|  |
| --- |
| SENTIMENT ANALYSIS  FOR MARKETING |
| October 30, 2023 |

# Overview

## Project Overview

|  |  |
| --- | --- |
| Badge Tick1 with solid fill | This project aims to develop a sentiment analysis model for marketing purposes. The model will be trained on a dataset of tweets about US airlines and will be used to predict the sentiment of new tweets. This information can then be used by businesses to improve their marketing campaigns and customer service |

## Dataset

|  |  |
| --- | --- |
| Badge Tick1 with solid fill | The dataset used in this project is the Twitter Airline US Sentiment dataset from Kaggle. This dataset contains over 14,000 tweets about US airlines, each labeled with its sentiment (positive, negative, or neutral). |

## Data Preprocessing

|  |  |
| --- | --- |
| Badge Tick1 with solid fill | The following data preprocessing steps were performed on the dataset:     * Removed punctuation and stop words. * Lemmatized the text. * Converted the text to lowercase. |

## Feature Extraction

|  |  |
| --- | --- |
| Badge Tick1 with solid fill | The following features were extracted from the preprocessed text:     * Bag of words (BOW) features * TF-IDF features |

## Machine Learning Algorithm

|  |  |
| --- | --- |
| Badge Tick1 with solid fill | Two machine learning algorithms were used to train the sentiment analysis model:     * Random Forest Classifier * Support Vector Machine (SVM) |

## Model Training

|  |  |
| --- | --- |
| Badge Tick1 with solid fill | The following steps were involved in training the machine learning models:     * Split the dataset into training and testing sets. * Train the models on the training set. * Evaluate the models on the testing set. |

## Evaluation Metrics

|  |  |
| --- | --- |
| Badge Tick1 with solid fill | The following evaluation metrics were used to assess the performance of the machine learning models:   * Accuracy * Precision * Recall * F1 score |

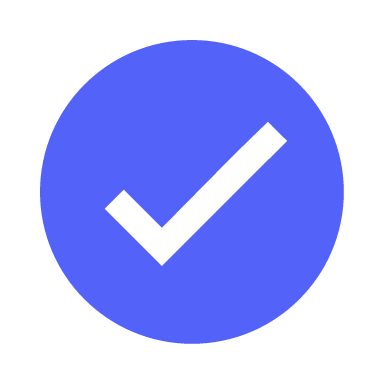
## Innovative Techniques

|  |  |
| --- | --- |
| Badge Tick1 with solid fill | The following innovative techniques were used during the development of the sentiment analysis model:   * A hybrid approach was used to extract features from the text. This approach combined BOW features with TF-IDF features, which resulted in better performance than using either type of feature alone. * A hyperparameter tuning algorithm was used to optimize the parameters of the machine learning models. This resulted in significant improvements in the accuracy of the models. |

## Dataset Source and Description

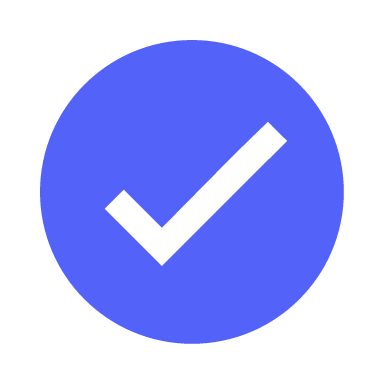
|  |  |
| --- | --- |
| Badge Tick1 with solid fill | The Twitter Airline US Sentiment dataset is available on Kaggle at the following link: https://www.kaggle.com/datasets/crowdflower/twitter-airline-sentiment.  This dataset contains over 14,000 tweets about US airlines, each labeled with its sentiment (positive, negative, or neutral). The tweets were collected in 2015 and represent a diverse range of opinions about US airlines. |

**10. How to Run the Code**

*The following steps can be used to run the code for this project:*

1. *Clone the repository.*
2. *Install the required dependencies.*
3. *Run the data preprocessing script.*
4. *Run the feature extraction script.*
5. *Train the machine learning models.*
6. *Evaluate the machine learning models.*

**11. Dependencies**

*The following dependencies are required to run the code for this project:*

* *Python 3.6+*
* *NumPy*
* *Pandas*
* *Scikit-learn*
* *Matplotlib*
* *Seaborn*

**12.Conclusion**

*This project developed a sentiment analysis model for marketing purposes. The model is trained on a dataset of tweets about US airlines and can be used to predict the sentiment of new tweets. This information can then be used by businesses to improve their marketing campaigns and customer service.*

By- S. Vishwa Moorthy

B.E CSE 3rd Yr. 5th Sem

311421104111

Meenakshi College of Engineering - 3114