CSCE 5222-Feature Engineering

**DETECTION OF TROLL QUESTIONS IN GOOGLE SEARCH ENGINE BY USING MACHINE LEARNING TECHNIQUES**

GITHUB:[PROPOSAL LINK](https://github.com/Saibhargav2001/FE-PROJECT)

**Project Proposal Description:**

**1a) Project Title:**

We have decided to propose the project of **Unethical and Bad questions detection by using machine learning techniques** like k-means, random forest and support vector algorithms.

**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 communities

**Goals and Objectives:**

**Goals:**

1)Here, we will 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 k-means, random forest, and support vector algorithms.

3.Here by using the models we can group words of different categories

4. for doing above we use sentimental analysis technique one of natural language processing techniques.

5.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

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

**Objectives:**

1.Atfirst, we will preprocess the data which we give as an input.

2.We will compare different models based on accuracy and processing speed we will choose best of them by watching results.

3.Then we will 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 Google 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 combination of advanced machine learning algorithms like k-means, support vector machine and random forest algorithms along with text preprocessing techniques to run model fast.

2.Though we have 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 Google search engine.

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

**Literature Survey:**

1.Depression detection using emotional artificial intelligence and machine learning

Relevance to project: Here, in this project we are going take the techniques like unsupervised learning techniques for grouping different words in this article they used it for grouping different conditions of human beings.

2.A review on sentiment analysis and emotion detection from text Pansy Nandwani1 · Rupali Verma

Relevance to project: Here, in this project we are going to use sentimental analysis to detect troll questions in our project.

Here, we also used general natural language processing techniques for cleaning of data and some unsupervised learning techniques

Link: <https://rdcu.be/dpujk>

**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 an input the words in the question which are negative are used as features, we extract them in the project, in fact we are working to give that ability to Google Search engine.

**Expected Outcome**:

Here, we are having these outcomes,

1)A software system combined with Google 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.

References to Literature Survey:

1.Depression detection using emotional artificial intelligence and machine learning

Link : <https://www.sciencedirect.com/science/article/abs/pii/S2214785322005430>

2.A review on sentiment analysis and emotion detection from text Pansy Nandwani1 · Rupali Verma

Link: <https://rdcu.be/dpujk>