From Manual Analysis to Predictive Automation: A Case Study in Financial Forecasting

Eduardo Menezes de Souza Amarante¹*, Laverty Diderone de Assis Ladeia¹
¹SENAI CIMATEC University; Salvador, Bahia, Brazil

This paper presents the design and implementation of a business intelligence solution developed to automate, optimize, and scale financial forecasting processes in a multi-organization environment. The initiative was applied within the FIEB system, encompassing entities such as SENAI, SESI, IEL, and CIEB. The proposed solution integrates a data lake hosted in Microsoft Azure, a complete statistical and machine learning modeling pipeline, and an interactive dashboard built in Power BI to support data-driven decision-making and strategic financial planning. Initially, historical financial records were processed through a robust cloud-based ETL pipeline developed with Azure Data Factory. This pipeline performed automated cleaning, standardization, type correction, and monthly aggregation of accounting transactions. To define appropriate modeling strategies for each time series, we applied statistical analyses including signal-to-noise ratio (SNR) calculations and outlier smoothing techniques such as winsorizing and moving averages. Based on these insights, ARIMA and SARIMA models were trained to predict future values for each accounting item across the institutions. A custom API was developed to deliver prediction results and evaluation metrics to the Power BI dashboard, which updates automatically every month. The dashboard aggregates the accounting items by economic nature and displays six-month rolling forecasts, enabling financial managers to make proactive and informed decisions. As a result, the solution reduced manual workload, increased forecast accuracy, incorporated seasonal variations, and cut execution time by 90%. This project exemplifies how cloud computing, automation, and statistical modeling can modernize financial workflows and improve decision-making quality in complex organizational ecosystems.
Keywords: Cash Flow Predictions. Items Account. Automated Process. Data-Driven Decision Making. Machine Learning Models (ML Models). Microsoft Azure.

Data-driven decision-making (DDDM) constitutes a methodological approach in corporate finance whereby strategic and operational decisions are derived from systematic data analysis, as opposed to intuition or subjective reasoning [1]. This approach is very important to identify opportunities for improvement and critical points in business. Companies realized that data is a strategic asset with potential drive innovation and can optimize resource allocation. Business Intelligence changed the management process, bringing value to the company, adding reliability as the information is produced constantly and automatically through an integrated system [2].

However, DDDM requires not only an appropriate technological infrastructure, but also an organizational culture that recognizes data as a strategic asset. In this study, we discuss the development of the infrastructure required to implement DDDM at FIEB in the context of corporate finance management, and we present the findings of this innovative approach within the company.

Materials and Methods

The original data source of the company contains historical financial data from five organizations, such as FIEB, SENAI, IEL, SESI and CIEB. The data contains transactions related to accounting items such as personnel expenses, consumable materials, transportation and travel, other payments, service revenues, direct collections, etc.

First, we built a data lake on Azure Cloud with four layers to store raw and preprocessed data, following best practices of data governance.

Second, we developed a pipeline for extracting, transforming, and loading data using Azure Data Factory (ADF). During the transformation process, we cleaned the data and applied business rules to ensure process reliability, data quality, and to reflect the operational reality of the company. A monthly incremental data extraction from the same source was configured using the same Azure service.

In addition, we developed a machine learning pipeline to train one predictive model for each account item in each organization. In total, forty-eight models were created. This pipeline was also responsible for training, evaluating, and saving the machine learning models in the Azure repository.

Moreover, we developed a custom API from scratch, which plays a central role in enabling data integration and automation within the business intelligence pipeline. It was developed to feed the end-user dashboard with not only the predictions results, but also the historical data and the evaluation metrics. Both ML model training pipeline and the API were coded in Python and deployed using Azure Container. Finally, we created a Power BI dashboard to support DDDM for managers. We automated the entire process, from data extraction to dashboard updates, to meet the business requirement of delivering monthly updated forecasting of cash-flow for six months ahead. Table 1 summarizes and compares the characteristics of BI solution with procedures before the solution. Figure 1 presents the architecture of the BI solution.

Table 1. Comparison between methods.

Method	Traditional	Creative Innovation
Procedures	Manual	Automated
Staff Involved	One or more	None
Forecasting Method	Simple average	Machine learning
Seasonal Events Inclusion	Not	Yes
Time Consumption	Several days	Few hours

Figure 1. Architecture of the BI solution.
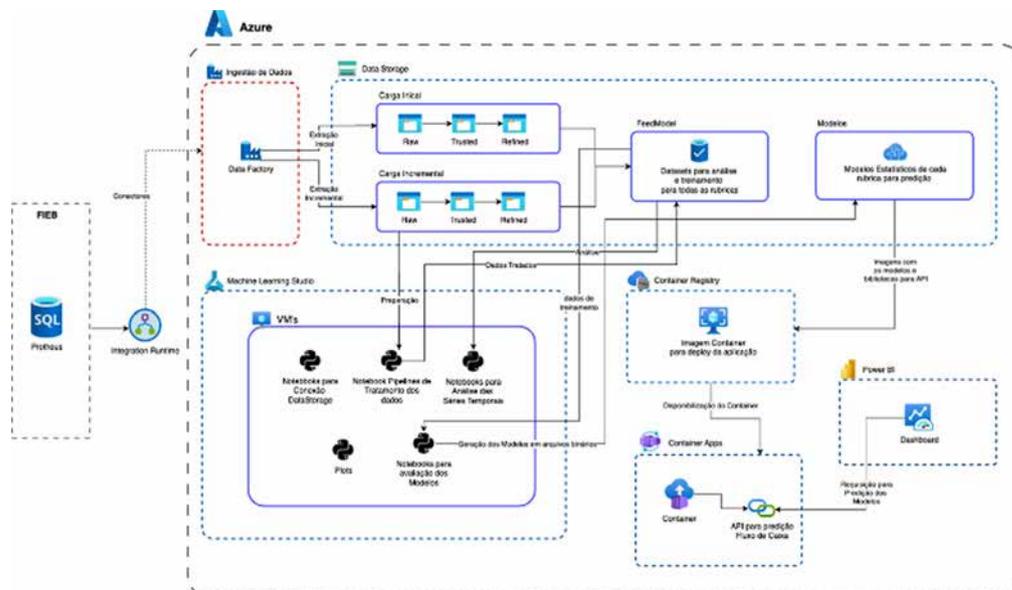
Preprocessing Data

We initially cleaned the raw data, deleted duplicated records, removed useless columns, and corrected their data types. Secondly, we aggregated the records and summed the transactions' values by month. The results were stored in the following data lake layers. All the process was done using a data pipeline with ADF.

Statistical Analysis

We performed statistical analysis to understand each data collection and to find patterns and outliers. We calculated the signal-to-noise ratio (SNR) for each accounting classification using the following formula:

$$SNR = \frac{VAR(Signal)}{VAR(Noise)}$$

In the formula above, VAR means a variance, signal means the sum of values of seasonality and trend for each time series, and noise means the unpredictable component of time series obtained after the decomposition of time series [3,4]. This type of analysis allowed us to identify which rubric would be modeled using either the whole data or only the trend component. The ratio of amplitude of the signal to error is called the signal-to-noise ratio (SNR); the larger the SNR, the easier it is to detect the signal. SNR > 1 means signal well defined, and we can model the time series using whole information, and SNR < 1 means more noise than signal; we need to extract the trend to perform the model.

We also identified and smoothed the outliers using several statistical techniques such as moving average, median, and winsorizing.

Machine Learning Training Pipeline

Due to the amount of data for each accounting item, which was between sixty and eighty, we decided to use statistical methods such as ARIMA and SARIMA to develop each model. During the exploratory and prior predictive analysis, we created a file that contains each characteristic of time series, which is used to guide the training, such as if we would use the trend or all components, ARIMA or SARIMA, expenses or revenue, etc. The pipeline follows the instructions given by the file guide. At the end of the pipeline, each model is saved in a repository on Azure.

ARIMA is the acronym for AutoRegressive Integrated Moving Average. This type of model is widely used to analyze and predict short time series. There are three main components in ARIMA model: AR (AutoRegressive – p), which captures the linear dependency between the current value and past values; I (Integrated – d), which indicates the number of times the data must be differentiate to achieve stationarity; and MA (Moving Average – q), which represents the linear dependency between the current value and past error values. SARIMA is an extension of the ARIMA model that incorporates seasonal components, meaning that certain patterns occur in regular cycles [5].

Integration

We created an API from scratch to feed Power BI dashboard with the results of predictions. Besides, API also carries the historical data and metrics of evaluation of models, such as root mean square error, mean square error, mean absolute percentage error. This API downloads the models into the application to avoid delays during requests. It was designed to update automatically the models once a month after their updating, and a security redundance layer was added to ensure high availability and fault tolerance.

Power BI Dashboard

We developed a dashboard designed to meet business needs by applying data visualization best practices, supporting data-driven decision-making, and ensuring alignment with user requirements and organizational objectives. We grouped and aggregated the accounting items by economic nature to calculate the cash flow and predicted the

cash flow for six months ahead. So, the decision maker can anticipate and make decisions to manage better resources company, aiming to improve profitability and minimize risks.

Results and Discussion

After implementing the solution, we improved the quality of the predictions, by increasing accuracy and incorporating seasonal components for some accounting items, since the methodology used initially was the simple average of the last months. In addition, the solution reduced the time of analysis due to all steps being automated reducing not only the execution time to achieve the results but also operational costs reduction. The execution time was reduced by approximately 90%. Finally, the forecasting horizon expanded from ten days to six months.

Conclusion

The solution built has proven to be useful and highly important for the decision-making process, as it reduced both execution time and costs. All automated steps ensure that task are executed consistently and accurately, saving days due through faster execution, reducing man-hours for repetitive tasks, providing easy scalability to handle large volumes of data or operations, and allowing professionals to focus on analysis, innovation, and decision-making rather than operational activities.

References

1. Zaitsava M, Marku E, Di Guardo MC. Is data-driven decision-making driven only by data? When cognition meets data. Eur Manag J. 2022;40(5):656–70.
2. Boalento GP, Costa SA. A influência do Business Intelligence para a gestão das empresas na perspectiva de gestores e controllers. In: 22º USP International Conference in Accounting; 2022 Jul 27-29; São Paulo.
3. Chatfield C. The Analysis of Time Series: An Introduction. 6th ed. Boca Raton, FL: Chapman & Hall/CRC; 2003.
4. Shumway RH, Stoffer DS. Time Series Analysis and Its Applications: With R Examples. 4th ed. New York: Springer; 2017.
5. Peixeiro M. Time Series Forecasting in Python. Shelter Island, NY: Manning Publications; 2022. 456 p. ISBN 978-1-61729-988-9.