



**MACHINE LEARNING FOR CONTROLLERS
USE CASES FOR FORECASTING,
PLANNING, AND SIMULATION.**

A HIGH LEVEL OVERVIEW.

Content

01 INTRODUCTION	2
02 IMPROVED DATA COLLECTION	4
03 TIME SERIES-BASED SALES FORECASTING	5
04 SALES PIPELINE FORECASTING	7
05 COST FORECASTING	9
06 CONTRIBUTOR ANALYTICS	10
07 DRIVER BASED SIMULATION	11
08 CONCLUSION AND NEXT STEPS	12
09 ABOUT THE AUTHORS	13

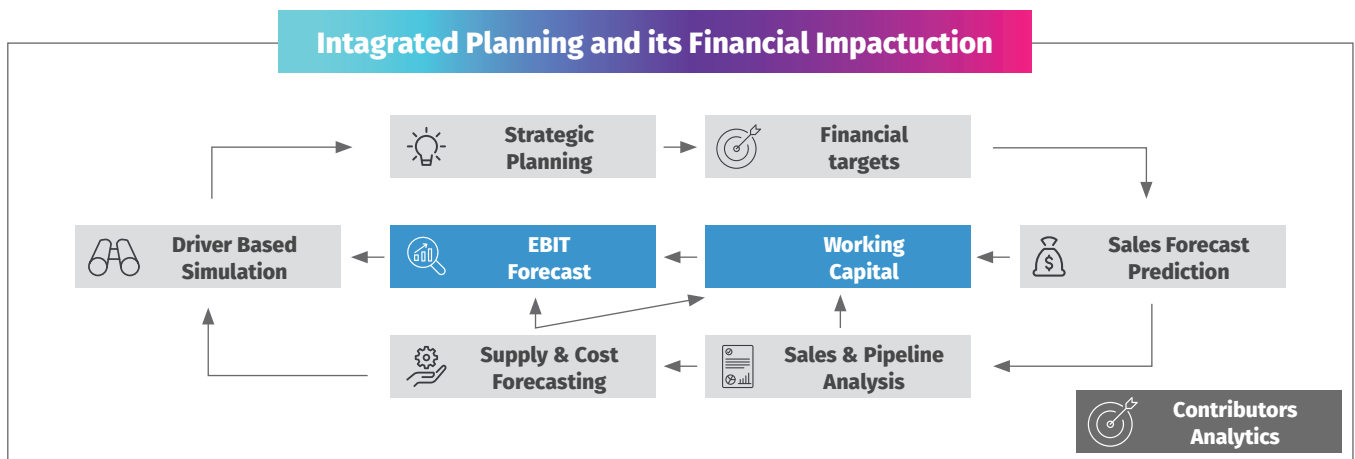
Introduction

Data-driven decision making is evolving very fast with new technologies like artificial intelligence and in particular with machine learning. In this whitepaper we would like to take a practical approach on how to get started in the Office of Finance. For this a number of use cases are described starting from relatively simple to more advanced as a path for growth and taking away pain points often experienced and mentioned in the use cases below.

A big issue for the Office of Finance is often the speed to deploy new developments. In 2020 the Office of Finance needs to make sure to have the foundation in place to be able to make next steps (Ventana Research speaks about a fast follower approach). This foundation consists of a centralized storage of required financial and operational data with high data quality, a modern enterprise performance management platform and the right people with advanced analytical skills.

As a next step various use cases can be implemented, for which we provide some examples:

1. Use Machine Learning in data collection to improve data quality;
2. Sales volume or revenue forecasting
3. Sales Pipeline forecasting
4. Cost forecasting
5. Contributor analytics
6. Driver based simulations



Why we start with these use cases?

As you can see there is a connection between the use cases from an integrated planning and financial results approach.

These use cases reflect typical pain points addressed by business experts in a survey from e.g. German BARC Research. Use Case 1 is a precondition for 2 to 6. Sufficient data quality is the foundation for a successful predictive analytics project. Usually the most impactful use case is about sales forecasting, as this is a very important driver for revenue. Starting with sales volume all planning activities more or less are based on these predictions. Sales forecasting is divided into forecasting for anonymous mass markets and customer relationships.

Cost forecasting is used to identify possible cost drivers to better understand overhead costs. As cost forecasting is mainly based on internal data, it might be an easy starting point.

The quality of contributed forecast data is often an issue, as being e.g. too optimistic or too pessimistic.

Based on a workflow an interactive outlier detection and analysis process can be set up.

Various correlations and potential causal effects are discovered through predictive analytics.

This knowledge can be used to improve simulations and scenarios.

Improved data collection

The first use case is about creating the foundation for all business-oriented use cases. In order to apply machine learning successfully to large data sets it is important to bring both financial and operational data together. Validated data has to be available in a structured manner. Data from different sources and with varying quality needs has to be brought together and refreshed periodically

How can machine learning help to improve the acquisition and the quality of the data? Established ETL-Tools provide functions like recoding, filtering or pivoting. What are typical inaccuracies which cannot be handled with conventional ETL?

- Duplicates (often quite complex to find out)
- Data fields are not completely filled
- Data is not properly assigned to periods and other dimensions, which causes a risk for outliers in the data
- Structural breaks limit the length of time series and bias driver relationships

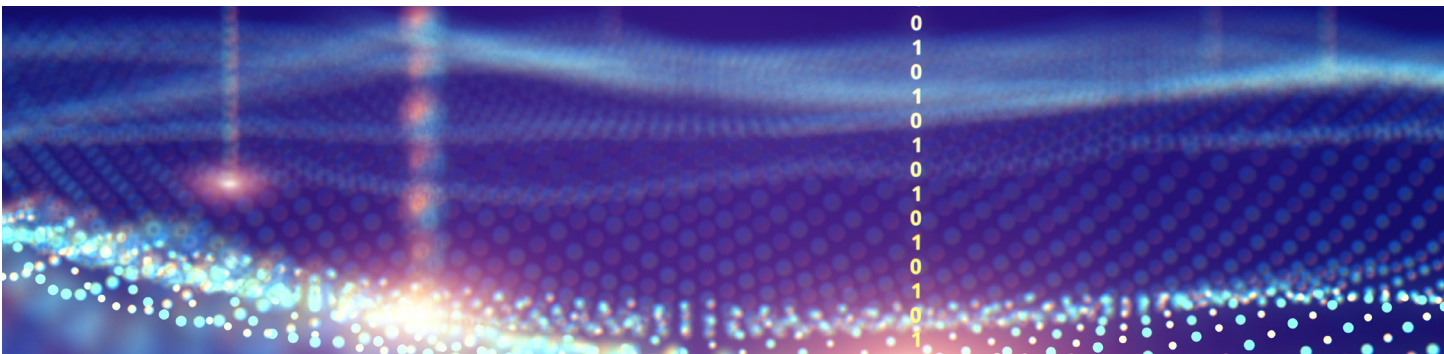
Outliers have to be detected to train the software based on every field provided from the data source. When sophisticated enterprise performance management software is used for planning, consolidation and reporting processes, validations are used to check the quality of the data. Most-ly, this is at an aggregated level for the periodic consolidation. Now with more granular financial and operational data, different validations are needed to make sure the data is correct (meaning registration of transactions is done correctly). Machine learning should be applied to train the software on outlier detection and to automatically apply translation rules to address data to the right spot in the data model needed for advanced analytics.

Some items to specify:

- Machine Learning (ML)-Supported
 - _ Imputing
 - _ Outlier detection and adjustments
 - _ Intelligent assignments
 - _ Text crawling

Important sources

- _ Financial and management accounting;
- _ Customer / prospect data
- _ Economic data;
- _ Supply Chain data.



Time series-based Sales Forecasting

Predicting sales in an accurate way is essential for a good enterprise planning system. Sales forecasting is usually the start of planning efforts and thus all succeeding steps are based on realistic sales predictions.

Sales Forecasting based on time series is a well investigated topic. However, with machine learning, new opportunities for time series analysis arise. First, modern tools simplify usage. Second, by capturing more data, it is possible to consider a broader range of drivers. And third, method restrictions like linearity can be overcome.

Quite often sales forecast is performed by rather simple extrapolations or manual data entry.

Companies which rely on this way of forecasting face problems:

- The core problem is inaccuracy in sales forecast usually measured by indicators like mean average percentage error (MAPE). In consequence increasing planning problems like inventory shortage, capacity shortage, may occur. Even smaller problems in volume forecasting accuracy lead to significant profit and cash variances;

- Furthermore, there is usually a lack of transparency. Unknown drivers or known drivers not used for sales prediction limit forecasting ability;

- High effort resulting from manual forecasting because domain experts are asked for future assessments. The results must be accurately consolidated. Usually manual data is biased and often should be adjusted to reflect realistic assessments.

A broad range of companies can use time series-based forecasting. A certain history of data must be available, of course. Volatility shouldn't be too high, unless it is related to meaningful seasonality.

To use automated forecasting in planning processes, it must be an integrated part of the planning process and easy to set up. Limited technical and statistical knowledge should be needed. Methods like ARIMA, vector auto regression (statistical), random forest or neuronal nets (machine learning) can be used to predict the future on different products or customers hierarchical levels.

To provide sufficient flexibility, forecast methods should be selected directly from input forms and also be applied in batch processing. The system should check whether a method is appropriate and suggest the right input driver and parameters. Measures of significance and bandwidths of confidence should be provided to express the forecast quality. The results of the forecast can be directly used in the preceding planning steps.

The outcome of an integrated and automated forecast is obvious:

- Accuracy in sales forecast leads to a better planning foundation;

- Forecasting accuracy provides stakeholders trust in management capabilities;

- Costs for manual forecast can be reduced by automation;

- The learning effect should not be underestimated: Better understanding of sales drivers, for instance discounts, marketing expenses (including time lags) can be used for simulation, activity planning, variance analysis, etc.;

- The forecasted numbers are uncertain, even with the best forecasting method. Revealing of confidence corridors can be of the foundation of a following risk analysis (see last use case) and help to calculate value at risk resulting from forecast errors.

Sales Pipeline Forecasting



Usually sales are volatile when there is no mass business. No stable development, at least in the short-term can be provided. Time series analysis wouldn't work well in this environment. However, accurate forecasting is necessary as already mentioned.

A lot of details about opportunities and leads are stored in CRM – systems like Salesforce or SAP. Together with marketing information systems and web tracking systems a broad foundation for analysis is usually available. This can help to identify leads/opportunities with a high probability to buy. However, very often the assessment of buying likelihood is processed manually and the (fi-nance) forecast is done the same way by asking the sales representatives.

Some of the pain points of a manual assessment:

- Human bias (like overconfidence) results in high variances between forecasting and actual sales
- Limited learning about buying decision drivers
- The effect of activities to reach a certain target cannot be assessed properly

Several companies can profit with predictive analytics from their sales forecasting pipeline. Pre-conditions are a significant number of concrete sales projects, availability of historical data and sales forecasting data from usually a CRM system. Furthermore, a structured sales process (funnel) with a systematic data collection should be available. A classification of the status of sales cycles can be a helpful basis for accurate forecasting.

The aforementioned information can be used to predict the buying decision and the buying date. Machine learning methods like random forest, gradient boosting, neuronal nets, etc. help to find attributes, that have a significant impact on the buying decision. Attributes like length of customer relation, seller's experience, touch points, reached sales state, etc. are usually good indicators.

But the result of this prediction is not a periodical forecast. In a second step, it is necessary to transform this prediction into a periodical forecast to get a clear picture of the profit situation for a certain period. For instance, between now and the chosen horizon (e.g. one year), the sales pipeline does not remain static. New, yet unknown leads must be considered and included, because they also bear a probability to be successfully closed within the regarded horizon. For this pipeline, transformation is necessary.

What is the outcome?

- More accurate sales forecasting by using high impact attributes for prediction. Human bias can be reduced
- Clear understanding of levers. How strong do the used attributes affect the buying decision?
- Assessment of effectiveness of sales activities because driver impact can be quantified. With an idea of the impact, activities can be prioritized. Focusing on leads with a higher probability could be a result. Effects on discounts, higher visit rate or customer experience touchpoints, etc. can be assessed properly

Cost Forecasting

Cost planning is usually a time-consuming task: resource consumption must be assessed. Multiplied with prices (often a further prediction is necessary) it provides costs. For an accurate cost planning, it is needed to identify cost drivers. Reality, particularly in non-production areas, looks different. In overhead areas, cost drivers are often unknown. The consequences are:

- Costs are usually extrapolated without knowing whether resource consumption is really necessary, due to unknown drivers. This so-called “black holes” in overhead costing are more or less unmanageable due to limited transparency;
- Trials to enhance the transparency usually lead to huge effort for analytical planning. Each cost type has to be analyzed identifying potential drivers.
- In consequence, there is limited support for assessing cost management activities: un-realistic simulations are often the consequence.

Companies with a significant percentage of overhead costs and a history of operational data, which contains potential drivers, can be combined with already available driver structures, like bill of material, work breakdown structures, routings, overhead, and analysis, etc.

As costs are mostly about internal and sizeable data, this is a good spot to get started. For forecasting the development of an operational plan, e.g. sales and operations planning is needed. The starting point, (predicted) sales volumes must be broken down into resource consumption using bill of material, work routing or any other available structure. Possible constraints must be considered. But also, indirect or hard to measure effects have to be assessed. A good example are costs resulting from complexity. How does the number of product variations affect purchasing costs, planning costs etc.? Information about drivers’ dependencies can be used as a starting point to assess costs. Machine learning methods can identify impact and possible time lags. Together with the price, accurate cost can be forecasted.

Cost controlling can be improved by better understanding the cost dependencies:

- Higher accuracy of cost planning, particularly for overhead areas
- Higher transparency of costs and drivers for proactive cost management. Methods like activity-based costing can be supported to manage overhead costs
- Time-saving by automation



Contributor Analytics



Planning and forecasting usually contain manual entered data, even in highly automated environments, because planning is an expression of human will. But planning commitments have to be consistent within the environment.

Integrity can be checked by validation at least for simple cases. For instance, a sudden high increase of the sales costs might be suspicious and can be detected by conventional means, best directly after manual entry. But what about more multivariate outliers? For instance, a moderate sales decrease might not be remarkable. But in comparison with peers or external figures, showing an increase, it could be a significant outlier.

What are the typical problems with manual or oversimplified validation?

- A lot of inconsistencies are usually hidden in contributed data. There is usually too much data to find them efficiently. An aggregated view, used to reduce complexity, often balances out detailed problems
- Limited mutual understanding in contributor processes. Planning behavior like sand-bagging, etc. remains undiscovered

Companies with a decentralized planning or forecasting process can use sophisticated outlier detection methods. The central units usually have a certain degree of freedom to decide. But it is likely that they do not see the overall picture. It is possible to let machine learning algorithms check the contributed data for inconsistencies. Methods, identifying fraud and transactional data, can be used to detect suspicious patterns. For instance, the contributed information is often too optimistic/pessimistic to be realistic. The contributed data can be compared with peers, historical data, external data, etc. to check the realism.

However, outlier detection is only the first step. The results reveal only exceptions and not necessarily inconsistencies: maybe there are reasons for a detected outlier. A sudden cost increase can be explained for example, because of a structural change. In the following interaction between contributor and manager, reasons for the outliers must be documented and exchanged. Learning algorithms are able to adapt the results to improve consistency checks for the succeeding planning rounds.

Typical benefits are

- Increased planning quality by detecting inconsistencies and controlled adjustments
- Open up discussions to avoid adverse behavior like sandbagging, etc
- Better mutual understanding about targets, expectations and limitations.

Driver Based Simulation

Forecast is the more passive way to draw a realistic view of the future. However, decisions can be made to influence future developments. Before making those decisions, an assessment of possible effects is crucial. Simulation is helpful when defining bundles of activities (the plan). The quality of the resulting scenarios depends on accuracy in modeling the dependencies.

Too often, managers trust oversimplified simulation models. For instance, the effect of sales price variation is only considered on revenue but the effect on volume is often neglected. This can be risky:

- Wrong decisions due to over/under estimation of developments
- Only a partial view of the decision is considered. For example, tax effects on cash are not or only roughly included
- No understanding of effects of potential risks on profit and cash. Average calculations usually hide potential risks

All management decisions about complex considerations need some kind of quantification. It is helpful, if a core financial model is available and can be part of the simulation.

The same machine learning methods detecting forecast figures can be used to identify cause and effects for simulation models. The aforementioned drivers derived from sales and costs forecasting can be used to enrich simulation models. A simple example is the effect of discounts on sales. If it is quantifiable a more realistic what-if analysis can be created. In consequence, financial effects of discount policies for instance can be evaluated. Bandwidths are also usually provided by machine learning and can be used to analyze the potential risks (and chances).

The results can improve the management capabilities for accurate decisions. Results from predictive analytics can be used to improve the quality of the scenarios.

- Creating meaningful business cases by better assessment of activities/projects
- Integrated view of potential risks
- More realistic view on your planning results by providing bandwidths

Conclusion and next steps

Implementing use cases for data driven decision making in the Office of Finance requires a holistic approach, as it affects not only technology, but also people and acquiring analytical skills. This is the foundation to successfully apply machine learning uses cases. By providing some examples we want to create awareness on where and how to get started.

If you didn't get started yet, make sure to create the foundation in 2020 by identifying the right people and put them together in a (innovation) project. Please make sure to take an integrated and finance approach to deliver data-driven decision making the right way, which need to include the financial impact of your decisions. A modern software platform that includes both granular operational and financial data needs to be implemented in order to forecast faster, manage better and act faster in the most reliable fashion.

A final remark: to gain momentum for data-driven decision making, data storytelling is very important. Business analysts and controllers need the communication skills and tools to have the data compose the story for them. Start small and build on the foundation to create business value for the future.



About the authors



Prof. Dr. Karsten Oehler

Prof. Dr. Karsten Oehler is a full professor of management accounting at Provdia School of Inter-national Management and Technology in Frankfurt and business development manager with Wolters Kluwer CCH Tagetik, a leading corporate performance management company based in Lucca. With more than twenty-five years working as a consultant for several international companies (IBM, Oracle, SAP, INFOR) he gained deepest insights in financial accounting processes and various software tools.

He published ten books on IT systems for financial applications, OLAP and advanced analytics (e.g. with Wiley, McGraw-Hill) and more than 160 articles on software systems for managerial accounting. He is a frequent speaker at conferences and seminars. Very recently he released the first book in German on advanced analytics for controllers.



Drs. Marco van der Kooij

Marco van der Kooij is Managing Director of ForSight Consulting and Domain Expert Digital Transformation in the CFO Office. He has over 25 years of experience in enterprise performance management and business intelligence. During his career he has been involved in more than 250 projects across Europe. Marco worked for several consulting firms and software vendors like Exact, Hyperion and Wolters Kluwer CCH Tagetik in various research & development, consulting, sales and management positions.

ForSight Consulting supports the CFO office to align its Enterprise Performance Management (EPM) capabilities with scalable finance technology to achieve the goal of sustainable value creation and to improve agility. We do this by helping them benefit from an enhanced understanding of EPM practices, knowledge of enabling and emerging technologies and leverage a network of more than 170 EPM experts. ForSight Consulting helps customers to uncover the opportunities created by an effective and efficient EPM function optimizing financial close, compliance, planning, forecasting, profitability analysis, performance management & advanced analytics to drive improved financial and operational results. ForSight Consulting also supports the CFO office orchestrate their finance & operational data and technology strategies to drive sustainable value creation through continuous modernization.