

*Dedicated to the scientific community*

# Declaration

I certify that

- The work contained in the dissertation is original and has been done by myself under the general supervision of my supervisor.
- The work has not been submitted to any other Institute for any degree or diploma.
- I have followed the guidelines provided by Tezpur University in writing the thesis.
- I have conformed to the norms and guidelines given in the Ethical Code of Conduct of the University.
- Whenever I have used materials (data, theoretical analysis, and text) from other sources, I have given due credit to them by citing them in the text of the dissertation and giving their details in the references.



Ashwani Gupta  
08.09.25



Department of Computer Science and Engineering  
Tezpur University

Napaam, Tezpur- 784028, Assam, India.

Dr. Utpal Sharma  
Professor

Phone: +91-3712-275107  
Fax: +91-3712-267005/6  
E-Mail: utpal@tezu.ernet.in

### Certificate of Supervisor

This is to certify that the thesis entitled “**Aspect-based sentiment analysis for Hindi document**” submitted to Tezpur University in the Department of Computer Science and Engineering under the School of Engineering in partial fulfillment of the award of the degree of Doctor of Philosophy in Computer Science and Engineering is a record of research work carried out by **Ashwani Gupta** under my supervision and guidance.

All helps received by him from various sources have been duly acknowledged. No part of this thesis has been submitted else where for award of any other degree.

Signature of Supervisor

(Utpal Sharma)

Professor

Department of Computer Science and Engineering

Tezpur University

Assam, India-784028



## Certificate

This is to certify that the thesis entitled “**Aspect-based sentiment analysis for Hindi document**” submitted by **Ashwani Gupta** to Tezpur University in the Department of Computer Science and Engineering under the School of Engineering in partial fulfillment of the requirements for the award of the degree of Doctor of Philosophy in Computer Science and Engineering has been examined by us on ..... and found to be satisfactory.

The Committee recommends for award of the degree of Doctor of Philosophy.

Signature of Principal Supervisor

Signature of External Examiner

# Acknowledgments

At the very beginning, I would like to mention that this thesis is an outcome of combined efforts made by many people. They have actually helped me a lot in materializing the works compiled in this thesis. I take this proud privilege to express my sincere obligations and gratitude towards all of them. This thesis would have never been finished with a great satisfaction without their constant support, advice and companionship.

I would like to express my sincere gratitude to my supervisor **Prof. Utpal Sharma** for his patience, guidance and admirable support throughout the entire process of writing this dissertation. I have learnt a lot in my time as his student not only about computer science but life in general. I could not have imagined having a better advisor and mentor for my PhD study.

I am indebted to the Department of Computer Science and Engineering of Tezpur University and its faculty members and non-teaching staff. I am particularly grateful to Prof. Sarat Saharia, Dr. Debojit Boro, Dr. Sarangthem Ibotombi Singh and Dr. Arindam Karmakar, who have been members of my doctoral committee.

It is a great pleasure to offer my sincere gratitude to Prof. Krishna Pal Singh, Hon'ble Vice Chancellor, MJP Rohilkhand University, Bareilly, UP and Faculty members, Department of Computer Science and Information Technology, Faculty of Engineering and Technology, MJP Rohilkhand University, Bareilly, UP, for their valuable supports. I am especially thankful to some of my colleagues of

---

Faculty of Engineering and Technology, MJP Rohilkhand University, Bareilly, UP, Brajesh Kumar, Asim Rehman Ansari and Ashish Sankhwar for their encouragement and friendly advice.

I am indebted to the All India Council for Technical Education (AICTE) for providing me with a National Fellowship (NF) under Quality Improvement Programme(QIP). I would also like to thank all the anonymous reviewers of the published works and the thesis for their constructive comments.

I am very fortunate that I met my friends Abhijit, Amit, Anurag, Devajit, Dilbar, Manash, N Rana Singha, Praganur, Prasiddha, Rahul, Rajesh, Rishang and Sunil who made my stay at Tezpur University pleasurable.

I am extremely grateful to my parents and family members for their blessings and continuous encouragement to achieve this big goal. Without their constant love, guidance, and support, I would not have been in a position to complete this research work.

Finally, I am highly obliged to the supreme power of the God and the elders for their blessings with me at every step of this journey because of their grace I have achieved my goal.

**Ashwani Gupta**

# List of Figures

3-1	Proposed methodology . . . . .	40
3-2	Evaluation for TU-HSA dataset(TF-IDF) . . . . .	51
3-3	Evaluation for TU-HSA(Count Vectorizer) . . . . .	52
3-4	Evaluation for TU-HSA(English)(TF-IDF) . . . . .	53
3-5	Evaluation for TU-HSA(English)(Count Vectorizer) . . . . .	54
4-1	Block diagram for ATE model . . . . .	57
4-2	Hindi sentence distribution between domains . . . . .	59
4-3	LSTM unit architecture . . . . .	60
4-4	BiLSTM unit architecture . . . . .	61
4-5	Loss function vs epochs . . . . .	65
5-1	Aspect category detection model . . . . .	71
5-2	Hindi sentence distribution between categories . . . . .	72
5-3	Hindi sentence distribution between categories . . . . .	73
5-4	Multi-label assignment . . . . .	73

## List of Figures

---

5-5	Accuracy versus epochs . . . . .	78
5-6	Loss versus epochs . . . . .	78
5-7	Aspect category-wise performance concerning precision, recall and F1-score . . . . .	79
5-8	Accuracy obtained across different mini-batch sizes and learning rates	80
5-9	F1-score comparison of the proposed model with state-of-art results	81
6-1	Multitask learning workflow . . . . .	86
6-2	Dataset snippet . . . . .	89



# List of Tables

1.1	ABSA . . . . .	4
1.2	Confusion matrix . . . . .	13
2.1	Summary of related work on Aspect Sentiment Classification . . .	26
2.2	Summary of multitask learning approaches for aspect-category de- tection and aspect sentiment classification . . . . .	28
2.3	Summary of Hindi datasets . . . . .	28
2.4	TU-HSA dataset details . . . . .	29
2.5	Domain-wise Dataset Details and Sources . . . . .	32
2.6	Some basic statistics . . . . .	33
2.7	IITP-I labelled structure . . . . .	34
2.8	Aspect categories corresponding to distinct domains . . . . .	34
2.9	IITP-II labelled structure . . . . .	35
3.1	Evaluation Metrics (TF-IDF) . . . . .	51
3.2	Evaluation Metrics (Count Vectorizer) . . . . .	51
3.3	Evaluation Metrics (TF-IDF) . . . . .	53

3.4	Evaluation Metrics (Count Vectorizer) . . . . .	54
4.1	Review and predicted labels-example I . . . . .	58
4.2	Some basic statistics . . . . .	58
4.3	Hyper-parameters for the ATE model . . . . .	64
4.4	Review and predicted labels-example II . . . . .	66
4.5	Review and predicted labels-example III . . . . .	66
4.6	Review and predicted labels-example IV . . . . .	66
4.7	Comparison of aspect term extraction methods for Hindi . . . . .	68
5.1	Aspect categories by domain . . . . .	72
5.2	Dimensions and optimal hyper-parameters for the aspect category detection model . . . . .	79
5.3	Comparison of F1-scores of Aspect Category Detection methods for Hindi . . . . .	82
6.1	Evaluation metrics for MLC . . . . .	94
6.2	Evaluation metrics for MCC . . . . .	95
6.3	Comparison of accuracy for multitasking model with English . . . .	95
A.1	Sample TU-HSA dataset . . . . .	115

# List of Abbreviations

ABSA	Aspect Based Sentiment Analysis.
ACD	Aspect Category Detection.
ASC	Aspect Sentiment Classification.
ATE	Aspect Term Extraction.
BiLSTM	Bidirectional Long-Short Term Memory.
BIO	Begin-Inside-Outside.
CNN	Convolutional Neural Network.
CPU	Central Processing Unit.
CRF	Conditional Random Field.
DL	Deep Learning.
DT	Decision Tree.
FNN	Forward Neural Network.
GCN	Graph Convolutional Network.
GPU	Graphical Processing Unit.
GRU	Gated Recurrent Unit.
GNB	Gaussian Naive Bayes.
IIT	Indian Institute of Technology.
IndicBERT	Indic Bidirectional Encoder Representations from Transformers.
IT	Information Technology.
KNN	$k$ -Nearest Neighbour.
LR	Logistic Regression.
LSTM	Long-Short Term Memory.
mBERT	multi-lingual Bidirectional Encoder Representations from Transformers.
ML	Machine Learning.
MLC	Multi-Label Classifier.
MNB	Multi-nomial Naive Bayes.
MOO	Multi-objective optimization.
MuRIL	Multilingual Representations for Indian Languages.
MCC	Multi-Class Classifier.
NB	Naive Bayes.
NER	Name Entity Recognition.
NLP	Natural Language Processing.
NN	Neural Network.
POS	Part-Of-Speech.
RAM	Random Access Memory.

ReLU	Rectified Linear Unit.
RF	Random Forest.
RNN	Recursive Neural Network.
SA	Sentiment Analysis.
SVM	Support Vector Machine.
TU-HSA	Tezpur University-Hindi Sentiment Analysis.
www	World Wide Web.
XGBoost	eXtreme Gradient Boosting.
XML	Extensible Markup Language.
XOR	Ex-OR.