

REG No:513421106042

NAME:SUGANYA K BE(ECE) III Yr

COLLEGE NAME: University College of engineering kancheepuram

SENTIMENT ANALYSIS FOR MARKETING

Phase 1: Problem Definition and Design Thinking OBJECTIVE:

Sentiment analysis can do wonders for any marketer. By understanding what your target audience is thinking on a scale that only sentiment analysis can achieve, you can tweak a product, campaign, and more, to

meet their needs and let your customers know you're listening.

Sentiment analysis is an artificial intelligence technique that uses machine learning and natural language processing (NLP) to analyse text for polarity of opinion (positive to negative).

DATA COLLECTION:

Depending on your research goal and scope, you can choose from different data sources and methods for

sentiment analysis. Some common data sources are online reviews, social media platforms, blogs, forums, news

articles, surveys, and emails. Some common methods are web scraping, APIs, surveys, and online tools.

DATA PREPROCESSING:

- The code begins by importing the necessary libraries, including pandas for data handling, matplotlib and seaborn for visualization, and scikit-learn for machine learning.

- The airline tweet dataset is loaded from a CSV file.

For cleaning the data, we will do the following:

Combine both test and training set so we can preprocess both together

Remove redundant characters- numerics, special characters (not hashtags), short words, usernames(@user)

Tokenise the processed tweet

Stemming- strip suffixes to get the root word.

SENTIMENT ANALYSIS TECHNIQUES:

Sentiment analysis is used to analyse raw text to drive objective quantitative results using natural

language processing, machine learning, and other data analytics techniques. It is used to detect positive or

negative sentiment in text, and often businesses use it to gauge branded reputation among their customers.

There are various types of sentiment analysis where the models focus on feelings and emotions, urgency, even

intentions, and polarity. The most popular types of sentiment analysis are:

- A. Fine-grained sentiment analysis
- B. Emotion detection
- C. Aspect based sentiment analysis
- D. Multilingual sentiment analysis

Sentiment analysis is critical because it helps businesses to understand the emotion and sentiments of their

customers. Companies analyze customers' sentiment through social media conversations and reviews so they

can make better-informed decisions. The Global Sentiment Analysis Software Market is projected to reach

US\$4.3 billion by the year 2027. Between 2017 and 2023, the global sentiment analysis market will increase by

a CAGR of 14%.

FEATURE EXTRACTION:

In sentiment analysis, we detect tweets that have negative sentiment, i.e, racist, sexist or general hate speech. Here, tweets with a label '1' denote a negative tweet, while '0' denotes the absence of hate speech in the

tweet.

We extract features using the following:

1. Bag of Words Features
2. TF-IDF features
3. Word Embedding's

VISUALIZATION:

- The code creates a histogram to visualize the distribution of airline sentiments.
- It also creates a pie chart to visualize the sentiment distribution using percentages.

We will analyse the text of the tweet and its relation to the sentiment with the following:

Wordcloud: Most used words (have bigger fonts), for positive and negative tweets. Reference.

Hashtags: Analyse the effect of hashtags on the tweet sentiment.

```
all_words=' '.join([text for text in combine['tidy_tweet']])  
  
from wordcloud import WordCloud  
  
word_cloud=WordCloud(width=800, height=500, random_state=21,max_font_size=110,colormap=  
generate(all_words)  
  
plt.figure(figsize=(10,  
7)) plt.imshow(word_cloud)  
  
plt.axis('off') plt.show()
```

INSIGHTS GENERATION:

Sentiment Analysis is a natural language processing (NLP) technique used to determine the sentiment of data, i.e. whether the data is positive, negative or neutral. NLP is a branch of artificial intelligence concerning

linguistics, more specifically how a computer understands, processes, and analyses large amount of natural

language data. It has a vast amount of other use cases such as text classification, speech recognition, chat-bots

and more, however the main focus for this project will be Sentiment Analysis. Customer Insights One of the

main applications of Sentiment Analysis is for Customer Insights. A customer insight is an interpretation used

by businesses to gain a deeper understanding of how their audience feels towards their product or business,

allowing them to better understand their consumers needs and improve their product/service accordingly.

```
# Import Libraries import  
pandas as pd  
  
import matplotlib.pyplot as plt import  
seaborn as sns  
  
from sklearn.feature_extraction.text import TfidfVectorizer
```

```
from sklearn.linear_model import LogisticRegression from
sklearn.metrics import roc_auc_score, confusion_matrix from
sklearn.model_selection import train_test_split
```

visualize:

```
# Visualize the distribution of airline sentiments using a pie chart sentiment_counts
= df['airline_sentiment'].value_counts()
plt.figure(figsize=(8, 8))
plt.pie(sentiment_counts, labels=sentiment_counts.index, autopct='%1.1f%%', startangle=140)
plt.title('Distribution of Airline Sentiments') plt.axis('equal') # Equal aspect ratio ensures that
pie is drawn as a circle.
```

```
# function to collect hashtags def
```

```
hashtag_extract(x):
```

```
    hashtags = []
```

```
    for i in x:
```

```
        ht = re.findall(r"#(\w+)", i)
```

```
    hashtags.append(ht)
```

```
    return hashtags
```

```
In [22]: linkcode
```

```
#hashtag list for non negative tweets
```

```
HT_non_negative = hashtag_extract(combine['tidy_tweet'][combine['label'] == 0])
```

```
#hashtag list for negative tweets
```

```
HT_negative = hashtag_extract(combine['tidy_tweet'][combine['label'] == 1])
```

```
#unnest list
```

```
HT_non_negative = sum(HT_non_negative,[])
```

```
HT_negative = sum(HT_negative,[])
```