



**Department of Electrical and Computer Engineering  
North South University**

## **Senior Design Project**

# **Bangla Hate Speech Detection: Comparative Analysis of Machine Learning Models and Recurrent Neural Networks**

**Abdullah Al Shafi                      ID# 19111515642**

**Miraj Hosain Shawon                ID# 1911329042**

**Rafiad Rahman Badhon              ID# 1722044642**

**Faculty Advisor:**

**Mohammad Monirujjaman Khan**

**Associate Professor**

**ECE Department**

**Spring, 2023**

# LETTER OF TRANSMITTAL

June, 2023

To

Dr. Rajesh Palit  
Chairman,  
Department of Electrical and Computer Engineering  
North South University, Dhaka

**Subject: Submission of Capstone Project Report on “Bangla Hate Speech Detection: Comparative Analysis of Machine Learning Models and Recurrent Neural Networks”**

Dear Sir,

With due respect, we would like to submit our **Capstone Project Report on “Bangla Hate Speech Detection: Comparative Analysis of Machine Learning Models and Recurrent Neural Networks”** as a part of our BSc program. The primary aim of this project is to demonstrate a Comparative Analysis between ML Models and RNN regarding the Detection of Hate Speech in the Bengali language. The dataset employed for this investigation is a privately acquired corpus obtained from various social media platforms. We tried to the maximum competence to meet all the dimensions required from this report.

We will be highly obliged if you kindly receive this report and provide your valuable judgment. It would be our immense pleasure if you find this report useful and informative to have an apparent perspective on the issue.

Sincerely Yours,

.....  
Abdullah Al Shafi  
ECE Department  
North South University, Bangladesh

.....  
Miraj Hossain Shawon  
ECE Department  
North South University, Bangladesh

.....  
Rafiad Rahman Badhon  
ECE Department  
North South University, Bangladesh

# APPROVAL

Abdullah Al Shafi (ID #1911515642 ), Miraj Hossain Shawon (ID # 1911329042), and Rafiad Rahman Badhon (ID # 1722044642) from Electrical and Computer Engineering Department of North South University, have worked on the Senior Design Project titled “Bangla Hate Speech Detection: Comparative Analysis of Machine Learning Models and Recurrent Neural Networks” under the supervision of Dr. Mohammad Monirujjaman Khan partial fulfillment of the requirement for the degree of Bachelors of Science in Engineering and has been accepted as satisfactory.

## Supervisor’s Signature

.....

**Dr. Mohammad Monirujjaman Khan**

**Associate Professor**

Department of Electrical and Computer Engineering

North South University

Dhaka, Bangladesh.

## Chairman’s Signature

.....

**Dr. Rajesh Palit**

**Professor**

Department of Electrical and Computer Engineering

North South University

Dhaka, Bangladesh.

# DECLARATION

This is to declare that this project is our original work. No part of this work has been submitted elsewhere partially or fully for the award of any other degree or diploma. All project-related information will remain confidential and shall not be disclosed without the formal consent of the project supervisor. Relevant previous works presented in this report have been properly acknowledged and cited. The plagiarism policy, as stated by the supervisor, has been maintained.

Students' names & Signatures

-----

**1. Abdullah Al Shafi**

-----

**2. Miraj Hossain Shawon**

-----

**3. Rafiad Rahman Badhon**

## ACKNOWLEDGEMENTS

The authors would like to express their heartfelt gratitude towards their project and research supervisor, Dr. Mohammad Monirujjaman Khan, Associate Professor, Department of Electrical and Computer Engineering, North South University, Bangladesh, for his invaluable support, precise guidance, and advice pertaining to the experiments, research and theoretical studies carried out during the course of the current project and also in the preparation of the current report.

Furthermore, the authors would like to thank the Department of Electrical and Computer Engineering, North South University, Bangladesh for facilitating the research. The authors would also like to thank their parents for their countless sacrifices and continual support.

## ABSTRACT

# **Bangla Hate Speech Detection: Comparative Analysis of Machine Learning Models and Recurrent Neural Networks**

The spread of hate speech on the internet has been related to an increase in violent acts committed against minority groups all across the world. Nowadays, usage of social media is at its peak, and so is hate speech on online social media. Nearly every continent has reported incidents. A majority of the world's population uses Facebook alone, and many people increasingly converse on social media. Although multiple hate speech detection paper has been done based on the Bengali language, there is no paper has been done based on comparative analysis of machine learning and recurrent neural networks regarding Bengali hate speech detection. Therefore, this study aims to train different models that can detect Bengali hate speech on different social media platforms and do a comparative analysis of the models. Bengali hate speech is on the rise on social media platforms, threatening the general public's mental health. Thus, detecting and preventing it from posting on social media is a suitable approach to preventing hate speech from spreading. Using statistical methods, the collected data were analyzed. Also, manually labeled the collected data based on sentiment. To remove punctuation marks, a punctuation remover is used, and regular expressions are used to remove foreign languages from the dataset. Moreover, Bangla natural language toolkit was used to remove Bangla stopwords from the data. A label encoding method is used to make the dataset machine readable. Natural language processing toolkit porter stemmer is used for tokenization and for feature extraction, term frequency-inverse document frequency is used for training and testing the dataset, hold-out validation approach was used. Several machine learning and recurrent neural network models, decision tree (DT), K-nearest neighbor (KNN), random forest (RF), support vector machine (SVM), multinomial naïve bayes (MNB), long short term memory (LSTM), bidirectional long short term memory (Bi-LSTM), and convolutional long short term memory (CNN-LSTM) were implemented. Among machine learning models, support vector machine gained 0.96 accuracy, and among recurrent neural network (RNN) models, bidirectional long short term memory gained 0.94 accuracy.