

Project Proposal

Toxic Comment Detection



Submitted by

Abu Sufian Emon

ID: 0562210005101023

Spring-22

Submitted to

Razorshi Prozzwal Talukder

Lecturer

Dept. of Computer Science and Engineering

Introduction:

As online platforms grow, moderating harmful comments becomes increasingly difficult. Manual moderation is slow and unscalable. This project proposes a deep learning-based model that can automatically detect toxic content across multiple categories, such as threats and hate speech. The system is designed for real-time use and includes an interactive Gradio interface for testing.

Objective:

To develop an AI-powered model that detects various types of toxicity in user comments—such as toxic, obscene, threat, and identity hate—using deep learning, enhancing online content moderation.

Motivation:

Online platforms are plagued by harmful comments, making moderation essential. Automating this process with AI ensures faster and scalable protection for users.

Tools & Technologies:

- **Languages/Libraries:** Python, TensorFlow, NumPy, Pandas, Matplotlib, Gradio
- **Platform:** Google Colab

Expected Outcomes:

- A trained deep learning model capable of identifying multiple toxicity labels.
- A web-based interface (Gradio) to test comments interactively.