Individual Final Assignment – Analysis of Wish.com "Summer"-Products: How to enter new Markets based on Product Likeability

Lucas Hagelstein (559020)

04/01/2021

Dr. Michel van de Velden

Dr. Anastasija Tetereva

Dr. Carlo Cavicchia

Seminar in Data Science for Marketing Analytics

MSc in Data Science and Marketing Analytics

Erasmus School of Economic

Table of Contents

1. Introduction

The customer is an established Chinese clothing manufacturer who plans to expand his product portfolio by a summer-line. Further, instead of selling this line Business to Business (B2B), the idea is to implement a new sales channel via wish.com to gain some spill-over effects: First, a potentially new target group might be discovered by entering the U.S. market. Second, they expect to be less dependent on only a few business customers by selling Business to Customer (B2C) as well.

According to the Forbes Magazine (2019), Wish.com is an ultra-bargain e-commerce B2C shopping platform whose customers are part of the American working-class with only little money available. Therefore, products listed on wish.com are no-name products mostly shipped by Chinese merchants but the platform yields high awareness and is the third biggest American marketplace in terms of sales. As most of the items offered on wish.com are very cheap and of low quality the customer satisfaction (in terms of product rating) is expected to be the main driver to successfully offer goods on the B2C-platform wish.com in the long-run.

Kaushik et al. (2018) have shown that negative customer reviews in the early stage have largely negative effects on the upcoming product performance, thus it is important to understand the mechanisms of e-commerce platforms first, before offering any products. Additionally, Bhatti et al. (2020) determine a global up trending e-commerce market in response to the COVID-19 outbreak, thus it might be of increasing importance to position well in the market to be competitive in upcoming periods.

Therefore, the research question is two-folded:

> *First, is it possible to create distinctive subcategories of summer products on wish.com by using a list of keywords.*
>
> *Second, to what extent is it possible to predict a product's likeability and subsequently determine the drivers of the prediction outcome.*

The categorisation is based on the idea that products which contain the same search tags appear in the same two-dimensional subspace. This will be done by applying a combination of correspondence analysis and a non-hierarchical partitioning algorithm. This method is called iterative factor clustering on binary data (i-FCB) proposed by Iodice D' Enza, and Palumbo (2013) and is the preferred method because it calculates the Euclidean Distance based on statistical units which are described by the binary variables.

The prediction task is solved by applying a Decision Tree method, Bagging and Random Forest which are especially suitable in the case of tabular data with a mixture of numerical, categorical and binary features. In the ongoing analysis it is demanded to understand the decisions made by the model which is done easily using Random Forest compared to other more complex models. Finally, Random Forest cannot overfit the data by ensuring uncorrelated trees.

2. Data

The investigated dataset “Sales of summer clothes in E-commerce Wish” was published by Jeffrey Mvutu Mabilama and can be accessed via kaggle.com. The initial dataset has 43 variables and 1,573 observations which are products listed in the “summer”-rubric on wish.com, collected in August 2020. The cleaning procedure required to select variables that contain information content as well as the feature engineering of potentially useful variables. For example, the variable “product_color” contains 102 colors whereas many colors are either duplicates with different spelling or variations from a main color. The subsequent cleaning reduced the variation from 102 colors to 14 levels without dropping any observations. The variable “product_variation_size” was reduced from 102 to 67 levels. As this report focusses only on clothing, observations that provide unknown clothing sizes as shoe sizes and other “wrong” inputs have been dropped. The number of observations shrinked from 1,573 to 1,505 and the levels of this variable from 67 to 14. Filtering reliable ratings is done by setting a minimum threshold of 15 ratings per product, thus a dataset with 1,211 observations is obtained. Finally, a column with a “taglist” is provided which provides key terms of how the merchant labels a certain product. The tags are extracted and replaced with a dummy variable per tag.

The average product in the summer rubric has 370 customer reviews, is sold 1,000 times at the price of $8 dollars exclusive shipping costs of $2.40 dollars (shipped to 40 countries) and is offered at a discount of 30%. Only limited products have badges for fast shipping (1%), local products (2%) and product quality (9%) but 30% have an urgency banner which indicates a limited stock. Further, merchants use ad boosts for 43% of summer products. Interestingly, only 16% of merchants even have a profile picture at an average customer rating count of 10,000.

Products and Merchants, both, are rated by customers on a scale ranging from 1 to 5 whereas summer products yield an average rating of 3.8. Merchants, in contrast, have average ratings of 4.0. This indicates summer products to be rated below the average but as information is only available for summer products nothing specific can be said about the discrepancy. The average distribution of the unique rating options is 48% for five stars, 20% for four stars, 15% for three stars, 7% for two stars and 11% for one star. 96.5% of summer products are offered by Chinese merchants whereas American merchants are the second most frequent seller on the wish.com summer product market with a fraction of only 2% (see Table 1).

Table 1 Descriptive Statistics

N = 1,211	Price (in $)	Shipping Price (in $)	#Sales	#Ratings
Mean	8.5	2.4	5,400	1,100
SD	3.6	0.94	10,000	2,200
Median	8.0	2.0	1,000	370
[Min; Max]	[1.0; 26]	[1.0; 7.0]	[50; 100,000]	[15; 21,000]

3. Methodology

First, the iterative Factor Clustering on Binary data (i-FCB). Van de Velden et al. (2017) determine i-FCB as non-symmetric correspondence analysis (NSCA). It is used as dimension reduction technique to create clusters from many binary variables. The i-FCB method is two-fold: First, obtain a NSCA solution for the cross-tabulation of the cluster allocation with binary variables. Second, allocate the obtained units in a spanned space whereby the sum of variances between clusters is maximized. Second, Random Forest (RF) is used to predict the likeability of summer products offered on wish.com. RF is understood as an ensemble learning technique that combines several decision trees which are grown by bootstrapped data samples combined with randomized feature selection. Decision trees cast a vote, thus the majority vote wins the final prediction.

3.1. i-FCB Method

According to Iodice D' Enza, and Palumbo (2013), the dependent variable X determines if observation i is in the k-th cluster $X_k (k = 1, \dots, K)$ whereas the categorical variables $Z_j (j = 1, \dots , p)$ are binary $\{z, \bar{z}\}$ which indicates presence or absence. These events are given in the two-way cross-classification table $\boldsymbol{F}$ with elements f_{kh}. The count of category k (row margin of $\boldsymbol{F}$) is written f_{k+} and the count of category h (column margin of $\boldsymbol{F}$) f_{+h}. The maximization problem can be written as

$$\sum_{j=1}^{p} \left(G(X) - G(X|Z_j) \right) = \sum_{j=1}^{p} \left(\frac{1}{n} \sum_{k=1}^{K} \sum_{h}^{2} \frac{f_{kh}^2}{f_{+h}} - \frac{1}{n} \sum_{k=1}^{K} \frac{f_{k+}^2}{n} \right),$$

where G(X) is the Gini-index that describes the variance of X. $G(X|Z_j)$ denotes the average of the Gini-indexes for $G(X|z_j)$ and $G(X|\overline{z_j})$ of the categorical variable j. Intuitively, the equation describes the sum of variances explained by each categorical variable. As it is not feasible to solve this equation an iterative approach is explained using the algebraic formalization. Therefore, the re-written first-order condition (with respect to **U**) is written as

$$\mathbf{D} = \mathbf{\Lambda U} = \frac{1}{n} \left[\mathbf{X}^T \mathbf{Z} (\Delta)^{-1} \mathbf{Z}^T \mathbf{X} - \frac{p}{n} (\mathbf{X}^T \mathbf{1} \mathbf{1}^T \mathbf{X}) \right] \mathbf{U},$$

where the two-way cross-table $\boldsymbol{F}$ is determined by $\mathbf{X}^T \mathbf{Z}$ with $\mathbf{Z}$ being equal to the $(n\ x\ 2p)$ block matrix consisting of $\mathbf{Z}_j$ $(j = 1, \dots, p)$ which is a $n\ x\ 2$ matrix that indicates if the j-th of p features is present or not (one or zero). Hence, $\Delta = diag(\mathbf{Z}^T \mathbf{Z})$ to be a diagonal matrix with ones on its diagonal entries whereas $\mathbf{1}$ is a vector of ones. The matrix $\mathbf{X}$ $(n\ x\ K)$ allocates each observation into the k-th cluster. This equation can be referred to as an eigenanalysis problem with eigenvalues defined as diagonalmatrix $\mathbf{\Lambda}$ and the eigenvectors in matrix **U**.

To obtain the clusters of each observation an eigen-decomposition is conducted on matrix **D**. The decomposed matrix is defined as

$$\mathbf{\Psi} = \left(\mathbf{Z} (\Delta)^{-1} \mathbf{Z}^T - \frac{p}{n} \mathbf{1} \mathbf{1}^T \right) \mathbf{X U \Lambda}^{\frac{1}{2}},$$

where $\mathbf{\Psi}$ represents each statistical unit in the space by the columns of **U** (eigenvectors). Next, the randomly created matrix **X** is updated by the Euclidean Distance to conduct the partitioning cluster analysis using the information of matrices **Ψ** and **X** which results in the centroid matrix **G**. The procedure of updating the matrix **X** can be summarized as follows: Firstly, decomposing the resulting matrix **D**. Second, the eigenvectors **U** and eigenvalues **Λ** are extracted from **D** to receive matrix **Ψ**. Third, the centroid matrix **G** is determined with help of **Ψ** and the current matrix **X**. The final step of the current iteration is then to update matrix **X** with the new Euclidean Distance of the partioning cluster algorithm using the statistical unit projection matrix **Ψ** and the centroid matrix **G**. This iterative procedure is conducted until the increase of the value given by $\sum_{j=1}^{p}\left(G(X) - G\left(X \middle| Z_j\right)\right)$ is not significant.

3.2. Decision Tree and Random Forest

Second, the Random Forest (RF) classifier is used to predict the likeability of products. Breimann (2001) defines RF as ensemble of tree-based classifiers $h_k(x) = h(x, \Theta_k)$ where $\Theta_k(k = 1, \dots, K)$ is a random vector that is created independently from the previous random vectors $\Theta_1, \dots, \Theta_{k-1}$. Decision Trees are grown by using Θ_k and the input training data x.

Song and Ying (2015) describe the basic concept of a binary decision tree classifier as a supervised learning method that predicts an output variable Y that has two possible outcomes (0 or 1) regarding the features $X_j(j = 1, \dots, p)$ of the trainig data. A Decision Tree consists of nodes and branches. It is grown by splitting nodes, stopping the node split or pruning the tree afterwards.

Nodes are divided into root node (subdivides records into mutually exclusive subsets), internal nodes (represents choices at a certain point in the tree) and leaf nodes (represent the final classification).

Branches can be understood as pathways that hierarchically connect root nodes or internal nodes with the leaf nodes and represent outcomes that can be read as a decision rule.

Splitting is the process of dividing observations regarding a feature X_j into bins. The Gini-Index is used to evaluate the purity of a node split and is defined by James et al. (2013) as

$$G = \sum_{k=1}^{K} \hat{p}_{mk}\,(1 - \hat{p}_{mk}),$$

where $\hat{p}_{mk}$ determines the proportion in the m-th region from the k-th class. Therefore, the Gini-Index is a measure of total variance across the K classes and also known as node purity. A small value of the Gini-Index is given if $\hat{p}_{mk}$ is either close to zero or close to one. Such a node contains mostly only observations of one class.

Stopping is the counteract of splitting. It balances the bias-variance trade-off to prevent the tree from over-fitting the training data. In other words, stopping determines stopping rules like a minimum node size, a minimum number of observations in each leaf node or the maximum depth of a tree.

Pruning is an alternative concept of stopping, thus the largest possible tree is grown and pruned subsequently by removing nodes that lack in information gain regarding some information criterion (e.g. Gini-Index).

A decision tree is a high variance method. In other words, several decision trees might end up with highly varying solutions as they split the training data differently. This can be solved by Bagging which bootstraps the initial training data to fit a decision tree to each bootstrapped dataset. The high variance of each single decision tree can be reduced by aggregating. In the context of classification each tree casts a vote for its prediction $\hat{f}^b(x)$ $(b = 1, \dots, B)$ whereas the prediction outcome with the highest number of votes is chosen. This process is called majority vote and is formally written as

$$\hat{f}_{bag}(x) = \underset{y \in Y}{\arg max} \sum_{b:\hat{f}^b(x)=y} 1,$$

where the final prediction $\hat{f}_{bag}(x)$ is equal to the outcome y that yields the highest number of votes given the predictions of each classifier $\hat{f}^b(x)$ on the B bootstrapped samples. Due to the reduced variance by aggregating decision trees, in contrast to single decision trees, bagged decision trees are not pruned. A further advantage of Bagging lies in the bootstrap process. Empirically, $\frac{1}{3}$ of the data is not used in the bootstrapped datasets, thus these out-of-bag (OOB) observations can be used to test the model performance (James et al., 2013).

Finally, RF is the combination of Bagging and a random selection of features. Each tree in the RF uses m out of p features. An optimal number of random features have been shown to be $m = \sqrt{p}$. Especially if a particularly strong feature is present, single trees are likely to be structured all the same. Random feature selection decorrelates the trees and ensures low variance as aggregating highly correlated trees would not lead to a large decrease in variance. The higher the correlation between predictors, the lower the number of randomly selected features m per tree to ensure uncorrelated trees in the RF. Hyperparameters in RF are the number of trees, the minimum size of nodes in each tree, the splitting rule, the number of random features and the sample size drawn per tree. Probst et al. (2019) determine the brier score to perform well while tuning the classification model to specify the tuned hyperparameters that optimize the RF model. Brier (1950) defined this formally as

$$BrierScore = \frac{1}{n} \sum_{r=1}^{R} \sum_{i=1}^{n} (f_{ir} - C_{ir})^2,$$

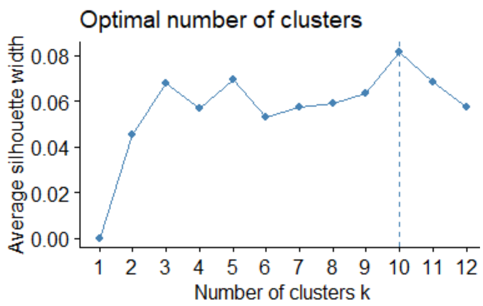
where C_{ir} takes values one or zero to indicate if observation i is in the actual class r out of R or not, whereas f_{ir} is the predicted probability that observation i belongs into a certain class r. A low value of the brier score therefore indicates less discrepancies between the predicted probabilities that an observation belongs into a class or not and the actual class.

Ensemble learning methods sacrifice interpretability in exchange to an improved prediction performance. Therefore model-agnostic methods are at hand to understand how the model makes its predictions. A basic global model-agnostic method is Variable-Importance (VIP). In RF, VIP is the summed relative importance of feature j over all internal nodes per tree which is subsequently averaged (Breimann, 2001). Unfortunately, nothing can be said about the direction of a variable influence. In the focus of local methods an expansion of the Partial Dependence Plots which investigate the marginal effect of feature j on the predicted outcome (Friedmann, 2001) is used. Namely, the Accumulated Local Effects (ALE) specified by Apley and Zhu (2020): Intuitively, ALE shows how feature j , isolated of all other features, changes locally if j is varied. This is done by accumulating the averaged changes in predictions. This method is favourable because it eases the assumption of uncorrelated features that is made for PDP.

4. Results

4.1. Product Clustering using i-FCB

Figure 2 Silhouette Method



The first research aim is to create subcategories for the listed summer products by using provided search tags. As described in section 3.1 the i-FCB method is used to combine a correspondence-analysis approach with a non-hierarchical partitioning algorithm. First, binary variables for each distinct tag are created. Second, tags that have less than ten counts are dropped as they are said to have too less information due to their rareness. Third, three tags are dropped as they occur in about 75% of products and therefore discriminate too less between products. From the initial 2047 tags, 244 remain in the analysis. Subsequently, the silhouette method in Figure 2 determines the optimal number of clusters K as $K = 10$. Further, each product is assigned to 1 out of 10 clusters which are described more detailed in Table 2. The corresponding biplot in Figure 1 plots the products on the first and the second dimension. The distances between observations relate to the similarity/dissimilarity between observations, whereas close observations are similar to each other and far observations are dissimilar to each other. As expected, search tags regarding dresses are rarely used in combination with products of the categories "Swim Wear", "Men/Sportswear" or "Sleeveless Tops" whereas all categories regarding different types of dresses appear in the same sub-space. Further, "T-Shirts/Blouses" appear closer to "Sleeveless Tops" than "Pants/Shorts" or the subspace of dresses. Overall, the allocation of products in the corresponding clusters is in line with the previous expectations, thus the clusters are used in the subsequent analysis.

Figure 1 ClusMCA [Method: i-FCB]

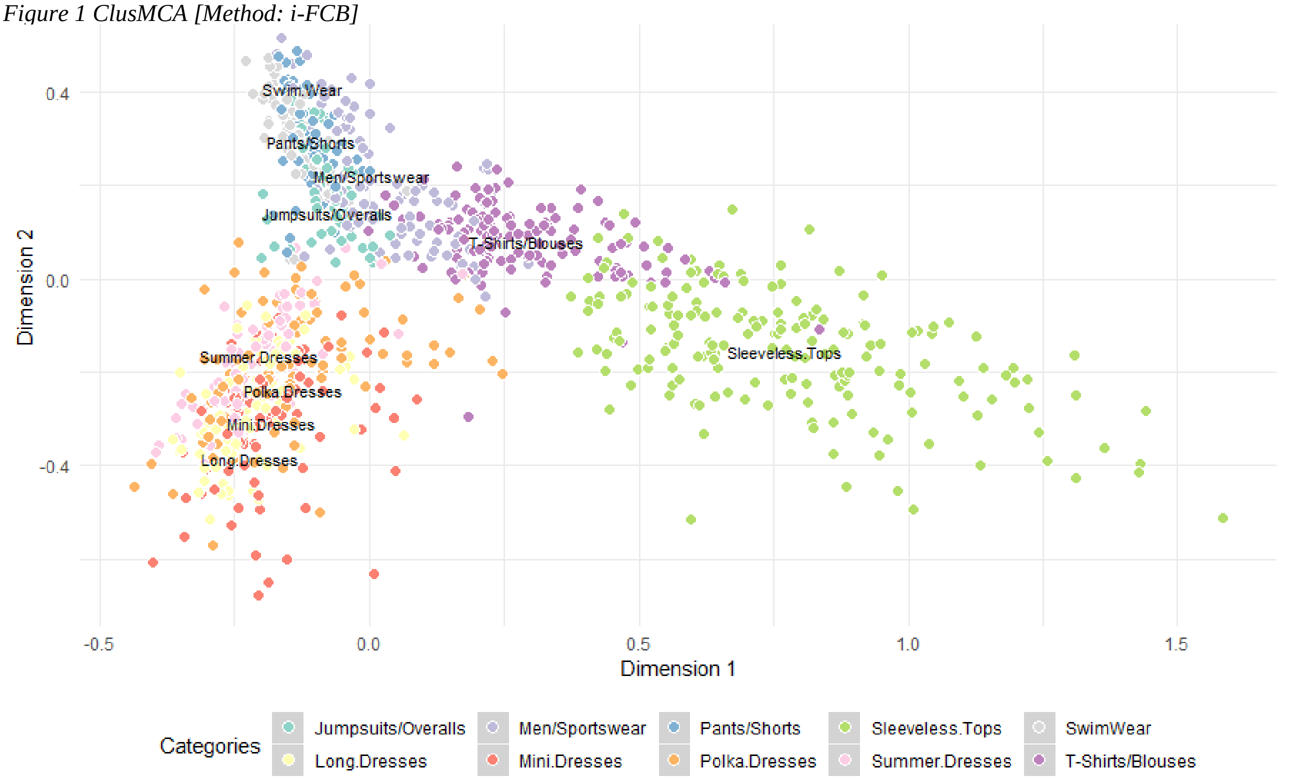


Table 2 Description of Cluster-Tags

	Cluster	Size	Influential Tags (1=Yes, 0=No) Negative Tags	Positive Tags
1	Sleeveless Tops	208	tops.0, tank.0, tank.top.0, dress.1, shorts.1	tank.top.1, tank.1, vest.1, tops.1, sleeveless.tops.1, sleeveless.1, sleeveless.shirt.1, women.vest.1
2	Polka Dresses	148	dress.0, tank.1, tops.1	dress.1, loose.1, round.neck.1, necks.1, neck.dress.1, polka.dot.1, womens.dresses.1, mini.1, polkas.1
3	T-Shirts/Blouses	145	sleeve.0, short.sleeves.0, dress.1	t.shirts.1, sleeve.1, short.sleeves.1, women.t.shirt.1, summer.t.shirt.1, blouse.1
4	Men/Sportswear	125	dress.1, plus.size.1, sleeveless.1	men.1, men.s.fashion.1, sport.1, men.s.shorts.1, yoga.1, fitness.1
5	Pants/Shorts	105	short.0, pants.0	short.pants.1, pants.1, jeans.1, waist.1
6	Long Dresses	103	dress.0	maxi.dress.1, long.dress.1, party.dress.1
7	Summer Dresses	100	dress.0	white.1, sundress.1, cocktail.1
8	Jumpsuits/Overalls	96	jumpsuit.0, tops.1	jumpsuit.1, rompers.1, overalls.1, trousers.1, long.pants.1
9	Mini Dresses	93	-	casual.dress.1, mini.dress.1, mini.1, slim.dress.1, bodycon.dress.1,
10	Swim Wear	88	swimsuit.0, bikini.0, swimwear.0	swimsuit.1, swimwear.1, bikini.1, swimming.1, tankinis.1, onepiece.1

4.2. Prediction of Product Likeability

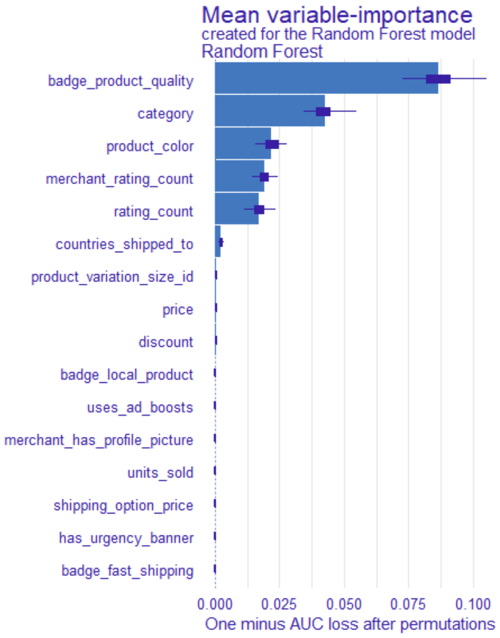
In order to predict the likeability of the listed products a variable "LIKE" has been created which is based on a product's average rating. The average of all product ratings is 3.8. As customer ratings are restricted to whole numbers a product is considered as liked if the rating exceeds the threshold of 4. To compare different prediction approaches the data is divided into training (65%) and test (35%) data. Further, as the classes are imbalanced between "DISLIKE" (67%) and "LIKE" (33%) the split is done by stratification to obtain an equal split between both outcomes in training and test data. Next, 16 features have been selected to predict the outcome variable "LIKE". They, firstly, describe the products' popularity (number of ratings, total units sold), secondly, the products' characteristics (price, shipping price, type of discount, if the product earned badges [local product, fast shipping and quality], product color, clothing size, if a product is boosted by advertisments and if a product indicates a limited stock) and thirdly, the merchants' characteristics (amount of shipped countries, merchant rating counts, and if the merchant has a profile picture or not). Finally, the variable "category" is considered as well, which enables to consider the categories defined in section 4.1.

Table 3 Prediction Results

Prediction Model	Decision Tree		Bagging		Random Forest	
	Dislike	Like	Dislike	Like	Dislike	Like
Dislike (Negative)	245	82	266	75	280	81
Like (Positive)	40	56	19	63	5	57
Accuracy	71%		78%		80%	
Cohen's Kappa	0.28		0.43		0.46	
Optimal Parameters	-		-		$\boldsymbol{ntrees^* = 500}$, $\boldsymbol{mtry^* = 7}$, $\boldsymbol{min.node^* = 4}$, $\boldsymbol{sample.split^* = 0.7086}$	

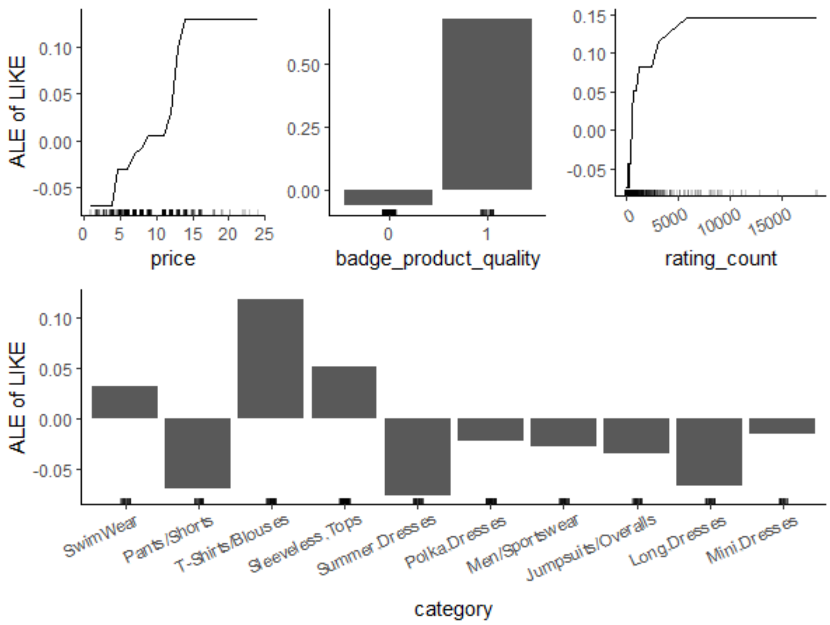
Cohen's Kappa is used for model evaluation as it is robust to imbalanced data. Hence, the Random Forest with 7 randomly selected features per tree, a minimum node size of 4 observations and a sample split per node of 0.7086 has a prediction performance of 80% accuracy and a Cohen's Kappa of 0.46, which is determined as fair agreement. Random Forest improves over Bagging in terms of decreasing FP's in exchange for an increase in FN's which may be favourable as the related costs for predicting that a product will be liked even though it will probably be disliked might be higher than vice versa. (see Table 3)

Figure 3 Variable Importance (VIP)



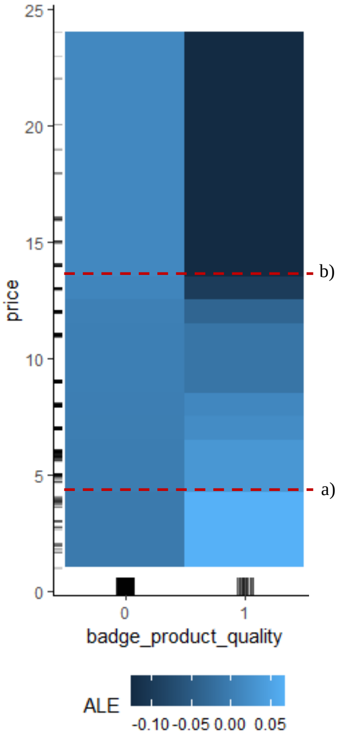
Next, the RF model is investigated to find drivers of the predictions. First, the VIP indicates quality badges, the products' categories and colors as well as the total number of ratings as substantial drivers of the prediction outcome. Interestingly, price and sold units have only limited impact on predictions (see Figure 3). Accumulated Local Effects (ALE) of the RF model are investigated to elaborate the local effects if the variables price, quality badges, product ratings and categories are varied. The model suggests an upward trending but non-linear relationship with a stepwise increase in the average prediction as prices increase. Three major price segments can be approximated. First, the low-priced segment \$1 - \$5 dollars. Second, the mid-priced segment \$5 - \$13.50 dollars. Third, the high-priced segment where prices are higher than \$13.50 dollars. The positive relationship between higher prices and a higher predicted likeability could imply more expensive products to be of better quality. The substantial driver, quality badge, influences the prediction outcome positively if a product has a quality badge and slightly negative if the badge is missing. The predicted probability of satisfied customers increases with the total number of ratings which might also imply a confirmation of a corresponding quality. Interestingly, the predicted probability increases largely until reaching a threshold of about 800 - 1,000 ratings. Product categories, however, impact the prediction ambiguously. "Swim Wear", "T-Shirts/Blouses" and "Sleeveless Tops" indicate positive effects on the predicted likeability, whereas the remaining categories show negative impacts. (see Figure 4).

Figure 4 ALE Plots: Price, Quality Badge, Ratings and Categories



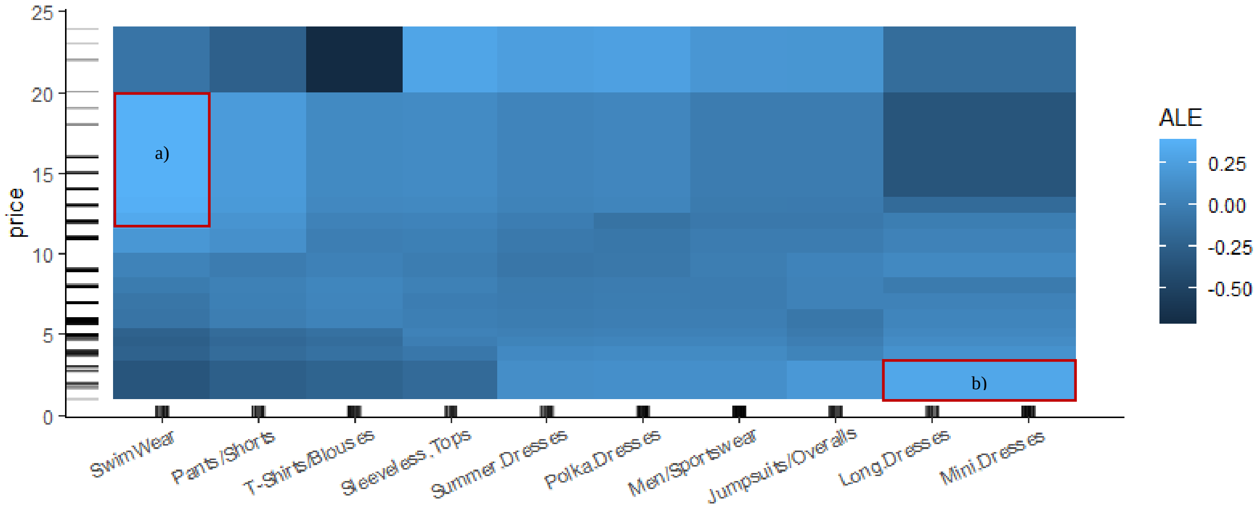
As product quality might be a substantial driver of the prediction outcome its effect is further investigated by plotting its second-order ALE in combination with price which represents a change in the main effect that has been shown in Figure 4. It discovers products of the low-priced segment in combination with a quality badge having substantial increase in the main effect on the predicted probabilities compared to products in the same segment without quality badges (see Figure 5, below line a). Further, a reversed effect is detected in the mid- to high-priced segment in which products without quality badge but higher prices have slightly increased main effects on predicted probabilities of being liked (see Figure 5, above line b). Surprisingly, products of the same segment but with quality badges experience a decrease in the main effects, thus they have reduced but still positive effects (see Figure 5). A possible explanation of the reversed effect could be the expectation management of customers. Expectations may be higher for high-priced products with quality badges compared to high-priced products without quality badges.

Figure 5 ALE Plot: Price and Quality Badge



In the same manner, the variation of product categories was investigated in combination with prices to elaborate if certain categories should be offered in combination with a low-, mid- or high-priced strategy. Swim wear in the mid-priced segment from \$10 - \$20 dollars experienced a substantial increase in the main effects on predicted likeability (see Figure 6, a). Apart from that dresses within the mini and long category are found to have increased main effects on the predicted probabilities of likeability in combination with a low-price strategy (see Figure 6, b). Despite their main effects are somewhat negative, they end up with a positive effect in the low-priced segment. However, such a product should not exceed a price of \$12 dollars (see Figure 6).

Figure 6 ALE Plot: Price and Category



5. Conclusion

Relating back to the initial research questions, products have been clustered into categories using the i-FCB method with the optimal number of categories being determined statistically as 10. The solution has been confirmed in the business context by labelling the categories with the help of tags that have been found to have high impacts on specific clusters.

In the prediction task for the given set of variables a fine-tuned Random Forest outperforms a single Decision Tree as well as Bagging in terms of accuracy and Cohen's Kappa. Despite only one-third of variables seeming to be the main drivers of the predicted outcome (see Figure 3, page 8) all variables have been kept to take advantage of random feature selection in RF and therefore guarantee uncorrelated trees and avoid overfitting. Building up on the results, a next step would be to elaborate the costs related to false positive and false negative predictions to adjust the model in the business context. Negative reviews are made if a product is disliked even though the customer expected to actually like it. Therefore, this number (false negatives) has been minimized as much as possible in this analysis. Further investigation is needed to translate the proposed solution into costs to solve this trade-off.

Wish.com products are generally cheap (average price of $8 dollars and average shipping costs of $2.40 dollars) but its customers do not want to sacrifice quality. Total ratings, higher prices as well as badges for good quality have been found to have positive upward sloping main effects on the predicted probability of product likeability and are therefore suggested to be indicators for product quality.

Two alternative market entry strategies are proposed: First, quality badges have been found to increase likeability of products even in the low-priced segment ($1 - $5 dollars). In this segment products within the mini and long dresses categories are favourable to ensure product likeability. Their negative main effects are supplemented with large increases in the effect on product likeability. Second, in the case that the product badge cannot be earned a mid- to high-priced strategy ($13.50 - $20 dollars) should be preferred as the higher price itself implies product quality for the customer. The substantial category in this segment to ensure product likeability is "Swim Wear", whose main effect on likeability increases largely by offering the product in the price range of $13.50 - $20 dollars.

These strategies are proposed based on data collected during the COVID-19 crisis, thus customer behavior on e-commerce platforms might differ in post-crisis periods. The results are therefore limited to the current circumstances.

Unfortunately, nothing can be said about the product-specific target group as no customer-level information is available. Further, the expectation management of customers could help to enhance the understanding of product likeability by considering the product rating in relation to the actual expectation of customers before the product was ordered. Future research should therefore focus on understanding the expectation management of online shoppers as the process is substantially different from visiting shops. As Kaushik et al. (2018) elaborated, ratings and reviews received in the early stage of a product on e-commerce platforms are of substantial importance and may decide on a product's success or failure.

6. References

Greenwell, B.M., & Boehmke, B.C. (2020). Variable Importance Plots—An Introduction to the vip Package. R Journal, 12, 343-366.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. Annals of statistics, 1189-1232.

Brier, G. W. (1950). Verification of Forecasts Expressed in Terms of Probability. MWRv, 78(1), 1.

Bhatti, A., Akram, H., Basit, H. M., Khan, A. U., Raza, S. M., & Naqvi, M. B. (2020). E-commerce trends during COVID-19 Pandemic. International Journal of Future Generation Communication and Networking, 13(2), 1449-1452**.**

Kaushik, K., Mishra, R., Rana, N. P., & Dwivedi, Y. K. (2018). Exploring reviews and review sequences on e-commerce platform: A study of helpful reviews on Amazon. in. Journal of Retailing and Consumer Services, 45, 21-32.

D'Enza, A. I., & Palumbo, F. (2013). Iterative factor clustering of binary data. Computational Statistics, 28(2), 789-807.

Van de Velden, M., D'Enza, A. I., & Palumbo, F. (2017). Cluster correspondence analysis. psychometrika, 82(1), 158-185.

Song, Y. Y., & Ying**, L. U. (2015)**. Decision tree methods: applications for classification and prediction. Shanghai archives of psychiatry, 27(2), 130.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning (Vol. 112, p. 18). New York: springer.

Probst, P., Wright, M. N., & Boulesteix, A. L. (2019). Hyperparameters and tuning strategies for random forest. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 9(3), e1301.

Apley, D. W., & Zhu, J. (2020). Visualizing the effects of predictor variables in black box supervised learning models. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 82(4), 1059-1086.

Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.

Forbes Magazine (2019). Daily Cover. March 13, 2019, 06:00am EDT. Accessed: 17.12.2020. https://www.forbes.com/sites/parmyolson/2019/03/13/meet-the-billionaire-who-defied-amazon-and-built-wish-the-worlds-most-downloaded-e-commerce-app/?sh=60785d170f52