# Detailed Summary on "Abusive Language Detection from Social Media Comments Using Conventional Machine Learning and Deep Learning Approaches"

Author: Saurabh Sunil Mishra  
Roll No: 2023201034  
Professor: Radhika Mamidi  
Laboratory: Natural Language Processing

## Abstract

The increasing prevalence of abusive language on social media platforms necessitates robust automated detection systems. This paper evaluates the efficacy of various conventional machine learning (ML) and deep learning (DL) models in detecting abusive comments, particularly in the Urdu and Roman Urdu languages, which are underrepresented in computational language studies. The research reveals that deep learning models, specifically Convolutional Neural Networks (CNN), outperform traditional machine learning models in accuracy and efficiency.

## 1. Introduction

With the ubiquity of social media, instances of abusive language are on the rise, affecting individuals and communities. Previous detection efforts have primarily focused on English, leaving languages like Urdu with significant linguistic resources and research gaps. Urdu, with its complex script and extensive morphology, presents unique challenges for text processing, which this study addresses by deploying a range of ML and DL models tailored to its nuances.

## 2. Background and Significance

### 2.1 Linguistic Characteristics of Urdu

Urdu is written in the Nastaleeq script and comprises several dialects, each with unique phonetic and syntactic traits that complicate text analysis. The Roman Urdu variant, popular on digital platforms, lacks standardization, further complicating computational processing.

### 2.2 Need for Advanced Detection Techniques

Automated tools for detecting abusive language must be sensitive to linguistic subtleties and scalable across vast datasets, a requirement that deep learning fulfills more adeptly than traditional machine learning due to its capacity for feature extraction without explicit programming.

## 3. Dataset Description

### 3.1 Composition and Collection

The study utilized two datasets:  
- A Roman Urdu dataset consisting of 10,000 comments, balanced between abusive and non-abusive content.  
- An Urdu dataset featuring 2,171 comments, also balanced, sourced from various social media platforms and annotated for abuse presence.

### 3.2 Challenges

The primary challenges include the informal and inconsistent use of language in Roman Urdu and the script complexity of Urdu, which impacts the training efficiency of ML models.

## 4. Methodology

### 4.1 Machine Learning Models

Five conventional ML models were employed:  
- Naive Bayes (NB): Probabilistic model, effective for baseline comparisons.  
- Support Vector Machine (SVM): Utilized for its robustness in high-dimensional spaces.  
- K-Nearest Neighbors (IBK): A non-parametric method used for its simplicity.  
- Logistic Regression: For binary classification problems typical of abusive language detection.  
- JRip: A rule-based approach that generates sets of decision rules.

### 4.2 Deep Learning Models

Four deep learning architectures were tested:  
- Convolutional Neural Network (CNN): Optimal for pattern recognition within data, adjusting to the intricacies of language structures.  
- Long Short-Term Memory (LSTM): Addresses the issue of learning long-term dependencies.  
- Bidirectional LSTM (BLSTM): Expands on LSTM by processing data in both forward and backward directions to better capture context.  
- Convolutional LSTM (CLSTM): Combines CNN and LSTM models to leverage both local feature extraction and sequence dependency learning.

## 5. Experiments and Results

Extensive experiments were conducted to assess the models’ performance across both datasets, with CNN demonstrating superior accuracy and efficiency. The results highlighted the strengths of deep learning in managing the contextual and morphological complexities of the Urdu language.

## 6. Discussion and Conclusion

The study confirms that deep learning models are more adept at processing and classifying complex linguistic data compared to traditional machine learning models. CNN's ability to effectively handle both character and word-level features makes it particularly valuable for languages like Urdu.

## 7. Future Work and Recommendations

Future research should expand the datasets to include more diverse social media platforms and possibly integrate multilingual data to enhance model robustness. Additionally, exploring hybrid models that combine the strengths of both ML and DL could offer improvements in accuracy and processing speed.

## 8. Acknowledgments

Acknowledgments are extended to the funding bodies and research assistants who contributed to the dataset preparation and preliminary analyses.

## 9. References

A comprehensive list of references supports the research, drawing from foundational texts in machine learning, deep learning, and natural language processing, as well as previous studies on abusive language detection.