For the People

By

16ET415: Shikhar Maheshwari

16ET430: Kaiwalya Patil

Guided by

Prof. Amit H. Choksi

(Assistant Professor)

A Project Submitted to
Birla Vishvakarma Mahavidyalaya (Engineering College), An Autonomous
Institution affiliated to Gujarat Technological University in Partial Fulfillment of the
Requirements for
the Bachelor of Technology with Specialization in *Electronics and Communication*
Engineering

July, 2020

BIRLA VISHVAKARMA MAHAVIDYALAYA

(ENGINEERING COLLEGE)

AN AUTONOMOUS INSTITUTION

Vallabh Vidyanagar – 388120

GUJARAT, INDIA

COMPLIANCE CERTIFICATE

This is to certify that the research work embodied in this Project entitled **"For the People"** was carried out at Birla Vishvakarma Mahavidyalaya (Engineering College) An Autonomous Institution for partial fulfillment of Bachelor of Technology with Specialization in Electronics and Communication Engineering degree to be awarded by Gujarat Technological University. He / She has complied with the comments given by the Mid Semester Project Reviewer.

Date :

Place :

Name of the Students	ID No. of the Students	
1.Shikhar Maheshwari 2.Kaiwalya Patil	1. 16ET415 2. 16ET430	
Prof. Amit H. Choksi	Project Coordinator Dr. Darshan C. Dalwadi Prof. Anish A. Vahora	Head, *(EC Dept.)* *Dr. Bhargav C. Goradiya*

BIRLA VISHVAKARMA MAHAVIDYALAYA

(ENGINEERING COLLEGE)

AN AUTONOMOUS INSTITUTION

Vallabh Vidyanagar – 388120

GUJARAT, INDIA

DECLARATION OF ORIGINALITY

I hereby certify that I am the sole author of this Project and that neither any part of this Project nor the whole of the Project has been submitted for a degree to any other University or Institution.

I certify that, to the best of my knowledge, my Project does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations or any other material from the work of other people included in my Project, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Indian Copyright Act, I certify that I have obtained a written permission from the copyright owner(s) to include such material(s) in my Project and have included copies of such copyright clearances to our appendix.

I declare that this is a true copy of Project, including any final revisions, as approved by my Project review committee.

Date:

16ET415: Shikhar Maheshwari

16ET430: Kaiwalya Patil

BIRLA VISHVAKARMA MAHAVIDYALAYA

(ENGINEERING COLLEGE)

AN AUTONOMOUS INSTITUTION

Vallabh Vidyanagar – 388120

GUJARAT, INDIA

Abstract

People's feedback is an important aspect of democracy and this often gets overlooked in India. In this project, we aim to get feedback of willing people by making certain online platforms that are open to all. We plan to closely analyze these inputs from the people and extract important parameters, like the problem, location, which department/ministry it is falling under, frequency of submission. This data would be then submitted to the authority of city/state and hence they can be notified about the persisting issues under their locality. Hence by getting such feedbacks, analyzing them closely and solving problems, we aim to help people resolve their problems and work as a stepping stone towards perfect democracy.

The services which are provided by the government with an assurance "For the People" are built and never looked upon unless there is wreckage in their daily functioning. Since we live in a democracy, it becomes our right to enjoy every service in its best possible way. Hence, we aim to solve this problem by creating online platforms immersed with data analytics.

Acknowledgement

First, we would like to thank our Project guides, **Prof. Amit H. Choksi** for their guidance, help and support. And also, for encouraging, motivating us to design and build this Project. We would like to thank all the other faculties of our EC Department for their support and guidance. We also thank **Dr. Bhargav C. Goradiya,** Head of EC department for his support.

Finally, we would like to thank our college, "**Birla Vishvakarma Mahavidyalaya**" Engineering college for this opportunity to work on this project.

Table of contents

Appendices

Progress Report

Plagiarism Report

Poster

List of Figures

Chapter 1: Introduction

- The central intuition behind the working of this system is to let people enjoy real democracy. We observed that a lot of recent movements and bills targeted the citizen's privacy or were against the will of citizens.

- And lots of influential voices and opinions were left unheard. And this continues and keeps getting deep-rooted with every passing day.

- Other than that, the genal grievances of the people remain unheard by the intuitions that were established to solved them

- The major reason behind all the above being lack of a system that can meaning fully collect, organise and represent to the above-mentioned authorities

1.1: Our approach

- To overcome all these problems that are being faced, we decided to work on a project, where people can open up their problems, report other issues relating to a physical problem etc.

- We have made an Android based multilingual application, for getting inputs of people and a website to display all of the carefully analyzed values and outcomes.

- Behind all of this is our core algorithm which works on NLP and entity recognition. Meaning, the inputs that we get from people are automatically classified and given individual identities, example water logging is given identity of infrastructure, theft is given identity of crime etc.

- We also have a voting page, where people can vote on the government bills, so if someone is again a certain bill like GST, the person can vote. And same goes if someone is with the idea of bill being passed.

- We display all these stats on our website, which has various sections like reports page, voting report display page, how many positive or negative feedbacks on voting etc.

- On top of this we have an option to manually classify the entity, which our algorithm failed to classify, example the term corona virus was less known until 2020, so a human can manually add this for better further execution.

1.2: Machine Learning tools

1.2.1 spaCy:

spaCy is a NLP framework designed upon industry standards to make using NLP easier and more efficient .The Power of spacy is best utilized in large scale information extraction tasks as it uses Cython to maximize speed and minimize memory usage .It also provides many pretrained models to help with the Natural Language processing. spaCy is used my almost all the tech giants like Microsoft, Amazon and Google.Amazon and Google.

1.2.2 en_core_web_sm:

It is a English multi-task CNN trained on OntoNotes. Assigns context-specific token vectors, POS tags, dependency parse and named entities.

1.2.3 NLTK:

Natural language tool kit is a leading platform for building Python programs to work with human language data. It is easy to use and has a corpora of 50 and lexical resources such as WordNet, along with a suite of text processing libraries for classification, tokenization, stemming, tagging, parsing, and semantic reasoning, wrappers for industrial-strength NLP libraries.

Chapter 2: Scope and Application

2.1 Public grievance reporting system:-

The Public grievance reporting system is one of the applications we have backed into our current project. It uses the NLP engine to process human text and language into keywords so that the grievance can be classified into a department and also quantitively measure the number of a grievance for a specific problem. We have multi lingual support including Hindi, Hinglish, English, Gujarati and Marathi.
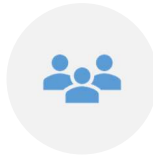
2.2 Anonymous crime reporting Fasttrack: -

This is another application of our system which we have not built for the scope of this project, but this same system can be used for anonymous reporting of crime while classifying it into type of crime for easier analysis of this data by the police .

2.3 Quantitively analysis of Individual govt. bodies: -

Currently there is no metric available to check performance of government bodies.

Our system being an open system can quantitatively track the number of grievances related to a specific organization and how much of them are resolved. Using these basic metrics we can derive a lot of conclusions regarding underfunding or over funding of an organization its efficiency and much more.

2.4 Macro Analysis of individual Grievances report to help better allocate budget of Country: -

The macro Analysis system is present to provide a quantitative measurement of the different sector of govt so that the budget allocation of the country can be on direct metrics.

2.5 Expansion on the E-Governance: -

The Current Government headed by PM Modi has taken significant initiatives to bring government bodies online and has succeeded in doing so . We thought of further improving that concept by allowing people to vote on existing bills in parliament. giving govt. direct feedback on the will of the people so they can make better decisions .

2.6 Outbreak Prediction: -

Just like on our current pandemic situation of COVID-19 the Arogya Setu app is using location data form people's phone and Bluetooth to predict locality outbreaks. we will use the grievances to detect and report possible outbreaks. This method is more effective for less contagious diseases like malaria. hence the local municipal corporation can act and curb the spread of the disease.

Chapter 3: System Software

3.1 Python

Python is used successfully in thousands of real-world business applications around the world, including many large and mission critical systems. Python is an interpreter based high level programming language Created by Guido van Rossum and first released in 1991, Python has recently gained popularity to become the most widely used language in the world. And it as one of the best community support which is a contributing factor to this meteoric rise.

Fig 3.1 Python

3.2 spaCy

spaCy is a NLP framework designed upon industry standards to make using NLP easier and more efficient .The Power of spacy is best utilized in large scale information extraction tasks as it uses Cython to maximize speed and minimize memory usage .It also provides many pretrained models to help with the Natural Language processing. spaCy is used my almost all the tech giants like Microsoft, Amazon and Google.

Fig 3.2 Spacy

3.3 Flask

Flask is a micro web framework written in Python. It is classified as a microframework because it does not require particular tools or libraries. It has no database abstraction layer, form validation, or any other components where pre-existing third-party libraries provide common functions. However, Flask supports extensions that can add application features as if they were implemented in Flask itself. Extensions exist for object-relational mappers, form validation, upload handling, various open authentication technologies and several common framework related tools. Applications that use the Flask framework include Pinterest and LinkedIn.

Fig 3.3 Flask

3.4 Selenium Webdriver

Selenium is a framework for testing applications and scraping the web using scripting languages like python, ruby, Perl and groovy. It runs on windows, Linux and mac os and is distributed under an open source software license

Fig 3.4 Selenium

3.5 Flutter

Flutter is Google's UI toolkit for building beautiful, natively compiled applications for mobile, web, and desktop from a single codebase. Flutter uses dart as its coding language which is a compile time language giving it speed and efficient memory management. Flutter unlike other multi-platform framework uses its own rendering engine written for each platform which gives it native performance while the code is written only a single time.

Fig 3.5 Flutter

3.6 Firebase

Firebase is a mobile and web application development platform developed by Firebase, Inc. in 2011, then acquired by Google in 2014. Firebase is an example of backend as a service and provides the fastest NO-SQL Database in the world.

Fig 3.6 Firebase

3.7 Dart

Dart is a client-optimized programming language for apps on multiple platforms. It is developed by Google and is used to build mobile, desktop, server, and web applications. Dart is an object-oriented, class-based, garbage-collected language with C-style syntax.


Fig 3.7 Dart logo

3.8 GitHub

Git hub uses the git versioning system and acts as the remote origin to save backup and share codebase repositories. It is particularly useful when you want to reference to a older state of you project and also remote collaboration and both of these tools proved to be extremely useful for our team to remotely collaborate.

Fig 3.8 GitHub logo

Chapter 4: Block Diagram
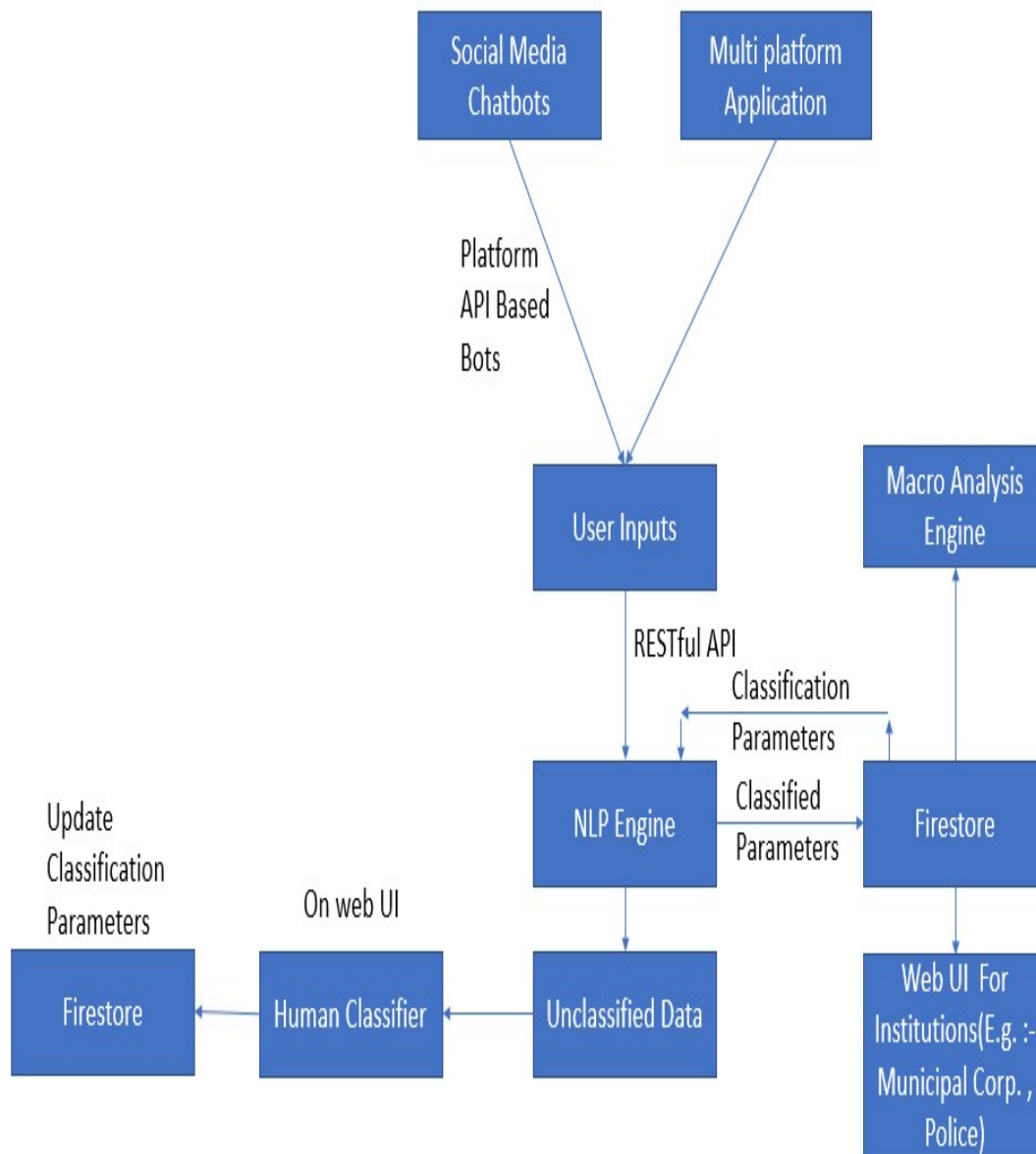
The above flowchart shows the major parts and how they interact with each other in our overall system.

1. **Social Media Chat Bots**: Theses are bots on different social media platforms which receive data from users and sed it to our system.

2. **Multi-platform Application:** - A direct application for our E-governance platform and direct grievance reporting.

3. **NLP engine**: - The NLP engine takes care of processing and classifying the data. It retrieves the latest classification parameters firestore and then uses them to build the engine and classify the given user inputs and then again store the results to the firestore db.

4. **Firestore: -** Firestore acts as the central repository for all communication between all the components of our system and as a DB of storing all data . It also serves as the user authentication platform for users of the application.

5. **Human classifier:** - The human classifier is a person which handles all the user inputs our system could no classify and manually classifies it so that similar inputs are classified from next time directly by our system. The human classifier serves as a feedback loop to continuously improve our systems classification capabilities.

6. **WEB UI:** - The web UI is present for easy access to our classified data and human classifier without the need for giving access to the database to any one.

Chapter 5: Flowcharts

5.1 Natural Language Processing engine
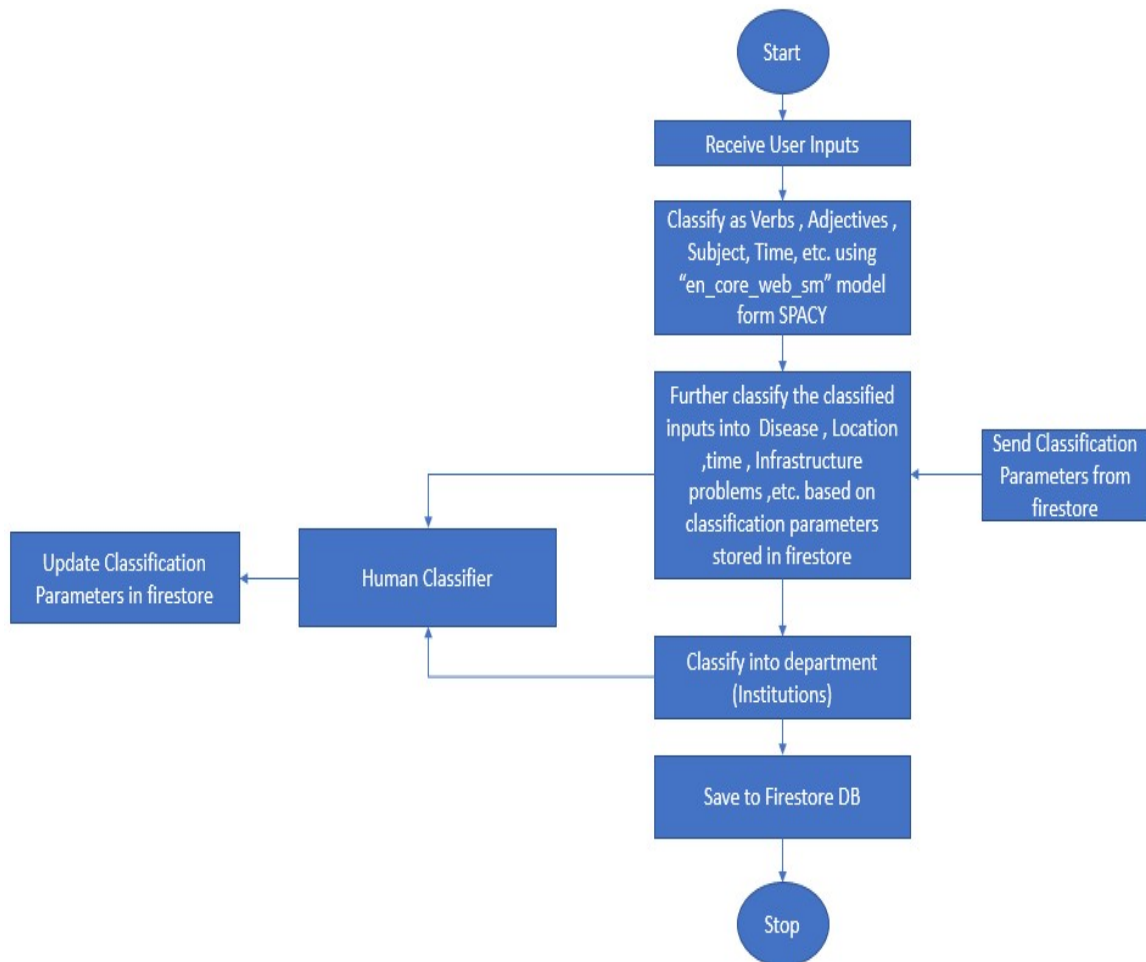
Description: -

The NLP engine is based upon the en_core_web_sm model from spaCy and built upon them are classification pipelines . The data to build these pipelines are saved as a document array pairs in the firestore database under a collection called NLP.

After building the pipelines the engine takes in the input data and processes it. After which there is a decision engine which decides if input is sufficiently classified or not.

If it is deemed sufficiently classified the output is stored to the firestore DB under classified reports collection if not then the same is stored under weekly classified which is then later taken up by the human classifier to reclassify.

When the human classifier classified it, it is sent to reports classified and the new classification parameter are stored in the NLP collection which will be used in all further classifications hence improving our System.
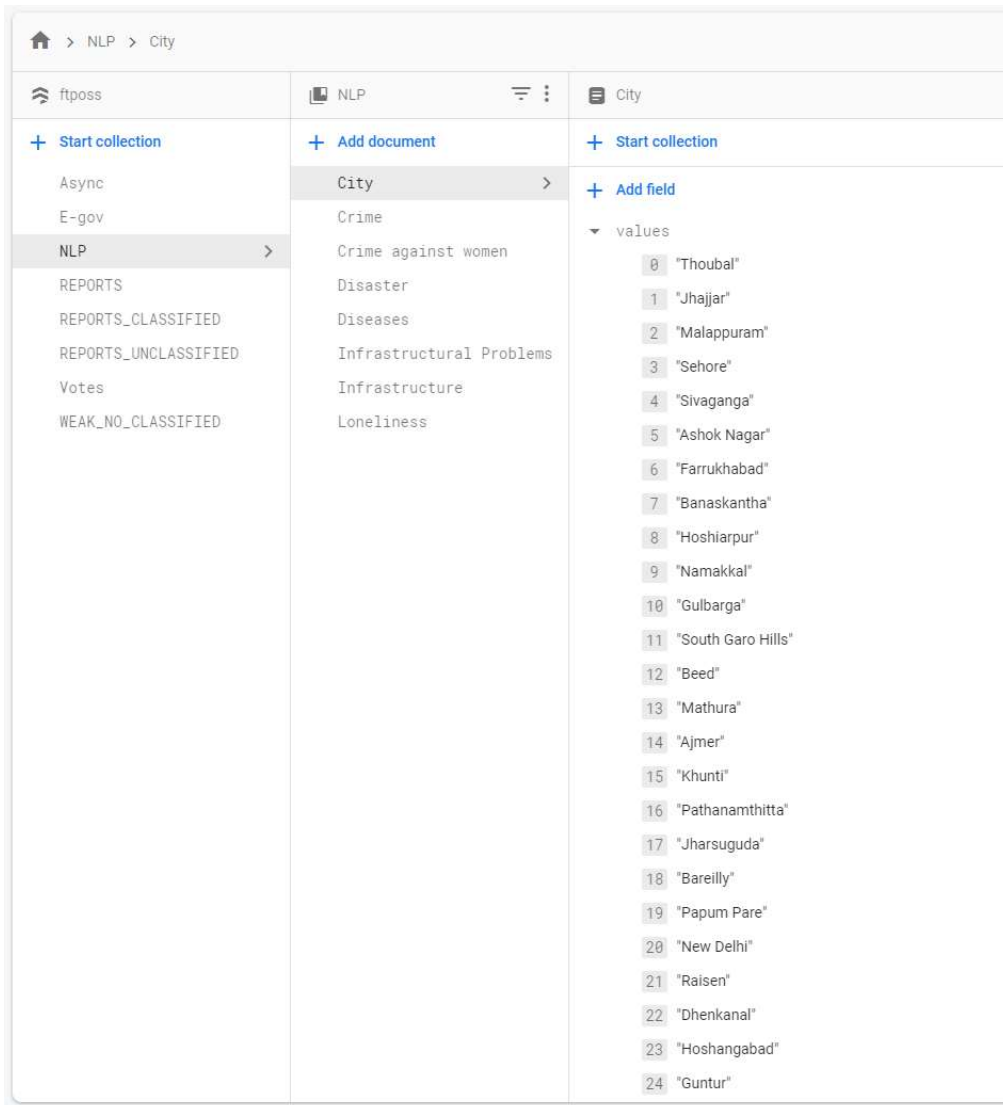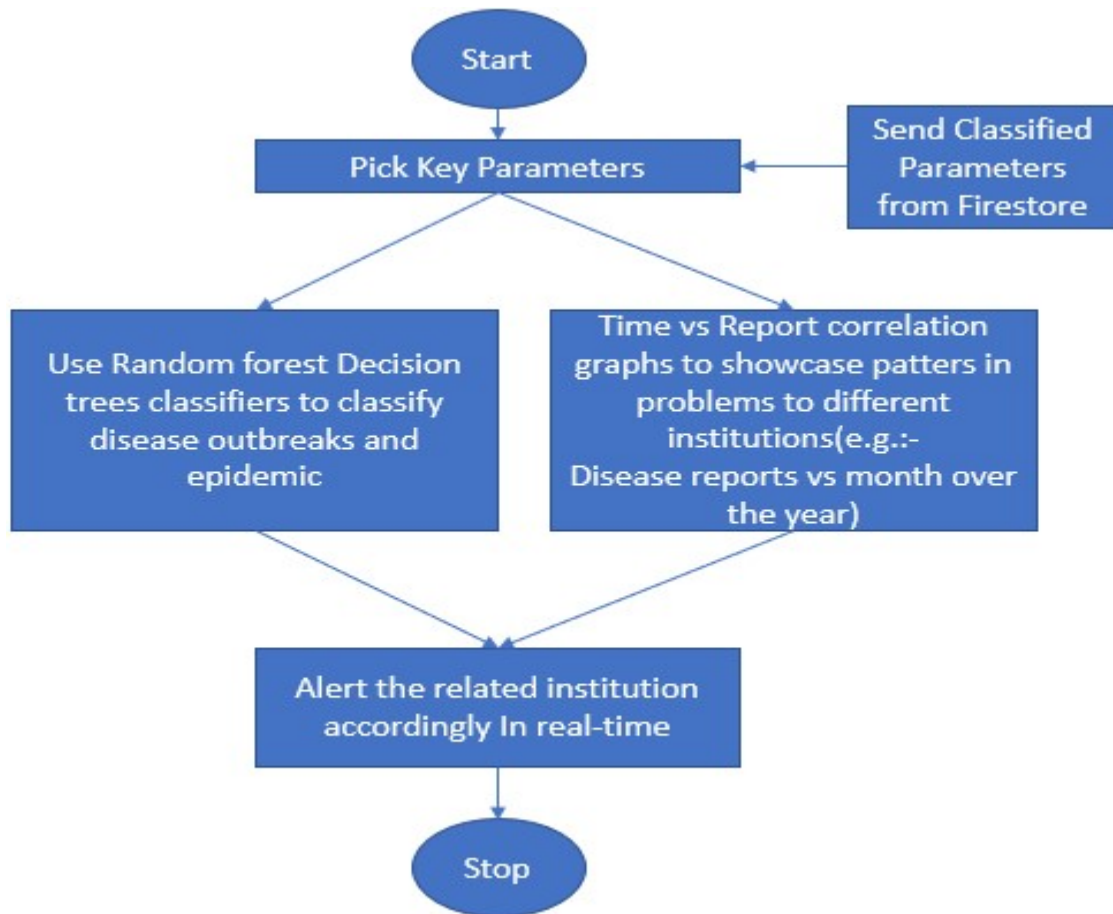

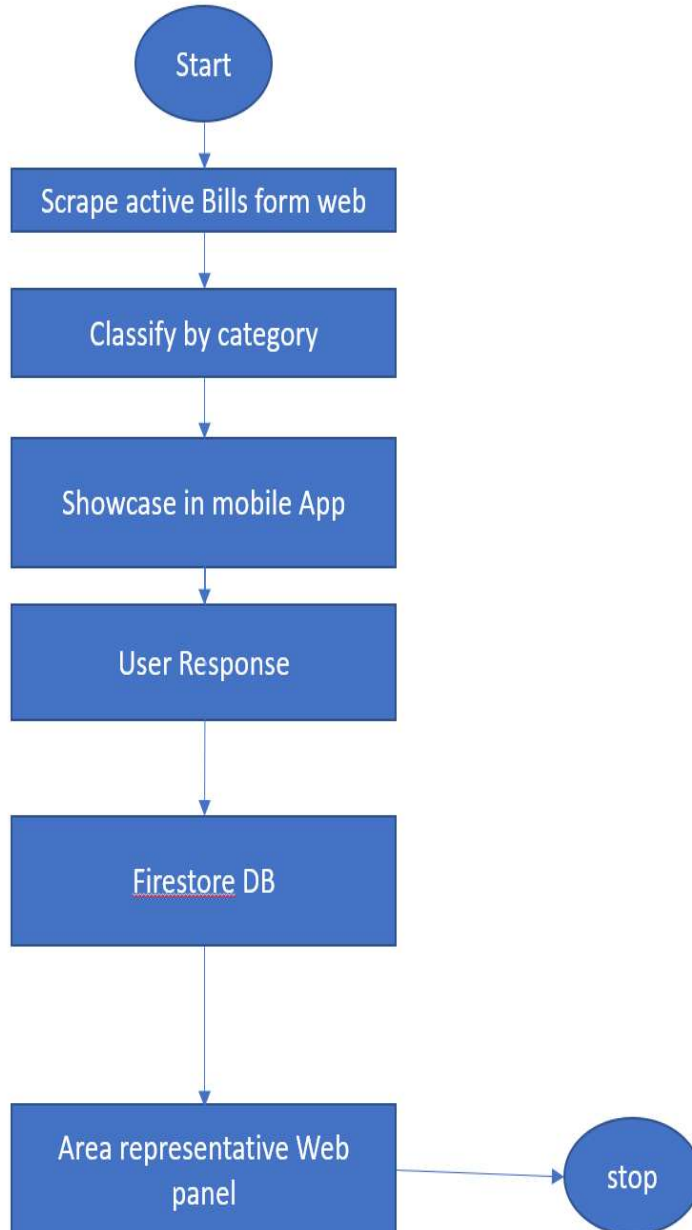Fig 5.1 Snapshot of Firestore DB

5.2 Macro analysis engine

Description: -

The primary job of the Macro analysis engine is to perform local outbreak prediction in and area and showcase data in a graphical manner for easier interpretation.

5.3 E-Governance Engine

Description: -

The basic agenda of E-governance module is to allow people to vote on existing bills in parliament. giving govt. direct feedback on the will of the people, so they can make better decisions.

We do this by first scraping the web and storing all bills into our DB (this is regularly updated). And then fetch this data from the db to our app.

Displaying the data to the user and then the user can vote and we save those responses in our DB again.
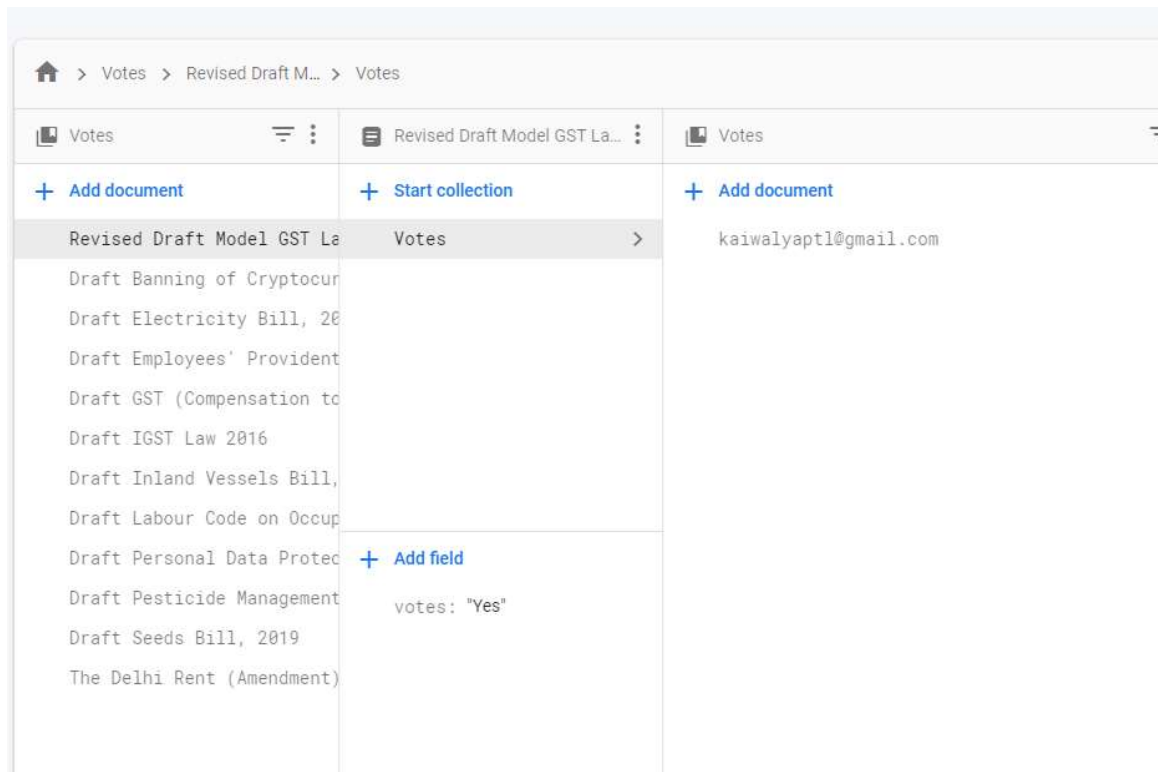

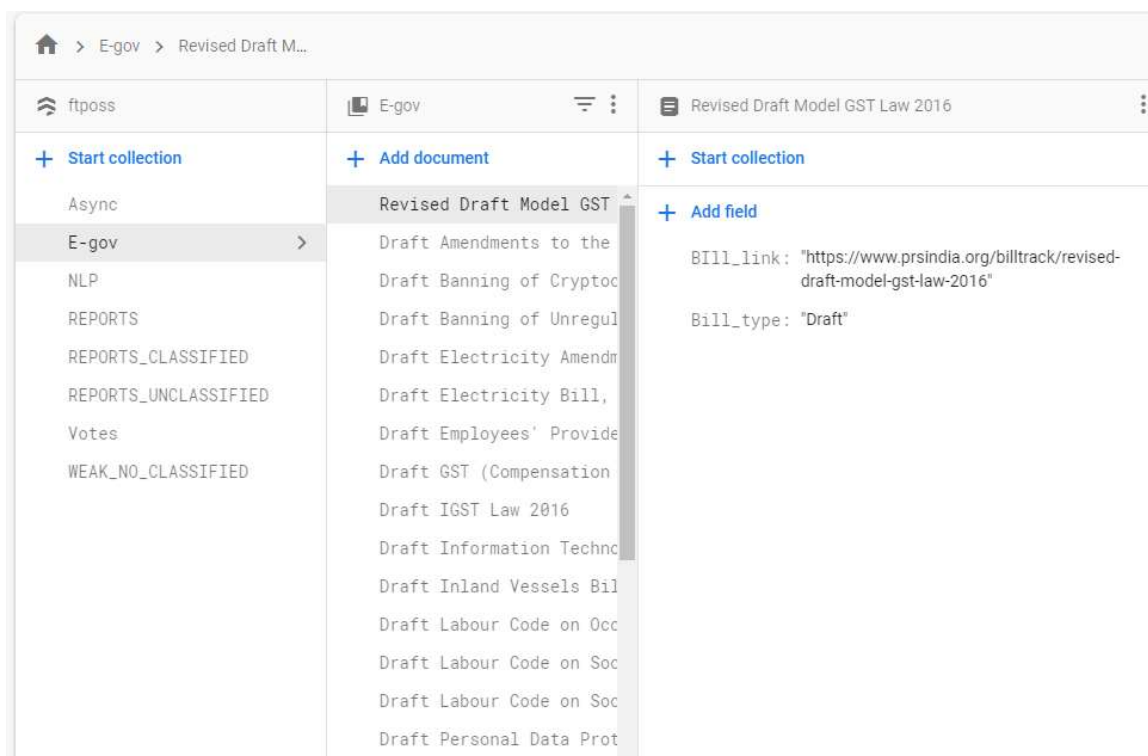Fig 5.3.1 Voting in APP

Fig 5.3.2 Storage of votes in DB

Fig 5.3.3 Storage of Bills in DB

Chapter 6: Project Outcomes

6.1 Application:

6.1.1 Sign In page:-

This page allows the user to lo into the app via a google account

Fig 6.1.1 Sign in

6.1.2 Grievance reporting and main page:-

The Grievance reporting page allows the user to report and feed their grievances directly into our system.
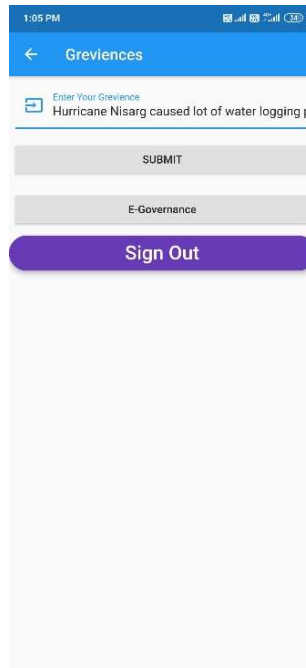

Fig 6.1.2 Grievance reporting

6.1.3 E-governance pages page:-

The E-governance pages in app where people can view and vote on bills which is then stored in Firestore DB

Fig 6.1.3.1 Bills to view and votes

Fig 6.1.3.2 Voting: support or against page

6.2 Web UI:

6.2.1 Home page:-

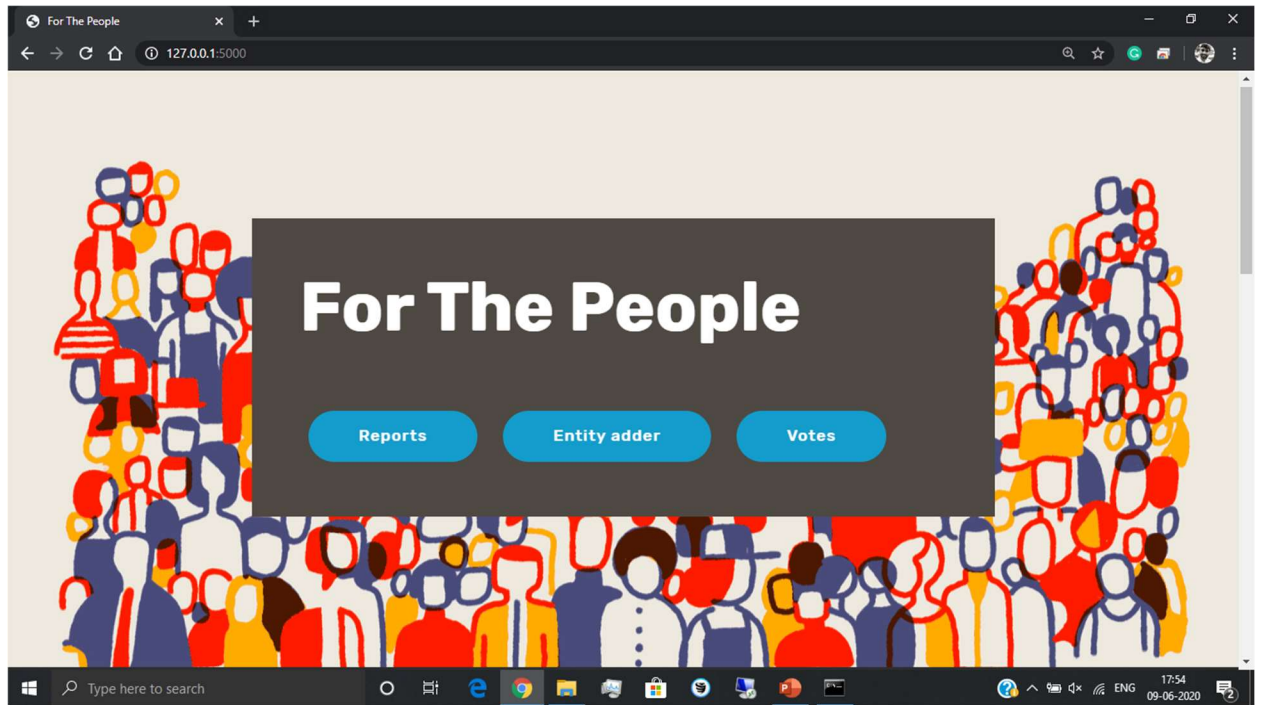The Home page of the web UI acts as a central navigation page to all the important ages

Fig 6.2.1 Website: Home Page

6.2.2 Classified reports page:-

The Website classified reports page displays the results of all the classifications on a webpage as counts so that the authorities can identify their most pressing problems.
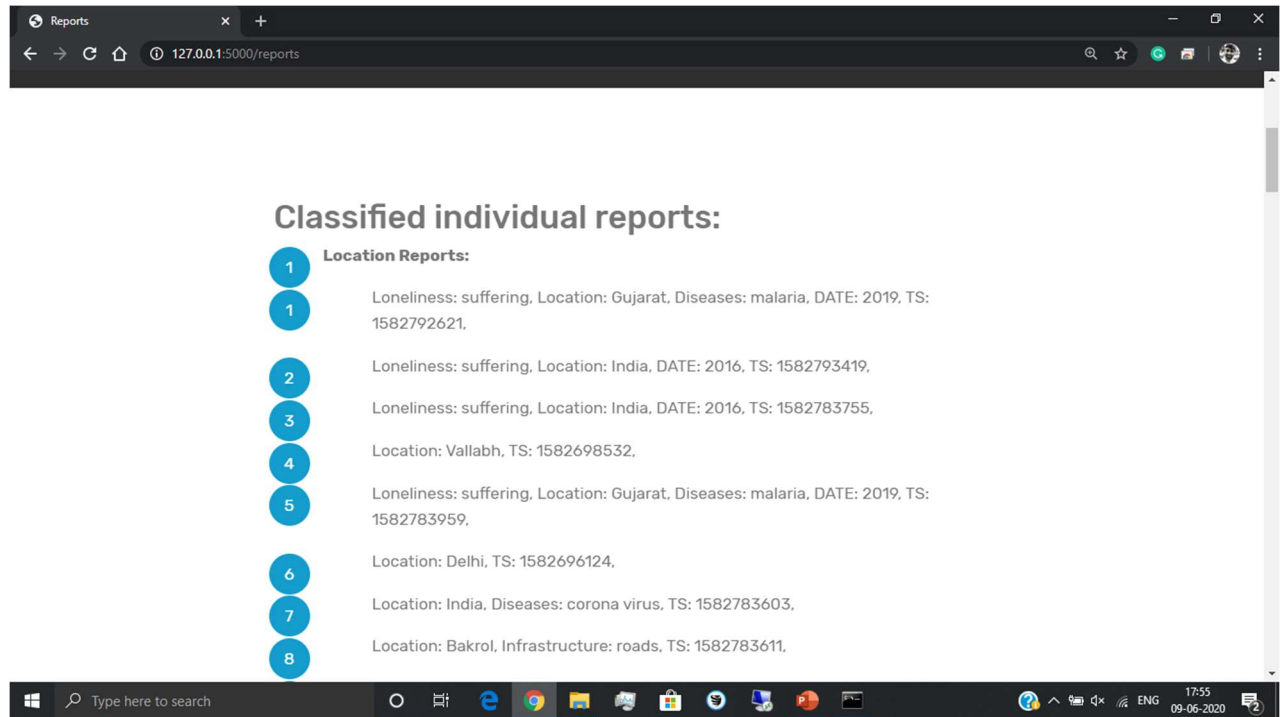

Fig 6.2.2 Website: Classified reports page

6.2.3 Votes analysis page:-

This page showcases the no. of votes counted for and against a bill from our app
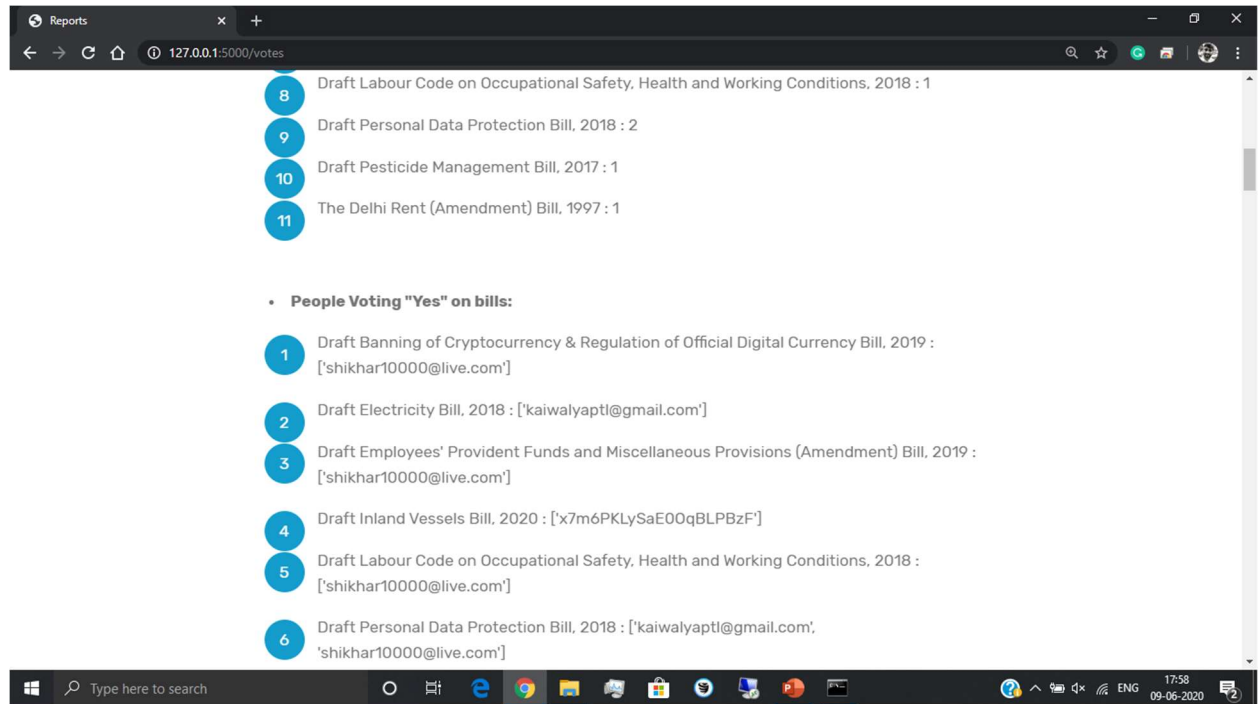

Fig 6.2.3 Website: Votes analysis page

6.2.4 Entity adder page:-

The Entity adder page is for our human classifier. This can be used to add entities to our NLP engine pipelines collection in our DB so as to included those search terms from the next run .
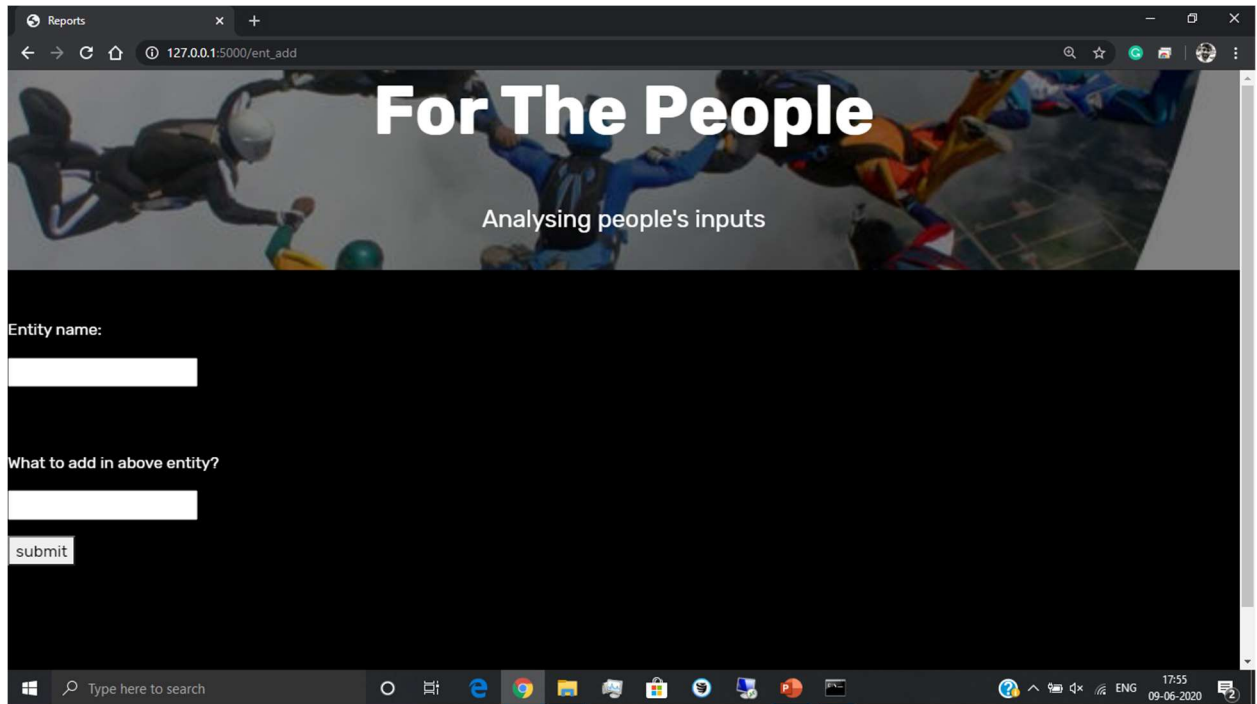

Fig 6.2.4 Website: Entity adder page

Chapter 7: Conclusion

Thus, by using Natural Language Processing (NLP), to classify the user inputs into proper entities, we have achieved our goal of achieving a better democracy. We have used NLP to classify various important words from the input sentence. On top of this we have added our own entities, so that no word remains unmatched and a perfect idea about what user wants to convey is obtained. We have made the platform to submit the data easily accessible to everyone by integrating it with an android application. Besides this, we have also added a section where users can read and vote on the current bills which are yet to pass or already been passed, this voting data helps us understand better about what people are thinking. All of the data that is classified and recognized is shown on website, so at a glance, we can get insights about people's opinions. Thus we have made an all rounder system, which helps an individual to express his/her views in a better and open way and also in an anonymous manner.

Chapter 8: Future work

- Adding various other public opinions to be voted upon.
- Relating the existing data of a certain regions with psychology to get a deeper insight of the impacts of the data on economy.
- Adding an option to take peoples input from voice. We can also analyze the sentiment of the voice. This can be used extract any such information.
- Adding an option to take people's input from images, example people have to click images about theft or water logging or any such problems. And the algorithm automatically detects the scene with the reported location and time.
- We can scrape newspapers or articles over the internet and check if the reported problems are actually legit. If they are, we can use NLP for sentiment analysis and getting the depth of problem.
- Instead of human classifier, the algorithm googles the unclassified data and allocated a label based on search and automatically updates the database.
- We Plan to make our system open source after submission of our project we plan to make it an opensource repository so that anyone can further the development on it from its current state.

Chapter 9: References

- Fix-my-Street, Brazil
- Zubair Bhatti's work using pro-active citizen feedback
- UK's 'nudge unit' used in organ donation
- NLP based entity recognition and disambiguation: https://patentimages.storage.googleapis.com/5e/35/6d/03deafc588b4a2/US8594996.pdf
- A survey of named entity recognition and classification: https://nlp.cs.nyu.edu/sekine/papers/li07.pdf
- BBMP Sahaya app, used in Banglore: https://bbmp.sahaaya.in/
- BMPP IFMS
- Named Entity Recognition (NER): https://arxiv.org/abs/2002.08902
- Named Entity Recognition (NER) for relation extraction: https://arxiv.org/abs/1912.13415
- Emotion cause pair extraction: https://arxiv.org/abs/1906.01267