

Profit Optimization

Quinn Burzynski, Lydia Frank, Zac Nordstrom, and Jake Wolfe

Dr. Song Chen and Dr. Chad Vidden, UW-LaCrosse Mathematics Department

ABSTRACT

Each branch store of Fastenal is responsible for the growth of business and sales in each of their local markets, where local pricing and quoting practices guide growth. In this paper we created and validated a model to optimize profit through quoting practices with new customers. We began by utilizing a random forest to analyze quoting factors and other patterns in quoting success vs. failure. We have discovered optimal discounts per product categories to maximize profit when selling to new customers.

PROBLEM INTRODUCTION

Fastenal partnered with the PIC Math¹ course from the Mathematics department at UW - La Crosse. They asked us to analyze data sets, answer questions, and procure results pertaining to a specific project.

Fastenal, founded 50 years ago by Bob Kierlin, is now a \$13 billion company and is based out of Winona, Minnesota. They employ over 19,000 people to work at 2,500 stores in 21 different countries. With Kierlin's belief of "Growth through customer service" Fastenal has made 4 billion in net sales. Fastenal focuses their sales on nuts and bolts and all other fastening devices along with safety gear and work tools. Additionally, they build and distribute vending machines for each of the aforementioned items of which they have installed more than 62,000.

Challenge

We have been asked to discover the key factors that influence a quote's success or failure and to develop our own quoting strategy, or pricing guide, to maximize sales and profitability. To clarify, a quote is an estimated price on a single good or a bundle of goods. In this case, a quote may be delivered verbally, in a storefront; or via phone, email, or memo.

Our first step is to determine the factors involved in quoting. These factors can range from the brands being sold, to the customers they're selling to, to the location of the store, to the price of the quote itself, etc. Our next step is to use the factors we find to develop a pricing guide that will offer better bids for Fastenal's stores to its customers. From our guide we will be presenting a 'what-if' scenario for Fastenal of how much profit they will see using our guide verses sticking with their current practices.

Preliminary Results

For simplicity we initially chose only four key factors to consider in our model: customer name, category ID, quantity, and unit price, per Fastenal's suggestion. These four factors directly affect the customer and their likelihood to purchase a product. As the project developed, we realized that our focus would need to lie with wholesale price, unit price, discount, and quantity purchased.

¹ PIC Math stands for Preparation for Industrial Careers in Mathematical Sciences. This course engages students in a semester-long group project, a research problem coming directly from a company in industry, that focuses on honing problem solving, critical thinking, independent thinking, and communication skills. With guidance from professors, students are given a problem to address, solve, and present as a small group. These activities are expected to increase awareness among mathematical sciences undergraduates about non-academic career options. Additionally, this course provides each student with valuable research experience while working on a real world problem.

We subset our data according to price and quantity ranges and created individual random forest models for each small group of the data. This separation allows for a higher accuracy (around 70% - see **Figure 5**) in the model and for each subset to stand on its own. We used a simulation to determine the discount rate to maximize the profit. Reference **Figure 1** as an example of our results displayed. For instance, a quote that falls into the category of being priced between \$0 - \$1.32 and having a quantity of 0 to 2, to maximize profit for this quote, it should be presented to the customer with a 7% discount. Note that the choice of variables used to subset the data can be very flexible and the model can also be easily adapted to maximize other target values, such as quoting success, rather than profit.

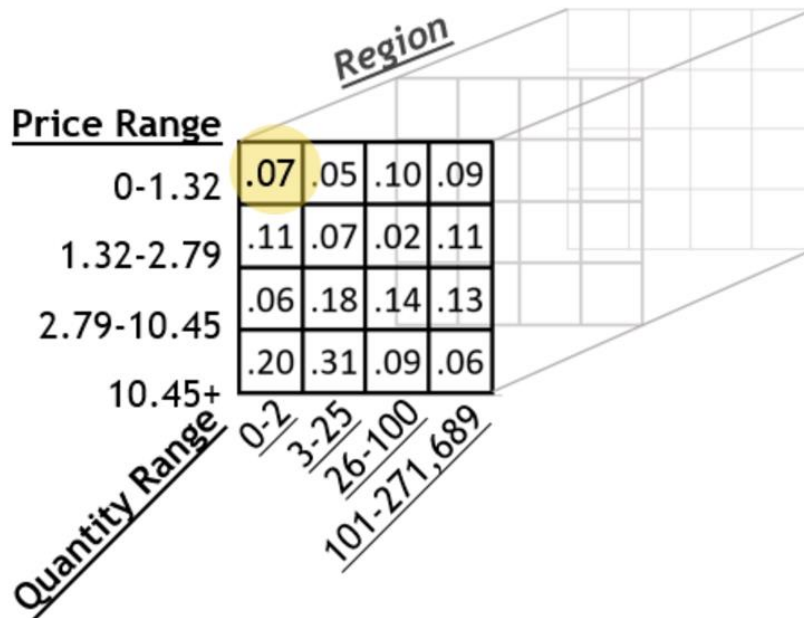


Figure 1. The horizontal axis breaks down by quantity range and the vertical axis breaks down by price range. To find the optimal discount percent, find the corresponding quantity and price ranges for a quote and follow the column/row to the intersecting matrix square. There you will find the optimal percent discount.

PROBLEM APPROACH

Problem Formation

Our overall goal is to identify any inefficiencies in Fastenal's quoting practices and make adjustments to create growth in sales and profit using a predetermined quoting guide in stores.

All of the models we worked with are *classification algorithms*. Classification problems map input variables into discrete categories of either 1 or 0, success or failure. For our project, the classification problem produces the probability that our initial quote will succeed or fail. If the probability yields anything less than 0.6, the model outputs a 0 or “failure”. If the hypothesis yields anything more than 0.6, the model outputs a 1 or “success”.

The 0.6 probability split was chosen through a cross validation process to maximize the model's accuracy. The cross-validation process is a “trial and error” cycle with various probability splits.

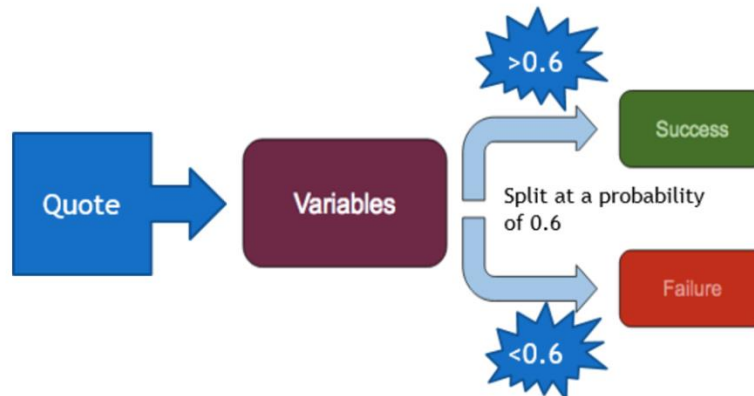


Figure 2. A graphical representation of the flow of a general classification model.

The steps of our approach to the whole problem can be visualized in **Figure 3**. Our first step was to clean our data (discussed in Data Cleaning) and to merge together the useful tables. The next step was to pick the appropriate classification model. We tried logistic regression, support vector machine, neural network², and random forest. We chose to use a random forest over other classification models for a few reasons. Firstly, random forest out-performs other models on accuracy. Secondly, the random forest is a simple model for implementation - R Studio offers a very sophisticated random forest package, which is very easy to tune to fit our data. Lastly, the random forest model can handle unknown categories for existing variables whereas other models cannot. This feature was important to our project because of the “category ID” variable. If a new category ID is created, that doesn't impede the model from running properly, which is very important for a thriving, growing business like Fastenal.

After completing the construction of the classification model, we ran our model with a subset of the merged data table to predict the profit increase within those subsets. The purpose of using only a subset to begin with is to minimize the time associated with each run of our program and to remove outliers that may sway our results. After establishing confidence in our model, we picked a second subset of data to run through the model as a simulation. The purpose of choosing a separate grouping of data is to test the model's accuracy in a variety of scenarios, rather than just with the data used to construct it.

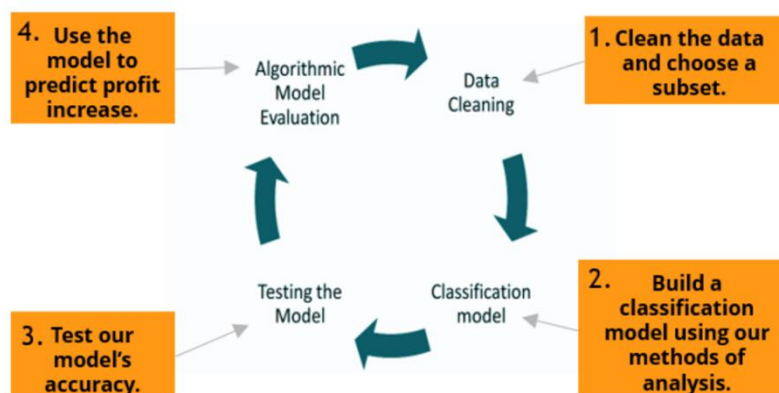


Figure 3. An overview of our steps of the problem.

Throughout the simulation process, we may need to go back and tweak parts of our model. This is dependent on the results we receive after running the separate subset(s) of data. We will repeat this process with multiple different subsets of data.

² We do see potential in the use of a neural network. There is more information pertaining to neural networks in section 3.2.

Once we have made our final conclusions on this project, we will be able to present a “what if” scenario; We will be presenting our last report to show what the company would gain, and in which areas, by using our model.

Data Cleaning

In the data, there are more than 50 variables associated in the quoting process. In our project, we are choosing to focus more specifically on the variables that directly affect the customer's decision (i.e. wholesale price, unit price, discount, and quantity being purchased).

For our initial subset, we cut out any items that fall into the following categories...

- Items with a discount percent less than 0% or greater than 100%;
- Any item where the unit price is less than the cost of the good, because this would yield no profit;
- Customers who have a success rate less than 10% which are the people who almost never buy, no matter what quote they're offered;
- Customers with a success rate greater than 90%, which are the customers who almost always buy, so there isn't room for improvement;
- Any incomplete sections of data, for example if something just had a missing piece of information or an entry of “NA”.

The purpose of eliminating the aforementioned pieces of data are to avoid using outliers which may skew or bias our model and to shrink the data set slightly to allow our model to run faster.

Feature Engineering

Utilizing the variables in the given data, we created two new variables: discount percent and percent success. Particularly, the discount percent is a variable that we use in the final results of the project. The calculations are shown below in equations (1) and (2).

$$\text{Discount Rate} = \frac{(\text{Wholesale Price} - \text{Unit Price})}{\text{Wholesale Price}}$$

$$\text{Percent Success} = \frac{\# \text{ of Successful Quotes}}{\# \text{ of Total Quotes}}$$

The percent success, equation (2), was calculated per customer.

Random Forest

We used a random forest as our model. Essentially, a random forest model is made up of an ensemble of decision trees. Therefore, to explain a random forest, we will first go through the functionality of a decision tree.

Decision trees are a decision support tool. They use a tree-like graphic, true to their name, to display outcomes such as probability, cost, consequences, etc. At each “decision point,” the *burst node* - a.k.a. split node - represents a test on a specific attribute of the entered data. The branch that follows then displays the outcome of that previous test.

For our project, we took the aforementioned subset of data and allowed for the tree to split the data by the most significant factors. Our model uses a decision tree as part of the calculations; we do not get a decision tree as a final product. The random forest we are using is made up of many many decision trees just like the one in **Figure 4**.

Let us reference **Figure 4** as an example of a decision tree. We began with a random subset of data. As the data hits each node, if it responds ‘yes’ to the test, that data flows to the left, and if it responds ‘no’ to the test, that data flows to the right.

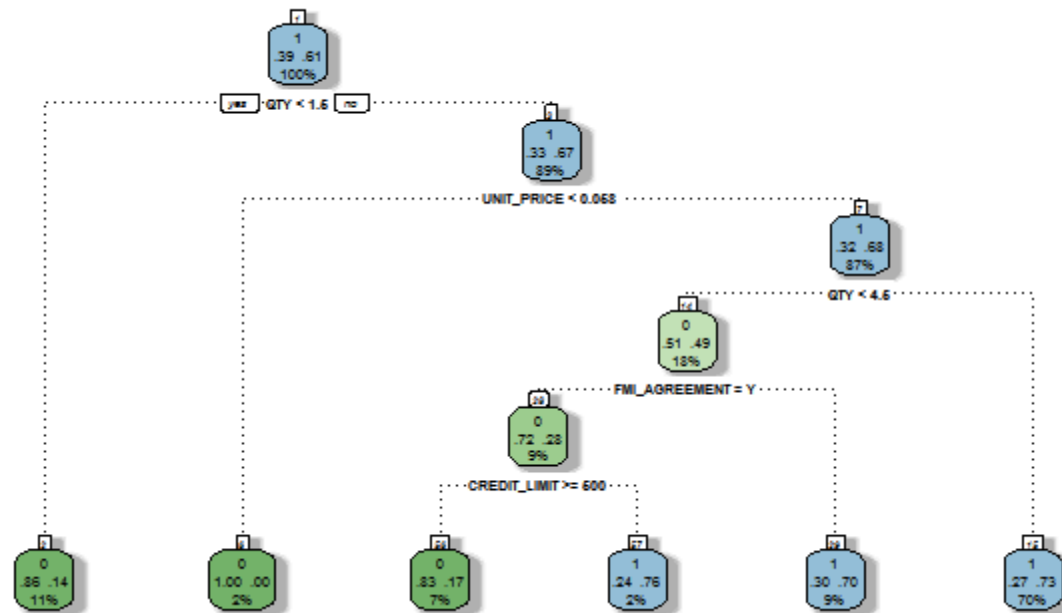


Figure 4. This is a decision tree where the green indicates an overall failed quote and the blue indicates an overall successful quote.

We follow the tree to the right, with the data that responded ‘no’ to the first test. The data here is checked to be less than a unit price of \$0.068, where the specific value is determined by the model. If the data responds ‘no’, in other words has a unit price greater than \$0.068, then it continues to the right, which we will again follow.

The third split node is checking the quantity in each quote to be less than 4.6. For the data that responds ‘no’ to this prompt, being that the quote has a quantity more than 4.6, we follow to the right. We see in the rightmost, bottom node that the data we are left with contains quotes with a unit price greater than \$0.068, and a quantity greater than 4.6. This data makes up 70% of the initial subset and has a quoting success rate of 73%.

	Reference	
Prediction	0	1
0	30229	16830
1	9811	40872

Accuracy : 72.74%
 No Information Rate : 59.04%

Figure 5. A confusion matrix displaying the accuracy of our results and our model.

We created a confusion matrix to check the accuracy of our model, which is shown to have an accuracy of about 73%. **Figure 5** shows that our model is a 13% improvement to just quoting with no rhyme or reason. This kind of improvement over the “no information rate” tells us that this is a good model, particularly to use for new customers. A no information rate would be like if you were presented with 100 quotes, 60 of which were successful. If you closed your eyes, waved your hand and said that they would all be successful, you would be 60% accurate, i.e. you would have a no information rate of 60.

Results

We began by using our method of sub-setting to split the data. Then we created individual models for each subset of data. The results took the form of a two-dimensional matrix with the ability to add a third dimension (see **Figure 1**).

With the information in **Figure 1**, we created the plots shown in **Figure 6** and **Figure 7**. **Figure 6** shows each calculated discount percent plotted against the profit it yields for a *subset* of data. This is how one would plot and visualize the data received from the model. This is a helpful representation because it is clear at which discount percent the peak profit occurs for this subset.

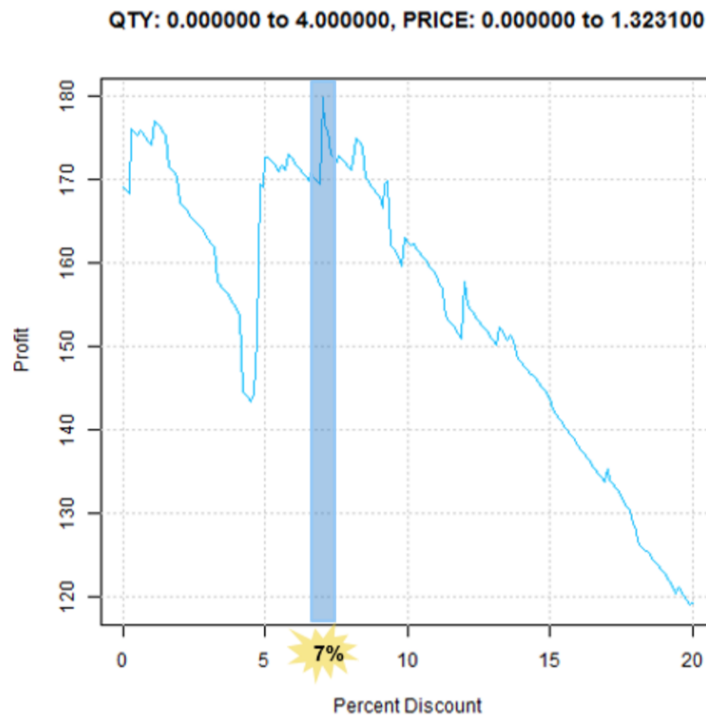


Figure 6. Profit plotted against percent discount for a subset with a quantity from 0 to 4 and a price from \$0 to \$1.32. The highlighted peak indicates the optimal percent discount for this subset of data.

Figure 7 shows the graphical representation of one specific quote taken from the subset displayed in **Figure 6**. Specifically, **Figure 7** displays the profit against discount for “Quote 91994,” which comes from the aforementioned subset with a quantity of 0 to 4 and a price from \$0 to \$1.32. This quote has a peak of profit around 7% as the matrix in **Figure 1** indicates. While this is just the visual of one quote from our model, this same idea applies to every quote we dealt with.

This model is a great starting point for dealing with new customers. With returning customers, there exists extra variables that are not considered in our model. However, with our model's flexibility (specifically in the third dimension of the matrix in **Figure 1**), a good, working model for returning customers would be simple enough to construct.

Another feature of this model is that it is incredibly flexible. The most obvious point of flexibility is the available third dimension of the matrix in **Figure 1**. Additionally, the variables of sub-setting the data could be altered (discussed previously in Preliminary Results). Lastly, the numerical values on axes of the matrix in **Figure 1** could be altered to get different results. We chose to run the model in its current standing because it yielded the highest accuracy in order to maximize profit. While maximizing profit is helpful, this model could also be used to optimize any other variable that might be important to Fastenal (for example, quoting success).

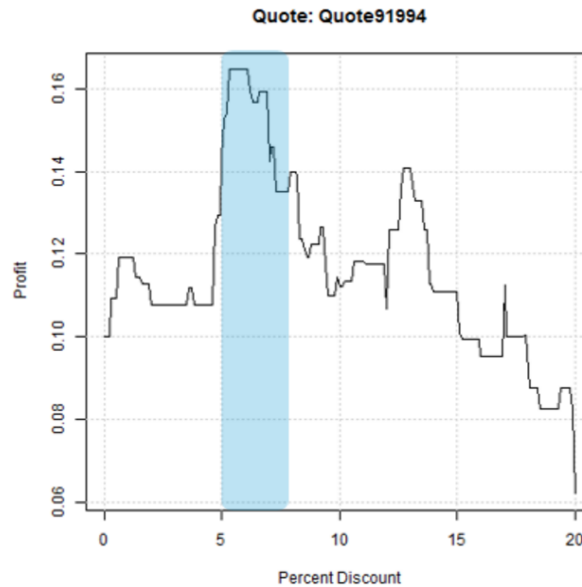


Figure 7. A plot of profit against discount percent against for one specific quote from the subset discussed in **Figure 1**.

CONCLUSION AND FUTURE WORK

Business Action Plan

To implement our model, Fastenal would take the following four steps...

1. Pick separation points; Here is where you would choose which variables you would like to use to separate the data. In our model we chose price range and quantity to be our separation variables.
2. Build a random forest model for each individual subset of data; Here is where we built the two-dimensional matrix to display the optimal discounts that correspond to the previously chosen separation points (see **Figure 1**).
3. Calculate the profit at the various percent discounts and plot them (see **Figure 6**).
4. Observe the resulting plots to determine the level of discount percent where profit peaks (see **Figures 6 and 7**).

Neural Networks

In the future, we would like this project to have a neural network as a solution. In the current conclusion, we can see, graphically, that the percent discount per region is a linear, choppy transition (reference **Figure 6**). With a neural network, you would be able to get similar results with a more fluid curve, which would allow for more accurate analysis on a deeper level. Recall as well that our Random Forest model had an accuracy of 73%. An additional benefit of utilizing a Neural Network is that we could improve on this accuracy. Note, however, that a Neural Network is very time-consuming and challenging, so given the time constraints on this project, the Random Forest model was our best option, yielding the highest possible accuracy.

A neural network is a computer program that essentially mimics the function of a human brain; at its core it is a classification algorithm. In your brain, there are *neurons* that connect in a massive string and they send signals throughout your body, instructions, reactions, etc. In a neuron, there are three main pieces, the dendrite, the axon, and the cell body. A neuron in a neural network, referred to from here on out as a *node*, has these same three basic body parts with similar functions: the input wire, the output wire, and the body, respectively. In a neural network, nodes can connect an output wire to a separate input wire to create layers upon layers of computations, just like in the brain. The in between layers are called *hidden layers*.

Note that for classification hypothesis', we will use $h_{\theta}(x) = g(z)$, where the function $g(z) = \frac{1}{1+e^{-z}}$, and $z = \theta^T x$. This $g(z)$ is called the sigmoid function, or the logistic function.

Overall, a neural network is analogous to a sophisticated logistic regression model. The aforementioned hidden layers utilize the *sigmoid function* to calculate the weight for each layer of internal calculations, and the output wire uses a basis hypothesis like in logistic regression to receive its output.

ACKNOWLEDGEMENTS

PIC Math is a program of the Mathematical Association of America (MAA) and the Society for Industrial and Applied Mathematics (SIAM). Support is provided by the National Science Foundation (NSF grant DMS-1345499).

We would like to thank Brian Keeling from Fastenal for his cooperation and collaboration with our group and for access to the data and knowledge of the company. Additionally, we would like to thank our faculty advisors, Drs. Song Chen and Chad Vidden, for their guidance throughout this semester.

REFERENCES

- Ng, Andrew, Prf. "Logistic Regression." Lecture. Accessed February 6, 2017.
http://www.holehouse.org/mlclass/06_Logistic_Regression.html
- Ng, Andrew, Prf. "Neural Network Representation." Lecture. Accessed March 7, 2017.
<https://www.coursera.org/learn/machine-learning/supplement/jtFHI/lecture-slides>
- Stergiou, Christos, and Dimitrios Siganos. "Neural Networks." Neural Networks. Accessed March 31, 2017.
https://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html
- "Decision tree." Wikipedia. April 07, 2017. Accessed April 19, 2017. https://en.wikipedia.org/wiki/Decision_tree