statworx®

Whitepaper

Effective Forecasting:

Technical Methods, Profitable Application & Challenges in a Corporate Environment

Forecasting transforms historical data into reliable predictions about the future. Discover how forecasting can be technically implemented and how businesses can benefit from it in practice.

2024 – 11

About the Whitepaper and the Authors

Discover what this whitepaper offers and learn about the authors behind it.

In this whitepaper, you will learn what forecasting is and how businesses can technically implement and profitably utilize various forecasting methods. We present real-world use cases and discuss the challenges companies face during implementation.

Marlon Ziegler
Senior Data Scientist

Marlon Ziegler is a Senior Data Scientist specializing in time series data, forecasting, and anomaly detection. He leads the internal „Forecasting“ working group, aimed at promoting knowledge exchange and development.

Daniel Lüttgau
Head of AI Development

Daniel Lüttgau is responsible for AI Development at statworx and is an expert in leveraging AI to create value for businesses.

statworx®

statworx GmbH
Hanauer Landstr. 150
60314 Frankfurt am Main

www.statworx.com
info@statworx.com
+49 (0)69 6783 067 - 51

Content

Entry

Introduction

From order intake and sales forecasting for improved stock management to customer development models and the prediction of liquidity time series, forecasting enables companies to prepare for changes, allocate resources efficiently, and optimize profitability.

Forecasting is a central task in almost all departments of a company. Typical use cases include:

SALES & MARKETING:
Sales and revenue forecasts, budget planning

FINANCE:
Financial planning, liquidity planning

PRODUCTION:
Production & procurement planning, maintenance needs forecasts

CUSTOMER SERVICE:
Support request forecasts, resource planning

FACILITY MANAGEMENT:
Maintenance forecasts, energy consumption forecasts

LOGISTICS & TRANSPORT:
Warehouse capacity planning

HR:
Workforce planning, turnover forecasts

IT:
Capacity planning, resource utilization

In all listed cases, data-driven forecasting can provide significant added value. This value can arise in two ways:

1) Either the forecast itself is the goal and offers financial benefit, or

2) the forecast is an indispensable part of a larger process contributing to value creation.

This whitepaper provides companies with an overview of various forecasting models and demonstrates which models are suitable for which challenges. It also illustrates how the integration of data-driven forecasts in various areas such as sales, finance, production, and logistics leads to more efficient resource allocation and increased profitability.

The whitepaper delves into different methodological approaches–from classical statistical methods to modern deep learning and foundation models–and discusses the challenges and solutions in implementation. The goal is to provide a comprehensive overview of how companies can effectively use data-driven forecasts to achieve strategic clarity and operational excellence.

Continue on the next page →

1

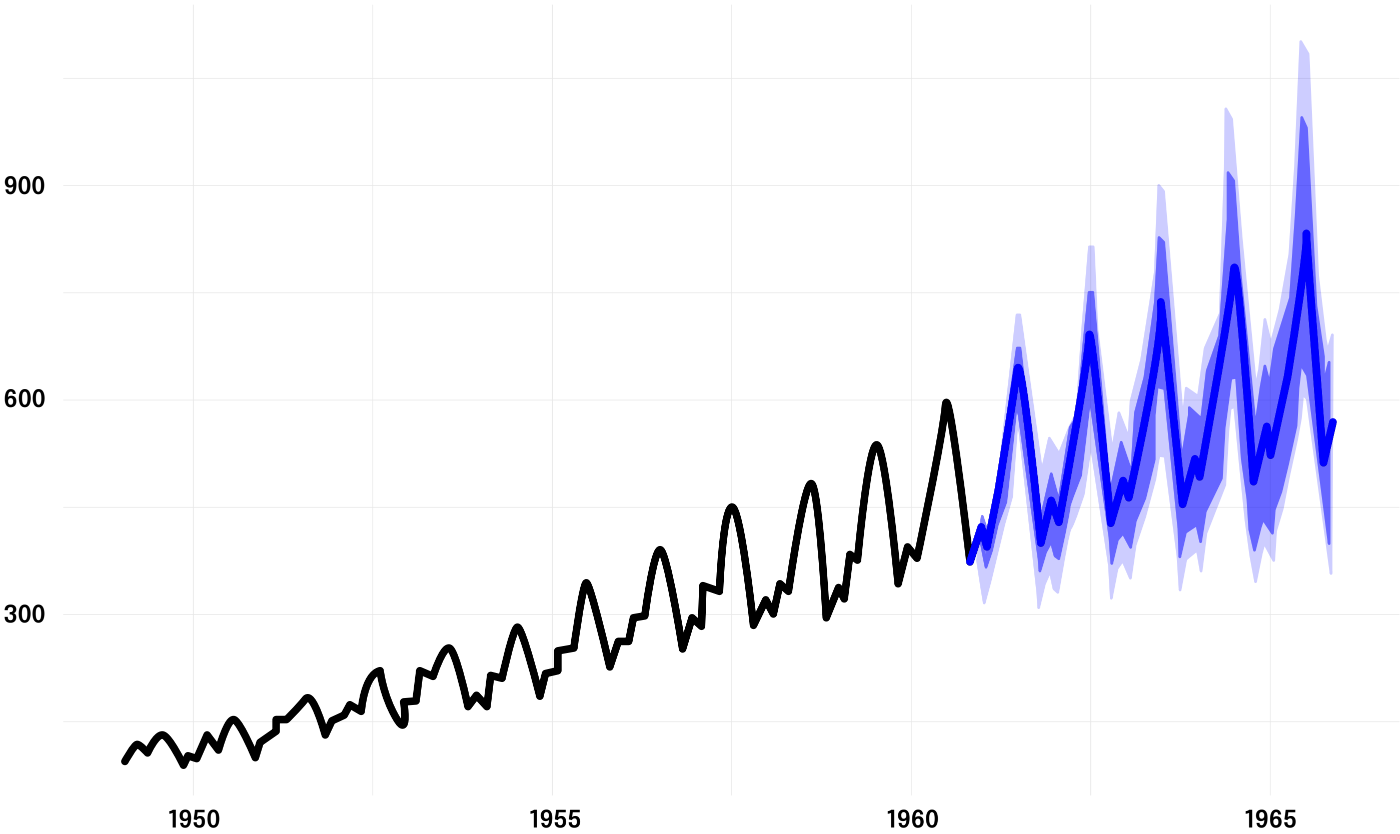
Definition: Forecasting

Definition & Overview

What is Forecasting?

Forecasting refers to the systematic process of predicting future events, trends, or developments based on historical data and corresponding models. Forecasts are usually based on the temporal sequence of data, but they can also be based on cross-sectional data or structural models. This whitepaper focuses on time-series forecasting, which involves predicting future events or values based on previously observed data. A typical application is temperature forecasting: patterns can be detected from a list of daily measured temperatures, allowing estimates of future temperatures to be made. The aim is to learn from the past to make informed predictions for the future.

A *time series* is a chronologically ordered sequence of observations of a numerical variable, such as monthly sales figures for a particular product, collected at consistent intervals. Time series can be broken down into various components, such as trend and seasonality.



Textbook example of a time series by Box & Jenkins: Monthly total numbers of passengers in international air traffic, 1949 to 1960. Forecast over 5 years.

Definition & Overview

A forecast is an estimate of how the sequence of observations will continue in the future. The methodological range is as diverse as the application cases: from classical statistical, univariate methods to established multivariate machine-learning models to complex deep-learning approaches. The choice of the appropriate approach depends on the characteristics of the input data and the requirements of the forecast.

The goal of forecasting is not to completely exclude experts with domain knowledge from the equation and replace them with data-driven insights. On the contrary, domain knowledge and intuition still have their deserved place. In some cases, they even offer a decisive advantage over structured available data, as models cannot account for all potential influencing factors that an experienced person can incorporate into their forecast. Rather, forecasting aims to provide experts with additional knowledge input based on data-driven pattern recognition alongside their experiential values.

Integrating human knowledge and machine forecasting solutions presents a challenge for organizations. Experts must understand precisely what information is contained in the forecasting models and what is not. Only this additional information enables them to adjust and refine predictions. For example, experts can correct forecasts when they know that current supply chain bottlenecks are not captured by the available data and therefore not considered in the forecasts. This additional human information complements the machine-generated predictions and contributes to increased accuracy.

Effective change management is essential to achieve such synergy between humans and machines. Organizations must ensure that their employees receive the necessary training and tools to use the new technologies efficiently. This includes not only technical training but also developing an understanding of the limits and possibilities of the forecasting models used. Additionally, communication plays a crucial role. There must be continuous dialogue between the developers of the forecasting solutions and the users. This ensures that the models are constantly improved and adapted to the changing requirements of the organization.

2 Forecasting Use Cases

Use Cases

Application Examples of Forecasting

Forecasting takes many forms and is used in various industries to predict future developments. This chapter presents different practical examples that showcase the versatility and practical benefits of forecasting in the business world. From sales forecasts in the fashion industry to price and discount adjustments to liquidity forecasts, the examples illustrate how precise predictions can support strategic decisions and add value to companies.

Practical Example: Forecasting in the Fast Fashion Industry

CHALLENGE

Our client, a leading sports goods company in the European market, faced the challenge of improving sales forecasts for an extensive and complex product portfolio. We developed an automated forecasting engine capable of generating explainable and probabilistic predictions for future product sales over periods of up to 18 months.

SOLUTION

The forecasting engine is based on an advanced, explainable deep-learning model capable of processing large amounts of data and providing detailed predictions.

RESULT

The engine is highly flexible and modular, allowing future algorithms to be easily integrated. The Backlink LP Forecasting Engine offers the following advantages:

Increased Accuracy: Manual forecasts were surpassed by an average of 10 percentage points. This reduces unnecessary inventory costs and decreases missing capacities.

Time Savings: Monthly planning time was reduced from weeks to hours.

Wide Applicability: The engine can forecast over 500 products across online, retail, and wholesale channels throughout the European market.

Support for Cold-Start Products: New products without historical data can be forecasted based on product images and master data of existing products.

Scalability: Thanks to a fully scalable cloud implementation and data pipelines utilizing distributed computing, the solution can be expanded to up to 20,000 products.

Use Cases

Practical Example: Forecasting for Optimal Capital Utilization

CHALLENGE

Our client, a large retail company, faced the challenge of optimally utilizing a defined budget for discounts to maximize profits, optimize inventory, and achieve sales targets. We developed a solution based on a machine-learning approach to calculate the price elasticities of products and accurately predict future sales, allowing the budget to be allocated efficiently even for products without historical data.

SOLUTION

We developed a comprehensive solution consisting of three central components:

Price Elasticities: Based on a machine-learning model for modeling the likelihood of an order, price elasticities are determined. This helps identify products with low price sensitivity to set higher prices or lower discounts, while highly price-sensitive products receive larger discounts.

Forecast: A reliable basis for optimization is created through the forecast of future sales. A machine-learning model generates accurate predictions under unchanged conditions, even for new products without historical data, by leveraging data from similar predecessor products as well as seasonality and trend patterns.

Optimization: Based on the calculated price elasticities and predicted sales, the budget is optimally distributed across the products. Exceeding the budget is not allowed, while aiming for underspending to achieve the set goals often with a lower budget.

RESULT

The solution enabled flexible adjustment of prices and discounts, significantly improving both margins and sales and inventory levels. Several goals were set for optimization, which can be applied as needed: increasing margins, achieving sales targets, or a mix of both. The solution offers several advantages:

Increased Profit Margins: Targeted price strategies improved the margin for products with low price sensitivity, while sales of highly price-sensitive products were significantly increased through larger discounts, leading to an overall increase in total profit.

Optimized Inventory Management: Precise sales forecasts enabled better inventory planning, minimizing overstock and stockouts and improving capital binding.

Efficient Budget Utilization: By targeted distribution of discounts, the company could achieve its sales targets while using the budget more efficiently.

Scalability: The solution is fully implemented in the cloud and utilizes distributed computing to scale to a larger number of products. It also supports new products without historical data by analyzing product images and master data of existing products. The technical challenge of including new products without historical data in the forecasts was met by using data from similar predecessor products as well as seasonality and trend analyses.

Use Cases

Practical Example: Liquidity Forecast for Efficient Interest Reduction

CHALLENGE

The project's goal was to create the most accurate forecast of liquidity (inflows and outflows) at various hierarchy levels for multiple companies belonging to a holding. The client's hypothesis was that human planning experts have a certain bias in estimating cash flow and tend to be conservative. This means they tend to underestimate inflows and overestimate outflows. A data-driven and thus unbiased machine-learning approach was expected to predict actual values more accurately.

SOLUTION

The solution involved developing an ML-based approach to predict liquidity.

Data Classification and Modeling: Individual time series were classified into different categories based on their heterogeneity in terms of history, seasonality, and variance. Subsequently, they were predicted using different modeling approaches. For short time series, naive forecasts or simple statistical models were used. With sufficient data, a LightGBM model enriched with exogenous information was employed.

Hierarchical Consistency: Since the time series were hierarchically structured, it was important to ensure consistency across the various aggregation levels. The „Optimal Reconciliation" approach by Hyndman and Athanasopoulos (2014) was used. This approach involves predicting all-time series separately and then optimally reconciling them to achieve the best results.

RESULT

The evaluation showed that the ML model regularly provided better forecasts than human experts in companies without active expert planning. In companies with active expert planning, the model performed worse. These results demonstrate that data-driven solutions are not necessarily intended to replace existing human resources. Instead, experts should be supported by empirically based information. Simultaneously, the added value of an ML-based forecast can be scaled to companies without specialized planning experts.

ROI AND INTEREST COST AVOIDANCE

The ROI of the ML-based approach can be determined by avoiding interest costs. Each euro of ostensibly necessary liquidity provided by expert planning but not needed resulted in interest losses. Each euro not provided due to more accurate data-based forecasting and ultimately not needed resulted in avoiding interest costs. This avoidance was defined as a KPI, enabling a clear assessment of the financial advantage of the new approach.

3 Challenges for Organizations

Challenges

Current Challenges for Organizations

Many companies handle forecasting use cases in isolation within various departments. This leads to so-called silos, systems that structurally separate employees who should work and communicate together. Silos create barriers that hinder cross-team collaboration and communication, reduce efficiency, and impede information flow.

A typical example is sales forecasting conducted by both sales and production. This redundant work is inefficient, causes inconsistent results, and complicates the overall overview. If each department uses its own technologies and frameworks, maintenance efforts increase, unnecessarily burdening IT resources. The fragmentation caused by silos complicates collaboration and leads to inefficient processes and higher costs. A central data basis is lacking, increasing the risk of data inconsistencies and erroneous decisions.

Different data sources and their varying interpretations in departments lead to inaccurate forecasts that impair strategic planning. Varying technologies and frameworks cause compatibility issues and hinder information exchange. The increased training effort for employees constantly adapting to new systems is another problem. Therefore, companies should create a harmonized and centralized data infrastructure that consolidates forecasts and provides a unified data overview. Only through close collaboration and standardized processes can efficiency and forecast quality be sustainably improved.

Continue on the next page →

4 Forecasting Models: Technology

Forecasting Models

Technical Overview

The methods used to create forecasting models are diverse, ranging from classical statistical models to machine-learning methods to highly complex deep-learning architectures and the currently often-discussed foundation models. Each of these methods has strengths and weaknesses. The choice of the right method depends heavily on the specific requirements of the application and the available data.

Time Series Models

Classical time-series models like ARIMA, SARIMA, and ETS have long been proven tools for predicting univariate time series with pronounced seasonal and trend-related patterns. These models are particularly valuable for benchmarking and applications where the forecast is based on a single variable. Their strengths lie in their simplicity and ability to provide clear and understandable results. Below are various time-series models with their strengths. Hyndman and Athanasopoulos (2021) offer detailed explanations for these and other time-series models.

(S)ARIMA (AutoRegressive Integrated Moving Average):

ARIMA models combine autoregression (AR), differencing (I) for stationarization, and moving averages (MA) to model temporal dependencies and trends. SARIMA extends this methodology by integrating seasonal components to account for periodic patterns.

ETS (Error, Trend, Seasonal):

ETS models encompass a family of models considering errors, trends, and seasonality. This method uses an additive or multiplicative combination of these three components to capture temporal patterns. It is particularly useful for data with clear seasonal and trend-related structures, allowing flexible adaptation to different time-series characteristics.

In addition to the two classical time-series models, there are time-series models offering higher complexity.

TBATS (Trigonometric Seasonal Components, Box-Cox Transformation, ARIMA Errors, Trend, Seasonal Components):

TBATS models are especially suitable for complex seasonal patterns and long seasonal cycles that are difficult to capture with classical models. They combine several techniques: Box-Cox transformation for variance stabilization, trigonometric functions for capturing seasonal patterns, ARMA models for modeling short-term dependencies, and trend components for accounting for long-term trends. TBATS is particularly useful for time series with complex seasonality in the form of multiple seasonalities.

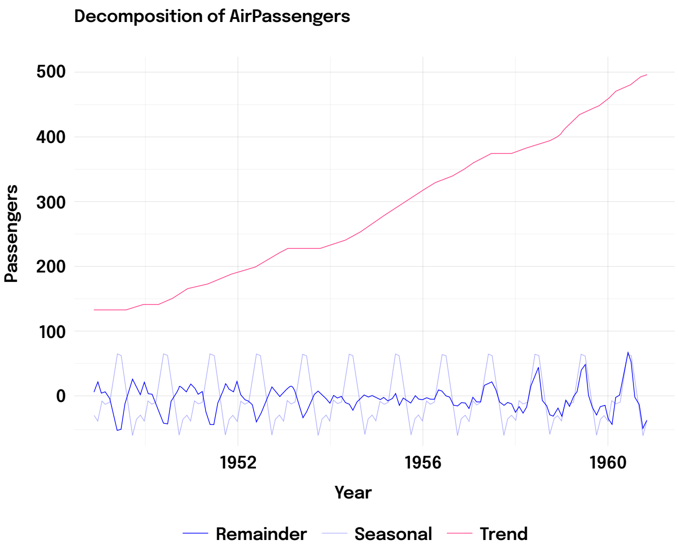
Prophet:

Developed by Meta, Prophet offers a robust solution for data with strong seasonal effects and missing values. It uses additive modeling, defining separate components for trend, seasonal effects, and holiday effects. This makes it particularly user-friendly and adaptable. Additionally, external features can be incorporated, providing an advantage over most other time-series models that cannot account for factors like weekdays, holidays, or marketing campaigns.

At statworx, we typically start a project involving time-series forecasts with classical time-series models and naive methods like the mean or the last known observation. Especially in the early phases of a project, these simple approaches allow a quick assessment of the data and often deliver surprisingly good results. Due to their simplicity and low computational requirements, they are excellent starting points before more complex and computationally intensive models are employed. This allows us to quickly make initial predictions and flexibly adjust

Forecasting Models

the modeling strategy without unnecessarily investing time and resources. Moreover, we sometimes integrate these classical models into ensemble models to optimally leverage their strengths. Particularly for time series with pronounced seasonality, they offer significant added value: For part of the modeling, we remove the seasonal component from the time series and explicitly model the seasonality with a time-series model.



Tree Based Modelle (ML)

Tree-based models such as LightGBM, XGBoost, and CatBoost have been popular for many years due to their efficiency. They can process large amounts of exogenous features efficiently and model complex non-linear relationships. These methods are particularly useful when it comes to incorporating a wide variety of influencing factors into predictions. Additionally, techniques like SHAP (Shapley values) provide high transparency and explainability of model decisions, which is an important criterion for trust and acceptance of the models in practice.

Basic Concept of a Tree-Based Model: Tree-based models are based on decision trees, where the data is recursively split into smaller subsets. Each node in the tree represents a decision based on a feature, leading to further splits until a terminal condition (leaf node) is reached. The leaf nodes represent the prediction values. This structure allows the models to capture complex non-linear relationships in the data.

Gradient Boosting und Ensemble

- **Gradient Boosting:** Gradient boosting algorithms, such as LightGBM and CatBoost, sequentially build multiple trees, with each new tree attempting to correct the errors of the previous trees. This results in a robust, composite model that can deliver precise predictions.

- **Ensemble:** In contrast, the Random Forest algorithm generates many decision trees in parallel by using random samples of the data and features. The final prediction is derived from the average or majority of the individual predictions of the trees. The random selection of features and the sampling of training data sources independent errors, contributing to avoiding overfitting.

When developing models, algorithms based on gradient boosting offer several advantages: Gradient boosting methods like LightGBM, XGBoost, and CatBoost typically provide higher efficiency and better performance compared to Random Forest, especially when hyperparameters are tuned. Modern GBM algorithms incorporate advanced features that enhance their scalability and efficiency. Additionally, CatBoost specifically offers the advantage of efficiently and natively incorporating categorical features into the modeling process.

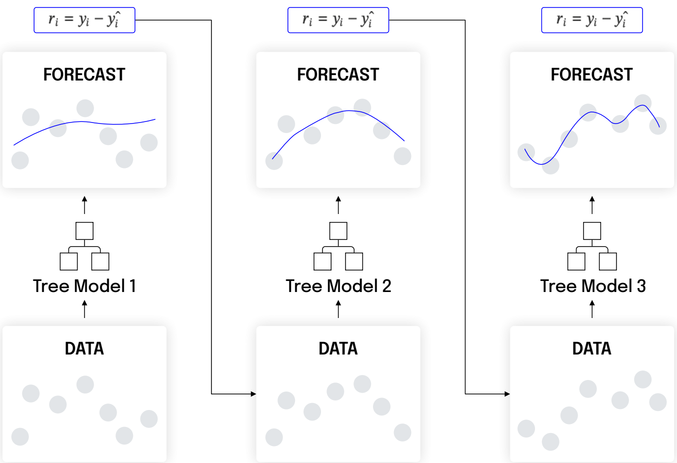
However, caution is advisable when using tree-based models. These models make predictions based on splits in the data determined by input features, limiting their ability to predict trends outside the training data. To ensure extrapolation for tree-based models, the following strategies can be used:

- **De-Trending:**
  Removing the trend from the time series allows the model to focus on cyclical and seasonal patterns. The trend is estimated separately and added back after the model's predictions. This approach is easy to implement but heavily depends on the quality of the trend estimation.

Forecasting Models

- **Differencing:**
  This method calculates the difference between successive observations to stabilize the time series. Although simple, excessive differencing can lead to loss of information and complicate the modeling of the time series.

- **Use of the Linear Tree Method:**
  Integrating a linear component into the tree structure allows the model to capture both linear and nonlinear patterns, enabling trend extrapolation without transforming the time series to be predicted. This approach is model-specific and can increase complexity.



Deep Learning

Deep learning models have the potential to recognize and utilize very complex and non-linear patterns in the data. These models typically require large datasets and complex implementation but offer significant advantages when making predictions in highly dynamic and complex environments. Model architecture plays a crucial role, and optimizing these models requires specialized knowledge and experience.

The use of deep learning in forecasting is not new. Early models such as Recurrent Neural Networks (RNN), Convolutional Neural Networks (CNN), and Long Short-Term Memory (LSTM) models were developed as early as the 1990s. In recent years, however, deep learning has gained increasing attention in the forecasting domain, particularly due to Transformer models. By employing a special architecture known as Attention, these models can better capture and process contextual dependencies in the data.

Temporal Fusion Transformer (TFT): The TFT is a specialized deep learning architecture designed to efficiently process time series data and make accurate predictions, particularly in scenarios with complex and dynamic environments (Lim et al., 2021). The TFT architecture combines several concepts, including „Attention“ modules, to capture both temporal dependencies and relationships between different features. Its key technical characteristics include:

- **Multi-Head Attention:**
  To capture long-term dependencies, the TFT employs a Multi-Head Attention module. This module allows the model to consider multiple past points simultaneously and determine their importance for the current prediction. Multi-Head Attention means that the model uses several heads (parallel attention layers) to analyze different aspects of the data. This parallel consideration enables capturing both detailed local patterns and broader, global dependencies over longer periods.

- **Gating Mechanisms:**
  The TFT uses specialized gating mechanisms to control the flow of information through the network. One of the key mechanisms is the „Variable Selection Network“ gate, which determines at each step which input variables are relevant for the prediction. This ensures that the model is not overwhelmed by irrelevant or noisy features but focuses on essential information.

- **LSTM Layers:**
  A central component of the TFT architecture is the Long Short-Term Memory (LSTM) layers, which capture temporal dependencies in the data. LSTMs are particularly good at learning long-term dependencies by storing information about previous time steps and using it for future predictions. In the TFT, these LSTM layers are used to model the dynamics in the time series data, both for past and future time points.

Forecasting Models

- **Static Covariate Encoders:**
Static features, or static covariates, are features that do not change over time, such as geographic location, certain demographic information, or specific product characteristics. These features are often crucial for modeling as they provide essential contextual information that can influence the dynamic processes in the time series. In the TFT, these static features are integrated into the model through static covariate encoders. These encoders transform the static features into context vectors that influence the entire network, serving as a condition for the temporal dynamics in the model.

Many other Transformer models have been developed to leverage the architecture's properties to address forecasting challenges:

- **Informer:** Designed for efficient long-sequence and time series prediction, it introduces a special form of self-attention to reduce computational complexity. Through self-attention distillation and a generative decoder module, long sequences can be predicted in a single pass (Zhou et al., 2021).

- **TimesNet:** Treats time series as a combination of seasonal patterns and uses „frequency-temporal attention" to decompose the time series and capture dependencies on multiple levels (Wu et al., 2022).

- **PatchTST:** Adapts vision-transformer concepts for time series prediction by dividing input sequences into non-overlapping patches. More complex multivariate data are processed independently per channel. Additionally, the model can be trained through self-supervised pretraining (Nie et al., 2022).

- **Crossformer:** Captures short-term and long-term dependencies in time series data through segmental self- and cross-scale attention. A multiscale encoder processes inputs at different temporal levels, enabling efficient processing of both local and global patterns (Wang et al., 2021).

- **iTransformer:** Utilizes attention and feed-forward networks on inverted dimensions. Time points of individual time series are embedded into variate-tokens, used by the attention mechanism to capture multivariate correlations. Simultaneously, the feed-forward network is applied to each variate-token to learn non-linear representations of the time series (Liu et al., 2024).

Models like DLinear, N-BEATS, and NHITS offer less complex architectures that are both flexible and efficient. Particularly for less complex data, these models can serve as suitable alternatives to classical time series models and the previously mentioned architectures.

One advantage of deep learning methods is their inherent ability to automatically extract and transform features. This reduces the need for manual feature engineering often necessary for traditional ML methods. Techniques such as batch normalization, dropout, and other regularization methods can effectively counteract overfitting and achieve better generalization on unseen data.

Deep learning models are generally more complex and require significantly greater computational power and longer training times. This complexity increases not only the effort in implementation but also in model maintenance and optimization. Furthermore, deep learning models are often more sensitive to overfitting, especially when trained on small or insufficiently representative datasets. While techniques like dropout and regularization can help, the risk remains that the model will too closely fit the training data, resulting in poorer performance on new, unseen data.

In practice, deep learning models can be particularly useful when the complexity of the data structure overwhelms traditional tree-based approaches or when prediction accuracy needs to be optimized for volatile and dynamic datasets. Despite recent successes of Transformer-based models in time series forecasts, research shows that even simple architectures like one-layer linear models can outperform complex Transformer models in certain cases. These surprising results by Zeng et al. (2023) suggest that temporal relationships in the data may sometimes be better captured by less complex models. Especially when combined with transfer learning and advanced optimization algorithms, these simpler models can offer tailored solutions for a variety of use cases. Therefore, the decision to use deep learning models should be made carefully, as the higher implementation effort does not necessarily lead to a better solution.

Forecasting Models

Foundation Models

Foundation models represent a relatively new advancement in machine learning. They are currently considered one of the most intensive research areas in forecasting. A prominent example is ChatGPT, probably the most well-known large language model (LLM). These large-time-series models (LTSM) are characterized by their enormous performance, as they are trained on vast amounts of data and can handle a variety of tasks. In forecasting, they offer the advantage of being able to be used for both car sales forecasts and production capacity forecasts simultaneously. The main difference from other model classes is that they can be used directly without requiring separate and often very time-consuming training.

The underlying model architecture is currently mostly based on transformers, a deep learning technology similar to large language models. The potential of this model class is immense: Companies can create their own forecasts in seconds without developing features, training models, evaluating, and selecting them. This efficiency and versatility make foundation models a promising technology. Another advantage is that companies often have small datasets in the context of time series, so the benefits of deep learning forecasting models often remain unused, and they must rely on other model classes.

Currently, there are several providers of these models. Nixtla's TimeGPT-1 (NIXTLA, 2024) is the first paid forecasting foundation model. However, there are also open-source alternatives such as TimesFM by Google (Das et al., 2023) or MOMENT , a collaboration between the University of Pennsylvania and Carnegie Mellon University (Goswami et al., 2024). The application of these models is simple; datasets can be sent to an interface in a certain structure along with setting parameters, which then returns a forecast and sometimes even uncertainty intervals. For some models, the forecast horizon is freely selectable, meaning it can predict as far into the future as desired. Of course, prediction quality decreases with longer horizons, as with all other model classes.

Similar to LLMs, LTSMs can also be fine-tuned. This means that the model can not only be applied but also adapted and trained on the company's own data. LLMs can, for example, be adapted to the company's internal communication (e.g., corporate language) and technical terminology. In the context of LTSMs, this refers to specific characteristics of a company's data, such as different seasonal sales cycles in the fashion industry. Unlike large language models, fine-tuning is more cost-effective as LTSMs are significantly less complex than LLMs. Garza and Mergenthaler-Canseco (2023) have shown that prediction accuracy can be significantly improved through fine-tuning compared to the pure application of LTSMs.

Time series often contain highly sensitive company data, so the use of foundation models must be internally coordinated. Secure use of these models can be ensured by hosting them in the company's own cloud or on-premises. This way, the data remains within the familiar work environment and is securely processed within the company's infrastructure.

Foundation models for time series forecasting represent a user-friendly and promising alternative to traditional model classes. With the increasing number of providers, more foundation models will come to market, leading to continuous improvement in quality and performance in the future. LTSMs are suitable as reference models because they can be implemented with minimal effort and deliver high prediction quality. However, it is important to emphasize that they are by no means a panacea for all use cases in time series forecasting. They should always be compared with classical statistical methods, as these can often be implemented with minimal effort.

Despite their versatility and performance, the use of LTSMs requires careful data preparation. The quality and consistency of the data remain the decisive factors for the accuracy of the forecasts. Additionally, the interpretability of the models should not be overlooked. While LTSMs are powerful tools, the transparency of decision-making processes is often limited. Therefore, it may be beneficial to use supplementary explainable models to gain insights into the underlying patterns and relationships within the data.

5

Challenges in Implementation

Challenges

Typical Implementation Challenges

The demand for service-based forecasting solutions, allowing users to generate predictions at the push of a button based on simple data input, is growing. Many established providers offer standard solutions for this purpose. However, depending on the industry and project, the complexity of forecasting needs can exceed the capabilities of such solutions. To fully realize the value of forecasts, complex business processes must be considered: For example, lead times and procurement processes must be included in sales forecasts. A forecasted sales volume should be realistic in terms of deliverability to avoid bottlenecks or overstock. This means that forecasts must be aligned with production and order cycles to ensure that the required products are available on time. The type of product also has a significant impact on forecasting: How do you create a forecast for a completely new product without historical data („cold start“)? Conversely, what impact does a production stop and the resulting continuous reduction of remaining stock have on the sales development of this product, and to what extent do „raw“ model forecasts need to be corrected accordingly?

These are just a few of the many questions that a high-quality, sustainable forecasting solution must consider. Therefore, services specializing in numerical predictions often only provide part of the solution. Forecasting fundamentally has two application types. Either the forecast itself is the solution, embedded in one of the business processes outlined above and providing its own added value. Or the forecast is a means to an end in another data-driven use case. Typical examples include achieving sales targets or optimizing inventory levels through targeted price adjustments (intelligent pricing). When implementing a comprehensive solution that delivers tangible added value, companies often face specific challenges. These can be due to both external factors and the requirements of the solution itself. Typical challenges repeatedly occur in projects. Here, we will examine the most common problems and how they can be addressed.

Exogenous Shocks Like Pandemics and Supply Chain Issues

Core Issue: Unpredictable external events such as the COVID-19 pandemic or supply chain disruptions can significantly impact the accuracy of forecasting models. Such shocks are rarely predictable through available data and lead to abrupt and often drastic changes in the data. These structural breaks, which do not correspond to the usual market environment, can cause significant distortions in trained models if not appropriately accounted for.

Solution Approach: The problem of exogenous shocks can be addressed through model selection and data processing to best compensate for negative impacts on performance.

- **Data Preparation:** In data processing, it is first necessary to identify all anomalies and time periods associated with exogenous shocks. This can be done through experts and qualitative approaches. Especially in the context of COVID-19, the periods in the data are usually clearly recognizable. Algorithms can also be used to identify anomalies and structural breaks. After identification, the affected data must either be corrected or removed from the training set. One option is to replace these data with simulated or estimated values that reflect normal conditions.

- **Model Selection:** The use of robust models that are resistant to outliers and anomalies is particularly helpful in minimizing the impact of exogenous shocks. Tree-based models such as LightGBM and Random Forest provide mechanisms to handle anomalies and improve model accuracy. LightGBM uses a leaf-wise

Challenges

growth algorithm and extensive regularization options to control overfitting and the impact of outliers. Random Forest reduces sensitivity to outliers through bootstrap aggregation and feature randomness by training multiple decision trees on random data samples and aggregating their results.

Product Launches

In many industries, new products are regularly introduced, requiring forecasts. Since there is no historical data for new products, they cannot be explicitly considered during model training. To still provide a forecast for such products, the following approaches are particularly suitable:

- **Naive Approaches:** Especially in areas where there are predecessor products, naive approaches are suitable. Based on predecessor information, the future product's sales level can be inferred. Seasonal components and the sales development of previous ramp-ups can also provide additional value for the forecast.

- **Transfer Learning:** Even if training with explicit consideration of future products is not possible, existing models can be used to create forecasts for new products. This involves using models and features that were used for similar products or predecessors.

In both approaches, it is assumed that conclusions about new products can be drawn by analyzing related or similar products. However, it is important to note that forecasts for new products are generally associated with higher uncertainty than for already established products with extensive history. A crucial success factor for these methods is the timely availability of information about upcoming product launches.

Explainability

In many projects, it is crucial that forecasts are understandable and transparent. Since complex models are often perceived as black boxes, there is a risk that their results may not be accepted. To increase the acceptance and comprehensibility of forecasts, it is therefore important to find ways to „explain“ the models and their predictions. There are two potential approaches for this goal. First, the model can be generally explained. This involves examining which information is important for the model and how this information influences the predictions. Second, individual predictions can be explained. This shows how each piece of information influenced the respective prediction. The following are some common methods explained in more detail:

- **SHAP (SHapley Additive exPlanations):**
  SHAP is a method based on Shapley values from game theory (Lundberg & Lee, 2017). It calculates the contribution of each feature to a prediction by repeatedly masking individual features and analyzing how the model responds. SHAP provides detailed explanations both at the individual (local) and aggregated (global) levels. Additionally, SHAP considers interactions between features, resulting in stable explanations. Particularly for tree-based models like LightGBM and CatBoost, SHAP has established itself as an efficient standard.

- **LIME (Local Interpretable Model-agnostic Explanations):**
  LIME creates local explanations by fitting a simple, interpretable model (e.g., linear regression) around the prediction of a complex model. This helps to make the prediction for individual data points more understandable. LIME can be applied to a variety of models, including complex models like deep learning networks and tree-based models. Unlike SHAP, LIME is faster and less computationally intensive, although it is debated whether LIME reliably produces true explanations, as it is less mathematically grounded than SHAP.

Permutation Feature Importance: Permutation Feature Importance is a technique that maps the importance of a feature for a model using a score. The importance of a feature is measured by the change in model performance when the values of this feature are randomly permuted. This method provides a global explanation of the model by showing which features have the most significant influence on the predictions.

Challenges

Permutation Feature Importance is highly efficient as it can be conducted without deep modifications to the model. However, this method is based on the assumption that the features are independent of each other, which can lead to misestimations of a feature's actual importance. If there are interactions between features or high correlation, the determined feature importance can be misleading. Therefore, it should be used as an initial indication, while for substantiated statements, methods that consider dependencies between the features should be employed.

Partial Dependence Plots (PDPs): PDPs display the influence of one or more features on the prediction while keeping other features constant. The advantage of PDPs is that they provide a global view of how changes in a feature affect model predictions. It is important to note that the foundation of PDPs is actually Individual Conditional Expectation (ICE) plots, which offer a local perspective. Both types of explanations–global through PDPs and local through ICE plots–are possible. Additionally, Accumulated Local Effects (ALE) are often preferred over PDPs as they better account for feature interactions.

Local Explainability:

Local explanations are created for individual predictions or data points. These include methods such as SHAP and LIME, which provide detailed insights into how individual features influence the prediction for a specific case. These explanations are particularly helpful in understanding individual decisions made by the model.

Global Explainability:

Global explanations describe the overall functioning of the model. This includes general insights into which features are important in the model (e.g., Permutation Feature Importance) and how features affect predictions on average (PDPs). These explanations are useful for gaining a general understanding of the model.

Hierarchical Forecasts

Hierarchical forecasts are predictions organized in a hierarchical structure, where forecasts for different levels of a hierarchy, such as geographic regions or product categories, are conducted. This type of forecasting is particularly relevant for organizations that need their forecasts at various levels of aggregation, such as national sales, regional sales, and local sales. The challenge is to ensure that the predictions are consistent and coherent across all levels of the hierarchy. There are various approaches to creating these hierarchical forecasts, ranging from classical approaches like Bottom-Up and Top-Down to approaches like Optimal Reconciliation.

- **Bottom-Up:** The Bottom-Up approach starts with forecasting at the lowest level of the hierarchy and aggregates these forecasts to generate predictions for higher levels. This allows for a detailed view of the lower levels but can lead to inconsistent forecasts at higher levels.

- **Top-Down:** In contrast to the Bottom-Up approach, the Top-Down approach starts at the highest level of the hierarchy and distributes the forecasts to the lower levels based on historical proportions or other methods. This ensures consistency at the higher levels but can impair accuracy at the lower levels.

- **Middle-Out:** This approach is a hybrid that combines Bottom-Up and Top-Down. Here, a middle level of the hierarchy is chosen as an anchor. Forecasts above this level are calculated using the Bottom-Up method, while the lower levels use the Top-Down method. This method is particularly useful when the middle level offers a good balance between detailed and aggregated data.

- **Optimal Reconciliation:** This approach aims for the best possible consistency between the different hierarchy levels. Known as Optimal Combination, this method uses a generalized least squares estimator considering the covariance matrix of the coherence errors. This ensures that the final forecasts are coherent and optimally combined at both lower and higher levels. This approach was proposed by Hyndman et al. (2011) and is considered one of the most effective methods for consolidating hierarchical forecasts.

Challenges

Uncertainty in Forecasts

Point estimates, where a single value is given as a prediction, can be misleading as they do not account for the uncertainty in the data and models. This is especially problematic in dynamic and complex environments where many unknown factors can influence the prediction. A single point value can lead to a false sense of security, resulting in suboptimal decisions.

Incorporating uncertainty into forecasts offers significant advantages. Instead of predicting just a single value, interval forecasts provide a range of possible outcomes. This allows decision-makers to better manage risk and make more informed decisions, particularly in scenarios with high volatility or uncertainty.

There are various methods to incorporate uncertainty into forecasts:

- **Confidence Intervals:** When using time series models like ARIMA or ETS, confidence intervals are available by default. These intervals quantify the prediction's uncertainty by defining a range within which the future value is likely to fall with a certain probability (e.g., 95%).

- **Quantile Regression:** For tree-based models like XGBoost, uncertainty can be modeled using quantile regressions. Instead of delivering just a point forecast (median), the model calculates predictions for different quantiles, such as the 5th and 95th percentiles. These quantiles then provide the bounds of an interval reflecting the forecast's uncertainty.

- **Monte Carlo Dropout:** For deep learning models, interval forecasts can be achieved using Monte Carlo Dropout. In this method, a portion of the neurons is randomly deactivated (dropout) during the prediction process, allowing multiple different predictions. The distribution of these predictions provides insight into the uncertainty. The model thus delivers a distribution of outcomes instead of a single point value.

Conclusion & Outlook

Added value on several levels

Forecasting offers companies enormous potential to achieve strategic advantages and increase operational efficiency. In an increasingly data-driven world where the availability of information grows exponentially, the ability to make accurate predictions becomes a crucial competitive factor. Companies that effectively use forecasting can optimize their resources, better manage risks, and base their decision-making on a well-founded, data-driven basis.

However, the true value of forecasting lies not only in the ability to predict future developments but also in the flexibility it offers companies. Whether it is about early detection of liquidity shortages, creating sales forecasts for targeted marketing strategies, or making the entire supply chain more efficient–forecasting can create significant added value in almost every area of a company. It is a tool that can generate both direct and indirect benefits, whether through immediate savings or long-term process optimizations.

It is crucial to understand that there is no „one-size-fits-all" solution in forecasting. Each organization is unique and has unique requirements, goals, and challenges. The choice of the appropriate forecasting method depends on various factors, such as the availability and quality of data, specific business requirements, and expected external influences. Standard solutions often cannot provide the necessary precision and adaptability to be successful in complex and dynamic environments.

The requirements and challenges addressed in the whitepaper, such as explainability, external shocks, and product life cycles, should fundamentally be considered when implementing forecasting use cases. Additionally, there are many other requirements shaped primarily by company- or industry-specific business logics. As long as these requirements can be quantified and the necessary data is available, solution concepts for implementation can also be developed and implemented.

Even today, forecasts in many companies are still created by experts with high manual effort. This is inefficient and quickly reaches its limits with a high number of products or complex properties. Data-driven forecasts can significantly increase efficiency and objectively consider complex relationships. Experts can still remain important parts of the process to ensure the quality of the forecasts with their expertise.

By considering individual requirements and applying tailored solutions, forecasting can not only address operational challenges but also secure long-term strategic advantages. Companies that recognize and leverage these potentials are better equipped to succeed in an ever-changing world.

Creating value from data & AI.

statworx is one of the leading consulting and development companies for data and AI in the DACH region.

→ About us

As a leading consulting and development company for data and AI, we support you in all aspects of digital transformation - from strategic AI consulting and targeted AI training to the development of cutting-edge AI solutions.

→ Numbers & Facts

500+ Data & AI Projects successfully completed	**100+** Customers from 10 industries	**80+** Employees from 17 fields of study	**10+** Years of experience in Data Science & AI

→ Some of our clients

Start your journey into the future with Data and AI

Our AI Potential Workshop is the optimal starting point for your AI transformation - understandable and practical.

AI Potential Workshop

One of the big challenges in introducing AI to enterprises is identifying value-added use cases. With our AI Potential Workshop we offer companies the opportunity to quickly and effectively find exciting use cases for AI technology and test them.

BOOK A DATE

MORE INFORMATION

Procedure

→ **Step 1: Assessment**
- Coordination meeting to assess the situation and challenges

→ **Step 2: Preperation**
- Use cases with AI potentials are prepared
- Includes assessment of data and its utility to increase efficiency

→ **Step 3: Workshop**
- One-day workshop
- Evaluation and development of agile and interactive strategies for DS, ML and analytics use cases
- Focus on identifying and implementing the greatest opportunities

→ **Step 4: Prototype**
- Development of a prototype
- Development budget: 10 person days

Your contact

Daniel Lüttgau
Head of AI Development

Daniel Lüttgau is responsible for AI Development at statworx and is an expert in leveraging AI to create value for businesses.

Sources

References and further information on the content of this whitepaper.

Das, A., Kong, W., Sen, R., & Zhou, Y. (2023). A decoder-only foundation model for time-series forecasting. arXiv preprint arXiv:2310.10688.

Garza, A., & Mergenthaler-Canseco, M. (2023). TimeGPT-1. arXiv preprint arXiv:2310.03589.

Goswami, M., Szafer, K., Choudhry, A., Cai, Y., Li, S., & Dubrawski, A. (2024). Moment: A family of open time-series foundation models. arXiv preprint arXiv:2402.03885.

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in neural information processing systems, 30.

Hyndman, R.J., & Athanasopoulos, G. (2021) Forecasting: principles and practice, 3rd edition, OTexts: Melbourne, Australia. OTexts.com/fpp3

Hyndman, R. J., Ahmed, R. A., Athanasopoulos, G., & Shang, H. L. (2011). Optimal combination forecasts for hierarchical time series. Computational statistics & data analysis, 55(9), 2579-2589.

Hyndman, R. J., & Athanasopoulos, G. (2014). Optimally reconciling forecasts in a hierarchy. Foresight: The International Journal of Applied Forecasting, (35), 42-48.

Kong, Y., Wang, Z., Nie, Y., Zhou, T., Zohren, S., Liang, Y., ... & Wen, Q. (2024). Unlocking the Power of LSTM for Long Term Time Series Forecasting. arXiv preprint arXiv:2408.10006.

Lim, B., Arık, S. Ö., Loeff, N., & Pfister, T. (2021). Temporal fusion transformers for interpretable multi-horizon time series forecasting. International Journal of Forecasting, 37(4), 1748-1764.

Liu, Y., Hu, T., Zhang, H., Wu, H., Wang, S., Ma, L., & Long, M. (2023). itransformer: Inverted transformers are effective for time series forecasting. arXiv preprint arXiv:2310.06625.

Nie, Y., Nguyen, N. H., Sinthong, P., & Kalagnanam, J. (2022). A time series is worth 64 words: Long-term forecasting with transformers. arXiv preprint arXiv:2211.14730.

NIXTLA. (2024). Meet TimeGPT. https://www.nixtla.io/timegpt

Wang, W., Yao, L., Chen, L., Lin, B., Cai, D., He, X., & Liu, W. (2021). CrossFormer: A versatile vision transformer hinging on cross-scale attention. arXiv 2021. arXiv preprint arXiv:2108.00154.

Wu, H., Hu, T., Liu, Y., Zhou, H., Wang, J., & Long, M. (2022). Timesnet: Temporal 2d-variation modeling for general time series analysis. arXiv preprint arXiv:2210.02186.

Zeng, A., Chen, M., Zhang, L., & Xu, Q. (2023). Are transformers effective for time series forecasting?. In Proceedings of the AAAI conference on artificial intelligence (Vol. 37, No. 9, pp. 11121-11128).

Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., & Zhang, W. (2021, May). Informer: Beyond efficient transformer for long sequence time-series forecasting. In Proceedings of the AAAI conference on artificial intelligence (Vol. 35, No. 12, pp. 11106-11115).

statworx

statworx GmbH
Hanauer Landstr. 150
60314 Frankfurt am Main

www.statworx.com
info@statworx.com
+49 (0)69 6783 067 - 51