ERASMUS UNIVERSITY ROTTERDAM
ERASMUS SCHOOL OF ECONOMICS
MSc Economics & Business
Specialization Data Science & Marketing Analytics

An image is worth a thousand words

A machine learning approach to improve customer engagement on social media for eco-friendly skincare brands

Author: Lize Appel
Student number: 509714
Thesis supervisor: Dr. A Tetereva
Second reader: T.B.A.
Finish date: July 2023

Abstract

With the rise of social media, people are more connected than ever. From all over the world, and of all ages, people can connect and engage. But there is more, users can also create content and encourage others to do or buy something. This makes social media a good marketing tool. At the same time, consumers are becoming more aware of the impact that their purchasing decisions have on the environment. Consumers are shifting more to sustainable and eco-friendly products, trying to bring less harm to others, animals, and nature. These two major changes, marketers using social media more often as marketing tool and consumers shifting towards more sustainable products, are subject of this study. This paper aims to find what eco-friendly brands should feature in their social media marketing. Data is obtained by web scraping Instagram pages of five eco-friendly skincare companies for a period of 6 months. Image features are extracted using convolutional neural networks. The relationship between the image features and customer engagement is uncovered using machine learning algorithms. The results indicate that monotonic photos, including hands and packaged products, that are related to influencers and are posted for advertisement purposes, receive the highest customer engagement. These findings suggest that marketeers of eco-friendly skincare products should use little color and focus on the placement of hands and products over faces. This is contrary to the findings for non-eco-friendly products, which are best promoted using color and faces. When promoting products, regardless of being eco-friendly, relating influencers to your brand is beneficial to your campaign. The comparison between eco-friendly and non-eco-friendly marketing strategies is what makes this paper valuable, for both academics and practice.

Keywords: Machine learning, web scraping, image classification, customer engagement, social media.
JEL Classifications: C40, C45, C53, M31.

Table of contents

1 Introduction

This study finds that eco-friendly products are best promoted on social media using monotonic images that feature invisible consumers. Supplementary, posting for selling purposes and associating your brand to influencers, is also recommended based on the findings of this paper. The results of this paper, on how to market eco-friendly products best, differ from the findings for non-eco-friendly products. Non-eco-friendly products are best promoted using color and visible consumers. Posts that entertain social media users, perform better for non-eco-friendly products. There is a similarity between the marketing of eco-friendly products versus non-eco-friendly products, relating influencers to your brand or post positively influences customer engagement for both. The results are found using machine learning models, which predict customer engagement on social media. With the use of interpretation methods, the importance of the different features and their relationship to customer engagement is measured.

Conscious consumerism is an emerging trend in which consumers prioritize sustainability, human rights and social responsibility when making purchasing decisions. Sustainable products have a 32% growth share and dominate 17% of the market already. Products that are marketed as sustainable grow 2.7 times faster than those that are not (Ruiz, 2023). However, many people believe that marketing and sustainability are incompatible since they are each other's opposites. One is about selling more, while the other is about consuming less. There is growing interest in the relationship between the two. Charter, Peattie, Ottman & Polonsky (2002) already recognized that sustainability is going to have a big influence on marketing. Being sustainable requires far more than creating and producing sustainable. Sustainability takes marketers outside their traditional frame, having to adapt to consumer's renewed preferences. The last decade this change in marketing has been visible, store shelves have been filled with greener looking products and cosmetics are promoted more often as being cruelty free.

Peter Fisk stresses the value of researching sustainability in relation to marketing, arguing that sustainability is one of the key trends shaping marketing (Jones, Clark-Hill, Comfort, & Hillier, 2008). According to Schaefer (2005), the sustainable development is the most difficult problem that marketing is currently facing. Jones et al. (2008) acknowledge that the relationship between marketing and sustainability is going to be high on the agenda for years to come. Indeed, many researchers have focused on capturing the relationship between sustainability and marketing. Kong, Witmaier & Ko (2021) study how and when to communicate sustainability. Their findings indicate that sustainable communication is most effective for non-luxury brands in a cultural setting. Banytė, Brazionienė & Gadeikienė (2010) identify the new, so called green, consumer. Their results indicate that green consumers are mostly female of the age 30 to 44, have higher education and receive higher than average income. Barbarossa & de Pelsmacker (2016) examine the different purchasing patterns of green and non-green consumers. Altruistic motives are more important for green consumers, while negative ego-centric motives affect non-green consumers more. Banytė et al. (2010) & Barbarossa & de Pelsmacker (2016) both touch on an interesting subject of research. With the rise of green consumption, a new group of consumers emerged. These green consumers may be different in their needs and preferences. It is interesting to find in what ways their needs differ from those of non-green consumers. Accompanying, if these new customer needs ask for different marketing strategies. This is an interesting topic that has been little subject of research. Dos Santos, de Brito Silva, da Costa & Batista (2023) did study how green products should be marketed. For vegan cosmetics they find that digital influencers help shaping purchasing intentions. However, the paper does not link back to non-vegan cosmetics. This is a missed opportunity since the comparison would be interesting. Yan, Hyllegard & Blaesi (2010) also miss this opportunity. The researchers examined the effect of environmentally friendly marketing claims, through brand name and message explicitness, on student's attitudes towards advertisement. Especially message explicitness is a useful way to market eco-friendly products. The researchers guess that message explicitness may also raise awareness and acknowledge the benefits of eco-friendliness for those who have lower environmental commitment. Like dos Santos et al. (2023), the researchers do not link back to results found for non-eco-friendly products. Presumably, brand name placement and information provision does not change advertisement attitudes the same way for non-eco-friendly products. Like dos Santos et al. (2023) and Yan et al. (2010), other researchers have investigated what to feature in eco-friendly advertisement (Testa, Iraldo, Vaccari & Ferrari, 2013; Lu, Bock & Josep, 2013; Song, Qin & Qin, 2020). None of them compared their results to those found for non-eco-friendly products. This paper fills the gap in the literature by comparing the findings for eco-friendly products to those for non-eco-friendly products.

This paper studies how the marketing of eco-friendly products differs from that of non-eco-friendly products. This paper uses customer engagement to measure the performance of social media campaigns for eco-friendly skincare companies. Green skincare or eco-friendly skincare uses ingredients from botanical sources and that are manufactured by preserving the integrity of the ingredients (Chin et al., 2018). Based on the findings of this study, recommendations for promoting eco-friendly skincare products are presented. Additionally, a comparison to the results found for non-ecofriendly products is made. The corresponding research question is:

What attributes in social media posts lead to higher customer engagement for eco-friendly skincare products?

This study makes important academic and practical contributions. The outcome of this paper is valuable for companies selling eco-friendly products. Eco-friendly companies can improve their social media marketing by using more monotonic images, featuring invisible consumers, and relating the posts to influencers. This would result in higher customer engagement on social media, which translates to more sales (Gill et al., 2017).

The paper also adds value to academics. The paper fills a gap in the literature by comparing the marketing of eco-friendly products to the marketing of non-eco-friendly products. Second, the literature is enriched by studying customer engagement for eco-friendly and sustainable brands. Lastly, as few papers studied customer engagement using advanced machine learning methods, this study adds value by doing so.

2 Literature review

The last decades, consumers are becoming more conscious of their ecological footprint. This trend of environmental together with health awareness, led consumers to change to natural cosmetics more (Amberg & Fogarassy, 2019). Women believe that natural cosmetics harmonize their self-image, lower health risks, and stand for feminism (Pudaruth, Juwaheer & Seewoo, 2015). Women not only chance to natural cosmetics but try to adapt a more sustainable lifestyle in general. Kim & Chung (2011) find that a more organic lifestyle is reflected in an individual's consumption pattern and that past experiences with other organic products have a significant positive impact on the purchase intention of organic personal care products. Sukato & Elsey (2009) find that not only women, but also men condition their purchase behavior of skincare on self-image, beliefs, and normative influences.

This shift towards more green consumption has been subject of research. Hsu, Chang & Yansritakul (2017) research what drives consumers to buy green skincare products. It is mostly the subjective norm that has a significant positive effect on purchase intention. Also, consumers with a high perceived behavioral control are the ones buying green skincare more frequently. Jaini, Quouquab, Mohammad & Hussin (2020) also research what influences the green purchase behavior of cosmetic consumers. The study presented personal norms, altruistic value, and hedonic value as the key drivers of green purchasing behavior. Johri & Sahasakmontri (1998) find that Oriental Princess performs better in Thailand than the Body Shop, because of the materials and ingredients they use. This is consistent with the findings of Johri & Sahasakmontri (1998), because consumers care more about the environmental aspect rather than package and marketing. Prothero & McDonagh (1992) and Grappe, Lombart, Louis & Durif (2021) also support the findings of Jaini et al. (2020) and Johri & Sahasakmontri (1998). Moslehpour, Chaiyapruk, Faez & Wong (2021) surprisingly find the contrary, namely that environmental concerns are not the main driver of green purchasing intentions for personal care products. The paper argues that people's attitude towards green packaging and green marketing has the strongest effect on purchasing intentions. A positive attitude towards these two components, leads to a stronger purchasing intention. Dewi, Avicenna & Meideline (2020) study the purchase intentions of Instagram users for green skincare products. They conclude that once consumers consider environmentally friendly products, they may seek to identify green products. After seeking information, they prioritize the products over others.

Samaraweera, Sims & Homsey (2021) dove deeper into promoting green products. They study consumers of a grocery store in a field setting to test if green and nature images increase willingness to pay. Surprisingly, the results show that customers are willing to pay more for white-toned labels. This creates a clean and high-quality look, which better conveys the story of sustainability. But it also builds to the idea of sustainable products being healthier. The paper finds that nature images do not have a significant effect on willingness to pay. However, the findings of Samaraweera et al. (2021) are refuted by Lavuri, Jabbour, Grebinevych & Roubaud (2022). Lavuri et al. (2022) find that green ads and green brand images positively influence consumers' attitude towards green beauty products. Seo & Scammon (2017) confirm that using the color green on packaging enhances the message of environmental awareness of brands. The authors do note that this perception can be misused. Xue & Muralidharan (2015) confirm that using green visuals leads to better perception of brand's environmental efforts and that it creates positive advertisement responses. However, they conclude that textual sustainability claims are even more effective.

Takahashi (2021) examines if information provision about eco-friendly coffee improves purchasing behavior. He finds that informing consumers in social places increases coffee sales by 7%. This effect is found to be significant. Takahashi argues that this could be due to consumers trying to build a green reputation for others. This experiment supports the findings of Hsu et al. (2017), who suggest that it is mostly the subjective norm that influences green purchase intentions positively.

Todd (2004) examines the best green marketing strategy for eco-friendly personal care products. His recommendations fall in line with the idea that using sustainability claims helps increasing purchase intentions. Todd (2004) advises companies to emphasize the eco costs and impact on ecological integrity of their products to consumers to optimize marketing efforts.

The difference between the marketing of green versus non-green products has been little subject of research before. Ewe & Tjiptono (2023) conclude that when consumers are more familiar with an eco-friendly brand than a non-eco-friendly brand, that their buying intention and willingness to pay are significantly higher than for non-eco-friendly brands. Tripathi & Pandey (2018) study the difference in pricing eco-friendly products relative to non-eco-friendly products. The researchers find that consumers prefer zero-ending prices for green products. Round digits enhance the perception of green products being high-quality products. Contrary, consumers prefer nine-ending prices for low-priced and utilitarian products. Although the difference between the marketing of eco-friendly versus non-eco-friendly products has been subject of research before, there are still gaps in the literature. This study tries to fill these gaps by comparing social media marketing strategies for eco-friendly products versus non-eco-friendly products. This study examines which features to include in social media marketing for eco-friendly products by measuring what drives customer engagement on social media. Multiple studies have examined what drives customer engagement on social media for non-eco-friendly products. These results will be compared to the outcome of this study, to identify the differences and similarities between social media marketing for eco-friendly versus non-eco-friendly products.

Social media has completely changed the way people communicate and engage, which makes it an interesting subject of research. Social media transformed consumers from passive observers of marketing to active participants who can interact with the content. Consumers can even co-create through these interactions (Dolan, Conduit, Frethey-Bentham, Fahy & Goodman, 2019). This led to major changes in the world of marketing. Social media is a marketing tool that can be used for other means than selling. Brands can create awareness, extend the customer experience, or anticipate on the customer's needs. The definition of customer engagement also extended with the rise of social media because consumers have new ways of interacting with companies. On social media, customers can interact and engage with brands by liking, commenting, or sharing their social posts (Aichner, 2019). Before diving deeper into the already existing literature on customer engagement, the term "customer engagement" is conceptualized.

Hollebeek, Glynn & Brodie (2014) define customer engagement as "consumers' positively valenced brand-related cognitive, emotional and behavioral activity during or related to consumer/brand interactions". Leek, Houghton & Canning (2019) refine this to "a psychological state resulting from specific interactive episodes that a customer experiences with a focal agent or object". Calder, Malthouse and Maslowska (2016) have a more expanded version of the definition of Leek et al. (2019): "psychological state that occurs by virtue of interactive, co-creative customer experiences with a focal agent/object, under a specific set of context-dependent conditions, and exists as a dynamic, iterative process in which other relational concepts are antecedents and/or consequences". Shawky, Kubacki, Dietrich & Weaven (2020) define customer engagement as "an interactive and dynamic process between multiple actors that supports the development of enduring and long-term relationships". Barger, Pelteir & Schultz (2016) define customer engagement on social media: "a set of measurable actions that consumers take on social media in response to brand-related content". This paper follows the definition of Barger et al. (2016), since it is simple and clear.

Van Doorn, Lemon, Mittal, Nass, Pick, Pirner & Verhoef (2010) discuss the value of customer engagement marketing as research topic for students. It is important to understand what customers are interested in and how to maintain engagement. To maintain or improve customer engagement, a measure for customer engagement needs to be established. In the already existing literature, there is dissensus on how to measure customer engagement on social media. Cortez, Johnston, Ghosh & Dastidar (2023) measure customer engagement as five components. The first one is called the impressions; this is the baseline for all other forms of customer engagement. An impression is simply the act of paying attention to a post. The second component is reactions, which has six types of behavior, a sign of like, celebration, support, love, insightfulness, or curiosity. The third component is a share, which is the opportunity to share the post of another user. Clicks refer to the action of a user clicking on a post. And lastly, commenting. This is a customer responding to a post. Further, the paper studied the short- and long-term effects of types of posts, website visits and new followers. The paper finds that sales post increases the number of followers and social posts increase engagement.

Zhao, Zhang, Ming, Niu & Wang (2023) measure engagement as likes, shares, sentiment, and cognitive engagement. Cognitive engagement indicates customers' cognitive awareness when interacting with brands (Hollebeek, Glynn & Brodie, 2014). Sentiment measures the emotional engagement of customers through reviews/comments. The paper studies how image richness affects customer engagement using deep learning. The results indicate that image richness is positively related to emotional engagement (sentiment) and behavioral engagement (likes and shares), but negatively to cognitive engagement.

Corone et al. (2021) measure customer engagement as likes and comments placed under Instagram photos. The paper finds that the theme of slogans significantly impacts the effect of social presence on customers' engagement in terms of likes, but not comments. Their findings show that text can complement images.

Zhang, Lee, Singh & Srinivasan (2017) investigate if quality of property images matters for the occupancy of Airbnb residences. Airbnb properties with verified images, compared to properties with images taken by the host, had a 9% higher occupancy, meaning that the quality of images matters. The paper used convolutional neural networks to classify the quality of the photos. The paper shows that the quality of pictures makes a difference in the way customers review products. Li & Xie (2020) also find that high-quality photos lead to higher customer engagement. Their findings support the results of Zhang et al. (2017). Li & Xie (2020) also find that colorfulness in images leads to higher customer engagement for airlines and sport utility vehicle companies.

Hartmann et al. (2019) measure customer engagement as likes and comments on Instagram and Twitter. The paper studies a quarter million brand-images of 185 brands. Convolutional neural networks are used to divide the brand-images into three categories: consumer selfies (featuring a visible customer), brand selfies (invisible customers holding a branded product) and pack shots. The study finds that consumer selfies, featuring consumers' faces, generate the highest level of customer engagement. However, brand selfies lead to higher purchasing intentions than consumer selfies.

Lou, Tan & Chen (2019) follow both Corone et al. (2021) and Hartmann et al. (2019) by defining customer engagement on Instagram as the number of likes and comments. The paper finds that influencer-promoted ads enjoy significantly higher engagement than brand-promoted ads for the top 50 apparel companies in the United States. Their findings suggest that influencer association is an important factor that needs to be accounted for.

Aichner (2019) measures customer engagement as likes, comments, and shares. The paper studies if certain posts lead to higher customer engagement. The paper finds that social media users like and comment independently of the content of the post. This is contrary to the findings of Liu et al. (2019), who find that post related to entertainment, interaction, and trendiness result in higher customer engagement. The findings of Corone et al. (2021) and Luarn et al. (2015) are also opposite to the findings of Aichner (2019). Corone et al. (2021) find that the theme of a post impacts the number of likes and comments. Luarn et al. (2015) make it more concrete, according to their paper remuneration posts receive more likes.

Most studies measure customer engagement on social media as likes, comments, and shares. Unfortunately, data on sharing is not publicly available for Instagram and therefore not considered in this paper. Customer engagement on Instagram is measured as the number of likes and comments.

There are already tons of literature about customer engagement and how to measure customer engagement on social media. Some studies even focus on how to improve customer engagement (Hartmann et al., 2019; Zhao et al., 2023; Lou, Tan & Chen, 2019). However, the analyses are mostly performed for luxury brands or high-end products, such as vehicles and vacation homes. No study has focused on common products that are bought regularly. Besides that, the literature has also ignored the trend of environmental awareness. This paper fills both gaps in the literature by examining eco-friendly skincare brands. Table 1 provides a summary of literature related to customer engagement.

The term customer engagement is conceptualized and the different ways to measure customer engagement on social media have been explained. Next, the literature relating customer engagement to different attributes in social media posts will be discussed.

Tellis, MachInnis, Tirunillai & Zhang (2019) research what influences digital sharing for 79 non-eco-friendly brands. Positive emotions, for example amusement, excitement and warmth positively influence sharing. Positive emotions can be created using babies, animals, celebrities, or the element of surprise. Early or late placement of the brand could potentially harm sharing. Hartmann et al. (2019) examine what drives likes and comments on Instagram for food and drink consumption. Their findings suggest that marketing photos using consumers faces lead to more customer engagement relative to product photos with (or without) an invisible consumer holding the product. This confirms earlier findings of Delbaere, McQuarrie & Phillips (2011), Poor, Duhachek & Krishnan (2013), Xiao & Ding (2014) and Sajjacholapunt & Ball (2014) who all suggest that human faces enhance advertising effectiveness. Hartmann et al. (2019) enrich the existing literature by confirming that this idea also holds for social media marketing. This study will examine if this statement also holds for eco-friendly products. The first hypothesis states:

Photos featuring human faces have higher customer engagement relative to photos featuring products and hands.

Color plays an important role in marketing. It influences brand personality, likability, and familiarity (Labrecque & Milne, 2012). Samaraweera et al. (2021) argue that eco-friendly products can best be promoted using white-toned labels. This reflects the high-quality of the products. Lavuri et al. (2022), Seo & Scammon (2017) and Xue & Muralidharan (2015) argue differently. The findings of their studies indicate that green helps convey the message of environmental awareness and impact of the brand's. Yu, Xie & Wen (2020) study color psychology and its relation to Instagram post popularity for destination photos of travel guides. Their results suggest that individuals are more likely to respond to brighter and saturated photos. Li & Xie (2020) support these findings partly. They concluded that brighter images lead to higher customer engagement for airlines and sport utility vehicle companies. The second hypothesis holds:

The brighter the colors in a photo, the higher customer engagement.

Lou et al. (2019) find that influencer-promoted ads enjoy significantly higher engagement over brand-promoted ads for the top 50 apparel companies in the US (non-eco-friendly). Dos Santos et al (2023) also conclude that influencer have a positive effect on customer engagement. Their results show that consumers are susceptible to influencers promoting vegan cosmetic products. This is because a customer is attracted to the persona of the influencer, and therefore beliefs the product fit their beliefs and desires. Tellis et al. (2019) also confirm that influencers have a positive effect on customer engagement for 79 brands from different (non-eco-friendly) industries. This led to the third hypothesis:

Posts related to influencers receive higher customer engagement.

Rietveld, van Dolen, Mazloom & Worring (2020) conclude that the message content matters for customer engagement on Instagram. Liu et al. (2021) study content for luxury brands and find that they should focus on entertainment, interaction, and trendiness marketing on Twitter. Posts related to these topics lead to higher customer engagement in terms of likes and comments. Luarn et al. (2015) confirm that post having interaction purposes lead to higher customer engagement for products in the personal care and food consumption industry. The results of this study show that remuneration posts, posts that are of benefit to the customer with the purpose to attract attention, exhibit high engagement through likes. Tellis et al. (2019) find that information-focused content has a significantly negatively effect on sharing. On the contrary, Aichner (2019) find that social media users like and comment independently of the content of the post. Again, all studies are performed for non-eco-friendly products. This study will measure is the literature also holds for eco-friendly products. The corresponding hypothesis is:

Posts for entertainment and interaction purposes receive higher customer engagement.

The hypotheses are the foundation for the conceptual framework of this paper. A visualization of the conceptual framework is provided in Figure 1.

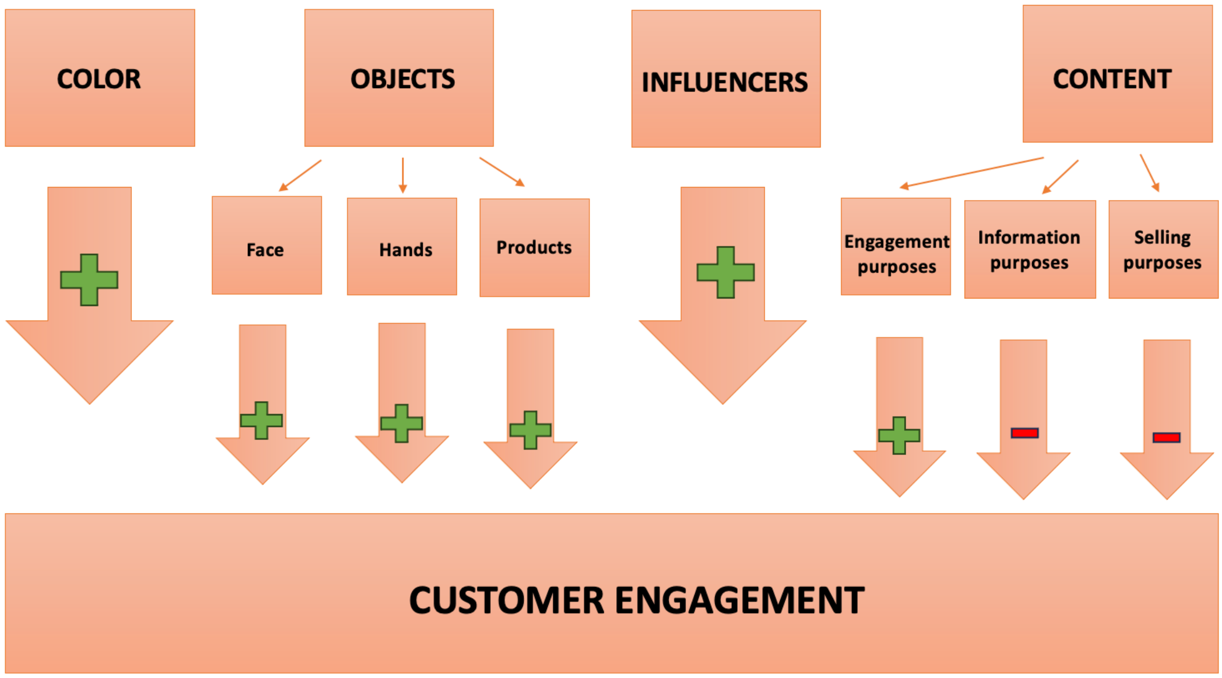


Figure 1: Conceptual framework

Notes: This paper studies what image features lead to higher customer engagement, which translates to more sales (Gill et al., 2017). The four main variables of interest are: color, objects, influencers, and content type. Objects include human faces (visible customer), hands (invisible customer) and products. Content can be posted for various reasons, to engage customers, to inform customers or to sell to customers. The hypotheses are formulated based on existing literature for non-eco-friendly products. This paper studies if the hypotheses should be accepted or rejected for eco-friendly products. The conceptual framework is a representation of all hypotheses.

Customer engagement on social media is a trending topic, given the amount of (new) literature written on this. Many definitions for customer engagement are formulated. This paper follows the simple and clear definition of Barger et al. (2016): “a set of measurable actions that consumers take on social media in response to brand-related content”. Most studies measure customer engagement on social media as likes, comments and shares. Unfortunately, data on sharing is not publicly available for Instagram. Therefore, this paper measures customer engagement as likes and comments on Instagram. The existing literature mainly focuses on expensive products. This paper will fill the gap in the literature by providing results for a more common sector, the skincare industry. This research also follows the trend of environmental awareness by studying eco-friendly skincare. The next section describes the data collection process.

Table 1. A literature overview.
Notes: The table provides a summary of academic papers that are related to the subject of this paper. The first column states the author(s) and publication year. The second summarizes the data used in the papers. The third column elaborates on the methods used by the authors. The fourth column summarizes the results of the papers.

Author(s) and publication year	**Data**	**Methodology**	**Results**
Archak, Ghose & Ipeirotis (2011)	Dataset from Amazone, containing sales and customer review data for digital cameras and camcorders over 15-month period (2005-2006).	Text mining. Clustering textual opinions based on pointwise mutual information and using externally imposed review semantics.	Consumers' preferences can be derived from reviews. Research incorporates product attributes, which product attributes are most preferred?
Luarn, Lin & Chiu (2015)	Timeframe: March 1 until May 1, 2014. 1030 photos from Facebook of 10 popular official brand pages.	ANOVA.	Vividness may significantly influence online engagement of users. Remuneration posts exhibit high engagement through likes. Photos are assigned a type of content (four categories: Coyle and Thorson 2001; De Vries et al., 2012; Ryan et al., 2013).
Zhang, Lee, Singh & Srinivasan (2017)	Airbnb dataset containing 7423 properties over 16-month period. Panel data from 13000 listings with 510000 images.	Deep learning and difference-in-difference analyses. Amazon Mechanical Turk to classify quality of pictures. Pixel-level is than used in convolutional neural network (CNN) to classify photos.	Examine if verified photos lead to higher revenue. Study finds that properties with verified images had 8.98% higher occupancy than properties without verified images (images taken by the host).
Aichner (2019)	Analyzing postings of 78 European football clubs and user reactions via social media.	Three different models. 1. Corporate social media use model to analyze 20954 Facebook, YouTube, Instagram and Twitter postings. 2. Categorization for social media	Measures customer engagement for social media posts by football clubs. Social media users like, comment and share independently of the content of the posting.

		postings using ANOVA and Scheffè test. 3. Experiment by creating Facebook campaign with a fictional hedonic low-involvement product.	
Hartmann, Heitmann, Schamp & Netzer (2019)	43.585 brand images from Instagram.	Transfer learning and CNN to classify the different types of brand images (3 types: consumer selfie, brand selfie and packshots). Also use text mining transformer-based language models to identify buying behavior in comments.	Consumer selfies generate the highest level of sender engagement (likes and comments), whereas they lead to lower levels of brand engagement (intention to purchase) relative to brand selfies. Brand selfies have higher brand engagement.
Liu, Shin & Burns (2019)	Luxury brands should focus on entertainment, interaction and trendiness of marketing to increase customer engagement.	NLP (Neuro-linguistic programming) is used to categorize big textual data. MySQL to aggregate to monthly panel data. Fixed-effects model is used to find insights and Unit root test to check for non-stationarity.	Luxury brands should focus on entertainment, interaction and trendiness of marketing to increase customer engagement. Focusing on customization does not increase customer engagement.
Lou, Tan & Chen (2019)	Instagram. Top 50 apparel companies in US ((Apparel Magazine 2016). March 1 – May 31, 2017.	Instagram Application Programming Interface (API) for extracting date, likes, comments and sponsorship. Support Vector Machine (SVM) and Topic Modeling (LDA).	Influencer-promoted ads enjoy significantly higher engagement (likes and comments) than brand-promoted ads.

Tellis, MacInnis & Zhang (2019)	Branded video ads on YouTube. From November 2013 – March 2014 (5 months). 1962 video ads. Multiple brands.	Manually assigned content.	Information content has significantly negative effect on sharing, except in risky contexts. Positive emotions have a positive effect on sharing prominent placement of brand names hurts sharing.
Shawky, Kubacki, Dietrich & Weaven (2020)	32 interviews held between November 2018 and February 2019. The interviews were recorded. Afterwards they were transcribed, in a way that reflects the meaning of the interviewees.	The interviews were coded using NVivo.	This research distinct four different levels of customer engagement on social media. Also, the study establishes measurement rules for these different levels of customer engagement. Four themes of interviews emerged from the interviews: connection, interaction, loyalty, and advocacy.
Li & Xie (2020)	Dataset of social media posts from Twitter (59.755 tweets) and Instagram of airlines and sport utility vehicle brands.	Google Cloud Vision API to categorize the images. Bivariate Zero-Inflated Negative Binomial Model.	Image content mere presence has a positive effect on user engagement. High-quality and professionally shot photos lead to higher engagement. Effect of colorfulness varies per product category.
Yu, Xie & Wen (2020)	Photos of Instagram travel guides that share about GBA (Greater Bay Area) locations. Octoparse to web scrape for past 3 years.	Image Color Summarizer, open-source tool from Genome Sciences Centre, K-means clustering and regressions.	Generally, individuals are more likely to respond to bright and saturated destination pictures on Instagram.
Corone, Nanne & van Miltenburg (2021)	1761 Instagram post brand-generated of Shanty Biscuits. InstaLooter to retrieve Instagram data.	Python Image Library to find pictures with exact same pixels to exclude double posts. MANCOVA.	Theme of slogan in post impacts the effect of social presence on customers' engagement in likes.

Cortez, Johnston & Dastidar (2023)	106 weeks of data from a Latin American B2B company operating in Chile and Peru.	VAR model with exogenous variables (VARX). Granger causality is used to test endogeneity and Philips-Perron test to evaluate stationarity.	What are the effects (short and long-term) of types of post, website visits, new followers and engagement over time? Social posts, new followers and increasing sales have a positive effect on engagement.
Dos Santos, de Brito Silva, da Costa & Batista (2023)	190 questionnaires about vegan consumption of Brazilian population. Mostly closed questions.	Cross-sectional survey method and Structural Equation Modelling.	Digital influencers can shape consumer intentions, even if the products do not represent their way of life (in this case: vegan cosmetics).
Zhao, Zhang, Ming, Niu & Wang (2023)	Data crawling from Sina Weibo.	Deep learning.	How image richness affects customer engagement. Image richness is positively related to emotional engagement and behavioral engagement, but negatively related to cognitive engagement.

3 Data

This section describes the data collection process and provides a more detailed look into the data itself. The data is collected using web scraping and applying convolutional neural networks. Five sustainable skincare companies are subject of this study: Alaffia, Biossance, Facetheory, Meow Meow Tweet and REN skincare. Data is collected from November 2022 until April 2023. A deeper look into the data is given by examining summary and descriptive statistics.

The choice of platform for this study is Instagram. Instagram is known as the most popular image-sharing platform of all social media platforms (Hartmann et al., 2019). Instagram is also the second most used social media platform by marketers in 2022. This is no surprise since Instagram has more than two billion active users monthly (Statista, 2023). The popularity among consumers and wide use of Instagram by marketers has ensured Instagram as the social media platform of this study. This paper focuses on brand-generated photos from eco-friendly skincare companies posted on Instagram feed (Luarn, Liu & Chiu, 2015; Corone et al., 2021).

Not all eco-friendly skincare companies are considered. Only companies of substantial size are considered, meaning companies that have at least 50,000 followers on Instagram (Shen & Bissell, 2013). To be considered eco-friendly, companies must sell green products. Green products are products that are produced using toxic-free materials and environmentally friendly measures, and which are certified as such by recognized organizations (Kumar & Ghodeswar, 2015). Companies who profile themselves as eco-friendly but really are not, are excluded. For example, The Body Shop and Kiehl's are excluded since both use plastic packaging and their products contain a lot of damaging ingredients such as petrochemicals and parabens.

Five brands are subject of this research: *Alaffia, Biossance, Facetheory*, *Meow Meow Tweet* and *REN skincare*. The big difference between these companies and other skincare companies is that these brands are vegan, low-waste and cruelty-free, bringing no harm to animals. Multiple skincare brands are considered to avoid one brand misrepresenting the eco-friendly skincare industry. Although panel data might be preferred over pooled data, pooled data is better to generalize the results, provide a higher estimate of variance, and it might yield higher accuracy on the test data. To control for possible differences between the brand's, fixed affects are added. The fixed affects relate to the brand itself and its popularity on Instagram. This will be discussed in more detail later.

The data for this study is collected by the researcher herself. The data is collected for 6 months, ranging from November 1st, 2022, until April 30th, 2023. In total 616 images are web scraped and analyzed. The data on customer engagement, likes and comments, is collected by web scraping the Instagram pages of the five brands. Web scraping is performed using the Instaloader package from Python. Also, the posts, including both photos and videos, of the brands are collected by web scraping. The videos are transformed into single photos (automatically by the Instaloader package) to make analyzing easier. Of course, this leads to some bias, since one picture cannot fully capture the content of a video. This bias is bigger for some brands than others. 74% of the posts posted by Biossance are videos, while only 14% of the posts posted by Meow Meow Tweet are videos (Table 2). The same problem arises for photo slides, posts containing multiple photos (an example can be found in Appendix A). Only the first photo of the photo slide is analyzed to make things easier. Again, this leads to some bias since this single picture does not capture all content. This bias is expected to be smaller since companies post less photo slides than videos (Table 2).

For all posts, the caption is analyzed to see if the brand name is mentioned in the caption, together with the number of hashtags (#) and tags (@) used. The captions are analyzed using the grepl() and str_count() functions from the base and stringr packages in R. The brand name variable is included to test the assumption of Tellis, MacInnis & Zhang (2019) who suggest that prominent brand name placement can hurt customer engagement. The number of hashtags is included to account for engagement obtained through these hashtags. The number of tags is included to recognize influencer-promoted posts. The brightness of a photo is measured as the sum of the three top color's pixel percentages (Li & Xie, 2020). The percentage of red, green, and blue indicates the color variation in an image. A low pixel percentage suggests a colorful photo, whereas a high pixel percentage suggests a monotonic photo. The lower the pixel percentage, the brighter the photo is. The pixel percentage per image is collected using a self-built formula in R.

The other variables related to the content of the post are (mostly) collected using deep learning. Convolutional neural networks, also referred to as CNN or ConvNet, are used to extract the image features. CNN was first introduced by Fukushima (1980) who called it Neocognitron. Deep learning is the most frequently used machine learning method to analyze text and image data in marketing studies (Ma & Sun, 2020; Simonyvan, Vedaldi & Zisserman, 2013). CNN is a class of neural networks that is specialized in processing images. Neural networks are a series of algorithms that is inspired by the human brain, mimicking the way that biological neurons signal to each other. The network has multiple layers who are known as "deep" networks and are used for deep learning algorithms. A CNN typically has three layers: the convolutional layer, the pooling layer, and the fully connected layer (Figure 2). The convolutional layer is the key block. This layer performs a dot product between the kernel (matrix one) and matrix two, which is a restricted portion of the receptive field. The pooling layer reduces the size of the representation. This allows the computation and weights to decrease. Lastly, the fully connected layer is there to map the representation between the input and the output.

Google Cloud Vision API is used as the image classification machine learning algorithm to classify the objects in the images. Google Cloud Vision API is used in combination with Python. Google Cloud Vision API is the preferred algorithm because of its high accuracy (Singh, Wheeler, Fong & Chaudhary, 2019). Google Cloud Vision API uses pre-trained machine learning models to understand images and provide different features such as detection of faces, logos, and landmarks. The algorithm is trained in a similar way as how the human brain learns, through trial and error (Philp, Jacobson & Pancer, 2022). For this study, Vision API is used to detect faces, to classify facial expression, detect text and to detect objects and assign labels to them. An illustration on how Vision API works is provided in Figure 3-5. Objects are for example persons, animals, or goods (Figure 4). Labels classify persons and things, for example body parts, products, and general themes. Only labels with a confidence score of 50% or higher are presented by Vision API (Figure 3). Facial expressions are rated on a scale from 1 (very unlikely) to 5 (very likely) (Figure 5). A facial expression is accepted from 3 (possible) onwards.

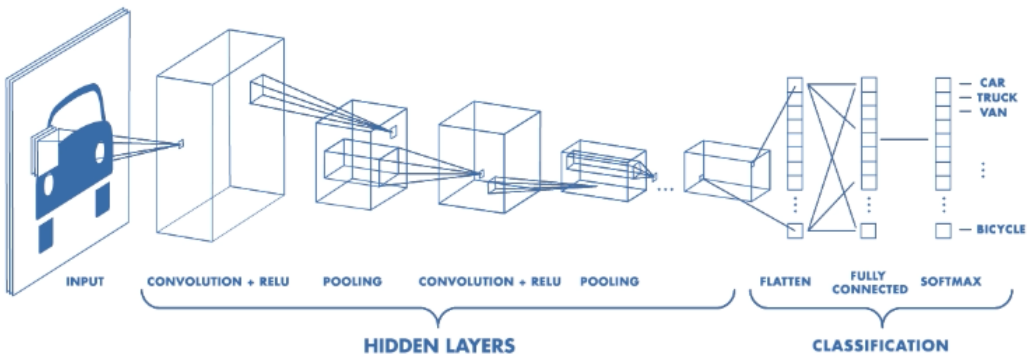


Figure 2 (Mishra, 2020). CNN layers.
Notes: A visual representation of the different convolutional neural network layers. The first layer is the convolutional layer, the second the pooling layer and the last layer the fully connected layer.

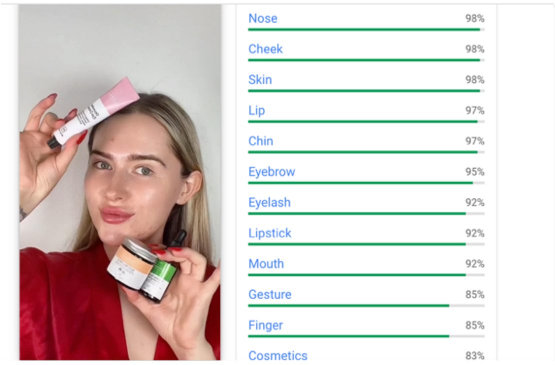


Figure 3. Label detection.
Notes: Google Cloud Vision API label detection detects objects and assigns labels to them.

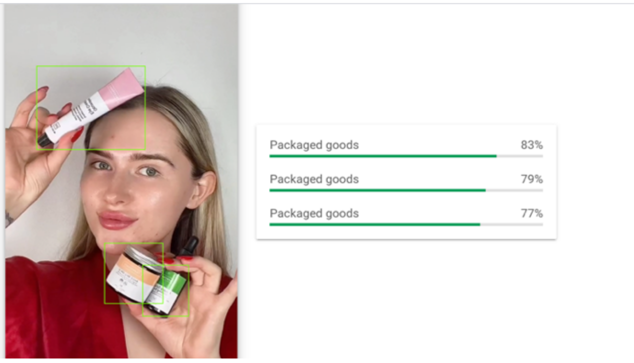


Figure 4. Object detection.
Notes: Google Cloud Vision API detects objects and indicates confidence scores.

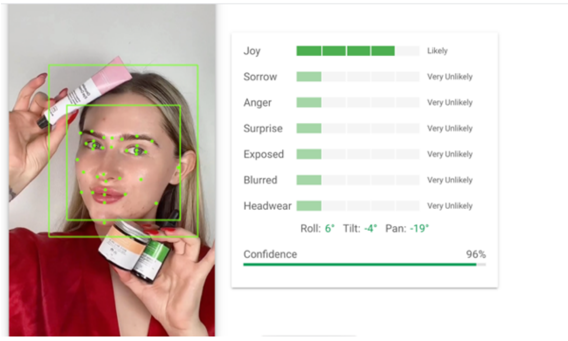


Figure 5. Face detection.
Notes: Google Cloud Vision API recognizes facial expressions and scores them on a scale from 1-5 (very unlikely to very likely).

The image features extracted using CNN are faces, facial expressions, hands, products, packaged goods, text, and the label "personal care". Face is a dummy variable indicating if the post contains at least one human face, meaning if there is a customer visible. The facial expressions are also obtained. For every face detected, all facial expressions are ranked on a scale from 0 (very unlikely) to 5 (very likely). Google Cloud Vision can detect four emotions: joy, anger, surprise and sorrow. A facial expression is accepted from 3 (possible) onwards. The only available facial expression in the dataset is joy, therefore only a dummy for this facial expression is included. It is not surprising that only joyful faces appear since all Instagram pages are (mostly) used for marketing purposes. The variable product indicates if there is a product visible. The variable packaged goods counts the number of packaged goods recognized by the algorithm. Hand indicates whether hand/arm/finger/thumb is visible. This variable indicates if there is an invisible customer (a customer holding a product, without the appearance of a face). Text refers to the post having text other than the text on the packaging. Personal care is a dummy variable that has value 1 if the photo is labelled "personal care" by the algorithm.

The content of the post is analyzed manually since the categories of the used algorithm are not relevant for this study. Cloud vision API only detects the following content: adult, spoof, medical, violence and racy. These types of content are not related to (eco-friendly) skincare and therefore not used. Instead, the type of content is manually assigned to all posts. Although this a little tight, is it has been done before (Luarn et al., 2015; Aichner, 2019). Every post is assigned one of four categories: informational post, entertainment post, remuneration post or advertisement post (Luarn et al., 2015). Informational posts provide information about products, ingredients, stores and how to use the products (De Vries, Gensler & Leeflang, 2012; Muntinga, Moorman & Smit, 2011). Entertainment posts do not refer to the brand or product but have the purpose of engaging people. Think for example about humor, wordplay, or anecdotes (Cvijikj & Michahelles, 2011). Remuneration posts contain information that is associated with benefits for the consumer. These include promotions, giveaways, special offers and other offers that attract attention (Cvijiki & Michahelles, 2011; Wood, Ray & Messinger, 2013). Lastly, advertisement posts refer to posts that aim to promote products/services.

Four control variables are added to control for the popularity of the brand and the time of the post being online (Philp, Jacobson & Pancer, 2022). These four variables are the brand name, number of followers for the brand (on May 2nd), the average number of likes for the brand and the number of days the post is online. The information for these variables is collected using web scraping. The variables are built using R. An overview of all the variables can be found in Table 3.

Table 2. Bias overview.

Notes: The percentage of posts being videos and photo slides per brand are presented. The content of the videos and photo slides is transformed into one single photo, which could cause some bias. This table indicates how much of the data per brand is subject to bias.

Brand	**Videos**	**%**	**Slides**	**%**
Alaffia	47	40	16	14
Biossance	140	74	33	24
Facetheory	60	57	16	15
Meow Meow Tweet	6	15	10	24
REN skincare	84	52	5	3

Table 3. Variable overview.
Notes: This table presents all variables used in this study. The first column contains the variable name. The second column gives a short explanation of the variable. The third column indicates what kind of variable it is.

Variable		
Brandname	The name of the brand that posted the photo	Nominal
Caption	Caption of the photo	Nominal
Days online	Days past since the photo is posted	Continuous
Followers	The brand's number of followers (on May 2^{nd})	Continuous
Average likes	Average number of likes for brand	Continuous
Likes	Number of likes photo	Discrete
Comments	Number of comments photo	Discrete
Hashtags (#)	Number of hashtags in caption	Discrete
Tags (@)	Number of tags in caption	Discrete
(Caption) brandname	Whether or not caption contains brand name	Nominal
Type of post	If the post is a video or photo	Nominal
Photo slide	How many photos and videos a post consists of	Discrete
Type of content	The type of content	Nominal
Face	Whether the photo includes a human face	Nominal
Joy	Whether at least one facial expression is joy	Nominal
Product	Whether the photo contains a product	Nominal
Packaged goods	The number of packaged goods in the photo	Continuous
Hand	Whether the photo contains a hand/arm/finger/thumb	Nominal
Text	Whether the photo contains text other than packaging	Nominal
Brightness	Sum of three top colors' pixel percentages	Continuous
Personal Care	Whether the photo is labelled "personal care"	Nominal

Some interesting insights into the data will be covered. The descriptive statistics for the brands are presented in Table 4. The descriptive statistics describe the total number of posts per brand and the number of followers per brand (on May 2^{nd}). If a brand is verified and has a business account is also checked. Biossance has the highest following (539,702), while Meow Meow Tweet has the lowest (57,822). Interesting to note is that Biossance has about 20 times more posts than Alaffia (4,316 respectively 226). This could be because Biossance posts more regularly or that the page is live for a longer time.

To compare the brands better, the statistics for this study's time span are presented in Table 5. This study's time period runs from November 1^{st} until April 30^{th}. The number of posts and average number of likes and comments per brand for the chosen time span are presented. The last column (Table 5) refers to the number of posts that are in collaboration with other Instagram accounts (influencers or other brands). Biossance has the highest customer engagement in terms of likes (1,348 on average), while REN skincare has the highest customer engagement in terms of comments (197 on average, Table 5). Alaffia has the least number of likes on average (65), while both Alaffia and Meow Meow Tweet have the least number of comments on average (6 respectively 5). Why these numbers differ so much, is studied in this paper.

Table 4. Descriptive statistics.

Notes: Descriptive statistics per brand retrieved by web scraping Instagram pages. The data is collected on May 2nd. Any posts or likes after that date do not count towards statistics in this table.

Brand	**Total number of posts**	**Number of followers**	**Verified**	**Business Account**
Alaffia	226	69,169	True	True
Biossance	4,316	539,702	True	True
Facetheory	967	174,062	False	True
Meow meow tweet	2,917	57,822	True	True
REN	3,061	323,868	True	True

Table 5. Summary statistics.

Notes: Summary statistics per brand from November 2022 until April 2023 retrieved by web scraping Instagram pages. The last column indicates the number of posts that are in collaborations with other brands and/or influencers.

Brand	**Number of posts**	**Average number of likes**	**Average number of comments**	**Number of collaborations**
Alaffia	117	65	6	3
Biossance	190	1,348	114	3
Facetheory	106	315	54	2
Meow meow tweet	41	129	5	0
REN	162	546	197	0

For all discrete variables (Table 3), the mean, standard deviation, minimum, and maximum are presented in Table 6. The number of average likes is much higher than the average number of comments. This is not surprising since liking a post takes least effort (Zipf's law of least effort).

A summary of the image features per brand is give in Table 8 (column 2-3). Facetheory focuses a lot on incorporating hand and text into their images, while showing less products. Meow Meow Tweet scores below average for all features. REN skincare and Biossance both feature a lot of faces and products relatively to the other brands. The content purposes of the brands are summarized in Table 8 (column 4-5). All brand pages focus on advertisement (49-78%), while posting little remunerational (0-7%). Alaffia features a lot of entertainment (25%) respectively to the other brands (Biossance 7%, Facetheory 16%, Meow Meow Tweet 7% and REN 2%). The Instagram page of Biossance and REN mainly focusses on ads (72%

respectively 78%). Facetheory, Meow Meow Tweet and REN skincare have almost no remunerational posts (2% respectively 0% respectively 1%).

Table 7 provides insights into the correlation between the numeric variables. As expected, the average number of likes and followers are highly correlated. The correlation between these variables is almost 1 (0.98). The number of hashtags in the caption is negatively related to the number of followers and average number of likes (-0.41 respectively -0.39). Likes and comments are also correlated to each other (0.23), but not as much as expected.

Table 6. Discrete variables.
Notes: Summary statistics for the discrete variables from November 2022 until April 2023. The mean, standard deviation, minimum and maximum values for the variables are presented.

Variable				
Discrete variables	**Mean**	**SD**	**Min**	**Max**
Likes	625	2,780	0	63,337
Comments	96	515	0	6,895
Hashtags (#)	1.2	2	0	15
Tags (@)	0.6	1	0	24
Brightness	1.64	0.50	0.52	2.79

Table 7. Correlation matrix.
Notes: A correlation matrix displaying the correlation coefficients for all numeric variables. The sign indicates if the correlation is positive (+) or negative (-) A correlation of 1 indicates perfect positive correlation, -1 perfect negative correlation.

Variable							
Numeric	**Followers**	**Average likes**	**Likes**	**Comments**	**Days online**	**Hashtags**	**Tags**
Followers	1	0.98	0.18	0.08	-0.01	-0.41	0.06
Average likes	0.98	1	0.18	0.06	0.00	-0.39	0.04
Likes	0.18	0.18	1	0.23	0.02	-0.04	0.10
Comments	0.08	0.06	0.23	1	-0.07	-0.02	0.29
Days online	-0.01	0.02	0.02	-0.07	1	-0.12	-0.01
Hashtags	-0.41	-0.39	-0.04	-0.02	-0.12	1	-0.02
Tags	0.06	0.04	0.10	0.29	-0.01	-0.02	1

Table 8. Photo features and content per brand.

Notes: This table presents a summary of the content per brand. Columns 2 and 3 indicate the proportion of photos containing a specific image feature. The four image features are: faces, products, hands, and text. One feature does not rule out another feature. Columns 4 and 5 indicate the proportion of photos being a specific content type. Only one content type is selected per photo, meaning the percentages add up to 100.

Variable				
Nominal variables	Feature	Proportion	Content	Proportion
Alaffia				
	Faces	28%	Advertisement	49%
	Product	67%	Remuneration	7%
	Hand	13%	Information	19%
	Text	41%	Entertainment	25%
Biossance				
	Faces	36%	Advertisement	72%
	Product	69%	Remuneration	7%
	Hand	17%	Information	14%
	Text	38%	Entertainment	7%
Facetheory				
	Faces	33%	Advertisement	64%
	Product	63%	Remuneration	2%
	Hand	23%	Information	18%
	Text	52%	Entertainment	16%
Meow Meow Tweet				
	Faces	5%	Advertisement	71%
	Product	61%	Remuneration	0%
	Hand	10%	Information	22%
	Text	29%	Entertainment	7%
REN				
	Faces	38%	Advertisement	78%
	Product	69%	Remuneration	1%
	Hand	17%	Information	19%
	Text	35%	Entertainment	2%

4 Methods

To answer the research question, this paper uses several methods. First, two simple analyses are performed. Second, more advanced machine learning methods are applied.

To obtain initial insights, the relationship between customer engagement and all other variables is measured by constructing two relatively simple methods. A linear and lasso model are built to predict likes and comments. Second, the more advanced methods, decision trees and random forest, are applied.

All methods are applied twice since there are two measures of customer engagement. In the first regression, the number of likes a post receives is the dependent variable, while in the second regression the number of comments a post receives is the dependent variable. The model descriptions are as follows:

$$Likes = Brightness + Followers + Average\ likes + Comments + Hashtags + Tags + Caption\ brand\ name + Brand\ name\ + Type\ post + Photo\ slide + \ Type\ content + Face + Facial\ expression\ joy + Packaged\ goods + Product + Hand + Text + Personal\ care + Days\ online \ (3.1)$$

$$Comments = \ Brightness + Followers + Average\ likes + Likes + Hashtags + Tags + Caption\ brand\ name + Brand\ name + Type\ post + Photo\ slide + Type\ content + Face + Facial\ expression\ joy + Packaged\ goods + Product + Hand + Text + Personal\ care + Days\ online \ (3.2)$$

4.1 Linear regression

Linear regression is a supervised learning method. Supervised learning is known as a subcategory of machine learning and is the training of algorithms to predict outcomes by using labeled datasets. A linear regression is a simple approach, nevertheless still widely used. It is a good starting point for newer approaches (James, Witten, Hastie & Tibshirani, 2013). A quantitative response variable (Y) is predicted on the basis of a predictor variable (X).

This study uses the multiple linear regression, which is an extension to the simple linear regression. The multiple linear regression considers more than one predictor variable. The regression coefficients are estimated using:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_p x_p \quad (3.3)$$

The parameters are estimated minimalizing the residual sum of squares. This is called the least squares approach.

A linear regression has four assumptions (Freedman, 2009). There must be a linear relationship between the dependent variable and the predictor variables. The residuals should be normally distributed and independent. The residuals need to have a constant variance at every level of X. All assumptions will be checked. The fit of the model will also be examined. The model fit can be examined by studying the R^2. The R^2 indicates how much of the variance in the response variable is explained by the predictor variables.

4.2 Lasso regression

To select a subset of predictors, shrinkage methods are available. These models use a technique that constrains or regularizes the coefficient estimates so that the coefficient estimates shrink towards zero. This can reduce variance and fix problems with multicollinearity. This study uses a lasso regression to reduce variance and avoid multicollinearity problems. Lasso stands for least absolute shrinkage and selection operator (Tibshirani, 1996). For lasso, the coefficient estimates are estimated by minimizing:

$$\sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{p} \left| \beta_j \right| \quad (3.4)$$

where λ is the tuning parameter that is equal to or bigger than zero.

Lasso uses a ℓ_1 penalty, which means that the coefficient estimates can be forced to zero when the tuning parameter becomes large (James et al., 2013). Lasso therefore does variable selection. This helps identifying the most important variables and reduces the complexity of the model.

The lasso model will be used to predict customer engagement. The model is used as a benchmark, to compare overall model performance. The performance of all models is measured by the root mean square error (RMSE). The RMSE is the standard deviation of the residuals (predicted errors):

$$\sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}} \quad (3.5)$$

where $\hat{y}_i$ represents the predicted values and y_i the actual values.

4.3 Tree based models

To predict customer engagement, two more complex machine learning methods will be used as well. Machine learning is a subset of artificial intelligence. Artificial intelligence is the branch of computer science that focuses on the simulation of human intelligence (Luger, 2005). Machine learning is programming computers in such a way that they can perform tasks without explicit programming. Machine learning methods can be supervised or unsupervised. This study makes use of supervised machine learning methods since the models are trained on labelled datasets. The two more complex methods used are decision trees and random forests.

Decision trees are non-parametric supervised learning algorithms. Decision trees can be used for both regression and classification problems. This study solves a regression problem. The decision tree takes its name from its hierarchical, tree structure. The tree always starts with a root node, which splits the data in two. After that, multiple internal nodes can follow. At the end of the tree, the terminal (or leaf) nodes appear. These hold the prediction values.

For selecting the features at each split, the decision tree uses a top-down, greedy approach, known as recursive binary splitting. It is a top-down approach because it starts at the top and works all the way down. It is greedy because it determines the best split not by looking into the future. The splits within the tree are selected based on minimizing the residual sum of squares. For each possible split, the residual sum of squares is calculated. The split that minimizes the residual sum of squares is selected. This process continues until a certain stopping criterium is reached (James et al., 2013).

A problem can occur when following these steps. The tree might predict good on the training set but poor on the test set. If this occurs, it is probably due to overfitting. The decision tree is too complex and is grown too much in depth. To avoid this problem, decision trees can be pruned. Variance and interpretation will improve at the cost of (a little) bias. Cost complexity pruning reduces the size of the tree. The cost complexity parameter with the lowest cross-validated error is selected to prune the tree.

Decision trees often outperform regression models because linearity is not assumed (Sadeque & Bethard, 2019). The downside of decision trees is that they are prone to overfit. To avoid this, ensemble methods have been introduced (Breiman 1996; Breiman 2001). Bagging and random forest are ensemble methods that combine multiple decision trees to reduce overfitting. This comes at the expense of interpretability, and it is computationally more expensive.

Ensemble methods combine several models to improve accuracy. Bootstrapping aggregation, also referred to as bagging, was first introduced by Breiman in 1996. The idea is to combine multiple simple models (decision trees) into one powerful model (James et al., 2013). Decision trees suffer from high variance and are prone to overfit. Bagging averages a set of observations to reduce this variance. Multiple training sets are taken from the data set with replacement (bootstrapping). For each bootstrapped training set (B), a decision tree is built, and the average outcome of all predictions is taken:

$$\hat{f}_{bag}(x) = \frac{1}{B}\sum_{b=1}^{B} \hat{f}^{b}(x) \quad (3.6)$$

A big disadvantage of bagging is that all trees are highly correlated, since all trees tend to be similar. The trees tend to make the same splits since they share the same features. A strong predictor that reduces the residual sum of squares a lot will appear in the top of most trees. To avoid this problem, Breiman introduced random forests in 2001. This model builds forth on the idea of bagging. This method still builds trees based on multiple bootstrapping samples. However, now only a random sample of m predictors is considered at each split. M is the parameter that controls for the number of predictors considered at each split. The default is using $m = \frac{p}{3}$, where p is the number of predictor variables. However, the parameter can also be tuned using a grid search or via cross-validation (Gareth James, 2013).

The performance of an ensemble method can be measured as the out-of-bag error (OOB-error). For all data points, predictions are made for all trees that do not have this data point included in the bootstrap sample. The overall OOB error can be computed from this. The OOB error is an estimate of the test error of the model.

4.4 Interpretation methods

Machine learning methods can be classified into two groups, white-box, and black-box models. These groups refer to the extent to which the models can be interpreted. When choosing between these groups, it is a trade-off between accuracy and interpretability (Pintelas, Livieris & Pintelas, 2020). White-box models are "models whose inner logic, workings and programming steps are transparent and therefore it's decision-making process is interpretable" (Pintelas et al., 2020). An example of a white-box model is a single decision tree. On the other hand, black-box models are models whose inner workings are not known and therefore it is harder to interpret the output of these models. Black-box models only provide outcome/predictions. An example of a black-box model is a neural network or support vector machine. Also, the ensemble methods bagging, and random forest are considered black-box models. In this study, the regressions and decision trees are white-box models, while the random forests are black-box models.

The black-box models tend to outperform the white-box models in terms of accuracy. This comes at the loss of interpretability. Since the black-box models only provide predictions, it is unclear how these predictions are obtained. Of course, accuracy is necessary to be certain of predictions made. But interpretability can be very valuable as well. Solely being able to predict customer engagement is not meaningful for marketing teams. The marketing team wants to know how to improve customer engagement and what it exactly is that drives customer engagement. Interpretation methods help understanding why certain images perform better than others and what causes this.

Post-hoc interpretation methods are developed to explain black-box models. The interpretation methods are not part of the black-box algorithms. Because of this, the interpretation methods can be biased. These post-hoc interpretation methods can be classified into two groups: global model-agnostic methods and local-model agnostic methods. Global model-agnostic models are used to describe the model in general. Local model-agnostic methods are used to explain why a certain prediction is made by the model. The most common interpretation methods are permutation feature importance (global), partial dependence plots (global) and local surrogate (local).

This study is interested in the global working of the models, not why specific predictions are made. Therefore, this paper only uses global interpretation methods. The permutation feature importance and partial dependency are measured. Interaction between the variables could bias the results of these two methods. Therefore, the interaction between all variables is also examined.

Permutation feature importance is used to detect the most important features in a model. Breiman (2001) introduced the method for random forest models and Fisher, Rudin and Dominici (2018) generalized the idea for all black-box models. The method shuffles a single feature value to measure how much the model score decreases. The drop in model performance indicates how much the model relays on this feature and thus reflects the feature's importance. Feature importance is computed on the test data (Molnar, 2019).

Partial dependence plots uncover the relationship between the dependent variable and one (or more) predictor variables (Friedman, 2001). The partial dependence function for regressions is defined as:

$$\hat{f}_s(x_s) = E_{xc}\,[\hat{f}\,(x_s, X_c] = \int \hat{f}(x_s, X_c) d\mathbb{P}(X_c) \text{ (3.7)}$$

Where x_s are the features for which the function should be plotted and X_c the other features in the model.

A one-way partial dependence plot displays the relationship between the dependent variable and one predictor variable. These graphs are easiest to interpret and therefore used in this study. On the y-axis the dependent variable, in this case customer engagement, is measured and on the x-axis the predictor variable.

Feature interactions estimates the interaction in a model. It is measured via the decomposition of the prediction function. If any variance of a predictor on the response variable cannot be explained, this is assigned to interaction strength. The interaction strength is the proportion of variance of the two-dimensional relationship that is not explained by the sum of the two one-dimensional relations. (Molnar, 2019). Interaction is measured by Friedman's H-statistic and takes a value between 0 (no interaction) and 1 (complete interaction) (Friedman & Popescu, 2008).

A wide range of methods is selected because there is no best method. Depending on the data, it may differ which method performs best. In most cases, more advanced methods yield higher accuracy because of their complexity (James et al., 2013). However, these more advanced methods are harder to interpret. Simpler models are easier to interpret, but often yield lower accuracy. There is a trade-off between accuracy and interpretability when choosing a model. When simpler tools perform (almost) as well as more advanced methods, these methods are preferred because of their simplicity and interpretability. To discover which method best describes the data, both advanced and simpler methods are used. The more advanced methods are decision trees and random forests. Decision trees tend to outperform regressions when the relationships are nonlinear and complex (James et al., 2013). Random forest is selected since it is an extension to decision trees. It is chosen over bagging because it builds trees that are uncorrelated (in comparison to bagging where the trees are correlated). For this paper, constructing a neural network is considered, because it well-suits nonlinear and complex data and is being used a lot in recent studies. However, because of the size of the data (only 431 images in the training set), this method is not performed. A neural network contains to many parameters to train successfully, considering the small size of the data.

5 Results

The last section discussed the methods used. This section presents the results of the different methods. The results of the multiple linear regression and lasso regression are elaborated on first. Next, the main models are discussed. The main models used to predict customer engagement are decision trees and random forests. All methods are applied twice, first to predict the number of likes, second to predict the number of comments. The measures of customer engagement are on a different scale. They are not standardized, to avoid loss of interpretability. This paper aims to find what drives likes and comments for eco-friendly social media campaigns, therefore interpretability is preferred over comparison.

5.1 Linear regression

The relationship between customer engagement and the predictor variables is first modelled using a multiple linear regression. Two models are built to measure both the relationship between likes and the predictor variables and comments and the predictor variables. The models are built together, but output is provided separately. The model descriptions can be found in 3.1 and 3.2.

In the first model, the dependent variable is the number of likes a post receives. The independent variables are all photo features, such as brightness and face, all caption related variables and the control variables (3.1). In the second model, the dependent variable is the number of comments a post receives. The independent variables are again all photo features (3.2).

Table 9 provides the coefficients of the linear models. The first two columns belong to the model predicting likes, the last two columns to the model predicting comments. Only significant variables are interpreted. For the first model, predicting likes, it is found that the number of tags and whether a hand or product are displayed have a significant effect on the number of likes. The number of tags a caption contains has a positive effect on the number of likes, meaning that the more tags are included in the caption, the higher the number of likes is. Displaying a product has a significant negative effect on the number of likes, with a coefficient of -880.1. This means that on average, a product being displayed lowers the number of likes by around 880. On the other hand, a hand, finger, or arm being displayed has a significant positive effect on the number of likes, with a coefficient of 865.0. On average, a hand, finger, or arm being visible, leads to 865 more likes.

The second model predict the number of comments. The corresponding coefficients are presented in the fourth column (Table 9). The number of followers, tags, packaged good, product, hand, and days online all have a significant effect on the number of comments. The number of followers has a very small positive effect on the number of comments (0.001) and is significant at 0.05 significance level. The number of tags again has a positive effect on customer engagement and is statistically significant at 0.1% level. For every tag in the caption of a post, the post on average receives 94.8 more comments. Packaged goods also have a positive effect on customer engagement and is statistically significant at 0.01 level as well. For every packaged good displayed, the number of comments on average increases by 53.1. A hand, arm or finger displayed also leads to (178.8) more comments. The number of days a post is online has a negative effect on the number of comments. This is interesting, since it is illogical to have the number of comments decrease over time. However, it could be that recent posts happen to have more comments.

The coefficient of determination, also referred to as R^2 is a measure of goodness of fit. It measures how much of the variation in the dependent variable is accounted for by the independent variables. The R^2 of the first model is 7.4% and the adjusted R^2 is 4.4%. The second model has a R^2 of 16.4% and adjusted R^2 of 13.8%. The adjusted R^2 corrects for significancy. Both R^2 indicate a relatively poor model fit. The standard errors also indicate the model fit. Since there is heteroskedasticity (Figure B5 and B6, Appendix B), robust standard errors are used. The standard errors are still quite high, which indicates a relatively poor model fit as well.

All assumptions of the linear models are checked (Appendix B). For the model predicting likes, it holds that there is a linear relationship (Figure B1, Appendix B). For the model predicting comments, it is harder to assume a linear relationship (Figure B2). Normality of the residuals can be assumed for the likes model (Figure B3). It is harder to assume for the comments model (Figure B4). For both models, homoskedasticity is hard to assume since the variance in the residual error is not completely constant (Figure B5 and B6). The residuals have a mean around 0 and are spread out (Figure B7 and Figure B8). The correlation between the continuous variables is checked (Figure B9). There is a strong correlation between the number of followers a brand has and the average number of likes. This correlation is expected to cause problems in the linear model. To control for this correlation, a lasso regression is performed as well. There is most likely multicollinearity since number of followers and average likes have a variance inflation factor value of 34 for both models (Table B1). This means that the regressors are not independent. Since there is possibly multicollinearity in the data, a lasso regression is performed to correct for this.

Table 9. Linear models.

Notes: The table presents the coefficients of two linear models. The dependent variable is the first model is likes. The coefficients corresponding to this model are found in column 2. The dependent variable of the second model is comments. The corresponding coefficients can be found in column 4. Robust standard deviations are show in between brackets.

Linear regression			
Likes	Coefficients	**Comments**	Coefficients
Intercept	170.7 (383.5)	Intercept	-65.4 (113.1)
Brightness	121.2 (200.0)	Brightness	50.9 (55.7)
Followers	-0.0 (0.0)	Followers	0.0* (0.0)
Average likes	1.2 (0.8)	Average likes	-0.4 (0.2)
Hashtags	39.1 (25.1)	Hashtags	-3.8 (7.8)
Tags	176.9* (92.6)	Tags	94.8*** (41.9)
Type post (video)	-76.9 (118.0)	Type post (video)	-86.1 (61.8)
Photo slide	-80.3 (41.5)	Photo slide	-14.6 (9.7)
Caption brand name (true)	73.9 (115.1)	Caption brand name (true)	84.3 (51.4)
Type content (remuneration)	-87.6 (182.0)	Type content (remuneration)	-82.3 (85.1)
Type content (informational)	-276.1 (224.6)	Type content (informational)	-39.0 (29.1)
Type content (entertainment)	-516.5 (343.1)	Type content (entertainment)	-11.0 (26.7)
Face (true)	-185.0 (218.0)	Face (true)	4.8 (95.0)
Facial expression joy (true)	621.5 (475.3)	Facial expression joy (true)	-37.0 (87.4)
Packaged goods	65.1 (28.8)	Packaged goods	53.1*** (22.5)
Product (true)	-88.0** (627.4)	Product (true)	-121.4* (49.3)
Hand (true)	865.0** (707.7)	Hand (true)	178.8** (82.6)
Text (true)	-270.1 (215.8)	Text (true)	-28.3 (40.5)
Personal care (true)	40.1 (126.2)	Personal care (true)	39.4 (52.1)
Days online	1.1 (1.9)	Days online	-90.9* (0.4)

* Significant at 5% level, ** significant at 1% level, *** significant at 0.1% level.

5.2 Lasso regression

Lasso is a type of linear regression that adds a penalty for nonzero coefficients. This means that variable selection is applied. This type of regression well-suits data dealing with multicollinearity. Since there is multicollinearity in the data (Table B1, Appendix B) and the predictors are correlated (Figure B9), lasso regressions are performed. Again, the first model explains the number of likes based on the predictor variables (3.1) and the second model explains the number of comments (3.2). The lasso regressions are used as benchmark models, to compare the performance of all models. The penalty parameter (λ) is equal to 199.4 for model one and 25.53 for model two. The first model keeps relatively few variables compared to the second model (Table 10). For both models it holds that the use of tags in the caption leads to higher customer engagement. Also, featuring hands in the photos leads to higher customer engagement for both regressions. Some coefficients in the second model are surprising. A higher pixel percentage, meaning less color, indicates more comments. Also, the coefficient for days online is negative suggesting that the more recent photos have received more comments.

Five test sets are created to measure the performance of the model. The first model, predicting likes, has an average RMSE of 167.6. The model predicting comments an average RMSE of 20.8. These measurements will apply as benchmark.

The problem of multicollinearity is overcome. The two variables causing multicollinearity are average likes and number of followers (Table B1, Appendix B). In the first model, one of the two variables (followers) leading to the multicollinearity in the linear model is dropped. In the second model, the other variable (average likes) is dropped.

Table. 10 Lasso models.

Notes: The coefficients of two lasso regressions are presented. The second column presents the coefficients for predicting likes. The fourth column present the coefficients for predicting comments.

Lasso regression			
Likes	Coefficients	**Comments**	Coefficients
Intercept	634.8	Intercept	93.4
Average likes	299.5	Brightness	11.2
Tags	76.6	Followers	3.2
Facial expression joy (true)	67.8	Tags	115.5
Hand (true)	73.8	Packaged goods	78.4
		Hand (true)	26.3
		Personal care (true)	2.2
		Days online	-17.8

5.3 Tree based models

To predict customer engagement, two more complex machine learning methods are performed. The first one being a single decision tree. The second method is an ensemble method, a random forest. Both methods are applied twice, first to predict the number of likes (3.1), second to predict the number of comments (3.2).

The data is split into a train and test set. The train set consist of 70% of the data and is used to train the models. The test set consists of the remaining 30% of the data and is used to measure the performance of the models. The createDataPartition function from the caret package in R is used to keep the distribution of customer engagement about equal in all sets. To avoid the results relying on the split in the data too much, five test sets are created. The performance of the models is measured as the average performance on these 5 test sets. All models are trained on the same training set and the performance is measured on the same five test sets. This is done to allow comparison of the models. Model performance is measured using the RMSE. This metric is also used to compare the models.

First, the number of likes a post receives is predicted using a decision tree. The single decision tree with a cost complexity parameter equal to zero has an RMSE of 171 likes on average. The decision tree is too big to visualize and interpret. The decision tree is pruned to improve predictions. The cost complexity parameter with the lowest cross-validation error is 0.014 (Figure G1, Appendix G). The root node splits the data into two groups, posts with a pixel percentage of 2.6 or higher and posts with a pixel percentage lower than 2.6 (Figure 6). If the pixel percentage is 2.6 or higher, the number of likes is on average 3,313 (4th terminal node). If the pixel percentage is lower than 2.6, the brand and number of packaged goods displayed determine the number of likes. If the post is posted by Alaffia, Facetheory or Meow Meow Tweet, the number of likes is lowest (on average 179, 1st terminal node). If Biossance or REN posted the photo and it is featuring 6 or more packaged goods, this will lead to 1763 likes on average (3rd terminal node). If one of these two brands posted the photo, with less than 6 packaged goods displayed, the number of likes is 646 on average (2nd terminal node). The RMSE of the pruned tree slightly improves relative to the single decision tree. The RMSE drops from 171 to 165.

In the same manner, two decision trees are built to predict comments. The single decision tree, with a cost complexity parameter equal to zero, is again too complex to visualize and interpret. The tree has a RMSE of 16. To improve the predictions, the tree is pruned with a cost complexity parameter equal to 0.012 (Figure G2, Appendix G). The root node of the tree is the number of tags used in the caption of the post (Figure 7). If a post has 5 or more tags in the caption, the average number of comments is 1,377 (4th terminal node). If a post has less than 5 tags, the number of packaged goods displayed determines the number of comments. Less than 5 tags in combination with less than 5 packaged goods displayed, means low customer engagement (on average 45 comments, 1st terminal node). If there are more than 5 packaged goods displayed, the number of hashtags is determinative. A caption that includes more than 1 hashtags leads to a high number of comments on average (1,000, 3rd terminal node), while less than 1 hashtags is leading to less comments (40, 2nd terminal node). The RMSE of the pruned tree is improved compared to the single decision tree. The RMSE drops from 16 to 15.8.

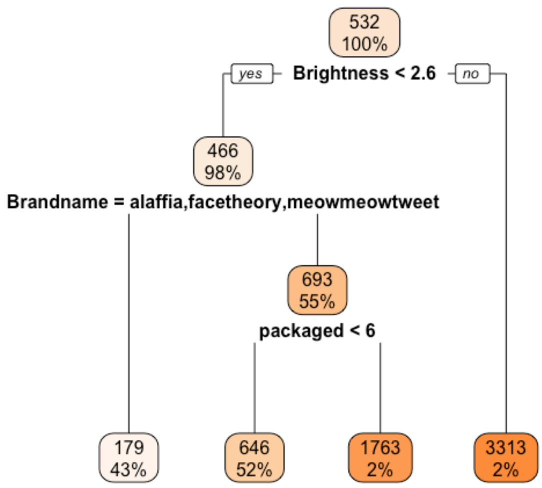


Figure 6. Pruned decision tree likes.
Notes: Visualization of the pruned decision tree predicting likes. The root node splits the data into two groups. The other splits divide the data further until the terminal nodes. The terminal nodes contain the predictions.

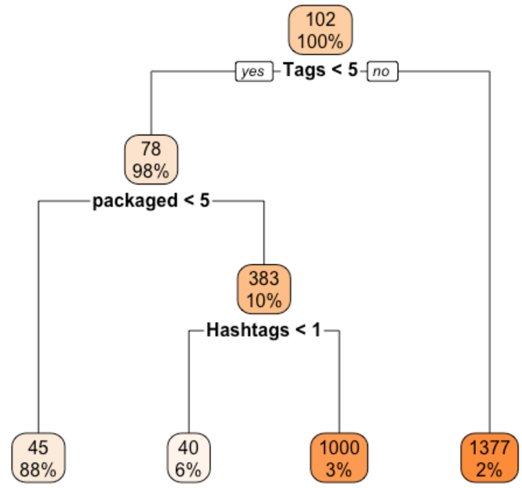


Figure 7. Pruned decision tree comments.
Notes: Visualization of the pruned decision tree predicting comments. The root node splits the data into two groups. The other splits divide the data further until the terminal nodes. The terminal nodes contain the Predictions.

Random forest is an ensemble method that constructs multiple decision trees at once. The method is used for a regression problem. The number of likes is predicted first (3.1), before predicting the number of comments (3.2). A grid search is used to find the optimal number of variables considered at each split (M). The optimal number of features is 2 for both models. An M (mtry parameter) of 2 results in the lowest RMSE for both the prediction of likes and comments (Figure C3 and C4, appendix C). This means that at each split, only 2 features are considered. The number of trees (ntree parameter) is equal to the default of 500 for both models since this provides the lowest error (Figure C1 and C2).

Since random forest is a black box model, it is unclear what happens internally in the model. How did the model predict the outcome values? To interpret the models, black box interpretation methods are used. To describe how the model behaves in general, feature importance, partial dependency and interaction is examined.

When predicting the number of likes, the variance explained by the random forest is equal to 9.1%. The mean of squared residuals is 1,764,029. The RMSE is 180. The feature importance presents the most important variables for predicting the number of likes (Figure 8). The number of hashtags in the caption is the most important feature when predicting likes, but also brightness, the number of followers and the average likes of the brand are important. The type of post (photo/video) and if a face is included are the least important features in the random forest model.

To examine the relationship between the predictor variables and the dependent variable, partial dependence plots are constructed. First, the features with the highest importance are plotted. Second, the photo features that are of main interest. If a caption includes more hashtags, more likes are received up until 2.5 hashtags (Figure 9). After that, there is a small drop before flattening. There is not sufficient data to conclude if using more than 6 hashtags is recommended. Brightness is an indication of the pixel percentage in the image, a lower percentage means a more colorful photo. There is little data on images with a pixel percentage lower than 1.0 or higher than 2.5 (Appendix E). Therefore, these pixel percentages are not considered when measuring the relationship between color brightness and customer engagement. There is not a strong relationship between color brightness and customer engagement in terms of likes (Figure 10). It can be noted that a pixel percentage higher than 2.0 increases customer engagement slightly. Both the number of followers and average likes have a positive relation with the number of likes a new post receives, which is not surprising.

The main variables of interest are products, hands, faces, tags, and type of content. Product is a variable indicating if a product is visible. Images including products have a slightly higher number of likes (from 521 to 527, Figure 11). Images including hands, arms or fingers have a much higher number of likes than images without (from 510 to 600, Figure 12). Photos that feature a human face, have about 20 more likes than photos without (Figure 13). Tags are indicators of influencer-related posts. The more tags are linked to a photo, the higher the number of likes is (Figure 14). Also, the type of content displayed is defining the number of likes. Advertorial and informational posts receive relative more likes than remunerational and entertainment posts. Advertorial post on average 530 likes, informational post on average 525, remunerational post on average 508 and entertainment posts below 490 on average (Figure 15).

Feature importance and partial dependency can be influenced by interaction between the predictor variables. To account for interaction between the predictor variables, the overall interaction between the variables is plotted (Figure 16). The number of hashtags highly interacts with other variables (> 60%). For this feature, the interaction relative to all other variables is also examined. Hashtags mostly interacts with the average number of likes and followers a brand has (Appendix C5).

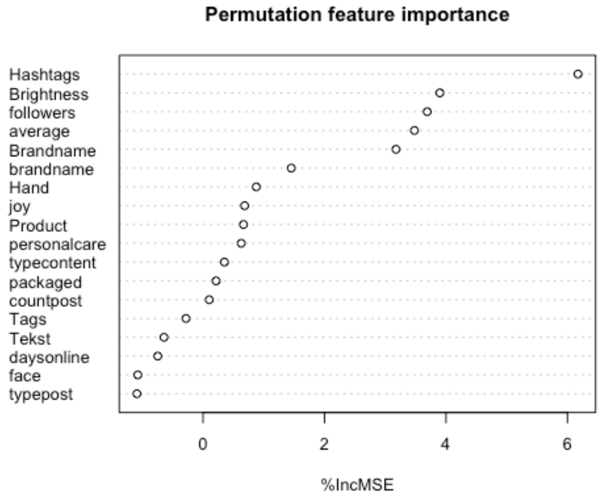
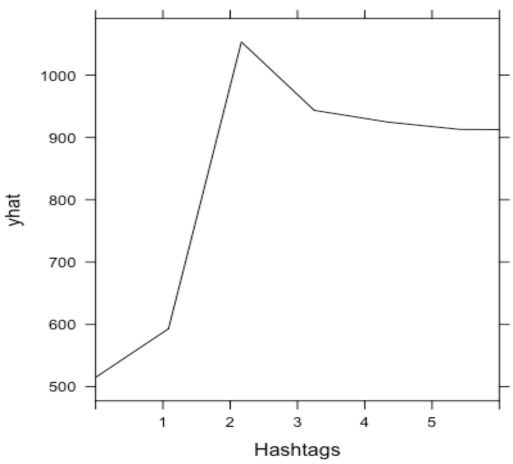


Figure 8. Feature importance.
Notes: This figure displays the variable importance of the predictor variables of number of likes. The variable importance is obtained by applying interpretation methods on the random forest model. The method measures variable importance as the decrease in model score after shuffling the value of one feature. The drop in performance indicates the importance of that one feature.



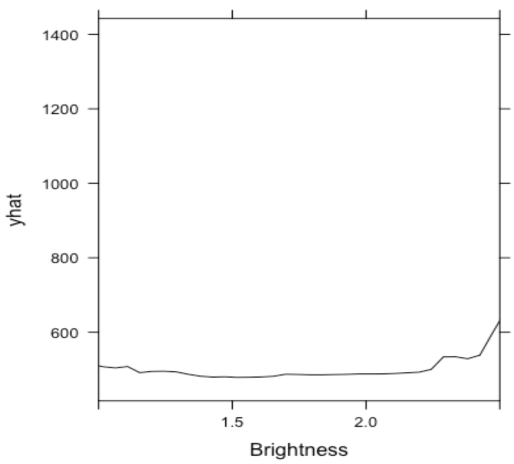


Figure 9. Partial dependence plot.
Hashtags plotted against likes.

Figure 10. Partial dependence plot.
Color brightness plotted against likes.

Notes: The relationship between multiple predictor variables and the dependent variable, number of likes, are uncovered. The y-axis represents the dependent variable, the x-axis the predictor variable. The plots are partial dependence plots, a form of interpretation methods for black-box models.

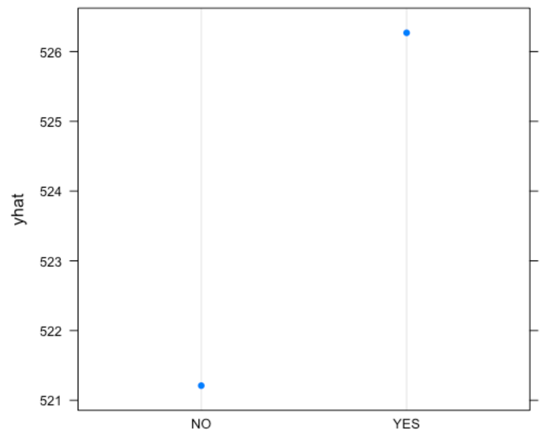


Figure 11. Partial dependence plot.
Product plotted against likes.

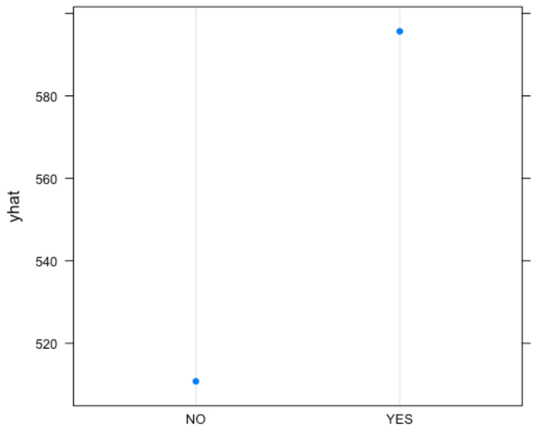


Figure 12. Partial dependence plot.
Hands plotted against likes.

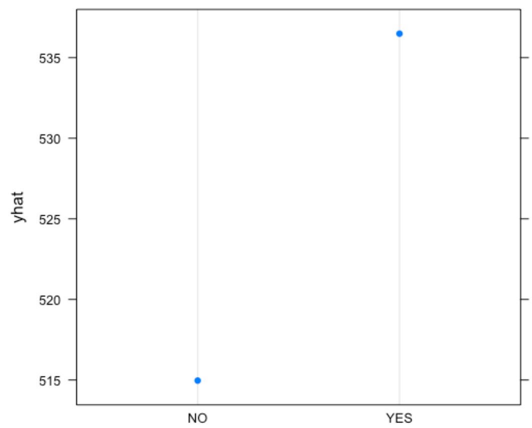


Figure 13. Partial dependence plot.
Face plotted against likes.

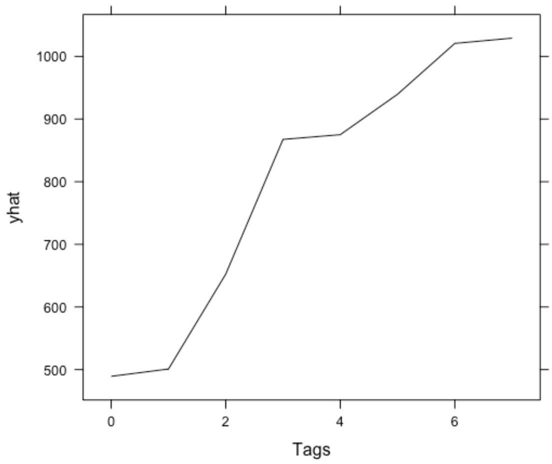


Figure 14. Partial dependence plot.
Tags plotted against likes.

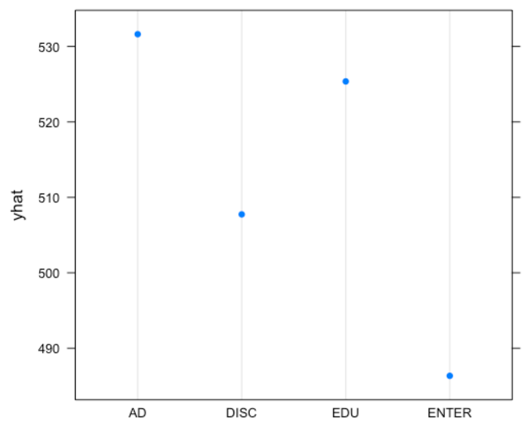


Figure 15. Partial dependence plot.
Content type plotted against likes.

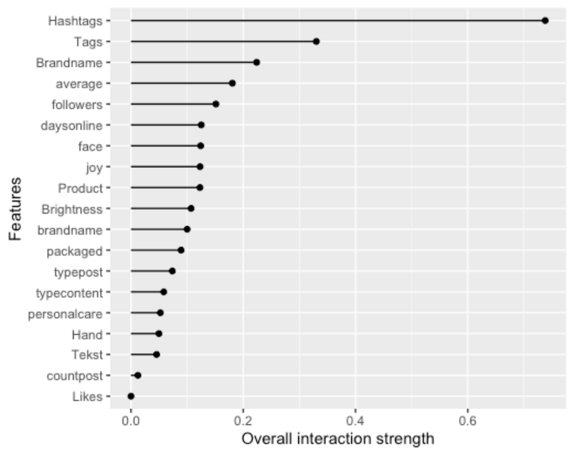


Figure 16. Interaction plot.
Interaction predictor variables.

The random forest that models the number of comments has a variance explained of 14.9%. The mean of squared residuals is 213,461. The RMSE is 8 on average. Again, global interpretation methods are applied. Feature importance indicates that the brand name in the caption and tags are the most important features when predicting the number of comments (Figure 17). The number of average likes, followers and hashtags are also important. When predicting comments, the type of post is relevant, contrary to when predicting the number of likes (Figure 8). The variables of least importance are product, face, and packaged goods.

The relationship between the predictor variables and the dependent variable is examined by plotting the partial dependency. Including the brand name in the caption leads to more customer engagement in terms of comments (Figure 18). It increases the comment count by 50 comments on average. The number of followers and average likes again have a positive effect on customer engagement. Again, there is little data on images with a pixel percentage lower than 1.0 or higher than 2.5 (Appendix E). Therefore, these pixel percentages are not considered. For comments it holds that a pixel percentage above 2.0 leads to high customer engagement, meaning colorless images attract response (Figure 19).

Relating an image to an influencer, by using tags, causes the comment count to increase (Figure 20). The placement of a hand or product also increases the number of comments; however, a human face decreases the number of comments. The placement of a hand, arm or finger is expected to increase comments by 60 (Figure 21). The placement of a product is expected to increase comments by 8 (Figure 22), while a face decreases comments by 4 (Figure 23). For comments it holds that posts that are intended for advertisement or remunerational purposes are performing best. It might seem surprising that remunerational posts are performing this good, but often a condition to receive value is to comment on the post (Figure 24).

The interaction between all predictor variables is considered again (Figure 25). Average number of likes interacts most with all other variables, especially with number of hashtags (Appendix C6). However, the interaction is of relatively small size (< 35%).

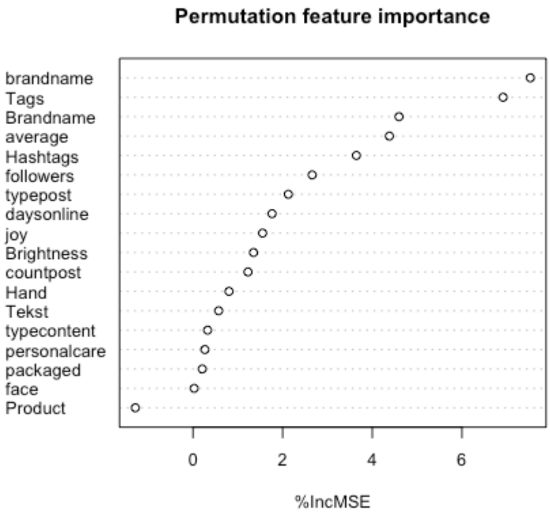


Figure 17. Feature importance

Notes: This figure displays the variable importance of the predictor variables of number of comments. The variable importance is obtained by applying interpretation methods on the random forest model. The method measures variable importance as the decrease in model score after shuffling the value of one feature. The drop in performance indicates the importance of that one feature.

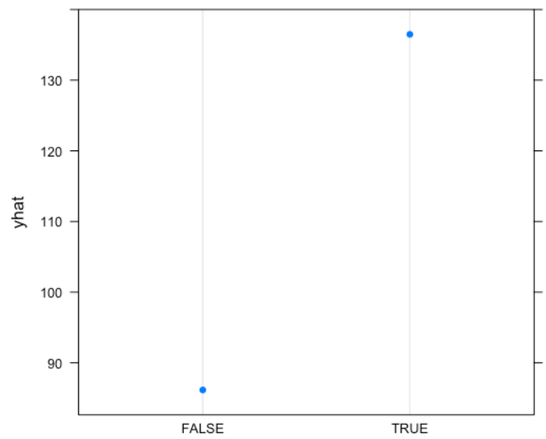


Figure 18. Partial dependence plot.

Brand name in caption plotted against comments.

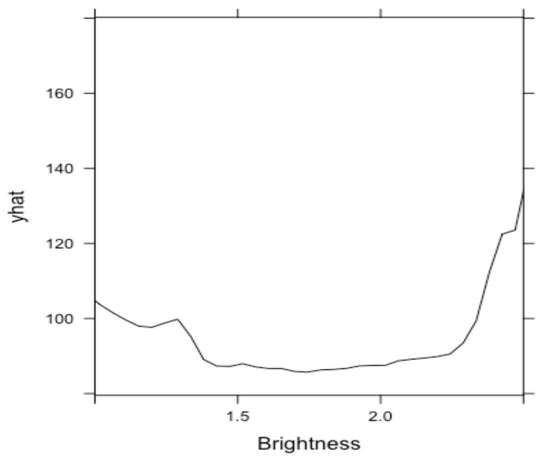


Figure 19. Partial dependence plot.

Brightness plotted against comments.

Notes: The relationship between multiple predictor variables and the dependent variable, number of comments, are uncovered. The y-axis represents the dependent variable, the x-axis the predictor variable. The plots are partial dependence plots, a form of interpretation methods for black-box models.

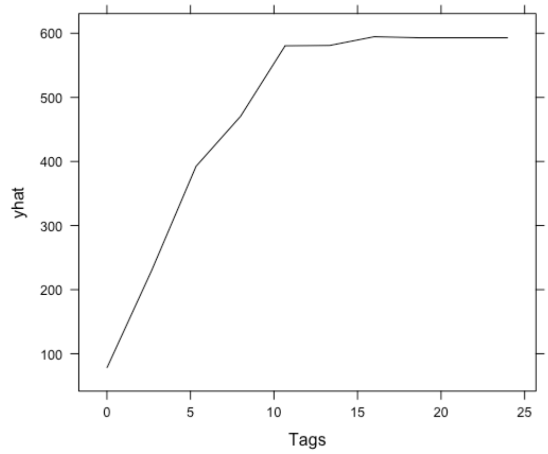


Figure 20. Partial dependence plot.
Tags plotted against comments.

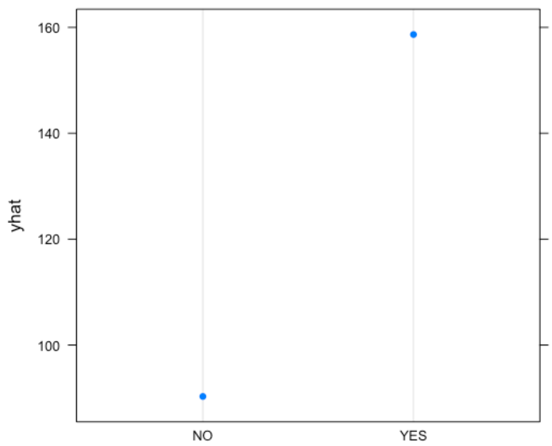


Figure 21. Partial dependence plot.
Hand plotted against comments

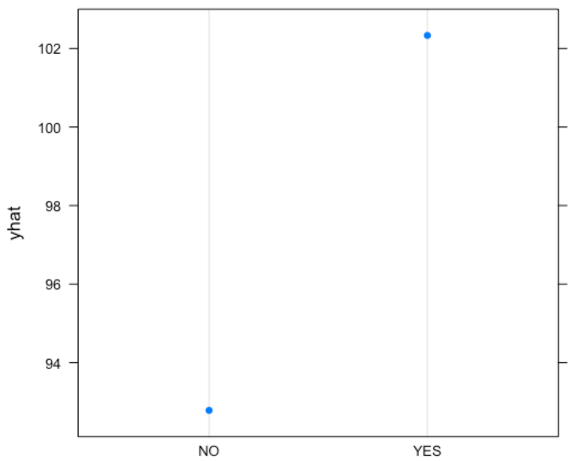


Figure 22. Partial dependence plot.
Product plotted against comments.

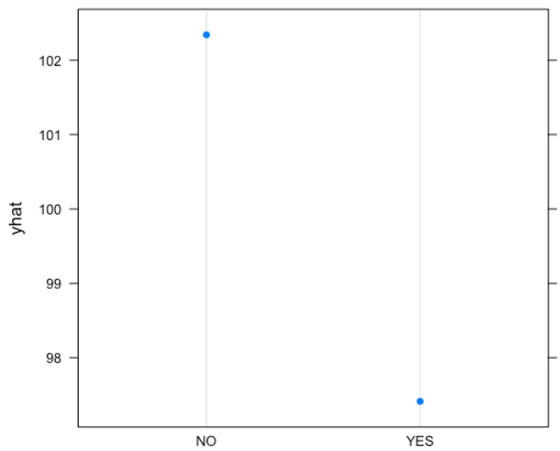


Figure 23. Partial dependence plot.
Face plotted against comments.

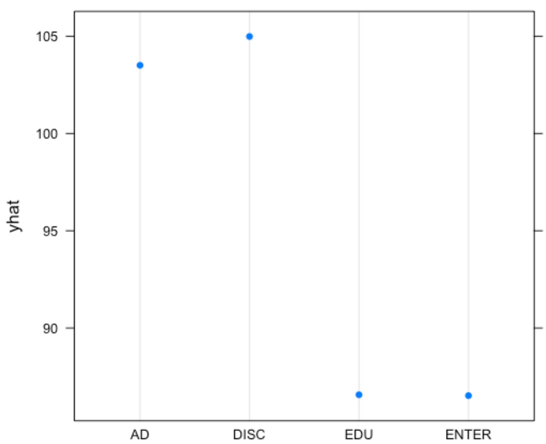


Figure 24. Partial dependence plot.
Content type plotted against comments.

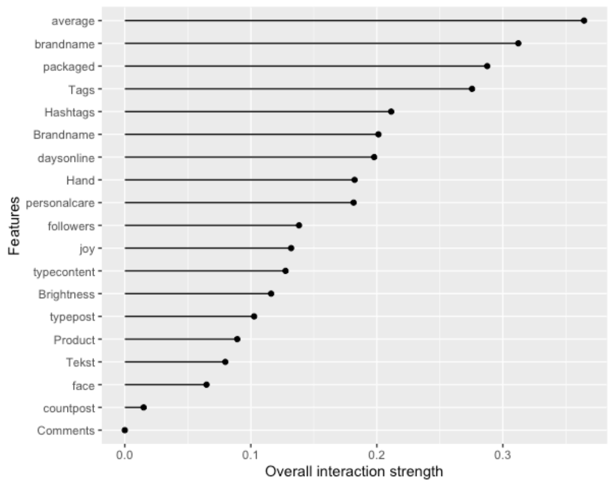


Figure 25. Interaction plot.
Interaction predictor variables.

5.4 Interpretation methods

The main goal of this study is to measure which image features influence customer engagement positively for eco-friendly products. The decision trees together with the global interpretation methods best describe the relationships between the predictor variables and customer engagement.

Brand popularity, measured as average number of likes and followers, has a positive relation to customer engagement for eco-friendly products. This result indicates that building a brand name is valuable for eco-friendly companies. This finding supports the finding of Ewe & Tjiptono (2023), who suggest that when consumers are more familiar with an eco-friendly brand, that their buying intention and willingness to pay is significantly higher relative to non-eco-friendly brands. The featuring of the brand name in the caption is also positively related to customer engagement for eco-friendly products. This is opposite to the findings for non-eco-friendly brands, where prominent brand name placement hurts the brand (Tellis et al., 2019).

For customer engagement in terms of likes, it is found that a pixel percentage of 2.0 or higher slightly increases the number of likes. A higher pixel percentage means less color and thus colorless images receive more likes. Colorless pictures (pixel percentage ≥ 2.0) also receive a lot of comments, as well do colorful images (pixel percentage < 1.5). This study concludes that monotonic images receive more customer engagement, especially in terms of comments. This result is contrary to the findings for non-eco-friendly products (Li & Xie, 2020; Yu, Xie & Wen, 2020). However, it is not as surprising. Samaraweera et al. (2021) find that white-toned labels represent a high-quality look and increases willingness to pay for eco-friendly products. This high-quality look could be the reason why consumers respond better to monotonic images.

Displaying products and packaged goods is positively related to customer engagement for eco-friendly products. Also, featuring hands, arms and/or fingers has a positive effect on customer engagement for eco-friendly products. Displaying human faces positively influences the number of likes, but negatively influences comments. This study finds that for eco-friendly products, displaying hands, arms and fingers is most effective, followed by products. Companies can use this insight to their advantage, considering that the featuring of hands, arms and fingers is currently quite low (Table 8). This result is opposite to the findings for non-eco-friendly brands. For non-eco-friendly products, it holds that human faces increase customer engagement most relative to products and hands (Hartmann et al., 2019).

Influencer-related posts, measured through tags, receive more likes and comments for eco-friendly products. This confirms the findings for non-eco-friendly products (Lou et al., 2019; Tellis et al., 2019) and eco-friendly products (dos Santos et al., 2023). Lou et al (2019) found that influencer-promoted ads perform significantly better in terms of customer engagement for the top 50 apparel companies in the US. These companies are mostly from the clothing industry. Dos Santos et al. (2023) found the same results for vegan consumption, while Tellis et al (2019) did for 79 brands from different non-eco-friendly industries.

This study finds that advertorial and informational posts perform best to attract likes. Advertorial and remunerational posts do for comments. Remunerational posts have interaction purposes and often require a comment to participate. This could explain the result of remunerational posts leading to higher comments. Overall, it can be suggested that post intending to encourage and/or engage customers (through ads or remunerations) lead to higher customer engagement. There is a lot to gain here for these eco-friendly skincare brands, considering almost no content is remunerational (Table 8). For non-eco-friendly products, it is found that entertainment posts and remuneration posts receive more customer engagement (Luarn et al., 2015; Liu et al., 2019; Cortez et al., 2023). The results for eco-friendly versus non-eco-friendly products partly overlap, namely for both product categories remunerational posts perform well.

5.5 Comparison of the models

Four different machine learning models are built to predict customer engagement. The predictive performance of the decision trees and random forests are visualized in Figure 26 and 27. Figure 26 shows the actual values versus predicted values for the predictions of number of likes. Only predictions between 0 and 1000 likes are presented, to facilitate visualization. The predictions for over 1000 likes, can be found in the Appendix (Figure F1, Appendix F). The pruned decision tree only predicts four values (Figure 6), two being over 1000 (1762.0 and 3313.2). There seems to be little difference in the performance of the three models (Figure 26).

For the prediction of comments, the pruned decision tree also predicts four values (Figure 7). Again, two of these values (999.9 and 1376.8) are outside the scope of the figure (Figure F2, Appendix F). Again, there seems to be little difference in the performance of the three models (Figure 27). The figures visualize the predictions made earlier, which helps understanding the methods and results. Although the figures are quite interesting, it is not possible to derive the best suited methods. Because all models are trained on the same training and test sets, it is possible to compare the performance of the models by comparing the RMSE.

The models are tested on five test sets, that are the same for all models, to avoid the results being dependent on the distribution of the data.

The first model, the lasso regression, operates as the benchmark model. The performance of all other models is compared to this benchmark. The lasso regression predicts likes on average with an error of 167.6 (Table 11). This is quite high, considering the average likes being 625 (Table 6), which means that the error is on average 27% of the mean. The lasso regression predicts comments on average with an error of 20.8 (Table 12), only 20% error of the mean (Table 6). The second model is a single decision tree. The single decision tree predicting likes does not outperform the benchmark. The single decision tree has a RMSE of 171.4 on average compared to 167.6 for the benchmark. The single decision tree predicting comments outperforms the benchmark. The single decision tree is outperformed by the pruned decision tree both times (Table 11 and 12). The pruned decision tree predicts likes on average with an error of 165.3 (Table 11) and comments with an average error of 15.8 (Table 12). The pruned decision tree also outperforms the benchmark twice. When predicting likes, the pruned decision tree is the best suited model with the lowest average RMSE (165.3, Table 11). This means that on average the predicted values are 165 likes from the actual values. This is still quite high, being 26% of the mean (Table 6). The pruned decision tree is considered a simpler method, which makes it easier to interpret. Since both the accuracy and interpretability are good, the pruned decision tree is the preferred model to predict likes. The best model to predict comments is a random forest. This model predicts the number of comments with an average error of 8.1 (Table 12). This is a relatively good prediction model, only predicting comments with an error of 8% of the mean (Table 6). The results of this study show that predicting the number of comments is done better, with an error of only 8% of the mean compared to 26% when predicting likes.

Table 11. Modelling likes.

Notes: All models are trained and tested on the same sets. This allows for comparison of the models. The models predict the number of likes a photo receives. The benchmark model is a lasso regression. The other models are a single decision tree, pruned decision tree and random forest. The performance of the models is measured using RMSE. This metric is also used to compare the models. This table summarizes the performance of all models.

Model		**RMSE**	
Likes	**Min**	**Average**	**Max**
Benchmark (lasso)	52.9	167.6	282.1
Single decision tree	9.9	171.4	284.7
Pruned decision tree	14.9	165.3	290.75
Random Forest	3.2	180.2	322.0

Table 12. Modelling comments.

Notes: All models are trained and tested on the same sets. This allows for comparison of the models. The models predict the number of comments a photo receives. The benchmark model is a lasso regression. The other models are a single decision tree, pruned decision tree and random forest. The performance of the models is measured using RMSE. This metric is also used to compare the models. This table summarizes the performance of all models.

Model	**RMSE**		
Comments	**Min**	**Average**	**Max**
Benchmark (lasso)	10.2	20.8	47.3
Single decision tree	2.4	16.3	23.8
Pruned decision tree	7.9	15.8	22.9
Random Forest	0.5	8.1	19.6

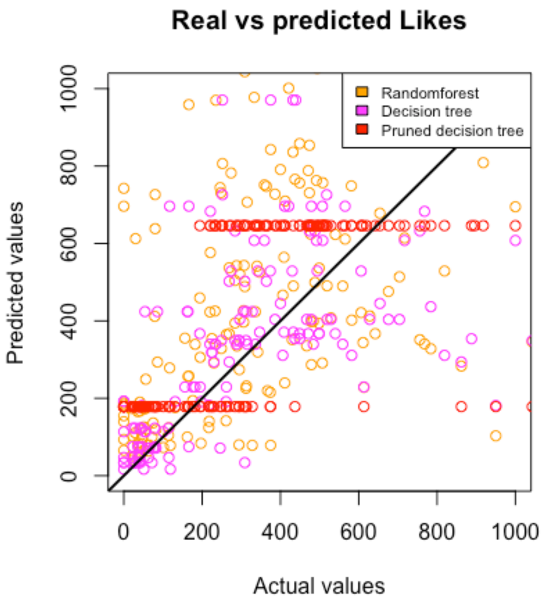


Figure 26. Predicted values likes.

Notes: The graph shows the actual values plotted against the predicted values. The x-axis presents the actual values. The y-axis presents the predicted values. The line indicates predictions with a 100% accuracy. The orange dots represent the predictions of the random forest, the pink the decision tree and the red the pruned decision tree.

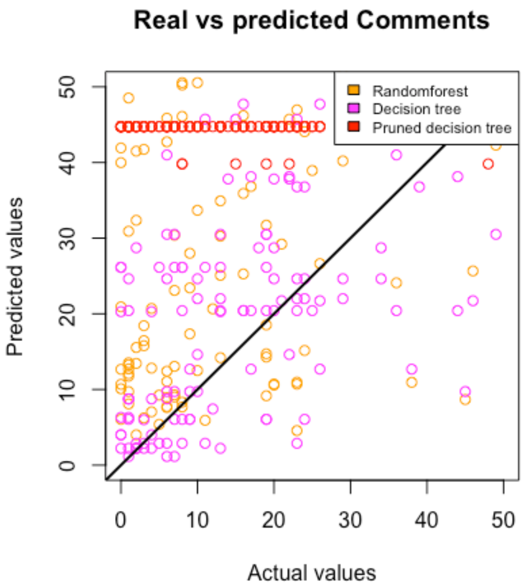


Figure 27. Predicted values comments.
Notes: The graph shows the actual values plotted against the predicted values. The x-axis presents the actual values. The y-axis presents the predicted values. The line indicates predictions with a 100% accuracy. The orange dots represent the predictions of the random forest, the pink the decision tree and the red the pruned decision tree.

To test if the models statistically differ, an Analysis of variance test and Kruskal-Wallis test are performed (Yusuf, 2020; Chavaltada, Pasupa & Hardoon, 2017). To get some first insights, the spread of model performance is visualized (Figure 28 and 29). From the boxplots, it can be assumed that the models are relatively close in performance. To check if the performance of the models is statistically different, statistical tests are performed.

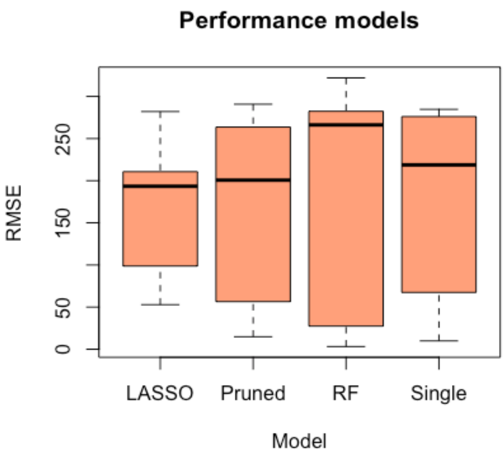


Figure 28. Boxplot likes.
Notes: The spread of model performance for all models that predict likes. The model performance is measured using RMSE.

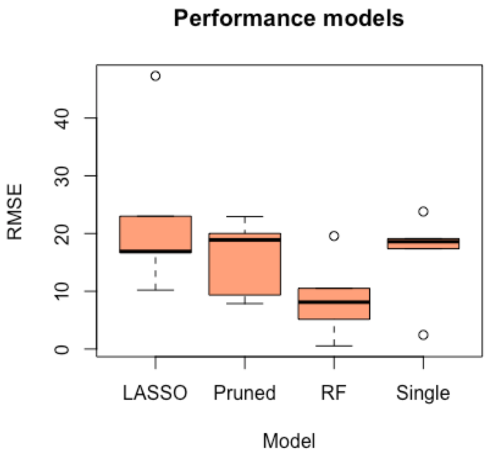


Figure 29. Boxplot comments.
Notes: The spread of model performance for all models that predict comments. The model performance is measured using RMSE.

Analysis of Variance, also referred to as ANOVA, is performed to test if the models statistically differ. Before performing the analysis, the assumptions of ANOVA are checked. The first assumption is that each factor, must be randomly sampled. If this assumption is met, is debatable. The second assumption states that all factors must be independent. This assumption is satisfied. The third assumption requires the data to be normally distributed with equal variances. To check this assumption, three methods are used: a density plot, Q-Q plot, and Shapiro-Wilk's test. All methods are applied to the prediction of likes and to the prediction of comments. The density plot and Q-Q plot indicate that the RMSE's of the models predicting likes are not normally distributed (Appendix D). The hypotheses for the Shapiro-Wilk's test are as follows:

H0: The data is normally distributed.
Ha: The data is not normally distributed.

The p-value of the Shapiro-Wilk's test is 0.008, which is smaller than the significance level of 0.05. This means that the null hypothesis, that there is a normal distribution, is rejected.

The density plot and Q-Q plot corresponding to the predictions of comments indicate that the data is not normally distributed (Appendix D). The p-value of the Shapiro-Wilk's test is smaller than 0.05 (0.002) meaning that the null-hypothesis is rejected again.

For both measures of customer engagement, the RMSE's are not normally distributed. A violation of the assumptions is not necessarily problematic. What matters is if the validity of the results is affected (Glass, Peckham & Sanders, 1972). The violation of normal distribution mostly effects the type I errors (false-positive). This means that a not-normal distribution can cause a researcher to reject the null hypothesis too soon (Lix, Keselman & Keselman, 1996). To control for this possibility, Kruskal-Wallis tests are performed as well. This is a nonparametric test that does not require a normal distribution. The ANOVA test and Kruskal-Wallis test have the same hypotheses:

H0: There is no statistically significant difference between the performance of the models.
Ha: There is a statistically significant difference between the performance of the models.

For the models that predict the number of likes, the p-value of the ANOVA test is equal to 0.0004 (< 0.05). Therefore, the null hypothesis is rejected, which states that there is no statistically significant difference between the performance of the models. To control for possible false positives, the Kruskal-Wallis test is performed as well. The p-value of this test is equal to 0.0009 (< 0.05), meaning that again the null hypothesis is rejected again. Both tests indicate that there might be a statically difference between the performance of the model's predicting likes.

Both tests are also applied to compare the performance of the model's predicting comments. The p-value of the ANOVA test is equal to 0.002 (< 0.05). The null hypothesis is rejected. The p-value of the Kruskal-Wallis test is 0.003 (< 0.05). Again, the null hypothesis is rejected. It can be assumed that there is a statistically difference between the performance of the model's predicting comments.

This section discussed the results of four different methods: a linear model, lasso model, decision tree and random forest. The pruned decision tree is the best method to predict the number of likes. The number of comments is best predicted using a random forest. The performance of the models is assumed to be statistically significant different. The average RMSE of the pruned decision tree predicting likes is 165.3. The average RMSE of the random forest predicting comments is 8.1. The image features that impact customer engagement positively are hands, arms and/or fingers, products, and influencers. Faces only positively influence likes, not comments. The number of likes increases most when posting advertisements or informative images. The number of comments increases most when stimulating users to interact (remunerational posts) and encouraging to buy (advertisements). Color also affects customer engagement, monotonic photos receives more customer engagement relative to colorful images.

6 Discussion

The aim of this study is to compare the best social media marketing strategy for eco-friendly products relative to non-eco-friendly products. This study finds what attributes in social media posts lead to higher customer engagement for eco-friendly skincare companies on Instagram. This study finds that monotonic photos, displaying hands, that are influencer-related and are posted for advertisement purposes by popular brands, have the highest expected customer engagement for eco-friendly products. For non-eco-friendly products, it is assumed that using color and visible consumers, yields the highest expected customer engagement. Posts that entertain social media users, perform better for non-eco-friendly products. There is a similarity between the marketing of eco-friendly products versus non-eco-friendly products, relating influencers to your brand or post positively influences customer engagement for both.

The first hypothesis, stating that images featuring human faces receive higher customer engagement relative to images featuring hands and/or products, must be rejected for eco-friendly products. For both measures of customer engagement, it holds that hands, arms and/or fingers increase customer engagement more than faces and products do. This differs from the result found for non-eco-friendly foods, which states that featuring consumer faces is more effective compared to hands or products (Hartmann et al., 2019).

The second hypothesis suggests that brighter images receive higher customer engagement. This study finds that for eco-friendly skincare products it holds that monotonic images yield the highest customer engagement, especially in terms of comments (pixel percentage ≥ 2.0). The hypothesis must therefore be rejected. This result is opposite to the findings for destination photos in travel guides, airlines, and utility vehicle companies (Yu, Xie & Wen, 2020; Li & Xie, 2020).

The third hypothesis states: "Posts related to influencers receive higher customer engagement". The third hypothesis is accepted, meaning that for both eco-friendly and non-eco-friendly products, influencers are positively related to customer engagement.

The last hypothesis suggests that posts for entertainment and interaction purposes receive higher customer engagement. The results of this study indicate that posts with selling purposes (ads) receive the highest customer engagement for eco-friendly products. This does not align with the hypothesis, that is based on the findings for non-eco-friendly products. For luxury brands, selling non-eco-friendly products, it is found that entertainment and interaction marketing are most effective (Liu et al., 2021). For foods, it is found that seeking interaction with customers, is most effective (Luarn et al., 2015). This study did find that posts for interaction purposes lead to high comments, not necessarily likes. The hypothesis is therefore only partly accepted.

The research question is: "*What attributes in social media posts lead to higher customer engagement for eco-friendly skincare products?*". To answer this question, it can be stated that posts posted by popular brands on Instagram including less colorful images, that feature hands, are influencer-related and are posted for selling purposes, yield the highest expected customer engagement. These are the features eco-friendly and sustainable skincare companies should focus on when building new social media marketing campaigns. This is different to non-eco-friendly products for whom colorful images, that feature faces, and are posted for entertainment purposes receive the highest customer engagement. However, the different product types do agree on two things, namely influencer-relatedness influences customer engagement positively and posts for interaction purposes (remunerational posts) perform very well.

This paper fills a gap in the literature by comparing social media marketing for eco-friendly products relative to non-eco-friendly products. The paper did find some differences between the marketing of eco-friendly versus non-eco-friendly. The discovery of these differences (and similarities) can be of great value to marketers. Eco-friendly brands can now shape their advertisements to the preferences of their (green) consumers. They should focus on more clean images, showing only hands holding products. Additionally, monotonic images perform better since they represent quality and a high-end look. This paper thus not only contributes to academics, but also to practice.

Like any research, this research has some limitations. When collecting the data, videos are transformed into single photos. A single photo cannot capture the full content of an entire video and therefore this can cause some bias. For slideshows (Appendix A), the same problem arises. Only the first photo of a slideshow is considered. This can cause some bias as well. The content of the posts is assigned manually. Although the researcher had clear divisions, it can also cause some bias. Another limitation of this research is that customer engagement can still grow or less common but possible, decline. The number of likes and comments is always subject to change. Given that the number of average likes and comments is relatively low for the brands used, it is not expected to cause issues.

For the methods, all methods applied are supervised learning methods, meaning that the outcome is needed. The decision trees lack smoothness and are unstable. Random forest is a more robust method. However, this method is harder to interpret. The methods used to interpret random forest are sensitive to highly correlated features. When features highly correlate, the results of feature importance and partial dependence plots can be biased. Interaction is checked and expected not to cause big issues. Also, one should be careful drawing conclusions from these interpretation methods. The methods give a general overview rather than a detailed look. The results show that overall variance explained for by the models is low.

Some recommendations for future research are to use a bigger and less recent dataset. The variance explained of the models is relatively low. Future research could focus on what else contributes to the prediction of customer engagement. Future research could also add content or sentiment analysis of the comments to get a broader idea of the content of the comments. One last recommendation for future research is to measure the direct effect between sales and image features. For this, the researcher must have access to sales data of companies.

References

Aichner, T. (2019). Football clubs' social media use and user engagement. *Marketing Intelligence & Planning*, *37*(3), 242-257.

Amberg, N., & Fogarassy, C. (2019). Green consumer behavior in the cosmetics market. *Resources*, *8*(3), 137.

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

Banytė, J., Brazionienė, L., & Gadeikienė, A. (2010). Investigation of green consumer profile: a case of Lithuanian market of eco-friendly food products. *Ekonomika ir vadyba*, (15), 374-383.

Barbarossa, C., & de Pelsmacker, P. (2016). Positive and Negative Antecedents of Purchasing Eco-friendly Products: A Comparison Between Green and Non-green Consumers. *Journal of Business Ethics, 134*, 229-247.

Barger, V., Peltier, J. W., & Schultz, D. E. (2016). Social media and consumer engagement: a review and research agenda. *Journal of Research in Interactive Marketing*, *10*(4), 268-287.

Breiman, L. (1996). Bagging predictors. *Machine learning*, *24,* 123-140.

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

Calder, B. J., Malthouse, E. C., & Maslowska, E. (2016). Brand marketing, big data and social innovation as future research directions for engagement. *Journal of Marketing Management*, *32*(5-6), 579-585.

Chapman, C., & Feit, E.M. (2015). *R for marketing research and analytics*. New York, NY: Springer.

Charter, M., Peattie, K., Ottman, J., & Polonsky, M. J. (2002). Marketing and sustainability. *Centre for Business Relationships, Accountability, Sustainability and Society (BRASS) in association with The Centre for Sustainable Design, April*, *324*.

Chavaltada, C., Pasupa, K., & Hardoon, D.R. (2017, June 21-26). *A comparative study of machine learning techniques for automatic product categorisation* [Paper presentation]. IEEE 14th International Symposium, Hokkaido, Japan.

Chin, J., Jiang, B. C., Mufidah, I., Persada, S. F., & Noer, B. A. (2018). The investigation of consumers' behavior intention in using green skincare products: a pro-environmental behavior model approach. *Sustainability*, *10*(11), 3922.

Cochrane, J.H. (2005). Writing tips for Ph. D. students.

Corone, A., Nanne, A. J., & van Miltenburg, E. (2021, October 28-29). *Controlling Social Media Data: a Case Study of the Effect of Social Presence on Consumers' Engagement with Brand-generated Instagram Posts.* Conference on Computer-Mediated Communication CMC and Social Media Corpora, Nijmegen, Netherlands.

Cortez, R. M., Johnston, W. J., & Dastidar, A. G. (2023). Managing the content of LinkedIn posts: Influence on B2B customer engagement and sales? *Journal of Business Research*, *155*(3), 113388.

Cvijiki, I.P., & Michahelles, F. (2011, December 12-14). *Monitoring trends on facebook* [Paper presentation]. IEEE Ninth International Conference on Dependable, Autonomic and Secure Computing, Sydney, Australia.

De Vries, L., Gensler, S., & Leeflang, P.S. (2012). Popularity of brand posts on brand fan pages: An investigation of the effects of social media marketing. *Journal of interactive marketing, 26*(2), 83-91.

Delbaere, M., & McQuarrie, E.F., & Phillips, B.J. (2011). Personification in advertising. *Journal of Advertising, 40*(1), 121-130.

Dewi, W. W. A., Avicenna, F., & Meideline, M. M. (2020). Purchase intention of green products following an environmentally friendly marketing campaign: results of a survey of instagram followers of InnisfreeIndonesia. *Asian Journal for Public Opinion Research*, *8*(2), 160-177.

Dolan, R., Conduit, J., Frethey-Bentham, C., Fahy, J. & Goodman, S. (2019). Social media engagement behavior. A framework for engaging customer through social media content. *European Journal of Marketing, 53*(10), 2213-2243.

dos Santos, R. C., de Brito Silva, M. J., da Costa, M. F., & Batista, K. (2023). Go vegan! digital influence and social media use in the purchase intention of vegan products in the cosmetics industry. *Social Network Analysis and Mining*, *13*(1), 49.

Dwivedi, Y.K., Ismagilova, E., Hughes, D.L., Carlson, J., Filieri, R., Jacobson, J., Jain, V., Karjaluoto, H., Kefi, H., Krishen, A., Kumar, V., Rahman, M.M., Raman, R., Rauschnabel, R.A., Rowley, J., Salo, J., Tran, G.A., & Wang, Y. (2021). Setting the future of digital and social media marketing research: Perspectives and research propositions. *International Journal of Information Management, 59,* 102168.

Ewe, S.Y., & Tjiptono, F. (2023). Green behavior among Gen Z consumers in an emerging market: eco-friendly versus non-eco-friendly products. *Young Consumers, 24*(2), 234-252.

Fisher, A., Rudin, C., & Dominici, F. (2019). All Models are Wrong, but Many are Useful: Learning a Variable's Importance by Studying an Entire Class of Prediction Models Simultaneously. *J. Mach. Learn. Res., 20*(177), 1-81.

Freedman, D.A. (2009). *Statistical models: theory and practice.* Cambridge University Press.

Friedman, J.H. (2001). Greedy function approximation: a gradient boosting machine. *The annals of Statistics, 29*(5), 1189-1232.

Friedman, J.H., & Popescu, B.E. (2008). Predictive Learning via Rule Ensembles. *The Annals of Statistics*, *3*(2), 916-954.

Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological cybernetics, 36*(4), 193-202.

Gareth James, D. W. (2013). *An introduction to statistical learning: with applications in R.* New York: Springer.

Giglio, S., Pantano, E., Bilotta, E., & Melewar, T. C. (2020). Branding luxury hotels: Evidence from the analysis of consumers' "big" visual data on TripAdvisor. *Journal of business research*, *119*, 495-501.

Gill, M., Sridhar, S., & Grewal, R. (2017). Return on engagement initiatives: A study of a business-to-business mobile app. *Journal of marketing*, *81*(4), 45-66.

Glass, G.V., Peckham, P.D., & Sanders, J.R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. *Review of educational research, 42*(3), 237-288.

Grappe, C. G., Lombart, C., Louis, D., & Durif, F. (2021). "Not tested on animals": how consumers react to cruelty-free cosmetics proposed by manufacturers and retailers? *International Journal of Retail & Distribution Management*, *49*(11), 1532-1553.

Hartmann, J., Heitmann, M., Schamp, C., & Netzer, O. (2020). The power of brand selfies in consumer-generated brand images. *Journal of Marketing Research, 58*(6), 1159-1177.

Hsu, C. L., Chang, C. Y., & Yansritakul, C. (2017). Exploring purchase intention of green skincare products using the theory of planned behavior: Testing the moderating effects of country of origin and price sensitivity. *Journal of retailing and consumer services*, *34*, 145-152.

Hollebeek, L. D., Glynn, M. S., & Brodie, R. J. (2014). Consumer brand engagement in social media: Conceptualization, scale development and validation. *Journal of interactive marketing*, *28*(2), 149-165.

Jaini, A., Quoquab, F., Mohammad, J., & Hussin, N. (2020). “I buy green products, do you…?”: The moderating effect of eWOM on green purchase behavior in Malaysian cosmetics industry. *International Journal of Pharmaceutical and Healthcare Marketing*, *14*(1), 89-112.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning*. New York: Springer.

Johri, L. M., & Sahasakmontri, K. (1998). Green marketing of cosmetics and toiletries in Thailand. *Journal of consumer marketing*, *15*(3), 265-281.

Jones, P., Clarke-Hill, C., Comfort, D., & Hillier, D. (2008). Marketing and sustainability. *Marketing intelligence & planning*, *26*(2), 123-130.

Jung, N., & Im, S. (2021). The mechanism of social media marketing: influencer characteristics, consumer empathy, immersion, and sponsorship disclosure. *International Journal of Advertising, 40*(8), 1265-1293.

Luger, G. F. (2005). *Artificial intelligence: structures and strategies for complex problem solving*. Pearson education.

Kim, Y.H., & Chung, J. E. (2011). Consumer purchase intention for organic personal care products. *Journal of consumer Marketing*, *28*(1), 40-47.

Kong, H.M., Witmaier, A., & Ko, E. (2021). Sustainability and social media communication: How consumers respond to marketing efforts of luxury and non-luxury fashion brands. *Journal of Business Research, 131*, 640-651.

Kozinets, R.V., De Valck, K., Wojnicki, A.C., & Wilner, S.J. (2010). Networked narratives: Understanding word-of-mouth marketing in online communities. *Journal of marketing, 74*(2), 71-89.

Kumar, P., & Ghodeswar, B. M. (2015). Factors affecting consumers’ green product purchase decisions. *Marketing Intelligence & Planning*, *33*(3), 330–347.

Labrecque, L.I., & Milne, G.R. (2012). Exciting red and competent blue: the importance of color in marketing. *Journal of the Academy of Marketing Science, 40*(5), 711-727.

Lamberton, C., & Stephen, A. T. (2016). A thematic exploration of digital, social media, and mobile marketing: Research evolution from 2000 to 2015 and an agenda for future inquiry. *Journal of marketing*, *80*(6), 146-172.

Laroche, M., Bergeron, J. and Barbaro-Forleo, G. (2001), "Targeting consumers who are willing to pay more for environmentally friendly products", *Journal of Consumer Marketing, 18(6),* 503-520.

Lavuri, R., Jabbour, C. J. C., Grebinevych, O., & Roubaud, D. (2022). Green factors stimulating the purchase intention of innovative luxury organic beauty products: Implications for sustainable development. *Journal of Environmental Management*, *301*, 113899.

Leek, S., Houghton, D., & Canning, L. (2019). Twitter and behavioral engagement in the healthcare sector: An examination of product and service companies. *Industrial Marketing Management*, *81*, 115-129.

Liu, X., Lee, D., & Srinivasan, K. (2019). Large-scale cross-category analysis of consumer review content on sales conversion leveraging deep learning. *Journal of Marketing Research*, *56*(6), 918-943.

Liu, X., Shin, H., & Burns, A. C. (2021). Examining the impact of luxury brand's social media marketing on customer engagement: Using big data analytics and natural language processing. *Journal of Business research*, *125*, 815-826.

Li, Y., & Xie, Y. (2020). Is a picture worth a thousand words? An empirical study of image content and social media engagement. *Journal of Marketing Research*, *57*(1), 1-19.

Lix, L.M., Keselman, J.C., & Keselman, H.J. (1996). Consequences of assumption violations revisited: A quantitative review of alternatives to the one-way analysis of variance F test. *Review of educational research, 66*(4), 579-619.

Lou, C., Tan, S. S., & Chen, X. (2019). Investigating consumer engagement with influencer- vs. brand-promoted ads: The roles of source and disclosure. *Journal of Interactive Advertising*, *19*(3), 169-186.

Lu, L., Bock, D., & Joseph, M. (2013). Green marketing: what the Millennials buy. *Journal of Business Strategy, 24*(6), 3-10.

Luarn, P., Lin, Y. F., & Chiu, Y. P. (2015). Influence of Facebook brand-page posts on online engagement. *Online Information Review*, *39*(4), 505-519.

Ma, L., & Sun, B. (2020). Machine learning and AI in marketing–Connecting computing power to human insights. *International Journal of Research in Marketing*, *37*(3), 481-504.

Mahmoud, T. O. (2018). Impact of green marketing mix on purchase intention. International *Journal of Advanced and applied sciences, 5*(2), 127–135.

Mishra, M. (2020, August 26). *Convolutional Neural Networks, Explained.* Towards data science. https://towardsdatascience.com/convolutional-neural-networks-explained-9cc5188c4939

Molnar, C. (2020). *Interpretable machine learning.* Lulu.com.

Moslehpour, M., Chaiyapruk, P., Faez, S., & Wong, W. K. (2021). Generation Y's Sustainable Purchasing Intention of Green Personal Car Products. *Sustainability*, *13*(23), 13385.

Muntinga, D.G., Moorman, M., & Smit, E.G. (2011). Introducing COBRAs: Exploring motivations for brand-related social media use. *International Journal of advertising*, *30*(1), 13-46.

Paladino, A. (2006), "Understanding the green consumerism: an empirical analysis", *Journal of Customer Behaviour*, *4*(1), 69-102.

Pansari, A., & Kumar, V. (2017). Customer engagement: The construct, antecedents, and consequences. *Journal of the Academy of Marketing Science*, *45*, 294–311.

Philp, M., Jacobson, J., & Pancer, E. (2022). Predicting social media engagement with computer vision: An examination of food marketing on Instagram. *Journal of Business Research, 149*, 736-747.

Pintelas, E., Livieris, I.E., & Pintelas, P. (2020). A grey-box ensemble model exploiting black-box accuracy and white-box intrinsic interpretability. *Algorithms*, *13*(1), 17.

Poor, M., Duhachek, A., & Krishnan, H.S. (2013). How images of other consumers influence subsequent taste perceptions. *Journal of Marketing, 77*(6), 124-139.

Prothero, A., & McDonagh, P. (1992). Producing environmentally acceptable cosmetics? The impact of environmentalism on the United Kingdom cosmetics and toiletries industry. *Journal of Marketing Management*, *8*(2), 147-166.

Pudaruth, S., Juwaheer, T. D., & Seewoo, Y. D. (2015). Gender-based differences in understanding the purchasing patterns of eco-friendly cosmetics and beauty care products in Mauritius: a study of female customers. *Social responsibility journal*, *11*(1), 179-198.

Rietveld, R., Van Dolen, W., Mazloom, M., & Worring, M. (2020). What you feel, is what you like influence of message appeals on customer engagement on Instagram. *Journal of Interactive Marketing, 49*(1), 20-53.

Ruiz, A. (2023). *51 Huge Environmentally Conscious Consumer Statistics*. The Roundup. https://theroundup.org/environmentally-conscious-consumer-statistics/

Sadeque, F., & Bethard, S. (2019). Predicting engagement in online social networks: Challenges and opportunities. *ArXiv:1907.05442.*

Sajjacholapunt, P., & Ball, L.J. (2014). The influence of banner advertisements on attention and memory: Human faces with adverted gaze can enhance advertising effectiveness. *Frontiers in psychology, 5,* 166.

Samaraweera, M., Sims, J. D., & Homsey, D. M. (2021). Will a green color and nature images make consumers pay more for a green product? *Journal of Consumer Marketing*, *38*(3), 305-312.

Seo, J.Y., & Scammon, D.L. (2017). Do green packages lead to misperceptions? The influence of package colors on consumers' perceptions of brands with environmental claims. *Marketing Letters, 28*, 357-369.

Schaefer, A. (2005). Some considerations regarding the ecological sustainability of marketing systems. *Electronic Journal of Radical Organisation Theory*, *9*(1), 40-51.

Shawky, S., Kubacki, K., Dietrich, T., & Weaven, S. (2020). A dynamic framework for managing customer engagement on social media. *Journal of Business Research*, *121*, 567-577.

Shen, B., & Bissell, K. (2013). Social media, social me: A content analysis of beauty companies' use of Facebook in marketing and branding. *Journal of Promotion Management*, *19*(5), 629-651.

Simonyan, K., Vedaldi, A., & Zisserman, A. (2013). Deep inside convolutional networks: Visualising image classification models and saliency maps. *arXiv preprint arXiv:1312.6034*.

Song, Y., Qin, Z., & Qin, Z. (2020). Green marketing to gen Z consumers in China: Examining the mediating factors of an eco-label-informed purchase. *Sage open, 10*(4), 2158244020963573.

Statista (2023) Instagram – Statistics & Facts. Consulted on March 23, 2023 via https://www.statista.com/topics/1882/instagram/

Sukato, N., & Elsey, B. (2009). A model of male consumer behaviour in buying skin care products in Thailand. *ABAC journal*, *29*(1).

Takahashi, R. (2021). How to stimulate environmentally friendly consumption: Evidence from a nationwide social experiment in Japan to promote eco-friendly coffee. *Ecological Economics, 186,* 107082.

Tellis, G. J., MacInnis, D. J., Tirunillai, S., & Zhang, Y. (2019). What drives virality (sharing) of online digital content? The critical role of information, emotion, and brand prominence. *Journal of Marketing*, *83*(4), 1-20.

Testa, F., Iraldo, F., Vaccari, A., & Ferrari, E. (2013). Why Eco-labels can be Effective Marketing Tools: Evidence from a Study on Italian Consumers. *Business Strategy and the Environment, 24*(4), 252-265.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological), 58*(1), 267-288.

Todd, A. M. (2004). The aesthetic turn in green marketing: Environmental consumer ethics of natural personal care products. *Ethics and the Environment*, 86-102.

Tripathi, A., & Pandey, N. (2018). Does impact of price endings differ for the non-green and green products? Role of product categories and price levels. *Journal of Consumer Marketing, 35*(2), 143-156.

Van Doorn, J., Lemon, K. N., Mittal, V., Nass, S., Pick, D., Pirner, P., & Verhoef, P. C. (2010). Customer Engagement Behavior: Theoretical Foundations and Research Directions. *Journal of Service Research*, *13*(3), 253-266.

Vivek, S. D., Beatty, S. E., & Morgan, R. M. (2012). Customer engagement: Exploring customer relationships beyond purchase. *Journal of Marketing Theory and Practice, 20*(2), 127–145.

Wood, C.A., Ray, S., & Messinger, P. (2013). Leaving the tier: an examination of asymmetry in pricing patterns in online high tech shops. In J. Järveläinen, H. Li, A. Tuikka, & T. Kuusela (Ed*). Co-created Effective, Agile, and Trusted eServices* (pp. 63-73). Springer Berlin Heidelberg.

Xiao, L, & Ding, M. (2014). Just the faces: Exploring the effects of facial features in print advertising. *Marketing Science, 33*(3), 338-352.

Xue, F., & Muralidharan, S. (2015). A green picture is worth a thousand words?: Effects of visual and textual environmental appeals in advertising and the moderating role of product involvement. *Journal of Promotion Management, 21*(1), 82-106.

Yan, R., Hyllegard, K.H., & Blaesi, L.F. (2012). Marketing eco-fashion: The influence of brand name and message explicitness. *Journal of Marketing Communications, 18*(2), 151-168.

Yu, C. E., Xie, S. Y., & Wen, J. (2020). Coloring the destination: The role of color psychology on Instagram. *Tourism Management*, *80*, 104110.

Yusuf, R. (2020, December 14-15). *Comparing Different Supervised Machine Learning Accuracy on Analyzing COVID-19 Data using ANOVA test* [Paper presentation]. IEEE 6th Conference on Interactive Digital Media (ICIDM), Bandung, Indonesia.

Zhang, S., Lee, D., Singh, P. V., & Srinivasan, K. (2017). How Much Is an Image Worth? Airbnb Property Demand Analytics Leveraging A Scalable Image Classification Algorithm. *Management Science, 68*(8), 5644-5666.

Zhao, L., Zhang, M., Ming, Y., Niu, T., & Wang, Y. (2023). The effect of image richness on customer engagement: Evidence from Sina Weibo. *Journal of Business Research*, *154*, 113307.

Appendix A Example of photo slide.

The idea behind a photo slide is that multiple photos are featured in one post but can only be viewed separately. This feature was launched by Instagram in 2017. An example can be found below for Biossance on 15 April 2023. One single post includes 3 different photos that have number of likes (and comments) combined into one measure.

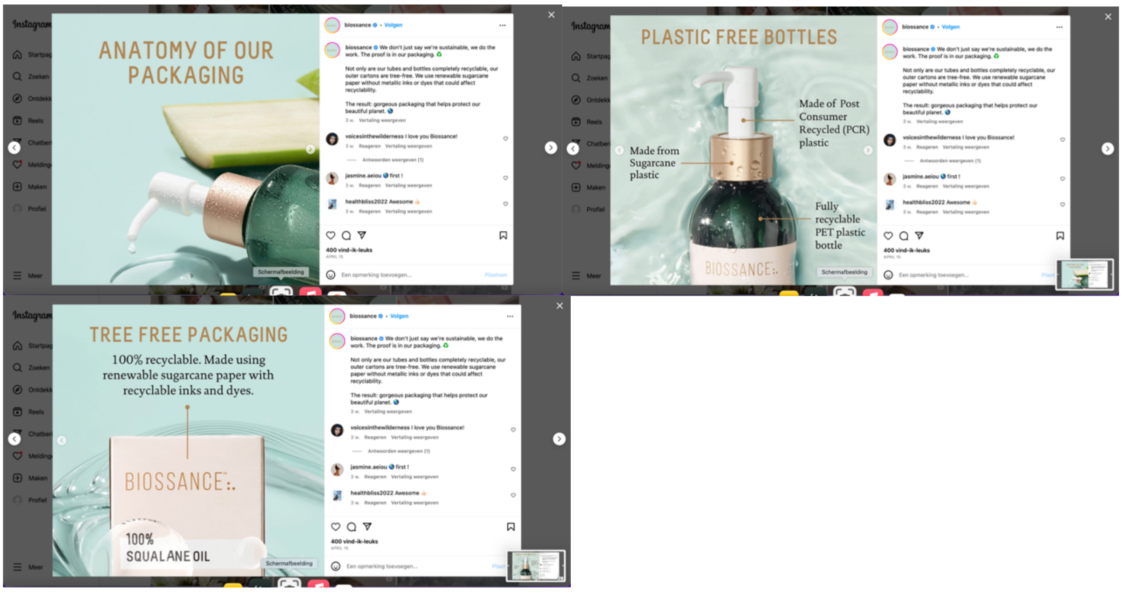


Figure A1. Image example of photo slide.

Notes: An example of a photo slide on Instagram. This is a type of post which allows for multiple photos in one post. The content of these type of posts is narrowed down to one picture (the first slide).

Appendix B Linear assumptions checked.

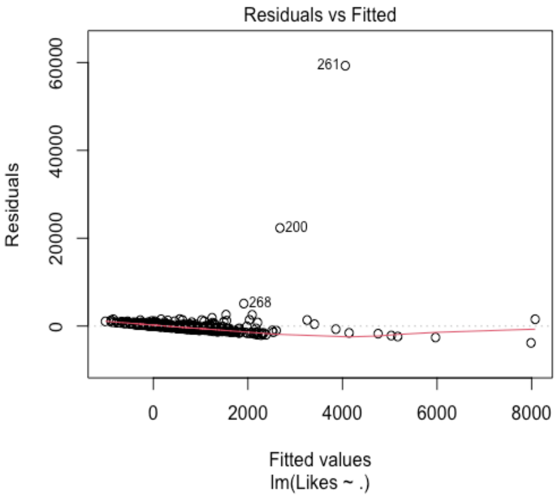


Figure B1. Residuals vs fitted values.

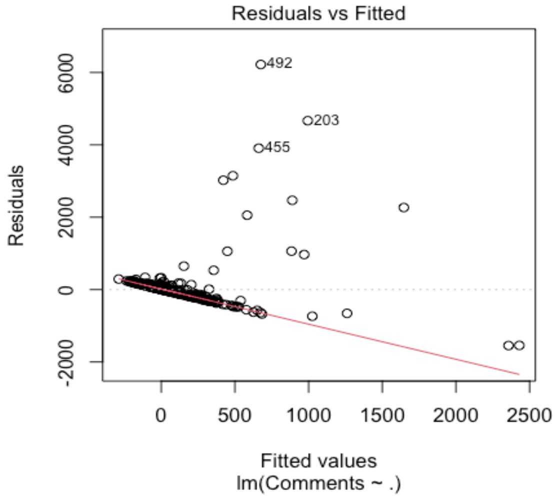


Figure B2. Residuals vs fitted values.

Notes: To check the assumptions of the linear regression, the residuals are plotted against the fitted values. The x-axis presents the fitted values, the y-axis the residuals. The plots are used to linearity.

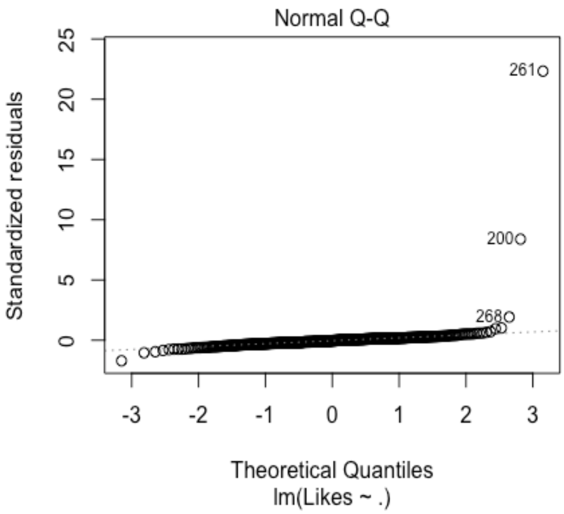


Figure B3. Distribution of residuals.

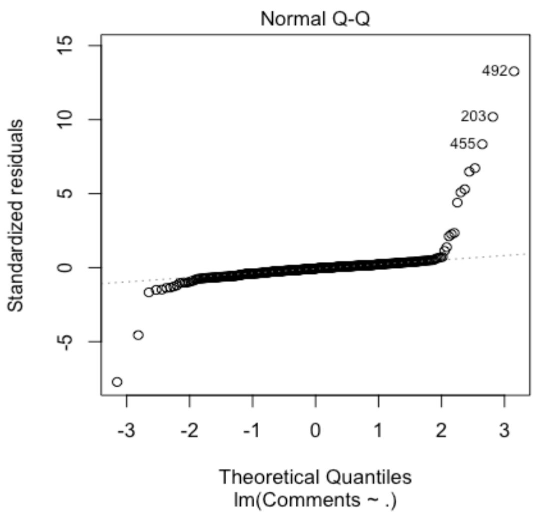
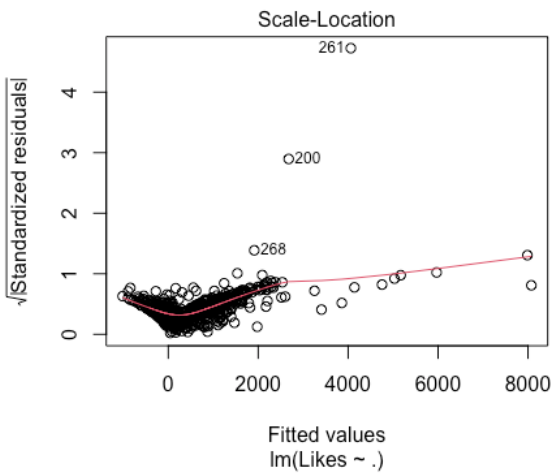


Figure B4. Distribution of residuals.

Notes: To check the assumption of normality, a Q-Q plot is constructed.



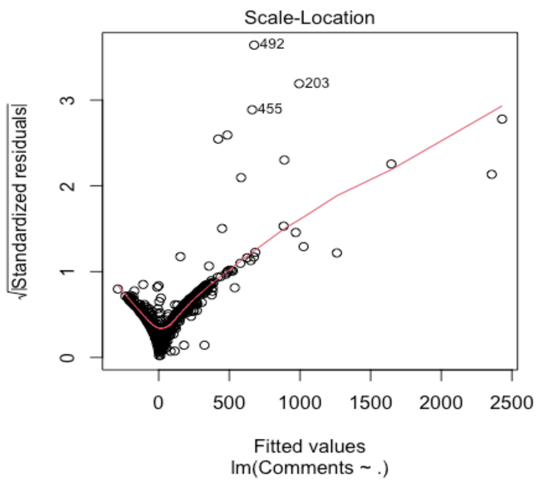
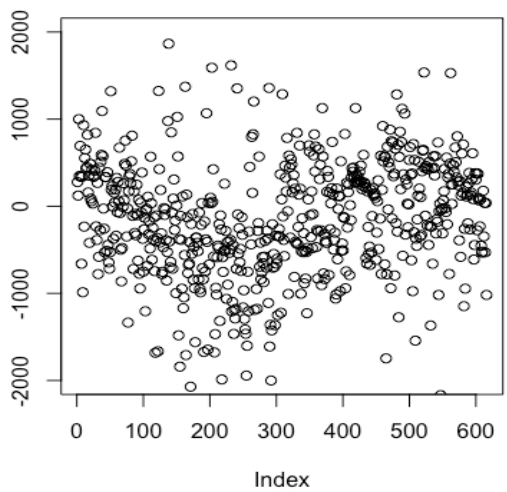


Figure B5. Homogeneity of variance of residuals.

Figure B6. Homogeneity of variance of residuals.

Notes: To check the assumptions of the linear regression, the variance in the residuals is examined. If the variance is constant, homoskedasticity is assumed.



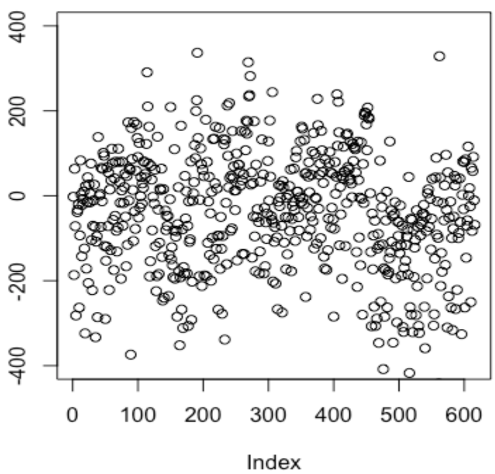


Figure B7. Residuals.

Figure B8. Residuals.

Notes: The residuals are plotted to test is the errors have mean zero, conditional on the covariates.

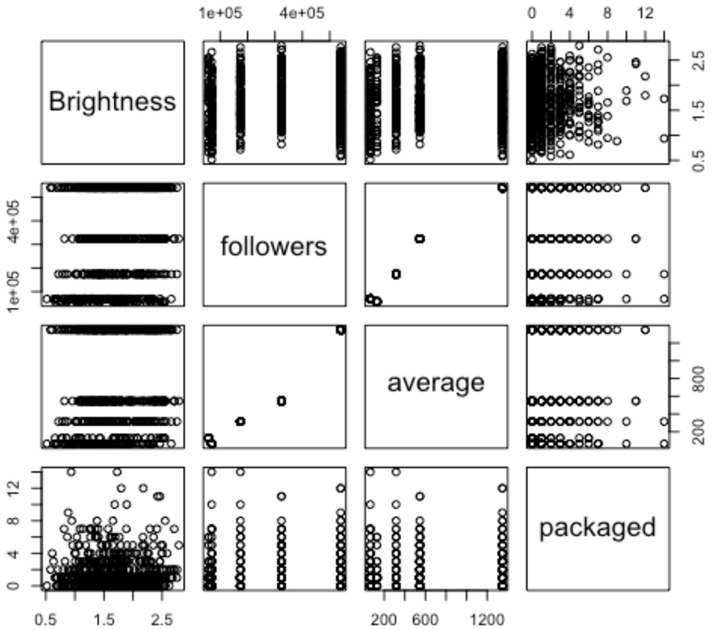


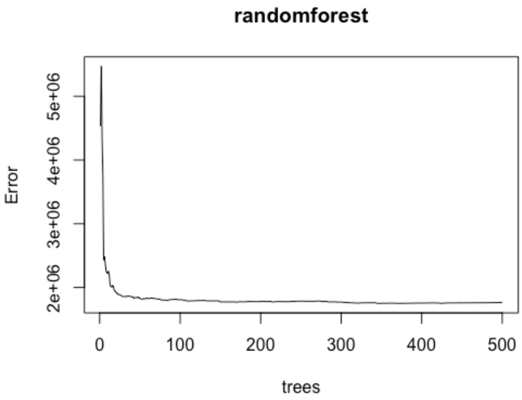
Figure B9. Correlation between covariates.
Notes: Four continuous variables are plotted against each other to check for correlation. The four variables are brightness, average likes, followers, and packaged goods.

Table B1. VIF-scores.
Notes: The variance influence factor (VIF) score for all variables in the linear model are presented. The first two columns present the results for predicting likes. The last two columns present the results for predicting comments. The VIF-score indicates the increase in the variance of a regression coefficient as a result of collinearity. One or more high VIF-scores (> 10) indicate that there is multicollinearity in the data.

Linear regression			
Likes	VIF	**Comments**	VIF
Brightness	1.1	Brightness	1.1
Followers	34.5	Followers	34.2
Average likes	34.7	Average likes	34.5
Hashtags	1.5	Hashtags	1.5
Tags	1.2	Tags	1.1
Type of post	1.4	Type of post	1.4
Post count	1.2	Post count	1.2
Brandname	1.5	Brandname	1.5
Type of content	2.0	Type of content	2.0
Face	2.6	Face	2.6
Facial expression joy	2.3	Facial expression joy	2.3
Packaged goods	1.7	Packaged goods	1.6
Products	1.9	Products	1.9

Hand	1.1	Hand	1.1
Text	1.4	Text	1.4
Personal care	1.4	Personal care	1.4
Days online	1.1	Days online	1.1

Appendix C Machine learning methods.



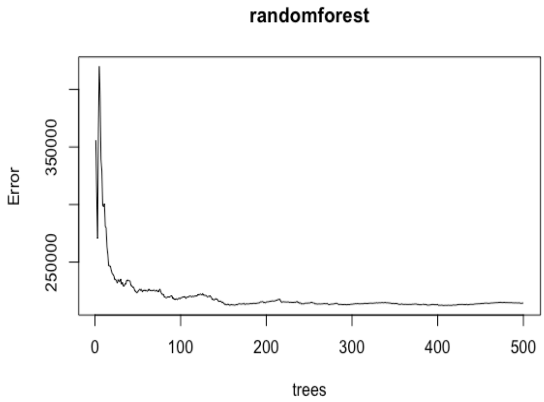
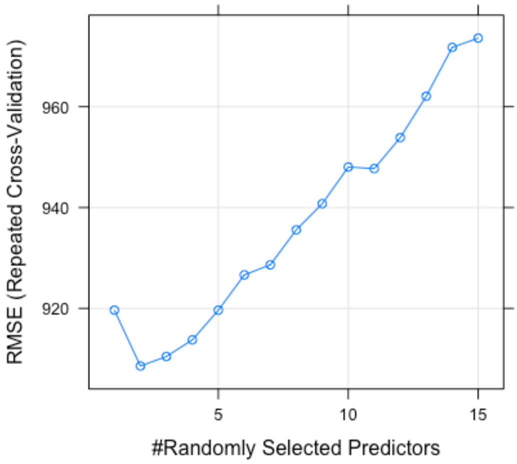


Figure C1. Ntree likes. Figure C2. Ntree comments.

Notes: The number of trees plotted against the error term. A lower error is preferred.



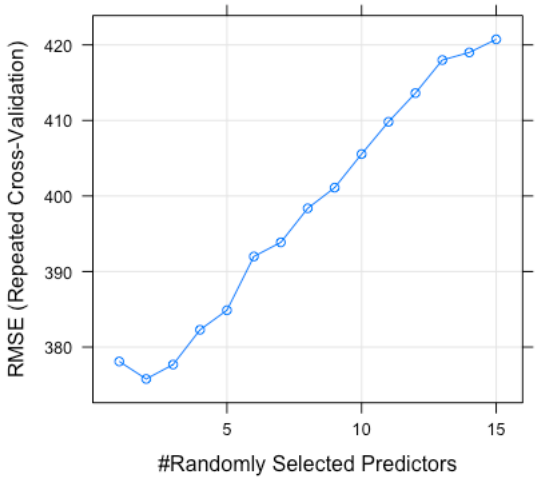


Figure C3. Mtry tuning. Figure C4. Mtry tuning.

Notes: A grid search is used to find the optimal features considered at each split in the random forest. This parameter is also referred to as Mtry. The optimal Mtry is the number with the lowest RMSE.

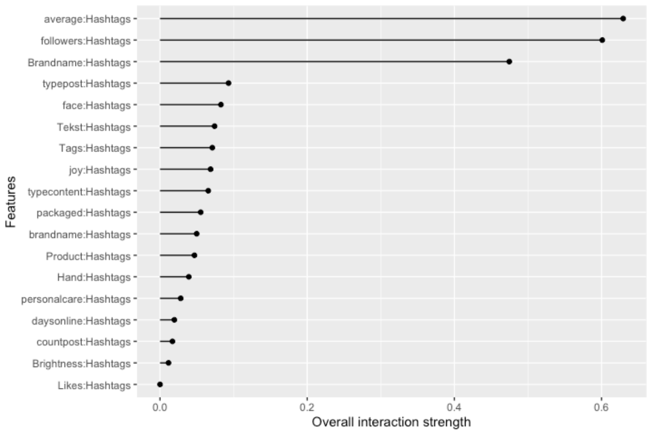


Figure C5. Hashtag interaction (likes).
Notes: The interaction between the number of hashtags used in the caption and all other predictor variables. These interactions hold for the prediction of likes.

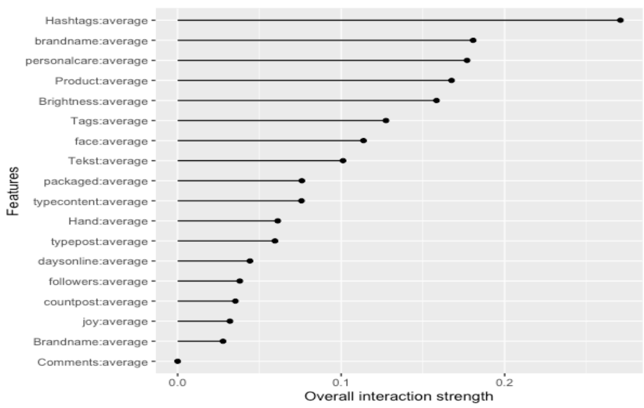
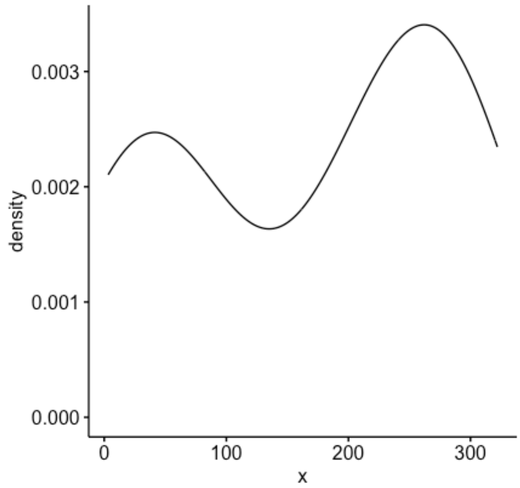


Figure C6. average likes interaction (Comments).
Notes: The interaction between the average number of likes and all other predictor variables. These interactions hold for the prediction of comments.

Appendix D Assumptions ANOVA checked.



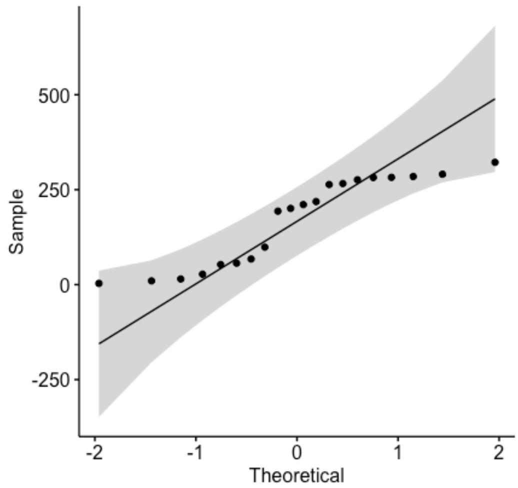
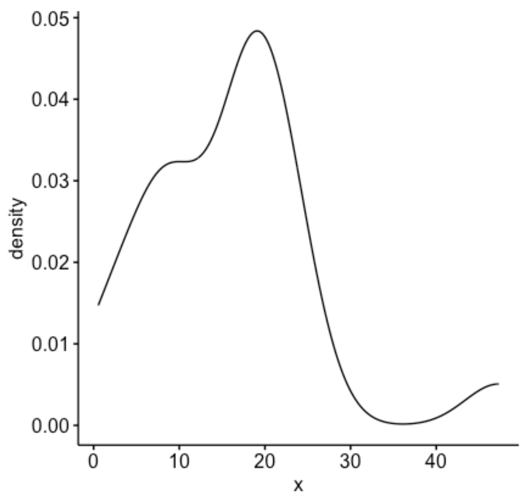


Figure D1. Density plot for likes.

Figure D2. Q-Q plot for likes.

Notes: To check the assumptions of ANOVA, normality of the model performance (RMSE) is examined. The density plot and Q-Q plot are constructed for the models predicting likes.



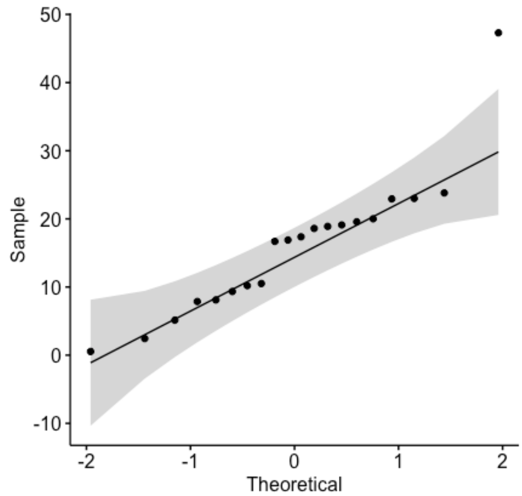


Figure D3. Density plot for comments.

Figure D4. Q-Q plot for likes.

Notes: To check the assumptions of ANOVA, normality of the model performance (RMSE) is examined. The density plot and Q-Q plot are constructed for the model's predicting comments.

Appendix E Distribution of color brightness.

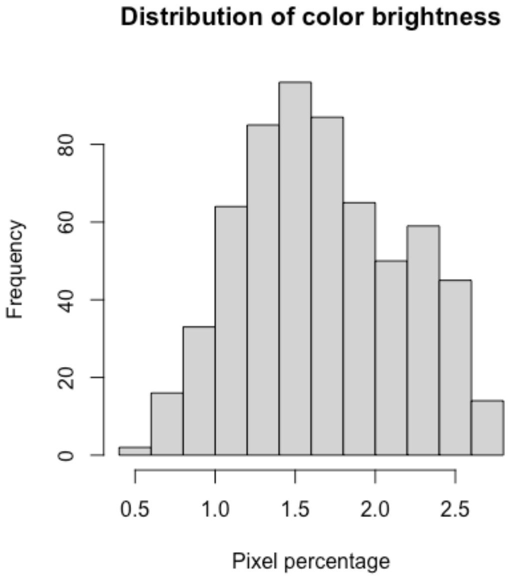


Figure E1. Distribution of color brightness.
Notes: A closer look into the distribution of color brightness. There is little data on images with a pixel percentage lower than 1.0 and higher than 2.5. These images are excluded when constructing partial dependence plots, to avoid bias view.

Appendix F Real vs predicted values.

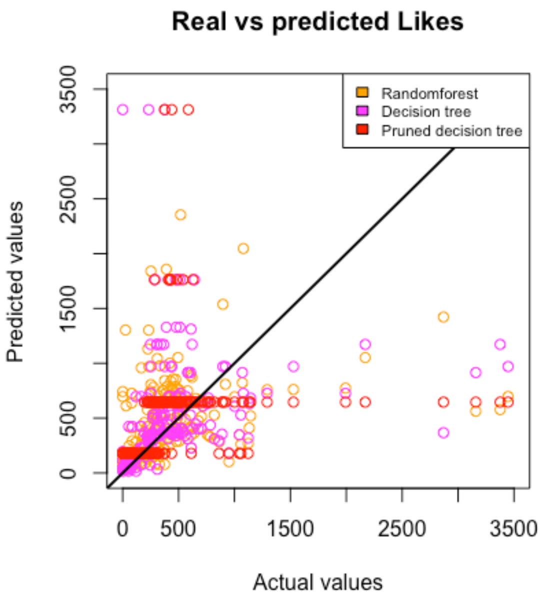


Figure F1. Predictions of likes.
Notes: A zoomed out view of the real versus predicted values of likes. The x-axis represents the actual values, the y-axis the predicted values. The yellow dots belong to the random forest, the pink to the decision tree and the red to the pruned decision tree. This zoomed out view shows that the pruned decision tree predicts only four values.

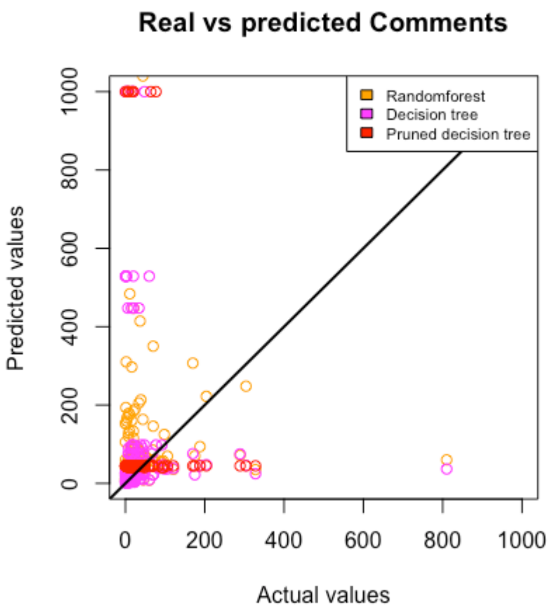
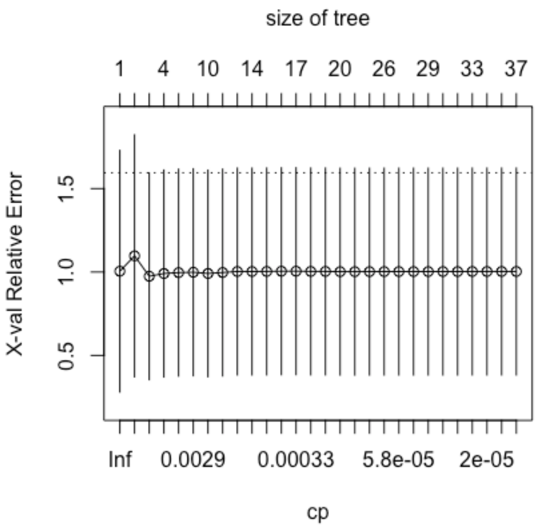


Figure F2. Predictions of comments.
Notes: A zoomed out view of the real versus predicted values of comments. The x-axis represents the actual values, the y-axis the predicted values. The yellow dots belong to the random forest, the pink to the decision tree and the red to the pruned decision tree. This zoomed out view shows that the pruned decision tree predicts only four values.

Appendix G Pruning decision tree.



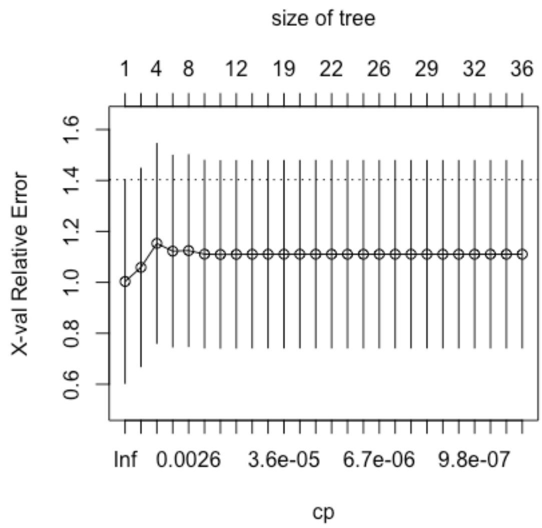


Figure G1. Pruning decision tree likes. Figure G2. Pruning decision tree comments.

Notes: The number of splits with the corresponding cost complexity parameter is plotted against the cross-validated error. The number of splits with the lowest error, is the optimal depth of the tree.