

Tweet Tracker Project

Comprehensive Analysis using VADER and TextBlob

Project Overview

Dataset:

Sentiment140 (1.6M tweets)

Methods:

VADER & TextBlob Sentiment Analysis

Analysis:

Comparative Performance Study

Agreement Rate:

66.85%

Processing:

Multi-stage text preprocessing pipeline

Project Overview & Dataset

This project implements a comprehensive sentiment analysis system for Twitter data using **VADER** and **TextBlob**. The system processes tweets from the Sentiment140 dataset, applying advanced preprocessing techniques and comparing the performance of both sentiment analysis approaches.

Key Objectives

Classify tweets into positive, negative, or neutral sentiments

Compare effectiveness of VADER and TextBlob analyzers

Analyze agreement patterns between different tools

Visualize sentiment distributions and word patterns

Dataset Information

1.6M

TOTAL TWEETS

CSV

FORMAT

Latin-1

ENCODING

Source: Sentiment140 Dataset from Kaggle

Structure: Contains columns for target, id, date, flag, user, and text

Quality: Pre-labeled dataset suitable for supervised learning approaches

Installation & Setup

```
# Core libraries
pip install pandas numpy matplotlib seaborn

# Text processing and sentiment analysis
pip install nltk vaderSentiment textblob wordcloud snsrape
```

Data Preprocessing Pipeline

The preprocessing pipeline ensures high-quality input through multiple cleaning stages:

```
def clean_tweet(text):
    """Comprehensive text cleaning function"""
    t = str(text).lower()
    t = re.sub(r"http\S+|www\S+|https\S+", "", t)
    t = re.sub(r"@w+", "", t)
    t = re.sub(r"#", "", t)
    t = re.sub(r"^\w\s", "", t)
    return t.strip()
```

Sentiment Analysis Methods

VADER Analysis

VADER is designed for social media text, excelling at handling emoticons, slang, capitalization emphasis, and punctuation-based sentiment intensification.

VADER Advantages

- Social media optimized
- Handles emoticons and slang
- Considers capitalization emphasis
- Compound score: -1 to +1

TextBlob Analysis

TextBlob provides polarity and subjectivity scores, offering a nuanced view of sentiment with both orientation and objectivity measures.

TextBlob Features

- **Polarity:** -1 to +1 scale
- **Subjectivity:** 0 to 1 scale
- Simple, intuitive API
- Versatile text processing

Implementation Code

```
# VADER
analyzer = SentimentIntensityAnalyzer()

def get_label(score):
    if score >= 0.05:
        return "Positive"
    elif score <= -0.05:
        return "Negative"
    else:
        return "Neutral"

# TextBlob
def analyze_sentiment(text):
    blob = TextBlob(str(text))
    return pd.Series({
        'polarity': blob.sentiment.polarity,
        'subjectivity': blob.sentiment.subjectivity
    })
```

Classification Criteria

- **Positive:** Score ≥ 0.05
- **Negative:** Score ≤ -0.05
- **Neutral:** $-0.05 < \text{Score} < 0.05$

Results & Analysis

66.85%

AGREEMENT RATE

33.15%

DISAGREEMENT RATE

1.6M

TWEETS ANALYZED

Sentiment Distribution

| Model | Positive | Negative | Neutral |
|----------|----------|----------|---------|
| VADER | 48.26% | 24.39% | 27.35% |
| TextBlob | 41.20% | 18.82% | 40.00% |

Disagreement Examples

Case 1: "my whole body feels itchy and like its on fire"

- TextBlob: Positive | VADER: Negative

Case 2: "Need a hug"

- TextBlob: Neutral | VADER: Positive

Word Pattern Analysis

Common Indicators

Positive: love, lol, haha, good, great, awesome, happy, thanks

Negative: hate, bad, upset, sad, angry, terrible, disappointed

Key Insights & Future Work

Applications

- Brand sentiment tracking
- Customer feedback analysis
- Crisis monitoring

Limitations

- Only English tweets
- No sarcasm detection
- Only 3 sentiment categories

Future Enhancements

Technical

- BERT, RoBERTa integration
- Streaming support

Analytical

- Emotion classification
- Aspect-based sentiment

Conclusion

This project demonstrates the complementary power of VADER and TextBlob in analyzing social media sentiment. With a 66.85% agreement rate, it lays a solid foundation for more advanced, multilingual, and deep learning-based models in the future.