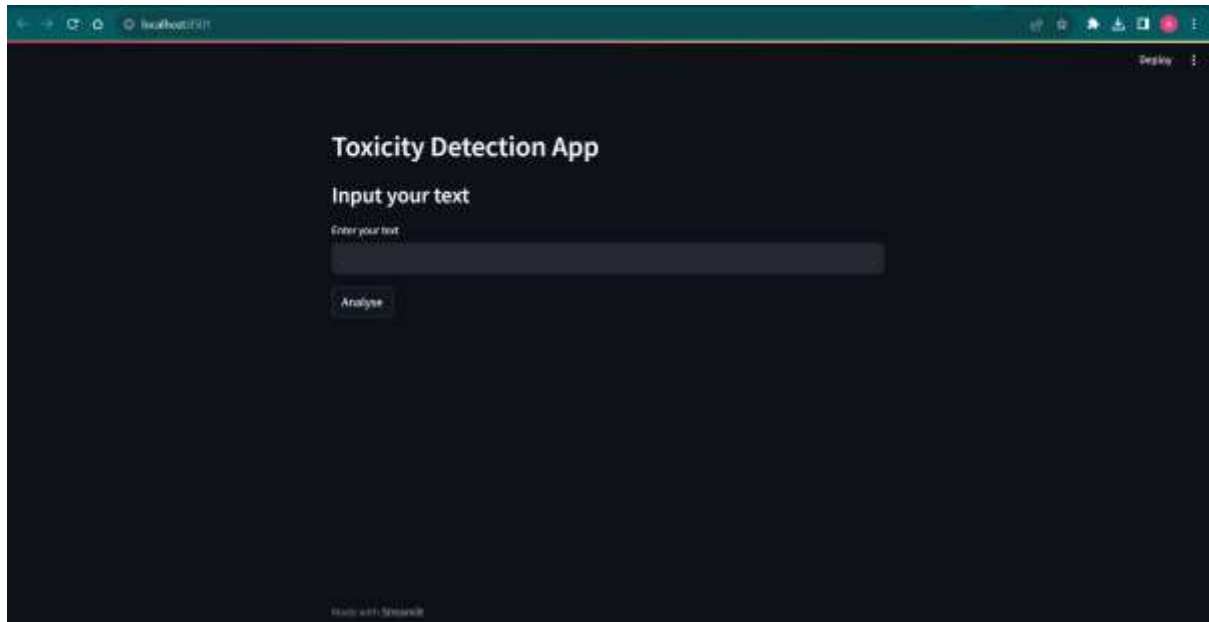


## Toxic classifier (MVP) – Amritha Prakash

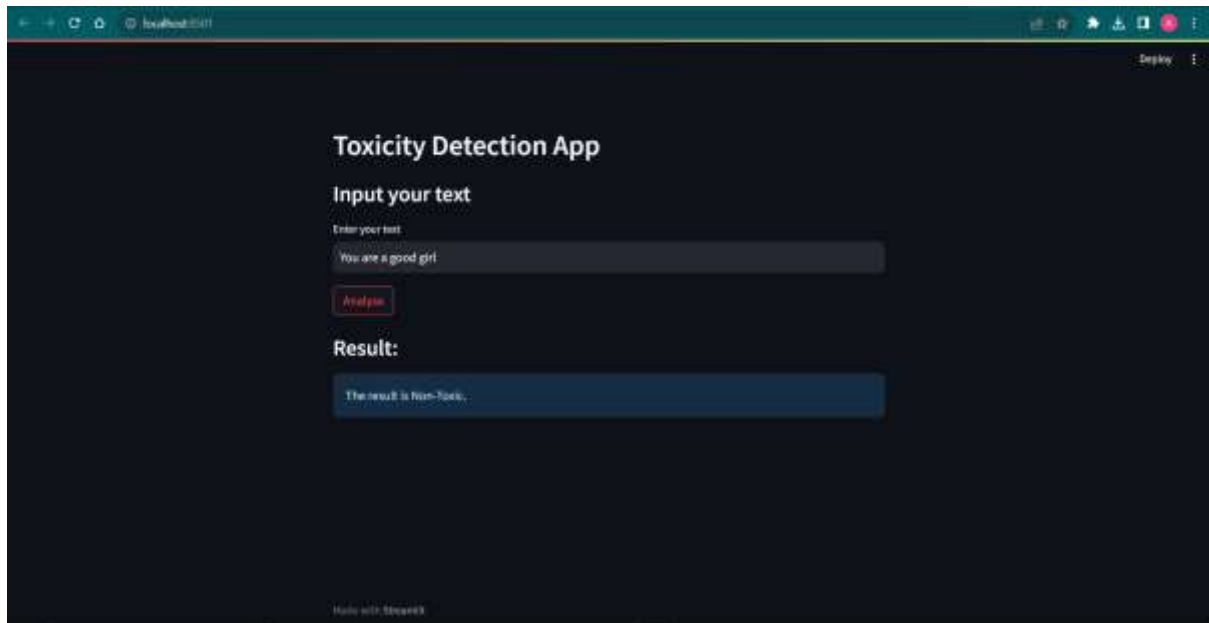
### Demonstration :

Toxic Classifier -

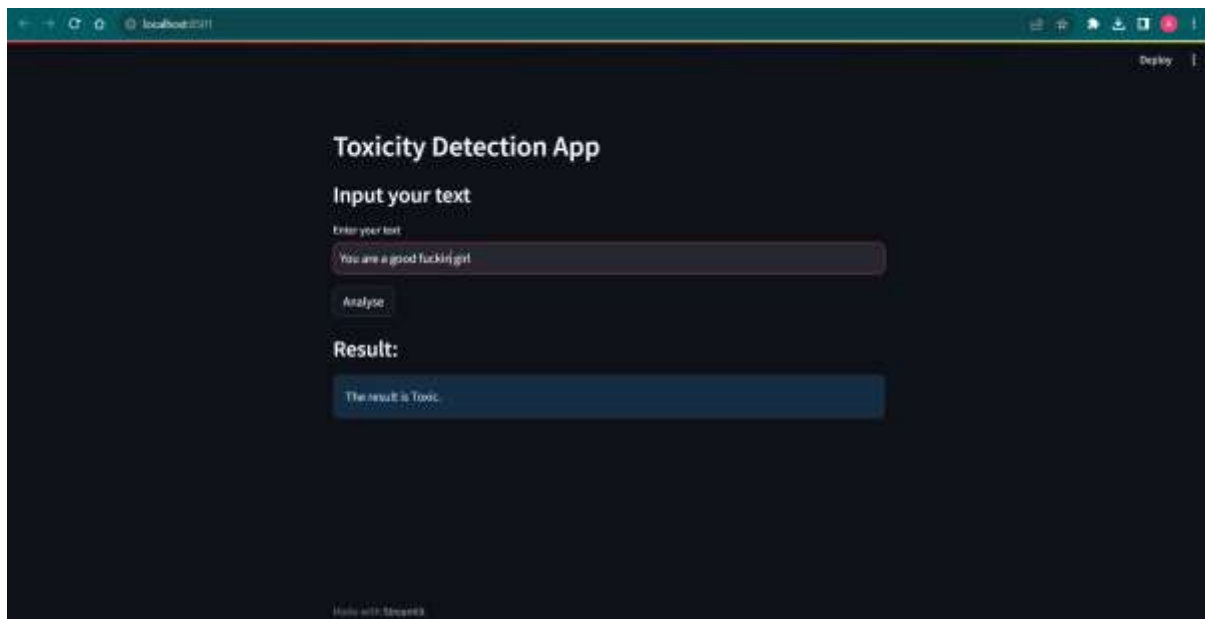


Testing :

Non- toxic comment -



Toxic comment –



### **Work done :**

Since my project is a combination of both NLP and Predictive model.

Following are the steps which I have followed -

1. Read the train Data and complete Data Cleaning
2. Exploratory Data Analysis over the Data
3. Pre-processed the data, by removing unnecessary format of texts
4. Visualized the preprocessed data to find the most used words which can categorize to tell if the sentence is toxic
5. Split the data into Train and Test, further make use of Vectorizer (TFIDF) to split and create words based on ngram. Also removed stopwords at this stage.
6. Ran the model through various Classifiers to identify the Accuracy and F1 score. As accuracy helps to understand whether the predicted values match the actual values of the target field within the incertitude due to statistical fluctuations and noise in the input data values. While F1 score computes the average of precision and recall, where the relative contribution of both of these metrics is equal to F1 score. The best value of F1 score is 1 and the worst is 0.

Made use of following classifiers -

- Logistic Regression
- Knn classifier (giving errors)
- Bernoulli Naïve Bayes
- Multinomial Naïve Bayes
- Support Vector Machine Classifier (SVM)
- Random Forest Classifier (giving errors)

7. As of now can see SVM having good Accuracy and compared to other models better f1 score.

8. So have created a sample code for the streamlit app with SVM, to predict toxicity

**Conflicts:**

- Certain Classifiers are not working, need to check on that error, to predict best classifier.

**Next Steps:**

- To redo the classifier check to understand best of all models.
- Also, try a few more vectorizers.
- Test a bit more, to see good results. As right now I see certain certain words misspelt. Maybe check to improve that for the model.
- Make use of pre-trained models to see improvement in results.
- Look for more ways to improve the application in all forms (accuracy and UI)

**GITHUB URL:** [https://github.com/AmrithaP/Toxic\\_Classification](https://github.com/AmrithaP/Toxic_Classification)