Experiment No: 01

Group Members:

Name	Roll No
Jignesh Gurav	16
Jayesh Ingale	18
Prathmesh Thakare	61

Title: Toxic Comment Classification using BERT

Abstract:

Social media is an online media that functions as a platform for users to participate, share, create, andexchange information through various forums and social networks. The rapid increase in social mediaactivity causes an increase in the number of comments on social media. This is prone to triggeringdebate due to the easy formation of open discussions between social media users. However, thedebateoften triggers the emergence of negative things, causing great fights on social media. Social mediausers often use comments containing toxic words to argue and corner a party or group. This studyconducted an experiment to detect comments containing toxic sentences on social media usingaPreTrained Model. This study performed a multilabel classification and evaluated the classificationresults generated by the BERT.

Methodology:

1. **Data Collectio**n: Gather comments labeled as toxic or not from various sources, creating a dataset.

2. **Preprocessing**: Tokenize and format comments for compatibility with BERT, adding special tokens as needed.

3. **Loading Pretrained BERT Model**: Download and import a pretrained BERT model into your environment.

4. **Fine-Tuning**: Develop a classification model on top of BERT, training it on your labeled data.

5. **Evaluation**: Measure model accuracy and effectiveness using validation and test sets and appropriate metrics.

6. **Deployment and Maintenance:** Put the model into use, continuously monitor its performance, and update as necessary.