Leveraging the Power of Images in Predicting Product Return Rates

Daria Dzyabura*

Stern School of Business, New York University, New York, NY 10012

ddzyabur@stern.nyu.edu

Siham El Kihal

Frankfurt School of Finance and Management, 60322 Frankfurt am Main, Germany

s.elkihal@fs.de

Marat Ibragimov

New Economic School, 143026, Russia, Moscow

mmibragimov@nes.ru

July 2018

*Authors listed alphabetically

Leveraging the Power of Images in Predicting Product Return Rates

Abstract

Compared with offline retail channels, online channels are challenged by high product return rates. Processing and refurbishing returned items is costly, so a product's return rate has a large impact on its profitability. Therefore, in retailing decisions, such as product line optimization, it is important to predict the return rate prior to a product's launch.

Based on the sales and return data of a large apparel brand, we demonstrate that consumers' demand for products differs online and offline, such that some products attain higher market shares online than offline or vice versa. Furthermore, controlling for online demand, products with lower offline demand have higher return rates. This relationship is not simply due to missing information on the touch-and-feel attributes of the product online. We show that a machine learning algorithm can accurately predict the return rate using the product image, which consumers see online. We extract several visual features using image processing techniques, and predict return rates using a gradient boosted regression tree model. Incorporating image data improves the models' predictive accuracy by up to 37%. Firms can use this approach to better forecast an individual product's profitability and make effective merchandising and retailing decisions.

Keywords: machine learning, image processing, product returns

1. Introduction

As online retail has become more common, many firms now sell their products through both online and offline channels. The online channel has many advantages over the offline channel, including broader reach, lower travel costs for consumers, and saved costs of renting and operating retail space. However, a large cost for online retailers, relative to traditional brick and mortar retailers, is the cost of processing product returns. Nick Robertson, CEO of the UK's largest fashion retailer, ASOS, stated that a 1% drop in ASOS' return rate could increase the firm's bottom line by an impressive 30% (Thomasson 2013). Even Amazon struggles with managing product returns (Wall Street Journal 2018).

Product returns are vastly more common online than offline. According to our data, the average return rate per item is 53% online but only 3% offline, for the same set of products. Also, the cost of processing an individual return is higher online. Offline, the customer brings the item back to the store and a sales associate evaluates its condition and processes the return. However, online, the firm pays to ship the item to a return processing center, and an employee opens the box, evaluates the item, and issues a refund to the customer. Additionally, the firm accrues a refurbishing fee and many returned items are discarded or sent to outlet stores. The resulting costs range between $6 and $20 per returned item (The Economist 2013). As a result, retailers' profits from online channels are highly sensitive to product returns, and even small changes in return rates result in large profit improvements.

Several factors differentiate online purchase decisions from offline ones. First, when consumers make purchase decisions online, an important part of products' utility, such as the touch and feel product attributes, is only revealed upon physical inspection. If the product turns out to be poor on these attributes, the consumer may choose to return the product. In an offline environment, these attributes are revealed at the time of the purchase decision, so a product with poor on touch and feel attributes will simply not be purchased offline. Second, in a lab setting, Dzyabura et al. (2018) demonstrate that consumers place different weights on attributes online than they do offline, and therefore some products are more attractive when evaluated online than offline and vice versa. Finally, the consumer search literature has shown that consumers are more likely to search products with high variance, even if they have lower expected utility (Weitzman 1979). An online purchase decision made when consumers know that they may return the product may be treated more like a decision to search rather than a decision to purchase.

We obtain a large data set from a European apparel manufacturer and retailer, which includes over 1.8 million online and offline transactions over four years involving near 10,000 unique fashion products. We find large discrepancies between individual products' performance online and offline that are systematically related to return rates; the better a product sells offline, the fewer returns are made, and vice versa.

Next, we show that it is possible to predict return rates with high accuracy based on a product's online description: price, category, and product image. This finding is interesting for two reasons. First, it is very useful for managers; the return rate of a product significantly affects net product profitability, and the ability to forecast it before launching the product allows a firm to more efficiently make decisions regarding optimization of product lines and prices. Second, it suggests that returns happen not only because of touch and feel attributes that are only revealed to the consumer upon physical examination of the product. Rather, our model predicts the rate of product returns based solely on information available to the consumer online.

Methodologically, we demonstrate the power of product images in predicting products' profitability. We use a machine learning approach that combines several image feature extraction tools, including color histograms, Gabor filters, and convolutional neural networks, to quantify information from images. We use this information in a gradient boosted regression tree prediction model. Incorporating the visual features of a product considerably improves the accuracy with which the model can predict return rates by an impressive 37% compared to models using only non-image characteristics. Incorporating product images can be therefore applied to other key economic variables in addition to a product's return rate, such as clicks, demand, or even consumer search models.

Finally, we show that having data concerning the first two weeks of a product's offline performance helps further improve predictions of online return rates. Some omnichannel retailers may be able to launch products earlier in offline channels, and due to their ability to more accurately predict return rates, make informed decisions about launching products online.

2. Related Literature

This paper draws upon, and contributes to, three streams of marketing literature: leveraging unstructured data, product returns, and online and offline retail.

2.1 Leveraging Unstructured Data

We use methods from the computer vision and machine learning literature to predict return rates based on unstructured data, specifically, product images. An emerging stream of marketing research is dedicated to using unstructured data to better inform decision making.

Several marketing studies have used text data to generate managerial insights. For example, Timoshenko and Hauser (2018) use convolutional neural networks to extract customer needs from reviewed texts, and Liu and Toubia (2018) infer consumer preferences based on search query texts. Other studies use textual consumer-generated content to predict demand, sales, financial performance, and consumer engagement or determine market structure (e.g., Chevalier and Mayzlin 2006; Lee and Bradlow 2011; Archak et al. 2011; Onishi and Manchanda 2012; Tirunillai and Tellis 2012; Netzer et al. 2012). For example, Toubia and Netzer (2017) automatically rate the creativity of an idea based on unstructured text.

More recently, computer science and marketing research started using images to provide valuable information to marketers. A few studies use images to create recommendations regarding clothing styles and substitutes and more accurate personalized rankings (McAuley et al. 2015; Lynch et al. 2015; He and McAuley 2016). More recent research by Liu et al. (2018) uses images posted by consumers to measure how brands are portrayed in social media. Zhang and Luo (2018) use images and text from Yelp reviews to predict restaurants' survival. Zwebner et al. (2017) demonstrate that it is feasible to predict a person's name based on an image of their face.

We contribute to this literature by extracting products' features from product images and using them to improve forecasts about the product's performance prior to launch.

2.2 Products Returns

Marketing research on product returns focuses on either firms' return policies or customers' return behavior. Some research examines the impact of different policy components (e.g., fees, deadlines) on both purchases and returns (e.g., Shulman et al. 2011; see Janakiraman et al. 2016 for a review). In this paper, we consider aspects of return policies to be exogenous. European law mandates very lenient return policies, and most US retailers now also offer such policies to stay competitive.

Another stream of research focuses on understanding and managing the product return behavior of individual customers. For example, Nasr-Bechwati and Schreiner (2005) investigate the effects of pre-purchase choices on customers' probability of returning a product. El Kihal, Erdem, and Schulze (2018) show that customer return rates increase over the course of their relationship with online retailers.

Our research differs from both streams of literature because we focus on individual products and their features to forecast return rates prior to a product's launch.

2.3 Online and Offline Retail

Much of the literature on the interplay between online and offline retail has examined spillover or cannibalization between channels using total sales or other aggregate metrics. For example, Pauwels and Neslin (2015) demonstrate that introducing a physical store in a geographic region cannibalizes catalogue sales but has less impact on internet sales and results a net increase in overall sales. Wang and Goldfarb (2017) find that the presence of a physical store increases customer acquisition in online channels. Ansari et al. (2008) investigate customer channel migration and the short- and long-term effects of channel usage on channel selection and demand. In a lab experiment, Dzyabura et al. (2018) analyze consumers' product evaluation in the online and offline channel and show that there can be large differences between how consumers evaluate the same individual product online versus offline.

We build on this literature by studying the demand for individual products in online and offline channels and linking the discrepancy between them to product returns.

3. Online and Offline Demand and Product Returns

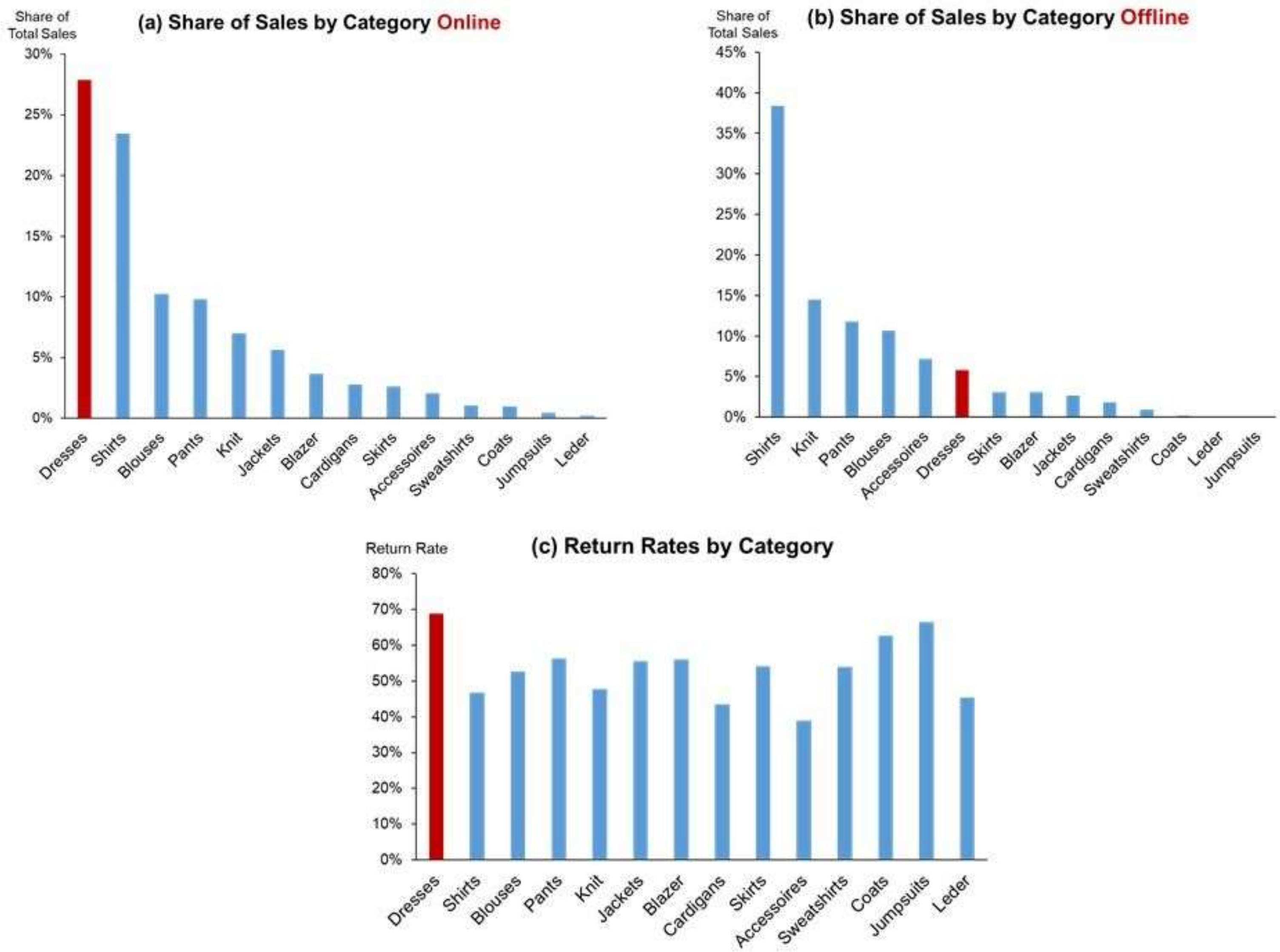
We obtain data from a large European apparel manufacturer and retailer. The company has a large online operation, which accounts for 22% of its sales, and a network of 39 retail locations in Germany. The company specializes in women's apparel. The data contains 1,838,851 transactions, including sales and returns, that occurred through both online and offline channels during the observation period (2013–2016). We exclude non-fashion items such as perfume or gift cards. We also truncate the purchase data by date from above, deleting purchases for which the return deadline fell outside our observation period. We also exclude transactions with nonzero delivery costs to remove noise due to delivery fees and their impact on return decisions. Finally, we exclude items that were sold fewer than 20 times in the online channel. The resulting data set consists of 9,307 distinct products from 15 different product categories, as categorized by the retailer.

The retailer has a lenient return policy: customers can return any purchased product for any reason within 14 days. While the average product return rate for products purchased through the online channel is 53% (ranging from 0–94%), it is only 3% for products purchased through the offline channel. The return rate, RR_i, of product i is computed as the ratio of the number of returned items to the number of purchased items online. The difference in return rates between the two channels may arise from differences in consumers' decision making in the two channels.

When purchasing online, consumers make two evaluations of the product, first online, to determine whether to purchase, then offline, to determine whether to keep or return. We expect that a product that sells well online and less well offline is more likely to be returned if purchased online. Moreover, we know from consumer search literature that consumers are more likely to search products with high variance, even if they have lower expected utility (Weitzman 1979). In the presence of return policies, consumers may treat an online purchase decision like a decision to search rather than a decision to purchase (Anderson et al. 2009). This, again, would lead to a systematic shift in demand online relative to offline, as well as return rates.

We first present some model-free evidence that items that sell well online and poorly offline have high online return rates. This relationship is particularly evident at the category level. Figures 1a and 1b show the share of sales in each category through online and offline channels, respectively. The best-selling category online is dresses, which account for 27% of online sales. Offline, dresses comprise the sixth-best-selling category, accounting for only 8% of offline sales. At the same time, dresses have the highest return rate of all categories: 72% on average. Figure 1c provides the average return rate of products by category.

Figure 1: Online and Offline Performance and Return Rates by Category



We see a similar pattern at the product level. Specifically, we estimate the following model:

$$RR_i = \beta_0 + \beta_1 shareOn_i + \beta_2 shareOff_i \ , \qquad (1)$$

where $shareOn_i$, $shareOff_i$, and RR_i are product i's online and offline market shares relative to category (e.g., shirts, dresses) and return rate (continuous between 0 and 1), respectively. The estimates match our expectations. In addition, the coefficient of online market share is positive—β_1=0.039 ($p<10^{-15}$)—capturing the relationship with products' popularity online. The more an item sells online, the more often it is returned. However, products that sell better offline are returned less often (β_2 =-0.026, $p<10^{-27}$), consistent with our hypothesis that the discrepancy between online and offline consumer decision processes corresponds to high return rates.

This systematic relationship between product online and offline demand, and online returns is consistent with our understanding of the differences between the two channels. Upon receiving the product and examining it physically, the consumer gains additional information, based on which she makes a decision to keep or return the product. Offline, this information is taken into account when the purchase decision is made.

However, as we will see in the next section, it is possible to accurately predict a product's return rate based on the product's price, category, and image (that is, based solely on information available to consumers when purchasing online). This means that the decision to purchase and then return a product is not driven solely by product information that is not available online, such as comfort and fit. Rather, there are certain product features that are available to consumers prior to an online purchase that make the product more likely to be purchased and then returned. These may be features that consumers systematically value more online than offline.

.

4. Model, Estimation, and Prediction

In order to make effective merchandizing decisions regarding, for example, the design of a product line, a manager would need to know an item's return rate prior to launching the product, and thus before observing online or offline demand. We now present a machine learning model for predicting the return rate of a new product. For each product, i, we observe the following characteristics:

- Price (p_i), a continuous variable in Euro.
- Category (c_i), a dummy variable that takes one of fifteen values, which correspond to the apparel categories defined by the retailer.
- Color labels (cl_i), a dummy variable that takes one of fourteen values, which correspond to the labelled color of the product, as categorized by the retailer. For example, the dress in Figure 2a has the color label "Black/Red," and the pants in Figure 2b have the color label "Green."
- Product image, which appears on the retailer's website. Products are pictured against a white background (never on a model). We use the images depicting the front of the products. Figure 2 presents two examples of such images. Features extracted from the product images, which we describe below, are denoted by I_i.

Figure 2: Examples of Images of Two Products

The goal of the prediction task is to obtain a mapping from the above product information to the online return rate:

$$RR_i = f(p_i, c_i, I_i)\ , \tag{2}$$

We face two methodological challenges. The first is quantitatively capturing the information contained in the product image through vector I_i. The second is estimating the function $f(\cdot)$. We describe our approach to these challenges below.

4.1 Image Feature Extraction

Raw image data cannot be incorporated into the model because such data do not contain quantitative information. Therefore, to use an image in a model, we must first quantify the information in the image. Such quantitative information is known in the economics and computer science literature as *image features*. In this section, we describe the approaches we used to extract the image features of the clothing images.

Color histograms.

Images are stored as two-dimensional matrices of elements called pixels. Each pixel contains an array of numbers that describe the particular point of the image (e.g., color, brightness). The most common and simple representation is RGB, in which each pixel is captured by a three-dimensional array corresponding to red, green, and blue channels, each containing an integer between 0 and 255. For example, a standard image with 600 x 600 = 360K pixels is represented by over 1 million numbers. The standard approach to reducing dimensionality is to use a histogram. The histogram is usually constructed using three dimensional bins. The 256 x 256 x 256 dimensional RGB space is split into cells of shape b_1 x b_2 x b_3, and each pixel lies in one of the bins. We then calculate the number of pixels in each bin, resulting in a feature vector of $\frac{256}{b_1} \cdot \frac{256}{b_2} \cdot \frac{256}{b_3}$ values.

Including this feature in our model enables us to capture the systematic dependence of a product's return rate on its color.

Gabor features.

A Gabor filter captures the pattern and texture characteristics of an image. It isolates periodicity in the image by modeling the image as a set of superimposing sinusoidal waves of different frequencies. It is effective for processing images with a quasi-periodic structure; for example, it can identify whether a clothing item has a striped or checkered pattern. The key parameters are the range of wave frequencies and wave direction. Similar to Liu et al. (2018), we implement Gabor features by applying the following transformation to the image:

$$g(x, y; \lambda, \theta, \sigma) = e^{\left(-\frac{\widetilde{x^2}+\widetilde{y^2}}{2\sigma^2}\right)} \cos\left(2\pi \frac{\tilde{x}}{\lambda}\right) \quad , \tag{3}$$

$$\begin{pmatrix} \tilde{x} \\ \tilde{y} \end{pmatrix} = \begin{pmatrix} \cos\theta & sin\theta \\ -sin\theta & cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

where x and y are the pixel coordinates of the original image; λ and θ are frequency and direction parameters, respectively; and σ is the size of the Gaussian smoothing filter. For example, a horizontal striped item corresponds to one value of θ, and a checkered item corresponds to two different values of θ since the lines go in two directions.

We calculate the output after applying a discretized version of the filter in Equation 3:

$$I^{output}(i,j;\lambda,\theta,\sigma) \propto \sum_{a=1}^{N}\sum_{b=1}^{N} I^{input}(a,b;\lambda,\theta,\sigma)g(i-a,j-b). \quad (4)$$

Each filter has a strong response to images with patterns close to the frequency and orientation parameters of the filter. We apply B filters with different combinations of the parameters $(\lambda,\theta,\sigma)_b$ to an image and construct the feature vector, $(\mu_1, s_1^2, \ldots, \mu_B, s_B^2)$, by aggregating the filters over the output image as follows:

$$\begin{aligned} \mu_b &= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} I^{output}(i,j;(\lambda,\theta,\sigma)_b) \\ s_b^2 &= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(I^{output}(i,j;(\lambda,\theta,\sigma)_b)-\mu_b\right)^2 \end{aligned} . \quad (5)$$

Deep-learned features.

Deep learning has been successfully used in many unstructured data applications. A distinguishing property of deep learning algorithms is that they are based on feature/representation learning. This means that they not only make predictions based on a set of features but also learn the optimal features for a particular prediction problem from unstructured data. The most popular deep learning models for image processing are convolutional neural networks (CNNs). Starting from the raw image pixels in the first layer, each layer of the neural network transforms the features of the previous layer to obtain a new set of features. This way, through a series of simple nonlinear transformations, a neural network is able to capture a highly complex nonlinear transformation, which can approximate the function that maps an image to a target variable. Each transformation convolves the input with a matrix. Consider, for example, an l_1 x l_2 matrix with elements $\omega_{a,b}$. The output of the transformation would be as follows:

$$output_{i,j} = \sum_{a=-l_1/2}^{l_1/2} \sum_{b=-l_2/2}^{l_2/2} \omega_{a,b} \cdot input_{i+a,\, j+b} \,, \tag{6}$$

where the matrix elements $\omega_{a,b}$ are unknown parameters that have to be estimated. A typical CNN consists of several “layers” of transformations, usually 10 or more, with each “layer” containing between 60 and 500 different transformations.

Our data set is not large enough to train our own deep learning model to extract optimal features for prediction. We therefore use one of the layers of a pre-trained CNN as a feature vector. This is a common approach in economics and marketing research. We also used the VGG-19 network (Simonyan and Zisserman 2014), which was the winner of the ImageNet Large Scale Visual Recognition Challenge. This network was trained on the ImageNet data set, which contains 1.3 million images of about 1,000 classes of objects. The VGG-19 network consists of nineteen layers, the last of which is the output layer. We use the 17^{th}, or second-to-last, layer of the network, which contains 4,096 features.

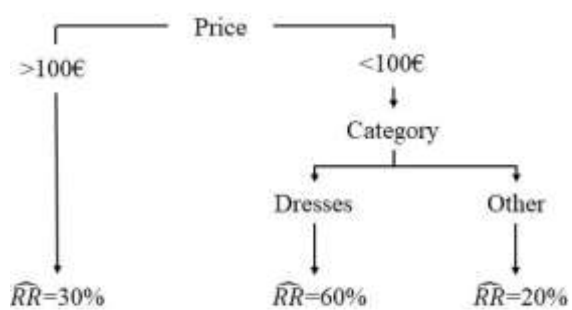
Of course, the ImageNet data set consists of images that are different from our clothing images, and the target prediction task is different: the VGG-19 network is optimized for object recognition and detection, while we are trying to predict return rates. This means that the optimal predictive features extracted from the data set may not be optimal for our specific task.

4.2 Prediction Model

We next consider the function $f(X_i)$, where X_i captures product j’s characteristics, p_i,c_i, and I_i as described above and maps them onto the product return rate. We include the product image features described above and then compare the predictive accuracy of the resulting models.

We train a gradient boosted regression tree (GBRT) model, which is an advanced version of a regression tree-based model. A regression tree model partitions an input space for explanatory variables into multiple regions and predicts a single value for all points in a region of this space. This prediction rule is formed like a tree; at each node, the algorithm divides the space into two parts based on one explanatory variable. Figure 3 shows an example of a simple regression tree model with three leaves, or output regions, on our data. This example tree first determines whether the price of a product is less than or equal to 100€. If it is, it predicts a return rate of 30%. If not, it then asks if the category is "dresses." If it is, the model predicts a return rate of 60%. Otherwise, it predicts a return rate of 20%.

Figure 3: Example of a Regression Tree



The advantage of tree-based models is their ability to capture higher-order interactions between features by making different predictions for different regions of the feature space rather than explicitly writing them in the model. It is important to set a parameter for regression tree models to limit the depth of a tree. If there is no rule preventing the addition of more splits, the model could minimize errors by splitting the input space into enough regions that each region only contains one data point. One common improvement of a single tree is a random forest model, which consists of a set of regression trees, each trained on a bootstrapped subsample of the original data. The predicted value for a given X_i is the average of the predictions of each tree in the ensemble.

The gradient boosted regression tree (GBRT) model is a further advance of the random forest models. Boosting is similar to ensembles in the sense that the model prediction is a combination of several simpler models. While ensembles separately train individual models, such as trees, on different subsamples of data, boosting add trees greedily, with each tree trained on the residual of the current model. The best existing boosted tree algorithm, XGBoost, was introduced by Chen and Guestrin (2016) and is the one we use in this paper. It outperformed random forest regression in Kaggle machine learning competitions. Boosted trees have been shown to work well in marketing applications such as predicting clicks (Rafieian and Yoganarasimhan 2018) or customer churn (Neslin et al. 2006).

4.3 Results

We train the models on 75% of products (K_{train}), and we report the R-squared predictive accuracy of the remaining 25% of products (K_{test}):

$$R^2_{model} = 1 - \frac{\sum_{i \in K_{test}} \left(RR_i - \widehat{RR}_i^{model}\right)^2}{\sum_{i \in K_{test}} \left(RR_i - \widehat{RR}_i^{random}\right)^2}, \tag{7}$$

where $\widehat{RR}_i^{model}$ is the predicted return rate of product *i* according to the *model* and the *random* model predicts the average return rate for all products. We create 100 random partitions of the data set into training and test subsets, compute the holdout R-squared metric for each, and report the average and standard deviations of the resulting R-squares. The results of all models are reported in Table 1.

Table 1: Holdout Predictive Accuracy of the Return Rates for Different Sets of Product Features.

Model	Product Features Included	Image Features Included	Predictive Accuracy (out of sample R^2)	Improvement in Predictive Accuracy
Model 1 (Benchmark)	p_i, c_i	None	33.88 (2.93)	-
Model 2	p_i, c_i, Color labels (cl_i)	None	35.51 (2.79)	4.81%
Model 3	p_i, c_i	RGB	44.12 (2.44)	30.22%
Model 4	p_i, c_i	Gabor	42.16 (2.57)	24.44%
Model 5	p_i, c_i	Deep-learned	44.64 (2.45)	31.76%
Model 6	p_i, c_i	RGB+ Deep-learned	46.26 (2.49)	36.54%
Model 7	p_i, c_i	RGB + Gabor	44.74 (2.47)	32.05%
Model 8	p_i, c_i	Gabor+ Deep-learned	45.32 (2.51)	33.77%
Model 9	p_i, c_i	RGB+ Gabor+ Deep-learned	46.38 (2.44)	36.89 %

Notes: (1) We create 100 random partitions of the data set into training and test subsets, compute the holdout R-squared metric for each, and report the average and standard deviations of the resulting R-squares (in parentheses). (2) Improvements in predictive accuracy are relative to the benchmark model (Model 1).

Table 1 compares the predictive accuracy of the model, when trained with different image features, and product features quantified by the retailer. All nine models use products' price and category information. The benchmark model (Model 1) uses no additional information. Model 2 adds the retailer's color labels, and we can see that the prediction improves. Next, Models 3, 4, and 5 add the RGB, Gabor, and deep-learned features, respectively, one at a time. All three models greatly outperform Models 1 and 2, which do not use the product image for prediction (i.e., the predictive accuracy improves by around 30%). This finding highlights the importance of using product images to model consumer behavior and demonstrates the potential of image processing methods for econometric modeling. In addition, the models trained on RGB and deep-learned features perform similarly well. This may be because the entire prediction is captured by color, and anything additional, captured by the CNN is not relevant for predicting product return rates. To test this, we combine the RGB and deep-learned features and find that the resulting model (Model 6) outperforms models that include these features one at a time. This means that the RGB and deep-learned features contain different relevant information, which is not captured by the other feature. The neural network captures the non-color properties of the product but does not capture color in as much detail as RGB. Surprisingly, the Gabor features, which captures patterns, did not improve performance beyond the RGB or deep-learned features (Models 7 and 8). Although pattern is a seemingly important attribute of clothing, it did not improve predictions of the product return rate. Model 9, which adds the Gabor features to RGB and deep-learned features, does not improve predictive accuracy beyond Model 6.

Being able to predict products' return rate prior to launch will allow managers to make better-informed retailing decisions. Since a firm's profit in the online channel is highly sensitive to a product's return rate, an accurate estimate of the return rate for an individual item is key for profitable assortment planning.

Conceptually, it is interesting that the online description of a product is enough to obtain a good prediction of the return rate. This means that consumers return products not only because of touch and feel attributes, which can only be evaluated offline: after all, our prediction is based only on information available to the consumer prior to purchase. One possible explanation is that consumers systematically over- and under-weight certain product attributes online relative to offline. A deeper behavioral study would be necessary to pin down the mechanism.

Recall that Section 3 stated that a product's performance in an offline channel is related to the return rate. This suggests that offline demand may have predictive ability as well and can help to improve predictions even further. We explore this next.

Observing offline channel first.

While our task is to predict the return rate of a product prior to launch, some retailers might be able to launch products first in the offline channel and decide whether to launch them in the online channel based on their offline performance over a short period. To investigate the benefit of observing offline sales for a short period, we include in our models data concerning the first two weeks of a product's offline sales. Adding the first two weeks' offline sales data to the best performing model, results in predictive accuracy of 47.64 (2.36), which is higher than the best performing model in Table 1. Still, while there is an improvement in the prediction of return rates, it is not large.

4.4 Robustness Tests

To test the robustness of our findings, we estimated the following models (available from the authors upon request):

- **Sales threshold variation:** The main models include products that were sold online at least 20 times. Changing the threshold to 10, 30, and 50 did not change our results.
- **Transactions with nonzero delivery costs:** The main analysis excludes transactions with nonzero delivery costs to remove noise due to different delivery fees and their impact on return decisions. Including these, however, did not change the results.
- **Color histogram variation:** We used the HSV (hue, saturation, value) color histogram instead of the RGB histogram to capture image color. This resulted in similar predictive accuracy compared to the RGB histogram.

- **Reducing dimensionality of image features:** We also estimated models using versions of the image features with reduced dimensionality. We used PCA to reduce the dimensionality of all features, extracted via a Gabor filter, HSV and RGB color histograms, and CNN-based features. Still, the image features extracted via color histograms, Gabor filters and CNN-based features perform better than the baseline models.

5. Summary and Discussion

Product returns generate huge costs for online retailers are growing as online retail grows and as consumers get used to purchasing and returning items through online channels. It is important to manage this problem preemptively, by designing and marketing products that will not have high return rates in the online channel. In this paper, we show how a firm can forecast the return rate for completely new products based on product images. A firm can further improve this forecast by testing the product out offline first. Working with a large apparel retailer allowed us to observe products' online and offline demand and return rates for a large set of products over a long period of time.

First, we find that products that sell well online but less well offline have high return rates. Second, we show that including product information contained in the product image improves considerably the model. We hope that future research will further leverage product images to improve models in a broad range of contexts, not only fashion but also hotels, groceries, furniture, real estate, etc. We show that it is possible to predict return rates with high accuracy using only product information that is available during consumers' online purchases, suggesting that returns do not happen only because of the touch and feel attributes. Third, methodologically, we contribute to the rapidly growing literature on machine learning in marketing. We use the extreme gradient boosted regression tree model on image data and show how different combinations of image features outperform models that do not include visual product features.

6. References

Anderson ET, Hansen K, Simester D (2009) The option value of returns: Theory and empirical evidence. *Marketing Science. 28*(3): 405–423.

Ansari A, Mela CF, Neslin SA (2008) Customer channel migration. *Journal of Marketing Research*. *45*(1):60–76.

Archak N, Ghose A, Ipeirotis, PG (2011) Deriving the pricing power of product features by mining consumer reviews. *Management science. 57*(8):1485–1509.

Chevalier JA, Mayzlin D (2006) The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, *43*(3):345–354.

Chen T, Guestrin C (2016) Xgboost: A scalable tree boosting system. *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. ACM.

Dzyabura D, Jagabathula S, Muller E (2018) Accounting for discrepancies between online and offline shopping behavior. *Marketing Science*. Forthcoming.

El Kihal S, Erdem T, Schulze C (2018) Is how you start how you finish? Customer return rate evolution over time. Working Paper.

Janakiraman N, Syrdal HA, Freling R (2016) The effect of return policy leniency on consumer purchase and return decisions: A meta-analytic Review. *Journal of Retailing*. 92(2):226–235.

He R, McAuley J (2016) VBPR: Visual bayesian personalized ranking from implicit feedback. *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*.

Luo L (2011) Product line design for consumer durables: An integrated marketing and engineering approach. *Journal of Marketing Research.* 48(1):128–139.

Lee TY, Bradlow ET (2011) Automated marketing research using online customer reviews. *Journal of Marketing Research.* 48(5):881–894.

Liu J, Toubia O (2018) A semantic approach for estimating consumer content preferences from online search queries. *Marketing Science.* Forthcoming.

Liu L, Dzyabura D, Mizik NV (2018) Visual listening in: Extracting brand image portrayed on social media. Working Paper.

Lynch C, Aryafar K, Attenberg J (2015) Images don't lie: Transferring deep visual semantic features to large-scale multimodal learning to rank. arXiv preprint arXiv:1511.06746.

McAuley J, Targett C, Shi Q, van den Hengel A (2015) Image-based recommendations on styles and substitutes. arXiv preprint arXiv:1506.04757

Nasr-Bechwati N, Schneier Siegal W (2005) The Impact of the prechoice process on product returns. *Journal of Marketing Research.* 42 (3):358–67.

Netzer O, Feldman R, Goldenberg J, Fresko M (2012) Mine your own business: Market-structure surveillance through text mining. *Marketing Science.* 31(3):521–543.

Onishi H, Manchanda P (2012) Marketing activity, blogging and sales. *International Journal of Research in Marketing.* 29(3):221–234.

Pauwels K, Neslin S (2015) Building with bricks and mortar: The revenue impact of opening physical stores in a multichannel environment. *Journal of Retailing.* 91(2):182–197.

Rafieian O, Yoganarasimhan H (2018) Targeting and privacy in mobile advertising. Working Paper.

Neslin S, Gupta S, Kamakura W, Lu J, Mason C (2006) Defection detection: Measuring and understanding the predictive accuracy of customer churn models. *Journal of Marketing Research* 43(2):204–211.

Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. *Proceedings of International Conference on Learning Representations*.
The Economist (2013) Return to Santa – E-commerce firms have a hard core of costly, impossible-to-please customers. *The Economist*. (December 21), https://www.economist.com/news/business/21591874-e-commerce-firms-have-hard-core-costly-impossible-please-customers-return-santa
Thomasson E (2013) Online retailers go Hi-Tech to size up shoppers and cut returns. Reuters (October 2), http://www.reuters.com/article/net-us-retail-online-returns-idUSBRE98Q0GS20131002
Timoshenko A, Hauser JR (2018) Identifying customer needs from user-generated content. *Marketing Science*. Forthcoming.
Tirunillai S, Tellis GJ (2012) Does chatter really matter? Dynamics of user-generated content and stock performance. *Marketing Science.* 31(2):198–215.
Toubia O, Netzer O (2017) Idea generation, creativity, and prototypicality. *Marketing Science.* 36(1):1–20.
Wall Street Journal (2018) Banned from Amazon: The shoppers who make too many returns. WSJ (June 11), https://www.wsj.com/articles/banned-from-amazon-the-shoppers-who-make-too-many-returns-1526981401
Wang K, Goldfarb A (2017) Can offline stores drive online sales? *Journal of Marketing Research. 54* (5):706–719.
Weitzman,ML (1979) Optimal search for the best alternative. *Econometrica.* 47(3):641–654.
Zhang M, Luo L (2018) Can user generated content predict restaurant survival: Deep learning of Yelp photos and reviews. Working Paper.
Zwebner Y, Rosenfeld N, Sellier A, Goldenberg J, Mayo R (2017) We look like our names: The manifestation of name stereotypes in facial appearance. *Journal of Personality and Social Psychology.* 112 (4): 527–554.