

Safe Discord Bot Using ML

Kunal Ingawle

Dept.Of Computer Engineering
BRACIT VIT, Pune, India
kunal.ingawle21@vit.edu

Shraddha Bagav

Dept.Of Computer Engineering
BRACIT VIT, Pune, India
shraddha.bagav21@vit.edu

Suraj Gaikwad

Dept.Of Computer Engineering
BRACIT VIT, Pune, India
suraj.gaikwad21@vit.edu

Om More

Dept.Of Computer Engineering
BRACIT VIT, Pune, India
om.more21@vit.edu

Pragati Hiwale

Dept.Of Computer Engineering
BRACIT VIT, Pune, India
pragati.hiwale21@vit.edu

Guided By : Prof . Dr. Ashutosh Mrathe
Vishwakarma Institute of Technology, Pune 411037

Abstract—Discord is a popular real-time chat platform with comprehensive bot support. Bots are common on Discord and offer a variety of services such as moderating aid, games, music, internet searches, money processing, and more. We attempted to create a simple variation in the Discord community by deploying a bot in this study. The purpose of this study is to implement a safe bot in Discord that identifies chat emotions and automatically eliminates chats that are not acceptable for community platforms. Using Machine Learning and sentiment analysis the bot is able to classify the sentences and classify them into different categories of emotions and automatically eliminates the sentences which are categorized as unusual. This helps to make your chat safe and helpful for the community. In the first instance, sentiment analysis finds emotional tones in real-time behind your chat messages and deletes them automatically if any negative correlation is detected in them.

Keywords— Raspberry pi, Discord bot, ML.

I. INTRODUCTION

Every day, Chat Data generates a large amount of data containing various sentiments. Discord is the most used platform by Various communities to chat and share their views. As every social media platform contains spamming, negative messages and hated messages Discord has it too. So, to determine it and to help the community to be safer with minimum spamming and unwanted negative chats. The proposed analysis system uses data to learn and predict the sentiment of the chat and deletes it automatically if the sentence could be negative, threatening, or fearful. Behind bars, the machine learning model is trained with different sentences based on sentiments, so the model learns and predicts on the basis of that chats. The server is deployed on the Raspberry pie which is a tiny computer with a System on Chip.

II. LITERATURE SURVEY

This section includes prior research and information about Discord. The bot's purpose is to make the Discord server a bit more fun by making it interactive and interesting[5].Users communicate via voice calls, video calls, text messaging,

media, and files in private chats or as members of "servers". But this generates a huge amount of text and messages. . Sentiment analysis or classification of this user generated data is very useful in knowing the opinion of the crowd[2]. Authors of the study "Text classification using Machine Learning and Deep Learning Models"[1] suggest that if the data is more organized, the better is the analysis, and eventually the decisions would be better and also authors deep dives into different text classification algorithms, discussing their advantages and disadvantages. Neethu M S and Rajasree R [2] suggest Machine learning to be a useful approach that makes use of a training set to develop a sentiment classifier that classifies sentiments. In the proposed system for training purposes Dataset with negative, fearful, rude used for training so it could help in testing on the sentences from discord. A number of machine learning techniques like Naive Bayes (NB), Maximum Entropy (ME), and Support Vector Machines (SVM) are used to classify reviews [3]. It can likewise get the inter-dependency conditions or the more profound associations between the measures of the observed information. Also same researched done by authors of Sentiment Analysis in Twitter using Machine Learning Techniques[2] they used a Machine Learning strategy to analyze Twitter posts, which entailed obtaining a Twitter dataset and cleaning, preprocessing, and creating the ML model with the help of the Twitter API. Finally, tweets are classified as joyful, sad, afraid, or neutral using this method, among other factors.

So, similar to what we're attempting in the discord servers with the assistance of the discord bot.

III. PROPOSED SYSTEM

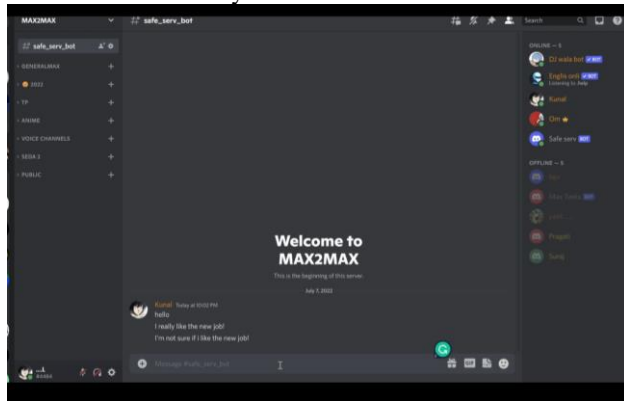
Our Proposed system uses Machine Learning model to delete the negative, spammed and threatful statements form Discord servers. Any one can easily clone our code and run the scripts in their server and get their bot setup on their communities. Proposed system also aims to provide configs which can be

used to deploy easily on any Linux server like raspberry pi so with that privacy of users is always protected because servers will be controlled on moderator.

After turning on the Raspberry Pi, the Safe Serv bot appears. As a result, we conducted some testing

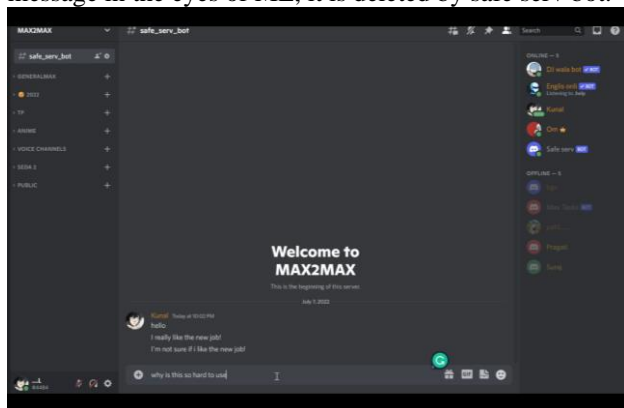
[i] Normal Messages

These are some examples of normal and positive messages that are not deleted by our bot.

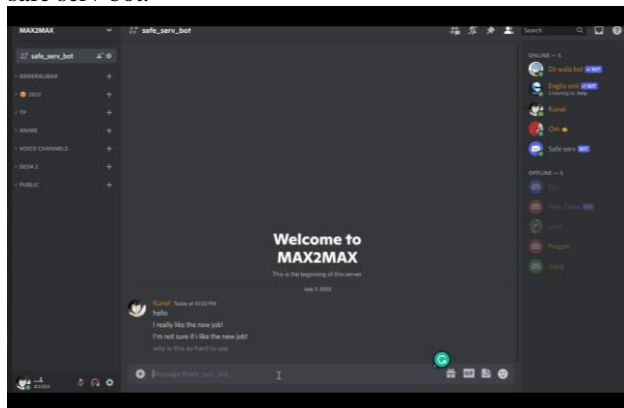


[ii] Negative Message

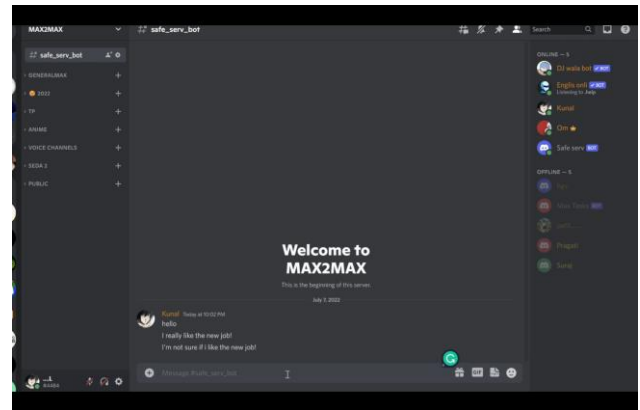
"Why is this so hard to use ?" Because this is a negative message in the eyes of ML, it is deleted by safe serv bot.



The negative messages will be deleted immediately by the safe serv bot.



[iii] After Negative Message Disappearance



IV. TOOLS AND TECHNOLOGY

Discord:

Discord provides tools and services through which a community can be formed and managed. It is perfect for communities and small groups since it is rich in helper bots, moderation options and features with paid boosts. Therefore we are creating a bot to further increase the engagement and trust of communities in their respective motives.

Machine Learning:

In order to get the context and react with proper negatives and positives about message Machine Learning is used.

With Machine Learning we have trained models with Sentiment Analysis and tested it on samples of negatives and positives and tweaked it to get right results as possible

Using Machine Learning there is no need for human work and thus human errors are avoided and also decreasing overheads from admins and moderators.

Sentiment Analysis:

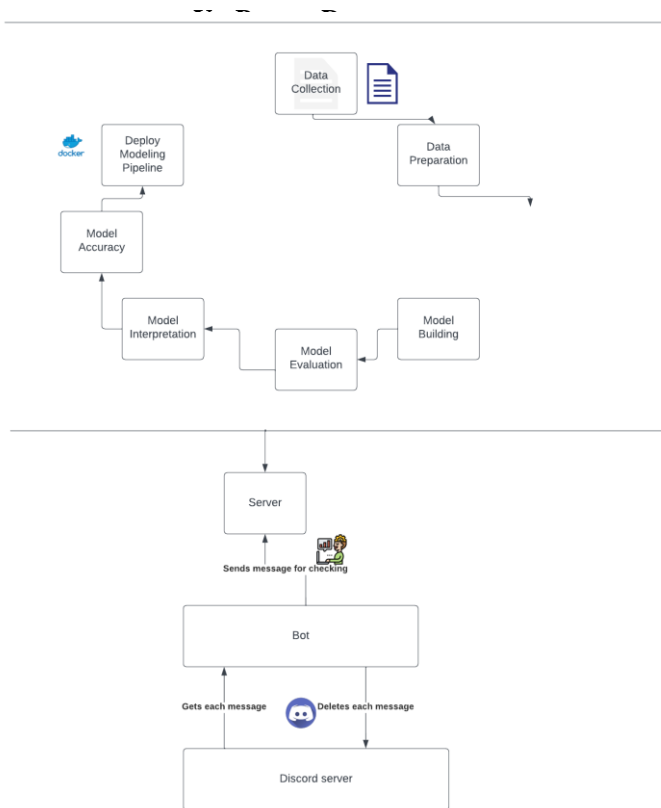
We needed to get sentiment from the text and messages sent by users to properly scan them as positive and negative. Therefore we have implemented the NLP technique of sentiment analysis. This technique is popular in business and analysis of documents and using it provides great results for understanding the user's tone in a message and whether it should be flagged as hate or spam.

Python:

We have used Python since due to its rich set of AI libraries and ability to plug in the database with few simple lines. The support for platforms is also huge, like for example ARM with raspberry PI 4 has first class support for Linux OS like Ubuntu, due to which the server can also be coded using Python and hosted anywhere in the platform/servers.

Raspberry Pi:

Raspberry Pi is the perfect device to host servers and databases, since by hosting these critical services privacy is protected, since the admins will have full contract on the server hence the data. The Raspberry Pi is powerful, affordable and small in size which makes it very efficient for this operation.



[3]A machine learning approach to sentiment analysis in multilingual Web texts,Erik Boiy & Marie-Francine Moens,26 September 2008.

[4]Machine Learning Techniques for Sentiment Analysis,Munir Ahmad, Shabib Aftab, Syed Shah Muhammad and Sarfraz Ahmad.

[5]A Comprehensive Review on Bot - Discord Bot Anirudh Verma 1 , Shashikant Tyagi1 , Gauri Mathur2

VI. CONCLUSION & FUTURE SCOPE

The Bot created efficiently helps Discord communities and users to filter spam and harmful content.

These services include a Machine Learning model made using python which can take messages and show results which can be future used for classifying if the message is positive or negative. Then it includes a python server which imports models and then verifies for messages on the discord server and immediately checks if it is positive and negative and then deletes it if it is negation.

The configs which can be used to deploy our whole light weight stack on the raspberry server and then run it as service which will then start the bot automatically when that raspberry pi starts (or restarts).

The Future scope is to improve our model accuracy and introduce mechanisms where it will store the new data and process the data for improved results.

We will also enable users to report false positives so users can simply react with thumbs down emoji and we will collect that data as negative labels and further process that so that false positives will be less in future.

VII. REFERENCES

[1]Text classification using machine learning and deep learning models. Johnson Kolluri, Shaik Razia1, and Soumya Ranjan Nayak.

[2]Sentiment Analysis in Twitter using Machine Learning,Neethu M S Rajasree R.