AI Based : Events and Festival Planner

Md Shahul pasha, M S Mubarak basha, Kushal R, Manoj T N, Archana Das
1,2,3 Department of Information Science and Engineering, T John Institute of Technology, Bengaluru, India.
4 Assistant professor, Department of Information Science and Engineering, T John Institute of Technology, Bengaluru, India.

Student, Department of Information Science and Engineering, T John Institute of Technology, Karnataka, India
Student, Department of Information Science and Engineering, T John Institute of Technology, Karnataka, India
Student, Department of Information Science and Engineering, T John Institute of Technology, Karnataka, India
Student, Department of Information Science and Engineering, T John Institute of Technology, Karnataka, India
Faculty, Department of Information Science and Engineering, T John Institute of Technology, Karnataka, India

ABSTRACT

Information sharing is one of the huge topics in social media platform regarding the daily news related to events or disasters happens in nature or its human-made. The automatic urgent need identification and sharing posts and information delivery with a short response are essential tasks in this area. The key goal of this research is developing a solution for management of disasters and emergency response using social media platforms as a core component. This process focuses on text analysis techniques to improve the process of authorities in terms of emergency response and filter the information using the automatically gathered information to support the relief efforts. Specifically, we used state-of-art Machine Learning (ML), Deep Learning (DL), and Natural Language Processing (NLP) based on supervised and unsupervised learning using social media datasets to extract real-time content related to the emergency events to comfort the fast response in a critical situation. Similarly, the blockchain framework used in this process for trust verification of the detected events and eliminating the single authority on the system. The main reason of using the integrated system is to improve the system security and transparency to avoid sharing the wrong information related to an event in social media.

I. INTRODUCTION:

Disasters are part of the daily news in social media during the past few years. There is various type of disasters such as earthquake, flood, typhoon, pandemics of diseases and similarly, human-made disasters, e.g., incidents of terrorism and industrial accidents . The number of social media networks and their activity increasing with a high-speed day by day and daily information sharing and user-generated contents is passing hand by hand between millions of internet users. The user-generated content mainly focuses on the daily events and news, which are the current discussed topics in the real world. Internet platforms consider a powerful communication environment between people for information exchanging in a large variety of daily events. The use of social networking and information sharing in an emergency type of events and dangerous disasters is the research challenge for event detection and tracking it in the early stage. Recently, the extensive connection and increase of social media platforms give the opportunity for the management of crises based on crowdsourcing. One of the famous tools of crowd-sourcing is Ushahidi [5], which visualize the reports of crowd-sourced it's a perfect example for improving the awareness of various social networks. There are various ways to share information in recent developments, e.g., national security agencies, media outlets, civil defense, etc. The social media potentiality caught the attention through the crisis for higher management quality. The capability of the limited generalization reason is the level of

micro-blogging, which is a changeable topic in terms of abbreviations, informal language, limitation of characters, etc. The recent novel approach proposed by Kruspe et al. regarding the Twitter detection based on clustering method and event detection proposed by Fedoryszak et al. based on full Twitter firehose demonstrating the contextual information value by aggregating and sentiment the microblog messages.

What is Blockchain?

- ❖ Blockchain is a continuous sequential chain of blocks containing information built according to certain rules.
- ❖ Blockchain is a distributed database where storage devices are not connected to a common server. This database stores an ever-growing list of ordered records called blocks. Each block contains a timestamp and a link to the previous block.
- ❖ Blockchain is a promising technology and is becoming predominant for solving many problems in the field related to security under the control of public and private sectors.

II. EXISTING SYSTEM :

- Most of the discussed contents are related to a visual display of crisis-related information in social media based on the thematic, temporal, and spatial aspects for awareness of the situation.
- The main elements show the various computations between capabilities e.g content extraction regarding special criteria and using Natural Language Processing (NLP) techniques, applying Named Entity Recognition (NER) and other concepts.
- Some of the social media platforms point to making actionable reports for the relief activity and supporting disaster response. To do this, creating a report requires tagging the pre-defined categories in cloud-source.
- Similarly, there is a lack of related documents to extract the information for creating a report for a possible response.
- The centralised control on the application can violate the Right-to Speech or may support wrong messages.

Disadvantages :

- One hand control over the application
- Fake messages circulates easily
- Hard to believe the real messages

III. PROPOSED SYSTEM :

- The main focus of this system is creating a blockchain and AI based environment for the management of the crisis in social media using social media analytics.
- The key point of the developed this process activates the authority of the related disaster management for integrating and internet-based data access based on applying semantic analysis for action generating and content responses.
- The collected results can be used to monitor the related emergency and management of disasters, early warning, risk mitigation, and assessments.
- This architecture has four main components: event identification, automatic reasoning, incident monitoring and block chain.
- The event identification uses real-time data from social networks. Automatic reasoning extracts the information and knowledge from accessible data using intelligent techniques.

- Incident monitoring, processes the knowledge-based professional emergency using the sensory interfaces and blockchain framework analyse the security and transparency of system and
  similarly the proof-of-authority for having the secure and stable system based on trust.

Advantages :

- No centralized control
- Blockchain ensures security and integrity
- AI to decide the message category

IV. OBJECTIVES :

- During the emergency or disaster situation fake news may spread in the social media platform.
- Single authority control on the social media platform.
- Right information may not be believed by most of the users.

The main concept of this research is to extract the right information which is sharing in social media during crisis by using the recent technologies which gives us the trustworthy and secure information to avoid fake contents and fake users.

V. FEATURES :

- Event detection.
- Natural Language Processing to process the text data.
- Etherium Blockchain to store the records.
- Using ML and deep learning like Naive Bayes, K-Nearest Neighbour, Support Vector Machine, Logistic Regression, XGBoost, cnn to train the data.

VI. IMPLEMENTATION :

K-NEAREST NEIGHBOR (KNN) ALGORITHM PSEUDOCODE :

Let (X_i, C_i) where i = 1, 2……., n be data points. X_i denotes feature values &C_i denoteslabelfor X_i for each i.Assuming the number of classes as 'c'$C_i \in \{1, 2, 3, \ldots\ldots, c\}$ for all values of i
Let x be a point for which label is not known, and we would like to find the label class using k-nearest neighbor algorithms.
Calculate"$d(x,x_i)$"i=1,2,….., **n**; where **d** denote the Euclidean distance between the points.
Arrange the calculated **n** Euclidean distances in non-decreasing order.
Let **k** be a +ve integer, take the first **k** distances from this sorted list.
Find those **k**-points corresponding to these **k**-distances.
Let $\mathbf{k}_i$ denotes the number of points belonging to the i^{th} class among **k** points

i.e. $k \geq 0$

If $k_i > k_j \ \forall \ i \neq j$ then put x in class i.

REGRESSION :

Regression analysis is one of the most important fields in statistics and machine learning. Thereare many regression methods available. Linear regression is one of them.

What Is Regression?

Regression searches for relationships among variables.

For example, you can observe several employees of some company and try to understand how their salaries depend on the **features**, such as experience, level of education, role, city they work.

Similarly, you can try to establish a mathematical dependence of the prices of houses on their areas, numbers of bedrooms, distances to the city center, and so on.

Generally, in regression analysis, you usually consider some phenomenon of interest and have a number of observations. Each observation has two or more features. Following the assumption that (at least) one of the features depends on the others, you try to establish a relation among them.

In other words, you need to find a function that maps some features or variables to others sufficiently well.

The dependent features are called the dependent variables, outputs, or responses. The independent features are called the independent variables, inputs, or predictors.

Regression problems usually have one continuous and unbounded dependent variable. The inputs, however, can be continuous, discrete, or even categorical data such as gender, nationality, brand, and so on.

It is a common practice to denote the outputs with y and inputs with x. If there are two or more independent variables, they can be represented as the vector $\mathbf{x} = (x_1, \ldots, x_r)$, where r is the number of inputs.

When Do You Need Regression?

Typically, you need regression to answer whether and how some phenomenon influences the other or **how several variables are related**. For example, you can use it to determine if and to what extent the experience or gender impact salaries.

Regression is also useful when you want **to forecast a response** using a new set of predictors. For example, you could try to predict electricity consumption of a household for the next hour given the outdoor temperature, time of day, and number of residents in that household.

Regression is used in many different fields: economy, computer science, social sciences, and so on. Its importance rises every day with the availability of large amounts of data and increased awareness of the practical value of data.

LINEAR REGRESSION :

Linear regression is probably one of the most important and widely used regression techniques. It's among the simplest regression methods. One of its main advantages is the ease of interpreting results.

Problem Formulation

When implementing linear regression of some dependent variable y on the set of independent variables $\mathbf{x} = (x_1, \ldots, x_r)$, where r is the number of predictors, you assume a linear relationship between y and $\mathbf{x}$: $y = \beta_0 + \beta_1 x_1 + \cdots + \beta_r x_r + \varepsilon$. This equation is the **regression equation**. β_0, β_1,

Linear regression calculates the **estimators** of the regression coefficients or simply the **predicted weights**, denoted with b_0, b_1, …, b_r. They define the **estimated regression function** $(\mathbf{x}) = b_0 +$

$b_1 x_1 + \cdots + b_r x_r$. This function should capture the dependencies between the inputs and output sufficiently well.

The **estimated** or **predicted response**, $(\mathbf{x}_i)$, for each observation $i = 1, \ldots, n$, should be as close as possible to the corresponding **actual response** y_i. The differences y_i - $(\mathbf{x}_i)$ for all observations $i = 1, \ldots, n$, are called the **residuals**. Regression is about determining the **best predicted weights**, that is the weights corresponding to the smallest residuals.

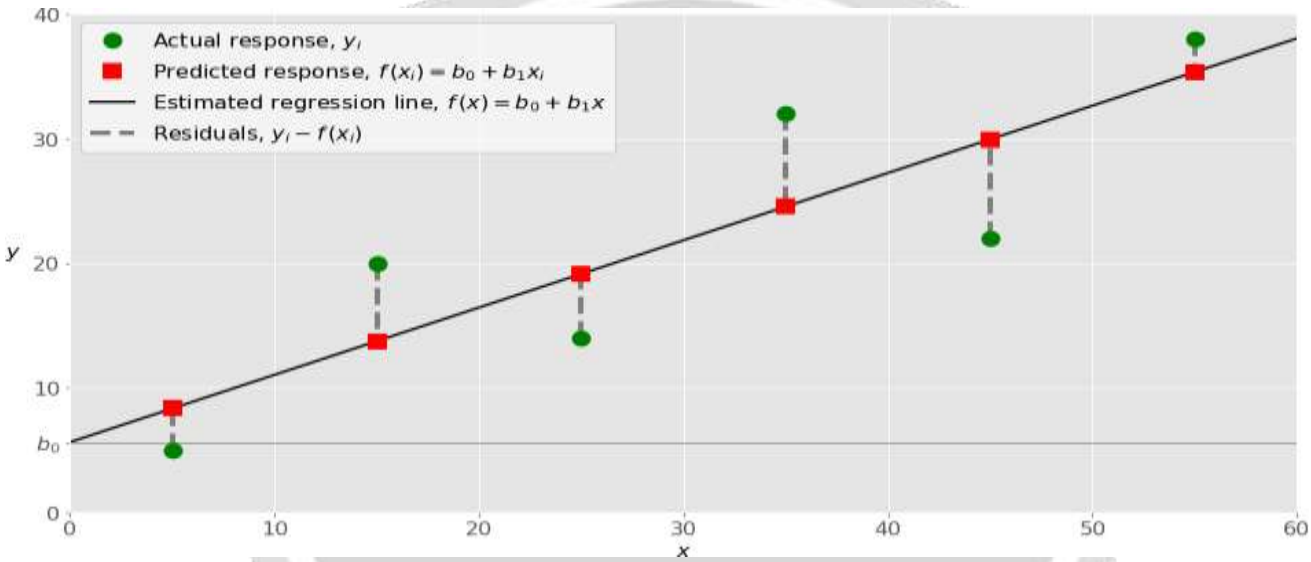
To get the best weights, you usually **minimize the sum of squared residuals** (SSR) for all observations $i = 1, \ldots, n$: $\mathrm{SSR} = \Sigma_i (y_i - f(\mathbf{x}_i))^2$. This approach is called the **method of ordinary least squares**.

REGRESSION PERFORMANCE :

The variation of actual responses y_i, $i = 1, \ldots, n$, occurs partly due to the dependence on the predictors $\mathbf{x}_i$. However, there is also an additional inherent variance of the output.

The **coefficient of determination**, denoted as R^2, tells you which amount of variation in y can beexplained by the dependence on **x** using the particular regression model. Larger R^2 indicates a better fit and means that the model can better explain the variation of the output with different inputs.
The value $R^2 = 1$ corresponds to SSR = 0, that is to the **perfect fit** since the values of predicted and actual responses fit completely to each other.

SIMPLE LINEAR REGRESSION :
Simple or single-variate linear regression is the simplest case of linear regression with a single independent variable, $\mathbf{x} = x$.
The following figure illustrates simple linear regression:

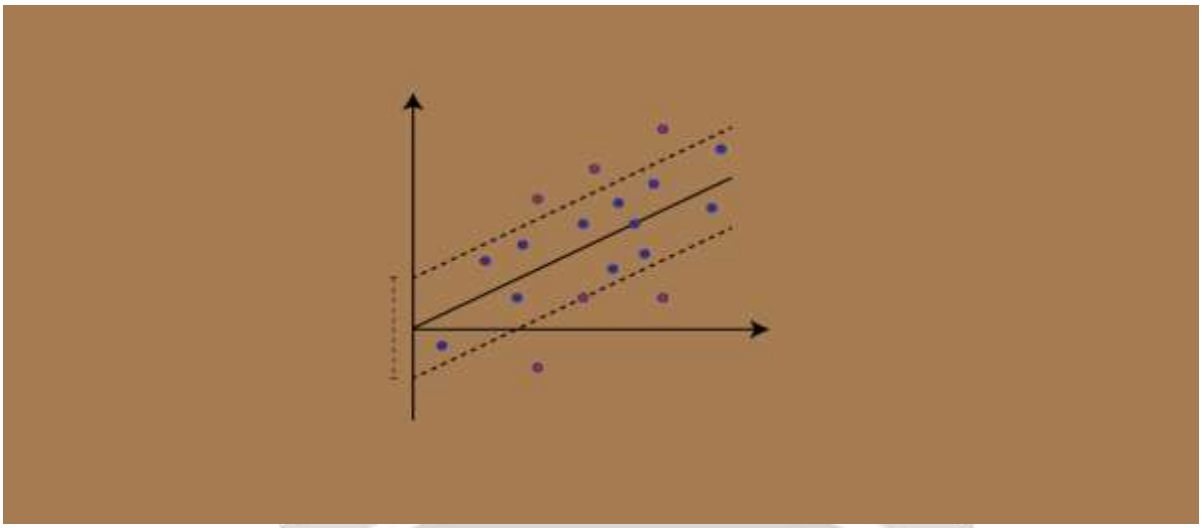


SIMPLE LINEAR REGRESSION

SUPPORT VECTOR REGRESSION ALGORITHM :

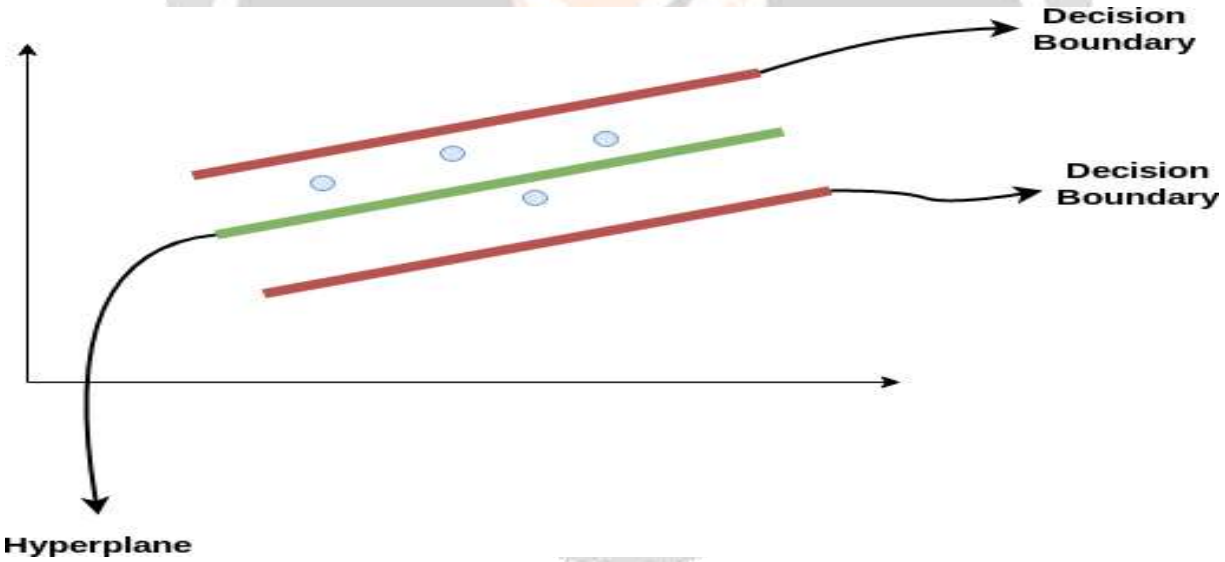
Support Vector Machines (SVM) are popularly and widely used for classification problems in machine learning. I've often relied on this not just in machine learning projects but when I wanta quick result in a hackathon.

But SVM for regression analysis? I hadn't even considered the possibility for a while! And evennow when I bring up "Support Vector Regression" in front of machine learning beginners, I often get a bemused expression. I understand – most courses and experts don't even mention Support Vector Regression (SVR) as a machine learning algorithm.



SUPPORT VECTOR REGRESSION ALGORITHM

The problem of regression is to find a function that approximates mapping from an input domainto real numbers on the basis of a training sample. So let's now dive deep and understand how SVR works actually.



SUPPORT VECTOR REGRESSION

The problem of regression is to find a function that approximates mapping from an input domain to real numbers on the basis of a training sample. So let's now dive deep and understand how SVR works actually.

Consider these two red lines as the decision boundary and the green line as the hyperplane. Our objective, when we are moving on with SVR, is to basically consider the points that are within the decision boundary line. Our best fit line is the hyperplane that has a maximum number of points.

The first thing that we'll understand is what is the decision boundary (the danger red line above!).Consider these lines as being at any distance, say 'a', from the hyperplane. So, these arethe lines that we draw at

distance '+a' and '-a' from the hyperplane. This 'a' in the text is basically referredto as epsilon.

Random forest is a type of supervised machine learning algorithm based on ensemble learning. Ensemble learning is a type of learning where you join different types of algorithms or samealgorithm multiple times to form a more powerful prediction model. The random forest algorithm combines multiple algorithm of the same type i.e. multiple decision trees, resulting in a forest of trees, hence the name "Random Forest". The random forest algorithm can be used for both regression and classification tasks.

Assuming that the equation of the hyperplane is as follows:

Y = wx+b (equation of hyperplane)wx+b= +a
wx+b= -

-a < Y- wx+b < +a

Thus, any hyperplane that satisfies our SVR should satisfy

To solve this regression problem we will use the random forest algorithm via the Scikit-Learn Pythonlibrary

What is a hyperplane?

As a simple example, for a classification task with only two features (like the image above), you canthink of a hyperplane as a line that linearly separates and classifies a set of data. Intuitively, the further from the hyperplane our data points lie, the more confident we are that they have been correctly classified. We therefore want our data points to be as far away from the hyperplane as possible, while still being on the correct side of it. So when new testing data is added, whatever side of the hyperplaneit lands will decide the class that we assign to it.

How do we find the right hyperplane?

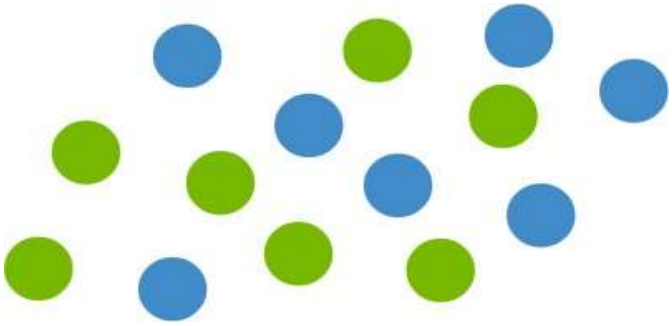
In other words, how do we best segregate the two classes within the data?

The distance between the hyperplane and the nearest data point from either set is known as the margin. The goal is to choose a hyperplane with the greatest possible margin between the hyperplane and any point within the training set, giving a greater chance of new data being classified correctly.

RIGHT HYPER PLANE

But what happens when there is no clear hyperplane?

This is where it can get tricky. Data is rarely ever as clean as our simple example above. A dataset will often look more like the jumbled balls below which represent a linearly non separable dataset.

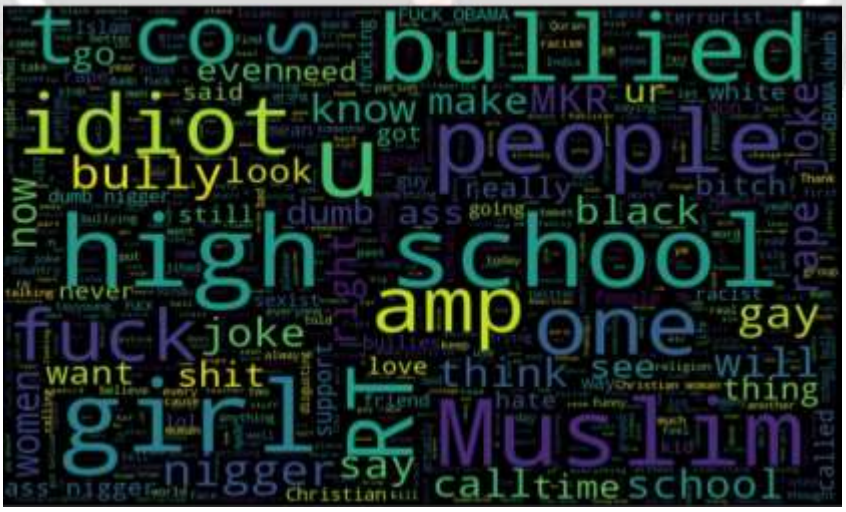


NO CLEAR HYPERPLANE

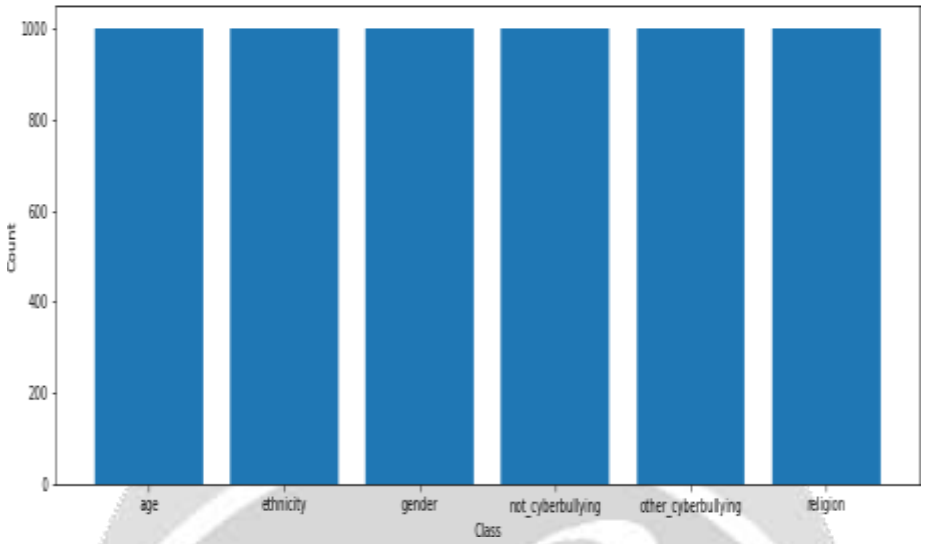
In order to classify a dataset like the one above it's necessary to move away from a 2d view of the data to a 3d view. Explaining this is easiest with another simplified example. Imagine that our two sets of colored balls above are sitting on a sheet and this sheet is lifted suddenly, launching the balls intothe air. While the balls are up in the air, you use the sheet to separate them. This 'lifting' of the balls represents the mapping of data into a higher dimension. This is known as kernelling.

Because we are now in three dimensions, our hyperplane can no longer be a line. It must now be a planeas shown in the example above. The idea is that the data will continue to be mapped into higher and higher dimensions until a hyperplane can be formed to segregate it.

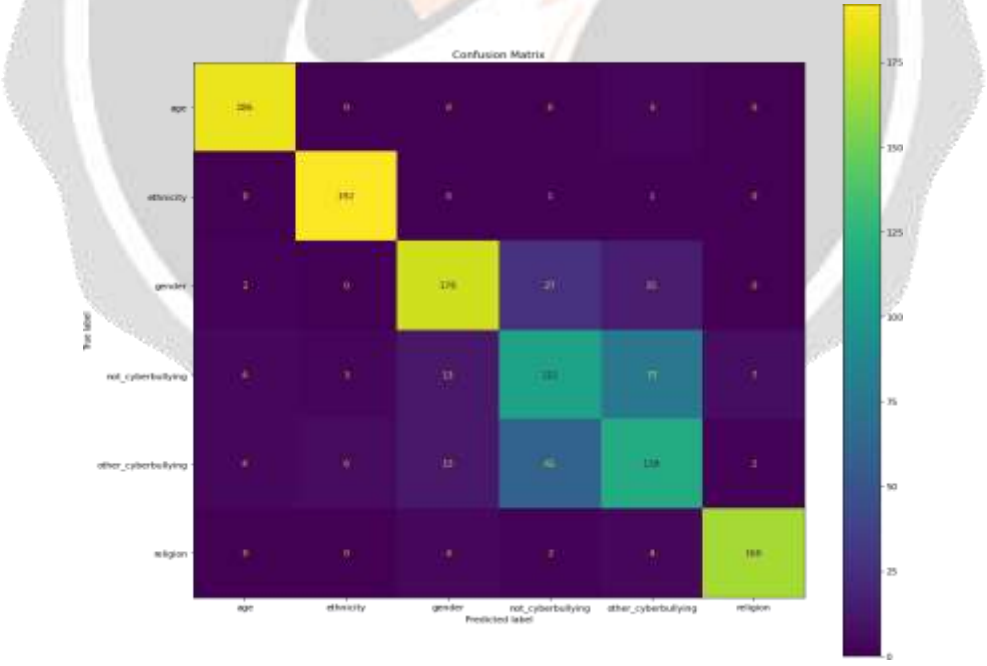
VII. RESULTS :



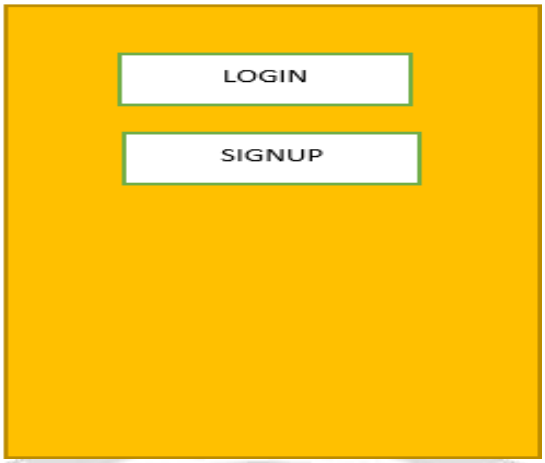
Cloud of Words



Class Balance Graph

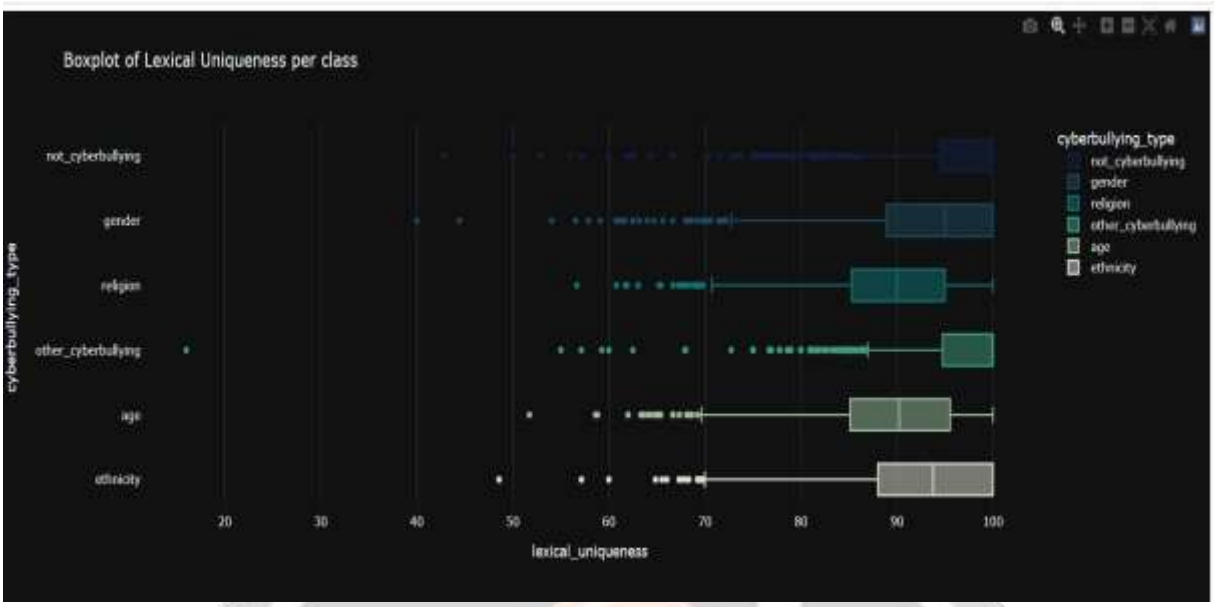


Confusion Matrix

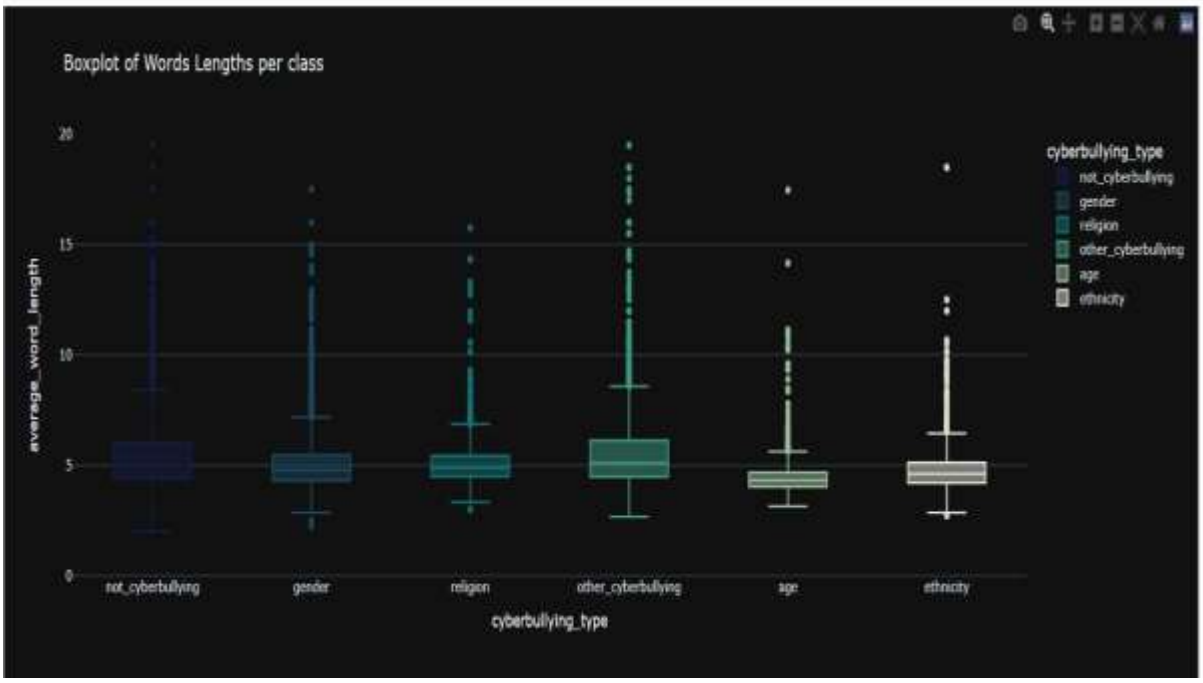


Main Activity

Sign Up



Uniqueness per class



Length per class

VIII. CONCLUSION:

From this system we make people get knowledge about the daily news and crisis. This helps to give the society a better knowledge and have great use of the social media platforms. Also it useful to detect fake news so that the circulation of fake news can be stopped and make people know the news which would be

helpful . Blockchain helps in having the data security as well as the data is kept in nodes which will easy to fetch. Natural language processing helps in making the language easier so that the people can understand easily. At the end it is more beneficial for the society due to which no problems are caused and keeps the people updated.

IX. REFERENCES:

[1] J Abbas, D Wang, Z Su, A Ziapour. 2021 . "The role of social media in the advent of COVID-19 pandemic: Crisis management, mental health challenges and implications."

[2] Mateusz Fedoryszak, Brent Frederick. 2019 ."Real-time Event Detection on Social Data Streams."

[3] Faisal Jamil 1 , Lei Hang 1 , KyuHyung Kim 2 . 2019 ."A novel medical blockchain model for drug supply chain integrity management in a smart hospital"

[4] Faisal Jamil 1 , Lei Hang 1 , KyuHyung Kim 2 . 2019 ."A novel medical blockchain model for drug supply chain integrity management in a smart hospital"

	Student at T john Institute of technology Department of information Science and engineering
	Student at T john Institute of technology Department of information Science and engineering
	Student at T john Institute of technology Department of information Science and engineering

	Student at T john Institute of technology Department of information Science and engineering
	Stud at T john Institute of technology Department of information Science and engineering