SAGI RAMAKRISHNAM RAJU ENGINEERING COLLEGE(A), BHIMAVARAM
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

HINDI FORMALITY STYLE TRANSFER

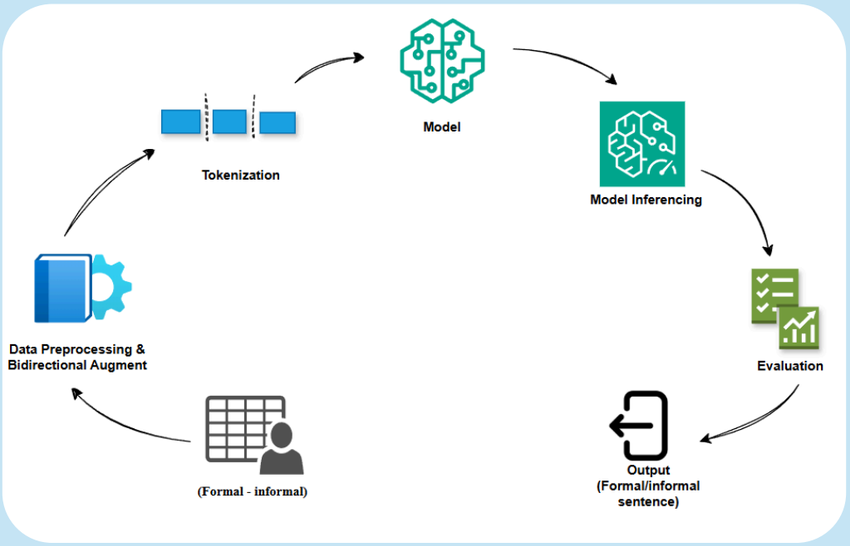
SUPERVISOR: Dr B N V NARASIMHA RAJU

B 2

Abstract

This work introduces a bidirectional Hindi formality style transfer model using IndicBART and mT5, addressing challenges like data scarcity and linguistic complexity.Back-translation-based data augmentation supports effective formalization and informalization within a unified framework.The model outperforms baselines like GPT-2 and mBART in semantic preservation and stylistic accuracy. It enables impactful applications in social media, professional writing, and AI-assisted localization.

Architecture



Methodology

Data Collection

A manually curated parallel corpus of Hindi sentence pairs was loaded (formalize: or informalize:) and vice-versa

Preprocessing

Tokenization
T5Tokenizer (for mT5) and AutoTokenizer (for IndicBART)

Model Selection

Hindi Language models: mT5-Small,IndicBART with Fine Tuning of mC4 and IndicCorp based on the parallel formality data then Ensembling both models using GPT-2

Evaluation

Final predictions are evaluated using:
BLEU Score
BERTScore
Perplexity

Results

Model	BLEU Score	BERTscore F1	BERTscore Precision	BERTscore recall	Perplexity
mT5	**23.31**	**0.85**	**0.85**	**0.85**	8.25
IndicBART	17.84	0.81	0.82	0.81	5.36
Ensemble Model	21.25	0.84	0.84	0.84	6.20
GPT-2	0.26	0.60	0.58	0.62	3.13
IndicGPT	20.01	0.63	0.62	0.65	41.77
mBART	8.91	0.77	0.80	0.74	15.07

Conclusion

Adapting textual formality is crucial for effective communication, especially in languages like Hindi. Hindi formality style transfer faces challenges due to complex grammar, honorifics, and the lack of high-quality datasets. By fine-tuning mT5 and IndicBART on a 50k-sentence corpus using bidirectional prompting, strong improvements were achieved. Evaluation using BLEU, BERTScore, and Perplexity confirmed better accuracy, fluency, and semantic preservation. Overall, mT5 outperformed other models, setting a new benchmark for Hindi formality style transfer.

SUBMITTED BY:
K VARALAKSHMI (21B91A6128), P L PRASANNA (22B95A6104)
V NARASIMHA RAJA (21B91A6161), K DEEPTANUSH (21B91A6130)