Demand Forecasting with Machine Learning

by

Chi Sheng Tseng
BSc, Resources Engineering, National Cheng Kung University

and

Turgay Turkmen
MSc, Management, Vrije Universiteit Brussel

SUBMITTED TO THE PROGRAM IN SUPPLY CHAIN MANAGEMENT
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE IN SUPPLY CHAIN MANAGEMENT
AT THE MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2024

Signature of Author: ______________________________

Chi Sheng Tseng
Department of Supply Chain Management
June 20, 2024

Signature of Author: Turgay Turkmen

Turgay Turkmen
Department of Supply Chain Management
June 20, 2024

Certified by: ______________________________

Dr. Juan Carlos Piña Pardo
Postdoctoral Associate, MIT Megacity Logistics Lab
Capstone Advisor

Accepted by: ______________________________

Prof. Yossi Sheffi
Director, Center for Transportation and Logistics
Elisha Gray II Professor of Engineering Systems
Professor, Civil and Environmental Engineering

Demand Forecasting with Machine Learning

by
Chi Sheng Tseng
and
Turgay Turkmen

Submitted to the Program in Supply Chain Management
on June 20, 2024, in Partial Fulfillment of the
Requirements for the Degree of Master of Applied Science in Supply Chain Management

ABSTRACT

Our capstone project focuses on forecasting sales of the sponsor company's Heat-not-Burn (HNB) consumables in Italy. We have monthly data between 2015 and 2023. Our research objective is forecasting future sales with machine learning (ML) models. As a methodology, our approach is first to understand the business problem and available data. Then, we forecast using traditional methods as a baseline. After that, we apply different machine learning models. Finally, we compare the models' accuracy to understand the value of ML models over traditional forecasting methods. Our key findings are that Prophet is the best forecasting model, beating traditional forecasting methods (such as the Holt-Winters method and Auto-Regressive Integrated Moving Average) and other ML models tested. This is because Prophet excels in capturing the complex patterns and seasonality in the historical sales data. We also apply hyperparameter tuning to Prophet to identify the optimal parameter setting to predict HNB consumables sales.

Capstone Advisor: Dr. Juan Carlos Piña Pardo
Title: Postdoctoral Associate

ACKNOWLEDGMENTS

The journey of capstone completion is a collaborative one. Many individuals, to whom we owe a great deal of gratitude, were involved in our achievement of success.

First and foremost, we express our profound appreciation to Dr. Juan Carlos Piña Pardo. Without his invaluable advice, we could not be where we are today. Dr. Juan Carlos' significant attention to details and academic vision shaped our project. In addition, his patience with us during the time spent and solving the difficulties we have faced both at the technical and personal levels were priceless. We express our deep thanks to Dr. Juan Carlos' contribution to our personal and professional growth.

Our gratitude also extends to the entire SCM management team. Their decision to admit us to the study program laid the foundation for the research presented in this capstone. Their support and belief in the value of our work have been essential to our journey. Classes we have taken such as SCM.256 where we had the chance to learn best practices of Data Science and Machine Learning were a game changer for us in completing our project. We thank Dr. Elenna Dugundji for her teachings in the Spring semester.

We hereby thank our sponsor company for giving us the opportunity to work on a very interesting problem. Our gratitude to the sponsor company management for always being present in our meetings, giving timely responses, and explaining in detailed their business processes. Without their continuous input and efforts, we would not be able to develop a satisfactory capstone.

TABLE OF CONTENTS

1. Introduction

Successful global organizations are those that accurately estimate market needs. To do so, these companies rely on demand forecasting. Indeed, demand forecasting is considered the “backbone” of organizations seeking to ensure a profitable and sustainable business model (Petropoulos et al., 2022). However, it is difficult for companies to achieve a certain level of forecasting accuracy. Low accuracy levels in demand forecasting would result in two main problems. The first is *surplus events* (i.e., supply exceeds demand), which cause problems such as high inventory holding costs, limited free cash flow, and potential risk of product obsolescence. The second is *stockout events* (i.e., supply cannot meet customer demand), which lead to loss of revenue and margins, loss of customers to competition, and potential damage to brand images.

Demand forecasting challenges also differ among industries. Mature or low-tech products generally have stable demand. Therefore, demand forecasts for such products tend to have acceptable accuracy (Småros, 2002). However, the challenge becomes much greater for new or high-tech products, especially new products that are disruptive to the existing markets where it is difficult to predict consumer demand (Chern et al., 2010). An example is the tobacco industry, which has faced similar disruptions in recent years (Edwards et al., 2022). Recently, there has been a trend of industry leaders aiming to offer less harmful alternatives to cigarettes for their customers (Smith et al., 2023). Tobacco companies have invested significant resources in research and development, engineering, and manufacturing to develop these smoke-free alternatives, including vapors, nicotine pouches, and heat-not-burn (HNB) products (Foundation for a Smoke-Free World, 2022), usually composed by devices and consumables. Furthermore, their sales and marketing teams commercialize these alternatives towards adult smokers to encourage those who would not quit smoking to convert to smoke-free products. However, their supply chain teams struggle to minimize stockout or overstocking of these items, as predicting demand for new products is difficult.

The market of smoke-free alternatives is growing rapidly as the product portfolio of HNB products expands globally into new markets. However, since customer behavior is unknown during the first years of market introduction, it is a great challenge to forecast short and medium-term (0–18 months) demand in non-mature markets. Therefore, our sponsor company, a leading international tobacco company, has an enormous interest to understand relevant drivers that impact the forecast accuracy of HNB consumables and how these drivers can be explained.

1.1. Problem Statement and Research Questions

The sponsor company currently uses univariate statistical forecasting software that auto-selects best-fitting statistical models to predict customer demand. While the tool works well for the demand forecasts of traditional tobacco products, it generates unsatisfactory results for HNB consumables in non-mature markets and lacks explainability. This results in the need to improve the current forecasting accuracy levels of HNB consumables to mitigate operational costs and increase explainability. In this context, our capstone project aims to focus on the following research topics:

1. Benchmark forecast accuracy and performance errors of the current forecasting model from the sponsor company.
2. Compare traditional forecasting approaches and machine learning (ML) approaches and select the model that best improves current forecast accuracy with the causes of improved accuracy.
3. Include external information if needed to yield better forecast accuracy.

1.2. Hypothesis

We have a demand forecasting accuracy problem for a breakthrough product. We believe that an ML model that considers not only historical sales data, but also considers external factors can help better predict customer demand for HNB consumables. External factors could include getting consumer insights from social media, web analytics platforms, third-party market research reports, and our sponsor company's internal reports. In addition to consumer insights, we intend to consider the impact of other external factors, including (but not limited to) Gross Domestic Product (GDP), average wages, allocated price levels, regulations of selected countries, taxation, and local market-specific elements.

1.3. Project Goals and Expected Outcomes

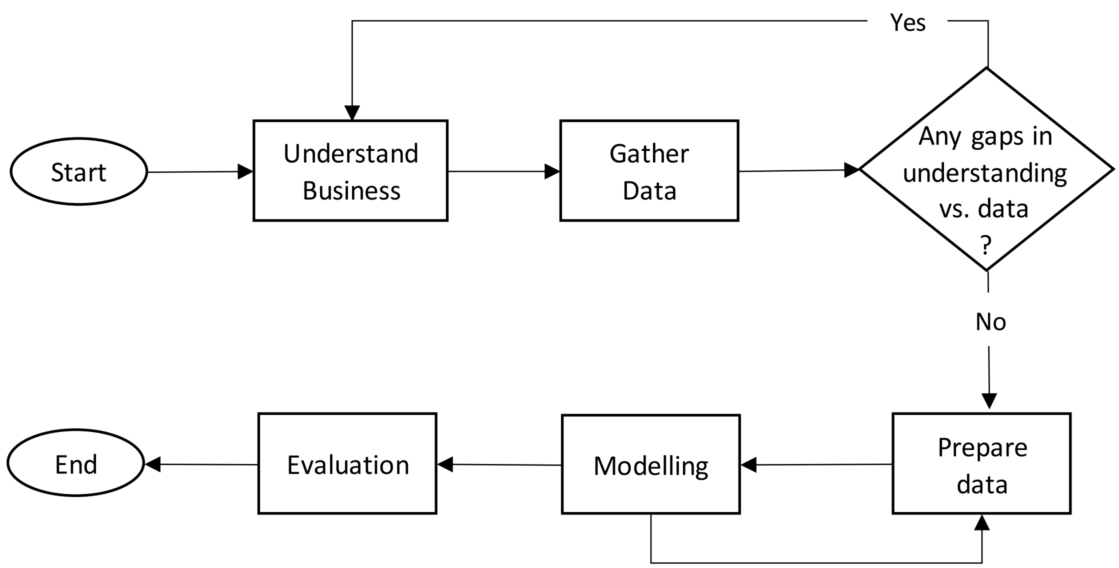
We aim to develop ML-based forecasting models to support our sponsor company in predicting customer demand in non-mature markets for the HNB consumables category. The minimum expectation is to improve the accuracy levels of the current tool used by our sponsor company. Utilizing modern ML methods and external factors, we aim to reach acceptable accuracy levels for our sponsor company. The project's deliverable is a handover of a working ML model that considers external factors. In addition, we intend to identify the main demand drivers for the assigned markets and explain our model's explainability. Lastly, we provide the codes and explanations of these drivers to the sponsor company for them to amend/update the model for their future needs.

1.4. Work Plan

Our work plan is based on the "Cross Industry Standard Process for Data Mining" (CRISP-DM) methodology, a *de facto* standard in data mining (Schröer et al., 2021). This methodology provides a structured approach to tracking the project's progress and helps us ensure the process's quality. Figure 1 shows a high-level flow chart of the capstone project steps.

Figure 1

Plan of Work based on CRISP-DM Methodology



2. State of the Practice

This chapter surveys various demand forecasting methods prevailing in various industries. We start by giving an overview of the forecasting challenges organizations in the industry encounter. Then, we examine and compare two types of forecasting methods. The first type, *traditional methods*, includes judgmental, experimental, causal, and time series approaches. The second type, *advanced methods*, uses ML algorithms together with consumer data. Finally, we explain why incorporating ML algorithms with consumer data can help our sponsor company further improve demand forecasting accuracy.

2.1. Forecasting Challenges

Demand forecasting is crucial in driving broad aspects of supply chain implementations (Boone et al., 2019). Applications of demand forecasting range from long-term capacity investment to short-term production planning (Nowadly & Jung, 2020). Poor demand forecasts can lead to both stock-out and overstock events, affecting organizations' ability to realize maximum revenue and maintain financial health (Trapero et al., 2023; Steenbergen & Mes, 2020). For example, a leading global pharmaceutical company suffered severe stock-outs in its e-commerce business due to unprecedented demand growth from the COVID-19 outbreak in 2020. These demand spikes disrupted the company's supply chain and caused a huge loss of sales revenue (Izaguirre & Chao, 2020). Conversely, a company in the consumer-packaged goods industry experienced an overstock event due to the global economic recession. Consequently, the company was motivated to improve its demand forecasts to overcome the financial difficulty during the recession (Uriarte, 2010). Therefore, to remain competitive in changing economic and business conditions, companies must understand current state-of-the-art forecasting methods and how to use them to improve demand forecasts.

2.2. Forecasting Methods

Traditional forecasting methods include subjective and objective methods. (Petropoulos et al., 2022). Subjective methods are used when historical data is unavailable (e.g., a new product launch). They can be classified as being judgmental (often relying on expert knowledge) or experimental (such as sampling customers via surveys to make predictions), whereas objective methods are used to build time series and causal approaches when historical data is available (Nowadly & Jung, 2020). Time series approaches are objective methods that essentially match the patterns observed in the data over time with models such as moving averages or exponential smoothing (Caplice & Ponce, 2023). In traditional forecasting, the Holt method is well known to give accurate predictions when a dataset has a trend. When there is both trend and seasonality, the Holt-Winter method gives the best outcome with lower error metrics (Caplice & Ponce, 2023). Multiple industries have adopted time series approaches for their simplicity and typically satisfactory results (Eiskowitz, 2021). Causal approaches, such as linear regression, are used when there is an underlying relationship among variables (Seyedan & Mafakheri, 2020).

Although traditional forecasting methods often work well for products with a long demand history and low technical requirements, they might not be as appropriate for new products due to the lack of actual demand data (Goodwin et al., 2014; Lynn et al., 1999). However, introducing new technological products is essential for companies to remain competitive and improve their revenue and profit levels (Yan & Hu, 2023). Therefore, ensuring demand forecasting accuracy for new product types is key to a company's success.

Forecasting techniques for new products vary between new and current markets (see Table 1). These include Delphi sessions, which consider expert opinions for future prediction, and Bass diffusion models, which predict demand based on different phases of a product's life cycle for new markets. For current markets, principally, techniques such as "looks-like" (e.g., similar product that was launched before) or analogous models are widely adopted by searching historical launches of similar products to generate sales records for prediction (Kahn, 2002).

Table 1

Forecasting for New Product

Market	**Forecasting Techniques**
New market	Delphi Sessions, Bass Diffusion Models
Current Market	Looks-like or Analogous Analysis

Note: This table is adapted from Caplice and Ponce (2023).

2.3. Machine Learning Applications in Forecasting

Demand forecasting has been one of the successful applications of ML-based models (see, e.g., Smirnov & Sudakov, 2020; Hamoudia & Vanston, 2023; Amar et al., 2022). The ability of ML models to outperform traditional forecasting methods has been widely demonstrated by several academic studies (see, e.g., Mia et al., 2021; Pavlyshenko, 2019) and consulting reports (Amar et al., 2022). Also, large companies such as Amazon (AWS, 2021) and Walmart (Silverstein, 2020) have already established ML models for their demand forecasts.

Some use cases where ML solves pressing industry problems include reducing obsolescence due to poor forecasts (Jennings et al., 2016) and forecasting new product demand (Smirnov & Sudakov, 2020; Hamoudia & Vanston, 2023). Although new product forecasting difficulties are recognized in academic studies—for instance, Kahn (2002) found that only 58% forecast accuracy has been observed in new products—ML applications generally yield better outcomes, because ML models are not limited by the constraints of traditional forecasting approaches.

In addition, ML forecasting plays an important role in a company's financial success. For example, NXP Semiconductors developed an ML model to estimate sales even before the new product release and achieved a financial advantage (Boutane et. Al., 2023). Therefore, companies in the industry are able to attract more investment by applying ML approaches to ensure a good forecast (Mattei & Platikanova, 2017).

Our capstone aims to have high forecast accuracy and good explainability. Therefore, during our research, we also investigate advanced ML applications. A good example is the work of Pavlyshenko (2019). The author focuses on retail store sales data and applies multiple methods such as Extra Tree, Random Forest, Lasso, and Neural Network. In addition to the listed ML models, Pavlyshenko (2019) applied the stacking technique and achieved a high-accuracy forecast. Another example is the work of Mia et al., (2021), who show that a Back Propagation Neural Network (BPNN) model can achieve high-accuracy forecasts in the retail industry.

The use of external data is an important element of our capstone. Paruthipattu and Haycock (2021) suggest using external data for better forecasting accuracy. Iftikhar and Khan (2020) highlight the benefit of ML demand forecasting with external factors in the apparel industry. The authors reviewed consumer insights from Facebook and Twitter data as external factors for forecasting Nike sales. In their research, they see a correlation between social media posts and sales of the products. Using social media input, the ML model generated an improved forecast accuracy. Similarly, our research focuses on consumer insights such as social media data as an external factor to drive better forecast accuracy.

In summary, we have seen evidence that ML forecasting improves accuracy, resulting in better financial positions for companies. We understood that advanced models could offer even better accuracy; however, we observed problems with their explainability. For example, while the ML models developed by Pavlyshenko (2019) and Mia et al. (2021) achieved high-accuracy forecasts, they lack explainability, which is impractical for our purposes. Lastly, we reviewed the importance of external data. In the next section, we focus more on explainability and compare different ML models with each other.

2.4. Explainable Machine Learning Forecasting Models

Our sponsor company is mainly interested in understanding how their sales data and external factors influence prediction results. The explainability of an ML model means that the outputs of the model can be explained by the inputs. Some ML models lack explainability and remain a black box to their users. Explainability is critical for many company executives to trust model outputs (Misheva et al., 2021; Nimmy et al., 2022). A McKinsey study suggests that "people use what they understand and trust" (Grennan et al., 2022). Based on our literature review, we identified ML methods such as XGBoost, LightGBM, and Random Forest as having good explainability features. A comparison of popular ML models and their specific characteristics is listed in Table 2.

Table 2

Models Comparison Matrix

Synthesized from the works of Géron (2019), Susmita (2019), Singh et al. (2016), Angelov et al. (2021), Jha & Pande (2021), *and Hamoudia et al. (2023)*

ML Model	**Definition**	**Pros**	**Cons**	**Explainability**
Linear Regression	Supervised learning, fits a straight line to dataset	Performs well if relationship between variables is linear, easy to understand	Oversimplifies real world problems, limited use-case	High
XG Boost	A gradient boosting algorithm that uses decision trees as base learners.	Accurate and efficient, can handle large datasets and non-linear relationship	Requires a lot of data to train	High
Light GBM	Similar to XG Boost but faster and more memory efficient	Fast, can handle large datasets	Fewer features	High
Random Forest	An ensemble learning algorithm averages the predictions of decision trees.	Easy to interpret. Well handles large datasets. Robust to outliers	Slow on large datasets, less accurate vs. other ML models	High
Prophet	An open-source forecasting tool introduced by Meta	Capture temporal components of trend and seasonality well, easy to interpret	Performance could be poor when dealing with external factors	High
Support Vector Machines	Supervised learning, hyperplane separates data into two classes	Good performance on complex problems and can handle semi-structured and structured data	Performance is poor for large datasets, sensitive to outliers	Medium
Neural Networks	Inspired by human brain. Applicable for problems with non-linear relationships	Performs well with complex problems	Difficult to interpret and expensive to train large data	Low
K Nearest Neighbors	Supervised learning, that classifies data based on their similarities	Easy to understand and implement. Flexible method and cheap	Noisy features decrease accuracy	Low

2.5. Our Findings

During our literature research, we identified several ML models that can be used in our capstone and listed them in Table 2. Our findings show that the listed ML models are already being used in many industries for forecasting. Considering our focus in this capstone is accuracy and explainability, we tested some of the models from Table 2 that best fit the goal of the capstone.

3. Methodology

Our capstone aims to identify and apply the most relevant ML model to improve the sponsor company's current forecasting model. Our methodology, based on Géron (2019), is summarized in Figure 2.

First, we look at the big picture and define the scope and business objectives of the problem. We understand from the sponsor company the key variables, underlying assumptions, and constraints of the received data and current model. Second, based on our understanding, we transform the received data to fit the purpose of the capstone project by applying techniques such as expanding data and creating new features. We then plot the transformed data and confirm our findings with our sponsor company. Third, we prepare the data and see if further data processing is needed, such as amending missing values and feature scaling. Now, the data is prepared for modeling. We split the data into training and test sets and fit the trained data into multiple models we identified based on the literature review and test and evaluate the results (see Table 2 in Section 2.4). Finally, we select the best-performing model, optimize the results by adjusting parameters, and present our findings to the sponsor company.

Figure 2

Capstone Methodology Steps – Synthetized from Géron (2019)

Look at the big picture	Transform data	Prepare data	Train and evaluate models	Optimize selected model and present results
• Data/ Business understanding • Error metrics selection • Assumption clarification	• Data expansion • Feature extracting • Data visualization	• Data processing • Missing values amending • Feature scaling	• Training and test sets splitting • Results comparison • Model evaluation	• Model selection • Parameters fine-tuning • Solution presentation

3.1. The Big Picture

Our process starts with understanding the sponsor company's business needs. After a series of discussions with the sponsor company, the scope of our project is to develop an ML model with a time series forecast to provide an explainable model.

Then, we discuss how to measure the performance of our model. Currently, the sponsor company does not have a fixed measurement methodology and uses a percentage bias method. Percentage bias is the relative difference between the predicted and actual values shown as percentages. Percentage bias is calculated by dividing predicted values by actual values. In our capstone, we measure the accuracy of our model with commonly used methods to cross-check the results between each error metric. Selected error measurement metrics include Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Squared Error (RMSE). We clarified our initial assumption that beating existing accuracy was not critical for the sponsor company. We aim to build a model with explainability and satisfactory accuracy.

3.2. Transform Data

A large dataset is favored when conducting a time series forecast as it provides more granularity for the models to pick up important signals from the time series. Given the small data set in our received data, we transform the original data by expanding and converting the monthly sales records to daily. In addition, we extract time features from time series to understand the trend and seasonality of the HNB consumables. Finally, we plot the transformed data and confirm our observations with the sponsor company.

3.3. Data Preparation

To prepare the transformed data for modeling, we review the data and see if further cleaning is required. First, we clean up the data by checking for any missing value that needs to be amended. Second, we look at the features and see if scaling needs to be applied. Scaling features is a common technique when preparing the data for modeling. This can ensure comparable features and prevent features with a large range from skewing the model prediction. Finally, the transformed data is prepared for the next step.

3.4. Train and Evaluate Model

Now, we split the prepared data into training and test sets and start training our selected models considering Table 2. We then compare the results among traditional and ML models based on the error metrics specified in section 3.1 and evaluate the models' performance using the test set. Ultimately, we can select a model that yields the best performance, whether a traditional or an ML model.

3.5. Test Model and Present Results

At this stage, we select the best model for our capstone project after comparing the performance of all the models we consider. To optimize the model's performance, we apply hyperparameter tuning and identify the key parameters that maximize performance. Finally, we interpret the model based on the selected parameters and present our findings to the sponsor company.

4. Data and Model Preparation

Our capstone project aims to study the relevant drivers that impact the forecast accuracy of HNB consumables. This section is organized as follows: Section 4.1 focuses on data understanding and limitations. Section 4.2 focuses on data expansion. Section 4.3 explains feature engineering and visualization of HNB consumables' sales with new features. Section 4.4 addresses how the best model is selected. Lastly, Section 4.5 discusses how the best-performing model can be further improved using data splitting and hyperparameter tuning approaches.

4.1. Data Understanding and Limitations

We have monthly sales data for HNB consumables with time series from January 2015 to September 2023 (105 rows). The data has 117 fields, including macroeconomic factors (e.g., unemployment rate, consumer price index, and disposable income), sponsor company-specific information, public competitors' information (e.g., pricing, launching period), and external factors (e.g., number of times being mentioned on social media platforms). Excluding the date (which is in time format), we have 44 categorical fields (mostly presented in binary format) and 72 numerical fields (either integer or float datatypes).

Time series information is critical in time series forecasts. However, limited by only 105 sales records in the received data, ML models' ability to make good predictions is hindered (Antonio et al., 2019). To solve this, we need to expand the monthly sales of HNB consumables and only consider time features extracted from time series. Therefore, the focus of the capstone becomes comparing the results of time series forecasts using traditional methods and ML methods.

4.2. Data Expansion

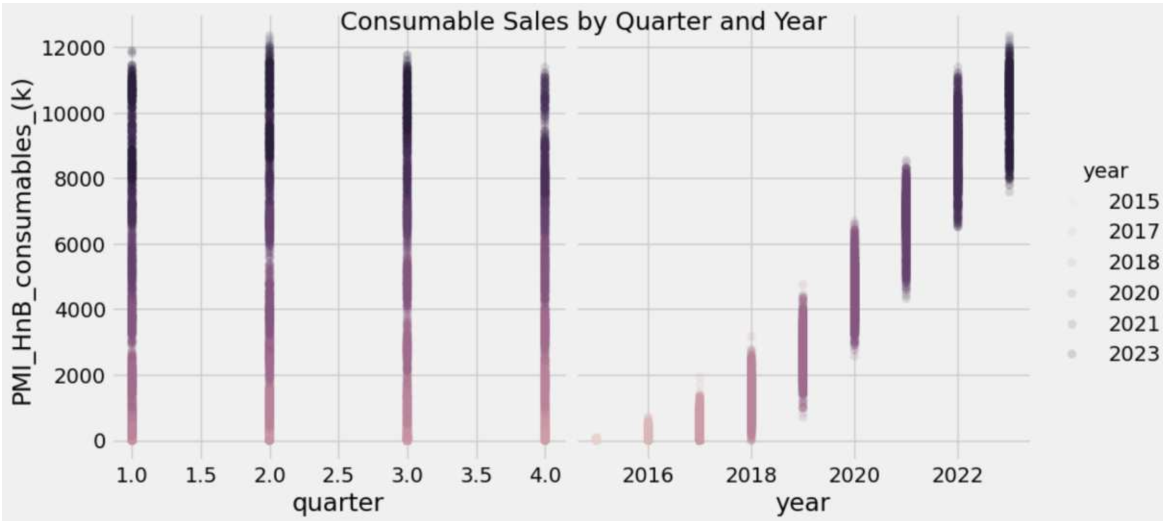
To address the limitation coming from the small dataset, we keep the important features in the original data: time series, consumable sales, and then we expand each monthly sales data by certain folds to convert original data into daily sales records (Yan et al., 2019). Our approach is based on the following assumptions and processes. First, we set the size of the expansion to 100 folds. This way, we will get multiple sales records each day coming from the original monthly sales data. Second, we assume all the newly generated 100 daily sales records follow a Normal distribution with a mean and standard deviation defined by the monthly ones divided by 100 and the square root of 100, respectively. Then, we make sure the sum of all 100 new daily sales is equal to the corresponding monthly sales. This process ends with a new expanded dataset with 10500 rows of daily sales records.

4.3. Feature Engineering and Data Visualization

Since we now focus on time series forecasts, extracting useful temporal features from the time series information in the expanded data is important. This can help us understand which selected models can pick up these features well in the training process and make a representative predictive model. These time features include information such as month, quarter, and year of the sales. To observe if there are any trends and seasonality underlying the data, we then plot the sales of HNB consumables against the extracted time features to get a visualization of the data. Figure 3 suggests an upward trend of fast-growing HNB consumable sales by year and quarterly seasonality, with the second quarter (summer) of the year being the highest and the fourth quarter (winter) being the lowest. We confirm the observation in Figure 3 with our sponsor company and get positive feedback. As a result, the temporal characteristics observed in the plots are factored in when selecting suitable models to test for our time series forecast. Now, we have our new data prepared with both the expanded data and the time features extracted from the time series for modeling.

Figure 3

Trend and Seasonality Visualization – HNB Consumables



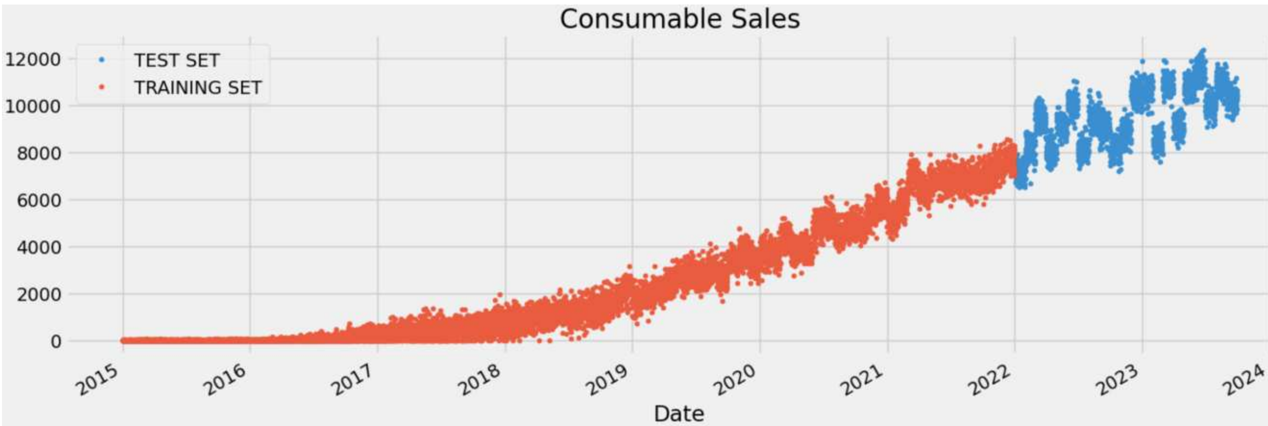
4.4. Model Selection and Error Metric Conversion

To shortlist the models that are desired for the time series forecast, we refer to Table 2 for ML models. The selected ML models are Prophet and XGBoost. We chose Prophet because of its outstanding ability to capture trends and seasonality in time series forecasts. On the other hand, as mentioned in Section 1.1, the current linear model the sponsor company currently uses does not produce satisfactory results. Therefore, we pick XGBoost as it serves the purpose well when handling non-linear patterns in the data. For the traditional forecasting models, the most commonly used is Moving Averages (MA). The application is simple to understand and is primarily an effective method. We apply a five-month MA to the sales data. In addition, we also use the Holt-Winter method (as it deals well with trend and seasonality; see Section 2.2) and ARIMA (as it performs well with large data and is also widely adopted).

At this stage, we have listed the models for either traditional or ML methods. Now, we need to test these models and compare the results based on all the error metrics mentioned in Section 3.1. First, we split the new data into 80% training set and 20% testing set. Figure 4 illustrates how the data is split between training and testing sets for HNB consumables. All the data before January 2022 represents the training set, and all the data after and including January 2022 accounts is the test set. We use the training set to fit the selected models and make the sales forecast out of the testing set with trained models. Since we now have the predictions of each model we test, daily MAE, MAPE, and RMSE can be computed. Ultimately, we produce a monthly sales forecast for HNB consumables. Therefore, before comparing models' results, a conversion between daily error metrics to monthly must be done. To do so, we group all the daily sales by month and sum them up for both actual values in the testing set and predicted values coming from the prediction. Finally, we can compare the results and select the best-performing model based on the monthly error metrics.

Figure 4

Split Between Training and Test Sets - Consumables



4.5. Data Splitting and Hyperparameter Tuning

To further improve the results according to the best-performing model, we create three types of splitting: 80%, 85%, and 90%. Since the prediction varies by different types of splits, so does the model's performance. With this approach, we can tell if the model performs better with a larger or smaller training set and identify the ideal split percentage that gives rise to the best model performance.

Another approach to optimize the model's results is applying hyperparameter tuning. With the help of parameter grid search, we can optimize the performance of the selected model by fine-tuning the hyperparameters underneath the architectures of the model (Abhishek et al., 2023). This approach allows the model to tune and select the optimal combination of the hyperparameters that yield the best prediction and, therefore, achieve the lowest error metrics. Taking Prophet as an example, below are the descriptions of the four most common parameters used to fine-tune a Prophet model (Lorenzo et al., 2021):

1. changepoint_prior_scale controls the flexibility of the trend. A higher value allows for more flexibility in fitting the trend to the data, potentially capturing more short-term fluctuations, while a lower value results in a more rigid trend.

2. changepoint_range controls the proportion of the data at which potential changepoints are placed. By default, it uses 0.8, meaning it will place changepoints in the first 80% of the data.
3. seasonality_prior_scale controls the strength of the seasonality model using Fourier series components. A higher value allows for more flexible seasonality patterns, while a lower value results in smoother seasonality.
4. seasonality_mode supports additive and multiplicative seasonality. By default, it uses additive seasonality, but you can switch to multiplicative if your data exhibits multiplicative seasonality patterns.

5. Results

In this section, we present the results according to the procedures detailed in Section 4. First, we compare the results of traditional and ML models specified in Section 4.4 based on monthly MAE, MAPE, and RMSE. Second, we select the best-performing model and test if the performance can be further improved based on different training and test split percentages. Lastly, we show the best results by applying hyperparameter tuning and provide explainability of the model.

5.1. Traditional and ML Models Comparison

The models we compare in this section can be divided into traditional time series and ML models. As mentioned in Section 4.4, we select moving average (MA), Holt-Winter, and ARIMA for traditional models, while for ML models, we select Prophet and XGBoost. The scores of error metrics for these models are shown in Table 3.

Table 3 shows that the best-performing model for consumables is Prophet. Its error metrics are significantly lower than that of the remaining models, whether traditional or ML models. Particularly, a MAPE lower than 8% suggests the Prophet's ability to pick up signals from time series and provide good predictions in the sales of HNB consumables. On the other hand, traditional models overall have good performance on HNB consumables with MAPE between 10% and 20%. Comparing all traditional models, we see that the Holt-Winter model produces the best results while the 10-month MA model has the least favorable results. XGBoost performs poorly on HNB consumables, as mentioned previously in Table 2, XGBoost performs well when the underlying pattern of the data is non-linear.

Table 3

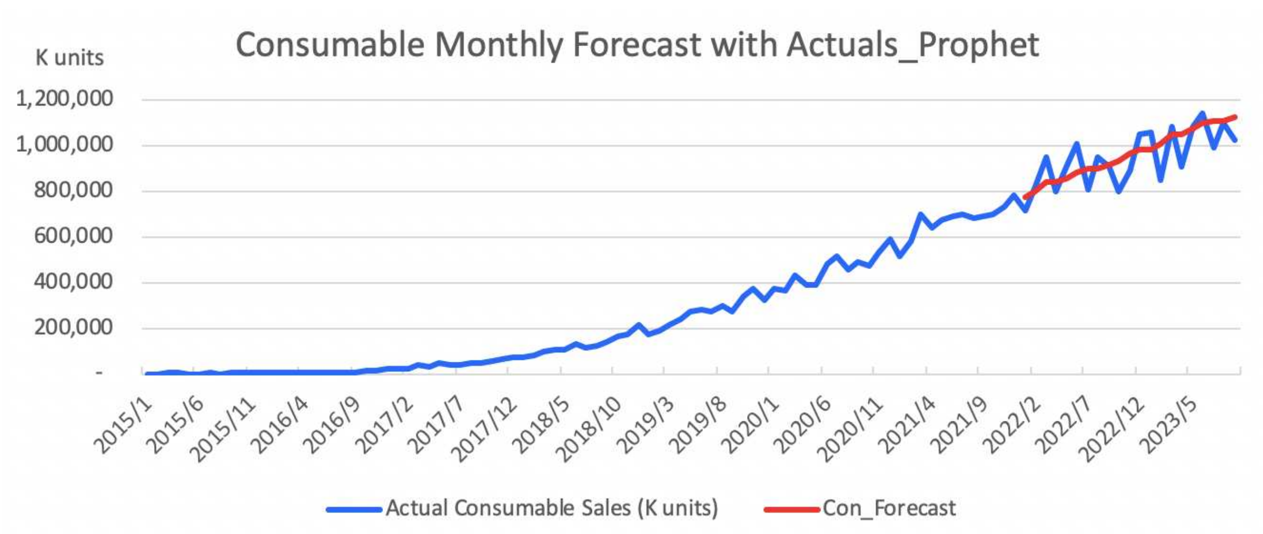
Error Metric Comparison Between Traditional and ML Models

Time Features	Consumables		
Models	**MAE**	**MAPE**	**RMSE**
Prophet	71,835.41	7.85%	84,541.67
Holt-Winter	104,513.13	10.44%	126,684.21
ARIMA	166,723.67	16.54%	198,406.67
MA (5 months)	172,024.91	17.08%	203,670.52
MA (10 months)	174,732.42	17.35%	206,334.02
XGBoost	227,117.69	23.08%	250,934.13

After comparing the results, we select Prophet as the best-performing model. To further improve the prediction of Prophet, we plot the predictions against the actual sales records for HNB consumables. In Figure 5, we observe that Prophet performs really well in predicting the trend in the test set, but fails to capture demand fluctuations. Therefore, in Sections 5.2 and 5.3, we introduce two approaches to further improve the performance of Prophet. The first is adjusting the percentage between training and testing sets. The second is to apply hyperparameter tuning.

Figure 5

Monthly Forecast with Actual Values – HNB Consumables



5.2. Adjusting Training and Test Split

In this section, we adjust the proportion of training and test sets to see if the Prophet performance from Section 5.1 can be further enhanced. As shown in Figure 5, the model can forecast the trend decently but fails to capture the sales fluctuation in the testing set. To improve the model's forecasting ability in capturing fluctuated demand, we try splitting the data into 80%, 85%, and 90% of training data. The results are shown in Table 4.

The results from Table 4 suggest that training more data does not yield better results for consumables, although there is a slight improvement when we increase the split percentage to 90%. Therefore, in the following, we apply hyperparameter to further improve our results.

Table 4

Prophet Performance Based on Different Split %

Prophet	**HNB Consumables**		
Result Optimization	**MAE**	**MAPE**	**RMSE**
80% Training	71,835.41	7.85%	84,541.67
85% Training	93,467.77	10.17%	114,738.37
90% Training	70,105.95	7.28%	84,760.72
Hyperparameter Tuning	66,420.05	7.05%	75,781.32

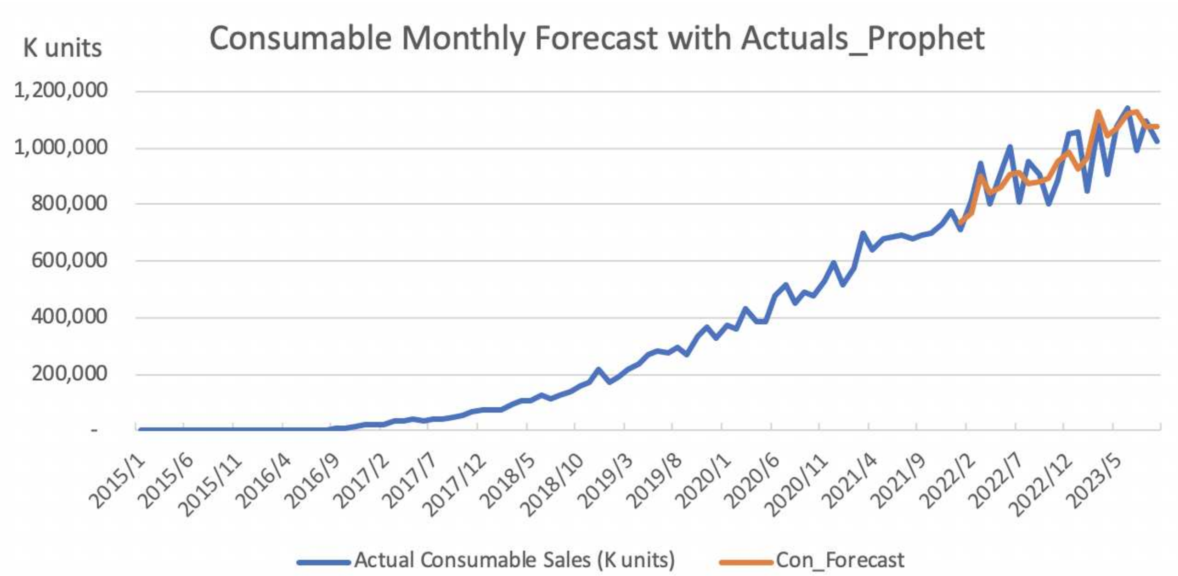
5.3. Hyperparameter Tuning and Model Explainability

As mentioned in Section 4.5, we can use grid search to find the optimal parameter setting and optimize the performance of a model. In addition, we can retrieve these best hyperparameters and use them to interpret the model. Table 4 shows the results of Prophet when a hyperparameter is applied. We can see that across all the selected error metrics: MAE, MAPE, and RMSE, the hyperparameter tuning approach successfully produces the best performance out of the Prophet model for HNB consumables.

To see the visualization of how well the hyper-tuned Prophet models have improved, we plot the predictions against the actual values in Figure 6 for HNB consumables. Unlike the forecast curves for HNB consumables with the untuned Prophet model (illustrated in Figure 5) where the model only captures the trend but fails to generalize the seasonality in the data, the fine-tuned models are able to effectively pick up both the trend and seasonality from the training set and make the best predictions to its ability on the testing set.

Figure 6

Monthly Forecast with Actual Values – HNB Consumables with Tuned Prophet Model



After implementing hyperparameter tuning for Prophet using grid search and optimizing the model's performance, we retrieve the optimal input values of these parameters based on different error metrics. Table 6 summarizes the combination of these inputs contributing to the lowest error metrics for HNB consumables.

Table 5

Best Combinations of Hyperparameters

Prophet	Consumables		
Hyperparameter	**Best MAE**	**Best MAPE**	**Best RMSE**
changepoint_prior_scale	0.01	0.1	0.01
changepoint_range	0.95	0.95	0.9
seasonality_prior_scale	5.0	1.0	10.0
seasonality_mode	multiplicative	multiplicative	multiplicative

In Table 5, the combination of hyperparameters suggests the model forecasts with rigid trends and a higher degree of seasonal effect to get the best MAE. It also indicates that the model focuses more on historical data, making it less adaptable to recent data and a multiplicative seasonality pattern is also observed along the trend in forecasting.

6. Managerial Implications

The findings from our capstone project have several significant managerial implications for decision-makers in the sponsor company regarding the future forecasting practice of HNB consumables.

The granularity of data impacts model performance. Our study shows the limitation of small datasets impeding the model's ability to make good predictions. We suggest the sponsor company collect as much sales data of HNB consumables as possible. In our opinion, the sponsor company should ideally have daily sales data instead of monthly. This can increase the granularity level in the dataset and enhance the model's ability to discover detailed patterns underlying HNB consumables sales and create better forecasts. With higher granularity, we expect to have more robust and accurate forecasting models. Managers can also consider investing resources in data expansion techniques to enhance the accuracy of sales forecasts, thereby facilitating better strategic decision-making.

Time series feature extraction is essential to time series forecasts. Our work suggests that considering temporal features extracted from time series data is essential to time series forecasts. Not only can the time features improve the model's performance, but they can also help managers develop comprehensive business strategies by leveraging the trends and seasonality observed in the data.

Understanding selected models is key to outcome comparison. The comparison of traditional time-series models and machine learning algorithms shows certain models' ability to capture the nuances of HNB consumables sales. Manager needs to understand how the selected models work before comparing the models' results. These insights can be used to select appropriate forecasting models tailored to the specific needs of the forecasting purpose. For instance, our study shows that ML models like Prophet perform better than traditional models, such as the moving average and the Holt-Winter methods. It also captures complex trends and seasonality patterns (see Figure 6), particularly in dynamic market environments where HNB consumables are present.

Model optimization reveals further insights into making robust forecasting. Our study highlights the importance of fine-tuning model parameters and optimizing data-splitting strategies to enhance forecasting accuracy further. Conducting sensitivity analysis using these approaches helps managers optimize model performance and ensure robustness across varying market conditions. This optimization process not only facilitates managers to derive actionable insights from sales forecasts but also enables proactive decision-making and strategic planning.

Hyperparameter tuning provides model's explainability. The analysis provides valuable insights into the explainability of model results, particularly in understanding the impact of hyperparameters on forecasting accuracy based on the Prophet model. Managers can use these insights to interpret model outputs, review optimal parameters, and better understand the key factors driving sales performance.

7. Conclusions

This capstone focuses on forecasting HNB consumables sales. Historical sales data show a strong trend because the product is not yet mature in the market. Sales data also show seasonality due to the tobacco industry's dynamics. It isn't easy to reach an accurate forecast for products with strong trend that are not mature in the market. Our approach is to solve this challenge by utilizing ML models and creating an advanced forecasting solution.

Our methodology starts with understanding our sponsor company's data and business dynamics. Since our sponsoring company provided us with a limited amount of data, we applied data expansion techniques before training our ML models. Further, we extract time features from data, then apply the ML models and compare them based on selected error metrics. In addition, we apply hyperparameter tuning and different train-test splits to improve the model performance.

The results we have are promising. We identified Prophet as the best-performing model, and the error metrics beat all traditional forecasting alternatives and other ML models. After hyperparameter tuning, we observe that the MAPE is only 7.05%. Considering the complexity of the business, which has both trend and seasonality, this is a satisfactory result.

In future work, the study should be repeated with a higher granularity of sales data. The sponsor company should collect daily sales data and repeat this study with the actual sales dataset instead of the expansion technique. We believe such work has the potential to provide even better forecasting results.

References

Amar, J., Rahimi, S., Surak, Z., & von Bismarck, N. (2022, February 15). *AI-driven operations forecasting in data-light environments*. Retrieved from https://www.mckinsey.com/: https://www.mckinsey.com/capabilities/operations/our-insights/ai-driven-operations-forecasting-in-data-light-environments

Amazon Web Services. (2021, December 3). *Predicting The Future Of Demand: How Amazon Is Reinventing Forecasting With Machine Learning*. Retrieved from www.forbes.com: https://www.forbes.com/sites/amazonwebservices/2021/12/03/predicting-the-future-of-demand-how-amazon-is-reinventing-forecasting-with-machine-learning/

Angelov, P. P., Soares, E. A., Jiang, R., Arnold, N. I., & Atkinson, P. M. (2021). Explainable artificial intelligence: an analytical review. *WIREs Data Mining and Knowledge Discovery*.

Boone, T., Ganeshan, R., Jain, A., & Sanders, N. R. (2019, January-March). Forecasting sales in the supply chain: Consumer analytics in the big data era. *International Journal of Forecasting*, pp. 170-180.

Caplice, C., & Ponce, E. (2023). *MITx MicroMasters Program in SCM Key Concepts.* Cambridge, MA, USA: MITx MicroMasters® Program in Supply Chain Management.

Carbonneau, R., Laframboise, K., & Vahidov, R. (2008). Application of machine learning techniques for supply chain demand forecasting. *European Journal of Operational Research*, 1140-1154.

Chern, C.-C., Ao Ieong, K., Wu, L.-L., & Kung, L.-C. (2010). Designing a decision-support system for new product sales forecasting. *Expert Systems with Applications*, Pages 1654-1665.

da Silva, L. C., & Teles, P. J. (2019). *Market Sizing for "Reduced-Risk Products" in the Tobacco Industry.* Porto: University of Porto.

Edwards, R., Hoek, J., Karreman, N., & Gilmore, A. (2022). Evaluating tobacco industry ‘transformation’: a proposed rubric and analysis. *Tobacco Control*, 31:313-321.

Eiskowitz, S. (2021). A Machine Learning Approach for Forecasting with Limited Data and for Distant Time Horizons. *Graduate Theses.* Massachusetts Institute of Technology.

Foundation for a Smoke-Free World. (2022, September 16). *2022 Index Ranking Report.* Retrieved from https://tobaccotransformationindex.org/: https://tobaccotransformationindex.org/wp-content/uploads/2022/09/2022-Index-Ranking-Report_September-2022-1.pdf

Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn Keras and Tensorflow Concepts.* O’Reilly Media.

Goodwin, P., Meeran, S., & Dyussekeneva, K. (2014). The challenges of pre-launch forecasting of adoption time series for new durable products. *International Journal of Forecasting*, 1082-1097.

Grennan, L., Kremer, A., Singla, A., & Zipparo, P. (2022, September 29). *Why businesses need explainable AI—and how to deliver it*. Retrieved from www.mckinsey.com: https://www.mckinsey.com/capabilities/quantumblack/our-insights/why-businesses-need-explainable-ai-and-how-to-deliver-it

Hamoudia, M., & Vanston, L. (2023). Machine Learning for New Product Forecasting. In M. Hamoudia, S. Makridakis, & E. Spil, *Forecasting with Artificial Intelligence* (pp. 77 - 106). Palgrave Macmillan.

Hamoudia, M., Makridakis, S., & Spiliotis, E. (2023). *Forecasting with Artificial Intelligence.* Palgrave Macmillan.

Iftikhar, R., & Khan, M. S. (2020). Social Media Big Data Analytics: Development and Case Implementation of an Innovative Framework. *Journal of Global Information Management, Volume 28*.

Izaguirre, d., Guillermo, F., & Chao, T.-N. (2020). Identifying the Root Causes of Stockout Events in e-commerce Using Machine Learning Techniques. *Supply Chain Management Capstone Projects.* Massachusetts Institute of Technology.

Jain, P. K., Pamula, R., & Sriva, G. (2021). A systematic literature review on machine learning applications for consumer sentiment analysis using online reviews. *Computer Science Review, Volume 41,*.

Jennings, C., Wu, D., & Terpenny, J. (2016). Forecasting Obsolescence Risk and Product Life Cycle With Machine Learning. *IEEE Transactions on Components, Packaging and Manufacturing Technology, vol. 6, no. 9*, 1428-1439.

Kahn, K. B. (2002). An exploratory Investigation of new product forecasting practices. *The Journal of Product Innovation Management 19*, 133 - 143.

Kahn, K. B. (2006). *New Product Forecasting: An Applied Approach.* USA: Routledge.

Linardatos, P., Papastefanopoulos, V., & Kotsiantis, S. (2021). Explainable AI A Review of Machine Learning Interpretability Methods. *Entropy*.

Lynn, G. S., Schnaars, S. P., & Skov, R. B. (1999, November). A Survey of New Product Forecasting Practices in Industrial High Technology and Low Technology Businesses. *Industrial Marketing Management*, pp. 565-571.

Mattei, M. M., & Platikanova, P. (2017). Do product market threats affect analyst forecast precision. *Review of Accounting Studies volume 22*, 1628–1665.

Mia, A. R., Yousuf, M. A., & Ghosh, R. (2021). Business Forecasting System using Machine Learning Approach. *2nd International Conference on Robotics* (pp. 314-318). DHAKA, Bangladesh: Electrical and Signal Processing Techniques (ICREST).

Misheva, B. H., Osterrieder, J., & Hirsa, A. (2021). Explainable AI in Credit Risk Management.

Nimmy, S. F., Hussain, O. K., Chakrabortty, R. K., Hussain, F. K., & Saberi, M. (2022). Explainability in supply chain operational risk management: A systematic literature review. *Knowledge-Based Systems, Volume 235*.

Nowadly, K. G., & Jung, S. (2020). Using Machine Learning Approaches to Improve Long-Range Demand Forecasting. *Supply Chain Management Capstone Projects.* Massachusetts Institute of Technology.

Paruthipattu, S. P., & Haycock, B. (2021). *Demand Forecasting Based on External Factors Using Clustering and Machine Learning.* Dublin: School of Computing, National College of Ireland.

Pavlyshenko, B. (2019). Machine-Learning Models for Sales Time Series Forecasting. *Data*, 1-11.

Petropoulos, F., & et al. (2022, July-September). Forecasting: theory and practice. *International Journal of Forecasting*, pp. 705-871.

Schröer, C., Kruse, F., & Gómez, J. M. (2021). A Systematic Literature Review on Applying CRISP-DM Process Model. *Procedia Computer Science*, Pages 526-534.

Seyedan, M., & Mafakheri, F. (2020, July). Predictive big data analytics for supply chain demand forecasting: methods, applications, and research opportunities. *Journal of Big Data*.

Silverstein, S. (2020, September 17). *Walmart uses AI to predict demand*. Retrieved from www.supplychaindive.com: https://www.supplychaindive.com/news/walmart-grocery-AI-demand-operations/585424/

Singh, A., Thakur, N., & Sharma, A. (2016). A review of supervised machine learning algorithms. *3rd International Conference on Computing for Sustainable Global Development* (pp. pp. 1310-1315). New Delhi: INDIACom.

Småros, J. (2002). *Collaborative Forecasting in Practice.* Birmingham, UK: Logistics Research Network 2002 Conference.

Smirnov, P. S., & Sudakov, V. A. (2020). Forecasting new product demand using machine learning. *Journal of Physics: Conference Series, Volume 1925.* IOP Publishing Ltd.

Smith, T. T., O'Connor, R. J., Cummings, M. K., & Liber, A. C. (2023). *Tobacco control strategies.* Elsevier.

Souad Boutane, Bart Zeeman, Ben Fridolin, Cornee Geenen, Ahsan Ali, Mehdi Noori, Huzefa Rangwala, Yifu Hu. (2023, April 6). *Predicting new and existing product sales in semiconductors using Amazon Forecast*. Retrieved from aws.amazon.com: https://aws.amazon.com/blogs/machine-learning/predicting-new-and-existing-product-sales-in-semiconductors-using-amazon-forecast/

Steenbergen, R. v., & Mes, M. (2020, December). Forecasting demand profiles of new products. *Decision Support Systems*.

Susmita, R. (2019). A Quick Review of Machine Learning Algorithms. *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing* (pp. pp. 35-39). Faridabad: COMITCon.

Trapero, J. R., de Frutos, E. H., & Pedregal, D. J. (2023, October 8th). Demand forecasting under lost sales stock policies. *International Journal of Forecasting*.

Uriarte, D. A. (2010). Developing a framework for dependable demand forecasts in the consumer packaged goods industry. *Graduate Theses.* Massachusetts Institute of Technology.

Yan, X., & Hu, H. (2023, August). New product demand forecasting and production capacity adjustment strategies: Within-product and cross-product word-of-mouth. *Computers & Industrial Engineering*.