Improving Profitability Using Predictive Analytics

Bhaskar Roy [1]
Debabrata Bera [2]
Praveen Kumar Tripathi [3]
S. K. Upadhyay [4]

Abstract

The objective of this paper was to explore a better pricing strategy by predicting contribution margin (CM) to drive wins at higher prices. This paper focused on the services industry, where macroeconomic factors play a key decisive role in arriving at contribution margin to win more deals in this competitive market. The paper incorporated prior research findings to develop a multidimensional and multifaceted framework depicting the methodology of formulating customer value for value-based pricing. The empirical portion of this paper contained a case study based on masked industry data from an industrial manufacturing company dealing with products and services. We discussed and highlighted the criticality of identifying and capturing the right features while creating the right pricing strategy using multiple linear regression and decision tree techniques. Applying the predictive analytics approach helped us estimate the contribution margin with a higher winning probability during contract negotiation. This paper would aid organizations to develop and implement an enterprise-wide strategic pricing discipline designed to bolster the value and impact of their products and service pricing.

Keywords : **Regression, predictive modeling, contribution margin, multicollinearity, value-based pricing, customer value, pricing strategy, business-to-business, decision tree, strategic pricing**

Paper Submission Date : January 10, 2021 ; Paper sent back for Revision : March 10, 2021 ; Paper Acceptance Date : May 7, 2021 ; Paper Published Online : August 30, 2021

Even a marginal pricing improvement, among other operational management levers, can boost profitability significantly. Effective pricing tactics and strategies can easily deliver a 2% – 7% increase in return on sales (Chan et al., 2014). Among all the existing operational management tools, pricing seems to be the most beneficial for achieving profit growth. The famous marketing mix has only one crucial element – 'pricing' that actively generates revenues, placing it firmly as an integral marketing component (Kotler & Keller, 2014). Pricing, therefore, is a business management mechanism that can have a very powerful impact.

Considering the business impact of pricing on profitability, one can conclude that a deep understanding of pricing strategy is a crucial competency within a business. Usually, there are three broad categories of pricing strategies : cost-based, competition-based, and value-based. Successfully implementing and practicing

[1] *Research Scholar (Corresponding Author),* Department of Statistics, Banaras Hindu University, Varanasi - 221 005, Uttar Pradesh & [1] *Vice President* - Genpact. (Email : bhaskar1974.br@gmail.com) ;
ORCID iD : https://orcid.org/0000-0003-3922-1583
[2] *MPhil,* International Institute for Population Science, Mumbai - 400 088, Maharashtra & [2] *Senior Manager* - Genpact. (Email : debabrata.bera@gmail.com)
[3] *Assistant Professor,* Department of Mathematics and Statistics, Banasthali Vidyapith, Banasthali - 304 022, Rajasthan. (Email : pkt2088@gmail.com)
[4] *Professor,* Department of Statistics, Banaras Hindu University, Varanasi - 221 005, Uttar Pradesh.
(Email : skupadhyay@gmail.com)

DOI : https://doi.org/10.17010/ijom/2021/v51/i8/165759

value-based pricing strategy within business operations will enable an organization to stand out from the rest. Industrial companies with a business-to-business (B2B) model often struggle with value-based pricing strategy as most of them fail to understand the concept behind it or, in other words, do not know what values to offer to their customers. The B2B concept requires organizations to understand the criticality of the buyer-seller relationship. It is the customer value that would eventually formulate a long-term product of the relationship.

Nowadays, most companies are analyzing their offerings and customers and identifying the most profitable ones. Contribution margin (CM) represents the revenue that contributes to profit after a company reaches its break-even point, at which it meets the fixed costs, and its sales become profitable. These profits contribute to paying overhead costs or generating further profits for the organization. Pricing strategy is not about proposing a price with a higher margin but more about identifying the sustainable and realistic CM that has the maximum winning probability in a competition.

During our research and detailed literature review, we find that almost all businesses wish to improve their profits and margins, but they struggle to create any optimum strategy due to a lack of a data-driven approach, which could potentially guide the business on focus areas. It is quite evident that multiple factors could drive profits, and it is imperative for an organization to understand the levers of profitability and its application using the correct methods and techniques.

This paper demonstrates the use of predictive analytics-driven approaches along with the traditional regression-based pricing model to estimate a CM-driven price against the user-defined attributes based on historical data.

Literature Review

Most businesses have varied marketing goals and objectives and could implement multiple pricing approaches for their products and services. Businesses use these approaches to gain market share in a specific geography or to defend their current prices in an existing market from new entrants. In this section, we have discussed various pricing approaches for a service or product business and their relative merits (or demerits) and identified an alternative approach to overcome the problem of pricing optimization. We have tried to briefly discuss the most used pricing techniques that are crucial for the understanding and continuity of this paper.

Traditional Approaches of Pricing

Traditional approaches utilize different pricing schemes to boost the profitability of a service (Simon & Fassnacht, 2018). Various approaches have been discussed in this section.

- **Cost - Plus Pricing :** This method requires a business to simply consider the cost of providing a product or service and the desired profit to determine the final cost. While calculating the total cost, businesses consider all the direct, indirect, and fixed costs associated with a product or service. They also build in their overhead costs along with a portion of the overall rent, utilities, and administrative costs and then arrive at the final cost of any product or service. This pricing model assumes that unit cost can be determined before price, and demand does not impact prices (ultimately affecting production costs). Thus, using the cost-plus pricing model could limit profitability because of inaccurate pricing (Armstrong et al., 2014).

- **Competitors' Pricing :** This method is essentially based on competitors' costing of a similar product or service in the market. Winning customers while competing via price depends entirely on customers' loyalty when they are receiving the same service at a lower price (Daly, 2002).

- **Price Skimming :** Under the price skimming strategy, a product or service is first offered at the highest price that customers are expected to pay. The businesses then drop the price sequentially until it reaches a level where it is viable for the long term. Well-known mobile handset manufacturers such as Samsung and Apple use this strategy while launching their new products (Spann et al., 2015).

- **Discounting :** This pricing strategy is used by businesses to sell low-priced products in large quantities. To remain in the competition, companies usually include cost-cutting as part of this business strategy. The discount pricing strategy proves beneficial while driving traffic and sales for a short-term duration. Usually, businesses avoid implementing this strategy for the long term to evade its adverse effects on their market position and brand loyalty. This strategy includes festival season discounts, pre-payment discounts, and volume discounts (Chan, 2017).

- **Price Bundling :** Sometimes, an 'all-inclusive reduced price' is offered for products or services that are a comprehensive package of several products or services. This can boost profits even though items are sold at a discounted price because this strategy promotes a simultaneous purchase of multiple items. An example of this pricing strategy can be software bundling containing multiple applications (Liberto, 2021).

The above-mentioned pricing strategies are not always beneficial, and there are some serious disadvantages of using them. For instance, even after applying the cost or competition-dependent pricing strategies, a business might not be able to match the value proposition expected by customers. The business may end up undercharging or overcharging its customers, thereby affecting its revenues and profit margin.

Scientific Approach : Value-Based Pricing

Through this pricing strategy, businesses are known for setting their product prices primarily, but not exclusively, based on the value perceived or estimated by customers rather than the products' cost. This strategy is called 'value-optimized pricing.' The value here pertains to the customers' perception of benefits received from the price paid. Thus, the value-based pricing is an actual win-win scenario for both the producers and customers. As this strategy is developed through the deep understanding of an offering's value from a customer's standpoint, manufacturers can use it for hiking product prices in a manner that would limit sales loss. The most suitable approach for a company aiming to optimize its profitability would be a value-based pricing strategy. A study conducted by Tatyana (2011) presented the price formation, based on product value, as a source of competitive advantage. The value and the product price for the customer are derived as the 'key factors' for success of the company's competitive strategy. The value of the product in the marketing and pricing is revealed by this method.

As per the Monitor group, the companies that developed and used effective pricing strategies based on value achieved 31% higher operating profit than companies that focused on price to gain more market share or higher target profit (Netseva - Porcheva, 2011).

Customer value is a multidimensional and multifaceted construct consisting of several dimensions such as operations, proficiency, strategic, symbolic, social, and environmental influences. Customers' buying behavior varies depending on the product price (Jain & Kamble, 2021). Product innovation attributes are the key influencers of customer satisfaction, which are significantly influenced by brand loyalty (Panigrahi et al., 2021). Wroblewski and Loss (2018) discussed value-based pricing to conceptualize the dimensions of customer value in a B2B context in their research paper. Customer value is somewhat similar when companies use the term 'value of the product' (Nagle et al., 2011 ; Panigrahi et al., 2018).

Although companies strive to apply value-based pricing as their main strategy, they often implement other pricing strategies due to the complexity of the value-based pricing process and lack of understanding of the customer perceived value of the product. There are currently limited references available of possible applications of value-based pricing in global organizations. As only fragmented research studies have been carried out on pricing, there is no reference within the existing knowledge about pricing from a strategic perspective (Björk & Forsberg, 2015). It is, thus, very pertinent to understand various frameworks of value-based pricing to create an effective pricing strategy (Armstrong et al., 2014).

➷ **Economic Value Estimation :** A product's reference value and differentiation value together constitute its economic value. To determine it, start by determining the price of the product as set by the next best competitor(s) and label it the reference value. For example, the reference value of a hotel room is the price charged for the next-best hotel in town. If we are referencing the value of a smartphone, it would be the price of a comparable smartphone of another brand. Similarly, the net benefits of a product or service delivered to customers, over and above those provided by competitors, constitute a differentiation value (Claessens, 2018).

➷ **Van Westendorp Price Sensitivity Meter :** The Van Westendorp pricing model, also known as the price sensitivity meter (PSM), is a method for gauging consumers' perception of a service or product value and is widely used throughout the market research industry. Westendorp encourages the use of PSM by combining open-ended questions with price and quality. Since price is assumed to reflect value or quality, this technique is unhelpful for a true luxury good (SurveyKing, 2021a).

➷ **Conjoint Analysis :** Whenever a business launches a new product or service, it must be aware of the product's or service's value as perceived by its customer. These days, most products and services are launched with a host of attractive features to influence customers. However, it is crucial to ascertain specific aspects of these attractive features for most customers. Hence, marketing leaders prefer to run conjoint analyses to understand customer preferences better. Conjoint analysis is a survey-based statistical technique aimed to determine customer perception of attributes such as features, function, and benefits in a product or service (SurveyKing, 2021b).

All these above-discussed frameworks are often used to understand customer insights and depict their willingness to pay. These frameworks focus on the primary attributes that drive customer value and establish a relationship between a product's attributes and its sales price.

Predictive Analytics Techniques Used in Value-Based Pricing

For almost all retail companies, pricing has become a key focus area, given that competitive pressures and pricing wars are driving today's markets. In the current scenario, pricing is the single biggest profit lever for retailers and the most powerful tool to define product and brand positioning. According to the PwC analysis on "Driving Retail Growth by Leveraging Analytics" (2016), pricing analytics may use any of the six different approaches based on the product life cycle : (a) price setting, (b) pricing structure, (c) profit leakage, (d) cross-/up-selling, (e) promotional pricing, and (f) markdown. Apart from these approaches, many marketing and pricing professionals use regression analysis to calculate price points to optimize profitability.

Predictive models can help retailers analyze customers' past performance to assess their likelihood of exhibiting a specific behavior, helping them improve their marketing effectiveness. Besides using data mining techniques, businesses can segment customers into groups with similar shopping behavior based on which each group can be targeted with discounts or rewards. Campaigns with tailored promotions are developed by gaining deeper insights into customers' baskets and determining their lifestyle and price sensitivity. By analyzing products' types and price ranges within a basket, businesses can group customers into a specific lifestyle segment.

A quantifiable relationship between attributes and price enables businesses to use price optimization techniques that match the products' value as perceived by the customer. The most commonly used method to obtain a quantifiable relationship between driver and dependent variables is regression analysis. Regression analysis can be used to determine the correlation of transaction price to each characteristic, such as physical characteristics and other external influencing elements that may add or subtract from the building value (Monson, 2009). Also, the price-response function can be estimated using econometric techniques such as linear or nonlinear regression. The price-response function specifies demand for the product of a single seller as a function of the price offered by that seller (Bodea & Ferguson, 2012).

Most researchers used famous regression techniques to establish this relationship between price and its factors in the past. One of the key assumptions of linear regression is normality assumption, which may be incorrect in several practical scenarios. In the current competitive market, price drivers are becoming multidirectional. In this paper, we will discuss how the integration of regression techniques, along with decision tree classification, can help determine the most optimized price and understand the most prominent features or attributes that significantly impact the profits of a product or service.

Research Methodology

Multiple Linear Regression Model to Determine CM and Its Drivers

It is often the case that one is required to think of a model that contains more than two variables because the univariate or bivariate models often reflect the simple and unrealistic situations with reference to the intended objective of exploring a better pricing strategy. Typically, the situations foresee the cases where we have a dependent variable and a set of independent variables. If we assume a linear relationship, this results into a multiple linear regression (MLR) model given by :

$$y = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_n x_n + \varepsilon \qquad \text{..........(1)}$$

where, b_0 is the intercept, ε is the Gaussian error term with mean zero and variance (say) σ, y is dependent variable, and x_1, x_2,..., x_n are independent variables with associated regression coefficients b_1, b_2,...,b_n. Obviously, the MLR model aims to infer about unknown regression coefficients and the error variance so that using the given values of independent variables, one can easily predict the dependent variable y. We may, of course, test the regression coefficients against zero, and if the assertion is found true in the case of some of the regression coefficients, the resulting model will be easy to deal with (Montgomery et al., 2012).

To build an effective pricing strategy, it is essential to understand the degree of impact different features have on pricing calculations or profitability, which is explicitly available in the regression analysis. We shall, however, be confined to the MLR model in the present paper. Moreover, in a realistic business situation, we often come across missing values or missing information which can significantly affect the outcome of regression analysis. Such missing information cannot be efficiently handled in the regression analysis. We have, therefore, relied upon an alternative prediction technique known as decision tree, which is one of the powerful machine learning techniques and can address the above challenges of missing values in a routine manner. This technique is based on the probability and payoffs of decisions and is helpful to identify the important factors. The technique provides a user-friendly visual output of the model for a marketing team to build an effective pricing strategy. When both MLR and decision tree are combined for pricing decisions, the combination can help obtain a more comprehensive and accurate output for the marketing and pricing professionals.

Decision Tree Technique

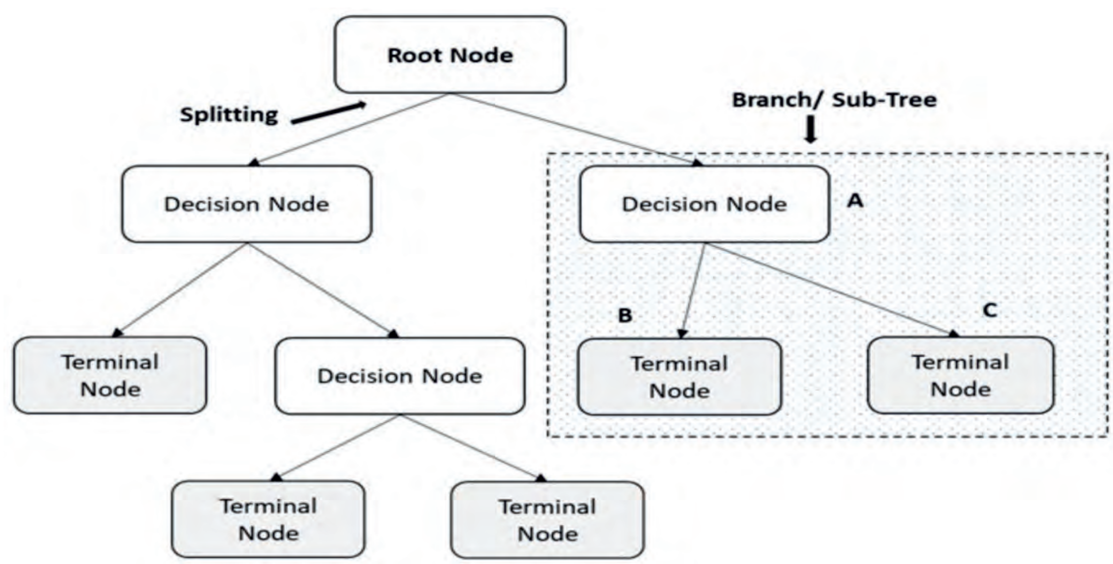
In current times, decision tree analysis has become very effective as a predictive analytics tool powered by data science techniques. This technique requires less data cleansing and is not influenced by outliers and missing values to a fair degree. It also helps capture the non-linearity in the data, which otherwise is not possible using linear regression. Decision trees often mimic human-level thinking ; it is simple to understand the data and make logical interpretations.

A decision tree is a tree-shaped graph that helps to determine a possible course of action. It is a type of supervised machine learning algorithm that has a pre-defined output variable. This technique divides the population or sample into two or more homogeneous groups based on the most significant differentiator in the input variables. The technique provides an effective structure in which alternative decisions and the implications of their choices can be assessed. As such, it helps the formation of fair, balanced risks and rewards arising from certain choices (Tofan, 2015).

There are two types of decision trees (depends on the type of output variable), namely the categorical variable decision tree (when the output variable is categorical) and the continuous variable decision tree (when the output variable is continuous).

In this study, we have considered a continuous output variable and, hence, a continuous decision tree approach is the ultimate choice. In the upcoming discussion, we shall provide some basic terminologies of decision trees, which are given for completeness. The complete framework of a simple decision tree is shown in Figure 1.

Figure 1. ***Decision Tree Framework***



Note. A is parent node of B and C.

Terminologies Related to Decision Trees

(i) Root Node : It represents the entire population or sample, and this further gets divided into two or more homogeneous sets.

(ii) Splitting : It is a process of dividing a node into two or more sub-nodes.

(iii) Decision Node : When a sub-node splits into further sub-nodes, it is called a decision node.

(iv) Leaf/Terminal Node : Nodes that do not split are called leaf or terminal node.

(v) Pruning : It reduces the size of decision trees by removing sections of the tree that provide little power to classify instances.

(vi) Branch/Sub-Tree : A subsection of the entire tree is called a branch or sub-tree.

(vii) Parent and Child Node : A node, which is divided into sub-nodes, is called the parent node of sub-nodes; whereas, sub-nodes are the children of the parent node.

Decision trees use multiple algorithms to decide on splitting a node into two or more sub-nodes. The algorithm selection is also based on the type of target variables. The four most commonly used algorithms in the decision tree are Gini index, chi-square, information gain, and reduction in variance. Amongst all, the Gini index is the most frequently used approach (Baykara, 2015 ; Kantardzic, 2011 ; Tangirala, 2020). It is a metric to measure how often a randomly chosen element would be incorrectly identified and is defined by :

$$Gini\ Index = 1 - \Sigma^{n}_{j=1}(p^{2}_{j}) \qquad \text{............ (2)}$$

where p_j is the probability of an object being classified to a particular class. Obviously, the Gini index is calculated by subtracting the sum of the squared probabilities of each class from unity. It favors large partitions. The variable that has a low Gini index may be used for splitting.

Data Source

To illustrate our methodology, we collected data from a leading manufacturing company's spare parts sales data for 2018 – 2019. Data were segregated as per two major technologies (Type1 and Type2). As 2020 was an exception year for business where the sales were down in most geographies due to the pandemic effect, we haven't

Table 1. ***Various Categories of Variables for the Classification of Data***

Variable (Variable Name in Code)	Description
Technology	Different types of technologies of Turbine products. e.g. Type1 (0), Type2 (1).
Contract Period *(Contract_period)*	Duration for which the contract is being signed, for example, 5 years, 10 years, etc.
Minimum Sale *(minimum_sale)*	Minimum business value ($MM) that the client is going to provide.
Prime Customer *(prime_customer)*	Refers to the type of customer. An exclusive customer is one who is going to deal only with the client, that is, the client will be the only vendor for any parts or services requirements (Yes = 0, No = 1).
Inflation *(Inflation)*	Annual percent increase in the price of parts and services.
Material Percent *(material_percent)*	The estimated percentage of parts that the client needs to provide.
Sales Volume *(sales_volume)*	Total amount of deal in million dollars.
Competitors *(Competitors)*	A dummy variable which holds value Yes (0) - if competitors are present at the time of bid negotiation and No (1) - if there is no competitor.
Loyalty with Customer *(Loyalty)*	0 - No penetration partner ; 1 - High penetration partner.
Customer History *(Customer_history)*	0 - Pays lower than the list price ; 1- Pays higher than the list price.
Margin (CM %) *(Margin)*	The historic contribution margin percent for the contract (1.0 = 100%).

used any 2020 data in our study. The data collected were arranged according to various categories of variables under consideration (Table 1). These categories are considered as the main drivers of the value for the customers and are used for the required analysis.

In Table 1, we have considered CM as the dependent variable and all other variables as independent variables. Also, contract period, minimum sale, material percent, sales volume, and inflation are continuous and quantitative variables, and the rest are qualitative. The qualitative variables were converted to quantitative equivalents for using them in further analysis. Characteristics like technology, prime customer, competitors, loyalty, and customer history were turned into quantitative data using binary classification, for example, 0 for No and 1 for Yes. This process can actually be considered as dummy coding (refer to Kuhn & Johnson, 2019).

Demographics

To continue with the analysis, we converted the real data into masked data to maintain the confidentiality of the business as per the data privacy policy. The subjects involved in the experiment were the CMs of different contracts for technology Type1 and Type2. The dataset consisted of a total of 65 contracts, with 55% belonging to Type1 technology and 45% to Type2 technology. In Table 2, we have given a snapshot of only 10 data points used for the numerical illustration. The complete data set is shared on https://github.com/bhas-1974/Bhaskar-Roy/tree/main as filename : *Data Analysis for Pricing*.

Table 2. ***Snapshot of the Data Used for Numerical Illustration***

Technology	Contract Period (in years)	Minimum Sale	Prime Customer	Inflation	Material Percent	Sales Volume ($MM)	Competitors	Loyalty	Customer History
Type1	14	0	No	3%	74%	39	No	High penetration partner	Pays lower than list price
Type2	14	0	No	2%	25%	36	No	High penetration partner	Pays higher than list price
Type1	5	0	No	2%	67%	1802	No	High penetration partner	Pays lower than list price
Type2	12	15	Yes	2%	43%	24	No	No penetration partner	Pays lower than list price
Type2	10	0	No	3%	39%	11	No	High penetration partner	Pays higher than list price
Type1	12	0	No	2%	46%	18	No	High penetration partner	Pays higher than list price
Type1	11	15	Yes	2%	62%	19	No	High penetration partner	Pays higher than list price
Type2	11	25	Yes	3%	79%	40	Yes	High penetration partner	Pays higher than list price
Type1	9	4	Yes	0%	82%	8	No	High penetration partner	Pays lower than list price
Type2	12	0	No	2%	20%	14	Yes	No penetration partner	Pays lower than list price

Empirical Analysis and Results

The sales data for 2018 – 2019 were analyzed using statistical methods and Python software. The analysis was initiated first with correlation analysis to measure the association between different variables as we have a combination of both continuous and dichotomous variables. Then we did multiple linear regression analysis followed by decision tree analysis. We have conducted a similar analysis for both technologies (Type1 and Type2) separately and analyzed the results separately.

Correlation Matrix and Multicollinearity

A point bi-serial correlation analysis has been conducted to measure the association between different variables as we have a combination of both continuous and dichotomous variables. The point bi-serial correlation coefficient is a special case of Pearson's correlation coefficient, and this has been applied for the scenario where we have one variable as continuous and the other as dichotomous. However, in the case of two continuous variables, we have used Pearson's correlation coefficient. The results are given in Table 3. It can be seen that out of all 10 independent variables, technology and material percentage have the strongest correlation with CM as compared to other pairs (Table 3). Correlation between different features and dependent variables ranges between –0.36 to 0.36. Weak correlation (within +/–50%) indicates that there appears no indication of multicollinearity between different independent variables. We have also calculated variance inflation factor (VIF) for individual variables, which are

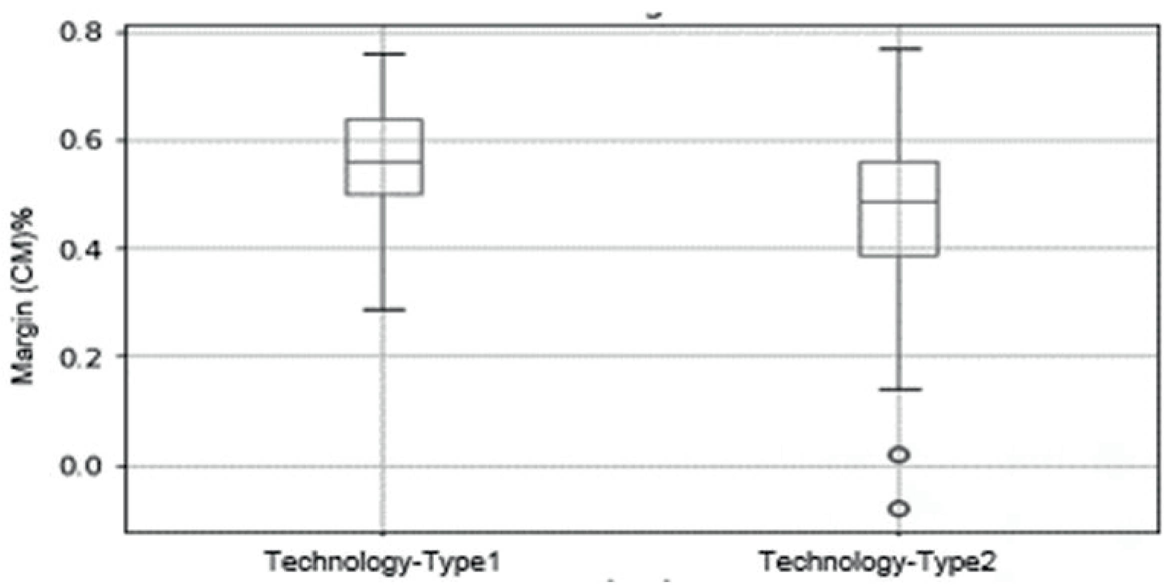
Table 3. ***Correlation Matrix Between Different Pairs of Variables***

	Technology*	Contract Period (in years)	Minimum Sale	Prime Customer*	Inflation	Material Percent	Sales Volume	Competitors*	Loyalty*	Customer History*	Margin**
Technology*	1.00										
Contract Period (in years)	0.13	1.00									
Minimum Sale	–0.14	–0.11	1.00								
Prime Customer*	0.19	–0.09	–0.03	1.00							
Inflation	0.11	–0.04	–0.13	0.12	1.00						
Material Percent	–0.07	–0.19	0.17	0.08	0.04	1.00					
Sales Volume	–0.19	–0.10	0.36	0.13	–0.10	0.15	1.00				
Competitors*	–0.18	0.06	0.00	0.13	–0.10	–0.01	0.09	1.00			
Loyalty*	0.14	–0.11	0.01	0.25	0.04	0.09	0.06	0.36	1.00		
Customer History*	0.06	–0.01	0.02	0.10	0.05	–0.10	–0.16	0.12	0.31	1.00	
Margin**	–0.36	–0.19	–0.08	–0.22	–0.11	0.30	0.02	0.01	–0.01	0.27	1.00

Note. * Dichotomous independent variables and the corresponding point bi-serial correlations.

** Continuous dependent variables and corresponding Pearson correlations.

Figure 2. *Box Plots of Two Different Technologies with Respect to CM*



1.11, 1.22, 1.06, 1.10, and 1.30, respectively for the contract period, minimum sale, inflation, material percent, and sales volume. Obviously, the reported values of VIF indicate an absence of multicollinearity (Senaviratna & Cooray, 2019 ; Tay, 2017).

While building the data science models, a question arises if we should build one combined technology, that is, the MLR model for both technologies together or create separate models for two different technologies. To get the answer, we use independent sample *t*-test along with the box plot features. Based on independent sample *t*-test, it is found that the two technologies have different behavior with respect to CM (*t*-statistics = 2.85 and the corresponding *p*-value = 0.0067). Also, it is quite evident from the box plots (Figure 2) that the median value of CM for Type1 technology is higher than the corresponding value for Type2 technology. Since the box plots reflect almost symmetric densities for the two technologies, a similar comment can be given for mean values as well. Obviously, based on the two conclusions, it is desired to build separate MLR models for separate technologies to avoid misleading results in a true sense.

Results Based on Regression Analysis

MLR analysis is performed on the dataset to yield a quantifiable value for each independent variable on the dependent variable (CM). The analysis is performed on the data set in three different scenarios. In the first case, we combined the two technologies and used the entire dataset, treating technology as one of the independent variables. We considered the Type1 and Type2 technologies separately in the other two cases and analyzed the corresponding datasets. Based on the results of MLR analysis, it is observed that a number of associated coefficients in all the three cases are insignificant, and thus, the corresponding independent variables could be considered to have insignificant effects on the dependent variable. The remaining regression coefficients where the associated independent variables have significant effects on the dependent variable (CM) are reported in Table 4. The dashed cells in Table 4 correspond to those variables which have no significant effect on the dependent variable and are dropped from the model.

Table 4. ***Results of Regression Analysis for the Three Considered Cases***

Variables	Coefficients – Considering Both Technologies Together	Coefficients Technology-Type1	Coefficients Technology-Type12
Constant	0.3578	0.5007	0.4081
Technology	–0.1117	-	-
Contract Period	-	–0.0136	-
Minimum Sale	–0.0004	-	-
Prime Customer	–0.0748	-	–0.1611
Inflation	-	-	–3.8084
Material Percent	0.3125	0.1963	0.1613
Sales Volume	-	-	-
Competitors	-	-	-
Loyalty	-	-	-
Pricing History with Customer	0.1231	0.0737	0.1991

Table 4 clearly establishes that material percentage is a significant variable in both Type1 and Type2 technologies. Apart from this, contract period and pricing history are also significant for Type1 technology, which means less contract duration and deeper customer history help drive better margins in the case of Type1 technology. Similarly, inflation, prime customer, and pricing history also play a role for Type2 technology. Also, one can easily predict the corresponding CM% with the help of Table 4. All these results profoundly impact business decisions, and hence, the MLR analysis is hugely popular for pricing professionals across industries.

Results Based on Decision Tree Analysis

In a typical business problem, the linear relationship between CM and other features is a major assumption to carry the regression analysis, and the same has been done in the previous sub-section using the MLR model. It is also observed that all independent variables are not significantly affecting the dependent variable. Since some variables are continuous and some are dichotomous, it is generally difficult to understand their impact through a regression model. Also, business strategy requires some best-case scenarios to understand the intersection and relationship between CM and other independent variables to identify the pricing strategy better.

We have used feature importance scores to calculate the importance matrix. The entire computation was done by using a self-developed Python code. It may be noted that the feature importance scores play an important role in a predictive modeling project. The scores can be used to get insights into the data and the model as well as act as the basis for dimensionality reduction and feature selection that can finally improve the efficiency and effectiveness of a predictive model on the problem. This importance is calculated explicitly for each attribute in

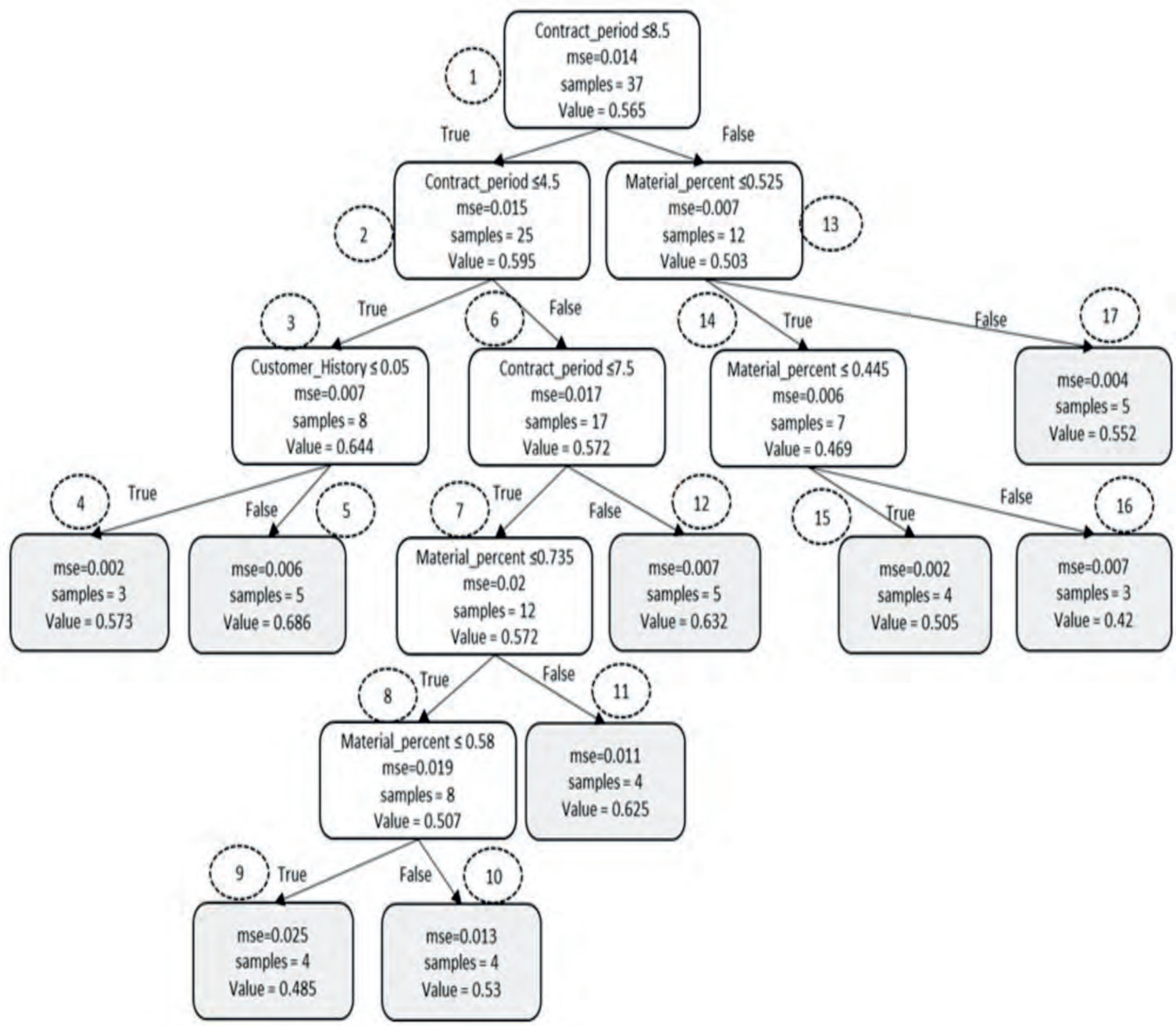
Table 5. ***Feature Importance Matrix for Type1 Technology***

Variable	Importance
Contract Period	56%
Material Percent	34%
Customer History	11%

Table 6. *Feature Importance Matrix for Type2 Technology*

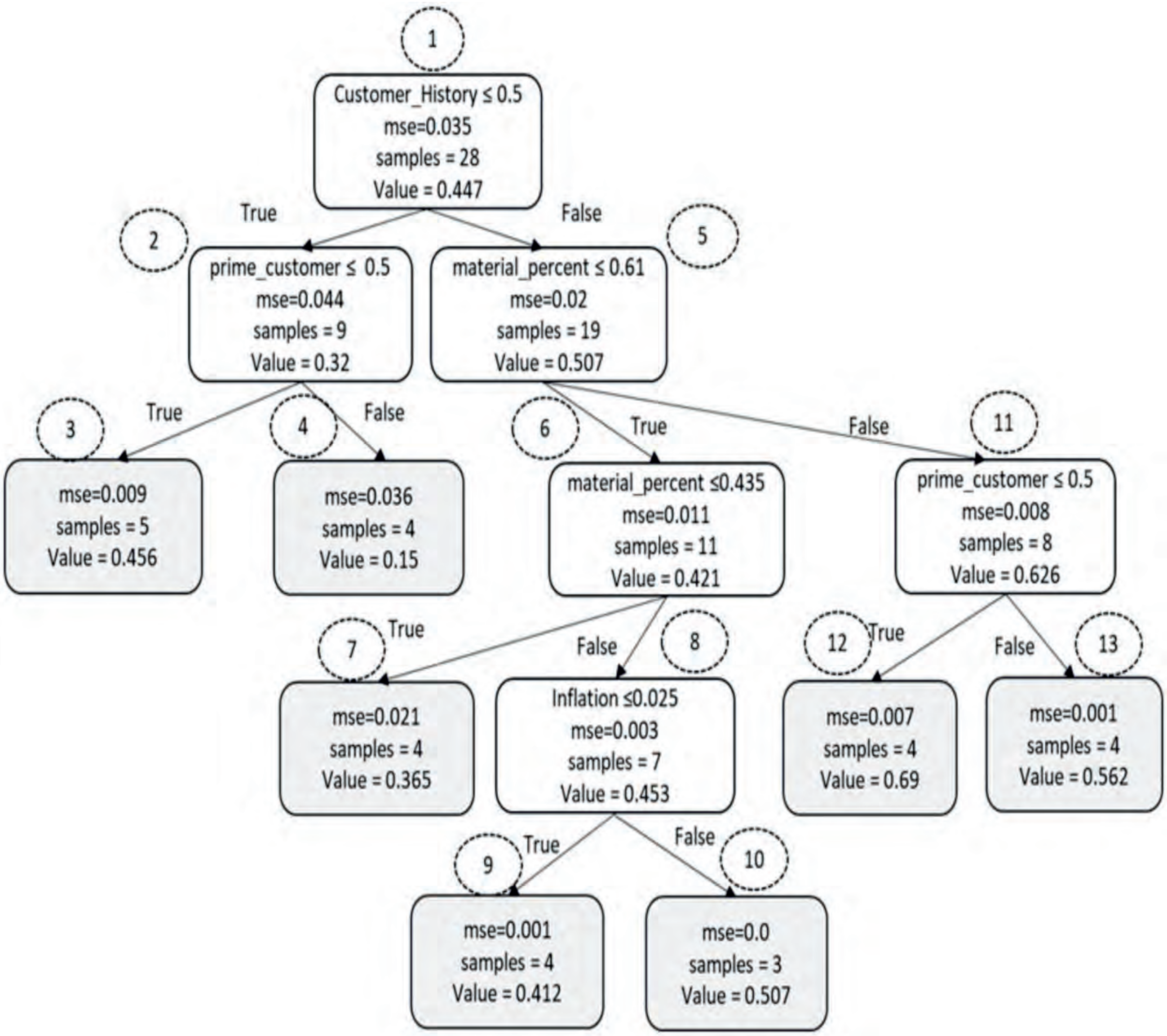
Variable	Importance
Prime Customer	35%
Material Percent	31%
Customer History	31%
Inflation	2%

Figure 3. ***Decision Tree Model 1***



the dataset, allowing attributes to be ranked and compared to each other (Brownlee, 2020). The feature importance scores (Table 5 and Table 6) are discussed to understand the most influential independent variables for our dependent variable (CM).

Figure 4. ***Decision Tree Model 2***



In the case of Type1 technology, it is clear that contract period, material percent, and customer history are the top three statistically significant features derived from feature importance scores. In the case of Type2 technology, however, the statistically significant three important features are prime customer, material percent, and customer history. It can be seen that material percent and customer history are present in both the technologies and hence, can be considered as significant features for price setting. The business analysts should focus on these two product-specific and customer-specific features to identify the right strategy to derive the maximum margin. Moreover, if we combine the results obtained using MLR and decision tree, we see almost similar behavior, and hence, it is vital for the businesses to have a combination of the two methods to derive the exact pricing strategies.

After identifying the important features, we proceed to build a decision tree using Python software. A decision tree's nodes help us to identify the splitting points towards decision making. Based on optimum (minimum) mean squared error (MSE), the first split took place for contract period followed by material percentage for Type1 technology and customer history followed by prime customer and material percent for Type2 technology. Our decision tree is built using the feature/threshold combinations where MSE is minimized.

In Type1 technology, the root node uses the variable contract period with a threshold of 8.5 to initially divide the samples. We can see that at the root node, we have 37 samples, which we divided into 25 (left child node) and 12 (right child node). The right child node uses the variable material percent with a threshold of 0.5, and the left child node uses the same contract period with a threshold of 4.5, resulting in some additional child nodes that finally terminated at leaf nodes. This is how the decision tree is built and used for interpreting different scenarios.

In the case of Type1 technology (Figure 3), node 5 clearly indicates that whenever the contract period is less than 4.5 years and the customer has a history of paying higher than the list price, it becomes clear-cut indicative of a profitable case for the business (CM% = 69%). On the other hand, node 16 indicates that for a long duration contract (>8.5 years) and moderately large material percent (>45%), the business is most likely to get a very low margin (CM% = 42%). The former situation can be considered as the best-case scenario ; whereas, the latter case may be treated as the worst-case scenario.

In the case of Type2 technology (Figure 4), nodes 4 and 12 offer minimum and maximum margin scenarios. It can be seen that the CM% is very small for the 4th node and the maximum (69%) for the 12th node. Obviously, the former results into the worst-case scenario; whereas, the latter can be considered to provide the best-case scenario. Moreover, different nodes provide appropriate wining CM% for different scenarios to commercial teams for their price estimation. The business leaders and analysts must understand the contributing factors and adjust the prices accordingly to enjoy a high CM%.

Conclusion

Based on the available literature, it is pretty evident that the modern pricing strategies evolved from traditional methods to value-based methods over a period of time. This paper has focused on how pricing analytics can be transformed with the help of predictive analytics (combining both multiple linear regression and decision tree analyses) to build a unique pricing strategy to optimize the contribution margin (CM).

This empirical case study helps us explore how decision tree and regression analysis can assist commercial teams in building better pricing strategies by predicting CM to drive wins at higher prices in the future using historical information. This research suggests building a separate model for each technology since pricing drivers are not equal for all the technologies or products. Some technologies are generic, and some are niche. Our study, however, concludes that for a technological company, material percent, customer history, and contract period are the key deciding factors in arriving at winning price points to win more and more deals in this competitive industry. For a better pricing strategy considering compound scenarios, this research strongly recommends using decision tree analysis along with regression instead of a standalone regression model or standalone decision tree technique. Truly speaking, decision tree helps us identify the right avenues that will drive better margins that require special focus. Moreover, the extensive use of Python software for the considered predictive analysis provides an easy and user-friendly interface.

Implications

Theoretical Implications

The study has focused on the role of analytics in 'pricing methodology,' which is a critical element of marketing strategy. We focused on three key contributing factors in pricing strategy. Firstly, we have provided a holistic understanding of different pricing approaches for the B2B industry. Secondly, we have shown how to use predictive analytics to estimate CM% to derive an optimum pricing strategy. And lastly, we have demonstrated how the data science approach can help organizations get actionable insights and identify key features that are the driving forces of CM. Hence, this study is quite relevant for the pricing research field from both theoretical and practical aspects.

Managerial Implications

Predictive analytics for pricing strategy is one of the most critical functions for marketing and commercial decision-making processes in the current competitive market. To build an effective pricing strategy, predictive analytics is becoming very crucial. This study highlights how businesses can capture the right features and incorporate those, as per their magnitude, into business strategy. Applying the predictive analytics approach helps us estimate the CM% and arrive at a probable winning price during bid negotiations.

This study should help businesses move from traditional pricing strategies to more data-driven scientific approaches, thus creating a win-win situation for both the organization and the clients. The pricing leaders and commercial managers would be able to manage and control the pricing levers more effectively to improve their profits and margins for the business.

Limitations of the Study and Scope for Future Research

In this study, we have used limited data points; large data points definitely help to generalize the solutions to a larger extent with greater confidence. Based on limited data points, we have established the approach to provide guidance to businesses for a better and effective pricing strategy. There are some other areas of research, particularly on bringing external features like competitor details, macro-economic factors, geographical distribution, and some additional key indices like leading and lagging indicators, etc. We have used multiple linear regression and decision tree to estimate CM% to identify important features that could potentially drive CM. This study has not focused much on the accuracy of CM% estimation. Accuracy can be optimized using large data sets, additional features, and exploring multiple techniques. Also, the research can be extended to further micro-level to build models for various geographies and product specifications which is purely dependent on data variety and availability.

Authors' Contribution

Bhaskar Roy conceived the idea and developed the quantitative design to undertake the empirical study. Debabrata Bera generated concepts relevant to the research design and extracted relevant research papers from online databases. Dr. Praveen Kumar Tripathi conducted proofreading of the paper and further helped to modify the sections. Prof. S. K. Upadhyay verified the analytical methods and supervised the study. The quantitative data collection and numerical computations were done by Bhaskar Roy and Debabrata Bera jointly using Python software. Finally, Bhaskar Roy and Debabrata Bera wrote the manuscript in consultation with the co-authors.

Conflict of Interest

The authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

Funding Acknowledgment

The authors received no financial support for the research, authorship, and/or for the publication of this manuscript.

References

Armstrong, G., Adam, S., Denize, S., & Kotler, P. (2014). *Principles of marketing.* Pearson Australia.

Baykara, B. (2015). *Impact of evaluation methods on decision tree accuracy* (Msc Thesis). School of Information Sciences, University of Tampere, Finland. https://core.ac.uk/download/pdf/250137461.pdf

Björk, H., & Forsberg, L. (2015). *How to succeed with value-based pricing.* KTH Industrial Engineering and Management Industrial Management, Stockholm, Sweden.

Bodea, T., & Ferguson, M. (2012). *Pricing segmentation and analytics* (1st edition). Business Expert Press Marketing Strategy Collection.

Brownlee, J. (2020). *How to calculate feature importance with python.* Machine Learning Mastery. https://machinelearningmastery.com/calculate-feature-importance-with-python/

Chan, K., Jubas, J., Kordes, B., & Sueling, M. (2014). *Understanding your options : An overview of proven pricing strategies and how they work.* McKinsey & Company.

Chan, M. (2017, December 5). *Discount pricing strategy explained.* https://www.unleashedsoftware.com/blog/discount-pricing-strategy-explained

Claessens, M. (2018, June 27). Economic value estimation - How to estimate economic value ? *Marketing Insider.* https://marketing-insider.eu/economic-value-estimation/

Daly, J. L. (2002). *Pricing for profitability.* Wiley.

Jain, S., & Kamble, S. C. (2021). An analytical study of buyers' behaviour towards automatic fuel pricing and price smoothing in India. *Prabandhan : Indian Journal of Management, 14*(2), 40–59. https://doi.org/10.17010/pijom/2021/v14i2/157692

Kantardzic, M. (2011). *Data mining : Concepts, models, methods, and algorithms.* John Wiley & Sons.

Kotler, P., & Keller, K. (2014). *Marketing management* (15th edition). Prentice Hall.

Kuhn, M. & Johnson, K. (2019). *Feature engineering and selection : A practical approach for predictive models.* Taylor & Francis Group. https://bookdown.org/max/FES/creating-dummy-variables-for-unordered-categories.html

Liberto, D. (2021). *Bundling.* https://www.investopedia.com/terms/b/bundling.asp

Monson, M. (2009). Valuation using hedonic pricing models. *Cornell Real Estate Review, 7,* 62 – 73.

Montgomery, D.C., Peck, E.A., & Vining, G.G. (2012). *Introduction to linear regression analysis* (5th edition). John Wiley & Sons.

Nagle, T., Hogan, J., & Zale, J. (2011). *The strategy and tactics of pricing : A guide to growing more profitably* (5th edition). Pearson.

Panigrahi S. K., Azizan, N. A., & Al - Shamsi, I. R. (2021). Product innovation, customer satisfaction, and brand loyalty of using smartphones among university students : PLS – SEM approach. *Indian Journal of Marketing, 51*(1), 8 – 25. https://doi.org/10.17010/ijom/2021/v51/i1/156931

Panigrahi, S. K., Azizan, N. A., & Khan, M. W. (2018). Investigating the empirical relationship between service quality, trust, satisfaction, and intention of customers purchasing life insurance products. *Indian Journal of Marketing, 48*(1), 28 – 46. https://doi.org/10.17010/ijom/2018/v48/i1/120734

PwC. (2016). *Driving retail growth by leveraging analytics.* https://www.pwc.in/assets/pdfs/publications/2016/profitable-growth-for-retail-businesses-online.pdf

Senaviratna, N. A., & Cooray, T. M. (2019). Diagnosing multicollinearity of logistic regression model. *Asian Journal of Probability and Statistics, 5*(2), 1 – 9. https://doi.org/10.9734/ajpas/2019/v5i230132

Simon, H., & Fassnacht, M. (2018). *Price management.* Springer International Publishing.

Spann, M., Fischer, M., & Tellis, G. (2015). Skimming or penetration ? Strategic dynamic pricing for new products. *Marketing Science, 34*(2), 235 – 249. https://doi.org/10.1287/mksc.2014.0891

SurveyKing. (2021a.). *A guide to the Van Westendorp price sensitivity meter.* https://www.surveyking.com/help/van-westendorp-analysis

SurveyKing. (2021b.). *An interactive conjoint survey tutorial with sample data.* https://www.surveyking.com/help/conjoint-analysis-explained

Tangirala, S. (2020). Evaluating the Impact of GINI Index and information gain on classification using decision tree classifier algorithm. *International Journal of Advanced Computer Science and Applications, 11*(2), 612–619. https://doi.org/10.14569/IJACSA.2020.0110277

Tatyana, N. - P. (2011). Value-based pricing : A success factor in the competitive struggle. *Marketing, 42*(4), 227 – 236. https://doi.org/10.5937/markt1104227N

Tay, R. (2017). Correlation, variance inflation and multicollinearity in regression model. *Journal of Eastern Asia Transportation Society, 12,* 2006 – 2015. https://doi.org/10.11175/easts.12.2006

Tofan, C.A. (2015). Decision tree method applied in cost-based decisions in an enterprise. *Procedia Economics and Finance, 32,* 1088 – 1092.

Wroblewski, J., & Loss, S. E. (2018). *Understanding value-based pricing : Conceptualising customer value in a B2B industrial context.* School of Economics and Management, Lund University. http://lup.lub.lu.se/luur/download?func=downloadFile&recordOId=8949083&fileOId=8949087

About the Authors

Bhaskar Roy is MSc (Statistics) and MBA (Marketing). He is currently working as Vice President and Global Operations Leader in Genpact. He has more than 22 years of diverse experience in analytics & data science, consulting, and intelligent operations. He is Academic Advisor for several B-schools as well as an Industry Speaker. He is also a Research Scholar and currently working towards completing his PhD thesis in Predictive Analytics Applications from Banaras Hindu University, Varanasi.

Debabrata Bera has completed MSc in Statistics and MPhil in Population Sciences from International Institute for Population Science, Mumbai. He has worked as a Statistician on multiple social research projects and has two publications to his credit. He is currently working as Sr. Manager in Genpact and driving analytics with data science, predictive analytics, and digital technologies.

Dr. Praveen Kumar Tripathi is an Assistant Professor at the Department of Mathematics & Statistics, Banasthali Vidyapith, Rajasthan. He has been awarded his PhD degree in Statistics from Banaras Hindu University, Varanasi. His areas of research include statistical data analysis, Bayesian inference, time series analysis, and econometrics.

Prof. S. K. Upadhyay is a Professor at the Department of Statistics, Banaras Hindu University. Prof. Upadhyay is one of the world-renowned researchers in Bayesian statistics and computation, statistical reliability, biostatistics, etc. Prof Upadhyay has over 38 years of teaching and research experience. He has been honored by Imperial College, London, and has been elected fellow of the Royal Statistical Society (FSS), London. He has authored and published more than 150 papers and written several books. Prof. Upadhyay obtained MSc and PhD degrees from Banaras Hindu University, Varanasi.