মেট্রোপলিটন ইউনিভার্সিটি
Metropolitan UNIVERSITY

A Thesis Report

Submitted To

The Department of Computer Science and Engineering, Metropolitan University, Sylhet in partial fulfillment of the requirements for the degree of Bachelor of Science in Computer Science and Engineering.

ON

Intent Detection in Bangla Dataset for the Food Domain using BERT and Machine Learning Techniques

Course Title: Final Year Project
Course Code: CSE-435 AND CSE-436

Submitted By

Badar Uddin Ahmad Ridoy
ID: 201-115-143
Batch: 50th

Dibbyasree Dey Mumu
ID: 201-115-106
Batch: 50th

Supervisor

Md. Shamihul Islam Khan Limon
Lecturer
Department of Computer Science and Engineering
Metropolitan University, Sylhet

Department of Computer Science & Engineering
Metropolitan University

Date of Submission: 26 May, 2024

মেট্রোপলিটন ইউনিভার্সিটি
Metropolitan
UNIVERSITY

Declaration

We are declaring that this thesis is our original work and we have written it completely By ourselves. We acknowledge all the sources of information which has been used in this Thesis duly. Previously the thesis has not been submitted for any degree in any University.

Ridoy 26/05/24

Badar Uddin Ahmad Ridoy (201-115-143)

Mumu 26/05/24

Dibbyasree Dey Mumu (201-115-106)

Department of Computer Science & Engineering
Metropolitan University

Recommendation Letter from Thesis Supervisor

The thesis entitled "**Intent Detection in Bangla Dataset for the Food Domain using BERT and Machine Learning Techniques**" submitted by the students:

1. Badar Uddin Ahmad Ridoy
2. Dibbyasree Dey Mumu

Is a record of research work carried out under my supervision and I, hereby, approve that the report be submitted in partial fulfillment of the requirements for the award of their Bachelor's Degrees.

26.05.24

Signature of the Supervisor

Md. Shamihul Islam Khan Limon

Lecturer,

Department of Computer Science and Engineering
Metropolitan University, Sylhet.

Certificate of Acceptance of the Thesis

The undersigned hereby certify that they have read and recommended to the controller of examination for acceptance of the project entitled- **"Intent Detection in Bangla Dataset for the Food Domain using BERT and Machine Learning Techniques** "by **Badar Uddin Ahmad Ridoy** and **Dibbyasree Dey Mumu**. It has been defended in front of the following members of the project committee in 2024. The members have accepted this project as the partial fulfillment of the requirement for the degree of Bachelor of Science in Computer Science and Engineering.

Science and Engineering.

26.05.24

Head of Department of CSE

Md. Mahfujul Hasan

Assistant Professor, Dept. of CSE,

Metropolitan University, Sylhet.

26.05.24

Md. Shamihul Islam Khan Limon

Lecturer, Dept. of CSE,

Metropolitan University, Sylhet.

Reviewers:

26.05.24

Md. Mushtaq Shahriyar Rafee

Senior Lecturer,

Dept. of CSE,

Metropolitan University, Sylhet

26.05.24

Archi Arani Basak

Senior Lecturer,

Dept. of CSE,

Metropolitan University, Sylhet

Department of Computer Science & Engineering
Metropolitan University

Acknowledgments

Firstly, we would like to thank the Almighty Allah or God for completing and submitting the final year Thesis on time. Regarding the outcome of our thesis, we would like to express our most profound sense of gratitude and respect to our supervisor **Md. Shamihul Islam Khan Limon**, **Lecturer**, Department of Computer Science and Engineering, Metropolitan University, Sylhet for his suggestion to select the topic and constant guidance, excellent supervision, advice, and other fruitful help throughout our thesis. His dedication, collaboration, and interaction were key factors in the success of our thesis. Especially Thanks our honorable Md. **Mushtaq Shahriyar Rafee Sir** and **Archi Arani Basak Madam** for their assistance, support and advice that were very helpful in this research.

We would also like to express our deepest appreciation to **Md. Mahfujul Hasan** Head of the Department Of CSE, Metropolitan University, and all the faculty members of the department of CSE. All of them helped us gain knowledge throughout these 4 years of the journey which encouraged us to create this thesis work. We are forever grateful to each one of them.

We also wish to acknowledge our parents, other teachers, and some of our close friends for their continued support throughout the thesis.

Finally, we would like to thank Almighty Allah from the deepest of our hearts for making us able to complete this thesis work.

Abstract

Natural Language Processing (NLP) is pivotal in today's digital world, with Natural Language Understanding (NLU) gaining popularity for its cognitive and AI applications. Spoken Language Understanding (SLU) is crucial for goal-oriented dialogue systems, involving word recognition and intent detection. Although SLU research has advanced rapidly, intent detection in low-resource languages like Bangla remains underexplored. This paper addresses this gap by introducing a benchmark dataset for intent detection in the Bangla language within the food domain. The dataset comprises approximately 4,539 user comments from Facebook and YouTube. We meticulously processed the raw text data through cleaning, filtering, and preprocessing to facilitate intent extraction. We utilized several models, including BERT (Bidirectional Encoder Representations from Transformers) and traditional machine learning models like Support Vector Machine (SVM) and Logistic Regression. The test accuracies achieved were 98.02% for BERT, 98.67% for SVM, and 97.68% for Logistic Regression. This thesis presents a robust solution for intent detection in Bangla, contributing significantly to the field by leveraging Transformer-based architecture and machine learning techniques. The creation of this dataset aims to promote further research in Bangla intent detection.

Keywords: **Natural Language Processing (NLP),Spoken Language understanding (SLU), Intent Detection, Slot Filling, Food Domain, Machine Learning, Bangla Dataset.**

Contents

List of Tables

List of Figures