CLASSIFICATION OF ONLINE TOXIC COMMENTS USING MACHINE LEARNING ALGORITHMS

INTRODUCTION

- With the rapid growth of online platforms and social media,toxic comments have become a serious concern,often leading to harmful effects on users and communitues.
- These comments may include hate speech,threats,insults,or other forms of abusive language.
- To address this issue ,machine learning techinques can appled to automatically detect and classify toxic comments.

EXISTING PROBLEM DRAWBACKS

Here are some drawbacks:

- ❖anual moderation
- ❖Keyword-based filters
- ❖Lack of context understanding
- ❖Bias in models
- ❖Limited language coverage

PROPOSED METHOD ADVANTAGES

- Automated detection
- Improved accuracy
- Context awareness
- Scalability
- adaptability

METHODOLOGY

- Here are some methodologies:
- Upload toxic comments data set
- Preprocess data set
- Apply count vectorizer
- Run svm algorithm
- Run logistic regression algorithm
- Run decision tree algorithm
- Run random forest algorithm
- Run knn algorithm
- Accuracy comparision graph
- Predict toxic comments from test data

OUTPUT EXPLANATION

The output of the toxic comment classification system is

❖ Binary classification output

- Toxic
- Non-toxic

❖ Multi-class or multi-label output

- Toxic
- Severe toxic
- Threat
- Insult
- Identity hate

❖ Probability scores

EX: TOXIC:0.85

INSULT:0.60

CONCLUSION

- The classification of online toxic comments using machine learning algorithms offers an effective solution to the growing problem of harmful content on digital platforms.
- We discussed six machine learning techinques i.e.,logistic regression,naïve bayes,decision tree,random forest,knn classification,and SVM classifier.
- Our project team got 100%result from random forest algorithm when compared to other algorithms.