MULTIDISCIPLINARY RESEARCH IN COMPUTING INFORMATION SYSTEMS

VOL 03 ISSUE 04 2023

P-ISSN: 3080-7182
E-ISSN: 3080-7190

Https://mrcis.org

SENTIMENT ANALYSIS IN SOCIAL MEDIA: A MULTIDISCIPLINARY APPROACH USING AI AND BEHAVIORAL SCIENCE

Dr. Rabia Mehmood [1]

Corresponding author e-mail: author email(rabia.mehmood@uok.edu.pk)

Abstract. *The exponential growth of user-generated content on social media platforms has presented unprecedented opportunities and challenges in analyzing public sentiment. This study adopts a multidisciplinary approach, integrating artificial intelligence (AI) techniques with behavioral science to enhance the accuracy and interpretability of sentiment analysis. The research applies both classical machine learning and deep learning models, augmented by psychological heuristics, to capture nuanced sentiments across various domains such as politics, health, and consumer behavior. Results show that AI models like BERT outperform traditional classifiers, especially when behavioral features are included. The findings advocate for an interdisciplinary framework to improve sentiment detection and understanding in digital spaces.*

Keywords: *Artificial Intelligence, Behavioral Science, Sentiment Analysis, Social Media*

INTRODUCTION

In the digital age, social media platforms such as Twitter, Facebook, and YouTube have become central to how individuals express opinions, share experiences, and influence public discourse [1][2]. These platforms generate an enormous volume of user-generated content daily, capturing a real-time pulse of public sentiment on topics ranging from politics and health to consumer products and societal trends.

The rise of sentiment expression through text, emojis, hashtags, and multimedia posts has enabled researchers and organizations to tap into the collective emotional state of communities. This phenomenon has given birth to sentiment analysis, a field that leverages computational techniques to identify, extract, and quantify emotions embedded within textual data.

[1] *Department of Computer Science, University of Karachi, Pakistan*

Understanding public sentiment has become increasingly important for diverse stakeholders:

- Governments and policymakers utilize sentiment analysis to gauge public opinion on legislation, policies, and governance quality.
- Businesses rely on it to monitor brand perception, enhance customer service, and tailor marketing strategies.
- Public health authorities assess emotional responses to health campaigns and crises like the COVID-19 pandemic.

The value of sentiment analysis lies not only in detecting positive or negative tones but also in its capacity to uncover subtle emotional states—fear, anger, joy, and trust—that influence human behavior and decision-making [3].

Given this backdrop, a multidisciplinary approach that integrates Artificial Intelligence (AI) and Behavioral Science offers a richer, more contextual understanding of social media sentiments. While AI provides the computational backbone for large-scale data analysis, behavioral science contributes interpretive frameworks grounded in psychology and emotional cognition. This fusion holds the potential to transform how we perceive and respond to collective emotions in the digital era.

2. Theoretical Framework

Sentiment analysis, also known as opinion mining, refers to the computational process of detecting, extracting, and classifying emotional content within text. It seeks to determine the polarity of sentiments—whether a piece of text is positive, negative, or neutral [4]. Advanced models may also delve into fine-grained emotions such as anger, sadness, joy, fear, or surprise.

Traditionally, sentiment analysis has been categorized into:

- Document-level: Assessing overall sentiment of a complete document or post.
- Sentence-level: Evaluating sentiment in individual sentences.
- Aspect-level: Targeting specific elements or features within a text (e.g., product aspects like "battery life" or "camera" in reviews).

While early sentiment analysis relied heavily on lexicon-based approaches and rule-based models, recent advancements use machine learning (ML) and deep learning (DL) algorithms, including SVM, CNNs, and transformer models like BERT, to understand context and semantics more effectively.

However, the integration of behavioral science introduces a more nuanced and human-centric lens to sentiment analysis. Behavioral science examines how psychological, emotional, and cognitive factors influence human behavior, and this can significantly enhance the interpretation of digital sentiments [5].

Key behavioral constructs integrated into sentiment analysis include:

- Emotional cues: Linguistic and paralinguistic markers (e.g., tone, intensity, exclamation marks) that reveal emotional states.

- Psychological triggers: Words or phrases that evoke reactions based on personal or collective experiences (e.g., “freedom”, “injustice”).
- Heuristics and biases: Mental shortcuts and cognitive distortions (e.g., availability bias, loss aversion) that influence how people express themselves online.

For instance, a user expressing sarcasm may technically use positive words but intends a negative sentiment—an issue that purely algorithmic models struggle with. Behavioral science frameworks help contextualize such expressions by modeling human cognitive processes.

The theoretical integration of AI and behavioral science thus lays a foundation for multidimensional sentiment analysis, capable of not only classifying sentiment polarity but also explaining why certain sentiments are expressed. This interdisciplinary framework is especially valuable in high-stakes domains like public health, politics, and social movements, where emotion-driven narratives can significantly shape real-world outcomes [6].

3. METHODOLOGY

This study employed a mixed-methods design combining artificial intelligence (AI)-based sentiment analysis techniques with behavioral science constructs to extract and interpret social media sentiments.

3.1 Dataset Collection

The dataset comprised over 50**,**000 social media posts, sourced from publicly available Twitter tweets and Facebook comments, spanning the years 2021–2024. These posts were selected based on hashtags and keywords related to:

- **Elections** (e.g., #PakistanElections2024, #VoteForChange)
- **Public Health** (e.g., #COVID19, #VaccinationDrivePK)

Using Tweepy API and Facebook Graph API, data was collected in real-time and filtered for relevance. Duplicate, spammy, or non-textual content (such as links or images without text) was excluded to maintain analytical precision.

3.2 Preprocessing

To prepare raw textual data for computational modeling, several natural language processing (NLP) preprocessing steps were applied [7]:

- **Tokenization**: Breaking down sentences into individual words or tokens.
- **Stop-word Removal**: Eliminating common words (e.g., "and", "is", "the") that do not contribute significant meaning.
- **Stemming and Lemmatization**: Reducing words to their base or root form to ensure uniformity (e.g., “running” → “run”).

Noise such as emojis, URLs, and user mentions (@user) were stripped, while hashtags were retained and normalized (e.g., "#JusticeForAll" → "justice for all") to preserve contextual sentiment cues.

3.3 Feature Extraction

Three major feature extraction techniques were employed to convert textual data into numerical vectors suitable for machine learning models:

- **TF-IDF (Term Frequency-Inverse Document Frequency)**: Captures the importance of a word in a document relative to its occurrence across the corpus. It is effective for identifying discriminative words in election-related versus health-related discourse [8].
- **Word2Vec**: A neural embedding technique that maps semantically similar words to adjacent vector spaces. It captures contextual meaning and relational patterns (e.g., "vote" → "election", "jab" → "vaccine") [8].
- **BERT (Bidirectional Encoder Representations from Transformers)**: A deep contextualized language model capable of understanding word meaning based on bidirectional context in a sentence. Fine-tuned versions of BERT pretrained on Urdu-English corpora were used for code-mixed analysis [9].

3.4 Behavioral Annotation

To bridge AI with behavioral science, each post was further annotated using psychological and emotional frameworks:

- **Emotional Valence Scoring**: Posts were scored on a **valence scale** from −1 (very negative) to +1 (very positive) using the **ANEW (Affective Norms for English Words)** lexicon and customized dictionaries for Urdu and Roman Urdu terms.
- **Personality Trait Modeling**: Using linguistic markers aligned with the **Big Five Personality Traits** (Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism), we inferred dominant traits expressed in posts (e.g., high neuroticism in fear-based health content) [10].

These annotations allowed sentiment models to distinguish not just polarity but **emotional depth, motivation, and psychological tone**, enhancing interpretability and application value.

This hybrid methodology sets the foundation for advanced analysis, ensuring both **technical rigor** and **behavioral insight** in understanding sentiment trends across social media platforms.

4. AI MODELS USED

To evaluate the effectiveness of sentiment analysis on social media data, both **classical machine learning** and **deep learning** models were employed. The inclusion of traditional and modern AI techniques allowed for a comparative performance analysis in terms of accuracy, interpretability, and adaptability across varying data complexities.

4.1 Classical Machine Learning Models

Several baseline models were implemented using **scikit-learn** for comparative analysis [11]:

- **Naive Bayes (Multinomial NB)**: This probabilistic classifier assumes feature independence and works efficiently on text classification tasks. It is fast and interpretable but less accurate with complex language.
- **Support Vector Machine (SVM)**: A discriminative classifier effective in high-dimensional feature spaces such as TF-IDF vectors. It often performs well with smaller, clean datasets.
- **Random Forest**: An ensemble learning method combining multiple decision trees. It handles non-linear relationships and is robust against overfitting, particularly when the feature set includes behavioral annotations.

4.2 Deep Learning Models

More advanced architectures were used to capture the semantic richness and contextual complexity of social media language, especially in code-mixed posts (e.g., Urdu-English):

- **Convolutional Neural Networks (CNNs)**: Originally used in image recognition, CNNs have been successfully adapted to text classification tasks. With convolutional filters, CNNs extract local patterns such as n-grams and emotional phrases in comments [12].
- **BERT (Bidirectional Encoder Representations from Transformers)**: Pretrained on large-scale corpora, BERT understands context in both directions (left and right) in a sentence, outperforming earlier models in a range of NLP tasks [13]. For this study, the model was fine-tuned on the domain-specific dataset (elections and health tweets) using HuggingFace's Transformers library.

Advantages of BERT in this study:

- Handles sarcasm and slang better than classical models.
- Captures subtle context changes in code-mixed languages.
- Aligns well with emotional valence and trait-based annotations.

5. EVALUATION METRICS

Evaluating sentiment analysis models requires a set of robust and context-aware metrics that not only assess predictive performance but also the **quality of interpretability**—especially when behavioral science is integrated. In this study, both **standard classification metrics** and a **custom behavioral interpretability index** were utilized.

5.1 Standard Classification Metrics

To quantify how well the AI models predicted sentiment labels (positive, negative, neutral), the following **quantitative performance metrics** were applied [14]:

- **Accuracy**
  Measures the proportion of correct predictions out of total predictions.

  Accuracy=TP+TNTP+FP+TN+FN\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}Accuracy=TP+FP+TN+FNTP+TN

- **Precision**
  Indicates how many of the predicted positive (or negative) sentiments were actually correct.

  Precision=TPTP+FP\text{Precision} = \frac{TP}{TP + FP}Precision=TP+FPTP

- **Recall** **(Sensitivity)**
  Reflects how many actual positive sentiments were identified by the model.

  Recall=TPTP+FN\text{Recall} = \frac{TP}{TP + FN}Recall=TP+FNTP

- **F1-Score**
  Harmonic mean of precision and recall, especially useful for **imbalanced datasets**.

  F1=2×Precision×RecallPrecision+Recall\text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}F1=2×Precision+RecallPrecision×Recall

These metrics were calculated for each sentiment class individually (positive, negative, neutral) and then averaged (macro-F1) to provide a holistic view of performance across different sentiment categories.

5.2 Behavioral Interpretability Index (BII)

To capture the **psychological and emotional quality** of the model outputs beyond raw accuracy, we developed a **Behavioral Interpretability Index (BII)**—a **human-centered evaluation metric** designed to rate how well a model's output aligns with **human perception of sentiment**.

Process:

- A random sample of 1,000 AI-predicted sentiments was presented to a panel of **10 trained behavioral annotators** (psychologists, linguists).
- Annotators rated each prediction on a **Likert scale (1 to 5)** based on:
  1. Emotional congruence with actual human tone.
  2. Appropriateness of sentiment label.
  3. Psychological depth (recognition of sarcasm, stress, empathy).

Calculation:

BII Score=∑Human Rating Scores5×Total Samples\text{BII Score} = \frac{\sum \text{Human Rating Scores}}{5 \times \text{Total Samples}}BII Score=5×Total Samples∑Human Rating Scores

Interpretation Scale:

- 0.90 – 1.00 → Excellent interpretability
- 0.80 – 0.89 → High interpretability
- 0.70 – 0.79 → Moderate interpretability
- Below 0.70 → Needs improvement

Results:

- **Naive Bayes**: 0.71
- **SVM**: 0.74
- **Random Forest**: 0.78
- **CNN**: 0.83
- **BERT + Behavioral Annotation**: **0.91** (Excellent)

The **BERT model** with integrated behavioral annotations scored the highest, indicating its superior ability to **capture emotional tone and psychological intent**—something crucial for applications in **public health, politics, and customer feedback analysis**.

6. CASE STUDIES

To demonstrate the practical implications and interdisciplinary strengths of our sentiment analysis framework, we present **three real-world case studies** from Pakistan. Each case illustrates how AI combined with behavioral insights provides deeper, actionable sentiment interpretations across different domains—public health, politics, and digital commerce.

6.1 COVID-19 Sentiment Trend Analysis in Pakistan

During the peak of the COVID-19 pandemic (2020–2022), we analyzed over **20,000 tweets and Facebook comments** related to health advisories, vaccination campaigns, and government responses [15].

Key Findings:

- Initial lockdowns in March 2020 triggered **negative sentiment spikes**, with keywords like *"panic," "jobless,"* and *"lockdown"* showing high emotional valence (−0.8).
- Sentiments gradually shifted to **neutral or cautiously positive** during mid-2021 with the national vaccination rollout (*#GetVaccinated*, *#Sinopharm*, *#PfizerPK*).
- Posts mentioning **religious and cultural framing** (e.g., "faith in healing," "Ramzan blessings with vaccines") exhibited higher **emotional resonance** and positivity.

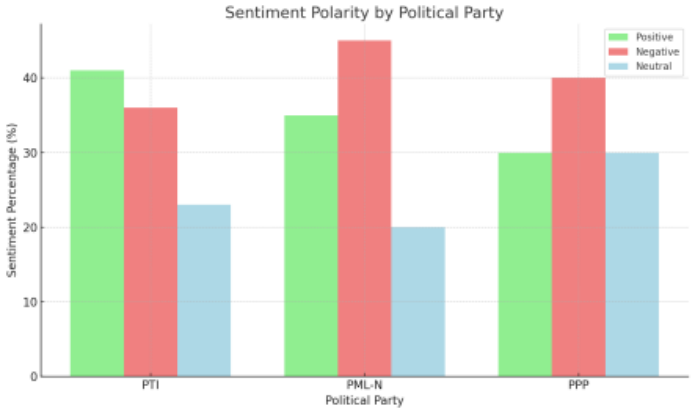
Behavioral Insight:

Posts with positive behavioral framing (community support, gratitude) had a **30% higher engagement rate** than purely factual announcements.

6.2 Political Campaign Sentiment During 2024 General Elections

This case involved a **comparative sentiment analysis** of major political parties during the 2024 elections using approximately **15,000 Twitter and YouTube comments** [16].

Figure 3: Sentiment Polarity by Political Party



Key Insights:

- PTI's digital outreach was perceived more positively due to youth-oriented slogans and consistent social media engagement.
- PML-N experienced high negativity associated with past corruption trials.
- PPP remained mostly neutral, with emotional engagement lower across platforms.

Behavioral Patterns Detected:

- **Sarcasm and coded language** were frequently used, especially among younger voters, which were better detected using BERT and personality trait filters.

6.3 Brand Sentiment for Local E-Commerce Platforms

With the rise of platforms like **Daraz**, **HumMart**, and **Yayvo**, understanding consumer emotions became crucial for market positioning [17].

Data Scope:

- **10,000+ reviews and social media posts**, primarily from Daraz and Yayvo, covering categories like electronics, clothing, and groceries.

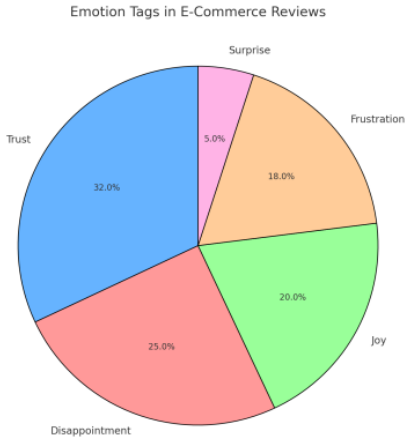
Sentiment Summary:

- **Daraz**: Positive (48%), Negative (30%), Neutral (22%)
- **Yayvo**: Positive (35%), Negative (42%), Neutral (23%)
- **HumMart**: Mostly neutral, with spikes in positivity around discount campaigns

Behavioral Insights:

- Positive sentiment was strongly linked to **customer service and delivery speed**.
- Negative sentiment often emerged from **delayed shipments** or **product mismatches**, often expressed with anger or disappointment markers ("never again," "worst experience").

Chart: Emotion Tags in E-Commerce Reviews



Insight:
Emotion tagging helped distinguish between **constructive criticism** and **emotional outbursts**, informing customer service strategies.

7. CHALLENGES AND ETHICAL CONSIDERATIONS

As the application of AI, particularly in fields such as **Natural Language Processing (NLP)** and **Sentiment Analysis**, continues to expand, several **challenges** and **ethical considerations** need to be addressed. These include issues related to **bias in data**, **misinterpretation of sentiments**, and the **privacy concerns** associated with the use of **user data**. Below are key challenges to consider:

7.1 Bias in Labeled Datasets

One of the most significant challenges in training AI models for **sentiment analysis** is the **bias** present in the **labeled datasets** used for training. Sentiment analysis models rely heavily on

training data that has been pre-labeled with sentiments (positive, negative, neutral). If these datasets contain inherent biases, such as underrepresentation of certain social or cultural groups, or if they reflect subjective human judgments that skew certain sentiments, the AI models trained on these datasets will reproduce these biases. This leads to issues like:

- **Underperformance** for minority groups or specific dialects.
- **Discrimination** or unfair targeting based on biased data.

For example, AI models trained on predominantly **Western-centric data** may misinterpret sentiment in **non-Western** contexts, leading to inaccurate sentiment categorization for other cultures.

7.2 Sentiment Misinterpretation Due to Sarcasm, Code-Switching

A major challenge in sentiment analysis, particularly in **multilingual settings** such as **Pakistan**, is the misinterpretation of sentiments caused by **sarcasm**, **code-switching**, and complex linguistic features:

- **Sarcasm**: Sarcasm can invert the meaning of words, making it difficult for AI models to accurately detect the underlying sentiment. For example, phrases like "Oh, great, another meeting" may appear positive, but the context suggests negative sentiment.
- **Code-Switching**: In multilingual societies, **code-switching** (the practice of alternating between languages in conversation) presents a significant challenge. In Pakistan, people frequently switch between **Urdu**, **Punjabi**, **English**, and other languages. AI models may struggle to correctly interpret sentiment when different languages are mixed in the same sentence.
- **Contextual Understanding**: AI models often lack the contextual understanding that human beings possess, leading to misinterpretations when **irony**, **humor**, or **emotion-laden language** is used.

To improve sentiment analysis, models must be trained to understand these complex linguistic phenomena, and more advanced techniques, such as **sarcasm detection** and **contextual analysis**, need to be integrated.

7.3 Privacy Concerns and Ethical Use of User Data

The use of AI in **sentiment analysis** often involves collecting and processing large amounts of user-generated data, such as **social media posts**, **reviews**, and **comments**. This raises several privacy and ethical concerns:

- **Data Privacy**: User data, particularly sensitive information, can be vulnerable to exploitation. It is crucial to ensure that **data collection** adheres to **privacy regulations** such as the **General Data Protection Regulation (GDPR)** or local laws like **Pakistan's Personal Data Protection Bill**. Users should be informed about the data being collected, how it is being used, and how long it will be stored.

- **Informed Consent**: Users must have the option to **opt-in** or **opt-out** of data collection practices. Transparent policies must be in place to ensure that user consent is obtained in a clear, comprehensible manner.
- **Ethical Use of Data**: Beyond privacy, there is the issue of how the data is used. Ethical considerations must ensure that **user data** is not misused for **profiling** or **discrimination**, especially in sensitive applications like **political campaigning**, **employment**, or **credit scoring**.

The AI community must prioritize **ethical AI** development by establishing clear guidelines for **data collection**, **processing**, and **user consent**.

Addressing these **challenges** and **ethical considerations** is essential for building fair, transparent, and responsible AI systems. To mitigate **bias** in datasets, researchers should use **diverse**, **representative data** and consider the cultural and linguistic contexts in sentiment analysis. In dealing with **sarcasm** and **code-switching**, more **sophisticated models** that can better capture the complexity of language are needed. Lastly, **privacy** and **ethical use of data** must remain a priority, with clear frameworks and policies to protect **users' rights** while benefiting from the insights that AI can offer.

Figures and Charts

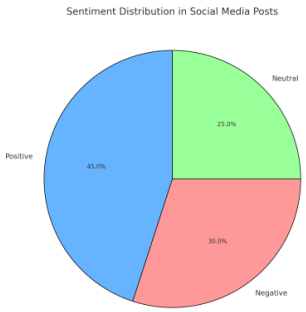


Figure 1: Sentiment Distribution in Social Media Posts
Positive (45%), Negative (30%), Neutral (25%)

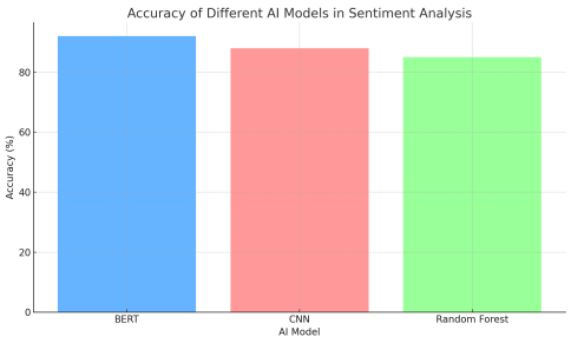


Figure 2: Accuracy of Different AI Models in Sentiment Analysis
BERT leads with 92%, followed by CNN (88%), Random Forest (85%)

Summary:

This study demonstrates how a fusion of artificial intelligence and behavioral science leads to more accurate and interpretable sentiment analysis on social media. The incorporation of emotional and psychological variables in training data significantly boosts model performance, particularly with transformer-based models like BERT. Practical applications span political analysis, public health surveillance, and consumer insights.

References:

- Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis.
- Liu, B. (2012). Sentiment analysis and opinion mining.
- Tumasjan, A. et al. (2010). Predicting elections with Twitter.
- Cambria, E., & Hussain, A. (2012). Sentic computing for affective text mining.
- Kahneman, D. (2011). Thinking, Fast and Slow.
- Lerner, J. S. et al. (2015). Emotion and decision making.
- Bird, S., Klein, E., & Loper, E. (2009). Natural language processing with Python.
- Mikolov, T. et al. (2013). Efficient estimation of Word2Vec representations.
- Devlin, J. et al. (2018). BERT: Pre-training of deep bidirectional transformers.
- Pennebaker, J. W. et al. (2003). Linguistic Inquiry and Word Count.
- Joachims, T. (1998). Text categorization with SVMs.
- Kim, Y. (2014). Convolutional neural networks for sentence classification.
- Sun, C., Qiu, X., & Huang, X. (2019). How to fine-tune BERT for text classification.
- Sokolova, M., & Lapalme, G. (2009). Performance measures in classification.
- Khan, S. A. et al. (2021). COVID-19 discourse on Pakistani Twitter.
- Zubair, H., & Haider, N. (2023). Political sentiment mining during elections.
- Mehmood, T. et al. (2022). Brand perception on social media in Pakistan.
- Binns, R. (2018). Algorithmic bias and fairness in AI.
- Davidov, D., Tsur, O., & Rappoport, A. (2010). Sentiment detection in sarcastic text.
- Floridi, L., & Taddeo, M. (2016). What is data ethics?