Exploration of the Most Preferred Social Media for the Fashion Business Practices

MSc Research Project
Data Analytics

Apurva Kumari
Student ID: X21118175

School of Computing
National College of Ireland

Supervisor: Mr Bharat Agarwal

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Apurva Kumari
Student ID:	X21118175
Programme:	Msc Data Analytics
Year:	2022
Module:	Research Project
Supervisor:	Mr Bharat Agarwal
Submission Due Date:	15/12/2022
Project Title:	Exploration of the Most Preferred Social Media for the Fashion Business Practices
Word Count:	6685
Page Count:	17

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Apurva Kumari
Date:	13th December 2022

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Exploration of the Most Preferred Social Media for the Fashion Business Practices

Apurva Kumari
X21118175

Abstract

This study looks into the problem to find out how social media affects the fashion industry. In the last ten years, social media have come a long way. Companies are using social media as a way to market their products. Social media is the best and cheapest way for the fashion industry, which is one of the industries with the most growth, to talk to people. Text processing is one of the most natural applications for machine learning and deep learning models. There are, however, few research that combine the two. This study uses a combination of sentiment analysis and text mining to find out how customers behave and how happy they are with different fashion businesses on social media, where customers are very active in sharing their opinions in real time. This study uses Twitter data in many ways, including retrieval, cleaning, feature selection, and classification using three different machine learning algorithms (Naive Bayes, XGBoost, and SVM) and one Deep ML technique (BERT). After the raw data has been cleaned up and the key features for a classification algorithm have been extracted, classification and model validation are done. This study shows that, compared to other classification models, the BERT model has the highest classification accuracy and the highest weighted value of precision, with 99.0% accuracy. The results of this study say that almost 61% of the texts collected had positive things to say about different fashion labels. When it comes to the number of tweets per country, Kampala, Uganda is at the top of the list.

1 Introduction

1.1 Background

People's lives have changed a lot in the last 10 years because of social media. What was first used for talking, staying in touch, and getting back in touch with old friends has changed how business is done in a big way. The fashion industry is one area of business that has been changed a lot by the rise of social media (Arrigo; 2018). Due to the growth of social media, information that was once only available to a small group of insiders is now easily accessible to everyone. For example, thanks to a designer's Facebook page,Twitter feed, or Instagram account, people who weren't interested in high fashion before can now watch live videos from New York City's Mercedes-Benz Fashion Week. To keep up with technology, all kinds of businesses need to change, but fashion shops need to pay special attention.

Some things about fashion are changed by the environment. In other words, fashion is very flexible, and the latest trends change all the time. From the customer's point of view, it is easy to see how social media can be used to speed up the spread of these trends. But from a retailer's point of view, it's important to be connected and knowledgeable to better meet customer needs

(Arrigo; 2018). Like fashion, social media is a way to show how you feel. But you can find examples of this phrase online. There are a number of threads that connect the two. For example, both channels show how important it is to have a certain persona from a psychological point of view. Some people dress a certain way to show the public how they want to be seen, just like they might set up their Facebook page, Twitter account, or Instagram account (Arrigo; 2018).
People can also show that they want to fit in and be liked by how they use social media and what they wear. Social networking seems to be a great fit for one industry, fashion, in a way that makes sense. As a part of the fashion industry, social media is used to direct social networks and other digital sites that let fashion companies talk to their customers using the latest social networking technology (Batrinca and Treleaven; 2015). Social media is used by the fashion industry to study trends and predict how customers will act. It has become more popular because it uses social media. Since social networking sites like Facebook and Twitter are so popular, the Factionists now have a new way to get ideas (Batrinca and Treleaven; 2015). The designers and companies with a lot of followers are the ones who really get what social media can do for them. Also, fashion bloggers are becoming more important in the industry.
Upmarket businesses are more interested in viral marketing and word-of-mouth promotion now that more people are using social media sites like Facebook and Twitter. Word of mouth (WOM), also called customer-to-customer interaction, is one of the most important ways for consumers to learn about the market (Park et al.; 2020). No matter how good the content is, WOM that spreads online has the potential to be very powerful. Followers are having a harder time figuring out what is true and what is not because the information is so short and there is no way to tell if it is true or not. Because information gets around, people gather in public places to talk about their ideas.
Thanks to the internet, everyone can now find out what is in style. Thanks to blogs, customers now have almost unlimited space on the Internet to say what they want. Blogging is the term for a personal website that is usually run by one person who regularly adds comments, explanations of events, or other content like pictures or videos. The postings are usually shown in reverse chronological order (Park et al.; 2020). Unlike fashion-focused newspapers and TV shows, there are millions of fashion magazines all over the world, and they are often updated with the latest fashion trends. The blog is so popular because it has a strong, personal, broad, and elitist point of view. Readers can say what they think and talk to the fashion critics, which is a fun experience.

1.2 Research Questions

1. How do social media helps fashion businesses to attract consumers?
2. What is the result of social media on the fashion industry?

1.3 Research Objectives

1. To explore social media platforms in order to improve the fashion business.
2. To analyze the consequences of social media on the fashion industry.

2 Related Work

(Wadhe and Suratkar; 2020) and (Chatterjee et al.; 2021) are two of the recent studies that have focused on sentiment classification, text data, and data analysis. Numerous publication types were reviewed between 2015 and 2021 for the proposed research. There are two sections: 2.1

discusses how social media has affected the fashion industry, and 2.2 details the various facets of online networking sites. Two research papers are summarized in Table 1 within Overview section.

2.1 The consequence of social media on the fashion industry

The most recent fashion industry trend is social networking. Social networking platforms are becoming increasingly popular and accepted within businesses (Wu and Song; 2019). The majority of companies have seen a rise in their social media presence over the past year. In contrast to other apparel categories, the presence of fashion shops on social networking platforms has increased dramatically, according to (Lee et al.; 2018) from one of the most reputable and regarded sources of digital market research. According to (Lee et al.; 2018), 28% of respondents talked about brands in internet forums, and 19% posted brand-related content on the home pages of their favourite social media sites. In an effort to create their brand personas in real-time on global digital platforms, fashion couture companies, designers, and merchants frequently tweet, weblog, and update their profiles (Laurell et al.; 2019). To give the brand a human voice, the designers use social media to share films, advertisements, behind-the-scenes footage, off-screen content, and design presentations. According to (Lee et al.; 2018), fashion houses, companies, and stores use social networking platforms to promote in-the-moment client involvement. Customers are made to feel as though they are an integral part of both the brand and the broader company family. The interactive component just makes this connection stronger. Social media platforms offer the advantage of allowing businesses to take part in discussions about their own branding (Abbas et al.; 2019).

According to (Yu and Hu; 2020), since it demonstrates that they are connected with their clients, launching corporate accounts on social media platforms like Facebook aids businesses in developing their brands (Pedersen et al.; 2019). Designers have the opportunity to display their authenticity and vision while showcasing their cutting-edge looks and trends on social media. Traditional business practices provide a solid basis for the fashion industry. However, thanks to social media, firms in the fashion industry may now depend more on technology than employees, which is unquestionably a speedier approach to establishing a brand. According to (Yu and Hu; 2020), retailers and clothing producers may build and maintain a positive public image with the help of consultants and professionals in modern marketing relations and public relations. A few of the fashion promotional campaigns and PR techniques for architects and brands comprise functioning with the media, well-known media kits with marketed sets of advertising material like pictures and background information, scheduled mainstream press tours or desk sides, celebrity credit advancement, fashion events that include details about fashion exhibits and fashion week, and sales meetings. According to (Yu and Hu; 2020), the fashion business was fragmented before social media was developed and widely used, consisting of retail stores, aspiring stylists, and clothing designers. Directing retailers and brands have developed as an outcome of digital PR developments and the alternation between brands and newspapers (Jacobson; 2020).

2.2 Different factors related to social media platforms

Social media is assisting in the development of social fashion in the fashion business, where knowledge is used to spur creativity. Bloggers and independent stylists are acknowledged as essential components of marketing and promotion plans for businesses. Making purchases from reputable businesses when shopping online gives customers confidence. According to

(Pentina et al.; 2018), social media has introduced unique elements to the online shopping experience and may be a catalyst for growing sales by enabling businesses and sellers to show their items in front of thousands of prospective customers in a friendly environment (Sivarajah et al.; 2020). Social networking sites make it simpler to connect with the target audience and potential customers. In 2009, high-end businesses rapidly boosted their use of social networking platforms. Customers are encouraged to connect with businesses due to technology. According to (Pentina et al.; 2018), these consumer interactions advance the brand by raising awareness, participation, and engagement, which in turn improves brand memory and raises sales. Fashion businesses may engage with customers via tweets, blogs, and social media sites like Facebook, Twitter, YouTube, Instagram, and Pinterest. Although many fashion firms first thought social media would erode their relationships with customers, it is now seen as a way to create these connections and eventually reach a larger audience (Sukumar et al.; 2021).
According to (Romão et al.; 2019), famous bloggers are seen by brands as new reporters and influencers. The emergence of blogging agencies is evidence of the importance that fashion blogs are gaining. Style bloggers, formerly thought of as fashion-obsessed amateurs, have changed into trendsetters and now demand four- and five-figure sums from businesses. Fashion and lifestyle bloggers are represented by new companies like Digital Brand Architects in New York, which acts as an intermediary endorsement agreement with apparel designers, secures sponsors, and, in certain circumstances, arranges lucrative television advertising. Influential bloggers are getting to be represented by even sizable companies like Creative Artists Agency (Wu, and Song, 2019). According to (Romão et al.; 2019), despite the benefits of these business models, if businesses want to continue expanding, they must always come up with fresh approaches to luring customers, creating trustworthy connections, and promoting social participation. If businesses wish to engage with consumers through cutting-edge platforms, formats, or income sources while still fostering strong customer connections and brand loyalty, social media must be integrated into new, creative business models (Mariani and Wamba; 2020). According to (Romão et al.; 2019), luxury marketers must try to increase their presence on social media on the platforms they now use while also trying to add new channels in order to connect with new customers. Additionally, luxury marketers should think about how they may customise each social media channel to meet the demands of their crucial audience (Pedersen et al.; 2019).
According to (Athwal et al.; 2019), given the accessibility of a sizable audience on various social media sites and a change in consumer behaviour, it is not surprising that certain fashion companies have used social media as a marketing and communication medium. Fast-fashion shops like Zara and H&M, sportswear manufacturers like Nike, online-only enterprises, and even fashion SMEs are adopting social media more and more for marketing and communication (Acharya et al.; 2018). This is because social media is accessible, reasonably priced, and gives businesses the chance to interact and connect with potential customers more often. After initially shunning it because they were unclear on how the enhanced accessibility offered by social media could be compatible with the exclusive attributes associated with luxury, luxury fashion companies are increasingly communicating and interacting with their potential clients on social media. According to (Athwal et al.; 2019), by doing this, you will be able to keep current with the ever-expanding online community while also reaching the huge audience that social media platforms have to offer. Luxury apparel firms have increasingly altered their marketing approaches to make use of social media's potential. Social media, which enables both corporate and consumer participation, has significantly changed the marketing landscape. According to (Bazi et al.; 2020), fashion designers now have new options to engage with clients in a virtual setting that is accessible to companies of all sizes thanks to social media. By having allowed peer-reviewed papers, ideas, labels, internet blogs, thought leaders, and other techniques

that aren't under the control of corporations, social media has significantly shifted power from fashion companies to consumers. These techniques may have both positive and negative effects on a brand's reputation (Brydges; 2021).

2.3 Overview

The previous research conducted by (Wadhe and Suratkar; 2020) and (Chatterjee et al.; 2021) is utilized here for the most part. The findings of both investigations are summarized in Table 1, which can be found below. Although the word feature accuracy is low as a result of insufficient data cleaning and preprocessing, several classification machine learning methods are applied and compared in the two studies. These limitations may be alleviated if additional research was conducted to improve the precision of models that employ text mining and sentiment analysis.

Author	Models used	Dataset
(Wadhe and Suratkar; 2020)	SVM,Naive Bayes	Tourism Websites
(Chatterjee et al.; 2021)	SVM,XGBoost	Twitter Dataset

Table 1: Overview of Literature Review

3 Research Methodology

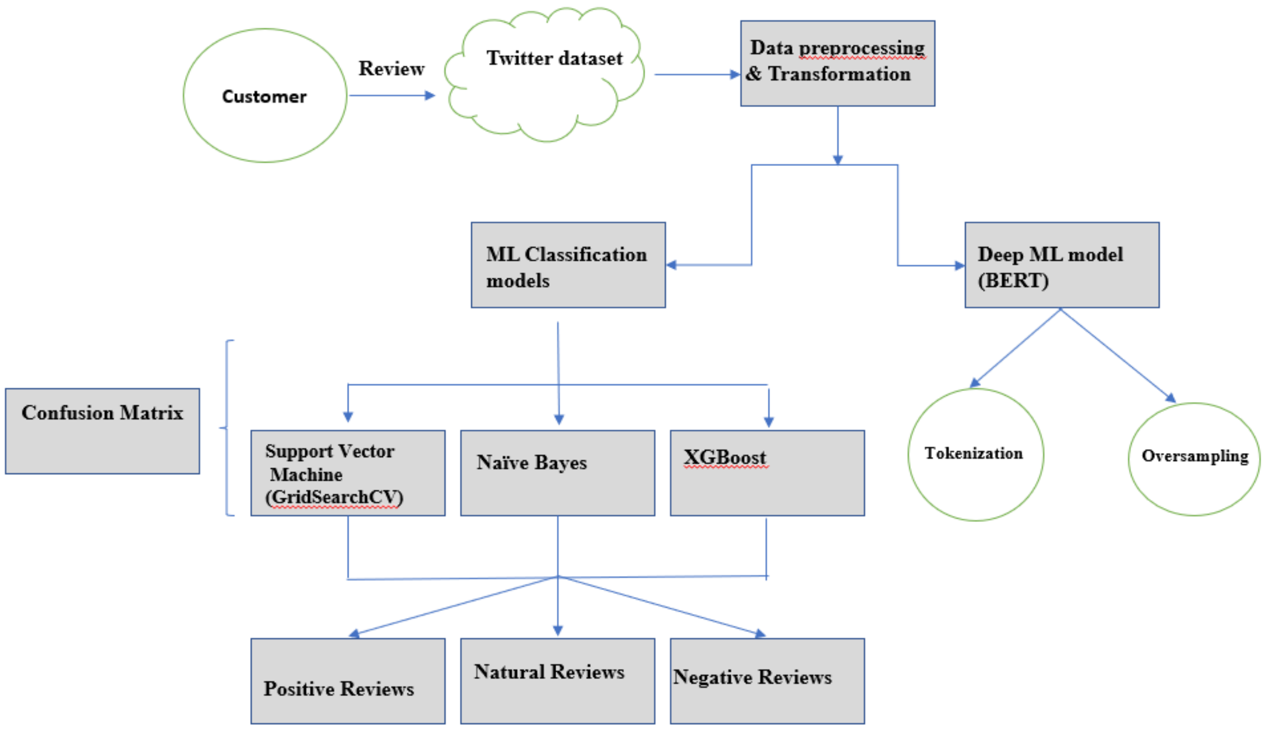

Figure 1: Flow Diagram of Research Methodology

This section of the study details the many procedures that were implemented during the study. Data is gathered from the Twitter dataset, processed (both in terms of tweets and customer reviews), classified (using various machine learning models), clustered (to create groups of reviews that are similar), and evaluated (in terms of model performance).The development of research methods is shown in Figure 1.

3.1 Data Collection using Twitter API Dataset

This study's data originated from two websites that adhere to moral and ethical criteria and are publicly accessible. The Twitter data set was collected from the authenticated Twitter[1]. Streaming API by creating a new account and logging in with Twitter-supplied consumer and API tokens. Tweepy is a Python-based library used to retrieve data and execute a number of operations. This information is supplied in JSON format.

3.2 Data Selection, Data Pre-processing and Data Transformation

Four Twitter datasets were imported using the read csv function of the Pandas package. These four tweet datasets are fashion, fashionbrand, onlineshopping, and clothingbrand. The collection includes 1000 x 10, 110 x 10, 47 x 10, and 1000 x 10 shapes. Each dataset was imported separately at first. After integrating these four datasets, you will obtain 3119 x 10. The final dataset includes followers, total tweets, retweet counts, descriptions, hashtags, and locations. Both sets of JSON data have been transformed to a structured tabular representation for preliminary processing. Missing data was erased afterward. Twitter Sentiment was built using the textblob library as a target variable. This variable has positive and negative variables. Data preparation is required before sentiment analysis. Generating Word Count Vectors, Producing the Word Bag, Determining What Percentage of Words Should Be Used, Eliminating Stop Words and Tokenizing Text, and Counting Words Used to Express Different Emotions. After importing the dataset, Natural language Toolkit(NLTK[2]) performed sentiment analysis. Various technologies were used to clean the datasets. Stop words, punctuation, HTML elements, digits, and white space have been removed. Lowercase letters and stems were added. Lemmatization and tokenization were used to modify words and sentences.

3.3 Using Document Term Matrix and Vectorization to Transform Data

Processing text data has resulted in several modifications to subsequent data. Following the tokenization procedure, stop words and punctuation were removed from the final text. The frequency-inverse document frequency feature is a method for determining the significance of each word in a corpus. Knowing the frequency with which a particular word appears in a customer review provides valuable context. Sometimes, a document term matrix is used to characterize the list of terms resulting from this operation. This crucial stage of creating a document term matrix has been completed. It displays the frequency with which each word appears in each document. Every every word is considered a feature. The training vector has a total of (3119, 7786) characteristics for the purposes of this investigation.

[1] `https://developer.twitter.com/en`

[2] NLTK is a set of Python tools and programs for symbolic and statistical natural language processing in English.

3.4 Customer review classification using XGBoost, Naive Bayes and Support Vector Machine

In order to classify the customer reviews within the dataset, the XGBoost classifier, the support vector machine, and the Naive Bayes model were all employed. The outcomes of various models have been compared to determine which model is the most trustworthy. 20% of the data was used to test these models in their actual form. In order to quickly identify positive and negative consumer behaviour, a model was trained and the data was categorized based on the ratio of positive to negative tweets that contribute to the target variable in the Twitter data set. This technique has been developed to categorize consumer input so that predictions can be made regarding the extent to which it will pique the interest of prospective buyers. Imblearn[3] was utilized to implement oversampling, which increased the accuracy of these models. The goal of this experiment was to assess whether or not prior studies conducted on the same type of paper by other researchers could be regarded valid.

3.5 Customer review classification using BERT model classifier

BERT uses natural language analysis to accomplish the phrase pair classification problem. The BERT is a highly adaptable tool for natural language processing (NLP). In this case, TensorFlow python library is used for BERT which further implements NLP. This model is cutting-edge in deep learning for NLP applications (NLP). We employ maximum series length after tokenization, tweet texts, and sentiment annotation columns. To compare machine learning model accuracy, change the network optimizer, number of epochs, batch sizes, etc.

3.6 Performance and Evaluation for Classification models

Using 20% of the data set, models have been tried out and their accuracy, precision, and recall have been measured. This was done so that the SVM, naive Bayes, and XGBoost algorithms could be judged on how well they worked. The confusion matrix has shown what happened as a result. For classifying Twitter data, SVM has shown that it is 92% accurate, multinomial Naive Bayes has shown that it is 86% accurate, and XGBoost has shown that it is 80% accurate. But in the end, when the dataset was run through a BERT classifier that had already been trained, the model's accuracy was 99%, which was the highest and most accurate of all the results.

4 Design Specification

The framework demonstrates how this research will evolve as a whole in order to predict the amount of consumer satisfaction. This is performed using clustering and classification techniques. Figure 2 presents the structure's framework, which you may view here.

[3]Imbalanced-learn (Imblearn) is a Python library that creates data sets with equal class counts.

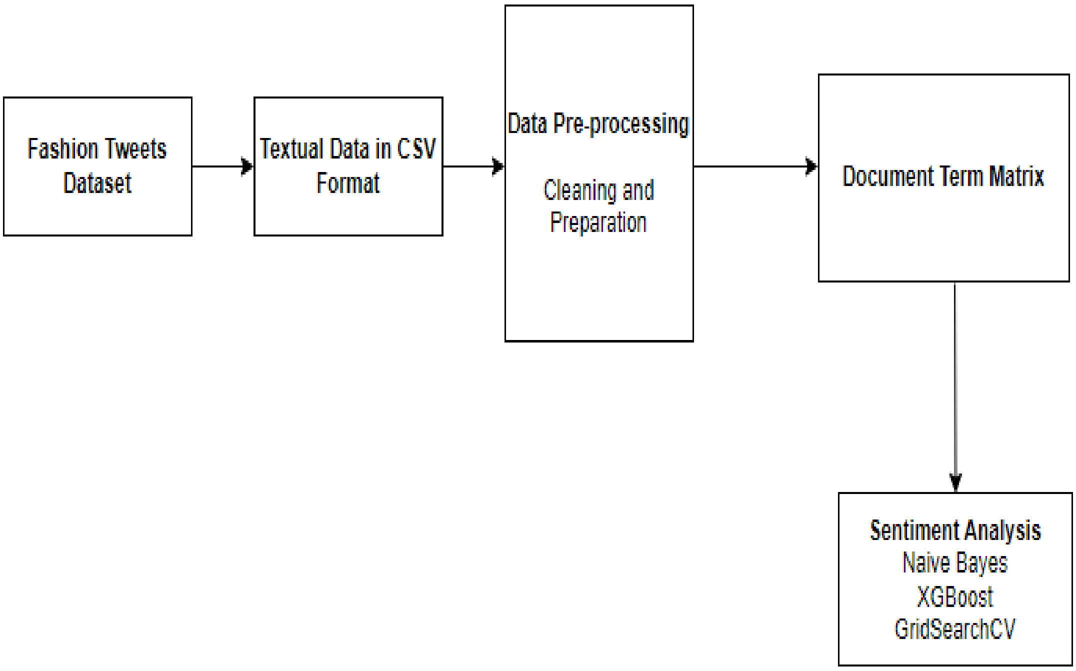

Figure 2: Framework of Fashion Tweet Dataset Classification

The first step is to create a database where a dataset full of Twitter information will reside. In the following step, we pulled the Twitter dataset from the DB and stored it in an unstructured JSON file. Also, the textual information that was previously unorganized has been transformed into a DataFrame. Followers, total tweets, retweet counts, descriptions, hashtags, and locations were only some of the data points included in the table that accompanied the dataset. The next step involves pre-processing and cleaning the dataset before transforming it into a document term matrix (DTM). In the most recent stage of research and development, sentiment analysis has been carried out by means of classification and clustering methods.

5 Implementation

Python, a general-purpose programming language, and Jupyter Notebook, a general-purpose programming environment, were used to carry out the implementation of this research. Both the Twitter dataset and the Yelp dataset have been subjected to extensive preprocessing and modification with the help of a wide variety of Python tools. The datasets from Twitter and Yelp, which had previously been stored in an unstructured JSON format, have now been converted into a structured format. In addition, the SVM, Nave Bayes, random forest, and Deep ML BERT classifier machine learning models are created with the help of data transformation techniques like document term matrices, and vectorization.

5.1 Using Document Term Matrix and Vectorization to Transform Data

After collecting the final dataset, a sentiment analysis determined the fraction of negative, positive, and neutral tweets. The word clouds for good, negative, and neutral tweets have been determined. Sentiment analysis involves tokenization, stemming, and stop word deletion. (Polignano et al.; 2019) has multiple steps. All of these steps are done sequentially. BERT is

used with TensorFlow to tokenize textual data. Following tokenization, each sentiment was classified using the tokenized sentence. The tweets' stop-words were eliminated, and the Label Encoder assigned a numerical value to their thoughts. X characteristics is countvectorizer tweets and Y is sentiments. After creating X and Y labels from the textual dataset, supervised machine learning algorithms classified tweets as positive, negative, and neutral. Nave Bayes, XGBoost, SVM, Deep ML, and BERT classification methods were employed.
Term frequency-inverse document frequency analyzes every word in a corpus for its relevance. It counted how often a word appeared in consumer feedback. A document term matrix was created to track how often terms appeared in different documents. Vectorization was done using sklearn package methods. The size of the Twitter vector used in this investigation is (3119, 7786).

5.2 XGBoost, Support Vector Machine and Naïve Bayes Classifier

Model development and application are next in the strategy's implementation. Each model uses 20% of available data for evaluation and 80% for training. In this research, we used three classifiers to categorize customer comments, learn how customers felt, and measure our success. The Naive Bayes model categorizes reviews and tweets by tone (Salmi and Rustam; 2019). It is an algorithm-driven machine learning model. Independent variable: clients' tweets. Textblob is a sentiment analysis of tweets is the study dependent variable. Second-generation models are SVMs. This is a non-human-instructed machine learning model. GridSearchCV (Zhao et al.; 2020) classifies reviews and tweets as good, negative, and neutral. Third model: XGBoost. Using supervised learning, new trees are generated to anticipate previous trees' errors and residuals. These trees are averaged for the final forecast (Jiang et al.; 2019). All these models use sklearn library.

5.3 BERT model Classifier

The Twitter dataset has seen applications of the Naive Bayes, XGBoost, and SVM machine learning models for categorization. The BERT method, on the other hand, is the most recent Deep ML technique within the NLP framework. The model has made use of 80% of the data that it was trained on and 20% of the test data. We load the BERT Classifier, Tokenizer, and Input modules after the Transformers library has been installed because the model has already been trained. The data is modified by first removing the text of emojis and hashtags, then eliminating punctuation, links, mentions, and rn characters, and then classifying the data that is left after this process. Tokenization is carried out, in addition to indexing the texts and tokenizing them individually. Oversampling is done in order to rectify the uneven data, and after that, sentiment analysis is carried out on the data that has been completely categorized. In spite of the fact that the processing time required by the model is more than that required by other models, it delivers the most accurate performance of any model.

6 Evaluation

The classification model is evaluated based on classification precision and other performance measures. The sentiment analysis performed in this work and the ratio of positive to negative tweets, and neutral words in the dataset are also investigated in Figure 3.

Figure 3: Ratio of Positive, Negative, and Neutral Words

The ratio of the positive tweets are highest in the selected tweets dataset with almost 61% of total. The lowest proportions of the negative tweets have been found and there is also significant number of neutral tweets.

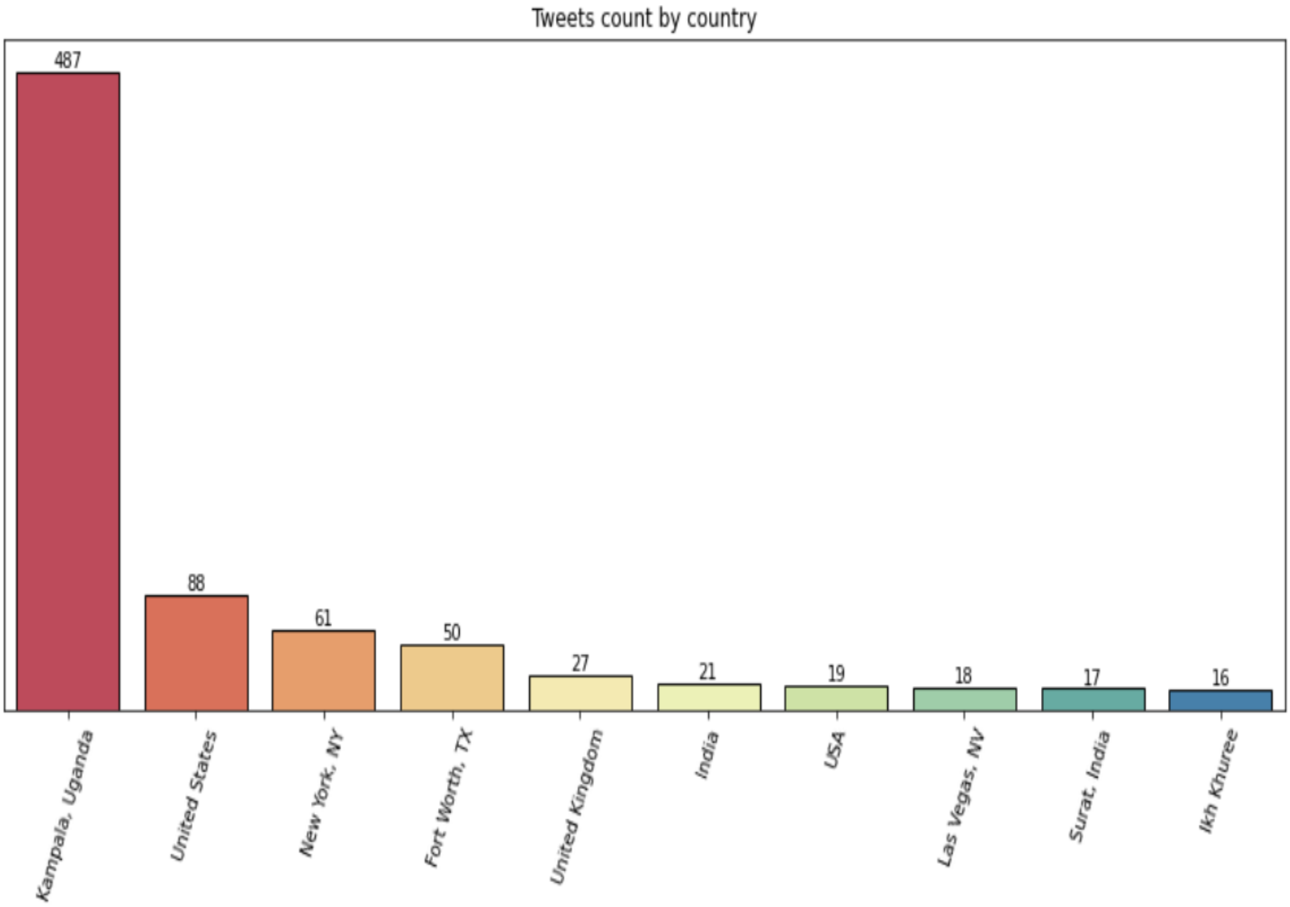

Figure 4: Top 10 Tweets Counts by Country

In Figure 4, bar plot demonstrates the tweets count based on the country. The highest tweets count is from the Kampala, Uganda and the lowest tweets counts are from the country Ikh Khuree.

6.1 Word Clouds based on Sentiments

Word clouds illustrate the specific word in a bag of words that are consequently used in the tweets. The word that is most frequently used is illustrated in word cloud is highlighted with bold and big fonts (Budiharto and Meiliana; 2018). There are a total of four word clouds have been generated. The word cloud shown below demonstrates the word cloud for all data. The most frequent words in the total word cloud are Fashion, truwears, exciting and dedicated. The word clouds for the sentiments mainly positive and negative are generated and shown below in Figure 5 and Figure 6.

Figure 5: Word Cloud for Positive Words

The terms in the dataset that indicate a positive emotion are depicted in Figure 5's word cloud. The most frequent words in the above word cloud are Fashion, truwears, exciting and clothes.

Figure 6: Word Cloud for Negative Words

In Figure 6, the word cloud for the negative sentiment the **words** that are frequently used in tweets are onlineshopping, clothingbrand, “fashion” and “amp”. There are no negative words have been recognized in the above word cloud captured for the negative sentiment.

6.2 Machine Learning Classification Models

After the study has been put into action, it is necessary to evaluate how well the models have been put into action. Within the scope of this investigation, four models were implemented. Classifiers such as Support Vector Machine, Nave Bayes, and XGBoost were utilized in each of the three experiments that were conducted in order to classify the customer evaluations for the purpose of predicting consumer satisfaction. The confusion matrix, accuracy, precision, and recall are the criteria that are utilized in their evaluation. The Deep ML classifier, BERT model is used in the fourth experiment, which utilizes the same evaluation measures as the previous three experiments.

6.2.1 Naïve Bayes Classification Model

The classification summary of the Naïve Bayes classification model is demonstrated in the below picture that includes the accuracy of sentiment classification, weighted value of recall, precision and f1 score. The sentiments are classified as Neutral(0), Positive(1) and Negative(2).

Evaluation Metrics	**Precision**	**Recall**	**F1-Score**
Neutral(0)	0.34	0.91	0.50
Positive(1)	0.97	0.76	0.85
Negative(2)	0.93	0.90	0.91
Accuracy	86%		

Table 2: Performance Metrics of Naive Bayes Model

The overall sentiment classification accuracy of the Naive Bayes classification model is 0.86. The generated classification model is capable of classifying the tweets in three categories positive, negative and neutral with 86% accuracy. The weighted value of the precision parameter is 0.97 for positive tweets that demonstrates that among 100 tweets, 97 tweets are positively classified in the class where they actually belong.
The purpose of developing this model was to replicate the results found by (Wadhe and Suratkar; 2020). It has proven both precision and accuracy at a rate of 88%. It has showed performance that is inferior to that of SVM, but superior to that of (Wadhe and Suratkar; 2020). In the experiments conducted by (Wadhe and Suratkar; 2020), there is an absence of preprocessing, as well as an improvement of unbalanced text input, both of which have been shown to boost the performance of models. Table 2 shows a performance analysis of how well Naive Bayes algorithms work.

6.2.2 SVM Classification Model

Evaluation Metrics	Precision	Recall	F1-Score
Neutral(0)	0.91	0.61	0.73
Positive(1)	0.86	0.90	0.88
Negative(2)	0.94	0.95	0.95
Accuracy	92%		

Table 3: Performance Metrics of SVM Model

Table 3 shows summary of performance metric is for the GridSearchCV that has been used for classification of the tweets. To use grid search to optimize the hyperparameters of an SVM, first specify the range of values that want to search over for each hyperparameter. For example, need to specify a range of values for the learning rate and a range of values for the regularization term. Support vector machines (SVMs) are a type of supervised learning algorithm that can be used for classification or regression tasks. The main idea behind SVMs is to find the hyperplane in a high-dimensional space that maximally separates the classes. This is done by solving a quadratic optimization problem that seeks to maximize the margin between the classes. From the trained model the best parameters for the SVM calculated are C=100, gamma=0.001, kernel=sigmoid.
The GridSearchCV classification model's total sentiment classification accuracy is 0.92. With 92% accuracy, the generated classification model is capable of classifying tweets into three categories: positive, negative, and neutral. The precision parameter has a weighted value of 0.86 for positive tweets, indicating that 86 tweets out of 100 are correctly categorised in the class to which they belong.

6.2.3 XGBoost Classification Model

Evaluation Metrics	Precision	Recall	F1-Score
Neutral(0)	0.00	0.00	0.00
Positive(1)	0.95	0.56	0.70
Negative(2)	0.77	0.98	0.86
Accuracy	80%		

Table 4: Performance Metrics of XGBoost Model

Table 4 summarizes overview of performance metrics for XGBoost, which was utilised for tweet categorization. The XGBoost classification model has a total sentiment classification accuracy of 0.80. With an accuracy of 80%, the categorization algorithm that was developed is able to place tweets into one of three categories: neutral, negative, or positive. The precision metric has a weighted value of 0.95 for positive tweets, suggesting that 95 tweets out of 100 are accurately classified in the class to which they belong.

6.2.4 BERT Classification Model

The below Table 5 demonstrates the classification summary of the performed BERT model for the sentiment classification. When it comes to positive tweets, all of the performance metrics factors, such as precision, recall, accuracy, and F1 score, have a value of 0.98. The accuracy as a whole works out to be 99%. When compared to all of the other classifiers that were utilized for the categorization of the tweets dataset, the BERT model had the highest accuracy in terms of classification techniques.

Evaluation Metrics	**Precision**	**Recall**	**F1-Score**
Neutral(0)	0.99	1.00	1.00
Positive(1)	0.98	0.98	0.98
Negative(2)	0.98	0.97	0.98
Accuracy	99%		

Table 5: Performance Summary of the BERT Classifier Model

7 Results

From the sentiment analysis the ratio of positive to negative tweets and neutral tweets have been found. The ratio of the positive tweets are highest in the selected tweets dataset. The lowest proportions of the negative tweets have been found and there is also significant number of neutral tweets. The count of tweets from different countries are also computed and it is found that the highest tweets count is from the Kampala, Uganda and the lowest tweets counts are from the country Ikh Khuree. The generated word cloud demonstrates the most frequent words in the tweet dataset are Fashion, truwears, exciting and dedicated. The first three machine learning classification algorithms are implemented for classification of sentiment in the tweet dataset. These experiments were performed to compare performance and accuracy with two recent base papers (Wadhe and Suratkar; 2020) and (Chatterjee et al.; 2021). The results were better when comapred to base papers results. The highest classification accuracy is found with BERT classifier. The lowest classification accuracy is obtained with the Naïve Bayes Classifier. To meet the limitations of both articles, the dataset was preprocessed, cleaned, and balanced with extra features that helped improve performance. But when the model's results were looked at, it was decided that Twitter data doesn't contain enough useful information for business analysis. Tweets about customer reviews are either not available or only accessible through certain hashtags. Instead, the page has many different kinds of reviews and comments. Multiple hashtags are used to get more records, and an extra Deep ML method is used to get the most accurate results. This study isn't as good as it could be because the Twitter dataset doesn't give enough information and because a lot of text data needs to be preprocessed, which takes even more time.

By putting customer ratings into groups, it's possible to predict how customers will feel about different fashion businesses and brands. Based on this research, fashion companies can make better business decisions when they know how much pleasure their customers get from their clothes. Positive and negative reviews can help you understand what your customers think, and negative reviews can be used to improve the quality of clothing in a fashion company based on what the customers want. According to the results of this argument, the study work has made a contribution to the field of data analytics by improving past experiments, combining XGboost with oversampling methods, and getting the best results by using a pre-trained Deep ML BERT model.

8 Conclusion and Future Work

The report examines the tweets dataset using the sentiment analysis in Python. The tweet dataset used for performing the sentiment analysis is fashion dataset. There is a total of four dataset have been collected and imported to Python environment for performing the sentiment analysis. The ratio of positive to negative tweets and neutral sentiment is investigated and the word clouds are generated for each sort of sentiment class. The most frequent words in the above word cloud are Fashion, truwears, exciting and dedicated. For tokenization and building the classification models for the classification of the tweets in different sentiment class, BERT architecture has been followed in this work. In this work, supervised machine learning methods have been developed. The classification algorithms used in this work are the Naïve Bayes, XGBoost, GridSearchCV and BERT Classifier. The performance metrics for each classification algorithm is interpreted in the report. The highest classification accuracy and the weighted value of precision are obtained for the BERT classifier model. The classification accuracy of the BERT model is 0.99. The created classification model is capable of classifying tweets into three categories with an accuracy of 99% percent. These categories are neutral, negative, and positive, respectively.The precision parameter has a weighted value of 0.99, indicating that 99 tweets out of 100 are correctly categorised in the class to which they belong. For the classification of the sentiment categories four machine learning algorithms have been implemented and the highest obtained classification accuracy is 0.99. The BERT classifier has shown the highest classification accuracy. Random forest and other classification algorithms are considered to provide high classification accuracy with tuning the hyper-parameters of the model. So, in future work, different other ML classification algorithms and Deep ML techniques like CNN etc can be used with tuned parameters to enhance the classification accuracy.

References

Abbas, J., Mahmood, S., Ali, H., Raza, M. A., Ali, G., Aman, J., Bano, S. and Nurunnabi, M. (2019). The effects of corporate social responsibility practices and environmental factors through a moderating role of social media marketing on sustainable performance of business firms, *Sustainability (Switzerland)* **11**.

Acharya, A., Singh, S. K., Pereira, V. and Singh, P. (2018). Big data, knowledge co-creation and decision making in fashion industry, *International Journal of Information Management* **42**.

Arrigo, E. (2018). Social media marketing in luxury brands: A systematic literature review and implications for management research, *Management Research Review* **41**.

Athwal, N., Istanbulluoglu, D. and McCormack, S. E. (2019). The allure of luxury brands' social media activities: a uses and gratifications perspective, *Information Technology and People* **32**.

Batrinca, B. and Treleaven, P. C. (2015). Social media analytics: a survey of techniques, tools and platforms, *AI and Society* **30**.

Bazi, S., Filieri, R. and Gorton, M. (2020). Customers' motivation to engage with luxury brands on social media, *Journal of Business Research* **112**.

Brydges, T. (2021). Closing the loop on take, make, waste: Investigating circular economy practices in the swedish fashion industry, *Journal of Cleaner Production* **293**.

Budiharto, W. and Meiliana, M. (2018). Prediction and analysis of indonesia presidential election from twitter using sentiment analysis, *Journal of Big Data* **5**.

Chatterjee, D. P., Mukhopadhyay, S., Goswami, S. and Panigrahi, P. K. (2021). Efficacy of oversampling over machine learning algorithms in case of sentiment analysis, Vol. 1175.

Jacobson, J. (2020). You are a brand: social media managers' personal branding and "the future audience", *Journal of Product and Brand Management* **29**.

Jiang, Y., Tong, G., Yin, H. and Xiong, N. (2019). A pedestrian detection method based on genetic algorithm for optimize xgboost training parameters, *IEEE Access* **7**.

Laurell, C., Sandström, C., Berthold, A. and Larsson, D. (2019). Exploring barriers to adoption of virtual reality through social media analytics and machine learning – an assessment of technology, network, price and trialability, *Journal of Business Research* **100**.

Lee, J. E., Hur, S. and Watkins, B. (2018). Visual communication of luxury fashion brands on social media: effects of visual complexity and brand familiarity, *Journal of Brand Management* **25**.

Mariani, M. M. and Wamba, S. F. (2020). Exploring how consumer goods companies innovate in the digital age: The role of big data analytics companies, *Journal of Business Research* **121**.

Park, M., Im, H. and Kim, H. Y. (2020). "you are too friendly!" the negative effects of social media marketing on value perceptions of luxury fashion brands, *Journal of Business Research* **117**.

Pedersen, E. R. G., Earley, R. and Andersen, K. R. (2019). From singular to plural: exploring organisational complexities and circular business model design, *Journal of Fashion Marketing and Management* **23**.

Pentina, I., Guilloux, V. and Micu, A. C. (2018). Exploring social media engagement behaviors in the context of luxury brands, *Journal of Advertising* **47**.

Polignano, M., Basile, P., de Gemmis, M., Semeraro, G. and Basile, V. (2019). Alberto: Italian bert language understanding model for nlp challenging tasks based on tweets, *CEUR Workshop Proceedings* **2481**.

Romão, M. T., Moro, S., Rita, P. and Ramos, P. (2019). Leveraging a luxury fashion brand through social media, *European Research on Management and Business Economics* **25**.

Salmi, N. and Rustam, Z. (2019). Naïve bayes classifier models for predicting the colon cancer, *IOP Conference Series: Materials Science and Engineering* **546**.

Sivarajah, U., Irani, Z., Gupta, S. and Mahroof, K. (2020). Role of big data and social media analytics for business to business sustainability: A participatory web context, *Industrial Marketing Management* **86**.

Sukumar, A., Jafari-Sadeghi, V. and Xu, Z. (2021). The influences of social media on chinese start-up stage entrepreneurship, *World Review of Entrepreneurship, Management and Sustainable Development* **17**.

Wadhe, A. A. and Suratkar, S. S. (2020). Tourist place reviews sentiment classification using machine learning techniques.

Wu, Y. and Song, D. (2019). Gratifications for social media use in entrepreneurship courses: Learners' perspective, *Frontiers in Psychology* **10**.

Yu, S. and Hu, Y. (2020). When luxury brands meet china: The effect of localized celebrity endorsements in social media marketing, *Journal of Retailing and Consumer Services* **54**.

Zhao, S., Mao, X., Lin, H., Yin, H. and Xu, P. (2020). Machine learning prediction for 50 anti-cancer food molecules from 968 anti-cancer drugs, *International Journal of Intelligence Science* **10**.