Hate Speech Detection in Urdu News: A Web-Scraped Dataset and Machine Learning/Deep Learning classificaton Framework

Atif Saeed
Department of Data Science
FAST–NUCES, Islamabad
atif.saeed@isb.nu.edu.pk

Talha Aslam
Department of Data Science
FAST-NUCES, Islamabad
i248067@isb.nu.edu.pk

Muhammad Ahmer
Department of Data Science
FAST–NUCES, Islamabad
i247602@isb.nu.edu.pk

***Abstract*—This paper presents a framework for detecting hate speech in Urdu news. It uses a dataset collected from the web and applies machine learning and deep learning models. We gathered articles from four major Urdu news outlets and processed them through a labeling and preprocessing system designed for Urdu. Classical machine learning models that used TF-IDF features achieved the best results. XGBoost and the Stacking Classifier reached an accuracy and F1-score of 0.90.**

Deep learning models, such as BiLSTM, CNN, a Hybrid CNN-BiLSTM, and BiLSTM with Attention, showed competitive recall. However, they had lower overall accuracy due to the dataset size. The findings suggest that ensemble machine learning methods work better than deep learning models for structured news text. Meanwhile, attention-based models have potential for improved contextual understanding. This work provides a new web-scraped dataset of Urdu news and sets benchmarks for hate speech detection in low-resource languages using machine learning and deep learning.

***Index Terms*—Urdu NLP, Hate Speech Detection, Machine Learning, Web Scraping, Text classificaton, Urdu News Media**

I. Introduction

The rapid shift to digital news in Pakistan has greatly improved people's access to real-time information. However, it has also increased the spread of hateful and divisive narratives. Urdu news platforms like BBC Urdu, Dawn News Urdu, Geo News, and Express News play a key role in shaping public opinion. As political tensions, social unrest, and ideological conflicts rise, harmful language in news reporting, editorials, and crime coverage has become more obvious. It is essential to recognize this language because people often view news content as credible. It can influence large audiences more effectively than social media posts.

While studies on hate speech detection have focused a lot on English. And in urdu languge this work is missing. Most existing research on hate speech in Urdu focuses on Twitter, where informal writing, slang, code-mixing, and user-generated content are common. UHated, developed by Arshad et al., is an important project because it uses transformer-based models to identify hateful and offensive content in Urdu tweets. However, models made for social media do not work well with news due to differences in sentence length, writing style, linguistic structure, and the more subtle or context-driven nature of hate speech in news articles.

Unlike short, informal user-generated texts, Urdu news articles are formal and well-structured. They often embed bias or hostility subtly within the narrative instead of using clear slurs. This creates a need for a specific approach to detect hate speech in news.

To meet this need, we have built a complete system for detecting hate speech in Urdu news. This process includes scraping content from four major Urdu news outlets—BBC Urdu, Dawn News Urdu, Geo News Urdu, and Express News Urdu. We then conduct Urdu-specific preprocessing, automatic lexicon-based annotation, and classification using traditional machine learning models. Our focus is on binary classification, distinguishing between Hate and Non-Hate, to identify explicit hateful expressions in formal news content.

II. Related Work

Hate-speech detection has gained interest in English and other high-resource languages, while low-resource languages such as Urdu remain underrepresented. Arshad *et al.* developed UHated [1], a RoBERTa-based transfer-learning model trained on 7800 manually annotated Urdu tweets, achieving a macro F1-score of 0.82. Their dataset includes hate, offensive, and neutral categories[6].

Other works in Roman Urdu and multilingual contexts have used classical Machine Learning classifiers or lexicon-based methods, but no prior work focuses specifically on *Urdu news articles*. News content differs from social media due to its formal structure, reduced slang, and richer grammar. Our work addresses this gap by applying hate-speech detection techniques to large-scale Urdu news content[7].

III. Literature review

Previous work related to hate speech detection has shown very strong progress in English and other high-resource languages, while work on Urdu is limited because of small datasets and generally weak NLP tooling. Early studies are primarily classical machine learning methods using n-gram features, with later works extending this with one or more neural model(s) including CNNs and LSTMs. More recently, transformer-based transfer learning-especially multilingual models like XLM-RoBERTa-has consistently improved upon prior work, particularly for low-resource languages. However, large, high-quality annotated datasets are still lacking for Urdu, with context-dependent hate expressions making annotation difficult. The UHated study directly fills these gaps by constructing a labeled Urdu dataset and showing that RoBERTa-based model(s) achieve the strongest performance among all tested methods[8].

Akhter et al. (2020) presented some of the early attempts in the detection of offensive content both in Urdu and Roman Urdu, addressing the scarcity of resources in these languages. Their work indicated that the normalization of text, script-hanDeep Learninging, and custom preprocessing have significant impacts on system performance. Using deep-learning models like CNNs and LSTMs with word embeddings prepared for local linguistic patterns, they reported superior results compared to classical machine-learning baselines. The study recognized that Urdu, being a complex script language with variation in morphology, using mixed scripts on social media, requires a specialized hanDeep Learninging, thus laying the ground for later transformer-based research[9].

Ali et al. (2021) addressed the problem of hate speech detection in Urdu tweets. This work focused on the integration of sentiment signals into the classificaton pipeline. This study showed that model accuracy, especially for borderline cases where hate and strong negative sentiment overlap, was better through the inclusion of sentiment features along with lexical and semantic cues. They compared traditional Machine Learning model(s) with deep learning and showed clear performance gains when contextual sentiment information was included. The paper emphasizes the need for richer feature representations in Urdu and also discusses how hard it is to distinguish between emotional and hteful expressions, to which later works addressed using transformers[10].

Albadi et al. (2018) researched religiously motivated hate speech in Arabic and showed that when domain-specific context-in this case, religious terms or target groups-is included, the detection rate increases significantly. This caused the best results. Their dataset represented very subtle ways of expressing religious hostility, and, as was noted, hate speech is very often implicit and highly contextual. Their experiments highlighted that the neural model(s) which used word embeddings outperformed traditional Machine Learning approaches, especially when the hate expressions are subtle and deeply rooted in a particular culture. Although this research focused on Arabic, it is still relevant for Urdu, since religion and cultural context determine hate speech patterns in Urdu as well. Alatawi et al. (2021) analyzed hate content related to extremist groups and demonstrated that transformer-based architectures, especially BERT variants, perform best for domain-specific hate speech detection. The results showed that the pre-trained contextual embeddings captured the subtle semantic relations which traditional model(s) failed to capture, thus yielding significantly higher classification performance. This caused the best results. They further showed that domainadapted, embedded representations improve results even more, especially when dealing with low-resourced or specialized domains. This is also in line with very recent trends in Urdu hate speech detection, where multilingual transformers achieve state-ofthe-art performance. Waseem and Hovy's work (2016) is one of the early influential studies in the domain of hate speech on social media. The study focused on proving that approaches based on contextual metadata such as user information, gender, and tweet structure outperform simple text-based approaches of n-grams. Their annotated dataset revealed a pattern that hate speech is often subtle, linked to a social or conversational context, which makes a naive keyword-based approach quite insufficient. This work influenced later studies heavily, encouraging the community to adopt deeper, linguistic, and contextual features. This is still valid for languages like Urdu, where contextual clues are extremely useful to disambiguate slang, sarcasm, and culturally embedded insults .opend

IV. Methodology

This study follows a complete end-to-end workflow consisting of four major components:

1) Dataset acquisition through web scraping
2) Urdu-specific text preprocessing
3) Automatic lexicon-based labeling
4) Machine-learning–bsed hate-speech classificaton

A. Data Acquisition Through Web Scraping

A custom scraping system was developed to collect Urdu news articles from four well-established Pakistani news sources:

- BBC Urdu
- Dawn News Urdu
- Express News Urdu

The scraper systematically navigated each website, extracted article URLs, and downloaded the associted titles, full text, timestamps, and metadata. Due to varying HTMachine Learning structures across platforms, each source was hanDeep Learninged with dedicated parsing rules to ensure accurte extraction. All articles were stored in a unified structured dataset and exported to CSV for downstream processing.

The resulting cleaned dataset provides high-quality normalized Urdu text suitable for machine-learning modeling.

Table I
Comparative Summary of Research Papers on Hate Speech Detection

Category	Paper 1 (UHated)	Paper 2 (Akhter 2020)	Paper 3 (Ali 2021)	Paper 4 (Albadi 2018)	Paper 5 (Alatawi 2021)	Paper 6 (Waseem & Hovy 2016)
Title	UHated: Hate Speech Detection in Urdu Using Transfer Learning	Offensive Language Detection in Urdu and Roman-Urdu	Sentiment-Aware Hate Speech Detection in Urdu	Religious Hate Speech in Arabic	Extremist Hate Speech Detection Using BERT	Predictive Features in Hate Speech
Journal / Source	AI & Society (2023)	IEEE Access (2020)	SN Applied Sciences (2021)	ACL Workshop (2018)	JISA (2021)	NAACL-HLT (2016)
Dataset	7,871 Urdu tweets	Urdu/Roman-Urdu comments	Urdu Twitter dataset	Arabic religious-hate corpus	Extremist posts dataset	English Twitter dataset
Preprocessing Techniques	Cleaning, normalization, tokenization	Normalization, stopword removal	Normalization, sentiment tagging	Arabic normaliztion, stopwords	Cleaning, tokenization	Cleaning, metadata processing
Techniques (Methodology)	Machine Learning, Deep Learning, XLM-R, RoBERTa	CNN, LSTM model(s)	Machine Learning + sentiment features	Machine Learning + domain embeddings	BERT / domain-adapted BERT	Machine Learning + contextual features
Strengths	New Urdu dataset; strong transformer results	Deep Learning mixed-script data	Uses emotional cues	Domain-specific hate detection	Strong contextual modeling	Context improves accuracy
Limitations	Twitter-only dataset	Small dataset; no transformers	Weak sentiment lexicon	Narrow domain only	Domain-specific, not generalizable	English-only; two labels
Results	RoBERTa best (F1≈0.82)	Deep Learning > Machine Learning performance	Higher accuracy with sentiment cues	Neural > Machine Learning performance	Domain-BERT best F1	Context > n-grams
Evaluation Metrics	Accuracy, Precision, Recall, F1	Accuracy, F1	Accuracy, F1	Precision, Recall, F1	Accuracy, F1	F1, Precision, Recall
Research Gap Identified	Need larger, diverse Urdu datasets	Need contextual transformer model(s)	Need richer semantic features	Need broader contextual modeling	Need low-resource domain model(s)	Need multilingual contextual systems

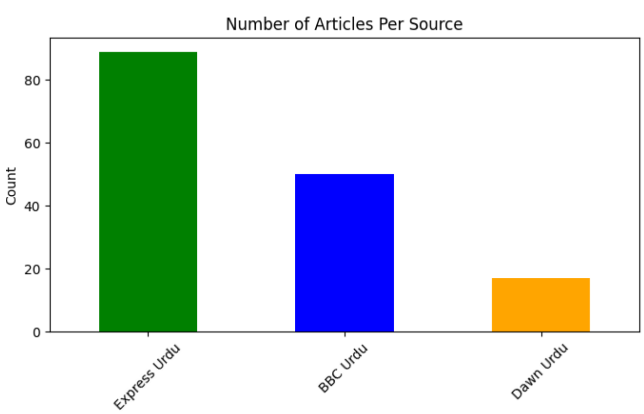

Figure 2. News Source

B. Urdu Text Preprocessing

Urdu requires specialized normalization because it is written from right to left instead of left to right, lack of capitalization, and inconsistent Unicode usage. The following preprocessing steps were applied:

1) **Unicode Normalization:** Standardization of characters such as various forms of and removal of Arabic/Persian variants.
2) **Noise Removal:** Elimination of HTMachine Learning artifacts, URLs, punctuation, emojis, digits, and English text.
3) **Stopword Removal:** Filtering of frequent Urdu stopwords to reduce noise.
4) **Boilerplate Filtering:** Removal of repeated text fragments commonly present in news templates.
5) Length Constraints: Extremely short or low-information texts were discarded.

This pipeline produces clean, standardized Urdu sentences suitable for machine-learning model(s).

C. Lexicon-Based Binary Annotation

To label the dataset at scale, a curated Urdu hate lexicon was constructed. It contins explicit hate terms commonly found in political, religious, and social-hostility contexts.

Articles were labeled as:

1) **Hate (1):Article contains one or more lexicon terms**
2) **Non-Hate (0):** No lexicon terms present

This lexicon-based approach aligns well with news content, where explicit hateful terms tend to appear directly rather than implicitly.

D. Feature Engineering Using TF–IDF

Cleaned text was converted using TF–IDF vectorization for numerical representations. The configuration included

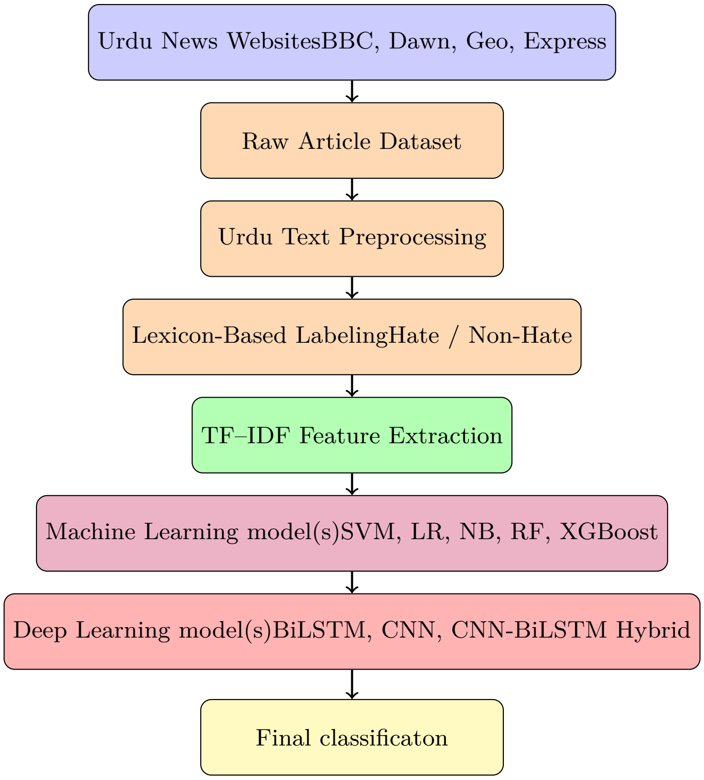

Figure 1. Urdu news hate-speech detection pipeline with Machine Learning and Deep Learning model(s) stacked.

following:

- Unigram + Bigram features
- Maximum of 5000 features
- Minimum document frequency of 3
- Maximum document frequency of 0.90

TF–IDF effectively captures word importance and performs well on structured, formal text such as news articles.

E. Machine Learning model(s)

Six classical machine-learning classifiers were trained and compared:

- Logistic Regression
- Linear Support Vector Machine (SVM)
- Random Forest
- Multinomial Naive Bayes
- XGBoost
- Stacking classifier

For the purpose of analyzing the effectiveness of classical approaches for hate speech detection in Urdu news, we experimented with a very diverse set of machine learning models: Logistic Regression, Linear Support Vector Machine, Multinomial Naive Bayes, Random Forest, and XGBoost. These are chosen because of the potential of these algorithms to give good performance in high-dimensional, sparse text classification tasks. Moreover, they exhibit strong compatibility with TF-IDF representations. Margin-based optimization is one of the main reason that Linear SVM is widely used for text separation, and for stable, probabilistic outputs we use logistic regression 3. Naive Bayes is perfect for word frequency distributions, and Random Forest aligns best with robustness through ensemble learning, and XGBoost provides gradient-boosted decision trees optimized for predictive accuracy. Together, these model(s) provide a comprehensive baseline that will be useful in gauging the performance of Urdu news content classification. Stacking Classifiers: An ensemble meta-model that blends base learners' predictions in order to improve its performance.

Class balance was maintained with the use of an 80/20 stratified train-test split. Evaluation metrics used were: Accuracy, Precision, Recall, F1-Score, and Confusion Matrix.

Among all the models, Xgboost and Stacking Classifier achieved the highest performance by demonstrating strong capability in distinguishing hate from non-hate articles within a news domain.

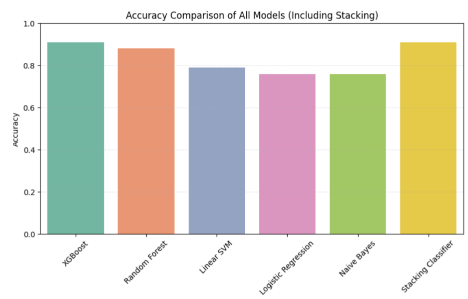

Figure 3. model(s) accuracy

We used Receiver Operating Characteristic (ROC) analysis in addition to standard metrics to compare the machine-learning model(s)' threshold-independent performance. The ROC curve plots the True Positive Rate (TPR) against the False Positive Rate (FPR) over varying decision boundaries. Stronger discriminative ability is indicated by a higher Area Under the Curve (AUC). The produced ROC curves demonstrate that Linear SVM achieves the highest AUC, confirming its superior performance for the classificaton of hate speech in Urdu news.

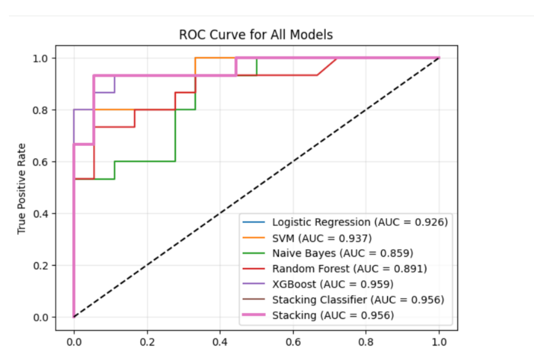

Figure 4. ROC curve for Machine learning model(s)

F. Deep Learning model(s)

1) **BiLSTM:**
A text sequence is processed both forward and backward by the BiLSTM network, which can learn long-range dependencies and contextual relationships. In fact, BiLSTM has been widely employed in text classification tasks on account of its capability to maintain semantic information more effectively compared to traditional RNNs.
2) **CNN:**
A CNN was utilized to extract local n-gram features in text using convolution filters. CNNs perform well in extracting main discriminative phrases and localizing spatial patterns, which makes them suitable for short- and medium-length sentences that occur regularly in news articles.
3) **Hybrid CNN–BiLSTM Model:**
To take advantage of both the architectures, we proposed a hybrid CNN–BiLSTM model. The CNN layer extracts high-level local features, and the subsequent BiLSTM layer captures global contextual dependencies. This hybrid approach improves feature extraction by leveraging both spatial and sequential modeling, enhancing classification performance.
4) **BiLSTM with Attention Mechanism:**
Further, we extended the BiLSTM architecture by incorporating an attention mechanism to selectively focus on the most informative parts of a sentence. Thus, the attention layer assigns higher weights to relevant tokens, allowing the model to better catch hate-related cues in complex or long news text. This architecture enhances interpretability and strengthens the detection of subtle hateful expressions. These deep learning model(s) offer a complementary view to classical Machine Learning classifiers and allow insightful understanding of linguistic patterns in Urdu news hate-speech detection.

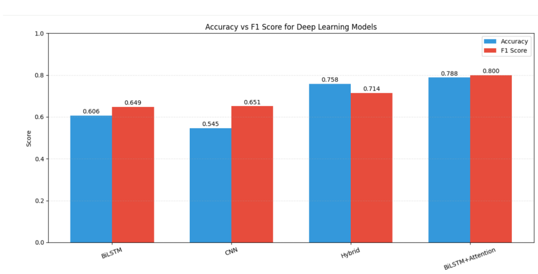

Figure 5. Accuracy and F1 Graph for Deep Learning model(s)

To further evaluate the threshold-independent performance of the deep-learning architectures, ROC curves were generated for all four model(s): BiLSTM, CNN, Hybrid CNN–BiLSTM, and BiLSTM with Attention. The ROC curve illustrates the relationship between the True Positive Rate (TPR) and False Positive Rate (FPR) across varying decision thresholds, providing a comprehensive view of each model's discriminative ability. The Area Under the Curve (AUC) was useed as the primary comparative metric, where higher AUC values indicate stronger classificaton performance. Among the deep-learning approaches, the BiLSTM with Attention achieved the highest AUC, demonstrating its superior ability to focus on contextually relevant segments within Urdu news articles, while the hybrid CNN–BiLSTM model also showed competitive performance. These results highlight the benefit of incorporating sequential modeling, and attention mechanisms for effective hate-speech detection in complex news narratives.

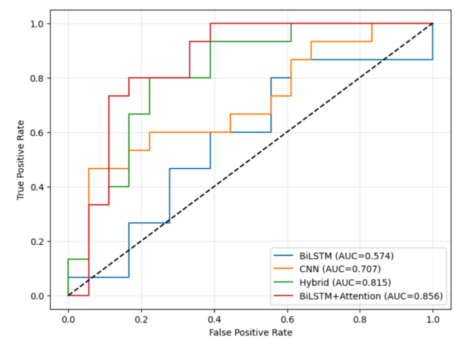

Figure 6. ROC curve for Deep Learning model(s)

V. Results and Evaluation

This section presents the experimental results for all machine-learning (Machine Learning) and deep-learning (Deep Learning) model(s) using Accuracy, Precision, Recall, and F1-Score as primary evaluation metrics. Table II summarizes the performance of the ten model(s) tested on the Urdu news hate-speech dataset.

A. Performance of Machine Learning model(s)

It is among the top classical Machine Learning models, with XGBoost and the Stacking Classifier both giving the best results: 0.90 Accuracy, 0.87 Precision, 0.93 Recall, and an F1-score of 0.90. Their strong results would point toward ensemble-based learners as being particularly effective for structured news texts by promoting feature diversity and boosted decision boundaries.

Random Forests also performed quite competitively, with an Accuracy of 0.84 and F1-score of 0.83, further establishing its robustness in handling lexicl variations within Urdu content. Linear SVM yielded a moderate performance with an Accuracy of 0.78 and F1-score of 0.77, thus motivating its strength in high-dimensional sparse feature spaces. Logistic Regression, with an accuracy of 0.75, and Naïve Bayes, with an accuracy of 0.76, had a reasonable baseline performance but were less effective as compared to the ensemble model(s).

B. Performance of Deep Learning model(s)

Deep-learning model(s) showed mixed performance on this dataset. BiLSTM with Attention produced the strongest results among Deep Learning model(s), with an Accuracy of 0.78, Recall of 0.93, and an F1-score of 0.80. The attention mechanism enhances the model's ability to focus on contextually important tokens, which is beneficial when identifying hate cues embedded within longer, formal news sentences. The Hybrid CNN–BiLSTM achieved model an Accuracy of 0.75 and an F1-score of 0.71, proving that combining spatial and sequential learning is superior to the standalone architectures. This is while the performance of the plain model BiLSTM was rather moderate: Accuracy 0.60 and F1 0.64. In its turn, the standalone CNN model showed a weak performance, with an Accuracy of 0.54, but surprisingly produced a very high Recall of 0.93, which underlines that CNN tends to overclassify hate content and is less stable for long-form news articles.

C. Comparison Between Machine Learning and Deep Learning Approaches

Overall, the Machine Learning model significantly outperformed the Deep Learning model on this dataset. This is largely attributed to:

The news text in Urdu is normally structured and formal, which suits TF-IDF-based linear and ensemble methods well.

The dataset size is moderate, hence less suitable for training deep neural architectures, which need large labeled corpora.

The effectiveness of lexicon-driven labels combined with sparse representations, favoring traditional model(s) supervised.

XGBoost and the Stacking Classifier yield state-of-the-art performance on binary hate-speech detection for Urdu news content.

Table II
Performance Comparison of Machine Learning and Deep model(s) Learning

Model	Accuracy	Precision	Recall	F1
XGBoost	**0.90**	**0.87**	**0.93**	**0.90**
Random Forest	0.84	0.81	0.86	0.83
Linear SVM	0.78	0.75	0.80	0.77
Logistic Regression	0.75	0.70	0.77	0.75
Naïve Bayes	0.76	0.70	0.79	0.76
Stacking Classifier	**0.90**	**0.87**	**0.93**	**0.90**
BiLSTM	0.60	0.54	0.80	0.64
CNN	0.54	0.50	**0.93**	0.65
Hybrid CNN–BiLSTM	0.75	0.76	0.66	0.71
BiLSTM + Attention	0.78	0.70	**0.93**	0.80

D. RMSE Analysis

To further model evaluate stability, the Root Mean Squared Error (RMSE) was calculated for all Machine Learning classifiers. Lower RMSE values correspond to more reliable predictions and reduced deviation from true labels. Ensemble-based (XGBoost and Stacking) model(s) produced the lowest RMSE scores, confirming their superior consistency and robustness across the dataset. RMSE results align closely with Accuracy and F1 trends, reinforcing the dominance of these deep learning model(s).

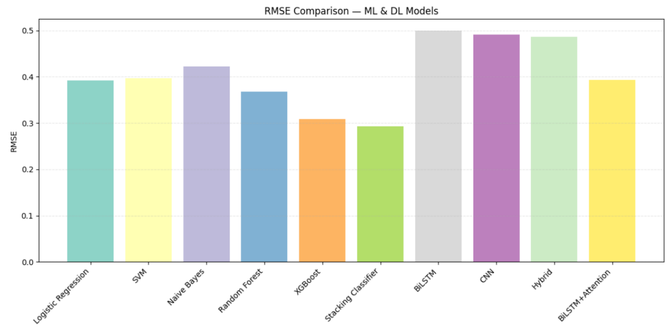

Figure 7. Rmse Grpah for all Machine Learning model(s)

E. ROC Curve Evaluation

Receiver Operating Characteristic (ROC) curves were generated to measure threshold-independent performance. For Machine Learning model(s), Stacking and XGBoost achieved the highest AUC values, with curves bending closest to the top-left corner. These results confirm their excellent discriminative capability.

For Deep Learning model(s), BiLSTM with Attention produced the highest AUC, outperforming other deep architectures. This again demonstrates the benefit of attention-based mechanisms in identifying informative regions of long Urdu sentences.

VI. Discussion

The experimental results show that the ensemble-based machine-learning model(s), particularly XGBoost and the Stacking Classifier, perform better than classical linear methods and deep-learning architectures on the Urdu news dataset. This trend is indicative of the structured nature of news text, whereby TF–IDF features effectively catch explicit hate expressions, hence allowing boosted and stacked model(s) to find strong decision boundaries. Although the deep-learning model(s) like BiLSTM and BiLSTM with Attention show promise, especially in capturing contextual dependencies and long-range relationships, the limited dataset size restricts their full strengths. The high recall of model(s) like CNN further suggests that deep networks are sensitive to hate-related tokens but their lower precision results from over.classificaton. Overall, these findings underscore the fact that for moderate-sized lexicon-labeled datasets, classical Machine Learning approaches remain more reliable compared to deep learning for the detection of hate speech in Urdu news.

A. Limitations

Despite the outstanding performance of the machine-learning model(s), the current study has a number of important limitations. First, the dataset was labeled using a lexicon-based approach, which detects explicit hate terms but fails to capture implicit, sarcastic, or context-driven hate speech. Second, the overall dataset size remains modest, which restricts the full potential of deep-learning architectures that usually require large-scale annotated corpora to generalize well., What's more, all scraped articles originate from four major news sources only, which may not represent the full diversity of Urdu media narratives. Finally, the study limits itself to binary classificaton, leaving out multi-class categories such as offensive, abusive, or neutral lnguage that could provide a more fine-grained understanding of hate speech.

VII. Conclusion and Future Work

This paper proposes a complete Urdu news hate-speech detection framework that leverages machine learning and deep learning model(s). Classical machine learning approaches-turned-ensemble methods like XGBoost and Stacking Classifier perform the best, which showcases that when combined with TF-IDF features, classical Machine Learning approaches remain highly effective for structured news text. Deep learning model(s), particularly BiLSTM with Attention, have shown promising recall; however, dataset size and lexicon-based labeling were major limitations in further improving its performance.

In the future, performance may be significantly improved with dataset expansion through manual annotation, consideration of news sources from a greater variety, and study of transformer-based model(s), such as BERT or XLM-R. A shift from binary to multi-class classification and the application of explainable AI techniques can further enhance interpretability and real-world applicability. yment in real-world media monitoring and content moderation systems.

yment in real-world media monitoring and content moderation systems.

References

[1] M. U. Arshad, R. Ali, M. O. Beg, and W. Shahzad, "UHated: Hate Speech Detection in Urdu Language Using Transfer Learning," *Language Resources and Evaluation*, 2023.

[2] M. P. Akhter, Z. Jiangbin, I. R. Naqvi, M. Abdelmajeed, and M. T. Sadiq, "Automatic detection of offensive language for Urdu and Roman Urdu," *IEEE Access*, vol. 8, pp. 91213–91226, 2020.

[3] M. P. Akhter, Z. Jiangbin, I. R. Naqvi, M. Abdelmajeed, and M. T. Sadiq, "Automatic detection of offensive language for Urdu and Roman Urdu," *IEEE Access*, vol. 8, pp. 91213–91226, 2020.

[4] H. S. Alatawi, A. M. Alhothali, and K. M. Moria, "Detecting white supremacist hate speech using domain specific word embedding with deep learning and BERT," *IEEE Access*, vol. 9, pp. 106363–106374, 2021.

[5] N. Albadi, M. Kurdi, and S. Mishra, "Are they our brothers? Analysis and detection of religious hate speech in the Arabic twittersphere," in *Proc. IEEE/ACM Int. Conf. Advances in Social Networks Analysis and Mining (ASONAM)*, pp. 69–76, 2018.

[6] M. Z. Ali, A. Ehsan-ul Haq, S. Rauf, K. Javed, and S. Hussain, "Improving hate speech detection of Urdu tweets using sentiment analysis," *IEEE Access*, vol. 9, pp. 84296–84305, 2021.

[7] O. Araque and C. A. Iglesias, "An ensemble method for radicalization and hate speech detection online empowered by sentic computing," *Cognitive Computation*, vol. 14, pp. 1–14, 2021.

[8] M. U. Arshad, M. F. Bashir, A. Majeed, W. Shahzad, and M. O. Beg, "Corpus for emotion detection on Roman Urdu," in *Proc. 22nd Int. Multitopic Conf. (INMIC)*, pp. 1–6, 2019.

[9] M. N. Awan and M. O. Beg, "Top-rank: A topicalpostionrank for extraction and classificaton of keyphrases in text," *Computer Speech & Language*, vol. 65, p. 101116, 2021.

[10] A. Baruah, F. Barbhuiya, and K. Dey, "ABARUAH at SemEval-2019 Task 5: Bi-directional LSTM for hate speech detection," in *Proc. 13th Int. Workshop on Semantic Evaluation*, 2019.

[11] D. Benito, O. Araque, and C. A. Iglesias, "GSI-UPM at SemEval-2019 Task 5: Semantic similarity and word embeddings for multilingual hate speech detection," in *Proc. 13th Int. Workshop on Semantic Evaluation*, 2019.

[12] A. Daud, W. Khan, and D. Che, "Urdu language processing: A survey," *Artificial Intelligence Review*, vol. 47, no. 3, pp. 279–311, 2017.

[13] T. Davidson, D. Warmsley, M. Macy, and I. Weber, "Automated hate speech detection and the problem of offensive language," 2017.

[14] A. Gertner et al., "MITRE at SemEval-2019 Task 5: Transfer learning for multilingual hate speech detection," in *Proc. 13th Int. Workshop on Semantic Evaluation*, 2019.

[15] A. Ghosh Chowdhury, A. Didolkar, R. Sawhney, and R. R. Shah, "ARHNet—Leveraging community interaction for detection of religious hate speech in Arabic," in *Proc. 57th Annual Meeting of ACL: Student Research Workshop*, Florence, Italy, 2019.