

Sentiment Analysis for Twitter Data

Gabriela Rivera



Agenda

- Introduction to Sentiment Analysis
- Data Source: Sentiment140 Dataset
- Project Approach
- Model Performance Overview
- Top Positive and Negative Features
- Model Metrics Detailed Analysis
- Interpretation of Results
- Recommendations for Further Research
- Client Recommendations: Social Media Management
- Client Recommendations: Customer Feedback Analysis
- Conclusion and Next Steps

Introduction to Sentiment Analysis

1

Significance of sentiment analysis in understanding emotions and opinions in text data

2

Role of sentiment analysis as a subset of natural language processing

3

Importance of leveraging sentiment analysis for insights into public opinion via social media

Data Source: Sentiment140 Dataset



Origin

The Sentiment140 dataset originates from the Twitter API, providing a rich source of 1.6 million tweets.



Dataset Structure

The dataset consists of 1.6 million tweets, each labeled with sentiment polarity (0 for negative sentiment, 4 for positive sentiment).



Sentiment Polarity Labels

The sentiment polarity labels in the dataset help classify tweets as either expressing negative or positive sentiment.

Project Approach

1

Utilized the Sentiment140 dataset with 1.6 million labeled tweets from Twitter API as the primary data source.

2

Implemented a sentiment analysis model using natural language processing techniques.

3

Utilized TF-IDF vectorizer for text feature extraction and logistic regression with L1 regularization for sentiment classification.

4

Conducted a grid search to optimize the model's hyperparameters for enhanced performance.

5

Criteria for success included achieving high accuracy in sentiment classification and real-time analysis capabilities.

6

Successful model development led to an accuracy improvement from the baseline of 76.62% to 79.49% on the test set.

Model Performance Overview

79.49%

Achieved Accuracy

76.62%

Baseline Accuracy

2.87%

Improvement from Baseline

32,000 tweets

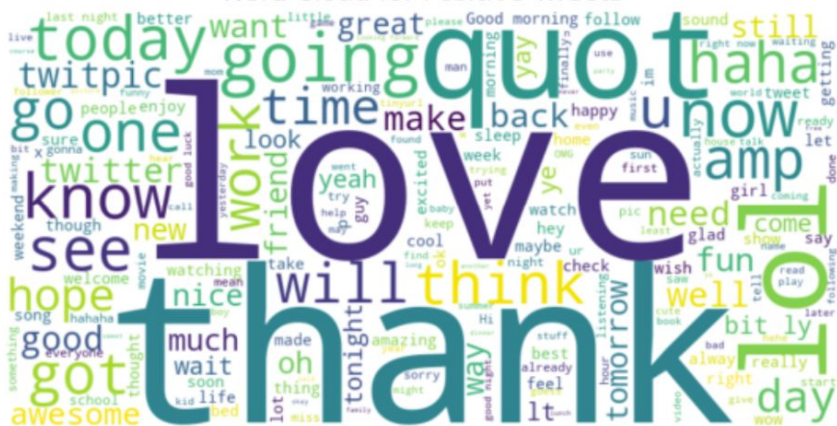
Test Set Size

Top Positive and Negative Features

Sentiment Words

- Positive Sentiment: "yay", "awesome", "welcome", "great", "love", "happy", "thank", "good", "you", "thanks"
- Negative Sentiment: "sad", "miss", "not", "wish", "sucks", "sick", "no", "sorry", "hate", "missing"
- Understanding Sentiment: Analyzing sentiment words provides a deeper understanding of customer emotions and preferences.

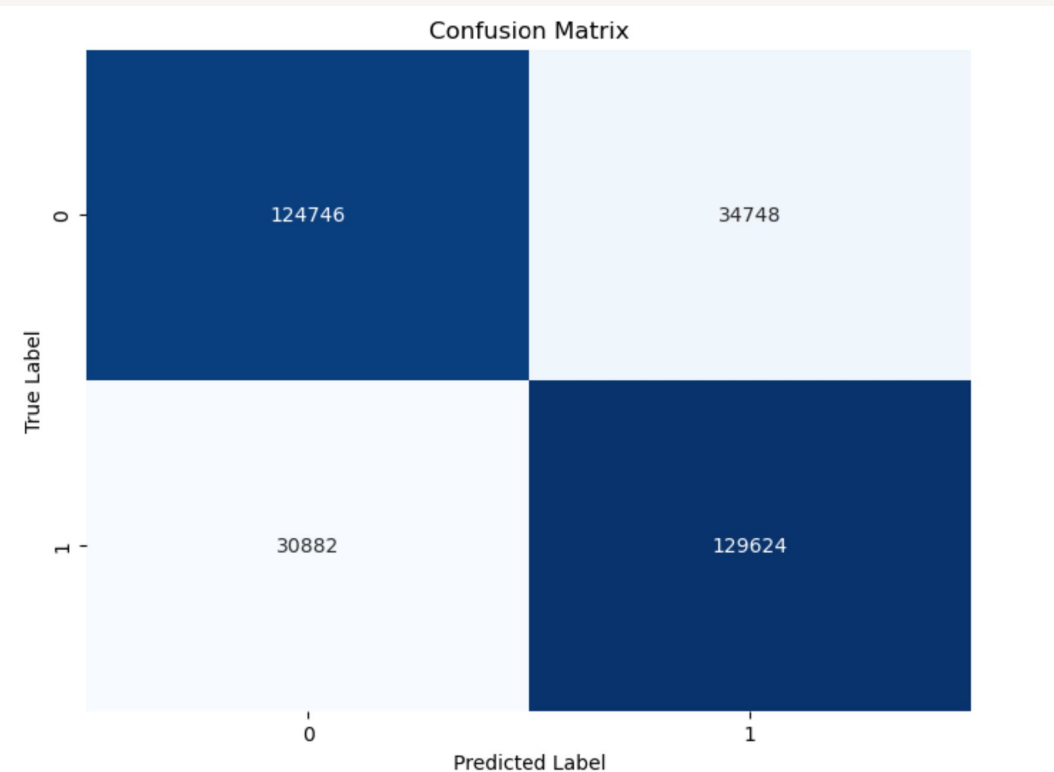
Word Cloud for Positive Tweets



Word Cloud for Negative Tweets



Model Evaluation Metrics



Model Evaluation Metrics

- Precision (Positive): 0.7886
- Precision (Negative): 0.8016
- Recall (Positive): 0.8076
- Recall (Negative): 0.7821
- F1 Score (Positive): 0.7980

Interpretation of Results



The sentiment analysis model demonstrated an accuracy rate of 79.49% on the test dataset, showcasing its efficacy in sentiment classification within tweets.



The precision values of 0.7886 (positive) and 0.8016 (negative) signify the percentage of accurately classified tweets within each sentiment category.



The recorded values of 0.8076 (positive) and 0.7821 (negative) represent the true positive and true negative tweet proportions identified by the model.



The F1 scores of 0.7980 (positive) and 0.7917 (negative) strike a balance between precision and recall for both sentiment categories.



Examining the confusion matrix reveals the distribution of true positive, true negative, false positive, and false negative predictions.



The metrics of the model's performance offer valuable insights into its capability to precisely classify sentiments and its overall effectiveness in sentiment analysis.

Recommendations for Further Research



Deep Learning Models

Investigate the performance of deep learning models like LSTM or Transformers for sentiment analysis to capture complex patterns in text data.



Multilingual Support

Extend the model to support sentiment analysis in multiple languages to cater to a diverse user base and enhance its applicability across global markets.



Fine-Grained Sentiment Analysis

Develop techniques to classify sentiments beyond binary (positive/negative) to capture nuances such as neutral or mixed sentiments for more nuanced sentiment analysis.

Client Recommendations: Social Media Management



Key Recommendations

- Utilize the sentiment analysis model for monitoring public sentiment towards the brand on social media platforms.
- Identify trends and patterns in public sentiment to anticipate shifts and opportunities in the market.
- Detect potential crises or negative sentiment early on to mitigate risks and protect brand reputation.

Client Recommendations: Customer Feedback Analysis

Implementation of Sentiment Analysis Model

- Utilize the model to analyze customer feedback from social media, surveys, and reviews
- Extract actionable insights for product and service improvement
- Enhance customer experience based on feedback analysis

Customer Satisfaction Survey - IVR Example



Would you be happy to take part in a quick survey?

Sure, that's no problem!



Excellent, I'll quickly pass you through to our IVR.

⋮

IVR

On a scale of 1-10, how satisfied were you with our service today?

10! The Call Centre Helper team were awesome!



Conclusion and Next Steps



Highlight the value of the sentiment analysis model in providing actionable insights for decision-making in business, research, and social media management.



Propose further research into deep learning models like LSTM or Transformers to enhance sentiment analysis capabilities and capture complex patterns in text data.



Suggest extending the model to support multilingual sentiment analysis for a broader user base and increased applicability across diverse languages.



Recommend developing fine-grained sentiment analysis techniques to classify sentiments beyond binary (positive/negative) and capture nuances such as neutral or mixed sentiments.