# Print Demand Forecasting with Machine Learning at HP Inc.

M Harshvardhan
University of Tennessee, Knoxville, harshvar@utk.edu

Cara Curtland
HP Inc., Vancouver, WA, cara.curtland@hp.com

Jerry Hwang
HP Inc., Palo Alto, CA, jhwang@alumni.stanford.edu

Chuck VanDam
HP Inc., Boise, ID, vandam@hp.com

Adam Ghozeil
HP Inc., Corvallis, OR, adam.ghozeil@hp.com

Pedro A. Neto
HP Inc., San Francisco, CA pedro.neto@hp.com

Frederic Marie
HP Inc., Grenoble, France frederic.mr.marie@hp.com

ChuanRen Liu
University of Tennessee, Knoxville, cliu89@utk.edu

HP Inc. manufactures and sells more than 18,000 print-related products in over 170 countries. Accurate forecasting of the heterogeneous and dynamic demand is vital to support supply planning decisions for manufacturing, inventory management, shipment scheduling, and ultimately customer satisfaction. Forecasting higher or lower than actual demand results excess or shortage that reduces profitability and impacts on-time delivery to customers. Historically, the supply planning depended on (1) consensus demand forecasting approach, which requires manual collection and integration of information by the forecasting experts, and (2) statistical time-series forecasting models. The consensus forecasting approach also requires frequent corrections if some uncertainties in the demand are not accounted for when releasing the forecasting results. While traditional time-series models can work automatically without frequent correction, their forecasting performance is unsatisfactory due to oversimplified modeling inputs and assumptions. In this project, we document the process of using Machine Learning (ML) techniques across all print products at HP Inc., worldwide. Our aim is to automate the forecasting process with high accuracy and to integrate those results into a human-in-the-loop process that merges the strengths of ML, statistical, and consensus forecasting. Our tree-based forecasting model reduced systematic errors in comparison to existing approaches such as the consensus and statistical forecasting approaches and was deployed as an integrated part of HP Inc.'s forecasting process. Furthermore, our ML framework establishes strong foundation for further methodological improvements in the ML algorithm. We report extensive empirical evidence guiding our methodology design and demonstrating the business implications of our project. We also share several important principles we have applied to manage team-based collaboration for an enterprise-scale project and to ensure the success of our ML-based demand forecasting.

*Keywords:* Printers and Electronics, Data-Driven Decision Making, Demand Forecasting, Machine Learning.

# Introduction
## Background

HP Inc. manufactures and sells over 18,000 Stock Keeping Units (SKUs) of print products that are sold in over 170 countries. They include home printers, office printers, ink, toner, and other services such as 3D and large-format printing. Specifically, home printers are targeted to consumers looking to buy standalone printers. They're usually sold through channel partners including retailers like Walmart and Amazon. Office printers are usually sold via business contracts through managed account deals. The consumables, Ink and Toner, are sold to existing printer-owners. 3D Printing offers a portfolio of additive manufacturing solutions and supplies to help customers with unique or experimental demands. Additionally, HP also offers large-format printing solutions and supplies through industrial products. Beyond these five top-level categories, products are further classified based on their technology and platform, resulting in over 18,000 SKUs. Building on this portfolio breadth, HP operates on a global scale with markets organized into three world regions: Americas (AMS), Europe, Middle-East and Africa (EMEA), and Asia-Pacific (APAC). Countries in each world region are grouped by geographical proximity and the demand forecasting is needed for each SKU in each Group of Countries (GOC).

Given its diverse product portfolio and extensive global reach, accurate demand forecasting is a crucial component of operational strategy for an international company like HP. Indeed, accurate forecasts are critical to planning and operational decisions such as strategically allocating resources, managing inventory, and aligning production schedules with consumer demand (Gardner 1990, Ritzman and King 1993, Lee 2002, Seifert et al. 2015). Furthermore, past studies have highlighted that effective forecasting can not only support business operations, but can also lead to cost savings and improved efficiency throughout the supply chain (Simatupang and Sridharan 2005, Seifert et al. 2015, Fildes et al. 2022). With the advancement of machine learning (ML) technologies, there's been a significant interest from academics and practitioners in applying ML methods for these forecasting tasks. This paper discusses the challenges and solutions to deploy an ML-based framework to forecast product demand for a Fortune-500 technology company like HP.

**Current Practices**

Before implementing ML-based models, we relied on *Statistical* and *Consensus* forecasts for demand forecasting. The *Statistical* forecasts leverage historical demand data and uses conventional time-series models, such as autoregressive (AR), moving averages (MA), ARMA, ARIMA, and exponential smoothing (ETS) models (Hyndman and Athanasopoulos 2018). While these models are cost-effective and easy to implement, they often lack the nuance required for accurate forecasting due to oversimplified modeling assumptions. *Statistical* models are also 'local' in nature, training with a single time-series whereas ML-based models are 'global', incorporating details from multiple time-series. We also found that local models struggle with short product life cycles whereas global model gets to learn from similar products. A common attempt to handle this is through predecessor-successor mapping, but such information isn't always readily available to forecasters (Manary et al. 2019). In contrast, the *Consensus* forecasts incorporate quantitative information such as historical demand and current inventory levels, as well as qualitative demand signals and contextual information, with the *Statistical* forecast also serving as an input. Particularly, the Consensus forecasters heavily leverage *soft data* like customer demand sentiments and deal progress (Fildes et al. 2009, Petropoulos et al. 2018). 'Soft data' includes qualitative knowledge on upcoming promotions offered by channel partners to their customers, deal stage for bulk corporate orders, subjective opinions from market insiders and experts, and networking insights through deep business relationships, among others. Though soft data is challenging to include and maintain, its strategic advantages in capturing transient market conditions make it invaluable to forecasting, especially contributing to robustness of planner forecasts. However, recent research also shows that human-based forecasts struggle to effectively filter out noise in the inputs. In fact, forecasters tend to reproduce the noise in a time-series in their forecasts rather than filter it out (Petropoulos and Siemsen 2023). Figure 1 depicts and compares the different demand forecasting solutions, where our focus is to develop the new machine learning forecasts as shown in the orange box.

To bridge these gaps and develop a unified approach, an expert group was tasked with developing and deploying an ML-based framework for demand forecasting. When the accuracy of the ML model outperforms the traditional *Statistical* forecasts in terms of accuracy, it is beneficial to use ML forecasts as the basis for the *Consensus* forecasts; that is, choosing the best analytical forecast by product and geography. Armed with ML knowledge

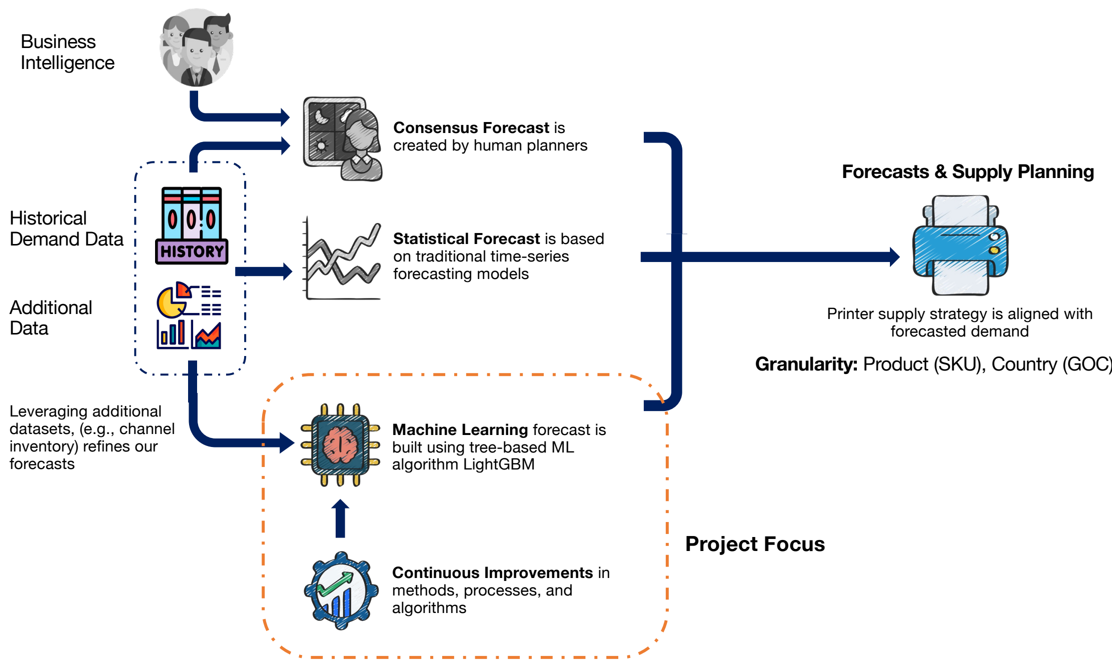


**Figure 1** **Overview of the forecasting process.**

*Notes:* Our approach leverages historical and additional data to create robust statistical and machine learning forecasts. These forecasts are then refined by consensus planners, serving as the crucial human element in the loop, to formulate a comprehensive forecast that informs granular supply planning. The focus of this work is “ML Forecasting”.

and domain expertise, Strategic Planning and Modeling (SPaM) group at HP Inc. utilized tree-based ML models for forecasting Print demand, and deployed the model for forecasting at scale.

### Strategic Planning and Modeling Group (SPaM)

Formed in 1994, SPaM is a team of OR specialists, data scientists, and external collaborators who provide internal support to HP product divisions to improve their efficiency, cost-effectiveness, and profitability (Laval et al. 2005). SPaM has developed and adapted many supply chain models for specific applications at HP (Cargille and Branvold 2000). For example, Ward et al. (2010) documents the team’s work in transforming product portfolio management: developing a new framework for screening new products using custom return-on-investment calculators, and a revenue-coverage-optimization tool to manage product variety after introduction. Similarly, Billington et al. (2004) documents how efforts from SPaM helped HP create a standard process for analyzing and designing supply-chain networks.

**Challenges**

Now we will discuss some major challenges in adopting ML techniques for the product demand forecasting task. *First*, demand for products in different markets can be impacted by the complex interplay of various factors, such as economic conditions, seasonal trends, and regional variations. Given the scope of the problem involving a wide range of products being sold across numerous market regions, it is a non-trivial endeavor to develop one versatile model to incorporate all the factors that can generalize well while still being tailored to individual products and regions. *Second*, adaptability to market fluctuations and external factors is essential for accurate predictions in the face of demand shifts, supply chain disruptions, or unforeseen events. A model that can automatically incorporate dynamic changes will be much easier to manage than statistical and planner forecasting approaches. *Third*, the availability and quality of historical demand data also play a crucial role in the prediction performance (Cortes et al. 1994). Addressing data quality issues, such as inconsistent, inaccurate, outdated, and missing information, is crucial to ensure that the forecasting model is robust and reliable to support planning and operational decisions.

In addition to the three technical challenges in developing the ML models for demand forecasting, a robust *project management strategy* is pivotal for the successful deployment of such project. To achieve that, our team coordinated efforts from data scientists, production planners, and external experts, in addition to the Consensus and Statistical forecasting team. We developed a framework that adeptly captures the intricate relationships and patterns among various input factors to support accurate demand forecasts. The depth and precision of our approach are fine-tuned through iterative model design and rigorous experimental validation. We will discuss more details of our project management strategy and efforts in later sections.

Our ML-based forecasts complement, rather than replace, existing methods. Our goal is to create analytical forecasts using more advanced models than simplistic statistical models, which demonstrates higher accuracy than planners' forecasts. However, it is not imperative to predict everything better. As shown in Figure 1, we make a decision to choose between Statistical and ML forecasts for each product and geography, every month depending on their historical performance and expected future performance. Combining analytical forecasts with planner forecasts is known to increase forecast accuracy (Lawrence et al. 1986, Armstrong 2001). This system, akin to human-in-the-loop, allows planners to automate usage of best-performers as final forecasts, freeing them to focus on critical products. This collaborative approach enables all three forecasts to continue to improve in parallel, while shifting the work to higher-value-add analytics as the ML forecast improves.

**Contributions**

Addressing the above limitations in traditional demand forecasting by creating an accurate, automated ML model is a challenging yet valuable endeavor. Implementing this at HP, i.e. forecasting 18,000+ products across 170 countries, provides a scalable case study for other businesses.

Our key contributions include:

1. **Scalable ML-based Forecasting Framework:** We demonstrate the effectiveness of tree-based models, specifically LightGBM, in addressing enterprise-scale product demand forecasting challenges across diverse products and countries.

2. **MLOps for Forecasting:** We stress the need of robust project management and maintenance strategies, specifically aligning with principles outlined in Curtland et al. (2022). Our framework values reproducible analysis with parametrized notebooks, and using advanced experiment tracking for sustaining the performance and reliability of models over time with MLFlow (Zaharia et al. 2018).

3. **Case Study at HP Inc.:** Our work serves as a comprehensive guide for both practitioners and researchers attempting to tackle similar enterprise-scale forecasting challenges in other industries or contexts. This work establishes a strong foundation for further ML model improvements and operationalizing them at scale.

## Literature Review

In this section, we first provide some related methodology papers, followed by model selection, and lastly papers discussing implementation of demand forecasting at an organization. A summary of works of similar nature is provided in Table 1.

The methodology frameworks for demand forecasting have significantly evolved over the last few decades. Traditionally, classic time-series models such as AR, MA, ARMA, ARIMA, and ETS were used for demand forecasting tasks, which only utilize lagged demands as the input (Hyndman and Athanasopoulos 2018). ML models can accommodate nonlinearity and handle a broader range of inputs, such as unstructured and high-dimensional data of various types. In recent years we have seen huge potential of ML algorithms in demand

| **Paper** | **Input** | **Model** | **Evaluation Metric** |
|---|---|---|---|
| Dodin et al. (2023) | Lagged demands, demand statistics, seasonality components, region and month index, average age of shipped products | Improved LightGBM, Elastic Net | RMSSE |
| Gardner (1990) | Lagged demand | Exponential-smoothing Model (ETS) | Investment and Delay Time |
| Makridakis et al. (2018) | M-3 data | MLP, BNN, RBF, GRNN, KNN, CART, SVR, GP, RNN, LSTM, SES, ETS | sMAPE, MASE |
| Qi et al. (2023) | Lagged demand, inventory | End-to-end Model (Dynamic Programming, RNN, MLP) | Stockout rate, turnover rate, total inventory management, holding, and stockout costs |
| Deng et al. (2023) | Lagged demand, inventory, among others | DeepAR, N-BEATS, Prophet | WMAPE |
| Sagaert et al. (2018) | Lagged demand, macroeconomic indicators | LASSO Regression | MAPE |

**Table 1 Summary of related research papers.**

forecasting tasks due to their better data fitting capabilities. Particularly, the Makridakis (M-series) competition has been a key test-bed for evaluating different forecasting models, such as multilayer perceptrons, Bayesian neural networks, radial basis functions, generalized regression neural networks (also called kernel regression), K-nearest neighbor regression, CART regression trees, support vector regression, and Gaussian processes(Makridakis and Hibon 2000, Ahmed et al. 2010, Makridakis et al. 2018, 2021). LightGBM (Ke et al. 2017), which is an advanced tree-based model, is notable for its fast and efficient training and prediction, and was used by all of the top-50 performers in the M-5 competition (Makridakis et al. 2022). LightGBM's accuracy has been validated by several other research studies for predictive modeling (Deng et al. 2021, Bandara et al. 2020, Zhang et al. 2020). Motivated by these studies, results from M-5 competition and our own experiments, we adopted the LightGBM algorithm for our task.

Incorporating additional data into ML-based forecasting models has been shown beneficial to improve forecasting performance. For instance, Sagaert et al. (2018) use LASSO regression and a large number of macroeconomic indicators from Federal Reserve Economic Data (FRED) to enhance tactical forecasts. In supply chain, private data creates information asymmetry; lack of information sharing hinders abilities to adequately harmonize manufacturer's activities to align with customers (Simatupang and Sridharan 2002). Information shared by suppliers and customers can also improve accuracy of demand forecasting.

For example, Hartzel and Wood (2017) show that demand forecasts benefit heavily from point-of-sale reporting. Kurtuluş et al. (2012) show that such forecast (called 'collaborative forecast') can be helpful for customers as well as suppliers, depending on the contractual obligations of both parties. Under the Newsvendor model setting, Taylor and Xiao (2010) show that the manufacturer benefits from selling to a better-forecasting retailer if and only if the retailer is already a good forecaster. These studies guide us to use demand and inventory information reported by our supply chain partners as part of input to our forecasting model to further improve the forecasting performance.

In addition to selecting the appropriate ML-based forecasting algorithm and the necessary data input, there are also significant implementation challenges in an enterprise setting. Deng et al. (2023) outlined the implementation of a comprehensive omnichannel retail infrastructure by Alibaba, which was a 2022 Edelman Award finalist. The infrastructure integrates demand forecasting with inventory management and price optimization, driven by product recommendations. Their implementation leverages deep learning models like DeepAR (Salinas et al. 2017), Prophet (Taylor and Letham 2018), Wavenet (Oord et al. 2016), and N-BEATS (Oreshkin et al. 2019) to generate demand forecasts. Dodin et al. (2023) showcased a pragmatic application of LightGBM models in forecasting the demand of parts at Bombardier. Ferreira et al. (2016) utilized a regression tree-based model for demand forecasting in the pipeline for price optimization. Although these studies report ML-based implementation of demand forecasting models in companies, there are few detailed discussions on their project management, deployment pipeline, and continuous performance monitoring. Based on previous work (Zaharia et al. 2018, Curtland et al. 2022), this paper will share some generalizable lessons on management strategies for the enterprise-scale implementation of ML-based demand forecasting. We believe our *project management strategy* would be useful for the reader as such issues are non-trivial in practice.

## Problem Formulation and Methodology

We now describe our problem setting. Our work addresses the problem of predicting demand for a product $p$ in a specific country $c$ at time $t$. Given a dataset of historical demand data among others, our goal is to train a model that can forecast the demand for future time periods. The historical data includes information about the actual demand $y_{t,c,p}$ and a set of associated features $X_{t,c,p}$. These features represent various aspects of the time, market, and product, as well as lagged demand for up to 15 months prior to the forecasting month.

Complete model formulation is provided in the Appendix. Details of the model inputs are provided in Table 2 in the later section.

While the existing process had predominantly been utilizing statistical forecasting methods such as ETS and ARIMA, our focus was on machine learning models. In the process of model selection, we rigorously evaluated a range of algorithms including XGBoost, LightGBM, Prophet, ARIMAX, ETS, and multilayer perceptrons, utilizing the Python `Darts` library for a unified and methodologically consistent comparative framework (Herzen et al. 2022). Empirical evaluations, emphasizing predictive accuracy and computational efficiency, revealed the superiority of tree-based models. These models excel in constructing predictive frameworks from training data through recursive bifurcation based on feature values, outperforming others in handling multiple products and countries' demand forecasting due to their interpretability, non-linear relationship modeling, and complex data structure delineation.

Utimately, LightGBM, a gradient boosting model developed by Microsoft in 2017 (Ke et al. 2017), stood out, particularly for its forecast precision and swift training. The straightforward optimization of LightGBM parameters, like learning rate and iteration count, simplifies generalizing across diverse datasets. And, the reduced memory footprint and expedited training time relative to other tree-based models are essential for our scale. This rationale, supported by data, affirmed our choice of LightGBM, given its proven effectiveness with datasets similar to ours in structure and complexity.

**Iterative Forecasting Algorithm**

Our *Iterative Forecasting Algorithm* is outlined in Algorithm 1 in the Appendix. The algorithm employs the LightGBM model as its core predictive engine, although it's adaptable to other algorithms. It's designed to forecast demand iteratively over a time window $T$, which allows for dynamically updating forecasts. The model begins by preprocessing the data, which includes data cleaning and feature engineering. Optimal hyperparameters are identified using `Hyperopt` parameter tuning for the LightGBM model by considering the last month's information for validation.

Our approach allows us to capture both the seasonality and trends in the demand while benefiting from the efficiency and scalability of LightGBM. Moreover, the iterative nature of this algorithm allows for frequent model updating, leveraging the most recent one-month data for cross-validation. This ensures that the model stays responsive to any significant changes in the underlying data patterns. Storing the serialized model in MLFlow, we are able to ensure repeatability and continuity for future efforts, detailed in *Project Management Principles* section below.

**Model Input Features**

Our machine learning-based models for demand forecasting distinguish themselves from conventional time series models by their ability to incorporate a diverse set of features including categorical, numeric, and more. These features are carefully selected to not only capture historical demand data but also to offer insights into the multifaceted nature of demand generation and fulfillment. A summary of these features is provided in Table 2 for ease of reference.

*Types of Features*

1. *Lag Demands:* Demand from the previous $m$ months are factored in, with $m = 15$ for products with intermittent demand and annual buying cycles.

2. *Rolling Demand Features:* These are statistical measures—mean, coefficient of variation, and outlier counts—computed over rolling windows of 3, 6, and 12 months, capturing both recency and variability in demand.

3. *Product and Geography-based Statistics:* Summary statistics are categorized by product and geography to model unique trends and attributes within these dimensions.

4. *Seasonal Fluctuations:* Binary indicators for each fiscal quarter are included to capture seasonal demand patterns. Additionally, a monthly integer representing month of the quarter is also included.

5. *Product Life Cycle (PLC):* Calculated as $(M - m)/M$, where $M$ is the total expected lifetime of product, and $m$ is the current forecasting month, this feature considers a product's remaining lifespan, enriching the model's temporal context. Typically, products introduced to the market experience a surge in demand initially, attributable to their innovative features and promotional efforts, followed by a gradual decline in sales as they progress through their product life cycle.

6. *Channel Metrics:* Features such as 'Channel Partner Inventory' and 'Sell-through' provide a nuanced understanding of real-time market demand and potential future orders with direct inputs from our distribution channel partners (customers in B2B setting). Channel partner inventory refers to the SKU-level inventory that our channel partners report monthly, while Sell-through represents the sales by our partners to their customers.

| **Feature Name** | **Description** | **Granularity** | **Utility for Forecasting** |
|---|---|---|---|
| Lagged Demand | Size of demand from previous m months, m varies per product group | Month ($t$) | Captures influence of past on future trends |
| Rolling Demand Features | Statistics of demands within n-month rolling window (mean, coefficient of variation, outliers) | Month ($t$) | Assesses recent trend, variability |
| Product-based Statistics | Mean and coefficient of variation of lagged demand and rolling features, per product category | SKU ($p$) | Specific trends in product categories |
| Geography-based Statistics | Mean and coefficient of variation of lagged demand and rolling features, per country | Country ($c$) | Location-specific trends |
| Seasonal Fluctuation | Binary indicator for each fiscal quarter and integer month within a quarter | Month ($t$) | Captures seasonal effects |
| Product Life Cycle | Proportion of product life cycle left, calculated as $(M-m)/M$ | SKU, Country ($p,c$) | Stage of the product in its life cycle |
| Channel Partner Inventory | Inventory reported by distribution channel partners | SKU, Country, Month ($p,c,t$) | Indicates potential reordering |
| Sell-through | Sales to distribution channel partners | SKU, Country, Month ($p,c,t$) | Reflection of down-stream demand (to channel partners' customers) |

**Table 2** **Input features to the forecasting model, encompassing over 100 variables after incorporating all calculated statistics.**

*Feature Selection* We considered two algorithms for our feature selection strategy: the Fast AI method based on Howard (2019) and the Quadratic Programming Feature Selection (QPFS) technique as proposed by Rodriguez-Lujan et al. (2010). Fast AI's approach involves generating a correlation matrix followed by a dendrogram of all features. This guides the systematic pruning of correlated features, thus honing the feature set down to those that are most informative. QPFS, on the other hand, uses quadratic programming to balance feature importance against redundancy. From our comparative analysis between these methods, we discovered that QPFS produced high variance in each cycle's feature importance results, while Fast AI method led to a stable set of features. Given this, we chose the Fast AI method for our production code.

## Project Management Principles

To enhance value creation and streamline the entire ML project lifecycle, data scientists and managers utilize several DevOps concepts (Chen et al. 2020). Large-scale projects with numerous collaborators and users necessitate robust coordination and maintenance tools. As outlined by John et al. (2021), the MLOps framework proves indispensable for tracking of data for ML development, validation of ML models, release of ML models, and storage of serialized models for replication and future applications.

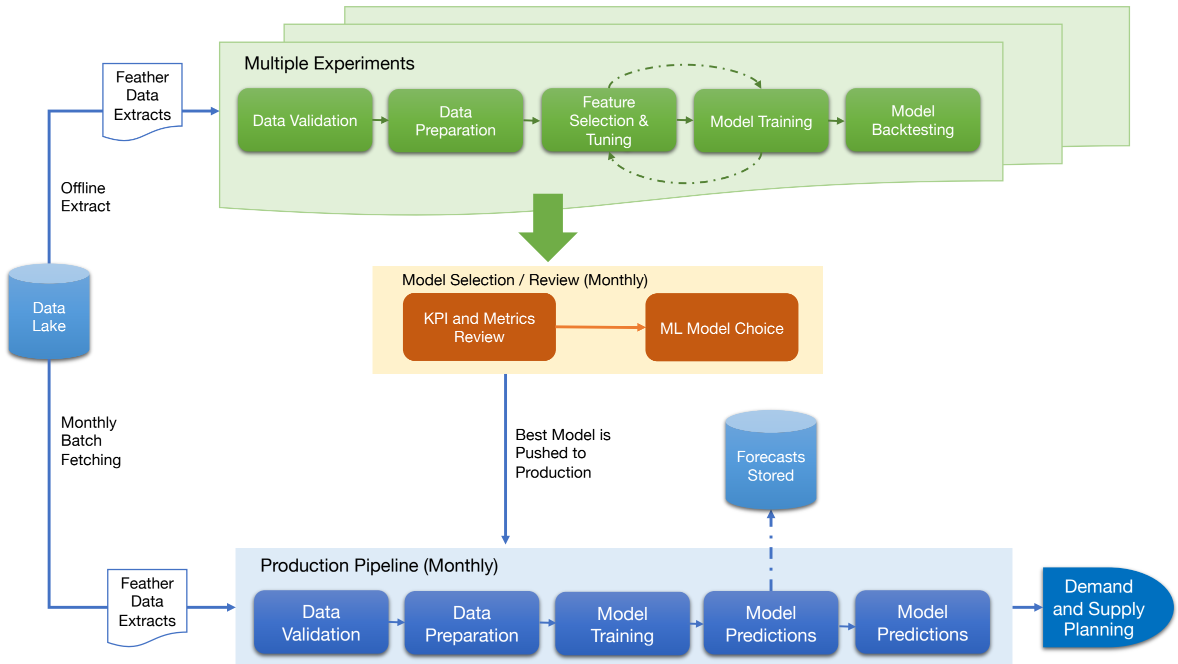


**Figure 2** **Project management for continuous deployment pipeline.**

The majority of machine learning enhancements are driven by *experimentation*. This involves exploring multiple datasets, variable transformations, model architectures, software libraries, and more. These experiments not only have diverse inputs and outputs but must also be efficiently timed. Given the reliance of model performance on input data and training, *reproducibility* becomes paramount. In our project, before each month starts, one model gets selected for *deployment* and producing ML forecasts to support operational decisions. Yet, experimentation persists to further refine our models for future months. We show our project management strategy through a flowchart in Figure 2.

We adopted various open-source tools in our project management strategy:

1. **Experimentation and Reproducibility:** `MLflow`, an open-source ML platform Zaharia et al. (2018), tackles challenges linked to experimentation, reproducibility, and deployment. It provides extensive experiment tracking, covering parameters, metrics, code, input data, and more, accessible through an API and an interactive dashboard. We opted for `MLflow` due to its self-hosting capabilities, which streamlined our workflow, at no additional cost to HP.

2. **Documenting Results:** Jupyter Notebooks aid reproducibility, allowing detailed annotations on processes, inputs, and outputs using markdown cells. These notebooks can

be parameterized, turning their execution into function calls with the `papermill` library, a tool that enables operating one notebook from another notebook similar to a function call. For recurring tasks like monthly time-series forecasting, parameterized notebooks paired with the *Script Monkey* technique prove invaluable to rapid pilot-phase coding efforts. We added a keyword like "Monkey," allowing quick navigation for necessary adjustments before re-running scripts. When the operations were pushed to production, these efforts were automated through parametrized notebooks.

3. **Model Serialization:** Once we move experimental models to production, serializing and storing them for future reference becomes essential, which is where `MLflow` becomes indispensable. The models can be used later for warm-starting future training, comparing accuracy, and results can be reproduces when necessary.

4. **Data Storage:** Data storage demanded substantial disk space. Initially, we used Python's Pickle for data snapshots. However, due to fundamental issues with Pickle, such as corruption from version changes and ballooning file sizes, we transitioned to Apache Feather. Feather boasts a powerful compression algorithm resulting in vastly reduced file sizes compared to CSV, and native compatibility with tools like `pandas`. Crucially, Feather maintains forward and backward compatibility, ensuring hassle-free file accessibility across versions.

5. **Rapid Testing with `FLAML`:** Our expansive data set made it impractical to run full experiments each time. Thus, preliminary assessments were vital. We relied on `FLAML`'s efficient evaluation mechanisms, leveraging its automatic Bayesian hyperparameter search and cross-validation. This allowed for targeted improvements within our time budgets, ensuring only the most promising strategies proceeded to in-depth testing. Such rapid tests were foundational; more exhaustive experimentation followed once a direction was determined, culminating in integrating findings into our primary model, as presented in Algorithm 1.

## Performance Evaluation and Empirical Validation

Ultimately, the accuracy of the new algorithm dictates the final usage of the newly minted ML forecasting pipeline. We validate the performance of ML-based forecasts and compare it to the existing *Statistical* and *Consensus* forecasts. This empirical validation serves two main purposes. First, prior to enterprise-scale deployment for each product in every geography, we must demonstrate that the accuracy obtained by the ML-based pipeline is acceptable, and equal or more reliable than the existing methods. Second, we must also evaluate the judicious use of the additional project management machinery (see Figure 2) as these require significant investment. When the first purpose is achieved, we will be able to justify the use of additional resources.

**Evaluation Metrics**

Our evaluation incorporates three key metrics: bias, weighted mean absolute percentage error (wMAPE), and root mean squared error (RMSE) which are defined in the Appendix. RMSE is symmetric and continuously differentiable, thus the preferred choice for ML model training. Bias and wMAPE are the key performance indicators (KPIs) used by planners and managers for evaluation due to their ease of interpretation and focused action. For additional insights, we also direct the interested readers to Hyndman and Koehler (2006) for a useful discussion and side-by-side comparison of various accuracy metrics, as applied to M-3 forecasting.

RMSE is used for ML model training due to its mathematical properties, striking a balance by being sensitive to larger errors and scale-dependent. RMSE's sensitivity to outliers is a double-edged sword: it helps in emphasizing significant errors, which can be beneficial in certain contexts, but it can also be a drawback as it may skew the model's performance evaluation in datasets with extreme values. Additionally, RMSE's advantages include being differentiable, symmetric, and its widespread use in theoretical statistical modeling.

These metrics are calculated on a cumulative basis, i.e., the error is calculated for the next $k$ months. $k$-month cumulative actuals for the month $t$ would be calculated as $\sum_{i=t}^{t+k-1} y_i$. Similarly, cumulative forecasts are $\sum_{i=t}^{t+k-1} \hat{y}_i$. For example, if we are forecasting in January, the three month cumulative forecast ($k = 3$) will be the sum of forecasts in January, February, and March. The appropriate cumulative forecast horizons depend on the specific supply chain lengths and decisions for which the forecast will be used.

Measuring and improving the forecast over different lead times is important due to practical business reasons. Supply chains have lead times associated with them for building and shipping new products. Typically, some inventory is held close to the customer to support the expected variability in demand over lead time. For a three month lead time supply chain it is important to measure and reduce the error in the three month cumulative forecast in order to reduce the inventory needed to maintain high product availability.

| Model | CM1 | | | CM3 | | | CM6 | | |
|---|---|---|---|---|---|---|---|---|---|
| Metric | Bias | RMSE | wMAPE | Bias | RMSE | wMAPE | Bias | RMSE | wMAPE |
| Value (SD) | -3.08% | 13.09 | 15.92% | -1.08% | 32.76 | 9.25% | 2.42% | 57.29 | 9.08% |
| | (7.05%) | (3.38) | (5.62%) | (4.68%) | (6.66) | (3.28%) | (3.96%) | (11.51) | (2.64%) |
| | 1.17% | 11.87 | 12.33% | 1.25% | 31.03 | 5.25% | 3.75% | 60.28 | 5.08% |
| | (8.92%) | (4.87) | (6.69%) | (7.34%) | (9.43) | (4.39%) | (5.26%) | (16.83) | (2.91%) |
| | 2.67% | 13.71 | 16.75% | 1.08% | 34.55 | 9.33% | 2.08% | 62.47 | 9.17% |
| | (10.14%) | (2.89) | (4.99%) | (5.00%) | (6.03) | (3.42%) | (5.38%) | (9.08) | (1.90%) |

**Table 3 ML Forecasting accuracy metrics (bias, RMSE, and wMAPE) for cumulative forecast horizons (CM1, CM3, CM6) with Mean (Standard Deviation).**

Further, an optimized supply chain will pool inventory and size factory capacity based on the appropriate lead time and forecast performance.

**Forecast Accuracy Evaluations**

We first present a comparative visualisation of forecasting performance for a select business segment (1,484 products) from all three methods: consensus (ConsFcst), machine learning (MLFcst), and statistical (StatFcst), evaluated at cumulative forecast horizons of one (CM1), three (CM3), and six (CM6) months. Although the scales have been adjusted for anonimity, the actual trends remain the same. Results from all product lines are not presented due to data sensitivity, and accuracy results vary across business segments.

We utilize three key metrics — Bias, wMAPE, and RMSE — to gauge the performance of each forecasting method, offering a multi-dimensional view of their strengths and weaknesses. A summary of accuracy results are provided in Table 3. These metrics are also presented as a dumbbell plot in Figure 3 with center point as 12 month averages and whiskers as one standard deviation in either direction. Additionally, Figure 4 visually represents these metrics over all 12 months, further elucidating the monthly accuracy trends of the corresponding methods. Finally, a statistical test of comparing the difference between the metrics over 12 months using paired t-test is presented in Table 4.

The ML forecast method demonstrates considerable strengths in its forecasting accuracy as compared to the statistical method, particularly in the metrics of wMAPE and RMSE. We observe that wMAPE for ML forecast is better than the other two in all three cumulative periods. In fact, at CM3 and CM6, that is for longer range forecasts, our model has wMAPE almost half of the other two methods. When looking at statistically significant differences, we find statistically significant difference between ML and STAT models with positive t-statistic and p-values less than 0.05. These strongly suggest the statistical superiority of

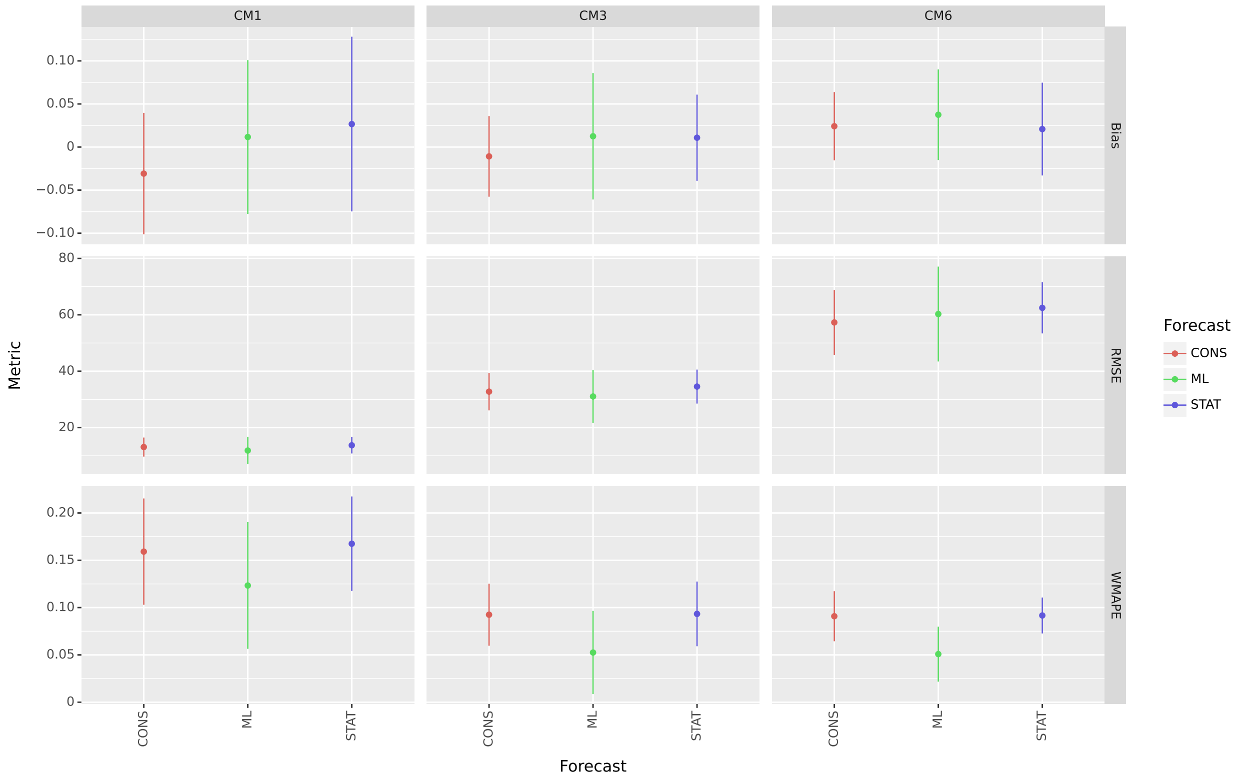


**Figure 3** **Dumbbell plot visualizing the mean (center point) and one standard deviation (vertical lines) of Bias, RMSE, and WMAPE for three forecasting methods (Consensus, Machine Learning, and Statistical) over cumulative forecast horizons of one month (CM1), three months (CM3), and six months (CM6).**

*Note:* Each color represents a different forecasting method, illustrating the variability and central tendency of the forecast accuracy metrics across different periods.

| Cumulative | Comparison | Bias | RMSE | WMAPE |
|---|---|---|---|---|
| CM1 | CONS vs ML | -1.295 (0.209) | 0.716 (0.482) | 1.421 (0.169) |
| | STAT vs ML | 0.385 (0.704) | 1.128 (0.272) | 1.832 (0.080) |
| CM3 | CONS vs ML | -0.929 (0.363) | 0.518 (0.610) | 2.528 (0.019) |
| | STAT vs ML | -0.065 (0.949) | 1.089 (0.288) | 2.541 (0.019) |
| CM6 | CONS vs ML | -0.701 (0.490) | -0.507 (0.617) | 3.526 (0.002) |
| | STAT vs ML | -0.767 (0.451) | 0.399 (0.694) | 4.074 (0.001) |

**Table 4** **Forecasting Accuracy Metrics: Bias, wMAPE, RMSE Comparison for CONS, ML, and STAT Methods.**

*Note:* The accompanying table presents t-statistics and p-values (in brackets) for an in-depth assessment across various cumulative forecast horizons.

the ML forecast in wMAPE. It also attests to the model's alignment to business objectives; wMAPE is commonly used to track business objectives at HP.

The higher accuracy of ML model in wMAPE is particularly surprising since it was trained with RMSE as the loss function. In the case of RMSE, which is sensitive to large

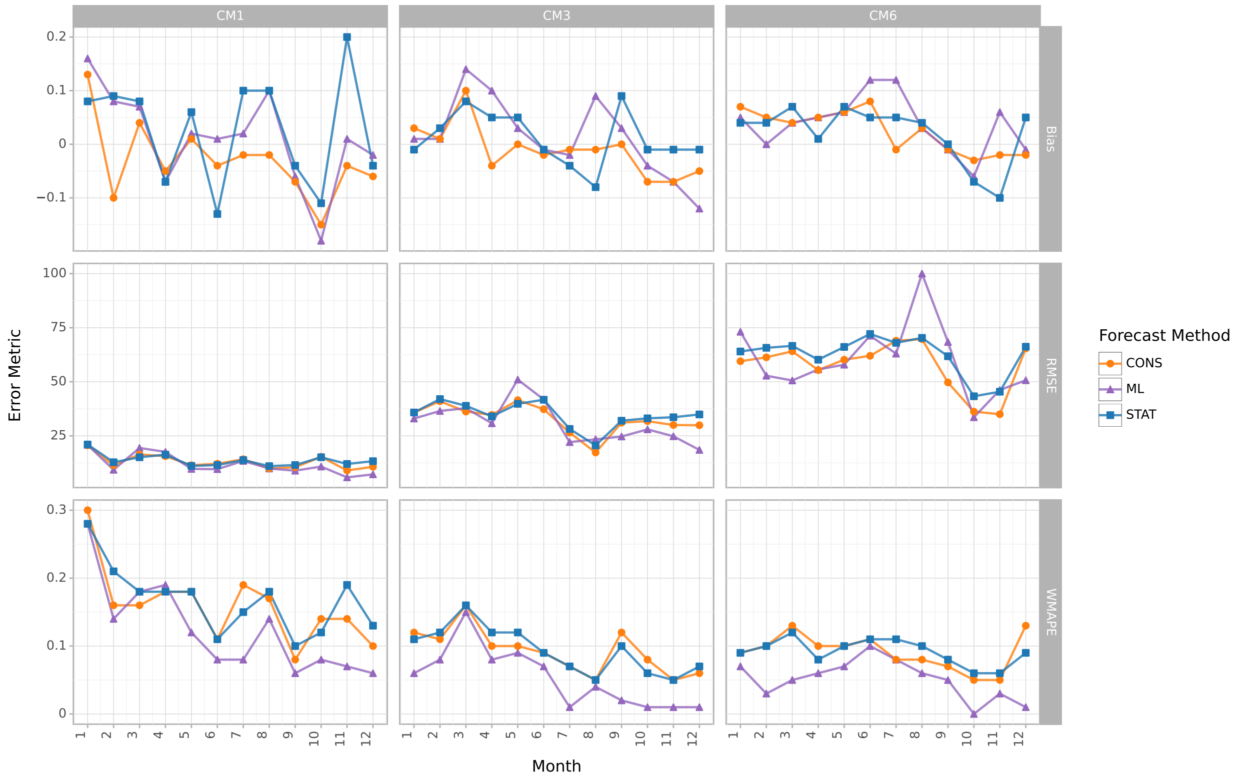


**Figure 4 Bias, WMAPE and RMSE metrics over 12 months show that ML model is consistently among the top performer of the three models.**

*Note:* CM1 is point forecast, while CM3 and CM6 are three and six months cumulative forecasts, respectively.

forecast errors, the ML forecast again proves to be more adept than others, though not statistically significant.

However, the ML forecast does not consistently dominate across all metrics and comparisons. When considering Bias, which reflects the systematic error in forecasts (either as overestimation or underestimation), the ML method does not exhibit a statistically significant difference from the statistical or consensus forecasts in any of the cumulative periods (CM1, CM3, and CM6), as evidenced by p-values greater than 0.05. Our model bias are also higher than the Consensus and Statistical model results. We observed that ML model had a strong tendency to underforecast, especially in the longer time horizons.

These outcomes suggests that in certain scenarios, especially those involving longer-term predictions, the aggregated judgments represented by the consensus forecast can yield more accurate results than the current ML modeling method. This contrast highlights the ML forecast's relative strengths in specific areas, enabling the modeling team to focus on improvements and the business team to choose the best performing model for each product-geo and time horizon combination. The data presented in Table 4 and the trends observed in Figure 4 collectively bolster the case for the ML model's adoption, in addition to STAT and CONS models at HP as part of an integrated effort to improve overall forecasting performance.

### Dashboard of Results

As an essential advancement in disseminating forecasting analytics, the incorporation of Analytical Dashboards serves to share results to a wider audience, including planners and decision-makers. This powerful tool does not only exhibit the performance of varying models, but it also provides an avenue to scrutinize their historical accuracies and pertinent details. Constructed with customizable Key Performance Indicators (KPIs), the dashboard extends the capacity to inspect product hierarchies from different lenses, thereby promoting informed business strategies and policies.

We share Bias, wMAPE and RMSE as metrics to compare historical performance of algorithms. For the planners to choose the best model, we create a heatmap of best forecast as measured by wMAPE. The heatmap covers all HP Print product categories by time period, which enables planners to visualise relative performance of different methods over time.

## Lessons Learned and Business Implications

Implementing a robust machine learning (ML) based demand forecasting system at a global scale highlighted several key lessons and challenges, crucial for business leaders in similar domains. Work on the project started in 2019 and remained in pilot-phase for a year. In 2020, the results were published in the standard KPI dashboards and available for manual use within the Statistical and Consensus forecasting modeling processes depicted in Figure 1. In summer of 2023, SKU-level forecasts for every Print product by geography were systematically loaded into the data pipeline for use in the business forecasting process. Inclusion of the analytical forecast in the business KPI dashboards led to wide-scale adoption of our work.

While previous efforts to implement ML-based models were scattered and unsuccessful, our solution was implemented at scale due to its inclusive approach. This was accomplished both because our model choice performed well across the print portfolio, but also because we partnered with the business team to preemptively address management of change challenges that plague other large-scale efforts. The adoption of LightGBM models for their capability to handle large datasets and complex patterns was crucial. LightGBM demonstrated remarkable adaptability in handling market fluctuations, including during the pandemic. This adaptability to market dynamics was key in ensuring the model's effectiveness in dynamic, supply-constrained environments.

A unified, holistic approach in which only one model architecture with various data sources was used proved essential in overcoming the complexities of large-scale forecasting. Incorporating our signal into the standard business process in conjunction with effective dashboard visualizations and KPIs enabled successful implementation. The importance of MLOps, achieved through MLflow, in effective management of models became apparent to us. It was instrumental in streamlining ML development, facilitating the tracking of experiments, packaging code, and ensuring the use of the most current and precise models.

An important aspect of our solution is that our final forecast is not confined to being either human- or machine-produced. With our human-in-the-loop architecture, the human forecaster exercises judgment about when and how to apply layers on top of the modeled output. Over time, we expect our solution to improve by supporting the human with prescriptive drift and anomaly detection, along with AI-enhanced dashboards. This will build upon the existing explainability and causality capabilities of the solution, creating better insight generation and enhancing model itself.

Ensuring reproducibility was also a significant challenge, addressed by using MLflow for maintaining serialized models and Jupyter Notebooks, particularly parametrized with `papermill`. This enabled linking analysis to outputs and ensured comprehensive documentation. Due to the scale of implementation, computational resource optimization became necessary. This involved employing high-performance workstations and adopting efficient data storage and retrieval methods, like the Apache Feather format, which provided significant improvements in data handling and processing efficiency.

The quality of data plays a crucial role in the effectiveness of machine learning (ML) models. Because HP executed a digital transformation in parallel with our modelling efforts, we faced several real-world data challenges in our project. These included addressing missing data, reconciling unlinked datasets, managing irregular database updates, and navigating limited documentation. We also had to consider the proper handling of historical data. Another significant aspect was the integration of 'soft data', which required manual analysis.

Successfully overcoming these hurdles was key to improving the accuracy and reliability of our forecasting models. Our solutions also helped seed the solutions employed in HP's digital transformation.

In summary, our experience at HP underscores the importance of a well-integrated, adaptive ML-based approach in demand forecasting. Addressing these challenges was pivotal in optimizing the ML models for supply chain management, leading to more efficient decision-making and operational management. The insights gleaned offer a valuable template for business leaders facing similar challenges in large-scale demand forecasting. Further, our collaborative, agile development model is expected to deliver improved accuracy as our backlog of additional modelling ideas are implemented.

## Concluding Remarks

In this paper, we detailed our implementation of ML-based demand forecasting system at HP, implemented for all print products worldwide. We demonstrated the successful application of LightGBM models, integrated with MLOps for consistent model maintenance and tracking. Our approach leveraged a unified framework, ensuring reproducibility and adaptability in our forecasting methods. The key insights from this endeavor highlight the importance of computational resources, robust data management, and a proactive stance towards market fluctuations and data quality.

Our experience offers a template for business leaders facing similar challenges in demand forecasting. The lessons learned underscore the value of a comprehensive, adaptable, and rigorously managed ML-based forecasting approach. This not only facilitates more accurate and efficient demand predictions but also fosters an agile and responsive business environment.

# Appendix
## Problem Formulation

We formulate the problem as a supervised learning task where we aim to minimize the forecasting loss over the dataset $D$, consisting of pairs of input features $X_{t,c,p}$ and corresponding demands values $y_{t+1,c,p}$. That is,

$$D = \{(X_{t,c,p}, y_{t+1,c,p}) : \forall c, p, t_{first} \leq t < t_{now}\}, \tag{1}$$

where $t_{first}$ is the first period when we have enough observations to create all features, especially the lagged features. The training process minimizes the forecasting loss (RMSE):

$$\ell(f|D) = \sqrt{\mathbb{E}_{X,y \in D} \left(f(X) - y\right)^2}, \tag{2}$$

in addition to necessary regularization terms.

In this context, our model $f(\cdot)$ learns to predict future demand based on the input features. Once trained, the model can be applied to forecast demand for future time periods $t \geq t_{now}$.

We use $F_{t,c,p} \in \mathbb{R}^T$ to represent forecasts for $T$ periods starting with $t_{now}$:

$$F^T_{t,c,p} = (\hat{y}_{t+1,c,p}, \cdots, \hat{y}_{t+T,c,p}). \tag{3}$$

## Iterative Forecasting Algorithm

Here we describe our iterative forecasting algorithm. For each time step $t_\alpha$, the algorithm constructs a training dataset $D_\alpha$ using all available data up to that point in time. Identified hyperparameters are used with $D_\alpha$ to train the LightGBM model $f(\cdot)$, which is optimized to minimize the Root Mean Squared Error (RMSE). Once trained, the model generates $T$ future forecasts for each time step $t_\alpha$. The LightGBM model is then either incrementally updated (i.e. warm started from best results from last month) or retrained from scratch, providing flexibility in handling significant changes in underlying data distribution.

## Evaluation Metrics

Let $n$ is the number of data points, $y_i$ represents the actual value, $\hat{y}_i$ represents the predicted value.

**Bias** measures the weighted percentage error in forecasts, signified by a positive or negative value indicating over or underforecasting, respectively. Bias is calculated using the formula:

$$Bias = \sum_{i=1}^{n} \frac{y_i - \hat{y}_i}{y_i}.$$

**wMAPE** represents the weighted mean of absolute percentage errors, a metric easily understood even by non-technical stakeholders as percentage deviation from actuals. It is expressed as:

$$wMAPE = \frac{\sum_{i=1}^{n} |\hat{y}_i - y_i|}{\sum_{i=1}^{n} y_i}.$$

**RMSE** (Root Mean Squared Error) is defined as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}}.$$

**Algorithm 1** Enhanced training and forecasting algorithm with LightGBM

**Preprocess the data:** Data cleaning and feature engineering.

**Determine optimal hyperparameters:** Use grid search or random search for the LightGBM model.

**Initialize forecast horizon** $T$ (e.g., 7).

**for** $t_\alpha$ in $(t_{\text{first}} : t_{\text{now}})$ **do**

  Create the training data:

$$D_\alpha = \{(X_{t,c,p}, y_{t,c,p}) : \forall c, p, t_{\text{first}} \leq t \leq t_\alpha\}$$

  Perform time-series cross-validation on $D_\alpha$ and train the LightGBM model $f(\cdot)$ with optimal hyperparameters, minimizing loss (RMSE):

$$\ell(f|D) = \sqrt{\mathbb{E}_{X,y\in D}\left(f(X) - y\right)^2}$$

  With the fitted model, create $T$ forecasts for $t_\alpha + 1$ to $t_\alpha + T$:

$$F^T_{t_\alpha,c,p} = \Big(f(\hat{X}_{t_\alpha+1,c,p}), f(\hat{X}_{t_\alpha+2,c,p}), \cdots, f(\hat{X}_{t_\alpha+T,c,p})\Big)$$

  Update the LightGBM model incrementally by warm starting from last month's best results if possible, or retrain it from scratch.

**end for**

**Perform Backtesting:** Apply the trained model to a historical dataset $D_{\text{historical}}$ to simulate past predictions. Evaluate its performance using appropriate metrics (e.g., RMSE, MAE).

**Store Forecasts:** Save the generated forecasts $F^T_{t_\alpha,c,p}$ to a dedicated database or file storage for future evaluation, comparison, or direct usage.

**Log Model:** Serialize the LightGBM model, hyperparameters, and performance metrics for future reference or retraining using MLFlow.

## References


Ahmed NK, Atiya AF, Gayar NE, El-Shishiny H (2010) An empirical comparison of machine learning models for time series forecasting. *Econometric Reviews* 29(5–6):594–621, ISSN 0747-4938, URL `http://dx.doi.org/10.1080/07474938.2010.481556`.

Armstrong JS (2001) Combining forecasts. *Principles of forecasting*, 417–439 (Springer).

Bandara K, Bergmeir C, Smyl S (2020) Forecasting across time series databases using recurrent neural networks on groups of similar series: A clustering approach. *Expert systems with applications* 140:112896.

Billington C, Callioni G, Crane B, Ruark JD, Rapp JU, White T, Willems SP (2004) Accelerating the profitability of hewlett-packard's supply chains. *Interfaces* 34(1):59–72.

Cargille B, Branvold D (2000) Diffusing supply chain innovations at hewlett-packard company: Applications of performance technology. *Performance Improvement Quarterly* 13(4):6–15.

Chen A, Chow A, Davidson A, DCunha A, Ghodsi A, Hong SA, Konwinski A, Mewald C, Murching S, Nykodym T, et al. (2020) Developments in mlflow: A system to accelerate the machine learning lifecycle. *Proceedings of the fourth international workshop on data management for end-to-end machine learning*, 1–4.

Cortes C, Jackel LD, Chiang WP (1994) Limits on learning machine accuracy imposed by data quality. *Advances in Neural Information Processing Systems* 7.

Curtland C, Neto P, Ghozeil A (2022) Hp inc.. advanced analytics powers technology in the service of humanity. URL `https://pubsonline.informs.org/do/10.1287/orms.2022.02.18/full/`.

Deng T, Zhao Y, Wang S, Yu H (2021) Sales forecasting based on lightgbm. *2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE)*, 383–386 (IEEE).

Deng Y, Zhang X, Wang T, Wang L, Zhang Y, Wang X, Zhao S, Qi Y, Yang G, Peng X (2023) Alibaba realizes millions in cost savings through integrated demand forecasting, inventory management, price optimization, and product recommendations. *INFORMS Journal on Applied Analytics* 53(1):32–46.

Dodin P, Xiao J, Adulyasak Y, Alamdari NE, Gauthier L, Grangier P, Lemaitre P, Hamilton WL (2023) Bombardier aftermarket demand forecast with machine learning. *INFORMS Journal on Applied Analytics* .

Ferreira KJ, Lee BHA, Simchi-Levi D (2016) Analytics for an online retailer: Demand forecasting and price optimization. *Manufacturing & service operations management* 18(1):69–88.

Fildes R, Goodwin P, Lawrence M, Nikolopoulos K (2009) Effective forecasting and judgmental adjustments: an empirical evaluation and strategies for improvement in supply-chain planning. *International journal of forecasting* 25(1):3–23.

Fildes R, Ma S, Kolassa S (2022) Retail forecasting: Research and practice. *International Journal of Forecasting* 38(4):1283–1318.

Gardner ES (1990) Evaluating forecast performance in an inventory control system. *Management science* 36(4):490–499.

Hartzel KS, Wood CA (2017) Factors that affect the improvement of demand forecast accuracy through point-of-sale reporting. *European Journal of Operational Research* 260(1):171–182.

Herzen J, Lässig F, Piazzetta SG, Neuer T, Tafti L, Raille G, Pottelbergh TV, Pasieka M, Skrodzki A, Huguenin N, Dumonal M, Kościsz J, Bader D, Gusset F, Benheddi M, Williamson C, Kosinski M, Petrik M, Grosch G (2022) Darts: User-friendly modern machine learning for time series. *Journal of Machine Learning Research* 23(124):1–6, URL `http://jmlr.org/papers/v23/21-1177.html`.

Howard J (2019) Practical deep learning. Online Course.

Hyndman RJ, Athanasopoulos G (2018) *Forecasting: principles and practice* (OTexts).

Hyndman RJ, Koehler AB (2006) Another look at measures of forecast accuracy. *International journal of forecasting* 22(4):679–688.

John MM, Olsson HH, Bosch J (2021) Towards mlops: A framework and maturity model. *2021 47th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*, 1–8 (IEEE).

Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, Liu TY (2017) Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems* 30:3146–3154.

Kurtuluş M, Ülkü S, Toktay BL (2012) The value of collaborative forecasting in supply chains. *Manufacturing & Service Operations Management* 14(1):82–98.

Laval C, Feyhl M, Kakouros S (2005) Hewlett-packard combined or and expert knowledge to design its supply chains. *Interfaces* 35(3):238–247.

Lawrence MJ, Edmundson RH, O'Connor MJ (1986) The accuracy of combining judgemental and statistical forecasts. *Management Science* 32(12):1521–1532.

Lee HL (2002) Aligning supply chain strategies with product uncertainties. *California management review* 44(3):105–119.

Makridakis S, Hibon M (2000) The m3-competition: results, conclusions and implications. *International journal of forecasting* 16(4):451–476.

Makridakis S, Spiliotis E, Assimakopoulos V (2018) Statistical and machine learning forecasting methods: Concerns and ways forward. *PLoS ONE* 13.

Makridakis S, Spiliotis E, Assimakopoulos V (2021) The m5 competition: Background, organization, and implementation. *International Journal of Forecasting* .

Makridakis S, Spiliotis E, Assimakopoulos V (2022) M5 accuracy competition: Results, findings, and conclusions. *International Journal of Forecasting* 38(4):1346–1364.

Manary MP, Wieland B, Willems SP, Kempf KG (2019) Analytics makes inventory planning a lights-out activity at intel corporation. *INFORMS Journal on Applied Analytics* 49(1):52–63.

Oord Avd, Dieleman S, Zen H, Simonyan K, Vinyals O, Graves A, Kalchbrenner N, Senior A, Kavukcuoglu K (2016) Wavenet: A generative model for raw audio. *arXiv preprint arXiv:1609.03499* .

Oreshkin BN, Carpov D, Chapados N, Bengio Y (2019) N-beats: Neural basis expansion analysis for interpretable time series forecasting. *arXiv preprint arXiv:1905.10437* .

Petropoulos F, Kourentzes N, Nikolopoulos K, Siemsen E (2018) Judgmental selection of forecasting models. *Journal of Operations Management* 60:34–46.

Petropoulos F, Siemsen E (2023) Forecast selection and representativeness. *Management Science* 69(5):2672–2690.

Qi M, Shi Y, Qi Y, Ma C, Yuan R, Wu D, Shen ZJ (2023) A practical end-to-end inventory management model with deep learning. *Management Science* 69(2):759–773.

Ritzman LP, King BE (1993) The relative significance of forecast errors in multistage manufacturing. *Journal of Operations Management* 11(1):51–65.

Rodriguez-Lujan I, Elkan C, Santa Cruz Fernández C, Huerta R, et al. (2010) Quadratic programming feature selection. *Journal of Machine Learning Research* .

Sagaert YR, Aghezzaf EH, Kourentzes N, Desmet B (2018) Temporal big data for tactical sales forecasting in the tire industry. *Interfaces* 48(2):121–129.

Salinas D, Flunkert V, Gasthaus J (2017) Deepar: Probabilistic forecasting with autoregressive recurrent networks. arxiv 2017. *arXiv preprint arXiv:1704.04110* .

Seifert M, Siemsen E, Hadida AL, Eisingerich AB (2015) Effective judgmental forecasting in the context of fashion products. *Journal of Operations Management* 36:33–45.

Simatupang TM, Sridharan R (2002) The collaborative supply chain. *The international journal of logistics management* 13(1):15–30.

Simatupang TM, Sridharan R (2005) An integrative framework for supply chain collaboration. *The international Journal of Logistics management* 16(2):257–274.

Taylor SJ, Letham B (2018) Forecasting at scale. *The American Statistician* 72(1):37–45.

Taylor TA, Xiao W (2010) Does a manufacturer benefit from selling to a better-forecasting retailer? *Management Science* 56(9):1584–1598.

Ward J, Zhang B, Jain S, Fry C, Olavson T, Mishal H, Amaral J, Beyer D, Brecht A, Cargille B, et al. (2010) Hp transforms product portfolio management with operations research. *Interfaces* 40(1):17–32.

Zaharia M, Chen A, Davidson A, Ghodsi A, Hong SA, Konwinski A, Murching S, Nykodym T, Ogilvie P, Parkhe M, et al. (2018) Accelerating the machine learning lifecycle with mlflow. *IEEE Data Eng. Bull.* 41(4):39–45.

Zhang Y, Zhu C, Wang Q (2020) Lightgbm-based model for metro passenger volume forecasting. *IET Intelligent Transport Systems* 14(13):1815–1823.

## About the authors

**Harshvardhan** is a PhD Candidate at the Haslam College of Business at the University of Tennessee, Knoxville, advised by Dr. Chuanren Liu. He is broadly interested in developing and deploying machine learning algorithms for useful applications. During his PhD, he also worked with Strategic Planning and Modeling (SPaM) team at HP for over a year in bringing this project to action. Harsh earned his BA and MBA at the Indian Institute of Management Indore.

**Cara Curtland** is a Supply Chain Data Science Strategist in the Strategic Planning & Modeling (SPaM) team at HP Inc. As a 28-year HP veteran, Cara has held roles in manufacturing, writing systems, and supply chain operations within a wide variety of region, business unit, and global organizations. Her experience spans manufacturing, R&D, planning, forecasting, supply chain design, complexity management, and inventory & cash flow optimization. Cara earned B.S. and M.S. degrees in Industrial Engineering from Purdue University.

**Jerry Hwang** is a Data Scientist with over 20 years of experience in supply chain analytics. He has contributed to demand forecasting initiatives at HP and Uber. He has also worked on projects encompassing procurement risk management, inventory optimization, and supply chain network design. He holds a MSc degree in Management Science and Engineering from Stanford University.

**Chuanren Liu** is an Associate Professor and Melton Faculty Fellow at the Haslam College of Business, University of Tennessee, Knoxville. He holds a PhD from Rutgers University. His research interests include data mining and knowledge discovery. He has published papers in journals and conference proceedings, such as IEEE Transactions on Knowledge and Data Engineering, INFORMS Journal on Computing, and European Journal of Operational Research.

**Chuck VanDam** is a Supply Chain Data Scientist in the Strategic Planning & Modeling (SPaM) team at HP Inc. Over the past 25 years, Chuck has held operations management, engineering, and consulting roles within HP, Verigy, Agilent Technologies, and Cisco Systems, with a focus on forecasting, planning, inventory and working capital management, network design, and complexity management. Chuck holds MBA and Manufacturing Systems Engineering M.S. degrees from Stanford University.

**Adam Ghozeil** is a Principal Data Scientist in the Digital and Transformation Office at HP Inc. His focus is on developing data pipelines, AI models, and digital tools to drive efficiency and accuracy. Adam earned a B.S degree in Electrical Engineering from UC San Diego, and has 28 years of experience across R&D, manufacturing, and business functions.

**Pedro A. Neto** is a Supply Chain Data Scientist in the Strategic Planning & Modeling (SPaM) team at HP Inc. With more than 12 years of industry experience, Neto is passionate about leading and implementing advanced projects in data science and supply chain analytics, as well as creating models that facilitate complex decision-making processes and provide actionable insights to stakeholders. He currently serves as a member of the Board of Directors at the Association for Supply Chain Management (ASCM). He holds a B.S. in Industrial Engineering and a Ph.D. in Operations Research and Industrial Engineering, both from Penn State.

**Frederic Marie** is a Supply Chain Data Scientist in the Strategic Planning & Modeling (SPaM) team at HP Inc. He enjoys using Operations Research and Artificial Intelligence to answer business questions. He loves to create intuitive visualizations and explore fusion cuisine. Frederic holds an Advanced Engineering Degree in Computer Science, a M.S. in Applied Mathematics from Joseph Fourier University in Grenoble, and an MBA from the Wharton School of the University of Pennsylvania.