Explorer

How to use this site

Topics

Global attitudes towards GBV
World Bank
Information Hub
Female Reproductive Health
Tools and Data

Welcome to **Tools and Data** documentation

This documentation is not required to use the global attitudes towards gender based violence pages. This documentation is aimed at providing a technical documentation of all resources, models and functionality used. That being said, it will also provide high level details which are helpful in understanding the technical components, including guides on how to use this section.

For an entry level guide on all pages and using the main components of each page, please see How to use this site at the top of the left-hand pane.

Section summary

Tools and data section is geared towards putting the power of data and data science back into the hands of the community. Those who are wanting to leverage the raw data sets, or data sets which have been cleaned or make use of experimental tools along with source code will find this section valuable.

Overall, the tools and data section has 3 sub-sections:

- Documentation (this page)
- Data Portal
- Tools

These pages draw on the following resources:

- All data assets provided by (Zindi)https://zindi.africa/competitions/afd-solutions-for-gender-based-violence-challenge/data are made available
- Approved world bank shape files are also provided
- Tools code is provided in one the page. The can alo be found in the project root folder `text_classfier.py`

Data Portal

The data portal is all about datasets together for use outside of this website. Although this website provides a rich way to interact with data, such as interactive visual that provide a wealth of information, it's often helpful to have access to the raw data. In addition, the raw data is scattered about the internet and difficult to source (hence the first leg of this GBV competition which had the aim of sourcing information). Furthermore, many components of this website use transformed or cleaned data. Therefore, the data portal provides:

- Raw data sets
- Cleaned data sets
- all in one place!

There is not much technical documentation on the data portal, as this is contained in other sections when discussing the data. However, to use the data portal you will see:

- Several collapsible sections (e.g. World Bank Statistics or Acts and Laws)
- Several data sets per section - once a collapsible section is expanded it will reveal data sets
- Download buttons to download the data
- Brief descriptions of the data sets

Tools

The tools section is aimed at putting tools into the hands of users and also providing the code to build these tools. The aim is that data science tools for good become **accessible** to less technical audiences who can harness their power to address GBV. In addition, technical data scientists can leverage and improve on source code provided for models. This will allow the crowd sourcing and building of next generation tools to address GBV. Through community involvement, these improved tools can then be integrated back into our website for all to use.

There are two primary tools available:

- Tweet classifier
- GBV image classifier

Tweet classifier

The tweet classifier has been specifically developed to classify tweets into one of several categories, which include:

- Rape
- Physical force
- Humiliation
- Deny /lost job despite qualified
- Lost /denied/job due to refused sex
- Sex without consent
- Defile
- FGM
- Insult
- Child marriage
- Abuse
- Spank buttocks

The primary use of the model would be to define the type of GBV, given GBV is present. This may be helpful where a chatbot or a human assistant is available provide the correct types of assistance in different situations. However, it is still applicable in finding examples of GBV in a range of tweets regardless of whether GBV is present. This can help proactively find where GBV may be an issue.

There were several approaches considered for text classification. However, since the purpose of the website was to be accessible we chose to use the Fast AI approach to text classification. Fast AI have provided an easy to use library that abstracts a lot of the technical details out into an easy to use API. In particular, the model uses transfer learning. That is, a model is pretrained on a corpus of data (wikipedia articles). This allows the model understand english. Then the model is trained to learn the particular details of our tweet data. Now that the model has a grasp of language and the language of our tweets we then train the model to classify tweets.

Particular details of the model are:

- Tweet set 3 was used to train the model `TwitterDataSets/Set3/Tweets.csv`
- Validation set is 20% of the data
- Model achieves a near 100% accuracy on this validation set (this is a concern discussed below)
- Secondly, the model is restricted to the above categories (based on data available) and therefore does not classify texts as a non-GBC related category.

Fitting the model:

- language model trained in 3 epochs (i.e. learning the language of tweets)
- Then we fit one cycle using a learning rate of 0.002
- Then we unfreeze to layer -2 and and fit another cycle using a learning rte of 0.002
- Then we unfreeze to layer -4 and and fit another cycle using a learning rte of 0.002

Then the model is saved to the disk for later use

Accuracy concern

- Training seems to hone in on keywords. For example, only tweets with 'spank buttocks' will be classified as 'spank buttocks'
- Therefore, the model can identify obvious examples of abuse
- False positives in data set. There presence of keywords does not imply the presence of GBV. This will require a large amount of manual effort
- Overall, i think this explains really high accuracy.

Future improvements

- Ensure training data that has non-GBV related tweets. This allows the model to disregard non-GBV tweets before honing in on a specific type of abuse.
- Ensure the training data is not just keyword specific but can identify categories of GBV where keyword is not present. For example 'He took advantage of me sexually' can be considered rape even though the word rape is not prese nt.