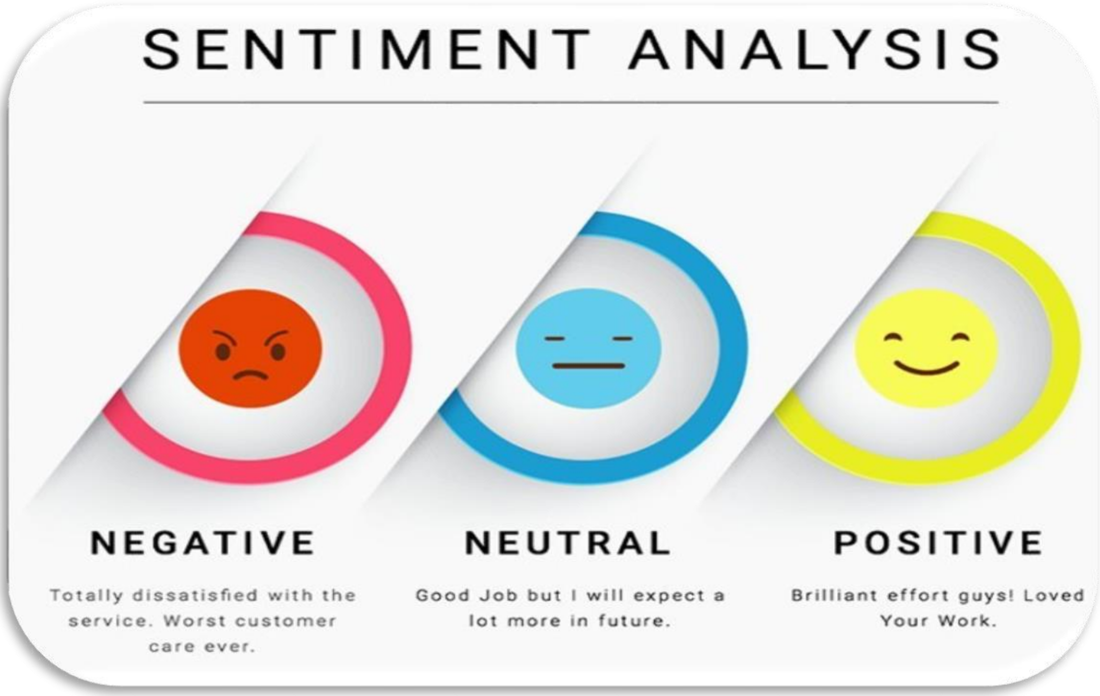
SENTIMENT ANALYSIS FOR MARKETING

PHASE 4 Submission Document

Phase 4: **Development Part 2**

Topic: **Continue building the Sentiment analysis for marketing model by feature engineering, model training and evaluation.**



Introduction

- Sentiment analysis can be defined as analysing the positive or negative sentiment of the customer in text. The contextual analysis of identifying information helps businesses understand their customers' social sentiment by monitoring online conversations.
- As customers express their reviews and thoughts about the brand more openly than ever before, sentiment analysis has become a powerful tool to monitor and understand online conversations.
- Recent advancements in machine learning and deep learning have increased the efficiency of sentiment analysis algorithms. You can creatively use advanced artificial intelligence and machine learning tools for doing research and draw out the analysis.

Abstract

Sentiment analysis or opinion mining is the computational study of people's opinions, sentiments, attitudes, and emotions expressed in written language. It is one of the most active research areas in natural language processing and text mining in recent years. Its popularity is mainly due to two reasons. Whenever we need to make a decision we want to hear others' opinions. Second, it presents many challenging research problems, which had never been attempted before the year 2000. Part of the reason for the lack of study before was that there was little opinionated text in digital forms. It is thus no +surprise that the inception and the rapid growth of the field coincide with those of the social media on the Web.

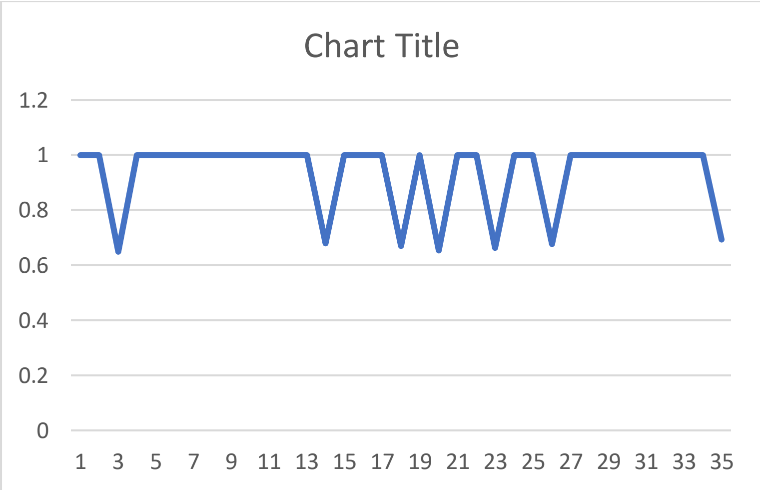
Dataset Link

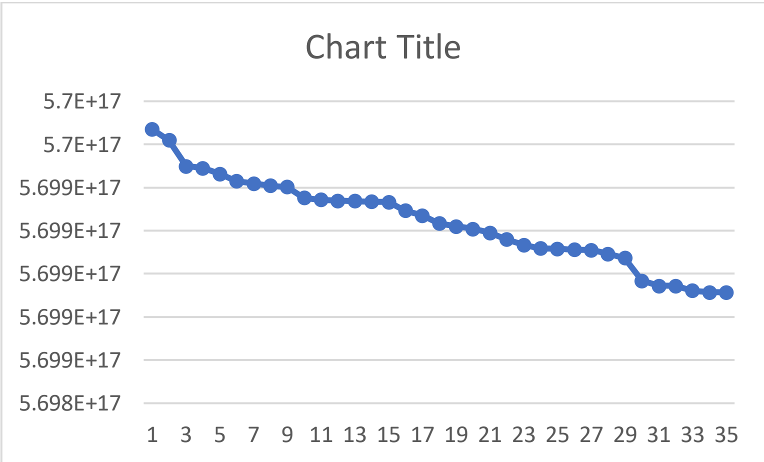
https://www.kaggle.com/datasets/crowdflower/twitter-airline-sentiment

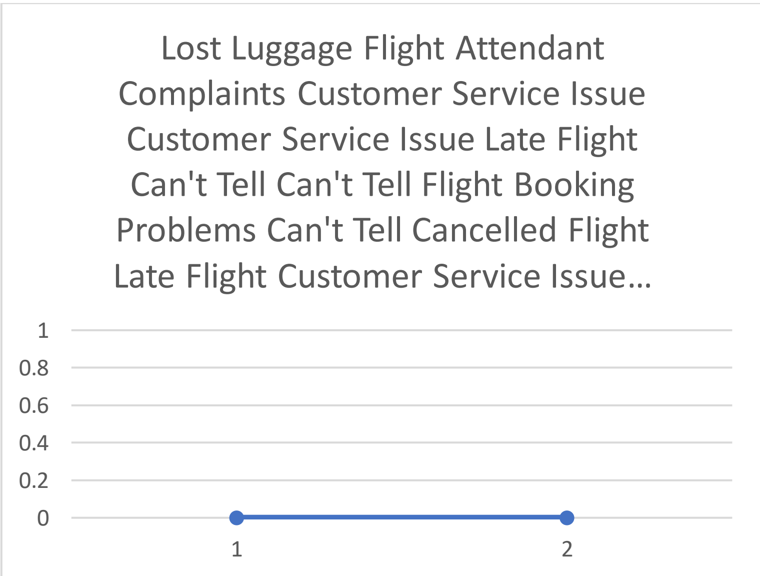
Sentiment dataset

tweet_id	airline_ser	airline_ser	negativere	negativere	airline	airline_ser	name	negativere	retweet_c	text	tweet_coo	tweet_cre	tweet_loc	user_timezone	
5.7E+17	negative	1	Lost Lugga	1	Virgin America		gianagon		0	@VirginAm	[40.64137	########	New York	Eastern Time (US & Canada)	
5.7E+17	neutral	1			Virgin America		bxchen		0	@virginamerica Need		########	San Franci	Eastern Time (US & Canada)	
5.7E+17	neutral	0.6492		0	Virgin America		seimatrun		0	@VirginAmerica I ema		########	Los Angeles		
5.7E+17	neutral	1			Virgin America		jamied7		0	@VirginAmerica hi I ju		########	London, Er	London	
5.7E+17	negative	1	Flight Atter	0.3516	Virgin America		seimatrun		0	@VirginAmerica your		########	Los Angeles		
5.7E+17	positive	1			Virgin America		mrmichaellay		0	@VirginAm	[36.08457	########	Floridian fi	Eastern Time (US & Canada)	
5.7E+17	positive	1			Virgin America		TaylorLumsden		0	@VirginAmerica awes		########	Dallas, Tex	Mountain Time (US & Canada)	
5.7E+17	neutral	1			Virgin America		campusmoviefest		0	@VirginAmerica Or w		########	USA	Eastern Time (US & Canada)	
5.7E+17	neutral	1			Virgin America		TaylorLumsden		0	@VirginAmerica first t		########	Dallas, Tex	Mountain Time (US & Canada)	
5.7E+17	negative	1	Customer	1	Virgin America		meme_meng		0	@VirginAmerica what		########			
5.7E+17	neutral	1			Virgin America		kyle_romanoff		0	@VirginAmerica what		########			
5.7E+17	negative	1	Customer	1	Virgin America		GunsNDip		0	@VirginAmerica why		########		Pacific Time (US & Canada)	
5.7E+17	positive	1			Virgin America		artisticwritr87		0	@VirginAmerica I've a		########	Seattle, W	Pacific Time (US & Canada)	
5.7E+17	negative	0.6792	Late Flight	0.3477	Virgin America		arieldaie		0	@VirginAmerica you'r		########	Los Angeles		
5.7E+17	negative	1	Can't Tell	1	Virgin America		GunsNDip		0	@VirginAmerica I hav		########		Pacific Time (US & Canada)	
5.7E+17	negative	1	Can't Tell	1	Virgin America		GunsNDip		0	@VirginAmerica it wa		########		Pacific Time (US & Canada)	
5.7E+17	negative	1	Flight Bool	1	Virgin America		jsatk		0	@VirginAm	[0.0, 0.0]	########	Lower Pac	Pacific Time (US & Canada)	
5.7E+17	neutral	0.6705		0	Virgin America		serenaklal		0	@VirginAmerica Can't		########	Chicago	Eastern Time (US & Canada)	
5.7E+17	neutral	1			Virgin America		openambit1		0	@VirginAmerica Rand		########			
5.7E+17	neutral	0.6545		0	Virgin America		cabowine		0	@VirginAmerica I <		########	Los Cabos,	Arizona	
5.7E+17	negative	1	Can't Tell	0.6513	Virgin America		MaryAnnTaylorT		0	@VirginAmerica Why		########	New York,	Arizona	
5.7E+17	neutral	1			Virgin America		RamotControl		0	@VirginAmerica "You		########		Pacific Time (US & Canada)	
5.7E+17	neutral	0.6639		0	Virgin America		losermelon		0	@VirginAmerica hi, i c		########			
5.7E+17	negative	1	Cancelled	1	Virgin America		AlisonK33774854		0	@VirginAmerica I like		########			
5.7E+17	negative	1	Late Flight	1	Virgin America		GunsNDip		0	@VirginAmerica just l		########		Pacific Time (US & Canada)	
5.7E+17	neutral	0.6765		0	Virgin America		yazdanagh		0	@VirginAmerica why		########			
5.7E+17	negative	1	Customer	0.6863	Virgin America		MerchEngines		0	@VirginAmerica Is it r		########	Los Angele	Arizona	
5.7E+17	negative	1	Customer	0.6771	Virgin America		ColorCartel		0	@VirginAmerica I can		########	Austin, TX	Mountain Time (US & Canada)	
5.7E+17	negative	1	Can't Tell	0.659	Virgin America		MustBeSpoken		0	@VirginAmerica - Let		########			
5.7E+17	negative	1	Flight Bool	0.6714	Virgin America		mattbunk		0	@virginamerica What		########	Sterling He	Eastern Time (US & Canada)	
5.7E+17	negative	1	Customer	1	Virgin America		louisjenny		0	@VirginAmerica is an		########	Washingto	Quito	
5.7E+17	neutral	1			Virgin America		STravelsW		0	@VirginAmerica trying		########	Manhattan Beach, CA		
5.7E+17	negative	1	Late Flight	0.6882	Virgin America		GunsNDip		0	@VirginAmerica why		########		Pacific Time (US & Canada)	
5.7E+17	neutral	1			Virgin America		joeyrenagade		0	@VirginAmerica chec		########	Greater Los Angeles		
5.7E+17	negative	0.6925	Late Flight	0.3521	Virgin America		mrmichaellay		0	@virginam	[0.0, 0.0]	########	Floridian fi	Eastern Time (US & Canada)	

Pre-Processing data set:







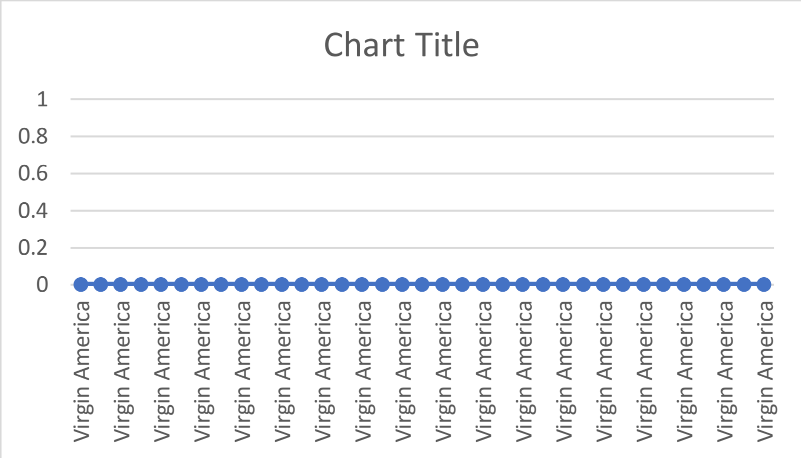


Chart Title
1
0.8
0.6
0.4
0.2
0
Virgin America

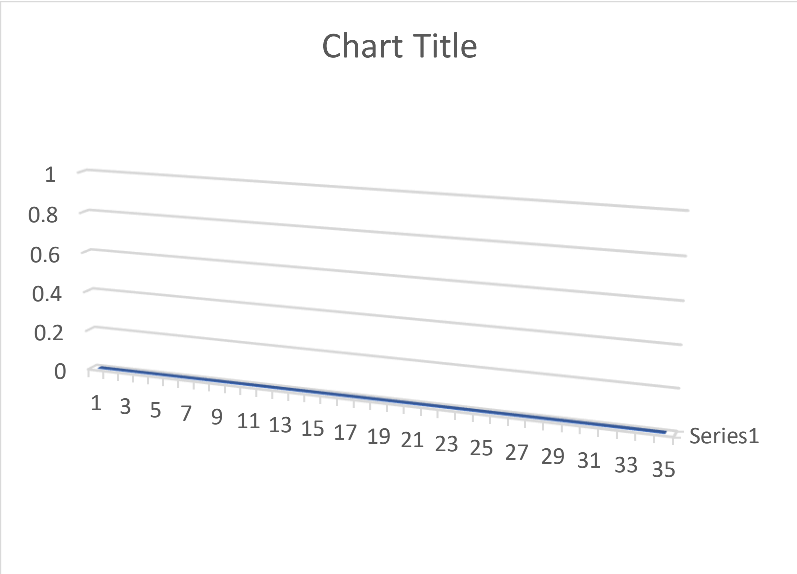
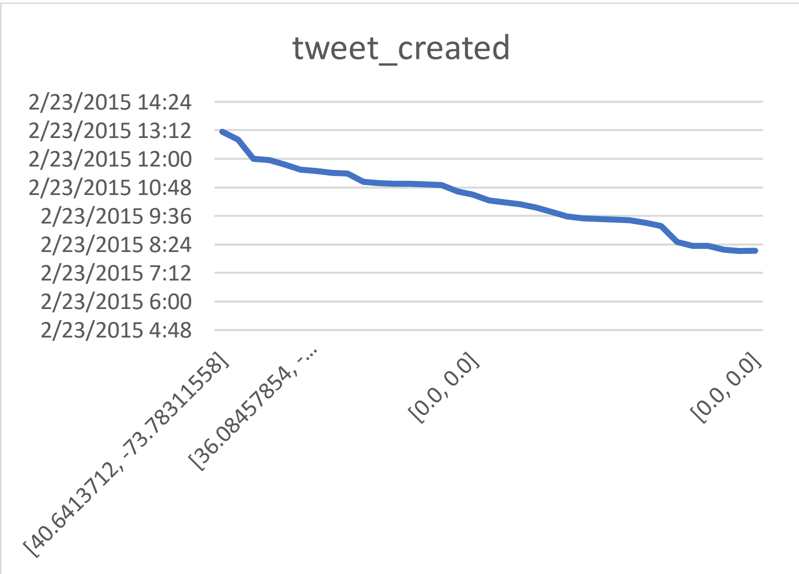


Chart Title
1
0.8
0.6
0.4
0.2
0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35
Series1


tweet_created
2/23/2015 14:24
2/23/2015 13:12
2/23/2015 12:00
2/23/2015 10:48
2/23/2015 9:36
2/23/2015 8:24
2/23/2015 7:12
2/23/2015 6:00
2/23/2015 4:48
[40.6413712, -73.78311558]
[36.08457854, -...
[0.0, 0.0]
[0.0, 0.0]

Data Collection:

Identify a dataset containing customer reviews and sentiment about competitor products.

Data Preprocessing:

Clean and preprocess the textual data for analysis.

- Step 1: Delete duplicate data.
- Step 2: Remove irrelevant items.
- Step 3: Check for outlier data.
- Step 4: Correct typos and structural mistakes.
- Step 5: Check for missing data.
- Step 6: Validate your data.
- Discover More: Complete Sentiment Analysis Process.
- The data pre-processing techniques includes five activities such as Data Cleaning, Data Optimization, Data Transformation, Data Integration and Data Conversion.

Sentiment analysis techniques:

Employ different NLP techniques like Bag of words, word embeddings, or transformer models for sentiment analysis.

Feature extraction:

Extract features and sentiment from the text data.

Visualization:

Create visualization to depict the sentiment distribution and analyze trends.

Insights Generation:

Extract meaningful insights from the sentiment analysis result to guide business decisions.

Sentiment analysis techniques:

- NLP model: A rule-based system uses a set of human-crafted rules to help identify subjectivity, polarity, or the subject of an opinion.
- These rules may include various NLP techniques developed in computational linguistics, such as: Stemming, tokenization, part-of-speech tagging and parsing.
- bag-of-words model in sentiment analysis: bag-of-words model is a way of extracting features from text so the text input can be used with machine learning algorithms like neural networks.
- Each document, in this case a review, is converted into a vector representation.

Proposed system:

- Performing sentiment analysis on social networks such as Twitter is considered as a significant tool to gather information about the opinion or emotions of the public in real-time applications.
- The proposed work helps in extracting the sentiment of the tweet posted by Twitter users in various situations and the proposed method have the capability to recognize the emotion from the text.
- The machine learning method is used for analysing the sentiment which helps to gain the capability automatic learning

to the model. The block diagram for sentiment analysis using in Twitter data using machine learning method.

Benefits

Sentiment analysis helps companies communicate better with customers and develop more relevant messages. By identifying the users' emotions, you can get a better idea of their experience and provide better customer service, which eventually leads to a decrease in customer churn

- Infer meaning from unstructured data
- Take quick action against poor customer experience
- Boost business performance and strategy
- Develop empathetic chatbot
- Reduce customer churn

Feature Engineering

Feature engineering plays a crucial role in sentiment analysis for marketing, as it helps transform raw text data into meaningful features that machine learning models can effectively utilize. By extracting relevant information and representing it in a structured format, feature engineering enhances the performance of sentiment analysis models and enables them to make accurate predictions about the sentiment expressed in marketing-related text.

Here are some key feature engineering techniques commonly employed in sentiment analysis for marketing:

- Bag-of-Words (BoW): This method represents text as a collection of words, disregarding word order and grammar. Each word is assigned a numerical value based on its frequency in the text.

    - N-grams: N-grams capture sequences of n consecutive words, preserving some context and word order information. Common n-gram sizes include bigrams (n=2) and trigrams (n=3).

    - Part-of-Speech (POS) Tagging: POS tagging assigns grammatical categories to words, such as nouns, verbs, adjectives, and adverbs. This information can help identify sentiment-bearing words and phrases.

- Sentiment Lexicons: These are pre-built dictionaries that assign sentiment scores to words, indicating their positive, negative, or neutral connotation.

- Negation Handling: Negation words like "not" or "never" can reverse the sentiment of a phrase. Feature engineering should account for negation to accurately capture the overall sentiment.

- Emoticons and Emojis: These symbols often convey sentiment and can be incorporated as features.

- Capitalization and Punctuation: Excessive capitalization and exclamation marks may indicate strong sentiment and can be used as features.

- Word Embeddings: These represent words as vectors in a high-dimensional space, capturing semantic relationships between words. Word embeddings can improve sentiment analysis accuracy.

- Topic Modeling: This technique identifies latent topics in text, which can be used to contextualize sentiment analysis.

By carefully selecting and engineering relevant features, sentiment analysis models can effectively capture the opinions, emotions, and attitudes expressed in marketing-related text, providing valuable insights for marketing campaigns, product feedback, and customer satisfaction analysis.

Model training

Sentiment analysis is a valuable tool for marketing teams as it allows them to gather insights from customer feedback, social media conversations, and online reviews. By understanding the sentiment of these interactions, marketers can identify areas for improvement, track brand reputation, and measure the effectiveness of their campaigns.

To train a sentiment analysis model for marketing, you'll need to follow these steps:

- Gather and prepare data: Collect a large dataset of text samples relevant to your marketing domain. This could include social media posts, product reviews, customer feedback, and other forms of online communication.

- Label the data: Manually or using automated tools, label each text sample with its corresponding sentiment, such as positive, negative, or neutral. This labeled dataset will serve as the training data for your model.

- Choose a machine learning algorithm: Select an appropriate machine learning algorithm for sentiment analysis, such as Naive Bayes, Support Vector Machines (SVM), or deep learning models like Recurrent Neural Networks (RNNs) or Convolutional Neural Networks (CNNs).

- Train the model: Using the labeled dataset, train the chosen machine learning algorithm to identify and classify the sentiment of new text samples.

- Evaluate and optimize the model: Evaluate the performance of the trained model using metrics like accuracy, precision, recall, and F1-score. If necessary, fine-tune the model's parameters or try different algorithms to improve its performance.

- Deploy and monitor the model: Once satisfied with the model's performance, deploy it into your marketing workflow to analyze new data and generate insights. Continuously monitor the model's performance and retrain it as needed to maintain accuracy.

Here are some additional tips for training a sentiment analysis model for marketing:

- Consider the context: Sentiment can vary depending on the context, so ensure your model is trained on data that reflects the specific context of your marketing efforts.

- Handle negation and sarcasm: Train your model to recognize negation and sarcasm, which can significantly alter the sentiment of a statement.

- Account for domain-specific language: If your marketing domain involves specific jargon or slang, ensure your model is trained to understand these terms and their sentiment implications.

- Use a variety of data sources: Gather data from multiple sources, such as social media, reviews, and surveys, to provide a comprehensive view of customer sentiment.

By following these steps and considering these tips, you can train a sentiment analysis model that provides valuable insights for your marketing campaigns and helps you make data-driven decisions to improve customer satisfaction and brand reputation.

Evaluation

Sentiment analysis plays a crucial role in evaluating marketing efforts by providing insights into customer opinions, brand perception, and campaign effectiveness. It helps marketers understand how their messages are resonating with the target audience and identify areas for improvement.

To effectively evaluate marketing using sentiment analysis, follow these steps:

- Define evaluation goals: Clearly define the objectives of your evaluation. Do you want to assess overall brand sentiment, measure campaign effectiveness, or identify customer pain points?

- Gather relevant data: Collect data from various sources, such as social media mentions, product reviews, customer surveys, and online forums. Ensure the data is relevant to your evaluation goals.

- Apply sentiment analysis: Utilize sentiment analysis tools or techniques to classify the collected data into positive, negative, or neutral sentiment categories

- Analyse sentiment trends: Track sentiment trends over time to identify changes in customer perception and evaluate the impact of marketing campaigns.

- Drill down into specific topics: Analyse sentiment for specific topics, such as product features, customer service, or pricing, to gain granular insights.

- Identify sentiment drivers: Determine the factors driving positive or negative sentiment, such as specific product attributes, customer service experiences, or marketing messages.

- Integrate with other metrics: Combine sentiment analysis with other marketing metrics, such as engagement rates, conversion rates, and brand awareness, to gain a comprehensive view of marketing performance.

- Communicate insights: Share sentiment analysis findings with relevant stakeholders, including marketing teams, product development teams, and customer service teams, to drive action and improvement.

Here are specific examples of how sentiment analysis can be used for marketing evaluation:

- Measuring campaign effectiveness: Analyse sentiment before, during, and after a marketing campaign to assess its impact on customer perception.

- Identifying customer pain points: Analyse sentiment in customer feedback to identify common issues or frustrations that need to be addressed.

- Tracking brand reputation: Monitor sentiment across social media and review platforms to gauge overall brand perception and identify potential reputational risks.

- Optimizing marketing messages: Analyse sentiment towards specific marketing messages to determine which resonate best with the target audience.

- Improving customer service: Analyse sentiment in customer service interactions to identify areas for improvement and enhance customer satisfaction.

By effectively utilizing sentiment analysis, marketers can gain valuable insights into customer opinions, evaluate the effectiveness of their campaigns, and make data-driven decisions to improve their overall marketing strategy.

Types of Sentiment Analysis

Various types of sentiment analysis can be performed, depending on the specific focus and objective of the analysis. Some common types include:

- **Document-Level Sentiment Analysis**: This type of analysis determines the overall sentiment expressed in a document, such as a review or an article. It aims to classify the entire text as positive, negative, or neutral.

- **Sentence-Level Sentiment Analysis:** Here, the sentiment of each sentence within a document is analyzed. This type provides a more granular understanding of the sentiment expressed in different text parts.

- **Aspect-Based Sentiment Analysis:** This approach focuses on identifying and extracting the sentiment associated with specific aspects or entities mentioned in the text. For example, in a product review, the sentiment towards different features of the product (e.g., performance, design, usability) can be analyzed separately.

- **Entity-Level Sentiment Analysis:** This type of analysis identifies the sentiment expressed towards specific entities or targets mentioned in the understand the sentiment associated with different entities within the same document.

- **Comparative Sentiment Analysis:** This approach involves comparing the sentiment between different entities or aspects mentioned in the text. It aims to identify the relative sentiment or preferences expressed towards various entities or features.

Sentiment Analysis Use Cases

We just saw how sentiment analysis can empower organizations with insights that can help them make data-driven decisions.

Social Media Monitoring for Brand Management: Brands can use sentiment analysis to gauge their Brand's public outlook.

Product/Service Analysis: Brands/Organizations can perform sentiment analysis on customer reviews to see how well a product or service is doing in the market and make future decisions accordingly.

Stock Price Prediction: Predicting whether the stocks of a company will go up or down is crucial for investors.

Program

SENTIMENT ANALYSIS FOR MARKETING

```
# Input: Customer review
review = input("Enter a customer review: ")

# Analyze sentiment
analysis = TextBlob(review)
sentiment = analysis.sentiment.polarity

# Output: Sentiment classification
if sentiment > 0:
    print("Positive sentiment")
elif sentiment < 0:
    print("Negative sentiment")
else:
    print("Neutral sentiment")
```

Output

Positive sentiment

Sentiment strength: 0.8

Code for Sentiment Analysis Using Vader:

```
from vaderSentiment.vaderSentiment import
SentimentIntensityAnalyzer

sentiment = SentimentIntensityAnalyzer()
```

```
text_1 = "The book was a perfect balance between wrtiting style and plot."
text_2 =  "The pizza tastes terrible."
sent_1 = sentiment.polarity_scores(text_1)
sent_2 = sentiment.polarity_scores(text_2)
print("Sentiment of text 1:", sent_1)
print("Sentiment of text 2:", sent_2)
```

Output

```
Sentiment of text 1: {'neg': 0.0, 'neu': 0.73, 'pos': 0.27, 'compound': 0.5719}
Sentiment of text 2: {'neg': 0.508, 'neu': 0.492, 'pos': 0.0, 'compound': -0.4767}
```

Code for Sentiment Analysis using Bag of Words Vectorization Approach:

```
#Loading the Dataset
import pandas as pd
data = pd.read_csv('Finance_data.csv')
#Pre-Prcoessing and Bag of Word Vectorization using Count Vectorizer
```

```
from sklearn.feature_extraction.text import CountVectorizer
from nltk.tokenize import RegexpTokenizer
token = RegexpTokenizer(r'[a-zA-Z0-9]+')
cv = CountVectorizer(stop_words='english',ngram_range = (1,1),tokenizer = token.tokenize)
text_counts = cv.fit_transform(data['sentences'])
#Splitting the data into trainig and testing
from sklearn.model_selection import train_test_split
X_train, X_test, Y_train, Y_test = train_test_split(text_counts, data['feedback'], test_size=0.25, random_state=5)
#Training the model
from sklearn.naive_bayes import MultinomialNB
MNB = MultinomialNB()
MNB.fit(X_train, Y_train)
#Caluclating the accuracy score of the model
from sklearn import metrics
predicted = MNB.predict(X_test)
accuracy_score = metrics.accuracy_score(predicted, Y_test)
print("Accuracuy Score: ",accuracy_score)
```

Output:

Accuracy Score: 0.9111675126903553

Code for Sentiment Analysis Using Transformer based models:

```
from transformers import pipeline
sentiment_pipeline = pipeline("sentiment-analysis")
data = ["It was the best of times.", "t was the worst of times."]
sentiment_pipeline(data)
```

Output:

[{'label': 'POSITIVE', 'score': 0.999457061290741}, {'label': 'NEGATIVE', 'score': 0.9987301230430603}]

Air lines using sentiment analysis

```
import pandas as pd
from textblob import TextBlob

# Input: Airline customer review
review = input("Enter an airline customer review: ")

# Analyze sentiment
analysis = TextBlob(review)
sentiment = analysis.sentiment.polarity

# Output: Sentiment classification
if sentiment > 0:
    print("Positive sentiment")
elif sentiment < 0:
    print("Negative sentiment")
else:
    print("Neutral sentiment")
```

Output:

 Sentiment strength
print("Sentiment strength:", abs(sentiment))

# Identify sentiment drivers
keywords = analysis.tags

print("Keywords:", keywords)

Training dataset

X_train=["JavaTpoint provides best tutorial for students",
"It is a great platform to start off your IT career",
Concepts are explained very well",
The articles have some interesting examples",
 "Some tutorials are bad",
"Their content can confuse students"]
y_train=[1,1,1,1,0,0] #1-Positive, 0 -Negative
X_train

Output

```
[84] X_train=["JavaTpoint provides best tutorials for students",
            "It is a great platform to start off your IT career",
           "Concepts are explained very well",
           "The articles have some interesting examples",
           "Some tutorials are bad",
          "Their content can confuse students"]

     y_train=[1,1,1,1,0,0] #1-Positive, 0 -Negative

     X_train

     ['JavaTpoint provides best tutorials for students',
      'It is a great platform to start off your IT career',
      'Concepts are explained very well',
      'The articles have some interesting examples',
      'Some tutorials are bad',
      'Their content can confuse students']
```

In the given function, we are performing tokenization and stop word removal at the same time.

```
def getCleanedText(text):
 texttext=text.lower()
 #tokenize
 tokens=tokenizer.tokenize(text)
new_tokens=[token for token in tokens if token not in en_stop
words]
stemmed_tokens=[ps.stem(tokens) for tokens in new_tokens]

 clean_text=" ".join(stemmed_tokens)
 return clean_text
```

Output

```
[90] def getCleanedText(text):
         text=text.lower()

         #tokenize
         tokens=tokenizer.tokenize(text)
         new_tokens=[token for token in tokens if token not in en_stopwords]
         stemmed_tokens=[ps.stem(tokens) for tokens in new_tokens]
         clean_text=" ".join(stemmed_tokens)
         return clean_text
```

Advantages of Sentiment Analysis

- product review monitoring – monitoring which of your products receive a higher rate of positive comments.
- market research – discovering attitudes of internet users toward the research target.
- search engines/recommender systems – enhancing performance by better understanding what users meant by the content of a query.

Conclusion

Sentiment analysis can be a very useful tool for user response monitoring. Its most significant advantage is the introduction of the possibility to use direct user feedback with minimal human supervision while still being able to scale up easily.