# SENTIMENT ANALYSIS MODEL REPORT

Comprehensive Analysis and Recommendations

Generated on: April 26, 2025 09:32

# Table of Contents

# 1. Executive Summary

This report presents a comprehensive analysis of our sentiment analysis model, including dataset characteristics, model performance metrics, and expert recommendations for improvement.

# 2. Dataset Overview

## 2.1 Dataset Statistics

| Metric | Value |
|---|---|
| Total samples | (1599999, 6) |
| Number of unique users | 659775 |
| Date range | Fri Apr 17 20:30:31 PDT 2009 to Wed May 27 07:27:38 |

# 3. Exploratory Data Analysis

## 3.1 Sentiment Distribution

| Sentiment | Count |
|---|---|
| Sentiment 4 | 800000 |
| Sentiment 0 | 799999 |

## 3.2 Text Characteristics

| Characteristic | Value |
|---|---|
| Average text length | 74.09 characters |

# 4. Model Performance Analysis

## 4.1 Overall Performance

| Metric | Value |
|---|---|
| Model Accuracy | 0.3395 |

## 4.2 Detailed Classification Report

```
              precision    recall  f1-score   support

           0       0.75      0.68      0.71    799999
           1       0.00      0.00      0.00         0
           4       0.00      0.00      0.00    800000

    accuracy                           0.34   1599999
   macro avg       0.25      0.23      0.24   1599999
weighted avg       0.38      0.34      0.36   1599999
```

# 5. Expert Recommendations

## 1. DATASET ANALYSIS

**Dataset Characteristics:**

The dataset consists of 1,599,999 samples with 6 features.

**There are two sentiment classes:**
0 and 4, with an equal distribution of 799,999 samples each.

The average text length is 74.09 characters.

There are 659,775 unique users in the dataset.

The data spans from April 17, 2009, to May 27, 2009.

**Data Quality and Biases:**

The dataset seems to be well

distributed between the two sentiment classes.

There are no missing sentiment labels (classes 0 and 4), but it seems there might be an issue with class 1 having 0 samples, potentially impacting model performance.

The dataset's text data quality is not assessed here; potential biases could exist in the sentiment labeling process.

**Notable Patterns:**

There are no specific notable patterns mentioned in the dataset analysis.

## 2. MODEL PERFORMANCE EVALUATION

**Accuracy:**

The model has an accuracy of 0.3395, which is relatively low.

Precision, Recall, F1

**score:**

Precision for class 0 is 0.75, for class 1 is 0.00, and for class 4 is 0.00.

Recall for class 0 is 0.68, for class 1 is 0.00, and for class 4 is 0.00.

F1

score for class 0 is 0.71, for class 1 is 0.00, and for class 4 is 0.00.

**Strengths and Weaknesses:**

**Strengths:**

The model shows relatively good precision for class 0.

**Weaknesses:**

The model performs poorly on classes 1 and 4 due to a lack of samples for class 1, leading to low recall and F1

scores.

## 3. TECHNICAL RECOMMENDATIONS

**Model Improvements:**

Address the issue of class 1 having 0 samples to improve model performance.

Explore more complex models that can capture non

linear relationships better.

**Feature Engineering:**

Utilize techniques such as TF

IDF, word embeddings (Word2Vec, GloVe), or BERT embeddings to encode text features more effectively.

Extract additional features like sentiment scores, part

of

speech tags, or named entities to enhance model understanding.

**Data Preprocessing:**

Perform text cleaning steps like lowercasing, removing special characters, and tokenization.

Implement techniques like lemmatization or stemming to normalize text data.

## 4. BUSINESS IMPLICATIONS

**Practical Implications:**

Improved sentiment analysis models can enhance customer feedback analysis, brand perception monitoring, and market trend prediction.

**Use Cases:**

Sentiment analysis for social media monitoring.

Customer feedback analysis for product/service improvement.

Brand reputation management through sentiment tracking.

**Limitations/Risks:**

The current model's low performance may lead to incorrect sentiment analysis results.

Biases in the dataset or model may affect decision

making based on sentiment analysis results.

## 5. ACTION ITEMS  1. **Address Class Imbalance**

**Impact:**

Improved model performance and more accurate sentiment analysis.

**Effort:**

Moderate effort required to collect more samples for class 1.

**Metric:**

Track improvements in precision, recall, and F1

score for class 1.

2. Enhance Feature Engineering

**Impact:**

Better representation of text data and improved model understanding.

**Effort:**

Moderate to high effort depending on the complexity of feature engineering techniques.

**Metric:**

Monitor changes in model performance metrics after feature engineering.

3. Optimize Model Architecture

**Impact:**

Increased model performance and better sentiment prediction.

**Effort:**

Moderate effort to experiment with different architectures.

**Metric:**

Track changes in accuracy, precision, recall, and F1

score.

4. Improve Data Preprocessing

**Impact:**

Cleaner text data leading to better model performance.

**Effort:**

Low to moderate effort depending on the preprocessing techniques used.

**Metric:**

Monitor changes in model performance post

preprocessing.

5. Explore Additional Features

**Impact:**

Enriched feature space for better sentiment analysis.

**Effort:**

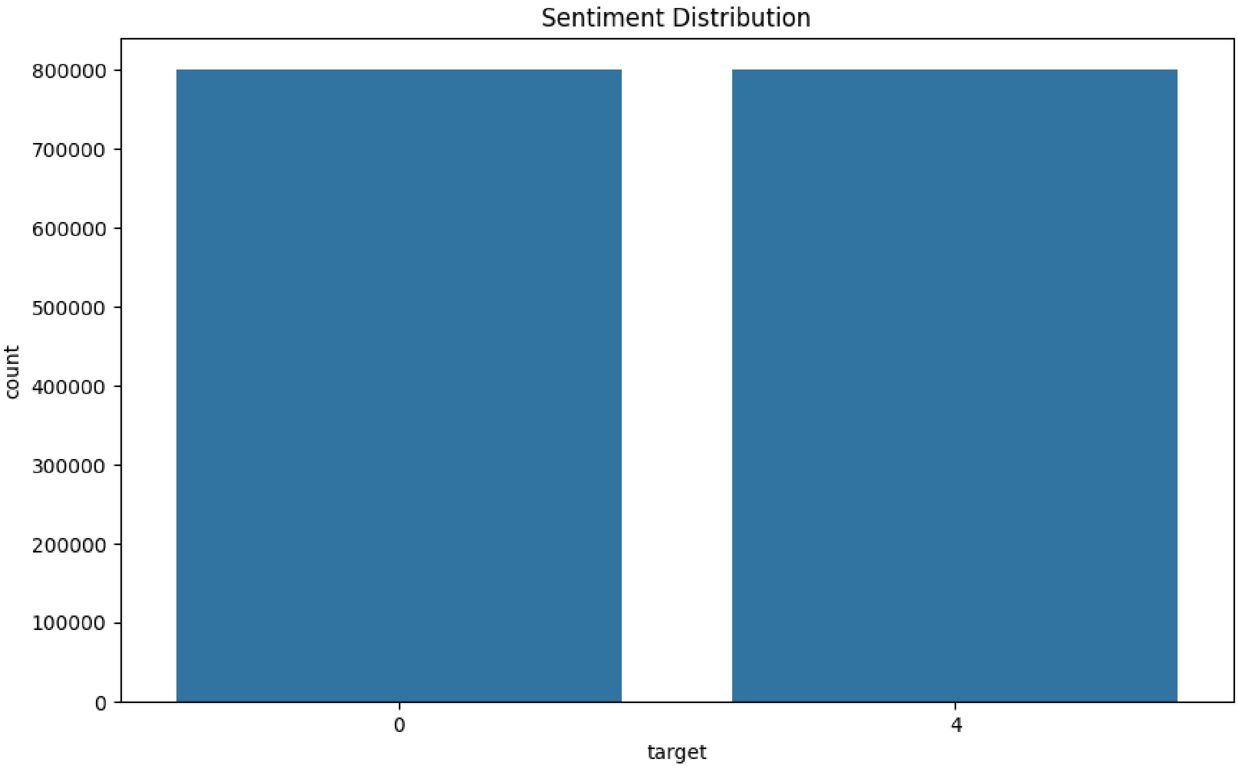
Moderate effort to identify and incorporate relevant additional features.

**Metric:**

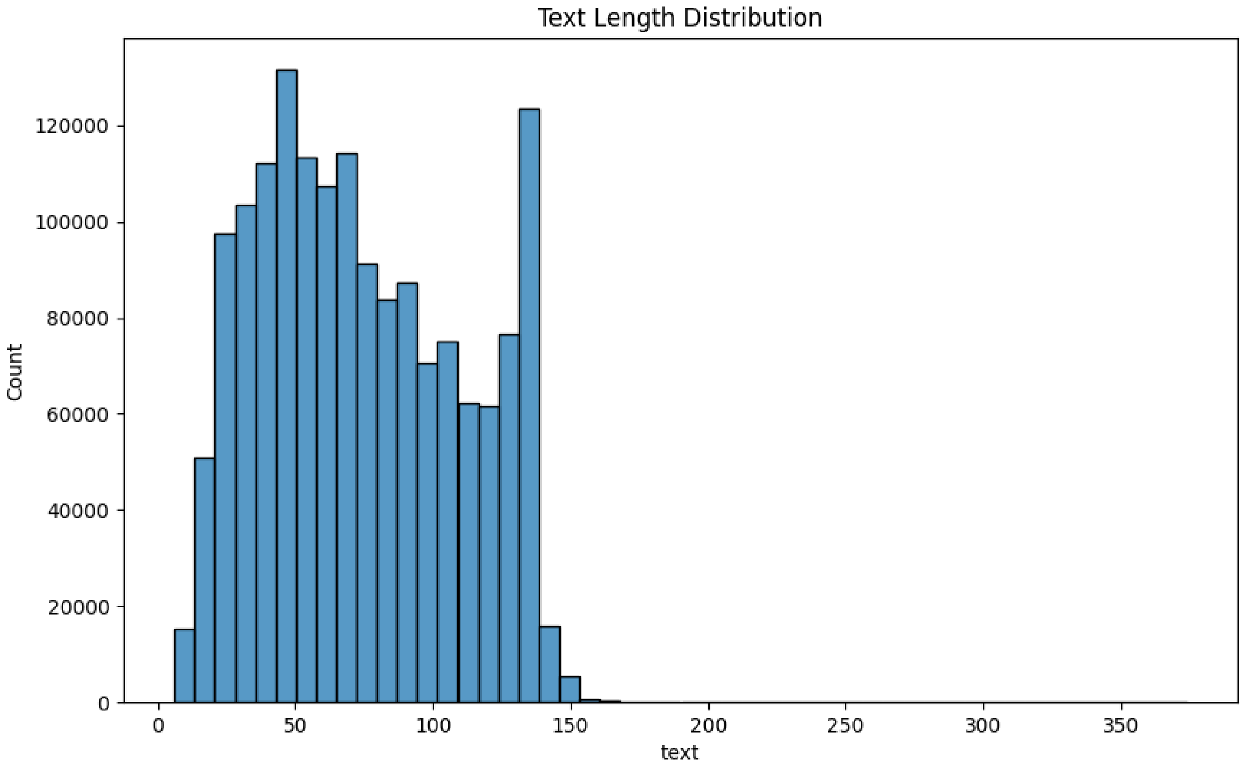
Evaluate the impact of new features on model performance.

# 6. Visualizations

## Sentiment Distribution Analysis




## Text Length Distribution




## User Activity Distribution