Sentiment Analysis

By Jeff Kiarie

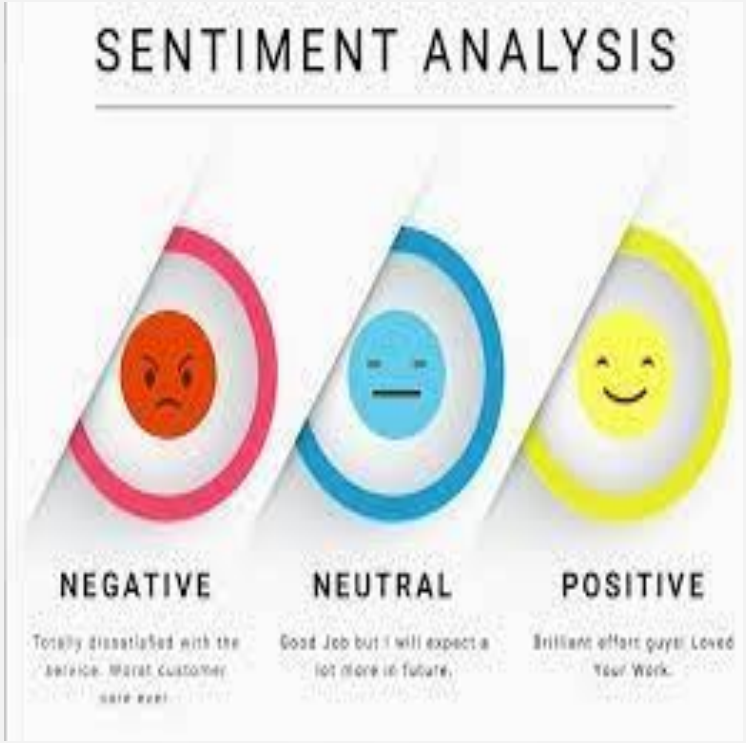
Overview

1. Introduction
2. What is Sentiment Analysis?
3. Business Understanding
4. Data Understanding and Data Preparation
5. Modeling and Evaluation
6. Recommendations
7. Next Steps

Introduction

Welcome to this presentation on Sentiment Analysis,a powerful tool for understanding the emotions and opinions hidden within text data

Today we will explore the world of sentiment analysis,its significance and its wide-ranging application across various industries.



What is Sentiment Analysis?

Sentiment analysis also known as opinion mining is a natural language processing
technique used to determine the emotional tone or subjective opinion expressed in a piece of text such as a review comment or an news article.

It involves analyzing the text to understand whether the expressed sentiment is positive,negative or neutral.

Ways Sentiment Analysis can be used

1.Customer Feedback and Satisfaction Analysis

2.Brand Monitoring and Reputation Management

3.Product and Service Development

4.Social Media Marketing

5.Competitor Analysis

Business Understanding

In today's hyperconnected world, social media platforms like Twitter have become powerful channels for consumers to express their thoughts and emotions about products and brands. For tech giants like Apple and Google, understanding the sentiment and emotions expressed in tweets related to their products is of paramount importance.

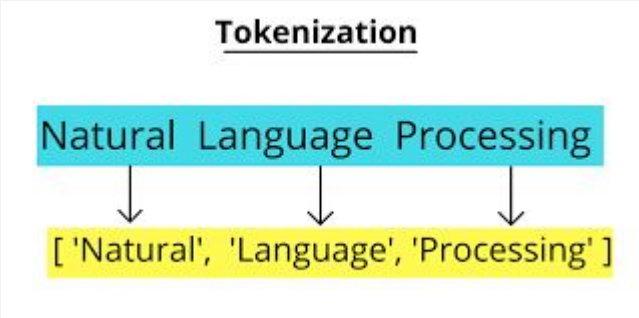
This Twitter sentiment analysis project aims to provide valuable insights into the online sentiment surrounding Apple and Google products.

Data Understanding and Data Preparation

To discern the sentiment of a tweet, we can train a model using a dataset of tweets alongside their associated sentiments. This training process equips the model with the ability to categorize a tweet into one of three categories: positive, negative, or neutral

Luckily,we've been allocated such a dataset that can be found.here.The dataset comprises three columns, representing tweets, the targeted brand, and associated emotions, totaling 8,721 entries.But before training the model,the dataset should be processed and here are the steps followed:

1.Tokenization: Split the text into individual words or tokens. Tokenization is a crucial step to convert text into a format suitable for analysis.

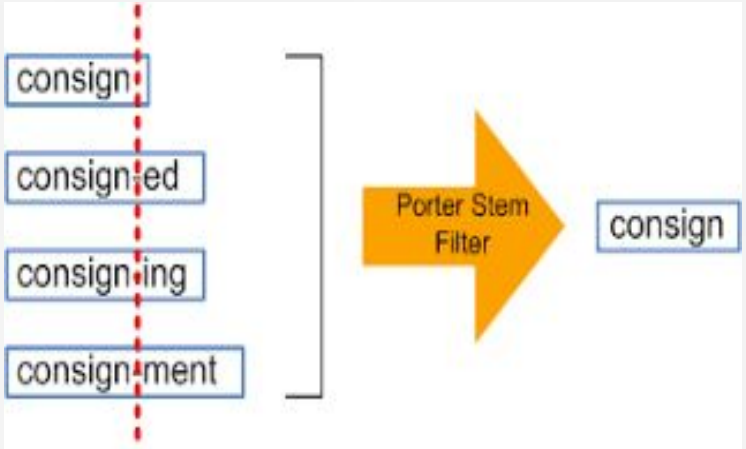


2.Data Cleaning:

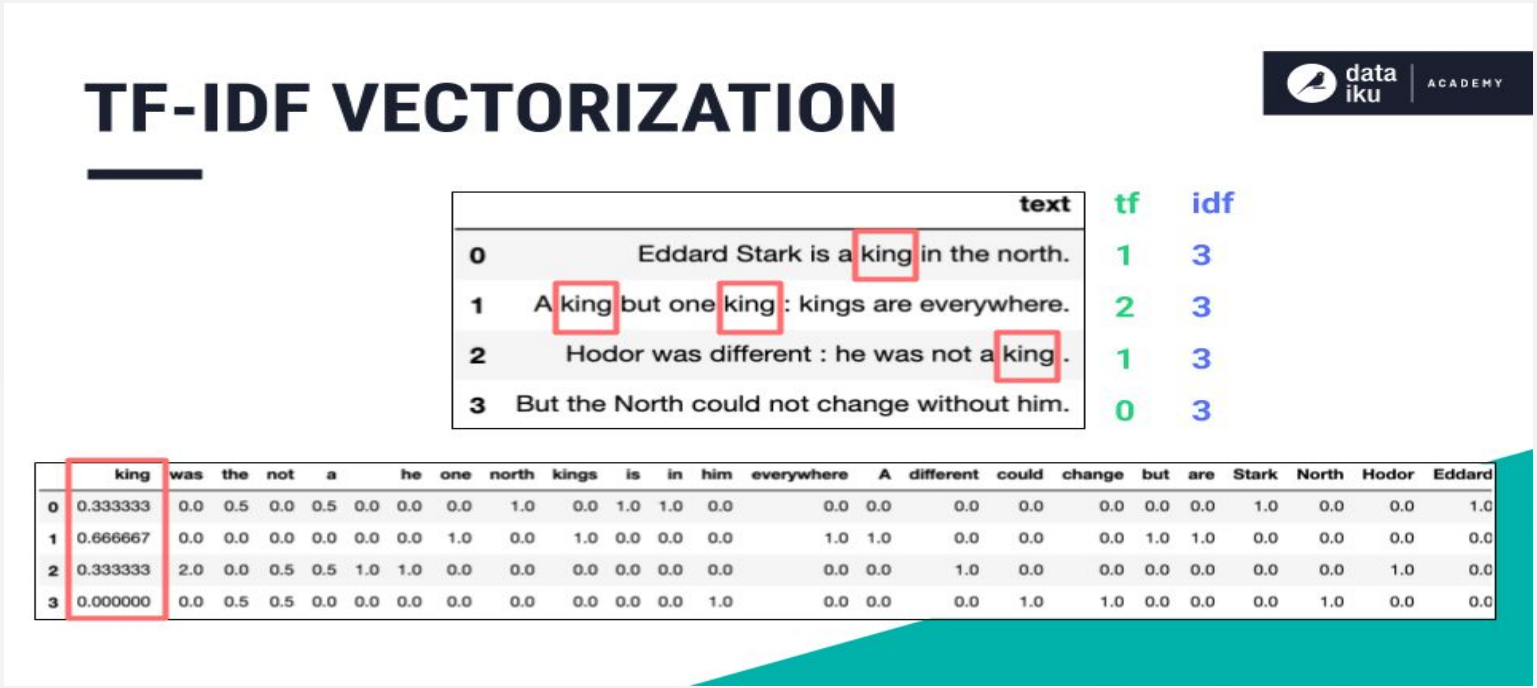
- Text Cleaning: Remove any unnecessary characters, special symbols, or formatting artifacts.
- Lowercasing: Convert all text to lowercase to ensure uniformity.
- Handling Missing Data: Address any missing or null values in the text data appropriately.

3.Stop Word Removal: Remove common words (stop words) like "and," "the," "in," etc., that do not carry significant meaning for your analysis. This helps reduce noise in the data.

4.Stemming or Lemmatization: Reduce words to their base or root form to normalize the text. This step helps in grouping similar words together.



5.Text Vectorization: Convert the tokenized text data into numerical vectors suitable for feeding into models.



	king	was	the	not	a		he	one	north	kings	is	in	him	everywhere	A	different	could	change	but	are	Stark	North	Hodor	Eddard
0	0.333333	0.0	0.5	0.0	0.5	0.0	0.0	0.0	1.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0
1	0.666667	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	1.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0
2	0.333333	2.0	0.0	0.5	0.5	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0
3	0.000000	0.0	0.5	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0

6.Data Splitting: Split the preprocessed data into training, validation, and test sets, ensuring that they are representative of the overall dataset.

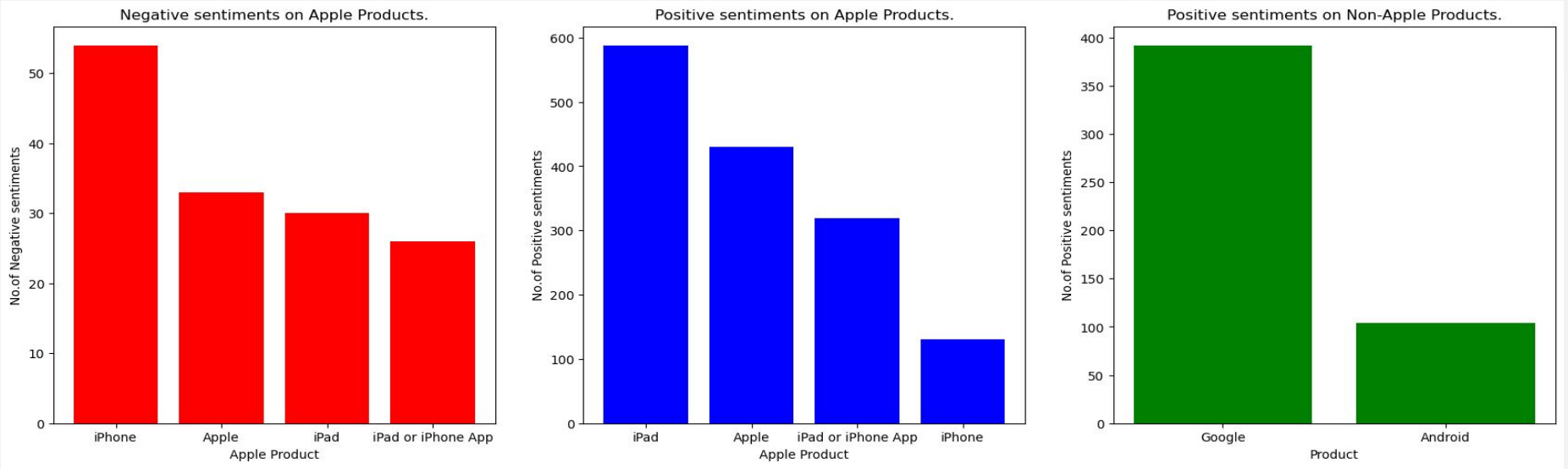
Modeling and Evaluation

In modeling the models tries to learn the patterns and relationship within the dataset.In our project we used sklearn’s Logistic Regression and Support Vector Machine which are models for multiclass classification.We also built a neural network that could be used likewise.

To assess how well the models fits the data and generalize to new data,we used evaluation metrics such as accuracy and cross validation score.The model achieved a cross validation score of roughly 68% which is fair considering we do not want it to overfit.

Recommendations

When we predicted the sentiments of various tweets using our model,we saw that the more negative comments toward Apple were regarding iphones.Google had their fair share in comments both positive and negative but they were in low quantities.



With that here are some of the recommendations we could make:

1.Investigate the common issues or concerns raised by users of iPhones in negative reviews and take proactive steps to address them.

2.Consider expanding the iPad product line or introducing new versions to maintain positive customer sentiment.

3.Leverage the favorable feedback received by Google products and services to bolster brand reputation and customer loyalty.

4.Understand the factors contributing to Apple's overall higher positive reviews compared to Google and utilize these insights to maintain a positive image.

Next Steps

1.Implement Sentiment Analysis in Your Organization:Consider how sentiment analysis can be integrated into your business processes, such as customer support, marketing, or product development.

2.Explore Multimodal Sentiment Analysis:Consider incorporating other data types like images, videos, and audio for a more comprehensive understanding of sentiment

3.Stay Updated with NLP Advances: - Continuously keep up with the latest developments in NLP and sentiment analysis to leverage new techniques and technologies.

Thank you

Any Questions?

Just as knowledge makes human intelligent,data makes software intelligent.

Amarpreet kalkat