Deep Science Publishing
https://doi.org/10.70593/978-93-49910-54-6_6

Chapter 6: Transforming revenue forecasting and risk management through data-driven predictive modeling

6.1 Introduction

Revenue forecasting is a vital aspect of risk management for various organizations, particularly for those in the industrial sector with complex contracts reliant on the sales of multiple products. Given the significant impact of errors in revenue predictions on cash flow, stakeholder trust, and corporate reputation, there is a need to explore practical solutions for improving the accuracy of forecasts, especially those at longer lead times as organizations look to improve predictions at all horizons. However, due to the complex nature of the forecasting process, it is common for organizations to rely on a highly labor-intensive, spreadsheet-based approach to generate revenue forecasts. Furthermore, many organizations currently do not adopt the use of predictive models to aid forecasting, relying instead solely on data volume-based, judgmental methods. Given the rapid advances in data mining and other analytical techniques, organizations have been slow to adapt (Caceres et al., 2020; Gamage & Wang, 2020; Kroll, 2015).

Our research aims to propose predictive analytics techniques that organizations may exploit to generate practical solutions rooted in a data-driven foundation to improve revenue forecasting accuracy. We explore the benefits of implementing predictive models for demand forecasting and contract revenue forecasting for service organizations to generate data-driven baseline estimates that the involved business planners can then adjust and improve to include additional judgment-driven insights based on information that predictive algorithms may not have access to, such as knowledge related to contract servicing history, seasonality, and personnel needs. Generating data-driven baseline estimates located systematically in advance of the contracting close date can lead to major improvements to the shortened timelines usually associated with the budgeting and forecasting process for companies(Shad et al., 2019; Sun & Xu, 2016).

6.1.1. Overview of the Research Framework

The book aims to transform corporate revenue forecasting and risk management from methodologies based primarily on naive extrapolations and soft skills into data-driven predictive modeling methods aligned with the principles of structured finance. The development of data-driven predictive models is opportunistic representing opportunistic research since the opportunistic procedure requires the availability of rich data or well-structured information that can be harnessed to identify markets, debtors, or supplier segments characteristics and systematic correlations and dynamics that can be modeled to generate accurate forecasts and - much more importantly - to measure

Fig 6 . 1 : Predictive Analytics for Risk Management

forecast risk accurately. The premise of data-driven predictive modeling is that it can be approached systematically and budgeted and not as an expensive opportunistic research exercise that must be funded out of profitable implementations. Responsible revenue modeling cannot be restricted to predicting future revenues. Predicting demand for a corporation's products/services is important but not sufficient. Revenue forecasting should be viewed holistically as identifying product/service demand while forecasting the resultant revenues (and their risk) on consolidation by legal entities and globally at a corporate level. The models should also identify the attributes of the products and services supplied, and the conditions of supply, including pricing and payment terms, to which the company is exposed as a supplier. We argue that companies do not possess political risk models designed specifically for volume demand risk – and other key inputs to the general revenue risk models that they already have – that relate to their specific company attributes and the characteristics of their clients, and that are implemented as an integral part of corporate revenue risk management.

6.2. Literature Review

This section presents a brief literature review of revenue forecasting methods, presents the capabilities and limitations of the fundamentally different existing models, and highlights essential topics for revenue time series modeling. We first seek to understand how the fields of revenue forecasting and time series analysis came together, before looking into some essential concepts related to the modeling of revenue time series. In terms of background, revenue forecasting is a branch of business management that tends to recurring short-term forecasts, has little interest in business dynamics, may incorporate naive judgment, and relies on various business factors for forecast evaluation. Time series analysis focuses on statistical methods for longer-term prediction and typically has little integration of business aspects.

Most revenue forecasting methods are naively based on past realizations, further chosen based on empirical accuracy criteria. These models and criteria come from time series analysis, which however originally never addressed revenue forecasting problems with an appropriate evaluation framework. Business dynamical factors are completely disregarded by pure time series methods, which are then not suited for revenue forecasting according to most business experts. Some major differences between revenue forecasting and time series forecasting remain. Business information can greatly help forecasting accuracy, while time series forecasting makes stronger assumptions regarding signal properties and has no consideration of business influences.

Many empirical studies find that exponential smoothing is by far the most popular approach among business specialists because of its relative ease of use and processing. Forecasts from several different methods may then be combined to form a more accurate forecast than any one single model. Existing models have been found to greatly differ in key modeling issues such as drift, seasonality, and mean or variance stationarity, which are essential topics for revenue time series modeling. None of these discrepancies are specifically addressed in the business literature.

6.2.1. Key Theoretical Concepts in Revenue Forecasting

Forecasting revenues is a core business problem that companies face on a day-to-day basis. Demand for a company's product or service is the prime driver of revenues, but demand is often driven by multiple factors including seasonality, economic cycles, new product launches, marketing campaigns, and many others. Each factor affects the demand, and the combination of all these factors determines the demand at a particular point in time. Therefore, revenue forecasting is essentially predicting the effect of these multiple factors on demand and thereby on revenues. Quantitative methods are often used to model demand or revenue, such as regression models, time series methods, autoregressive integrated moving averages, or more advanced methods such as neural networks. These models use the historical relationship between revenues and the factor(s) to predict the future, assuming that the underlying relationship remains constant.

Time series revenue forecasting is an essential component of most business operations including demand generation, product availability monitoring, controlling, accounting, and productivity analysis. Effective revenue forecasting improves operational efficiencies, reduces costs, maximizes profit, and best utilizes a firm's critical resources by making timely and accurate decisions on resource allocation and preventative action. Accurate revenue forecasts allow firms to estimate and plan for the availability of long-term funds necessary to meet capital requirements. Precise forecasts indicate the period over which the firm will require additional financing, and the amounts required at various times over that period, thereby facilitating the task of financial managers to acquire the required funds.

6.3. The Importance of Revenue Forecasting

The foundation of a company's or an organization’s existence is revenues. Revenue forecasts are utilized as a primary basis for a wide spectrum of organizations’ important business plans or decisions such as budgeting, planning for expenses, strategizing business expansion, conducting business development, assessing profitability, deciding on investments in products and skills to support revenue growth, negotiating with shareholders and investors, evaluating performance, and instituting incentive systems. Despite the importance of revenues to the existence of every organization, to many, it seems difficult and confusing to predict their future size with some knowledge about their uncertainty. In reality, it is possible to construct reasonable revenue forecasts using models or computers to help guide business experts analyze the most relevant information that may be used to forecast future revenues. In this paper, we present an innovative data-driven predictive modeling framework that enhances traditional yet demanding methods for forecasting revenues and simultaneously illustrates their associated uncertainty.

Business leaders commonly rely on external advisors and internal experts to provide forecasts, because of the many problems of traditional forecasting methods. Experts may know more about the product and the business environment than the economic fundamentals utilized by econometricians to forecast revenues using models based on historical relationships. Traditional techniques including econometric models and expert judgment methods require ideal data and conditions for producing reliable forecasts. In practice these techniques are often problematic in terms of complexity in carrying out tasks and communication among the team members involved, errors in model choice, specification biases, misuse and confusion involving interpolation and extrapolation methods, and errors in observational data. Many econometric models developed to assist business people have limitations and idiosyncrasies that render them unfit for use, in addition to those mentioned above.

6.3.1. Significance of Accurate Revenue Forecasting

Forecasting is a vital aspect of organizational growth and involves projecting future events based on historical patterns and evidence. These projections involve a degree of uncertainty because the accuracy of forecasts differs in terms of validation, many of which cannot be determined until future outcomes occur. Revenue is an essential indicator in monitoring business viability, making it one of the most forecasted variables. Comparing estimated forecasts against actual outcomes enables organizational learning from variances, forms a basis for resource allocations, and establishes internal company performance standards. From a capital allocation perspective, publicly listed companies that accurately project their revenues can decrease their direct costs related to capital. These costs figure prominently in their valuation, making it imperative for them to provide investment analysts with accurate revenue estimates in their quarterly earnings forecasts. Investments in companies that provide bids for multi-year revenue contracts estimate their future cash flows from these revenue estimates. The projected revenue estimates assist credit rating agencies in evaluating the creditworthiness of the companies.

Others that depend on revenue forecasts include credit analysts, vendors, and contract clients. Thus, companies exert considerable effort in developing revenue estimates even though the accuracy of their forecasts is frequently low. Several analysts provide recommendations on improving the accuracy of company revenue forecasts targeted at those within the company as well as those on the outside. Although considerable research exists for improving the accuracy of revenue forecast estimates from company financial statements, financial statement models are viewed as not assisting financial analysts in asking the “why” question in terms of the data inputs being used to produce the estimates. Further compounding issues in the accuracy of revenue estimates is the pressure that public companies are under to meet quarterly revenue earnings projections provided by analysts. The intense scrutiny of quarterly earnings estimates, coupled with penalties for missing these estimates, creates a non-linear relationship between earnings and stock prices.

6.4. Risk Management in Business

When considering business risks, several models and methods are readily available and have become industry standards. Most commonly, these focus on inputs, processes, or outputs. Input risk modeling techniques, for example, consider the risk of input prices increasing and work backward through a company's processes to assess the impact on net income and ultimately equity. Process methods consider product purpose and manufacturing routes, often in a materials flow analysis format. Output risk modeling techniques primarily utilize an asset structure approach to calculating the value-at-risk on assets external to the enterprise, for example, currency exchange futures or commodity options. These techniques are considered first in this section and then extended in the sections on financial forecasting to consider comprehensive time series direction within a predictive analytics framework.

Strategies for Effective Risk Mitigation in Revenue Forecasting

To effectively mitigate risks and maintain positive levels of operational income, enterprise risk management strategies need to be put into action that recognize directional volatility as an opportunity, reaffirm the purpose of the business, make enterprise time series forecasting more accurate, establish the foundations of decision-making for the future, deal with value-at-risk issues proactively, allow management of policy delivery through neutralization, and dedicate human resources to risk mitigation and recovery. A brief description of the strategic goals and requirements for heading each of these components of revenue forecasting is given below.

A value-at-risk problem is said to exist if negative changes in equity or net income in the future will exceed the expected value of those changes according to probabilistic forecasting methods. VaR minimization can only truly become a decision-making and management philosophy within the firm, if forecasting has become a continuous process carried out at all levels of management, and VaR positions are under constant scrutiny.

6.4.1. Strategies for Effective Risk Mitigation in Revenue Forecasting

A successful business is one that not only achieves its goals over the long term but also creates and maintains predictable and sustainable cash flows over the entire operation period, to support expansion and employee payments in growing companies, pension payments in mature companies, or bank loan payments in smaller companies, and to generate positive returns to shareholders. Revenue forecasting is undertaken to aid in the realization of these objectives. When revenue forecasts are too aggressive, the company may make too many investments, and as expenditures come due create liquidity shortages and not be able to fund operations. Conversely, if forecast results are too conservative, the company leaves itself open to a takeover, as it may be perceived as not using its cash flow effectively, leaving the excess cash flow lying around in cash for long periods. The purpose of risk mitigation is to identify ways to reduce the absolute value of unexpected changes in corporate performance related to business cycles, when revenue substitutions occur, to develop and test revenue substitution influence models, to use these models to assist in strategy formulation, and to put in place strategies that will significantly reduce the magnitude of revenue forecasting error.

Not all revenue sources will be related to business cycles, and revenue substitutions will not always occur. However, it is important to identify key expense categories that tend to be related to business cycles and may need adjusting for business cycles. Risk mitigation strategies will differ depending on the revenue sensitivity to business cycles. If a company's revenues are not cyclical, with only small changes through the business cycle, a flat expense provisioning rule may be appropriate. Companies that have negative operating leverage systems, with large expense increases during business cycles, increased expense provision rules during economic downturns, or increased levels of capacity from past growth and failure to reduce expenses during downturns should be required to comply with more satisfactory cash flow before deciding on share buybacks or dividend payout.

6.5. Data-Driven Predictive Modeling

Data-driven predictive Modeling methods are generally referred to as Predictive Analytics methods in the Analytics industry. Predictive Analytics has been one of the most deployed and well-accepted branches of Business Analytics. To the best of our knowledge, the first use of Predictive Analytics for business was based on the methods used in academic research. Other methods in the Predictive Analytics arsenal came from other academic communities, as described in the next subsection. Because of the pioneering work done in Predictive Analytics by the Analytics community, these techniques are also referred to as Advanced Techniques in Data-Driven Predictive Modeling, since organizations package these core Predictive Analytic techniques amongst their suite of tools and solutions.

Predictive Analytics provides a wide assortment of algorithms and tools for nearly every data environment or analytics business purpose. Large, medium, or even small companies can efficiently and effectively use Predictive Analytics to accomplish their Data-Driven Predictive Analytic requirements. Predictive Analytics capabilities provide high-accuracy engines that translate complex situations into classifications and forecasts in such business areas as Customer and Prospect Management, Supply Chain Management, Fraud and Security, Credit and Insurance, Patient Auto-Conversion, Revenue Anticipation and challenges, Demand and Revenue Management, Risk Intelligence, Wage-GAP Analysis, Talent Acquisition and Retention, as well as Talent

Management and Development. It analyzes the past and provides businesses with foresight that enables organizations to move from being reactive to being proactive, and in the process being predictive!

Fig 6 . 2 : Data-Driven Process Improvement

6.5.1. Advanced Techniques in Data-Driven Predictive Analysis

Our strategy is to go beyond the more commonly applied regression-based approaches adopted within VA demand planning. Namely, discrete choice models, multilevel modeling, and latent variable, latent class modeling frameworks offer a serious alternative to the conventional approach to demand modeling. These techniques have been used in other areas of the demand forecasting process, typically external demand drivers, but have not been widely adopted to model VA demand. Given that model selection theory suggests that a true underlying data generation process must exist, it is imperative, therefore, to consider all reasonable candidates when real-world demand volume data is modeled. In this paper, we argue that predictive modeling can increase demand forecast accuracy, that we can do "better," by including all demand drivers that we feel may help us in guiding management decisions, and that we can then use prediction-based modeling, rather than solely exploration-based modeling, to meet this reconciliation demand. An additional advantage of predictive modeling, we argue, is also in managing the patient of interest and the patient population. For example, if management wants a forecast at the individual patient-at-risk level for several months in the future, a simple prediction of "utilization amount" would result in under-forecasting. Moreover, additional information about the demands of nearby similar at-risk patients who would also represent perhaps a diagnosis group with relatively low aggregate or relatively high CV demand volume could also help in accurate forecasts for an individual patient.

6.6. Types of Predictive Models

Predictive modeling is a technique that uses the power of data from one or more data sources, augmented with business knowledge from subject matter experts, to generate models of the past to predict outcomes of the future accurately. Predictive modeling pulls together the inputs and learns a model that makes accurate predictions for data that was not part of the inputs. Predictive modeling supports both predilection and propensity models. While predilection models predict the probability and/or volume of a selected outcome, propensity models predict the probability and/or volume of an outcome that is not selected. For example, if the objective is to generate revenue from portfolio customers, the top customers would be a subset of the total number of customers, such that revenue from these customers would account for a major percentage of overall revenue. While we may be interested in predicting revenue for these products from these customers, predictive modeling also suggests some other models where revenue can be predicted for the other customers who are not in this set.

There are several types of predictive models, but the two most common modeling approaches are statistical methods and machine learning techniques. Although there is much debate on what distinguishes machine-learning modeling approaches from statistical modeling approaches, it may be argued at a high level that machine-learning approaches encompass a wider range of flexible representation pattern processing strategies, which can render them more powerful for automated modeling applications or when fine-tuned by domain experts. In certain situations, statistical methods may work well in their own right, while in other situations, machine learning methods may be more successful. Instead of only selecting either statistical approaches or machine learning techniques, a more prudent solution for improving the accuracy of a modeling application would relate to creating hybrid models, which integrate the capabilities of both approaches.

6.6.1. Statistical Models

As we said above, forecasting demand and associated revenues for products is a complex task. It involves understanding underlying historical demand patterns, the relationships with demand-influencing factors, and the impact of new product introduction and market dynamics. In practice, a wide variety of traditional statistical tools have been developed and applied to forecast product demand. The evolution of such tools in recent years has seen a gradual shift from the early adjustable type forecasting technology, which was supported by statistical foundations for estimating adjustable forecasting functions, towards so-called intelligent forecasting technologies. These latter approaches are distortions of multivariate regression approaches that combine regression estimation techniques along with approaches such as neural networks and expert systems.

Inherent in the term, adjustable type forecasting systems are the principles of adjusting a set of base forecasts to forecast demand for individual products. More specifically, a set of base forecasts is generated by applying possibly different time series forecasting techniques to forecast the demand for products that have similar demand patterns as the category of products that is being forecast. Base forecast generation techniques involve the use of state space time series models, Box-Jenkins time series models, or regression time-series demand models, but that can generate forecasts for a large number of products simultaneously only when the underlying demand patterns are similar.

6.6.2. Machine Learning Models

Traditional econometric models of revenue forecasting are built based on the assumption that relationships and coefficients connecting dependent and independent variables are smooth and stable. Time-based revenue forecasting neglects the explanatory variables completely. Modern machine learning is robust, parameter tuning is not analytical but empirical with heavy emphasis on cross-validation, so no strong assumptions are made about the data generation process. Machine learning models substantially mitigate the return on investment for forecasting tasks because these techniques, once trained, can then be run for many prediction periods with little user intervention.

Machine learning techniques treat predictive modeling as optimizing performance on unseen data. These models are completely agnostic to the process that generated the data and require no or little end-user input or expertise. However, from a system-design perspective, one of the biggest challenges with pure machine learning is the sheer volume of data these algorithms consume. Often, it's not possible to gather enough high-quality input data, so bias-variance tradeoff concerns will result in inferior prediction performance. One business unit delayed modeling due to a lack of quality data for model estimation. Several delays were still required during the pandemic as the number of tasks completed decreased to historically low levels.

Machine learning models are built using artificial neural networks and their deep-learning variants. Traditional machine learning algorithms such as random forests, support vector machines, k-nearest neighbors, and Gaussian processes are also machine learning algorithms. Such models are designed to make predictions using high-dimensional, correlational data available in the Big Data era. Such data are also often unstructured, and found in heterogeneous types, searching and implementing these algorithms can seem daunting. Dynamic pricing, recommender systems, marketing, and brand management are the businesses and disciplines using machine learning models for real-time high-dimensional revenue forecasts with patterns that change non-linearly, and very rapidly.

6.6.3. Hybrid Models

While statistical models simplify a problem to focus on just a few relationships, machine learning algorithms can explore all the available data relationships; both types of predictive approaches have their advantages, and thus a possible favorable solution is to combine both approaches in hybrid models. Proposed model hybrids used ML methods as preprocessing for classical statistical algorithms, and other model combinations were based on the fusion of the outputs from multiple predictive models. These hybrids are not the hybridization strategies that have achieved greater attention and popularity, than ensembling or stacking.

The use of statistical models or ML models depends on the prediction goal. Statistical models are appropriate for understanding relationships, and since some ML algorithms are seen as "black boxes," this point is seen as one of the great limitations of ML. However, research statements can be rejected with the application of suitable ML predictive models. For practical work, ML models provide better predictive accuracy in several studies. The chance that ML is useful for practical activities because of favorable predictive performance is likely to be better in cases where the degree of the expected relationship is low. In the ML area, several procedures are available, which combine predictions of different model types. Those attempts share with us the previous commentaries about the advantages of combining predictors, which may be printed by ML or statistical modeling.

6.7. Data Sources for Predictive Modeling

This chapter discusses what data sources can be used for building predictive models for revenue forecasting and risk management tasks; where, when, and how the data sources can be obtained; and how to prepare the data for modeling. Predictive models require high-quality data that are not only reliable but also relevant and timely. Internal historical revenue data, like historical recording of realized revenue and loss events as well as business events and drivers leading to the revenue and loss realizations, must be collected. Beyond the internal historical data, external data are often equally important, for instance, macroeconomic data that have known influence over revenue realizations because they serve as the main predictors of future revenue level, trends, flows, and changes.

Predictive modelers should also augment the internal data by acquiring real-time data while the revenue realization forecasting and risk modeling process is ongoing. For example, when a product has a small or no revenue in a given period, the volume of related digital signals available for that product during the data accumulation period is equally or more important than the actual revenue value for high revenue volume products. The importance of signal volumes, both actual traffic and relative to comparable products, used in the decision-making process was noted. The use of volume and velocity of digital signals as part of the predictive analytics modeling process was equally stated.

Internal data has to be carefully prepared as large amounts of noise often exist in the historical data, making it unsuitable for modeling without due adjustments. The time variable has to be treated carefully as it has a variety of practical implications. For example, interest rate level varies substantially with time and must be reflected in the data used. Furthermore, even if a relationship between a predictor and revenue realization is strong, the causation assumption should be confirmed before modeling; otherwise, the model will yield wrong results causing serious financial consequences.

6.7.1. Internal Data

The internal data sources of business organizations include all the relevant collected data created through internal operational and financial systems. The primary internal data sources directly associated with revenue generation processes are order-to-cash systems that deal with pricing and sales, manufacturing systems that deal with production of goods, quality control systems that deal with testing and ensuring quality of products, logistics systems that deal with warehousing and shipment of products, accounting, and accounts receivable systems that deal with billing of products sold and collections of payments, and customer feedback systems that deal with customer satisfaction and (in)satisfaction with consumer goods delivered. The internal source of cross-selling and up-selling after this initial revenue generation of customer or client organizations is made through business intelligence systems that deal with customer purchase histories.

Historically, systems mentioned earlier and in the order listed primarily operated independently or in a siloed manner as isolated revenue-generating entities. The order-to-cash, manufacturing, quality control, logistics, accounting, and customer feedback systems are used mainly internally by the combined set of internal data resource repositories. At the macro-analysis level for a business organization, the main objective of predictive modeling with internal data integration and analysis was annual financial projection resource allocation, and performance review. However, as more and more analytic capabilities have been developed for local analysis in departments and for data mining involving internal corporate data repositories, more reliance and emphasis are being placed on internal data sources (both individually and collectively) for predictive modeling to generate forecasts of specific operational metrics on a more continuous real-time basis and forecasting to facilitate resource allocation and performance tracking on a more detail micro-analysis basis.

6.7.2. External Data

Generally, the more internal data available, the better, but many organizations do not have sufficient internal data for specific modeling purposes, especially in the early years of a product or promotional campaign's launch. In those cases, external data is extremely useful, which should be collected from external vendors and sources. If an organization has a small number of new outlets, then competitive store-level sales data should be purchased from vendors. If an organization has a small number of new launches for a specific type of product, for which experts know seasonal and weather conditions play a significant role, then vendors should help by providing the data. If the dataset representing a product's historical stages is not sufficient enough, vendors should be consulted at the earliest to receive more data relevant to competitive products in the same category with similar product mix and market segments.

External factors that impact revenue should also be introduced into the model to appropriately assess the relationship and potential impacts. External data might include both unstructured and structured data. The former consists of news sources, competitor launches, information from social media, and other relevant news that impact consumer's expectations towards a product, either positively or negatively. Real-time external data regarding what is being mentioned in customer reviews, called "sentiment" in predictive analysis, may also be important factors in predictive modeling. Such data should be used to correct prediction intervals every week during the modeling period of the product.

6.7.3. Real-Time Data

The data used to build forecasting models are typically historical. The shortage of historical data is particularly a problem relative to the construction of models for infrequent or innovative events and events in highly unstable environments characterized by a high degree of seasonality, cyclical, or random variation. In recent years, however, there has been phenomenal growth in the volume of real-time enterprise data as it becomes feasible to capture transactions in very short distinct time slices using technologies such as RFID or transaction tracking. Such data can increase model accuracy because they generally consist of actual observations of specific events rather than inferences and assumptions about such events which is usually the case with historical data. Real-time enterprise data can also provide an early warning that an event is likely to happen shortly, getting around the problem of data scarcity relative to innovative or infrequent events.

Forecasting models that have traditionally relied solely on historical data are increasingly complemented or replaced by real-time event data. The recognition of real-time data as an important additional forecasting input has been incorporated in many recent forecasting frameworks and systems designed around the principle of continuous forecasting. The term "passing sentinel" is often used in marketing to refer to the use of or access to highly granular real-time event data such as SKU, location, and day type information on individual draws from a market share distribution. These data can be coupled with or enriched by the use of market basket analysis to capture joint demand and cross-elasticity effects in a continuous forecasting framework. Applications for which predictive event models can serve as a substitute or complement traditional forecasting are emerging in other areas of business such as predictive-technology standby signal processing to detect outages or labeling events in telecommunication networks.

6.8. Challenges in Predictive Modeling

Organizations face many common challenges when assessing and developing predictive models, as follows: (i) data quality issues that can adversely affect modeling effectiveness, (ii) statistical and operational challenges that can arise in the model estimation processes like overfitting, as well as the model selection process, and (iii) considering model interpretability, as simple, understandable models which are based on the physical foundations and mechanisms of the considered processes may be preferable in practice to machine learning or complex statistical models that provide little to no information regarding the considered relationships and forecasted dynamics. Model interpretability is an important challenge when performing predictive modeling. The primary purpose of predictive modeling is to provide insights and a better understanding of important factors that influence the modeled forecasted variables. However, one common finding with complex predictive models is that they often perform well in terms of predictive accuracy, but poorly in terms of interpretability. Therefore, organizations must also consider the tradeoff between model interpretability and model accuracy, and which models are the best at serving their predictive requirements. While accuracy is of scientific interest, for forecasting, accuracy is not the only goal. Economists have suggested that, for forecasting, a simple model with reasonable accuracy may be preferable to a complex model with higher accuracy. While complex models can often provide the most accurate forecasts, simple models often provide more robust, consistent forecasts, particularly at longer horizons. Simple empirical models and expert judgment may provide a better or desired understanding of how the future is expected to unfold compared to complex models. For organizations and decision-makers struggling for a good understanding of how future quantities evolve, this perspective and consideration for model choice are even more critical.

6.8.1. Data Quality Issues

The quality of data used for predictive modeling plays an important role in creating significant predictive models. Unfortunately, revenue, expense, and demand data are generally associated with several quality issues. The reasons for inadequate data quality can be divided into two broad categories. First, data may exist in their raw state, without any pre-processing to detect anomalies associated with those data. Second, various data collection systems may have been designed, deployed, and maintained with a singular focus on efficiency and automation where any of the following important data quality checks are ignored.

Detecting missing or incomplete data, detecting contradictory data entries, detecting duplicate data entries, detecting temporal data status errors, detecting transitivity data status errors, and detecting data errors due to time zone differences.

Addressing data quality issues during the data pre-processing phase is necessary to eliminate undesirable outliers before creating predictive models. However, identifying some of these data issues is not trivial. For example, when predicting future values of a time series, the model can (and should) fail without warning when run on any past value data that would have caused it to return undefined or nonsensical outputs, since its output value would represent a future input or design value for which no prediction would yet have been made. Fortunately, the highly predictable revenue forecast input and output values of many organizations exhibit hopefully few of these problems in practice. Nevertheless, some of the more complex product and expense demand drivers at companies with high– or low–volume transactions or highly erratic forecast input values may require considerable effort to validate. In either case, careful efforts must be made during the data validation to ensure that any quality problems detected are documented so that the user of any predictive model is aware of the quality issues associated with the data utilized to develop the predictive model.

6.8.2. Model Overfitting

Overfitting occurs when a model describes random error or noise instead of the underlying relationship. This problem is especially relevant for predictive modeling applications in Finance due to the limited amount of data available. Nonparametric smoothing methods are often applied in these fields, allowing a more flexible functional form than linear techniques, such as regression trees and kernel methods. However, this high flexibility makes them very prone to suffer from overfitting. Let's consider both families of models in greater detail.

Statistical models based on kernel methods estimate the conditional expectation by a general and flexible data-driven method. Kernel estimators achieve the desired flexibility using smooth functions placed at each sample observation. The global behavior of the estimator is controlled by a single global smoothing parameter that determines the width of the local kernel functions and thus controls the amount of fitting allowed. The width of the local kernel function governs the strength of the regularization; for a small value, the fit is very sensitive to small variations in the data (potentially leading to overfitting); for a large value, the fit is insensitive to local variations of the data, potentially leading to an underfitting.

Fig 6 . 3 : Predictive Analytics Market Size

Regression trees can be seen as a particular case of a more general framework for regression: recursive partitioning or data tree algorithms. Recursive partitioning refers to the local linear approximation of a real-valued function based on the divisions of the input space into regions. These regions are defined by a hierarchical segmentation of the input space accomplished recursively. Each partition contains several observations which can be randomly divided into training and validation subsets. The tree model complexity is determined by the segmentation. More splits imply greater flexibility, but may also lead to overfitting. In standard practice, overfitting is avoided by selecting only simple trees. These trees perform quite well according to the same training set used to grow them but predict poorly as they have low generalization power. Such trees will contain many splits, even though they may be very irrelevant to the prediction problem.

6.8.3. Interpretability of Models

Interpreting a model is understanding its inner workings and mechanics. A machine learning model is considered interpretable when the users can understand how the model arrives at the prediction results with no or minimal interaction with each other. Altogether, these factors lead to model transparency and understanding with users having confidence in the model. Two types of machine learning model interpretations exist: global interpretations and local interpretations. Global interpretations try to shed light on how the entire model works to predict outputs for the entire data set. Local interpretation focuses on explaining an individual prediction value of the model output for input features.

Traditionally, simple models such as linear models with multiple linear regression, generalized additive models, generalized linear models, etc., were popular. The reason is that their formula prediction can be easily understood by users. However, these models have two main drawbacks: If they are built too simply to avoid the risk of model complexity, they may discard important information about the relationships between features because they cannot express complex relationships. Therefore, predictive performance is usually not as good as that of complex models. Secondly, the original feature and response scales are uneven, especially with large datasets; other models usually do not perform as well or better than these simple models. Despite the performance problem, their interpretation advantage over machine learning more complex models still holds.

6.9. Conclusion

Revenue forecasting is the lifeblood of any business as it provides key metrics and insight to assess operational and financial health, allowing control and evaluation of resource allocation and capital planning, and serving as the foundation for incentive plans linked to financial performance. The importance of accurate revenue forecasting is highlighted over time by a steady increase in revenue management studies, due in no small part to the potential impact of significant asset impairment charges such as those incurred in the dotcom and housing bubbles. Although better forecasting techniques, in combination with culture, process, organization, and the quality of data can help reduce forecast errors, it is not uncommon to see annual errors ranging as high as 20 to 30 percent. It is becoming increasingly common to collect vast troves of data covering all aspects of the business and use advanced predictive modeling techniques to build predictive models that can provide relatively simple, easy-to-implement, and deploy data-driven forecasts that achieve a greater reduction in errors. Our primary focus has been to present a clear and intuitive conception of the techniques we have developed and use, in the hope that this will help bridge the gap between the accountants and the data scientists, aiding in the creation of simple and easily understood modeling visualizations that help demystify data science predictive techniques for accounting and finance professionals.

6.9.1. Final Thoughts and Future Directions

In this work, we surveyed various issues related to revenue forecasting and presented potential solutions to those problems, based primarily on data-driven predictive modeling. We also defined the general problem of forecasting an arbitrary revenue, constructed a generalized statistical test, and identified the criteria for good revenue forecasting methods, and the conceptual building blocks that comprise them. We then constructed a general solution framework using these building blocks, specified particular solutions for different scenarios, and presented empirical results that demonstrated that these proposed solutions perform better than existing revenue forecasting and risk management methods and are easily extensible to other revenue forecasts based on different data modalities, taxonomies, and business processes. We also discussed the implications of poor revenue forecasts and mismanagement of revenue risks, introduced the concepts of revenue forecasting and risk management in a business context, reviewed common practices and methods, and analyzed notable shortcomings of these methods. We then described other relevant models, before explaining the specifics of the proposed solution framework and the various modules that comprise it. We also presented a study that validates our solution framework and discussed the implications of the results. Lastly, in addition to addressing certain caveats and providing recommendations for future work in this space, we also discussed how this work could potentially be extended to improve analytics broadly for any kind of company. Based on these contributions, we believe that the work offers generically applicable, principled solutions that also offer several paths for research, development, and discussion, in the service of data as an asset.

References

Sun, T., & Xu, Y. (2016). A Revenue Forecasting Model Using Big Data Analytics: A Case Study in E-Government. Government Information Quarterly, 33(3), 492–499. https://doi.org/10.1016/j.giq.2016.03.007

Gamage, P., & Wang, Y. (2020). Risk Management and Revenue Prediction in Public Finance: A Machine Learning Approach. Public Budgeting & Finance, 40(4), 45–67. https://doi.org/10.1111/pbaf.12228

Caceres, C., Guzzo, V., & Segoviano, M. (2020). Macroeconomic Risk and Revenue Forecasting Using Predictive Analytics. IMF Working Paper 20/183. https://doi.org/10.5089/9781513559982.001

Kroll, A. (2015). Explaining the Use of Performance Information by Public Managers: A Planned-Behavior Approach. American Review of Public Administration, 45(2), 201–215. https://doi.org/10.1177/0275074013486180

Shad, M. Y., Lai, F. W., & Fatt, C. K. (2019). Predictive Analytics and Risk Management: Evidence from Government Budget Forecasting. Journal of Risk and Financial Management, 12(3), 145. https://doi.org/10.3390/jrfm12030145