Sentiment Analysis of Twitter Data for Brand Monitoring

Introduction

Objective:

- Utilize sentiment analysis and entity recognition to monitor and analyze public sentiments towards brands using Twitter data.

Purpose:

- Help brands understand public opinion trends and adapt strategies accordingly.

Motivation

Why Twitter?

Twitter generates a vast amount of real-time data on public opinions.

Valuable for tracking brand perception and customer feedback.

Strategic Impact:

Insight into public sentiment dynamics.

Enhance marketing tactics and brand management.

Preliminary Research

Focus Area	Key Findings	Reference
Machine Learning Models for Sentiment Analysis	Naïve Bayes, Maximum Entropy, and SVM were effective in sentiment classification. Feature selection and preprocessing are critical for accuracy.	Agarwal et al. (2011)
Lexicon-based and Machine Learning Combination	A blend of lexicon-based and machine learning approaches improved sentiment accuracy for entity-based sentiment analysis.	Saif et al. (2012)
Named Entity Recognition (NER)	Introduced NER approaches for noisy social media text, confirming their effectiveness for brand/entity identification.	Ritter et al. (2011)
Temporal Sentiment Trends	Analyzed sentiment fluctuations over time using time series data, aiding in sentiment trend forecasting.	Golder and Macy (2011)
Real-time Sentiment Analysis	Demonstrated the potential of social media for live sentiment tracking to monitor brand perception.	Balahur et al. (2014)
Sentiment Features in Tweets	Emoticons, hashtags, and internet slang improved sentiment classification accuracy.	Kouloumpis, Wilson, and Moore (2011)
Sentiment Mapping	Explored real-time sentiment mapping with timeline charts, emphasizing the importance of visualization in brand monitoring.	Diakopoulos and Shamma (2010)
Sentiment Analysis of Brand Mentions	Used Twitter sentiment data for brand monitoring, focusing on customer satisfaction and reputation analysis.	Mostafa (2013)

Methodology Overview

Dataset:

Twitter Entity Sentiment Analysis Dataset (Kaggle).

Key attributes: tweet text, sentiment, brand mentions, and user information.

Workflow:

Data Collection and Preprocessing.

Brand and Sentiment Extraction.

Sentiment Trend Analysis.

Visualization Development.

Progress Update # 1

Completed Tasks:

Dataset imported and analyzed.

Preprocessing initiated: stop-word removal, tokenization, and text normalization.

Findings:

Diverse data for sentiment analysis.

Challenges with imbalanced sentiment classes (neutral vs. positive/negative).

Progress Update #2

Feature Engineering

- Extracted key features to enhance model training:
  - **Text Features**: Length of tweets, frequency of hashtags, mentions, and URLs.
  - **Sentiment Features**: Polarity scores calculated using VADER.
  - **Entity Features**: Brand identification and categorization using NER tools (spaCy).

Insights Gained:

- Positive sentiments: Product quality and satisfaction.
- Negative sentiments: Service-related issues and delays.

Observations

Feature Engineering Outcomes:

Enriched dataset with meaningful features for sentiment classification.

Improved understanding of sentiment categories and patterns.

Challenges Identified:

Noise in NER results due to abbreviations and informal language in tweets.

Model Development

Sentiment Classification:

Algorithms: Naïve Bayes, SVM, and advanced deep learning models.

Metrics: Accuracy, Precision, Recall, F1-Score.

Brand Recognition:

NER tools like spaCy and NLTK.

Sentiment Trend Forecasting

Objective:

Predict future sentiment trends using ARIMA or Prophet.

Benefits:

Proactive brand strategy adjustments.

Enhanced customer engagement.

Visualizations

Planned Outputs:

- Temporal sentiment trends (line charts).
- Word clouds for frequent terms.
- Heatmaps for sentiment peaks.

Challenges and Mitigation

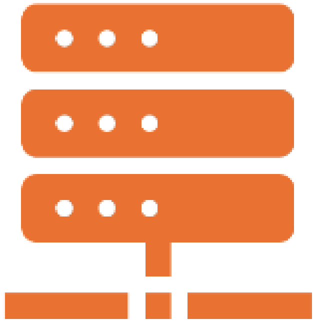
Challenges Identified:

- Data Noise: Abbreviations, emojis, and slang.
- Class Imbalance: Dominance of neutral/positive sentiments.
- Model Overfitting.

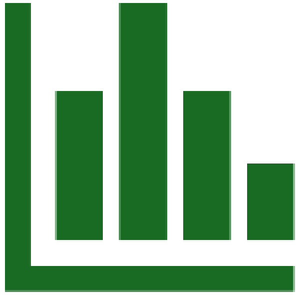
Mitigation Strategies:

- Text filtering, SMOTE for balancing, cross-validation for generalization.

Next Steps



Integrate engineered features into baseline machine learning models (e.g., Naïve Bayes, SVM).



Evaluate initial model performance and refine based on results.

Expected Outcomes

REAL-TIME SENTIMENT ANALYSIS OF TWITTER DATA.

BRAND SENTIMENT TRENDS VISUALIZED FOR ACTIONABLE INSIGHTS.

ENHANCED BRAND MANAGEMENT STRATEGIES.

References

[1]. Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. (2011). *Sentiment analysis of twitter data*. Proceedings of the Workshop on Languages in Social Media (pp. 30-38). Association for Computational Linguistics.

[2]. Pak, A., & Paroubek, P. (2010). *Twitter as a corpus for sentiment analysis and opinion mining*. Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC'10) (pp. 1320-1326). European Language Resources Association (ELRA).

[3] Balahur, A., Steinberger, R., Kabadjov, M., Zavarella, V., van der Goot, E., Halkia, M., ... & Belyaeva, J. (2014).

[4] E. Kouloumpis, T. Wilson, and J. Moore, "Twitter sentiment analysis: The good the bad and the omg!," *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, 2011, pp. 538-541.

[5] A. Ritter, S. Clark, O. Etzioni, and Mausam, "Named entity recognition in tweets: An experimental study," *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 2011, pp. 1524-1534.

[6] S. A. Golder and M. W. Macy, "Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures," *Science*, vol. 333, no. 6051, pp. 1878-1881, 2011.

[7] A. Go, R. Bhayani, and L. Huang, "Twitter sentiment classification using distant supervision," *Stanford University Technical Report*, 2009.

[8] N. A. Diakopoulos and D. A. Shamma, "Characterizing debate performance via aggregated Twitter sentiment," *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2010, pp. 1195-1198.

[9] H. Saif, Y. He, and H. Alani, "Semantic sentiment analysis of Twitter," *International Semantic Web Conference*, 2012, pp. 508-524.

[10] M. M. Mostafa, "More than words: Social networks' text mining for consumer brand sentiments," *Expert Systems with Applications*, vol. 40, no. 10, pp. 4241-4251, 2013.