A. Streamlit Web Application User Manual

1. System Requirement

To run the Toxic Word Detection Streamlit web application, the following softwares are required to install on the Google Colab platform.

Table 1. Software Requirement of running code

Operation Platform	Google Colaboratory (TPU)
Essential Library	Transformers (version 3.5.1)
Essential Library	Streamlit (latest version)
Essential Library	Pytorch_pretrained_bert
Essential Library	Ngrok-stable-linux-amd64
Other Python Packages	Pandas, Numpy, Spacy, etc (listed in source code).

2. Running Instruction

The source code of Streamlit web application has been loaded on Google Colab. User can follow the steps below to run it.

a) Load source code file (Webapp.ipynb) on Google Colab through url: https://colab.research.google.com/drive/1d2vUyAMVnLSk7lGdYYmNa2eRG5vpKbg7?usp=sharing

b) Install latest version of Streamlit.

```
[7]   1 !pip  install  streamlit  -q
```

c) Install latest version of Ngrok.

```
[11]   1 !wget  https://bin.equinox.io/c/4VmDzA7iaHb/ngrok-stable-linux-amd64.zip
       2 !unzip  -qq  ngrok-stable-linux-amd64.zip
```

d) Install pytorch version of pretrained bert model.

```
[ ]   1 !  pip  install  pytorch_pretrained_bert
```

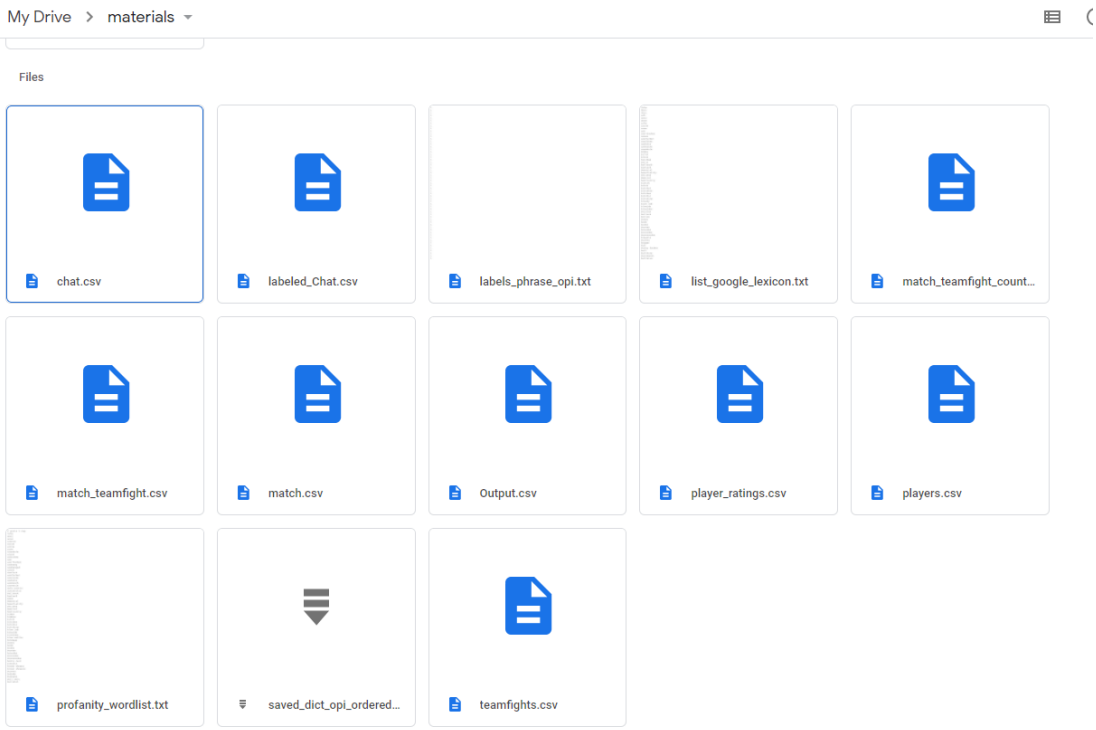
e) Install Huggingface transformers version 3.5.1

```
[ ]  1 ! pip install transformers==3.5.1
```

f) Load Google Drive which contains all configuration files and data files.

```
1 from google.colab import drive
2 drive.mount('/content/gdrive')
```

All files used and the folder hierarchy should be set as shown in following figure. The required files has been placed in the same folder.



g) Execute command that writing the main python file into local environment.

```
1 %%writefile web.py
2
```

```
Writing web.py
```

h) Execute command to get an accessible url through Ngrok.

```
get_ipython().system_raw('./ngrok http 8501 &')
! curl -s http://localhost:4040/api/tunnels | python3 -c \
    "import sys, json; print(json.load(sys.stdin)['tunnels'][0]['public_url'])"
```

i) Execute command that running Streamlit at backend

```
!streamlit run web.py
```

```
You can now view your Streamlit app in your browser.

Network URL: http://172.28.0.2:8501
External URL: http://35.232.246.119:8501
```

j) Use the public url generated by Ngrok to access the Toxic word detection web page.

```
get_ipython().system_raw('./ngrok http 8501 &')
! curl -s http://localhost:4040/api/tunnels | python3 -c \
    "import sys, json; print(json.load(sys.stdin)['tunnels'][0]['public_url'])"
```

https://a5a7edfa017f.ngrok.io

3. Use Instruction

After opening the ngrok public url, it will load 3 features of the web application.

a) Toxic word detection function

The 'Detection' function is displayed at the main page by default. User can select any function by using the sidebar (red box 1). User type any text that they want to test in the input field (red box 2) and press keyboard button 'ENTER' to run the prediction function. The result (Toxic or Non-toxic) will be displayed in the blue box filed at the bottom of input field.

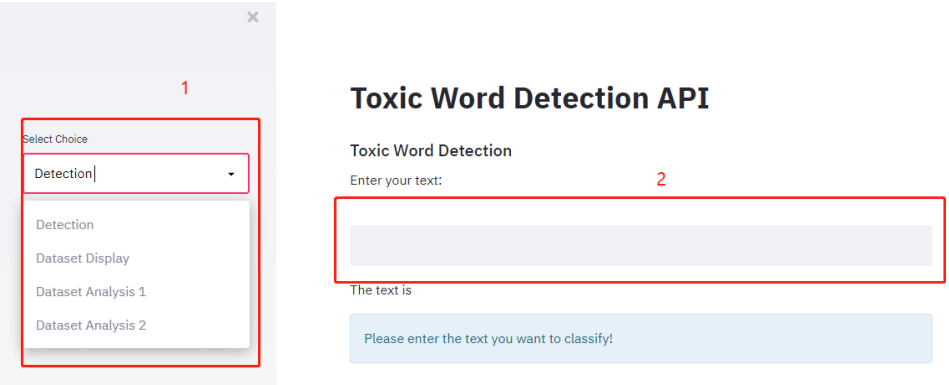


Figure 1. Detection UI

b) Dataset display function

The dataset displayed here is the Dota2 Chat dataset after labelling with 1 (toxic) or 0 (non-toxic). By using the slider (red box 1), user can filler the dataset based on their needs. User can also click the field name in red box 2 to sort the data.

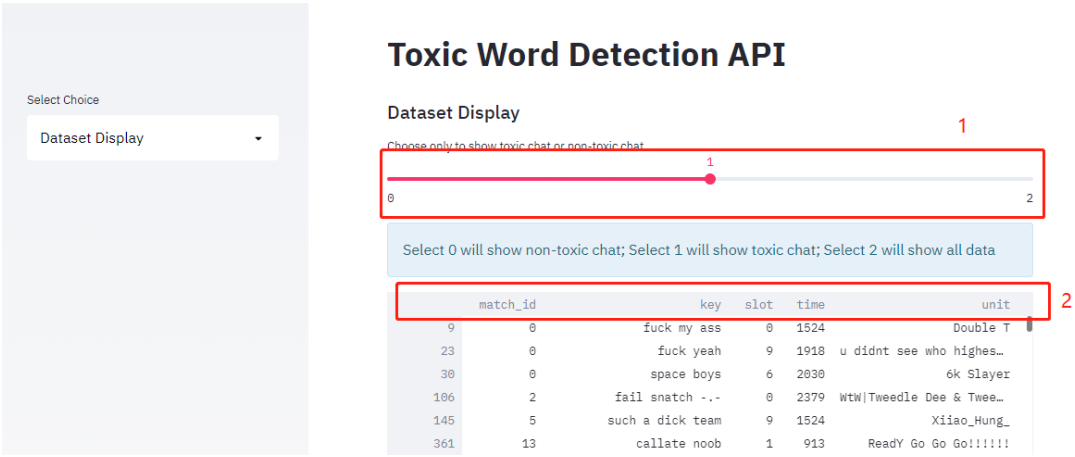


Figure 2. Dataset display UI

c) Dataset analysis visualization function

This function mainly shows the visualization of our data analysis work. User can have a look on the figure and check the analysis below. By clicking the selection box (red box 1), user can choose to hide or display the figure. By clicking the button in red box 2, the figure will be display as fullscreen. By clicking the button in red box 3, user can check the analysis based on the figure.

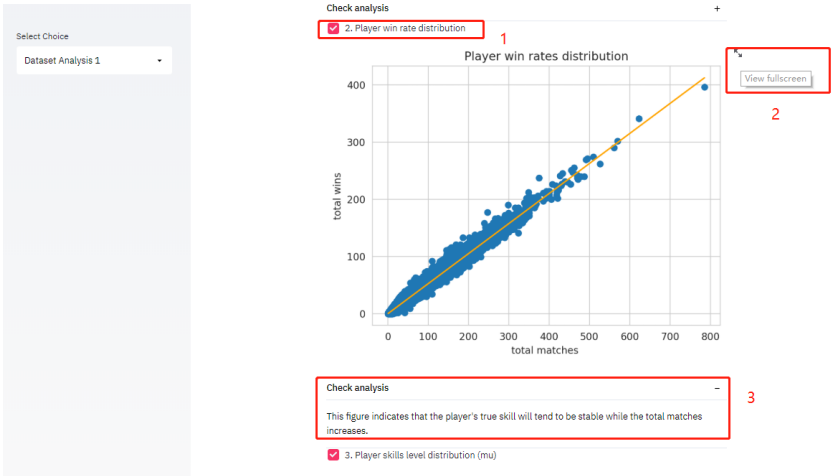


Figure 3. Dataset analysis UI