ECONSTOR
Make Your Publications Visible.

A Service of
ZBW
Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Singh, Vinay; Choubey, Bhasker; Sauer, Stephan

Article — Published Version

Liquidity forecasting at corporate and subsidiary levels using machine learning

Intelligent Systems in Accounting, Finance and Management

Provided in Cooperation with:
John Wiley & Sons

Suggested Citation: Singh, Vinay; Choubey, Bhasker; Sauer, Stephan (2024) : Liquidity forecasting at corporate and subsidiary levels using machine learning, Intelligent Systems in Accounting, Finance and Management, ISSN 1099-1174, Wiley, Hoboken, NJ, Vol. 31, Iss. 3, https://doi.org/10.1002/isaf.1565

This Version is available at:
https://hdl.handle.net/10419/306232

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

http://creativecommons.org/licenses/by/4.0/

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

Received: 12 June 2022 | Revised: 11 June 2024 | Accepted: 13 July 2024
DOI: 10.1002/isaf.1565

RESEARCH ARTICLE

WILEY

Liquidity forecasting at corporate and subsidiary levels using machine learning

Vinay Singh[1] | Bhasker Choubey[2] | Stephan Sauer[1]

[1]BASF, Ludwigshafen, SE, Germany
[2]Universitat Siegen, Siegen, Germany

Correspondence
Vinay Singh, BASF SE, Ludwigshafen, Germany.
Email: vinaysingh.bvp@gmail.com

Summary
Liquidity planning and forecasting are essential activities in corporate financial planning team. Traditionally, empirical models and techniques based on in-house expertise have been used to navigate the numerous challenges of this forecasting activity. These challenges become more complex when the forecasting activities are extended to subsidiaries of a large firm. This paper presents a structured approach that utilizes 240 covariates to predict net liquidity, customer receipts, and payments to suppliers to improve the accuracy and efficiency of liquidity forecasting in subsidiaries and at the corporate level. The approach is empirically validated on a large corporation headquartered in Germany, with average annual revenue from 6 to 7 billion Euro spanning 80 countries. The proposed approach demonstrated superior performance over existing methods in six out of nine forecasts using the data from 2014 to 2018. These findings suggest that a firm's classical approach to liquidity forecasting can be effectively challenged and outperformed by the algorithmic approach.

KEYWORDS
corporate finance planning, forecast, liquidity planning, machine learning algorithms

1 | INTRODUCTION

Firms maintain liquidity to meet the net cash outlays, including unanticipated investment opportunities and expenses. Failure to manage liquidity properly can lead to a firm delaying payments, selling assets, forgoing profitable investment opportunities, or financing at unfavourable terms (Emery & Cogger, 1982). Historically, research in liquidity has been concentrated at financial institutions like central banks, with extensive literature on empirical and algorithmic approaches for liquidity forecasting in these institutions (Johnson & So, 2018; Cornett et al., 2011). In contrast, non-financial corporations, particularly manufacturing firms, on the other hand, often consider liquidity as residual figures. Hence, there is a gap in the literature liquidity forecasting for typical business firms (Wang et al., 2022). However, this perspective has changed since the financial crisis of 2008 (Adrian et al., 2017). Currently, liquidity calculation and monitoring are becoming critical in determining investment in strategic projects and are often necessary for a firm's survival, leading to growth in literature on liquidity management (Bouslimi et al., 2024; Maurin et al., 2023).

Nevertheless, liquidity forecasting is still very challenging for large corporations with multiple subsidiaries across various countries (Teece, 2014). These challenges are further exacerbated when considering currency-related forecasts, with subsidiaries working in diverse national and international financial and political regimes (Petropoulos et al., 2022). Considering these complexities, several large firms still rely on trained human experts utilizing spreadsheet-based calculations for predicting liquidity, including the one utilized as a case study in this paper. While experienced, these human predictors have low validation capabilities and typically suffer from limited accuracy and

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2024 The Author(s). Intelligent Systems in Accounting, Finance and Management published by John Wiley & Sons Ltd.

trustworthiness. Computational approaches can automate this process and enhance the quality of forecasts by leveraging emerging artificial intelligence algorithms (Paul et al., 2023).

In this paper, we propose a novel methodology for automatic liquidity forecasting. To the best of our knowledge, this approach is the first of its kind, featuring an input-blind system, which learns to identify the necessary data from a sea of inputs utilizing machine learning-based techniques. We also present a case validating our approach for liquidity forecasting with data from a large global manufacturing firm headquartered in Germany. We argue that several critical factors in liquidity management can be understood through a framework that can help financially distressed firms ensure efficient investment in the future (Almeida et al., 2014; Hirunyawipada & Xiong, 2018). Our contribution extends the literature on cash flow prediction, filling a gap left by recently developed literature on liquidity management, wherein most of them have focused on bankruptcy prediction rather than cash flow prediction. The structured liquidity forecast (FC) showcased in this paper would be valuable for both small and large firms conceptualizing liquidity and provide a platform for future research in corporate liquidity studies. The practical implications of our work are multi-fold. First, we demonstrate that a machine-based liquidity forecasting approach, considering various indicators can outperform classical forecasting approaches. This provides substantial business advantages as more accurate liquidity forecasts can save money and help in better financial planning as well as investment decisions. In the case being reported, an improvement of just 10% in liquidity forecasting for a large firm can lead to a substantial monthly liquidity increase of € 100 million.

Before presenting our new methodology, we review the current understanding of liquidity forecasting in the next section. Our new computational methodology is described in Section 3. Typical numerical predictors identify trends or relationships between the required information and the collected data inputs (Lepenioti et al., 2020). However, one must identify and then collect the inputs required for such prediction before utilizing any computational approach. These can be challenging even for a single firm operating under one financial climate, as financial figures contributing to liquidity, while better understood than past, are still an open area of research. With large corporations having multiple subsidiaries under multiple financial climates, data identification and collection scale significantly in difficulty. Sections 4 and 5 describe the experimental landscape required for validating our new technique in the German manufacturing firm. This organization employs over a hundred thousand people and generates income in the range of 6–7 Billion Euros. With seven functional subsidiaries having footprints in more than 80 countries, it provides an ideal scenario to validate our techniques for complex liquidity forecasting. Sections 6 and 7 present the results of applying our technique to the firm's data. Our study aimed to identify a superior approach for liquidity planning vis-à-vis their existing methodology and suggest some areas of future research. Our primary concern is to deliver a high-level and abstract liquidity forecast, as requested by the corporate finance team. Finally, the last section outlines direction for future work.

2 | LITERATURE REVIEW AND CONCEPTUAL BACKGROUND

We present a comprehensive literature review from multiple angles relevant to this study. We first analyse the importance of liquidity management for firms. We then review studies that quantitatively analyse various aspects of corporate liquidity. There is extensive literature on corporate planning in financial institutions, including central banks, where several machine-learning tools have also been applied. We provide a brief review of these studies to form the foundation of our paper.

2.1 | Corporate liquidity

Keynes (1937), in his early seminal work, argued that a firm's liquidity planning is irrelevant if financial markets function well, as we typically assume. Hence, he fundamentally linked liquidity management to financing constraints (Kurowski & Sussman, 2011). Consequently, his and other studies focused the liquidity argument on cash holdings as firms' only option to manage liquidity. This optimal cash holding has been calculated as a simple product of the firm's investment and liquidity shortfall. In addition, the criticality of having cash for a firm depends on the probability that the firm will face a liquidity crunch and, therefore, must use the cash to finance investments (Almeida et al., 2011). Thus, by controlling the level of investment, optimal cash holding can be increased with the size of the liquidity shortfall (Box et al., 2016).

Emery and Cogger (1982) modelled a firm's liquidity position as a Wiener process and identified two liquidity measurement methods. Their first method calculated the likelihood of insolvency using a probability distribution function of insolvency, utilizing a normal distribution of cash flow over time with initial liquid reserve. They further extend it with another metric measuring relative liquidity across firms. Subsequent studies have proposed several solvency measures showing liquidity as exceptional cases of Emery's probability distribution function and relative liquidity metrics (Gryglewicz, 2011; Mukhtarov et al., 2022). These methods and most derivative works focus on qualitative predictions of the probability of insolvency. However, they provide limited insights on required quantitative predictions like cash flow figures.

For example, Gray (2008) discussed the challenges in forecasting the central bank's Balance Sheet. He uses his experience from forecasting the autonomous factors for central banks to individual big firms. He highlighted that if banks hold either too much or too little liquidity, they will respond in a way that may be detrimental to the central bank's goals. The same applies to large corporates and their subsidiary. However, Gray's analysis suffers from the limited quantitative forecast. Marti and co-workers have proposed clustering cash flow predictions in large firms (Marti et al., 2021) in financial markets to identify predictable clusters. Letizia and Lillo have also taken a similar approach, with applications in corporate payment networks and credit risk rating (Galar et al., 2016; Letizia & Lillo, 2019). However, these again do not provide individual liquidity forecasts.

In addition to the limitations, these traditional approaches for measuring liquidity emphasize a firm's working capital position, using the current and quick ratios as indicators of short-term solvency (Ibendahl, 2016). These ratios have been employed in numerous credit evaluation and failure prediction studies. However, they are seldom used as significant variables by practitioners in the industry (Prasad et al., 2019; Walter, 1957). One possible explanation is that such traditional ratios do not accurately describe the firm's liquidity position. They are static measures that ignore cash flows. With the optimal cash holding approach, firms have a target level of cash, which varies depending on the value of the firm's investments and the likelihood that the firm will not finance these investments without the retained cash (Azar et al., 2016).

2.2 | Statistical measures and time series analysis

While corporate liquidity forecasting has not been extensively studied, attempts have been made to FC various components of liquidity. Future incoming cash is a significant aspect of liquidity. One way to estimate it is to utilize data on cash collection from previous periods. This is regularly undertaken and often referred to as account analysis (Laik & Mirchandani, 2023). The analysis of this past data can potentially form a reliable basis for collection forecasting (Fortsch et al., 2022). However, the internal data analysis for one or few months alone does not provide any insight into external microelectronic or macroeconomic factors. This is further affected by typical limitations in data collection methodologies, including market conditions, currency fluctuations, commodities pricing, weather patterns, and geopolitical considerations. Alternatively, statistical techniques like multiple regression have been proposed to estimate the cash collection and related payment proportions with large data series (Bahrami et al., 2020; Kim & Kang, 2016; Stone, 1976). However, these techniques suffer from typical time lags between a credit sale and cash realization point. In addition, the lagged effect of credit sales and cash inflows is often distributed over several periods, which further restricts the accuracy of these techniques (Acharya et al., 2013). Nevertheless, these two methods provide a reasonably good estimation of both collectible and uncollectible cash patterns. Hence, they have been used as simple methods for forecasting liquidity (Elliman, 2006).

These time series analysis techniques have been further enhanced by statistical measures like the *Z* score, the degree of relative liquidity, or the lambda index (Shim, 2009). For example, model developed by Altman et al. (1977) uses multiple discriminant analyses to present a relatively good predictor of whether a firm will go bankrupt within the next five years. However, it should be evaluated over several years and hence, is difficult to be used as the sole basis of liquidity. Another work of Altman proposed the degree of the relative liquidity model, to evaluate a firm's ability to meet its short-term obligations (Altman, 1968). This model also uses discriminant analysis by combining several ratios to derive a percentage figure that indicates the firm's ability to meet short-term obligations (Cheng & Hüllermeier, 2009). However, this model does not incorporate the timing and variances in cash flows, as it assumes them to be uniform and continuous (Almeida et al., 2004). As a result, while correctly identifying an improved or deteriorated liquidity position, it cannot suggest exact causes of change. Alternatively, the widely popular, lambda index model evaluates the firm's ability to generate or obtain cash on a short-term basis to meet current obligations and therefore predict solvency (Altman, 1968; Altman et al., 1977). This index gauges a firm's liquidity. However, forecasted figures on other influencing macroeconomic and microeconomic parameters must be used to calculate this index, making the final lambda value suspect in several cases. Simonton (1977) developed alternative analytical procedures for cross-sectional time series in large sample sizes and relatively small number of observations per case. Interrupted time series, equivalent time samples, and multiple time series are treated within a multiple regression framework. Additional studies like those by Chordia et al. (2000), Domowitz et al. (2005), and Hasbrouck (2001) suggest that the components used in liquidity forecasting are typically similar. Chordia et al. (2000) even find liquidity co-movement across asset classes and analyse commonality across small and large firms.

It is also worth recalling that most formal studies have focused on bankruptcy prediction and avoidance rather than cash flow prediction (Mckee, 2000). Our research aims to develop such a system to assist decision-makers at the board level in appreciating their firm's liquidity and its relation to subsidiaries. Such a technique is highly desirable, though unfortunately, not yet available to most corporations. More importantly, the past liquidity risk determinants and forecasts were mainly related to close financial aggregates and indexes from financial institutions' statements. Hence, it is probable that a financial crisis originating from a single financial institution could propagate to the whole economic system, becoming a systemic crisis in due course. Therefore, we hypothesize that the financial components are not the only determinants of liquidity. One should also utilize macroeconomic determinants, which influence the liquidity quantity and quality in the market. Therefore, we consider multiple macroeconomic factors and traditional variables for a better liquidity forecast, systematically including them in a quantitative model. While some of these macroeconomic factors could have been included in some manual forecasts, it is difficult, if not impossible, to include them all in any spreadsheet-based analysis. The exact nature of the relationship between liquidity and each of these individual variables of a firm's financial data and macroeconomic parameters is difficult, if not impossible, to derive. Hence, we assume a black-box model wherein we train a system to learn the best possible relationship between input data and output liquidity. However, a completely independent learning system is mathematically a hard construct. Therefore, such systems need to be provided with known prior intelligence for faster convergence in training and better results during operation.

Furthermore, earlier studies have used the concept of clustering to segregate industry types. We also extend this to liquidity forecasting (Marques & Alves, 2020). Finally, we build on recent advances in deep learning and intelligent systems to better analyse the available data and build a robust model that improves future predictions. We will first present the quantitative details of our algorithm, followed by a description of our data collection and prediction strategy.

2.3 | Macroeconomic variables for liquidity forecasting

The few studies on corporate finance present in literature, primarily use internal data for liquidity analysis. At the same time, macroeconomic variables have long been recognized for their effectiveness in forecasting liquidity in financial institutions. For instance, studies by Bhattacharya and Patnaik (2016) show that inflation, exchange rates, and interest rates significantly impact liquidity forecasts. Numerous country-specific studies, including Cheng and Wu (2017), confirm the importance of macroeconomic variables such as gross domestic product growth, inflation, exchange rates, and interest rates in liquidity prediction.

Additionally, several studies, including Chordia et al. (2000, 2005), Hasbrouck and Seppi (2001), and Domowitz et al. (2005), provide strong evidence of commonality in liquidity, suggesting a systematic component. Extending beyond industry clustering, our approach integrates multiple macroeconomic factors for more robust liquidity forecasts. Recognizing that financial crises can be systemic, we argue that macroeconomic determinants influence market liquidity quantity and quality, not just financial components. This study, hence, fills a gap by focusing on cash flow prediction rather than bankruptcy prediction, as observed in prior research.

2.4 | Machine learning and liquidity in financial institutions

Recently studies have applied machine-based approaches around liquidity planning and forecasting (Guerra et al., 2022; Singh et al., 2022; Du & Lian, 2021). Barongo and Mbelwa (2024) overview the various machine learning techniques used for liquidity forecasting and evaluate their performance. Cai et al. (2021) propose a machine learning-based approach using a combination of clustering and regression techniques. Wang (2021) used a neural network model to predict liquidity risk. Similarly, Zhu et al. (2020) proposed a deep learning–based approach to liquidity risk management that uses a convolutional neural network (CNN) to FC liquidity risk. Zhao et al. (2023) experimented with solvency ratios using machine learning algorithms such as decision trees and random forests for Chinese listed companies, finding that these algorithms outperformed traditional statistical models in predicting these ratios. Similar results were observed by Corrêa and Saito (2018) for Brazilian listed companies. Lee et al. (2020) applied decision trees and random forests to predict the liquidity risk of Taiwanese commercial banks, wherein learning algorithms outperformed traditional statistical models. Finally, Shafiee and Mohamad (2019) challenge the traditional statistical models with an algorithm-based approach to Malaysian listed companies. However, most models in these prior studies have focused on bankruptcy prediction rather than cash flow prediction.

2.5 | Limitations of past approaches

The literature reveals several limitations in existing machine learning based approaches in liquidity. In addition to focussing on bankruptcy prediction rather than cash flow prediction, they suffer from limited feature set used to train machine learning models. Liquidity forecasting requires a comprehensive set of features, including macroeconomic, market, and financial data with complex dependencies between these instruments, which is difficult to capture using traditional machine learning models relying on linear relationships between variables. Finally, the lack of historical data also limits the effectiveness of machine learning models.

3 | ANALYTICAL FC METHODOLOGY

We aim to employ a machine learning–based system for liquidity forecasting, as these systems have shown superior performance in data analysis than pure statistical measures. Several algorithmic techniques have been suggested for forecasting on historical data (Chen & Boylan, 2007; Clements & Hendry, 1998; Collopy & Armstrong, 1992; Hanssens et al., 2003). Inspired by the statistical literature, we initially utilized several time series algorithms for forecasting. The simplest of these is a Naïve forecasting method, which uses the cash flow values from the previous period as the prediction value for the next period. It has the advantage of being quite simple, requiring no computational overhead. However, it has little or almost no ability to long-term or even short-term FC with reasonable accuracy. We then utilized statistical techniques like the auto-regressive integrated moving average (ARIMA) to represent widely used statistical techniques (Li et al., 2020).

However, as expected and described in our literature review section, these statistical techniques yielded limited success. This necessitated a shift towards a non-analytical black-box type solution. We, therefore, utilized applied machine learning techniques to train a system blindly from input data. Many learning algorithms exist in the literature with different training efforts and output accuracies. We present the best-case scenario observed till date with our firm without delving into various algorithms' individual advantages and disadvantages. We will report results based on the Cubist algorithm (Kuhn et al., 2012; Quinlan, 1992) as a representative to build a liquidity forecasting model.

Cubist machine learning algorithm belongs to the class of rule-based regression algorithms, combining elements of decision trees and linear regression (Pratama & Choi, 2017). It builds a series of regression trees and applies a "cubist modeling" process to derive rules from these trees. These rules are used to make predictions by combining linear models based on different combinations of input

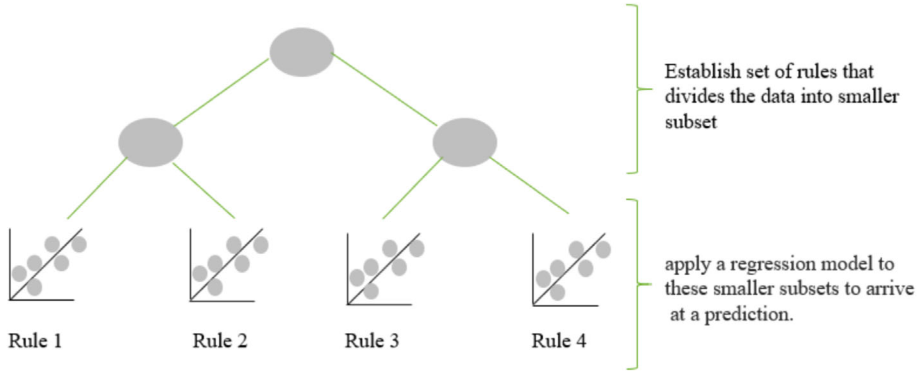


FIGURE 1 Two steps in Cubist models.

variables. Cubist is particularly effective in handling complex datasets with interactions and nonlinear relationships. CART (Classification and Regression Trees) algorithms are also decision tree-based algorithms; however, they differ in their approach and functionality from Cubist (Carmona & Sáez, 2016) (please refer to Appendix A for further details).

Cubist splits the time series data into previous predictors to create a tree. This tree has implicit regime identification as each branch is a rule utilizing intermediate linear models at each tree step. At the terminal node of this tree, a linear regression model is used for prediction, which is then "smoothed" by considering the linear model's prediction at the previous node. This recursive process reduces the tree to a set of "if-after-after" rules, each associated with a multivariate linear model. If the set of covariates satisfies the rule's conditions, the corresponding model is used to calculate the predicted value. These rules initially form paths from top to bottom and are then pruned and combined for simplification. Figure 1 shows the two-step process: Step1 establishes a set of rules that divide the data into smaller subsets, and Step 2 applies a regression model to these smaller subsets to arrive at a prediction. The Cubist algorithm allows multivariate linear models to overlap while not requiring them to be mutually exclusive (Gleason & Im, 2012). This is useful for liquidity forecasting, as we do not have prior knowledge of any relationship between these models. It is worth noting that this is unlike the well-known random forest model, as predictions are being pruned at each node from other similar training cases or nearest-neighbour model results (Kuhn et al., 2012; Quinlan, 1992). In addition, the Cubist algorithm can also predict values beyond the range covered by the samples compared with the random forest.

4 | DATA COLLECTION

The previous section briefly outlined the algorithmic foundation for liquidity prediction. However, any data analysis technique is only as effective as the quality of data collected and refined. The success of Cubist or any other machine learning algorithm depends on the data quality. For liquidity forecasting in large corporations with multiple subsidiaries, identifying relevant data to collect and categorizing it appropriately can be challenging even before addressing computational issues. Therefore, we developed a suitable integrated framework for our research, which can also be utilized by other corporations. We will now discuss the data requirements, including defining and categorizing cash flow clusters, selecting and adding indicator time series, and data cleansing. We will then analyse this data by our model and optimize it. Finally, we will benchmark our results with current practices in large firms and traditional statistical methods. Please refer to the Appendix C for the experimental landscape.

4.1 | Cash flow categorization and predictable cluster identification

As a first step, we needed to define the types of cash flows to be collected as time series. From an algorithmic perspective, it is beneficial to categorize our data into various clusters (Marti et al., 2021). It is worth recalling that typical cash flow is segregated by its directions: inflows and outflows. Customer payments mainly define inflows. We categorize these into the time-series cluster of "customer receipts." The outflow is separated into different clusters of "payments to suppliers," "payments to employees," and "dividend payments" due to their different patterns. For example, employee payment often peaks in certain months. In Europe, this is typically in December due to a bonus payment for Christmas. On the other hand, the shareholder meeting date defines the dividend payment. Furthermore, as we are interested in subsidiaries as well as the corporate headquarters, we also consider data breakdown from the global level by regional and business area levels. A firm's subsidiaries can differentiate cash flow behaviour from their parent organization, due to different business areas or geographical locations. The related economic situation with the subsidiaries and its divergent customer payment behaviour would impact the overall customer receipts of the parent organization. Therefore, we studied different cash flow clusters for each subsidiary, *focussing on* net cash flow, customer receipts, and supplier payment (refer to Appendix B). We *excluded* traditional components like dividend payments or capital investments from consideration, as these are driven by manual decisions within the company and, therefore, *are* already known (Acharya et al., 2012). By focusing on net cash flow and its components, including customer receipts and payments to suppliers, we gain a better understanding of the firm's operational

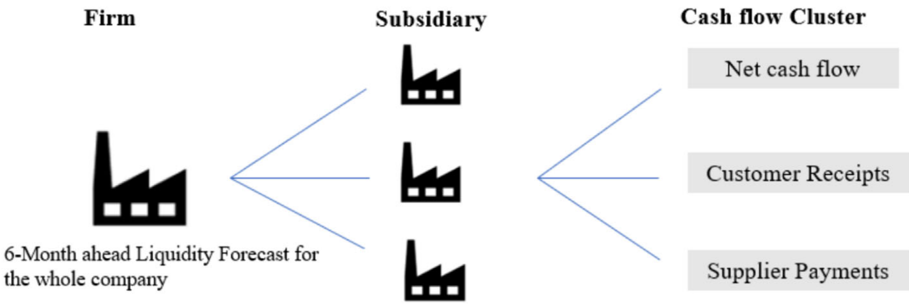


FIGURE 2 Clustering cash flows by levels.

liquidity. This categorizes an overall firm over the intermediate level regions and business area to cash flow clusters. A typical schematic level representation for such clustering is shown in Figure 2.

4.2 | Data acquisition: Source definition for cash flows and indicators

After clustering the cash flow, one must define the data source and characteristics of previously categorized cash flow clusters and the indicator time series. We will demonstrate this through the activities undertaken to validate our model in this paper. First, we collected monthly cash flow data from January 2014 to December 2018. This means time series for each 48 monthly data points (12 × 4 years). We then divided these data into three clusters: net cash flow, customer receipts, and payment to suppliers. These defined cash flow clusters are our target time-series variables. One can typically obtain this data from the global enterprise resource planning (ERP) system within their firm, and we did the same. ERP data is aggregated in several more detailed in- and out-flows cash positions in these three cash flow clusters. Based on that, we further split cash flow data into seven subsidiaries to separate regional and business area–related effects.

One can also cluster the data based on market, business area, and financial constructs like future, spot, or oligopolistic. However, we decided not to do so, as it was immaterial for decision-making based on overall liquidity. Furthermore, very sparse cash flow data was available at the corporate level, which introduces further complications. This insufficient data granularity mandated a breakdown to the subsidiaries' level. Furthermore, in the case of some specific subsidiaries, the business model has also changed in the recent past. Therefore, data points before such a change were removed and not considered in the time series. Finally, we derived time series per subsidiary and cash flow cluster. This shows typical problems at any corporation, where similar strategy can be used.

Additionally, as described above in our hypothesis, we desire to use internal and external micro and macroeconomic indicators in liquidity forecasting (Jordan & Messner, 2020). To identify the indicators having a significant predictive impact on the FC accuracy, we undertook deliberations with many internal business experts (Kuhn & Johnson, 2013). We utilized internal figures that have a significant impact on our cash flow clusters. Hence, we considered the conversion chain from sales (turnover) to cash flow (customer receipts), as shown in Figure 3.

Sales and account receivables are "accounting" parameters and appear in a company's profit and loss statement and the balance sheet before they become an actual monetary inflow, that is the customer receipts (Lubinska, 2020). We observed an identical behaviour on the outflow, i.e., payments to suppliers. Purchases and account payables are a firm's future payments to suppliers. These internal indicator data were extracted from an internal OLTP (online transaction processing) system like ERP system and an OLAP (online analytical processing) like business warehouse by executing queries and filters.

In addition to these internal factors, several external influencing indicators are of interest in a large multinational multi-business firm (Swieringa & Weick, 1987). The optimum external indicators required for liquidity forecasts are still a matter of research; however, we hypothesized that the following should affect our liquidity, utilizing the opinion of many in-house practitioners in the firm. First, exchange rates influence the time a client buys a product. An increased rate makes products more expensive for clients belonging to another currency zone and can postpone the moment they buy products due to speculations for a lower exchange rate. Therefore, it is reasonable to assume that it will affect the liquidity FC of any multinational firm.

Additionally, the economic environment significantly impacts consumer consumption and production. In favourable economic conditions, higher consumption and production result in higher payments and employment, which in turn further stimulates higher consumption. Conversely in dire economic situation, consumption and production decline, reducing payments and employment. Therefore, the economic situation directly influences a company's cash inflows and outflows, particularly for firms selling day-to-day used products. To capture these effects, we utilized macroeconomic indices reflecting the current conditions and stock indices predicting future trends. We also differentiated these indicators between a subsidiaries location or its consuming industries.

Furthermore, commodities prices, particularly for essential materials like oil and gas, play a crucial role in a company's purchasing decision and hence, define the price of produced goods and their customer demand (Chase, 2013; Chen et al., 2023). However, these external indicators can vary between subsidiaries' business and geographical regions. Therefore, different subsidiaries may require other indicators (Lahmiri, 2020). Fortunately, learning-based algorithmic systems can manage significant indicators and prioritize those needed for

FIGURE 3 Conversion chain from sales to cash flow.

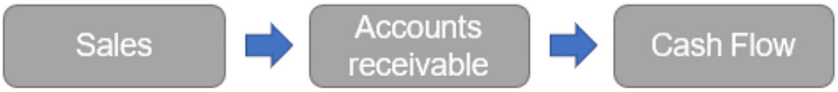


TABLE 1 Sample Indicator time series used for a subsidiary of the firm reported in the study, with all data points collected monthly from its operation in Germany.

Origin	Area	Indicator
Internal	Controlling	Turnover, account receivables, account payables
Internal	Procurement	Purchase expenses
External	Foreign exchange rates	USD/EUR
External	Stock indices per country	DAX (German stock index)
External	Stock indices per industry	TecDax (stock index for 30 largest German companies from the technology sector)
External	Commodity prices	Oil, gas and naptha
External	Macroeconomic indices per country	Germany's Consumer Confidence Index, USA's Industrial Production Index, German Production Index (Overall), German Car Manufacturing Index.

any subsidiary. For example, Table 1 shows the factors utilized for data analysis of one of the subsidiaries of the firm, with operational data in Germany.

We also examined boundary conditions like market psychology, driven by economic fundamentals and commodity prices (like Oil prices). Although our study focused on an industry heavily influenced by government interventions, we did not include this aspect due to the difficulty in quantifying it at this stage. We collected macrofinancial data from Bloomberg and other European financial agencies for this study. Fortunately, our firm's cash flow and indicator time series data are structured similarly to the market data, including the date, type, timestamp, and value. Nevertheless, one may argue that these data structures could differ from those of other firms. Hence, one needs to undertake an intermediate step of restructuring the data, potentially leading to data sparsity that must be accounted for.

For our study, all time-series data were collected at monthly periods. It still required further data cleaning as we needed to match the granularity of historical data, indicator databases, and FC horizon. The indicator time series and the predicted cash flow time series both spanned at least 48 data points over the same time range. Besides, the indicator time series can be extended by one year (12 data points) for most cases. Hence, we cover a range from January 2013 to December 2018 for the indicator data, but the prediction data range remains from January 2014 to December 2018. This approach helped us appreciate lagging indicators. One may recall that a lagging indicator could be any measurable or observable variable that changes its direction with any change in the target variable. Hence, these lagging indicators confirm any trends or changes in trends in the time series. These lags are especially useful in time series analysis because they tend to correlate the values within a time series with previous copies of itself. In our research, we have considered a lag of 3 months, as shown in Figure 4.

The data collection led us to create 12-time series with a lag from 0 to 12 out of one indicator time series. Using this, we even consider effects up to 12 months in the past and build future data points correlating with the target time series. This mechanism is applied to all our indicator data.

In conclusion, we have a mix of 12 different internal and external indicators for the considered example. Finally, the data input is defined by applying a lag up to 12 per indicator time series, i.e., 144-time series plus three cash flow cluster time series, resulting in 147-time series and a minimum total of 7.056 data. It is worth noting that the amount of used indicator time series and the resulting data point number can vary depending on the subsidiary.

4.3 | Data cleansing and enrichment: Outliers and seasonality

One may recall that liquidity traces the cash flow, whose different types vary in their volatility. For example, customer receipts can range from 1 to 6 months depending on the agreed payment terms. Often, the payment takes place after (overdue) or before (due to the current unattractive cash investment conditions) the due date. Such volatility can sometimes show a strong seasonal pattern or is a one-time effect influenced by unexpected or unknown factors (Sun & Ding, 2020). However, any collected real data could also have periodicity due to external factors and outliers. Hence, before undertaking liquidity forecasting, we need to undertake further input data enrichment by considering seasonality and removing one-time effects of outliers, which could be due to internal as well as external factors. Furthermore, it is highly improbable to get all data; hence, missing data is a fact of life in most data collection exercises, requiring date pre-processing.

In our data collection, we encountered several such data effects. For example, Figure 5 shows the customer receipts of one subsidiary of the firm, as collected from one country, in this case, Germany, from January 2014 till December 2018. One can readily appreciate the single strong peak anomaly in this data. However, these peaks need careful consideration to determine whether they are outliers or data errors. In our case, we discovered that some customers paid for their purchases at once by the end of 2016 due to an exclusive discount on the payment terms instead of monthly payments. Therefore, this would not be a good metric to reflect in typical forecasting activity. We, therefore, removed this and other similar outlier values, replacing them with values fitting the general trend of data. Similar outliers

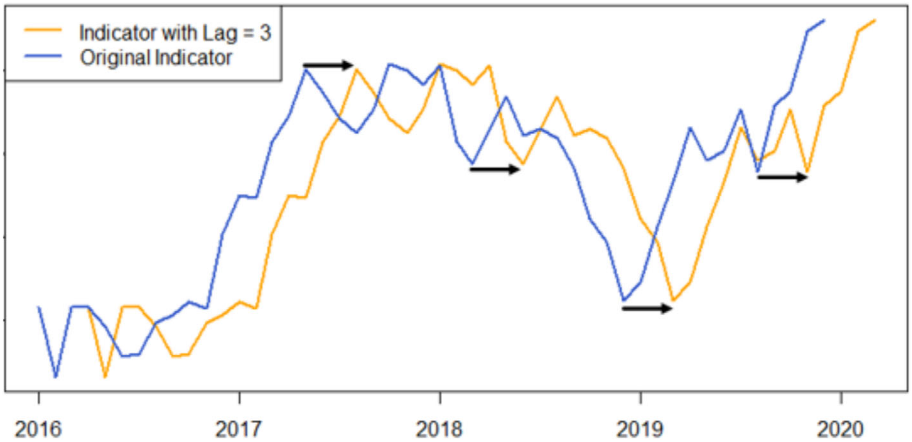


FIGURE 4 Sample time series lagged by 3 months.

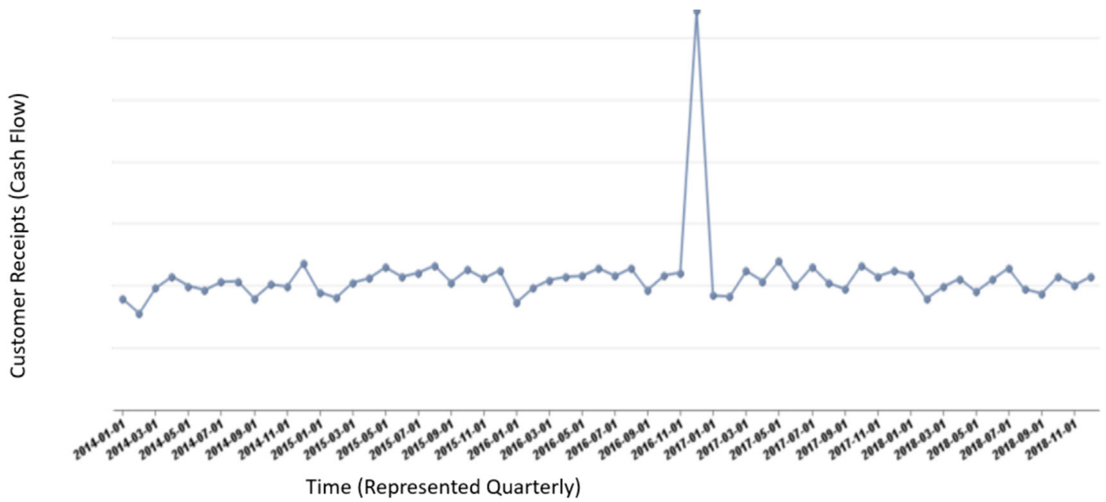


FIGURE 5 Cash flow time series with anomaly.

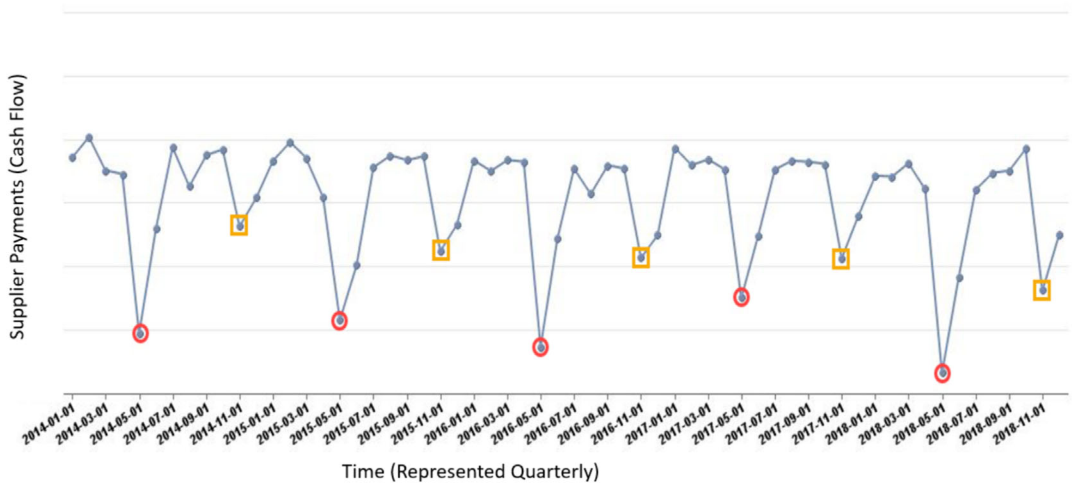


FIGURE 6 Cash flow time series with frequent outliers (seasonality).

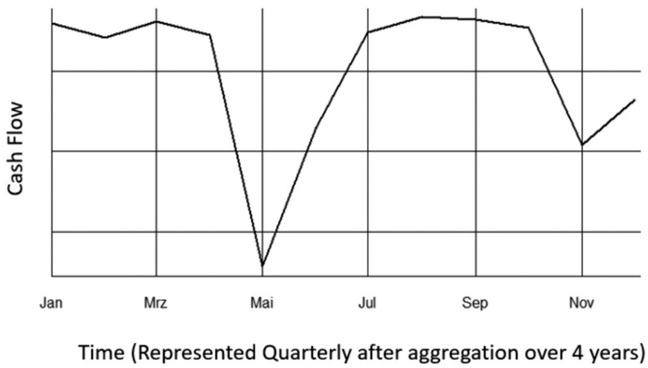


FIGURE 7 Decomposed seasonality in cash flow over 4 years.

were detected in other time series due to factors like extraordinary payment behaviours and incorrect classifications of underlying cash flows.

Simultaneously, one often also comes across outliers in small time windows, which would not necessarily be incorrect when considering over longer durations. For example, seasonal spikes appear as outliers over a short period. However, one would see periodicity in their occurrence over several years, Hence, they need not be considered outliers but preferably utilized in liquidity forecasting (Chordia et al., 2005). For example, Figure 6 shows seasonal effects in the data from one subsidiary, with recurring peaks in May and November, marked red and yellow. These peaks were therefore not removed from the analysis.

More importantly, we actively trained our model to compensate for seasonality, as illustrated in Figure 7 showing decomposed seasonality of the previous four years of data. The time-series data may also have multiplicative or additive components. For example, a decreasing or non-repeating cycle with repeating seasonality components can follow an increasing trend. Decomposition helps us analyse the data better and explore different ways to solve the problem.

As mentioned in Section 4.2, we extended the pre-selected indicators by seasonality. As presented in the previous section, the input data are enriched with decomposed seasonality time series and cleaned outliers. After cleaning extraordinary outliers, we also enriched our pre-selected indicators and the input data with decomposed seasonality.

5 | DATA ANALYSIS AND VALIDATION

5.1 | Liquidity metric

Before the current research was proposed and undertaken, our firm in question employed the simple Naïve liquidity FC method. This was rule-based on historical results and the business experience of expert personnel employed to undertake this. To compare this traditional approach with our new method, we needed a mathematical metric. We selected the mean absolute error (MAE) between the predicted and measured liquidity as a simple metric for model comparison and a more subjective average error measure. We understand that this is not the only qualitative metric, and one can also use root-mean-squared errors (RMSE). However, unlike RMSE, MAE is unambiguous (Willmott & Matsuura, 2005) and provides a clear interpretation of the deviation between a prediction and an actual value. Moreover, it treats positive and negative errors symmetrically.

It is important to note that firms often do not symmetrically handle positive and negative liquidity errors. For example, overestimating net cash flow is riskier than underestimating it, as firms prefer to have excess cash rather than risk bankruptcy (Glaum et al., 2018). Therefore, one can argue that a metric that emphasizes negative values more than positive ones could be a better metric to compare. However, the firm in discussion for this work, assessed positive and negative errors equally based on the findings from cross-validation. Hence, our study is restricted to symmetric treatment. It is worth recalling that the mean absolute error (MAE) is:

$$MAE = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n}$$

where n is the number of data points, y_i is the predicted value, and x_i is the actual reported value. To express the MAE as a percentage, one can use the mean absolute percentage error (MAPE), which calculates the relative error. However, it has a highly skewed distribution, particularly with outliers caused by high prediction error, with a low observation value (Goodwin & Lawton, 1999). Alternatively, a median-based metric is also of limited use due to the potential of zero values in the cash flow time series, which would lead to a divide by zero error (Goodwin & Lawton, 1999). Considering this, we selected relative MAE (RMAE) as a better representative measure. This is calculated as

$$RMAE = \frac{MAE}{\frac{\sum_{i=1}^{n} x_i}{n}}$$

The FC accuracy of the prediction can, then, be calculated by $(1 - RMAE)$.

5.2 | Cross-validation and model comparisons

Before analysing the results of our liquidity forecasting, it is worth revisiting the traditional human expert–based approach used in our firm and many other corporations. Such Naïve forecasts typically rely on recent observation data to predict the future values of potential liquidity. To improve predictions, a weighted factor approach is often undertaken. This factor is generally derived from an expert's business knowledge and experience. As a typical example, while predicting Jan 2019 forecasts, one could use Dec 2018 observation (e.g., 10 M EUR) and apply a multiplication factor (e.g., 0.7, based on experience) to predict 7 M EUR.

We analysed the liquidity forecasts from past time-series data as predicted by the traditional Naïve approach, computational methods like the statistical ARIMA, and our machine-learning-based black box system. These evaluations were performed on the target time series without considering any covariates or further tuning of parameters. We used the previously described cash flow time series from our firm consideration. Our firm uses a 6-month rolling FC strategy. This means that it executes a prediction of six FC steps every month, depending on the current month. For example, Figure 8 shows known data points as blue squares. The six FC steps are then predicted for future month data points. Brown gradations visualize these at the end of the series.

The various approaches to liquidity FC were evaluated using a nested cross-validation methodology, as shown in Figure 8. We collected firm performance and liquidity data from January 2014 to December 2018. We consider the data range from January 2017 to December 2018 for validating the proposed system. We build separated predictions for those time points per FC step. The training data in the figure is shown in blue. The next 6 months' data is the validation data. The training data was increased by a month for every month of data. This means the start date for all training scenarios was fixed at January 2014. However, the last available data point in every case would be different. For example, in the first training scenario, we had data until December 2016, thereby of 24 months. It was used to predict liquidity for the next 6 months. We were then able to compare this data with the actual liquidity recorded.

In the following training scenario, we used the data from January 2014 till January 2017, increasing the input data to 25 months. Then, we again predicted the next six months. Therefore, the distance between the predicted FC step and the last available training data is always equal, even though the length of training data differs in different iterations. For example, the iteration highlighted with the red box in Figure 9 shows the historical data till March 2017 being used for training. The predictions are now being undertaken for the next six months, from April 2017. Hence, the evaluation of the three approaches happens in chronological order. We start with Jan 2017 and predict next 6 months. We then repeat the process 24 times for every FC step independently till the last data point of Dec 2018 is reached. Please note here that liquidity forecasts are already undertaken for 6 months in future. In the first set, we consider the case when we are in 07.2016. One must predict for 01.2017, i.e., 6 months in advance, so the FC is done; as soon as we get the data for next month, i.e., 08.2016, we include that into our input data set. This allows us to have a rolling FC of the previous month's data set (actuals) to better our FC for the next 6 months. An aggregated metric each FC step was calculated based on residuals from these

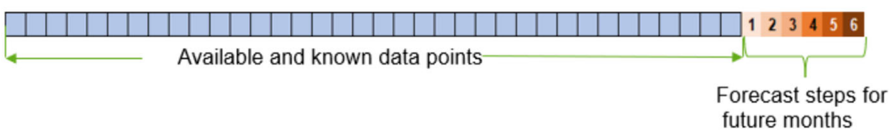


FIGURE 8 Six forecast steps for future month based on available and know data points.

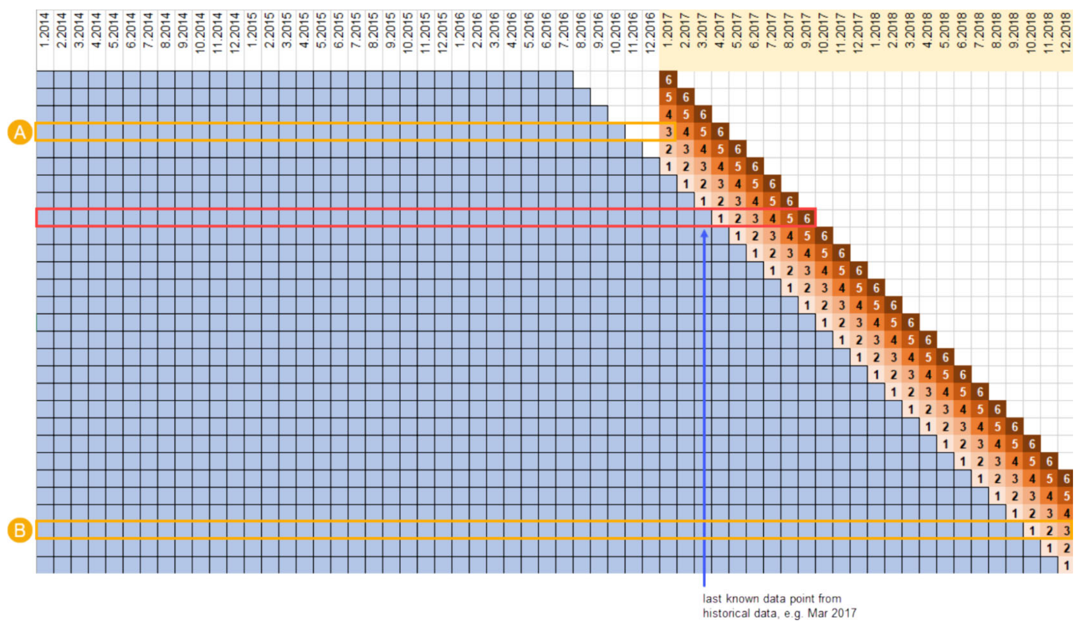


FIGURE 9 Nested cross evaluation.

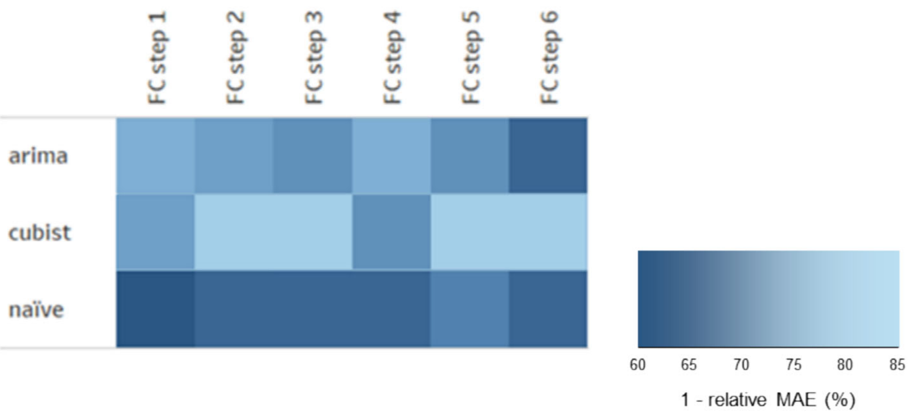


FIGURE 10 Liquidity forecast model evaluation: Naïve vs. Arima vs. Cubist. *FC = forecast.

24 predicted data points and the related recorded values. Figure 10 shows the results of these predictions, as measured through FC accuracy of (1 − RMAE).

One can observe that the machine learning approach of the Cubist algorithm consistently provides the least error among three methods in all FC steps. On the other hand, the simple naïve approach never performs as well. One reason for this could be that it cannot manage seasonality and trend inside the target time series. While statistical ARIMA works better than Naïve, the Cubist outperforms it in all cases. With the confidence that Cubist outperforms ARIMA, we forecasted the firm's actual liquidity and compared it to the human techniques applied previously.

6 | ANALYSIS OF MACHINE LEARNING–BASED LIQUIDITY FORECAST

We now undertake further analysis of our machine learning approach to liquidity forecasting and the data used for training. It is important to note that we propose a learning-based system, which requires careful optimization to achieve effective results within a reasonable timeframe and with finite computing power. This section details how the cubist algorithm was optimized for liquidity forecasting. This optimization was not tailored to the firm or data collection type making it applicable to other firms as well. Figure 11 shows a snippet (without any claim to completeness) of the input data used for training, model evaluation, and prediction. Each column refers to a different time series, interpreted as features by the cubist algorithm. The input data contains lagged indicator data like Industrial Production Index–Lag 0, 1, 2, 3, and 4, where each number represents lag by that many months. It is further extended by features like seasonality and its lags, applied after decomposing the target time series. Each row in this table represents data with single observations for every time series. This example refers to a prediction for January 2019, with a forecast step of one, which has been marked in orange. This hence uses the training data from January 2014 till December 2018. Every time series, with no future data points for our prediction month, is not considered. For example, the grey marked column labelled “Industrial Production Index–Lag 0” is ignored, as the time series used for training ends in December 2018 and does not provide any observation for predicting subsequent months. All other indicators time series contribute to potential prediction for January 2019 and therefore are used as features.

After providing the input data set, we again applied the nested cross-validation approach with machine learning to evaluate the best-performing features and parameters of the Cubist algorithm, like “ensemble classifier with schema called committees,” which were further optimized. The process is illustrated in Figure 12. It starts with a target time series and several indicator time series. In the second step, indicators are lagged and enriched, broadening their feature range. Finally, we reduced the number of features to those predicting results with the lowest metric values, after evaluation and feature elimination. This process is repeated for every forecast step independently. At the end of each forecast step, the set of lagged indicator time series and parameters with the lowest metric value is used to predict a future data point of our target time series. The results are discussed in Section 7.

6.1 | Forecast optimization

Machine learning algorithms, particularly Cubist, can be optimized for a range of desirable outcomes, such as forecast accuracy and forecast stability. Forecast stability is affected by variance in the model parameters between different planning periods. Fixing these parameters would help in stabilizing a forecast between periods. On the other hand, the forecast accuracy can be influenced by calculating the estimates in different ways and aggregating the results. For example, Figure 13 shows one such technique where the required forecast is calculated directly and indirectly by forecasting lower levels. Ideally, this approach must be applied to each node, such as at the level of the net cash flow per subsidiary. Each planning period should then be individually re-evaluated. This rule adapts to feature selection and hyperparameter tuning too. These two can then be averaged together to increase the forecast accuracy.

It is also worth noting that sales and customer receipts have lagged correlations due to several factors like applicable payment terms, typically observed in large firms (Namazi et al., 2016). Hence,

Date	Industrial Production Index - Lag 0	Industrial Production Index - Lag 1	Industrial Production Index - Lag 2	Industrial Production Index - Lag 3	Industrial Production Index - Lag 4	Seasonality Lag 0	Seasonality Lag 1	Seasonality Lag 2
01.01.2014	0.7837302	NA	NA	NA	NA	1.201551212	NA	NA
01.02.2014	0.8026166	0.7837302	NA	NA	NA	0.416186301	1.201551212	NA
01.03.2014	0.8904406	0.8026166	0.7837302	NA	NA	0.809218946	0.416186301	1.201551212
01.04.2014	0.8986424	0.8904406	0.8026166	0.7837302	NA	0.463781611	0.809218946	0.416186301
01.05.2014	0.9663175	0.8986424	0.8904406	0.8026166	0.7837302	0.247147384	0.463781611	0.809218946
01.06.2014	1.0290753	0.9663175	0.8986424	0.8904406	0.8026166	0.16179263	0.247147384	0.463781611
01.07.2014	1.0179148	1.0290753	0.9663175	0.8986424	0.8904406	-0.003689963	0.16179263	0.247147384
01.08.2014	0.9450935	1.0179148	1.0290753	0.9663175	0.8986424	-1.172273738	-0.003689963	0.16179263
01.09.2014	0.9502273	0.9450935	1.0179148	1.0290753	0.9663175	0.046889041	-1.172273738	-0.003689963
01.10.2014	0.7325368	0.9502273	0.9450935	1.0179148	1.0290753	0.094150807	0.046889041	-1.172273738
01.11.2014	0.859342	0.7325368	0.9502273	0.9450935	1.0179148	0.08884097	0.094150807	0.046889041
01.12.2014	0.8620951	0.859342	0.7325368	0.9502273	0.9450935	-2.876576112	0.08884097	0.094150807
01.01.2015	0.8437001	0.8620951	0.859342	0.7325368	0.9502273	1.201551212	-2.876576112	0.08884097
01.02.2015	1.0440419	0.8437001	0.8620951	0.859342	0.7325368	0.416186301	1.201551212	-2.876576112
01.03.2015	1.0340682	1.0440419	0.8437001	0.8620951	0.859342	0.809218946	0.416186301	1.201551212
01.04.2015	1.0570627	1.0340682	1.0440419	0.8437001	0.8620951	0.463781611	0.809218946	0.416186301
01.05.2015	1.0960874	1.0570627	1.0340682	1.0440419	0.8437001	0.247147384	0.463781611	0.809218946
01.06.2015	1.0483418	1.0960874	1.0570627	1.0340682	1.0440419	0.16179263	0.247147384	0.463781611
01.07.2015	0.923187	1.0483418	1.0960874	1.0570627	1.0340682	-0.003689963	0.16179263	0.247147384
01.08.2015	0.7539525	0.923187	1.0483418	1.0960874	1.0570627	-1.172273738	-0.003689963	0.16179263
01.09.2015	0.5071523	0.7539525	0.923187	1.0483418	1.0960874	0.046889041	-1.172273738	-0.003689963
01.10.2015	0.6361063	0.5071523	0.7539525	0.923187	1.0483418	0.094150807	0.046889041	-1.172273738
01.11.2015	0.752814	0.6361063	0.5071523	0.7539525	0.923187	0.08884097	0.094150807	0.046889041
01.12.2015	0.6710157	0.752814	0.6361063	0.5071523	0.7539525	-2.876576112	0.08884097	0.094150807
01.01.2016	0.3901128	0.6710157	0.752814	0.6361063	0.5071523	1.201551212	-2.876576112	0.08884097
01.02.2016	0.3804498	0.3901128	0.6710157	0.752814	0.6361063	0.416186301	1.201551212	-2.876576112
01.03.2016	0.5725592	0.3804498	0.3901128	0.6710157	0.752814	0.809218946	0.416186301	1.201551212
01.04.2016	0.6727065	0.5725592	0.3804498	0.3901128	0.6710157	0.463781611	0.809218946	0.416186301
01.05.2016	0.6735258	0.6727065	0.5725592	0.3804498	0.3901128	0.247147384	0.463781611	0.809218946
01.06.2016	0.666216	0.6735258	0.6727065	0.5725592	0.3804498	0.16179263	0.247147384	0.463781611
01.07.2016	0.6718229	0.666216	0.6735258	0.6727065	0.5725592	-0.003689963	0.16179263	0.247147384
01.08.2016	0.803496	0.6718229	0.666216	0.6735258	0.6727065	-1.172273738	-0.003689963	0.16179263
01.09.2016	0.8165425	0.803496	0.6718229	0.666216	0.6735258	0.046889041	-1.172273738	-0.003689963
01.10.2016	0.803723	0.8165425	0.803496	0.6718229	0.666216	0.094150807	0.046889041	-1.172273738
01.11.2016	0.7816642	0.803723	0.8165425	0.803496	0.6718229	0.08884097	0.094150807	0.046889041
01.12.2016	0.9059847	0.7816642	0.803723	0.8165425	0.803496	-2.876576112	0.08884097	0.094150807
01.01.2017	0.9887334	0.9059847	0.7816642	0.803723	0.8165425	1.201551212	-2.876576112	0.08884097
01.02.2017	1.0739093	0.9887334	0.9059847	0.7816642	0.803723	0.416186301	1.201551212	-2.876576112
01.03.2017	1.123265	1.0739093	0.9887334	0.9059847	0.7816642	0.809218946	0.416186301	1.201551212
01.04.2017	1.1644155	1.123265	1.0739093	0.9887334	0.9059847	0.463781611	0.809218946	0.416186301
01.05.2017	1.2225903	1.1644155	1.123265	1.0739093	0.9887334	0.247147384	0.463781611	0.809218946
01.06.2017	1.2963772	1.2225903	1.1644155	1.123265	1.0739093	0.16179263	0.247147384	0.463781611
01.07.2017	1.3544274	1.2963772	1.2225903	1.1644155	1.123265	-0.003689963	0.16179263	0.247147384
01.08.2017	1.3575065	1.3544274	1.2963772	1.2225903	1.1644155	-1.172273738	-0.003689963	0.16179263
01.09.2017	1.5118387	1.3575065	1.3544274	1.2963772	1.2225903	0.046889041	-1.172273738	-0.003689963
01.10.2017	1.6624939	1.5118387	1.3575065	1.3544274	1.2963772	0.094150807	0.046889041	-1.172273738
01.11.2017	1.7735843	1.6624939	1.5118387	1.3575065	1.3544274	0.08884097	0.094150807	0.046889041
01.12.2017	1.7937357	1.7735843	1.6624939	1.5118387	1.3575065	-2.876576112	0.08884097	0.094150807
01.01.2018	1.9997337	1.7937357	1.7735843	1.6624939	1.5118387	1.201551212	-2.876576112	0.08884097
01.02.2018	1.786689	1.9997337	1.7937357	1.7735843	1.6624939	0.416186301	1.201551212	-2.876576112
01.03.2018	1.7061851	1.786689	1.9997337	1.7937357	1.7735843	0.809218946	0.416186301	1.201551212
01.04.2018	1.6654096	1.7061851	1.786689	1.9997337	1.7937357	0.463781611	0.809218946	0.416186301
01.05.2018	1.7144591	1.6654096	1.7061851	1.786689	1.9997337	0.247147384	0.463781611	0.809218946
01.06.2018	1.7042258	1.7144591	1.6654096	1.7061851	1.786689	0.16179263	0.247147384	0.463781611
01.07.2018	1.6856531	1.7042258	1.7144591	1.6654096	1.7061851	-0.003689963	0.16179263	0.247147384
01.08.2018	1.7294228	1.6856531	1.7042258	1.7144591	1.6654096	-1.172273738	-0.003689963	0.16179263
01.09.2018	1.7418289	1.7294228	1.6856531	1.7042258	1.7144591	0.046889041	-1.172273738	-0.003689963
01.10.2018	1.4426184	1.7418289	1.7294228	1.6856531	1.7042258	0.094150807	0.046889041	-1.172273738
01.11.2018	1.3495517	1.4426184	1.7418289	1.7294228	1.6856531	0.08884097	0.094150807	0.046889041
01.12.2018	1.1015442	1.3495517	1.4426184	1.7418289	1.7294228	-2.876576112	0.08884097	0.094150807
01.01.2019	NA	1.1015442	1.3495517	1.4426184	1.7418289	1.201551212	-2.876576112	0.08884097
01.02.2019	NA	NA	1.1015442	1.3495517	1.4426184	NA	1.201551212	-2.876576112
01.03.2019	NA	NA	NA	1.1015442	1.3495517	NA	NA	1.201551212
01.04.2019	NA	NA	NA	NA	1.1015442	NA	NA	NA

FIGURE 11 Sample input data from our firm in question.

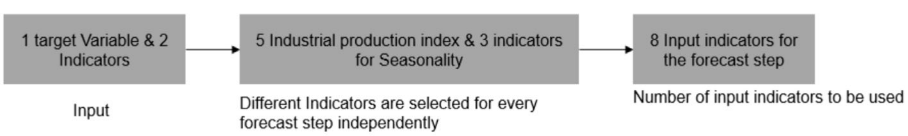


FIGURE 12 Sample procedure for indicator selection.

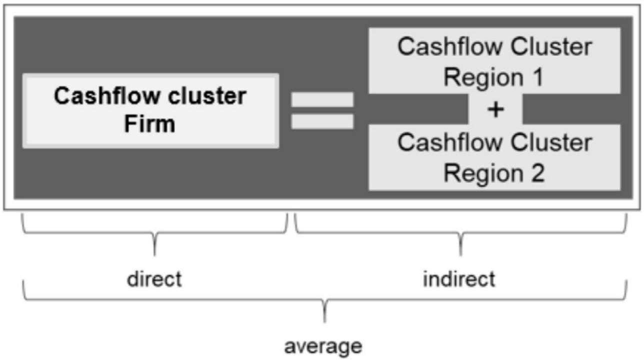


FIGURE 13 Variants of hierarchical forecast optimization.

one can often observe recurring behaviour due to seasonality, even in customer receipts or payment to suppliers, like those seen in sales. To account for the seasonality in payment to suppliers, we aggregate the sum of receipts and payments as a net cash flow forecast component.

It enabled us either to predict the net cash flow directly or indirectly through separate predictions of receipts and payments, with subsequent aggregation. A similar approach can be applied to a direct forecast for the entire group or an indirect forecast through every subsidiary individually, followed by aggregation. We decided to forecast for every subsidiary and for every cluster separately to validate both the direct and indirect net cash flow approaches. However, we do not predict the entire group's forecast directly due to different business areas and regional effects that would be lost in such

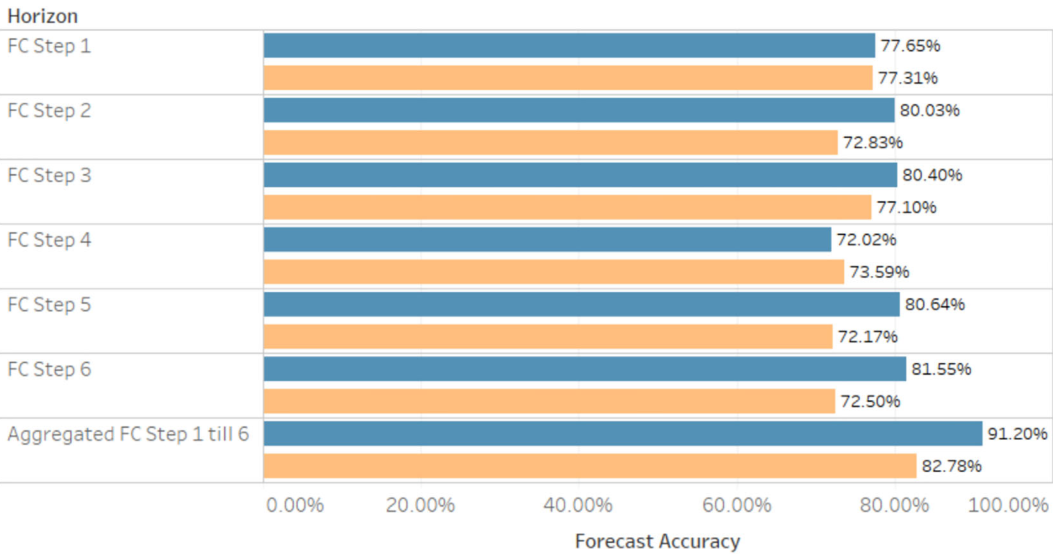


FIGURE 14 Performance of net cash flow (direct approach) — human based–machine based.

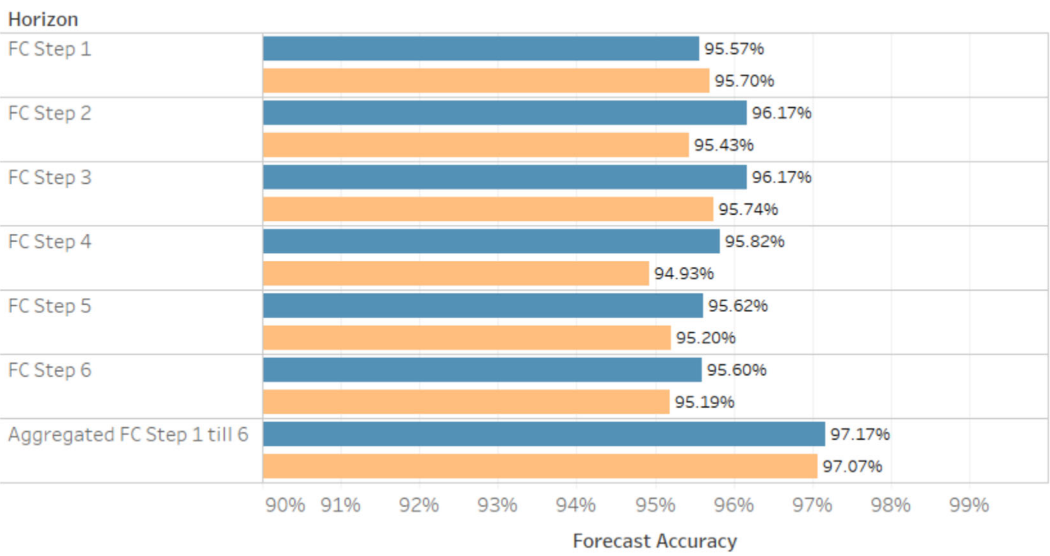


FIGURE 15 Performance of customer receipt forecasts (at subsidiary level).

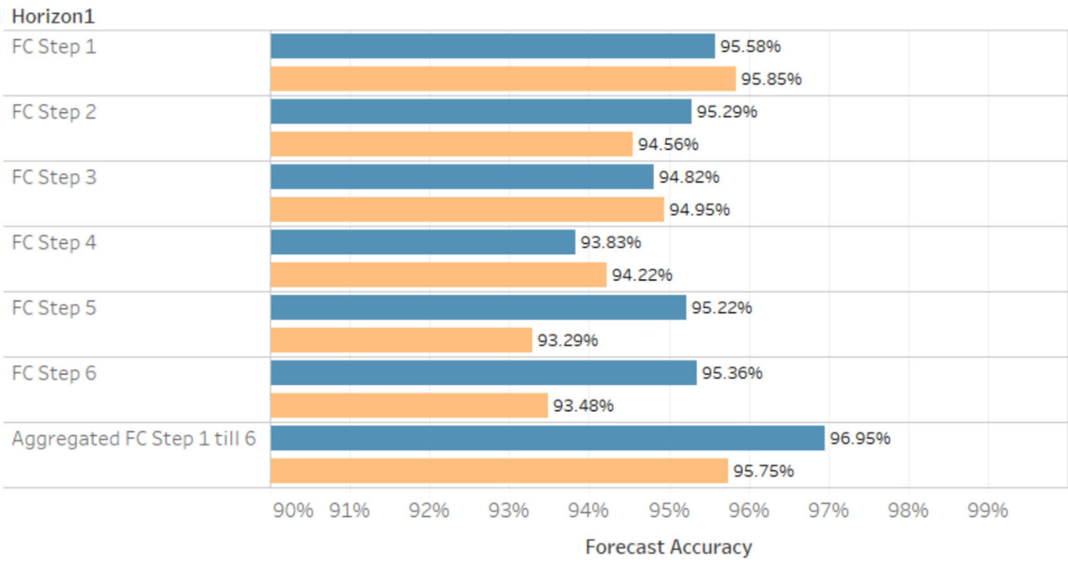


FIGURE 16 Performance of supplier payment forecasts (at subsidiary level) — human based–machine based.

forecasts. Subsidiary-level forecasts provide more control and understanding of various effects, further explored in Appendix D.

7 | FURTHER RESULTS

In our Cubist-based liquidity forecast approach, we used financial ratios (such as current ratio, quick ratio, and cash ratio), cash balances, accounts receivable, and accounts payable. We also considered parameters such as sales data over time (identify patterns and trends in firms' trading activity, which can inform liquidity forecasts), trading volume (i.e., the total number of contracts that the firm had during a given time frame), expenses; market volatility; sentiment analysis and macroeconomics indicators gathered from external sources such as economic reports and industry reports. Table 1 in Section 4.2 shows the sample indicator time series used in the study. Liquidity forecasts were conducted using the acquired data and suggested mathematical metrics. From the data collected and described in Section 4, the target time-series data from 2014 to 2016 and indicator time-series data from 2013 to 2016 was used for training. The last 2 years (2017–18) data was used for cross-validation. The training data had enough data points to detect any seasonal pattern. Cross-validation of the results consisted of 24 points, iterated so that the following period's training set is extended.

Figure 14 compares the performance of the human-expert-based Naïve prediction and machine learning–based models of the net cash flow forecast of one of the subsidiaries. Similar findings were also undertaken in other subsidiaries. In our comparative analysis with data from the last 24 months, each FC (forecast) step denotes a month's forecast of the 6 months rolling period. It can be observed that the machine-based forecast outperformed aggregated forecast over six individual forecast steps. However, for a one step (month) FC, classical human-expert forecast performs as well as the machine forecast, if not slightly better. With longer forecasting steps, the learning system outperformed the conventional approach, demonstrating a balancing approach due to averaging positive and negative deviations and reducing total error compared to individual errors from forecast steps. The aggregated forecast accuracy of approximately 91% surpasses the classical performance by about 9%. Similar behaviour was also observed in the forecast performance comparison in the customer receipts and supplier payments, shown in Figures 15 and 16. One can also observe that the machine-based approach leads to maximum improvement when comparing net cash flow. Further reduction in RMAE was often obtained by direct or average approach on the single subsidiary level.

8 | CONCLUSION AND FUTURE WORK

The aggregated forecast encompassing all forecast steps shows a significant improvement in accuracy compared to the classical forecast. While the classical method benefits from the unique insights of the cash flow planner for short-term forecasts, it falls short in long-term accuracy compared to machine learning systems. Our algorithm demonstrates a notable 9% improvement over human experts. Although this may seem modest, a 10% improvement in planning quality can lead to a substantial monthly liquidity increase of 100 million Euros and annual savings of 100 thousand Euros for the firm. The structured forecast outlined here, focusing on recurring cash flow clusters and minimizing human influence, offers valuable insights for firms of all sizes. This methodology enhances automation and accuracy in liquidity forecasting. Given the similarities in corporate liquidity factors among manufacturing-based multinational firms, this approach, while applied to one firm, suggests broader applicability. Furthermore, the validation across seven subsidiaries implies relevance to multiple independent firms, effectively equating to validation in at least seven different firms. Additional considerations for end-users and planners during liquidity planning are detailed in Appendix D.

Our study considered 240 covariates to predict net liquidity, customer receipts, and payments to suppliers to improve the accuracy and efficiency of liquidity forecasting at both subsidiaries and corporate level. For future work, the described methodology can be applied to additional cash flow time series and clusters, which have other characteristics, such as employee salaries, interest expenses, and taxes. Alternative forecasting methods, including emerging machine learning algorithms, should also be considered. Additionally, the availability of indicators can be extended and verified for continued relevance. Finally, making it challenging to identify the specific variables and factors driving the predictions. This lack of transparency can hinder understanding and acceptance. Therefore, future research should investigate explainable components in these approaches (Singh & Kar, 2022). Enhancing interpretability would help users understand how the model arrives at its predictions, thus increasing its acceptability and effectiveness in decision-making.

DATA AVAILABILITY STATEMENT

Research data are not shared.

ORCID

Vinay Singh https://orcid.org/0000-0002-8401-1123

REFERENCES

Acharya, V., Almeida, H., & Capello, M. (2013). Aggregate risk and the choice between cash and lines of credit. *Journal of Finance*, *68*, 2059–2116. https://doi.org/10.1111/jofi.12056

Acharya, V., Davydenko, S., & Strebulaev, I. (2012). Cash holdings and credit risk. *Review of Financial Studies*, *25*, 3572–3609. https://doi.org/10.1093/rfs/hhs106

Adrian, T., Fleming, M., Shachar, O., & Vogt, E. (2017). Market liquidity after the financial crisis. *Annual Review of Financial Economics*, *9*(1), 43–83. https://doi.org/10.1146/annurev-financial-110716-032325

Almeida, H., Campello, M., Cunha, I., & Weisbach, M. S. (2014). Corporate liquidity management: A conceptual framework and survey. *Annual Review of Financial Economics*, *6*(1), 135–162. https://doi.org/10.1146/annurev-financial-110613-034502

Almeida, H., Campello, M., & Hackbarth, D. (2011). Liquidity mergers. *Journal of Financial Economics*, *102*(3), 526–558. https://doi.org/10.1016/j.jfineco.2011.08.002

Almeida, H., Campello, M., & Weisbach, M. (2004). The cash flow sensitivity to cash. *Journal of Finance*, *59*, 1777–1804. https://doi.org/10.1111/j.1540-6261.2004.00679.x

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, *23*(4), 589–609. https://doi.org/10.1111/j.1540-6261.1968.tb00843.x

Altman, E. I., Haldeman, R. G., & Narayanan, P. (1977). ZETATM analysis: A new model to identify bankruptcy risk of corporations. *Journal of Banking & Finance*, *1*(1), 29–54. https://doi.org/10.1016/0378-4266(77)90017-6

Azar, J. A., Kagy, J.-F., & Schmalz, M. (2016). Can changes in the cost of carry explain the dynamics of corporate cash holdings? *Review of Financial Studies*, *29*, 2194–2240. https://doi.org/10.1093/rfs/hhw021

Bahrami, M., Bozkaya, B., & Balcisoy, S. (2020). Using behavioral analytics to predict customer invoice payment. *Big Data*, *8*(1), 25–37. https://doi.org/10.1089/big.2018.0116

Barongo, R. I., & Mbelwa, J. T. (2024). Using machine learning for detecting liquidity risk in banks. *Machine Learning with Applications*, *15*, 100511. https://doi.org/10.1016/j.mlwa.2023.100511

Bhattacharya, R., & Patnaik, I. (2016). Financial inclusion, productivity shocks, and consumption volatility in emerging economies. *The World Bank Economic Review*, *30*(1), 171–201.

Bouslimi, J., Hakimi, A., Zaghdoudi, T., & Tissaoui, K. (2024). The complex relationship between credit and liquidity risks: A linear and non-linear analysis for the banking sector. *Humanities and Social Sciences Communications*, *11*(1), 471. https://doi.org/10.1057/s41599-024-02951-4

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2016). *Time series analysis: Forecasting and control* (Fifth ed.). John Wiley & Sons, Inc.

Cai, W., Pan, Z., & Wang, Y. (2021). Uncertainty and the predictability of stock returns. *Journal of Forecasting*, *41*(4), 765–792. Portico. https://doi.org/10.1002/for.2832

Carmona, C. J., & Sáez, J. A. (2016). Comparing Cubist and random forest models for credit scoring: A replication study and extension. *Expert Systems with Applications*, *63*, 381–394.

Chase, C. (2013). *Demand-driven forecasting: A structured approach to forecasting* (Second ed.). Wiley. https://doi.org/10.1002/9781118691861

Chen, H., & Boylan, J. E. (2007). Use of individual and group seasonal indices in subaggregate demand forecasting. *Journal of the Operational Research Society*, *58*(12), 1660–1671. https://doi.org/10.1057/palgrave.jors.2602310

Chen, S. S., Li, T. L., Wu, Y. C., & Singh, V. (2023). An algorithm-based approach for mapping customer journeys by identifying customer browsing behaviors on E-commerce clickstream data. 2023 5th International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), IEEE, pp. 1–8.

Cheng, J. C., & Wu, R.-. S. (2017). Internal capital market efficiency and the diversification discount: The role of financial statement comparability. *Journal of Business Finance & Accounting*, *45*(5–6), 572–603. https://doi.org/10.1111/jbfa.12307

Cheng, W., & Hüllermeier, E. (2009). Combining instance-based learning and logistic regression for multilabel classification. *Machine Learning*, *76*(2–3), 211–225. https://doi.org/10.1007/s10994-009-5127-5

Chordia, T., Roll, R., & Subrahmanyam, A. (2000). Commonality in liquidity. *Journal of Financial Economics*, *56*(1), 3–28. https://doi.org/10.1016/S0304-405X(99)00057-4

Chordia, T., Sarkar, A., & Subrahmanyam, A. (2005). An empirical analysis of stock and bond market liquidity. *Review of Financial Studies*, *18*(1), 85–129. https://doi.org/10.1093/rfs/hhi010

Clements, M. P., & Hendry, D. F. (1998). *Forecasting economic time series*. Cambridge University Press. https://doi.org/10.1017/CBO9780511599286

Collopy, F., & Armstrong, J. S. (1992). Rule-based forecasting: Development and validation of an expert systems approach to combining time series extrapolations. *Management Science*, *38*(10), 1394–1414. https://doi.org/10.1287/mnsc.38.10.1394

Cornett, M. M., McNutt, J. J., Strahan, P. E., & Tehranian, H. (2011). Liquidity risk management and credit supply in the financial crisis. *Journal of Financial Economics*, *101*(2), 297–312. https://doi.org/10.1016/j.jfineco.2011.03.001

Corrêa & Saito. (2018). A machine learning approach to liquidity forecasting. *Journal of Financial Management, Markets, and Institutions*, *6*(2), 79–102.

Domowitz, I., Hansch, O., & Wang, X. (2005). Liquidity commonality and return co-movement. *Journal of Financial Markets*, *8*(4), 351–376. https://doi.org/10.1016/j.finmar.2005.06.001

Du, Y., & Lian, G. (2021). Machine learning for liquidity risk management: A review of current approaches. *Journal of Risk Research*, *24*(3), 333–352.

Elliman, D. (2006). Pattern recognition and financial time-series. *Intelligent Systems in Accounting, Finance and Management*, *14*(3), 99–115. https://doi.org/10.1002/isaf.279

Emery, G. W., & Cogger, K. O. (1982). The measurement of liquidity. *Journal of Accounting Research*, *20*(2), 290–303. https://doi.org/10.2307/2490741

Fortsch, S. M., Choi, J. H., & Khapalova, E. A. (2022). Competition can help predict sales. *Journal of Forecasting*, *41*(2), 331–344. https://doi.org/10.1002/for.2818

Galar, M., Fernandez, A., Barrenechea, E., & Bustince, H. (2016). A comparative study of statistical and machine learning models for credit scoring. *Journal of Business Research*, *69*(9), 3635–3641.

Glaum, M., Schmidt, P., & Schnürer, K. (2018). Processes and accuracy of cash flow forecasting: A case study of a multinational corporation. *Journal of Applied Corporate Finance*, *30*(2), 65–82. https://doi.org/10.1111/jacf.12301

Gleason, C. J., & Im, J. (2012). Forest biomass estimation from airborne LiDAR data using machine learning approaches. *Remote Sensing of Environment*, *125*, 80–91. https://doi.org/10.1016/j.rse.2012.07.006

Goodwin, P., & Lawton, R. (1999). On the asymmetry of the symmetric MAPE. *International Journal of Forecasting*, *15*(4), 405–408. https://doi.org/10.1016/S0169-2070(99)00007-2

Gray, S. T. (2008). *Liquidity forecasting*. Centre for Central Banking Studies, Bank of England.

Gryglewicz, S. (2011). A theory of corporate financial decisions with liquidity and solvency concerns. *Journal of Financial Economics*, *99*(2), 365–384. https://doi.org/10.1016/j.jfineco.2010.09.010

Guerra, P., Castelli, M., & Côrte-Real, N. (2022). Machine learning for liquidity risk modelling: A supervisory perspective. *Economic Analysis and Policy*, *74*, 175–187. https://doi.org/10.1016/j.eap.2022.02.001

Hanssens, D. M., Parsons, L. J., & Schultz, R. L. (2003). *Market response models: Econometric and time series analysis* (2nd ed., 2. print). Kluwer.

Hasbrouck, J. (2001). Common factors in prices, order flows, and liquidity. *Journal of Financial Economics*, *59*(3), 383–411. https://doi.org/10.1016/S0304-405X(00)00091-X

Hasbrouck, J., & Seppi, D. J. (2001). Common factors in prices, order flows, and liquidity. *Journal of Financial Economics*, *59*(3), 383–411.

Hirunyawipada, T., & Xiong, G. (2018). Corporate environmental commitment and financial performance: Moderating effects of marketing and operations capabilities. *Journal of Business Research*, *86*, 22–31. https://doi.org/10.1016/j.jbusres.2018.01.002

Ibendahl, G. (2016). Using solvency ratios to predict future profitability. *Journal of ASFMRA*, *2016*, 195–201.

Johnson, T. L., & So, E. C. (2018). A simple multimarket measure of information asymmetry. *Management Science*, *64*(3), 1055–1080. https://doi.org/10.1287/mnsc.2016.2608

Jordan, S., & Messner, M. (2020). The use of forecast accuracy indicators to improve planning quality: Insights from a case study. *European Accounting Review*, *29*(2), 337–359.

Keynes, J. M. (1937). The general theory of employment. *The Quarterly Journal of Economics*, *51*(2), 209.

Kim, J., & Kang, P. (2016). Late payment prediction models for fair allocation of customer contact lists to call center agents. *Decision Support Systems*, *85*, 84–101. https://doi.org/10.1016/j.dss.2016.03.002

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling*. Springer. https://doi.org/10.1007/978-1-4614-6849-3

Kuhn, M., Weston, S., Keefer, C., & Coulter, N. (2012). Cubist models for regression. R Package Vignette R Package Version 0.0, 18, 480.

Kurowski, L., & Sussman, D. (2011). *Investment project design: A guide to financial and economic analysis with constraints*. John Wiley & Sons. https://doi.org/10.1002/9781118267103

Lahmiri, S. (2020). A predictive system integrating intrinsic mode functions, artificial neural networks, and genetic algorithms for forecasting S&P500 intra-day data. *Intelligent Systems in Accounting, Finance and Management*, *27*(2), 55–65. https://doi.org/10.1002/isaf.1470

Laik, J., & Mirchandani, P. (2023). Effect of seasonality, sales growth rate, and fiscal year end on cash conversion cycle. *Decision Sciences*, *54*(1), 43–63.

Lee, T. H., Lin, K. J., & Chang, W. Y. (2020). Forecasting liquidity risk of commercial banks with machine learning algorithms. *Journal of Banking & Finance*, *110*, 105751.

Lepenioti, K., Bousdekis, A., Apostolou, D., & Mentzas, G. (2020). Prescriptive analytics: Literature review and research challenges. *International Journal of Information Management*, *50*, 57–70.

Letizia, E., & Lillo, F. (2019). Corporate payments networks and credit risk rating. *EPJ Data Science*, *8*(1), 21. https://doi.org/10.1140/epjds/s13688-019-0197-5

Li, Z., Han, J., & Song, Y. (2020). On the forecasting of high-frequency financial time series based on ARIMA model improved by deep learning. *Journal of Forecasting*, *39*(7), 1081–1097. https://doi.org/10.1002/for.2677

Lubinska, B. (2020). *Asset liability management optimization: A practitioner's guide to balance sheet management and remodelling* (First ed.). Wiley. https://doi.org/10.1002/9781119635529

Marques, B. P., & Alves, C. F. (2020). Using clustering ensemble to identify banking business models. *Intelligent Systems in Accounting, Finance and Management*, *27*(2), 66–94. https://doi.org/10.1002/isaf.1471

Marti, G., Nielsen, F., Bińkowski, M., & Donnat, P. (2021). A review of two decades of correlations, hierarchies, networks and clustering in financial markets. In F. Nielsen (Ed.), *Progress in information geometry: Theory and applications* (pp. 245–274). Springer International Publishing. https://doi.org/10.1007/978-3-030-65459-7_10

Maurin, V., Robinson, D. T., & Strömberg, P. (2023). A theory of liquidity in private equity. *Management Science*, *69*(10), 5740–5771. https://doi.org/10.1287/mnsc.2022.4612

Mckee, T. E. (2000). Developing a bankruptcy prediction model via rough sets theory. *International Journal of Intelligent Systems in Accounting, Finance & Management*, *9*(3), 159–173. https://doi.org/10.1002/1099-1174(200009)9:3<159::AID-ISAF184>3.0.CO;2-C

Mukhtarov, S., Schoute, M., & Wielhouwer, J. L. (2022). The information content of the Solvency II ratio relative to earnings. *Journal of Risk and Insurance*, *89*(1), 237–266. https://doi.org/10.1111/jori.12354

Namazi, M., Shokrolahi, A., & Sadeghzadeh Maharluie, M. (2016). Detecting and ranking cash flow risk factors via artificial neural networks technique. *Journal of Business Research*, *69*(5), 1801–1806. https://doi.org/10.1016/j.jbusres.2015.10.059

Paul, S., Gupta, A., Kar, A. K., & Singh, V. (2023). An automatic deep reinforcement learning based credit scoring model using deep-Q network for classification of customer credit requests. In *2023 IEEE International Symposium on Technology and Society (ISTAS)* (pp. 1–8). IEEE.

Petropoulos, F., Apiletti, D., Assimakopoulos, V., Babai, M. Z., Barrow, D. K., Ben Taieb, S., Bergmeir, C., Bessa, R. J., Bijak, J., Boylan, J. E., Browell, J., Carnevale, C., Castle, J. L., Cirillo, P., Clements, M. P., Cordeiro, C., Cyrino Oliveira, F. L., de Baets, S., Dokumentov, A., … Ziel, F. (2022). Forecasting: Theory and practice. *International Journal of Forecasting*, *38*(3), 705–871. https://doi.org/10.1016/j.ijforecast.2021.11.001

Prasad, P., Sivasankaran, N., Paul, S., & Kannadhasan, M. (2019). Measuring impact of working capital efficiency on financial performance of a firm: An alternative approach. *Journal of Indian Business Research*, *11*(1), 75–94. https://doi.org/10.1108/JIBR-02-2018-0056

Pratama, M. R., & Choi, T. M. (2017). A comparative study of random forest and Cubist for predicting sales targets. *Decision Support Systems*, *101*, 12–22.

Quinlan, J. R. (1992). Learning with continuous classes. 343–348.

Shafiee, M., & Mohamad, M. S. (2019). Liquidity prediction using machine learning techniques. *Journal of Risk and Financial Management*, *12*(3), 121.

Shim, J. K. (2009). *Strategic business forecasting: Including business forecasting tools and applications*. Global Professional Pub.

Simonton, D. K. (1977). Cross-sectional time-series experiments: Some suggested statistical analyses. *Psychological Bulletin*, *84*(3), 489–502. https://doi.org/10.1037/0033-2909.84.3.489

Singh, V., Chen, S.-S., Singhania, M., Nanavati, B., Kumar kar, A., & Gupta, A. (2022). How are reinforcement learning and deep learning algorithms used for big data based decision making in financial industries–A review and research agenda. *International Journal of Information Management Data Insights*, *2*(2), 100094. https://doi.org/10.1016/j.jjimei.2022.100094

Singh, V., & Kar, A. (2022). *When to choose ranked area integrals versus integrated gradient for explainable artificial intelligence — a comparison of algorithms*. An International Journal.

Stone, B. K. (1976). The payments-pattern approach to the forecasting and control of accounts receivable. *Financial Management*, *5*(3), 65. https://doi.org/10.2307/3665133

Sun, W., & Ding, Y. (2020). Corporate social responsibility and cash flow volatility: The curvilinear moderation of marketing capability. *Journal of Business Research*, *116*, 48–59. https://doi.org/10.1016/j.jbusres.2020.05.016

Swieringa, R. J., & Weick, K. R. (1987). Management accounting and action. *Accounting, Organizations and Society*, *12*(3), 293–308. https://doi.org/10.1016/0361-3682(87)90042-0

Teece, D. J. (2014). A dynamic capabilities-based entrepreneurial theory of the multinational enterprise. *Journal of International Business Studies*, *45*(1), 8–37. https://doi.org/10.1057/jibs.2013.54

Walter, J. E. (1957). Determination of technical solvency. *The Journal of Business*, *30*(1), 30. https://doi.org/10.1086/294129

Wang, F., Mbanyele, W., & Muchenje, L. (2022). Economic policy uncertainty and stock liquidity: The mitigating effect of information disclosure. *Research in International Business and Finance*, *59*, 101553. https://doi.org/10.1016/j.ribaf.2021.101553

Wang, Q. (2021). Machine learning-based liquidity forecasting. *Journal of Risk and Financial Management*, *14*(11), 556.

Willmott, C., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, *30*, 79–82. https://doi.org/10.3354/cr030079

Zhao, Q., Xu, W., & Ji, Y. (2023). Predicting financial distress of Chinese listed companies using machine learning: To what extent does textual disclosure matter? *International Review of Financial Analysis*, *89*, 102770. https://doi.org/10.1016/j.irfa.2023.102770

Zhu, J., Li, L., & Hou, Z. (2020). Deep learning for liquidity risk management. *Journal of Financial Engineering*, *7*(2), 2050010.

How to cite this article: Singh, V., Choubey, B., & Sauer, S. (2024). Liquidity forecasting at corporate and subsidiary levels using machine learning. *Intelligent Systems in Accounting, Finance and Management*, *31*(3), e1565. https://doi.org/10.1002/isaf.1565

APPENDIX A: COMPARING CUBIST WITH CART A

In our approach, we have considered the comparison based on our specific liquidity needs, focusing on modeling techniques and how they handle variables, and the interaction during the handling of variables.

A.1 | Modeling technique

CART employs a binary recursive partitioning algorithm to construct decision trees. It recursively splits the dataset based on a selected attribute and threshold, optimizing for the purity (homogeneity) of the resulting subsets. On the other hand, Cubist utilizes a combination of decision trees and linear models. It first builds a series of regression trees using the CART algorithm and then applies a "cubist modeling" model to generate rules and linear models from these trees.

A.2 | Rule-based modeling

An essential distinction between Cubist and CART is the incorporation of rule-based modeling in Cubist. After creating the regression trees, Cubist applies a rule induction process to derive rules from the trees. These rules are based on the tree structure and the linear models associated with the terminal nodes. As a result, the rules offer a more interpretable representation of the relationships between input and target variables.

A.3 | Handling of continuous variables

CART handles continuous variables by selecting an appropriate threshold to split the data. Cubist, however, handles continuous variables differently. It uses "constructing a centre" to convert continuous variables into discrete variables by creating intervals or ranges. This allows Cubist to treat continuous variables as ordinal categorical variables during tree-building.

A.4 | Handling interactions

Cubist is specifically designed to handle interactions between variables, capturing both additive and multiplicative interactions by considering combinations of variables in the linear models associated with the rules. This enables Cubists to model complex relationships and interactions between variables more effectively than CART, which focuses on recursive splits based on individual variables.

Cubist extends CART's capabilities by incorporating rule-based modeling, handling continuous variables differently, and capturing interactions between variables. These differences make Cubist a robust algorithm for regression tasks, especially when dealing with complex datasets with nonlinear relationships and interactions.

APPENDIX B: KEY DEFINITIONS B

B.1 | Net cashflow

Net cash flow in financial liquidity planning refers to the difference between cash inflows and outflows over a specific period and is a crucial measure in liquidity planning, helping assess the availability of cash to meet short-term obligations and fund daily operations.

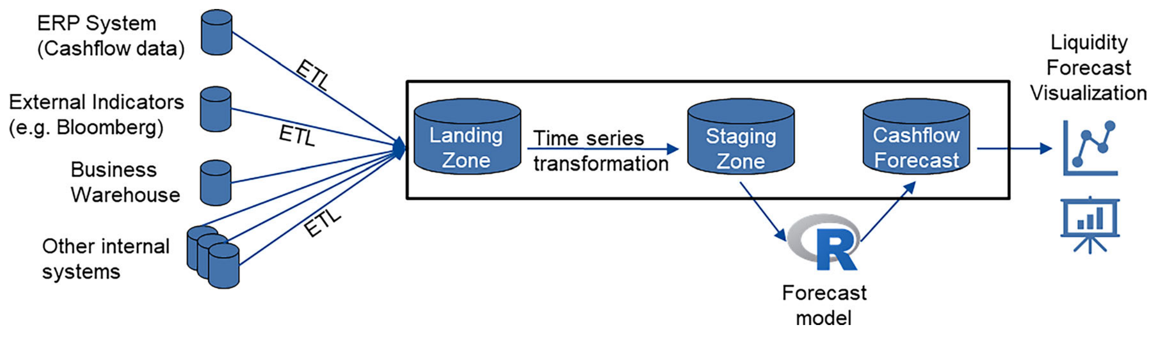
B.2 | Customer receipts

Customer receipts refer to the cash inflows received from customers or clients for the products or services provided by a business. It represents the payments the business receives from its sales or accounts receivable. It directly contributes to the cash available for meeting financial obligations and funding ongoing operations. We have considered the customer receipts features of timing of receipts, cash conversion cycle, dependency on cash flow forecasting, and cash conversion cycle.

B.3 | Supplier payments

Supplier payments refer to the cash payments a business makes to its suppliers or vendors for goods or services. It represents the outflow of cash from the business to settle its outstanding payables or trade debts. In our study, we have considered the payment terms and prioritization, discounts, and rebates.

APPENDIX C: EXPERIMENTAL SYSTEM LANDSCAPE C



ETL, extract, transform, and load.

APPENDIX D: VITAL CONSIDERATION BY END-USER OR PLANNER DURING LIQUIDITY PLANNING D

Key consideration by end-user or planner during liquidity planning may include:

- How to use a forecast?
  The effectiveness of a forecast hinges on its practical utility for cash flow planners. Therefore, forecasts should be easily visualized and interpretable, seamlessly integrated into the planning process to facilitate informed decision-making.
- How will the current planning process be influenced?
  Cash flow planners may not plan those cash flow clusters predicted using machine-based algorithms. Hence, they can focus on a stepwise approach to not-yet machine-based forecasted clusters and some human-made decisions. For example, it could require a monthly view of the latest actuals from cash flow clusters of the machine-based planning for data cleansing or reporting of effects.
- How could the current process be further digitalized?
  Introducing machine-based liquidity planning offers significant opportunity for digitalization. By integrating machine-based forecasts with manually planned cash comprehensive view of the firm's liquidity can be presented, promoting transparency and collaboration among cash flow managers across subsidiaries.
- Cost–benefit analysis of switching to a machine learning model.
  While machine learning projects inherently involve some level of experimentation, they offer the potential for reduced human effort and cost savings in the long run. By improving planning quality and avoiding unnecessary borrowing or missed investment opportunities, machine-based forecasting can yield positive monetary effects.
- Can this result at one firm to show that machine-based approach outperformed the classical approach be generalized?
  The effectiveness of a machine-based approach compared to classical techniques may vary depending on the characteristics of the firms and the timeseries of data collected and being analysed. Factors such as volatility, seasonal patterns, and data availability can influence the performance of these models, impacting their generalizability.
- Which firm characteristics make ML models unsuitable for liquidity planning?

Machine learning models may not be suitable for a firm with minimal monetary impacts due to low liquidity numbers or a small amount of forecasted time series. The same holds for firms with a moderate forecast accuracy improvement due to weak and continuous liquidity, e.g., 1 Mio (± 1%) every month.