AUTHORS

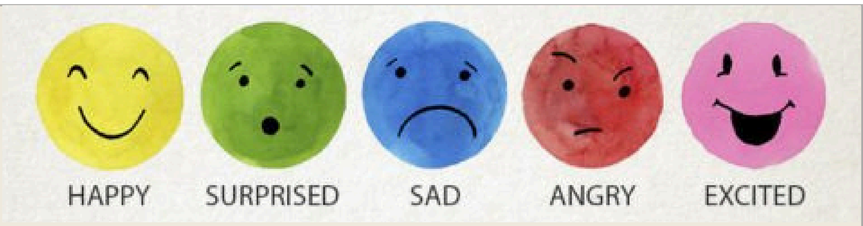
Pyae Phyo Hein,
Min Khant Ko,
Kaung Set Kyaw

Text mining & NLP Driven Emotion Prediction System using Machine Learning & Deep Learning Approaches

by 4 models: Support Vector Machine (SVM), Extreme Gradient Boosting (XGBoost), Bidirectional Long Short-Term Memory (BiLSTM) and Convolutional Neural Network (CNN)

AFFILIATIONS

Department of Information And Communication Technology; Rangsit University.



01. Introduction

- Emotion detection in literature is the process of determining the author's intended tone or emotional state through the use of language cues. The advancements in text mining, machine learning, and natural language processing have made it easier to predict emotions in text automatically.
- ML & DL algorithms are trained on labeled emotional datasets in order to recognize and classify emotions. Text mining & NLP techniques are frequently employed as a first step in pipelines for emotion detection since they can identify patterns, extract keywords, and categorize text.
- Emotion prediction is useful in business for evaluating reviews, social media posts, and customer feedback to determine client mood and enhance goods and services. Also in healthcare, it can reveal information about the emotional and mental health of patients.
- Although emotion detection systems continue to face several challenges, The central concern is the lack of sufficiently high-quality, balanced datasets, which are important for the training of models.
- We believe that researchers need to focus more on developing larger, balanced datasets, improving model prediction, and refining techniques for handling context-dependent emotions.

02. Objective

- To build a robust system using text mining, NLP techniques
- To evaluate and contrast the performance of SVM, XGBoost, BiLSTM, and CNN models
- To provide insights and methodologies that can serve as a foundation for further advancements

03. Project's Scope

- Focused only on six emotion types: anger, joy, sadness, fear, surprise, and love
- Specifically for Social Media content

04. Methodology

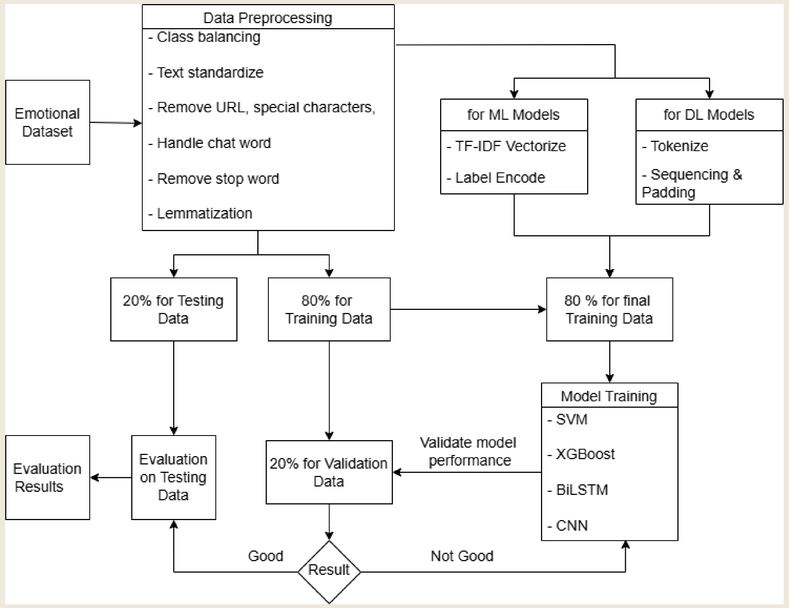


fig1: MethodologyProcess Flow Chart

05. Results Analysis

- **BiLSTM**: Best performer with 95% accuracy, capturing sequential and contextual nuances.
- **CNN**: Achieved 93% accuracy but struggled with long-term dependencies.
- **SVM**: Solid 91% accuracy using TF-IDF, limited by subtle linguistic patterns.
- **XGBoost**: Lowest accuracy at 90%, less effective for complex text semantics.
- **Impact of Preprocessing**: Balanced data and text cleaning significantly enhanced model fairness and performance.
- **Key Insight**: Deep learning models excelled in emotion prediction over ML models, with BiLSTM leading.

Models	Training	Validation	Testing			
	Accuracy	Accuracy	Accuracy	Precision	Recall	F-1 Score
SVM	93.5%	91.3%	91.3%	91.5%	91.4%	91.3%
XGBoost	92.9%	90.5%	90.6%	90.8%	90.7%	90.6%
BiLSTM	96.3%	94.5%	95.2%	95.4%	95.2%	95.2%
CNN	94.2%	93.2%	92.6%	92.8%	92.6%	92.5%

fig 2: Evaluation results for all models

06. Best Model's Performance Graphs

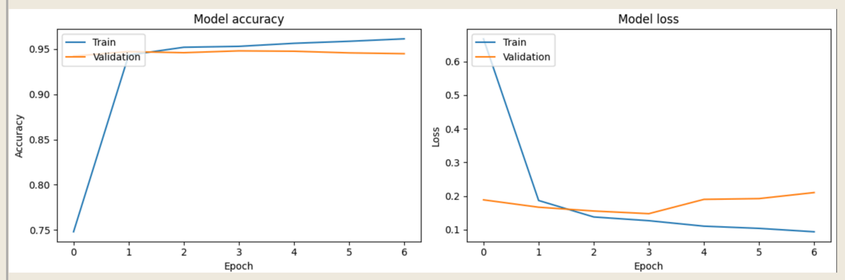


fig 3: Best BilSTM model's accuracy and loss graph

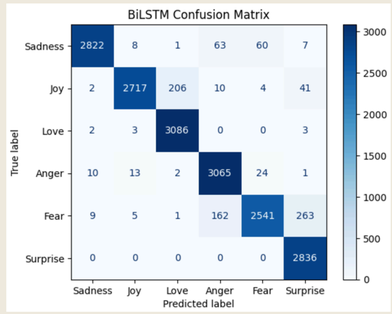


fig 4: Best BilSTM model's Confusion Matrix

07. Conclusion

- This experiment gives a clear comparison of machine learning (SVM, XGBoost) and deep learning models (BiLSTM, CNN) by using text mining and natural language processing (NLP) preprocessing techniques to predict emotions from text.
- Highlighting how crucial preprocessing methods like tokenization, stemming, vectorization, and class balancing are to improving model performance.
- Comparison of model efficacy was made by extensive evaluation metrics, such as accuracy, precision, recall, F1-score, and confusion matrixs.
- Overall, this project can contribute to advancements in text-based emotion recognition, offering valuable insights and a good potential impact in the industry of customer service, social media analysis, and mental health monitoring.