

Social Media Impression Prediction Using Machine Learning and Sentiment Analysis

NAME: THARINLM

MAIL ID: m.s.tharinii24@gmail.com

Phone number:8122514096

Abstract

This work aims to create a strong framework for predicting the impressions of social media posts via machine learning algorithms. Through a combination of post metadata (share, comment, like, etc.) and sentiment analysis of the captions of the posts, we seek to make predictions about post reach and efficiency. This project also investigates correlation between engagement indicators and user activities on platforms including Instagram and X (Twitter). The model results will assist content creators and companies in optimizing their content strategy for maximum reach.

Introduction

With the digital marketing era, it is important to know how and why some content gets more impressions. Impressions are how often a post is shown to users, so it is an important measure of reach. Most research centers on classification tasks such as sentiment polarity or detecting fake news, but fewer have addressed numeric prediction tasks such as impressions. This research fills that gap through supervised learning methods.

Objectives

- To collect and preprocess social media data with engagement metrics.
- To analyze the sentiment of post content using NLP techniques.
- To train and compare various regression models for impression prediction.
- To provide visual insights using heatmaps, correlation matrices, and trend plots.
- To propose improvements based on sentiment and engagement trends.

Literature Review

Recent studies (2023–2025) in social media analysis have used models such as Random Forest, LightGBM, and XGBoost to predict user engagement. Sentiment has been demonstrated to have a direct influence on engagement, particularly on X and Instagram. Nevertheless, real-time impression prediction via hybrid models (engagement + sentiment) is not fully explored.

Example papers:

- “A Sentiment-Based Forecasting System for Social Media” – 2024, Elsevier
- “Impact of Visuals and Text Sentiment on Instagram Reach” – 2023, Springer
- “Real-Time Engagement Prediction Using Hybrid NLP Models” – 2025, IEEE Access

Methodology

1. Data Collection

- Instagram & Twitter/X datasets (historical + real-time scraped using APIs).
- Engagement metrics: likes, comments, shares, followers, etc.
- Sentiment analysis using VADER/TextBlob/BERT.

2. Feature Engineering

- Caption length, number of hashtags, posting time, and sentiment score.
- Engagement rate derived from combined metrics.

3. Modeling

- Train models: Random Forest, LightGBM, XGBoost, Linear Regression.
- Evaluate using metrics: MAE, MSE, RMSE, and R² Score.

4. Visualization

- Correlation heatmaps
- Scatter plots for predicted vs. actual impressions
- Time series trend analysis

Expected Outcome

- Accurate prediction of impressions with R² above 0.85.
- Identification of key contributing features (e.g., sentiment score, engagement rate).
- A user-friendly dashboard or Streamlit app for prediction and analysis.

Significance of Study

- Helps digital marketers and influencers improve content strategies.
- Supports academic research in social media analytics and behavioral trends.
- Bridges the gap between textual sentiment and visual engagement analytics.

Future Scope

- Extend to video-based platforms like YouTube and TikTok.
- Integrate deep learning (LSTM, Transformers) for better sentiment understanding.
- Real-time campaign monitoring with feedback loops.
- Multilingual sentiment analysis across geographies.

References

1. **Tolebay Nurlanuly, A. (2025).** *Sentiment analysis of texts from social networks based on machine learning methods for monitoring public sentiment.* arXiv preprint arXiv:2502.17143. arXiv
2. **Agustini, M. A. (2021).** *Sentiment Analysis on Social Media using Machine Learning-Based Approaches.* University of Twente. University of Twente Student Theses
3. **Brandwatch. (2024).** *Social Media Sentiment Analysis: Decoding Public Opinion in 2024.* ResearchGate+4Brandwatch+4Determ+4
4. **Sprout Social. (2024).** *The social media metrics to track in 2025 (and why).* Influencer Marketing Hub+3Sprout Social+3Socialinsider+s3
5. **Influencer Marketing Hub. (2025).** *Social Media Listening in 2025: Global Trends, AI Advances.* Influencer Marketing Hub+1AMA Triangle+1
6. **American Marketing Association Triangle. (2025).** *Marketing Analytics in 2025: Trends, Insights & What Marketers Need to Know.* AMA Triangle
7. **Socialinsider. (2025).** *Social Media Trends For 2025 According To Experts.* Socialinsider