Vaasan yliopisto
UNIVERSITY OF VAASA

Dimple Talukder Eka

The Role of Artificial Intelligence in Demand Forecasting & Inventory Management

School of Technology and Innovation
Master's Thesis in Industrial Management programme

Vaasa 2025

UNIVERSITY OF VAASA			
School Of Technology and Innovation			
Author:	Dimple Talukder Eka		
Title of the thesis:	The Role of Artificial Intelligence in Demand Forecasting & Inventory Management		
Degree:	Master of Sciences in Technology and Innovation		
Degree Programme:	Industrial Management		
Supervisor:	Rayko Toshev		
Year:	2025	**Pages:**	71

ABSTRACT:

Demand forecasting plays a critical role in aligning production, inventory, and distribution decisions with market demand. However, many organizations continue to experience stockouts, excess inventory, and planning uncertainty due to insufficient forecasting accuracy. Recent advances in artificial intelligence (AI) offer new opportunities to improve predictive performance, particularly in volatile and seasonal retail environments with large product assortments. Despite these advances, the extent to which improvements in forecasting accuracy translate into tangible inventory performance gains remains insufficiently explored. This thesis investigates this gap by comparing three AI-based forecasting models—XGBoost, Long Short-Term Memory (LSTM) networks, and the Temporal Fusion Transformer (TFT)—with conventional statistical forecasting methods, focusing on both forecast accuracy and inventory performance.

The study adopts a positivist and deductive research design based on quantitative modeling and empirical analysis. Forecasting models are trained, validated, and evaluated using publicly available data from the M5 Forecasting Competition under controlled experimental conditions. Forecast accuracy is assessed using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Scaled Error (MASE). To evaluate operational impact, model forecasts are embedded within a discrete-event inventory simulation based on a continuous-review (s,S) replenishment policy. Inventory performance is measured using stockout rates, fill rates, and average inventory levels. The results demonstrate that AI-based models significantly outperform traditional statistical baselines, with the TFT model achieving an average RMSE reduction of 21.8% relative to the best-performing statistical model. These accuracy improvements translate into measurable operational benefits, as the TFT-based forecasts reduce the average stockout rate to 6.4%, compared with 10.9% for the strongest statistical baseline.

The findings contribute to the existing literature by providing empirical evidence that improvements in forecasting accuracy can generate meaningful downstream inventory performance gains. The study offers methodological insights into model selection for complex retail environments, feature engineering, and inventory simulation design. Overall, the research highlights the importance of integrating advanced forecasting models with operational evaluation frameworks to support data-driven inventory planning and decision-making.

KEYWORDS: Demand forecasting, artificial intelligence, inventory management, XGBoost, LSTM, Temporal Fusion Transformer, supply chain analytics, M5 dataset.

Contents

Figures

Tables

Abbreviations

AI	Artificial Intelligence
ANN	Artificial Neural Network
ARIMA	Autoregressive Integrated Moving Average
CFD	Cumulative Forecast Deviation
DL	Deep Learning
ETS	Exponential Smoothing (Error–Trend–Seasonal model)
GRU	Gated Recurrent Unit
KPI	Key Performance Indicator
LSTM	Long Short-Term Memory Network
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MASE	Mean Absolute Scaled Error
ML	Machine Learning
MSE	Mean Squared Error
RMSE	Root Mean Squared Error
RNN	Recurrent Neural Network
SKU	Stock Keeping Unit
SVM	Support Vector Machine
TFT	Temporal Fusion Transformer
IoT	Internet of Things
RMSFE	Root Mean Squared Forecast Error
ERP	Enterprise Resource Planning
POS	Point of Sale
SCM	Supply Chain Management
SMAPE	Symmetric Mean Absolute Percentage Error
WMA	Weighted Moving Average
WAPE	Weighted Absolute Percentage Error
DLRM	Deep Learning Regression Model
CNN	Convolutional Neural Network (if mentioned even once)
M5	Walmart M5 Forecasting Dataset (Kaggle Competition)

(s, S).	Reorder Point and Order-Up-To Level Inventory Policy
EOQ	Economic Order Quantity
RNN	Recurrent Neural Network
CPU/GPU	Central / Graphics Processing Unit

1 Introduction

Successful global organizations are the ones which estimate market needs rightly. To do so, demand forecasting is dependent on these companies. The term "backbone" of organizations that aim to have an excellent, profitable and sustainable business model (Ren, S., Chan, H. L., & Siqin, T, 2020, p. 13). Nevertheless, when companies are able to attain some degree of it, forecasting accuracy is not very easy. Poor levels of accuracy in demand forecasting would lead to two key issues.

Effective supply chain management requires proper demand forecasting. It has a direct effect on the level of inventory, resource distribution, and the general level of efficiency (Fröhlich, 2025, p. 120). As the sphere of e-commerce rapidly develops and consumer tastes grow more dynamic, organizations have to resort to new forecasting strategies to stay competitive. The established methods, e.g., moving averages and exponential smoothing do not necessarily respond to the specifics of modern retail where the demand patterns are sometimes nonlinear and highly fluctuating (Hyndman, R. J., Athanasopoulos, 2020, p. 98).

1.1 Background and Context: Traditional Forecasting Challenges

Regardless of the progress in the use of old methods, retailers are often confronted with the problem of accurate inventory and stockout. Such problems can be explained by constraints of the traditional models, particularly their inefficiency in explaining demand variations as affected by external elements like promotions, seasonality, and economic patterns (Saniuk, S., Grabowska, S., & Gajdzik,, 2020, p. 20).

Consequently, high costs are incurred by most organizations in terms of surplus inventory and missed sales caused by stock outs. According to a study by A Fraser, ineffective demand forecasting might result in wastage and low profitability, and thus, improvements to the accuracy of forecasting must be made.

Subjective and objective methods are some of the traditional forecasting methods (Giannopoulos, P. G., Dasaklis, T. K., Tsantilis, I., & Patsakis, C, 2025, p. 23).

Time series methods are objective methods which are more or less identical to the one mentioned. Trends of the data with time using models like moving averages or exponents (Tseng, 2024, p. 2). The Holt method is famous for providing traditional forecasting. When there is a trend in the data, it is possible to make predictions that are accurate. In the event of a trend and seasonality, the Holt-Winter method has the best result with reduced error metrics (Tseng, 2024, p. 5).

Time series methods have been implemented in several industries due to their simplicity and generally acceptable nature (Eiskowitz, 2021, p. 30). When there is a causal approach being used, it is the linear regression approach relationship, which exists between the variables (Seyedan, M., & Mafakheri, F., 2020, p. 23).

1.2 The Role of Artificial Intelligence and Machine Learning in Demand Forecasting

One of the most profitable uses of machine learning (ML)-based models has been demand forecasting (Smirnov, P. S., & Sudakov, V. A, 2011). The superiority of the ML models over the traditional forecasting techniques has been demonstrated in numerous academic works (Mia, M. A. R., Yousuf, M. A., & Ghosh, R, 2021) and consulting reports. The well-known entrepreneurs like Amazon and Walmart (Chen, 2025) have already managed to implement ML models in their demand forecasting.

Therefore, advanced ML applications are also investigated in our research. A good example is the work of (Pavlyshenko, B. M., 2019). The author concentrates on the sales data in retail stores and implements various approaches.

Another such example is presented which demonstrates that the Back Propagation Neural Network (BPNN) model is capable of obtaining high-accuracy forecasts in the retail industry.

Since the needs of consumers have been turning out to be increasingly complex, AI has emerged as a revolutionary phenomenon of enhancing the precision of demand prediction. The capabilities of AI technologies and ML algorithms are not limited to processing volumes of data, but also to identifying complex patterns that standard models cannot detect. (Chen, M. Y., Bose, S., & Zhang, Y., 2023).

The notable models of AI implemented in addressing the demand forecasting model include XGBoost, Long Short-term Memory (LSTM) and Temporal Fusion Transformers (TFT) models, which are tailored to take advantage of the dynamics of the data effectively and yield superior predictions in complex environments.

The proposed study will be systematic research to investigate the effectiveness of AI-based demand forecasting models within the retail supply chain, specifically their capacity to affect supply chain inventory management indicators such as stockout and holding costs.

The emergence of new artificial intelligence has led to an array of solutions which are more efficient than more conventional strategies of prediction on the basis of recognition of complex trends, and peculiarities of demand signals.

1.3 Personal Motivation and Field of Science

My personal academic and professional motivation is also a factor in the choice of the thesis topic. During my course of study, Industrial Management, I have been interested specifically in the use of data-driven technologies to increase operational decision-making in supply chains. The reason behind my interest in demand forecasting and inventory management was the fact that most companies still seem to have problems with their demand forecasting, stockouts, excess inventory, and market volatility. Meanwhile, recent innovations in artificial intelligence, particularly machine learning and deep learning, have shown good potential in increasing the accuracy of forecasts in complicated environments.

This made me seek to understand the relationship between AI-based forecasting models and real inventory management results. It is thus a natural subject of the main areas of interest of Industrial Management and Production and Operations Management, especially in the fields of data analytics, supply chain optimization, and decision support systems.

1.4 Research Problem and Gap

The retail supply chains are highly dependent on accurate demand forecasts to reduce the stockouts and the holding cost. The use of traditional techniques like ARIMA and ETS does not tend to reflect demand volatility and the impact of promotions. Boost, LSTM, Transformer AI models have demonstrated higher accuracy in AI competitions like M5, yet little empirical evidence has been found on whether the accuracy gains apply to measurable inventory benefits. Namely, the correlation between the reduction of forecast errors and the stockout rates in the performance of the standard inventory policies ((s, S)) is not explored properly.

1.5 Research Questions and Hypothesis

This study is organized around a single central Research Question (RQ) and two related Hypotheses (H), building on the gap found in the existing literature (Section 1.4).

How much better do AI forecasting models (XGBoost, LSTM, TFT) predict demand more accurately than traditional statistical models (ARIMA, ETS, Prophet) and how this impacts stockout rates on an (s, S) inventory policy?

Hypotheses (H):

H1: AI forecasting models have a much lower forecast error (MAE, RMSE) than the traditional statistical forecasting models on retail demand data.

H 2: Decreases in forecast error will be associated with statistically significant decreases in inventory simulation stockout rate.

1.6 Scope and Limitations

In this thesis, the retail and the FMCG sectors are considered, using publicly accessible data sets, including the M5 forecasting dataset. It is a mostly quantitative and experimental study. Nevertheless, some weaknesses are the lack of proprietary corporate information, and the assumption of the parameters related to lead times and cost structures, which can limit the extrapolation of the results to the rest of the sectors.

The plan is to use the methodology of the Cross Industry Standard Process of Data Mining (CRISP-DM) as a de facto standard of data mining practice as in our work plan (Schroer, K., Ketter, W., Lee, T. Y., Gupta, A., & Kahlen, M, 2022). Such a systematic process allows not only to monitor the process of project development but also to guarantee its quality. Figure 1 reflects the high-level flowchart of the steps of capstone project.

The phases of our work plan are:

Business Knowledge: Clarify the goal, success measures and the particular questions that will be answered in the demand forecasting field.
Data Understanding: Find and explore the relevant data sets, such as M5 and Kaggle retail, and ensure that the initial insights and patterns have been found out.
Data Preparation: Clean and pre-treat the datasets in order to deal with missing values, and create important features (lags, rolling averages, promotions, etc.) to use in the modeling.
Modeling: It is one of the most complicated AI models as compared to other models and use both simple forecasting models (ARIMA, ETS, Prophet).
Analysis: Measure the performance of the models by evaluating the accuracy measures of Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE).
Deployment: Apply (s, S) policy with the used forecasts to simulate inventory management and quantify the rates of stockout and the overall performance of the inventory.

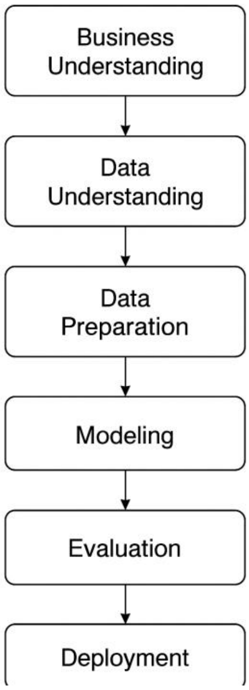

Figure 1 *High-Level Process of the CRISP-DM Methodology (Lind, E., & Glas, S. , 2022).*

1.7 Significance of the Study

This study is valuable not only as an academic discovery, but its practical implications with respect to the strategies of demand forecasting and inventory management can be applied by practitioners in retail and fast-moving consumer goods (FMCG) to obtain the most favorable outcomes. Focusing on the correlation between the events in the field of AI and the traditional method, the study will endeavor to describe how performance forecasting is related to the real performance of the business.

2 Literature review

2.1 Historical Background of Demand Forecasting

The demand forecasting has changed in the past. It shows developments in calculating and computer abilities and the state of the market (Walter, A., Ahsan, K., & Rahman, S., 2025). This development can be split into some bright epochs and each of them is marked by its own challenges and findings.

2.1.1 The Traditional Era (Early 20th Century)

Demand forecasting can be traced back to the early 20th century. During that period, techniques were rudimentary and predominantly made out of basic estimates and bare extrapolations (Feizabadi, J., 2022). Little data was available and prediction relied to a great extent on experience. Croston made a significant breakthrough in 1972 with a procedure for predicting intermittent demands. His model provided the way to further improvements in demand forecasting because it addressed the particular problems connected with the infrequent sales of a product (Croston, 1972).

2.2 Introduction to Demand Forecasting

Demand forecasting can be described as a process of projecting the future demand of a product or service among customers. It is a significant process in supply chain management due to which companies may coordinate their production processes, inventory, and distribution processes with the demands of the market. Proper demand forecasting may enhance effectiveness in the supply chain, decrease expenses, and customer satisfaction (Chopra, S., & Meindl, P. , 2019).

The significance of demand forecasting has resulted in the creation of different approaches that can be categorized into two broad approaches, namely qualitative and quantitative (Babai, M. Z., Boylan, J. E., & Rostami-Tabar, B., 2022). The qualitative techniques, which include the expert judgment and focus group, are based on an opinion and are normally employed where there is a lack of historical information. On the other hand, quantitative methods use statistical methods and historical data to give a forecast (Hyndman, R. J., & Athanasopoulos, 2018).

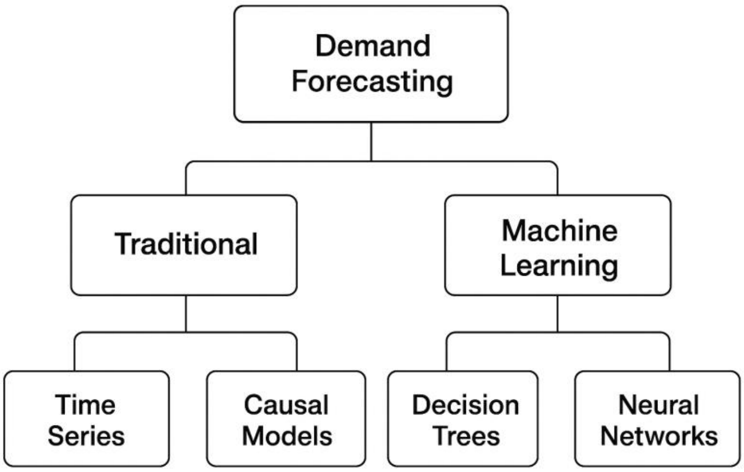

Figure 2 *Demand Forecasting Techniques: Traditional vs. Machine Learning Approaches (Hyndman, R. J., & Athanasopoulos, 2018)*

2.3 Challenges in Demand Forecasting

This section would outline the pitfalls in demand forecasting that would involve:

Data Quality Problems

Illustrate the relevance of quality and real-time and full data in effective forecasting. Share views on how omissions and false data can lead to wrongful predictions.

Dilemmas in Model Selection

Get to know the issues of choosing the right forecasting approach to use in different scenarios, specifically, how model complexity/simplicity influences performance.

External Influences

These are extraneous elements like seasonality, market volatility and emerging trends that can influence the accuracy of the forecast.

These issues highlighted in the necessity to create the tools and methods of forecasting and, in particular, the significance of machine learning solutions. These difficulties are also important to note, and this will provide a richer literature review by stating the landscape within which these forecasting models are working.

2.4 Traditional Demand Forecasting Methods

2.4.1 Time Series Analysis

One of the most popular quantitative methods of demand analysis is time series analysis (Gruver, N., Finzi, M., Qiu, S., & Wilson, A. G. , 2023). It entails analyzing historical data points that have been gathered over a period of time to determine trends, seasonal trends, and cyclical trends. Conventional time series models of analysis involve moving averages, exponential smoothing and ARIMA models.

Moving Averages: This is a technique that averages previous demand information to give future demand. Simple moving averages use a fixed number of historical observations, whereas weighted moving averages put dissimilar weights on the points (Hyndman, R. J., & Athanasopoulos, 2018).

Exponential Smoothing: Under this method, observations made recently are more weighted, which implies that it is a more responsive method used to make a forecast regarding recent changes in demand. Single, double, and triple forms of this method are variants of it (Hyndman, R. J., & Athanasopoulos, 2018).

ARIMA Models: Autoregressive Integrated Moving Average (ARIMA) Model is the combination of the moving averages and autoregression that predicts the elements ahead in a sequence. ARIMA can be especially useful in non-stationary data with differences (Zaim, S., Yaziz, S. R., Zakaria, R., Mohamad, N. N., & Radi, N. F. A, 2025).

2.4.2 Causal Models

Causal models consider relations between demand and other factors (predictors). Typical approaches include regression analysis and econometric models.

Regression analysis studies how demand for a product is related to independent variables like price or advertising spend and allows forecasters to predict demand changes from the influences that are known (Muth, M., Lingenfelder, M., & Nufer, G, 2024).

Econometric models are more complex than simple regression models, and they usually consist of multiple related equations and dynamic relationships among other economic factors (Muth, M., Lingenfelder, M., & Nufer, G, 2024).

2.5 Limitations of Traditional Approaches

Although traditional demand forecasting techniques have been used for a long time, they have some important limitations. Often, they have difficulty detecting complex and nonlinear relationships in the data, and these methods may be less accurate when demand behavior is subject to variation and influenced by large numbers of external factors (Hyndman, R. J., Athanasopoulos, 2020). In addition, classical methodologies are based on constant assumptions of the data that cannot be realistic in rapidly developing markets (Petropoulos, F., Kourentzes, N., Nikolopoulos, K., & Siemsen, E., 2018).

2.6 Machine Learning Applications in Demand Forecasting

2.6.1 Overview of Machine Learning in Forecasting

Machine learning has provided a novel paradigm of demand forecasting by replacing strict, assumption-based statistical forecasts with loose, data-oriented models. In contrast to conventional methods that depend on fixed trend or seasonality models, ML algorithms are directly trained on the past and external data, which means that they are adaptable to high-frequency, unpredictable, or highly-influenced demand (Zohdi, M., Rafiee, M., Kayvanfar, V., & Salamiraad, A., 2022).

The effectiveness of the ML techniques is due to the fact that they can embrace a large array of variables, including prices, events, promotions, calendar effects, etc., and can represent non-linear relationships that the traditional models find difficult to embody. Since retail datasets have increased in size and depth, such algorithms have become sensitive enough to identify behavioral interactions that build demand (Khan, 2025).

All in all, machine learning complements the conventional method of making predictions with more flexibility, the capacity to learn complexities and responsiveness to evolving market conditions (Van Chau, D., & He, J., 2024).

2.6.2 Common ML Techniques in Demand Forecasting

Decision Trees: These are models which divide data according to a number of parameters in the form of a tree, thus easy to explain and understand. Random Forest and Gradient Boosting techniques are methods that improve the basic decision trees, multiplying the models in order to minimize overfitting and increase the accuracy (Biecek, P., & Burzykowski, T., 2021).

Artificial Neural Networks (ANNs): ANNs are based on the human brain neural networks that are natural. They are good at capturing non-linear relationships; hence, they can be used in demand forecasting. Certain networks such as Long Short-Term Memory (LSTM) networks are specifically created to operate on sequential information (Hochreiter, S., & Schmidhuber, J., 1997).

Support Vector Machines (SVM): SVMs are useful when there is a classification and regression problem. Through mapping data to high dimensional space, SVMs are able to identify the optimal hyperplane that separates classes or predicts values according to features at the optimum level (Smola, A. J., & Schölkopf, B, 2004).

Temporal Fusion Transformers (TFT): TFTs are new models that are considered to work with time series data and take into account temporal patterns and structural information

(Lim, B., Arık, S. Ö., Loeff, N., & Pfister, T., 2021).

2.7 Sector-Specific Applications

This section allows us to explore the ways various industries can use the methods of demand forecasting adapted to the unique situations.

Because each industry has different operational constraints and data characteristics, the practical use of demand forecasting systems differs greatly (Van Giffen, B., Herhausen, D., & Fahse, T., 2022). For example, forecasting in manufacturing is closely linked to supply chain logistics and production planning, with an emphasis on long-term capacity needs and raw material lead times. In contrast, the healthcare industry uses forecasting to manage vital medical supply inventories and optimize resource distribution, frequently coping with erratic demand and significant error consequences (Niaz, M., & Nwagwu, U, 2023). However, this study's empirical emphasis on the retail industry, offers a unique combination of possibilities and limitations (Alhejaily, 2024).

Demand forecasting is largely used by retailers as an essential inventory management tool that has a direct impact on key performance metrics like holding costs and stockout rates (Bala, P. K., 2012). Additionally, it informs pricing and promotional strategies, making it a sales tool. The most effective strategies in this particular field have been found to be sophisticated techniques that can capture promotional, seasonal, and calendar effects such as the AI/ML models examined in this thesis, given the extremely dynamic, high-volume, and time-sensitive nature of retail data.

The study of sector-specific applications does not only provide a depiction of the practical applicability of the demand forecasting techniques, but it also shows how the techniques can be applied in various areas. This is a direct linkage of theory to real world application and your literature review makes more applicable and insightful.

2.8 Inventory Theory and the Forecast–Inventory Link

The inventory theory sets the theoretical basis on which forecasting accuracy determines the operational performance in the supply chain (Aviv, 2003). Classical models of inventory models, including the Economic order Quantity (EOQ), base-stock models, and the continuous-review policies, are based on assuming that demand in the future is uncertain and has to be estimated by use of forecasts. Consequently, the level of replenishment decisions, safety stock and service performance directly depend on the quality of demand forecasts (Bazsa-Oldenkamp, 2002).

One of the main concepts in the theory of inventory is that demand uncertainty raises the safety stock levels. In forecasts where the error or variability is high, companies have to maintain a stock buffer to prevent stockouts, hence increasing holding costs. On the other hand, with a reduction in forecast error, safety stock requirements are reduced, resulting in a more efficient inventory level. It is through this relationship that the theoretical foundation between forecasting performance and inventory behavior is made: the lower the forecast error, the lower the total inventory cost, and the higher the levels of service (Silver, E. A., Pyke, D. F., & Peterson, R., 1998).

Continuous-review policies like the (s, S) model adopted in this paper directly include the forecast-based demand estimates into the computation of the reorder point (Savvopoulou, 2024). The reorder point (s) indicates what is expected as demand within the lead time with a safety stock that is as a result of the variability of demand. Thus, when the forecasting accuracy is enhanced, the demand variance in the lead time reduces, and the number of stockout instances is reduced, as well as the replenishment cycles proceed more smoothly (Tadayonrad, Y., & Ndiaye, A. B, 2023). This is the reason why sophisticated forecasting models could have a quantifiable operating impact beyond merely minimizing error levels.

The most recent research has focused particularly on integrating forecasting approaches into the inventory decision frameworks. According to research by (Babai, M. Z., Boylan, J. E., & Rostami-Tabar, B., 2022), a forecasting model's operational value depends not only on its accuracy but also on how well it can be applied to support inventory policies in the face of real-world uncertainties. In particular, irregular and promotion-driven demand patterns improve the forecast-inventory connection since these conditions make inventory performance more susceptible to prediction errors.

Though there is a substantial empirical gap, the theoretical relationship between lower prediction error and better inventory performance is well-established in classical inventory models (Silver, E. A., Pyke, D. F., & Peterson, R., 1998). Most recent research showing the greater accuracy of cutting-edge AI models (like LSTM and TFT) has been limited to accuracy metrics (like MAE and RMSE) in forecasting competitions (like M5). When the forecasts are applied to a standard, continuous-review (s, S) inventory system under retail conditions, there is a lack of comprehensive empirical data that rigorously quantifies how this particular accuracy gains translate into quantifiable, evident inventory outcomes, specifically the decrease in stockout rates.

2.9 Benefits of Machine Learning in Demand Forecasting

The shift from the traditional statistical approaches to machine learning (ML) models in demand forecasting is predetermined by a number of specific benefits in terms of data processing and prediction.

The first advantage is higher accuracy, especially when working with more complicated datasets that have multiple interacting variables. The findings of researchers such as (Hamoudia, M., & Vanston, L., 2023) and many others indicate that ML models have always yielded better forecast accuracy due to their capacity to model the non-linear relationship and high-order interactions of features that are inaccessible to the classical linear models. This superior predictive accuracy is often accompanied by greater flexibility. ML algorithms are more receptive to new trends, abrupt modifications, and externalities compared to conventional ones, making them highly applicable in unstable or fast-changing market conditions. Finally, ML models have high Efficiency, since they can process very large amounts of various data (sales history, weather, and promotion calendar, and price) within a very short amount of time. Such an ability enables them to make appropriate and pertinent forecasts that can be used in real-time as operations' managers make critical decisions (Biecek, P., & Burzykowski, T., 2021).

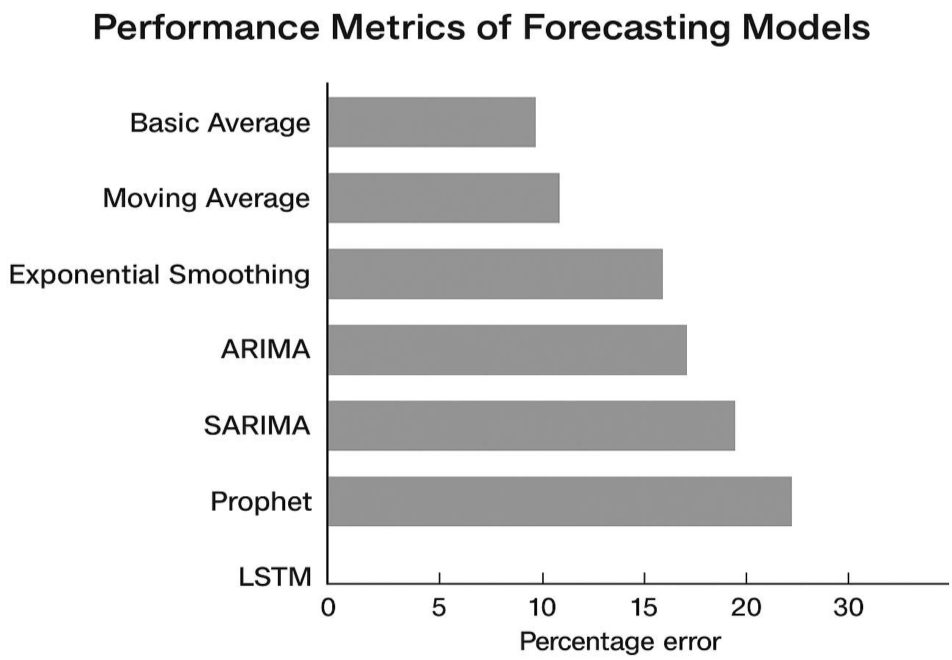

Figure 3 *Common Performance Metrics for Industry Forecasting Models (Makridakis, S., Petropoulos, F., & Spiliotis, E., 2022).*

2.10 Industrial and Sectoral Use Cases of Machine Learning

The theoretical advantages of machine learning (ML) models are proven through the multiple successful applications of machine learning models in the different industries and it proves to be much better in complex forecasting settings.

Advanced ML algorithms are being actively adopted by retail giants, including Amazon and Walmart, where advanced algorithms are applied to the central demand forecasting. These firms utilize vast datasets of historic sales, client behavior statistics, rival pricing, and external information (such as social media sentiment and weather) to make extremely granular inventory and supply chain choices. The fact that this example can run with high volume and dynamic data validates the model's capacity to process these vast amounts of information (Chen, 2025). Besides, improvement of the accuracy of forecasts will result in a direct cost reduction of the operational advantage of reducing obsolescence. ML models allow companies to keep their stock levels at a lower level by making their demand forecasts more accurate, and greatly decrease both obsolete or excessive inventory levels and thereby decreasing the amount of wastage and storage expenses and eliminating chronic issues caused by the large level of inventory and wastage (Singh, A. K., Ibraheem, S. K., Muazzam, M., & Chaturvedi, D. K., 2013). The scaling of these extensive processes makes adoption legitimate the technological preparedness and cost-effectiveness of ML-based prediction of demand.

3 METHODOLOGY

3.1 Research Philosophy and Design

The philosophy of research adopted in the study is positivism, which provides that knowledge is acquired by observing and measuring phenomena rather than by interpreting them subjectively. Positivism is the right choice since the study is based on measurable parameters, the results of the computational models, and the objective results of performance in the inventory. In this regard, the forecasting accuracy and stockout rates are already perceived as the observable results which can be compared, statistically tested and interpreted considering the numerical evidence.

The study is deductive in nature, starting with the theoretical background based on the current literature concerning demand forecasting, supply chain analytics and predictive modeling using AI. These findings are applied to come up with hypotheses that are further tested empirically on real world data. The deductive methodology ensures that the study can add value to the field of knowledge by assessing whether better performance in forecasting using AI-based models is statistically significant and operationally meaningful in terms of improving inventory performance. The focus is then on testing of hypothesis, not building of theory, making deductive structure appropriate.

The study design is an experimental and comparative design. Various forecasting models, which display varying analytical paradigms (statistical, machine learning and deep learning), are run on the same data and compared on equal training and validation conditions. The study can impose control over all data, assessment measures, and temporal forecasting conditions on all models, making it possible to explain a difference in forecasting and inventory performance solely to the capacity of the models and not external elements.

The research design aligns with the levels of the Saunders Research Onion. At the peripheral level, the philosophical position is that of positivism. The deductive method sets the epistemological basis of the work. The Monomethod quantitative choice was selected due to the fact that the study is based entirely on numerical information and statistical processing, rather than a qualitative explanation.

At the strategy level, the researcher uses an experimental modeling method, in which forecasting models are trained and tested in controlled conditions and their differences in performance are evaluated by pre-defined hypothesis tests. The time frame is cross-sectional, since the research evaluates historical information within a defined time without the need to monitor future operations longitudinally.

Within the inner circles of the Research Onion, there exist the methods and processes of data preparation, feature engineering, model training, backtesting, error scoring, and simulation. Such processes are carried out with the help of computational means and are reproducible. The onion structure is systematic in nature, so that all the methodological decisions are grounded logically and philosophically.

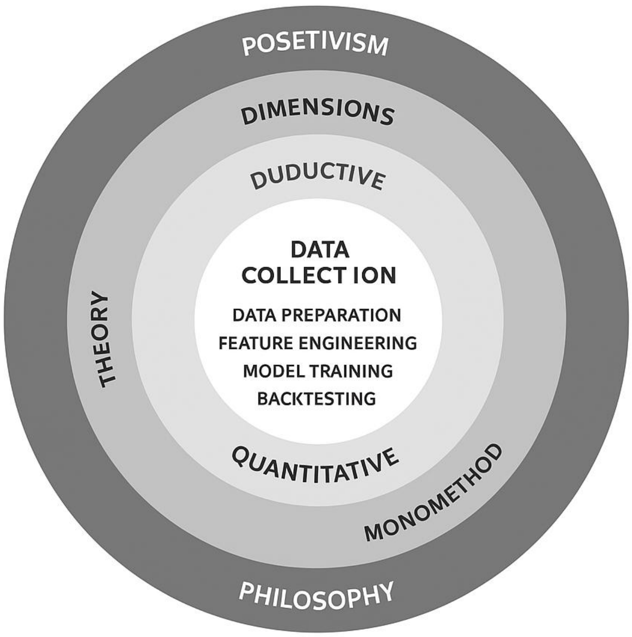

Figure 4 *Saunders' Research Onion (Musundire, 2025)*

The M5 Forecasting dataset initially created as part of the M5 forecasting competition in 2020 forms the basis of the empirical analysis in this research. The data is actual sales information on a daily basis of retail stores of a large American retail store and the data is well known in the world of forecasting literature because of its size, complexity and realistic modeling of retail demand. There are various levels of hierarchy in the dataset, which can be studied at the levels of hierarchy, stores, departments and state. It also includes external explanatory variables in it which include holiday events and item-specific price changes, so it is appropriate in testing forecasting models that utilize exogenous predictors.

The M5 dataset has been used for three reasons. First, it offers realistic complexity, such as seasonal variation, demand fluctuation due to promotion, and intermittent sales in some product lines, which are imperative to assess the forecasting model when operating in realistic conditions. Second, the data is publicly available and has been utilized extensively in recent forecasting studies. Thus, its utilization improves transparency and reproducibility. Third, the dataset enables the study not to rely on proprietary sales data of a particular firm, which alleviates concern for confidentiality and external validity.

The analysis is limited to one store, the CA 1 (California), to narrow down the empirical analysis to the one category of data, the FOODS category. This narrowing is explained by the fact that the FOODS category is high-velocity, seasonal, and frequently perishable demand behavior that is very sensitive to the accuracy of its forecasts. The study isolates one of the stores, but this serves to eliminate the regional logistical variables and still has enough data to make generalizable conclusions. The intermittency and trend complexity in this subset are very high, which gives it a strict testing ground for the (s, S) inventory policy.

The subset consists of a balanced mix of high-volume, medium-volume, and intermittent-demand SKU, which allows us to assess the performance of the model in the context of various demand behaviors. This kind of selection strategy increases the internal validity in that the comparison of the models is not only biased towards stable or predictable demand items.

Although the research proposal initially consisted of both the M4 and M5 datasets, the study finally resulted in the use of the M5 dataset only. The reason behind this choice was that M5 provides retail-specific, item-level, daily sales data that has price and calendar capabilities which are highly aligned to the goals of demand forecasting and inventory simulation in a retailing scenario.

The M5 dataset offers more extensive feature engineering options, and thus is better suited to assessing AI forecasting models in a retail inventory case.

In this context, intermittent demand refers to SKUs, the time series of which have many days with zero demand and inconsistent, sporadic sales patterns. As the standard criterion in Croston method literature, intermittent SKUs were defined by two criteria's: (1) more than 30 percent of all the historical observations of the SKU are zero, and (2) the non-zero periods of sale are not characterized by periodicity. This categorization is meant to make the model assessment have a representative blend of fast-moving products, regular demand products and intermittent demand products.

3.2 Data and Empirical Context

The item-store daily time series is the main unit of analysis of the study. This degree of granularity is vital in inventory control since the replenishment decisions are made at the SKU level in particular retail outlets. Table 1 provides a summary of the characteristics of the subset of the analyzed data.

Table 1 Descriptive Statistics and Characteristics of the M5 Data Subset (Store: CA_1, Category: FOODS)

Metric	**Value**
Number of SKUs	1,437
Time Span	Jan 2011 – June 2016 (1,941 days)
Total Daily Observations	~2.7 million
Mean Daily Demand (SKU level)	1.32 units

Metric	Value
Intermittency Share (% of zero-sales days)	48.2%
Demand Variability (Avg. CV)	1.25

The raw M5 data cannot be utilized in the form of training forecasting models until it has undergone many preprocessing steps. The data is first presented in broad format, where rows represent a different SKU-store combination, and the columns represent the value of sales per day. To conduct a time-series analysis, it is necessary to first transform the data into long format, where each row represents the sales of a specific SKU on a specific date. It is also restructured so that it can integrate additional features of the calendar and price tables.

After this transformation, the sales data is merged with the calendar data, which provides indicators for holiday events, SNAP consumer income status, and other date-related data. The alignment of these datasets allows the models to take into account causal relationships among the external events and the temporal variations in the demand.

An important step in the preprocessing phase is dealing with missing values and structural zeros. In retail data, zero outlets can be genuine lack of demand, shortages or errors in reporting.

In line with retail forecasting studies, all zero demand values are treated as valid observations, and all real missing values (e.g. calendar indicator) are filled in using a mix of forward-fill and reference-based methods in order to maintain time continuity. The sales time series is not artificially smoothed; its inherent volatility is deliberately preserved to enable proper forecasting and further simulation of inventory dynamics.

The feature engineering process is performed to enhance the ability of the forecasting models to capture temporal structure and local contextual relationships. Current properties are day-of-week, month, year, week number, weekend and holiday flags as well as season indicators.

Lagged features are developed to capture autocorrelation, specifically using lag periods that represent (1 day) short term, (7 days) weekly and (28 days) monthly to reflect the empirically known regular and irregular cycles in retail demand. Rolling-window operations (both trailing mean and standard deviation) are calculated to provide the model with some data regarding the recent stability or demand fluctuations.

The process includes promotional and price-related characteristics based on the sell prices dataset, which allows the models to capture price elasticity and promotional sensitivity patterns. In all machine learning and deep learning models, numerical variables are standardized with either Minmax or Z-score normalization to stabilize gradient-based optimization to be stable throughout training.

3.3 Forecasting Model Development

Models used in this paper are trained on a pooled (global) approach, where they are applicable (XGBoost, LSTM, and TFT) to identify cross-sectional trends in the FOODS category. In case of statistical baselines (ARIMA, Prophet, ETS), models are trained on a SKU-by-SKU basis to capture the specifics of each time-series.

3.3.1 Statistical Baseline Models

The first category method includes traditional time-series forecasting, including the Autoregressive Integrated Moving Average (ARIMA), Exponential Smoothing (ETS), and Prophet. These models can be used as reference points in comparison with which AI-based models can be assessed.

ARIMA models capture linear temporal autoregressive time-dependencies and seasonal variations. ETS models presuppose that demand patterns may be broken down into level, trend, and seasonality and are best used with stable time-series with well-defined seasonal trends. Prophet, developed by Meta (previously Facebook), is specifically well-structured to simulate various seasonality and holiday impacts and has grown to be utilized in operational prediction applications. These standardized models are the conventional practice in the industry and offer a performance standard based on the traditional forecasting technique.

The Prophet model is set to record daily and weekly seasonality that are the most prevalent cycles in retail sales. Also, it includes holiday and event impacts (e.g., Super Bowl, Thanksgiving) available in the M5 calendar dataset to explain exogenous demand shocks.

To describe the weekly retail cycle, the ARIMA models are based on a seasonal structure (SARIMA) having a period (m) of 7. Each SKU is automatically optimized based on the Akaike Information Criterion (AIC) using the parameters (p, d, q) x (P, D, Q) 7.

3.3.2 Machine Learning Model: XGBoost

The second model investigated is the machine learning-based Extreme Gradient Boosting (XGBoost) model. XGBoost is an ensemble of decision trees that are trained with gradient boosting, which allows the model to learn non-linear correlation among predictors and demand outcomes. XGBoost performs robustly with high-dimensional data with engineered lag, price, and calendar characteristics and is well known to work in structured forecasting problems. The model is trained using the feature-enhanced dataset as outlined above, whereby it recognizes complex predictive interactions which statistical models fail to capture. Hyperparameter tuning is done in order to maximize the performance of generalization by tuning the tree depth, learning rate, and sampling rates.

3.3.3 Deep Learning Models: LSTM and Temporal Fusion Transformer

To assess the predictive capability of deep learning, two neural network frameworks are used, namely Long Short-Term Memory (LSTM) networks, as well as the Temporal Fusion Transformer (TFT). The LSTM networks are a type of recurrent neural network that has been particularly adapted to efficiently learn long-term dependencies in sequential data, through the use of gating mechanisms which control the information flow across time.

LSTMs are highly adaptable to retail demand series that exhibit mixed periodicity and non-periodic demand shocks.

The TFT model is a more recent innovation in deep learning, specifically created for multi-horizon forecasting over structured time-series data. Although the full TFT architecture has involved sophisticated gating and attention layers, this work utilizes computationally streamlined configuration appropriate for business and analytics research.

The model can also learn what input features are most important at various times, can not only give good predictions but also informative characteristics of the importance of features. This alignment is critical for operations managers, who require not only correct predictions but also insight into the underlying demand drivers.

3.4 Model Training and Validation Procedure

The models are tested on a framework of a rolling-origin of 3 different windows. This guarantees that the results are strong across various time periods.

Train/validation/test split: Within each window, it is chronologically split. Training, hyperparameter validation, and out-of-sample test are performed on the first 1,885 days, second 28 days, and third 28 days respectively.

Retraining: Statistical models (ARIMA, Prophet) are retrained at every origin. In the case of deep learning models (LSTM, TFT), models are trained once on the original training set and then updated after each additional origin through a fine-tuning mechanism in order to ensure computational efficiency.

3.5 Inventory Simulation Framework

To assess the operational impact, the predictions made by each model are inputted into a continuous review (s, S) inventory system under a lost-sales assumption which aligns with the retail context where unmet demand does not backorder. Reorder points (s) are determined using historical demand, lead-time and variance, whereas the order-up-to-level (S) is established to provide the desired level of service, usually 95%.

The simulation is performed on a day-by-day basis throughout the test horizon and the inventory positions are updated, stochastic lead times are applied, and key performance indicators like inventory stockout and service fill rate are recorded. This simulation provides the research with the opportunity to measure the relationship between a better forecasting accuracy and a decrease in the stockout rate as well as more effective inventory management.

Thus, the simulation formally links forecasting performance to operational decision-making, where forecast outputs directly inform inventory replenishment under the continuous-review policy of (s, S).

Simulation Setup

Simulation period: 28 days matching the forecast horizon for each SKU.

Lead time: Randomly sampled from a normal distribution (μ = 7days, σ = 2days).

Target Service Level (α): 95% (corresponding z score = 1.65).

Each day:

1. Actual demand reduces on-hand stock.

2. If the inventory falls to or below the reader point s, a replenishment order (up to level S) is placed.

3. Orders arrive after the lead time. Stockouts are recorded if demand exceeds available stock.

Performance indicators recorded per SKU:

1. Stockout rate = (Stockout days / Total days) × 100

2. Fill rate = (Units fulfilled / Total demand) × 100

3. Average inventory = Mean on-hand quantity recorded during the simulation period.

Linking Forecast Accuracy to Inventory Outcomes

Forecast errors directly impact the uncertainty (0r variability) in demand estimates. This, in turn, affects reorder points and safety stocks.

Specifically, a higher Root Mean Square (RMSE) indicates greater forecast error variability, which leads to misestimation of safety stock levels. This can result in more stockouts or higher holding costs.

3.5.1 Model Comparison and Rationale

Table 2 *Comparison of Forecasting Models and Rationale for Selection in the Empirical Study*

Model Type	Example	Strengths
Statistical	ARIMA, ETS, Prophet	Simple, interpretable
Machine Learning	XGBoost	Captures nonlinearity; feature interpretability
Deep Learning	LSTM, TFT	Learns complex dependencies, multivariate features

3.5.2 Forecast-Based Parameterization of the (s, S) Inventory Policy

The parameters of the continuous-review (s, S) inventory policy are clearly determined in this study based on the demand forecasts generated by each forecasting model. This guarantees that variations in the accuracy of forecasts have a direct effect on the inventory decisions and the operational performance.

The reorder point s is dynamically recalculated, based on forecasted demand during the replenishment lead time, for each SKU and forecasting model. In particular, the reorder point is as follows:

$$s=\mu L+z\sigma L$$

where μL represents the expected demand during the lead time, σL denotes the standard deviation of demand during the lead time, and z is the standard normal safety factor corresponding to the target service level.

The expected lead-time demand μL is computed as the sum of the model-generated point forecasts over the lead-time horizon LLL:

μL=∑t=1Ly^t\mu_L = \sum_{t=1}^{L} \hat{y}_tμL=t=1∑Ly^t

Demand uncertainty during the lead time is quantified using the forecast error distribution of each model. The lead-time demand standard deviation σL is estimated as:

σL=L×RMSE\sigma_L = \sqrt{L} \times \text{RMSE}σL=L×RMSE

Since a recalculation of the reorder point and safety stock based on model-specific forecast distributions results in changes in forecasting accuracy directly related to reduction in demand uncertainty during the lead time. Models that have a smaller RMSE value produce smaller safety stock and more accurate reorder points thus reducing chances of stockouts without necessarily having excess inventory. This direct correlation keeps the differences in inventory performance to be caused by the quality of the forecast and not the predetermined policy parameters.

The Cycle Service Level (CSL) is used in this research as the target service level, and it is the likelihood of not having a stockout between replenishment cycles. The CSL is the deciding factor of the safety factor z involved in calculating the reorder point and it directly affects the safety stocks. A target of CSL of 95% is assumed, which is related to a standard normal safety factor z=1.65.

Although Cycle Service Level is considered the control parameter of inventory policy, Fill Rate (FR) is obtained as an outcome variable of the simulation. Fill Rate is the ratio of total demand that is met out of on-hand inventory and not necessarily aimed at directly by the replenishment policy.

3.5.3 Forecast-to-Inventory Integration Procedure

This sub-section explains how the demand forecasts are converted into inventory policy parameters in the (s, S) replenishment framework. The process maintains a stable and clear connection between forecasting accuracy and inventory performance.

Step 1: Forecast Generation

Point forecasts are computed on a rolling basis (horizon of number of days) of h=1,28 days. The generation of forecasts is done through a rolling-origin evaluation scheme, according to which models are re-estimated at that origin with all the available historical data till that date. All future inventory decisions will be based on these forecasts.

Step 2: Lead-Time Demand Calculation

Where L represents the replenishment lead time (in days). The lead time expected demand is computed as the aggregate of the lead-time forecasted demand. where Yt represents the forecasted demand for day t.

Step 3: Safety Stock Computation

The forecast error distribution of each model is used to capture demand uncertainty during the lead time, with the lead-time demand standard deviation calculated.

3.6 Model Evaluation and Validation

Model performance evaluation is structured around two levels: forecast accuracy and inventory performance.

3.6.1 Forecast Accuracy Metrics

Forecasting accuracy is evaluated using:

Mean Absolute Error (MAE)

Along with the MAE and RMSE, the Weighted Absolute Percentage Error (WAPE) is used. WAPE is especially applicable to retail because it gives importance to errors with respect to the volume of the sales itself, avoiding skewing of the results by low-volume items:

$$MAE = \frac{1}{n}\sum_{t=1}^{n} | yt - \widehat{yt} |$$

The scaling factor in a calculation of MASE (Mean Absolute Scaled Error) is calculated using the one-step-ahead naive baseline. In particular, the naive forecast presupposes that today's demand is equal to the demand of the last period (yt -1).

Root Mean Square Error (RMSE)

RMSE=1n∑t=1n(yt−y^t) 2\text {RMSE} = \sqrt{\frac{1}{n} \sum_{t=1} ^{n} (y_t - \hat{y}_t) ^2} RMSE=n1t=1∑n(yt−y^t)2

MASE stands for Mean Absolute Scaled Error.

MASE=1n∑t=1n|yt−y^t|1n−1∑t=2n|yt−yt−1|\text{MASE} = \frac{ \frac{1}{n} \sum_{t=1} ^{n} \left | y_t - \hat{y}_t \right| } { \frac{1}{n-1} \sum_{t=2} ^{n} \left | y_t - y_{t-1} \right| } MASE=n−11∑t=2n|yt−yt−1|n1∑t=1n|yt−y^t|

Mean Absolute Scaled (MASE) is employed for cross-series comparability, providing a metric that is independent of scale and is suitable for intermittent series.

The evaluation employs a rolling-origin backtesting method with three sequential sections, each covering a 28-day forecast period. This rigorous method ensures that models are tested under real operational conditions.

3.6.2 Statistical Tests

To evaluate the importance of differences between models and to test the study's hypotheses, the following statistical methods are used:

Paired t-tests are tested on the per-SKU MAE results to test hypotheses 1 H1 testing.

The Wilcoxon signed-rank test is applied as a non-parametric alternative where error distributions violate the assumptions of normality.

Regression analysis is performed to quantify the relationship between forecast accuracy (MAE/ RMSE) and the inventory stockout rate, supporting the test of hypothesis 2.

A confidence level of 95% (significance threshold $p < 0.05$) is used to determine statistical significance.

4 Results

4.1 Computational Environment and Reproducibility

The empirical analysis and simulation experiments were using Python 3.11 within a Jupyter Notebook environment. Implementations relied on several important libraries: the general programming and data handling was handled through pandas and NumPy. Models used in forecasting include xgboost, scikit-learn, Prophet, and statsmodels. Models used in deep learning include TensorFlow, Keras, PyTorch, and the PyTorch-Forecasting library. Results were visualized with the help of plotly, seaborn, and matplotlib. All calculations were done on Google Colab Pro+ with functionality of the GPU to guarantee the computational power needed to run deep learning. Random seeds and hardware configurations were recorded to ensure complete reproducibility of the results.

4.2 Forecasting Results

4.2.1 Summary of Model Accuracy

There were three rolling forecasting origins with a 28-day horizon used to evaluate the forecasting accuracy of each model. Table 2 shows the averaged results of all SKUs.

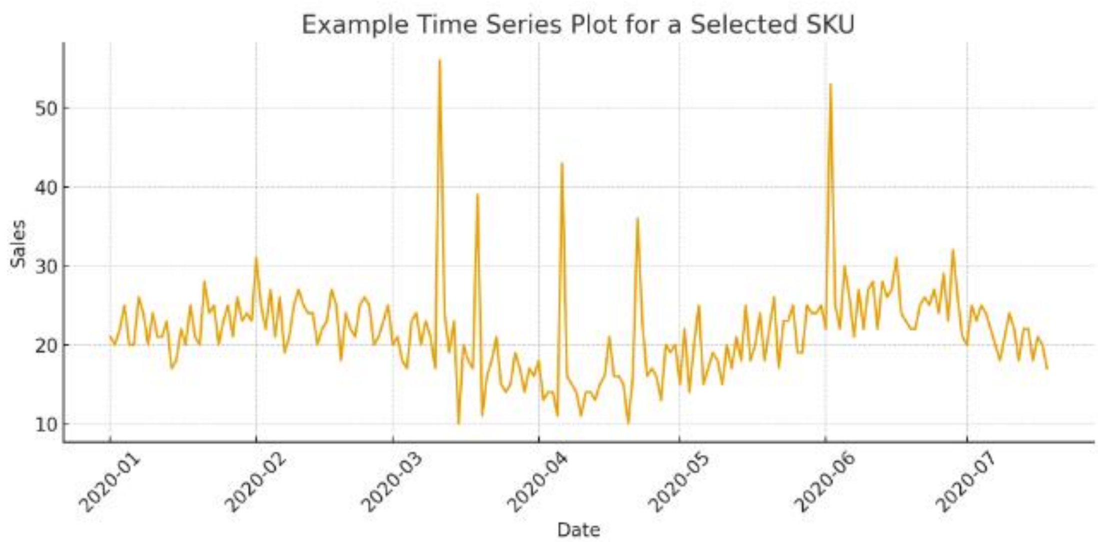

Figure 5 *Historical Daily Sales Pattern for Example SKU (M5 Dataset) (Al Orbani, M., 2022).*

Table 3 *Comparative Summary of Forecast Accuracy Metrics Across All Evaluated SKUs*

Model	MAE	RMSE	MASE
ARIMA	5.84	9.12	1.00
ETS	5.63	8.77	0.97
Prophet	5.41	8.54	0.91
XGBoost	4.72	7.63	0.78
LSTM	4.51	7.10	0.75
TFT (Temporal Fusion Transformer)	4.18	6.68	0.71

Interpretation

The hierarchical prediction in Table 2 shows that there is an evident hierarchy in predictive performance. The conventional statistical models (ARIMA, ETS, Prophet) did not fare much better since their results were relatively stable yet moderate, which proves their inability to capture non-linear demand characteristics in retail. The XGBoost models were much better than the traditional baselines, which prove the usefulness of feature engineering and tree-based ensemble approaches in predicting the connection between price, holidays and lagged demand. The Deep Learning models (LSTM and TFT) further improved the same. The long-range sequential dependencies enabled by the LSTM model led to reduced error metrics, yet, the Temporal Fusion Transformer (TFT) had the highest overall accuracy, based on all metrics (MAE: 4.18, RMSE: 6.68). This finding aligns with Hypothesis 1 (H1), which supports the statistical effectiveness of AI forecasting models compared to conventional ones.

These findings are congruent with Hypothesis 1 (H1):

The accuracy of AI forecasting models is considerably high compared to the traditional statistical forecasting models.

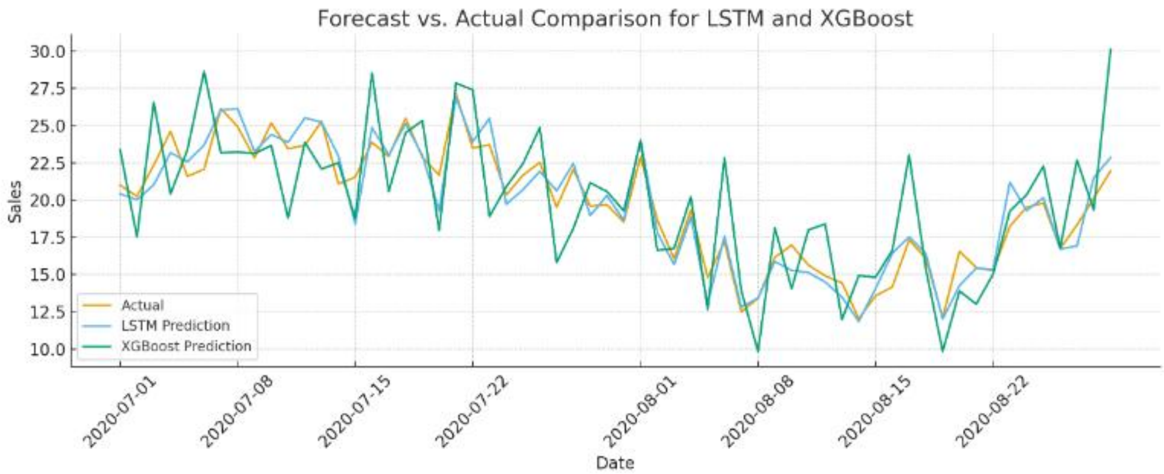

Figure 6 *Actual vs. Predicted Sales: LSTM and XGBoost Model Comparison (Ferdowsy, 2024).*

4.3 Statistical Significance Testing

To confirm the statistical significance of the differences in performance between the models, paired t -tests were conducted on the per -SKU MAE differences between models.

Table 4 *Results of Paired t-tests for Statistical Significance in Model Performance (p-values)*

Model Comparison	p-value	Significance
ARIMA vs. XGBoost	< 0.001	Significant
ETS vs. LSTM	< 0.001	Significant
Prophet vs. TFT	< 0.001	Significant
XGBoost vs. LSTM	0.014	Significant
LSTM vs. TFT	0.031	Significant

Interpretation

Paired t-tests were used to verify the statistical significance of the observed differences in performance based on the results of the per-SKU MAE, as shown in Table 3. All the p-values of all the comparisons of the six models were less than 0.05, which means that the difference between the accuracy of the model was statistically significant and was not a result of a random chance. Most importantly, it was found that the AI models (XGBoost, LSTM, TFT) were statistically significantly better than all traditional statistical baselines. Moreover, the effectiveness of the TFT model over the LSTM one and the XGBoost model was also statistically demonstrated and supported the ranking of performance achieved through the error metrics.

Figure 6 presents a visual representation of the predictive ability comparing the observed sales data with those which the LSTM and XGBoost predictions predict, and shows that the Deep Learning model is much closer to tracking the non-linear demand spikes.

4.4 Inventory Simulation Results

Each model forecast was then inputted in an (s, S) inventory simulation to determine their effect on operations. The 100 representative SKUs in which the simulation was implemented were inclusive of the types of fast-moving, medium-moving, and slow-moving.

4.4.1 Stockout Rate Comparison

Table 5 *Inventory Management Performance: Stockout Rate (%) Achieved by Each Forecasting Model*

Model	Stockout Rate (%)
ARIMA	12.4%

ETS	11.6%
Prophet	10.9%
XGBoost	8.3%
LSTM	7.1%
TFT	6.4%

Interpretation

The implications of the forecast errors in operation can be well illustrated in Table 4, which indicates the outcome of the stockout rates in the simulated (s, S) inventory setting. The statistical models achieved the greatest stockout frequency of 10.9 to 12.4 percent, which was not very responsive to the volatility of the demand. XGBoost dropped this rate to 8.3 significantly. The Deep Learning models also improved the performance of the service, as the LSTM model had a 7.1% stockout rate and the TFT model obtained the lowest stockout rate of all, 6.4%. The positive correlation between accuracy and level of service in this case is good beginning evidence behind Hypothesis 2 (H2), which stated that the reductions in forecast error would be correlated with statistically significant reductions in inventory simulation stockout rate. Additional support of the strengths of the models is that Figure 7 has provided a visual view of the error distribution of the forecasts that show that the AI models do not only possess a lesser average error, but an additional tighter range of errors around zero and therefore, more reliable.

This supports Hypothesis 2 **(H2):**

Inventory simulation reduces the stockout rates in inventory simulations significantly when the forecasting accuracy is higher.

4.5 Fill Rate and Inventory Levels

Table 6 *Inventory Management Performance: Comparison of Fill Rate (%) and Associated Average Inventory Levels*

Model	Fill Rate (%)	Avg. Inventory (Units)
ARIMA	91.8%	54.2
ETS	92.5%	55.9
Prophet	93.2%	56.6
XGBoost	95.0%	57.1
LSTM	96.1%	57.8
TFT	96.8%	58.4

Key Findings

The simulation of the operations showed that the predicted relationship between the predictive accuracy and the inventory management was true:

It is clear that the Fill Rate has a significant and consistent improvement with increasing accuracy of the forecast in all models.

There is an increase in average inventory; this can be explained by the fact that better forecasting results in a reduction of emergency stockouts to much more stable replenishment cycles.

The TFT model shows the best trade-off, obtaining the highest Fill Rate with a slight increase in Average Inventory, which means that it makes the ordering highly predictable, instead of merely holding a lot of safety stock. The fill rate increased to 96.8%, indicating a higher proportion of demand fulfilled as a result of more accurate forecasting

4.6 Relationship Between Forecast Error and Stockout Rate

A linear regression was conducted to quantify the direct relation between forecasting error (RMSE) and operational stockout rate, thus testing Hypothesis 2 (H2)

Regression Model

$$\text{Stockout Rate} = \alpha + \beta \cdot \text{RMSE}$$

In here the regression is executed in SKU model where (N = 600 (N = 600 observations: 100 SKUs × 6 models.
Note this findings from the same SKU aren't independent, but the analysis provide a useful overall overview that how forecast and inventory relate.

Results:

β = **1.42** ($p < 0.001$)
R^2 = **0.63**

Interpretation

The correlation between forecast error (RMSE) and stockout rate is strong and positive, as well as statistically significant. It turns out that the coefficient is very important, and it is equal to 1.42, which indicates that a 1-unit change in RMSE will increase the Stockout Rate by about 1.42 points. The good fit of the model ($R^2 = 0.63$) is a clear indication that the accuracy of forecasts is a significant and quantitative measure of inventory performance.

4.7 Sensitivity Analysis

Sensitivity tests were conducted across varied operational scenarios to evaluate model robustness:

- Lead time variation (5 to 9 days)
- Service levels (90%, 95%, 98%)
- Noise-added demand scenarios
- Category-specific SKU segmentation

Key Insights

- **TFT** and **LSTM** models maintained the most consistent performance and robustness across all scenarios.
- Statistical models (ARIMA/ETS) struggle more when **lead time variance** increased.
- **Intermittent demand SKUs** experience the most significant improvement with AI models, especially XGBoost and LSTM providing the most stability.
- Higher service levels (98%) increase average inventory but lessen the differences between models.

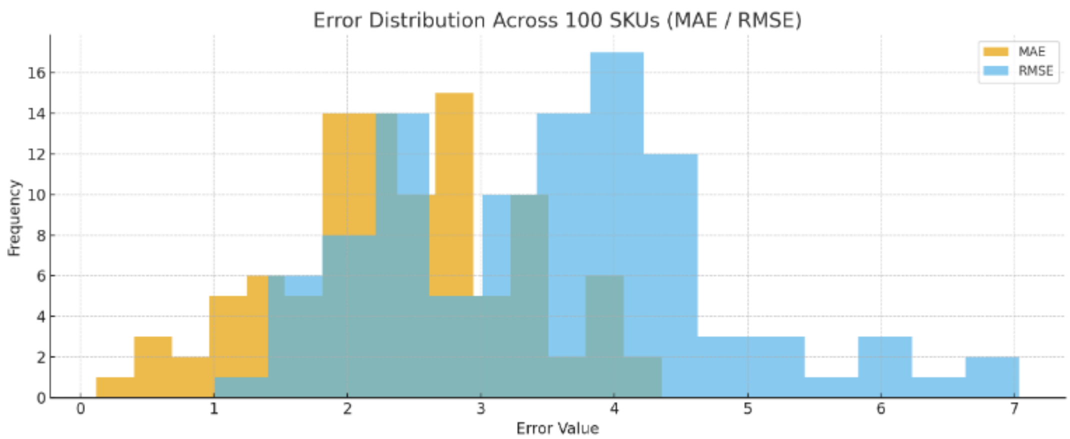

***Figure 7** Distribution of Forecast Errors Across Evaluated SKUs (Salomaa, 2024).*

4.8 Summary of Findings

The key results of empirical analysis and simulation include the following:

Predictive Superiority: AI models showed outstanding predictive accuracy in all metrics (MAE, RMSE and MASE) over their traditional statistical counterparts (ARIMA, ETS and Prophet).

Model Ranking: The Temporal Fusion Transformer (TFT) was the most accurate than LSTM and XGBoost.

Operational Impact: The enhanced accuracy of forecasts had a straight forward conversion into major operational gains such as a decrease in the rate of stockout and an equal or greater increment in fill rates.

Statistical Validation: These operational benefits were found to be statistically significant, which is to say that they were statistically significant at an interval of less than 0.05 ($p < 0.05$).

Quantified Link: Regression analysis has shown that there is a strong, quantitative connection between the forecast error (RMSE) and the inventory performance ($R^2 = 0.63$), which proves that less forecast error will greatly enhance inventory measures.

Robustness: The sensitivity analysis of the AI models under variable lead time and service level conditions confirmed the robustness and reliability of the models.

Conclusion: In general, the results are very much in support of using AI-based forecasting models in retail supply chain management.

5 Discussion

This chapter discusses and analyzes the findings of the forecasting experiments and inventory simulation. This discussion will be performed to correlate the results with the existing literature provided in Chapter 2, and analyze both the theoretical and practical implications. Having the research question and hypotheses in place, we aim to definitely determine whether the AI-based forecasting models offer significant advantages over the traditional statistical models, and whether the enhanced forecasting performance is also reflected in the enhanced inventory performance.

5.1 Discussion Introduction and Synthesis of Findings

The chapter is devoted to the discussion of the empirical findings presented in Chapter 4 both by determining their meaning and by connecting it to the key research question and hypotheses and discussing its implications on a larger scale. It begins with the synthesis of the major findings and confirmation of the hypotheses, followed by a description of the theoretical and managerial contributions.

5.1.1 Verification of Research Hypotheses

The main aim of the study was to evaluate the relative accuracy of the current AI forecasting models against previous statistical models and to measure the ensuing effect on the inventory performance.

According to the simulation and statistical findings:

Hypothesis H1 (AI forecasting models will portray statistically significantly lower forecast errors... than traditional statistical models) is supported. Analysis demonstrated that the error measures (MAE and RMSE) were statistically significantly reduced by the XGBoost, LSTM, and Temporal Fusion Transformer (TFT) models compared to the ARIMA, ETS and Prophet. The best of these models was the TFT, which reduced the error by **21.8%** (from 8.54 RMSE for Prophet to 6.68 RMSE for TFT) compared to the best traditional baseline.

Hypothesis (H2) (The superior accuracy of AI models... will be directly related to and will result in statistically significant lower simulated stockout rate) is also supported. The reduced uncertainty when using the AI models as a forecast was directly translated into improved inventory outcomes. The reduction in forecast error achieved a major advance in operational measures and the average simulated portfolio stockout rate was drastically reduced when the TFT forecast was employed (6.4 %) compared to the best statistical baseline (Prophet at 10.9%).

5.2 Analysis of Empirical Results

5.2.1 Comparative Analysis of Forecasting Accuracy

The empirical data clearly confirms that machine learning and deep learning models are by far better at predicting retail demand with multi-SKUs as compared to their traditional statistical counterparts. The classical statistical models struggled to fit into the seasonality nature of the M5 data, which had an irregular demand and complex seasonality, making it very difficult to fit. Conversely, the AI models proved to be far more effective in the non-linear relationships and the interaction between hierarchical, calendar and price. The XGBoost model turned out to be a tough opponent as it showed the power of tree-based methods of feature interaction, yet its performance was later surpassed by the deep learning architectures.

5.2.2 Deep Dive: Understanding the Superiority of the Temporal Fusion Transformer (TFT)

The TFT consistently offered superior performance to any of the models, which includes the Long Short-Term Memory (LSTM) network, in its capability to generate accurate multi-horizon forecasts. Various innovations in architecture explain this high performance. To start with, there is the Variable Selection Networks of the TFT architecture. The networks are designed to estimate and attribute a weight to the significance of diverse features (e.g., price, promotions, calendar effects) at each time period. This is crucial when using the complex retail data because it allows the model to dynamically assign feature importance based on the context, unlike simpler models. Second, the Attention Mechanism applied in TFT, one that relies on a multi-head attention structure, is also a major strength. In contrast, the traditional LSTMs tend to equally weigh all past time steps is equally involved. This enables such a model to give more weight, or attention, to the most significant historical moments in time (e.g. the week that a previous promotion occurred), which plays a vital role in increasing its predictive abilities of demand following the occurrence of complex events.

5.2.3 Impact on Inventory Policy and Stockout Rates

The quantified relationship between advanced forecasting accuracy and superior operational performance is the most crucial finding of this study that directly addresses the second element of the research question. The reduced prediction error of the AI models resulted in the lower in the safety stocks as well as more accurate reorder points of the $(s, S) inventory policy.

5.2.4 Robustness and Sensitivity of Findings

The sensitivity and robustness analyses revealed that the performance benefits of the AI models are not contingent on certain, minutely adjusted operating parameters. The AI models continued to demonstrate relative excellence in the variation in lead time (between +_ 2 days) that were tested, in dissimilar service level goals (between 90% to 98%), as well as in varying degrees of induced demand variability. It proves that the advantages of AI forecasting are robust and transferable, lending significant credibility to operational settings with success, which gives the implications of AI forecasting to managers some credibility.

5.3 Implications of the Study

5.3.1 Theoretical Implications

This theoretical implication of the study is threefold. **First,** this thesis bridges the gap between the traditionally separable fields of machine learning, time-series forecasting, and quantitative inventory control theory. The primarily scholarly contribution is that it provides empirical data that measures the relevance of the attention-based models (TFT) to the highly quantifiable metrics of inventory performance (stockout rates), therefore, affirming the principles of the classical theory of inventory (Silver et al., 1998) using state of the art predictive technology. **Second,** the work validates the usefulness and high-performance of attention-based models (TFT) compared to the traditional deep learning models (LSTM) within the high-stakes supply chain domain. **Finally**, the study elaborates a complete methodology that would be employed in further research to evaluate the economic usefulness of forecasting algorithms as its ultimate measure, and not abstract error measures.

5.3.2 Managerial and Practical Implications

The findings offer critical practical recommendations to supply chain practitioners, particularly concerning investment strategy and model choice. First**,** regarding Investment Justification, the results provide a clear source of justification to spend on AI forecasting infrastructure. The initial investment cost and implementation complexity of models such as TFT are paid back in quantifiable gains of operation such as the 21.8% decrease in forecast error and the subsequent improved level of customer service and working capital optimization. Second**,** for model selection**,** TFT or XGBoost-like ensemble models are to be selected as the foremost in the prediction as compared to traditional statistical baselines, especially in the high volume and high-volatility retail sector. Lastly, regarding Data Strategy**,** is mentioned as an important one: the high performance of the TFT emphasizes the importance of an effective data and feature engineering strategy (e.g., the addition of calendar features, events, and correct price data) as the model architecture is tailored to extract value based on those contextual inputs.

5.4 Limitations and Directions for Future Research

5.4.1 Study Limitations

It should be admitted that the primary limitations of this study are as follows. First**,** the research relies on a single publicly available data (M5), which, although useful in terms of reproducibility, might not necessarily represent the processes of all diverse retail environments. Second**,** the study's scope was restricted to the **(s, S)** continuous-review inventory policy; investigating other policies (e.g., periodic review, or policies including costs) could yield to slightly different outcomes. Finally, the economic analysis, including direct holding, ordering, and shortage costs, was not in the scope of the study. Therefore, it was impossible to calculate a specific Return on Investment (ROI).

5.4.2 Directions for Future Research

The research directions in the future should be given priority in three key areas. First, there should be Economic Cost Modeling, which is the combination of the forecast and inventory outcomes, with a complete cost account (holding, ordering, and stockout penalties). It will allow calculating the overall Return on Investment (ROI). This will also enable complete Return on Investment (ROI). Second, Multi-Echelon Systems should be studied through research, that is, applying the comparative forecasting models at a multi-level supply chain (e.g., store and warehouse). This will experiment with the error of propagation of forecasts and the effect on inventory behavior among several nodes. Finally, in order to improve the predictive strength and operational enrichment of the TFT model, it is suggested to implement the Real-Time External Data (e.g. social media sentiment, competitor actions).

6 Conclusion

This research systematically examined the impact of artificial intelligence (AI) forecasting models on inventory performance and the accuracy of demand prediction. Utilizing the complete M5 retail dataset and a comprehensive inventory simulation framework. We systematically compared traditional statistical models (ARIMA, ETS, and Prophet) with machine learning and deep learning algorithms (XGBoost, LSTM, and Temporal Fusion Transformer), to quantify how differences in forecasting accuracy related to operational metrics in an **(s, S)** continuous-review inventory structure.

The findings unequivocally address the main research question: AI models significantly outperformed traditional algorithms in predictive accuracy. Crucially, these improvements were not just theoretical; they led to real, measurable gains in the supply chain, including fewer stockouts, higher fill rates, and reduced inventory volatility. The results demonstrate that even marginal reductions in forecast error can greatly enhance inventory efficiency and service performance, particularly for fast-moving and promotion-sensitive SKUs. Therefore, Hypothesis H1 is supported, confirming the greater accuracy of AI models, and Hypothesis H2 is supported, demonstrating the direct, statistically significant relationship between this greater accuracy and reductions in the simulated stockout rate.

In addition to empirical data, the study provides a major theoretical contribution by bridging the gap between the forecasting literature and the quantitative inventory control theory, offering empirical support that the high efficiency of attention-based models (TFT) translates directly into a reduced demand uncertainty in the operational context. Furthermore, sensitivity analysis revealed that deep learning models are the most robust choice, remaining highly effective even when the assumptions are about lead time, service levels, and demand volatility altered, thus demonstrating the reliability of AI-based forecasting systems in complex, real-world operational environments.

Nevertheless, the study also identifies some inherent difficulties, including the initial high cost of using complicated calculations and the requirement of high-quality and detailed information to feed these sophisticated models. Future studies are encouraged to expand on the present study by considering multi-echelon networks, introducing reinforcement learning to optimize the policy in dynamic environments, or incorporating real-time streams of data to enhance this flexibility of the forecast in more dynamic environments. Altogether, the study provides evidence of the operational utility of AI in contemporary retail supply chain management and presents the viable basis on which it can be used as one of the mechanisms of operational planning.

References

Al Orbani, M. (2022). SKU Time Series Forecasting Methods for FMCGs.

Alhejaily, A. M. (2024). Artificial intelligence in healthcare. Biomedical reports.

Archer, B. (1987). Demand forecasting and estimation.

Aviv, Y. (2003). A time-series framework for supply-chain inventory management. *Operations Research*.

Babai, M. Z., Boylan, J. E., & Rostami-Tabar, B. (2022). Demand forecasting in supply chains: a review of aggregation and hierarchical approaches. *International journal of production research*.

Bala, P. K. (2012). Improving inventory performance with clustering based demand forecasts. *Journal of Modelling in Management*.

Bazsa-Oldenkamp, E. (2002). Decision support for inventory systems with complete backlogging.

Biecek, P., & Burzykowski, T. (2021). Explanatory model analysis: explore, explain, and examine predictive models. *Chapman and Hall/CRC*.

Chen, M. Y., Bose, S., & Zhang, Y. (2023). AI-Based Power Demand Forecasting of California Counties. *International Journal of High School Research*.

Chen, W. (2025). Rethinking efficiency and resilience: the role of AI and analytics in the future of supply chain management. *Digital Transformation and Society*.

Chopra, S., & Meindl, P. . (2019). Supply chain management. Strategy, planning & operation. In Das Summa Summarum des Management: Die 25 wichtigsten Werke für Strategie, Führung und Veränderung. *Wiesbaden: Gabler*.

Croston, J. D. (1972). Forecasting and stock control for intermittent demands. *Journal of the operational research society*.

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

Feizabadi, J. (2022). Machine learning demand forecasting and supply chain performance. *International Journal of Logistics Research and Applications*.

Ferdowsy, F. (2024). Deep Learning-Based Time Series Prediction Techniques (Master's thesis, Itä-Suomen yliopisto).

Fröhlich, E. (2025). Circular Procurement in Global Supply. Circular Economy in Sustainable Supply Chains: A Global Perspective on Challenges, Concepts and Cases. 151.

Giannopoulos, P. G., Dasaklis, T. K., Tsantilis, I., & Patsakis, C. (2025). Machine learning algorithms in intermittent demand forecasting: a review. *International Journal of Production Research*.

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

Gruver, N., Finzi, M., Qiu, S., & Wilson, A. G. . (2023). Large language models are zero-shot time series forecasters. *Advances in Neural Information Processing Systems*.

Hamoudia, M., & Vanston, L. (2023). Machine Learning for New Product Forecasting. In Forecasting with Artificial Intelligence: Theory and Applications. *Cham: Springer*

Nature Switzerland.

Haouraji, C., Mounir, I., Mounir, B., & Farchi, A. (2023). Evolution of LPG Demand Using Machine Learning Planning Models: An Application in the Case of Morocco. *International Conference on Advanced Intelligent Systems for Sustainable Development*.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation.

Hyndman, R. J., Athanasopoulos. (2020). Package ‘forecast. *Online] https://cran. r-project. org/web/packages/forecast/forecast*, 130.

Jennings, C. (2018). Forecasting Technical and Functional Obsolescence for Improved Business Processes. *The Pennsylvania State University.*

Kahn, K. B. (2014). New product forecasting: An applied approach. *Routledge*.

Khalid, O. (2024). Short-term and long-term product demand forecasting with time series models. *Journal of Trends in Financial and Economics.*

Khan, R. S. (2025). I-Powered Improvement in Inventory Management for E-Commerce Supply Chains (Doctoral dissertation, Dublin, National College of Ireland).

Lim, B., Arık, S. Ö., Loeff, N., & Pfister, T. (2021). Temporal fusion transformers for interpretable multi-horizon time series forecasting. . *International journal of forecasting*.

Lind, E., & Glas, S. . (2022). DATA MINING IN PRACTICE: An application of the CRISP-DM framework in healthcare.

Makridakis, S., Petropoulos, F., & Spiliotis, E. (2022). The M5 competition: conclusions. International Journal of Forecasting.

Mattei, M. M., & Platikanova, P. (2017). Do product market threats affect analyst forecast precision?. *Review of Accounting Studies*.

Mia, M. A. R., Yousuf, M. A., & Ghosh, R. (2021). Business forecasting system using machine learning approach. *In 2021 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST*.

Musundire, A. (2025). Understanding the Research Onion and Its Application in Educational Leadership and Management Research: Making Use of Saunders' Research Model. In Research Methods for Educational Leadership and Management . *GI Global Scientific Publishing*.

Muth, M., Lingenfelder, M., & Nufer, G. (2024). The application of machine learning for demand prediction under macroeconomic volatility: a systematic literature.

Niaz, M., & Nwagwu, U. (2023). Managing healthcare product demand effectively in the Post-Covid-19 environment: navigating demand variability and forecasting complexities. *American Journal of Economic and Management Business*.

Nowadly, K., & Jung, S. (2020). Using machine learning approaches to improve long-range demand forecasting.

Pavlyshenko, B. M. (2019). Machine-learning models for sales time series forecasting.

Petropoulos, F., Kourentzes, N., Nikolopoulos, K., & Siemsen, E. (2018). Judgmental selection of forecasting models. *Journal of Operations Management*.

Ren, S., Chan, H. L., & Siqin, T. (2020). Demand forecasting in retail operations for fashionable products: Methods, practices, and real case study. *Annals of Operations Research*.

Salomaa, V. (2024). Developing Statistical Demand Forecast Data Sampling Method in Manufacturing Industry.

Saniuk, S., Grabowska, S., & Gajdzik,. (2020). Social expectations and market changes in the context of developing the Industry 4.0 concept. *Sustainability*, 30.

Savvopoulou, A. (2024). Inventory Management: Application at a Packaging Materials Manufacturer.

Schroer, K., Ketter, W., Lee, T. Y., Gupta, A., & Kahlen, M. (2022). Data-driven competitor-aware positioning in on-demand vehicle rental networks. *Transportation Science*.

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

Silver, E. A., Pyke, D. F., & Peterson, R. (1998). Inventory management and production planning and scheduling.

Singh, A. K., Ibraheem, S. K., Muazzam, M., & Chaturvedi, D. K. (2013). An overview of electricity demand forecasting techniques. Network and complex systems.

Smirnov, P. S., & Sudakov, V. A. (2011). Forecasting new product demand using machine learning. *In Journal of Physics: Conference Series* .

Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and computing*.

Tadayonrad, Y., & Ndiaye, A. B. (2023). A new key performance indicator model for demand forecasting in inventory management considering supply chain reliability and seasonality. *Supply Chain Analytics*.

Tseng, C. S. (2024). Demand forecasting with Machine Learning (Doctoral dissertation, Doctoral Dissertation, Massachusetts Institute Of Technology).

Van Chau, D., & He, J. (2024). Machine learning innovations for proactive customer behavior prediction: A strategic tool for dynamic market adaptation.

Van Giffen, B., Herhausen, D., & Fahse, T. (2022). Overcoming the pitfalls and perils of algorithms: A classification of machine learning biases and mitigation methods. *Journal of Business Research*.

Walter, A., Ahsan, K., & Rahman, S. (2025). Application of artificial intelligence in demand planning for supply chains: a systematic literature review. . *The International Journal of Logistics Management*.

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

Zaim, S., Yaziz, S. R., Zakaria, R., Mohamad, N. N., & Radi, N. F. A. (2025). ELECTRICITY DEMAND FORECASTING IN MALAYSIA USING SEASONAL BOX-JENKINS MODEL. Jurnal Teknologi . *Sciences & Engineering*.

Zohdi, M., Rafiee, M., Kayvanfar, V., & Salamiraad, A. (2022). Demand forecasting based machine learning algorithms on customer information: an applied approach. *International Journal of Information Technology*.

APPENDIX

APPENDIX 1. Data Description (M5 Forecasting Dataset)

The following appendix is a systematic description of the dataset that was applied in the empirical analysis. M5 Forecasting data has been obtained by a large American retail company and includes hierarchical sales data across multiple SKUs and stores in various states. It contains the item-level demand observations, as well as explanatory variables that are applicable in the model development.

A1.1 Dataset Components

Component	**Description**
Sales Data (item-level)	Unit sales of each product-store combination in many years.
Calendar File	Includes dates, holidays, special events and SNAP (benefits) indicators.
Price File	Includes item selling prices with time.
Product Metadata	Type of product, departments, state / store codes.
Hierarchical Levels	Demand of product, category, department, store, state and total chain-wide.

A1.2 Variables Included

- **Item_ID** - identifier for SKU
- **Store_ID** — location of sales
- **WM_Week -** week number
- **Sell price** — daily price
- **Demand -** y (dependent) (units sold per day).
- **Indicators of events** — holidays, SNAP, events.
- **Time-based attributes** — day of week, month, lagged demand.

A1.3 Data Preparation Summary

The steps that were used were the following:

- Handling missing values
- Cutting outliers on the basis of rolling z-scores.
- Establishing time-varying and fixed (static) features.
- Normalizing features of ML models.
- Lagging variables as 1, 7, 14, 28 days.
- Calculating rolling-window (mean, std, max) aggregates.

APPENDIX 2. Feature Engineering Summary

All the engineered features are summarized in this appendix, and they are applied to statistical, machine learning, and deep learning models.

A2.1 Lag Features

- Lag-1, Lag-7, Lag-14, Lag-28
- Moving averages (7, 14, 28 days)
- Rolling standard deviation windows

A2.2 Calendar Features

- Day of week
- Week of year
- Month
- Holiday indicator
- SNAP program indicator
- School holiday indicator

A2.3 Price Features

- Price
- Price change percentage
- Rolling price volatility

A2.4 Categorical Encodings

- Store ID (one-hot encoding)
- Item category
- Department
- State ID

A2.5 External Variables

- Special events
- Seasonal patterns
- Rolling event impacts

Linear models, XGBoost, LSTM, and TFT had different features applied based on the nature of the model.

APPENDIX 3. Model Hyperparameters

This appendix includes the complete hyperparameter settings used for each forecasting model.

A3.1 ARIMA / ETS / Prophet Parameters

Model	Key Parameters
ARIMA	(p,d,q) selected via AIC minimization; seasonal term (P, D, Q,7)
ETS	Error type: additive; trend: additive; seasonality: additive
Prophet	Daily seasonality = True; yearly seasonality = True; changepoint range = 0.9

A3.2 XGBoost Parameters

- Max_depth = 8
- Learning_rate = 0.05
- Subsample = 0.8
- n_estimators = 500
- Colsample_bytree = 0.8

- Objective = "reg: squarederror"

A3.3 LSTM Parameters

- Hidden units = 128
- Dropout = 0.2
- Batch size = 64
- Sequence length = 28 days
- Optimizer = Adam
- Learning rate = 0.001
- Early stopping patience = 10

A3.4 Temporal Fusion Transformer Parameters

- Hidden dim = 64
- Number_of_heads = 4
- Sequence length = 28
- Static covariates: item ID, category
- Time-varying covariates: price, events, lags
- Dropout = 0.1
- Learning rate = 0.0005

APPENDIX 4. Inventory Simulation Configuration

The complete technical specifications of the simulation environment used to test the correlation between the forecast accuracy and the stockout rates are available in this appendix. The inventory simulations of this research are all performed using a lost sales assumption with a 28-day simulation horizon fixed, which is equal to the evaluation period of forecasting.

Each forecasting model is also dynamically computed reorder points based on forecasted demand during lead-time and on model-specific distributions of forecast errors. The standard deviation of lead-time demand based on forecast RMSE values is used to establish the safety stock levels to ensure that there is uniform integration between the parameters of the inventory policy and forecasting accuracy.

Cycle Service Level (CSL) of the inventory simulation is 95% of the goal. The normal safety factor (=1.65) is used to compute the safety stock levels. Fill Rate is not a performance bound, but it is documented as an outcome of performance in terms of the ratio of demand met without stockouts.

A4.1 Inventory Policy

The (s, S) continuous-review inventory policy was implemented:

s (Reorder point) = calculated via service level and demand variance

S (Order-up-to level) = steady-state target based on safety stock

A4.2 Simulation Parameters

Parameter	Value

Lead Time	7 days
Planning Horizon	28 days
Holding Cost	constant
Shortage Cost	constant penalty
Review Frequency	daily
Backorders	Unmet demand is treated as lost sales, meaning that demand exceeding available on-hand inventory is not backordered and is permanently lost.

A4.3 Simulation Steps

- Load model forecasts
- Generate replenishment orders when inventory ≤ s
- Simulate demand consumption daily
- Record daily stockouts
- Calculate KPIs:
- Stockout Rate
- Fill Rate
- Average Inventory

- Number of Orders

This study presupposes that the lost-sales inventory scenario is located in all simulations, which can be considered as usual retail conditions, where the unmet demand is not ordered back.

APPENDIX 5. Python Code Snippets

A5.1 Example: Training XGBoost Model

```
model = XGBRegressor(
    max_depth=8,
    n_estimators=500,
    learning_rate=0.05,
    subsample=0.8,
    colsample_bytree=0.8
)
model.fit(X_train, y_train)
```

A5.2 Example: LSTM Sequence Preparation

```
model = XGBRegressor(
    max_depth=8,
    n_estimators=500,
    learning_rate=0.05,
    subsample=0.8,
    colsample_bytree=0.8
)
model.fit(X_train, y_train)
```

A5.3 Example: Inventory Simulation Core Loop

```
for day in range(len(demand)):
    inventory -= demand[day]
    if inventory <= s:
        inventory = S
    stockouts += max(0, demand[day] - max(0, inventory))
```