Pricing Optimization using Machine Learning

L. Indira[1], T. C. Kevin Suchetan[2], MD. Shoieb Iqbal[3], A. Rohith[4], Preetham Shinde[5]

[1]Assistant Professor, Dept. of Computer Science and Engineering, VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad, India
[2, 3, 4, 5]Student, Dept. of Computer Science and Engineering, VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad, India

---***---

Abstract - *Today, the prices of goods and services change based on various situations and demands. Another important issue is how these prices are changed and how to find the pattern for their changes. Keeping the right prices for goods and services at the right time is very important. For that predicting the prices plays a major role. But price prediction needs to consider many aspects and conditions that affect the prices. This is one of the key problems with the current system. If we consider the current market, many businesses depend on price prediction for the profitability of the organization. Price prediction has developed a huge demand in the current times and optimizing prices will make the company profitable. The aim of this paper is to optimize generated revenues by defining a pricing algorithm able to predict and optimize daily prices in response to changing daily demand. The outcomes of this paper demonstrate machine learning's ability to be useful in this task.*

***Key Words*: Optimal price, Machine Learning, Demand, Price prediction, Market, Organizations.**

1. INTRODUCTION

Traditional marketers made most of their pricing decisions intuitively and paid little attention to consumer behavior, market trends, the impact of promotions, holidays, or how they affected how sensitively the items responded to price. Most businesses are utilizing big data technologies to optimize pricing decisions as a result of advancements in high computational capabilities that support analysis of enormous volumes of data over time. This is done in an effort to provide more competitive pricing while ensuring that maximum clearance/revenue/margin targets are met.

In order to achieve business goals like increased revenue and profitability, organizations must decide on the best price for their items. The single most important factor in deciding sales and revenue is product price. Therefore, to maximize income, firms must decide the ideal price for their items. In order to train and produce a demand curve, our model reads and analyses data of a product that was gathered from retail sources using an OLS linear regression model. Using the idea of price elasticity, one may forecast the product's ideal price at which the store will make the most money overall.

2. RELATED WORK

Rajan Gupta et al., [1] recommended means of predicting and anticipating the purchases done by online shoppers. An analysis of dynamically changing the price in e-commerce and offline businesses is presented in this paper. Dynamic pricing refers to the concept of offering goods at different prices depending on the customer's demand. Five pricing techniques were covered by the author: segmented pricing, service time pricing, peak pricing, purchase time pricing, and pricing for changing situations. Changing the price of the good or service in accordance with the customer's willingness to pay is known as segmented pricing. Peak user pricing is more frequently used in the airline and railroad sectors, where users are substantially taxed during peak hours. Service time pricing refers to charging high prices for short service periods or predetermined delivery deadlines. Purchasing period pricing speaks of the moment of purchase whenever the flight's take-off time is less. As a final step, changing conditions pricing is used when a product's market is uncertain.

Akhiro Yebe et al., [2] proposes a novel robust quadratic optimization framework for prescriptive price optimization. Robust quadratic programming was developed as a cautious upper-bound minimization as a result of statistical data showing that the estimation uncertainty in machine learning follows a matrix normal distribution. Two steps constitute the major contributions. Firstly, we demonstrate that, when the least square estimation is used, uncertainty in prescriptive pricing optimization may be represented in the form of matrix normal distribution. As a conservative lower-bound maximisation, this offers a naturally robust formulation of a price optimization. Second, we propose algorithms for robust quadratic optimization consisting of sequential relaxation to a non-robust counterpart that employs a non-robust algorithm as its sub-routine. Both practically and also theoretically, the sequential algorithms used for robust quadratic programming converge quickly, and they may be used to create non-robust pricing optimization techniques. The approach enables users to achieve both lucrative and safe pricing strategies in the prescriptive price optimization, according to experimental findings on both fake and real price data.

Winky K.O. Ho et al., [3] build a system, which is built on three machine learning algorithms: Support vector machine, random forest, and gradient boosting machine for the evaluation of real estate costs. This research is an experimental attempt to estimate house prices using these three machine learning methods, and then analyse these results. According to performance measures, sophisticated machine learning algorithms can estimate property prices accurately. First, the paper has demonstrated the potential of cutting-edge machine learning techniques for property analysts to employ in home price forecasting. These algorithms have some restrictions of their own. Second, compared to more established techniques like the hedonic pricing model, machine learning algorithms frequently require much longer computation times. The algorithm is selected depending on a number of variables, the amount of the data, the computational power of the tools, and the duration of availability of a waiting period for the entire results.

This research [4] introduces a progressive machine learning algorithm and optimization strategy, in order to find the ideal pricing point for each unique product in the fashion E-commerce sector. There are three main parts to it. First, a demand estimation model is adopted to forecast every product's demand, the following day at a specific discount rate. Next, by varying the discount percentage, the principle of price elasticity of demand is applied to obtain a number of demand values. For each product, as a result, multiple price-demand pairings are made, and one of among them is chosen for the further computing. E-commerce typically comprises millions of many products, therefore there are numerous arrangement possible. And for every combination, a different pricing point is set for all the products, adding up to a different income amount. Finally, one pricing range for each product is chosen using a linear programming optimization technique in order to maximise overall profit.

J. H. Zhang et al., [5] introduced a system that supports in decision-making for retail goods pricing and revenue optimization. The study made use of 2.5 years of sales information from well-known retailers in 45 distinct localities. Clustering and filtering are done using the R platform to redefine it, and the optimization model is applied. To forecast weekly demand, a machine learning system based on regression trees and random forests is used. Price, holidays, promotions, inventory, and other regional aspects too are taken into consideration while making decisions. The use of multiple trees in random forest reduced the scope of errors. Following that, a mathematical model of integer linear programming is used to calculate and include demand-price interdependencies for the optimal price allocation. Branch & Cut and Branch & Bound methods were used to maximize the expected revenue, after root node analysis was performed. By using heuristic techniques, the revenue is further optimized. When compared to branch and bound, the expected revenue from branch and cut is 108.24% higher, and after heuristic modification, it increased an additional 5.28% on average.

Giorgio Spedicato et al., [6] studied how machine learning techniques may replace traditional GLMs and increase policyholder retention and conversion estimation. The data for the analysis came from two-month individual motor liability insurance quotations. Open-source software has been used to make the study easily replicable, such as the H2O data mining programme (H2O.ai team 2017) and the R Core Team 2017. As an illustration, a conversion model focusses on the binary variable "Convert," which has two possible outcomes: Convert (Yes), Reject (No). XGBoost, Gradient Boosting Machine (GBM), Random Forest and Generalized Linear Model (GLM) are the models used and the metrics used for the comparison of the performance of models are Area Under Curve (AUC), Quote Nb (Naive Bayes) and Logloss. The computational time required for GLM is relatively low compared to that of most other machine learning models, such as GBMs. In terms of prediction accuracy, boosted models (GBM, XGBoost) perform best experimentally. On the AUC scale, the performance difference between Machine Learning methods and traditional GLM is greater than on the log-loss scale.

The authors [7] put forth a novel framework that is learning-based and primarily uses kernel regression. It was used and tested for several shop categories of a major European e-commerce business that focuses on family and children's products. Common pricing strategies such as competition-based pricing, pricing based on derivative-following Algorithm (DF) and model-optimizer algorithm were discussed. The proposed approach is also compared with these algorithms. The suggested framework obtains previous sales figures for the product whose new price is to be set. It calculates the probability for each historical price which is considered to be optimal. Kernel regression is used here to predict the new prices based on the past data in a robust manner. To deal with sparsity, statistics on past sales of comparable goods and competitor prices is considered. Higher-level predictions through prior function which uses decision trees are included in the prediction model. Finally, Metropolis-Hasting's algorithm is used to sample new pricing points. The framework designed is flexible, adaptable and generic which can be used for specific problems in hand. Revenue and profit improved progressively over time, reaching an increase of 28.04% and 20.64% from the first four months of measurement.

Massimiliano Moro et al., [8] developed a scalable revenue maximization system for an alcohol company, by observing the previous seven years of data. different models were developed for maximizing revenue. models were capable of handling the halo effects and cross effects. among the different statistical and automated machine learning models like Auto Arima, SARIMAX, and machine learning prophet best model was selected. these models were performing well than the basic naïve models. after using these models, revenue can potentially increase by 29.7%.

Rainer Schlosser et al., [9] studied the behaviour of two reinforcement learning algorithms for pricing a product in competitive online markets. The algorithms Soft Actor-Critic (SAC)and the DQA (Deep Q-Networks) were tested against each other. Both algorithms displayed strengths and weaknesses. These algorithms were tested in fixed, undercutting, two-bound pricing strategies. These algorithms are tested in duopoly and oligopoly settings; episodes for testing ranged from 50,000 to 1,000,000. The algorithms behave differently in each scenario. Both algorithms work fine in dynamic pricing, but these are tested one after the other which does not produce stable prices. To perform well, these algorithms require a large number of observations. Both algorithms performed well after 400k episodes of training.

Ruben van de Geer et al., [10] build a finite-mixture logit model, each customer in system is chosen using segment-specific parameters and the multinomial logit model. This model is to be developed towards a price optimization problem. where previously proposed price optimization models require exponential time for product count. but this finite mixture model is polynomial in products. The result is the heuristic approach of this model can be stuck into local optima. This algorithm runs efficiently for a wide range of instances.

3. BACKGROUND AND SYSTEM ARCHITECTURE

Here, we will discuss about the system architecture and explain about the sample dataset that we used to train our machine learning model.

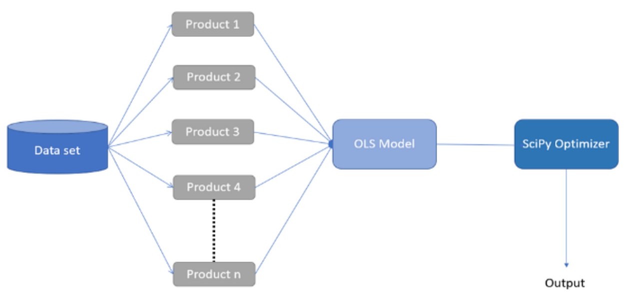

Fig. 1: System Architecture

It is a conceptual model that describes the structure, viewpoints, and behaviour of our system.

3.1 Machine Learning

The field of artificial intelligence includes machine learning, which is used to build some methods that perform some tasks. They follow a learning algorithm for training the model and aim for high accuracy by repeating the iterations. These methods have a wide variety of applications in various fields. This constant feedback improves the accuracy of the model and enhances the model's performance. It increases accuracy and enhances performance.

3.2 Data Set

The following attributes are included in the dataset:

Average Price/Unit: The price for which any product is sold in the market.

Cost/Unit: Product cost at the time of purchase.

Average Profit/Unit: The profit per unit.

Average units sold: Average number of units purchased per customer.

Incremental acquisition: The rate at which customer response increases for every 10% decline in the unit price.

Increase in sale volume: This is the increase in the sales of each product after decreasing the price by 10%.

3.3 OLS Model

OLS model stands for ordinary least squares model. It is a regression model. It is used to decrease the squares of distance between actual values and predicted values. In our model we establish the relation between the price and quantity using the OLS model which gives us the demand curve.

3.4 SciPy Minimizer

SciPy package is a core package of python which is also known as scientific python package has functions like minimizer that is used to fit the model under the circumstance of one or more than one variable. Using this package, the prices of the products are increased making sufficient profits according to the given constraints and criteria. We use SciPy minimizer to find the selling price of the product where the profit is maximum with less affecting the demand also by taking the constraints into consideration. In short, the SciPy minimizer is used to find the optimal price of the given product.

3.4 Matplotlib

Matplotlib package of python language is used to visualize the data and relation between the data variables. It supports many visualizing tools and helps to represent the summary/conclusions in different formats. This comprehensive library makes figures that are interactive and helps to generate quality plots. These visuals help us to understand the data and helps us to draw useful insights from the data.

4. IMPLEMENTATION AND PROPOSED SYSTEM

We have implemented our proposed system using the following steps:

Step-1: Collection and reading of the data set.

Explanation: The data set contains the product information that was obtained from a retail location. In our project, the data set is read and later this data is utilized for the training purpose of the model.

Step-2: Data pre-processing

Explanation: The data cleaning is done by removing the unrelated data, outliers and noisy data which would increase the accuracy of the final result.

Step-3: Data Visualization

Explanation: The graphs are drawn based on the collected data which gives us the better visualization about the collected data such as what is the range of the price and quantity of products in the data.

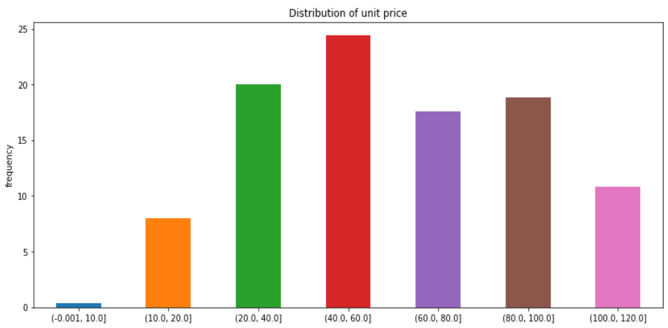

Fig. 2: Distribution of unit price

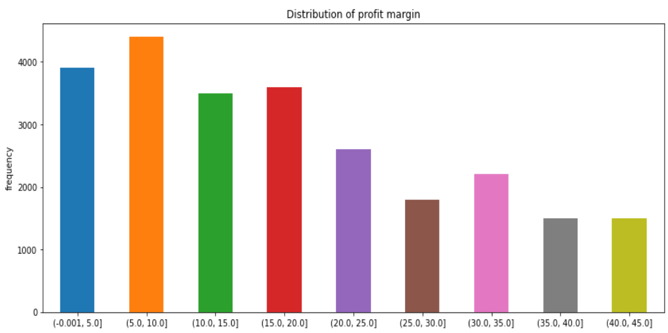

Fig. 3: Distribution of profit margin

Step-4: Training the OLS model to generate the demand curve.

Explanation: The price and quantity of product from the dataset is iteratively fitted into the OLS model to generate the demand curve for each product. The OLS model returns two parameters, the slope and the intercept of the demand equation.

Step-5: Profit maximization using SciPy Minimizer.

Explanation: The selling price is obtained from the demand equation at which the maximum profit can be obtained for each product. This price is called the optimal price. Setting the selling price to the optimal price will give us the maximum total profit.

5. RESULTS

The result shows the total profit before using our model and after using our model.

Profit before optimization: 3285.8999999999996

Profit after optimization: 6600.85076980578

The graph has been plotted between the price of a product before and after optimization.

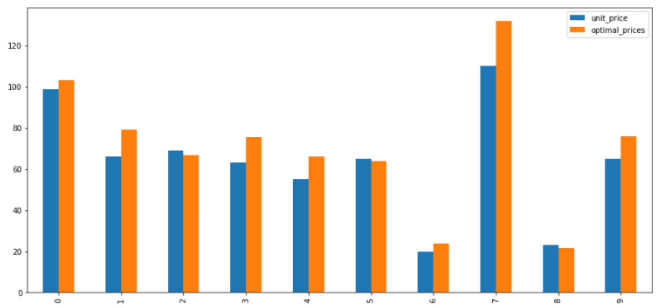

Fig. 4: Initial and Optimal Price of random products

6. CONCLUSION AND FUTURE SCOPE

With the rise in both the number of products and the number of users, companies need to compete with their competitors by giving optimal prices to their products. So, we can see how important price optimization is for the companies to stand in the market. Using a machine learning model makes predicting the prices for the product easy and simple by understanding the trends in the market and drawing patterns.

The Proposed solution considered the demand equation and has done the price predictions. But there can be many other factors that can influence the prices of the products. Analysing all required factors and incorporating those criteria will make a very efficient and robust model.

Future studies that can be extended to the system are

1. including other factors like seasonal offers, demands, salary of the customer, age of the customer etc.
2. predicting the prices using demand pricing.
3. Using more robust algorithm for training the model.

ACKNOWLEDGEMENT

We are very grateful to have L. Indira as our guide for this project who gave many useful ideas for the project and guided us to complete the project as part of our academics with great results in the end. Our special thanks to her for the immense support during the project.

REFERENCES

[1] Gupta, R. and Pathak, C., 2014. A machine learning framework for predicting purchase by online customers based on dynamic pricing. Procedia Computer Science, 36, pp.599-605.

[2] Yabe, A., Ito, S. and Fujimaki, R., 2017, January. Robust Quadratic Programming for Price Optimization. In IJCAI (pp. 4648-4654).

[3] Winky K.O. Ho, Bo-Sin Tang & Siu Wai Wong (2021) Predicting property prices with machine learning algorithms, Journal of Property Research, 38:1, 48-70, DOI:10.1080/09599916.2020.1832558

[4] Sajan Kedia,Samyak Jain,Abhishek Sharma (2020) Price Optimization in Fashion E-commerce, arXiv:2007.05216v2 DOI:10.48550/arXiv.2007.05216

[5] Qu, Tianliang & Zhang, Jianghua & Chan, Felix & Srivastava, R.S. & Tiwari, Manoj & Park, Woo Yong. (2017). Demand Prediction and Price Optimization for Semi-Luxury Supermarket Segment. Computers & Industrial Engineering. 113. 10.1016/j.cie.2017.09.004.

[6] Spedicato, Giorgio & Dutang, Christophe & Petrini, Leonardo. (2018). Machine Learning Methods to Perform Pricing Optimization. A Comparison with Standard GLMs. 12. 69-89.

[7] Josef Bauer, Dietmar Jannach, Optimal pricing in e-commerce based on sparse and noisy data, Decision Support Systems, Volume 106, 2018, Pages 53-63, ISSN 0167-9236, https://doi.org/10.1016/j.dss.2017.12.002. (https://www.sciencedirect.com/science/article/pii/S016792361730221X)

[8] Gupta, Nikhil; Moro, Massimiliano; Ayala, Kailey A.; and Sadler, Bivin (2020) "Price Optimization for Revenue Maximization at Scale," SMU Data Science Review: Vol. 3: No. 3, Article 4. Available at: https://scholar.smu.edu/datasciencereview/vol3/iss3/4

[9] Kastius, A., Schlosser, R. Dynamic pricing under competition using reinforcement learning. J Revenue Pricing Manag 21, 50–63 (2022). https://doi.org/10.1057/s41272-021-00285-3

[10] Van de Geer, Ruben and den Boer, Arnoud V., Price Optimization Under the Finite-Mixture Logit Model (December 23, 2019). Available at SSRN: https://ssrn.com/abstract=3235432 or http://dx.doi.org/10.2139/ssrn.3235432