

Project initiation Document

TOXIC COMMENTS CLASSIFICATION APP USING IBM

OVERVIEW:

Posting comments in online discussions has become an important way to exercise one's right to freedom of expression in the web. This essential right is however under attack: malicious users hinder otherwise respectful discussions with their toxic comments. A toxic comment is defined as a rude, disrespectful, or unreasonable comment that is likely to make other users leave a discussion. A subtask of sentiment analysis is toxic comment classification. we introduce a fine-grained classification scheme for toxic comments and motivate the task of detecting toxic comments in online discussions.

PURPOSE:

The goal of the multi-label **classification** task was to determine whether or not a **comment** is **toxic** or non-**toxic** and, if **toxic**, to determine what kind of **toxicity** this **comment** is (severe **Toxic**, obscene, threat, insult, and/or identityHate).

PROBLEM:

Recently, there is a significant number of research papers on the toxic comment classification problem, but, to date, there has not been a systematic literature review of this research theme, making it difficult to assess the maturity, trends and research gaps. In this work, our main aim was to overcome this by systematically listing, comparing and classifying the existing research on toxic comment classification to find promising research directions. The results of this systematic literature review are beneficial for researchers and natural language processing practitioners.

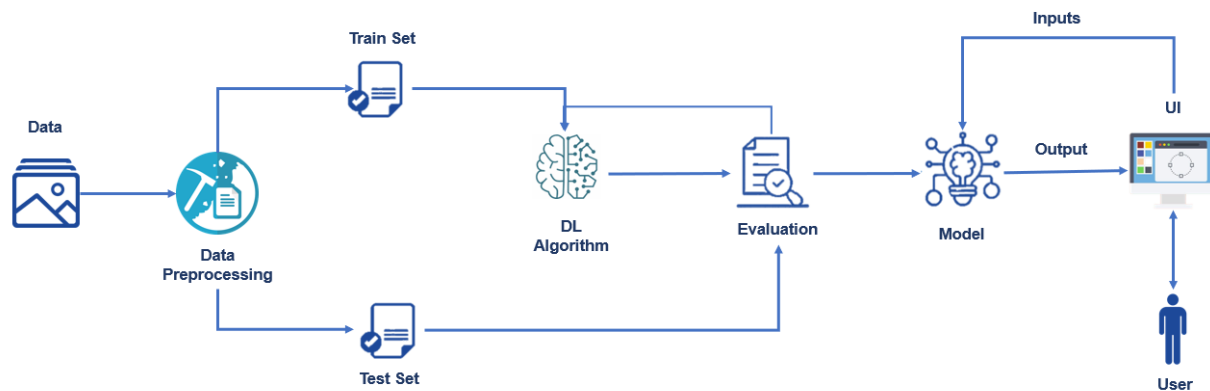
SOLUTION:

One of the major solution of such activity is automated detecting the toxic comments. Bag of words statistics and bag of symbols statistics are the typical source information for the toxic comments detection. Usually, the following statistics-based features are used: length of the comment, number of capital letters, number of exclamation marks, number of question marks, number of spelling errors, number of tokens with non-alphabet symbols, number of abusive, aggressive, and threatening words in the comment, etc. A neural network model is used to classify the comments.

Project initiation Document

TOXIC COMMENTS CLASSIFICATION APP USING IBM

Diagrammatic overview of Project:



In order to develop this project we need to install following softwares/packages:

Anaconda Navigator :

Anaconda Navigator is a free and open-source distribution of the Python and R programming languages for data science and machine learning related applications. It can be installed on Windows, Linux, and macOS. Conda is an open-source, cross-platform, package management system. Anaconda comes with great tools like JupyterLab, Jupyter Notebook, QtConsole, Spyder, Glueviz, Orange, Rstudio, Visual Studio Code. For this project, we will be using Jupyter notebook and Spyder

To build Machine learning models you must require the following packages

Sklearn: Scikit-learn is a library in Python that provides many unsupervised and supervised learning algorithms.

NumPy: NumPy is a Python package that stands for 'Numerical Python'. It is the core library for scientific computing, which contains a powerful n-dimensional array object

Pandas: pandas is a fast, powerful, flexible, and easy to use open-source data analysis and manipulation tool, built on top of the Python programming language.

Matplotlib: It provides an object-oriented API for embedding plots into application using general-purpose GUI toolkits

Project initiation Document

TOXIC COMMENTS CLASSIFICATION APP USING IBM

Flask: Web framework used for building Web applications.

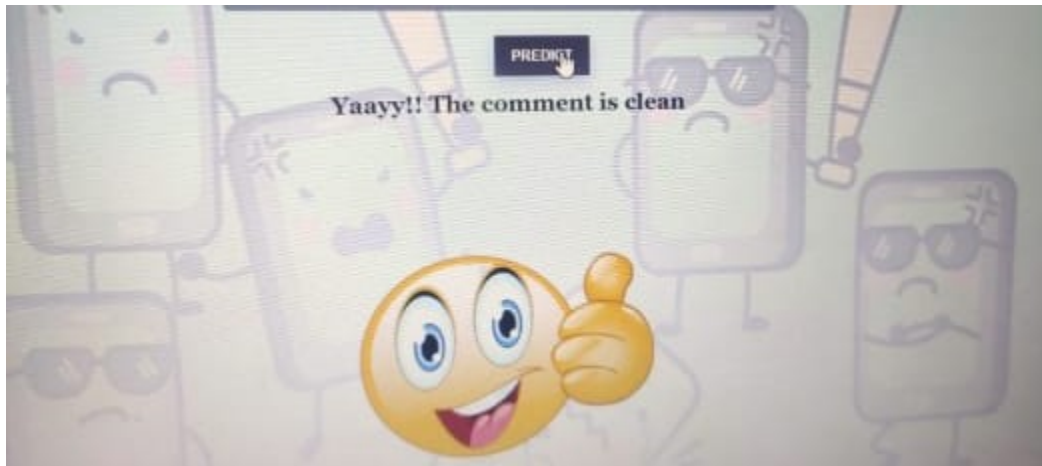
FLOW OF PROJECT

- Download the dataset.
- Preprocess the textual data.
- Classify the dataset into train and test sets.
- Add the neural network layers.
- Load the trained data and fit the model.
- Test the model.
- Save the model and its dependencies.
- Build a Web application using flask that integrates with the model built.

RESULT

comment 1 : " you are doing a great job!

output:

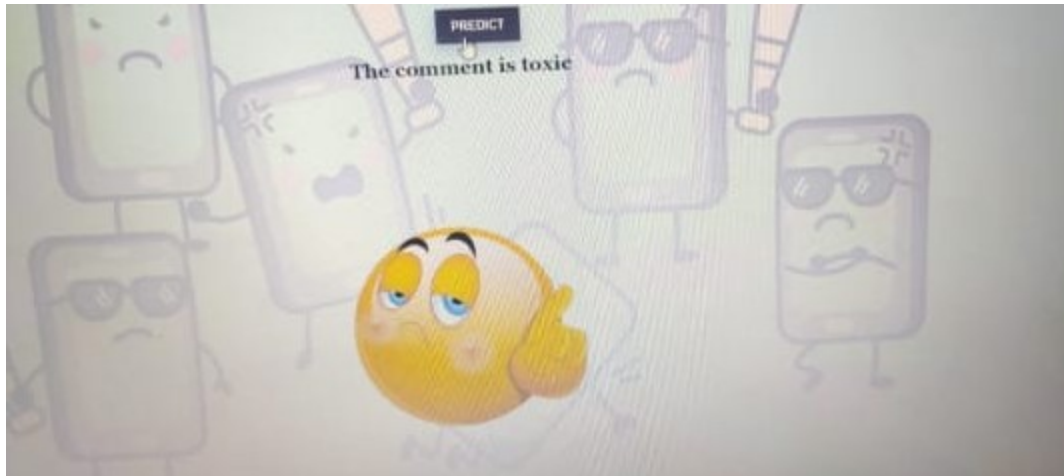


Project initiation Document

TOXIC COMMENTS CLASSIFICATION APP USING IBM

comment 2: " you are an idiot!"

output:



APPLICATION:

Hence this model can be used in various ways depending on the user.

This app is suitable for various social media platforms as it allows the user to decide the level of toxicity. This app can further be used by simple modifications by training more number of datasets, adding hate speech recognition feature.

CONCLUSION :

This concludes my work. Thank you!

You can have a look at the project proposal, report and simulations I created for this project on the [Github repository](#). I hope you learnt something new through this read!

Project initiation Document
TOXIC COMMENTS CLASSIFICATION APP USING IBM

REFERENCES:

DATASETS: <https://www.kaggle.com/fizzbuzz/cleaned-toxic-comments>

PROJECT FILES: <https://github.com/smartinternz02/Sl-GuidedProject-2292-1622047321>

PROJECT GUIDANCE: <https://smartinternz.com>