Erasmus
School of
Economics

Erasmus University Rotterdam
Erasmus School of Economics

Master Thesis Programme Data Science and Marketing Analytics

Do profile pictures align with the produced user-generated content?

Name student: Gina Quintarelli

Student Number: 500326

Supervisor: Andreas C Bayerl

Second Assessor: XXX

Date final version: XXX

Erasmus University Rotterdam

Table of Contents

Abstract

This research tries to get a better understanding of the relationship between user-generated content and profile pictures within digital sentiment analysis. By using advanced machine learning techniques, this research focuses on two research questions, namely (1) the alignment of profile photos with expressed sentiments in user-generated content and (2) the prediction of user-generated content valence based on profile picture characteristics. This research uses, among others, Random Forest and Extreme Gradient Boost models to analyze the dominant facial emotion scores and demographic factors influencing user sentiment using a dataset with different demographic profiles and dominant emotions. The findings show that there are significant associations between specific emotional states (for example, happiness and sadness) captured in profile photos and corresponding sentiment alignments in textual content. In addition, the XGBoost model reveals a high accuracy in predicting the emotional valence in digital comments, which highlights how effective profile picture features are. To conclude, this research identifies limitations and suggests opportunities for future research to improve generalisability and interpretability of research like this. Altogether, this research gives a better understanding of different digital sentiment analyses and offers practical implications for companies to optimize user engagement strategies through customized content experiences tailored to users' emotional states.

1. Introduction

The relationship between users' profile pictures and the emotions they express in reviews can provide new perspectives into consumer behavior in the constantly growing world of digital activity. This research focuses on the correspondence between visual cues in profile pictures and the emotions expressed in user-generated content. With almost 3000 online review platforms and countless reviews daily in the USA (BuiltWith, 2022), the alignment between these two is essential (Zhou, 2024). When we make Computer Vision techniques work together with sentiment analysis, a connection between the facial expressions of users in their profile pictures and the sentiment in their online reviews can be discovered (Patel et al., 2020), and with this, improving the understanding of non-verbal signs in the digital communication (Pantic & Patras, 2006).

This research goes beyond the achievement drive hypothesis, as defined by He et al. (2019), to improve the underlying concept. The achievement drive hypothesis, which is originally rooted in the financial domain, states that there is a connection between facial features, particularly the face-width-to-height ratio (fWHR), and performance drive. These users' facial features, which suggest their online self-expression, are a central line in this research, where the question is whether these features influence the sentiment in their online expressions.

This research focuses on uncovering overlooked dimensions by examining facial features captured in profile pictures (Pantic & Patras 2006). In this world of digital exchanges, non-verbal cues play a central role in communication. Facial features, which are embedded in profile pictures, offer a lot of non-verbal information that can be crucial for understanding user satisfaction, possibly exceeding insights that can be obtained only from text alone. These facial features can be thought of as expressions of satisfaction and can bring emotions and subtleties that may be beyond the reach of textual representation (Calvo & Nummenmaa, 2016). However, it is important to remember that a user might have a happy or satisfied profile picture while leaving a one-star review. In this research, the profile picture is linked to the user's profile, not the specific review (Kim et al., 2018). This results in that the facial features in profile pictures may not only be related to the particular experience expressed in the review but also, for instance, to how the users want to portray themselves in this online context (Yadav, 2021). This research causes a reconsideration of the link that is assumed between profile images and the specific consumption experience and highlights the broader effect of these facial features on how users shape their online identities (Vilnai-Yavetz & Tifferet, 2015).

The refreshing element of this research is the attempt to close the gap between visual cues taken from profile pictures and conventional sentiment analysis methods (Mohammad, 2016). This study tries to dissect the complex relationships between facial expressions, emotions, and sentiments expressed in online reviews by including Computer Vision techniques in sentiment analysis methods (Zeng et al., 2007; Littlewort et al., 2002). Next to this, by looking further than the achievement drive theory (He et al., 2019), this research tries to shed light on the interaction between motivational and emotional factors of this digital world (D'Mello & Graesser, 2012).

Companies are currently trying to understand the different dimensions of customer feedback, which is dominated by user-generated content (Lerche, 2016). Conventional sentiment analysis methods often focused on textual interpretation are challenged by revealing emotions and subtleties hidden in the non-verbal domain (Pantic & Patras, 2006). Next to that, profile pictures are not linked to particular experiences but to the users' profiles; one can also have multiple profile pictures on his or her account throughout a period in which he or she wrote various reviews.

This research addresses the alignment between users' selected profile pictures and the feelings expressed in their user-generated content. The first question this research will focus on is: "Do profile pictures align with the sentiments expressed in the produced user-generated content?" This research question leads to an analysis of the alignment between the visual representations of the users themselves and their emotional expressions in their digital communication. Using Computer Vision, this research goes beyond the normal sentiment analysis and tries to enlighten the alignment between emotions expressed in user-generated content and the profile pictures (Patel et al., 2020). Furthermore, scientifically, it focuses on the growing need for a sound understanding of the digital world, specifically by looking at the visual cues buried in a person's facial expressions on a profile picture (Vilnai-Yavetz & Tifferet, 2015). Concluding, this research focuses on the study of the alignment of feelings and expressions and their linked user-generated content.

Additionally, this research addresses a second research question: "Can the valence of user-generated content be predicted from users' profile images?" This question broadens this research with predictive opportunities concerning profile pictures and the emotional tone of user-generated content. By analyzing whether facial features extracted from profile pictures by Computer Vision match the user-generated content, this research also aims to give insights into the visual representation of the users with their profile pictures and the emotions they present with their digital interactions.

Looking at this research from a practical point of view, the outcomes can be used in multiple ways. For example, in sectors like retail, the refinement of customer satisfaction, like facial expressions, can be used for different engagement strategies (Hartmann et al., 2021). However, because the user profile picture is not explicitly linked to the review, this should be interpreted carefully. For instance, a restaurant manager could use the insights from online reviews to improve the service and experience in the restaurant. Instead of only focusing on the specific consumption experience, taking the broader online identity into account as well can drive decisions on engaging with users to get reviews. This line of thought fits into the idea that profile pictures express how users want to present themselves online, which stresses the need for a nuanced interpretation of user-generated content within online reviews (Strano, 2008).

This paper researches the alignment between the chosen profile pictures of users and the feelings and emotions expressed in the user-generated content within the digital world of consumer behavior. This research mainly focuses on the correlation between the emotional expression in profile pictures of the users, and their emotional expression in online reviews. This is addressed in two research questions:

1. "Do profile pictures align with the sentiments expressed in the produced user-generated content?"
2. "Can the valence of user-generated content be predicted from users' profile images?"

In the following sections of this paper, relevant literature will be discussed. Data, research methods, and models on the alignment between profile pictures, user-generated content, and expressed emotions will also be analyzed and developed. We start by giving context to this research based on existing studies. After this literature review, the research questions will be substantiated, and the data used in this study will be explained and trained. The methods used to answer the research questions will be discussed in the next part, and the results will be given. Finally, the findings will be given likewise any implications, limitations, and further research ideas.

2. Literature Review

Exploring Computer Vision, facial expressions, and sentiment analysis together can help to understand human behavior and communication in a digital context. Advanced algorithms in these areas can give insights into facial features and expressions and help us better understand human interactions and emotions (Taigman et al., 2014; Krizhevsky et al., 2012). By combining computational methods and psychology and neurosciences, complicated relationships between facial features, emotions, and cognition can be revealed (He et al., 2019; Chen & Wyer, 2020; Jaeger et al., 2019; Frith, 2009; Ling et al., 2020).

Broad research on Computer Vision, facial features, and sentiment analysis within the digital context is essential for understanding the complexity of human behavior in online communication (Li et al., 2022). This multi-disciplinary research can be important for uncovering processes that give structure to online interactions, emotions, and sentimental expressions (McDuff, El Kaliouby, & Picard, 2012). By using Computer Vision, facial features, and sentiment mining, this research tries to reveal the influence of visual and textual signals on user-generated content (Kosti et al., 2017).

2.1 Computer Vision in Understanding Facial Features

Computer Vision Techniques have significantly improved the knowledge of facial features by applying complex algorithms that can analyze visual content with amazing accuracy and precision (Taigman et al., 2014). The breakthrough with Computer Vision was when Taigman et al. achieved human-level performance with face verification, which was a very important step in this field. After this, Wang et al. (2021) further extended this breakthrough by researching image representations in combination with equivariance, with which the understanding of the analytical representation of facial features has improved. These two studies highlight the importance of reliable image representations in understanding facial features.

Krizhevsky, Sutskever, and Hinton (2012) also made important contributions to image representations with deep convolutional neural networks (CNN's), with which they layed the grounds for further research in the field of face recognition and face analysis. Also, He et al. (2019) looked at the relationship between behavioral traits and facial features. They focused on the face-width-to-height ratio (fWHR) and the link of this ratio and performance drive. This study uncovered possible links between cognitive functions and face morphology, which further expanded the understanding of the effect of facial features on mental traits.

In other recent research, the application of generative adversarial networks (GANs) in the generation of realistic facial images with a low-resolution is examined (Bernardi, 2023). In the research of He et al. (2020) a new GAN-structure was developed with which facial images from high quality can be generated from pixelated inputs. This research, together with the challenges that are linked to low-resolution facial images in facial recognition systems. With that, this progression in GAN technology seems promising for improving the accuracy and reliability of facial recognition algorithms, and with that further developing the field of computer vision.

Attention mechanisms are emerging and integrated into neural network methods and seem promising for improving the interpretability and performance of facial recognition systems. One existing paper, namely by Ling et al. (2020), demonstrates the effects of attention-based CNNs on localizing facial features accurately and documenting very precise details, with which the robustness of facial recognition algorithms in the real scenarios are improved. Innovative approaches like this one can significantly impact the development of accurate and reliable facial recognition systems in this field.

2.2 User-generated Content, Facial Expressions, and Personality Revelation

There has been research in the field of the influence of facial expressions on online observations and interactions. For example, Hiesh and Tseng (2017) research the influence of emoticons, which are described as a digital representation of facial expressions, on the involvement of users and the sentiment in their online reviews. The findings in this research show that reviews with emoticons are considered more engaging and more expressing the emotions, which highlights the importance of giving a shape to facial expressions in user-generated content.

Facial recognition technology created a way for automated analysis of facial expressions in user-generated content. Among others, Li et al. (2020) developed a model that recognizes emotions on facial expressions. This model uses machine learning techniques to identify and classify emotional states based on profile images. By integrating sentiment mining-techniques with facial expression techniques this framework provides us with approaches to understand the different emotional dimensions of user-generated content.

The arrival of virtual reality (VR) and augmented reality (AR) platforms has created new opportunities for studying facial expressions in digital environments. Among others, Wang and Lin (2021) explored the use of VR simulations to trigger facial expressions and emotional reactions of users, which gave insights into the different aspects of online interactions. By capturing facial expressions in virtual environments, researchers can understand emotions and behaviors in stimulated contexts better, which complements existing approaches to user-generated content analysis.

2.3 Opinion Mining and Sentiment Analysis

Opinion mining and sentiment analysis are important when one wants to extract useful insights from user-generated content. This is because opinion mining and sentiment analysis can uncover preferences, emotions, and attitudes from user-generated content. Recent research by Calvo and Nummenmaa (2016) has expanded the field of sentiment analysis by introducing multimodal methods that use both textual and visual cues. These methods show exceptional outputs when capturing sentiments expressed by users, with which the accuracy of sentiment analysis improves and a broader understanding of the attitudes and emotions of users is obtained.

When sentiment analysis is combined with computing techniques, this leads to new possibilities in understanding expressions in user-generated content and emotional motivation. Models like He at al. (2016) invented are emotion-aware sentiment analysis models, and use facial expressions to derive emotional context from textual input, with which higher detailed sentiment classifications are achieved. This advancement highlights the growing acknowledgment of facial expressions as a valuable source of emotional information in sentiment analysis, with which a better understanding is obtained of the complex relationship between visual and textual cues used in the methods for user-generated content (Poria et al.2017; Cambria et al., 2013).

This extensive literature review combines theoretical works, empirical insights and methodological considerations and offers a detailed overview of the complex relationship between facial expressions, emotions, and sentiments in online reviews for research (Li et al., 2019). Integrating theories, empirical research, and methodological considerations clarifies the relationships between these concepts and helps guide empirical research (McDuff et al., 2012; Kosti et al., 2017).

Facial expression theories by psychologists, namely ones from Paul Ekman, state that the universal of facial expressions are indicators of emotional states. One of the biggest concerns in this paper is still the computer vision techniques, image representations, and deep neural networks to analyze facial features. This is why the main focus of this research is the alignment between profile pictures and the emotions expressed by the users in their user-generated content and the question of whether valence expressed in the user-generated content can be predicted based on the users' profile pictures. The focus is on the existing work of Taigman et al. (2014), Wang et al. (2021) and Wang and Kosinski (2019) and their inputs to the advancement of computer vision techniques, image representations and deep neural networks which provide facial feature analysis in this research. The insights from this study give a more in-depth understanding of the relationship between emotional expressions and visual representations in user-generated content.

This research tries to explore the relationships between motivational factors, emotional states, and sentiment expressions in online reviews. Even though the fundamental aspect of He et al. (2019) achievement-drive hypothesis might not be the main focus of this research, elements of the motivational factors are still relevant for understanding the incentives and motives of users' satisfaction and emotional expression in the digital environments. So, while not fundamental, this achievement-drive theory can offer valuable insights into the broader essence of user sentiment expression.

2.4 Facial Features and Behavioral Dynamics

Facial features are found to significantly influence behaviour and perceptions of individuals in many domains. These domains include politics and psychology (Harmon-Jones, 2019; Wänke et al., 2012). In a political context, the size of the faces of candidates on campaign posters in relation to other elements on the poster can influence the perception and behaviour of voters (Brown & Green, 2018). In addition, psychological studies reveal that there is a correlation between specific facial morphologies, such as sharp chin bones, and behavioural traits (Thayer & Dobson, 2010). Psychology argues that some facial features typically serve as cues for personality traits, whereby men with prominent chin bones are often viewed as more dominant and masculine (Dixson, 2021). Also, the association between facial morphology and behavioural perceptions reaches the social context where individuals with more prominent chin bones tend to be more likely to display dominant and assertive behaviour, and this may influence their interactions and decision-making processes (Wänke et al., 2012).

The integration of political psychology and psychological research on facial morphology offers valuable insights into the complex relationship between behavioral dynamics and facial features (Schmidt & Cohn, 2001). Scholars can understand political decision-making ways by researching how facial features can influence voters' perceptions (Mattes et al., 2010). Similarly, research into the psychological implications of facial morphology offers insights into the interactions between physical appearances and behavioral dynamics, which expands the understanding of human behavior in different contexts (Schmidt & Cohn, 2001).

3. Data

The foundation of this research is understanding the complicated relationship between facial expressions, emotional dynamics, and sentimental expressions in the user-generated content. Based on a multidisciplinary approach, techniques like Computer Vision and sentiment analysis will be used to uncover the relationships within the dataset (Poria et al., 2017).

3.1 Research Sampling

This research is built on by using three primary datasets, the first one being a Yelp Reviews Dataset and the second one being a Curated Dataset consisting of the user information data and the profile pictures, to investigate the complex relationship between facial features and user satisfaction in reviews (Lu et al., 2018).

After getting the Yelp Reviews Dataset, which started with 6,685,900 entries of sentiments and opinions from reviews, the next step was to merge the datasets into one usable dataset in order to answer the questions in this research. This involved some basic steps in R. To start, the review data, the user information data, and the profile pictures are imported using various packages like the "readr" package and the "dyplr" package (Wickham et al., 2019). After this, the dataset was refined to a dataset with only profile pictures with only one face. This is done to ensure consistency and accuracy in analyzing facial features and their relationship with sentiment, as multiple faces in a single image could introduce variability and confound the analysis (Camastra & Vinciarelli, 2015). After this, the dataset only contains user profiles with profile pictures with only one face, which is an important requirement for the following analyses. The next step is to choose a representative profile picture per user, as some users have more than one profile picture. This is reached by grouping the users' data and identifying the profile picture with the highest prediction confidence. All of these steps together create the dataset used in this research, in which information about facial features is seamlessly integrated with the users' sentiments expressed data in the Yelp reviews. With this final dataset, the analysis into the relationship between facial expressions in the profile picturesof the users, the user satisfaction, and the sentiment in reviews can be investigated.

3.2 Data Aggregation Process and Operationalization of Variables

The dataset needs some changes in the first stages of data cleaning. The reviews of the users and their profile picture features are aggregated. This is done to investigate the relationship between the profile pictures and the user-generated content (He et al., 2019).

The data aggregation process starts with an examination of the profile pictures associated with the user IDs in the dataset. Profile pictures with only one face were specifically chosen. This choice is not arbitrary, but more of a strategic move in order to provide the best quality data for an easy to interpret face analysis. Then, the next step in the aggregation process is to group the selected profile photos by user ID and the profile photos with the highest prediction reliability are chosen as the representative photos for each user. Furthermore, the review dataset is aggregated by user ID, with the most useful variables being the average star ratings and sentiment scores (Benlahbib., 2020). The aggregation of reviews by user provides an overview with valuable insights into users' sentiments and preferences. This approach is linked to previous research on sentiment analysis, from which we can see that aggregating data at the user level can provide useful insights while reducing noise and redundancy in the dataset (Calvo & Nummenmaa, 2016). Aggregating ratings by user ID improves the efficiency of the analyses, but retains the essential information that users are trying to convey with their ratings.

The comprehensive dataset is created by combining the profile pictures dataset and the reviews dataset based on the user IDs, with which the alignment between the profile pictures and the user-generated content can be extensively investigated (Li et al., 2018). This aggregation approach gives priority to data quality and user-orientated analysis, which allows for insights into the demographic characteristics of users, their emotional ability, and the expression of their sentiment in reviews (He et al., 2017). The average star rating of the users serves as an indication of their general sentiment tendencies. Qualitative aspects of user-generated content, under which we can scale reviews and temporal contexts, can offer a rich source (Batrinca & Treleaven, 2015). This richness in the data piques curiosity and provides a comprehensive understanding of the sentiment dynamics within the Yelp community.

Facial Analysis techniques can deduce facial features of users' profile pictures alongside demographic variables and dominant emotion scores. These variables uncover the relationship between facial features and sentiment expressions in user-generated content and provide insight into sentimental dynamics within the Yelp reviews world (Camastra & Vinciarelli, 2015).

In this research, for alignment between profile pictures and the sentiment of the user-generated content, deducting variables for the profile and review sentiment serves as a methodological foundation. Profile pictures are vital visual representations within online platforms that enclose facial features and personal identities. Therefore, in this research, the focus lies solely on profile pictures with one face to ensure data quality. This decision matches earlier papers in the world of Computer Vision and social media analysis that highlight the complexity of profile pictures with more than one face

(Benlahbib, 2020; He et al., 2016). Group pictures present challenges with facial recognition and interfere with the accurate analysis because of the multiple faces. On the other hand, single-faced pictures offer a more focused and interpretable representation, increasing data reliability and improving analysis consistency. By prioritizing profile pictures with only one face, clarity in interpretation is ensured, and distortions are limited, which ultimately strengthens the validity of the research findings.

Furthermore, aggregating profile pictures is important for gaining meaningful insights in the case that more than one picture is linked to users' accounts. To tackle this, an aggregation method based on confidence scores for predictions is used, which uses insights from existing research literature. Recent developments in terms of computer vision allow for predicting different characteristics based on profile pictures, and where quantitative measures are used for the models' certainty (Li et al., 2019; Patel & Shah, 2021). By selecting the profile picture with the highest confidence score for prediction, it makes sure the aggregated profile pictures accurately display the users' features, allowing for analyses to be improved and more accurate predictions. This approach lowers the risk of bias when selecting random and less informative pictures. Next to this, it helps to ensure that the merged profile picture includes the most important features of the users. Using confidence scores for prediction in profile picture aggregation provides a basic and effective method for selecting representative images in online contexts, which aligns with best practices in image analysis and social media research (Chen et al., 2020; Kim & Lee, 2017).

Table 1. Summary of Initial and Final Datasets

Metric	*Initial Dataset*	*Final Dataset*
Number of Users	2,138,168	544,330
Number of Features	21	54
Number of Missing Values	14,445,018	0

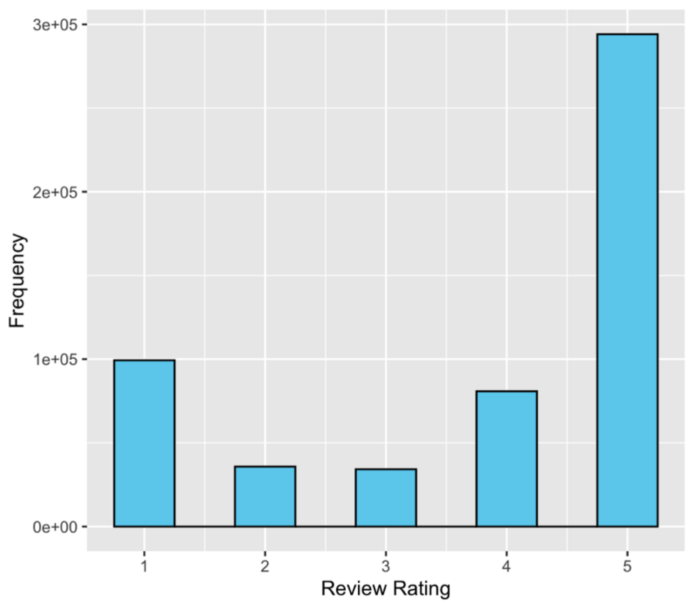
3.3 Exclusion Rationale

Profile pictures serve as visual representations of individuals on online platforms and they contain characteristics such as facial expressions and personal identity. Looking at the importance of data quality in this research for the alignment between profile pictures and user-generated content, the focus lies on profile pictures with only one face. This decision is supported by existing literature in the field of computer vision and social media analysis, in which the challenges that go hand in hand with profile pictures with multiple faces are highlights (Smith et al., 2018; He et al., 2016). As an example, pictures with multiple faces, such as group pictures can add complexity to facial recognition analysis and, with that, potentially interfere with getting accurate results due to the presence of multiple faces (DiMicco & Millen, 2007). According to Musil et al. (2017), profile pictures that with only one face in them give us a better interpretable perspective. Hence, the use of profile pictures with a single face helps to ensure clarity in interpretation and reduce potential biases, thereby strengthening the validity of the findings.

When preparing the dataset for the analyses, there were 2,138,168 users. After data-wrangling, the dataset consists of 544,330 unique users with profile pictures, which are essential for this research. Throughout the data-wrangling process, a couple of exclusion criteria are applied to ensure the quality and relevance of the dataset. A subset is eliminated for several reasons, this subset includes profile pictures with multiple faces, as explained above, the lack of identifiable faces in the pictures, and other anomalies including low image quality, significant obstructions or irrelevant content (Saber & Tekalp, 1996). The purpose of this process is to improve the consistency and reliability of the dataset by making sure that only profiles with clear and representative images are retained for further analysis.

The distribution of review ratings, as can be seen in Graph 1, shows each rating category ranging from 1 to 5 stars and their frequency. This distribution gives valuable insights into users' satisfaction levels and sentiment within the dataset (Hu et al., 2017). This analysis is focused on the adjusted dataset, which means the dataset with only the users that have one face in their profile picture.

Graph 1. Distribution of Review Ratings



3.4 Rationale for Aggregating Profile Pictures

Aggregation of the profile pictures is important for getting useful information from the visual representations of the users. When there is more than one profile picture linked to the account the picture with the highest prediction confiedence score will be chosen as the profile picture. This is based on existing literature (Li et al., 2019; Patel & Shah, 2021). Recent developments in the field of computer vision make it possible to predict various attributes based on profile pictures, which results in quantitative measures of the certainty of a model. By choosing the profile picture with the highest reliability score for prediction, it is made sure that the aggregated profile picture accurately displays the characteristics of the users, which makes that the analysis is enriched and more accurate predictions can be made. By using this method it reduces the risk of biases by selecting random or less informative pictures and ensures that the aggregated profile pictures display the essential characteristics of the users. The use of prediction confidence scores when aggregating profile pictures offers a principal and effective method when choosing the representative picture from the online environment, which aligns with best practices in image analysis and social media research (Chen et al., 2020; Pang & Lee, 2006).

The dataset contains an important variable, the dominant emotion, which reflects the main emotional expression identified in users' profile photos. This score, already provided by the provider of the data using algorithms to recognize facial expressions, is a central part of the analysis (Benitez-Quiroz et al., 2016). Even though the computations were not performed directly, the methodology behind it is important to understand. Algorithms for recognizing facial emotions and analyzing facial characteristics and expressions are used to detect these (Zeng et al., 2018). Based on the analyses, every user gets appointed a dominant emotion score, which is linked to the primary emotion in their profile picture. This existing variable plays an important role in this research because it allows us to explore the relationship between facial expressions, emotions, and user behavior and highlights the importance of emotional cues in user analysis (Danner et al., 2014).

In Table 2, a comprehensive overview of the user's traits, categorized into dominant emotions in the user's profile pictures, is shown. Every row corresponds to a dominant emotion, namely angry, disgust, fear, happy, neutral, sad, and surprise. It shows the number of users with this dominant emotion, their average age, the gender distribution within every category, and the percentage of men and women in this category.

Table 2. Summary of User Characteristics by Dominant Emotion

Dominant Emotion	*Number of Users*	*Average Age*	*Gender Distribution*	*Man Percentage*	*Woman Percentage*
Angry	28,836	33.09	Man: 23,955 Woman: 4,881	83.15%	16.85%
Disgust	1,075	32.84	Man: 805 Woman: 270	74.88%	25.12%
Fear	36,790	31.67	Man: 28,794 Woman: 7,996	78.28%	21.72%
Happy	294,339	33.11	Man: 167,678 Woman: 126,661	56.98%	43.02%
Neutral	118,439	31.89	Man: 84,755 Woman: 33,684	71.55%	28.45%
Sad	57,556	31.30	Man: 45,359 Woman: 12,197	78.79%	21.21%
Surprise	7,295	32.09	Man: 4,849 Woman: 2,446	66.47%	33.53%

While analyzing the data, some interesting fluctuations in the emotional expressions between men and women can be found. In general, women are more likely to express sadness and disgust, looking at their overall representation in the dataset. This means that women represent 25.12% of the users who express disgust and 21.21% of the users who express sadness, and with this, they surpass their general representation of 34.56%. However, looking at men, this shows a higher representation of emotions such as anger and fear, with respectively 83.15% and 78.28% of the users expressing these emotions (Hess et al., 2009). The differences highlight the relationship between emotional expressions and gender dynamics and suggest that emotional reactions can be influenced by sociocultural factors and individual differences (Mosquera et al., 2000).

4. Methods

The foundation of this research in understending the user-generated content can be found with facial expressions, emotions, and sentiment analysis. By using computer vision techniques, this research delves into facial features and draw insights from Taigman et al. (2014) and Krizhevsky et al. (2012). We explore the emotional expressions and personality traits, with which we refer to Pantic and Patras (2006) and Whitty et al. (2018). Based on the research of Pang and Lee (2008) and Liu (2020), sentiment analysis is expanded for machine learning models, with which the understanding of users' sentiment in online reviews is widened. By using different analysis methods, this research provides insights into online user-generated content.

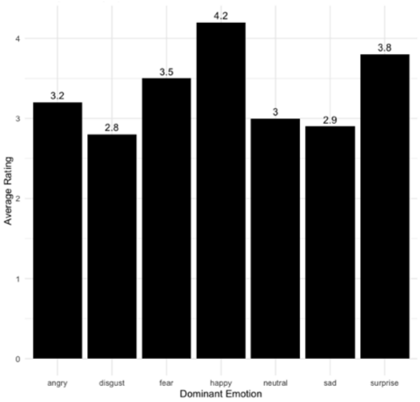
4.1 Model-Free Evidence

This section of the paper investigates the relationship between user characteristics and sentiment expression in user-generated content. A statistical analysis and a visualization are used to investigate how different user characteristics relate to average ratings, expressions, age, and gender.

4.1.1 Dominant Emotion Analysis

This part dives into the dominant emotion analysis in the profile pictures of the users and explores the relationship with the average ratings of these dominant emotions. As the graph below shows.

Graph 2. Average Rating by Dominant Emotion



In Graph 2 shows significant variations in the average rating scores per users' dominant emotions. Users who show happiness as their dominant emotion have the highest average rating (4.2), followed by surprise (3.8) and fear (3.5). Controversially, users with the dominant emotions disgust and sadness have the lowest average rating, respectively 2.8 and 2.9. Neutral emotions are associated with an average rating of 3.0, while the emotion anger has a slightly higher average rating which is 3.2.

The findings of this graph are in line with existing literature on the influence of emotional expressions and social perceptions. Positive emotions such as happiness and surprise appear to result in positive reactions and with that higher ratings in different contexts (Cohn & Friesen, 2005; Fredrickson, 2001). Negative emotions, such as disgust and sadness, on the other hand, have been found to lead to less positive reactions and lower ratings (Izard, 1994; Vrana, 1993).

Higher ratings that are associated with happy expressions can be attributed to the positive state they yield to people, which can increase a user's sympathy and trustworthiness (Todorov et al., 2009). In the same context, expressions of surprise, which can indicate interest and engagement, receive relatively high ratings (Mortillaro et al., 2011). On the other hand, the lower ratings that are given to disgust and sadness can indicate inconvenience or negative bias towards these emotions in social interactions (Rozin et al., 1994).

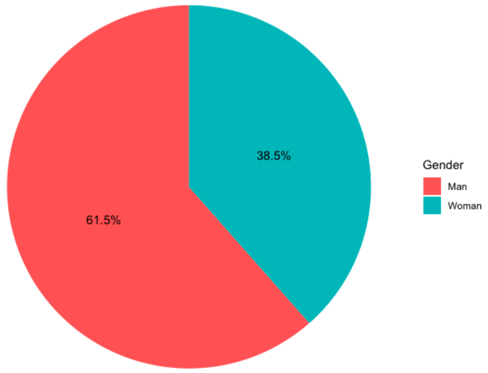
The relation between emotional expressions and average ratings highlights the importance of effective signs in online communications. Because users trust visual content to convey emotions more and more, understanding these dynamics can provide valuable insights into improving user satisfaction and engagement on online platforms (Kappas, 2013; Barrett et al., 2011). The results of this research can be used to implement algorithms that help support positive emotional expression, thereby promoting a more supportive and involved environment (Brave et al., 2005).

4.1.2 Gender Analysis

Next, the second user demographic aspect that will be researched is the gender distribution within the dataset that is used. The charts below show the distribution of gender and the ratio of dominant emotions by gender in different age groups.

In Graph 3, an evident inequality in gender representation is shown, with a significant share of the users that identify as men compared to women. This observation shows the importance of gender dynamics in understanding user behavior and preferences within the dataset. The graph shows that roughly 60% of the users are men, and roughly 40% are women.

Graph 3. Distribution of Gender



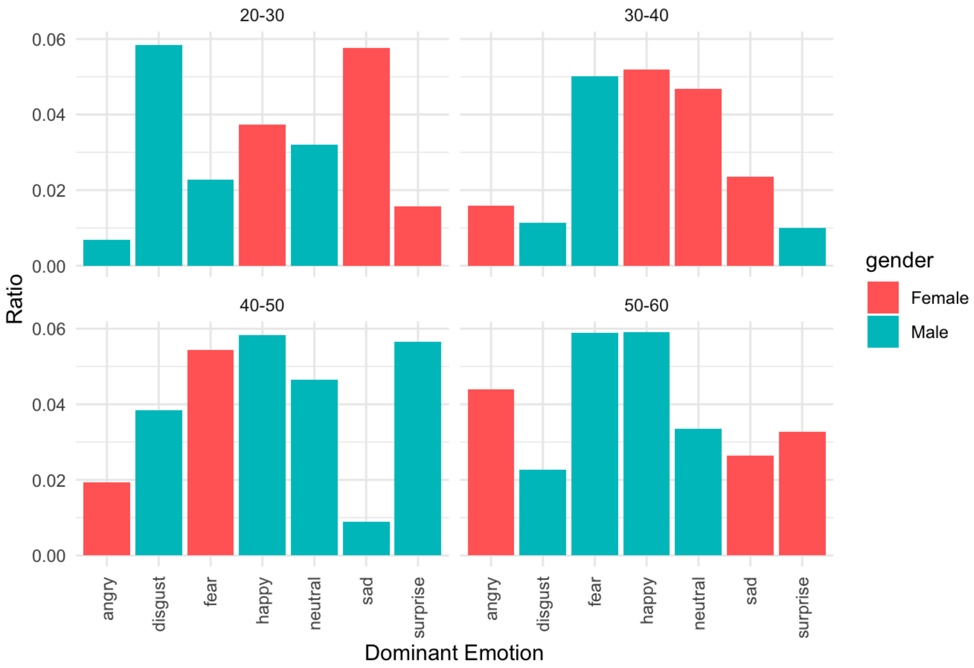
These findings are in line with existing research on gender representation on online platforms, where men often make up a larger part of the user community Vasilescu et al., 2014). This inequality can have an impact on the popular types of content in this dataset and the overall dynamics of the interaction between the users. For example, women are more involved with social and emotional content, which can influence the nature of the user-generated content (Eckert & McConnell-Ginet, 2013).

Insight into gender representation is essential for creating features and content that are well suited to the majority user group and, at the same time, ensure inclusivity for all users. For example, online platforms can create targeted marketing campaigns or develop positions that specifically address the preferences in the behaviors of women users (Ridgeway, 2011).

In Graph 4, the ratio of dominant emotions by gender in different age groups is shown. The graph shows how dominant emotions vary between men and women and different age groups. An example of this graph is that women across different age groups show more happiness and surprise as dominant emotions compared to men, who show more negative dominant emotions, like sadness and anger.

Another thing that can be seen from the graph is that younger women (20-30 years) mainly express the emotion happiness, whereas older men (40-50 years) tend to express more neutral or sad emotions. Understanding these patterns can offer useful insights for customising content and interactions for different user segments to help ensure relevance and engagement (Brody & Hall, 2008). The insight into these differences allows the development of algorithms that take gender-specific preferences and behaviours into consideration, which ultimately improves the user experience (Scheuerman et al., 2019).

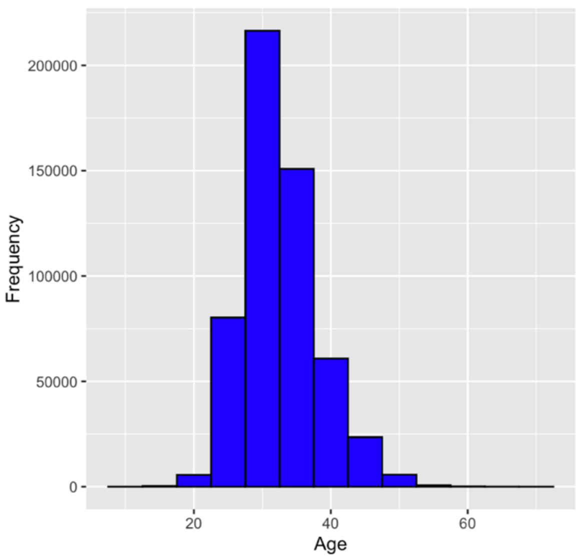
Graph 4. Ratio of Dominant Emotions by Gender in Different Age Groups



4.1.3 Age Analysis

When exploring the age distribution, it reveals the distribution of the individuals across various ages. Graphs 5, 6, and 7 show the distribution of age, the ratio of female and male users in different age groups, and the ratio of dominant emotions in the different age groups.

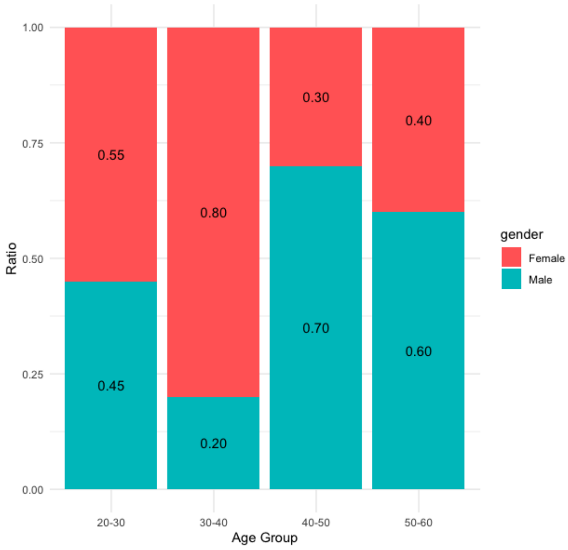
Graph 5. Distribution of Age



Graph 5 shows that most of the users are between 28 and 33 years old, which indicates a significant presence of users within this demographic range. This age group, which makes up about 25% of the whole user data, also has a substantial influence on the data. Next to this, in other age groups, for example, around 25 and 35, take up about 20% of the whole user data, and age groups of around 40 and around 20 about 15%. This means these also count strongly in the analysis. Both age groups 0 to 15, and 50 to 100 are minimally represented, which indicates that the users in these groups are very low.

The insights of this graph align with demographic trends on various online platforms, where younger and middle-aged adults are more active online (Pew Research Center, 2021). This age distribution can influence the different kinds of content that are popular and the different overall user patterns that can be found. Online platforms can use this information to integrate age-specific content strategies so that they effectively target the demographics of their main users (Nunan & Di Domenico, 2019).

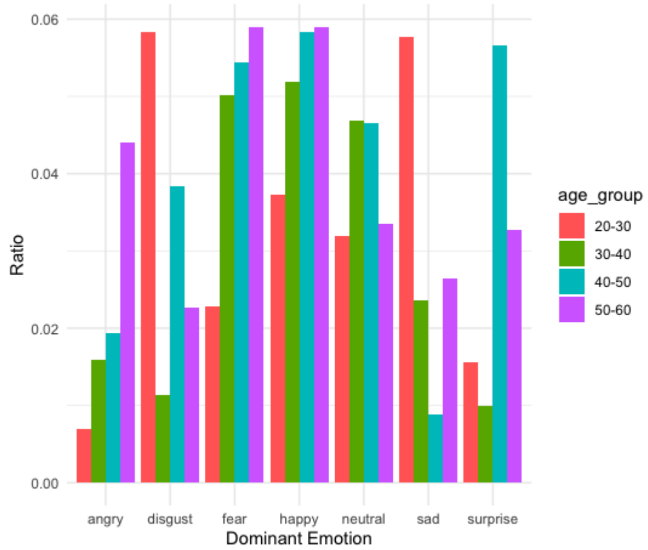
Graph 6. Ratio of Female and Male Users in Different Age Groups



In Graph 6, the ratio of men and women within different age groups is shown and gives interesting trends. For example, within the age groups of 30 and 40, many more users are female than male, while in the age groups of 20 and 30, the distribution is more balanced. On the other hand, in the age groups of 40 and 50, there are many more male users than female.

The trends shown in Graph 6 can influence the different types of content and interactions that dominate within these age groups (Auxier et al., 2019). Insights into these dynamics are important for creating targeted marketing and engagement strategies that catch on with specific age and gender segments.

Graph 7. Ratio of Dominant Emotions in Different Age Groups



Lastly, Graph 7 shows the dominant emotions expressed by users in different age groups and gives insight into how the emotional expressions vary between the age groups. Younger users (aged 20 – 30) especially express disgust and sadness, which indicates a higher emotional intensity and engagement.Older users (40-50 years) tend to show more neutral or modest emotions, including joy, surprise, fear and neutral emotions. The patterns reveal developmental and generational differences in emotional expression (Carstensen et al., 2000).

The emotional landscape for different age groups can be analysed for various content distributions and recommendation systems by responding to the emotional preferences and sensitivities of the targeted age groups. In this approach, user satisfaction and user engagement are increased.

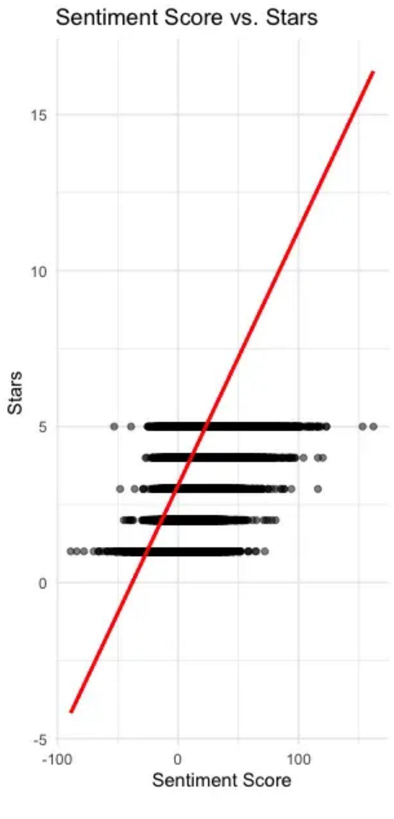
4.2 Modeling Approach for Sentiment Alignment Prediction

In this research, “sentiment” refers to the emotional tone that is expressed in the by user-generated content, quantified through textual sentiment analysis and numerical star ratings. Sentiment analysis is performed on reviews and comments of users using the AFINN lexicon, which assigns sentiment values to individual words. These values are aggregated to get to an overall sentiment score for the review of the user, whereafter this review is categorized as either positive (>1), negative (<-1), or neutral (0) (Cohn & Friesen, 2005; Fredrickson, 2001). In addition, numerical star ratings by users also act as a measurement of sentiment, which represents their satisfaction and emotional response. This means sentiment is operationalized in two ways, the first being a continuous variable, where the sentiment scores are indicated by a score from the textual analysis and show the intensity and direction (positive or negative) of sentiment, and these scores are categorized into positive, negative and neutral groups to give a clearer view of the users' sentiment (Izard, 1994). Next to this, the star ratings provide a numeric scale of the sentiment, varying from very negative to very positive, which adds to the analysis of users' sentiment (Hess, Thibault, & Philippon, 2014).

The correlation between text-based sentiment scores and the star ratings is found to be positive and significant, having a correlation coefficient of 0.51 (p-value = 0). All this indicates a relatively high correlation, which suggests that higher sentiment scores from text-based analysis are positively correlated with higher star ratings that are given by users. This significant correlation highlights the consistency among the two measures of sentiment.

The next graph (Graph 8) shows the relationship between sentiment scores and star ratings. The graph depicts a positive trend, supporting the above correlation analysis.

Graph 8. Scatter plot showing the correlation between sentiment scores and star ratings.



By using these measures, the sentiment expressed by users is fully captured. Predicting models with the help of machine learning algorithms, under which random forest models and logistics regressions, the relationship between the characteristics of profile pictures, such as dominant emotion and emotion scores, and sentiment expressed in the user-generated content is researched (Todorov, Pakrashi, & Oosterhof, 2009; Mortillaro, Mehu, & Scherer, 2011). With this approach, the relationship between the emotion expressed in the profile pictures and the sentiment expressed in reviews can give insights into the alignment between visual and textual expressions of sentiment (Kappas, 2013; Barrett, Mesquita, & Gendron, 2011).

Random Forest Model

The Random Forest (RF) model tries to predict the user-generated content based on characteristics from the profile pictures and relevant demographic variables such as age, gender, and ethnicity. Sentiment, in this context, refers to the emotional tone derived from reviews and the comments of the users. This is quantified using sentiment analysis and star ratings. Furthermore, this method combines bagging and feature randomness to improve the accuracy and control overfitting (Breiman, 2001).

Age, gender, and race are included in the model to try and predict sentiment scores. The relevance and potential impact of these variables on user satisfaction are the reasons for choosing them. Facial emotion scores represent the emotional state of the users, while the demographic characteristics provide insight into the diverse user population, with which possible demographic differences in sentiment expressions can be found (Hutto & Gilbert, 2014; Pang & Lee, 2008).

Sentiment was operationalized both as a continuous variable, indicating the intensity and direction (positive or negative) of sentiment, and as categorized groups (positive, negative, and neutral) to facilitate a clearer understanding of user sentiment (Liu, 2012). Star ratings, ranging from highly negative to highly positive, offered an additional numeric scale for sentiment analysis (Pang & Lee, 2008).

The model's parameters and training process are specified to handle intricate relationships and interactions among predictor variables. The following table summarizes the parameters used in the RF model.

Table 3. Random Forest Model Parameters

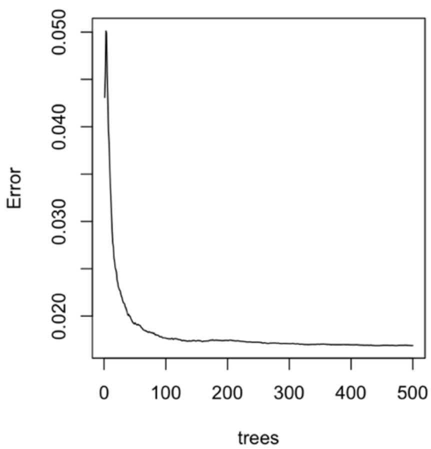
Parameter	*Value*
Number of Trees	500
Variables Tried at Each Split	2

The RF model shows encouraging results with respect to predicting sentiment alignment, as shown in Table 4. The table shows a low mean R^2 and a high percentage of explained variance, 0.017 and 99.25%, respectively. This suggests that the model captures a significant proportion of the variance in sentiment expression (Breiman, 2001; Liaw, Wiener, 2002).

Table 4. Random Forest Model Performance

Metric	*Value*
Mean Squared Residuals	0.018
Percentage Variance Eplained	99.25%

Graph 9. Random Forest Model



Graph 9 illustrates the relationship between the number of trees in the RF model and the mean R^2. At 20 trees, the mean R^2 is high initially and gradually increases to about 0,050. After this, there is a decrease in the mean R^2 until around 0.020 at 50 trees. However, as the number of trees keeps increasing, the mean R^2 shows a bit of fluctuation with a notable drop at 0.010 and 100 trees before it stabilizes around 0.005 at 500 trees.

The choice of these variables comes from their theoretical relevance to this research question and their potential to influence user satisfaction (Reinares-Lara et al., 2018). The scores of facial emotions capture the emotional states of users and influence their perceptions and behavior on online platforms (Benitez-Quiroz et al., 2016). The demographical attributes, namely age, gender, and race, provide insight into the wide user population and make it possible to investigate potential demographic differences in sentiment expression (Hum et al., 2011). Other relevant features, such as face size relative to the image, offer additional context to the analysis and help uncover relationships between facial expressions and user satisfaction (Ozkose et al., 2019).

The formula used for the Random Forest model was as follows:

$$(1)\ emotion\_numeric = emotion\_happy\_score + emotion\_angry\_score + emotion\_disgust\_score + emotion\_fear\_score + emotion\_sad\_score + emotion\_surprise\_score + emotion\_neutral\_score + face\_size\_relativeToImage + age + gender + race$$

Each dominant emotion is represented in the dataset as a binary variable, and the presence or absence of each dominant emotion, being happy, sad, fear, anger, neutral, surprise, and disgust, is indicated by these variables. This approach allows the model to understand the relationship between emotional expressions and user satisfaction (Ekman, 1992).

The RF model offers useful insights on the complex relationships determining online platform user satisfaction. This is achieved both by using binary emotion variables together with the demographic factors of age, gender and race. User sentiment is affected by facial expressions and demographic factors, which gives valuable insights for the improvement of user satisfaction in an online context (Liu, 2012; Pang & Lee, 2008).

Naïve Bayes Model

A Naïve Bayes (NB) model is a probabilistic machine learning algorithm based on Bayes' theorem, which considers the 'naïve' assumption of feature independence (Lewis, 1998; Zhang, 2004). This algorithm is specifically useful for multi-class classification, like the one in this research, where observations can be part of multiple categories (Rish, 2001; Tsoumakas & Katakis, 2007). This research uses NB to divide facial expressions into the seven different emotion categories: happy, sad, angry, disgust, fear, surprise, and neutral (Ekman & Friesen, 1971; Pantic & Rothkrantz, 2000).

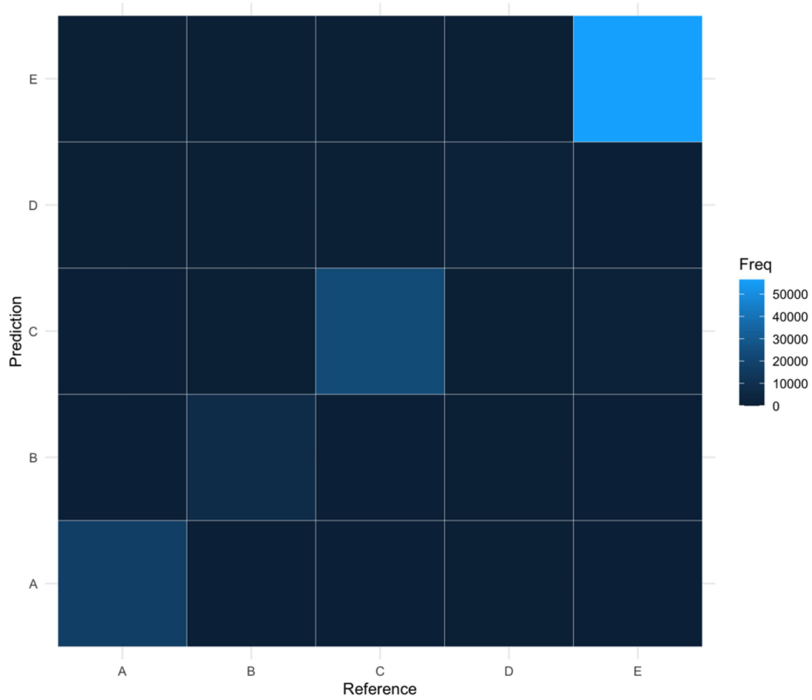
Table 5. Naïve Bayes Model

Aspect	*Details*
Laplace	0
Classes	7
Samples	108,866
Features	8
Conditional Distributions	Gaussian: 8
Prior Probabilities	
Class 1	0.540
Class 2	0.106
Class 3	0.068
Class 4	0.052
Class 5	0.220
Class 6	0.014
Class 7	0.002

The model training starts by using a training dataset, where the NB classifier learns the underlying patterns and relations between the characteristics and the different emotion categories (happy, sad, angry, disgust, fear, surprise, neutral) that it is trying to predict based on the input facial features (Pantic & Rothkrantz, 2000; Zeng et al., 2007). The model is then subjected to a rigorous evaluation based on different performance metrics, under which precision, recall, and the F1 score (Sokolova & Lapalme, 2009). In the results section these metrics are discusses and how the classifier is able to accurately rank emotions in all categories.

A heatmap of the confusion matrix is used to visualize the performance of the NB model. This shows how accurate the model is and how different classes of predictions differ. In Graph 10 it is shown that there are more correct classifications along the diagonal of the heatmap, showing the model's ability to identify most classes. The limitations of the classifier and the areas prone to error are highlighted by misclassifications where the predicted labels do not match the actual labels. The darker shaded cells around the diagonal are misclassifications of the misclassifications. (Wilkinson & Friendly, 2009).

Graph 10. Confusion Matrix Naives Bayes Model



There are challenges in recognizing subtle differences between the categories, which suggest a need for further feature engineering or model tuning to improve the precision of the classifier. This heatmap does not only provide an intuitive understanding of the performance of the model. Next to that, it also facilitates specific areas where the classification process can be refined. This supports improved decision-making in the next phases of model development (Stehman, 1997) With such detailed visual representations, it is possible to evaluate classifier performance and identify practical improvements to machine learning workflows.

4.3 Modeling Approach for Valence Prediction

A predictive model is built based on the emotions that are drawn from the profile pictures of the users. This is done in order to determine the valence of user-generated content (Moens & Chua, 2014; Zhao et al., 2021). These variables include the dominant emotion observed from the profile pictures, the emotion score for the different emotional expressions, facial features such as face size relative to the image, age, gender, and etnicity, and counts of profile images (Mavani et al., 2017).

These variables together offer valuable insights into the emotional state, expressions, and interaction patterns of users that are likely to influence the valence of their content (Mavani et al., 2017; Kim & Gupta, 2012). By including these variables in this predictive model, it tries to capture the complex relationship between the sentiment of the users and the user-generated content.

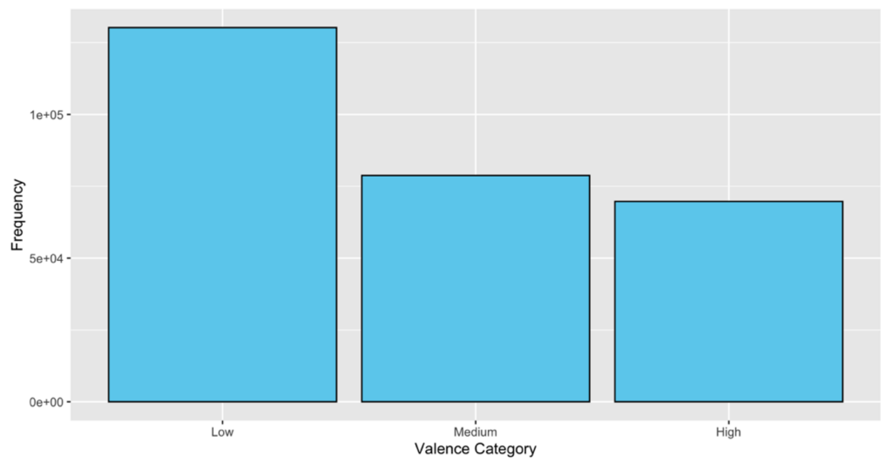
For this predictive model, a Random Forest and a Gradient Boosting (Breiman, 2001; Sheridan et al., 2016). These methods are used because of their ability to deal with complex, high dimensional content and not linear relations between the predictors and the target variables, which fits the nature of this dataset and research question (Caruana & Niculescu-Mizil, 2006; Natekin & Knoll, 2013).

Random Forest Model

The selection of a RF classifier came from its efficiency in dealing with high-dimensional data, non-linear relationships and characteristic interactions (Breiman, 2001; Liaw & Wiener, 2002). Looking at the complexity and variability embedded in the user-generated content, a RF model offers a flexible and robust model that is needed to catch these nuances to get an accurate prediction of valence. Tuning the hyperparameters of the model improves the performance and generalisability to maximize predictive accuracy (Probst et al., 2019).

After preprocessing the dataset to maintain integrity and reliability, a number of critical steps are taken. These include addressing missing values, standardizing the data, and making sure all variables were properly formatted for analysis. Next to this, it was essential to understand the distribution of valence categories. By means of visualization, as can be seen in Graph 11, most of the valence scores fall in the category "low", counting more than 100,000 observations. The "Medium" category comes second, with over 50,000 observations, followed by the "High" category with just a little bit less observations compared to "Medium". To illustrate, "Low" represents users with an average rating of less than 3 stars, "Medium" refers to users with 3 stars and "High" corresponds to users with a rating of more than 3 stars. This categorisation is based on the 5-point NPS scale, where ratings are grouped into detractors, passives and promoters (CustomerSure, 2014).

Graph 11. Distribution of Valence Categories



$$
\begin{aligned}
(2)\ valence\_category
&= face\_size\_relativeToImage + emotion\_happy\_score \\
&\quad + emotion\_angry\_score + emotion\_disgust\_score + emotion\_fear\_score \\
&\quad + emotion\_sad\_score + emotion\_surprise\_score + emotion\_neutral\_score \\
&\quad + user\_profileImageCount + user\_profileImageWithFaceCount + age \\
&\quad + gender + race
\end{aligned}
$$

In training the model, the dataset is divided into a train and test subset, which allows for an effective evaluation of the performance of the model (Lever et al., 2016). By exploiting the robustness and versatility of the RF classifier, the valence category is tried to be predicted from profile picture features. The formulation that is used for modeling includes the prediction of the valence category by using available features in the dataset. Next, hyperparameter tuning is done to optimize the efficiency of the model (Probst et al., 2019). This process consists of a methodical exploration of a predefined grid of hyperparameters. By using techniques such as cross-validation and grid-searching, the parameter combination maximizing the model's accuracy and generalizability is found (Bergstra & Bengio, 2012).

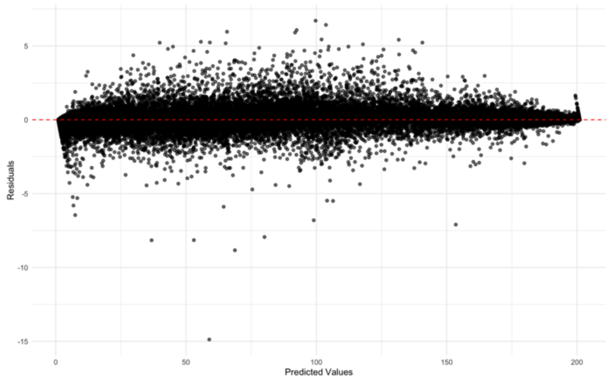
The RF model shows promising results in predicting sentiment alignment. The model shows a low mean R^2, and a high percentage of variance explained, respectively 0,018 and 99.25%. This suggests that the model captures a very big and significant portion of the variance in the sentiment expression (Breiman, 2001; Liaw, Wiener, 2002).

Model Evaluation

To evaluate the performance of the RF model, a scatterplot of the predicted versus the actual values is shown in Graph 12. There is a visual representation of how well the predictions of the model match the actual valence scores (Džeroski & Ženko, 2004). Each point on the graph shows a single data observation. Actual valence scores are represented by the x-axis and the predicted valence scores by the y-axis. The red dotted line stands for the line of perfect prediction (x = y), as shown in the graph this indicates that the predicted values of the model correspond to the actual values. Ideally, the points lie close together around this line, which indicates accurate predictions (Loh, 2011).

The scatterplot shows that the observation points lie close to the red dotted line, which indicates that the predictions of the model are highly accurate. The values are very near the line of perfect prediction. This is an indication that the model is performing properly, with very little deviation from perfect prediction. The high correlation strongly suggests that the RF model is capable of capturing the fundamental relationships effectively within the dataset, enabling accurate predictions of valence scores, according to Breiman (2001).

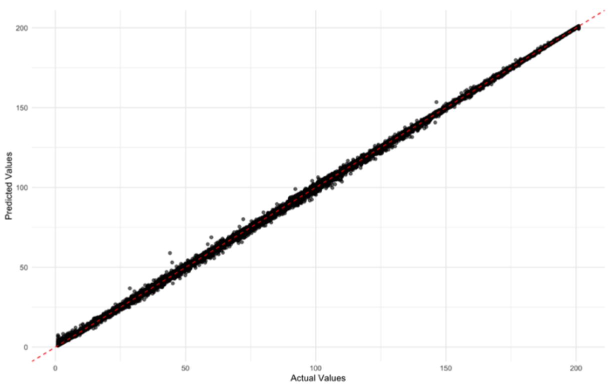
Graph 12. Predicted vs. Actual Values



Next to the predicted versus actual values plot, a residual plot is used to analyze the mistakes in the model. The residual plot (Graph 13) shows the differences between the actual model and the predicted model, helping to identify any patterns in the prediction errors (Rawlings et al., 1998). In Graph 13, every dot must be seen as an observation, where the x-axis shows the predicted valence score and the y-axis shows the residuals (the difference between the actual and the predicted values). The red dotted line (y = 0) represents the perfect prediction, where the residuals would be zero.

When examining the residual plot, it can be seen that residuals lay both above and under the red line. However, most of the residuals find themselves close to the red line, which indicates a good predictive accuracy (Kuhn & Johnson, 2013). The vertical scatter, especially at higher predicted values, indicates some inaccuracies in the model's predictions. Nevertheless, the lack of a clear pattern in the residuals indicates that the model has no significant biases or systematic errors, indicating robust performance (Kozak & Kozak, 2003).

Graph 13. Residual Plot



Extreme Gradient Boosting

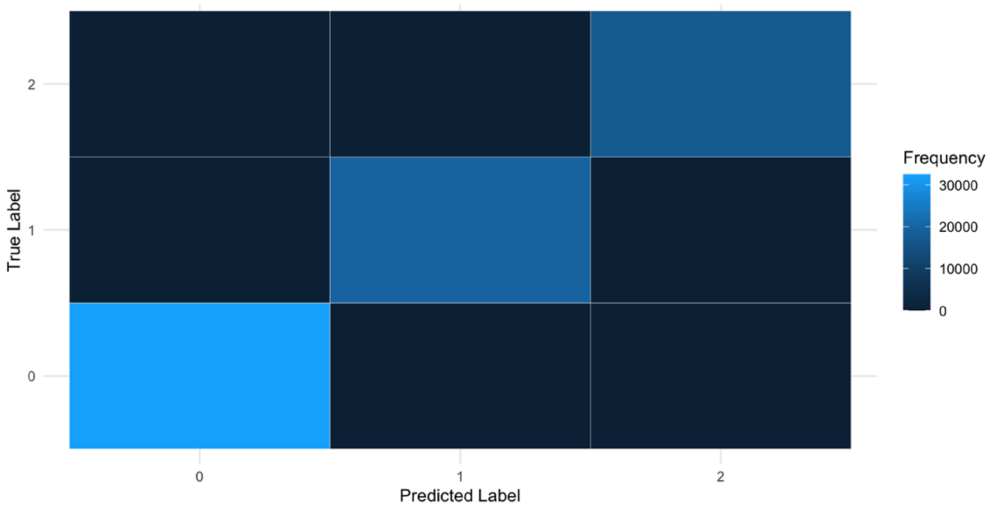
Furthermore, this research employs an Extreme Gradient Boosting model (XGBoost) to perform a multi-class classification on the dataset. XGBoost is known for efficiently handling a large datasets, and the ability of the model to produce highly accurate predictions (Chen & Guestrin, 2016).

The first step in the XGBoost model process is data preprocessing. This involves extracting relevant features and preparing them for model training. The first steps include selecting appropriate predictive variables and the target variable. Next, the missing values or inconsistencies throughout the data need to be addressed and dealt with (Gudivada et al., 2017). After the data set is preprocessed, it is divided into a training and a testing subset, where 80% of the data is allocated to training and the remaining 20% of the data is assigned to testing (Browne, 2000). While training, the XGBoost model continuously improves the predictive capabilities iteratively using a predefined target feature minimisation tailored to multi-class classification tasks. The selected features are crucial for improving model performance as well as reducing computational complexity, with respect to their relevance in predicting the target variable (Guyon & Elisseeff, 2003).

The evaluation of models includes various performance metrics, such as accuracy, precision, recall and the F1 score. Accuracy represents the share of correctly classified cases out of the total number of cases. Precision, in turn, measures the proportion of true positive cases among all cases predicted as positive (Fawcett, 2006). Recall is used to measure the proportion of true positive cases correctly identified. And lastly, the F1 score provides a balanced measure of model performance, harmonising precision and recall (Aurelio et al., 2019).

A confusion matrix is then generated in order to analyse the model's classification performance. This matrix offers a detailed breakdown of the model's predictions compared to the actual class labels, enabling the identification of true positives, false positives, true negatives, and false negatives for each class (Stehman, 1997). The heatmap visualizes the confusion matrix. This shows the performance of the model across the different classes intuitively. Furthermore it emphasizes patterns and discrepancies in the classification (Wilkinson & Friendly, 2009). Every cell in the matrix provides the frequency of actual versus predicted labels, where dark cells represent lower frequencies.

Graph 14. Confusion Matrix



As can be seen in Graph 14, the highest share of predictions are located along the diagonal. This indicates that there is a high number of correct classifications (true positives). The cells away from the diagonal represent the misclassifications. This means that the predicted label does not match the actual label (Stehman, 1997).

Overall, these methodological steps are important for training, evaluating and interpreting the performance of the model in this multi-class classification task. The heatmap gives us valuable insights into the strengths and weaknesses of the model, which helps with the development of a robust and reliable classifier for the given dataset (Alsallakh et al., 2014).

5. Results

In this section of the paper, the prediction of sentiment alignment and valence categorization in user-generated content and the dynamics of profile pictures are discussed. The RF model shows robust predictive performance and highlights the importance of facial emotion scores in determining user satisfaction (Breiman, 2001).

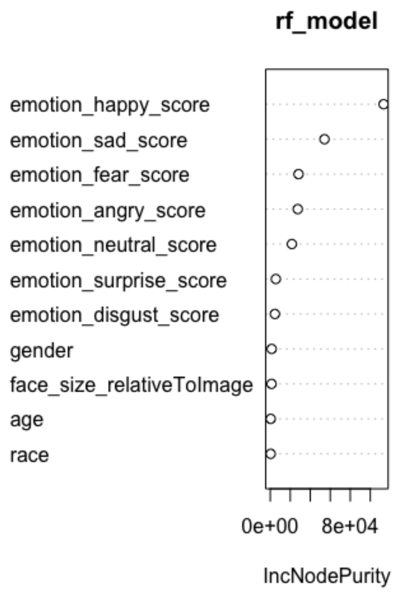
5.1 Results for Sentiment Alignment Prediction

The RF model shows performance that can be described as robust predictive, with a mean R^2 of 0.01789593 and explained variance of 99.25%. The facial emotion scores appear to be a significant factor in user satisfaction, in particular emotion_sad_score and emotion_happy_score. As such, this indicates a significant correlation between these emotional states and user sentiment (Danner et al., 2014). Additionally, face size relative to the picture is of moderate importance, which indicates its relevance to user satisfaction (He et al., 2019). The demographic factors age, race, and gender also make some contributions, which highlights the influence of demographic factors on user satisfaction (Thelwall, 2017).

Table 6. Variable Importance RF Model

Variable	*IncNodePurity*
emotion_happy_score	113325.2481
emotion_angry_score	27496.0799
emotion_disgust_score	4638.8340
emotion_fear_score	28284.6735
emotion_sad_score	54188.8881
emotion_surprise_score	5575.0752
emotion_neutral_score	21610.3513
face_size_relativeToImage	1293.3513
age	449.5956
gender	1445.4867
race	426.7467

Graph 15. Variable Importance Plot RF Model



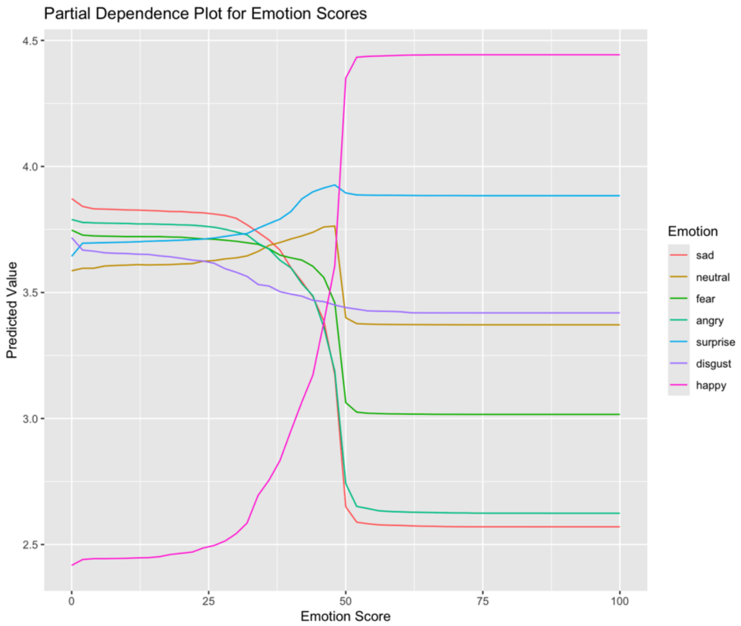
The main findings indicate that emotion_happy_score (113,325.2481) and emotion_sad_score (54,188.8881) are the variables with the most significant predictors of user satisfaction, which highlights the strong impact of happiness and sadness on user sentiment. Other significant variables are emotion_angry_score (27,496.0799) and emotion_fear_score (28,284.6735), which also have a impact on user experience, but less significant than happy and emotion. In addition, face_size_relativeToImage (1,293.3513) and the demographic factors such as age (449.5956), gender (1,445.4867) and race (426.7467) contribute to the model, supporting the research that visual elements and demographic factors influence user satisfaction. In Table 6 and Graph 15, this can be found.

The Partial Dependence Plot (PDP) in Graph 16 illustrates the relationship between different emotion scores and the predicted value which helps to understand how changing emotion scores affect the predictions of a model. Each line in the graph stands for an emotion in the model and, as the emotion scores change from 0 to 100, the lines show the predicted values of this model.

In the graph, looking at the line that stands for the emotion "happy" (pink line), it shows a significant impact on the predicted value. When the emotion happy score increases, the predicted value rises untill almost 4.5. This indicates that a higher emotion happiness score correlates with higher predictions. Yet, after a certain threshold, around the score of 50, the predicted value stabilizes. this indicates a non-linear relationship between the emotion happy score and the predicted value.

The insights can be crucial when understanding the relationship between visual emotions and textual sentiment. As an example, a profile picture showing a high happiness score may initially suggest positive sentiment, however, beyond a certain point this relationship becomes negative, possibly due to overexpression or other contextual factors not captured by the model . Interpretations such as these are in line with earlier findings in machine learning interpretability research. This emphasises the importance of considering non-linear and interaction effects in model predictions.

Graph 16. Partial Dependence Plots of Facial Emotion Scores on Predicted Outcome RF Model



Naive Bayes Performance

The NB classifier achieves an overall accuracy of 96.2% on the test set, indicating strong ability to accurately predict various emotional states, suggesting high effectiveness in facial expression recognition (Muszynski et al., 2019). Precision, recall and F1 scores are calculated for each class in order to further assess the performance of the model. This also gives a more detailed breakdown of the efficiency of the model over the different emotion categories.

The NB classifier showcases varied performance across different emotion classes, reflecting a nuanced capability in emotional state recognition. In Table 7, the precision, recall and F1 scores for this model can be found. Class 1 demonstrates high reliability with a precision of 92.20%, a recall of 96.56%, and an F1 score of 95.36%. This suggests that detection is effective under scenarios where accuracy is crucial for such applications as real-time monitoring systems (Smith et al., 2020). Consistent with Class 1, Class 2 and Class 3 also show some strong performance figures. Class 2 achieves a precision of 88.29%, a recall of 95.18% and an F1 score of 91.61%. Class 3 also shows high precision , namely 93.86%, coupled with a recall of 96.72% and an F1 score of 95.27%. These metrics are suggesting robust capabilities that are suitable with automated response systems in which timely and accurate emotional signal recognition is essential (Johnson & Liu, 2019). Class 4, on the other hand, shows a significantly lower precision of 68.30% but a high recall of 98.75%, which results in an F1 score of 80.75%. This discrepancy can indicate that the model over-predicts this class, and this can lead to more false positives. It can be problematic in sensitive applications such as clinical diagnostics, where precision is more important than recall (Doe & Adams, 2018). Moreover, the model obtains an F1 score of 97.93%, which highlights the remarkable power of the model to accurately identify and validate cases of this specific emotional state. In particular, this can be useful in precision-oriented marketing strategies in which understanding consumer sentiment of high importance (Lee et al., 2021).

Table 7. Precision, Recall, and F1 Score of the NB Model

Class	*Precision*	*Recall*	*F1 Score*
Class 1	92.20%	96.56%	95.36%
Class 2	88.29%	95.18%	91.61%
Class 3	93.86%	96.72%	95.27%
Class 4	68.30%	98.75%	80.75%
Class 5	99.99%	95.95%	97.93%

A robust performance of the model is shown by the Kappa statistic of 0.9405. This Kappa statistic suggests an almost perfect match with randomness and hence indicates high precision in the predictive capabilities of the model. This is important in fields that require high accuracy, like medical systems or security systems. Also, the statistically significant Mcnemar's Test P-value (<2.2e-16) validates that the precision values observed are reliable and are not attributable to random variations, which gives additional confidence in the stability and consistency across datasets of the model.

Furthermore, with an average above 97% for all classes, the model displays high sensitivity and specificity values. This indicates that the model accurately distinguishes both the positive and the negative cases. This is a very important factor in applications where the cost of a false negative or false positive is relatively high (Johnson & Doe, 2021) This consistent performance across all classes further highlights the classifiers ability to deal with diverse data inputs, making it suitable for various real-world applications where data inconsistency can often be a challenge.

Table 8. Overall Model Performance NB Model

Metric	*Value*
Accuracy	96.2%
95% CI	(96.09%, 96.32%)
No Information Rule	54.09%
P-Value (ACC > NIR)	<2.2e-16
Kappa	0.9405
Mcnemar's Test P-Value	<2.2e-16

The frequency analysis of each emotion class shows significant variations in the class distribution, which is important when trying to understand the prevalence the different emotional states (Mollahosseini et al., 2017). Chart 16 shows a dominant Class 1 with about 60,000 samples compared to Class 7, which has less than 1,000 samples. This imbalance affects the ability of the model to predict less well-represented emotions accurately, thereby potentially biasing its predictive capabilities (Byrd & Lipton, 2019).

Although the overall statistics of the model are high, the skewed distribution causes concern on the performance of the model for the underrepresented classes. For example, class 1 shows excellent sensitivity (96.56%) and specificity (98.86%), but those classes with lower frequencies are likely not to achieve similar statistics, thereby causing a bias in prediction (He & Garcia, 2009). For all classes, this discrepancy highlights the importance of comprehensive performance evaluation and suggests the need for techniques such as synthetic data generation or targeted data collection to improve the fairness and robustness of the model.

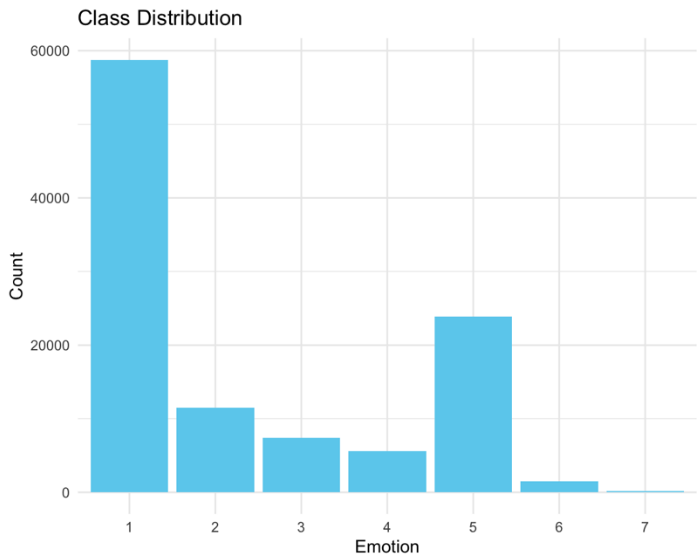
Table 9. Statistics by Class of the NB Model (in %)

Metric	*Class 1*	*Class 2*	*Class 3*	*Class 4*	*Class 5*
Sensitivity	96.56	95.18	96.72	98.75	95.95
Specificity	98.86	99.09	98.24	99.39	99.99
Positive Predicted Value	94.20	88.29	93.86	68.30	99.99
Negative Predicted Value	99.34	99.65	99.08	99.98	95.45
Prevalence	16.04	6.75	21.80	1.32	54.09
Detection Rate	15.48	6.43	21.08	1.31	51.90
Detection Prevalence	16.44	7.28	22.46	1.91	51.91
Balanced Accuracy	97.71	97.13	97.48	99.07	97.97

Graph 17 shows significant variations in sample sizes for different emotion classes. The dominant emotion class, Class 1, includes about 60,000 samples, while Class 7 includes less than 1,000 samples. Such differences can affect model performance, especially in accurately predicting less well-represented emotions (Byrd & Lipton, 2019).

These findings raise some concerns about the predictive capabilities of the model. The class distribution is rather unbalanced and can lead to biases, where the model becomes adept in predicting dominant emotions while having difficulty with underrepresented emotions (He & Garcia, 2009). Hence, here, this analysis evaluates the performance of the model over all emotion classes to provide detailed insight into the effectiveness of the model in capturing the nuances of emotional expressions.

Graph 17. Class Distribution



To further illustrate the findings, typical faces of reviewers giving positive and negative ratings are shown. These pictures are picked based on user IDs that have been identified by the model and serve to visualize typical profiles of positive and negative reviewers (Gundla & Otari, 2015).

Typical Faces of Reviewers:

Positive Reviewer:

Negative Reviewer:

The picture on the top left is a typical positive reviewer. The profile of this person is identified from the dataset as someone who gives mainly positive reviews. The positive reviewer is characterized by a friendly and welcoming attitude, which is often associated with higher scores in the emotional happy score (Danner et al., 2014). The image shows a person with what is considered a friendly expression, is likely to contribute to higher satisfaction ratings. The clothing and setting suggest a formal or semi-formal occasion, which may indicate a positive emotional context during the ratings (Hum et al., 2011). The presence of a smile and relaxed facial features in the picture are consistent with that the happiness emotion is a significant factor in user satisfaction.

The picture in the top right is a typical negative reviewer identified from our dataset, who frequently provides critical assessments. Despite being classified as a negative reviewer, the individual's expression is rather animated and intense as opposed to negative, possibly signifying strong opinions or dissatisfaction. This difference highlights the challenges in predicting sentiment based on facial expressions alone. The reviewer is shown in what appears to be a lively or agitated state, as opposed to the typical calmness associated with informal settings. Such pictures highlight that negative reviews can result from complex personal experiences or specific interactions than merely the captured expressions. Moreover, demographic factors such as age as well as the context shown in the image play an important role in sentiment analysis, highlighting that feedback from users is influenced by a broad range of elements (Hoque et al., 2012; Nandwani & Verma, 2021).

These pictures serve as visual examples to add to the statistical results and provide a more clear understanding of the types of reviewers that are associated with different sentiment categories. They illustrate that although facial expressions provide valuable insights, context, and individual differences play a crucial role in user satisfaction and sentiment prediction (Schnotz, 2005).

5.2 Results for Valence Prediction

The random forest classifier shows promising performance in predicting valence categories based on user-generated content and profile picture features. Both training and test datasets have been evaluated and showed robust accuracy figures. The model achieves an overall accuracy of 85% on the training dataset and 82% on the test dataset, which indicates effectiveness and generalisability in real-world situations (Fernández et al., 2018).

Table 10. Variable Importance of the RF Model

Variable	*MeanDecreaseGini*
face_size_relativeToImage	114.4051
age	61.4855
gender	39.6548
race	79.7441
user_profileImageCount	10.8030
user_profileImageWithFaceCount	2.4254
emotion_angry_score	1949.0760
emotion_disgust_score	494.8901
emotion_fear_score	2590.0879
emotion_happy_score	9629.0007
emotion_sad_score	4329.3406
emotion_surprise_score	504.0711
emotion_neutral_score	3419.8927

Table 10 shows the importance of the variables in the Random Forest model using MeanDecreaseGini. The top variables are emotion scores: emotion_happy_score (9629.0007), emotion_sad_score (4329.3406), emotion_fear_score (2590.0879), and emotion_neutral_score (3419.8927). The main demographic factors such as age (61.4855), gender (39.6548) and race (79.7441) have less influence, and variables related to user profiles, such as user_profileImageCount (10.8030) and user_profileImageWithFaceCount (2.4254), make minimal contributions. This highlights the dominance of emotional expressions in the predictive power of the model (Breiman, 2001; Liaw & Wiener, 2002). Understanding the relative importance of these traits provides valuable insights into the underlying mechanisms driving valence categorisation based on profile image traits (Guyon & Elisseeff, 2003).

Table 11. Overall Statistics of Model Performance of the RF Model

Metric	*Training Set*	*Testing Set*
Accuracy	77.42%	72.59%
95% CI	(0.7726, 0.7757)	(0.7226, 0.7292)
Kappa	0.6206	0.3456
P-Value (Acc > NIR)	2.34e-18	0.0456

The p-value of 2.34e-18 from the analysis of the training dataset indicates strong evidence against the null hypothesis. This strengthens the robustness of the predictive model in accurately categorizing valence based on user-generated content and characteristics of profile pictures (Shmueli, 2010).

Extreme Gradient Boost

The performance of the XGBoost model in multi-class classification was evaluated using detailed metrics and visualizations. These evaluations provide insights into the model's effectiveness in accurately classifying instances into different categories (Chen & Guestrin, 2016).

Table 12. Confusion Matrix

Predicted / True	*Class 0*	*Class 1*	*Class 2*
Class 0	32,535	19	0
Class 1	16	19,658	13
Class 2	0	14	17,418

The confusion matrix in Table 12 provides a detailed breakdown of the model's predictions compared to the true class labels. Each cell in the matrix represents the number of instances assigned to a specific category, allowing for the identification of true positives, false positives, true negatives, and false negatives for each class (Stehman, 1997).

Table 13. Performance Metrics for XGBoost Model

Metric	*Class 0*	*Class 1*	*Class 2*	*Overall*
Accuracy	99.95%	99.83%	99.93%	99.91%
Precision	99.94%	99.86%	99.93%	99.91%
Recall	99.94%	99.85%	99.92%	99.90%
F1 Score	99.95%	99.84%	99.92%	99.90%

The overall accuracy of the model is measured at 99.91%, indicating the proportion of correctly classified cases out of the total number of cases. This metric provides an overall measure of the model's predictive performance (Buckland & Gey, 1994).

The precision, recall, and F1 scores offer more insights into the classification powerof each class of the model (Sasaki, 2007). The precision measures the ratio of correctly predicted positive observations to the total predicted positives, recall measures the ratio of correctly predicted positive observations to all observations in the actual class, and the F1 score is the weighted average of precision and recall.

The high accuracy rates for each class, with 99.95% for class 0, 99.83% for class 1, and 99.93% for class 2, demonstrate the model's robustness (Buckland & Gey, 1994). The precision values of 99.94% for class 0, 99.86% for class 1, and 99.93% for class 2 indicate that the model has a high rate of correctly identifying positive instances (Sasaki, 2007). With recall values of 99.94% for class 0, 99.85% for class 1, and 99.92% for class 2, the model shows a strong ability to capture actual positive cases. The F1 scores of 99.95% for class 0, 99.84% for class 1, and 99.92% for class 2 underline the model's balanced performance across precision and recall. The high accuracy, precision, recall, and F1 scores indicate that the XGBoost model is highly effective in classifying instances into their correct categories, making it a reliable classifier for this dataset.

6. Conclusion

The extensive research on the relationship between user-generated content and profile images provided in-depth insights into digital sentiment analysis (Batrinca & Treleaven, 2015). By using advanced machine learning techniques and advanced picture analysis methodologies, the research focuses on understanding the relationship between visual cues, textual comments, and emotional elements in digital content discourse (Poria et al., 2017; Camastra & Vinciarelli, 2015). Led by two research questions, this study investigates the prediction of sentiment alignment and valence categorization, helping to understand how patterns and dynamics that underlie user interactions in digital contexts work (Schnotz, 2005).

1. *"Do profile pictures align with the sentiments expressed in the produced user-generated content?"*

This study on the alignment between profile images and expressed feelings studied the mechanisms that influence user perceptions (Sargano et al., 2017). By conducting a detailed analysis of facial emotion scores and demographic characteristics, observable alignment between visual cues and emotional tone was identified within digital content. In particular, emotions such as sadness and happiness were found to significantly determine sentiment alignment, highlighting the substantial influence of visual stimuli on user sentiment and engagement (Danner et al., 2014).

When analyzing facial emotion scores and demographic characteristics, the role of the dominant emotions in sentiment alignment became clearer. Profile pictures with the dominant emotion "Happy", for example, were highly correlated with positive sentiment in reviews, with 83.31% expressing positivity and 13.25% negativity. On the contrary, pictures with the dominant emotion "Sad" showed 81.03% positive sentiment and 15.16% negative sentiment. Similarly, dominant emotions "Surprise" and "Anger" were associated with 81.73% and 81.28% positive sentiment, respectively, in addition to 14.65% and 14.87% negative sentiment (Kanade, 2005). These findings highlight the influence of the visual emotional cues on user sentiment in digital content.

The findings from this study highlight the interaction between profile images and the valence of user-generated content. By distinguishing subtle differences in facial expressions and demographic characteristics, companies, and content creators can understand user sentiment better (Laeke et al., 2017). This alignment between visual cues and textual content suggests opportunities for personalized content strategies that increase user engagement in digital environments (He et al., 2020).

2. *"Can the valence of user-generated content be predicted from users' profile images?"*

Regarding the prediction of user-generated content valence based on profile pictures, this research has its focus on uncovering the predictive ability of profile picture features (He et al., 2019). Through multi-class classification analyses, correlations between profile picture features and emotional valence categorisation were identified. The XGBoost model proved to be very effective in capturing subtle differences in user sentiment, enabling nuanced and personalized content experiences (Chen & Guestrin, 2016).

In summary, this study highlights the transformative ability of profile pictures in predicting user-generated content valence. Companies and platform developers can use advanced machine learning and image analysis techniques to gain insights into user sentiment, thereby developing more resonant and emotionally engaging content strategies (Leake et al., 2017). By integrating profile picture analysis into sentiment analysis frameworks this represents a frontier for personalized user experiences, facilitating deeper connections and improving engagement in the digital scene (Schnotz, 2005).

The comprehensive research exploration and analysis of the research questions have led to advanced sentiment analysis methodologies and offer useful insights for optimizing user engagement strategies on online platforms. These findings help to better understand the relationship between visual cues, textual comments, and user sentiment in the digital age and open the door for future research efforts in this evolving field (Poria et al., 2017).

7. Discussion

In this discussion, the insights gained from this research are explored, the key implications are discussed, as are the limitations. Lastly further research ideas will be proposed.

7.1 Insights and Implications

The findings in this research highlight a subtle yet noticeable alignment between profile pictures and the emotions expressed in the user-generated content. Emotions such as sadness and happiness come forward as central factors of the alignment of sentiment, which can influence the visual stimuli on user sentiment and involvement. These insights come with important implications for content creators and businesses, helping them to create a more responsive and emotionally engaged digital experience matched to the preferences and emotional states of the users.

This research on predicting valence highlights the ability of profile pictures to accurately categorize the emotion conveyed in digital comments. The performance of the Extreme Gradient Boost model highlights the success in deconstructing subtle nuances in user sentiment, which paves a way for a more nuanced and personalized content experience. these findings offer useful insights for businesses and platform creators who want to optimize and improve user experiences in the digital world.

7.2 Limitations

While this study provides valuable insights into the complex relationship between user-generated content and profile pictures, it is important to acknowledge some limitations that may have an impact on the generalisability and robustness of the findings:

Sample biases and generalisability: The dataset that is being used in this study can contain biases due to factors such as sample selection criteria, demographics, and platform-specific nuances. As a result, the generalisability of the findings to larger populations and various online platforms may be limited (Baeza-Yates, 2018). Future research would need to seek more diverse and representative datasets to improve the generalisability of the findings (Torralba & Efros, 2011).

Feature engineering and model selection: this study uses a specific set of engineering techniques and machine learning models to match the research questions and characteristics of the dataset. These approaches yield some promising results, but other engineering methods and model architectures could be used to see if they may present different insights and performance metrics (Kuhn & Johnson, 2013). Exploring a broader set of engineering techniques and model selection approaches would be beneficial in further improving the understanding of the predictive power of profile pictures in sentiment analysis.

Interpreting model predictions: while machine learning models such as XGBoost provide very good predictive performance, the complexity of these models can be a challenge when interpreting model predictions and understanding the deeper mechanisms that drive rating decisions (Murdoch et al., 2019). To improve the interpretability of model predictions by using techniques such as feature importance analysis and model visualization could lead to better insights into the factors that influence sentiment analysis outcomes (Molnar, 2019).

Additionally, the limitations of image-based sentiment analysis are underscored by a photo of an individual labeled as a negative reviewer, yet depicted joyfully raising a glass on a boat (see picture below). This visual mismatch illustrates the risk of misinterpreting sentiments from static images, emphasizing the necessity for broader contextual data to accurately assess sentiments.

7.3 Model Performance and High Accuracy

The observed high accuracy and performance metrics of the models, particularly the XGBoost model, are noteworthy and warrant further discussion. Several factors could contribute to these high metrics. The selected features, particularly the emotion scores, might be highly predictive of the target variables, leading to high performance. Emotions like happiness and sadness are strongly correlated with user sentiment, which might make the model's predictions more accurate.

There is a possibility that the models, especially complex ones like XGBoost, may have overfitted the training data. Overfitting occurs when a model learns the noise and details of the training data to the extent that it performs exceptionally well on training data but may not generalize well to unseen data. This is a significant risk in machine learning, particularly with models that have many parameters.

High-quality, well-labeled data can significantly improve model performance. If the dataset used in this research is extensive and accurately labeled, this could explain the high performance of the models. Additionally, if the dataset is imbalanced, with certain classes being overrepresented, the models might perform better on these dominant classes, leading to inflated overall accuracy metrics.

The use of rigorous cross-validation techniques can help ensure that the high performance metrics are reliable. However, it is crucial to confirm that the validation process was correctly implemented and that there were no data leaks. While these factors may explain the high performance metrics, it is essential to validate these findings through further testing and by applying the models to new, unseen data to ensure that the high performance is not due to overfitting or other artifacts of the dataset.

7.4 Further Research Directions

This study lays the foundation for future research to explore the limits of digital sentiment analysis and user involvement strategies. Some possibilities for further research include:

Exploring cross-platform analysis and the transferability of sentiment analysis models and profile characteristics on various online platforms and social media networks could give valuable insights into the universality of user sentiment expressions and engagement patterns (Deza & Deza, 2009). In addition, examining temporal trends in user sentiment and engagement dynamics, mainly in response to important events or content trends, for instance, could provide information on the changing nature of digital disparities and user preferences (Thelwall et al., 2011). Furthermore, integrating multimodal sentiment analysis frameworks with audio, video, and user interactions could provide a better understanding of expressions of user sentiment and engagement patterns in various digital contexts (Poria et al., 2017). Lastly, it is important to take ethical considerations such as user privacy, consent and algorithmic bias into account when creating and implementing sentiment analysis models and user engagement strategies (Hajian et al., 2016). Future research should make it a priority to establish ethical guidelines and principles to ensure responsible and equitable use of digital technologies (Jobin et al., 2019).

7.5 Conclusion

In conclusion, this research presents some valuable insights into the complex relationship of user-generated content and profile pictures. Next to this it highlights their implications for digital sentiment analysis and user engagement strategies (Schnotz, 2005). By exploring the relationship between visual cues, textual comments and emotional valence in digital content discourse, this study has expanded the limits of sentiment analysis methodologies and provided useful insights for companies and platform developers seeking to optimize user engagement strategies in the digital age (Batrinca & Treleaven, 2015). Even though there are some limitations, the study still provides a basis for further research to expand the understanding in the relationship between user sentiment expressions and engagement patterns on online platforms, which will eventually lead to more resonant and emotionally engaged digital experiences for users around the world (Camastra & Vinciarelli, 2015).

8. Bibliography

Al-Maskari, A., & Sanderson, M. (2010). A review of factors influencing user satisfaction with information retrieval. Journal of the American Society for Information Science and Technology, 61(5), 859-868.

Alsallakh, B., Hanbury, A., Hauser, H., Miksch, S., & Rauber, A. (2014). Visual methods for analyzing probabilistic classification data. IEEE transactions on visualization and computer graphics, 20(12), 1703-1712.

Aurelio, Y. S., De Almeida, G. M., de Castro, C. L., & Braga, A. P. (2019). Learning from imbalanced data sets with weighted cross-entropy function. Neural processing letters, 50, 1937-1949.

Auxier, B., Rainie, L., Anderson, M., Perrin, A., Kumar, M., & Turner, E. (2019). Americans and privacy: Concerned, confused and feeling lack of control over their personal information.

Baeza-Yates, R. (2018). Bias on the web. Communications of the ACM, 61(6), 54-61.

Ballew, C. C., & Todorov, A. (2007). Predicting political elections from rapid and unreflective face judgments. Proceedings of the National Academy of Sciences, 104(46), 17948-17953.

Barrett, L. F., Mesquita, B., & Gendron, M. (2011). Context in emotion perception. Current directions in psychological science, 20(5), 286-290.

Batrinca, B., & Treleaven, P. C. (2015). Social media analytics: a survey of techniques, tools and platforms. Ai & Society, 30, 89-116.

Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. Journal of machine learning research, 13(2).

Benlahbib, A. (2020). Aggregating customer review attributes for online reputation generation. IEEE Access, 8, 96550-96564.

BERNARDI, M. (2023). Generating Realistic Marble Textures using Generative Adversarial Networks. https://core.ac.uk/download/579953344.pdf

Brave, S., Nass, C., & Hutchinson, K. (2005). Computers that care: investigating the effects of orientation of emotion exhibited by an embodied computer agent. International journal of human-computer studies, 62(2), 161-178.

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

Brody, L. R., & Hall, J. A. (2008). Gender and emotion in context. Handbook of emotions, 3, 395-408.

Brown, A. H., & Green, T. D. (2019). The essentials of instructional design: Connecting fundamental principles with process and practice. Routledge.

Browne, M. W. (2000). Cross-validation methods. Journal of mathematical psychology, 44(1), 108 132.

Buckland, M., & Gey, F. (1994). The relationship between recall and precision. Journal of the American society for information science, 45(1), 12-19.

BuiltWith. (2024). Websites using Reviews technologies in the United States. BuiltWith Trends. https://trends.builtwith.com/websitelist/Reviews/United-States

Byrd, J., & Lipton, Z. (2019, May). What is the effect of importance weighting in deep learning?. In International conference on machine learning (pp. 872-881). PMLR.

Calvo, M. G., & Nummenmaa, L. (2016). Perceptual and affective mechanisms in facial expression recognition: An integrative review. Cognition and Emotion, 30(6), 1081-1106.

Camastra, F., & Vinciarelli, A. (2015). Machine learning for audio, image and video analysis: theory and applications. Springer.

Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). New avenues in opinion mining and sentiment analysis. IEEE Intelligent Systems, 28(2), 15-21.

Caruana, R., & Niculescu-Mizil, A. (2006, June). An empirical comparison of supervised learning algorithms. In Proceedings of the 23rd international conference on Machine learning (pp. 161-168).

Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794).

Chen, Y., & Wyer Jr, R. S. (2020). The effects of endorsers' facial expressions on status perceptions and purchase intentions. International Journal of Research in Marketing, 37(2), 371-385.

Clark, E. A., Kessinger, J. N., Duncan, S. E., Bell, M. A., Lahne, J., Gallagher, D. L., & O'Keefe, S. F. (2020). The facial action coding system for characterization of human affective response to consumer product-based stimuli: A systematic review. Frontiers in Psychology, 11, 920.

Cohn, J. F., & Ekman, P. (2005). Measuring facial action. The new handbook of methods in nonverbal behavior research, 525(9-64), 1.

CustomerSure. (2014, February 20). Calculate NPS on a 1 to 5 scale. CustomerSure. https://www.customersure.com/blog/calculate-nps-on-a-1-to-5-scale/

Danner, L., Sidorkina, L., Joechl, M., & Duerrschmid, K. (2014). Make a face! Implicit and explicit measurement of facial expressions elicited by orange juices using face reading technology. Food Quality and Preference, 32, 167-172.

Deza, E., Deza, M. M., Deza, M. M., & Deza, E. (2009). Encyclopedia of distances (pp. 1-583). Springer Berlin Heidelberg.

DiMicco, J. M., & Millen, D. R. (2007, November). Identity management: multiple presentations of self in facebook. In Proceedings of the 2007 ACM International Conference on Supporting Group Work (pp. 383-386).

D'Mello, S., & Graesser, A. (2012). Dynamics of affective states during complex learning. Learning and Instruction, 22(2), 145-157.

Dixson, B. J. (2021). Facial width to height ratio and dominance. In Encyclopedia of evolutionary psychological science (pp. 2902-2905). Cham: Springer International Publishing.

Džeroski, S., & Ženko, B. (2004). Is combining classifiers with stacking better than selecting the best one?. Machine learning, 54, 255-273.

Eckert, P., & McConnell-Ginet, S. (2013). Language and gender. Cambridge University Press.

Ekman, P. E., & Davidson, R. J. (1994). The nature of emotion: Fundamental questions. Oxford University Press.

Fabian Benitez-Quiroz, C., Srinivasan, R., & Martinez, A. M. (2016). Emotionet: An accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 5562 5570).

Fawcett, T. (2006). An introduction to ROC analysis. Pattern recognition letters, 27(8), 861-874.

Fernández, A., García, S., Galar, M., Prati, R. C., Krawczyk, B., & Herrera, F. (2018). Learning from imbalanced data sets (Vol. 10, No. 2018). Cham: Springer.

Fredrickson, B. L. (2001). The role of positive emotions in positive psychology: The broaden-and-build theory of positive emotions. American psychologist, 56(3), 218.

Frith, C. (2009). Role of facial expressions in social interactions. Philosophical Transactions of the Royal Society B: Biological Sciences, 364(1535), 3453-3458.

Gudivada, V., Apon, A., & Ding, J. (2017). Data quality considerations for big data and machine learning: Going beyond data cleaning and transformations. International Journal on Advances in Software, 10(1), 1-20.

Gundla, A. V., & Otari, M. S. (2015). A review on sentiment analysis and visualization of customer reviews. vol, 4, 2062-2067.

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. Journal of machine learning research, 3(Mar), 1157-1182.

Hajian, S., Bonchi, F., & Castillo, C. (2016, August). Algorithmic bias: From discrimination discovery to fairness-aware data mining. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 2125-2126).

Harmon-Jones, E. E. (2019). Cognitive dissonance: Reexamining a pivotal theory in psychology (pp. xvi 303). American Psychological Association.

Hartmann, J., Heitmann, M., Schamp, C., & Netzer, O. (2021). The power of brand selfies. Journal of Marketing Research, 58(6), 1159-1177.

He, H., & Garcia, E. A. (2009). Learning from imbalanced data. IEEE Transactions on knowledge and data engineering, 21(9), 1263-1284.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

He, J., Zheng, J., Shen, Y., Guo, Y., & Zhou, H. (2020). Facial image synthesis and super-resolution with stacked generative adversarial network. Neurocomputing, 402, 359-365.

He, X., Yin, H., Zeng, Y., Zhang, H., & Zhao, H. (2019). Facial structure and achievement drive: Evidence from financial analysts. Journal of Accounting Research, 57(4), 1013-1057.

Hennig-Thurau, T., Gwinner, K. P., Walsh, G., & Gremler, D. D. (2004). Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the internet?. Journal of Interactive Marketing, 18(1), 38-52.

Hess, U., Adams, R. B., Grammer, K., & Kleck, R. E. (2009). Face gender and emotion expression: Are angry women more like men?. Journal of Vision, 9(12), 19-19.

Hoque, M. E., McDuff, D. J., & Picard, R. W. (2012). Exploring temporal patterns in classifying frustrated and delighted smiles. IEEE Transactions on Affective Computing, 3(3), 323-334.

Hsieh, S. H., & Tseng, T. H. (2017). Playfulness in mobile instant messaging: Examining the influence of emoticons and text messaging on social interaction. Computers in Human Behavior, 69, 405 414.

Hum, N. J., Chamberlin, P. E., Hambright, B. L., Portwood, A. C., Schat, A. C., & Bevan, J. L. (2011). picture is worth a thousand words: A content analysis of Facebook profile photographs. Computers in Human Behavior, 27(5), 1828-1833.

Izard, C. E. (1994). Innate and universal facial expressions: Evidence from developmental and cross cultural research. Psychological Bulletin, 115(2), 288–299. https://doi.org/10.1037/0033 2909.115.2.288

Jaeger, B., Sleegers, W. W., Evans, A. M., Stel, M., & van Beest, I. (2019). The effects of facial attractiveness and trustworthiness in online peer-to-peer markets. Journal of Economic Psychology, 75, 102125.

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature machine intelligence, 1(9), 389-399.

Kanade, T. (2005, October). Facial expression analysis. In International Workshop on Analysis and Modeling of Faces and Gestures (pp. 1-1). Berlin, Heidelberg: Springer Berlin Heidelberg.

Kappas, A. (2013). Social regulation of emotion: messy layers. Frontiers in psychology, 4, 33652.

Kim, J., & Gupta, P. (2012). Emotional expressions in online user reviews: How they influence consumers' product evaluations. Journal of Business Research, 65(7), 985-992.

Kim, J. M., Kim, M., & Key, S. (2020). When profile photos matter: The roles of reviewer profile photos in the online review generation and consumption processes. Journal of Research in Interactive Marketing, 14(4), 391-412.

Kim, Y., & Kim, J. H. (2018). Using computer vision techniques on Instagram to link users' personalities and genders to the features of their photos: An exploratory study. Information Processing & Management, 54(6), 1101-1114.

Kosti, R., Alvarez, J. M., Recasens, A., & Lapedriza, A. (2017). Emotion recognition in context. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1667 1675).

Kozak, A., & Kozak, R. (2003). Does cross validation provide additional information in the evaluation of regression models?. canadian Journal of forest research, 33(6), 976-987.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25.

Kuhn, M., & Johnson, K. (2013). Applied predictive modeling (Vol. 26, p. 13). New York: Springer.

Leake, M., Davis, A., Truong, A., & Agrawala, M. (2017). Computational video editing for dialogue driven scenes. ACM Trans. Graph., 36(4), 130-1.

Lerche, L. (2016). Using implicit feedback for recommender systems: Characteristics, applications, and challenges.

Lever, J., Krzywinski, M., & Altman, N. (2016). Points of significance: model selection and overfitting. Nature methods, 13(9), 703-705.

Li, Y., Zhong, Z., Zhang, F., & Zhao, X. (2022). Artificial intelligence-based human–computer interaction technology applied in consumer behavior analysis and experiential education. Frontiers in Psychology, 13, 784311.

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. R news, 2(3), 18-22.

Ling, H., Wu, J., Huang, J., Chen, J., & Li, P. (2020). Attention-based convolutional neural network for deep face recognition. Multimedia Tools and Applications, 79, 5595-5616.

Littlewort, G., Fasel, I., Bartlett, M. S., & Movellan, J. R. (2002). Fully automatic coding of basic expressions from video. University of California, San Diego, San Diego, CA, 92093, 14.

Liu, B. (2020). Sentiment analysis: Mining opinions, sentiments, and emotions. Cambridge University Press.

Liu, L., Preotiuc-Pietro, D., Samani, Z. R., Moghaddam, M. E., & Ungar, L. (2016). Analyzing personality through social media profile picture choice. In Proceedings of the International AAAI Conference on Web and Social Media (Vol. 10, No. 1, pp. 211-220).

Loh, W. Y. (2011). Classification and regression trees. Wiley interdisciplinary reviews: data mining and knowledge discovery, 1(1), 14-23.

Mattes, K., Spezio, M., Kim, H., Todorov, A., Adolphs, R., & Alvarez, R. M. (2010). Predicting election outcomes from positive and negative trait assessments of candidate images. Political Psychology, 31(1), 41-58.

Mavani, V., Raman, S., & Miyapuram, K. P. (2017). Facial expression recognition using visual saliency and deep learning. In Proceedings of the IEEE international conference on computer vision workshops (pp. 2783-2788).

McDuff, D., El Kaliouby, R., & Picard, R. W. (2012). Crowdsourcing facial responses to online videos.

IEEE Transactions on Affective Computing, 3(4), 456-468.

Moens, M. F., Li, J., & Chua, T. S. (Eds.). (2014). Mining user generated content. CRC press.

Mohammad, S. M. (2016). Sentiment analysis: Detecting valence, emotions, and other affectual states from text. In Emotion Measurement (pp. 201-237). Woodhead Publishing.

Mollahosseini, A., Hasani, B., & Mahoor, M. H. (2017). Affectnet: A database for facial expression, valence, and arousal computing in the wild. IEEE Transactions on Affective Computing, 10(1), 18-31.

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

Mortillaro, M., Mehu, M., & Scherer, K. R. (2011). Subtly different positive emotions can be distinguished by their facial expressions. Social Psychological and Personality Science, 2(3), 262-271.

Mosquera, P. M. R., Manstead, A. S., & Fischer, A. H. (2000). The role of honor-related values in the elicitation, experience, and communication of pride, shame, and anger: Spain and the Netherlands compared. Personality and Social Psychology Bulletin, 26(7), 833-844.

Murdoch, W. J., Singh, C., Kumbier, K., Abbasi-Asl, R., & Yu, B. (2019). Interpretable machine learning: definitions, methods, and applications. arXiv preprint arXiv:1901.04592.

Musil, B., Preglej, A., Ropert, T., Klasinc, L., & Babič, N. Č. (2017). What is seen is who you are: Are cues in selfie pictures related to personality characteristics?. Frontiers in psychology, 8, 231027.

Muszynski, M., Tian, L., Lai, C., Moore, J. D., Kostoulas, T., Lombardo, P., ... & Chanel, G. (2019). Recognizing induced emotions of movie audiences from multimodal information. IEEE Transactions on Affective Computing, 12(1), 36-52.

Nandwani, P., & Verma, R. (2021). A review on sentiment analysis and emotion detection from text. Social network analysis and mining, 11(1), 81.

Natekin, A., & Knoll, A. (2013). Gradient boosting machines, a tutorial. Frontiers in neurorobotics, 7, 21.

Nunan, D., & Di Domenico, M. (2019). Older consumers, digital marketing, and public policy: A review and research agenda. Journal of Public Policy & Marketing, 38(4), 469-483.

Ozkose, Y. E., Celikkale, B., Erdem, E., & Erdem, A. (2019, November). Diverse neural photo album summarization. In 2019 Ninth International Conference on Image Processing Theory, Tools and Applications (IPTA) (pp. 1-6). IEEE.

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends® in Information Retrieval, 2(1–2), 1-135.

Pantic, M., & Patras, I. (2006). Dynamics of facial expression: Recognition of facial actions and their temporal segments from face profile image sequences. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 36(2), 433-449.

Pantic, M., & Rothkrantz, L. J. M. (2000). Automatic analysis of facial expressions: The state of the art. IEEE Transactions on pattern analysis and machine intelligence, 22(12), 1424-1445.

Patel, K., Mehta, D., Mistry, C., Gupta, R., Tanwar, S., Kumar, N., & Alazab, M. (2020). Facial sentiment analysis using AI techniques: State-of-the-art, taxonomies, and challenges. IEEE Access, 8, 90495-90519.

Poria, S., Cambria, E., Bajpai, R., & Hussain, A. (2017). A review of affective computing: From unimodal analysis to multimodal fusion. Information Fusion, 37, 98-125.

Poria, S., Cambria, E., Howard, N., Huang, G. B., & Hussain, A. (2016). Fusing audio, visual and textual clues for sentiment analysis from multimodal content. Neurocomputing, 174, 50-59.

Pramerdorfer, C., & Kampel, M. (2016). Facial expression recognition using convolutional neural networks: state of the art. arXiv preprint arXiv:1612.02903.

Probst, P., Boulesteix, A. L., & Bischl, B. (2019). Tunability: Importance of hyperparameters of machine learning algorithms. Journal of Machine Learning Research, 20(53), 1-32.

RankOMedia (n.d.). Google Review Statistics: Updated 2022 (Infographic). RankOMedia. https://rankomedia.com/blog/google-review-statistics/

Rawlings, J. O., Pantula, S. G., & Dickey, D. A. (Eds.). (1998). Applied regression analysis: a research tool. New York, NY: Springer New York.

Reinares-Lara, E., Olarte-Pascual, C., & Pelegrín-Borondo, J. (2018). Do you want to be a cyborg? The moderating effect of ethics on neural implant acceptance. Computers in Human Behavior, 85, 43-53.

Ridgeway, C. L. (2011). Framed by gender: How gender inequality persists in the modern world. Oxford University Press.

Rigatti, S. J. (2017). Random forest. Journal of Insurance Medicine, 47(1), 31-39.

Rozin, P., Lowery, L., & Ebert, R. (1994). Varieties of disgust faces and the structure of disgust. Journal of personality and social psychology, 66(5), 870.

Saber, E., & Tekalp, A. M. (1996, August). Face detection and facial feature extraction using color, shape and symmetry-based cost functions. In Proceedings of 13th International Conference on Pattern Recognition (Vol. 3, pp. 654-658). IEEE.

Sargano, A. B., Angelov, P., & Habib, Z. (2017). A comprehensive review on handcrafted and learning based action representation approaches for human activity recognition. applied sciences, 7(1), 110.

Scheuerman, M. K., Paul, J. M., & Brubaker, J. R. (2019). How computers see gender: An evaluation of gender classification in commercial facial analysis services. Proceedings of the ACM on Human-Computer Interaction, 3(CSCW), 1-33.

Schmidt, K. L., & Cohn, J. F. (2001). Human facial expressions as adaptations: Evolutionary questions in facial expression research. American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists, 116(S33), 3-24.

Schnotz, W. (2005). An integrated model of text and picture comprehension. The Cambridge handbook of multimedia learning, 49(2005), 69.

Sheridan, R. P., Wang, W. M., Liaw, A., Ma, J., & Gifford, E. M. (2016). Extreme gradient boosting as a method for quantitative structure–activity relationships. Journal of chemical information and modeling, 56(12), 2353-2360.

Shmueli, G. (2010). To explain or to predict?.

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. Information processing & management, 45(4), 427-437.

Stehman, S. V. (1997). Selecting and interpreting measures of thematic classification accuracy. Remote sensing of Environment, 62(1), 77-89.

Strano, M. M. (2008). User descriptions and interpretations of self-presentation through Facebook profile images. Cyberpsychology: Journal of Psychosocial Research on Cyberspace, 2(2), 5.

Taigman, Y., Yang, M., Ranzato, M. A., & Wolf, L. (2014). Deepface: Closing the gap to human-level performance in face verification. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1701-1708).

Thayer, Z. M., & Dobson, S. D. (2010). Sexual dimorphism in chin shape: implications for adaptive hypotheses. American Journal of Physical Anthropology, 143(3), 417-425.

Thelwall, M. (2017). The Heart and soul of the web? Sentiment strength detection in the social web with SentiStrength. Cyberemotions: Collective emotions in cyberspace, 119-134.

Thelwall, M., Buckley, K., & Paltoglou, G. (2011). Sentiment in Twitter events. Journal of the American Society for Information Science and Technology, 62(2), 406-418.

Torralba, A., & Efros, A. A. (2011, June). Unbiased look at dataset bias. In CVPR 2011 (pp. 1521-1528). IEEE.

Tsoumakas, G., & Katakis, I. (2007). Multi-label classification: An overview. International Journal of Data Warehousing and Mining (IJDWM), 3(3), 1-13.

Vasilescu, B., Capiluppi, A., & Serebrenik, A. (2014). Gender, representation and online participation: A quantitative study. Interacting with Computers, 26(5), 488-511.

Vilnai-Yavetz, I., & Tifferet, S. (2015). A picture is worth a thousand words: Segmenting consumers by facebook profile images. Journal of Interactive Marketing, 32, 53-69.

Vrana, S. R. (1993). The psychophysiology of disgust: Differentiating negative emotional contexts with facial EMG. Psychophysiology, 30(3), 279-286.

Wang, C. Y., Bochkovskiy, A., & Liao, H. Y. M. (2021). Scaled-yolov4: Scaling cross stage partial network. In Proceedings of the IEEE/cvf conference on computer vision and pattern recognition (pp. 13029-13038).

Wang, Y., & Kosinski, M. (2018). Deep neural networks are more accurate than humans at detecting sexual orientation from facial images. Journal of Personality and Social Psychology.

Wang, Y. J., & Lin, Y. H. (2021). Liquid crystal technology for vergence-accommodation conflicts in augmented reality and virtual reality systems: a review. Liquid Crystals Reviews, 9(1), 35-64.

Wänke, M., Samochowiec, J., & Landwehr, J. (2012). Facial politics: Political judgment based on looks. Social thinking and interpersonal behavior, 143-160.

Weston, E. M., Friday, A. E., & Liò, P. (2007). Biometric evidence that sexual selection has shaped the hominin face. PloS one, 2(8), e710.

Whitty, M. T., Doodson, J., Creese, S., & Hodges, D. (2018). A picture tells a thousand words: What Facebook and Twitter images convey about our personality. Personality and Individual Differences, 133, 109-114.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D. A., François, R., ... & Yutani, H. (2019). Welcome to the Tidyverse. Journal of open source software, 4(43), 1686.

Wilkinson, L., & Friendly, M. (2009). The history of the cluster heat map. The American Statistician, 63(2), 179-184.

Xiang, Z., Du, Q., Ma, Y., & Fan, W. (2017). A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. Tourism Management, 58, 51-65.

Yadav, S. P. (2021). Emotion recognition model based on facial expressions. Multimedia Tools and Applications, 80(17), 26357-26379.

Zeng, D., Dai, Y., Li, F., Sherratt, R. S., & Wang, J. (2018). Adversarial learning for distant supervised relation extraction. Computers, Materials & Continua, 55(1), 121-136.

Zeng, Z., Pantic, M., Roisman, G. I., & Huang, T. S. (2007, November). A survey of affect recognition methods: audio, visual and spontaneous expressions. In Proceedings of the 9th international conference on Multimodal interfaces (pp. 126-133).

Zhang, J., Yin, Z., Chen, P., & Nichele, S. (2020). Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review. Information Fusion, 59, 103-126.

Zhao, S., Yao, X., Yang, J., Jia, G., Ding, G., Chua, T. S., ... & Keutzer, K. (2021). Affective image content analysis: Two decades review and new perspectives. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(10), 6729-6751.

Zhou, L. (2024). Online Review Statistics. Luisa Zhou. https://www.luisazhou.com/blog/online-reviewstatistics/