Predicting Brand Loyalty in Grocery Shoppers

Rafael Rivera-Soto
rivera43@stanford.edu

Daniel Gardner
dangard@stanford.edu

***Abstract*—The abstract goes here.**

I. Introduction

Brand loyalty is a chief concern for marketers of grocery products. Consumers will often buy the same brand of a household good for their entire lives. It is vital for product marketers to determine which consumers are especially brand loyal so that the marketers can target advertising and promotions toward them. For most grocery items, consumers are presented with the choice of purchasing a brand name product or a store brand product. Well-known national brands include Kelloggs cereal, Heinz ketchup, and Tide detergent, while store brands include Costcos Kirkland, Targets Market Pantry, and Walmarts Great Value. Despite a brand product often being more expensive than its store brand counterpart, many consumers prefer the reliability and known quality of the national brand. This brand preference or loyalty widely varies across different households and especially across different product types. We use machine learning techniques to predict a households preference for brand name products and determine what factors into that predilection. For a given product, we label households as brand loyal or not based on the brand ratio of their past purchases. Our algorithm takes as inputs the demographics of a household along with product-specific parameters. We then use logistic regression, support vector machines, and adaptive boosting to predict the households brand loyalty for that product. We also experiment with using the k-nearest neighbors algorithm to find similar product clusters and utilize these clusters as a feature in our final brand loyalty prediction.

II. Related Work

Brand loyalty has been extensively studied by economists and marketing researchers. Often their research focuses on a specific demographic group and observes whether this group has different behavior than the population at large. Bronnenberg et al. examines grocery purchases by consumers who are particularly well-informed about the homogeneity of certain brand and store-brand products and observes that they purchase the cheaper store brand more often than the average shopper. They are able to do this by matching employment information (choosing medical professionals and chefs) with domain-specific products (pain medication and baking goods). In a more recent paper, Bronnenberg et al. examines households that have lived in multiple regions of the United States in their lifetime and finds that the brand capital they have developed in the past makes them have different brand loyalties than similar consumers in their current region. The most common use of machine learning algorithms in consumer behavior research is to create market baskets, or products that are frequently purchased together. This is useful when trying to develop marketing campaigns to mesh multiple product categories, but it does little to explain which consumers might prefer to buy a brand of a particular product. The popular data science website kaggle.com has hosted multiple competitions to develop models related to grocery purchases. These include a problem posed by a marketing research firm to predict when shoppers will visit a store next and how much they will spend, as well as a problem from Walmart to classify different types of shopping trips. The previous approaches to brand loyalty have studied specific groups and how they behave differently, while machine learning in the grocery space has dealt mostly with clustering of substitute and complementary products. We will instead focus on what characteristics of the average consumer contributes to his or her brand purchasing choices.

III. Dataset and Features

We use the grocery purchases reported in the Nielsen Consumer Panel Dataset to determine brand loyalty, and then use the households demographic information found in the same dataset to make our model. We begin by examining particular products that have many occurrences in the dataset and also have a relatively balanced split in brand loyalty. We experimented with linear and logistic regression, finding much more success with the latter. It shows that the household size and whether a person is of hispanic or asian origin has an impact. We plan to expand our product basket as we perfect the method on our initial choices.

IV. Methods

A. Logistic Regression

Our initial efforts were concentrated upon whether consumers tend to buy more branded or non-branded products. This is a binary classification task for which we implemented a Logistic Regression model. A Logistic Regression squashes the output of the model in the range $y = \{0, 1\}$ using the sigmoid function (Eq. 1). The output values are then interpreted as the probabilities, thus any output greater than 0.5 is classified as belonging to the positive class and to the negative class otherwise (Eq. 2).

$$h_\theta(x) = g(\theta^T x) = \frac{1}{1 + e^{-\theta^T x}} \tag{1}$$

$$P(y=1|x;\theta) = h_\theta(x)$$
$$P(y=0|x;\theta) = 1 - h_\theta(x) \tag{2}$$

We based our initial predictions on products whose distributions between branded buyers and non-branded buyers was almost even. We labeled someone as being on the "branded-buyer" class if they're above the ratio cut-off value of 0.5, all other consumers were labeled as belonging to the"non-branded" class.

B. Support Vector Machines

The Support Vector Machine is a discriminative classifier that finds an optimal hyperplane with the largest margin between the classes. These models also allow us to implicitly map our input features into a high-dimensional feature space where the optimal hyper-plane might result in a better division between the classes. These feature mappings are called kernels. In our implementation, we used the Radial Basis Function (RBF) kernel (Eq. 3). SVM's use a particular choice of the loss function called the "Hinge Loss" (Eq. 4). To fit this model, we use an algorithm such as Stochastic Gradient Descent to adjusts our weights in such a way that the Hingle Loss is minimized. We used this model in the binary classification model described above and compare the effectiveness.

$$K(x, x') = exp(-\frac{||x - x'||^2}{2\sigma^2}) \tag{3}$$

$$\delta(z, y) = max\{0, 1 - yz\} \tag{4}$$

C. Boosting

The idea of this model is to take a weak learning algorithm, that is, any learning algorithm that does slightly better than random and transform it to a strong classifier that does much better than random. Roughly, this method begins by assigning every training example equal weight. It then receives a weak-hypothesis that does well according to the current weights. A weak hypothesis is an algorithm that takes as inputs some distribution (weights) p and outputs a weak learner that does better than random (Eq. 5). After evaluating the results after incorporating the new hypothesis, it re-weights the examples in such a way that incorrect classifications receive higher weights and correct classifications receives lower weights. In this way, boosting is able to create a strong hypothesis that generalizes well to new examples.

$$\sum_{i=1}^{m} p^{(i)} 1\{y^{(i)} \neq \phi_j(x^{(i)})\} \leq \frac{1}{2} - \gamma \tag{5}$$

D. Softmax Regression

After our initial efforts, we decided to extend our model to incorporate more than one class. In particular, we chose to divide consumers into three bins using cutoffs at 0.33 and at 0.66 (Eq. 6).

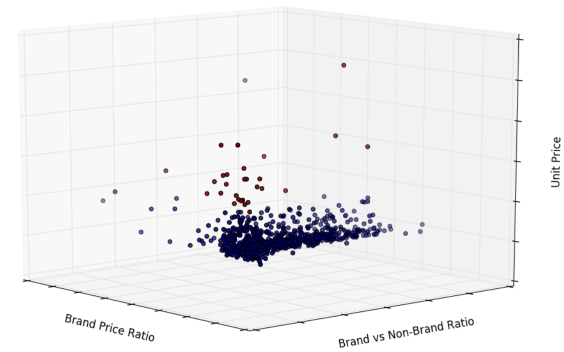

Fig. 1. Caption

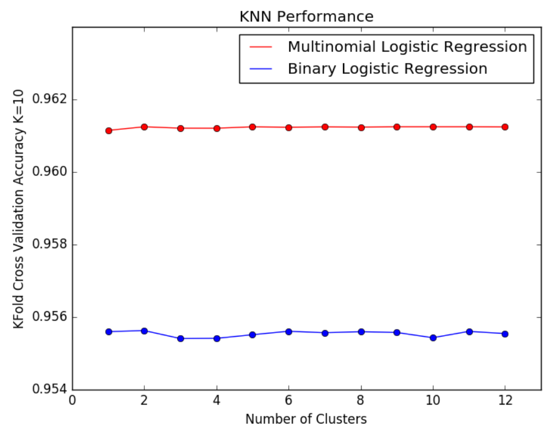

Fig. 2. Caption

$$C = \begin{cases} 0, \text{if } ratio < 0.33 \\ 1, \text{if } 0.33 \leq ratio < 0.66 \\ 2, \text{if } ratio \geq 0.66 \end{cases} \tag{6}$$

All of our previous models have been binary classifiers. In order to account for more classes we fit a Softmax Regression model. This is a generalization of the Logistic Regression models to multiple classes. In particular, Softmax Regression uses the Multinomial Distribution. The probability that our features take on a certain class is given by Eq. 7.

$$p(y = i|x;\theta) = \frac{e^{\eta_i}}{\sum_{j=1}^{k} e^{\theta_j^T x}} \tag{7}$$

V. Results

Results go here.

VI. Conclusion

The conclusion goes here.

References

[1] H. Kopka and P. W. Daly, *A Guide to LaTeX*, 3rd ed. Harlow, England: Addison-Wesley, 1999.