International Journal of Scientific Research & Engineering Trends
Volume 10, Issue 5, Sept-Oct-2024, ISSN (Online): 2395-566X

Sentiment Analysis of Customer Reviews Using Natural Language Processing

Ms. Jyoshna Butty[1], Ms. Ankita Gupta[2], Dr. Jasbir Kaur[3],
Assistant Professor Ms. Ifrah Kampoo[4], Assistant Professor Mr.Suraj Kanal[5]
Master of Computer Applications (MCA) at Guru Nanak Institute of Management Studies, Matunga, Mumbai, India[1,2]
Director GNIMS B-School, Head of Information Technology and HR at Guru Nanak Institute of Management Studies, Matunga, Mumbai, India[3]
Department of IT, Academic Co-ordinator for MCA Course, at Guru Nanak Institute of Management Studies, Matunga, Mumbai, India[4]
Guru Nanak Institute of Management Studies, Matunga, Mumbai, India[5]

Abstract- The purpose of this research is to use Natural Language Processing (NLP) to categorize customer reviews into three groups: favorable, negative, and neutral. We employ machine learning models to categorize sentiment by preprocessing textual data. Matplotlib is then used to illustrate the results using area plots, pie charts, and keyword-based analysis. Our investigation shows how sentiment analysis, which provides actionable insights generated from consumer feedback, can advise firms on how to improve customer satisfaction and experience.

Index Terms- based analysis, Area Plot, Matplotlib, Natural Language Processing, Customer Reviews, Sentiment Analysis, and Visualization.

I. INTRODUCTION

In today's business environment, client feedback is vital for determining corporate strategy and raising client satisfaction levels. Customer evaluations have increased dramatically as a result of the rise of internet platforms, providing businesses with never-before-seen access to user feedback. Nevertheless, it is wasteful and impossible to manually analyze these enormous amounts of data, which calls for the adoption of automated methods like Natural Language Processing (NLP).

The interaction of computers and human language is the focus of the artificial intelligence field of natural language processing, or NLP. Sentiment analysis, which seeks to ascertain the sentiment (positive, negative, or neutral) communicated in text, is one of the most important uses of natural language processing (NLP).

Since customer evaluations offer unbiased comments on goods and services, they are a perfect source of data for sentiment analysis. Gratitude

II. MAIN BODY SECTIONS

1 Data Preprocessing

For sentiment classification to be accurate, the text preparation step is essential. The following actions are involved:

- Tokenization is the process of dividing a text into discrete words or tokens.
- Stopword Removal: Eliminating frequently used terms that don't add context, like "is," "the," etc.
- Lemmatization is the process of breaking words down to their most basic forms such that phrases remain consistent (e.g., "running" becomes "run").

2. Sentiment Classification

Sentiment-based review classification forms the basis of the analysis. Reviews are divided into three sentiment categories: positive, negative, and neutral using the VADER model, which rates the sentiment of individual words.

For a more sophisticated classification, an additional machine learning model built on logistic regression is employed.

3. Visualization Using Matplotlib

Understanding sentiment trends and pinpointing areas of consumer happiness or dissatisfaction requires the use of visualizations. To display the findings, we employ the following plots:

- **Pie Chart:** Shows the percentage of favorable, unfavorable, and neutral evaluations.
- The analysis based on keywords displays the terms that appear most frequently in both favorable and negative reviews.

© 2024 IJSRET

- **Area Plot:** Monitors how consumer sentiment changes over time.

Figures and Tables

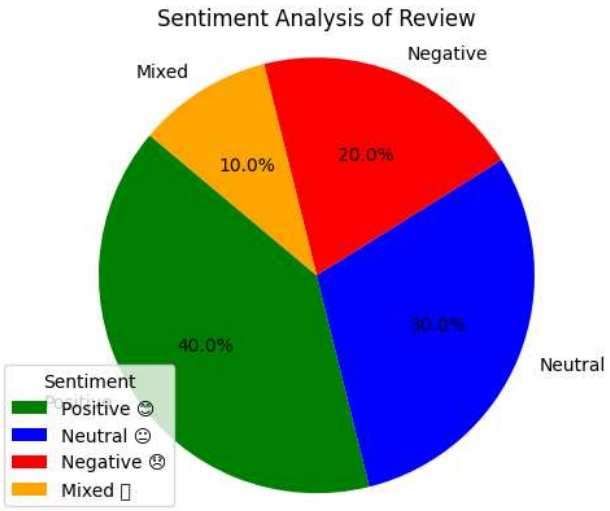
The pie chart provides an overview of the sentiment distribution.

```
!pip install vaderSentiment
from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer
file_path = "amazon.csv"
with open(file_path, 'r', encoding='latin-1') as file: text_data = file.read()

analyzer = SentimentIntensityAnalyzer()
sentiment_scores = analyzer.polarity_scores(text_data)
positive_score = sentiment_scores['pos']
negative_score = sentiment_scores['neg']
neutral_score = sentiment_scores['neu']

lelabs = ['Positive', 'Negative', 'Neutral']
sizes = [positive_score, negative_score, neutral_score]
colors = ['#4CAF50', '#FF5252', '#FFC107']
emojis = ['😊', '😞', '😐']

import matplotlib.pyplot as plt

# Define the data for the pie chart
sizes = [40, 30, 20, 10]  # Example sentiment distribution (e.g., positive,
neutral, negative, mixed)
labels = ['Positive', 'Neutral', 'Negative', 'Mixed']  # Categories for sentiments
colors = ['green', 'blue', 'red', 'orange']  # Colors for each sentiment category
emojis = ['😊', '😐', '😞', '😕']  # Emojis corresponding to each sentiment
# Create the pie chart
plt.pie(sizes, labels=labels, colors=colors, autopct='%1.1f%%',
startangle=140)

# Set the title of the chart
plt.title('Sentiment Analysis of Review')

# Create the legend with both the label and emoji
plt.legend(title='Sentiment', loc='best', labels=[f'{label} {emoji}' for label, emoji
in zip(labels, emojis)])

# Ensure the pie chart is a circle
plt.axis('equal')

# Show the plot
plt.show()
```

Link:https://drive.google.com/file/d/1O_jMBzCwnqiPPnogyhMbbee8lpgo0bhK/view?usp=sharing

Figure 1: Sentiment Distribution Pie Chart

Output



Using VADER from nltk, the code evaluates the sentiment of customer reviews and uses matplotlib to display the distribution of positive, neutral, and negative sentiments as a pie chart.

Keyword analysis shows the most frequently mentioned words in different sentiment categories.

```
import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv('/content/amazon.csv')

good_keywords = ['love', 'fantastic', 'great', 'amazing', 'excellent', 'satisfied',
'recommend']
bad_keywords = ['bad', 'poor', 'disappointed', 'hate', 'average', 'okay']
worst_keywords = ['worst', 'terrible', 'awful', 'horrible', 'disaster', 'nightmare']

def classify_comments(text):
  text = text.lower()
  if any(word in text for word in worst_keywords):
    return 'worst'
  elif any(word in text for word in bad_keywords):
    return 'bad'
  elif any(word in text for word in good_keywords):
    return 'good'
  else:
    return 'neutral'

df['comment_category'] = df['reviewText'].apply(classify_comments)

category_counts = df['comment_category'].value_counts()
```

© 2024 IJSRET

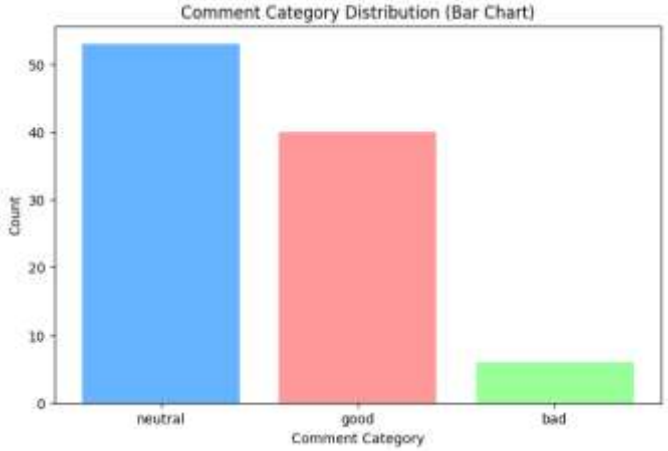
International Journal of Scientific Research & Engineering Trends
Volume 10, Issue 5, Sept-Oct-2024, ISSN (Online): 2395-566X

```
plt.figure(figsize=(8, 5))
plt.bar(category_counts.index, category_counts.values, color=['#66b3ff',
'#ff9999', '#99ff99'])
plt.title('Comment Category Distribution (Bar Chart)')
plt.xlabel('Comment Category')
plt.ylabel('Count')
plt.show()
```

Link:https://drive.google.com/file/d/1XXShqLNrmpW6zKRovHST2azayfAWmsaS/view?usp=sharing

Figure 2: Keyword-based Analysis

Output



The distribution of comment categories is displayed in a bar chart, where the y-axis indicates the number of comments in each category and the x-axis the categories (poor, good, and neutral). The graph shows that "neutral" comments make up the majority of comments, followed by "good," "bad," and so on.

An area plot depicts sentiment variation over time.

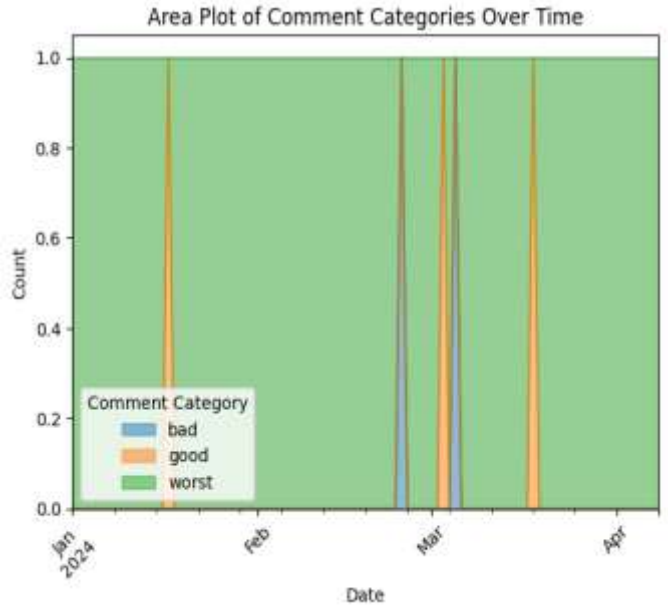
```
import pandas as pd
import matplotlib.pyplot as plt

# Load the data
df = pd.read_csv('amazon.csv')

# For demonstration, simulate dates and categorize comments
df['date'] = pd.date_range(start='2024-01-01', periods=len(df), freq='D')
df['comment_category'] = df['reviewText'].apply(lambda x: 'good' if 'love' in x
else 'bad' if 'bad' in x else 'worst')

# Count occurrences over time
df_counts = df.groupby(['date',
'comment_category']).size().unstack(fill_value=0)
```

```
# Create an Area Plot
plt.figure(figsize=(10, 6))
df_counts.plot(kind='area', alpha=0.5)
plt.title('Area Plot of Comment Categories Over Time')
plt.xlabel('Date')
plt.ylabel('Count')
plt.xticks(rotation=45)
plt.legend(title='Comment Category')
plt.show()
```

Link:https://drive.google.com/file/d/1BSl_iYiPfwbhVTiJkHIpmsHwxkiRzt89/view?usp=sharing

Figure 3: Sentiment Trends Area Plot

Output



The "good" category continuously dominates over time, according to the area plot that shows the temporal trends of the comment categories. The "bad" and "worst" categories, on the other hand, show scattered peaks that indicate a generally positive mood with occasional negative spikes.

III. EQUATIONS

Sentiment scores are calculated as the sum of the individual word scores, defined as:

$$\text{Sentiment Score}=i=1\sum n\text{Polarity}(wi)\times\text{Frequency}(wi)$$

Where wi is the i-th word , and its polarity is determined based on a predefined lexicon.

IV. CONCLUSION

This study used Natural Language Processing techniques to successfully apply sentiment analysis to customer feedback. A comprehensive comprehension of client sentiment trends was

© 2024 IJSRET

made possible by the combination of matplotlib's visual representations and VADER's rule-based sentiment classification. Area plots, pie charts, and keyword-based analytics offered insightful information about consumer experiences. In the future, sentiment categorization accuracy can be further enhanced by using larger datasets and more complex models, which will make it an even more useful tool for developing business strategies.

Acknowledgments
We sincerely appreciate the support and direction provided by the Guru Nanak Institute of Management Studies faculty and staff. We also thank our peers for their thoughtful criticism and recommendations during this study.

REFERENCES

1. Sindhura, K. (2023). Sentiment Analysis Using Natural Language Processing and Machine Learning. Journal of Data Acquisition and Processing, 38(2). DOI: 10.5281/zenodo.7766376.
2. Misal, S. R., & Mundhe, S. D. (2023). Customer Reviews Sentiments Analysis Using Natural Language Processing (NLP) and Deep Learning. Shodh Sanchar Bulletin.
3. Shiveshwarkar, R., Shende, O., Londhe, S., Ramane, S., & Khadkikar, P. A. (2022). Review on Sentiment Analysis on Customer Reviews. International Research Journal of Engineering and Technology (IRJET), 9(5), May 2022.
4. Fang, X., & Zhan, J. (2015). Sentiment analysis using product review data. Journal of Big Data, 2(5).
5. Sasikala, P., & Mary Immaculate Sheela, L. (2020). Sentiment analysis of online product reviews using DLMNN and future prediction of online product using IANFIS. Journal of Big Data, 7(33).
6. Panchal, D., Mehta, M., Mishra, A., Ghole, S., & Dandge, S. (2022). Sentiment Analysis Using Natural Language Processing. International Journal for Research in Applied Science & Engineering Technology, 10(V), 2262-2266.
7. Nikhil, N. D., Patil, P. V., Pawar, R. V., Phepale, R. B., & Waghmare, S. V. (2022). Sentiment Analysis Using Natural Language Processing. International Journal of Scientific Development and Research, 7(5), 458-462.
8. Aqlan, A. A. Q., Manjula, B., & Naik, R. L. (2019). A Study of Sentiment Analysis: Concepts, Techniques, and Challenges. Lecture Notes on Data Engineering and Communications Technologies, 28, 147-162.
9. Sultana, N., Kumar, P., Patra, M. R., Chandra, S., & Alam, S. K. S. (2019). Sentiment analysis for product review. ICTACT Journal on Soft Computing, 9(3), 1913-1919.
10. Harunasir, M. F. B., Palanichamy, N., Haw, S.-C., & Ng, K.-W. (2023). Sentiment analysis of Amazon product reviews by supervised machine learning models. Journal of Advances in Information Technology, 14(4), 857-862.

© 2024 IJSRET