Impact of Analytics on Pricing Decisions

Ghute Priyanka Rajendra[1], Dr. Manish Sinha[2]

[1,2]Symbiosis Centre for Management and Human Resource Development, SCMHRD, SIU, Symbiosis International (Deemed University), SIU, Hinjewadi, Pune, Maharashtra, India

[1]priyanka_ghute@scmhrd.edu, [2]manish_sinha@scmhrd.edu

ABSTRACT

Pricing strategy is a fundamental component of the marketing process. Pricing is a process where businesses decide the price at which the products or services should be sold. Some of the objectives of pricing are profit maximization, price stability and to prevent the competition. With the use of available data such as transactional data, sales data, these strategies can be implemented effectively. Pricing with analytics helps to leverage the data to increase the profit.

This paper explores approaches for price bundling and price forecasting strategies using machine learning algorithms. Bundling is about grouping products or services with the aim of offering a discount. The study uses the K-means clustering algorithm to determine the price of each bundle. Then the results are compared with the traditional approach of bundling where MS Excel Solver is used. Revenue is increased significantly using the new approach.

Further, it discusses an approach to forecast the prices based on qualitative data such as customer reviews, brand name, item condition. Natural Language Processing (NLP) methods are used to read and manipulate the text data and then prices are predicted using machine learning algorithms such linear regression, K-Means clustering. This method shows how textual data can be leveraged to make pricing decisions more efficient.

Keywords

Pricing strategies, machine learning, K-means clustering, NLP

Article Received: 10 August 2020, Revised: 25 October 2020, Accepted: 18 November 2020

Introduction

Typically, price acts as the main determinant of customer choice. Recently, there is a change in consumer behaviour with non-price factors also play a role in buying process. Still pricing remains the most important factor impacting buyers' decision. Out of all the elements of marketing mix, pricing generates revenue while other factors increase the costs. Also, price is the most adjustable component in marketing mix as it can be altered quickly unlike other factors such as new campaigns, changing product features. Pricing is used to target segment of customers. Considering the importance of pricing in a business, we focus on how pricing decisions can be made more efficient using analytics.

In many businesses, market research is not done from the start of the product design process. Hence there is a lack of understanding about how customers value feature and will actually pay for. Then common research methodologies such as customer surveys, competitor benchmarking and internal brainstorming are done to set the prices. This results in poor pricing strategy.

There are different methods used for pricing the products in market. Demand pricing refers to consumer demand of product or service as the main element for deciding the price. Competitive pricing is a method where prices set by competitors are referred. Cost-plus pricing considers all charges indulged in making the product like cost of raw materials and production. Penetration pricing method uses ow price to enter the market or while launching the product and then increases it gradually. Price skimming is a method where businesses enter market with high priced products to gain the most revenue. Economy pricing is a basic low- cost method where prices are kept low targeting price sensitive customer segment. In psychological pricing a minor different in a price impacts user buying decision. For example, a product priced at RS. 99 may be seen as much cheaper than the one priced at Rs. 100. Discount pricing method offers products and services at less price in seasonal sales or loyalty cards. In Geographical pricing, products are sold at different prices in different regions based on economy. In price bundling strategy product or services are sold as package. In this paper we have concentrated on price bundling and price forecasting.

With evolution in data analytics, we can leverage this technology to make better decisions. We propose two approaches here to improve pricing strategies. In first approach we focus on price forecasting. In second approach we discuss price bundling.

In first approach, we discuss how comments given by customers can be used while making the pricing decisions. We have used Natural Language Processing techniques to draw meaningful insights from the comments. Then we cluster the similar products together and decide prices as per the cluster. Then we have also shown how regression can be used on text data to predict the prices. Considering the customer's feedback to predict price is unique. The study forms a foundation for additional search.

In second approach we discuss how to improve price bundling strategy. We have referred to electronic services data. We propose unsupervised learning technique called clustering for bundling the products. The combination of products and various options for their prices are obtained. Then based on the revenue we decide the final prices for all the combinations. The results show that clustering method increases the revenue.

This paper is organized as follow: In the Literature Review subsection, previous studies on pricing methods are

outlined. In the Research Methods section, the data, the proposed methods and the evaluation of methods used in this study are described. In Results and analysis section, results are analysed. The conclusion and recommendations section present the conclusion, where the insights obtained, and future issues are summarized. Limitations section describes the limits of the research.

Literature Review

Price is the most important parameter that affects revenue significantly. Making changes in the prices according to market is called pricing strategy. Different strategies such as discount- strategy, market-segmentation or price-discrimination strategy, price-skimming, price- skimming, penetration pricing, high and low-price strategies, yield management are in practice. Forecasting competitor's reactions and price wars are main challenges in deciding pricing strategies. [1]

Supervised machine-learning techniques can be leveraged to forecast the prices of second-hand automobiles. The required data is taken from everyday newsprints. Various methods such as multiple-linear regression analysis, k-nearest neighbours, naïve-Bayes and decision trees are considered for the predictions and these models are compared based on accuracy to find the most accurate model. The decision trees and naïve Bayes cannot process output classes with numeric datatypes. The important limitation of the research is the less amount of data considered for modelling and should be tested for more data. [2]

Pricing and predicting the demand for product that is new in the market is challenging part for the retailer as it would directly affect the revenue. In this research, machine learning techniques are used to calculate loss in past years and calculate demand for the new products to be launched. A pricing decision tool developed using nonparametric regression which maps forecasts of demand into a pricing decision. Combining Regression-trees with bagging technique showed better results than other regression models compared based on various metrics. [3]

With evolvement in digital economy, dynamic pricing can be implemented using mathematical models. Various models such inventory-based models, data-driven models, auction-based models and machine-learning based models can be leveraged to formulate pricing strategy. In e-commerce markets dynamic-pricing strategies perform better than fixed-pricing strategies. How dynamic pricing can be implemented using reinforcement learning is illustrated using an example of single seller with non-linear pricing used for various quantities. [4]

Using statistical ad machine learning models, the purchase decisions that are made considering adaptive-pricing can be predicted. Data for modelling can be collected from purchase history, web data, visitor attributes and context-understanding using techniques such as data mining and big data. This paper focuses on purchasing behaviour of clusters of customers rather than individual customers. Here, logistic regression is used to predict if a costumer would buy a product or not given the cluster to which customer belongs and price range of the product. Output is obtained in binary form. The approach has been tested for large dataset of ecommerce firm and much better price ranges are obtained. The same framework ca be applied I various industries working online. [5]

When a seller has numerous items, manual pricing becomes tedious and costly. This is the case in most of the ecommerce businesses. Algorithms using Artificial-intelligence concepts are being leveraged to automate the saes process reducing labour cost and other operational charges. This paper explains an algorithm for determining the sale prices for online businesses which tries to maximize profit reducing sales time. The approach is tested for DVD movie sales in online market and has produced expected results. [6]

Retail industry applies forecasting in planning and managing to get effective outputs tin the business. Various forecasting methods has been used to predict the demand in fashion industry, to calculate the number of visitors, to project the advance in on-line retailing, to locate an ideal location for retail store, to forecast when to refill in the discount retail chain. This paper explains a two-fold approach to decide pricing strategy for semi-luxury supermarket commodities based on tree-based prediction model. It explains price demand inter-dependency while solving optimization problem. Bound based heuristics and recursive branch are used to solve the formulation. [7]

Digitization has created new opportunities in media industry such as changes to media products, new interactions across products and channels, availability of data to measure consumer behaviour and to make marketing strategies based on data and business intelligence. This paper talks about how business analytics can be used in media industry in the context of pricing and mixed bunding. The model is based on tiered pricing concept. [8]

This paper describes finest bundling approaches for a multi-product monopolist. It shows that bundling many un-associated information goods can be gainful. As bundling leads to more sales at optimal prices, this helps to get more sales, more economic productivity, and more revenue per item from a bundle of information goods than can be obtained when the same goods are traded individually. The result confirms that the tactic of vending a bundle of items for a single amount produces more revenue than selling them separately. [9]

This paper compiles and reviews some of the popular pricing strategies adopted by the modern businesses. It states pros and cons of each of the pricing strategies. It relates marketing management with managerial economics. It discusses in detail the importance of the pricing strategies, main objectives behind pricing strategies such as revenue maximization, various pricing strategies and their advantages and disadvantages and how the businesses should select the most suitable pricing strategy for themselves. [10]

In the Era of big data, prices are dynamic, and consumers can compare the prices using various applications before making buying decisions. This paper explains modified K-means clustering method which divides customers into clusters and prices are decided for each cluster. [11]

This paper describes insurance pricing and data science. It talks about general linear models, generalized additive models and credibility theory and then more advanced machine-learning techniques such as regression trees,

bagging and boosting techniques, random forest and neural networks. The research also provides methods for analysing telematics carriage driving data using unsupervised-learning. [12]

In ecommerce space, price of any product is extremely translucent. It is referred to increase margin and sales by several e-businesses as the price of competitors can be see easily. This paper discusses how the customers are benefitted by study of dynamic prices of the item. Many variables for instance attributes, buyer visit data and buying history of a buyer are considered for the research. [13]

Research Methods

We propose the methods using data analytics approaches for price suggestions and price bundling strategies as below:

I Price Suggestions

Prices are predicted for e-commerce website which sells a variety of products. The electronics industry has changing prices based on product-specifications and fashion industry shows cyclic pricing tendencies and is mainly controlled by brand tags.

Generally, prices are decided based on their specifications. Here we propose a method where customer's perception about the item is considered while pricing it.

The data is taken from Mercari, the shopping app [14]. It contains various details about the product such as brand name, category name, price, whether shipping is free or not, item description i.e., comments by customers. We have data of Mercari shopping app in Japan. The dataset has following fields: user-entered text descriptions of the products, product category name, brand name and item condition. A big challenge here is to operate on the text descriptions. It contains customers' comments such as their reviews about the product. These comments can be used while deciding the price. Here, we process the comments data using Natural Language Processing (NLP) techniques and then the comments are clustered to decide their prices.

Natural Language Processing is automatic manipulation of a natural language like text by software. A few examples of NLP that we use every day are Spell check, Autocomplete, voice to text messages, spam filtering, Siri, Alexa, Google Assistant. Natural language means the way, we humans, communicate with each other. We see a huge amount of text such as signs, menus, Email, Messages, SMS, web pages, newspaper and the list is endless. Considering the importance of this type of textual data, we must have methods to understand this language as we do for other types of data.

The recent developments in Artificial intelligence enable us to identify the words, phrases or response by context as any human would do. NLP performs better than humans when it comes to the amount of data and processing power. Consider an example of a brick-and-mortar store where employees and customers communicate before, during and after-sales. There are countless opportunities for NLP to improve how the store operates. The product or service can be improved by analysing thousands of customer interactions. It's impossible for humans to log and interpret all this data on their own but technology has made it possible.

Following steps are followed to implement the proposed approach:

Step 1: Processing of Text Data

The processing of comments of the customers regarding the product is done in the following steps:

1. Tokenization
2. TF-IDF Score
3. t-SNE
4. Clustering

Tokenization:

Tokenization is the task of dividing the document or sentence into tokens and at the same time removing unwanted characters such as punctuation. The first step is to tokenize the textual data. The main purpose of tokenization is to normalize the text. Here, we first divide the comments into sentences and the sentences into tokens. Then we delete punctuations and stop-words and convert the tokens to lowercase. The words having a length three or more are considered, in order to have a confined set of meaningful words to analyse. These tokens are passed to the next step. For example, "Adorable top with a hint of lace" would become ['Adorable', 'top', 'hint', 'lace'] after tokenizing.

The Category data is also tokenized to separate out the Main Category and the various Subcategories. The data in the Category column is split on the '/' character in order to get the individual categories. These categories are then put into different columns viz – Main_Category, Sub_Cat_1, Sub_Cat_2. Going forward, these categories will play a fundamental role in deriving the price of the product.

It is evident that if we consider the most common words by category, we can observe here that size, free and shipping words used frequently by the sellers, probably to attract customers.

We use the Word Cloud tool to visualize the frequency of the tokenized words across each category

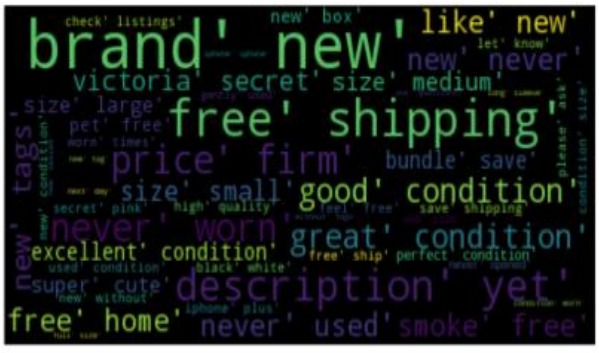

FIG. 1. WORD CLOUD FOR ITEM DESCRIPTION

Calculate TF-IDF score:

TF-IDF is abbreviation for Term Frequency- inverse document frequency. It quantifies the importance of a word based on term frequency i.e., occurrences of a term in a given text and inverse document frequency i.e., the

reciprocal number of times a token appears in a corpus of documents. TF-IDF score is calculated for all the words. If the word is mentioned frequently in all comments, its existence within a specific comment won't give us much specific information about the comment itself. Here, the second term is considered as a penalty term that penalizes common words such as "a", "the", etc. TF-IDF score indicates relevancy of a word in a specific text. TfIdfVectorizer is used to calculate the score. This outputs a tfidf matrix having the row count same as the total number of descriptions and the column count same as unique tokens across descriptions. The words such as "new", "brand", "size" got the fewer scores while the words such as "postnatal", "colour brow" got the higher scores. High scores indicate that the words are very specific and that by looking at them we can determine the category of the product. Following are the words with the highest and lowest TF-IDF scores.

TABLE 1. TFIDF TOP

Token	TF-IDF Score
Postnatal	13.195054
subdrip rda	13.195054
Lmt	13.195054
lbs length	13.195054
place step	13.195054
light volts	13.195054
thumb point	13.195054
Wedgwood	13.195054
novelty bill	13.195054
colour brow	13.195054

TABLE 2. BOTTOM SCORES SCORES

Token	TF-IDF Score
new	2.175653
size	2.330674
brand	2.75566
condition	2.799306
brand new	2.874418
free	2.903426
shipping	3.070592
worn	3.107882
used	3.16531
never	3.276901

Calculate t-SNE:

Given the high dimension of our tfidf matrix, we need to reduce their dimension using the Singular Value Decomposition (SVD) technique. And to visualize our vocabulary, we could next use t-SNE to minimize the dimension from 50 to 2. t-SNE is more suitable when we want dimensions of the dataset to be 2 or 3.

t-SNE (t- distributed stochastic neighbour embedding) is a method to reduce dimensions of a dataset. This technique takes a group of points in a multi-dimensional space and converts them in a lower-dimensional space. This works on probability distributions to find the structure within the data. Before applying t-SNE, dimensions are reduced using Singular Value Decomposition (SVD) to reduce the complexity. After SVD, dimensions are reduced to 50. Further, we reduce it to 2 using t-SNE technique.

After the application of SNE we get the dataset having the count of rows as the total number of comments and the count of columns as the total number of sole tokens across the comments.

Clustering:

Clustering is a type of unsupervised learning where we are not searching for any perceptions, but we are trying to find some structures within data without expecting an explicit outcome. This technique of recognizing and clubbing an analogous type of data is termed as clustering. Units in each cluster are similar entities of that group than those in other clusters. The goal is to combine sets with like properties and allocate them into clusters. Clustering can be alienated into two subcategories: Soft and hard clustering. Soft clustering refers to a likelihood or probability of being in a cluster that is assigned to each data point. In hard clustering, each data point either fits in a cluster or not.

In this example we have used the centroid model of clustering called K-means clustering. This is an iterative algorithm where similarity is calculated by the nearness of a point to the centroid of clusters. The clusters essential towards the end are supposed to be defined at the beginning, hence prior knowledge of the dataset is crucial. K -means work in five steps. Firstly, we stipulate the anticipated figure of clusters. Then arbitrarily allocate each data point to a cluster. Then cluster-centroids are computed. Now, based on Euclidean distance, each point is reassigned to the neighbouring cluster and the cluster-centroids are recomputed until global optima are obtained. The goal of K-means algorithm is to decrease the average squared Euclidean distance of description from their cluster centroids.

In this example, we cluster the item description together in order to predict the price for the cluster. We have created 30 clusters of the products based on the item description. Based on the item description, similar products are clubbed together in one cluster. We use the TF-IDF matrix formed earlier for clustering. The products with similar customer perspectives, item description and category are clubbed together. Now cluster wise price can be decided. In order to do this, we divide the dataset into train and test data. Then we build a K-means clustering model with the training data. This model divides the products in the training data within

30 clusters and assigns a price to each of these clusters. Then using the same model, we predict the cluster and thereby the price of the products in the test dataset. When a seller needs to decide the price, he can refer to the similar product's cluster and can accordingly decide its price. This approach will mark the products at optimum prices based on the customer's perspective.

Going further we can also apply topic modelling to the same dataset and form clusters of products based on the topics that we generate. Each topic is characterized as a distribution over words. Then, each document is then characterized as a distribution over topics. The probability distributions over words provided by the topics offer a sense of the diverse ideas confined in the documents. Latent Dirichlet Allocation (LDA) is an algorithm that can be used to notice the topics that exist in a corpus.

Step 2: Linear Regression

Linear regression explains the relationship among one or more independent variable(s) and one dependent variable. This algorithm is commonly used for predictive analysis and modelling. Linear regression is classified into two categories: simple linear regression and multiple linear regression. When there is only one dependent variable, it's called a simple linear regression and when dependent variables are more than one, it is referred to as multiple linear regression. For example, it can be used to predict salaries of employees based on age, education and gender.

Multiple Linear Regression is leveraged here to predict the prices of the products. Simple regressors can make a huge impact on the decision process. Here we have presented a linear model that uses item comments and category name to decide prices of the items. The comments and the category are first tokenized and then the TF-IDF scores for the same are calculated. We trained a simple tf-idf vectorizer on the combination of the item name and description. While doing this the stop words like 'a', 'the', etc are removed from the comments. After the TF-IDF matrix is formed, we run the linear regression to compute the price of the product using the two variables - Category and Comments. The equation of regression would be as below:

$$Price = b0 + b1(category\ name) + b2(comments) \quad (1)$$

Here b0, b1 and b2 are coefficients of regression.

Linear regression runs slow on our data as vocabulary is wide in the comments field. Hence, we have used SGDRegressor with ordinary least squares. Stochastic Gradient Descent (SGD) is an efficient approach to fit linear regressors under loss functions. SGD is applied when data is huge and if it's textual data. The gradient of the loss is calculated for each sample at a time and the model is updated with a diminishing strength schedule (learning rate). Learning rate is a hyperparameter used while training the model having a value between 0.0 and 1.0.

Step 3: Prediction on Test Data

The model is built on train data and is then applied to test data. MSE of 0.63 is obtained on test data. We use mean squared error (MSE) as the evaluation metric as negative values would crash mean squared log error. There is no acceptable range of error. It depends on the data.

This model can be used by sellers to predict the prices of items referring to the reviews of the customers. The predicted prices also consider category of the product.

In comparison, if we use simple linear regression on only the numeric data from the dataset, the predicted prices have more error than those predicted by the proposed approach. Using this model, the accuracy over the test data is low and the MSE is 0.86 which is significantly greater than the earlier model.

II Price Bundling

Pricing methods can be divided into two parts: linear pricing and non-linear pricing. When a buyer pays the equal price for each piece of a product, it called linear pricing. When the total amount a customer pays for a set of products is not equal to the sum of the individual products' prices, it is called nonlinear pricing. The most common method of nonlinear pricing is price bundling. We have discussed below how analytical methods can be used for bundling to maximize the profits. Bundling the products often makes the customer buy more products than they would buy without bundling.

Bundling strategy is used in various sectors such as telecommunications, financial services, health care services, etc. Automobile companies bundle keyless entry, airbags, navigation. Computer selling companies offer monitors, scanners and printers. The food industry combines separate food items into a meal combo. For example, McDonald's offer various combos of burger fries and a cold drink. It increases average customer purchase, increases margin, increases sales of high margined products such as fries and coke and creates a perception of saving in customer's mind. [15]

We have data on electronics sales services [16]. It contains the perceived value of customers for Internet, Television (TV), Mobile phone. The dataset contains records of 77 sample customers [14]. The perceived value determines a price that the buyer is ready to pay for a artefact or service. Here, we have assumed that customers make decisions based on consumer surplus. Consumer surplus is the difference between the real price of the product and the buyer's perceived value. For example, if the price of a mobile phone is $13 and the customer's perceived value is $16 then consumer surplus is (+) $3.

These individual products can be clubbed into combinations. Traditionally this is done using Excel. This approach is discussed below:

Traditional Approach

The evolutionary solver is used to fix the price of each product combination which will maximize the revenue. The evolutionary solver is used when target cell and constraints contain non-smooth functions such as if(), sum if(), etc. Revenue that can be generated for each set of possible combinations from sample customers is calculated. This is calculated by simple summation. Then we use the Evolutionary Solver to find the optimum prices for a set of products.

Step 1: Determine possible clusters

Seven combinations are possible as follow: Internet, TV, Mobile, Internet and TV, Internet and Mobile, Mobile and TV, all (internet, TV, mobile) [16]. Price of each combination is calculated by summation. For example, if the price of the TV is $20 and the price of the Internet is $5 then the value of a combination of TV and internet is $25. Similarly, the perceived values of customers for all the combinations are calculated. Now we have 77 rows and 7 columns. We also have original prices of all 3 products and based on that we calculate prices of combinations.

Step 2: Calculate Consumer Surplus

The consumer surplus is calculated for each combination by subtracting original prices from perceived values**.** Then we see for which combination, consumer surplus is maximum. This combination would be the best choice for the customer and he or she will buy that bundle. If a surplus is negative for all 7 options, he will not buy any of the combinations. If a customer's value for a combination is greater than the actual price, he will buy it.

Step 3: Find out Optimum price

Now we use an evolutionary solver to discover the optimum price for each of the combination. Prices obtained for each bundle are as below:

TABLE 1. PRICES FOR EACH BUNDLE

Product	***Internet***	***TV***	***Mobile***	***Internet + TV***	***Internet+ mobile***	***TV + mobile***	***All 3***
Price	74.35	35	82.16	69.99	69.99	69.99	89.94

Below table shows how many customers will buy each of the combinations.

TABLE 2. COUNT OF CUSTOMERS BUYING EACH BUNDLE

Product	***None***	***Internet***	***TV***	***Mobile***	***Internet + TV***	***Internet+ mobile***	***TV + mobile***	***All 3***
Count	25	0	19	0	8	2	1	22

Step 4: Revenue Calculation

As per the above frequency table, maximum customers will prefer All 3 bundle. 25 customers will not buy anything. No one preferred to buy internet or mobile alone. Hence seller should bundle these two products together along with TV. Revenue of $3413 is obtained using new prices.

Analytical Approach - Price Bundling using Clustering:

Clustering is the approach where the customers are segmented into different groups such that the customers in the same set are like each other and significantly different from the customers of other groups. This helps in forming discrete group of customers. This approach is mainly used to reduce the large number of customers to a fewer number of groups. Clustering helps in analysing similar customers in a way and allows the segregation of other methods to be used with the dissimilar customers. Clustering has a lot of applications across domains like market segmentation, recommendation engines, medical imaging, etc.

Step 1: Elbow method to find out number of clusters

K-means to Here we refer to the K-means clustering for the segmentation of the customers. Use of few clusters helps in reducing the customers into few easily interpreted pricing segments. We use elbow method to decide number of clusters.

FIG. 2. ELBOW METHOD OUTPUT

As per the above graph, curve flattens at K = 3, we assign **K = (3)** and divide the 77 customers into 3 clusters. K- means clustering is used to divide into three groups namely - low, medium and high payers.

Step 2: Assigning each Customer to Cluster:

This is done based on the amount that the customer is ready to pay for each of the 7 products mentioned above. The Squared Euclidean distance between each customer's perceived price and the mean price of the cluster is calculated. This is done by calculating the z-score for all the customer prices. The customer is assigned to the cluster for which it has the minimum squared difference between the respective z-scores.

Step 3: Analyse Clusters:

Finally, the Excel Solver finds the optimal means of the clusters such that the distance between the customers in same cluster is minimum and customers in other clusters are far away. According to the K-means clustering method, following is the distribution of the 77 customers in 3 clusters:

TABLE 5: CLUSTER WISE COUNT OF CUSTOMERS

Low payers	***Medium payers***	***High payers***
45	30	2

The Below figure visualizes the clusters created. Green, red and blue dots represent low, medium and high payers respectively. Yellow dots represent centroids of clusters.

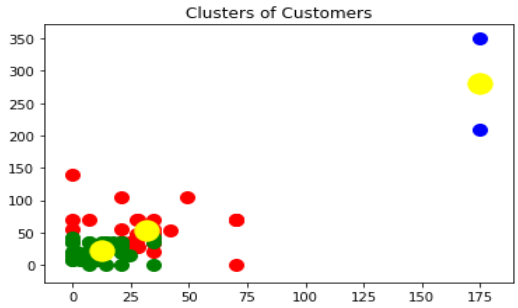

FIG. 3. CLUSTER VISUALIZATION

There are only 2 customers out of 77. Hence, we do not consider prices of high payers' cluster. This shows premium pricing would not be beneficial for this business. Premium pricing is strategy where businesses price a product higher than average market price to increase perceived quality.

Step 4: Surplus Calculation

For each cluster there is a mean price for each 7 products around which the cluster is actually formed. This mean price indicates the average price that the customers in the particular cluster are ready to recompense for the respective product. Average amount of low payers < Average price of medium payers < Average price of high payers. Every calculation henceforth revolves around this mean price.
Similar to the traditional approach the consumer surplus is calculated for each cluster based on its mean price for every product.

Step 5: Actual Buying Decision

Maximum positive surplus indicates that the customer will buy that product out of all available options. Then the total number of customers who actually buy any of the products is determined. The mean price of the products of the cluster from which maximum number of customers buy a product is taken as the optimal price for the products.

Step 6: Revenue Calculation

Based on optimal price the revenue is calculated by taking into consideration the customers from the other clusters whose mean price is higher than the optimal price. Use of this method results in a higher revenue than the traditional approach.
The prices of low payers' cluster yield maximum revenue of $3622.

TABLE 3. CLUSER-WISE PRICES AND REVENUE OF BUNDLES

		PRODUCT COMBINATIONS							
		Internet	**TV**	**Mobile**	**Internet+TV**	**Internet+ Mobile**	**Mobile+TV**	**All**	**Revenue**
CLUSTERS	**Low Payers**	58.53	35	14	87.74	35	90.39	56	3622
	Medium Payers	75.84	96.9	70.1	97.18	94.31	99.45	90	3138
	High Payers	58.34	100	0	100	33.29	92.04	100	200

RESULTS AND ANALYSIS

The results and its analysis are discussed below for proposed approaches:

I Price Suggestions

Results:

According to the proposed approach, the predicted price of a Coach bag with the following data is predicted to be 32.87, which is pretty close to the original price of **(33).**

TABLE 4. PREDICTION FOR AN ITEM

Id	**Name**	**Item condition id**	**Category name**	**Brand name**	**Shipping**	**Item Description**	**Price**
2	Coach bag	1	Vintage & Collectibles/Bags and Purses/Handbag	Coach	1	Brand new coach bag. Bought for [rm] at a Coac...	32.87

However, if we use simple linear regression on only two columns - item_condition and shipping cost, then the predicted price for the same bag is $22.46 which is poor as compared to the previous result.

Analysis:

Comparison Between Traditional and proposed approach:

Uniqueness:

When we use Natural Language Processing techniques like tokenization, TF-IDF scores, topic modelling before running linear regression, it is possible to incorporate the text data - item description and category in predicting the product prices, along with the numeric data. This helps in considering the customer perspective of the product and also the main category of the product which is a driving factor in the base price of the product most of the time.

Accuracy and Efficiency:

So, when the input of the linear regression is the item category, item description, item condition and shipping cost, together, the predicted price is closer to the actual price.
On the other hand, if we use a simple linear regression with only 2 numeric columns as the input - item condition and

shipping cost, the prediction is erroneous, and the model accuracy drops significantly.
The price predicted using traditional approach is $22.46 and by proposed approach is $32.87 which is close to actual price i.e., $33. R squared obtained using old approach is 0.57 while using new approach it increases to 0.69.

II Price Bundling

Results:

For price bundling strategy, we have used clustering algorithm and compared the results with traditional approach. Using the Analytical approach, revenue increased by $209. Hence, clustering yields more revenue than the traditional approach.

Analysis:

Comparison Between Traditional and proposed approach:

Uniqueness:

When amount of data is large, clustering outperforms the traditional method as the data is divided among the number of clusters and each cluster can be processed parallelly. Along with the prices for the combinations, the proposed approach outputs customer segmentation which can be used for positioning the products. In the discussed example, seller should target low and medium paying customers.

Efficiency:

The proposed approach yields more revenue. The traditional approach gives revenue equal to $3413 while new approach gives a revenue of $3622. Hence, the proposed approach is more efficient.

DISCUSSIONS

Analytical techniques such as NLP, machine learning are being used in many of the industries for various applications. The results show that these methods can significantly improve revenue if used in making pricing decisions.
For price forecasting, numeric data is used commonly. We have explored whether text data which is available easily and in huge amount can be leveraged. Textual data is used for sentiment score analysis to analyse whether customers are happy with the product or not. We have extended the use of this data in pricing applications. Also, generally supervised learning algorithms are used for forecasting. We have explored both supervised and unsupervised techniques for forecasting.
In marketing, clustering is used to segment the customers to target particular group of a customers. We have further extended the scope of clustering to determine the best bundles and their prices to maximize the revenue. The method proposed for bundling can be used in other business to customers markets where customers perceptions can be surveyed.

CONCLUSION AND RECOMMENDATIONS

I Conclusion

The developments in internet and data analytics have unlocked tremendous prospects in pricing strategies. Using historic data and customer's perceptions, prices can be improved to increase revenue significantly. In this paper we have discussed the following points:

- We have discussed various pricing strategies used in businesses. We have defined various terms and keywords regarding pricing and machine learning techniques.
- We have reviewed existing research in the field of dynamic pricing which includes machine learning models in pricing, use of dynamic pricing in various industries such as electricity, retails, insurance, etc.
- We have described price forecasting using text data. We have used Natural Language Processing techniques to process the data and then clustering and linear regression to forecast the prices. We have used data of ecommerce site.
- We have discussed how price bundling strategy is implemented using customer perceptions and how analytical approach can be leveraged to improve the results. Here we have used data of electronics services.

The main message of this paper is that machine learning algorithms can significantly increase the revenue if incorporated properly.

II Recommendations

We have limited our study to price forecasting and price bundling methods. The further research can be done to implement other pricing strategies discussed in introduction section using analytical or machine learning techniques.
In the price bundling approach, after clustering we used Excel solver to find the optimal prices. Further research can be done to find a method to replace the solver technique with machine learning algorithms. We have considered sample customers' perception value as the base for calculations. Here further research can be done do consider more parameters for bundling.

LIMITATIONS

Business to business (B2B) markets are beyond the scope of this study. Pricing is equally important element in B2B market. Due to the lack of availability of data, we haven't extended this research for B2B markets.
In price forecasting, we have focused mainly on dealing with text data. We have demonstrated one supervised and one unsupervised technique i.e., regression and clustering respectively. After NLP, other algorithms such as bagging, and boosting can be applied to further increase the accuracy.
In price bundling, we have used consumer surplus value to decide the prices. The data required to calculate consumer surplus is customer perception value which can be obtained from surveys and the actual price of the product which the business has already decided. Further studies should consider if there are other parameters that can be leveraged to improve prices using data analytics.

References

[1] J.-M. P. Alexandre Dolgui, "Pricing Strategies and Models," Elsevier, pp. 101-110, 2010.

[2] S. Pudaruth1, "Predicting the Price of Used Cars using Machine Learning," International Journal of Information & Computation Technology, pp. 753-764, 2014.

[3] B. H. A. L. D. S.-L. Kris Johnson Ferreira, "Analytics for an Online Retailer: Demand Forecasting and Price Optimization," Manufacturing & Service Operations Management, p. 69–88, 2015.

[4] C. V. L. K. R. S. Y NARAHARI, "Dynamic pricing models for electronic business," Sadhana Vol 30, Part 2&3, pp. 231-256, 2005.

[5] C. P. Rajan Gupta, "Machine Learning Framework for Predicting Purchase by online customers based on Dynamic Pricing," Procedia Computer Science 36 , p. 599 –605 , 2014.

[6] K. S. Hwang S.B., "Dynamic Pricing Algorithm for E-Commerce," Advances in Systems, Computing Sciences and Software Engineering, pp. 149-155, 2006.

[7] J. H. Z. F. T. S. C. R. S. S. M. K. T. W.-Y. P. T. Qu, "Demand prediction and price optimization for semi-luxury supermarket segment," Computers and Industrial Engineering Vol 113, 2017.

[8] Y. H. M. D. S. R. T. Brett Danaher, " An Empirical Analysis of Digital Music Bundling Strategies," anagement Science 60(6), pp. 1413-1433, 2014.

[9] E. B. Yannis Bakos, "Bundling Information Goods: Pricing,Profits, and Efficiency," Management Science© 1999 INFORMSVol. 45, No. 12, p. 613–1630, 1999.

[10] D. S. S. Deshpande, "Various Pricing Strategies," IOSR Journal of Business and Management (IOSR-JBM), pp. 75-79, 2018. s

[11] N. B. Nayak M., "Predicting Dynamic Product Price by Online Analysis: Modified K-Means Cluster.," Computational Intelligence in Pattern Recognition, pp. 1-15, 2020.

[12] C. B. Mario V. Wuthrich, "Data Analytics for Non-Life Insurance Pricing," Swiss Finance Institute Research, pp. 16-68, 2019.

[13] A. Shpanya, " 5 trends to anticipate in dynamic pricing," Retail Touch Points vol. 36, p. 599–605, 2014.

[14] M. Inc., Mercari price suggestion challenge., https://www.kaggle.com/c/mercari-price-suggestion-challenge, 2018.

[15] H. A. W. G. Fuerderer R., Introduction to Price Bundling, Berlin, Heidelberg: Springer, 1999.

[16] in Marketing Analytics, Wiley, 2014, pp. 107-122.