NOVA IMS Information Management School

MDDM

Master's Degree Program in
Data-Driven Marketing

Forecasting Demand in the Pharmaceutical Industry Using Machine Learning

João Aires Lancastre de Sousa Cabral de Ascensão

Internship Report

presented as a partial requirement for obtaining the master's degree Program in Data-Driven Marketing with a specialization in Data Science for Marketing

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

FORECASTING DEMAND IN THE PHARMACEUTICAL INDUSTRY USING MACHINE LEARNING

por

João Aires Lancastre de Sousa Cabral de Ascensão

Projeto apresentado como requisito parcial para obtenção do grau de Mestre em Data-Driven Marketing, com especialização em Data Science for Marketing

Orientador: Rui Alexandre Henriques Gonçalves

July 2023

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration. I further declare that I have fully acknowledged the Rules of Conduct and Code of Honor from the NOVA Information Management School.

Lisbon, 15th of July 2023

ABSTRACT

This study delves into the exploitation of three machine learning models, namely the Extreme Gradient Boosting (XGBoost), the Long Short-Term Memory (LSTM), and the novel Prophet algorithm, to surpass the challenge of demand forecast within the pharmaceutical industry. Following the CRISP-DM framework, we enabled accurate sales forecasting by studying, treating, transforming, and training a dataset containing historical sales data from a major Portuguese pharmaceutical company. Our findings align with the literature, underlying the robustness of the XGBoost and the inefficacy of the LSTM for the delineated task, considering the singularities of the provided data. Furthermore, this research highlights the potential of the Prophet for both its effectiveness and efficiency. This endeavor allowed us to reinforce the literature's conviction of the need for product-specific forecasting, showcasing that no single model achieves the best accuracy for all drugs.

KEYWORDS

Demand Forecasting; Pharmaceutical Sales; Machine Learning.

Sustainable Development Goals (SGD):

INDEX

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1 INTERNSHIP CONTEXT

Capgemini is a multinational Information Technology Consulting company founded in 1967 that operates in 50 countries and has over 325,000 team members worldwide. In Portugal, it has over 3,600 workers. The company has six different brands and focuses on seven distinct areas: Cloud, Cybersecurity, Customer Experience, Data and Artificial Intelligence, Enterprise Management, Sustainability, and Intelligent Industry. For the development of the internship report, the author joined Capgemini in June 2022, pursuing a 6-month internship within the Advanced Analytics team. During this period, the team was given a project with a pharmaceutical company and the opportunity to develop a tailor-made solution to assist the client in making demand forecasting and future market predictions more accurate and efficient.

1.2 STUDY RELEVANCE AND IMPORTANCE

In any industry, sales prediction is crucial to ensure supply chain efficiency. For a pharmaceutical company, the importance of this analysis is even more significant, because the company maintains an indirect consumer relationship that depends primarily on touchpoints that affect the end-to-end customer journey (Ascher et al., 2018). A discrepancy between supply and demand can lead to excessive inventory of short-shelf-life medicines (Keny et al., 2021) or a shortage of drugs that can result in the loss of market to the competition (Zhu et al., 2021).

Pharmaceuticals deal with many products represented by singular codes (referred to as AIM in the Portuguese market), organized in subcategories based on the common active agent into “homogeneous groups.” Each drug may have different dosages and delivery methods and, thus, a different demand record with singular behavior. This specificity of the pharmaceutical market means that the forecasting analysis should be univariate to achieve greater accuracy. While forecastings can be made through traditional tools such as Excel, being limited to this tool would result in unreliable results, a dependence on experiences, and thus inefficiency on top of suboptimal performance (Zhu et al., 2021).

Today, machine learning emerges as a crucial business tool for any business to reduce the mismatch between supply and demand, decrease sales cycles, identify clients at risk to churn, and design intervention campaigns to ensure the extension of the customer’s lifetime (Rosário et al., 2021). Numerous Pharmaceutical companies, such as Pfizer and AstraZeneca, are already exploiting the power of Machine Learning to design better-targeted multi-channel marketing campaigns and maximize optimal resource allocation and market growth (Khanna et al., 2020).

The developed work will contribute to the body as research, by showcasing how theoretical concepts can have a solid implementation within specific business cases. With an in-depth report on the transition from the academic mindset to the real business world, this project will prove that this methodology is not just relevant but essential for an up-to-date business to thrive in the long term.

1.3 PROBLEM AND MAIN GOALS

As a major pharmaceutical company within the Portuguese market, the client's product portfolio is vast. Until the project was assigned to Capgemini, the client's monthly demand forecasting, which concerned more than a thousand drugs, was solely relying on market assumptions and on the exploitation of Excel, leading to significant inaccuracy and efficiency loss. With a machine learning approach in mind, a project proposition was designed to address demand forecasting. The goal is to assist the client in reducing sales cycles and ultimately making deliberate business decisions in the future, allowing supply chain and marketing efficiency.

Throughout this study, we will build three machine learning models to predict the pharmaceutical's demand over a six-month period for each specific drug listed by the company. The three models will be evaluated based on historical data and we will analyze and select the best approach to efficiently meet the client's needs. Moreover, this will allow the supply and marketing teams to act in accordance with the forecast and refine every related business decision. The factory and the warehouse will be able to improve drug production and distribution management, and the marketing department to design strategies that sustain the customer lifetime value, reduce churn, and decrease seasonality. Both departments will approach the market with informed business decisions enabled by accurate knowledge (Rosário et al., 2021).

2. LITERATURE REVIEW

2.1 FORECASTING DEMAND IN THE PHARMACEUTICAL INDUSTRY

Demand forecasting can be defined as estimating future demand for goods or services based on past and present data and market trends. This process aims to assist in making strategic decisions, such as production planning, to ensure that the organization can meet its customers' needs (Khan et al., 2020). Forecasting establishes a relationship between a firm and its environment, which is essential for effective planning and decision-making. For instance, organizations can anticipate future challenges and take necessary measures by forecasting demand. Demand forecasting is also a fundamental aspect of many areas of supply chain management, in particular demand planning, supply control, production forecasting, and order fulfilment (Bonab, 2022).

Forecasting models can be organized into various categories based on their logic and application. One way to differentiate them is by their methodological approach, which can be either qualitative or quantitative. Qualitative methods use expert knowledge and verbal statements to make predictions, while quantitative methods use mathematical rules and principles to forecast future outcomes based on historical numerical data (Moroff et al., 2021). Another way to classify forecasting models is by the forecast period, which can be short-term, medium-term, or long-term. Quantitative forecasting methods are often used for short and medium-term forecasting and can employ statistics, machine learning, and deep learning techniques. Furthermore, forecasting models can also be distinguished by their level of complexity. Univariate models analyze a single variable, while multivariate models consider multiple variables that might influence the outcome (Moroff et al., 2021).

Companies face increasing pressure to optimize their supply chain performance in today's business environment. It is crucial for companies to adopt a proactive approach to demand forecasting to mitigate the challenges posed by the dynamic nature of market demands, which can lead to short-term changes in supply chain planning (Moroff et al., 2021). In order to make informed decisions, decision-makers must conduct a thorough assessment of the costs and benefits of alternative expenditures and investments. This assessment can be facilitated by utilizing decision support models that relate costs to customer purchase behavior and forecast the value of the customer portfolio. These models are instrumental in identifying the optimal allocation of resources to marketing and sales actions over time (Martínez et al., 2020).

Anticipating future customer behavior plays a pivotal aspect in effectively managing a business. In particular, it is essential for the sales and marketing departments as it provides vital information for allocating resources efficiently (Martínez et al., 2020). Moreover, machine learning methods appear ideal for predicting future demand and providing valuable supply chain management insights (Schreckenberg & Moroff, 2020).

Research has shown that machine learning models for time series forecasts can keep highly competitive results, often surpassing classical statistical models (Bonab, 2022). By utilizing advanced forecasting techniques such as statistical modeling or machine learning or incorporating real-time data into the forecasting process, companies can improve the accuracy and reliability of their demand forecasts (Moroff et al., 2021). Researchers have proposed specific machine learning outlines, such as a framework for customer purchase prediction in non-contractual settings (Martínez et al., 2020), and found that stockouts could be avoided by exploiting machine learning. When analyzing time-series data, a widely used approach is decomposing the data into its parts. Wold's decomposition theory states that each time series comprises four components: trend, seasonal, cyclical, and residual. These components can interact in various ways, resulting in diverse demand patterns. The combination of these components can be additive or multiplicative, ultimately determining the demand patterns' nature and enhancing the importance of choosing suitable methodologies and ensuring that the components are derived from reliable data. (Moroff et al., 2021).

Accurate data is crucial for precise forecasting, and any inaccuracies in the data collection process can lead to incorrect predictions and financial loss for the company. It is crucial to compare forecast data with real-time data to determine the accuracy of the predictions (Khan et al., 2020). Furthermore, several studies have revealed that different demand patterns can significantly impact the forecast quality of various models. Therefore, it is imperative to evaluate models based on forecast error and implementation efforts to identify the most appropriate model for a specific use case (Moroff et al., 2021). By utilizing analytical tools such as decision support models, businesses can make strategic decisions more likely to lead to favorable outcomes. Recent research has demonstrated that when evaluating forecasting models, it is paramount to consider both the forecast accuracy and the practicality of implementation. The research on this topic can provide a deeper understanding of the factors influencing customer behavior and the methods used to predict it (Martínez et al., 2020).

In addition to machine learning techniques, some studies have emphasized the importance of evaluating a company's maturity level before implementing machine learning-based forecasting methods. This is crucial for success when dealing with uncertainty in prices, markets, competitors, and customer preferences (Khan et al., 2020; Schreckenberg & Moroff, 2020).

Several analyses have been conducted using machine learning for demand forecasting in various industries. These studies have employed various approaches, such as the use of business intelligence empowered with machine learning (Khan et al., 2020) and the comparison of different machine learning models such as Seasonal Auto-Regressive Integrated Moving Average Extended (SARIMAX), Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Long-term short-term memory (LSTM) as well as hybrid models (Moroff et al., 2021). The results of these studies have consistently shown that machine learning plays a significant role in improving the accuracy of sales predictions and reducing the risk of stockouts, ultimately increasing customer satisfaction.

The supply chain of pharmaceutical products is known for its complexity, with limited and highly regulated channels for customer delivery. This complexity poses significant obstacles when improving performance and efficiency. Additionally, the potential negative impact on public health further emphasizes the importance of accurate forecasting and the avoidance of stockouts (Merkuryeva et al., 2019). Like other manufacturing sectors, the pharmaceutical industry relies on mass production to meet consumer demand. However, the industry also faces unique challenges in forecasting demand and managing inventory. Inaccurate demand calculation and inadequate preparation for pandemics can lead to loss of lives. Accurately forecasting the demand for life-saving drugs is of utmost importance, as it aligns with the directives mandated by regulatory authorities on a global scale. (Siddiqui et al., 2022).

The pharmaceutical industry is unique in its complex and intertwined supply chain activities. An effective forecast model is essential for pharmaceutical companies to match supply with changes in demand while minimizing inventory levels (Rathipriya et al., 2022). Achieving accurate demand forecasting in the pharmaceutical industry is a complex task influenced by various external factors. These factors encompass seasonal and epidemic diseases, fluctuating rates of active ingredients, human-related variables, market shares of competing products, and prevailing marketing conditions. The pharmaceutical sector is particularly vulnerable to forecast errors due to dynamic factors like regulatory changes, product introductions, technological advancements, and the preferences of healthcare professionals and pharmacists towards rival brands. Such inaccuracies in forecasting can lead to detrimental consequences, including stockouts and excessive inventory levels. (Siddiqui et al., 2022).

Researchers have tested several models, such as Simple Moving Average (SMA), multiple linear regressions, and symbolic regression with genetic programming (Merkuryeva et al., 2019), Polynomial Regression, and Support Vector Regression (SVR) (Keny et al., 2021) to improve inventory management and sales predictions, respectively. Proving that state-of-the-art forecasting techniques actually help companies accurately predict demand and make informed decisions about production and supply chain operations. This not only helps to overcome supply chain complexities but also provides a competitive advantage. (Rathipriya et al., 2022). The ongoing COVID-19 pandemic has further highlighted the importance of an accurate forecasting tool. Sudden outbreaks have led to rising demand, and lockdowns have affected supply predictability. In this context, efficient forecasting models become even more critical as they help companies stock or produce the correct quantity of products (Rathipriya et al., 2022).

For all the mentioned reasons, Demand Forecasting within the Pharmaceutical industry appears critical in achieving supply chain competitiveness and building relationships between suppliers and customers. Thus, the pharmaceutical industry must use advanced technologies and data-driven approaches to improve forecasting processes and supply chain efficiency to thrive in a competitive and complex industry (Siddiqui et al., 2022).

2.2 Selected Models

In the study "A machine learning framework for customer purchase prediction in the non-contractual setting" (Martínez et al., 2020), Lasso, Extreme Learning Machine, and XGBoost algorithms were compared. The results showed that XGBoost, specifically the gradient tree boosting algorithm, outperformed the other methods in accuracy and predictive power and these findings were consistent with the literature in the contractual setting, indicating that XGBoost's superior performance extends to both contractual and non-contractual customer purchase prediction scenarios.

In “Machine Learning and Statistics: A Study for assessing innovative Demand Forecasting Models” (Moroff et al., 2021), the data was analyzed with a monthly dataset, and the results led to the conclusion that between SARIMAX, ETS, Random Forest, XGBoost, LSTM, and MLP, all different models yield distinct forecast qualities and therefore different results. The LSTM and the MLP, together, had the best overall performance. However, the statistical and machine learning algorithms proved to be as efficient in getting the best forecast performance in half of the products analyzed. The machine learning methods proved to be much more effortless in implementation, while the Deep Learning algorithms resulted in more computational and data preparation effort.

In “Demand forecasting model for time-series pharmaceutical data using shallow and deep neural network model” (Rathipriya et al., 2022), it was also concluded from the conducted study that no single demand forecasting model is ideal for all the analyzed drugs. It also states that deep neural networks such as the LSTM did not show the best forecasting performance, possibly because the dataset used for this model was smaller.

From the mentioned literature, we can draw some hypothetical conclusions that we can further analyze. In light of these studies, we will examine whether a single model consistently outperforms others in predicting sales across different drugs. Moreover, it is crucial to evaluate the suitability of XGBoost for the task at hand and assess the value of LSTM despite its additional computational and data preparation requirements. In addition to XGBoost and LSTM, we will also consider the Prophet algorithm to gain insights into its potential strengths and suitability for sales prediction.

This comprehensive analysis will enrich our understanding of the value of machine learning in demand forecasting and enable us to draw meaningful conclusions. By delving deeper into the performance of these models, we can provide valuable insights for decision-making in the pharmaceutical industry and uncover the most effective approaches for accurate sales forecasting.

3. METHODOLOGY

3.1 CRISP-DM Framework

The Cross-Industry Standard Process for Data Mining (CRISP-DM) is a comprehensive, adaptable, and structured framework for data mining projects. This framework comprises six interrelated phases that offer flexibility while maintaining a logical sequence (Chapman et al., 2000). Over the years, CRISP-DM has maintained its prominence and relevance, being acknowledged as the most comprehensive methodology for meeting the needs of industrial projects and the de facto standard for analytics, data mining, and data science projects (Martinez-Plumed et al., 2021).

The first phase, Business Understanding, emphasizes comprehending a project's objectives and requirements from a business perspective. This understanding is then translated into a well-defined data mining problem and a preliminary plan. Subsequently, the Data Understanding phase involves the initial data collection and activities that familiarize practitioners with the data. These activities enable the identification of data quality issues, the acquisition of initial insights, and the discovery of exciting subsets.

Data Preparation comes into play as the process advances, during which raw data is transformed into the final dataset for modeling. This phase includes attribute and record selection, data transformation, and cleaning. Once data preparation is complete, the Modelling phase takes place. Various techniques are selected, applied, and calibrated in this phase for optimal performance. Revisiting the Data Preparation phase may be necessary if the selected techniques have specific data requirements.

The subsequent Evaluation phase assesses the model's quality and alignment with business objectives, identifying any overlooked critical business aspects. This assessment leads to a decision on whether to use the data mining results. Lastly, the Deployment phase focuses on organizing and presenting the acquired knowledge to the customer in a usable format, from generating a report to implementing a repeatable data mining process across an organization (Chapman et al., 2000).

In the context of demand forecasting and churn prediction models, the goal-oriented nature of CRISP-DM, along with its focus on processes, tasks, and roles, proves particularly effective for addressing the challenges associated with these specific applications. By prioritizing extracting value from data, CRISP-DM ensures practitioners can effectively address the challenges associated with demand forecasting models in a structured and scalable way (Martinez-Plumed et al., 2021).

TIMELINE

The project's timeline was carefully planned, considering the pharmaceutical's objectives, available data, and the CRISP-DM framework. The timeline structure is summarized in the following table, indicating the start and end dates and the duration in weeks for each stage. With this time frame, we were able to ensure the timely completion of each phase and guarantee a structured workflow throughout the project.

Stage	Start Date	End Date	Duration (weeks)
Business Understanding	2022-10-03	2022-10-14	2
Data Understanding	2022-10-17	2022-10-28	2
Data Preparation	2022-10-31	2022-11-25	4
Feature Engineering	2022-11-28	2022-12-09	2
Modeling	2022-12-12	2022-12-30	3
Model Evaluation	2023-01-02	2023-01-13	2
Validation Strategy	2023-01-16	2023-01-27	2
Deployment Strategy	2023-01-30	2023-02-10	2
Total	2022-10-03	2023-02-10	19

Table 1. Project's Timeline (Source: Author)

3.2 DEMAND FORECASTING MODEL

3.2.1 Business Understanding

Our client, a leading pharmaceutical company in Portugal, has requested assistance to improve their current demand forecast process. They face considerable challenges in accurately forecasting demand for their vast product portfolio of over 1000 products, relying heavily on industry knowledge and experience-based practice. This manual and time-consuming approach involves using Excel tools, which leads to inefficiencies and reduces overall productivity.

Our project aims to develop a data-driven approach using machine learning techniques that can accurately predict demand, reduce reliance on industry knowledge and experience-based forecasting, and eliminate the dependency on Excel tools. By doing so, we aim to improve production planning and inventory management and reduce production costs for our client. To achieve these goals, we will work closely with the client to comprehensively understand their current demand forecasting process, identify pain points, and design a machine learning model that meets their specific needs and ensures they remain competitive in the pharmaceutical industry.

3.2.2 Data Understanding and Preparation

The data understanding and preparation process was done in *Jupyter* notebooks using *Python* programming. We utilized a variety of libraries and packages, including *os, csv, numpy, pandas, matplotlib, seaborn, category_encoders, scikit-learn, xgboost, prophet, statsmodels, plotly, holidays, LSTM,* among many others. These tools enabled us to explore, analyse, and clean the data to prepare it for use in machine learning models.

An Excel file was provided by the client, containing structured data spread across seven columns: *'date'* (representing the dates), *'aim'* (representing product ID), *'gh'* (homogeneous group ID), *'sellout_value'* (sales in value), *'sellout_units'* (sales in units), *'stock_pharma'* (stock in pharmacies), and *'stockout'* (representing stockout events). The data set consists of 46,872 rows and seven columns, covering the period from January 2014 to January 2023. The last six months were designated the test set, while the remaining data was used for training. Of the over 1,000 products in the client's portfolio, 654 were included in the initial data set, with 30 products ultimately selected by the client to evaluate the machine learning model.

During the data cleaning process, we dropped 12 duplicates and feature-engineered the *'stockout'* variable into a binary variable, where 0 represents missing data and the inexistence of a stockout event, and 1 represents a stockout event. To ensure the machine learning model follows the sales trend, we replaced *'sellout_units'* with an average of sales every time there was a stockout event. We performed the ADF test to check the stationarity of the data, as a stationary time-series has statistical features that remain constant over time, making forecasting easier for some models (Ensafi et al., 2022). Initially, we tested the overall dataset and found that it was stationary, with an ADF statistic of -217.447 and a p-value of 0.000. However, when we tested individual products, we found that many of them had non-stationary data.

To better understand the underlying patterns and make more accurate predictions, we decomposed the time series data into its components: trend, seasonality, and noise. However, this was not performed individually for each product due to the number of products. It was instead performed on the 30 selected products.

All available features were used in the XGBoost and LSTM models, while only *'sellout_units'* and *'date'* were used for the Prophet algorithm. Through careful data understanding and preparation, we were able to ensure the suitability and quality of the data used in the machine learning models.

3.2.3 Modeling and Evaluation

3.2.3.1 Model Selection

The Prophet algorithm is a recent open-source time-series algorithm developed by Facebook. Based on the Bayesian curve fitting technique, this algorithm is by default an additive model that excels at generating high-quality forecasts. The model is represented by the additive regression equation $y(t) = g(t) + s(t) + h(t) + \varepsilon t$. The *y(t)* represents the output value, *g(t)* the trend function, *s(t)* depicts periodic changes, *h(t)* holiday effects, and *e(t)* the noise in the data (Chaturvedi et al., 2022).

The trend component, *g(t)* can be represented through a linear piecewise or a saturating growth model. While the seasonal component *s(t)* is modeled using a Fourier series that accounts for cyclic changes introduced during weekly, monthly, or annual cycles and is represented by the equation *s(t) = ∑N n=1 (an * cos(2πnt/P) + bn * sin(2πnt/P))* (Chaturvedi et al., 2022).

The determination of the holiday component, *h(t)*, within Prophet, is contingent upon the utilization of an indicator function. This function serves to identify whether a particular point in time, denoted as *t*, aligns with a holiday event. Accompanying each holiday is a dedicated parameter, *ki*, which assumes responsibility for modifying the forecast in a manner that aligns with the corresponding holiday's influence. By incorporating this mechanism, Prophet empowers its forecasting capabilities to accurately capture and accommodate the unique dynamics introduced by holiday occurrences (Chaturvedi et al., 2022).

The seasonal components and the different range of values that can be attributed to each parameter allow for greater prediction accuracy, by allowing the analyst to adapt the model in accordance with the data and the underlined patterns that characterize it. Selecting the seasonality mode that can be either additive or multiplicative, adjusting the seasonality prior scale within a range of 10 to 25, or modifying the Fourier component, allows a greater capacity of capturing seasonality (Ensafi et al., 2022).

The model's flexibility allows analysts to accommodate multiple seasonal periods and make different assumptions about trends. Its fast-fitting capability enables quick implementation of numerous specifications. Furthermore, the easily interpretable parameters empower specialists to modify the model to impose assumptions on forecasts, allowing them to build complexity based on their needs, preferences, or expertise (Taylor & Letham, 2018)

The Long Short-Term Memory (LSTM) model, in contrast to traditional feed-forward neural networks, employs a recurrent neural network architecture with feedback connections that facilitate the exchange of information from past instances to the current one. Its ability to compute derivatives relies on the backpropagation algorithm, which systematically propagates information from the final to the initial layer. This is particularly crucial as recurrent neural network architectures often face the vanishing gradient problem, where gradient information fails to propagate from the earlier layers to the current layer effectively (Chaturvedi et al., 2022). The LSTM architecture overcomes the vanishing gradient problem frequently encountered in traditional recurrent neural networks by efficiently learning and bridging long-time lags in the data. (Hochreiter & Schmidhuber, 1997).

LSTM units incorporate specialized components called "cells" or "memory blocks," which play a crucial role in capturing intricate dependencies among elements within an input sequence. These cells serve as a pivotal mechanism for information retention. To govern the flow of information, the model's gates are regulated using sigmoid activation functions. Specifically, the input gate manages the controlled influx of new values into the cell, while the output gate diligently governs the extent to which a value is retained and incorporated into the model's memory (Chaturvedi et al., 2022)

Throughout model training, the connections associated with the LSTM gates are assigned weights, some of which are recurrent. These weights are continuously updated over multiple training cycles (epochs) to enhance prediction accuracy. A forget gate and the addictive nature of cell state gradients enable LSTM to update connection weights effectively, significantly mitigating the vanishing gradient problem (Chaturvedi et al., 2022). This ability to handle long-term dependencies in time-series data makes LSTMs particularly relevant for forecasting tasks involving complex patterns and trends spanning significant periods.

The Extreme Gradient Boosting (XGBoost) is a powerful algorithm widely employed in various fields, that stands out as an exceptional tool for boosting prediction accuracy by combining multiple learning applications (Noorunnahar et al., 2023). Its underlying principle lies in the integration of weak models into a robust and high-performing predictor, achieved through iterative optimization using a gradient descent approach (Noorunnahar et al., 2023; Panarese et al., 2022).

The core of XGBoost lies in the objective function*, obj(y) = Σi L(^yi, yi) + Σk O(fk)*, which comprises two key components (Noorunnahar et al., 2023). The first term*, Σi L(^yi, yi)*, represents the loss function *(L)*, enabling the quantification of the disparities between predicted values *(^yi)* and observed values *(yi)* (Noorunnahar et al., 2023). This critical component serves as a measure to capture errors and refine the predictive accuracy of the model.

The second term, *Σk O(fk)*, denotes the regularization term *(O(fk))*, addressing the variation in output associated with newly integrated trees(Noorunnahar et al., 2023). By controlling the complexity and mitigating overfitting, this regularization term plays a pivotal role in shaping the model's performance. The optimization of the objective function empowers XGBoost to seamlessly amalgamate individual trees and minimize loss, ultimately enhancing the precision of predictions (Noorunnahar et al., 2023).

XGBoost exhibits remarkable mathematical prowess, enabling the development of a robust predictor while ensuring computational efficiency (Panarese et al., 2022). Notably, it possesses the capability to handle datasets with missing values, offering versatility in tackling various predictive tasks, including time-series forecasting (Chen & Guestrin, 2016). The fusion of XGBoost's performance and scalability makes it an invaluable resource for working with intricate datasets.

3.2.3.2 Feature and Parameter Selection

Prophet

As mentioned previously, when dealing with the Prophet algorithm, one must be aware that it can only recognize the target variable and a variable identifying the time factor. Accordingly, only the columns 'date' and *'sellout_units'* were selected for this model. The 'aim' variable was utilized to isolate each product in a new data frame and dropped along with the remaining columns, allowing the analysis to be tailor-made to each drug.

For the continuation of the modeling process, with the variables selected, the *'date'* defined as a date-time feature, the data was divided into the training and test sets, considering all historical data from July 2022 onwards, target values unseen by the training set where the model was fit. After fitting the model with the basic default parameters, some hyperparameter tunning was needed, in part due to some lack of experience in the field of time-series forecasting but primarily for efficiency reasons so the best combination of parameters could be found within a reasonable time, considering the number of products at hand. Considering the factors mentioned, a Random search allowed for finding a good set of parameters and creating some complexity in the model to improve its predicting power.

Parameter	Description	Value
Interval width	Width of confidence interval	0.95
Daily seasonality	Include daily seasonality	False
Growth	Method for time series growth	Randomly selected between "logistic" and "linear"
Changepoint prior scale	Scale of changepoint prior distribution	Randomly selected between 10^-3 and 1
Seasonality mode	Method for seasonality modeling	Randomly selected between 'additive' and 'multiplicative'
Holidays prior scale	Scale of holidays prior distribution	Randomly selected between 10^-1 and 10
Seasonality prior scale	Scale of seasonality prior distribution	Randomly selected between 10^-1 and 10
Seasonality period	Period of seasonality component	Randomly selected from [7, 30.5, 365.25]
Fourier order	Order of Fourier series components	Randomly selected between 1 and 20
Holidays	Holidays used for modelling effects	PT (Portugal)

Table 2. Prophet's selected parameters (Source: Author)

LSTM

In our deep learning model, we fed all features to the model to leverage its ability to find hidden correlations between variables and to observe if this gives the model an upper hand compared to univariate algorithms. We have carefully chosen a set of parameters to achieve optimal performance. The network architecture consisted of three LSTM layers, progressively decreasing in units (128, 64, and 32), complemented by a dense output layer. The ReLU activation function was consistently applied to effectively capture non-linear relationships. Dropout layers were excluded as their impact on performance proved negligible. To ensure seamless information propagation, the "Return Sequences" parameter was incorporated. Training the model involved a batch size of 32 across 200 epochs, implementing early stopping based on validation loss. The Adam optimizer, known for its reliable convergence, was employed, with the mean squared error (MSE) serving as the loss function. The project encountered challenges in analyzing multiple drugs, necessitating distinct neural network architectures for each drug type.

Layer	Output Shape	Number of Units	Activation Function	Return Sequences
LSTM (Layer 1)	(None, n_timesteps, 128)	128	ReLU	Yes
LSTM (Layer 2)	(None, n_timesteps, 64)	64	ReLU	Yes
LSTM (Layer 3)	(None, 32)	32	ReLU	No
Dense Output	(None, n_outputs)	-	None	-

Table 3. LSTM’s Architecture (Source: Author)

Parameter	Batch Size	Number of Epochs	Early Stopping	Optimizer	Loss Function
Value	32	200	Patience = 15	Adam	Mean Squared Error (MSE)

Table 4. LSTM’S additional parameters (Source: Author)

XGBoost

For the XGBoost algorithm, features were selected based on their pertinence to the research question and incorporated *'month'*, *'quarter'*, *'year'*, *'stock_pharma'*, *'stockout'*, *'sellout_units_gh'*, and *'sellout_ms'*.

A random search approach was employed to optimize the model's parameters for the parameter selection. This method allowed for efficient parameter space exploration, leading to a more robust model with improved performance. The chosen parameters included the *booster*, *n_estimators*, *objective*, *max_depth*, and *learning_rate*. Fine-tuning these parameters aimed to establish a balance between model complexity and predictive power, ultimately achieving a model that generalizes effectively to previously unseen data.

Parameter	Description	Value
Booster	Type of booster algorithm used for training	Randomly selected between gblinear, gbtree
n_estimators	Number of boosting rounds	Randomly selected between 500-2000
Objective	Loss function used for training	Randomly selected between reg:squarederror, reg:logistic
max_depth	Maximum depth of each tree in boosting process	Randomly selected between 2-10
learning_rate	Learning rate or step size for each boosting round	Randomly selected between 0.01, 0.1, 0.001
early_stopping_round	Number of rounds without improvement to trigger early stopping	Randomly selected between 20, 50, 100
eval_metric	Evaluation metric used to assess model's performence	Rmse, logloss

Table 5. Selected parameters XGBoost (Source: Author)

3.2.3.3 Evaluation

Model accuracy plays a crucial role in making sure the algorithms are properly designed and structured, with the optimal set of parameters that enable good generalizations of the data available (Ensafi et al., 2022). To ensure the model's reliability, we utilize four commonly used evaluation metrics, such as the Mean Absolute Error (MAE), the Mean Percentage Error (MPE), the Root Mean Squared Error (RMSE) that help us get an assessment on the magnitude of the error and the Mean Absolute Percentage Error (MAPE) which gives us the percentage deviation between the predicted and actual values (Noorunnahar et al., 2023). These four metrics allowed us to fine-tune the three models and unveil their behavior along the training process and ultimately contributed to the accomplishment of the best possible outcomes.

Metric	Formula	Description
MAE	Σ|Actual - Forecast| / n	Average absolute difference between predicted and actual values, measuring the magnitude of forecasting errors.
MPE	Σ(Actual - Forecast) / ΣActual * 100	Average percentage difference between predicted and actual values, quantifying the relative deviation in percentage terms.
RMSE	√Σ(Actual - Forecast)^2 / n	Overall measure of prediction error, considering both magnitude and direction of errors.
MAPE	Σ|(Actual - Forecast)| / ΣActual * 100	Average percentage deviation between predicted and actual values, indicating relative forecasting accuracy.

Table 6. Evaluation metrics description (Source: Author)

We will only consider 24 of the 30 products for the evaluation phase due to stockouts in the target months for 5 products and 1 product having substantial missing values, with 16 months of sales data missing, making it unsuitable for accurate forecasting. By focusing on the remaining 24 products, we can ensure a more reliable and accurate comparison of the forecasting models' performance. For interpretation purposes, we will focus on the MAPE metric throughout the Results and Discussion chapters, as it gives a clearer and more attainable reading of the output.

4. RESULTS

This chapter presents an in-depth examination of the results obtained from applying the three forecasting models—XGBoost, Prophet, and LSTM— to the 24 selected drugs for evaluation. The output data from two selected products, Product 14 and Product 9, are analyzed in greater detail to illustrate the trends observed in the results. A series of figures and graphs have been prepared to support the analysis.

Figure 1 displays the MAPE values for the three models across the 24 products, revealing a consistent pattern of high MAPE values for the LSTM model. This suggests that the LSTM model exhibits a lower accuracy in forecasting when compared to the XGBoost and Prophet models. The XGBoost model demonstrates a robust performance, achieving the lowest MAPE values for eighteen of the twenty-four drugs. Meanwhile, the Prophet model secures the lowest MAPE for six products. Figures 2 and 3, which showcase the MAPE values for Product 14 and Product 9, further substantiate this observation.

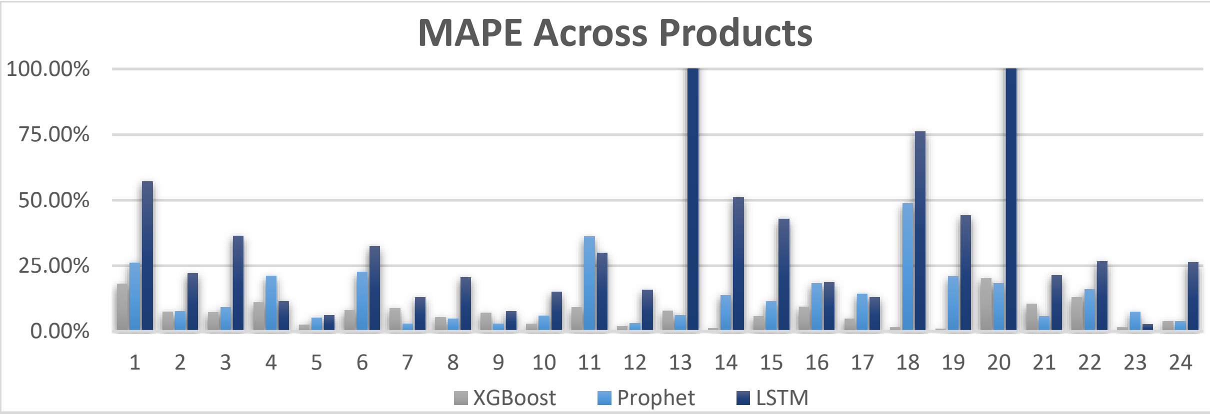

Figure 1. MAPE values across the 24 drugs (Source: Author)

Product	XGBoost	Prophet	LSTM
14	1.12%	13.76%	51.09%

Figure 2. Product nº 14- MAPE values (Source: Author)

Product	XGBoost	Prophet	LSTM
9	7.08%	2.89%	7.60%

Figure 3. Product nº 9- MAPE values (Source: Author)

Figures 4 and 5 visually represent the Prophet model's predictions for Product 14 and Product 9, respectively. These figures highlight the model's capacity to capture the seasonality and trends present in the sales data. Figures 6 and 7 display line plots of the actual and predicted values for each model applied to both products. These plots illustrate the variation in the performance of each model across different products, suggesting that the accuracy of the forecasting models may be influenced by specific product characteristics such as seasonality or demand patterns.

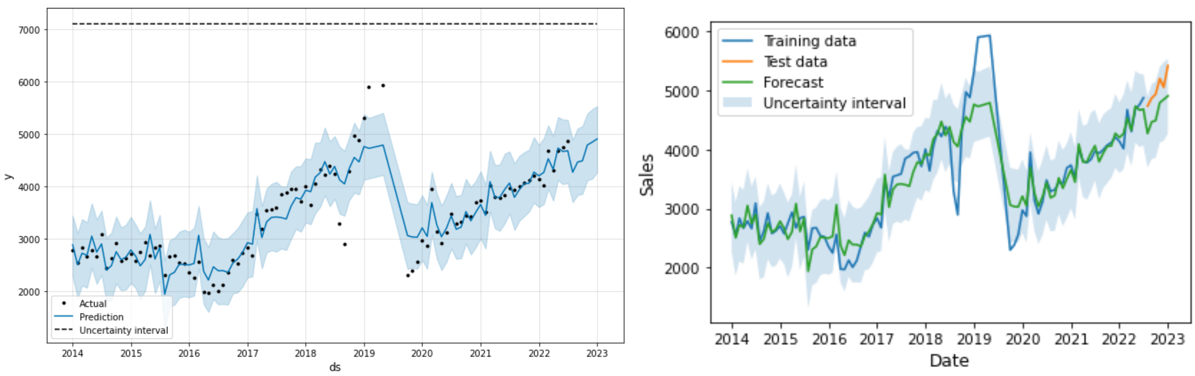

Figure 4. Prophet graphs- Product nº 14 (Source: Author)

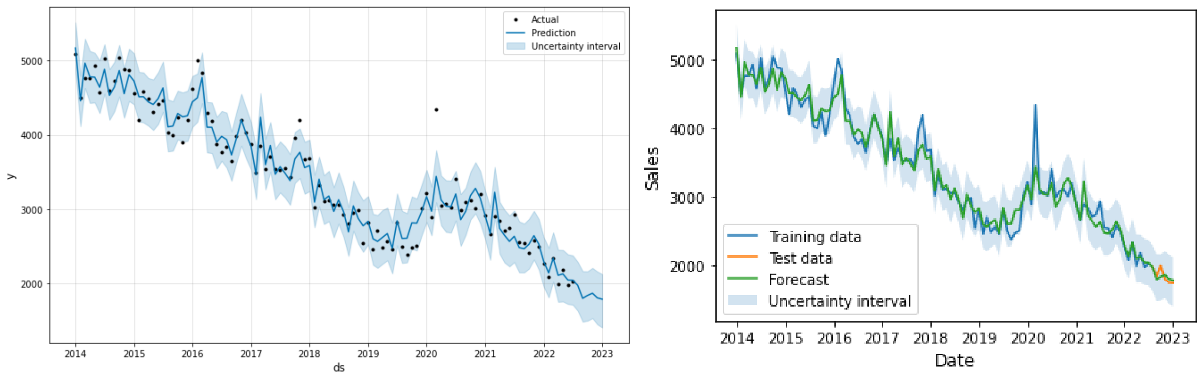

Figure 5. Prophet graphs- Product nº 9 (Source: Author)

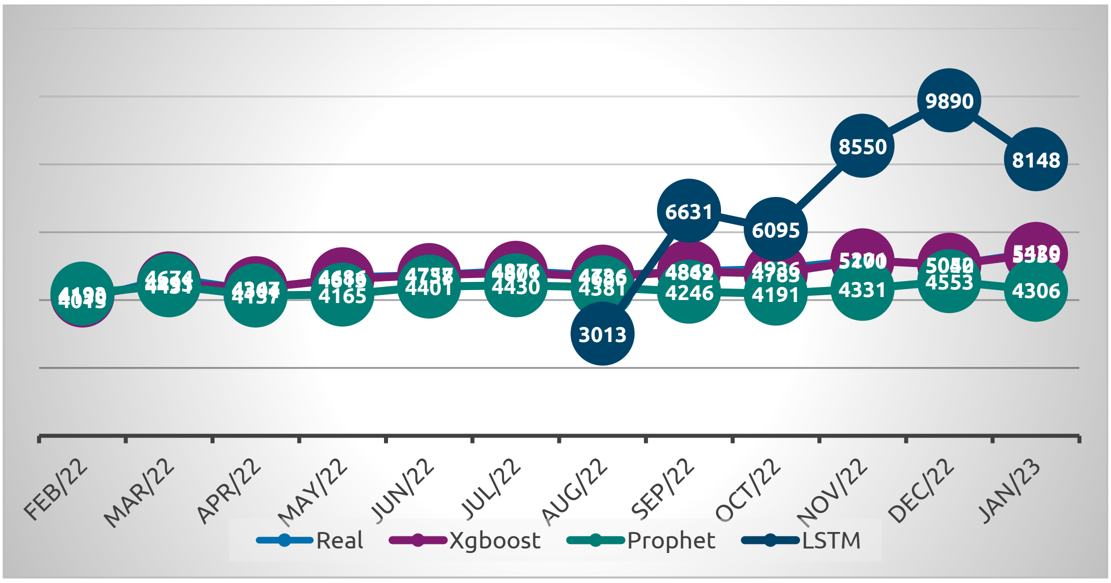

Figure 6. Product nº14 (Source: Author)

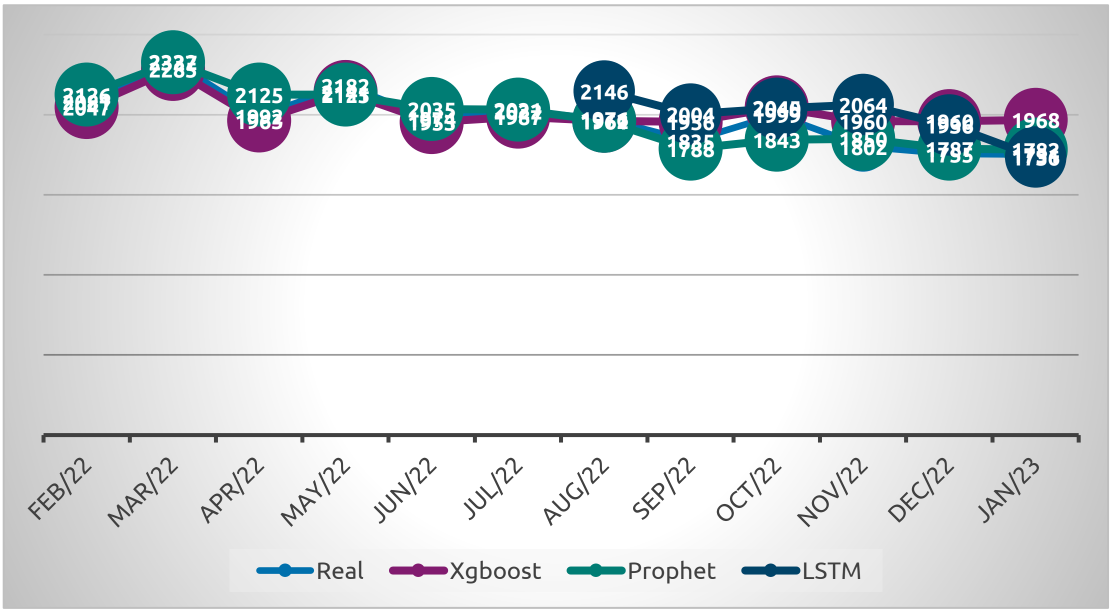

Figure 7. Product nº 9 (Source: Author)

A closer examination of the real units sold compared to the predictions for Product 14 and Product 9 is provided in Figures 8 and 9. These figures reveal the differences in the predictive capabilities of each model and the extent to which each model's predictions deviate from the actual sales values. This comparison emphasizes the necessity of assessing the performance of each model individually for every product, to gain a comprehensive understanding of their predictive capabilities in diverse contexts.

	Real	XGBoost	Prophet	LSTM
Aug-22	4736	4686	4381	3013
Sep-22	4869	4842	4246	6631
Oct-22	4936	4783	4191	6095
Nov-22	5201	5170	4331	8550
Dec-22	5052	5040	4553	9890
Jan-23	5420	5355	4306	8148

Figure 8. Product nº 14- Real units sold vs. Predictions (Source: Author)

	Real	XGBboost	Prophet	LSTM
Aug-22	1976	1961	1962	2146
Sep-22	1835	1956	1788	2004
Oct-22	1999	2045	1843	2040
Nov-22	1802	1960	1850	2064
Dec-22	1755	1960	1787	1936
Jan-23	1751	1968	1782	1736

Figure 9. Product nº 9- Real units sold vs. Predictions (Source: Author)

This chapter offers a comparative overview of the performance of the three forecasting models by thoroughly examining the results from the two selected products and the corresponding figures. The identified trends underscore the variations in the accuracy of the predictions generated by each model. The results are presented unbiasedly, with the analysis and discussion of the implications reserved for the subsequent chapter.

5. DISCUSSION

After displaying and reading the results from all three forecasting models developed, we can interpret them from a business perspective and extrapolate all the valuable insights. As we have seen from the three models, XGBoost has the lowest variation of MAPE across the selected drugs and generalized predictions quite accurately for most of them, proving to be a solid and robust algorithm for the dataset provided. This result is consistent with the literature (Martínez et al., 2020) and underlines the robustness of the XGBoost in forecasting sales, even in the pharmaceutical industry.

Despite the overall superiority of XGBoost, there were instances where Prophet outperformed it, offering an optimistic glimpse into what may be achieved with this reasonably new and unstudied model. These cases, although rare and maintaining the robustness of the results attained by the XGBoost, emphasize the importance of considering the unique characteristics and specificities of products in the pharmaceutical industry when selecting forecasting models. This aligns with existing literature highlighting the need for product-specific forecasting to ensure optimal algorithm selection based on underlying patterns and idiosyncrasies of the pharmaceuticals (Moroff et al., 2021; Rathipriya et al., 2022).

In line with the literature (Moroff et al., 2021; Rathipriya et al., 2022), the LSTM model proved to be incredibly demanding in computational effort compared to the other algorithms and did not perform as well as could be anticipated for encompassing a significant timeframe (2014-2023) with various sales fluctuations caused by the COVID-19 pandemic and geopolitical events. The LSTM model proved to be highly inefficient in our study context, revealing its inability to predict accurately and highlighting the complexity of correctly predicting pharmaceutical sales.

From the analysis performed on the results and considering the business needs and goals presented by the client, we conclude that the best approach is to build a hybrid forecasting model that exploits the advantages of both the XGboost and the Prophet algorithms to ensure the best possible predictions outcome. This decision is based on the importance of achieving the best accuracy possible when predicting demand but also based on some of the algorithm’s advantages when compared with the LSTM approach, such as the model’s complexity of implementation, its speed on training the data, and the ability to tailor the model along the way (Moroff et al., 2021).

From a marketing standpoint, the pharmaceutical company can utilize the accurate sales forecasts provided by the XGBoost and Prophet models to inform various marketing strategies. The company can optimize inventory management and distribution, ensuring that pharmacies maintain the required stock levels to meet customer demand (Zhu et al., 2021). This can lead to increased customer satisfaction and reduced stockouts, strengthening the company's reputation and competitiveness in the market.

Another potential marketing approach involves using the sales forecasting model to identify seasonal patterns and trends in product demand, enabling the company to develop targeted promotions and discounts during peak periods (Rosário et al., 2021). This strategy can help the company capitalize on periods of high demand, boosting sales and improving customer retention. Furthermore, the pharmaceutical company could consider constructing a churn prediction model to enhance its marketing efforts (Geiler et al., 2022). The company can create a more precise and actionable churn prediction model by integrating the existing sales forecasting model and incorporating new data, such as customer demographics, purchase behavior, and customer feedback.

Merging the sales forecasting model with churn prediction can assist the company in pinpointing potential high-risk accounts or pharmacies that may consider alternative suppliers. By proactively identifying these customers, the company can implement targeted retention strategies to prevent churn and maintain a robust customer base (Rosário et al., 2021).

In conclusion, developing a churn prediction model that leverages the sales forecasting model and additional customer information can give the pharmaceutical company valuable insights into customer behavior and preferences (Keny et al., 2021). By acting on these insights, the company can devise targeted marketing and retention strategies, ultimately improving customer satisfaction and business performance.

6. CONCLUSIONS AND FUTURE WORKS

The present study has undertaken a comprehensive evaluation of multiple algorithms, shedding light on the formidable potential of machine learning in forecasting demand, particularly within the pharmaceutical industry. To summarize our findings, we have compiled a table with an overview of the main strengths and weaknesses of each approach discussed in this study. Table seven offers therefore a consolidated set of valuable insights collected in the extent of the present work.

Model	Strengths	Weaknesses
XGBoost	• Low MAPE values for must drugs. • Robust and accurate predictions. • Handles missing values effectively. • Efficiently handles large datasets.	• Limited capturing of seasonality. • Difficulty in interpreting results in terms of underlying patterns. • Relatively longer training times
Prophet	• Ability to capture seasonality and trends. • Occasional outperformance. • Reasonably easy to compute. • Flexible and customizable model. • Handles outliers and missing data.	• Higher MAPE values in most cases. • Limited historical data for accurate long-term forecasting. • Limited support for complex feature engineering.
LSTM	• Ability to capture long-term dependencies. • Potential for accurate forecasting with extensive historical data. • Captures complex patterns and relationships.	• High computational requirements and expertise in implementation. • Lower accuracy compared to XGBoost and Prophet. • Difficulty in interpretability and model transparency.
Hybrid	• Utilizes strengths of both XGBoost and Prophet for enhanced predictions. • Improved accuracy and flexibility. • Tailored approach for better performance. • Provides ensemble benefits and model diversity.	• Potential complexity in combining and tuning the models. • Requires careful integration and parameter tuning. • Additional computational and implementation challenges.
Marketing Strategies	• Accurate sales forecasts to optimize inventory management and distribution. • Identifying seasonal patterns and trends for targeted promotions and discounts. • Churn prediction for customer retention.	• Integration challenges and data availability. • Requires additional resources for churn prediction integration. • Challenges in real-time implementation and adaptation.

Table 7. Main findings and conclusions (Source: Author)

By effectively bridging the theoretical underpinnings found within the literature and the practical implementation in the pharmaceutical industry, our investigation has yielded insightful conclusions and empirical evidence that corroborate previous findings, benefitting researchers, practitioners, and decision-makers in this field. Specifically, we have reinforced the robustness of XGBoost as a reliable forecasting tool for demand within the pharmaceutical industry. Furthermore, our study has shed light on the Prophet algorithm, a relatively recent addition to the forecasting landscape, revealing its potential for accurate predictions. Moreover, we have underscored the significance of tailoring forecasts to the unique characteristics and specificities of pharmaceutical products, encompassing variations in dosages and delivery methods. This emphasis on product-specific forecasting strategies aligns with existing literature advocating for the adaptation of different machine learning models accordingly with the product's singularities, thus enhancing forecast accuracy.

Through our practical implementation, we have facilitated the transition of the pharmaceutical company from reliance on traditional prediction tools, such as Excel, to leveraging advanced machine learning-based forecasts. Now the company can start making more informed marketing and general business decisions, improve inventory and distribution management, ultimately improving customer satisfaction, attraction, and retention.

While acknowledging the gathering of further expertise in LSTM architectures and specificities, we also recognize the vast potential for in-depth analysis of external factors that influence market dynamics within the industry. Factors such as the war in Ukraine, the ongoing COVID-19 pandemic, and regulatory changes warrant comprehensive exploration to enhance the predictive capabilities of forecasting models.

Building upon the results obtained and aligning them with the existing body of literature, our study propels the advancement of knowledge and practical implementation of forecasting models within the pharmaceutical domain. This knowledge expansion showcases Capgemini's expertise in the field of demand forecasting and predictive analytics, highlighting the company's ability to overcome complex business problems through advanced analytics. Likewise, it describes and follows the entire development of the project and demonstrates the company's ability to deliver tangible value and solutions to its clients. This advancement not only benefits industry stakeholders but also contributes to the broader body of knowledge in this area, by successfully integrating theoretical foundations with practical implementation in the pharmaceutical industry, providing valuable insights and empirical evidence that validate and reinforce previous findings.

BIBLIOGRAPHICAL REFERENCES

Ascher, J., Höglund, D., Ostojic, I., & Vancauwenberghe, M. (2018). *From product to customer experience: The new way to launch in pharma*.

Bonab, A. (2022). A Comparative Study of Demand Forecasting Based on ML. *Journal of Applied Research on Industrial Engineering*, *9*(3), 331–353.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). *Step-by-step data mining guide*. DaimlerChrysler.

Chaturvedi, S., Rajasekar, E., Natarajan, S., & McCullen, N. (2022). A comparative assessment of SARIMA, LSTM RNN and Fb Prophet models to forecast total and peak monthly energy demand for India. *Energy Policy*, *168*. https://doi.org/10.1016/j.enpol.2022.113097

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *13-17-August-2016*, 785–794. https://doi.org/10.1145/2939672.2939785

Ensafi, Y., Amin, S. H., Zhang, G., & Shah, B. (2022a). Time-series forecasting of seasonal items sales using machine learning – A comparative analysis. *International Journal of Information Management Data Insights*, *2*(1). https://doi.org/10.1016/j.jjimei.2022.100058

Ensafi, Y., Amin, S. H., Zhang, G., & Shah, B. (2022b). Time-series forecasting of seasonal items sales using machine learning – A comparative analysis. *International Journal of Information Management Data Insights*, *2*(1). https://doi.org/10.1016/j.jjimei.2022.100058

Geiler, L., Affeldt, S., & Nadif, M. (2022). An effective strategy for churn prediction and customer profiling. *Data and Knowledge Engineering*, *142*. https://doi.org/10.1016/j.datak.2022.102100

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Keny, S., Nair, S., Nandi, S., & Khachane, D. (2021). Sales Prediction for a Pharmaceutical Distribution Company. *International Journal of Engineering and Applied Physics (IJEAP)*, *1*(2), 186–191. https://ijeap.org/

Khan, M. A., Saqib, S., Alyas, T., Ur Rehman, A., Saeed, Y., Zeb, A., Zareei, M., & Mohamed, E. M. (2020). Effective Demand Forecasting Model Using Business Intelligence Empowered with Machine Learning. *IEEE Access*, *8*, 116013–116023. https://doi.org/10.1109/ACCESS.2020.3003790

Khanna, V., Ahuja, R., & Harvinder, P. (2020). ROLE OF ARTIFICIAL INTELLIGENCE IN PHARMACEUTICAL MARKETING_ A COMPREHENSIVE REVIEW. - EBSCO. *Journal of Advanced Scientific Research*, *11*(3), 54–61.

Martínez, A., Schmuck, C., Pereverzyev, S., Pirker, C., & Haltmeier, M. (2020). A machine learning framework for customer purchase prediction in the non-contractual setting. *European Journal of Operational Research*, *281*(3), 588–596. https://doi.org/10.1016/j.ejor.2018.04.034

Martinez-Plumed, F., Contreras-Ochando, L., Ferri, C., Hernandez-Orallo, J., Kull, M., Lachiche, N., Ramirez-Quintana, M. J., & Flach, P. (2021). CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories. *IEEE Transactions on Knowledge and Data Engineering*, *33*(8), 3048–3061. https://doi.org/10.1109/TKDE.2019.2962680

Merkuryeva, G., Valberga, A., & Smirnov, A. (2019). Demand forecasting in pharmaceutical supply chains: A case study. *Procedia Computer Science*, *149*, 3–10. https://doi.org/10.1016/j.procs.2019.01.100

Moroff, N. U., Kurt, E., & Kamphues, J. (2021). Machine Learning and Statistics: A Study for assessing innovative Demand Forecasting Models. *Procedia Computer Science*, *180*, 40–49. https://doi.org/10.1016/j.procs.2021.01.127

Noorunnahar, M., Chowdhury, A. H., & Mila, F. A. (2023). A tree based eXtreme Gradient Boosting (XGBoost) machine learning model to forecast the annual rice production in Bangladesh. *PLOS ONE*, *18*(3), e0283452. https://doi.org/10.1371/journal.pone.0283452

Panarese, A., Settanni, G., Vitti, V., & Galiano, A. (2022). Developing and Preliminary Testing of a Machine Learning-Based Platform for Sales Forecasting Using a Gradient Boosting Approach. *Applied Sciences (Switzerland)*, *12*(21). https://doi.org/10.3390/app122111054

Rathipriya, R., Abdul Rahman, A. A., Dhamodharavadhani, S., Meero, A., & Yoganandan, G. (2022). Demand forecasting model for time-series pharmaceutical data using shallow and deep neural network model. *Neural Computing and Applications*. https://doi.org/10.1007/s00521-022-07889-9

Rosário, A., Moniz, L. B., & Cruz, R. (2021). Data science applied to marketing: A literature review. In *Journal of Information Science and Engineering* (Vol. 37, Issue 5, pp. 1067–1081). Institute of Information Science. https://doi.org/10.6688/JISE.202109_37(5).0006

Schreckenberg, F., & Moroff, N. U. (2020). Developing a maturity-based workflow for the implementation of ML-applications using the example of a demand forecast. *Procedia Manufacturing*, *54*, 31–38. https://doi.org/10.1016/j.promfg.2021.07.006

Siddiqui, R., Azmat, M., Ahmed, S., & Kummer, S. (2022). A hybrid demand forecasting model for greater forecasting accuracy: the case of the pharmaceutical industry. *Supply Chain Forum*, *23*(2), 124–134. https://doi.org/10.1080/16258312.2021.1967081

Taylor, S. J., & Letham, B. (2018). Forecasting at Scale. *American Statistician*, *72*(1), 37–45. https://doi.org/10.1080/00031305.2017.1380080

Zhu, X., Ninh, A., Zhao, H., & Liu, Z. (2021). Demand Forecasting with Supply-Chain Information and Machine Learning: Evidence in the Pharmaceutical Industry. *Production and Operations Management*, *30*(9), 3231–3252. https://doi.org/10.1111/poms.13426