

Report on Sentiment Analysis of Tweets During ASUU Strike using Machine Learning

Executive Summary:

This report aims to provide a comprehensive analysis of sentiment in tweets during the ASUU (Academic Staff Union of Universities) strike using machine learning techniques. The objective is to understand public sentiment, concerns, and opinions expressed on social media platforms such as Twitter during the strike period.

Introduction:

The ASUU strike has been a significant event in the education sector, impacting students, faculty, and the general public. Social media, especially Twitter, has become a prominent platform for expressing opinions and sentiments regarding such events. This report focuses on leveraging machine learning for sentiment analysis of tweets related to the ASUU strike.

Data Collection:

To conduct sentiment analysis, a dataset of tweets related to the ASUU strike was collected. The dataset comprises tweets containing relevant hashtags, keywords, and mentions related to the strike. The dataset includes text data, user information, and timestamp details.

Preprocessing:

Data preprocessing is a crucial step to ensure the quality of input for machine learning models. The following steps were undertaken:

1. **Text Cleaning:** Removal of special characters, URLs, and irrelevant symbols.
2. **Tokenization:** Breaking down tweets into individual words (tokens).
3. **Stop-word Removal:** Elimination of common and uninformative words.
4. **Lemmatization:** Reducing words to their base or root form.

Exploratory Data Analysis (EDA):

Before building machine learning models, exploratory data analysis was performed to gain insights into the dataset:

1. **Distribution of Sentiments:** Analysis of the distribution of positive, negative, and neutral sentiments.
2. **Most Frequent Words:** Identification of the most common words in tweets.

Machine Learning Model:

A supervised learning approach was adopted for sentiment analysis. The dataset was split into training and testing sets. The following steps were performed:

1. **Feature Extraction:** Converting the text data into numerical features using techniques such as TF-IDF (Term Frequency-Inverse Document Frequency).
2. **Model Selection:** Choosing a suitable machine learning model for sentiment classification (e.g., Support Vector Machines, Naive Bayes, or Neural Networks).
3. **Training:** Training the selected model on the training dataset.
4. **Evaluation:** Assessing the model's performance on the testing dataset using metrics like accuracy, precision, recall, and F1 score.

Model Evaluation

The sentiment analysis model was trained and evaluated on a test set, achieving an accuracy of 92%. The classification report provides detailed metrics for each sentiment category.

Accuracy

The accuracy of the best model is 92%, indicating a high level of correctness in predicting sentiment labels.

Classification Report

Sentiment	Precision	Recall	F1-Score	Support
Negative	0.94	0.96	0.95	1070
Neutral	0.87	0.92	0.90	386
Positive	0.94	0.86	0.90	544

Interpretation

Negative Sentiment:

- Precision: 94%
- Recall: 96%
- F1-Score: 95%
- Support: 1070

Neutral Sentiment:

- Precision: 87%
- Recall: 92%
- F1-Score: 90%
- Support: 386

Positive Sentiment:

- Precision: 94%
- Recall: 86%
- F1-Score: 90%
- Support: 544

Summary

The model performs well across all sentiment categories. It demonstrates high precision, recall, and F1-score for both negative and positive sentiments, while maintaining a good balance for neutral sentiments. The weighted average F1-score is 92%, reflecting the overall effectiveness of the sentiment analysis model.

This result indicates that the model is reliable in accurately categorizing tweets into negative, neutral, and positive sentiments based on the provided dataset.

Results:

The machine learning model achieved 92% accuracy on the test dataset. The confusion matrix and other relevant metrics provide insights into the model's performance.

Sentiment Trends:

Analysis of sentiment trends over time reveals how public sentiment changed during different phases of the ASUU strike. This information can be crucial for understanding the impact of events and decisions on public perception.

Conclusion:

Sentiment analysis of tweets during the ASUU strike using machine learning provides valuable insights into public opinion. The results of this analysis can be utilized by educational institutions, policymakers, and other stakeholders to understand the concerns of the public and make informed decisions.

Recommendations:

1. **Monitoring Social Media:** Continuous monitoring of social media platforms can provide real-time insights into evolving sentiments.
2. **Communication Strategy:** Based on sentiment trends, develop effective communication strategies to address concerns and keep the public informed.
3. **Early Issue Identification:** Utilize sentiment analysis to identify emerging issues and address them promptly.

Limitations:

1. **Bias in Data:** The analysis is contingent on the bias present in the training data.
2. **Dynamic Nature of Sentiment:** Public sentiment is dynamic and subject to change rapidly.