UTM
UNIVERSITI TEKNOLOGI MALAYSIA

PROJECT PROPOSAL PRESENTATION

Sentiment Analysis and User Behavior Prediction in Social Networks

PRESENTED BY : LIU MINGJIE
MATRIC NO : MCS241019
LECTURER : ASSOC. PROF. DR MOHD SHAHIZAN BIN OTHMAN

INTRODUCTION

BACKGROUND

- Social media has become an important part of people's daily lives, generating a huge amount of user-generated content.
- This content contains a wealth of emotional and behavioral information, which is important for understanding user needs, developing marketing strategies, and more.
- Social media data, including text, images, and videos, is the primary source of information for video social networks. Sentiment analysis and user behavior prediction are mission-critical extractions from this data and are critical for many applications.

PROBLEM STATEMENT

- Disadvantages of single-modality analysis: At present, most sentiment analysis studies mainly focus on text modalities and ignore the information of other modalities such as images and videos, resulting in the analysis results may not be comprehensive and accurate.
- Challenges of multimodal fusion: Although multimodal sentiment analysis has become a research hotspot, how to effectively fuse information from different modalities and overcome the heterogeneity and complexity between modalities is still an urgent problem to be solved.

RESEARCH QUESTIONS & OBJECTIVES

Research Question

- **RQ 1**: How to construct an effective multimodal sentiment analysis framework to make full use of information in multiple modalities such as text and images?
- **RQ2**: What role does image modality play in sentiment analysis? How can I quantify the impact of images on sentiment analysis results?
- **RQ3** : Are there significant differences in the emotional expression of different brands on social media? What is the reason behind these differences?

Research Objectives

- This paper constructs and verifies a multimodal sentiment analysis framework, which can effectively integrate text and image information to improve the accuracy and efficiency of sentiment analysis..
- Quantify the influence of image modality in sentiment analysis, and determine the contribution of image features to the judgment of overall sentiment polarity through experimental analysis.
- Analyze the differences in the emotional expression of different brands on social media, and explore the reasons behind these differences, such as brand image, product characteristics, etc.

LITERATURE REVIEW

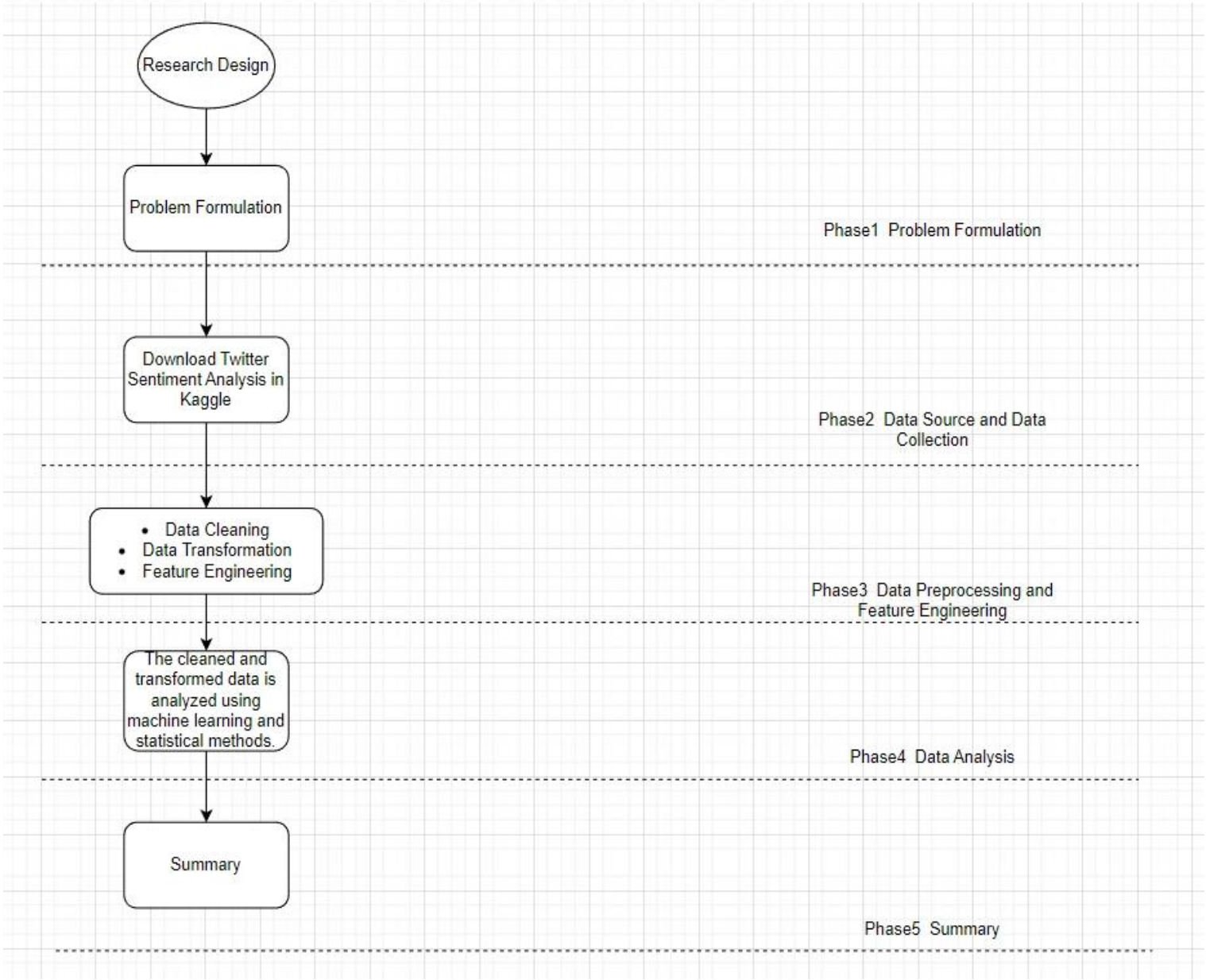
LITERATURE REVIEW

Title	Author	Method	Finding
Machine learning for text	Aggarwal, C. C.	Machine learning techniques, including Naive Bayes and SVM	Machine learning is an effective method for sentiment analysis, with Naive Bayes and SVM being commonly used models
Recursive deep models for semantic compositionality over a sentiment treebank	Socher, R., Perelygin, A., Wu, J. Y., Chuang, J., Manning, C. D., Ng, A., & Potts, C.	Recursive neural networks	Recursive neural networks are effective for sentiment analysis tasks

LITERATURE REVIEW

Title	Author	Method	Finding
Domain adaptation for sentiment classification with multiple sources	Piao, S., Wang, M., & Hu, J.	Domain adaptation techniques	Domain adaptation techniques can be used to improve the performance of sentiment analysis models in specific domains
A deep learning approach for cross-cultural sentiment analysis	Wang, Y., Kang, L., Zhao, Y., & Zhang, L.	Deep learning models	Deep learning models can be used for cross-cultural sentiment analysis and can effectively handle the challenges of language and cultural differences

METHODOLOGY



RESEARCH FRAMEWORK

Dataset

```
[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

```
*[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]:

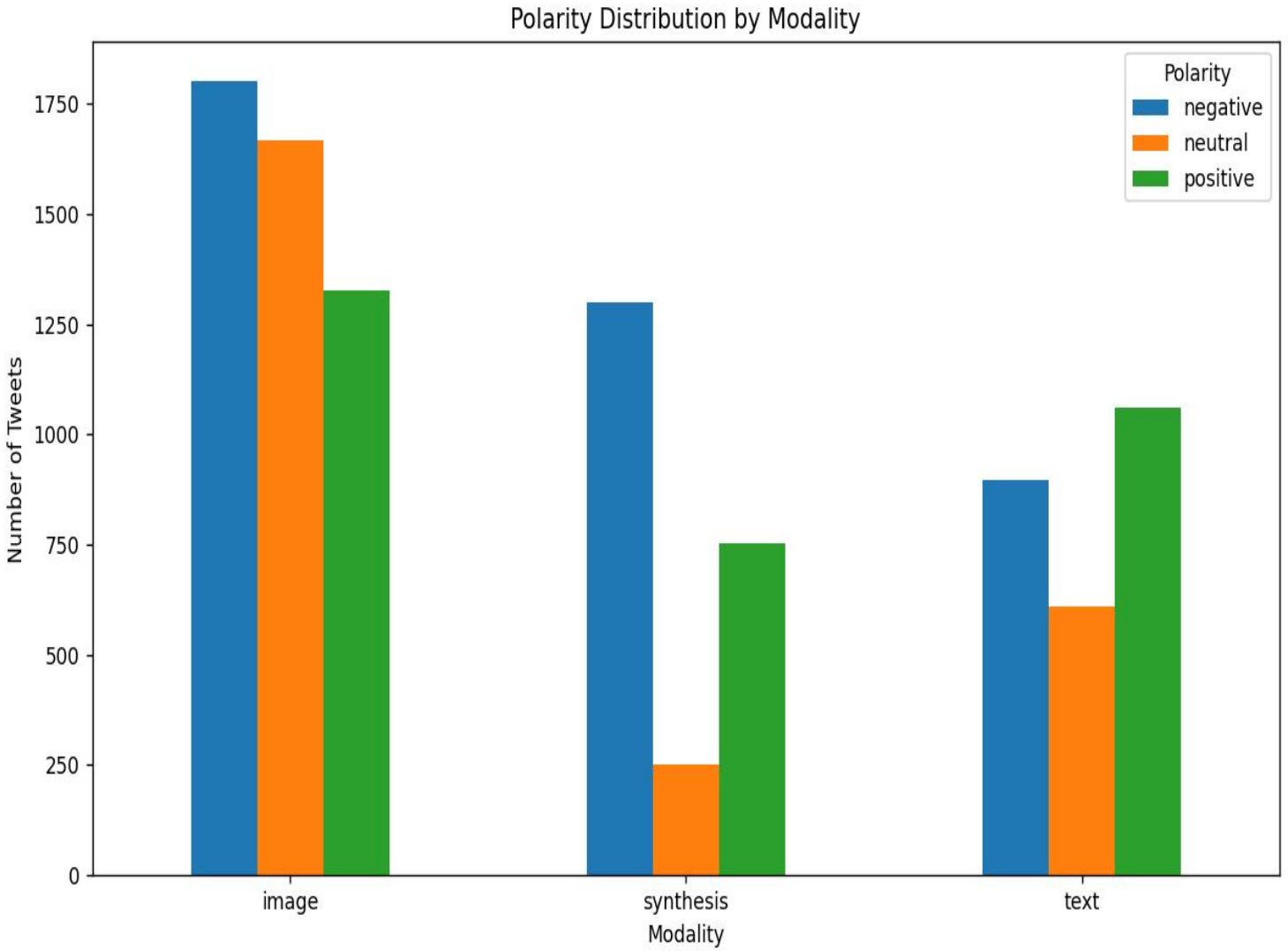
	Twitter ID	Tweet content	Image	Label the user ID	modality	polarity	emotion
0	1483637809449779200	kate and toby's smoker heating up to destroy t...	http://10.112.67.227:8666/img/download/1483637...	3	image	negative	["anger"]
1	1498163459749715973	@VIVIZ_staff All the best sinb, eunha, umji 🙂...	http://10.112.67.227:8666/img/download/1498163...	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

DATA CLEANING

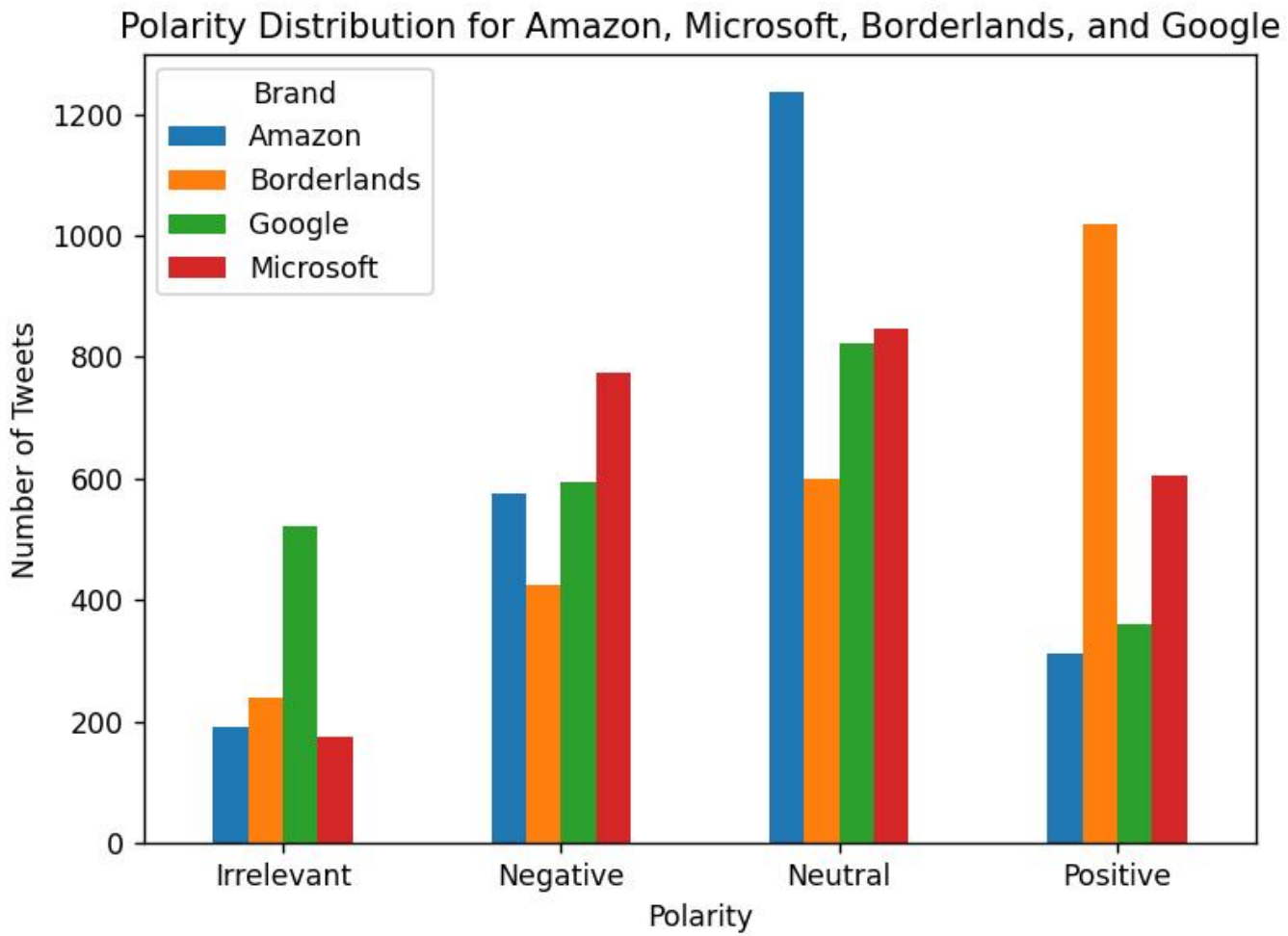
- Firstly, it is essential to identify and handle missing values within the dataset. This can involve imputation techniques to estimate missing values based on existing data or, in some cases, removing instances with missing values altogether. Once the missing values are addressed, the next crucial step is to identify and remove duplicate data entries. Duplicates can lead to redundancy and biased analysis results. By ensuring each data point is unique, we maintain the dataset's integrity and improve the accuracy of subsequent analysis.

Initial Findings



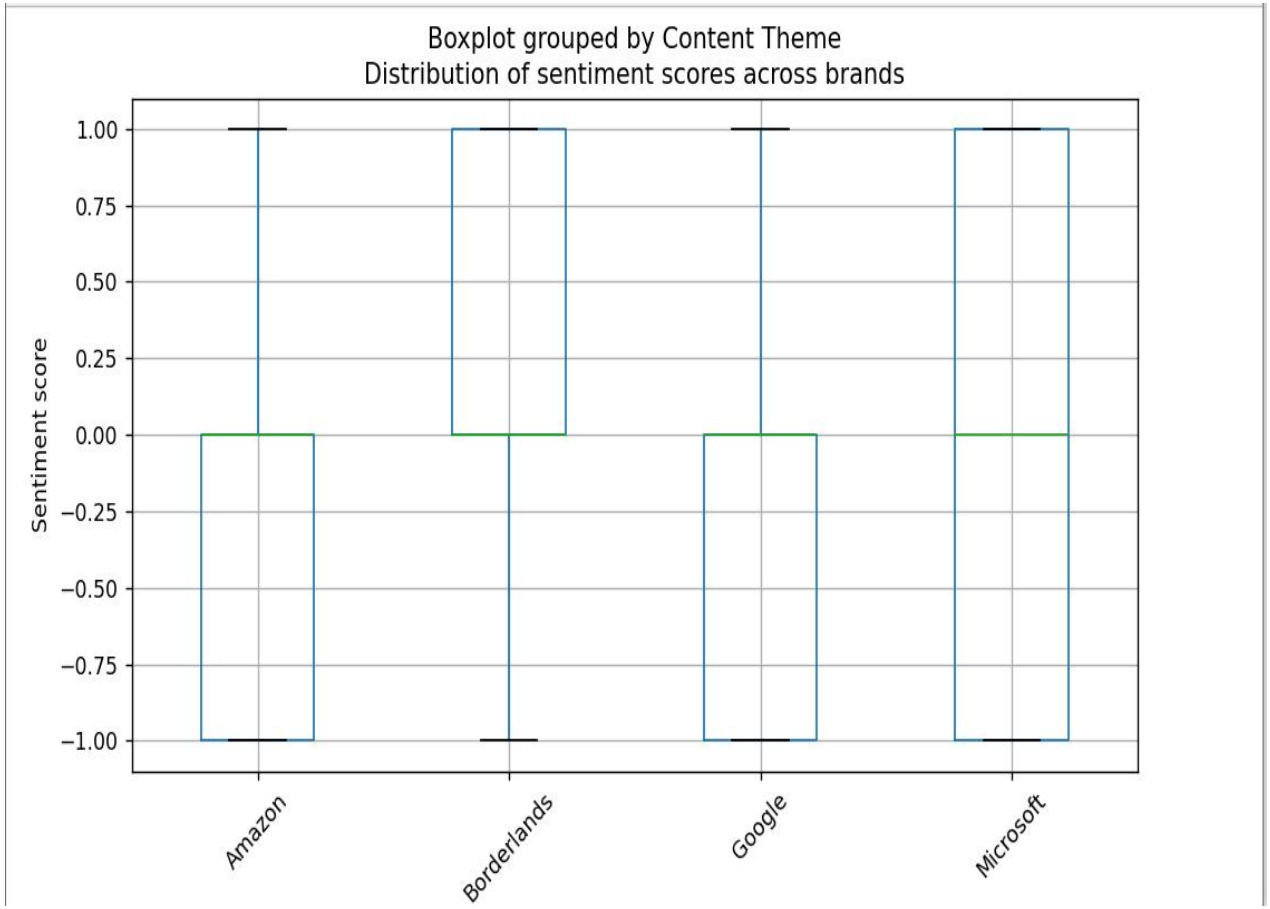
Polarity Distribution by Modality shows the distribution of emotional polarity across different modalities (image, compositive, text). As you can see from the graph, the image modality has the highest number of negative sentiment tweets, followed by the text modality. The synthetic modality has the highest number of neutral sentiment tweets, followed by the image modality. Overall, with the exception of the synthetic modality, the number of positive sentiment tweets in the other two modalities is higher than the number of negative sentiment tweets. This suggests that people are more inclined to express positive emotions than negative ones, whether in images or text.

Initial Findings



From the data, it can be seen that there are significant differences in the distribution of sentiment polarity on social media among different brands. For example, Borderlands has the highest proportion of positive emotional polarity, indicating that users have a high favorability towards it; Amazon and Google have a relatively high proportion of negative emotional polarity, indicating that users have more negative views about them. Microsoft's emotional polarity distribution is more balanced, but negative emotional polarity still accounts for a large proportion.

Initial Findings



The average sentiment score indicates the general brand perception on social media. For instance, Borderlands has a positive score, suggesting user favorability, while Amazon and Google have negative scores, indicating user dissatisfaction. Microsoft's score is neutral, leaning slightly negative.The standard deviation reflects sentiment stability. Amazon and Google exhibit low deviation, showing consistent user attitudes, while Borderlands and Microsoft have higher deviation, suggesting greater fluctuation due to events or campaigns.
These insights reveal brand image and provide valuable features for predicting user behavior.

THANK YOU

univteknologimalaysia utm.my utmofficial