Addressing gender-based violence together!

Harnessing data and technology to create a community

Together with Zindi and The Agence Française de Développement (AFD)

Welcome to our data driven community blog and forum! Our goal is to become the largest online community addressing gender-based violence (GBV) across Africa. We aim to achieve this goal through the use of data and technology. In doing so, we will be contributing to the UN's sustainable development goals, such as Goal 5: Achieve gender equality and empower all women and girls

To engage in the conversation, please sign-up and interact with our community by writing and replying to posts. If you prefer, you are also able to browse posts and explore data without signing-up. To get started, visit our How to use this site page which can also be found on the left-hand pane.

The home page has the following sub-sections:

- Understanding gender-based violence (GBV)
- Introduction: Our data driven community blog and forum
- Website index
- Models used
- Future vision
- Personal thoughts
- Competition goals: How we tried to meet them

§Understanding gender-based violence (GBV)

Before we can address gender-based violence, we first need to define and understand it. The University of Pretoria has provided the following description:

Gender-based violence (GBV) involves acts of violence committed against women, men, boys or girls, as a result of social norms about roles and behaviour expected of each gender. It often occurs within relationships. Although men and boys are also subjected to abuse and violence, women and girls are more often affected owing to power imbalances and the low social status that accorded to them by society. These factors often result in discrimination and in their being denied opportunities in various spheres of life.

Groups that are particularly vulnerable include:

- women and girls,
- children,
- older people,
- people living with disabilities, and
- lesbian, gay, bisexual, trans, queer/questioning, intersex, asexual (LGBTQIA+) people.

§Introduction: Our data driven community blog and forum

Now that we have a shared understanding of GBV, I would like to introduce our solution, one which has several sources of inspiration.

Firstly, I have recently read a book *Factfulness: The stress-reducing habit of only carrying opinions for which you have strong supporting facts* by Hans Rosling, an international public health expert. I believe that he would have been passionate about the addressing GBV. In addition, his approach to global problems such as GBV were to be **data driven**. To understand a problem using quantifiable measures, to see the big picture and to fight 'instincts' which detract from clear problem understanding and solving. These instincts and how to fight them can be found on his website Gap minder which has the goal of being "an independent educational non-profit fighting global misconceptions"

Secondly, globally we have seen social movements are possible on a grand scale through community involvement. The collective power of individuals with a shared goal has the ability to address GBV. In order to create a social movement, a forum used to address GBV needs to be **accessible** which I believe a website can perfectly achieve.

Thirdly and lastly, a big thank you to Zindi and the competition hosts, Agence Française de Développement, for making this possible. I now present **our** solution, a:

- Data driven
- Community blog and forum
- Made accessible to all

§Website Index

We will now provide a high level overview of the sections available on this website, in the order that they appear on the left-hand pane.

Global attitudes towards GBV

This section primarily looks at responses to a gender based violence survey undertaken in many countries around the world, in particular within Africa. The survey asked a variety of questions beginning with 'Is a husband justified hitting of beating his wife if ...'.

Although we know that hitting or beating is unacceptable, it is both chilling and eye opening to see how many people think it is acceptable. This section focuses on visualising the data to get a high level view of the state of gender-based violence before diving into trying to understand the issues further and what drives these issues. This page was powered by combining three separate data sets, namely, **World Bank data**, **World Bank shapefiles** and **gender based views survey data**. Then this data was further enhanced through machine learning to estimate survey results that were missing in order to provide a holistic view.

Sections include:

- Documentation
- Global attitudes, including 3 collapsible sections:
  - Map of gender based views
  - Data of gender based views
  - Understanding GBV drivers through the eyes of a model *(not that kind of model)*

World bank

The world bank provides a rich set of data relating to economic, education, social, contextual and health dimensions of countries. However, with hundreds of variables across 54 African countries this data can be difficult to use. This section provides 2 main tools to deal with this complexity:

- World bank interactive map and graphs to make the data more **accessible** as it's easier to use and understand. This utilises World Bank data that was cleaned to be easier to use.
- Country clusters: these are groups of countries which look similar through the lens of world bank data. They were derived using machine learning.

Sections include

- Documentation
- Country Clusters, including 3 collapsible sections:
  - World bank statistics map
  - World Bank data by country
  - World Bank data by cluster

Information Hub

One of the main goals of this website is to make information **accessible** to all. There is a wealth of information available relating to GBV which includes laws, publications and contact details to receive help for GBV related issues. However, There are a few impediments to accessing and using this data:

- Laws and policies may be in foreign languages or non-first languages
- Publications and academic articles are complex to review and when searching for information we require short excerpts to decide on whether certain publications are worth the time investment

Therefore, this section firstly brings together:

- Information sources from around Africa
- A variety of machine learning and data tools including text summarisation and text translation tools

The hope is that the the above will improve community **accessibility** to information.

Sections include:

- Documentation
- Laws
- Publications
- Support contacts

Female reproductive Health

Currently under construction

Tools and data

This section is again about making data and data science **accessible** to all. Accessibility is improved by providing both raw and clean data sets. Therefore, the community can benefit from the data cleaning of other users. In addition, since data is the new oil it's important to collect it, store it and refine it. This is the purpose of the data portal. The section also aims to make data science accessible, that's why the tools that have been built are also provided to users of the website and include source code to train the models. For example, one such tool is a GBV text classifier.

Sections include:

- Documentation
- Data portal
- Tools, including:
  - GBV Tweet classifier
  - GBV in images identification tool

§Models used

This solution leverages 4 unique models, with one currently under construction. Pre-trained models were utilised for two use cases and the other 3 were complete pipeline builds.

Estimation of attitudes towards GBV

Global attitudes towards GBV refers to how citizens of a country respond to questions that start with "Is husband is justified in hitting or beating his wife if ...", for example ending in "if she burns the food".

Only **38** of the 54 countries in Africa have at least some view into the extent of these GBV attitudes in their country. Furthermore, most countries do not have statistics across relevant all relevant demographic split, e.g. by sex, by age and by education level

Therefore, a model was created to estimate these attitudes for all African countries for all relevant demographics splits.

The outputs of these models can be used for:

- Identifying and hence investigating countries with unknown but high levels of estimated gender based violence
- Understanding the drivers and correlations of gender based violence. This allows for appropriate policies and programmes to be put in place to address GBV issues

With a better understanding of GBV and drivers we can develop data driven strategies to address issues.

Model to identify and group similar countries

When analysing issues and considering solutions it's often helpful to leverage learnings from other countries. However, "one size fits all" is often not the case when addressing societal issues. Therefore, if we are to benefit from the learnings of other countries it makes sense to look for countries which are similar in order to increase the likelihood of a solution's success.

Countries differ in many aspects, including:

- Culture
- Social beliefs
- Economic
- Education levels
- Existing legislation and policies

Therefore, in order to find similar countries, a clustering algorithm has been developed.

Text summarisation model

Academic papers can be cumbersome to access. In this competition's publications dataset, the accessibility was further reduced as there was no clear abstract provided. This makes the search for relevant publication difficult. However, since the dataset contains a substantial set of publications available in one place possible this improves the ability to access relevant publications.

Therefore, to overcome the above shortfall, machine learning models were used summarise publications thus improving the ability to search for relevant publications. Reduced time searching for information on problems and how to solve them can lead to more time addressing GBV issues.

Translation model

Standing on the shoulders of giants.

There are at least 5 languages in Africa with over 50 million speakers, therefore, for resources such as publications, laws and policies, and other media to be as effective as possible it is essential to have them available in other languages.

Leveraging the deep translator package, Google's deep learning translation model has been harnessed to translate Laws, Policies and other information on government action.

Violence detection text model

One of the largest growing datasets in the world is that of free text. It is also very difficult to work with due to the unstructured nature of the data and subtlety of connotation and denotation of texts. Therefore, a tweet classifier has been develop to identify several different types of GBV present in tweets. The model could solve several use cases:

- Identifying the type of GBV present in text, and therefore can provide direction to appropriate resources or appropriate responses by a chat bot or support line
- Proactively identifying GBV in tweets which might go unmissed. Allowing actions to be taken before situations escalate.

GBV image identifier *(Under construction)*

There are a few purposes for this module. Firstly, the tool may be used to identify and react to GBV. In addition, an issue in machine learning is data collection which reflects a real-world scenario. Since the training data comes from google scrapes I would be concerned that the original model may not be may not be trained on appropriate data. Therefore, images may be collected to train future models when users test this module. The benefit of 'real-world' images is that they are more likely provide training data where results can be replicated more closely in the future.

Prior to saving user's uploaded images, there will need to be consideration into how to protect privacy and how to ethically use the model.

§Future vision

This website is in it's infancy of potential. Potential improvements could be:

- Allowing notebooks for onsite interaction with new datasets
- Tutorials guiding those working with data sets
- Chat bots or chat functionality to assist in GBV queries
- FAQ section - this can be developed once the website has been deployed
- Campaigns and Donations page - to raise awareness and raise funds to address GBV
- Ability to follow users
- Ability to search posts of users

§Personal thoughts

I would like to start by thanking Zindi and The Agence Française de Développement (AFD) for making this challenge possible. This started off as a competition for me, however, it ended as a journey of learning. I was shocked to discover the extent to which GBV is present in Africa, in particular practices of Female Genital Mutilation. I was also comforted by insights received from the data. For example, those with higher levels of education find GBV less acceptable. Therefore, if we continue our current trajectory of education improvement we will slowly dismantle GBV. Data driven insights such as these are like the light at the end of a tunnel. We can see where we want to go and we have an idea of how to get there.

Summarising the above, we could say that the world is in a bad place, but has improved and is improving. The situation reminds me of a powerful quote from Hans Rosling's book **Factfulness**, the book mentioned in the above introduction:

Think of the world as a premature baby in an incubator. The baby's health status is extremely bad and her breathing, heart rate, and other important signs are tracked constantly so that changes for better or worse can quickly be seen. After a week, she is getting a lot better. On all the main measures, she is improving, but she still has to stay in the incubator because her health is still critical. Does it make sense to say that the infant's situation is improving? Yes. Absolutely. Does it make sense to say it is bad? Yes, absolutely. Does saying "things are improving" imply that everything is fine, and we should all relax and not worry? No, not at all. Is it helpful to have to choose between bad and improving? Definitely not. It's both. It's both bad and better. Better, and bad, at the same time. That is how we must think about the current state of the world."

Continuing this analogy, I would like to think of the data helping us track vital signs and this solution and other Zindi competitor solutions helping to improve the incubator and providing the much needed care the world (baby) still needs.

§Competition goals: How we tried to meet them

For this competition, we were challenged to create a solution that:

1. Has the potential to have an impact
2. Demonstrates a powerful use of the data sets provided
3. Is unique and creative
4. Is well documented
5. would knock your socks off!

Below documents my process in trying to achieve the above goals.

1. Has the potential to have an impact

Impact

- I believe a **large community** is one way to have a profound impact. There are many examples of social revolutions driven by technology (e.g. twitter and facebook). A website also has the potential to scale the impact it may have.
- In order to create a large community information needs to be **accessible**, both in terms of language and easy of use - which this website aims to achieve
- Hans Rosling, renowned for his work with the World Health Organisation and the UN believes change and understanding comes from being **data driven** - this website aims to nurture a data driven culture

2. Demonstrates a powerful use of the data sets provided

There are several different use cases for which tools have been developed. In particular, the data powers several models and interactive visualisations. All powered by data!

Data by itself is not powerful. It needs to be harnessed. My approach to harnessing the data to demonstrate the powerful use of datasets is as follows:

Taking many data sets > combining them > cleaning and enriching > building models > extracting insights > generating conversation > actionable insights + all in an accessible! I think this website captures and is an example of this entire process! To elaborate further:

We received many rich and complex datasets. Although relating to the same topic of GBV, these datasets have several different issues:

- Data sourced from multiple places means GBV information is fragmented across the internet does not provide a single view of the issue at hand
- Incomplete data, e.g. missing country statistics
- Difficulty connecting / joining data sets
- Lack of accessibility / understandability of data to a wider audience resulting from:
  - Language of text data is foreign to those who need it
  - Complexity of data

However, to be powerful we first need to make the complex simple. I hope to achieve this through interactive visualise that both simplify insights generation while remaining interactive.

- "The whole is greater than the sum of it's parts" - Therefore, by combing datasets we are able to get a holistic view. Combining data sets is done in several ways:
  - Joining datasets at a country level across different sources so get a full social, economic and cultural view
  - Centralising datasets into a single robust portal
- Secondly, the usage would need to add value to the process of addressing GBV. I believe this is done through:
- Enrichment of data sets
  - Imputing / estimating for countries of demographics which have not been captured
  - Translation of text data sets allows them to be more accessible and hence more powerful (e.g. laws in foreign languages)

3. Is unique and creative

Since there is no point in reinventing the wheel, we had to ask ourselves "Is my solution something that does not already exist?", "Is the solution innovative?" and "Does it demonstrate creativity"

This is why I decided to create a blog and discussion forum which focused on being data driven opposed to opinion driven. I believe this is unique as:

- I was unable to find a community blog and discussion forum **dedicated to addressing GBV**
- In addition, most forums do not have a direct link to **data driven conversations**
- The following are resources are examples of what I found to be closest to this solution:
  - https://www.saferspaces.org.za/blog
  - Medium articles -focus on education and blog posts opposed to community engagement
  - I was unable to find a reddit forum address GBV
  - South Africa's gender based violence website

A particular element of the website that I haven't seen much of on the web is that it makes data science solutions accessible to the general public. This is key to the theme of accessibility throughout the solution.

4. Is well documented

Since there is a lot of moving parts and a non-linear flow (you choose how to use the site!), i have attempted to provide documentation in two ways:

1. **How to use this site guide** can be found on the left hand explorer pane. This is high level documentation and provides details on how to use this website.
2. **Section documentation** can be found under all left hand side panel sections. This documentation is more detailed and technical. For example, details of models used.
3. **Commentary and examples** provided for each page to describe the purpose or use of different sections. For example, how to use a map and the available map views
4. **Hover information icons** ⓘ which document what specific components (e.g. Dropdowns) are used for

5. Would knock your socks off!

I will leave you to decide.

Some project statistics are:

- more than 5,000 lines of python code
- across python 67 files
- more than 1,500 lines of markdown
- across 41 files
- 242 lines of html code
- across 13 html files
- 1 sqllite database containing numerous user, their profiles and posts
- powering 5 models and numerous interactive and reactive components
- and ONE COMMUNITY