Pricex – An Intelligent Pricing Decision Support Tool

Pulkit Shukla, Prince Kumar, Sachit Bolisetty, Sonia Daryani, Matthew A. Lanham
Purdue University – Krannert School of Management
shukla30@purdue.edu; kumar355@purdue.edu; sboliset@purdue.edu; sdaryani@purdue.edu; lanhamm@purdue.edu

DATA STORY

Business Problem: Pricing is an important decision for all businesses. Price too high and you lose business to competitors, price too low and you lose margin. In collaboration with a national manufacturer, they seek to obtain a more analytics –based pricing model that can support their current pricing tool. The product offerings in their market are not differentiable via brand, quality, or other features, rather solely by price.

Analytics Problem: While the traditional cost-plus pricing strategy allows the manufacturer a certain degree of profits, data-driven price differentiation can provide an opportunity to convert the quoted bid price into a successful purchase, while not loosing too much margin. We frame the problem as a classification prediction problem, where the firm obtains the business as a quoted bid or not. Such a pricing model will work as a decision support tool to help the sales teams quote bid-winning prices while maximizing profits for the company.

Research Questions:

- How can the firm effectively predict a successful bid at a given price?
- How would the firm maximize their profits for using this model to set future prices?

Literature Review: A multitude of papers discuss the development of predictive models to assess wins and losses for different bidding opportunities. Our solution is modeled in a similar fashion to Agarwal and Ferguson, whom created a bid-response model for customized pricing and use customer characteristics and transaction attributes to predict the probability of winning a particular bid and then optimize the expected profit function to get a final price suggestion. How and which features we used are proprietary.

Figure 1) provides a diagram that supports the general idea of how features are input into a predictive model, and then that model is used in the objective function for a pricing optimization model.

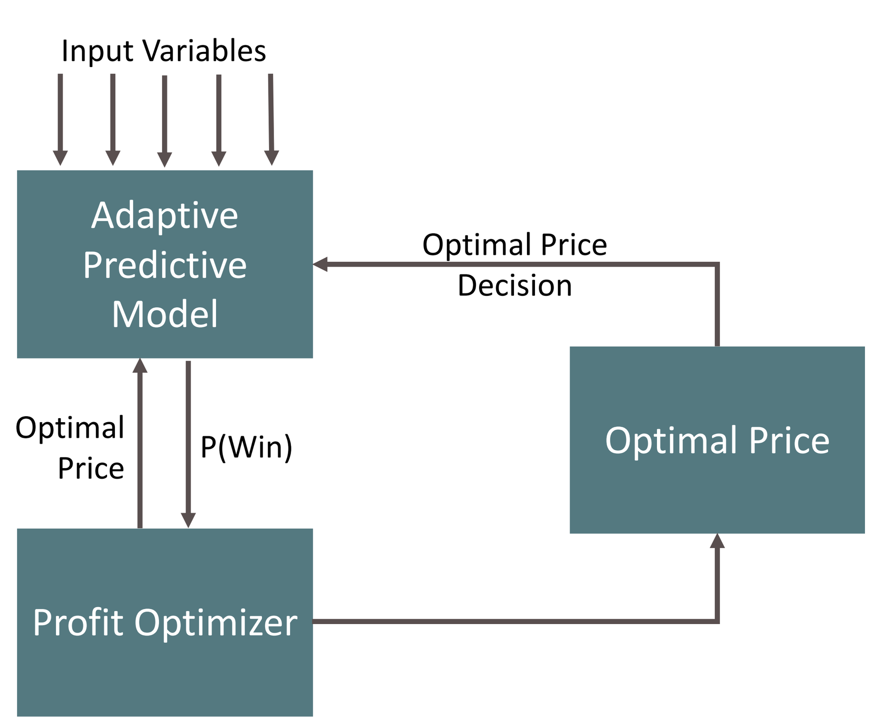


Figure 1: Model Overview

METHODOLOGY

Considering the size of the dataset size, LOOCV (Leave-One-Out-Cross-Validation) approach is implemented to train and gauge generalizability of the predictive model.

A multivariate logistic regression is used to model the relationship between a bid win/loss, price, and non-price factors. The probability of winning a bid and the related price is used to calculate and optimize the expected profit. A simplistic model form is below:

$$W_i = \beta_0 + \beta_1 p_i + \beta_2 Q_i + \cdots \quad [1]$$

$$\rho(p_i) = \frac{1}{1+e^{-W_i}} \quad [2]$$

$$Maximize\ E(\pi) = \rho(p_i) \times (p_i - c_i) \times Q_i \quad [3]$$

Figure 2: Model Equations

A key aspect of the model used to maximize the expected profit is the probability of winning a bid, which is a function of price (our decision variable). This probability can be calculated through multiple methods and to that effect multiple models were developed that differed in the equation used to measure p(pi) (as shown in Figure 2, Equation [2]) . The prediction models included Neural Network and Random Forests

CONCLUSION

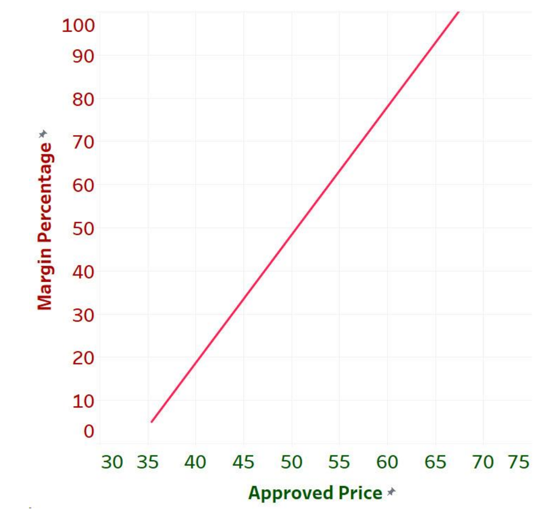


Figure 3: Profit vs. Price

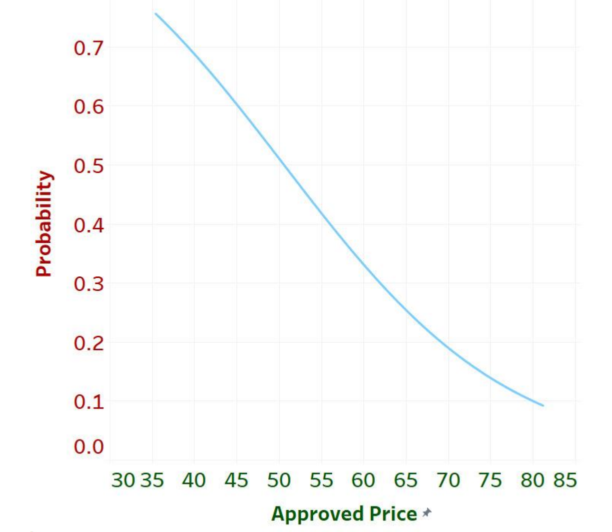


Figure 4: Winning Probability vs. Price

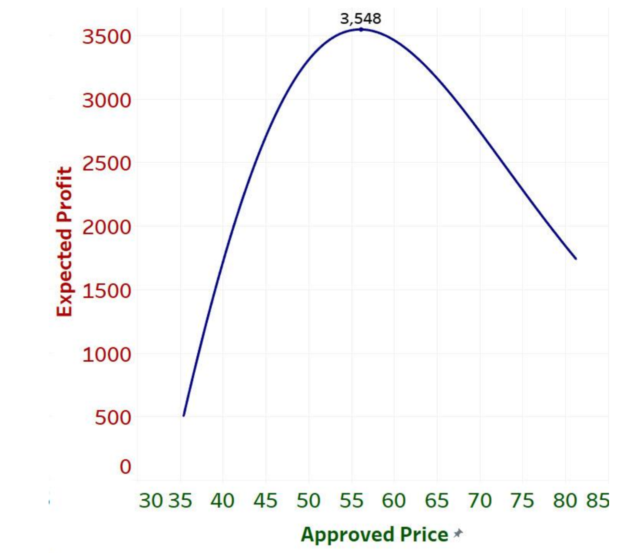


Figure 5: Expected Profit vs. Price

Figures 3 and 4 depict the linear increase in profit (i.e., difference of price and cost) and the decrease in winning probability, respectively, as price increases. The expected profit is an amalgamation of the two relations, and hence, is a curved graph (Figure 5) with the turning point constituting the optimal maximum profit.

Figure 6 depicts the accuracy achieved by the Logistic regression, Naive Bayes Classifier, Neural Network, and the Random Forest models on the train and test datasets (Train :Green , Test: Grey) . The highest classification accuracy of 77% is achieved in the Random Forest and Naive Bayes Classifier but based on interpretability and adoption constraints the final model chosen to predict the win probability in the logistic regression model.

The logistic model used for the optimization has good accu and is used for the final model as it allows for easy integrati with the optimization tool, and aids with the interpretability the results.

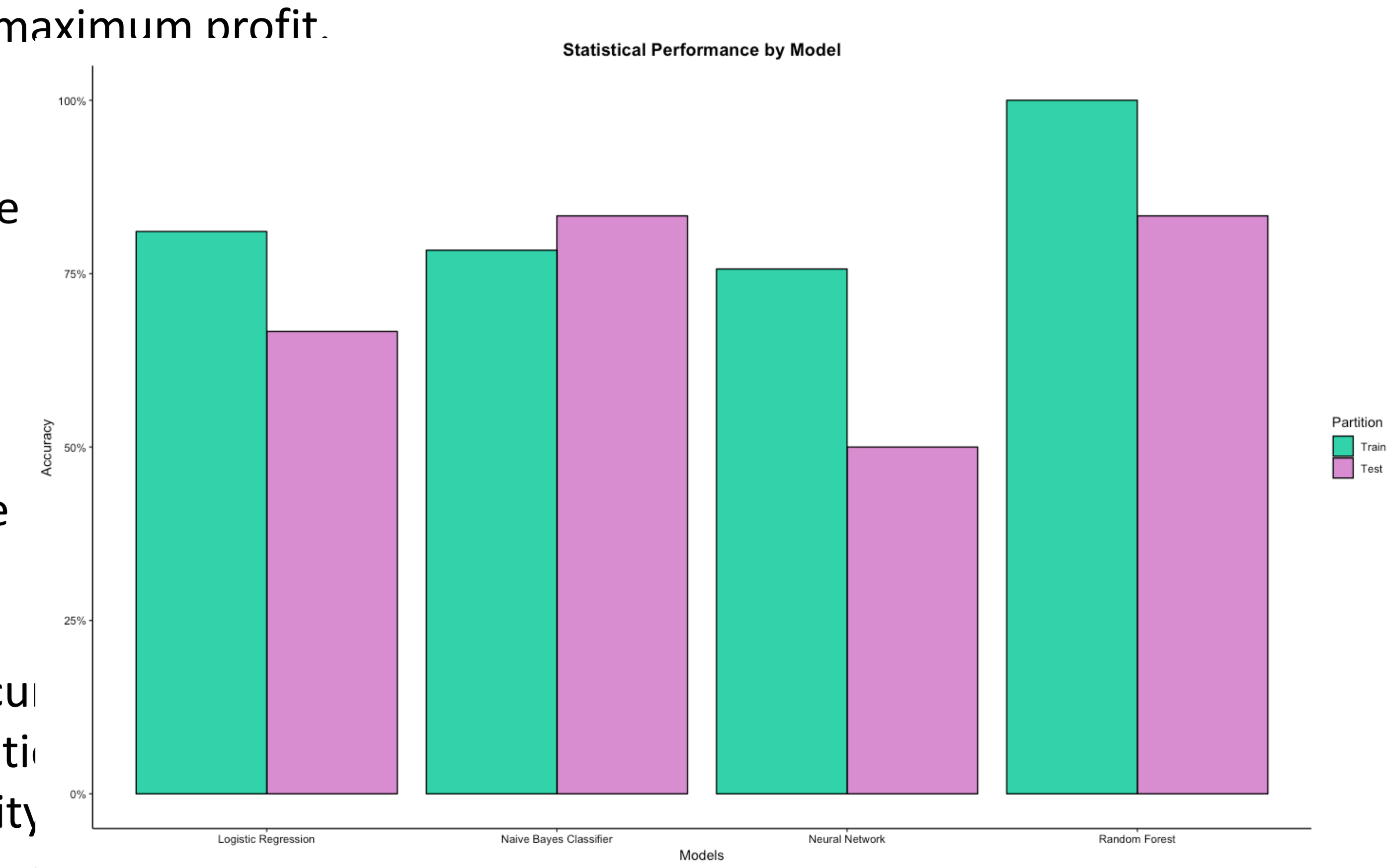
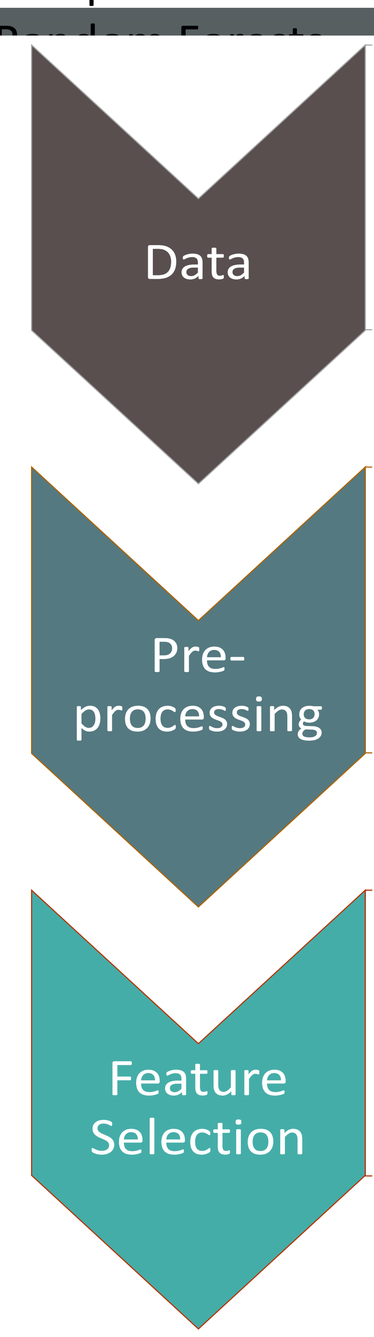


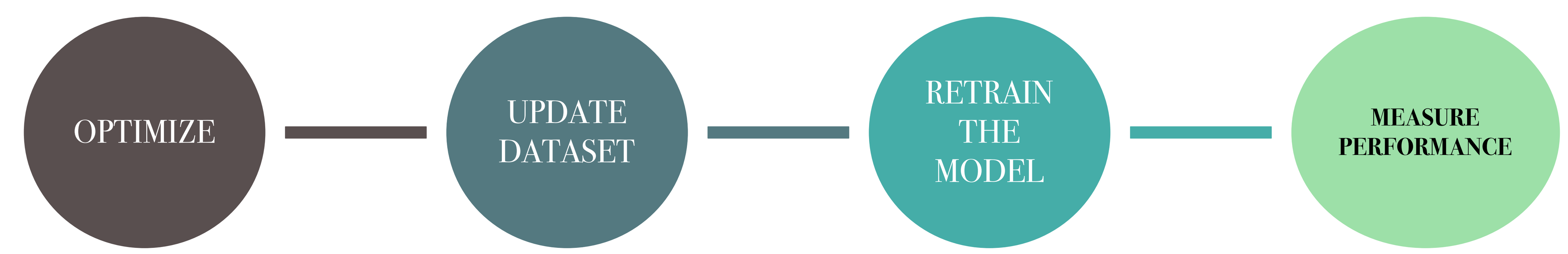
Figure 6: Accuracy vs. Regression Model



DATA DESCRIPTION

1. The dataset comprises of sales, price, bid (win/loss) information, and other non-disclosed features.
2. Missing values are removed as model-based data imputation was not a viable option.
3. Dummy variables are created for n-1 categories under each predictor variable.
4. Cost data is merged with bids data based on product types.
5. The post-processed data did not allow for any significant feature engineering – an age value is created as the difference between the start and end dates of the bids to track the change in market prices over that time period.
6. Final Variables: historical results of bids, along with related prices, quantities, customer segments, product types, relevant teams, opportunity types, and age.

IMPLEMENTATION ROADMAP



- Input new bid data
- Run predictive model
- Run solver to obtain suggested price given user constraints

- Maintain a consistent dataset.
- Capture additional predictor variables
- Update the dataset

- Retrain the model
- Export the retrained model
- Update model equation for probability computation

- Collect data for decisions made with and without the pricing tool
- Compare for performance improvement

ACKNOWLEDGEMENTS

We would like to thank Professor Matthew Lanham for his continued support throughout the project. Also, we would like to thank our industry partner for showing us how their current pricing model works, and allowing us to help them with their current pricing analytics initiatives.