Deep Science Publishing
https://doi.org/10.70593/978-93-49910-26-3_4

Chapter 4: Implementing artificial intelligence-powered predictive maintenance and inventory forecasting in retail supply chains

4.1. Introduction

To compete in the digital economy, retail industry firms must supply the right product, in the right quantity, at the right time, to the right place, and at the right price. Supply chains must be customer-centric and align with customer requirements. They are challenged to shorten cycle lead times as customer demand shifts to the just-in-time ordering, while increasing the inventory to meet service objectives in the face of increased demand volatility and unpredictability. Supply chain activities, especially warehousing and inventory management, are crucial to reduce total business costs, and technology-enabled decision tools are required to optimize these activities. Online channels are the most disruptive factor in today's retail environment, and consumers are increasingly using these channels to select their products while wanting to avoid delays in order fulfillment. This requires additional demand for fulfillment services from operation of distribution and retail branches with complex inventory policies (Choi et al., 2018; Duan et al., 2019; Ghosh et al., 2021). The disruptions have needed collaborative arrangements to reconceptualize the nature of core business, which involves the reconsideration of the product and service mix, the appropriate distribution channel, and the collaborative resources and capabilities required to support the service and product development process over the long term. Retail supply chains are now designed and managed by networks that represent more than just the flow of logistics in the form of distribution channels and suppliers, and new paradigms in performance assessment and resource planning are needed. Profit recovery is achieved through accurate prediction replacement part demand, particularly in demand spikes. Traditional demand forecasting methods use past demand history, but sales of many products exhibit seasonal, cyclical, or trend-stabilizing patterns, which are difficult to model. Time series, autoregressive statistical methods have been unable to provide accurate forecast results in terms of MSE or MAPD (Ivanov et al., 2019; Wang et al., 2021).

4.1.1. Setting the Stage for Retail Supply Chains: Current Landscape and Challenges

Retail supply chains exhibit unique characteristics. Retail supply chains distribute needed products to different retail outlets in different amounts at different times. A retail supply chain seeks to satisfy varying customer demand for quantities and assortment of products within a limited time frame. If a customer cannot find the desired product at the retail outlet at the desired time, the sale is lost. Selling the product is the goal of the supply chain because the customer will not seek out a competitor simply because the product is available. In contrast, for most other supply chains, the customer need not be concerned with the timing of product availability. While the supply chain usually desires to minimize the costs associated with holding inventory and transportation, if a product is not available to the consumer at the retail outlet at this critical point in time, the sale will be lost because competing retailers will be offering the same product for sale. This objective emphasizes the need for close coordination between the retail outlet and the upstream supply chain members that produce and transport the products.

Fig 4 . 1 : Retail Supply Chains

The retail world has been changing dramatically due largely to the rapid advances in information technology and telecommunications technology. Corresponding to these technological advances, business practices have also become more sophisticated and more efficient. Retailers are better able to put together strategic alliances with complementary manufacturers to offer leaner supply chains, but this has also shaped the way they have conducted their operations. Fulfilling customer demand is no longer the primary goal of retailers. Instead, they are very much concerned about minimizing the costs associated with out-of-stock situations while keeping required inventory investment and inventory holding costs as low as possible. They seek to maintain the secrecy concerning their retail data since sharing the information with upstream supply chain members has been perceived to jeopardize their competitive advantage in their quest for survival in the retail world.

4.2. Overview of Retail Supply Chains

Supply chain refers to a set of companies involved in the design, production, distribution, and sale of a product. For example, in a clothing supply chain, the suppliers provide fabric and zippers, the first-tier manufacturer assembles the clothing, the intermediary retailer provides service and support, and the end-user buys and uses the clothing. In retail supply chains, the retailer adds value by delivering the product closer to the consumer, and in return, the end-user pays a price higher than the product's cost at the retailer. Although supply chain partners operate in coordination, they compete with similar supply chain partners at each level; for example, there are different first-tier manufacturers and retailing intermediaries providing substitute products at varying prices. In practice, the end-user is often indirectly paying for the compounded cost from each level of the supply chain.

In the brick-and-mortar retail supply chain model, the manufacturer sells goods to retailers at wholesale prices, and retailers sell goods to customers at retail prices. Generally, the retail price is higher than the wholesale price, and the difference is the retailer's margin. Retailers typically keep the products in their inventory for some time, and customers demand the products according to a stochastic process. If the retailer's inventory level becomes zero, the customer has to wait until the retailer reorders and replenishes the product, or the customer switches to a competing product. Because the customers are unwilling to wait, the retailer will lose sales when the product is out of stock; that is, the retailer experiences a stockout, which is costly for both the retailer and the manufacturer.

4.2.1. Understanding Retail Supply Chain Dynamics and Trends

The major purpose of a retail supply chain is providing products when, where, and in the right quantities needed to satisfy customer needs at an optimum cost of service. Retail supply chains are different from other supply chains since demand evolution is shown to differ significantly, with retail being characterized by high market volatility, with rapid demand changes since being affected by frequently changing customer tastes and preferences, high levels of item variety with demand reflecting rules, or further increased due to the rise of private label brands. Supply chains are hierarchical consisting of core and extended members and several levels. Retail supply chains span wholesale suppliers, manufacturers, distributors, transportation services, and finally retail firms.

As opposed to pure manufacturer supply chains, higher levels of collaboration and optimization strategies can allow retailers to provide lower prices and better availability. The growing importance of sourcing optimized logistics and distribution networks reduces expediting demand variability through successful supplier management and resulting collaborative and trading partner efforts, using collaborative technology and information sharing. Enterprise marketing and distribution resource planning through collaborative sharing speed product time-to-market and provide accurate delivery schedules allowing better aligning distributor needs. Chains can provide more efficiency if decisions are done in a collaborative way. This extends to cross-docking policies and vendor-managed inventories applied to clearing orders consolidated by routing shipments across several stores to a single destination. But increased chain complexity further increased towards multi-channeling, deeper supply lines, brakes on stock and transport cost layers, and outsourcing further challenge the management of retail chain operations.

4.3. The Role of AI in Supply Chain Management

To establish efficient supply chain operations and effective logistics strategies, managing unpredictable behavior is a critical issue that can considerably enhance overall performance. Hence, attention has been paid by organizations to the integration of digital technologies such as AI. In digital transformation, AI is regarded as a powerful force to provide supply chains with real-time responses, predictive capabilities, enhanced decision-making, and improved people-machine collaboration. AI technology encompasses various technologies and services that can help organizations improve their operational processes, improve existing products and services, and develop new products and services. The use of AI in supply chain operations goes beyond that of automation. It includes product design and manufacturing, logistics, stocking of warehouses, inventory levels, controlling the flow of goods, and many other functions in the supply chain.

Traditionally, when managing complex and uncertain supply chain operations, traditional decision-making systems were deterministic models or prescriptive analytics at the upper stages of the decision level. In general, these levels are not directly affecting costs and performance on a short-term basis. Today, technological changes in IT systems make it possible to manage with AI others' decisions and information levels on neural networks to which data on costs and performance indicators are sent to bring them quickly and steadily within set performance ranges. In addition to reducing costs and improving performance, AI can facilitate predictive and preventive techniques and processes, provide training, add data in operational processes, and assist future decision-making at all levels.

4.3.1. Key Benefits of AI Integration in Supply Chain Operations

Artificial intelligence (AI) is becoming a crucial part of the future success of the retail supply chain. Until recently, retail supply chains have discussed Artificial Intelligence, but mainly in the form of business research. Different academic and commercial papers on business describe the benefits of AI but however only some leading companies have so far deployed such systems. The tendency until now has primarily focused on business risk instead of business advantages. New developments such as data availability, increased number of data sources that include Voice of Customer data, reduced model fitting time, and also enhanced computer processing open the door to many new applications for Artificial Intelligence in retail analytics. The most important benefit for a retail company is increased revenue.

The motivation for retail companies to invest in predictive and prescriptive AI systems is enhanced sales revenue that comes from enhanced sales prediction. Higher store sales in combination with real-time digital pricing will lead to increased customer loyalty while diminishing markdowns and price discrimination making more customers satisfied. Predictive AI tools will significantly impact the daily work of Merchandising Planning Associates at the corporate headquarters, as well as Inventory Planning Associates and Store Associates at the local level, who check every week product inventories in their stores. Store Associates are responsible for the replenishment of every product for every store, and predictive AI solutions will help them access the right data and make the right decisions in order to enhance product availability in the stores. Future predictive AI algorithms will ease these Associates' work and reduce the time of fitting models.

4.4. Predictive Maintenance in Retail

Over 1 Million retail stores exist worldwide, including large chains, tech chains, warehouse chains, and grocery megastores. Retail stores in industrialized ascension countries are expected to grow rapidly. These stores also need to maintain operations without interruptions from the failure of equipment such as point-of-sale registers, electronic billboards, credit processing terminals, securities devices, cameras, electronic surveillance systems, automatic substance dispensing devices, payment processing systems, and any equipment dependent on electricity and/or hardwired into a server. Creating and implementing an effective predictive maintenance strategy plan for repair and maintenance of these systems is crucial to ensuring that their operations run continuously without service failures.

To better understand the workshop methods of the predictive maintenance in retail supply chains, two exemplary small use cases are discussed. In both workflow models, we utilize machine learning techniques and embedded systems. The first use case is a POS register, which is the center point at a retail store where a monetary transaction occurs. An interruption of the operations of the POS system would result in huge losses if the payment cannot be processed. It is crucial to minimize the downtime of the point of sale register in case of failure. A second use case is an electronic billboard installed in a retail center. Electronic billboards attract customers and provide advertisement that information about the store enhances customer experience to create a positive imagine about the company. An interruption of the operation of an electronic billboard for a long time without displaying advertisements on it would result in huge losses.

4.4.1. Definition and Importance

Since the beginning of the 21st century, memorable increases in the investment in new technologies that allow retailers to develop new business strategies and models have been verified. Actually, recognizing the necessity to invest in innovation and research, governments have created partnerships with universities and private sectors in order to implement appropriate rules to govern the respective areas. This investment in technology has brought the development of new inventory management systems, distinct from traditional ones, which enable more rational decision making. Within these systems, predictive maintenance solutions make use of sophisticated mathematical models and new data analysis algorithms to anticipate and avoid possible equipment failures. Furthermore, companies that bet on better predictive maintenance management processes can be rewarded with reduced maintenance costs, reduced inventory levels, increased production capacity, improved product quality, and increased customer satisfaction, among others.

The enormous degree of Automation Industrial Revolution 4.0 and Industrial Internet of Things is heavily impacted by the drastically reduction of costs of secure cloud storage and real-time remote monitoring of information generated by millions of sensors in the majority of the production equipment used in the world, smart devices, enabling the development of sophisticated mathematical data and predictive analytics models since the last decade of the 20th century. With the drop of Big Data and advanced analytics techniques that deploy artificial intelligence and machine learning algorithms, which were commonly utilized to design predictive retail-centric maintenance solutions in the middle of the last decade, companies have accelerated the implementation of Industry

4.0 technologies towards Industry Digital Twins in order to develop self-optimizing factory solutions.

4.4.2. Technologies Used

The technologies used in predictive maintenance comprise the whole battery of tools and techniques of data analysis ranging from statistics to advanced artificial intelligence – mainly machine learning and deep learning – that may add value to the data. Statistical methods were the first to emerge, and, for many years, were the go-to techniques used in predictive maintenance. However, the capacity for handling large volumes of data, lack of prior assumptions about the probability distribution function of the data and the ability of machine learning and deep learning techniques for automatically capturing hidden patterns in the data made them more popular in the solutions of predictive maintenance deployed in recent years. Predictive maintenance methods can be classified according to the data they use and how they use them: Traditional methods rely on expert knowledge, machine learning / deep learning methods trained on historical failure data, and monitoring methods that detect anomalous patterns in the signals generated by the equipment while in operation. Additionally, methods may also be classified according to the stage of the process they are (1) one-time prediction of time to failure, (2) continuous prediction of time to failure, (3) maintenance scheduling, (4) work-order multi-objective optimization problem or (5) recently developed, closed-loop predictive maintenance. All these variations in data input and output make predictive maintenance a complex problem, with a variety of tools that may generate different outcomes.

4.4.3. Case Studies

The importance of predictive maintenance reflects on the studies that aim to evaluate it along with its implementation. Several studies influence such evaluation indicating that the implementation of predictive maintenance can minimize the downtimes and machine failures, influence equipment life cycle, among other impacts. Therefore, in this section, we explore some case studies that analyze the application of predictive maintenance in different situations.

A decision support system was developed that integrates various engineering models to identify the type of data warranting further collection and analysis. The system's primary goal is to develop an efficient predictive maintenance policy for systems composed of heterogeneous but dependent components to maximize profit. The approach is based on the principle of maximizing the profit generated by the combination of all maintenance and failure cost-related factors, including the influence of preventive maintenance on the operative costs during the time intervals between failures. A predictive maintenance method was applied in the main conveyor of a distribution center. The technique relies on a fault diagnosis method through decision trees created through the historical data of machine failure. After semiautomatic support for fault diagnosis, and classification of faults with respect to maintenance, predictive models are created based on time series regression between maintenance requisition and anomalies. Then, the results of the previous predictive models are assessed, selecting the best method. The method aims to bring benefits for companies by reduction of unplanned stop, cost reductions, increase of quality of service, and improvement of customer satisfaction.

An important contribution in the area of predictive maintenance applied to retail is found in a study that proposes the development of a software tool with integrated predictive maintenance support to help organizations in the retail sector to avoid equipment failures by providing indicators that help predict these failures. Such increased software tool advantages come essentially from the integrated support of several predictors and the analysis of historical data considering the periodic and seasonal nature of the retail business.

4.5. Inventory Forecasting Techniques

In many industries, the standard operational paradigm is that forecasted demand and inventory replenishment, normally triggered by some external factors, are required to meet the anticipated demand, following a flow-down process from the final product to its distribution centers, and then to different tiers of suppliers in hierarchy. To be more specific, the flow-down model uses information at the bottom of the supply chain, i.e., the demand at retail level, to forecast the needs at higher levels of the supply chain, then allocate these needs to various distribution and/or manufacturing locations to ensure acceptable service level. This requirement passes through the entire supply chain, leading to insufficient or excessive inventory at various levels of the supply chain and low resource utilization.

Most of the traditional demand forecasting models, such as time series trend analysis, exponential smoothing models, decomposition models, regression models as well as neural networks, mostly rely heavily on training historical data to estimate parameters and are not implementable when very little or no data is available at all, which is almost always the case at the higher levels of the supply chain. Moreover, these models impliedly assume that the underlying demand processes are deterministic, when the demand processes in reality exhibit stochastic behavior. In practice, any model of demand uncertainty used for inventory management should be based not on the economics of demand estimation, but rather on the entropic bias introduced by a limited demand history. More importantly, the intrinsic uncertainty in demand would in fact be, in most meaningful cases, considerably greater at the higher levels of the supply chain.

4.5.1. Traditional Methods

The science of inventory forecasting has a long history. Traditional methods include Qualitative Methods, Time Series Analysis, Causal Methods, Linear Regression, and Econometric Methods.

Qualitative Methods are suggestions from experts from different areas. These experts participate in juries or brainstorming, and underline what they think is more important. Experts can use previous data to try to quantify their suggestions. The advance of machine learning algorithms drastically increased the availability of decision variables. Words from customers, partners and employees can be easily transformed into quantitative variables to help experts to make better predictions. Text mining can transform opinions expressed on social media to reach workers that can improve predictions.

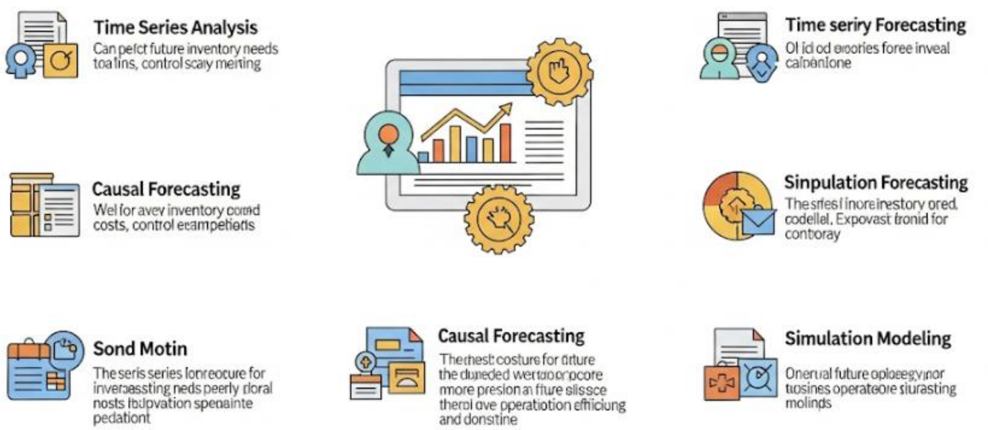

Fig 4 . 2 : Inventory Forecasting Methods

Time Series Analysis has been largely used in the past. They are based on collected sales data. The simplest method consists of using historical average sales in the time horizon to come but it assigns the same weight to all previous sales. Therefore, not considering possible recent changes in seasonality and trend. This brings us to other methods such as moving averages that also do not assign weights to influences throughout the history. The exponential smoothing method assigns a higher weight to recent sales but it is a linear function. Considering that sales may need to suffer heavier changes, in cases of an innovation, other approaches used in inventory theory such as the Holt-Winters method where trend and seasonality are linear as well as ARMA and ARIMAX models, which are dedicated ARMA for non-stationary series, have only linear solutions.

Causal Models are used to correlate sales that have been observed and independent variables. There are statistical models and simulation models. Simulation models are more advanced simulation software. For statistical models, consumers are assumed to be rational homogeneous individuals. Demands are inelastic and functions have to their theoretical properties. The truth is that those constraints can limit flexibility. Nevertheless, they are an excellent tool for demand forecasting when the conditions can be achieved.

4.5.2. AI-Driven Approaches

Furthermore, concerning the AI-supported techniques, it is suggested that, to better inform shoppers and help prices be more dynamic, the inclusion of five features in the prediction model, including seasonality, displayed promo, displayed state, peer promo, and whether promo, is beneficial. In our model, data had to be refined, scaled and was transformed into a supervised predictive model or translated into a supervised classification model. The first model applies to our non-disaggregated models, whereas the second model applies to our disaggregated models predicting whether or not a SKU was sold in each day of our prediction window. Starting from the existing demand data in the training window, we modeled the sales with one of two possible cases: one as a continuous quantity prediction model, including both sold and unsold units during promotions for a SKU on a specific day, with a constant price, and one as a classification model, signaling a binary SKU day case, which included the estimated units of the sold SKU at a specific day at the constant price, without being either displayed or peer promoted. Finally, both the models were validated against some benchmarks with different hyperparameter optimization approaches.

There is a need to review previous forecasting models, but there is no concern with utility functions in the overall forecasts. Similarly, after reviewing several forecasting techniques in terms of classification accuracy, predictive accuracy, and utility-based loss functions, it has been stated that, although the research into the sales forecasting process has been substantial, it has mainly explored the application of conventional statistical techniques. It is concluded that, despite the growing popularity of AI and ML techniques in many business areas, they do not seem to have had a similar impact on the sales forecasting process.

4.5.3. Comparative Analysis

A comparison matrix in Table 7 presents a summary of the reviewed methods on the following discussed criteria: capabilities, i.e., inventory demand characteristics that are addressed; computational efficiency, i.e., computational cost correlated to overhead; data requirements, i.e., amount and types of data needed for employing; training requirements, i.e., amount and types of data needed to train the models; and modeling capability, i.e., ability to capture dependencies between times series. Together with these attributes, we also add our perspective comments on the pros and cons of the different approaches, to help the reader get a sense of the most popular methods presented.

Traditional methods customize different strategies or mechanics to accommodate the characteristics of the problem domain in hand. Moving averages, exponential smoothing, or the Holt-Winters seasonal approach capture proven time-series features, while ARIMA-based models require an optimal selection of time-lagged forecast variables. Traditional methods rely on the past values, either from the target time series or covariates, and do not require training. In addition, they can generalize well when the models are adapted to the right settings.

Traditional methods can also tackle problems from limited data domains, provided they do not introduce selection biases in certain periods of time. They therefore require a limited amount of data, mostly historical, and can use external or internal forecasts at a lower cost and overhead relative to machine learning or AI. They can perform multi-sourcing by inherently including additional time series that are directly connected to the forecasting source or consider the sum of different products need to forecast the total demand. However, traditional methods certainly present several limitations. They rely on complex heuristics that are relatively ineffective at capturing the dependency patterns between product demand variations.

4.6. Data Collection and Management

This section describes data collection and management in detail. It discusses selected data sources, methods and tools used to develop, clean, store, and transform data to meet user requirements. Data management is an ongoing effort, which requires constant attention. Advantages of maintaining high-quality data include saving time by alleviating the data cleaning task, accelerating the data analysis and model calibration process, and preventing model failures due to inaccurate data. On the contrary, poor data quality may prevent systems from delivering the expected results. In the context of this thesis, retail and supply chain data integration and cleaning are paramount. This is particularly true for the third and sixth relevant use cases.

Historically, the adoption of Artificial Intelligence in supply chains has been slower than expected. Initially, the technology costs were too high. In addition, the required in-house expertise to implement tailored AI-based solutions was lacking. Only those companies willing to invest heavily on cloud computing were able to implement successful AI solutions. Technological progress and increased macro business data availability leveled the playing field and democratized access to data-based technological innovation, and particularly innovation driven by Machine Learning techniques. Today, due to the easy availability of cloud-based Machine Learning tools, even small and mid-size companies can benefit from machine learning-based solutions. Moreover, today's data-driven AI-based innovation goes beyond the company's internal data. Data from external, alternative sources has become an essential ingredient of AI-based business solutions.

Accordingly with machine learning best practice, data processing and data cleansing is a numerically important part of this work. In order for estimate errors to have a meaningful interpretation, high quality data should be used. In practice, raw data is not always available, and different versions exist. Different sources may use different encoding systems and communicational styles. Thus, using another source besides the one corresponding to the target variable would require translating and cleaning the variables to be compared and merged.

4.6.1. Sources of Data

Data collection is the starting point in the establishment of both predictive maintenance and inventory forecasting solutions. Consequently, understanding and agreeing upon the sources of data is essential, however, should not be chosen lightly, as this is a choice that can create long-lasting impacts. We can distinguish two main types of data sources that can be used in AI-PPM solutions: digitalized data sources and enhanced physical data sources. The digitalized data sources are data that exist in digital form in the company's systems that together form a digital backbone of a company. These databases are often built from the recording of salient events related to an asset or related to operations that contain timestamp identifiers and in many instances unique integer identifiers that relate to the asset for which the event is recorded. These data sources can be easily accessed, and the most common such sources are ERPs, which contain information on inventories and the supply chain in general, AMs, which contain a record of performed maintenance and repair actions, and IoT systems or sensors that track asset use over time. Other plausible digital databases are CRM systems, PLM databases, databases related to the company's ordering process, or even accident and incident databases if these are utilized in the company.

However, these data do not always exist, or do not contain the necessary granularity, or relevance for model building. In such instances, enhanced physical data sources should be considered. Enhanced physical data sources consist of data physically collected in a non-digitalized format that are created together with a thorough understanding and design of the desired modeling outcomes. Recordings can be made by either humans or smart devices. The enhanced physical data sources are often referred to as ground-truth data as they are manually collected data reflecting reality. An example of enhanced physical data sources utilized in an AI-PPM solution is the joint field expeditions where engineers collected activity data on asset use while encountering the difficulties of physically collecting these data.

4.6.2. Data Quality and Governance

The quality of data is critical, and systems should be in place to support consistency, quality control, integrity checking, sharing, and to avoid redundancy. For predicting maintenance requirements and forewarning potential issues within supply chain technology enabled systems, additional methods for ensuring that data is appropriate and usable would be useful and desirable. Big data standards in measuring quality in big data involve both subjective and objective measures. Subjective evaluation would involve an expert deciding on the quality of different truth enhancing data quality dimensions (for example, time dimension - semantic correctness related to the temporal extent of the data, or degree of aggregation; completeness; redundancy; dimensional consistency; precision – accuracy of data values considering their intended use); while objective evaluation processes could be implemented as well, where for example, integrity checking involves verifying that the data lies in the right domain.

For sensing systems, the data could be examined for each sensor, because input measurement error rates can differ considerably for different sensors. Also, monitoring these values or comparison to other sources could give an estimate of credibility. Testing involves either functionality testing via validation and comparison against other trusted data; or using confidence measures such as root mean square error, or the more commonly employed mean absolute error, to indicate validation threshold levels for sensor values. Although completeness is a major challenge, ongoing considerations and updates relative to the sensor technology can help improve and guarantee data quality.

Data governance involves multiple organizations working together to define policy and establish rules and standards supporting the creation, storage, movement, and consumer access to the data used throughout all of these collaborative organizations. This involves the establishment of procedures for creating useful metadata, and for data models, as well as the establishment of appropriate data sharing agreements and structures to provide user access for all of the synergies desired in attempting to gain better insight from the shared data.

4.7. Machine Learning Algorithms for Predictive Maintenance

The field of predictive maintenance has seen advancements stemming from novel sensor technologies and the vast amounts of data produced by sensors on industrial systems spanning various fields. Organizations are seeking to move past the traditional, time-based maintenance applied indiscriminately to all aspects of production and downtime management. Instead, work is focused on reducing operational risk by predicting equipment and production failures. The purpose of predictive maintenance is to better leverage the data being generated from machinery to perform efficient and strategic maintenance, improving KPIs such as cost, efficiency, and uptime. With the objective of reducing unplanned downtime, the decision of when to execute maintenance can be informed by understanding the underlying health of systems critical to the production lifecycle. The timing of maintenance can occur closer to operating thresholds using advanced data techniques, including machine learning. Performance prediction models can be constructed from information on physical, mechanical, and operational parameters.

The complex interdependency of physical systems and the uncertainty related to environmental parameters can be modeled, learned, and forecasted via machine learning techniques. In predictive maintenance, the lifecycle of predictive models is impacted by the characteristic lifecycle properties of machines. The performance of existing models deteriorates when they transition between different lifecycle phases, necessitating their periodic retraining. Companies are presented with challenges separate from the technical impediments to value generation through machine learning. There are short- and long-term difficulties around data strategy and governance; process, culture, and organizational design; technology management; expertise and skills transfer; and operational rigor. Many lifecycle changes and the impact on asset degradation and remaining life can be modeled with supervised machine learning techniques.

4.7.1. Supervised Learning

Many classification and regression ML algorithms belong to the category of supervised learning and the model learned by supervised algorithms can be used as an estimator to make predictions of failure events. In supervised learning problems, the model is trained using data which contain the inputs and the desired output(s). The given input-output pairs are referred to as labeled data. The labeled data used for supervised learning problems varies in characteristics such as reliability and amount. A large body of research has focused on predicting failure times or identifying failures using predictive models learned from reliable labeled and unlabeled data. Although such models can be powerful predictors of future failures, the construction of large, reliable history failure log databases can often pose significant challenges. Others have focused on enhancing the predictive accuracy of models trained on smaller amounts of labeled data by adopting semi-supervised learning techniques in which the model is first initialized with a model pre-trained with labeled data from a related dataset as a starting point and then fine-tuned using smaller amounts of labeled data. Others have considered the influence of performance measure selection on the predictive accuracy of a variety of weak classifier models trained on imbalanced datasets, varying the performance measures for sensitivity, specificity, overall accuracy, negative predictive accuracy, and Matthews Correlation Coefficient, and considered the use of the Matthews Correlation Coefficient as a performance measure in such imbalanced domains. Others have sought to identify and cluster failure events with similar temporal profiles and then share the related clusters among the classification models of different failure events using a multitask-learning framework.

4.7.2. Unsupervised Learning

Unsupervised learning allows the machine to decide what to learn and how. It examines the input data to find some inherent patterns, eliminating the need for human intervention in data tagging. Exploratory Data Analysis is an example of unsupervised learning that makes use of clustering techniques like K-means clustering or Hierarchical clustering. Advanced dimensionality reduction techniques like t-SNE and Auto-encoders use unsupervised learning to help in the visualization of data. In the context of time series data, clustering techniques can be used to discover some inherent patterns in the input data. Anomaly detection algorithms search for deviations from the normal pattern based on multivariate measures of distance, and Principal Component Analysis can model temperature effects and metadata de-noising in streaming sensor data.

Unsupervised learning is common in Predictive Maintenance applications where there is hardly any labelled data for training. Unsupervised learning has also been used for anomaly detection on data from industrial processes or condition monitoring of rotating machinery. Compared with supervised models, which are often less robust due to overfitting to historical anomalies, unsupervised anomaly detection models are more interpretable and easy to deploy. Cascaded anomaly detection models based on unsupervised learning can also help transfer knowledge from labelled data of similar failure modes in other products or regions. Additionally, unsupervised learning helps eliminate the challenges of classification-based architectures which can suffer from class imbalance.

4.7.3. Reinforcement Learning

Reinforcement learning (RL) finds its niche in situations where the actions to be taken aren't discretely defined within the environment – a non-stationary temporal context that isn't explicably bounded – but they can still be identified using the continual signal of accumulated rewards. RL, like the subfields of computer vision and natural language processing within machine learning, shaken loose a handful of very successful algorithms. In terms of core ideas, RL is more similar to econometric approaches than it is to any of the other machine learning subfields – such as supervised learning, classification and regression; or unsupervised learning, clustering and dimensionality reduction – because what RL emphasizes isn't broad training-sample use. It's that RL assumes that any decision-maker is optimizing an objective function.

In other words, each agent interacts with a complex environment and – by means of thousands or millions of random-thoughtful decisions – determines, through trial-and-error, an optimal course of action. The experience that the agent accumulates doesn't take the form of pairs of provided inputs and outputs, as is the case with supervised learning and successive iterations over a canonical training set. Rather, it comprises sequences of inputs and successive outputs, each decision leading to a change in the current state of the environment. As with supervised learning, the RL idea is to abstract from the details of those experiential sequences, and replace them with a multitude of simpler abstractions that can then be generalized to an abstract model of the decision-maker's environment. Importantly, the RL agent meshes with its environment in a continuous loop, providing experience that, in turn, changes the environment's response to the agent's inputs. In economic theory this is known as "strategic interaction." In RL, it's known as the policy-inference problem.

4.8. Integrating AI Solutions into Existing Systems

There is an increasing urgency to develop intelligent solutions in addressing the challenges of predictive inventory forecasting and predictive maintenance in retail supply chains. While some organizations would prefer investing in new or custom-build software applications, the majority would rather innovate within their existing systems, given the possible substantive investment that could come from a search for new software providers or risk protracted integrations with existing infrastructure. Hence, retail supply chains have a growing need for intelligent solutions that can seamlessly integrate into the existing system architecture. Enabling AI intelligence into solutions portfolios that have been used for years to address the problem of predictive maintenance and predictive inventory monitoring is an overlooked area in the development of intelligent enterprise systems frameworks and technological stacks for AI development.

Technological barriers prohibit many supply chain systems in retail from embodying ‘plug-and-play’ AI solutions. Technology vendors explore a small number of industries and offer applications for only the most common use cases. There are several barriers that contribute to the low usage of AI for predictive analytics even when highly relevant data is available. First are the challenges from internal stakeholder pressure in organizations. This collaboration can take a long time to reach expertise in the areas of knowledge management systems. Models may also be used which, if executed without suitable management, can lead to even greater obstacles for the user. However, the advantages of using AI techniques are such that, despite their limitations, they still present a far better alternative and should be taken into account for users. A common challenge of these models is the 'user look and feel'. When wrapped in a software system which has been designed to be easily used, their use becomes tougher, especially for complex models. These 'user-friendly' systems must appeal to the user portfolio in the organizations due to their complexity in some cases.

4.8.1. Challenges and Barriers

A very simple yet false assumption regarding AI solutions for business is that these are plug and play products that would yield immediate ROI when integrated in the company's business processes. Therefore, it is essential to better describe the barriers and challenges that organizations would typically experience when trying to integrate AI solutions into their existing business systems. There are challenges related to developing a complete understanding of the business processes by the AI developers and, vice versa, for the organizations to understand how best to utilize the capabilities and already defined limitations of the AI tools. Most existing business processes are very rich and rely on sophisticated knowledge and heuristics built over time by the staff based on their day-to-day experiences. However, most reporting business data in the organization, especially if batch oriented, are extremely limited in terms of what they convey about the knowledge required to build the predictive models that would be the essence of the AI solution.

Thus, with regards to the business processes designed by experts or used by employees for many years, there is usually a high possibility of the processes being undocumented at a level that would provide a solid understanding for documenting the key dependencies and variances used to create, modify and review any instance of these processes. Moreover, these processes often exhibit very high complexity with diverse structure and data, resources and people utilization. Simultaneously, the information that AI tools can process to learn the tasks typically exhibit low volume as they reflect the decisions made by the employees over long time spans. This can be minimized by using AI to review events and decisions over the past to help document the actual deviations from the expected work rules. This difficulty is heightened in retail by the need to merge many point of sales transactions spread over hundreds or thousands of stores. Even once actual tasks are documented, the task learning models may be difficult to build without proper design of the process and task documentation modules that support modeling the rich variety of different instances.

4.8.2. Best Practices

Integrating advanced machine learning techniques into existing enterprise-wide systems will require novel approaches, due to the nature of the data and techniques applied, but also due to the scale and operational requirements of real-world systems. In general, we recommend developing and testing advanced techniques in sandbox environments. Advanced AI models can then be incrementally integrated back into the integrated operational decision-making processes of corporate systems. In several cases, we recommend using supervised learning approaches due to the operational requirements of production environments, and leveraging the most advanced AI and machine learning techniques only in the learning steps of operational systems.

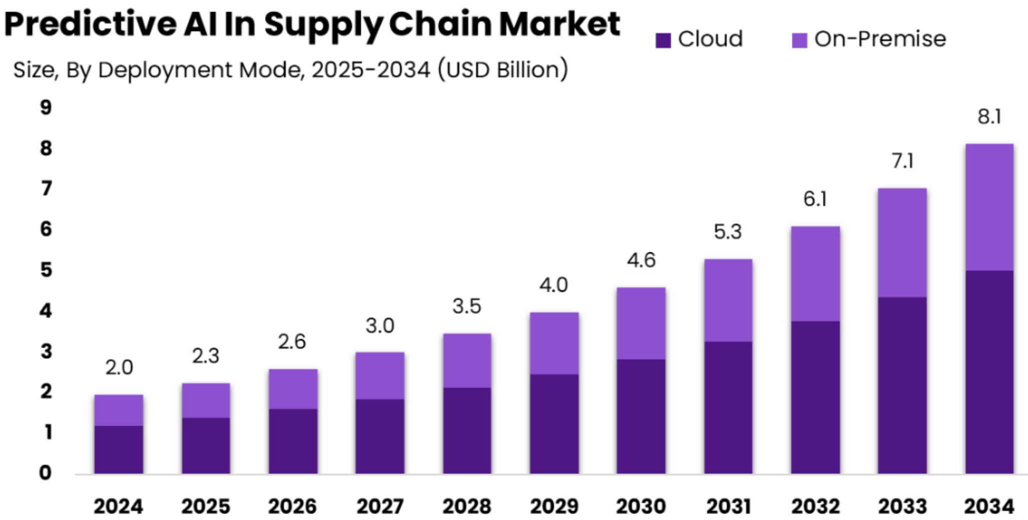

Fig 4 . 3 : Predictive AI In Supply Chain Market Size

Data is often collected in enterprise systems in order to solve decision-making problems, and not in general terms of extending support for modeling. Traditional exploratory analysis would often not be possible, and therefore, it will be necessary to leverage business domain results and large proven operational systems in order to identify the most promising data, modeling, and decision variables usable for advanced model leveraging. Domain modeling steps leading to supervised learning model development will often need a cyclical review and update process for parameter updates and recalibrations, due to the nature of enterprise systems and decision-making problem. More advanced techniques could be leveraged at the feedback stage, where additional iterations can be analyzed as control variables, if technically feasible for the specific enterprise decision.

Models developed with other learning approaches will often be categorical, due to the nature of the specific problem addressable and enterprise requirement of assigning optimal values for the corresponding extreme decision metric targets. In these cases, all possible combinations of model variable states will need to be processed in order to assign the necessary model weights for operational implementations and use.

4.9. Conclusion

Retail is an evolving enterprise which utilizes modern technology in many areas to enhance its abundance of products and services. Supply chain is a complex interplay of organizations, people, information, processes and products to deliver a consistent product to consumers. Superfluous inventory, ruptured supply, and exorbitant costs are serious challenges that retail chains cope with. No business can afford to hold excessive money in playing around SKUs and not being invested in more profitable ways. In this work, we discussed how to use AI technology, Business-Driven Predictive Analytics, applied in innovative ways, to enhance forecasting accuracy, hence improving other interrelated operational areas, such as inventory and merchandising, increasing the chance of correcting errors before they occur and streamlining the organization for better utilization of human resources.

Our Predictive Maintenance solution, along with the ability to train predictive machine learning models using a limited set of historical data, are both enablers to this goal. We are witnessing the advent and rise of Smart Stores, along with the fast-paced technological evolution in omnichannel logistics and ecommerce. Among many characteristics of Smart Stores, two are pertinent to the discussion: adaptation of store layouts to real-world consumer behavioral changes and automation of merchandising processes and logistics. Artificial intelligence, particularly business-driven predictive analytics and predictive maintenance capabilities, are instrumental in areas such as computer vision and conversational chatbots, empowering retailers to provide great customer experience when visiting stores and ease the handling of online order fulfillment either to be delivered directly to customers or picked up by customers at their local stores.

4.9.1. Final Thoughts on Advancing Retail Supply Chains Through Innovation

Revolution and disruption in retail are often used to characterize the last decade in the industry. However, at least in North America, the specific family of events has taken 30 years to unfold. It began with massive shifts in consumers' demographics and preferences, as well as overarching impacts from globalization, followed by the dot-com era. Retail modernization was slowly but steadily being built by innovators who utilized new technology, with credits given to companies such as Walmart and Amazon. However, the impact was muted until an unlikely pandemic turned out to be the proverbial spark that not only lit the fire but really intensified it. During the COVID crisis, consumers have embraced online shopping at a speed and scale that many believe has ushered in a new normal, one that is here to stay. For many retailers, it is now about survival, with no guarantee of prosperity ahead. This conclusion is written from the perspective of the prior content-breaking retail supply chain framework. We close with some final thoughts.

Advancing retail supply chains through the latest technology and business models is at the core of this effort. The value chain is essentially being turned inside out, with advanced collaborative, consumer-centric ecosystems fostering competitive advantage. In the post-pandemic new normal, smart supply chains need to connect the digital dots, thereby facilitating a customer-centric view of supply and product flows. There are a number of ways to accomplish this. This will require new and old players to work together. It will require innovation and a lot of it for real impact.

References

Ivanov D., Dolgui A., Sokolov B. (2019). The Impact of Digital Technology and Industry 4.0 on the Ripple Effect and Supply Chain Risk Analytics. International Journal of Production Research, 57(3), 829–846. https://doi.org/10.1080/00207543.2018.1488086

Duan Y., Cao G., Edwards J.S. (2019). Understanding the Impact of Business Analytics on Innovation. Information & Management, 56(1), 103–116. https://doi.org/10.1016/j.im.2018.06.003

Wang K., Zhang J., He W., Shen J. (2021). Deep Learning for Predictive Maintenance in Manufacturing: A Review. Journal of Manufacturing Systems, 59, 476–493. https://doi.org/10.1016/j.jmsy.2021.03.006

Choi T.M., Wallace S.W., Wang Y. (2018). Big Data Analytics in Operations Management. Production and Operations Management, 27(10), 1868–1889. https://doi.org/10.1111/poms.12838

Ghosh D., Kumar M., Bag S., De K. (2021). Impact of Artificial Intelligence on Supply Chain Performance: A Systematic Literature Review. Computers & Industrial Engineering, 161, 107675. https://doi.org/10.1016/j.cie.2021.107675