Erasmus University Rotterdam
Erasmus School of Economics
Master Thesis Data Science and Marketing Analytics

Analyzing Consumer Preferences in Sustainable Clothing through Text Mining of Online Reviews

Ànnia Urpinell Aynés (657504)

Supervisor:	Anastasija Tetereva
Second assessor:	Michiel Van Crombrugge
Date final version:	8th August 2023

Abstract

This master's thesis explores ways to extract customers' opinions and preferences on products by using user-generated content. The study establishes a framework to gain insights into customers' opinions about specific product attributes by analyzing online reviews for both sustainable and non-sustainable clothing products. Latent Dirichlet Allocation (LDA) is used to identify product attributes, and then Aspect-Based Sentiment Analysis is employed to identify the sentiment toward each attribute. The identified topics are later used to predict the overall rating in order to reveal attribute importance. Results show common attributes between sustainable and non-sustainable products, with varying importance, and that the overall sentiment toward attributes is positive. The study provides a straightforward framework to gain managerial insights from online reviews, aiding decision-making in marketing and product development for sustainable clothing products.

Keywords: Sustainable clothing, Online consumer reviews, Product attributes, Latent Dirichlet Allocation (LDA), Aspect-Based Sentiment Analysis, Attribute importance.

Contents

Chapter 1

Introduction

Given the rising awareness and concern about climate change, sustainable consumption has become an important focus for governments, organizations, and consumers alike. Over the last few years, there has been a noticeable increase in consumer awareness concerning the environmental consequences of their purchasing behavior, resulting in a growing interest in alternatives to fast-fashion. Companies have recognized the need to cater to these new demands by offering sustainable products that are aligned with these environmental concerns. However, despite this heightened awareness and the increasing number of alternatives, the fast fashion industry continues to dominate the market, and significant changes in consumption habits have yet to be seen. To address this challenge, understanding consumers' opinions and preferences regarding sustainable clothing products has become a crucial step that can aid in promoting and increasing consumption of more sustainable and environmentally conscious choices.

At the same time, the rapid growth of online retailing has opened up new opportunities for companies to gain insights into consumer behavior through user-generated content, particularly in the form of online reviews. Online consumer reviews have become a valuable source of data for businesses due to their cost-effectiveness and lack of biases often associated with more traditional market research methods. Consequently, companies are now starting to leverage the vast amount of available online reviews to extract valuable information that can assist in their marketing strategies, product design, and development efforts. The importance of online consumer reviews has also grown due to purchasing decisions relying more and more on the opinions and experiences of other consumers. User-generated content has become a trusted and influential resource for potential buyers, helping them make informed choices and creating a sense of community in the digital marketplace. For businesses, the value of online reviews lies in their user-oriented nature, providing authentic feedback that can directly impact their brand reputation and success.

This thesis brings these two ideas together and attempts to better understand consumer preferences when purchasing clothing online with the use of online consumer reviews. The objective is to identify the most important product attributes that resonate with consumers in the context of sustainable and environmentally friendly clothing products. By understanding consumers' preferences and requirements, businesses can enhance their product design and development processes, making green products more appealing and competitive in the market, and fostering a more sustainable and environmentally conscious future for the fashion industry.

1.1 Research Question

The goal of this study is to provide a framework to obtain insights on customers' opinions about specific product features using online customer reviews on sustainable and non-sustainable products. More specifically, it aims to identify product attributes and how they differ between sustainable and non-sustainable products, to provide an alternative to frequency-based attribute importance, and to assess customer's feelings towards each product feature. The research questions that this thesis will attempt to answer are the following:

1. Can online customer reviews be used to identify the most important attributes of clothing products and the customer's attitude towards each of them?
2. Can the importance of each of these topics be derived by measuring their impact on the overall rating?
3. What are the most important attributes valued by consumers when buying clothes online?
4. Are there significant differences in the most relevant features between sustainable and non-sustainable products?

1.2 Managerial Relevance

Text mining user reviews can help firms get a deeper understanding of consumers' opinions. Companies typically use ratings and sales data to help monitor performance, but the insights that can be obtained with user reviews are much richer. Companies can obtain information through online reviews that is similar to the one obtained in consumer surveys, but the cost is much lower. There are several potential benefits that can be derived from this study which can significantly impact managerial decision-making, namely: improve product design, as well as marketing communications, identify potential upgrades, benchmarking, and enhancement of customer retention.

To begin with, the findings of this study shed a light on what product features customers value the most. This information can be used by organizations to optimize their marketing communications, since messages highlighting these features will likely be more effective, as well as in the design of new products or lines.

Moreover, this analysis will also aid in identifying potential upgrades in products. Identify what are the aspects that customers don't like about the products, to improve both the product itself and the online image of products.

Furthermore, a company can obtain useful insights from not only analyzing reviews of their own products, but also products from competitors. This can help companies learn from their direct competitors, and also help sustainable brands to improve their products and marketing to reach fast-fashion customers.

Lastly, one of the main issues to tackle in marketing is customer retention. For this reason, obtaining real customer opinion right after a purchase can provide valuable information to create strategies to increase customer loyalty.

1.3 Academic Relevance

One of the main contributions of this study in the field of text mining is to provide a simple approach to topic modeling, identification of product attributes and overall rating prediction. While it is true that there have been several studies using state-of-the-art models to obtain insights from text reviews with a high level of accuracy (Noori, 2021; Chakraborty, Kim & Sudhir, 2022), these approaches typically require large amounts of labeled data, they are computationally expensive and the inner workings of the models are usually not known, making interpretations of the drivers of predictions difficult. This study provides a straight-forward framework that uses unlabelled data and that is easily interpretable, with the purpose of easy application by marketing managers.

As concluded by Chakraborty et al. (2022), deriving product attribute importance from the frequency of mentions in reviews can lead to biased conclusions, since the drivers of review writing are not necessarily a complete and objective evaluation of the product. This study adds literature to the line of research that uses overall rating as an indicator of customer satisfaction and predicts this score using the product attributes obtained through topic modeling. This approach has proven to be accurate with reviews from the restaurant sector, however, this study attempts to examine whether it also holds in the retail sector, more specifically, using clothing reviews.

The majority of existing literature on topic modeling and aspect-based sentiment analysis focuses on the restaurant sector, on accommodation websites such as hotels and only limited studies using retail and, more specifically, clothing products data (Jia, 2018; Archak, Ghose & Ipeirotis, 2011). Moreover, few studies have used text mining of online reviews to tackle consumers' opinion on sustainable products and green marketing. The existing literature on attitude towards sustainable products uses customer surveys as the primary source of data. Thus, my thesis adds literature on the line of research of attitude towards clothing products and sustainable alternatives, using user-generated content as a source of data.

Chapter 2

Literature Review

This section provides a review of the theory regarding attitude towards sustainable products, sustainability in clothing industry, and a review of studies using text reviews focusing on aspect-based sentiment analysis and rating prediction.

2.1 Attitude towards sustainable products

In recent years, with climate change becoming an imminent threat, the issue of how to live more sustainably has taken center stage. Public and private organizations, as well as individuals, are starting to make changes in the ways they operate to reduce negative impact on the environment. For private companies this has also created a new market: sustainable products. Sustainable products, also referred to as green products, are characterized by having positive social or environmental attributes (Luchs, Naylor, Irwin & Raghunathan, 2010). Then, individual consumers face the choice of opting for more sustainable options, or continuing with regular products. And the question arises of how consumers value these products as opposed to non-sustainable alternatives and what drives their choice to choose green products. There has been extensive research on sustainable products and customer behavior and purchasing decisions.

Several authors have concluded that environmental concern, i.e. the degree of consumers' awareness of environmental issues and their willingness to actively participate in solving them, is a key factor influencing purchase intention for green products (Alwitt & Pitts, 1996; P. C. Lin & Huang, 2012). However, this effect is indirect and it is mediated by specific attitudes about that product. For this reason it is important to communicate the sustainable aspects of green products in marketing strategies. There are other sources that argue that sustainable attributes don't always drive consumer choice (Rokka & Uusitalo, 2008) and that sustainable consciousness and positive attitude towards sustainable products doesn't always turn into actual purchases (Wiederhold & Martinez, 2018). Therefore, while it is true that generally there is a positive relationship between a sustainable attitude or concern and consumption of green products, there is still a gap between consumers who claim to be environmentally conscious and the share of consumption of green products (Peattie & Crane, 2005; Rex & Baumann, 2007). This points to the fact that there must be other relevant factors that influence the decision to purchase green products.

Based on the principles of utility theory, consumers will opt for the product that offers the highest utility, or in other words, the highest satisfaction. Marketing theory says that the utility of a product comes from the attributes or characteristics that the product possesses. Then, the attributes of a product act as qualifiers of the products themselves and during the decision-making process, these attributes determine the purchase decision. When deciding between several options, consumers assess each of the attributes and use them as criteria to decide (Lancaster, 1966; Srinivasan, 1979).

It is common for sustainable products to have higher prices compared to analogous non-sustainable options. Prices tend to be higher due to more expensive materials, due to higher manufacturing costs or even as a pricing strategy to establish green products as premium within the market (Peattie & Crane, 2005). For this reason, one of the product attributes that might have a negative impact on consumers choosing green products is their premium price. Overall, there seems to be a willingness to pay superior prices for green products. According to the European Comission (2013), approximately 75 percent of citizens claim to be willing to pay additional price for green products. But while there are several sources that report a willingness to pay more for sustainable products (Laroche, Bergeron & Barbaro-Forleo, 2001; Gomes, Lopes & Nogueira, 2023), other sources say that this willingness is modulated by the product category and the perceived benefits of the specific product (Essoussi & Linton, 2010). Yazdanpanah, Forouzani and Hojjati (2015) and Kovacs and Keresztes (2022) conclude that especially young people perceive the benefits of green products and are willing to pay more. P. C. Lin and Huang (2012) found that neither quality or price doesn't seem to have significant effects on consumer choice behavior, it seems that the benefits of purchasing sustainable products outweigh the additional costs that green products typically carry.

2.2 Sustainability in the clothing sector

Most of the research regarding purchase decisions of green products focuses on food, but limited literature on other sectors such as housing, automobiles and clothing can also be found (X. Zhang & Dong, 2020; Bangsa & Schlegelmilch, 2020).

X. Zhang and Dong (2020) studied the drivers of green product consumption and found that availability, product quality, packaging and origin are the most common. However, if we focus on the clothing sector they found that customers still pay attention to quality, but aesthetics and the appearance of the product play a bigger role in product choice. In the study conducted by Jung, Choi and Oh (2020), they found that aesthetic and utilitarian values, as well as social norms influenced purchasing intentions for green apparel products. Nonetheless, it seems that consumers will still choose functional performance over aesthetics in sustainable products when given the choice. In a recent study, Luchs and Kumar (2017) theorized that consumers make trade-offs between the sustainable aspect of a product and other product attributes. They found that consumers are more likely to trade-off hedonic value for sustainable value rather than utilitarian value. However, since clothing is used as a self-expression tool and as a representation of personality, aesthetics play a key role in purchase intention, and for that reason, customers might favor fashion alternatives that are cheaper because they offer more appealing offers and more options. Therefore, consumers will favor the need to self-express over an ethical fashion choice (Sudbury-Riley & Boltner, 2010). This needs to be researched further, since there are some discrepancies in literature. Battenfeld, Hörisch, Jacobs and Petersen (2018) found that hedonic values had a negative effect on sustainable clothing purchases and price and fashion consciousness, defined as interest in appearance and style, did not influence behavior in a significant way.

Within the clothing sector, authors have also found that environmental concern and a positive attitude towards ethical fashion has a positive impact on purchase intentions (Sudbury-Riley & Boltner, 2010). Still, there is still a gap between purchase intention and actual purchases, which is referred to as the attitude–behavior gap in sustainable consumption (Battenfeld et al., 2018; Park & Lin, 2020). Consumers themselves claimed the reasons were price and lack of ethical alternatives in the mainstream market and in the media. While that might be true, the data shows that shifting the responsibility to third parties was also a probable cause (Sudbury-Riley & Boltner, 2010).

2.3 Text mining online reviews

Companies carry out consumer surveys to obtain consumer opinions on their products, but these have biases, are costly and become outdated quickly (Chakraborty et al., 2022). But to learn consumers' opinions, user generated content such as online consumer reviews (OCR), offers a cheap, fast and reliable alternative. Additionally, it has become imperative for companies to be aware of their online image and what information and opinions will prospective buyers see in the media. Several studies have proven that product reviews have a significant effect on sales (Forman, Ghose & Wiesenfeld, 2008; Ghose & Ipeirotis, 2011).

For this reason, academics have been using text mining to extract information and derive managerial insights from user-generated content for several years (Netzer, Feldman, Goldenberg & Fresko, 2012; Tirunillai & Tellis, 2012). And more recently, there have been multiple studies using topic modeling to find latent topics in online reviews. There is a wide variety of methods being used for topic modeling. Deep learning models using word vectors have proven to be highly effective in text classification tasks, compared to more traditional supervised and unsupervised algorithms (Liang, Sun, Yunlei & Gao, 2017). However, these models are considered black-box models and therefore interpretability is not possible (Chakraborty et al., 2022).

There has also been extensive work on sentiment analysis for text reviews, using either lexicon-based approaches (Prakoso, Yananta, Setyawan & Muljono, 2018) or machine-learning based (Noori, 2021; Pang, Lee & Vaithyanathan, 2002). Sentiment analysis can be applied to the full review body to extract the general sentiment that the reviewer wants to express. However, this can prove to be complex, since in a lot of reviews multiple topics are mentioned and, while some of them are praised, others are criticized. In order to obtain an accurate picture of the sentiments that a review is expressing, sentiment should be extracted on an attribute-level. This is what is called Aspect-Based Sentiment Analysis and several authors have been applying it in their studies for several years (Tang, Fu, Yao & Xu, 2019; Xu, Liu, Wang & Yin, 2018; Y. Lin, Fu, Li, Cai & Zhou, 2021). Some of the latest applications use the bidirectional encoder representation from transformers (BERT), a language representation model introduced by Google in 2018 that has shown promising results (Devlin, Chang, Lee & Toutanova, 2019; Y. Zhang, Du, Ma, Wen & Fortino, 2021).

Büschken and Allenby (2020) proposed an improvement on the Latent Dirichlet Allocation model that accounts for topic carryover by including punctuation and conjunctions such as “and” and “but” in LDA. This improved model is still considered to use a bag-of-words approach, however, some sentence structure information is added, proving to be able to process better complex sentences.

There is another line of research that links latent topics and sentiment analysis to overall rating or other variables in order to assess the impact of topics and sentiments (Jia, 2018; Archak et al., 2011; Mahadevan & Arock, 2019; Chakraborty et al., 2022). Archak et al. (2011) used an econometric model using product reviews from Amazon to predict sales rank, using product features identified using an automated text-mining approach and a manual tagger approach. In this study, sentiment valence for specific product attributes is estimated and its impact on demand is evaluated. Another article in this field is the one published by Mahadevan and Arock (2019). They use different topic modeling techniques, including TF-IDF, LDA and NNMF, perform sentiment analysis and rating prediction using several different models. Chakraborty et al. (2022) developed a deep learning approach to obtain fine-grained attribute level sentiment scores. In their study, they account for spatial structure as well as for missing attributes when predicting ratings with the attributes and the sentiment scores. Their research on missing attributes concludes that attribute presence in reviews is not necessarily a proxy of its importance, since what drives review writing is usually the need to inform prospective buyers or to vent. Therefore, the frequency of mentions of a topic might lead us to wrongly estimate attribute importance.

As we have observed, the literature on consumers’ attitude towards sustainable clothing products is limited. Furthermore, research regarding product attributes is very broad and there is no literature addressing what specific features or attributes in clothing products are most important to customers. Consequently, this study aims to fill this research gap. Moreover, this study contributes to the literature on Aspect-Based Sentiment Analysis and provides a straight-forward framework to identify latent topics in reviews and assess consumers’ sentiment towards each of them. Then, applying a multi-attribute model where product evaluations are expressed as weighted sums of assessments of various product attributes, the importance of each attribute is derived. This offers an alternative approach to frequency-based methodologies, which have been proven to be inaccurate.

Chapter 3

Methods

The methodology used in this thesis can be divided into three different tasks: first, topic modeling, second, aspect-based sentiment analysis, and third, predicting overall rating. The following flowchart (Figure 3.1) represents the methodology of this study in broad terms.

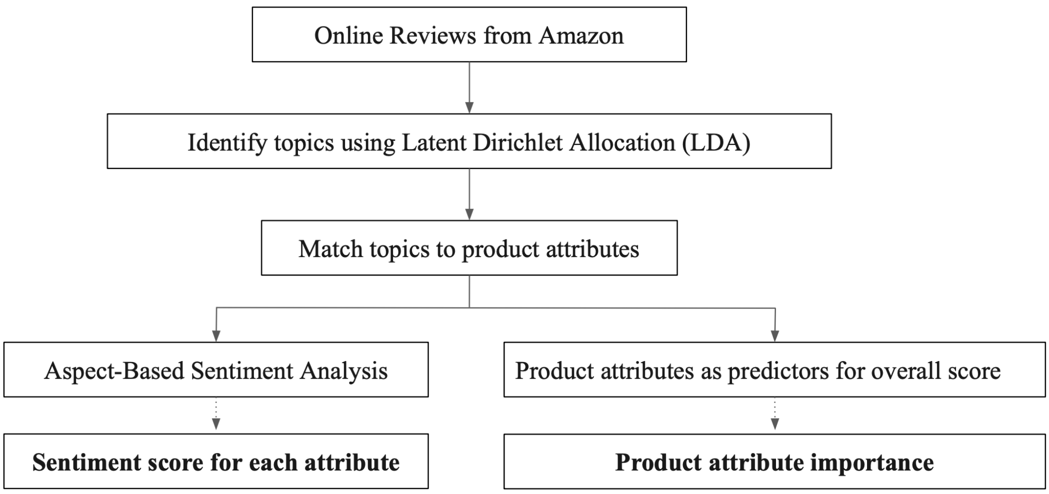

Figure 3.1: Methodology flowchart

3.1 Latent Dirichlet Allocation

The method used for topic modeling in this study is Latent Dirichlet Allocation (LDA), a generative probabilistic model commonly used in natural language processing to identify latent topics in textual data. It is an unsupervised machine learning model that identifies themes or subjects that tend to appear together in a set of documents, in this case, in reviews. It assumes that each document contains a variety of topics in different proportions and that each topic contains a set of words with different probabilities of occurrence. Since it allows for a document to be described by several topics with different probabilities, it is considered a soft clustering method. This method uses a bag-of-words approach, because it considers the set of words present in documents and their frequency, but not the order of those words.

To achieve that, the model assumes two prior Dirichlet distributions. First, the topic distribution for document i, $\theta_i \sim Dir(\alpha)$, where $i \,\epsilon\, \{1, ..., M\}$, with M denoting the total number of documents. And second, the word distribution for topic k, $\varphi_k \sim Dir(\beta)$, where $k \,\epsilon\, \{1, ..., K\}$, with K denoting the number of topics. Then, it attempts to find the most likely distributions by iteratively optimizing the parameters with an inference algorithm. In this case, LDA is implemented using *text2vec*, an R package that uses a parallelized Variational Bayes algorithm for inference.

A requirement of LDA is that k, the number of topics, needs to be determined in advance. Additionally, the document-topic prior and the topic-word prior also need to be specified beforehand. These parameters are tuned using grid-search, an algorithm that explores all possible combinations of a predefined set of values for each parameter, and chooses as optimal the combination that results in the best value of the chosen evaluation metric. In this case the evaluation metric is perplexity, a common measure for model fit that evaluates the performance of a probabilistic model by measuring how well the model can predict unseen data, therefore a lower perplexity score implies better model performance. Perplexity is calculated as:

$$Perplexity = exp\left(-\frac{\sum_{d=1}^{M} log\, p(w_d)}{\sum_{d=1}^{M} N_d}\right),$$

where w_d represents the words in document d, and N_d the number of words in document d.

However, to choose the number of topics it is also important to rely on the researcher's judgment. Depending on the purpose of the study, an appropriate number of topics based solely on perplexity score will not lead to a good choice in terms of interpretability or usability. Additionally, in this study it is important that the number of topics found in both of the product groups is the same for them to be comparable and identify common topics. Therefore, the number of topics is first determined using the perplexity score as well as the interpretability of each topic for both product groups, and then the document-topic prior and the topic-word prior are tuned using grid-search.

A key step in LDA is the interpretation of the topics obtained. In order to do that, the top words for each of the topics can be obtained and the researcher can find a common theme in the words or a title that captures the idea of most of the words in that topic. The top words per each topic can be obtained by sorting the probability of each word being present in the topic, however, this does not take into account the frequency of words in the corpus. Therefore, the top words are defined using relevance, a novel method for ranking words within topics introduced by (Sievert & Shirley, 2014). The relevance of term w to topic k is defined as:

$$r(w, k|\lambda) = \lambda\, log(\phi_{kw}) + (1 - \lambda) log\left(\frac{\phi_{kw}}{p_w}\right),$$

where ϕ_{kw} denotes the probability of term $w \,\epsilon\, \{1, ..., N\}$ for topic $k \,\epsilon\, \{1, ..., K\}$, p_w denotes the marginal probability of the term w in the corpus and λ the weight given to the probability of term w under topic k relative to its rank.

3.2 Aspect-Based Sentiment Analysis (ABSA)

Once the latent topics have been identified, we move on to performing sentiment analysis on each of these topics. The methodology used in this study is similar to the one reported by Chakraborty et al. (2022). First, all reviews are separated into sentences using the punctuation, more specifically, the characters used for separation are ".", "?", "!", and ";" . Secondly the LDA model used in the previous section is applied to each of the sentences. Then, a topic is assigned to each sentence using the document-topic distributions. For each review with at least one probability greater than 0, the topic with the highest probability for each review is assigned. Then, sentiment analysis (SA) is performed in each of the sentences. There's different approaches for sentiment analysis, which can be differentiated into lexicon-based and machine learning. In this study a lexicon approach is used, more specifically, a dictionary-based approach, which relies on a predefined list of words and related sentiments to derive sentiment scores from textual data (Medhat, Yousef & Mohamed, 2014). In this study, the *syuzhet* package, available in R, is used to obtain the sentiment score from each sentence, which uses a custom sentiment dictionary developed in the Nebraska Literary Lab (NRL). The SA model identifies in each sentence words that are present in the predefined list, and therefore have a sentiment valence assigned, which is either positive (+1) or negative (-1). Then, for each sentence the average sentiment valence is calculated. Following a similar methodology as Chakraborty et al. (2022), two assumptions are made: each sentence is associated with only one attribute, the topic with the highest probability, and the sentiment score of each attribute is the mean of the sentiment score of all the sentences associated with that attribute.

In the following graph (Figure 3.2), an overview of how Aspect-Based Sentiment Analysis is performed in this study is depicted.

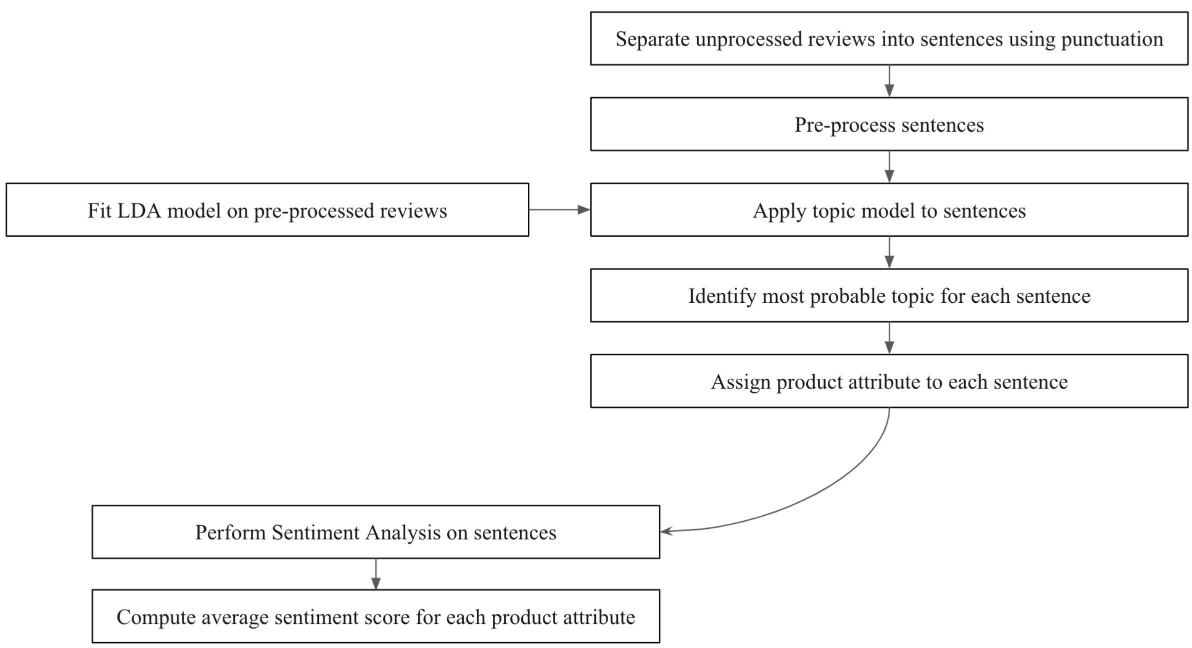

Figure 3.2: Flowchart representing how Aspect-Based Sentiment Analysis is applied

3.3 Multiclass Classification Tree

As a final step in the analysis, the importance of each product attribute is estimated by measuring its impact in the overall rating given by each reviewer using multiclass classification trees.

Decision trees are a supervised learning method that can be used for classification and for regression problems. In the case of a multiclass classification task, the algorithm creates a tree-like model of decisions based on the explanatory variables that predict the class of each observation. The basic idea behind the method is to continuously split the dataset based on the variable that best separates the target classes, i.e. that results in the most homogeneous groups possible. The splitting process is repeated until it reaches a stopping condition or until all end nodes contain observations of only one class, which are called pure nodes.

The decision tree uses a greedy algorithm to determine the best split at each step, i.e. it makes the best decision at each step without considering what is the best split globally. There are several possible criteria to determine the best split, such as Gini Index or Entropy. In this study the rpart function in R is used, which uses the CART algorithm and Gini Index as the splitting criteria, which measures the impurity of the nodes. A Gini index of 0 indicates perfect purity, which means that all the data points in the node belong in the same class. On the other hand, the higher the value of the Gini index gets, the higher the impurity. The objective is to create splits that generate nodes that are as pure as possible.

If a decision tree is allowed to grow until all nodes are pure, it will perfectly classify the training data but will perform poorly with new unseen data points, i.e. they will overfit the data. To prevent overfitting, several stopping rules can be applied. In this case, the size of the classification tree is controlled by the Complexity Parameter (cp), which uses a cost function to evaluate the quality of each split and balances tree complexity and the error rate. The cp acts as a threshold value to determine whether a split in the tree should be pruned, so the tree is recursively pruned back towards the root. The cp therefore controls the minimum improvement required for a split to be considered significant. This parameter is tuned with 10-fold cross validation, which randomly divides the observations into 10 groups, and while the model is fit on 9 folds it is validated in the remaining group. This process is repeated 10 times and the final evaluation metric is computed as the average of all 10 validations. In this case, the evaluation metric to estimate the optimal goodness of fit of the model is accuracy.

To evaluate the fit of the model, the confusion matrix is constructed. With the values in the confusion matrix, the accuracy and the balanced accuracy of each model is then calculated. The accuracy is computed as the sum of the true predictions (true positives and true negatives) divided by the total number of predictions, and the balanced accuracy is the average between sensitivity, the true positive rate, and specificity, the true negative rate.

After the model is built, the variable importance for each of the predictors can be obtained, which represents the contribution of each of the predictors in the model to the accuracy of the predictions of the decision tree. Variable importance is calculated as an extension of the Gini importance, and it considers the node impurity, the number of observations in each node, and the number of splits that involve each of the variables of the model. It is computed as the improvement in the overall Gini impurity achieved by each variable. The higher the gain in Gini impurity as a result of splits involving a variable, the more important that particular variable is to predict the dependent variable of the model.

Chapter 4

Data

This study uses online consumer reviews from clothing products on Amazon, obtained by performing web scraping. The online reviews belong to two different clothing brands: Amazon Aware and Amazon Essentials. Both brands are owned by Amazon, but the first one is marketed as a sustainable brand and the second as an affordable brand. The products used are men's and women's clothing items including: shirts, t-shirts, blouses, sweaters, sweatshirts and jackets. The reviews belong to 54 different products from Amazon Aware and 50 from Amazon Essentials.

A total of 18.858 reviews were extracted from the website, 2.690 belonging to Amazon Aware and 16.168 belonging to Amazon Essentials. The information scraped included the full review text, the price of the product, the total number of ratings given to that product and the overall rating given by each reviewer.

In order to analyze the review text, a lengthy data pre-processing is needed. The first step in pre-processing is to translate all the reviews that are not in English. This is done by DeepL, a translation service that uses artificial neural networks. Secondly, punctuation and all characters that are not letters are removed from the corpus. Then, stop words are removed using a premade stop word list from the SMART Information Retrieval System, developed by Cornell University. The list contains 571 words that frequently appear in text and that carry little information in text analysis, such as prepositions and articles. Removing these stopwords reduces dimensionality of the corpus, improving computational efficiency and increasing interpretability of results. Afterwards, some additional words are removed. The words for the different products, such as "sweatshirt", "shirt" or " jacket", do not provide any relevant information, therefore they are removed. Finally, all reviews are tokenized by words and the words are stemmed using Porter's stemming algorithm Snowball. Word stemming is a crucial step of pre-processing textual data. Stemming refers to the process of reducing words to their root or base in order to simplify analysis, since it reduces the dimensionality of the data and it converts words with different inflections or suffixes to the same form.

Prior to applying the text mining techniques described in the next section, the text data is required to be in Document-Text Matrix (DTM) form. In a DTM, each row represents a document, or review, each column corresponds to a unique term and the cell values represent the frequency of occurrence of each word in each document. Once the DTM for both datasets are created, the terms with less than 30 occurrences are removed in order to simplify the analysis and to maintain only the terms that are common across reviews. After data-preprocessing and the removal of infrequent words in the Document-Term Matrices, the number of reviews and unique terms left is slightly lower than in the raw data. In Table 4.1, some basic statistics of the data can be found.

	Sustainable Brand	**Non-sustainable Brand**
Time span of reviews	04/03/2022 - 03/05/2023	08/12/2018 - 22/05/2023
Number of reviews after pre-processing	2.090	12.245
Number of unique terms after pre-processing	379	663
Average price (in $)	31,59	21,28
Average rating (from 1 to 5)	4,47	4,16

Table 4.1: Summary statistics of the datasets

In Table 4.1 we observe that the average price is higher in the sustainable products, confirming the belief that, at least in these brands, sustainable products are more expensive. We also observe that the rating is between 4 and 4.5 for both groups, signaling that there might be class imbalance in the rating variable. If we take a closer look and visualize the distribution of the observations across rating scores (Figure 4.1), is is confirmed that in both groups the data is skewed to the right.

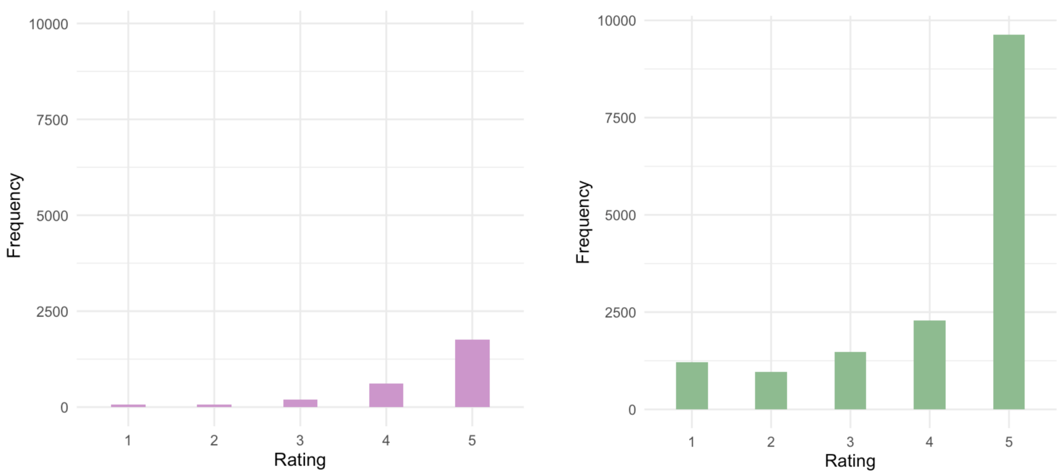

Figure 4.1: Frequency of Rating for sustainable products (left) and non-sustainable (right)

Chapter 5

Results

5.1 Topic Modeling

After a preliminary assessment of perplexity scores and an evaluation of the interpretability, the number of topics (k) for both datasets was set to 10. Afterwards, the document-topic prior and the topic-word prior were tunes using grid-search. For the sustainable products, the optimal values found were 0.4 for the document-topic prior and 0.025 for the topic-word prior, and for the non-sustainable products, the optimal values were 0.4 and 0.05, respectively. The perplexity scores after applying the LDA model with the parameter values specified were 170,41 and 170,61 for the sustainable and non-sustainable products, respectively.

Then, the top 10 words for each of the groups are obtained and used to give a name and an interpretation to each of the topics. The top words, as well as the given name for each topic, can be found in the following tables. For the sustainable products (Table 5.1), the first topic includes words such as “amazon” and “aware”, which is the name of the brand, and the word “brand” itself is also present, therefore, the first attribute is named Brand, referring to the overall image of the clothing line. The second topic is named Sleeves fit, since it includes words such as “sleev”, “arm” and “shoulder”, as well as words referring to the fit such as “long”, “length” or “short”. The third topic includes words like “dry”, “machin”, “washabl”, “water” and “polyest”, which indicate that this topic refers to the washing and maintenance of the product, which is why it it named Care. The fourth topic is named “Sustainability”, since it includes words related to the sustainable aspect of the products, such as “organ” from organic, “cotton”, “climat” and “neutral”, which is often found next to the word “carbon”. The next topic includes words like “layer”, “dress”, “jean”, as well as some weather words like “spring”, “summer” or “hot”, which suggest that this topic talks about when to wear this product and how, therefore it is named “Wear”. The following topic is named Size because most words refer to the sizing of the product, like “medium”, “small”, “larg”, “bigger” or “snug”. The next one is named Worth, because the top words seem to indicate whether the product and its price are worth it, for example with words like “price”, “recommend”, “worth” or “happi”. The next topic includes words such as “pink”, “green”, “dark” and “white”, therefore it is named Color. The following one includes words like “cut”, “seam”, “process” or “workmanship” that refer to how well the garment is made, which is why the topic is named Quality. And finally, the last topic is named Features because most words refer to special characteristics of the product, such as "pocket", "zipper", "hood" or "button".

Brand	**Sleeves fit**	**Care**	**Sustainability**	**Wear**	**Size**	**Worth**	**Color**	**Quality**	**Features**
amazon	sleev	dry	organ	layer	medium	colour	tee	cut	pocket
brand	long	polyest	cotton	dress	small	price	pink	seam	zipper
cloth	arm	machin	modal	spring	larg	design	green	process	front
item	length	dri	product	wool	run	recommend	hous	workmanship	inside
favorit	short	water	neutral	jean	overs	worth	dark	pleasant	button
awar	puff	recycl	sustain	summer	snug	rang	bra	smell	hood
check	shoulder	flat	tencel	knit	bigger	stitch	white	clean	warm
choic	hem	frame	certif	cardigan	slim	highli	cute	unpack	snap
piec	torso	washabl	climat	sweat	mean	happi	basic	fall	side
amaz	longer	low	manufactur	hot	refer	pleas	loung	unpleas	zip

Table 5.1: Top words (stemmed) and topic names for Amazon Aware dataset (sustainable products)

The first topic for the non-sustainable products (Table 5.2) is named Care, since it includes words referring to the washing and maintenance of the product, like "wash", "shrink", or "dry". The second topic includes words that refer to how the product fits overall, with words like "flatter", "cover", "waist" or "bodi", therefore it is named General fit. The third topic is similar to the previous one, because it also talks about the fit, but specifically for the sleeves in this case, therefore it is named Sleeves fit. The fourth topic includes words like "size" itself, "xl" or "smaller", which are very similar to the ones for the last topic, like "medium", "larg" or "small", therefore both topics are named Size. The fifth topic includes the brand name, with the words "amazon" and "essenti", but also words related to the price and quality of the product, so it most likely talks about whether the brand is worth buying, so it is named Brand worth. The next topic is named Material, since most words, like "cotton", "fabic", "material" or "feel", are related to this category. The next topic includes words such as "warm", "winter", "weather", "cozi" or "layer", which all refer to the climate or temperature, so the topic is named Weather. The following topic features words that are all specific characteristics of the product, like "pocket", "zipper" or "hood", so it is named Features. Finally, for the next topic, most of the words in the table are colors, so it is named Color.

Care	**General fit**	**Sleeves fit**	**Size**	**Brand worth**	**Material**	**Weather**	**Features**	**Color**	**Size**
wash	flatter	sleev	size	qualiti	cotton	warm	pocket	color	medium
shrink	bra	short	order	price	fabric	winter	zipper	blue	larg
wrink	cover	long	xl	amzon	feel	weather	hood	pink	small
dry	thin	length	larger	brand	neck	weight	water	pictur	lb
dress	hole	arm	smaller	essenti	materi	comfort	time	green	men
dryer	necklin	torso	run	product	100	light	insid	white	extra
jean	thumb	inch	xxl	surpris	made	soft	shrunk	navi	order
iron	clingi	longer	big	excel	turtleneck	layer	hand	red	review
casual	waist	waist	bigger	cloth	skin	cozi	think	black	size
work	bodi	bodi	usual	bui	felt	fleece	wash	dark	pound

Table 5.2: Top words (stemmed) and topic names for Amazon Essentials dataset (non-sustainable products)

In the following diagram (Figure 5.1), a comparison of analogous topics between sustainable and non-sustainable products is depicted. We observe some common topics between sustainable and non-sustainable products, like Sleeves fit, Care, Size, Color and Features. The topics Worth and Brand from the sustainable products together, are very similar to the topic Brand worth from the non-sustainable products. The main difference between the two groups is the presence of the Sustainability and the Quality topic in the sustainable brand. The non-sustainable brand seems to focus more on use of the product (with topics such as General fit and Weather), rather than on how the garment is made.

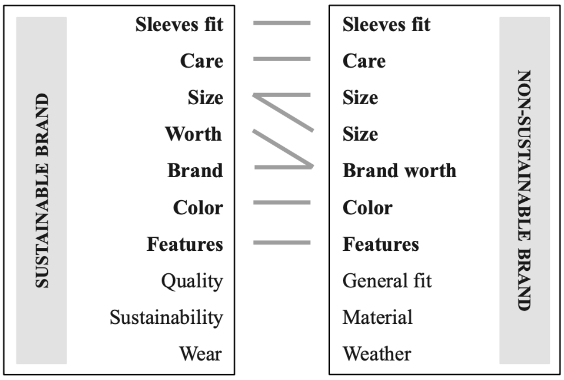

Figure 5.1: Relationship between topics in sustainable and non-sustainable products

5.2 Aspect-Based Sentiment Analysis

The average sentiment scores obtained for each of the product attributes through attribute-based sentiment analysis are found in the following tables (Table 5.3 and 5.4). The first, shows the sentiment scores, from highest to lowest, for the sustainable products and the second, shows the same information for the non-sustainable products. It can be noted that all the sentiment scores are positive, which indicates a positive sentiment associated with all attributes. This is not surprising because we previously observed that the average rating for both groups was above 4, indicating a general positive sentiment from customers. Additionally, the scores for the sustainable brand are higher than for the non-sustainable one, which is also inline with the information provided by the average ratings. This tells us that, in general, customers are happier with the product features of the sustainable brand than with the regular one, or at least, they talk more positively about them in the reviews. We can also observe that the product attribute with the highest sentiment score for both groups is Features, suggesting that customers usually praise the products for having certain features, such as pockets or zippers. We see that the feature Size has a much higher sentiment score in the non-sustainable products, 0.74, than in the sustainable ones, 0.53. Another attribute with relatively high sentiment scores in both groups of products is Color. And it is also worth noting that the Sustainability attribute has a relatively high score, 0.77, suggesting that customers often talk positively about the product being sustainable.

Topics for Sustainable Products	**Average Sentiment Score**
Features	0.84
Sleeves fit	0.8
Sustainability	0.77
Worth	0.76
Color	0.69
Quality	0.64
Brand	0.61
Wear	0.56
Care	0.55
Size	0.53

Table 5.3: Sentiment scores for each topic, or product attribute, of the sustainable brand

Topics for Non-Sustainable Products	**Average Sentiment Score**
Features	0.76
Size	0.74
Color	0.7
Brand worth	0.64
General fit	0.53
Care	0.46
Size	0.41
Material	0.4
Weather	0.37
Sleeves fit	0.19

Table 5.4: Sentiment scores for each topic, or product attribute, of the non-sustainable brand

The following diagrams (Figures 5.2 and 5.3) show examples of how two reviews, one from the sustainable brand and the other from the non-sustainable brand, are analyzed with the methodology of this section of the study. We start with the full review, then this review is broken into sentences. In this depiction, only the sentences where the model has identified a latent topic are kept. Next, the figure shows what attribute was identified by the topic modeling algorithm and finally the sentiment that was calculated for each of these sentences.

We can see that there are some inaccuracies in the topic identification, for instance in the second example review (Figure 5.3) the first sentence would be better summarized by the topic Material, however most of the sentences are very related to the topic identified by the model. In the sentiment classification, we also observe some inaccuracies, for instance the first sentence of the first review (Figure 5.2) is clearly positive, yet it is labeled as Negative. However, in general, the overall sentiment identified correlates with the meaning of the sentence.

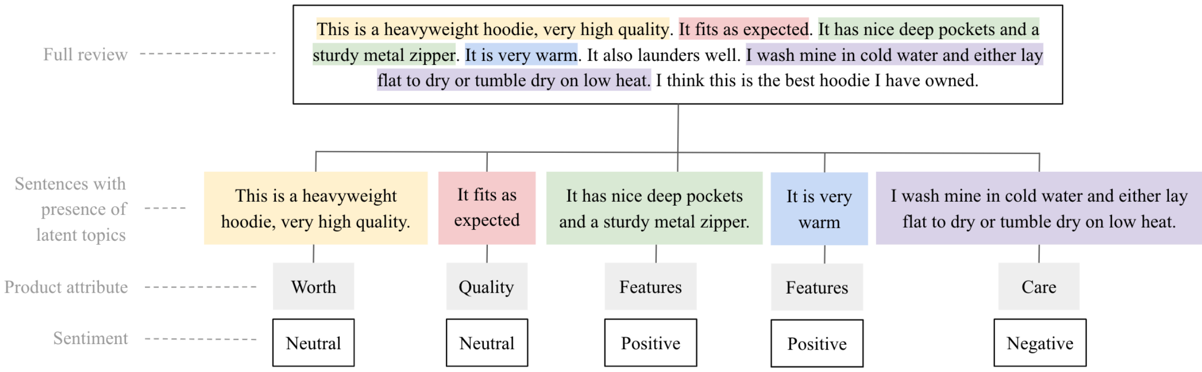

Figure 5.2: Example of Aspect-Based Sentiment Analysis for review number 1623 of the sustainable brand (Amazon Aware)

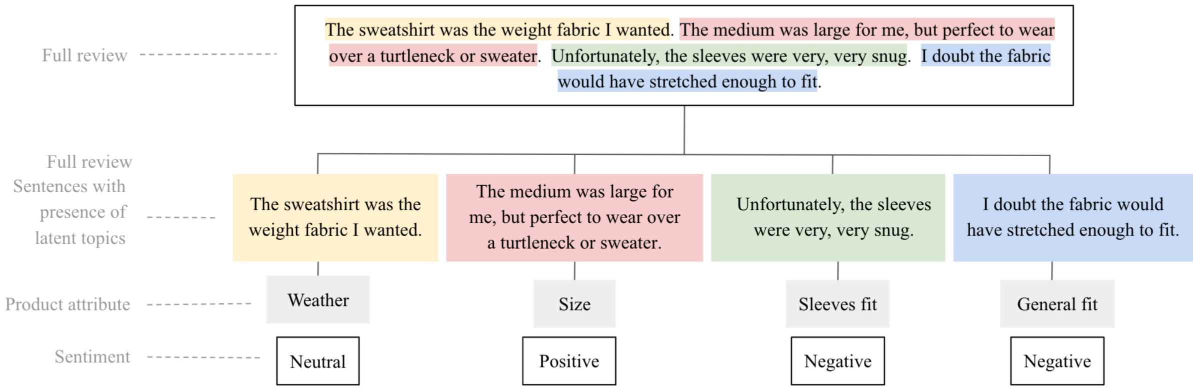

Figure 5.3: Example of Aspect-Based Sentiment Analysis for review number 4571 of the non-sustainable brand (Amazon Essentials)

5.3 Rating Prediction

In Figure 5.4 we can observe the plots with the values of the 10-fold cross validation performed to tune the Complexity Parameter (cp) for the Multiclass Classification Trees. The plot in the left, showing the results for the sustainable products, shows that the cp value resulting in the highest accuracy fit is 0.003, and on the right, the results for the non-sustainable products, which suggest that the best cp value is 0.005.

The variables used as predictors for the prediction of the overall rating are: the vectors with the probabilities of each topic for each review (10 in total) and the price of the product as a control variable. Due to high class imbalance in the target variable (rating), a sampling technique was used, more specifically, SMOTE (Synthetic Minority Oversampling Technique) was applied before modeling the decision tree.

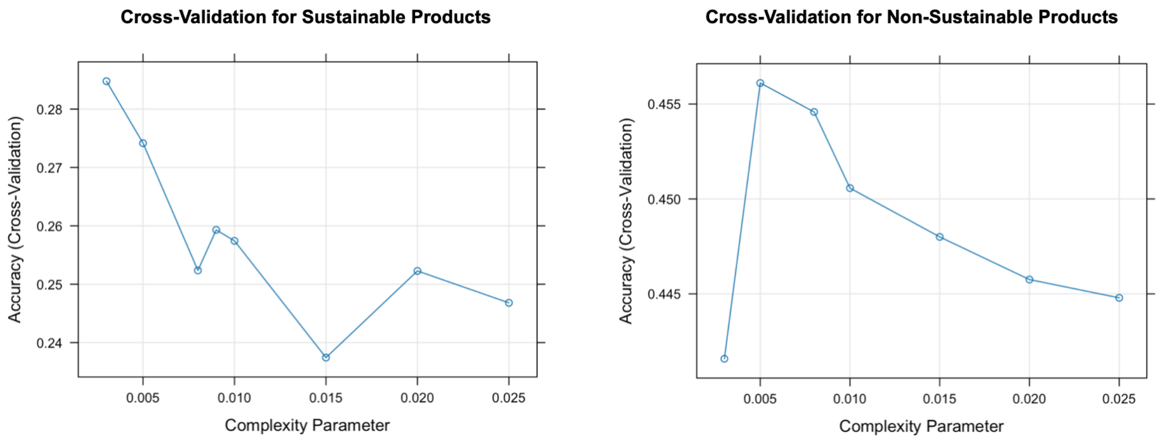

Figure 5.4: Cross-Validation values plot for sustainable brand (left) and for non-sustainable brand (right)

In order to assess the performance of the decision trees, an additional prediction method is applied to the same data. Specifically, a Multiclass Logistic Regression is implemented using analogous sampling techniques and the same predictor variables, which acts as a baseline method used for comparison. The accuracy scores for the Decision Trees and for the Logistic Regressions for the train and validation sets can be found in Table 5.5. We observe that for the sustainable brand, the accuracy of the Decision Tree is relatively low and that the accuracy for the test data is slightly lower than the one for the train data. However, this difference in accuracy does not seem to suggest overfitting. For the non-sustainable brand, the accuracy scores of the Decision Tree are higher and the values are more similar between the train and test sets. If we compare the accuracy scores with the ones obtained for the baseline model, we see that, although in general the values for the Decision Tree are higher, they do not seem to be significantly better because most 95 percent confidence intervals overlap. The only value where the difference is statistically significant is the one for the train accuracy for the non sustainable products. Despite the Decision Tree only slightly outperforming the baseline model, this model allows for a clear interpretation of attribute relevance through Variable Importance. Therefore, given the purpose of this study, the Multiclass Classification Tree is chosen as preferable.

		Sustainable products	**Non-sustainable products**
Decision Tree	*Accuracy (train data)*	0.3471 (0.3270, 0.3677)	0.4575 (0.4487, 0.4663)
	Accuracy (test data)	0.3178 (0.2787, 0.3590)	0.4652 (0.4475, 0.4829)
Baseline Model	*Accuracy (train data)*	0.3281 (0.3082, 0.3484)	0.3260 (0.3178, 0.3343)
	Accuracy (test data)	0.2602 (0.2236, 0.2995)	0.4363 (0.4188, 0.4539)

Table 5.5: Accuracy scores for multiclass classification tree and for multiclass logistic regression (baseline model), with the 95% confidence intervals in parentheses

Balanced accuracy is considered a more reliable evaluation metric for cases where there is a significant disparity in the number of samples in the different classes, i.e. class imbalance, and in this study, SMOTE was used to address this issue. However, to check its effectiveness, the balanced accuracy scores for each rating class are also computed and displayed in Tables 5.6 and 5.7. The balanced accuracy for the train data for the sustainable dataset is higher for classes 1 and 2. These are precisely the minority classes in the original dataset, therefore the higher accuracy might be a result of the synthetic observations generated by the sampling process, because in the test data we see that the balanced accuracy has similar values across all classes. If we focus on the non-sustainable dataset, we also observe a slightly higher accuracy for classes 1 and 2 in the train data and for classes 1, 2 and 5 in the validation set. However, in this group the values are, overall, more homogeneous across classes.

Sustainable products	**Rating values**				
	1	**2**	**3**	**4**	**5**
Balanced Accuracy for train data	0.7084	0.7480	0.6093	0.5847	0.5521
Balanced Accuracy for test data	0.4613	0.5601	0.5498	0.5321	0.5649

Table 5.6: Balanced Accuracy scores for multiclass classification tree

Non-sustainable products	**Rating values**				
	1	**2**	**3**	**4**	**5**
Balanced Accuracy for train data	0.5600	0.5598	0.5204	0.5174	0.5174
Balanced Accuracy for test data	0.5513	0.5924	0.5102	0.5102	0.6176

Table 5.7: Balanced Accuracy scores for multiclass classification tree

Next, the variable importance of the product attributes identified with LDA is assessed. The following plot (Figure 5.5) displays Variable Importance for both groups of products based on the Classification Tree, i.e. based on the impact of each variable in the overall rating score given by the reviewer. We observe that the most important features differ between product groups. For the sustainable products, the most important variable is Price, followed by Brand and Sustainability. On the contrary, for the non-sustainable products, Price is the least important variable and the most relevant are Features, Material and Size, followed by Weather and Brand Worth. Therefore, these results suggest that, even though consumers' opinions are highly influenced by the price of the product, the sustainability aspect of a product, as well as the brand image, will be the most important drivers of the overall satisfaction with a product. However, for the non-sustainable products, practical aspects such as features, size, and material will determine how the product is perceived and evaluated by customers.

A common method to determine the importance of a topic in textual data is to compute the average topic probability. However, as mentioned in the third chapter of this thesis, topic probability is not necessarily a proxy for topic importance. For this reason, the following scatter plots (Figures 5.6 and 5.7) provide a comparison of attribute importance based on the effect on overall rating (Variable Importance) and based on the presence of each topic in the reviews

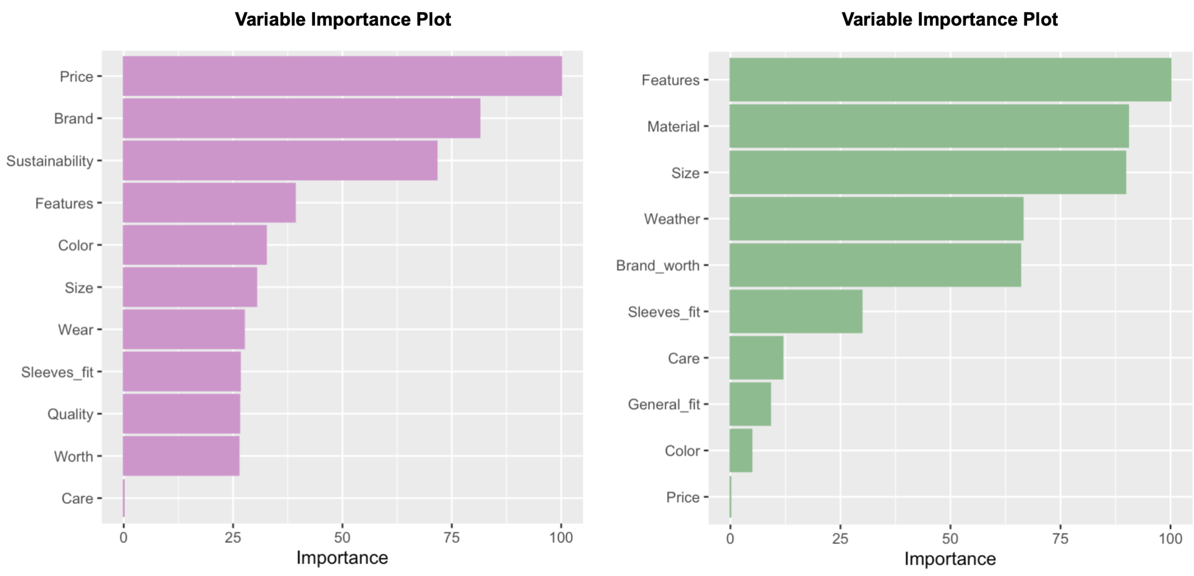

Figure 5.5: Feature Importance based on Variable Importance for the sustainable products (left) and the non-sustainable products (right)

(Average Topic Probability). Then, attributes can be described as impactful if their effect on the rating is significant, and as prevalent if they are mentioned frequently.

In Figure 5.6 we clearly see that for the sustainable brand, attribute relevance according to Variable Importance can be divided in three groups: those which are very relevant (Brand and Sustainability), those that are somewhat relevant (Sleeved fit, Wear, Color, Worth, Size, Features and Quality), and those which are not relevant at all (Care). However, while Sustainability is both impactful and prevalent and Care is neither, the relevance based on topic probability for the remaining attributes is inconistent. The attribute Brand, which is the most impactful, is not prevalent at all. And all the attributes which are only somewhat relevant have topic probabilities which vary from the second lowest (Sleeves fit) to the highest (Features and Quality).

In Figure 5.7, which displays the same measures for non-sustainable products, we identify two groups with different levels of impact on the overall rating: those with a high impact (Features, Size, Material, Weather and Brand Worth) and those with a lower impact (Sleeved fit, Care, General Fit and Color). Similarly to the sustainable products, these groups don't correspond with the importance based on prevalence. While Features is the attribute with the highest impact on rating, it is the least prevalent. The attributes Material and Brand worth are both impactful and prevalent, but Care, for instance, is at the same time one of the most prevalent and one of the least impactful.

Although it is worth noting that the values for the average topic probabilities are all clustered together between 0.075 and 0.12 for sustainable products and between 0.08 and 0.15 for the non-sustainable ones, there is a clear disparity between attribute relevance based on Variable Importance and based on Topic Prevalence. Therefore, the results seem to support the idea that the topics that are mentioned more often are not necessarily the ones that have the highest influence on overall rating.

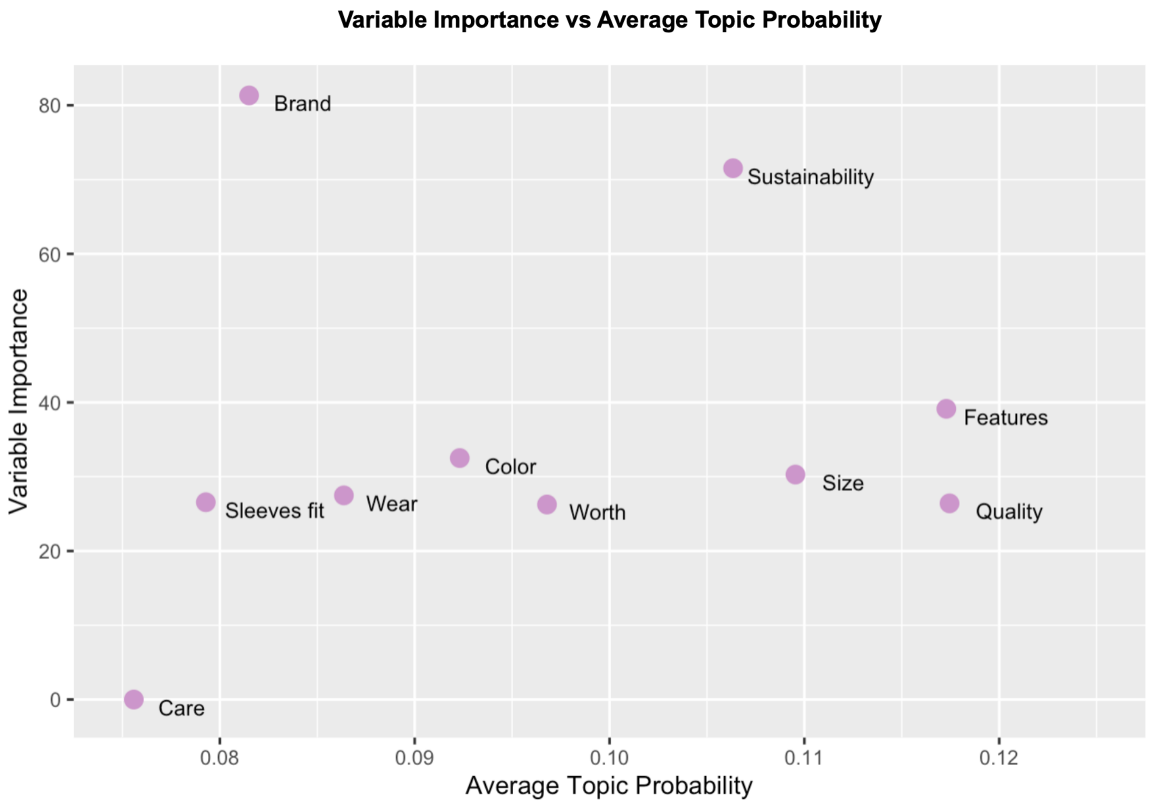

Figure 5.6: Comparison of attribute importance based on Variable Importance (y-axis) and Average Topic Probability (x-axis) for Sustainable Products

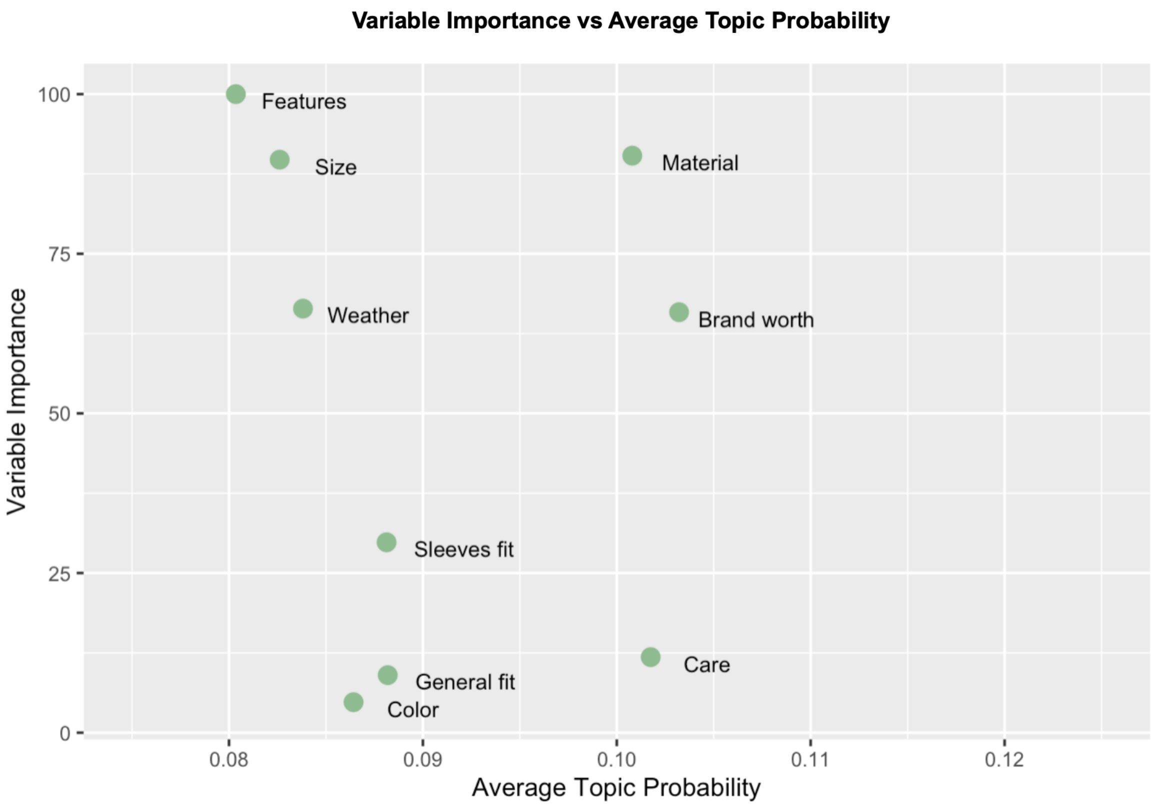

Figure 5.7: Comparison of attribute importance based on Variable Importance (y-axis) and Average Topic Probability (x-axis) for Non-Sustainable Products

Chapter 6

Discussion

6.1 Research Question

This study effectively used online reviews to provide insights on customers' feelings towards clothing products, more specifically, towards sustainable and non-sustainable clothing products. The methodology proposed provides a straightforward framework for summarizing review text, providing: first, a list of topics that can be linked to product attributes, second, the sentiment score for each of these attributes, and third, the importance of each attribute based on the effect on overall rating. The research questions that were presented at the beginning of this paper can be answered in the following way:

1. *Can online customer reviews be used to identify the most important attributes of clothing products and the customer's attitude towards each of them?* Yes, online customer reviews can serve as a useful source of information to derive the most important attributes in clothing. Using Latent Dirichlet Allocation (LDA), latent topics can be identified and then linked to specific product attributes looking at the top words for each topic and finding common themes. Then, the sentiment towards each of these attributes can be estimated using an Aspect-Based Sentiment Analysis algorithm, which revealed that the sentiment towards all product attributes is positive, and that, overall, it is more favorable towards sustainable products.

2. *Can the importance of each of these topics be derived by measuring their impact on the overall rating?* Yes, Variable Importance when predicting overall rating can be used as a measure of how much customers value each of the product attributes identified. This study showed that topic prevalence is not a proxy for topic importance, because not all attributes that are often mentioned in the reviews are then relevant for predicting overall rating, suggesting that even though consumers frequently mention some attributes in the reviews, these are not always relevant in the assessment of a product.

3. *What are the most important attributes valued by consumers when buying clothes online?* According to the results of this study, for sustainable clothing products, the most important attributes are Brand and Sustainability, with Price also being very relevant. However, for non-sustainable clothing products the attributes Features, Material and Size seem to be the most relevant.

4. *Are there significant differences in the most relevant features between sustainable and non-sustainable products?* Despite identifying common product attributes between both groups, there are relevant differences between the most important features in each of the groups. The features of the sustainable products are more focused on the sustainability aspect of the product and the image of the brand, whereas for the non-sustainable products the important variables seem to be more practical, such as what features does the product have, how it fits and what material is it made of.

6.2 Managerial Implications

The results of this study regarding sustainable products have significant managerial implications within the context of the Marketing Mix model, particularly impacting the Price, Promotion, and Product dimensions. In terms of Price, the study emphasizes that Price is a critical product attribute that influences customers' purchasing decisions. It is therefore clear that pricing strategies must balance the associated costs and desired margins, with consideration for the Price attribute's importance in influencing consumer preferences. The price of a sustainable product needs to be perceived as the cost of environmentally and socially conscious materials and manufacturing process, as well as a fair amount to pay for a high quality garment.

Regarding the Product dimension, the findings advocate for an attribute-focused design approach. Acknowledging the attributes most positively valued by consumers, as identified through sentiment analysis, becomes pivotal in shaping the product offering. At the same time, the attributes with the lowest scores provide improvement opportunities for upgrades and new product lines. Additionally, performing the same analysis on the brand's competitor's products will provide insights on how similar attributes are valued in other products or even make the company aware of attributes that their own products don't possess. Overall, aligning the product's attributes with consumers' preferences will aid in positioning the product in the market, increasing its appeal and fostering customer loyalty.

Furthermore, the Promotion aspect can leverage the study's insights to create compelling communication strategies. Effective promotion should not only emphasize the product's sustainable aspects but also contribute to the establishment of a strong brand image. The significance of the Brand attribute emphasizes the importance of fostering customer trust and belief in the brand's sustainability claims. By conveying the rationale behind the product's sustainable label and articulating its value proposition, promotional efforts can create a compelling narrative that resonates with conscious consumers.

In conclusion, the study's integration with the Marketing Mix model provides actionable recommendations for managers of sustainable clothing brands. The relationship between Price, Promotion, and Product attributes, aligned with customer preferences, lays the foundations for strategic decision-making in marketing and product development.

6.3 Academic Implications

This thesis adds literature on several techniques related to the field of Natural Language Processing of text data obtained from online consumer reviews. First, it contributes to the extensive literature regarding the application of Latent Dirichlet Allocation (LDA), proving once again its effectiveness as an intuitive and simple model. Second, it contributes to the literature on Aspect-Based Sentiment Analysis (ABSA), which at the moment mainly focuses on complex models, such as Neural Networks, that are highly accurate but its application is intricate and its interpretability is very limited. This study uses an algorithm that doesn't require labeled data or a complicated model tuning process, which makes it fast, easily applicable and straightforward to comprehend. Additionally, the ABSA algorithm proposed can be easily improved and tailored to specific requirements by applying a sentiment analysis method that caters to the specific needs of the researcher. Third, this study reaffirms the findings of Chakraborty et al. (2022), providing further evidence that attribute importance extracted from online reviews should be measured based on the impact on the rating, rather than with the frequency of said attribute, because topic probability is not an adequate proxy for topic importance. And finally, the main contribution of this study, is the application of all the aforementioned methods in the online sustainable clothing industry, a field with a limited number of studies dedicated to it.

6.4 Limitations and Further research

This study has several limitations, which can be divided into constraints regarding data and regarding the methodology used. The first limitation related to data collection is the limited amount of online reviews available for the sustainable products, in comparison to the reviews available for regular products. As a consequence, the sample might not be big enough to properly train the models and capture a full spectrum of customer opinions and topics, potentially leading to a lower degree of generalizability of the findings. Second, the data used in this study was not labeled in terms of topics and sentiments, which restricted the methods that could be used for topic modeling as well as for sentiment analysis. Regarding the limitations of the methodology used, a drawback of the method used for topic modeling is that it is a bag-of-words approach, and as such, it does not account for sentence structure. This limits the ability of the model to correctly interpret complex sentences and capture connections and links between words.

Finally, related to the limitations that this paper has, some areas for future research are proposed. A possible approach would be to apply a similar methodology as the one used in this study to a broader set of clothing products or even to other sustainable product industries, to see if the findings can be generalized. Another area for future research would be to apply a more complex topic modeling algorithm that is able to interpret complex sentences, such as state-of-the-art language models like BERT, which is able to capture semantic relationships and dependencies between words within a sentence thanks to its bidirectionality. And finally, a possible improvement would be to predict the overall rating using a larger set of variables, such as more control variables, as well as additional information extracted from the textual data itself, to improve the model fit and be able to extract variable importance more accurately.

References

Alwitt, L. F. & Pitts, R. E. (1996). Predicting purchase intentions for an environmentally sensitive product. *Journal of Consumer Psychology*, *5*(1), 49-64. Retrieved from https://doi.org/10.1207/s15327663jcp0501_03

Archak, N., Ghose, A. & Ipeirotis, P. (2011). Deriving the pricing power of product features by mining consumer reviews. *Management Science*, *57*(8), 1485-1509. Retrieved from https://doi.org/10.1287/mnsc.1110.1370

Bangsa, A. B. & Schlegelmilch, B. B. (2020). Linking sustainable product attributes and consumer decision-making: Insights from a systematic review. *Journal of Cleaner Production*, *245*, 118902. Retrieved from https://doi.org/10.1016/j.jclepro.2019.118902

Battenfeld, D., Hörisch, J., Jacobs, K. & Petersen, L. (2018). Green thinking but thoughtless buying? an empirical extension of the value-attitude-behaviour hierarchy in sustainable clothing. *Journal of Cleaner Production*, *203*, 1155-1169. Retrieved from https://doi.org/10.1016/j.jclepro.2018.07.320

Büschken, J. & Allenby, G. M. (2020). Improving text analysis using sentence conjunctions and punctuation. *Marketing Science*, *39*(4), 727-742. Retrieved from https://doi.org/10.1287/mksc.2019.1214

Chakraborty, I., Kim, M. & Sudhir, K. (2022). Attribute sentiment scoring with online text reviews: Accounting for language structure and missing attributes. *Journal of Marketing Research*, *59*(3), 600-622. Retrieved from https://doi.org/10.1177/00222437211052500

Comission. (2013). *Attitudes of europeans towards building the single market for green products.* Retrieved from https://europa.eu/eurobarometer/surveys/detail/1048

Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. (2019). *Bert: Pre-training of deep bidirectional transformers for language understanding.*

Essoussi, L. H. & Linton, J. (2010). New or recycled products: how much are consumers willing to pay? *Journal of Consumer Marketing*, *27*(5), 458-148. Retrieved from https://doi-org.eur.idm.oclc.org/10.1108/07363761011063358

Forman, C., Ghose, A. & Wiesenfeld, B. (2008, 09). Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. *Information Systems Research*, *19*, 291-313. Retrieved from https://doi.org/10.2139/ssrn.1026893

Ghose, A. & Ipeirotis, P. (2011, 10). Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *Knowledge and Data Engineering, IEEE Transactions on*, *23*, 1498-1512. Retrieved from https://doi.org/10.1109/TKDE.2010.188

Gomes, S., Lopes, J. M. & Nogueira, S. (2023). Willingness to pay more for green products: A critical challenge for gen z. *Journal of Cleaner Production*, *390*, 136092. Retrieved from https://doi.org/10.1016/j.jclepro.2023.136092

Jia, S. S. (2018). Behind the ratings: Text mining of restaurant customers' online reviews. *International Journal of Market Research*, *60*(6), 561-572. Retrieved from https://doi.org/10.1177/1470785317752048

Jung, H. J., Choi, Y. J. & Oh, K. W. (2020). Influencing factors of chinese consumers' purchase intention to sustainable apparel products: Exploring consumer "attitude–behavioral intention" gap. *Sustainability*, *12*(5). Retrieved from https://www.mdpi.com/2071-1050/12/5/1770

Kovacs, I. & Keresztes, E. R. (2022). Perceived consumer effectiveness and willingness to pay for credence product attributes of sustainable foods. *Sustainability*, *14*(7). Retrieved from http://dx.doi.org/10.3390/su14074338

Lancaster, K. J. (1966). A new approach to consumer theory. *Journal of Political Economy*, *74*(2), 132–157. Retrieved 2023-08-07, from http://www.jstor.org/stable/1828835

Laroche, M., Bergeron, J. & Barbaro-Forleo, G. (2001). Targeting consumers who are willing to pay more for environmentally friendly products. *Journal of Consumer Marketing*, *18*, 503-520. Retrieved from https://api.semanticscholar.org/CorpusID:16019909

Liang, H., Sun, X., Yunlei, S. & Gao, Y. (2017, 12). Text feature extraction based on deep learning: a review. *EURASIP Journal on Wireless Communications and Networking*, *2017*. Retrieved from https://doi.org/10.1186/s13638-017-0993-1

Lin, P. C. & Huang, Y. H. (2012). The influence factors on choice behavior regarding green products based on the theory of consumption values. *Journal of Cleaner Production*, *22*(1), 11-18. Retrieved from https://doi.org/10.1016/j.jclepro.2011.10.002

Lin, Y., Fu, Y., Li, Y., Cai, G. & Zhou, A. (2021). Aspect-based sentiment analysis for online reviews with hybrid attention networks. *World Wide Web*, *24*(4). Retrieved from https://doi.org/10.1007/s11280-021-00898-z

Luchs, M. G. & Kumar, M. (2017). "yes, but this other one looks better/works better": How do consumers respond to trade-offs between sustainability and other valued attributes? *Journal of Business Ethics*, *140*(3), 567–584. Retrieved 2023-07-29, from https://doi.org/10.1007/s10551-015-2695-0

Luchs, M. G., Naylor, R. W., Irwin, J. R. & Raghunathan, R. (2010). The sustainability liability: Potential negative effects of ethicality on product preference. *Journal of Marketing*, *74*(5), 18-31. Retrieved from https://doi.org/10.1509/jmkg.74.5.018

Mahadevan, A. & Arock, M. (2019). Review rating prediction using combined latent topics and associated sentiments: an empirical review. *Service Oriented Computing and Applications*, *14*, 19-34. Retrieved from https://doi.org/10.1007/s11761-019-00278-6

Medhat, W., Yousef, A. H. & Mohamed, H. (2014, 05). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, *5*. Retrieved from https://doi.org/10.1016/j.asej.2014.04.011

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. Retrieved from https://doi.org/10.1287/mksc.1120.0713

Noori, B. (2021). Classification of customer reviews using machine learning algorithms. *Applied Artificial Intelligence*, *35*(8), 567-588. Retrieved from https://doi.org/10.1080/08839514.2021.1922843

Pang, B., Lee, L. & Vaithyanathan, S. (2002, July). Thumbs up? sentiment classification using machine learning techniques. In *Proceedings of the 2002 conference on empirical methods in natural language processing (EMNLP 2002)* (pp. 79–86). Association for Computational Linguistics. Retrieved from https://aclanthology.org/W02-1011

Park, H. J. & Lin, L. M. (2020). Exploring attitude–behavior gap in sustainable consumption: comparison of recycled and upcycled fashion products. *Journal of Business Research*, *117*, 623-628. Retrieved from https://doi.org/10.1016/j.jbusres.2018.08.025

Peattie, K. & Crane, A. (2005). Green marketing: Legend, myth, farce or prophesy? *Qualitative Market Research*, *8*, 357-370. Retrieved from https://doi.org/10.1108/13522750510619733

Prakoso, A. A., Yananta, B. W., Setyawan, A. F. & Muljono. (2018). A lexicon-based sentiment analysis for amazon web review. In *2018 international seminar on application for technology of information and communication* (p. 503-508). Retrieved from https://doi.org/10.1007/978-3-030-37218-7_136

Rex, E. & Baumann, H. (2007). Beyond ecolabels: what green marketing can learn from conventional marketing. *Journal of Cleaner Production*, *15*(6), 567-576. Retrieved from https://doi.org/10.1016/j.jclepro.2006.05.013 (Sustainable Production and Consumption: Making the Connection)

Rokka, J. & Uusitalo, L. (2008). Preference for green packaging in consumer product choices – do consumers care? *International Journal of Consumer Studies*, *32*(5), 516-525. Retrieved from https://doi.org/10.1111/j.1470-6431.2008.00710.x

Sievert, C. & Shirley, K. (2014, June). LDAvis: A method for visualizing and interpreting topics. In *Proceedings of the workshop on interactive language learning, visualization, and interfaces* (pp. 63–70). Baltimore, Maryland, USA: Association for Computational Linguistics. Retrieved from https://doi.org/10.3115/v1/W14-3110

Srinivasan, V. (1979). Network models for estimating brand-specific effects in multi-attribute marketing models. *Management Science*, *25*(1), 11-21. Retrieved from https://doi.org/10.1287/mnsc.25.1.11

Sudbury-Riley, L. & Boltner, S. (2010). Fashion marketing and the ethical movement versus individualist consumption: Analysing the attitude behaviour gap. *Advances in consumer research. Association for Consumer Research (U.S.)*, 163-168.

Tang, F., Fu, L., Yao, B. & Xu, W. (2019). Aspect based fine-grained sentiment analysis for online reviews. *Information Sciences*, *488*, 190-204. Retrieved from https://doi.org/10.1016/j.ins.2019.02.064

Tirunillai, S. & Tellis, G. J. (2012). Does chatter really matter? dynamics of user-generated content and stock performance. *Marketing Science*, *31*(2), 198-215. Retrieved from https://doi.org/10.1287/mksc.1110.0682

Wiederhold, M. & Martinez, L. F. (2018). Ethical consumer behaviour in germany: The attitude-behaviour gap in the green apparel industry. *International Journal of Consumer*

Studies, *42*(4), 419-429. Retrieved from https://doi.org/10.1111/ijcs.12435

Xu, L., Liu, J., Wang, L. & Yin, C. (2018). Aspect based sentiment analysis for online reviews. In *Advances in computer science and ubiquitous computing* (pp. 475–480). Singapore: Springer Singapore.

Yazdanpanah, M., Forouzani, M. & Hojjati, M. (2015). Willingness of iranian young adults to eat organic foods: Application of the health belief model. *Food Quality and Preference*, *41*, 75-83. Retrieved from https://doi.org/10.1016/j.foodqual.2014.11.012

Zhang, X. & Dong, F. (2020). Why do consumers make green purchase decisions? insights from a systematic review. *International Journal of Environmental Research and Public Health*, *17*(18). Retrieved from http://dx.doi.org/10.3390/ijerph17186607

Zhang, Y., Du, J., Ma, X., Wen, H. & Fortino, G. (2021, 09). Aspect-based sentiment analysis for user reviews. *Cognitive Computation*, *13*, 1114–1127. Retrieved from https://doi.org/10.1007/978-3-030-37218-7_136