Predictive model building for driver-based budgeting using machine learning

Naveen Kunnathuvalappil Hariharan

University of the Cumberlands, United States

Abstract

Budgeting in the traditional sense is simply too slow and rigid to keep pace with the swiftly changing business environment. At the moment, there is far too much volatility, complexity, and uncertainty. A driver-based planning and budgeting model is more data-driven than a traditional budget model. This budgeting strategy shortens the time it takes to create a budget. Most driver-based planning and budgeting models center on predictions. One of the most difficult aspects of using driver-based planning, however, is identifying appropriate business drivers and predicting the impact of these drivers. Machine learning can assist driver-based budgeting processes in identifying the key drivers and predicting the impacts of these drivers. This study discusses the building of predictive modeling using machine learning. It illustrates stages from quantifying the budgeting issues to determining the best predictive mode for driver-based budgeting.

Keywords: *Driver-based budgeting, Machine learning, Model construction, Modelvalidation, Predictive model.*

Introduction

Budgeting is an essential part of financial success. Instead of reacting in a panic to unfavorable developments, many businesses now use the budgeting process to cautiouslyplan ahead of time for a variety of possible contingencies. Budgets are an essential component of management control systems, and when managed thoughtfully, they promote coordination and communication among company subunits, provide a framework for judging performance and facilitating learning, and motivate managers and other employees. (C Godley, 2016)

Every organization has plans and objectives that are derived from the organization's long- term strategy. The goal of budgeting is to assign financial values to those targets and plans, making progress easily measurable, and to translate strategic ideas into easy - to - understand operational actions (Réka, Ştefan and Daniel, 2014)

Traditional budgeting methods are inefficient and costly; they rarely focus on strategy and frequently contradict one another; they add little value; they prioritize cost reduction over value creation; and they strengthen vertical command and control. Although traditional budgets have evolved over time, it is now believed that they are incapable of dealing with changes in the economic and business environment and must be reshaped; alternative budgeting methods must be developed (Réka, Ştefan and Daniel, 2014).

Budgeting has come under increasing scrutiny in recent years. It takes far too long and thus costs far too much. However, it has additional flaws. Due to the rapid pace of change in manymarkets, the annual budget is obsolete almost immediately upon completion. That is why it iscritical to have the ability to re-forecast more frequently. Organizations must reassess the future on a regular basis; realign their operational plans and resources accordingly. Once a year, or even twice a year, is insufficient for the majority of organizations, and they are well aware of this (Barrett, 2007).

Budgeting, it has been argued, is a critical component of management accounting in the majority of organizations. Nevertheless, its contribution to the organization has been met withcriticism. While some believe it is necessary, others believe it is a pointless distraction. Libbyand Lindsay (2010) conducted additional research on the use of budgeting in organizations and its relevance and discovered that many organizations continue to use traditional budgeting despite widespread criticism from scholars (May, 2017).

Opponents of the traditional budgeting maintained that the traditional budgeting frequently results in dysfunctional behavior among employees, consumes management time, and impairsan organization's flexibility and adaptability in any given business environment. Additionally,it has been argued that budgets not only consume a significant amount of an organization's time and resources, but also fail to adapt to changes in a competitive business environment, thereby rendering the organization obsolete and uninteresting (May, 2017).

Driver-based budgeting

DBB (driver-based budgeting) is a technique that focuses on linking the performance of a business activity with a corresponding financial forecast. DBB improves overall transparencyby evaluating the business drivers that impact financial performance. This necessitates close collaboration between finance teams and business leaders in order to identify all relevant drivers affecting business performance.

Driver-based plans have an actual advantage over their simpler, trend-based counterparts in times of increased volatility and little historical precedent. While current popular movements may have sufficed in the past, results from the past may no longer be reliable predictors of future success in the increasingly unpredictable economy of today (Bahnub, 2010). When activity is heavily influenced by market forces or economic shocks, an increased reliance on external economic drivers should also be considered.

Driver-based budgeting models line item expenses using both non-financial and financial driver data. Drivers vary by industry and even by company, and it may appear as though thereis no one-size-fits-all definition of what it means to be a driver (Leon, Rafferty and Herschel, 2012). However, a working definition has been given as follows: In driver-based planning and budgeting, a driver is a piece of non-financial or financial data that, when modified, has adirect effect on either revenues or expenses, thereby affecting the forecast profit and loss account, cash flow, and balance sheet.

There are different types of drivers. Some of his instances have been reported below. Different types of drivers if we limit our attention to factors that have a direct impact onrevenues or line item expenses; we see that there are numerous different types of drivers usedin planning and budgeting. Among them are the following:

Rates of consumption, productivity, or cycle times; these indicate the quantity of resources required to meet demand or produce one unit of output. Examples: • Simple productivity ratios, • The ratio of staff to supervisors. Costs of unit resources; the average cost of a unit ofresource over a specified time period.

For instance, the cost of a liter of gasoline; the average salary of a particular grade of staff; and the anticipated cost to replace a desktop computer. Unit selling prices: The average price at which a product or service is sold. Examples include the average premium for a specific type of insurance policy, the anticipated fee for each consulting engagement, and the anticipated selling price for a specific product.

Quantitative indicators of demand; this includes both the forecast level of demand for the goods or services sold to clients and the level of demand experienced by specific departments. Examples: The number of customers who are currently active • Market size and market share • the quantity of a product sold • The quantity of items in a sales order.(Barrett, 2007), (Kale, 2014) , (Suveera, no date),(Saporito, 2014).

.

One advantage of driver-based planning and budgeting is that it shortens the time it takes to create a budget or re-forecast. This is without a doubt one of the most significant and appealing advantages, as there are frequently tangible cost savings associated with driver- based budgeting (Barrett, 2007).

Moreover, budgets based on driver aid in overcoming the calendar year fixation. Many of the relationships and rules in driver-based planning and budgeting models span multiple time periods (Collier and Agyei-Ampomah, 2005) . One of the benefits of systematically modeling resource requirements and comparing them to the amount of resource actually supplied is thatexcess capacity is instantly calculated and addressed directly across the organization (Rael, 2017).

According to the literature, another advantage of driver-based budgeting and planning is thatit saves resources (Rael, 2017), (Barrett, 2007). There may be some cost savings because the time needed to complete the yearly budgeting cycle or produce a mid-year re-forecast is decreased and may require less iteration. Typically, this occurs in the finance department, where either the quantity of extra hours needed during the annual budgeting cycle is reduced or one or more junior positions can be eliminated or redeployed to more beneficial tasks (Cokins, 2017) (Risk Management Institute Singapore, 2014) (Kale, 2016)

One of the most difficult aspects of using driver-based planning is identifying appropriate business drivers and understanding the impact of these drivers. In practice, determining specific drivers is often difficult due to the growing volume of business data, the complexity of today's business operations, and the rapid changes in the business environment (Risk Management Institute Singapore, 2014) . The identification of drivers and the determination of cause-and-effect relationships necessitates the provision of high-quality information from all organizational units (Andersen *et al.*, 2008), including financial and operational units, so that it can be used in an integrated manner to make accurate predictions.

A massive amount of data from various sources that is constantly growing (Lee, 2021), as well as the need for filtering, organization, and integration, may pose a significant barrier to identifying key drivers, conducting analysis, and reaching meaningful conclusions. Because this is critical to successfully implementing driver-based planning, it is clear that incorrectly defined drivers can result in poor budgeting and forecasting. When developing a driver-based model, it is critical to avoid overcomplicating and over-detailing. There is no reason toinclude variables that do not provide analytical benefits. It is more important to develop a model that is accurate, actionable, and focused on key performance drivers.

Most driver-based planning and budgeting models begin with a demand measurement. This could be a market-based model in consumer markets, with market size, market growth, and market share driving sales volumes and demand across the board. Organizations competing inbusiness-to-business markets may begin by viewing sales and marketing activity as the primary driver of demand for their model (Barrett, 2007).

Machine learning

Machine learning is becoming a progressively important tool in business, to the point where almost all employees will be impacted in some way over the next few years. Machine learning is concerned with learning the relationships between variables, making predictions, and making decisions in a changing environment using large data sets.

The amount of data available for machine learning is increasing at an exponential rate. It is estimated that over a two-year period, we generate nine times the amount of data that existed at the start of the two years. Companies now have more information about their customers and their purchasing habits than ever before (Maruster, 2003). Advances in computer processing speeds and lower data storage costs enable us to process this data and draw conclusions in ways that were previously impossible.

Machine learning is an artificial intelligence subfield (AI). AI is concerned with developing methods for machines to mimic and possibly improve human intelligence. Machine learningis the process of creating intelligence by learning from large amounts of data. It is arguably the most exciting advancement in AI, with the potential to transform virtually every aspect of a business.

There are societal benefits of replacing human decision-making with machines. One advantage is that it is faster. Machines can process data and reach conclusions much more quickly than humans. A machine's output is consistent and easily replicated on other machines. Humans, on the other hand, occasionally behave erratically, and training a human for a task can be time-consuming and costly (Sammut and Webb, 2010).

Building predictive models with machine learning

The first steps in developing a predictive model for driver-based budgeting using machine learning is to quantify the budgeting challenge, that is, to express what the company is seeking to accomplish as a single number of score. The first stage requires a detailed understanding of what the company wants, and then explains this in terms that can bereflected by the forecast scores provided by the model they create (Safar *et al.*, 2006).

The second stage of building predictive models for DBB involves conducting exploratory analysis to understand the organization's data and IT assets, how they work, and how they interact with each other and the rest of the business, in addition to determining how to articulate the business aim quantitatively. This includes the following (Laud and Ibrahim, 1995):

a)The type of databases the company is equipped with b) the purpose of each database, and how are they used in business processes. c) The contents of the databases i.e. what data pieces and how those data items are formatted. d) The methods used to update and maintain each database Most businesses have a mix of “real-time” client databases that are updated as soon as something changes and batch databases that are updated less regularly, such as at the end of the month. e) The systems of decision-making already in place. c) Determiningwhether the model be integrated into an existing system or necessitates the development of new functionality. f) The information the company currently utilizes to make decisions. Even if this data is used to feed an archaic manual process, it is likely to be incredibly relevant for machine learning purposes because the company has previously designated it as useful.

Stage three includes constructing the machine learning „‟development sample‟‟. The stage three occurs once a data analyst has a strong knowledge of what data a business has and how it is structured, they may

To develop predictive models, one naive method is to just put every piece of data anorganization has into the machine learning process. Much data is out-of-date, unstable, or otherwise improper in practice (Olivera *et al.*, 2017). As a result, it should be eliminated fromthe development sample if it degrades the machine learning process' performance.

The following are some common reasons for excluding data from the machine learning process: a) Data that is no longer valid. Data that is no longer relevant to today's world shouldnot be considered. Data on how people bought a particular product before an event isn't going to help understand how people buy the product today.

b) Data consistency. When the model is implemented, it is crucial to check whether the data utilized in the machine learning process is available. A data item should be excluded if it is not available at the time of implementation.

c) Inexplicability. A key premise of machine learning is that it is necessary to avoid developing models with data that is completely comprehensible. This is especially true for models that are regulated by the industry or are strategically crucial to a company's success. Ifa data analyst comes with a model based on data that they are unable to explain.

The process's fourth stage is data preparation. This is necessary because the raw data may be incorrectly formatted, contain errors, or do not provide the best possible representation of the data; a different depiction will make it much easier for the machine learning process to identify significant patterns in the data. During data preparation, four primary tasks occur (Barbieri and Berger, 2004):

1. Generating new data. For instance, for the majority of problems, if x is more predictive than y, as a result, if only the y is available, x may need to be calculated from y, if they are linearly related. If one has credit card transaction data for the customers, it is best to create summary variables that represent the average spend per transaction, the number of transactions last month, and the time since the last transaction, among other things.

2. Cleaning of data. There may be omissions or errors in the data during the Driver-based budgeting process. A primary goal of data preparation is to identify instances of this type of data and to either eliminate it (an exclusion) or reformat it; that is, to replace all incorrect/missing data with a standard value.

4. Numeric conversion. Algorithms for machine learning utilize numerical data. Instead of "Yes"/"No"/"Maybe," "Yes"/"No"/"Maybe" response data from consumers of a particular item would be converted to 0/1/2 flags. When dealing with complex text or speech, such as a Twitter feed for the customers of a product, one approach is to include flags indicating whether or not certain words appear, or to start generating counts of the number of times certain words show up.

The fifth stage to build a predictive model for DB budgeting involves pre - processing and preliminary variable choice once an appropriate data sample has been collected. Pre- processing data entails transforming it into a format that is most amenable to the machine learning algorithm being used. Typically, pre-processing entails two steps: standardization and transformation. Standardization processes normalize data so that it has comparable values(Cockburn, Gutwin and Greenberg, 2007).

Variable selection (variable reduction) is the process of determining which data elements to discard and which to retain and present to the machine learning algorithm. The primaryreason for this is that, despite the incredible power of today's computers, there is frequently too much data to process in a reasonable amount of time. Frequently, there will be thousands, if not millions, of individual data items created during data preparation and pre-processing. Typically, only a few of these prove to be materially significant in terms of generating accurate predictions for driver-driven budgeting.

Variable selection is typically a quick and straightforward process. Numerous statistical tests are run to determine the degree to which each piece of observation data is correlated with the predicted outcome. Data items are retained only if a correlation is established. Typically, this process retains between 1% and 10% of data items. Additionally, there are methods such as Principal Component Analysis (PCA) that attempt to reduce a large set of initial data items toa manageable number of new ones. These new data items are then used to construct the model.

Model construction is the sixth step in the process of making a predictive model for driver based budgeting. For a model building process, this is frequently the most notable feature of the machine learning process. However, that model construction accounts for only a small portion of the overall end-to-end process. Model construction begins with the presentation of development data to machine learning software, followed by the execution of the appropriate algorithm(s) and the generation of a predictive model. Typically, this would resemble scorecards, decision trees, and neural networks. Machine learning software in the modern era is highly automated. At the push of a button, a diverse array of algorithms is available. Numerous others include features for data preparation, pre-processing, and variable selection.Without the necessary technical background, it is possible that the data will not be properly prepared or preprocessed. Similarly, there are frequently a variety of different options available for a given algorithm, and the finance analyst ought to understand how each option affects the final solution. Similarly, one may be unable to interpret the software's outputs orbe unaware of them. However, it is entirely possible for the software to report (correctly) that the model created is extremely accurate when, in fact, the model is useless.

After developing a predictive model for budgeting using the development sample, additional work is required, at the seventh stage to evaluate the model's performance. From a technical standpoint, this primarily entails evaluating the model's predictive accuracy across different data samples. These should be completely unrelated to the data used to build the model in order to provide an accurate and unbiased assessment of the model's performance when applied operationally to new cases. If the development sample was created several months ago, the model should be evaluated using more recent data to ensure it continues to performas expected. While a data analyst may report model performance based on the development sample, all final model performance assessments should be based on one or more independent validation samples.

Other validation activities may be more qualitative in nature, to ensure compliance with any applicable legal restrictions on the use of specific types of data and to make sure that the data used during the model will be available when the model is applied; for example, if the model is going to be incorporated into one of the organization's systems, the data required to calculate the model's scores must be available (Carrion Schafer and Wakabayashi, 2012).

Iterative processes are required for machine learning once a predictive model for budgeting has been developed. The last stage involves these iterative activities. Often, several modelsare constructed using varieties of different algorithms and/or different representations of the data before settling on a final model. With additional time and resources, it is always possibleto improve predictive accuracy by a small margin. However, time and resources are limited inthe real world. As a result, a decision must be made regarding when to terminate the machine learning process.

It is crucial to review the developed models and choose one to be the final model. The final predictive budgeting model should be evaluated in terms of performance indicators and bottom line benefits from a business perspective. These should have been taken into account from the start.

Conclusion

It's no surprise that the traditional, out-of-date approach to budgeting and planning is in desperate need of an overhaul. While it is time-consuming and demanding, the true tragedy oftraditional budgeting and planning occurs when decision makers are presented with incomplete or inaccurate data. That is because business changes can render plans quickly obsolete and, in some cases, unusable.

These difficulties are compelling financial organizations to reconsider their approach tobudgeting and planning. And an increasing number are now adopting driver-based budgeting.Rather than taking a bottom-up approach, DBB takes a simplified top-down approach that focuses on key business drivers. These critical drivers serve as the information hub that enables leaders to make informed decisions efficiently.

Machine learning applications based on predictive models are increasingly being used to supplement and/or replace expert judgment and manual decision-making in a wide variety of fields. This is because predictive models are typically more precise and objective than human experts.

References

Andersen, E. *et al.* (2008) *Microsoft Office PerformancePoint Server 2007*. John Wiley &Sons.

Bahnub, B. J. (2010) *Activity-based management for financial institutions: Driving bottom-line results*. Nashville, TN: John Wiley & Sons (Wiley and SAS Business Series, 38).
Available at: https://play.google.com/store/books/details?id=bAssw0m0esIC.

Barbieri, M. M. and Berger, J. O. (2004) „Optimal predictive model selection", *The Annals ofStatistics*, 32(3), pp. 870–897.

Barrett, R. (2007) „Planning and Budgeting for the Agile Enterprise: A driver-based budgeting toolkit".

Carrion Schafer, B. and Wakabayashi, K. (2012) „Machine learning predictive modeling high-level synthesis design space exploration", *IET computers & digital techniques / IET*,6(3), p. 153.

C Godley, A. (2016) *Budgeting as a valuable tool in preventing unfavorable business developments*. Тернопіль, ТНЕУ. Available at: http://dspace.wunu.edu.ua/bitstream/316497/5055/1/Andrew_C._Godley.pdf.

Cockburn, A., Gutwin, C. and Greenberg, S. (2007) „A predictive model of menu performance", in *Proceedings of the SIGCHI Conference on Human Factors in ComputingSystems*. New York, NY, USA: Association for Computing Machinery, pp. 627–636.

Cokins, G. (2017) „Strategic business management: From planning to performance".

Collier, P. M. and Agyei-Ampomah, S. (2005) *Management Accounting-Risk and ControlStrategy*. Elsevier.

Kale, V. (2014) *Inverting the Paradox of Excellence: How Companies Use Variations forBusiness Excellence and How Enterprise Variations Are Enabled by SAP*. CRC Press.

Kale, V. (2016) *Enhancing Enterprise Intelligence: Leveraging ERP, CRM, SCM, PLM,BPM, and BI*. CRC Press.

Laud, P. W. and Ibrahim, J. G. (1995) „Predictive model selection", *Journal of the Royal Statistical Society*. Available at: https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1995.tb02028.x.

Leon, L. D., Rafferty, P. D. and Herschel, R. (2012) „Replacing the annual budget with business intelligence driver-based forecasts", *Intelligent information management*, 04(01), pp.6–12.

Maruster, L. (2003) *A machine learning approach to understand business processes*.Citeseer.

May, A. U. (2017) „Traditional budgeting in today"s business environment", *Journal ofApplied Finance & Banking*, 7(3), pp. 111–120.

Mohri, M., Rostamizadeh, A. and Talwalkar, A. (2018) „Foundations of machine learning".

Olivera, A. R. *et al.* (2017) „Comparison of machine-learning algorithms to build a predictivemodel for detecting undiagnosed diabetes - ELSA-Brasil: accuracy study", *Sao Paulo medicaljournal = Revista paulista de medicina*, 135(3), pp. 234–246.

Rael, R. (2017) *Smart Risk Management: A Guide to Identifying and Calibrating BusinessRisks*. John Wiley & Sons.

Réka, C. I., Ştefan, P. and Daniel, C. V. (2014) „TRADITIONAL BUDGETING VERSUS BEYOND BUDGETING: A LITERATURE REVIEW", *Annals of the University ofBucharest. Mathematical Series*.

Risk Management Institute Singapore (2014) *Global Credit Review - Volume 4*. WorldScientific.

Safar, J. A. *et al.* (2006) „Meeting business goals and managing office bandwidth: A predictive model for organizational change", *Journal of Change Management*, 6(1), pp. 87–98.

Sammut, C. and Webb, G. I. (2010) *Encyclopedia of machine learning*. 2010th edn. Editedby C. Sammut and G. I. Webb. New York, NY: Springer.

Saporito, P. L. (2014) *Applied Insurance Analytics: A Framework for Driving More Valuefrom Data Assets, Technologies, and Tools*. Pearson Education.

Suveera, G. (no date) *Cost and Management Accounting: Fundamentals and its Applications*.Vikas Publishing House.