### **Introduction**

Sentiment analysis on Twitter has become a popular research area due to the platform’s vast user base and real-time nature. Many studies focus on extracting polarity (positive, negative, neutral) from tweets using machine learning (ML) and deep learning approaches. However, challenges arise from Twitter’s character limits, informal language, and frequent use of slang and emoticons. My project builds on existing work by using a large-scale dataset—Sentiment140, which contains 1.6 million tweets labeled for polarity. The plan involves comprehensive data cleaning, exploratory data analysis (EDA), and experiments with both traditional ML methods and newer architectures like LSTM and transformer-based models. By comparing different approaches, this project seeks to identify which models perform best on a massive, real-world dataset. This literature review summarizes prior studies and highlights how my project extends or differs from them.

### **Body**

#### **1. Traditional Machine Learning Approaches**

**Neethu & Rajasree (2013)** explored sentiment analysis on Twitter by focusing on a relatively small dataset of product-related tweets. They showed that classifiers like Naïve Bayes (NB), Support Vector Machines (SVM), and Maximum Entropy (ME) achieve decent accuracy when combined with well-engineered features (e.g., emoticons, hashtags, keyword counts). Their feature-engineering approach was effective, but the dataset size was only 1200 tweets. In contrast, **my project** uses the much larger Sentiment140 dataset (1.6 million tweets), providing a better opportunity to train robust machine learning models. I will build on their emphasis on data preprocessing (removing URLs, slang, and repeated characters) but apply these methods at scale.

**Sarlan et al. (2014)** used both lexicon-based and ML-based approaches on tweets, implementing a system in Python. They highlighted the difficulties of handling informal language, emoticons, and real-time classification. While they aimed to build a web-based system using Django, they faced constraints related to deployment. For **my project**, I also plan to do preprocessing (tokenization, stopword removal, etc.) and compare various ML classifiers. However, I will go further by experimenting with advanced deep learning models and focusing on large-scale performance rather than a web-based deployment.

#### **2. Comprehensive Overviews and Key Challenges**

**Giachanou & Crestani (2016)** provided a broad survey of Twitter sentiment analysis methods, from machine learning to lexicon-based and hybrid approaches. They emphasized that feature engineering and deep learning significantly boost classification performance, but handling sarcasm, short text length, and domain dependency remain problematic. This survey confirms that **transformer-based architectures** and deep feature extraction can outperform simpler models. **My project** will incorporate these insights by testing transformer models (like BERT or similar) on the Sentiment140 dataset, which could offer improved context understanding despite Twitter’s inherent brevity.

#### **3. Deep Learning and Transformer-Based Methods**

**Teferra et al. (2024)** reviewed NLP approaches for depression screening, showing that transformer models (BERT, GPT, etc.) excel in extracting contextual clues from text. While their work focuses on clinical or depression-related tweets, it underlines how powerful transformer-based models can be in detecting subtle sentiment signals. **My project** aligns with this line of research because it will also explore deep learning methods—LSTM for sequence modeling and transformers for context-rich encoding. Even though my primary focus is not mental health screening, the same principles of advanced NLP can enhance general sentiment classification.

**Bokolo & Liu (2024)** did a comparative analysis of machine learning and transformer models for depression and suicide detection. They found that transformers usually outperform traditional ML classifiers when dealing with nuanced text, but simpler models sometimes work well in well-defined classification tasks. This is highly relevant because **Sentiment140**, while large, might contain straightforward positive/negative labels rather than the more subtle signs of depression. Nonetheless, their findings suggest that if the dataset has enough complexity, transformers can capture nuances that simpler ML models miss. **My project** will apply the same kind of comparative lens—testing traditional ML (Logistic Regression, Naïve Bayes, Random Forest) side by side with deep learning (LSTM) and transformers (e.g., BERT variants).

### **Conclusion**

Existing studies demonstrate that both traditional machine learning methods and deep learning models can produce good results for Twitter sentiment analysis. Early works (Neethu & Rajasree, Sarlan et al.) show the importance of effective preprocessing and feature engineering, but they often used smaller datasets. Surveys (Giachanou & Crestani) and specialized studies (Teferra et al., Bokolo & Liu) indicate that larger datasets and advanced deep learning approaches—especially transformers—usually deliver higher accuracy and better handling of nuances.

**My project** differentiates itself by:

1. Using a large-scale dataset (Sentiment140): Most prior works had smaller sets of tweets, potentially limiting generalizability.
2. Employing thorough EDA: Generating word clouds, n-gram analysis, etc., to understand the dataset’s quirks.
3. Comparing multiple model families: Traditional ML, LSTM, and transformer-based models will be evaluated, testing both performance and scalability.
4. Validating on unseen data: Moving beyond standard training/testing splits to assess how well the models perform on real-world, previously unseen tweets.

By building on previous research and applying state-of-the-art methods to a much larger dataset, this project aims to contribute new insights into the most effective techniques for sentiment analysis on Twitter.

### **References**

Bokolo, B. G., & Liu, Q. (2024). Advanced comparative analysis of machine learning and transformer models for depression and suicide detection in social media texts. *Electronics, 13*(20), 3980.<https://doi.org/10.3390/electronics13203980>

Giachanou, A., & Crestani, F. (2016). Like it or not: A survey of Twitter sentiment analysis methods. *ACM Computing Surveys, 49*(2), 1–41. <https://doi.org/10.1145/2938640>

Go, A., Bhayani, R., & Huang, L. (2009). Twitter sentiment classification using distant supervision. *CS224N Project Report, Stanford University*, 1(2009), 1–12. <https://www.kaggle.com/datasets/kazanova/sentiment140>

Neethu, M. S., & Rajasree, R. (2013). Sentiment analysis in Twitter using machine learning techniques. In *2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT)* (pp. 1–5). IEEE. <https://doi.org/10.1109/ICCCNT.2013.6726818>

Sarlan, A., Nadam, C., & Basri, S. (2014). Twitter sentiment analysis. In *Proceedings of the 6th International Conference on Information Technology and Multimedia (ICIMU)* (pp. 212–216). IEEE. <https://doi.org/10.1109/ICIMU.2014.7066632>

Teferra, B. G., Rueda, A., Pang, H., Valenzano, R., Samavi, R., Krishnan, S., & Bhat, V. (2024). Screening for depression using natural language processing: Literature review. *Interactive Journal of Medical Research, 13*, e55067. <https://doi.org/10.2196/55067>