the MSc degree in Business Analytics declare herewith that the present thesis titled "Portfolio cash liquidity forecasting with machine learning" is exclusively my own work, based on my research and only such external information as properly credited in notes and bibliography. I declare that no unidentified and illegitimate use was made of the work of others, and no part of the thesis infringes on any person's or institution's copyright.

I also declare that no part of the thesis has been submitted in this form to any other institution of higher education for an academic degree.

Vienna, 08 June 2025

Károly Takács