Department of Electrical and Computer Engineering
North South University

Directed Research

BANGLA PLAGIARISM CHECKER USING MACHINE LEARNING-MULTIPLE LINE APPROACH

MD SAIFUL ISLAM #1620578042

Faculty Advisor:

Dr. Nabeel Mohammed

Assistant Professor

ECE Department

Spring, 2023

LETTER OF TRANSMITTAL

24 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 Plagiarism Checker using Machine Learning-Multiple line Approach”.**

Dear Sir,

With due respect, I would like to submit my **Directed Research** on **“Bangla Plagiarism Checker using Machine Learning-Multiple line Approach”** as a part of our BSc program. The report deals with checking plagiarism for Bangla text, articles, books etc. This project was very much valuable to me as it helped me gain experience from practical fields and apply in real life. I tried to the maximum competence to meet all the dimensions required from this report.

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

Sincerely Yours,

....

Md Saiful Islam

ECE Department

North South University, Bangladesh

APPROVAL

Md Saiful Islam (ID # 1620578042),) from Electrical and Computer Engineering Department of North South University, have worked on the Senior Design Project titled "**Bangla Plagiarism Checker using Machine Learning-Multiple line Approach**" under the supervision of Dr.Nabeel Mohammed partial fulfillment of the requirement for the degree of Bachelors of Science in Engineering and has been accepted as satisfactory.

Supervisor's Signature

……………………………………

Dr. Nabeel Mohammed

Assistant 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 my 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

Md Saiful Islam

ACKNOWLEDGEMENTS

I would like to express my heartfelt gratitude towards my research supervisor, Dr. Nabeel Mohammed, Assistant 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, I would like to thank the Department of Electrical and Computer Engineering, North South University, Bangladesh for facilitating the research. I would also like to thank our loved ones for their countless sacrifices and continual support.

ABSTRACT

Bangla Plagiarism Checker using Machine Learning-Multiple Line Approach

Plagiarism is the practice of representing another person's ideas or works as your own, with or without that person's consent, by incorporating it into your work without giving it due credit. So, there are some excellent tools available to identify plagiarism, such Turnitin, Grammarly, etc. Even if there are many techniques available to detect plagiarism in a document, the most of them are domain-specific and designed to operate with English texts, but plagiarism is not limited to a single language only. With 300 million native speakers and 37 million second-language speakers, Bengali is the most frequently spoken language in both Bangladesh and India. In this project, we designed and developed a plagiarism detection method that makes use of Natural Language Processing (NLP) methods to deliver more precise results than traditional methods. We used some Bangla datasets which are mainly based on a lot of articles from Bangla newspapers and we also created a dataset using scarper. We processed the collected datasets and then merged it with the dataset that we created by scraping and categorizing it. Then we trained the dataset with different models and tested the result. After performance testing, we built an UI for Bangla Plagiarism Checker so that users can check the similarity based on our training dataset.