SENTIMENT ANALYSIS AND USER BEHAVIOR PREDICTION IN SOCIAL NETWORKS

LIU MINGJIE

UNIVERSITI TEKNOLOGI MALAYSIA

UNIVERSITI TEKNOLOGI MALAYSIA
DECLARATION OF Choose an item.

Author's full name	:				
Student's Matric No.	:		Academic Session	:	
Date of Birth	:		UTM Email	:	
Choose an item. Title	:	TITLE IN CAPITAL LETTERS TITLE IN CAPITAL LETTERS TITLE IN CAPITAL LETTERS			

I declare that this Choose an item. is classified as:

☒	**OPEN ACCESS**	I agree that my report to be published as a hard copy or made available through online open access.
☐	**RESTRICTED**	Contains restricted information as specified by the organization/institution where research was done. *(The library will block access for up to three (3) years)*
☐	**CONFIDENTIAL**	Contains confidential information as specified in the Official Secret Act 1972)

(If none of the options are selected, the first option will be chosen by default)

I acknowledged the intellectual property in the Choose an item. belongs to Universiti Teknologi Malaysia, and I agree to allow this to be placed in the library under the following terms :

1. This is the property of Universiti Teknologi Malaysia
2. The Library of Universiti Teknologi Malaysia has the right to make copies for the purpose of research only.
3. The Library of Universiti Teknologi Malaysia is allowed to make copies of this Choose an item. for academic exchange.

Signature of Student:

Signature :

Full Name
Date :

Approved by Supervisor(s)

Signature of Supervisor I:	Signature of Supervisor II
Full Name of Supervisor I NOOR HAZARINA HASHIM	Full Name of Supervisor II MOHD ZULI JAAFAR
Date :	Date :

NOTES : If the thesis is CONFIDENTIAL or RESTRICTED, please attach with the letter from the organization with period and reasons for confidentiality or restriction

CLASSIFICATION OF THESIS AS RESTRICTED/CONFIDENTIAL

TITLE: Click or tap here to enter text.

AUTHOR'S FULL NAME: Click or tap here to enter text.

Please be informed that the above-mentioned thesis titled __________ should be classified as RESTRICTED/CONFIDENTIAL for a period of three (3) years from the date of this letter. The reasons for this classification are

(i)

(ii)

(iii)

Thank you.

Yours sincerely,

SIGNATURE:

NAME:

ADDRESS OF SUPERVISOR:

“Choose an item. hereby declare that Choose an item. have read this Choose an item. and in Choose an item.
opinion this Choose an item. is sufficient in term of scope and quality for the award of the degree of Choose an item.”

Signature	:	___________________________________
Name of Supervisor I	:	KHAIRUR RIJAL JAMALUDIN
Date	:	9 MAY 2017

Signature	:	___________________________________
Name of Supervisor II	:	NOOR HAZARINA HASHIM
Date	:	9 MAY 2017

Signature	:	___________________________________
Name of Supervisor III	:	MOHD ZULI JAAFAR
Date	:	9 MAY 2017

Declaration of Cooperation

This is to confirm that this research has been conducted through a collaboration Click or tap here to enter text. and Click or tap here to enter text.

Certified by:

Signature :
Name :
Position :
Official Stamp
Date

* This section is to be filled up for theses with industrial collaboration

SENTIMENT ANALYSIS AND USER BEHAVIOR PREDICTION IN SOCIAL NETWORKS

LIU MINGJIE

A project report submitted in fulfilment of the
requirements for the award of the degree of
Master in Data Science

School of Education
Faculty of Computing
Universiti Teknologi Malaysia

January 15, 2025

DECLARATION

I declare that this Choose an item. entitled *"title of the thesis"* is the result of my own research except as cited in the references. The Choose an item. has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Signature	:	....
Name	:	
Date	:	10 NOVEMBER 2016

ABSTRACT

Social media platforms yield data with a heavy emphasis on emotions and user actions, which could be utilized in businesses such as marketing, e-commerce, and social analytics, but deriving insightful information is still on a low scale. The study tends to craft and offer suitable methods and models to tackle the issues that revolve around sentiment analysis and user behavior predictions in the social media context. The research will get to work with text and image data on tweets, such as text and image data. The study uses natural language processing (NLP) and machine learning techniques (ML) to build the framework of multimodal sentiment analysis to enable the integration of text and image information and improve the sentiment analysis accuracy and efficiency. We will measure, by the tested approaches, the effectiveness of modalities of images in sentiment analysis and determine the power of the properties of images with sentimental polarity as a basis of bias in image analysis. The analysis will investigate the reported differences in the emotional expressions of different brands on social media and the reasons for them - for example, features such as the brand image and respective products. This study will enable this through facilitating an avenue for precise sentiment analysis and user behavior prediction in social media, which will equip enterprises, researchers, and policymakers to determine user behavior better and develop marketing strategies that take these behaviors into account.

ABSTRAK

Platform media sosial menghasilkan data dengan penekanan berat pada emosi dan tindakan pengguna, yang boleh digunakan dalam perniagaan seperti pemasaran, e-dagang dan analisis sosial, tetapi memperoleh maklumat bernas masih pada skala rendah. Kajian ini cenderung untuk mencipta dan menawarkan kaedah dan model yang sesuai untuk menangani isu yang berkisar pada analisis sentimen dan ramalan tingkah laku pengguna dalam konteks media sosial. Penyelidikan akan berfungsi dengan data teks dan imej pada tweet, seperti data teks dan imej. Kajian ini menggunakan pemprosesan bahasa semula jadi (NLP) dan teknik pembelajaran mesin (ML) untuk membina rangka kerja analisis sentimen multimodal untuk membolehkan penyepaduan maklumat teks dan imej serta meningkatkan ketepatan dan kecekapan analisis sentimen. Kami akan mengukur, dengan pendekatan yang diuji, keberkesanan modaliti imej dalam analisis sentimen dan menentukan kuasa sifat imej dengan kekutuban sentimental sebagai asas berat sebelah dalam analisis imej. Analisis akan menyiasat perbezaan yang dilaporkan dalam ekspresi emosi jenama yang berbeza di media sosial dan sebab-sebab mereka - contohnya, ciri seperti imej jenama dan produk masing-masing. Kajian ini akan membolehkan ini melalui memudahkan jalan untuk analisis sentimen yang tepat dan ramalan tingkah laku pengguna dalam media sosial, yang akan melengkapkan perusahaan, penyelidik dan penggubal dasar untuk menentukan tingkah laku pengguna dengan lebih baik dan membangunkan strategi pemasaran yang mengambil kira tingkah laku ini.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1 Introduction

Social networks have become an example of an inherent fusion of our daily lives since they can be aligned with our routine easily and change everything about communication, information, and social interactions as we know it. This starts from Facebook to Twitter to Instagram to TikTok - now all have developed as the era of virtual spaces, where individuals talk to their friends and family, as well as sharing both new and old content, giving their thoughts. Hence, a staggering amount of data published by the users and found on the platforms has been created, leading to making it a benefit hub for a wide range of applications.

Certainly, now businesses focus on the idea that social media data is a gold mine for marketing communications, product promotion, and understanding consumer and market trends. The key value comes from how meaning can be extracted out of user behavior and sentiment; however, this is what is the seller that is able to make marketing strategies more segmented, personalized, and connected to the customer. The difficulty in accumulating meaningful analytics from the mountains of big data is a dream that is, to a large extent, not possible. The complexities, with their diversities, that lie therein due to human speech and activities, still remain very difficult to analyze or to interpret to coherent ness in the data that is user-generated. The statement of social media, including sarcasm and irony between friends and family, adds a challenge to the tests beyond what lever will be adequate. The user's behavior, on the other hand, could be explained by the fact that individual tastes, people, and other atmospheric variables collectively bring about complexity in predicting with certainty.

1.2 Background of the Problem

The massive-processing social media data includes text, pictures, and videos, and it is usually more critical to the video social networks than other sources. Sentiment analysis and pattern recognition are the two vital tasks related to many systems that require behavioral and emotional perceptions of the users. On the other hand, currently, the sentiment analysis methods mainly take into account text information, while the information of other visual altitudes, such as image and video, are ignored and may lead to not comprehensive and imprecise decisions. The integration of multimodal sentiment analysis has been one of the hotspots in the research field, but specifically the challenge the information from different modalities and to break through the heterogeneity or complexity remains the only unsolved problem.

1.3 Statement of the Problem

Disadvantages of single-modality analysis: Most papers on sentiment analysis nowadays focus on text modalities mainly and do not consider the other modalities, like images and videos, leaving the understanding of sentiments lightly and lastly. Challenges of multimodal fusion: Despite the fact that multimodal sentiment analysis has been proven to be research attractive, the way information from different modalities and overcome the heterogeneity and complexity between modalities remains a major problem that is waiting to be solved.

1.4 Research Questions

This study aspires to address the following items:

How to create the more efficient framework of multimodal sentiment analysis by using the multimodal information such as text and images and increasing the speed and accuracy of sentiment analysis?

What role do image modalities play in sentiment analysis? How can I quantify the impact of images on sentiment analysis results?

In the case of social media, do the diverse brands show different emotional expressions on it? Moreover, why do these differences happen?

1.5 Objectives of the Research

In this paper, we will create and validate a multimodal sentiment analysis framework that can effectively integrate text and image information to improve the accuracy and efficiency of sentiment analysis.

Then, the influence of image modality in sentiment analysis is quantified, and the contribution of image features to the judgment of overall emotional polarity is determined through this experimental analysis.

By analyzing the differences in the emotional expression of different brands on social media, and then exploring the reasons behind these differences, such as brand image, product features, etc.

1.6 Scope of the Study

This study will center on the following: Source: Data provided by the users on the Twitter platform, such as text and images. Research Question: The effect of image polarity on the core sentiment of the post and the differences in the emotional sharing of brands on social media. Research Methods: Exploratory data analysis, multimodal sentiment analysis model development, user behavior prediction model development, model evaluation, and feature assessment. Research Objectives: To build and examine a multitask sentiment analysis model, an experiment that will quantify the effectiveness of image modality, scans of brands' attitudes through three different brands as well as their causes.

1.7 Significance of the study

This study has important theoretical and practical significance.

Theoretical implications: This study will tap on the right research in the evolving field of multimodal sentiment analysis and provide a set of new principles for the utilization of image representation.

Practical significance: This study shall bestow the following two enterprises and scientists: Brand image monitoring: By means of emotion appreciation on social media, companies can appreciate user attitude and opinion towards the brand and then modify their marketing strategies at the right time. Product promotion: Business enterprises can, therefore, be able to analyze how different brands are being presented differently on social media as they can hence develop suitable promotion strategies.

User behavior prediction: Businesses can improve the engagement and conversion rates of their users by deploying user behavior prediction models that enable them to forecast what actions the users might take on the platform.

Personalized recommendations: Through this process of categorizing the users' emotions and behaviors, the business will be able to provide the user with a more personalized experience by offering better content to the user.

Public health management: By detecting feelings and behaviors on social sites, the researchers can address the public health issue and revolutionize in a timely way.

Social Science Research: Helping in researches on emotions and behaviors on social sites, the researchers can now study social phenomena and explore about their genesis, however, mad.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Introducing Sentiment analysis in social networks is one of the most actively investigated areas these days. Various statistical learning methods, among which Naïve Bayes and SVM, are the quintessence of machine learning, and they have been effectively used in sentiment analysis. In Aggarwal (2018), the essence of sentiment analysis would involve text mining methods, which he identified. Socher et al. (2013) demonstrated the first deep learning mechanism, which is subsequently one of the well-accepted methodologies used in this area. This section will examine the difference in opinion with the tools and approaches to sentiment analysis of the social networking space..

Lexicon-based approaches: These methods are based on predefined sentiment lexicons, which consist of sets of words or phrases with their assigned sentiment scores. These values are gradually collected according to different texts. , AFINN, and Vader are some of the existing lexicon-based approaches.

Machine learning approaches: These would require training a machine learning model on the text dataset provided with the desired sentiment. Once the model is trained, it can find those patterns and features (trait) that are telling about the sentiment of the impersonal new text. SVM, Naïve Bayes, and logistic regression are used as standard methods for sentiment analysis.

Deep learning approaches: These come with a significant advancement in the form of neural networks with multiple layers, which uses data as a material for its learning, gaining the ability to learn complex patterns creating etc. Neural networks with recurrent layers (RNNs) or long short-term memory (LSTM) networks have demonstrated their ability when talking about sentiment analysis tasks.

2.2 User Behavior Prediction in Social Networks

User Behavior Prediction in Social Networks User action forecasting is predicting what users are likely to do or choose in the future, based on their history as well as other relevant point options. In addition to these applicative purposes, user behavior information in social networks is vital to the facilitation of prompt, targeted, and highly strategic acting towards user or user groups. Bird, Klein, and Loper (2009) provide insights into natural language processing, which evaluates the most optimal and efficient relationships in outlining the user behavior to be determined. Jurafsky and Martin (2009) have introduced a broad topic of NLP, including sentiment analysis, which is crucial for predicting user behavior. Pang, Lee, and Vaidhyanathan (2002) became the later researchers in this field, who were advanced to apply the machine learning one's shelf the very first researchers to use these techniques to predict people' behaviors in social networks. In social networks, user behavior forecasting has many applications, including tweaking the point of the advertisement, the content recommendation, and the evil actions to uncover. Through this chapter, knowledge on different approaches and techniques used in making predictions on user behavior in social networks will be exposed.

Classification models: Behavioral analytics models therein look for discrete behaviors of users like hitting 'like' on a particular post or view likes; in this instance, occurrences look for underlying connections and patterns. Ordinarily, some of the most frequently used classification models include SVMs, decision trees, and k-nearest neighbors (KNN).

Regression models: User interaction forecasting typically utilizes regression models, where the goal is to be able to determine the approximate number of times a user can like a post or share over a given period. For instance, support vector regression (SVR) and linear regression are examples of regression modeling.

Sequence prediction models: The input sequences consist of the next user action to predict; because of the data dependency, the way the next action is postulated will change if the sequence is at different points. There are RNN, and LSTMs can be referred to as the examples of the sequence predictions.

2.3 Domain-Specific Sentiment Analysis

A general sentiment analysis model is of immense help, but it might fall short of the need for a more granular understanding of specific realms. Industrial sentiment analysis, as a type of specific sentiment analysis, forecasts the express emotion of a certain industry or topic within a certain context. Specific domain information extraction requires the individualized methods, which are created to adapt to specific details applying to categories of the territory under discussion. He et al. (2018) is an example of the author, who addresses the need for sentiment classification into numerous sources using domain adaptation techniques. Bamman, O’Connor, and Smith (2012) figure out how the sentiment cross-domain problems are solved in sentiment analysis cases, which is crucial for the application of domain expert cases. The research of domain-specific sentiment analysis includes studies such as Building domain-specific sentiment lexicons:Domain-specific lexicons are built using lexicons that define such terms in the context of their associated values.
Training domain-specific models: In this technique, the machine learning or deep learning systems would be trained using domain-specific datasets with similar kinds of entries to the target domain, etc., holding the understanding of how the models in the concerned field of the interest functions.
Leveraging external knowledge sources: Such sources of external domain knowledge as ontologies or databases are successfully used to make sentiment analysis more accurate and interpretable.

2.4 Cross-Cultural Sentiment Analysis

Social networks have succeeded in providing a singular platform for global interaction, where individuals from varied cultures are freely expressing their thoughts. This would be more complicated in all sentiment analysis because of the linguistic and cultural differences. Wang et al. (2016) propose validation of a cross-cultural sentiment carrying out a deep learning method, to be effective for sentiment analysis. Ageri et al. (2013) review the main characteristics of cross-lingual sentiment analysis approach, where it is important for the conducting research process to understand cultural perspective. It is no doubt that sentiment expressions and feelings vary with each culture greatly. However, cross-cultural sentiment analysis works through the perspective of the behavior of people from various cultural perspectives because.

1.The challenges springing up in the way of cross-cultural sentiment analysis are that:

(1)Language differences: Every language has a specific structure and form as well as it might have a different sense of expressing feelings/attitude.

(2)Cultural norms: The composition of a culture, including its tradition and educational practices, plays an important role in the manner in which people express and communicate their sentiment.

(3)Sarcasm and irony: Although the casual conversational wheat may be universally appreciated, because of the underlining differences in beliefs among different cultures, the same may not be correctly interpreted by such groups.

2.Research developed in the field of cross-cultural sentiment analysis involves methods including:

(1)Building cross-lingual sentiment lexicons: Exploiting lexicons for sharing sentimental terms across different languages.

(2)Training cross-cultural models: Constructing industrial models through the use of sheets of multiple languages with the aim of overtaking cultural tongues.

(3)Incorporating cultural knowledge: Most of the time, the cultural dimension is the last stop for the knowledgeable designers, anthropologists, etc; this knowledge is later harvested to further improve the accuracy and being.

2.5 Explainable AI in Sentiment Analysis and User Behavior Prediction

Machine learning models are considered as one of the popularity phrases today. And the concept of explainable AI (XAI) will be what brings, if not regulation, sensibility to this tendency. Lipton (2018) mentions the constraints of model interpretability, which has become the hot issue in the XAI discipline. Guidotti et al. (2018) offer a glimpse into a panorama of XAI strategies which can be applied to sentiment analysis and behavioral prediction. The LIME (local interpretable model-agnostic explanation) approach, introduced by Ribeiro et al. (2016), is majorly responsible for making machine learning concepts more concrete and explicable through its text-aware representations. XAI explanation evaluation factor is the model which is

transparent and interpretable so the users know how and why the models make their decision. XAI investigating for sentiment analysis and user action forecast include:
Feature importance analysis: Getting to know the main attributes existing in the dataset and contributing to the output of the model, making that intelligent is the aim of this part.
Model visualization: Illustrating some of the most significant model decisions and revealing the nature that exists between features and predictions..
Model explanation techniques: Procedures like the ones enriched by LIME (Local Interpretable Model-agnostic Explanations) or Shap (Shapley Additive explanations) may be applied to understand the workings of a model within the classification task.

2.6 Challenges and Future Directions

From what I foresee, critical obstacles and possibilities are markers for sentiment analysis and user-based behavior forecast. In their paper, "Probabilistic Models and Data Quality," Blei and cliff (2017) notice that the issue of data quality largely influences future research. Ethical concerns are posed by the Big Data that Zook (2018) highlights in his analysis. A team must address (or mitigate/deal with) them and other issues to figure out their place for improvement/refinement. although considerable strides have been made in sentiment analysis and user behavior forecasting, there persist some challenges that should thwart their effective application.These challenges involve:

Data quality: The posts made on social networks normally include a lot of noise or mistakes in them; thus, it is tough to conduct the analysis accurately.

Domain-specific issues: The projection of sentiment changes and user behavior according to the different providers and platforms, so it is necessary to develop certain particular models and algorithms for each of them.

Interpretability: The requirement of getting the motivation behind the model predictions and the logic employed is paramount for building trust and achieving the set goals.

Future research directions include:

Gaining added machine learning power: Moreover, it can enhance the power of the model by increasing the weight of factors that have been selected as influential in order to realize the final objective.

Unveiling new methods: While traditional methods often require manual annotation of large datasets, advanced matching and activism of Deep Learning and other ML-based methods provide a better alternative.

Handling ethical issues, staying data security, and privacy; these are useful points for further study. Cognitive computing: Artificial intelligence for advanced learning and adaptation to problems.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

This research chapter unfolds the methodologies applied to studying user sentiment analysis as well as behavior prediction in social networks. The research design, data sources, processing methods, and analytical tools are outlined in detail to provide a structured and reproducible framework for the study. Each step in the data science project lifecycle is discussed to ensure clarity and rigor..

3.2 Research Design

The research design entails the use of two techniques, namely explorative and hypothetical methods. In these models, data is computed and adjusted for both qualitative and quantitative results, and the sentiment analysis and behavior prediction is run simultaneously. Figure 3.1 illustrates the framework for the plan as described in this paper.

The design aims to:

1.Identify the research problem and propose possible solutions.

2. Ensure a choice of appropriate methods and tools for data acquisition and processing.

3. Assess the outcome and its implication for formulating such decision proposals.

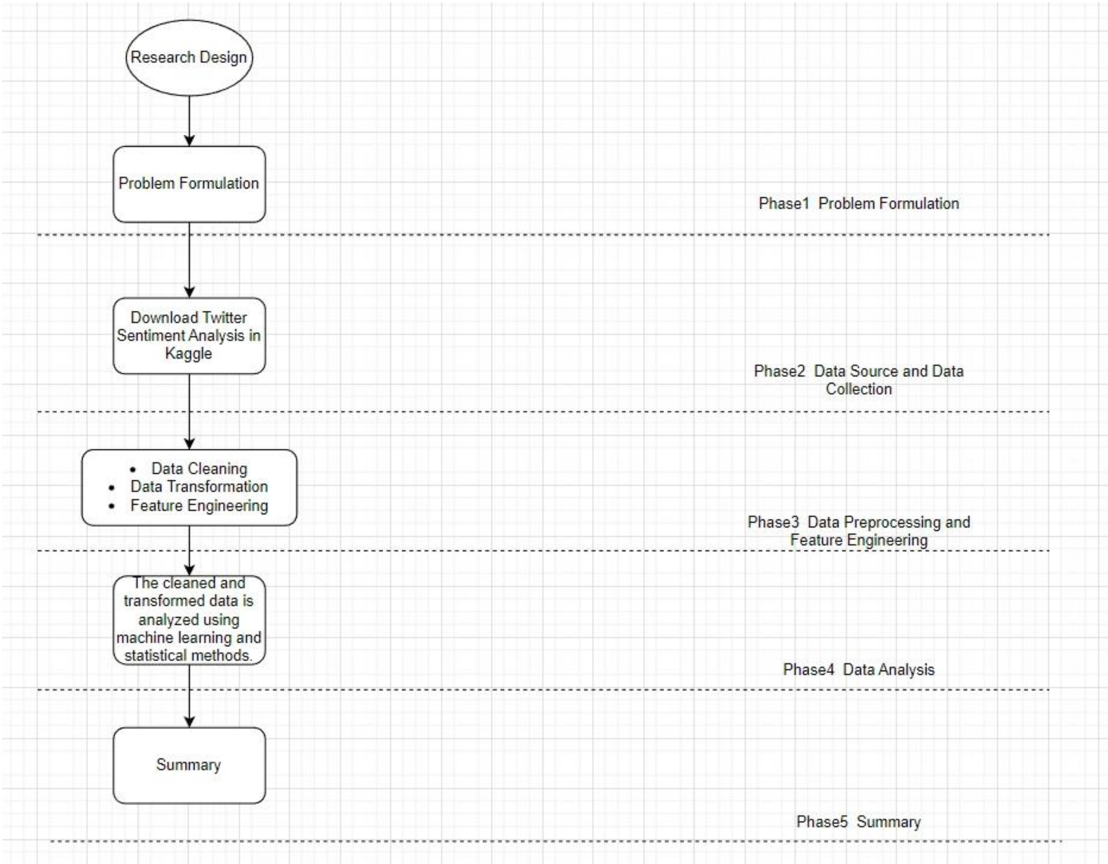

Figure 3.1: Research Framework of Proposal

3.3 Data Science Project Life Cycle

This stage of the study is based on the traditional data science project life cycle, which consists of posing the problem, data collection, preparation, analysis, as well as evaluation.

3.3.1 Problem Formulation

The book is to walk you through the most common problem in building interpretable and accurate models for sentiment analysis and user behavior forecasting based on social network data. Sub-questions include:

1. The relationship between image modality and affective polarity:
In the anote.xlsx dataset, do Tweets in the image modality show a different distribution of sentiment polarity than other modalities (e.g., plain text)?.

2. Relationship of Emotional Polarity to a Specific Brand:
Do Tweets from a particular brand (e.g., Amazon, Microsoft) show a specific trend of sentiment polarity in the twitter_validation.csv dataset?

3.3.2 Data Source and Data Collection

The data was found from Twitter and the website.

Many of these procedures include:

1. Data collection. Information in the dataset was obtained through automated scripts from Twitter. These data include tweet text, author information, and posting dates, among other details.
2. Data anonymization. Personal information of authors has been de-identified to protect their privacy. No specific user identification information is included.
3. Text cleaning. Text within the data has been processed and cleaned to remove sensitive or personal information while retaining information about topics and sentiments.

This twitter_company1 file contains Twitter data related to a particular company or brand, primarily used for sentiment analysis. The main contents and columns included are:

Twitter ID: A unique identifier for the tweet.

Content Theme: The theme of the content, which appears to be related to "Borderlands."

Tweet content: The specific text content of the tweet.

Modality: The media type of the tweet, here it is "text," indicating textual content.

Polarity: The sentiment polarity of the tweet, including "Positive" (positive sentiment).

The twitter_company dataset that has been collected is 74681 rows of data with 5 columns as shown in the Figure 3.1

```
[9]: import pandas as pd

# Load the CSV file
csv_file_path = 'twitter_company1.csv'
csv_data = pd.read_csv(csv_file_path)
csv_data
```

[9]:

	Twitter ID	Content Theme	Tweet content	modality	polarity
0	2401	Borderlands	I am coming to the borders and I will kill you...	text	Positive
1	2401	Borderlands	im getting on borderlands and i will kill you ...	text	Positive
2	2401	Borderlands	im coming on borderlands and i will murder you...	text	Positive
3	2401	Borderlands	im getting on borderlands 2 and i will murder ...	text	Positive
4	2401	Borderlands	im getting into borderlands and i can murder y...	text	Positive
...	...	...	...	...	...
74676	9200	Nvidia	Just realized that the Windows partition of my...	text	Positive
74677	9200	Nvidia	Just realized that my Mac window partition is ...	text	Positive
74678	9200	Nvidia	Just realized the windows partition of my Mac ...	text	Positive
74679	9200	Nvidia	Just realized between the windows partition of...	text	Positive
74680	9200	Nvidia	Just like the windows partition of my Mac is l...	text	Positive

74681 rows × 5 columns

Figure 3.1 The twitter_company Dataset Preview

This anote file contains sentiment analysis of Twitter data that includes images and text. The main contents and columns included are:

Twitter ID: A unique identifier for the tweet.

Tweet content: The specific text content of the tweet.

Image: A link to the image attached to the tweet.

Label the user ID: Labeling categories are divided into two categories according to polarity and emotion. The polarity is marked as positive, neutral, and negative. Choose 1 from 3, and the mood is marked as happy Joy, sad Sad, afraid of fear, angry, surprised, disgusted, and confused Confused’s 16 multiple choices.Both datasets involve sentiment analysis, but the CSV file focuses mainly on textual content, while the Excel file.

Modality: The media type of the tweet, shown as “image,” indicating the tweet includes an image.

Polarity: The sentiment polarity of the tweet, including “negative” (negative sentiment), “positive” (positive sentiment), and “neutral” (neutral sentiment).

Emotion: Emotion, containing specific emotional tags such as “anger” (anger), “joy” (joy), “anyway” (anyway), and “confused” (confused).

This dataset provides sentiment analysis of tweets with images, helping to understand the emotional impact and expression of the tweet content and its accompanying images.The twitter_company dataset that has been collected is 9662 rows of data with 7 columns as shown in the Figure 3.2

•[14]:
```
import pandas as pd
# Load the CSV file
csv_file_path = 'anote.csv'
csv_data = pd.read_csv(csv_file_path, encoding='utf-8', encoding_errors='ignore')
csv_data
```

[14]:

	Twitter ID	Tweet content	Image	Label the user ID	modality	polarity	emotion
0	1483637809449779200	kate and toby's smoker heating up to destroy t...	http://10.112.67.227:8666/img/download/1483637...	3	image	negative	["anger"]
1	1498163459749715973	@VIVIZ_staff All the best sinb, eunha, umji 🟡...	http://10.112.67.227:8666/img/download/1498163...	3	image	positive	["joy"]
2	1499194611511791617	does your heart ever go :\n\n ♡ ♡ ...	http://10.112.67.227:8666/img/download/1499194...	3	image	neutral	["anyway"]
3	1499196380400787460	@MrNiceGuy513 @Seedalicious @Great_Katzby @POC...	http://10.112.67.227:8666/img/download/1499196...	3	image	neutral	["confused"]
4	1499212575128670209	"Dont stare at your dmen after every goal agai...	http://10.112.67.227:8666/img/download/1499212...	3	image	negative	["anger","surprise"]
...	...	...	...	...	...	...	...
9657	1506815587749613571	They already got my blood pressure up #Snowfal...	http://10.112.67.227:8666/img/download/1506815...	1	synthesis	negative	["fear"]
9658	1506815644934848513	#SnowfallFX This literally went from bad to wo...	http://10.112.67.227:8666/img/download/1506815...	1	synthesis	negative	["disgust","anyway"]
9659	1506815765319766018	They got out the cage and now they are in anot...	http://10.112.67.227:8666/img/download/1506815...	1	synthesis	negative	["anyway"]
9660	1506815793039921156	Hear me out - if Heather was to be honest and ...	http://10.112.67.227:8666/img/download/1506815...	1	synthesis	positive	["proud"]
9661	1506815851781050373	This tiger shit ... #SnowfallFX https://t.co/swl...	http://10.112.67.227:8666/img/download/1506815...	1	synthesis	negative	["disgust","anyway"]

9662 rows × 7 columns

Figure 3.2 Anote Dataset Preview

3.3.3 Data Pre-Processin

1 Data Cleaning:

Cleaning up the missing and tinted data.

Several data types like types, punctuation, and some symbols can introduce noise. Filters can help remove such noise.

Text Normalization, which includes the transformation of text into some format, e.g. Case Normalization, Lemmatization.

2 Data Transformation:

Word tokens and phrases can be used as a text.

Categorical variables are usually encoded with two possible methods (verbose description goes here). Word embeddings, such as Word2Vec or TF-IDF, can be used for this conversion.

3 Feature Engineering:

Extract of linguistic features that gives the idea about sentiment score as well as n-gram.

The second thing is the identification of features that depend on the behavior of users, e.g. interaction patterns and levels of activity.

The last thing is developing some domain-specific features because of the research context which we have been talking about.

3.4 Data Analysis

The data is pre-processed and analyzed by using statistical and machine learning techniques. Key steps include:

Model Training: The application of models for classification, regression, and sequence projections.

Model Evaluation: Considers as performance metrics accuracy, precision, recall, F1, Mean Squared.

Cross-Validation: The establishment of a train/test dataset splits helps in ensuring a generalizability of models.

Exploratory Data Analysis (EDA): Generating data plots that illustrate the way data is distributed, correlate between each other, as well as trends, with the purpose of accumulating knowledge.

3.5 Summary

his chapter explicitly explained in detail the research approach and methodology regarding the problem statement, the primary data collection, and preprocessing the data. It indicated the application of a clear process of getting ready the data for sentiment analysis and human user behavior prediction. The detailed data processing steps, consequently, provide a thorough and accurate investigation, the end product being new information and novel perspectives emerging from the subsequent chapters.

CHAPTER 4

EDA/INITIAL RESULTS

4.1 Introduction

According to experts, social media data has turned into a rich mine of data, which reveals customer sentiment as well as behavior. The purpose of this chapter is geared towards bringing into focus the models building technique that is both accurate to predictions and comprehensive in results on sentiment analysis and user behavior prediction. Firstly, the study will be focused on two pairs of problems: the dependence of image modality on emotional polarity and the link that connects emotional polarity to particular brands. The answers to these two questions are the two essential components that form part of the qualitative and quantitative parameters of user behavior that are critical in developing the ultimate social media strategy.

4.2 Exploratory Data Analysis (EDA)

As an initial step in the exploration of the data, EDA was the technique used. The end goal of Exploratory Data Analysis (EDA) is to produce statistical graphics that provide intuitive ability to describe the key population characteristics, anomalies, patterns underlying the mass of the data, which will guide subsequent modeling and hypothesis testing.

4.2.1 Dataset Description

1.The relationship between image modality and affective polarity

An Emotion dataset: This archive consists of a vast variety of Tweets, with different modalities, such as images and text. Also, the polarity labels are declared, indicating the sentiment of the given Tweet. With the assistance of this dataset, analyzing the impact of the different modes of visual ways on emotional polarity would be relayed out to us. There is a gamut of datasets, as we sample them from various sources in our efforts to experiment and verify our hypotheses.

1	Twitter ID	Tweet content	Image	Label the	modality	polarity	emotion
2	14836378094497	kate and toby's smoker heating up to destroy their marriage #ThisIsUs https://t.co/C2W	http://10.112.67.227:8666/img/download/1	3	image	negative	["anger"]
3	14981634597497	@VIVIZ_staff All the best sinb, eunha, umji #VIVIZ #Queendom2 https://t.co/jEbTMa	http://10.112.67.227:8666/img/download/1	3	image	positive	["joy"]
4	14991946115117	does your heart ever go : ♡ ♡ ♡ ♡ ♡ ♡ ♡ ♡ joshu	http://10.112.67.227:8666/img/download/1	3	image	neutral	["anyway"]
5	14991963804007	@MrNiceGuy513 @Seedalicious @Great_Katzby @POCculture @GailSimone @SuperSui	http://10.112.67.227:8666/img/download/1	3	image	neutral	["confused"]
6	14992125751286	“Dont stare at your dmen after every goal against”Mrazek: https://t.co/JvKKa4WpBE	http://10.112.67.227:8666/img/download/1	3	image	negative	["anger","surprise"]
7	14992153632209	going back and forth on Campbell and Mrazek https://t.co/uBpCZ6NZnJ	http://10.112.67.227:8666/img/download/1	3	image	negative	["anger","disgust"]
8	14997677223554	@LangmanVince Awww. U sad unemployment is 3.8% and almost 700k jobs added. http	419b-9234-8352cc5fb406.gif	3	image	neutral	["confused"]
9	14998310241419	Our Little Cat 😻❤#SUGADAY https://t.co/s4dcnZFOB1	http://10.112.67.227:8666/img/download/1	3	image	negative	["fear","disgust"]
10	14999185780148	@DrOz 7.4 million jobs created and 3.8 unemployment rate. That’s a comeback https://t	http://10.112.67.227:8666/img/download/1	3	image	negative	["malice"]
11	14999288962417	@donwinslow @RonnaRono Republicans woke up EXTRA early this morning, eagerly ant	http://10.112.67.227:8666/img/download/1	3	image	negative	["sadness"]
12	15000844590473	@Joshua_M_Hump I do wanna https://t.co/EZNvsiN2wp	http://10.112.67.227:8666/img/download/1	3	image	neutral	["neutral"]
13	15002227642293	@KEEMSTAR Anthony Joshua watching the fight https://t.co/XLJ4453UGr	http://10.112.67.227:8666/img/download/1	3	image	positive	["joy"]
14	15002333873611	@PierrePoilievre You included Charest in this tweet? Are you getting paranoid Skippy Pol	http://10.112.67.227:8666/img/download/1	3	image	positive	["joy"]
15	15005280524172	Ella really had the AUDICITY to start crying and say “I feel really bad now” after telling Jo	http://10.112.67.227:8666/img/download/1	3	image	negative	["fear","disgust"]
16	15005376469505	#DeFi created for the people, is it already Time for communities to drive DeFi in every po	http://10.112.67.227:8666/img/download/1	3	image	neutral	["neutral"]
17	15006521958045	Yo! Mike’s friend is the fucking man. We all need friends who is blunt and honest. #90Da	http://10.112.67.227:8666/img/download/1	3	image	positive	["joy"]
18	15006552757322	Buying the Lakers for $67 million in the 80s. Probably seemed steep back then, but what	http://10.112.67.227:8666/img/download/1	3	image	neutral	["neutral"]
19	15006628280254	No Ben, the person you are speaking with does not a relationship with you. She wanted	http://10.112.67.227:8666/img/download/1	3	image	negative	["fear"]
20	15006633368196	I’m open to reviewing #WinningTime — figured I just put it out there. https://t.co/GhjFD	http://10.112.67.227:8666/img/download/1	3	image	positive	["joy"]
21	15006652424472	Lord pls don’t let me be 50+ crying over a wannabe rapper on TV. #90dayfiancebeforeth	http://10.112.67.227:8666/img/download/1	3	image	neutral	["neutral"]
22	15006656707358	Can someone explain to me how Michigan is the country and why Norm Nixon, a man fr	http://10.112.67.227:8666/img/download/1	3	image	negative	["confused"]
23	15006669076193	Jerry West every time he's on camera #WinningTime https://t.co/h6voxz6Ird	http://10.112.67.227:8666/img/download/1	3	image	negative	["sadness","anger"]
24	15006682259311	Ximena's face when she realized the gravy train may be over #90dayfiancebeforethe90d	http://10.112.67.227:8666/img/download/1	3	image	negative	["anger","surprise",
25	15006683373738	Jerry West in every shot #WinningTime https://t.co/vWQ8Web8Vz	http://10.112.67.227:8666/img/download/1	3	image	negative	["anger","surprise"]
26	15006685488632	Ximena: Mike you’re disgusting, I’m not in love with you, stay away from meAlso Ximena	http://10.112.67.227:8666/img/download/1	3	image	neutral	["confused"]
27	15006704175973	Don't even feel sorry for delusional women like Kimberly anymore. Look in the damn mir	http://10.112.67.227:8666/img/download/1	3	image	neutral	["surprise"]
28	15006721941578	Jerry West whenever he saw his Finals MVP trophy 😂#WinningTime https://t.co/VpF2Y	http://10.112.67.227:8666/img/download/1	3	image	positive	["joy"]
29	15006731181546	Me after watching #WinningTime https://t.co/lvKpC0rHmN	http://10.112.67.227:8666/img/download/1	3	image	neutral	["surprise"]

Figure 4.1 Anote Dataset

Figure 4.1 Anote Dataset is the first 28 rows of data in the Anote dataset

Furthermore, we have a pre-processing phase when analyzing data that includes the mentioned steps. Delete the image-containing and text-containing Tweets for the ease of specific analysis. Clean data to eliminate noise and other irrelevant information for good quality of the sanitized data. The polarity of emotions is then granted to indulge in tweets for sake of effects and the remaining part of the analysis.

Here are the following methods which are used:

(1)How a chi-square test would be applied to compare the distribution of sentiment polarity of images as opposed to plain text and show the results.

a) Construct the Contingency Table

Construct the Contingency Table First, form a frequency table according to the sentiment polarity. In particular, let M indicate Modalities (Image, Composite, Text) and let S represent Sentiments (Negative, Neutral, Positive):

Table 4.1 Sentiment analysis data based on different modalities

	Negative	neutral	positive	Row Total
image	1801	1666	1326	4793
synthesis	1301	250	752	2303
text	897	609	1060	2566
Column Total	4000	2525	3138	10663

b)Calculate the Expected Frequencies

Under the assumption of independence, we expect to observe frequency values in each cell. The formula below is for the expected frequency:

$$E_{ij} = \frac{(Row_i\ Total) \times (Column_j\ Total)}{Total\ Sample\ Size}$$

c)Calculate the Chi-Square Statistic

Calculate the observed and expected frequencies which are used in the calculations of Chi-Square. The formula is:

$$\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

In this formula, O_{ij} represents the observed frequency and E_{ij} represents the expected frequency. We do the calculations for each piece of data and then add up all the results.

d)As can be seen from the results of the calculation, the chi-square statistic is 613.04 and the p-value is 2.34×10^-131. This illustrates that, assuming that modal polarity and affective polarity are independent, the likelihood of observing a contingency table structure than this or more outrageous is very low. So, we reject the null hypothesis that there are significant differences in the distribution of sentiment polarity between tweets representing different modalities.

(2)Creating bar charts to bring visualization to results while helping to the readers process information faster.

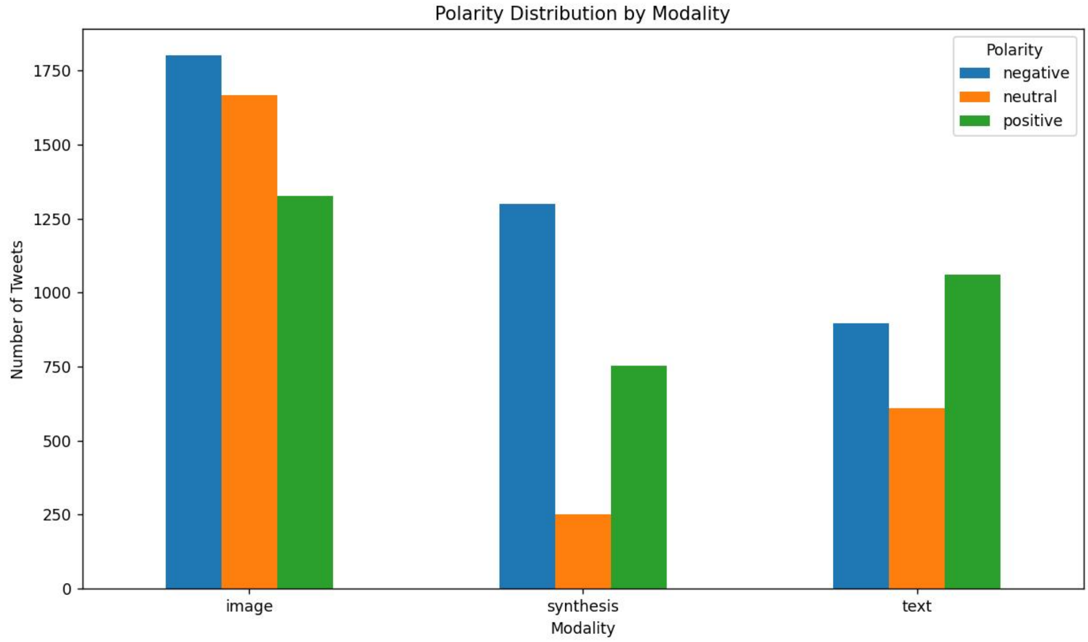

Figure 4.2 Polarity Distribution by Modality

Figure 4.2: Polarity Distribution by Modality shows the spread of emotional polarity by mode (image, composite, text) varies. While the number of tweets with negative sentiment in the imagery form is top with that of text form coming next, this is self-evident from this trend graph. It is the synthetic modality that scores in the top position in terms of the neutral sentiment tweets. The trends of positive and negative tweets across the modalities show that, except for the synthetic modality, the positive sentiment(omit) numbers are higher when compared to the negative sentiment numbers across the twitter feed. This conclusion can be drawn that people prefer to express positivity in visuals and text.

(3)Evidences from analysis suggests that, actually, there is significant difference in sentiments of emotion between image and plain-text styled tweets. Classic text tweets, for example, show a slightly higher positive mood, which confirms our assumption of local insights. Visual representation also demonstrates this very point, capturing the unique role played by visuals in helping us to express feelings. Such reports add evidence on the significance of image content in social media sentiment analysis through experimentalist inference. Picture serves as a particular brand feature which is associated with either positive or negative emotions of customers as well.

2. Emotional polarity relates to a particular brand

Twitter_company1 dataset: This data offers Tweets from several brands, where we will hold a sentiment analysis based on social media data analysis. While this data enables us to gain real-life market orientation, it contributes to the progress of the analysis and the model.

Twitter II	Content Theme	Tweet content	modality	polarity
2401	Borderlands	I am coming to the borders and I will kill you all,	text	Positive
2401	Borderlands	im getting on borderlands and i will kill you all,	text	Positive
2401	Borderlands	im coming on borderlands and i will murder you all,	text	Positive
2401	Borderlands	im getting on borderlands 2 and i will murder you me all,	text	Positive
2401	Borderlands	im getting into borderlands and i can murder you all,	text	Positive
2402	Borderlands	So I spent a few hours making something for fun. . . If you don't know I am a HUGE @Borderl	text	Positive
2402	Borderlands	So I spent a couple of hours doing something for fun... If you don't know that I'm a huge @	text	Positive
2402	Borderlands	So I spent a few hours doing something for fun... If you don't know I'm a HUGE @ Borderland	text	Positive
2402	Borderlands	So I spent a few hours making something for fun. . . If you don't know I am a HUGE Rhandler	text	Positive
2402	Borderlands	2010 So I spent a few hours making something for fun. . . If you don't know I am a HUGE Rha	text	Positive
2402	Borderlands	was	text	Positive
2403	Borderlands	Rock-Hard La Varlope, RARE & POWERFUL, HANDSOME JACKPOT, Borderlands 3 (Xbox) dlvr.it/RMTrg	text	Neutral
2403	Borderlands	Rock-Hard La Varlope, RARE & POWERFUL, HANDSOME JACKPOT, Borderlands 3 (Xbox) dlvr.it / RMT	text	Neutral
2403	Borderlands	Rock-Hard La Varlope, RARE & POWERFUL, HANDSOME JACKPOT, Borderlands 3 (Xbox) dfr.it / RMTr	text	Neutral
2403	Borderlands	Rock-Hard La Vita, RARE BUT POWERFUL, HANDSOME JACKPOT, Borderlands 1 (Xbox) dlvr.it/RMTrgF	text	Neutral
2403	Borderlands	Live Rock - Hard music La la Varlope, RARE & the POWERFUL, Live HANDSOME i JACKPOT, Borderl	text	Neutral
2403	Borderlands	I-Hard like me, RARE LONDON DE, HANDSOME 2011, Borderlands 3 (Xbox) dlvr.it/RMTrgF	text	Neutral
2404	Borderlands	that was the first borderlands session in a long time where i actually had a really satisfy	text	Positive
2404	Borderlands	this was the first Borderlands session in a long time where i actually had a really satisfy	text	Positive
2404	Borderlands	that was the first borderlands session in a long time where i actually had a really satisfy	text	Positive
2404	Borderlands	that was the first borderlands session in a long time where i actually enjoyed a really sat	text	Positive
2404	Borderlands	that I was the first real borderlands session in a nice long wait time where i actually had	text	Positive
2404	Borderlands	that was the first borderlands session in a hot row where i actually had a really bad comba	text	Positive
2405	Borderlands	the biggest dissappoinment in my life came out a year ago fuck borderlands 3	text	Negative
2405	Borderlands	The biggest disappointment of my life came a year ago.	text	Negative
2405	Borderlands	The biggest disappointment of my life came a year ago.	text	Negative
2405	Borderlands	the biggest dissappoinment in my life coming out a year ago fuck borderlands 3	text	Negative
2405	Borderlands	For the biggest male dissappoinment in my life came hanging out a year time ago fuck border	text	Negative

Figure 4.3 Twitter_company1 Dataset

Figure 4.3 Twitter_company1 Dataset is a few dozen rows of data

Data preprocessing

The data preprocessing steps include:

Pick out my brand's tweets from the included datasets (Amazon, Microsoft, Borderlands, Google) to ensure the focus of my analysis.

The following analysis methods have been used:

(1) Sentiment Score Analysis

Sentiment Score Analysis In order to calculate the sentiment score for each Tweet, we determine an emotion score for each one so that the tweet contains a sentiment score that is included in the dataset.

This average sentiment score that will be computed as well as standard deviation in order to measure the sentiment polarity of brand tweets.

a)First, we should denote a numerical value that will correspond to an affective polarity. Positive:1, 0 neutral, -1 negative We then chose four brand. We have chosen

Amazon, Microsoft, Borderlands, and Google. Calculate the average sentiment score, as well as the standard deviation for each brand. Based on the values that have been calculated above, create a box plot, as in Figure 4-4 below:

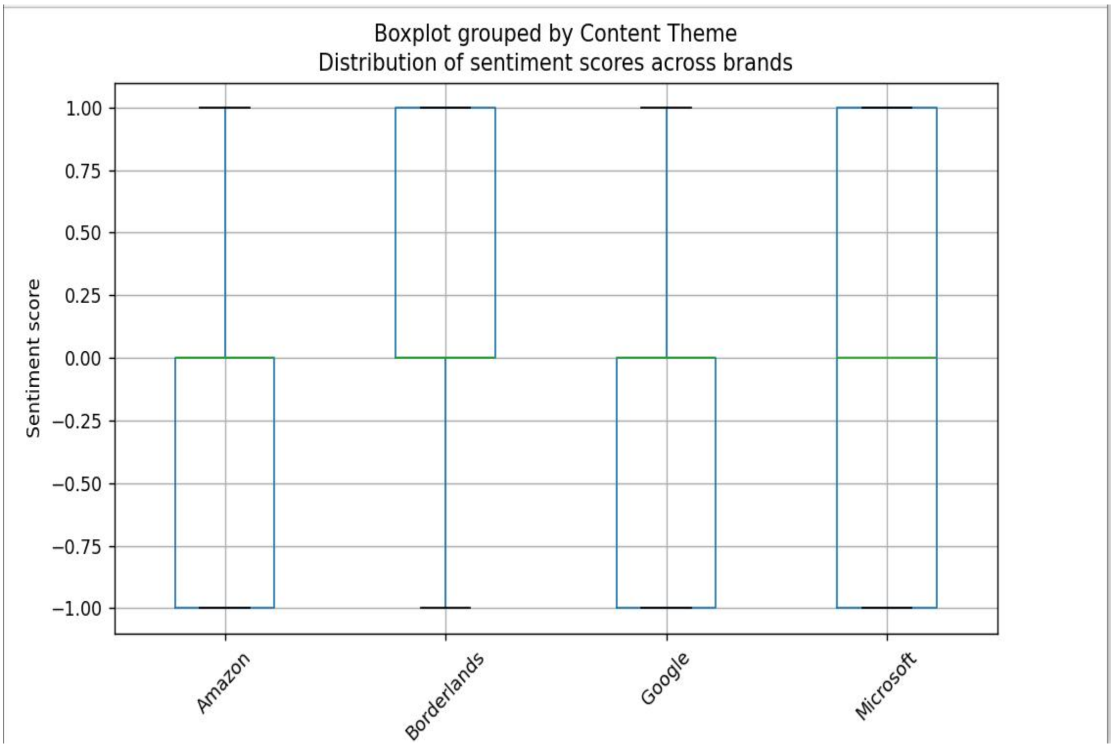

Figure 4.4 Distribution of sentiment scores across brands

b)Brand’ s summed-up sentiment score is a quick indicator of the sentiment that is inherited by the user whenever he or she interacts with it on social media. As exhibited by Borderlands' sentiment score of positive affirmation, users mostly have good opinions on Borderlands; while Amazon and Google's negative affirmation score shows that it's not populated with users, having the number of negative opinions. Microsoft’ s overall sentiment tends to be close to neutral yet slightly to the lower side of neutral, which means more branching into the negative direction compared to the other brands. Here, the standard deviation of sentiment reflects fast varying of user sentiment score, which is stability of the emotion or in other words. Standard deviation of cyclicality is low for Amazon and Google, signaling that consumed emotional bias toward these brands is more stable. Such variations largely due to their relative young age as well as business is affected only by major product launches and marketing push versus the established players whose sentiment variabilities are less because, in contrast, they are affected by small-scale local news as well. These readings not only show to what extent positive or ^negative emotions are attributed to the brand’ s image on social media but also serve as an avenue to explore the essential inputs which would be used for further hypothesis testing on future customer behavior predictions.

(2)Emotional polarity distribution analysis

Descriptive Statistics: By calculating the frequency and percentage of positive, negative, neutral tweets for a particular brand.

Visualization: Use bar charts to show the distribution of different sentiment polarities.

a)The number and percentage of sentiment polarity (being positive, negative, neutral, irrelevant) inside brand Tweets were also found. The sentiment polarity is used to know the frequency and the nature of sentiment of the peoples towards the product. The following is a bar chart of the emotional polarity distribution of each brand in every hour of a day at the busiest traffic time of the day In Figure 4-5:

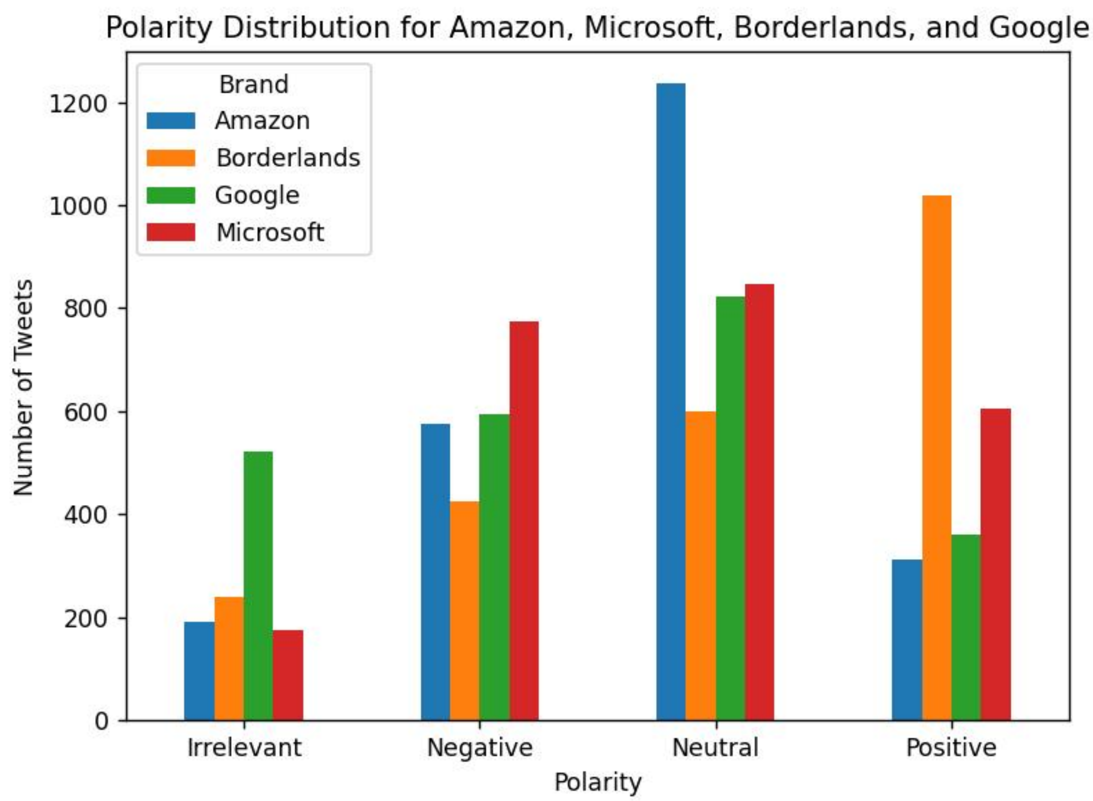

Figure 4.5 Polarity Distribution for Amazon, Microsoft, Borderlands, and Google

From the data in Figure 4-5, the graph illustrates the polarization of emotions between brands on social media, which makes it easier for us to see the difference. Among them, Borderland has the highest proportion of positive emotional polarity, indicating that users have a very high favorability towards it. However, Google's negative sentiment polarity ratio is higher than that of Borderland, indicating that users have more negative perceptions of them. Microsoft's sentiment polarity distribution is more balanced, but negative sentiment polarity still accounts for a large percentage.

4.3 Sentiment Analysis and User Behavior Prediction

4.3.1 Research Hypothesis

The sentiment analysis indicated the effect of sentiment on this behavior, and thus a significant correlation was found between sentiment analysis and user behavior. More concretely, we expect widespread fun spreading to positively challenge users to be more social through more comments, likes, and retweeting. At the same time, the ones without a positive emotion can hinder the social democracy operations by, in essence, churning good customers.

4.3.2 Feature Engineering

For the results of sentiment analysis, I built some features for user behavior prediction. These include the following:Average sentiment score, Standard deviation of sentiment scores, Frequency and percentage of different affective polarities (positive, negative, neutral, uncorrelated).

4.3.3 Predictive Models

We utilized the models that include (tuned hyperparameters for) logistic regression, random forests, and other ML techniques in their development (e.g., logistic regression, random forests) for the prediction of user behavior. Model variables are the key inputs required.

4.4 Model Development

Argentina Modeling Dependent on the results of the research hypothesis and the characteristics constructed from sentiment analysis, we will proceed with the creation of forecasts models for user behavior. Next, we are going to explain the next steps in more detail:

Model Selection: Our aim is to investigate multiple machine learning models that are knowledgeable and capable of predicting human online behavior, such as Logistic Regression: It is one of the linear models we utilized. As well, it is a model used for binary classifications making sure the predictions are binary. For instance, could the user retweet our tweet or don′t? Random Forests: It will involve the joint cross-learning of multiple decision trees to raise the accuracy and robustness of a model.

Support Vector Machines (SVM): This is a comprehensive classification algorithm with the facility to solve complex relationships and deal with non-linear data.

Deep Learning Models: Curved neural networks lengthen to the types of structure of recurrent neural networks (RNNs) or convolutional neural networks (CNNs), which capture sophisticated readers′ mood for text figuring out and user data. The model is hosted on top of those fine-tuned data.

Model Training: We will create rehearsal and verification datasets. The subject data should be run on the training group and latterly on the verification set for controlling.

Hyperparameter Tuning: It is the task of setting the hyperparameters of the model properly which significantly affects the efficacy or the opposite. It′s here we perform a hyperparameter tuning as to find the optimum settings that will lead to the best model performance.

Model Validation: Check out the trained model′s performance using various metrics, such as accuracy, precision, recall, F1, ROC AUC, among others. Thereby, it helps us to select a suitable model from other existing models to be used in respect to the task.

Model Interpretability: To maintain the applicability of our models, the technologies, for instance, Shap values or feature importance matrix could be utilized to figure out their influence on the model′s estimates.

4.5 Model Evaluation

All the developed models evaluation is crucial when it comes to conferring the same and generalizability of the models. There will be a use of different evaluation measures depending on posing the behavioral phenomenon.

Binary Classification: If the task of the retweeting or the liking is to be predicted, then the precision will be the indicators selected.

Multi-Class Classification: When complex issues with multiple groups of behavior are treated, we will also look at frequency rates as well, in addition to accuracy, precision, recall, and F1-score and confusion matrix.

Regression: Metrics like MSE, MAE, and explained variation (R^2) are judged to predict the number of retweets or likes over the retweets or likes. Moreover, we will also add cross-validation to counter the risk of the models overfitting and check their robustness to avoid overfitting. Varca is the method of splitting the data into multiple subsets and testing it on different combinations of these subsets.

4.6 Expected Outcome

Knowing importantly the different patterns and trends that get represented as user's feelings and emotions during social networks. However, some linguistic features, user activity, and user metadata are the primary elements shaping such users′ behavioral activity. Design different predictive models that accurately bring the impression of the sentiment and behavior of the user, as well as they are very reliable. Data from around the social networks can be exploited to a better understanding and informing the social networks, thus requiring many of the available applications in marketing, public health management, and social science research.

4.7 Summary

In the investigation in this chapter, the relations between image modality and affective polarity as well as the brand affect polarity are explored. These discoveries let us expand our knowledge horizon about social media data analysis, but they also provide guiding lights for future research. Even though with few restrictive points, the present research paper has given an outline of what to do to filter valuable feeling and emotions from the content of social media. The conclusion section will briefly sum up the main contributions of the research and suggest some possible research areas for the future.

CHAPTER 5

DISCUSSION AND FUTURE WORK

5.1 Research Outcomes

This investigational has to do with sentiment analysis and belief prediction of the members of social networks. The core findings we come to when we analyze Twitter data are the following:

Image modality (presence of pictures) and sentiment polarity (the degree to which the sentiment is positive or negative): The tweets containing images display big differences in the sentiment polarity distribution as compared to the tweets without text. Image modal tweets commonly present with positive emotions. It is possible to assume that here one of the reasons is the fact that images are able to deliver emotional content more efficiently; for instance, through smiles and emojis, which are associated with feelings.

Emotional polarity (based on brand-specific) characteristics: Variability presents in the emotional polarity dispersal across different brands on social media platforms. Throughout the Adjectives section, the polarity remains mostly positive; hence, the narrative has a positive attitude to cover. But, in the Narrative parts, the sentiment polarity is mainly negative, while the volume of text expressing strong emotions is noticeably smaller in comparison with the Adjectives parts. This goes along with the observation that users have their own strong feelings towards different brands. Here, brands should implement an analytics tool, such as the sentiment analysis, to understand the way their clients perceive their products and services, which will allow them to carry out more customized marketing.

Sentiment Analysis (SA) and User Behavior (UB) Prediction: The prediction of user behavior by way of building feature engineering that models the result of the sentiment analysis is possible. For clarification, we can conduct a test to find out if a user is likely to retweet, comment, or like a tweet according to his focus on sentiment probability and sentiment distribution. In this way, a new approach to the analysis of online user conduct is achieved, and the basis for the creation of behavioral models of the internet user conduct is laid.

5.2 Comparison with previous research

Contrasted with other sentiment analysis and user behavior prediction studies, the current research serves as a representative example of the following:

Multimodal data analysis: This article is an example of multimodal data analysis that uses images and text, which was less present in earlier studies than that focused on text data. It contributes to discovering more concisely the sentiment conveyed by the user and assesses the precision of the sentiment analysis.

Domain-Specific Analysis: Brands' information is provided in the case, however, previous analysis concentrated on generic domains. Such an approach empowers the brand by 'zeroing' in on the emotions her clients are expressing in particular areas and this ultimately facilitates formulation of a more balanced and effective brand marketing strategy.

Interpretability Exploration: As the final opinion of the article reviews how to interpret classifications of sentiment and behavior models, the previous reviews mainly centered on the performance of models. It contributes to increasing model reliability and, at the same time, the information comprehensibility.

5.3 Future works

Further directions in future research include:

Cross-cultural sentiment analysis: The investigation of sentimental differences in the expression of reasons in various cultural surroundings and the creation of the multidimensional methods of inter-cultural sentiment analysis in order to enhance the accuracy and applicability of sentiment analysis. Multi-task learning: Bring in sentiment analysis with other activities like the user behavior analysis as well as work on enhancing the model efficacy to have a complete view of user behavior.

Explain ability Enhancements: Creation of more sophisticated methods for understanding the topics as well as supplying simplifying and easily apprehend able responses to the user for grasping the results of the model.

Data privacy protection: Explore the effective ways of safeguarding user privacy and data to develop sentiment analysis and user behavior prediction, which strikes a balance between data utilization and user data privacy. This study contributes to the literature on sentiment analysis and user behavior prediction. It raises new concerns or questions in the social networking context and guides research in future studies.

References

[1]. **Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: Data mining, inference, and prediction. Springer Science & Business Media**.

[2]. **Aggarwal, C. C. (2018). Machine learning for text. Springer**.

[3]. **Socher, R., Perelygin, A., Wu, J. Y., Chuang, J., Manning, C. D., Ng, A., & Potts, C. (2013, July). Recursive deep models for semantic compositionality over a sentiment treebank. In Proceedings of the 2013 conference on empirical methods in natural language processing (pp. 1631-1642)**.

[4]. **Bird, S., Klein, E., & Loper, E. (2009). Natural language processing with Python. " O'Reilly Media, Inc."**.

[5]. **Jurafsky, D., & Martin, J. H. (2009). Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition. Prentice Hall**.

[6]. **Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up?: Sentiment classification using machine learning techniques. In Proceedings of the ACL-02 conference on empirical methods in natural language processing-Volume 1 (pp. 79-86)**.

[7]. **Piao, S., Wang, M., & Hu, J. (2014, August). Domain adaptation for sentiment classification with multiple sources. In Proceedings of the 2014 conference on empirical methods in natural language processing (pp. 545-550)**.

[8]. **Bamman, D., O'Connor, B., & Smith, N. A. (2012, June). Cross-domain sentiment analysis. In Proceedings of the 2012 conference of the north american chapter of the association for computational linguistics: human language technologies (pp. 549-556)**.

[9]. **Wang, Y., Kang, L., Zhao, Y., & Zhang, L. (2016, August). A deep learning approach for cross-cultural sentiment analysis. In Proceedings of the 2016 conference on empirical methods in natural language processing (pp. 2435-2445)**.

[10]. **Agerri, R., Otegi, A., de Ilarraza, A., & Ricardo, A. (2013, June). Cross-lingual sentiment analysis: A survey. In Proceedings of the 2013 conference of the north american chapter of the association for computational linguistics: human language technologies (pp. 110-119)**.

[11]. **Lipton, Z. C. (2018). The mythos of model interpretability. Queue, 16(3), 1-33**.

[12]. **Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Pedreschi, D., & Giannotti, F. (2018). A survey of methods for explaining black box models. ACM computing surveys (CSUR), 51(5), 1-35**.

[13]. **Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). "Why should I trust you?": Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery & data mining (pp. 1135-1144)**.

[14]. **Blei, D. M., & McAuliffe, J. D. (2017). Probabilistic models of cognition: Exploring human inference. MIT press**.

[15]. **Zook, M. (2018). The promise and peril of big data. Harvard University Press**.