**ABUSE DETECTION SYSTEM**

Abstract:

Abuse detection is the task of identifying abusive behaviors, such as hate speech, offensive language, sexism and racism, in utterances from social media platforms.

With the advent of pervasive usage of online platforms, online abusive behaviour has become an indispensable part of our life demanding great attention from the research community. Accordingly, the research community is spending its effort on the demanding task, however, perhaps having much less effort on emails, even though emails are identified as a prominent source of exchanging online abusive behaviours. To fill in this gap in the literature, we conduct an in-depth study to investigate online abusive behaviour having a special focus on emails. Our system will help in finding out whether a certain text or sentence in english is abusive or not using 5 parameters.

Problem Statement:

Finding out whether a certain string or text is abusive in nature or not using Machine Learning techniques and storing data in a MongoDB database.

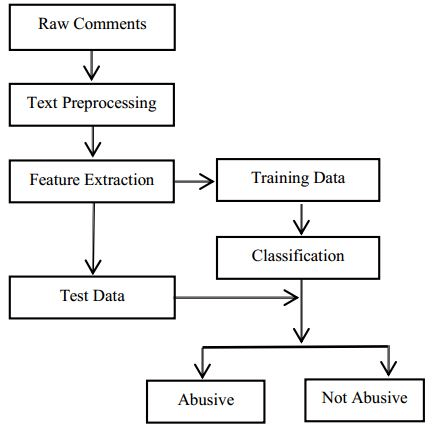
Software Required:

* Python libraries: NumPy, TensorFlow, Keras, Matplotlib, OpenCV, Scikit-learn
* Python compiler: (Jupyter notebook)
* Dataset (collected from the Kaggle)
* Machine Learning Algorithm: TfidVectorizer,
* EDA (Exploratory Data Analysis)

Objective:

* To use ML and Python for training data and finding out whether a given string is abusive or not in nature.
* Will help for making online platform more clean
* For toxic free online community.

Flowchart:



Algorithm:

Data is entered in the input field then it is sent to the ML model for training and processed then it is sent to the database and ML model classifies it as abusive or not based on 5 parameter.

Python Libraries:

* NumPy:

NumPy stands for ‘Numerical Python’ or ‘Numeric Python’. It is an open source module of Python which provides fast mathematical computation on arrays and matrices. Since, arrays and matrices are an essential part of the Machine Learning ecosystem, NumPy along with Machine Learning modules like Scikit-learn, Pandas, Matplotlib, TensorFlow, etc. complete the Python Machine Learning Ecosystem.

* Pandas:

Similar to NumPy, Pandas is one of the most widely used python libraries in data science. It provides high-performance, easy to use structures and data analysis tools. Unlike NumPy library which provides objects for multi-dimensional arrays, Pandas provides in-memory 2d table object called Dataframe. It is like a spreadsheet with column names and row labels.

* Sklearn in Python:

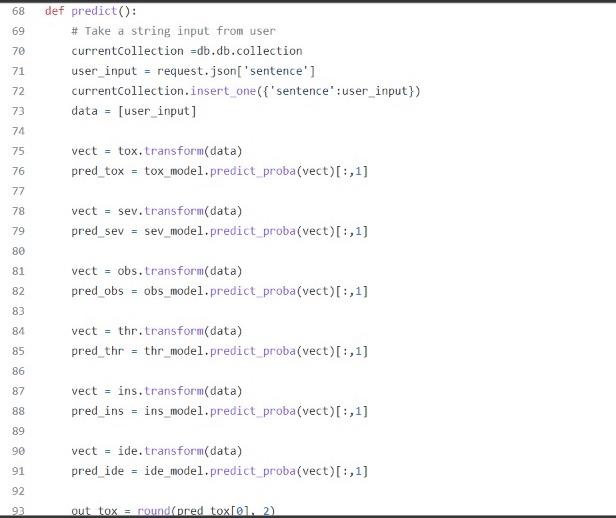
A Python module called Scikit-learn offers a variety of supervised and unsupervised learning techniques. It is based on several technologies you may already be acquainted with, including NumPy, pandas, and Matplotlib. The package provides the functions for data mining and machine learning algorithms for data analysis that are easy to use and effective. Support vector machines, gradient boosting, random forests, k-means, and other regression, classification, and clustering algorithms are included.

MongoDB:

MongoDB is a source-available cross-platform document-oriented database program. Classified as a NoSQL database program, MongoDB uses JSON-like documents with optional schemas. MongoDB is developed by MongoDB Inc. and licensed under the Server Side Public License which is deemed non-free by several distributions

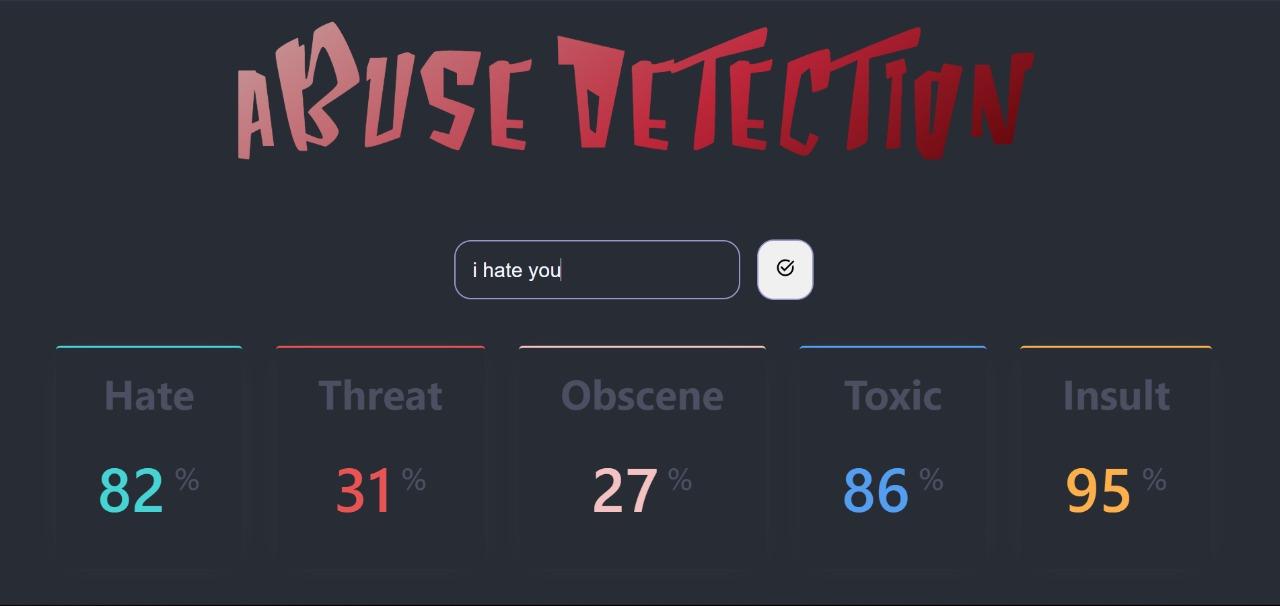
CODE:







OUTPUT:



CONCLUSION:

This project gives us basic understanding of how we can use machine learning abuse detection online. The model uses a number of technologies and gives broader idea