CUET-823 at SemEval-2026 Task 9: LoRA-Based Instruction Fine-Tuning of LLMs vs. Transformer Models for Bengali Polarization Detection

Arpita Mallik, Ratnajit Dhar
Department of Computer Science and Engineering
Chittagong University of Engineering and Technology, Bangladesh
{u2004023, u2004008}@student.cuet.ac.bd

Abstract

The rapid growth of social media has gone hand in hand with a sharp increase in heated public discussions, where debates about elections, conflicts, protests, and identity often turn into divisive and polarized rhetoric. In this paper, we present our system for SemEval 2026 Task 9 Subtask 1: Multilingual Text Classification Challenge-Polarization Detection, focusing specifically on the Bengali language. The task is a binary classification problem aimed at determining whether a social media post exhibits attitude polarization, such as intolerance, dehumanization, deindividuation, vilification, or stereotyping toward others opinions, identities, or beliefs. Among 49 participating teams, our approach ranked 2nd, achieving a macro-F1 score of 0.8582. We experimented with both transformer-based models and large language models (LLMs), and observed that LoRA-based instruction fine-tuned LLM-based approaches delivered the strongest performance in detecting nuanced and context-dependent polarization in Bengali text.

1 Introduction

Polarization refers to the expression of deeply divided views that present social or political issues as a clash between opposing groups. It frequently takes the form of "us versus them" terminology in online debates, where one group is presented as superior, moral, or right and the other as damaging, incorrect, or undeserving of respect. Stereotyping, demeaning, degrading, or rejecting the opinions and identities of others are examples of it.

Detecting polarization is particularly important in Bangladesh, where political divisions are closely tied to competing views of national identity and the legacy of the 1971 Liberation War (Rahman, 2019). Major political parties have propagated various accounts of the nation's origins since independence. These disagreements extend beyond elections and often shape broader social tensions. Because social media platforms have the ability to quickly magnify divided narratives and increase mistrust during delicate times, it is imperative to recognize polarized content on these platforms.

Label	Text
Non-Polarized	প্রিয় ভাই ও বোনেরা আমি ফুটবলের দেশ ব্রাজিলের পরিবেশ ও কালচার নিয়ে ভিডিও বানাই আমার স্বপ্ন পূরনের জন্য আপনাদের সাপোর্টের অনেক প্রয়োজন
Polarized	আরেক মিথ্যা বাদি কাওয়া তোরা ভোট চোর নির্লজ্জ বেহায়া লজ্জা শরম বলতে কিছু নেই ফেরাউন দলের নেতা কর্মীদের

Table 1: Example instances of polarized and non-polarized texts.

The aim of this work has been to identify polarized material in Bengali social media posts. The POLAR @ SemEval-2026 Task 9 (Naseem et al., 2026a) has introduced a dataset (Naseem et al., 2026b) under Subtask 1, which consisted of texts labeled as polarized (1) or non-polarized (0). The dataset included discussions about real world events like elections, conflicts, and social issues. It defined polarization not only by topic but also by how opinions are expressed, such as through stereotyping, vilification, or extreme language, making it easier to systematically identify divisive content.

We have experimented with a number of transformer-based models, such as mBERT, MuRIL, Bangla-BERT and XLM-RoBERTa, in order to accomplish our goal. We have also looked into ensemble combinations to improve performance. In parallel, we have evaluated a range of lightweight and adaptable LLMs, such as LLaMA-8B, Gemma-4B, Mistral-7B and Qwen3-14B, with both Bangla and English prompts. Our results have demonstrated that instruction fine-

Proceedings of the The 20th International Workshop on Semantic Evaluation (2026), pages 1128–1135

tuned LLMs have consistently outperformed fine-tuned transformer models in capturing nuanced and context-dependent polarization. The main contributions of our work are summarized as follows:

- We have conducted a systematic comparison of transformer models, ensembles, and large language models for Bengali polarization detection.
- We have demonstrated that LoRA-based instruction fine-tuned LLMs outperform traditional fine-tuning techniques.
- We have examined the effect of prompt language on model effectiveness.

Additional implementation details can be acquired via the GitHub repository. [1].

2 Related Work

Existing research on Bangla social-media opinion mining has mostly been framed as sentiment analysis (positive/negative/neutral) or stance (pro/against/neutral). While related, these settings are not equivalent to online polarization, which is better characterized by antagonistic framing, group-based positioning, and context-dependent rhetoric that may not align with sentiment labels. The POLAR @ SemEval-2026 (Naseem et al., 2026a) Task 9, Subtask-1 formalizes this distinction by defining polarization detection as a binary classification problem: deciding whether a post contains one or more polarized characteristics across multilingual and multi-event contexts.

Earlier Bangla work primarily relied on lexical feature pipelines. For instance, (Al Kaiser et al., 2021) uses TF–IDF n-grams with traditional classifiers for sentiment polarity in Facebook comments, and Rahman et al., 2023 applies TF–IDF word n-grams with Random Forest and SVM variants for tweet-level polarity classification. These approaches demonstrate the usefulness of surface lexical cues for sentiment-oriented objectives, but they are less aligned with polarization detection, where cues can be implicit and require broader contextual understanding.

More recent studies adopt transformer fine-tuning on Bangla political and social-media data. Hasan et al., 2023 show strong results with BanglaBERT on a Bangladesh YouTube dataset about the Russia–Ukraine war, supporting transformer encoders as competitive baselines for socio-political classification. Complementary dataset work such as SentNoB (Islam et al., 2021) highlights the difficulty of learning from noisy, informal Bangla, reinforcing that robust modeling remains challenging beyond clean sentiment benchmarks, an issue that also impacts binary polarization detection.

Beyond sentiment, closely related benchmarks emphasize that even strong models can behave inconsistently under ambiguity. BanglaBias (Lia et al., 2025) reports that LLMs perform better on explicitly opinionated articles than on harder, less overt cases, and Thapa et al., 2023 show that Bangla language models can encode political inclinations depending on training data. While our task uses only two labels (polarized vs. non-polarized), these findings still motivate careful evaluation on subtle cases where polarization is not expressed through explicit keywords.

In contrast to prior Bangla studies that primarily target sentiment/stance and focus on classical ML or standard transformer fine-tuning, none of the above explore parameter-efficient instruction fine-tuned LLMs for binary polarization detection. Motivated by this gap, our work has systematically evaluated transformer encoders against instruction fine-tuned, parameter-efficient LLMs, and compared Bangla and English prompts for detecting nuanced polarization in Bangla social-media text.

3 Dataset Analysis

We have used the Bengali dataset released for SemEval2026 Task 9, Subtask 1, formulated as a binary classification task for detecting polarization in social media text. The dataset comprises 5,000 instances, split into 3,333 training, 166 development, and 1,501 test samples. Approximately 42% of the instances have been labeled as polarized, and the class distribution has remained consistent across splits, indicating minimal label imbalance or distribution shift.

The dataset primarily consists of short-form posts, with most instances ranging between 1535 words. Lexical analysis has revealed significant vocabulary divergence between splits. The average rare-token ratio has increased from 0.11 in the training set to 0.17 in the test set, and approximately 45% of token types in the test set have not

[1] https://github.com/ArpitaMallik/SemEval-2026_Polarization-Detection-1

Split	Total	Non-Polarized	Polarized
Train	3333	1909	1424
Dev	166	96	70
Test	1501	868	633
Total	5000	2873	2127

Table 2: Label distribution across dataset.

appeared in the training data. These findings suggest that effective performance requires contextual generalization rather than reliance on surface-level lexical memorization.

4 Methodology

4.1 Transformer-based Training Setup

We have formulated polarization detection as a binary sequence classification task and have fine-tuned pre-trained transformer encoders (mBERT, XLM-RoBERTa, MuRIL, and Bangla-BERT) by adding a two-class classification head. The input texts have been tokenized using the respective model tokenizers with a maximum length of 256 tokens, applying truncation and padding.

We have performed an 85–15 stratified split of the training data for internal validation. The models have been trained for four epochs using AdamW with a learning rate of 2×10^{-5} and weight decay of 0.01. To address class imbalance, we have used class-weighted cross-entropy loss. We have selected the best checkpoint based on macro F1-score with early stopping, and have then retrained the model on the full training data before generating test predictions.

4.2 LLM-based Setup

We fine-tune instruction-following LLMs using LoRA-based supervised instruction fine-tuning. Each training instance is formatted as a chat-style prompt with a system instruction and user query, and the model is trained to generate a single-token label (0 or 1). We have used the Qwen3-14B model through the Unsloth framework with a maximum sequence length of 2048, and we have trained the model in 4-bit quantized mode to reduce memory usage while keeping the base model frozen.

We have applied Low-Rank Adaptation (LoRA) to update only a small set of trainable parameters on attention and MLP projection layers (q_proj, k_proj, v_proj, o_proj, gate_proj, up_proj, down_proj), using rank $r = 32$ and $\alpha = 32$ with no dropout. Training has been performed with SFT using a micro-batch size of 1 and gradient accumulation of 4, linear warmup (20 steps), AdamW in 8-bit mode, and a learning rate of 2×10^{-4} with weight decay 0.001. All LLMs reported in Table 4 (Qwen3-14B, LLaMA-8B, Gemma-4B, and Mistral-7B) were fine-tuned using the same LoRA-based supervised fine-tuning setup described above, with identical hyperparameters. During inference, we have generated exactly one new token with greedy decoding (temperature 0) to ensure the output has remained a single digit, and we have extracted the final prediction by matching {0,1} from the generated text. We experimented with both Bengali and English prompt templates across multiple LLM backbones to examine the effect of instruction language. The Bengali prompt template is shown below.

Bangla Prompt Template

System:
তুমি একজন বিশেষজ্ঞ, যার কাজ হলো বাংলা লেখায় পোলারায়িত (polarized) মতামত শনাক্ত করা। পোলারায়িত লেখা সাধারণত বিভাজন, দলভিত্তিক ঘৃণা, স্টেরিওটাইপিং, অপমান, অবমাননা, বা কোনো গোষ্ঠী বা মতাদর্শের বিরুদ্ধে তীব্র অসহিষ্ণুতা প্রকাশ করে।

User:
নিচের বাংলা লেখাটি পোলারায়িত (1) নাকি অপোলারায়িত (0) তা নির্ধারণ করো। শুধুমাত্র একটি সংখ্যা দেবে: 0 অথবা 1।
লেখা: {text}

5 Results and Analysis

5.1 Transformer-based Models

We begin with fine-tuned transformer encoders for binary polarization detection (Table 3). Overall, multilingual models have outperformed the Bengali-specific encoder, with MuRIL emerging as the strongest single model. This suggests that multilingual pretraining has helped the model capture the diverse ways polarization appears in social media text.

Combining all encoders through ensembling has yielded a modest improvement, increasing Macro F1 to 0.8407. The limited gain has indicated diminishing returns under standard fine-tuning, suggesting that encoder-based approaches may be nearing their performance limits in this setting.

Model	Macro F1	Accuracy	F1 (Polarized)	F1 (Non-Polarized)
XLM-RoBERTa-base	0.8303	0.8341	0.8050	0.8557
mBERT	0.7925	0.7981	0.7582	0.8268
MuRIL	0.8400	0.8441	0.8143	0.8657
Bangla-BERT	0.8031	0.8081	0.7718	0.8345
Ensemble (All Models)	**0.8407**	0.8675	0.8358	0.8456

Table 3: Performance of transformer-based models.

5.2 Large Language Models

We have next evaluated instruction fine-tuned LLMs with both Bengali and English prompts (Table 4). The strongest result has been achieved by Qwen3-14B under a Bengali prompt, which surpassed the transformer ensemble.

LLaMA-8B delivers competitive and stable performance across both prompt languages; with a slight advantage under English prompts, while Gemma-4B trailed somewhat behind, likely reflecting capacity differences.

Notably, Mistral-7B failed to detect polarized instances, effectively collapsing to majority-class predictions across both prompt languages. This behavior suggests that the model did not successfully adapt to the task during fine-tuning, possibly due to weaker multilingual capabilities or instability under low-resource LoRA-based training. This collapse highlights that instruction fine-tuning alone does not guarantee reliable cross-lingual generalization.

Unlike encoder models that depend on supervised fine tuning to learn representations, LLMs rely on instruction tuning and their ability to reason from context during inference. The balanced class-wise performance of Qwen3-14B suggests stronger handling of subtle or indirect polarization signals.

Overall, the results indicate that while prompt design influences performance, its effect is model-dependent. Bengali prompts improve performance for some models (e.g., Qwen3-14B and Gemma-4B), whereas others (e.g., LLaMA-8B) perform marginally better with English prompts. This variation likely reflects differences in pretraining data distribution and instruction-following behavior across models. Given the small performance gap, these differences should be interpreted cautiously.

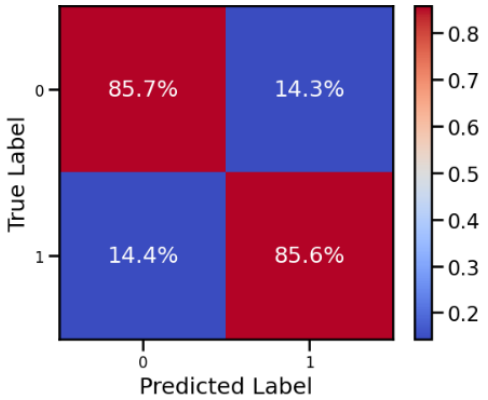

Figure 1: Confusion matrix of Qwen3-14B with Bengali prompt.

5.3 Error Analysis

We have compared our best-performing LLM, Qwen3-14B with a Bangla prompt, against our strongest transformer baseline, MuRIL. On the test set, Qwen3-14B has correctly classified 105 instances that were misclassified by MuRIL, while MuRIL has corrected 64 cases missed by the LLM.

As shown in Figure 1, the confusion matrix of Qwen3-14B has demonstrated balanced performance across both classes, with 85.7% of non-polarized instances and 85.6% of polarized instances correctly classified. The corresponding confusion matrices for both Qwen3-14B and MuRIL have been provided in Appendix A.3. Notably, Qwen3-14B has reduced false negatives to 91 compared to 143 for MuRIL, indicating improved sensitivity to polarized content. Although false positives increase slightly (124 vs. 113), the LLM maintains stable class-wise performance while improving recall for implicit and context-dependent polarization.

Qualitative analysis (Table 5) has further shown that Qwen3-14B better captures metaphorical dehumanization, indirect group targeting, and rhetorically framed political hostility. However, both

Model	Prompt	Macro F1	Accuracy	F1 (Polarized)	F1 (Non-Polarized)
Qwen3-14B	Bengali	**0.8582**	0.8608	0.8394	0.8770
Qwen3-14B	English	0.8541	0.8568	0.8289	0.8718
LLaMA-8B	Bengali	0.8473	0.8508	0.8245	0.8702
LLaMA-8B	English	0.8493	0.8521	0.8287	0.8699
Gemma-4B	Bengali	0.8246	0.8289	0.8041	0.8451
Gemma-4B	English	0.8192	0.8237	0.7984	0.8400
Mistral-7B	Bengali	0.3664	0.5783	0.0000	0.7328
Mistral-7B	English	0.3664	0.5783	0.0000	0.7328

Table 4: Performance of instruction fine-tuned LLMs with Bengali and English prompts.

Text	Label	MuRIL	Qwen3-14B	Observation
রাস্তার কুকুর চিৎকার করলে মানুষ তাকে লাঠি পেটা করে কোলে তুলে নাচে না আর এ শারীরিক ও বুদ্ধি প্রতিবন্ধীর কথা শুনে লোকে বিরক্ত বোধ করে এর বেশী কিছু না	1	0	1	Metaphorical dehumanization; LLM captures implicit insult
ভারতের তাবেদারি থেকে দেশকে বাচাতে এই দানব সরকারের বিদায় করতে হবে নয়তো দেশের সবাইকে চরম মূল্য দিতে হবে	1	0	1	Political framing with hostile rhetoric, LLM detects intense tone missed by encoder
আমার ভাবে খালেদা আর তারেকের কথা বললেন মনে হচ্ছে তারা দুইজন মা ছেলে ফেরেস্তা ছিলো একজন বাসায় এবং হসপিটালে চিকিৎসা নিচ্ছেন সাজা প্রাপ্ত আসামি হয়েও আর একজন চিকিৎসার নাম করে বিদেশে পালিয়ে আছে তাহলে আপনি একজন শিক্ষত মানুষ হয়েও কেন এতো মিথ্যা বলেন যে অন্যায় করছে সে সেই শাস্তি পাবেই এটাই স্বাভাবিক	0	1	1	Strong political criticism without explicit group-based polarization, LLM slightly over-predicts
খুব খারাপ চালিয়াতী অন্যের সাথে করতে করতে নিজদের মধ্যেও বাদ যায়না তাই বজ্জাতি বন্ধ করে ভাল হয়ে যাও	1	0	0	Hostile language with unclear out-group target; ambiguity affects both models

Table 5: Representative qualitative examples comparing MuRIL and Qwen3-14B (Bangla prompt). The LLM better captures implicit and rhetorically framed polarization but is slightly more prone to over-predict in politically charged contexts.

models struggled when strong sentiment lacked a clear target group, reflecting the inherent ambiguity of polarization detection.

6 Conclusion

We have presented a polarization detection system for SemEval2026 Task 9 (Subtask 1) in Bengali, formulated as a binary classification task on social media text. We have compared transformer baselines, including ensembles, with instruction fine-tuned LLMs using both Bangla and English prompts. Overall, instruction fine-tuned LLMs effectively capture nuanced and context-dependent polarization, with Qwen3-14B achieving the best performance (macro F1 of 0.8582). While Bengali prompts slightly improved performance for most models, LLaMA-8B performed marginally better with English prompts, suggesting that the effect of prompt language is model-dependent and influenced by pretraining characteristics. These findings suggest that efficiently adapted instruction-tuned models can be effective for polarization detection in low-resource settings.

For future work, we plan to evaluate cross-lingual generalization, improve prompt robustness and calibration, and incorporate interpretability and uncertainty estimation for more reliable deployment. We also aim to explore multimodal approaches and evaluate on newer, time-evolving datasets reflecting emerging topics and platform shifts.

Limitations

This study has several limitations. The model has remained text-only and has not incorporated multimodal signals such as audio, video, or images. Experiments have been conducted solely on Bengali data, leaving cross-lingual generalization uncer-

tain. Additionally, because polarization patterns evolve with new political events, social-media narratives, and public discourse, performance on future or previously unseen data may differ from the results reported here.

Model behavior has also been sensitive to prompt formulation and language choice. While the LLM has better captured metaphorical dehumanization and rhetorically framed hostility, it has occasionally over-predicted polarization in strongly critical but non-group-targeted statements. Both models have struggled with ambiguous cases lacking a clear out-group, highlighting the inherent subjectivity of polarization detection. Additionally, although our error analysis indicates complementary strengths between transformer-based models and LLMs, we did not explore hybrid ensembling approaches, which could potentially further improve performance. Finally, only open-source locally runnable models were evaluated due to cost constraints.

7 Ethics Statement

This work supports research and moderation for polarization detection, but such systems may be misused or reflect biases. We stress responsible use, clear reporting of limitations, and human oversight before deployment.

Acknowledgments

We thank the organizers of SemEval2026 Task 9 for organizing the competition and for providing the dataset used in this work.

References

Shad Al Kaiser, Sudipta Mandal, Ashraful Kalam Abid, Ekhfa Hossain, Ferdous Bin Ali, and Intisar Tahmid Naheen. 2021. Social media opinion mining based on bangla public post of facebook. In *2021 24th International Conference on Computer and Information Technology (ICCIT)*, pages 1–6. IEEE.

Mahmud Hasan, Labiba Islam, Ismat Jahan, Sabrina Mannan Meem, and Rashedur M Rahman. 2023. Natural language processing and sentiment analysis on bangla social media comments on russia–ukraine war using transformers. *Vietnam Journal of Computer Science*, 10(03):329–356.

Khondoker Ittehadul Islam, Sudipta Kar, Md Saiful Islam, and Mohammad Ruhul Amin. 2021. Sentnob: A dataset for analysing sentiment on noisy bangla texts. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 3265–3271.

Nusrat Jahan Lia, Shubhashis Roy Dipta, Abdullah Khan Zehady, Naymul Islam, Madhusodan Chakraborty, and Abdullah Al Wasif. 2025. Read between the lines: A benchmark for uncovering political bias in bangla news articles. In *Proceedings of the Second Workshop on Bangla Language Processing (BLP-2025)*, pages 61–79.

Usman Naseem, Robert Geislinger, Juan Ren, Sarah Kohail, Rudy Garrido Veliz, P Sam Sahil, Yiran Zhang, Marco Antonio Stranisci, Idris Abdulmumin, "Ozge Alacam, Cengiz Acar"urk, Aisha Jabr, Saba Anwar, Abinew Ali Ayele, Elena Tutubalina, Aung Kyaw Htet, Xintong Wang, Surendrabikram Thapa, Tanmoy Chakraborty, and 15 others. 2026a. SemEval-2026 task 9: Detecting multilingual, multicultural and multievent online polarization. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*. Association for Computational Linguistics.

Usman Naseem, Robert Geislinger, Juan Ren, Sarah Kohail, Rudy Garrido Veliz, P Sam Sahil, Yiran Zhang, Marco Antonio Stranisci, Idris Abdulmumin, Özge Alacam, Cengiz Acartürk, Aisha Jabr, Saba Anwar, Abinew Ali Ayele, Simona Frenda, Alessandra Teresa Cignarella, Elena Tutubalina, Oleg Rogov, Aung Kyaw Htet, and 24 others. 2026b. Polar: A benchmark for multilingual, multicultural, and multi-event online polarization. *Preprint*, arXiv:2505.20624.

Habibur Rahman, Md Munam Kazi, Md Samiul Islam, Jannatul Maoya, Md Monjurul Arif, Mahir Tagwar, Mahamudul Hasan, Md Sawkat Ali, Taskeed Jabid, and Maheen Islam. 2023. An analysis of bangla tweets on social media platform for polarity detection using machine learning algorithms. In *2023 4th International Conference on Big Data Analytics and Practices (IBDAP)*, pages 1–6. IEEE.

Tahmina Rahman. 2019. Party system institutionalization and pernicious polarization in bangladesh. *The ANNALS of the American Academy of Political and Social Science*, 681(1):173–192.

Surendrabikram Thapa, Ashwarya Maratha, Khan Md Hasib, Mehwish Nasim, and Usman Naseem. 2023. Assessing political inclination of bangla language models. In *Proceedings of the first workshop on Bangla language processing (BLP-2023)*, pages 62–71.

A Appendix

This appendix reports the full implementation details and hyperparameters for both transformer-based models and instruction-tuned LLMs, corresponding exactly to the configurations used in the main paper to ensure reproducibility.

A.1 Transformer Models

Table 6 summarizes the training configuration for the transformer-based models (XLM-RoBERTa,

Parameter	Value
Maximum sequence length	256
Train–validation split	85–15 (stratified)
Optimizer	AdamW
Learning rate	2×10^{-5}
Number of epochs	4
Training batch size	16
Evaluation batch size	32
Weight decay	0.01
Warmup steps	100
Evaluation steps	100
Early stopping	Patience = 3
Model selection metric	Macro F1
Loss function	Class-weighted cross-entropy
Precision	FP16 (CUDA)

Table 6: Training hyperparameters for transformer-based models.

Parameter	Value
Max sequence length	2048
Quantization	4-bit
Adaptation method	LoRA
LoRA rank (r)	32
LoRA alpha	32
LoRA dropout	0
Optimizer	AdamW (8-bit)
Learning rate	2×10^{-4}
Batch size	1
Gradient accumulation steps	4
Warmup steps	20
Weight decay	0.001
Scheduler	Linear
Decoding strategy	Greedy
Max new tokens	1

Table 7: Hyperparameters for the LLM models experiments.

mBERT, MuRIL and Bangla-BERT), including sequence length, data splits, optimization settings, and model selection criteria. Class-weighted cross-entropy addressed label imbalance, and early stopping based on Macro F1 prevented overfitting.

A.2 LLM Models

Table 7 summarizes the hyperparameters used for the LLM models, including LLaMA-3 (8B), Gemma-3 (4B), Qwen (14B), and Mistral (7B). The models were adapted using parameter-efficient fine-tuning with 4-bit quantization to reduce memory footprint while maintaining performance. The table details LoRA configuration, optimization settings, training schedule, and decoding strategy employed during evaluation.

A.3 Confusion Matrices

Figure 2 and Figure 3 present the confusion matrices for Qwen3-14B (Bangla prompt) and MuRIL, respectively. These visualizations highlight class-wise prediction behavior and provide insight into false positives and false negatives.

A.4 Lexical Analysis Visualization

Figures 4 and 5 visualize class-specific lexical patterns using word clouds generated from the top log-odds tokens for each class. The non-polarized class contains more neutral and informational terms, whereas the polarized class high-

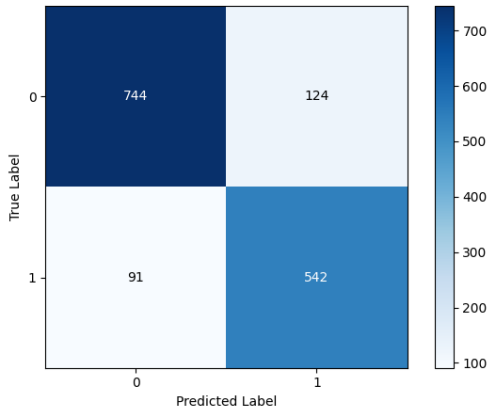

Figure 2: Confusion matrix of Qwen3-14B with Bangla prompt.

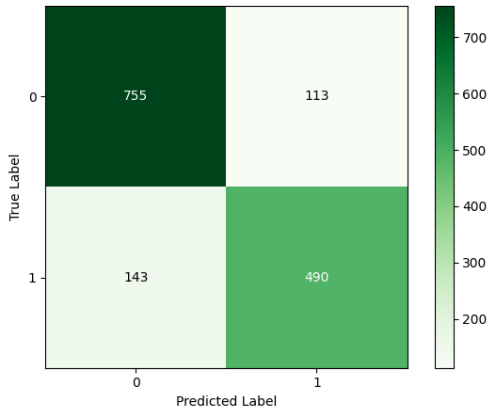

Figure 3: Confusion matrix of MuRIL.

Figure 4: Word cloud for the non-polarized class generated from top log-odds tokens. The vocabulary is predominantly neutral and informational in nature.

Figure 5: Word cloud for the polarized class generated from top log-odds tokens. The distribution highlights identity references and evaluative expressions characteristic of polarized discourse.

lights identity-related, evaluative, and politically charged expressions. These patterns support our observation that polarization is shaped not only by topic, but also by rhetorical framing and group-directed language.