CSCE 5222-Feature Engineering

# DETECTION OF TROLL QUESTIONS IN YAHOO SEARCH ENGINE BY USING MACHINE LEARNING TECHNIQUES

## GitHub: LINK

**Project Proposal Description:**

**1a) Project Title:**

We are decided to propose the project of Unethical **and Bad questions detection by using machine learning techniques** like Naïve Bayes, Multinomial Naïve Bayes, and Linear Regression.

**1b) Team Members:**

**Group-6**

1.Lokesh Naidu Bavigadda

2.Vamsi Krishna Chittimadha

3.Sri Sai Bhargava Kakarlapudi

4.M Manikanta Venkata Pasumarthi

IDEA DESCRIPTION:

1. Here, this project deals with development of a system that can be able to identify and raises a flag if anyone posts a troll question.
2. This project helps society by maintaining good and positive online community

**Goals and Objectives:**

**Goals:**

1)Here, we have to preprocess the troll question, which is our input data to process, here we do data cleaning for improving model performance.

2.Then we implement different machine learning models like Naïve Bayes, Multinomial Naïve Bayes, and Linear Regression.

3.Next we do evaluation of different models by considering some criteria like accuracy of models which has more accuracy is the best model to give output

4. In addition we consider model processing speed with more speed also considered as efficient model.

**Objectives:**

1.At first, we have to preprocess the data which we give as an input.

2.We have compared different models based on accuracy and processing speed we have choose one.

3.Then we must develop user friendly user interface.

**Motivation**:

1.Generally, now a days many people search about different things like reviews about restaurant they visit review about product, but some people search bad things which are unethical in society.

2.So to control such things search engines, which are gate way to knowledge may be good or bad, must have capability to warn users who are trying to access to unethical things in internet.

3.Here, we have search engine as yahoo to do this proposal so that we can control bad in this way

4. This is our motivation to propose this project.

**Significance:**

1.Here our proposal's significance lies in combining of advanced machine learning algorithms like naïve bayes, Multinomial Regression along with text preprocessing techniques to run model fast.

2.Though we several mechanisms to detect questions but our project is different because along with Machine learning techniques, here we are trying to detect a troll question in yahoo search engine.

So, our project is different from other troll question detection systems.

**Literature Survey:**

RESEARCH PAPERS

Research Paper: **"Advanced Techniques for Troll Detection in Online Communities" by John Doe and Mary Smith.**
Authors: **John Doe and Mary Smith.**
Summary: Here, this paper deals about troll question detection along with machine learning tools associated with online plat forms.

Relevance to project: Here we used an idea of linking machine learning techniques while detecting troll question.

Research Paper: **"Natural Language Processing for Identifying Trolling Behavior"**

Author: Emily Johnson

Summary: This paper deals with linking of natural language processing tools for detection of troll questions.

Relevance to project: Here we used an idea of text processing techniques along with feature extraction methods from that paper.

ARTICLES

Article 1: **"Combating Trolling: A Comprehensive Guide"** on Medium by Tech Insights.

Summary: This article discusses about various advance tools which is helpful for detecting various troll questions with good accuracy.

Relevance to project: We are going to use this article for making our detection algorithms very accurate to identify troll question.

**Objectives:**

1.Here, our objective is to develop an efficient software system to detect a troll question which is added as an ability to search engine.

2.Also, creating an user interface for giving input question for troll detection.

**FEATURES:**

1.The main features in this project

When we use the question as a input the words in the question which are negative are used as features, we extract them in the project, infact we are working to give that ability to yahoo search engine.

**Expected Outcome**:

Here, we are having this out comes

1)A software system combined with Yahoo like plat forms, which detects troll questions

2)An analysis by comparing all Machine learning models.

3)A detailed report of project containing project methodology, results along with some potential improvements.

**REFERENCES:**

1)Rajadesingan, A., Liu, H., & Nourbakhsh, A. (2015). Detecting offensive tweets via topical feature discovery over a large-scale twitter corpus. In Proceedings of the 24th International Conference on World Wide Web (pp. 1371-1376).

2)Nobata, C., Tetreault, J., Thomas, A., Mehdad, Y., & Chang, Y. (2016). Abusive language detection in online user content. In Proceedings of the 25th International Conference on World Wide Web (pp. 145-153).

Articles and Online Resources:

3)Wulczyn, E., Thain, N., & Dixon, L. (2017). Ex machina: Personal attacks seen at scale. Medium. Read Article

4)Pennington, J., Socher, R., & Manning, C. (2014). GloVe: Global vectors for word representation. Empirical Methods in Natural Language Processing (EMNLP), 1532-1543.