MES COLLEGE OF ENGINEERING, KUTTIPPURAM
DEPARTMENT OF COMPUTER APPLICATIONS
20MCA246 – MAIN PROJECT

PRO FORMAT FOR THE APPROVAL OF THE FOURTH SEMESTER MAIN PROJECT

(Note: All entries of the pro forma for approval should be filled up with appropriate and complete information. Incomplete Pro forma of approval in any respect will be rejected.)

Main Project Proposal No : _____ *(Filled by the Department)*	Academic Year : 2020-2022 Year of Admission : 2020

1. Title of the Project : CYBERBULLYING DETECTION & PREVENTION: DATA MINING IN SOCIAL MEDIA
2. Name of the Guide : Mr. MOHAMMAD JABIR C
3. Number of the Student: MES20MCA-2015
4. Student Details

Name (in BLOCK LETTERS)	Roll Number	Signature
1. DEEPIKA BALAKRISHNAN C	15	

Date:16/04/2022

Approval Status : Approved / Not Approved

Signature of Committee Members

Comments of The Mini Project Guide Dated Signature

Initial Submission :

First Review :

Second Review :

Comments of The Project Coordinator Dated Signature

Initial Submission:

First Review

Second Review

Final Comments: Dated Signature of HOD

CYBERBULLYING DETECTION & PREVENTION: DATA MINING IN SOCIAL MEDIA

DEEPIKA BALAKRISHNAN C

INTRODUCTION

This article is devoted to analysis of the articles on cyberbullying of children and adolescents and creating the methodology of working on the own work, which related especially to Kazakh social media. The article provides examples of similar works by foreign researchers, the United States, England and other European countries. Using the methods of theoretical, namely analysis, synthesis, and empirical: comparison and experiment, the analysis of works on this topic was carried out, and this work was aimed at analyzing the problem in the Kazakh space. In brief, we consider the issue of creating a parser, as well as collecting data for training machine learning and deep learning algorithms that could find and block texts containing a humiliating slope in real time. The article will be of interest to both novice specialists who are engaged in data analysis, and experienced ones to expand their horizons.

OBJECTIVES

Cyberbullying nowadays getting more popular, especially during the pandemic period. Issues of mental and actual brutality that were already just in the social climate have moved to the virtual one. From the start, it appears to be that the type of such provocation is innocuous. In any case, the contrasts among cyberbullying and conventional genuine harassing are because of the highlights of the Internet: secrecy, the presence of a wide crowd, the capacity to make assaults 24 hours per day The victims of cyberbullying are hesitant to impart their issues to grownups, in light of the fact that they believe that they might be denied of Internet access

TOOLS / PLATFORM, HARDWARE AND SOFTWARE REQUIREMENT

Hardware Requirements

- Input Device : Mouse, Keyboard
- Output Device : Monitor
- Memory : 4 Gb Ram (Minimum)
- Processor : Intel core i3 or above

Software Requirements

- Operating System : Windows 8 /10for Better Performance
- Front End : Python (Flask)
- Back End : MySQL
- Software Used : PyCharm

PROBLEM DEFINITION

Using the previous experience of the works done, we changed the course of our research. It was decided to take the direction of automating the process of finding texts containing a humiliating slope. Further, after that, we began to study whether there are similar works on the territory of Kazakhstan or at least on the territory of the former USSR countries. As it turned out, there are several similar works in concept, but even they did not meet our requirements. Since we decided to create an autonomous information system that could automatically detect and take actions to prevent them

BASIC FUNCTIONALITIES

- Exploring the media space where textual information is most used
- Use the methods of data collection in these portals and make an information system that makes it possible to analyze text information in real time.
- After the data has been collected, do manual marking of the data
- Experiment with conventional machine learning algorithms (Linear Regression, SVM, KNN...)
- Create a collection of words from the data set to create the core of the Kazakh language
- Collect a large amount of training data
- Make a training of deep learning algorithms for natural language processing