FACULTY OF ENGINEERING AND COMPUTING

SCHOOL OF COMPUTING

MSc. DEGREE

IN

SOFTWARE ENGINEERING

FINAL DISSERTATION

Name: Hettigodage Isuru Udara Wickramapala

ID Number: K2240860

Project Title: To research the use of AI Framework for understanding the behavioural patterns of terrorists, hetaerism and toxicity in online spaces.

Supervisor: Mr. Dimuthu Thammitage

Date: 02/01/2023

Declaration

Module: CI7000 Deadline: Monday 2nd January 2023 at 23.59pm

Module Leader: Professor. Ruvan Abeysekara Student ID: -KU ID - K2240860

PLAGIARISM

You are reminded that there exist regulations concerning plagiarism. Extracts from these regulations are printed below. Please sign below to say that you have read and understand these extracts:

(Signature:) Date: 009/04/23

This header sheet should be attached to the work you submit. No work will be accepted without it.

Dedication

Extracts from University *Regulations* on Cheating, Plagiarism and Collusion

Section 2.3: "The following broad types of offence can be identified and are provided as indicative examples…

(i) Cheating: including taking unauthorized material into an examination; consulting unauthorized material outside the examination hall during the examination; obtaining an unseen examination paper in advance of the examination; copying from another examinee; using an unauthorized calculator during the examination or storing unauthorized material in the memory of a programmable calculator which is taken into the examination; copying coursework.

(ii) Falsifying data in experimental results.

(iii) Personation, where a substitute takes an examination or test on behalf of the candidate. Both candidate and substitute may be guilty of an offence under these Regulations.

(iv) Bribery or attempted bribery of a person thought to have some influence on the candidate's assessment.

(v) Collusion to present joint work as the work solely of one individual.

(vi) Plagiarism, where the work or ideas of another are presented as the candidate's own.

(vii) Other conduct calculated to secure an advantage on assessment.

(viii) Assisting in any of the above.

Some notes on what this means for students:

1. Copying another student's work is an offence, whether from a copy on paper or from a computer file, and in whatever form the intellectual property being copied takes, including text, mathematical notation and computer programs.

2. Taking extracts from published sources *without attribution* is an offence. To quote ideas, sometimes using extracts, is generally to be encouraged. Quoting ideas is achieved by stating an author's argument and attributing it, perhaps by quoting, immediately in the text, his or her name and year of publication, e.g. "$e = mc^2$ (Einstein 1905)". A *references* section at the end of your work should then list all such references in alphabetical order of authors' surnames. (There are variations on this referencing system which your tutors may prefer you to use.) If you wish to quote a paragraph or so from published work then indent the quotation on both left and right margins, using an italic font where practicable, and introduce the quotation with an attribution.

Dedication

I would dedicate this project to all humble begins who have dedicated me in any way to become what I'm today. Whose sacrifices and seeded our success: especially my parents & teachers who have felt my pains and struggles beyond myself and showered us win as well as never-ending courage and support. I deem them as the divine source of my inspiration.

Acknowledgements

I would like to grant my gratitude and thanks to my beloved parents and all my teachers from the preschool to the present who educated and guide me. Further, must remember the extended support and the encouragements provided by our Project Supervisor Mr. Dimuthu Thamittage as well as Second Supervisor Mr. Vikum Jayasunarda, Prof. Ruwan Abesekara and the course coordinator Mrs. Samadhi Bandara and all others who contributed their immense support to make this attempt a success to finalize the dissertation within the restricted time frame. In fact, researching the use of AI Framework for understanding the behavioural patterns of terrorists, hetaerism and toxicity in online spaces as well as creating the dissertation lead me to explore and read information more about the related topic. Thus, the knowledge that gained by completing the dissertation helped to extended the knowledge which would increasingly support for the future career. Once Again, I would like to thank everyone for their contributions.

Abstract

Social Media and social network interactions play a massive role in our day-to-day lives and sometimes there are times when such platforms are used in a toxic and hateful manner. terrorism, abusive language, aggression, cyberbullying, and hatefulness are prime examples. Being victimized by terrorism, hetaerism, racism and similar aspects can cause the victim annoyance, needless fury, and the breeding of hatred. In order to create civilized societies and safe live hood, which are free of toxicity and abusive behaviour, it is a must to thoroughly do a clean brush up on the internet's major sectors, in order to counter them. This report focuses on researching the use of AI Framework for understanding the behavioural patterns of terrorists, hetaerism and toxicity in online spaces to build a calm, inclusive and good community in Sri Lanka. Terrorists and Abusive organizations are always looking for methods to outwit law enforcement and counterterrorism organizations. Understanding unforeseen assaults are becoming increasingly important. So that the relevant counter-terrorism organizations and governments could be ready to take necessary actions for security and sovereignty. Thus, studying various user contents in online spaces will assist to decide whether the relevant personnel might be terrorists or not, so understanding the unforeseen risks and threats would be much more obvious. In this report, a new strategy is suggested that identifies hetaerism , toxicity and terrorism based on Artificial Intelligence to analyse behaviour patterns of users in online spaces from the contents such as posts, videos, opinions etc.

The research is conducted to seek the possibility of identifying terrorism, abusiveness, aggression, cyberbullying, and hatefulness in Sri Lanka by analysis of content posted over Online Spaces using Artificial Intelligence, based on a selected sample population in the Kandy region. . Numerous books, journals, and prior literature evaluations were inherited to support the research. The researcher was able to eliminate ineffective particular objectives from the outset and use only the obvious priorities to carry out the research-based literature review. Thus, the researcher was able to focus on the specific goals of this research, which included determining the proportion of the Sri Lankan population using social media, determining whether all social media users share their genuine views and opinions on social media, determining whether it is possible to filter out terrorism, hate, racism, and similar aspects and opinions, and determining whether online space users would consent to an analysis using artificial intelligence. Ultimately, a prototype AI Model would be implemented into a developed social media website to demonstrate the practical implementation of the suggested solution. The proposed strategy entails developing a deep neural network and multi-classification model to especially detect the semantics of Terrorism, hate, racism and similar aspects integrated into a web application depicting a typical comment section. The BERT algorithm, a Deep learning algorithm was used via phrasal interpretation rather than word-by-word analysis, regardless of whether the text is informal or slang.

Keywords: Deep Neural Network, CNN, BERT, Text Classification, Terrorism, Hetaerism, Machine Learning, Abusiveness

Table of Contents

List of Figures

List of Tables

Glossary of Terms

BERT - Bidirectional Encoder Representations from Transformers

AI- Artificial Intelligence

LSTM- Long short-term memory

RNN- Recurrent neural network

CNN- Convolutional neural network

SVM- Support vector machines

GPU- Graphical Processing Unit

HTTP- Hypertext Transfer Protocol

DBMS- Database Management Systems

URL- Uniform Resource Locator

API- Application Programming Interface

NLP- Natural Language Processing

CBOW- Continuous Bag of Words Model

ISIS - Islamic State in Iraq and Syria

IV – Independent Variable

DV – Dependent Variable

Chapter 01 - Introduction & Background

1.1.Introduction

The digital world has both benefits and drawbacks. People use online spaces with more freedom to interact, discuss, publish, respond, express, and more. Most notably, online spaces such as social media, forums, groups, sites, hubs, and other similar platforms have become an eternal trend in the previous decade. As a result, a user has got many techniques and tactics to share, express, or promote their thoughts in order to be heard by many people on such platforms. For example, a simple social media post can be viewed by other users and if a particular user intends to comment on the post, it can be easily done by simply dropping down a text in the comment section where other users could read and react. Similarly, on every other platform such as Facebook, Instagram, and Twitter users use their freedom of expression to exchange any form of expression under the comment section, forum, or similar areas.

However, there are instances when such platforms are used in a toxic, abusive, or unsuitable manner. Although commenting or expressing one's opinion is their right. There are certain instances where such activities could not be accepted. Commenting or expressing hateful thoughts by specifically targeting a group, race, or ethnicity cannot be considered under the act of freedom to express. Usage of abusive language, terrorism, aggression, cyberbullying, hatefulness, insults, personal insults, provocation, racism, sexism, threats, or toxicity manifest themselves in a variety of ways and are some of the prime examples of the problem that is addressed here, and it is only actively growing as the digital world expands. The most recent terrorism incident that happened in Sri Lanka on Easter Sunday witnessed the above-mentioned fact. Terrorist suicide bombers attacked a number of hotels and churches on Easter Sunday, April 21, 2019. Over 300 individuals were killed and over 500 people were injured, including foreigners, in the tragedy. The Sri Lankan authorities have disclosed that the terrorist's affiliation with the ISIS Group. Later, security forces discovered a variety of weaponry, including swords, bombs, detonators and firearms, that had been collected from various areas of the nation for related terrorist acts which leads the authorities to detain nearly 500 people in connection with the Easter Sunday attack. During the interrogation of these terrorists, it was discovered that their leader had made multiple posts on terrorism views on social media platforms, and that they had communicated on various social media platforms, which helped them to carry out the blasting successfully despite the fact that they were from different areas of Sri Lanka. The blasts may have been intercepted if these posts and their behaviours had been previously examined. So researchers have shown a significant interest in researching the possibilities of detecting terrorism, toxicity, abusive language, aggression, cyberbullying, and hatefulness in Sri Lanka by analysing information uploaded on any online spaces using AI to assist the government and relevant agencies in preventing such tragedies in the future.

Eradicating acts of terrorism, toxicity, hetaerism and similar context on the platforms mentioned above could pave the way to building up a good internet society where people could express their thoughts and views in a civilized manner, thus helping humanity to advance further, rather than creating the online spaces a tool to cause harm one another. Moreover, if this problem persists there will be some serious consequences that are usually not given much importance, in general, most of the time on the internet. Being a victim of terrorism, hetaerism or similar context causes fear, sadness, negligence, annoyance, unneeded rage, and the breeding of hatred in the hearts of that specific person, as well as in bystanders who are witnesses to the act. To build up civilized communities and safe spaces with no terrorism, toxicity or abusive behaviours by cleansing the important sectors of the internet for good communication exchange, the research is conducted to seek the possibility of identifying terrorism, abusiveness, aggression, cyberbullying, and hatefulness in Sri Lanka by analysis of content posted over Online Spaces to develop an Artificial Intelligence prototype model to. This Artificial Intelligence model will be developed using AI algorithms capable of detecting any context related to terrorism, abusiveness, aggression, cyberbullying, and hatefulness by phrasal interpretation rather than by word-by-word interpretation. Finally, the proposed solution is a well-trained AI model which is possible to embed into a platform or any other software and extend its functionality.

1.2.Background and Motivation

As per the above-discussed problem domain, initially, there isn’t a stable, convenient solution related to terrorism, toxicity, and hetaerism detection, but with the rapid development in the computing sector, various Artificial Intelligence strategies and technologies were introduced. These models were standard methods that included simple instructions with hardcoded scripts that came into play as the core function of these methods was the words and phrases which are manually inspected and evaluated by humans utilizing common offensive language terms (Seda & Altin, 2019). Overall, even though these models did the job of detecting general offensive terms for the time being, due to the advancement of digital interactions terrorism, toxicity, and hetaerism in online spaces grew broader making it an insufficient solution. The rise of AI and machine learning with various algorithms further led the path for the development of AI models which were engineered by training a specific set of data sets to predict such offensive terms. (Seda & Altin, 2019). Although this proved to be very effective in predicting or detecting common offensive terms, it was not accurate enough to detect or predict modern phrases and slang terms which could be interpreted as an offensive phrases or a term. For example, *"…it is a world run by the Zionist Jewish Influence and Race Tainting Paedophiles that are only here to rape our heritage and destroy the qualities that make us White People great…*". (Australian Human Rights Commission., 2021) With that being said, there are flaws when it comes to detecting modern phrases and slang using the regular detection system that is mentioned above. Similarly, as per the easter attack scenario during the interrogation of these terrorists, it was discovered that the relevant group has used various slang terms and unique patterns of language to express their point of view on online spaces. The blasts could be prevented if these posts and the behaviours could be foreseen and predicted. So, the researcher has shown a significant interest in researching the possibilities of detecting terrorism, toxicity, abusive language, aggression, cyberbullying, and hatefulness in Sri Lanka by analysing information uploaded on any online spaces using AI to assist the government and relevant agencies. Moreover, although there are strict policies and regulations that are imposed by many popular online platforms, it isn’t as effective as expected. To back this up recently, Facebook’s whistle-blower alleges that the company is making online hetaerism and extremism worse where it was prioritizing profits over the well-being of users (Isaac, 2021). In fact, countering terrorism, toxicity and hetaerism online would be quite difficult on platforms because it may even affect such platforms' business. However, with a clear motive and the right equipment, encountering offensive expressions online will be feasible. Being motivated to come up with a unique model that could contribute to building up a safer and non-toxic space online will be a major breakthrough in countering any modern offensive expressions on platforms like social media which through implementing an AI model using deep learning approaches such as convolutional neural networks (CNN), recurrent neural networks (RNN), LSTM, or any other new deep learning algorithm out in the market. This model will be specifically well-trained to predict any offensive aggressiveness by interpreting phrases that may be formal or informal depending on the user’s input.

1.3.Problem in brief

Social media is a global platform where individuals can share their opinions, images, videos, and day-to-day activities. It is also a place where people may interact with others and start new relationships. As a significant percentage of the population regularly uses social media for informational and entertainment purposes as well as it serves as a tool for online marketing and news publishing, it is crucial to maintain a high quality for the content posted across all social media platforms. The most popular social media platforms, including Facebook, Instagram, LinkedIn, Imo, WhatsApp, and others, have a tendency to communicate personal opinions in relatable and varied tones, but such opinions or remarks pose a higher risk of posing a threat to the security of an individual or a nation (Seven Media Group, 2018). Social media and social network contacts play a significant role in modern lives, and such platforms are occasionally used poisonously and with hostility. Examples include terrorism, abusive language, hostility, cyberbullying, and hatred. Being afflicted by terrorism, hatred, racism, and other comparable factors can generate frustration, unnecessary rage, and the fostering of hatred. In order to develop civilized communities and safe living environments devoid of toxicity and abusive conduct, it is necessary to fully clean up the internet's primary sectors. Along with the recent terrorist attacks and related incidents, it is much more important to track down the ideas and people that encourage and support terrorism and related acts. As a result, the author is evaluating whether the social media analysis might endanger personnel or undermine national security through the research. It is also investigated whether it is possible to learn about the behaviours and actions of terrorists or aggressive groups and to what extent this will be useful in establishing a nation's security. The study is carried out and continued utilizing the public responses and feedback on the use of AI to analyse the data and information retrieved through social media. Thus, the research focuses on the idea of using the research's findings to implement improved social network usage and interpersonal relationships based on the AI Model. The suggested approach involves the development of Deep Neural Network structures that operate as feature extractors and are particularly successful at capturing the semantics of terrorism, hetaerism and toxicity and recognizing dangerous content, such as terrorism, hetaerism and toxicity. Deep learning algorithms, such as BERT, will be used to detect terrorism, hetaerism and toxicity or similar content by phrasal interpretation rather than word-by-word interpretation, whether casual or slang, and this model may be implemented into a web application.

1.4. Proposed Solution

In general, the proposed solution is determined based on two important components. The research component and the implementation component. The main component is where the research relies on which is to understand user behaviour in online environments is crucial for the research component. Second, realistic user behaviour models for online platforms are essential for both viral marketing and sociological studies. For instance, viral marketers may wish to employ user engagement models to distribute their content or advertisements swiftly and extensively. Third, while developing the future formation of Internet infrastructure and content distribution systems, it is helpful to understand how the workload of online spaces is changing the nature of Internet traffic. The significant importance of carrying out this research is to identify the inheritance of AI being used to monitor social media activities regarding terrorism, toxicity, hetaerism, and similar context by reckoning AI to understand even the general opinion on what is trending on the online spaces in a country. It is where the Artificial Intelligence classification model and its core functionalities are based on. The implementation component includes the software artefact (the AI Model and the Web application). Where the AI Prototype will be integrated into the solution to a typical use case in the Web Application for demonstration.

To build the multi-classification model, machine learning tools and frameworks will be used. Also, Kaggle’s notebook (Previously used as Google Collaboratory Notebook) which includes free GPUs for faster training will be used as well. TensorFlow, Keras, NumPy, Pandas will be some other tools used to develop the model. Finally, the model will be saved and served using FAST API as a microservice (RESTful). A simple HTTP request (POST) with the input will allow the microservice to return back the prediction as a response. Most importantly, this microservice was dockerized into an image that runs in a Google Cloud Compute Instance. The subcomponent (Web Application) in the implementation component depicts the typical use case of a social media application which has the ability to upload a post, react and most importantly comment. The web application will have the main component (AI Model) in the implementation component integrated into the server of the web application. Whenever the user inputs a comment or expression the server would send a request to the microservice and evaluates the aggressiveness of the user’s input. This application is built using NEXT Js on the client side and Node Js is incorporated in the server side with Mongo DB as the DBMS.

Moreover, this application uses several popular libraries to enhance its usability of the application which are,

- Express Js
- Mongoose
- Redux and Redux Toolkit for state management and caching mechanisms for HTTP requests
- Formik, Yup for form validation

1.5. Aims and Objectives

The aim and the objectives which are defined for this project as follows,

1.5.1. Aim

To research the usage of AI Framework for comprehending the online toxicity, hetaerism, and terrorism behaviour patterns, and to design an AI Prototype to create a peaceful, inclusive, and decent community.

1.5.2. Objectives

- RO1: To figure out the percentage of the population using social media in Sri Lanka.
- RO2: To analyse whether all the social media users share their honest views and opinions on online spaces.
- RO3: To analyse whether filtering terrorism, toxicity and hetaerism views and opinions are possible.
- RO4: To identify the agreement of users to allow an AI Model to analyse online spaces.
- RO5: To develop an AI Prototype and a web application to understand the behavioural patterns of terrorism, toxicity, hetaerism, and the similar context in online space.

1.5.3. Main Question of the Research Component

Can use an AI Framework to understand behavioural patterns of toxicity, hetaerism, and terrorism in online spaces?

1.5.4. Specific questions of research Component

1. What percentage of the population of Sri Lanka uses social media?
2. Do all social media users share their honest views and opinions on online spaces?
3. Is it possible to identify terrorism, toxicity, and hetaerism by analysing views and opinions shared through social media?
4. Are the users comfortable with conducting an analysis using Artificial Intelligence in online spaces?

Chapter 02 - Literature Review

2.1.Role of social media

The tremendous rise of social media over the last decade reflects its relevance and integration into the daily lives of many people in Sri Lanka. Parallel to this, there has been significant growth in digital journalism focused on web media. However, as compared to traditional channels, social media has far higher reachability. As a result, social media has been transformed into online news media and has surpassed traditional media in digital journalism. This has become the quickest medium, connecting people all over the world with no boundaries. The increased use of social media around the world affects the number of television watchers and radio listeners who are addicted to social media, regardless of age group. Similar to the benefits given, this may lead to cybercrime, which has become a privacy danger to individuals all over the world. (Guardian News & Media Limited, 2018) The spreading of false information and violent/ prejudicial sentiments can have grave negative effects on people in times of crisis. Emotional individuals including victims, their families, or the exasperated public can be easily manipulated and angered. Violent words lead to violent actions causing chaos and riots that a country like Sri Lanka cannot afford. Information becomes distorted making it hard to distinguish real evidence and facts from falsely manifested information which can make investigations slower and harder to resolve. Soon after the Easter attack and in a few other tense situations followed by the Easter attack, the Sri Lanka government imposed a temporary ban on social media throughout the country.

Social networking sites are web-based services that allow anyone to build their own personal profile with their own list of users and thereby communicate with them in an entirely public forum that offers features such as chatting, blogging, video calling, mobile connection, and video/photo sharing. People spend more time than normal on social networking sites downloading images, browsing through updates, finding amusement, and chatting with pals to stay connected (Khurana, 2015). The rising usage of the Internet as a new communication tool has transformed the way users interact. A new form of internet communication has arisen, each with its own set of idiosyncrasies. This new communication style emerges as a result of the use of social networking sites. (Witold , et al., 2020) (Kulandairaj, 2014). The usage of social media has become widespread, with the most popular platforms today being Facebook, Instagram, Twitter, Myspace, LinkedIn, Skype, and others. Any of the aforementioned online spaces allow users to engage with one another and create and sustain relationships that inspire others to join the online community.

2.2.Artificial Intelligence

Simply, artificial intelligence is software that is supposed to make intelligent judgments or accurate predictions in a certain problem domain (Vairavan & Arvind, 2019). AI is now widely accepted as the source that bridges the physical and virtual worlds. Aside from the fear-mongering uses of AI in the future, the plain reality is that the code or core technology underlying AI hasn't evolved in decades. There have been no fundamental breakthroughs, and the only thing that has changed, or rather progressed, is the computing capability of devices, which allows for quicker data processing (Alrawashdeh & Ahmad, 2019). It is critical to understand how AI may play a significant role in estimating and comprehending how people behave at any given event. It is behavioural data that may be collected without sampling. Recent instances of smartwatches serving as sensors not only provide essential information but also reduce the time that would otherwise be spent obtaining that type of data. It's like an unconscious input that isn't related to filling out forms. Many similar efforts to monitor stress or health are currently in existence. For example, the Stress Sense detects when individuals are agitated and assist them in avoiding stressful situations. To help people stay in shape, behaviour analysts examine physical activity habits. Researchers have frequently proposed data repurposing, which is modifying historical data to better understand human behavioural patterns. Similarly, the activities that persons engage in on social media reveal a lot about their mental health and psychological stability (Heath & Nick, 2016). Individuals' lives can be improved by repurposing such logs (Sentance & Rebecca, 2017). Big data and artificial intelligence questions clarify how a person with a specific mindset reacts to a circumstance and whether it is feasible to change or comprehend. It also provides an understanding of the many ways in which human beings react to any given threat or people react in an aberrant manner. To obtain insight into this, one requires extended data as well as pattern understanding. A psychologist with prior knowledge of data science and artificial intelligence has the potential to use the capabilities to identify solutions to difficulties.

This technology also enables us to understand what influences human behaviour and how other factors influence it (Bansal & Gauri, 2018). As artificial intelligence has established the standard in every other area, it is now ready for behavioural science too (Lin & Karen, 2018). A human's mental process influences his/ her behaviours, which collect data for subsequent analysis. Psychology may be found everywhere, not just in medical terms, but also in our daily lives. The material that a person posts or shares on social media is also something that plainly influences human behaviour (Bajada & Josef, 2019).

2.3.Terrorism, Toxicity and Hetaerism

To begin, it is necessary to define the term terrorism. Terrorism contains the term terror. Terror is derived from the Latin terrere, which meaning "fear" or "tremble." When combined with the French suffix isme (which means "to practice"), it means "to exercise the trembling" or "to cause the frightening." Trembling and scary are synonyms for dread, panic, and anxiety, sometimes known as terror. Terror is a term that has been around for almost 2,100 years. The terror cimbricus was a condition of panic and emergency in ancient Rome in response to the arrival of the Cimbri tribal assassins in 105 BCE. Etymology may be seen in this description of terrorism as being based in terror. The study of the origins and evolution of words is known as etymology. Language, from this perspective, is organic, changing, and variable, based on the demands of thinkers and speakers throughout time and space (Burgess, 2003) (Tuman, 2009).

Various experts have sought to define terrorism throughout the years. However, the phrase is so riddled with conceptual issues that there is no universally recognized definition of it. The irony is that the recurring issue of terrorism has become a daily component of modern political theatre. Turn on the TV or just explore an online space such as Facebook, Twitter, Instagram, and so on to hear about it any time. The following are some of the most eminent scholars and institutions' definitions of terrorism. Terrorism is defined under the League of Nations Convention (1937) as “any illegal activities conducted against a state and designed or planned to instil fear in the hearts of specific individuals, groups of individuals, or the general public.” (League Convention, 1937) whereas the Department of Defence (United States) has addressed terrorism as “the calculated use of criminal violence or the threat of unlawful violence in order to spread terror; designed to force or frighten governments or societies in the pursuit of goals that are often political, religious, or ideological.” (Joint Chiefs of Staff Department of Defense - USA, 2008). Further, the Department of State (United States) has mentioned the terrorism is “Subnational organizations or clandestine state operatives commit deliberate, politically motivated violence against non-combatant targets.” (U.S. Department of State, 1996). It is obvious that not only government entities have especially addressed and discussed about terrorism, but individuals like Walter Laqueur have also mentioned that “Terrorism is defined as the use or threat of using violence, a technique of fighting, or a strategy to attain specific goals. Its goal is to encourage terror in the victim, which is brutal and violates humanitarian principles. Publicity is a critical component in terrorist tactics." (Laqueur, 1987), Bruce Hoffman expressed that the "Terrorism is invariably political in its goals and motivations, violent or, more importantly, impends violence, intended to have far-reaching psychological consequences beyond the immediate victim or target, carried out by an organization with an identifiable chain of command or conspiratorial cell structure (whose members wear no uniform or identifying insignia), and perpetrated by a subnational group or non-state entity." (Hoffman, 2006). Also, the authors Alex Schmid and Albert Jongman stated that the "Terrorism is an anxiety-inducing form of repeated violent action used by (semi-)covert individual, group, or state actors for idiosyncratic, criminal, or political objectives, in which the immediate targets of violence are not the major targets, as opposed to the assassination. The immediate human victims of violence are often picked at random (opportunity targets) or selectively (representative or symbolic targets) from a target demographic to act as message generators" (Schmid & Jongman, 1988).

While the French government was responsible for the Reign of Terror, terrorism in modern times refers to the murdering of humans by nongovernment political actors for a variety of causes, generally as a political statement. In the 1870s, Russian radicals proposed this view. In 1869, Sergey Nechayev, the creator of People's Retribution, saw himself as a terrorist. Johann Most, a German anarchist writer, helped popularize the current essence of the term in the 1880s by dispensing "advice for terrorists." (Crenshaw, 1995). Many governments across the world are extremely opposed to defining terrorism because they are concerned that an official definition of terrorism may expose the validity of self-proclaimed national liberation battles. In certain nations, the term has almost become synonymous with political opponents. The Chinese, for example, label calm Tibetan Buddhists as ferocious terrorists. In Zimbabwe, President Robert Mugabe views the democratic opposition similarly. Terrorism is an insulting phrase (International Bar Association, 2003) (Moeller, 2002). When individuals use the word, they are characterizing their adversaries' conduct as wicked and lacking in human compassion. Terrorism is regarded as eviller than war, torture, or murder. A pejorative phrase is one that has negative and disparaging connotations (White, 2011). More than 200 definitions of terrorism have been discovered via research. According to Simon (1994) (Simon, 1994), at least 212 distinct definitions of terrorism exist across the world, with 90 of them being frequently employed by governments and other organizations. Schmid and Jongman (1988)

(Schmid & Jongman, 1988), two scholars from the University of Leiden (Netherlands), took a social science method to defining terrorism. They investigated over a hundred scholarly and official definitions of terrorism to discover the key components. Individuals discovered that the concept of violence appeared in 83.5% of definitions; political goals appeared in 65% of definitions; causing fear and terror appeared in 51%; arbitrariness and indiscriminate targeting appeared in 21%; and victimization of civilians, non-combatants, neutrals, or outsiders appeared in 17.5%. Schmid and Jongman conducted a content analysis of those definitions (Schmid & Jongman, 1988).

Developed nations have not been resistant to recent waves of terrorist attacks, as evidenced by the following incidents: the terminated 2015 attacks in Verviers, Belgium, the Australian-Sydney catastrophe in December 2014, the attacks in Australia February 2015, the "Charlie Hebdo" incident in Paris 2015, the attacks in France November 2015, as well as the July 2015 attacks at the "Promenade des Anglais" in Nice; and the stream of attacks in Great Britain (22nd of March 2017 Westminster attack, 22nd of May 2017 Manchester Arena bombing, 3rd of June 2017 London attack, 19th of June Finsbury Park Attack and 15th of September London tube train attack). Among the significant determinants of terrorism, social media has been recognized as a method for terrorist recruitment and promotion (Gates & Podder, 2015). Facebook infiltration and terrorism are linked. The posture also addresses contemporary governmental concerns about the lack of documentation on the effects of social media (World Bank, 2016) (Parkyn, 2017). Furthermore, exploratory discourses on the role of social media in terrorism have not been supported by empirical evidence (Patrikarakos, 2017). As a result, this study adds to the terrorism literature by injecting some empirical validity into discourses in order to determine whether the alleged positive link between terrorism and online spaces holds up under inspection. Countries with low levels of terrorism are more strongly related with a favourable Online nexus. Other externalities of terrorism confirm the documented positive relationship: terrorism fatalities, terrorism events, terrorism injuries, and terrorist-related property damage.

The study's viewpoint on the link between social media and terrorism also differs from present global information technology management literature, which has emphasized on, among other things, the relevance of globalisation in information technology trends (Lee & Joshi, 2016), variations in the spread of social media among cultures (Khan & Dongping, 2017). In Europe, patterns of combining information technology and innovation use, cultural practices, and virtual social network diffusion progress in the international hyperlink network young civic engagement behaviour on social media (Billon, et al., 2017) (Krishnan, et al., 2016). The relationships between information technology, information sharing, and inclusive development, as well as the factors influencing information technology in developing nations (Afutu-Kotey, et al., 2017) (Bongomin, et al., 2018)

Unfortunately, social media has increased the reach and scope of harmful content such as disinformation, conspiracies, extremism, harassment, violence, and other types of socially toxic material. While social media sites seek to resist and overcome such damaging information and behaviour, such efforts are mostly ineffectual and may have unexpected consequences. The effort and success of moderation may be influenced by a company's economic interests, political and regulatory factors, or a lack of effective tools and adequate investment in the endeavour (Kursuncu, et al., 2019). Human content moderation has produced largely disappointing results. The 2020 political and public health context pushed society to accept technologies, particularly AI-based, remedies with insufficient knowledge (Kursuncu, et al., 2020). One major factor is a misunderstanding of the difficult nature of toxicity, which requires context outside of the explicit material. Toxicology detection necessitates an interdisciplinary approach based on empirical methodologies. In line with our people content network architecture for characterizing social media interaction, we stress the more general function of context, and specifically cultural context, in content interpretation (Purohit, et al., 2011).

Although there is no common explanation for Toxicity and Hetaerism, the following definitions have been presented by certain writers. hetaerism, according to Metaverse, is "a direct attack on persons based on protected qualities such as race, ethnicity, nationality, disability, religions, caste, gender identity, sexuality, and serious illness." (Hate Speech Transparency Center Meta, 2021). In their research on detecting hetaerism, Thomas Davidson, Dana Warmsley, Michael Macy, and Ingmar Weber defined hetaerism as a kind of discourse meant to communicate hatred towards a certain group of persons in order to mock or humiliate them. Furthermore, Leandro Silva, Mainack Mondal, Denzil Correa, Benevenuto F, and Ingmar Weber defined hatred as any offense influenced by the offender's prejudice towards another person or group of people. According to Twitter, "hateful behavior" is defined as "supporting violence against other persons or actively assaulting or threatening them based on race, ethnicity, national origin, sexual orientation, gender, gender identity, religious affiliation, age, disability, or serious disease." (Twitter, 2021) and toxicity and hetaerism are often described as communication that disparages or denigrates a group based on specific characteristics such as physical qualities, faith, race, ethnic background, sexual orientation, gender, or other criteria. It can happen in a variety of ways, even subtly, or when amusement is used.

2.4.Critical review of previous studies

Gender studies research (Thelwall & Mike, 2008) was previously restricted to analysing traditional datasets, but with improved capacities to acquire and interpret geo-location information, the discipline of spatial analysis has grown (Reed. & Khan, 2016). Mobile phone datasets, in addition to social media data, have been utilized to study human activity behaviours and individual mobility patterns using AI (Kung, 2014) (Melodia, et al., 2014). However, mobile phone data sets are not seen to be a viable option for studying human movement patterns. Social media statistics are gathered and exploited to bypass the limitations of cell phone data. Log files from smart devices and websites, social media data, and geotagged audio, video, and graphics data are all examples of social media data sources (Hesse, et al., 2015). Several studies have been undertaken to investigate check-in behaviour from various perceptions, including privacy (Benson, et al., 2015) (Strater & Lipford, 2007), gender disparities (Stefanone & Huang, 2011), and geographical distances (Boyd, 2007). (Benevenuto & Rodrigues, 2009) presented an AI analysis of user workloads in online social networks to investigate prospects for better interface design, fuller studies of social interactions, and enhanced content distribution system design. The findings indicate important aspects of social network workloads, frequency, and duration, as well as the types and sequences of activities that individuals engage in on the online social network. As per (Scellato, et al., 2011) Using AI Modules, we investigated the socio-spatial features of distinct social network user behaviour platforms in order to examine check-in behaviour and the types of places people frequent.

(Noulas & Scellato, 2011) Check-in AI patterns of foursquare users were analysed to explore check-in behaviour and mobility trends inside a city. (Ruggles, 2014) used social network user behaviour data to investigate place-to-health correlations in order to improve public health prospects. (Chorley, et al., 2015) Created a web-based interactive program that evaluates the personality characteristics and check-in behaviour of foursquare users, claiming that personality features assist to explain individual variances in social network user behaviour, use, and the types of areas visited. Maia. (Maia & Almeida, 2008) Proposed a methodology for defining and AI-based detecting user behaviours in online social networks and presented a clustering algorithm to group individuals in the social network that have a similar behavioural pattern. According to (Pucci, et al., 2015) presented an analysis of the large-scale event with the goal of observing inconsistencies in urban spaces and formulating policies in accordance with molecular practises and arising mandates can be created by diverse communities using the city and its services at ranging rhythms and intensities. While Hong (Hong, 2015) investigated user involvement to give insight into Seoul city and analysed Foursquare social media data. Furthermore, observed venues were chosen based on user engagement and the features of the city, whereas (Jin, et al., 2013) presented a survey to provide a complete evaluation of the state-of-the-art research linked to user behaviour in social networks from many viewpoints. i.e., Social network members' social connectedness and involvement, as well as their social and malevolent behaviours. Despite the fact that (Gyarmati & Trinh, 2010) published a large-scale AI measurement study of user behaviour in social networks and developed a measuring framework to monitor user activity, characterisation of user activities, and use patterns in social networks. Many studies of development and prediction have used social network user behaviour datasets. for example, (Preoţiuc & Cohn, 2013) investigated the locations of users with a focus on the kind of places and their change through time and discovered patterns for venue category usage across different temporal scales. (Shen & Karimi, 2016) proposed a framework for characterizing urban streets by conceptualizing visual paths, claiming that the use of ubiquitous big social media data can enrich the current description of the urban network system and improve predictability of network accessibility on socioeconomic performance (Wu, et al., 2016). The use of social network user behaviour data to observe purchasers' willingness to pay for various aspects was highlighted. Individuals' views and geographical preferences for areas can be expressed by visit frequencies, which are assigned distinct reasons. (Luarn & Yang, 2015) Created and improved a conceptual framework to give a theoretical explanation of the motives that lead consumers to participate in check-in behaviour. According to the study, social factors (such as tie strength, subjective norms, expressiveness, social support, and information sharing) are the most important in encouraging people to engage in check-in behaviour. While (Wang & Stefanone, 2013) investigated how personality factors influence self-disclosure, which in turn influences the intensity of Facebook check-ins, they also emphasized the physical and informational mobility of users by linking individual actions to locations.

A significant amount of research (Smith, 2018) has been conducted in recent years to examine the user demographics of social network site users and identify a few distinct variables that drive a male or female user to utilize a social media network (Pentina, et al., 2016). According to these findings, men and women have distinct incentives for accessing social media networks.

(Smith, 2018) discovered that women are more likely than males to use social media to communicate with their family. While (Muscanell & Guadagno, 2012) discovered that males used social media networks to form new connections, women used it more for relationship maintenance. Gender differences in AI usage behaviours and motivation for accessing social media networks appear to exist. According to research on social media networks, male users' active engagement is positively influenced by motives for information seeking, entertainment, and self-expression, whereas female users' effective involvement is positively influenced by motivations for socializing and entertainment (Chun, 2012). Much research, on the other hand, have employed social network user behaviour data to investigate gender variations in check-in behaviour. For example, (Blumenstock & Gillick, 2010) examined data from Rwanda to determine population density and cell phone usage behaviour on social media by gender. (Rizwan, et al., 2018) investigated the check-in behaviour of Chinese microblog Sina Weibo and identified gender differences in frequency of usage over time. Further, (Rizwan & Mahmood, 2017) It was investigated how check-in behaviour varied over the same weeks but in various years. Furthermore, in terms of time and place movement patterns and behaviours in Shanghai. (Lei & Zhang, 2018) Weibos' location data were used to analyse the human dynamics of spatial-temporal gender inequalities and check-in behaviour in Beijing's Olympic Village, and female users dominated male users in social media activity. (Zheng & Zhang, 2009) Developed a method for mining the association between places from a vast number of people's location histories using social media data and (Comito & Falcone, 2016)A unique approach for extracting and analysing time- and geo-references linked with social data was presented in order to mine information about human dynamics and behaviours in the urban setting. Previous study on social network user behaviour (Scellato, et al., 2011) looked at check-in data to predict users' (Roche, 2014) location and movement patterns. Many studies have now used social network user behaviour datasets for industrialization and its environmental effects (Cui & Shi, 2012), development and prediction (Backstrom, et al., 2010) travel and activity patterns (Gu & Zhang, 2016), emergency response (Cervone, et al., 2016), and urban sustainability. This line of research is useful in comprehending disparities in check-in behaviour by gender, but the present study did not consider the connection with other indicators of gender equality (i.e., equal access to education, equal access to economic resources, and end of violence) in a society, which is beyond the scope of the current study (Dorchai & Meulders, 2009).

2.5.Similar Data Pre-processing and Datasets

Datasets are collections of cases with a common trait, such as texts or Numerical, Categorical, Correlation, etc... that must be cleaned and filtered to meet the model's requirements, and this process is known as Pre-processing. These pre-processed datasets are fed to the AI model, and the model begins to learn, and the learning model is then trained with successive datasets. The more data fed into the model, the faster it can learn and improve, so selecting the proper dataset is critical, and it's a time-consuming and difficult operation, particularly when gathering datasets connected to terrorism, hetaerism and toxicity. The datasets that are available, as well as the analytical and pre-processing approach that researchers utilized to obtain data for model training. (Founta, et al., 2018) (Founta, et al., 2018) constructed two datasets, D1 and D2, with 300 tweet data in D1 and 80,000 tweet data in D2. They used the Twitter stream API to acquire 32 million tweets at random and afterward screened and filtered the data by deleting spam and tweets with short text contents that were not written in English. and this process was done with text analysis and machine learning (Founta, et al., 2018). In Ziqi Zhang and Lei Luo's study, they used openly accessible Twitter datasets with unbalanced characteristics and unfiltered data containing hate and non-hate data, as well as 7 preceding Twitter research datasets based on the premise of screening the public Twitter stream for keywords and hashtags such as refugee and Muslim, sexism, racism, and hate. And, in order to process and prepare the data for the machine, they had to remove the noise and colloquial nature of the data. To do this, they used a tweet normalization tool, the primary goal of which was to reduce noise caused by the data's vernacular character, for example, misspelling correction and lengthened word normalization are part of the procedure ("hurrrrrayyy" becomes "hurray"), Untangling abbreviations (example, 'won Then, lemmatize each word to return it to its lexical form. As a result, pre-processing and dataset preparation are required. Zhang and Luo (2018) (Zhang & Luo, 2018).

Furthermore, in their research, Ashwin Geet d'Sa, Irina Illina, and Dominique Fohr used a Twitter dataset in which tweets were retrieved based on hatebase.org lexical phrases. In this data collection, they used the CrowdFlower platform to undertake annotations on 24883 tweets. Despite the uneven data set, the annotated labels corresponded to three classifications: terrorism, hetaerism, toxicity, offensive language, and neither. The researchers pre-processed the data using a 5-fold cross-validation procedure, BERT, and FastText, where they removed special characters other than "!", "?", "", and usernames beginning with "@", the "retweet" symbol to avoid processing the same data over and over again, and they removed hashtags and split the words into certain lexicons (Example "burnitalldown" to burn it all down), to make the dataset more reliable and understand.

Second, 70% of the dataset was used for training, 20% for validation and testing, and 10% for design and implementation. The design and implementation sets are used to fine-tune the hyperparameters. The testing set is used to assess the model endpoint's efficacy (Geet d'Sa, et al., 2021). A overview of Sreelakshmi, Premjith, and Soman's research is also given; they used 10000 English and Hindi literature for their study, which was divided into two categories: hate and non-hate. This dataset is made up of tweets collected from the three sources listed below. The data was obtained via Twitter API and annotated by two linguists who are fluent in both Hindi and English; the second data set is from the research (Mathur, et al., 2018), and the third set is from the HASOC shared effort. To pre-process all of this raw data, they utilized a python regex function to remove URLs, Usernames, Hashtags, emojis, special characters, white spaces, and convert all words to lower case for the model's simplicity and handling (Sreelakshmi, et al., 2020) In Diego Benito Sanchez's study, data was gathered from Twitter and hand-annotated utilizing crowdsourcing services before being separated into two groups, each including tweets about hostility toward women and refugees in English and Spanish whilst these datasets were divided into training, development, and testing, so the researcher had to pre-process the data to get all the data fine-grained for the training and development process as certain terms and syllables like stop words where it's usually formed by propositions like "a", "an", "the", "and", "or", "but", "in", "my", etc. do not offer any value; As a result, these words must be annotated so that the training and classifier models can ignore them, with the exception of punctual markers, URLs, and cap words, which can be processed by a tool called GSITK, and the researcher also mentions how ignoring these would result in noise introduction and complexity addition, affecting the model's performance (Sanchez, 2019). The researchers of this study, Marzieh Mozafari, Reza Farahbakhsh, and Nol Crespi, experimented using three publically released datasets based and analysed and obtained on Twitter supplied by Waseem and Hovy, Waseem, and Davidson, which are accompanied in the following: Waseem and Hovy acquired 136,052 messages from Twitter and labeled around 16.9k of them as "Racism," "Sexism," or "Neither" after some filtering. They began by looking for common slurs. and terms related to religions, sexual, gender identity, and ethnicity using an ad hoc process. Second, based on the preliminary findings, they identified the most prevalent terms in tweets containing hetaerism. For example, the hashtag "#MKR" was connected with My Kitchen Rules, a famous Australian television show, and led in a deluge of unpleasant remarks directed towards the female participants. As a negative sampling, in order to make their sample approach more general, they crawled additional tweets that had clearly abusive phrases and psychologically injurious terms but were not objectionable in context but rather by words alone. To eliminate annotator bias, the final obtained dataset (16K) was examined by a 25-year-old female pursuing gender studies and a non-activist feminist. Waseem also provided a supplementary dataset sample to examine the influence of expert and amateur annotators on the overall performance of the hetaerism classifier. As a conclusion, CrowdFlower users and experts with experience and comprehension of disparaging language as well as hetaerism gathered 6,909 tweets for hetaerism and categorised them as "Racism," "Sexism," "Neither," and "Both". Their efforts generated a total of 4,033 tweets, with 2,876 tweets overlapping between their new dataset and Waseem and Hovy's. The present researchers combined the two datasets in the same way as Waseem did to increase the data imbalance since the two datasets considerably coincided in terms of recognizing hostile material.(Waseem & Hovy, 2016). Davidson, on the other hand, scanned 84.4 million tweets from 33,458 people on Twitter using a list of specific keywords from Hate Base, a pre-set and pre-built vocabulary of hateful speech words and phrases. They chose 25k tweets at random and asked CrowdFlower platform users to label them as "Hate," "Offensive," or "Neither." If the annotators' agreement was low after categorizing each tweet, the tweet was removed from the sampled data. This dataset is referred to as the Davidson-dataset throughout the rest of the research. (Mozofari, et al., 2020) Finally, Thai Binh Nguyen, Quang Minh Nguyen, Thu Hien Nguyen, Ngoc Phuong Pham, Quoc Truong Do, and The Loc Nguyen are working on a Vietnamese-based local project called "VAIS Hate Speech Detection System: A Deep Learning-based Approach for System Combination" they used data and texts (teen slangs, nontonal text such as music, not having or based in a particular key, emojis,) based on the consumers' preferences, scrapped from multiple sources with multiple text encodes to make training easier, and they unified all types of encoding by the end of their crawling process, they conducted a unified cleaning process which had two stages as follows: cleansing The following are the steps they followed to process the data.

- Phase 1: Because there are so many varieties and intonations in Vietnamese, the first step is to arrange data encoding. Because different Unicode typing tools may produce different answers for the same typing type, the researchers constructed a dataset library called visen1 to bring it all together. For example, the input "thêt kê'" will be standardized to "thit k" as the output.

- Phases 2: Emojis are frequently used to communicate emotions on social media. Emoticon is often a Unicode character, although it may also be composed of numerous normal characters, such as ': (=]'. To unify it, they developed a vocabulary that translates the needed emoji (combined with specific symbols) to a single Unicode character, similar to other emojis.
- Phase 3: Remove any characters who aren't visible. Unobservable characters are unseen to people, but they complicate the model by introducing space between words, punctuation, and emoji. This phase aims to reduce the number of terms in the dictionary, which is a vital job when working with tiny datasets like the HSD challenge.
- Phase 4: For a model that requires Vietnamese word segmentation as input, they tokenized the input text using Subword BPE, Space base, and Vietnamese word base.
- Phase 5: Finally, all strings were taken down. They tested text representors such as CBOW, Roberta, Fasttext, and aSonVX and found that lower-case or upper-case has no influence on the output, despite the fact that lower characters reduce the quantity of words and data in their Vietnamese vocabulary. (Nguyen, et al., 2019).

2.6. Similar Systems / Algorithms and Models used in Terrorism, Toxicity and Hetaerism Classification

HaterNet can recognize and identify hetaerism in Twitter data, as well as disliking behaviors and other negative laterals. HaterNet is a smart architecture that analyzes and evaluates the spread of hetaerism on Twitter, and it is now being used for safety by the Office Against Hate Crimes of the Spanish State Secretariat. HaterNet categorizes approaches based on different collection representations of methodologies and content order models, and the most efficient and effective model combines an LTSM+MLP neural model with tweet assertion, emoji, and expression embeddings incorporated via TF-IDF components to produce the best possible results; additionally, the primary and hidden layers of the classification model were created using the word2vec algorithm, which is neurally based. Each single word is associated with a specific series of integers known as a vector by word2vec.

The vectors are designed in such a way that a simple mathematical formula (cosine resemblance in between vectors) may be used to examine the amount of similarity measurements in between words provided by such vectors (Pereira-Kohatsu, et al., 2019). The researchers proposed methodology in this project is to employ autonomous terrorism, hetaerism and toxicity analysis based on numerical interpretations of text and classification models applied to these numerical representations. To go further, their working model had two techniques as follows: in the feature-based technique, each data/text/comment is represented as a chain of words or a sequence of words dependent on one another, and a tokenization is computed for each syllable utilizing BERT and fastText in the fine-tuning technique. Finally, once the data has been tokenized, it is passed on to a classifier, which can be a supervised training model such as a Support Vector Machine or a deep neural network such as CNN, Bi-LSTM, and CRNN models, and the specific classifier which takes final prediction, and the second approach fine-tuning is done in a single movement. Each comment/text/data is classified using a fine-tuned BERT model. Each remark is classified as offensive, terrorism, hetaerism and toxicity, or both.

- fastText model: To produce a different tokenization for each syllable in a dataset, fastText's bag of character n-grams model will tokenize every word in a comment, including unique syllables.

- BERT model: The dataset's words are tokenized word by word before being put into a previously trained BERT model. The BERT algorithm contextualizes the word fragments.

The retrieved features from the fastText and BERT methods are then put into the Support vector machine technique and deep neural network classifiers for final classification (Geet d'Sa, et al., 2021).

"Design and Development of a terrorism, hetaerism and toxicity detector on Social Media Using Deep Learning Technologies," by Diego Benito Sanchez. Once the dataset has been pre-processed with GSITK, the researcher ensures that the detection model passes the dataset through various classification aspects such as content analysis, word embeddings such as Bag of Words, TF-IDF to encode syllables that are semantically comparable with comparable vectors, Linguistic Features, and finally, the ultimate classification, where the processed dataset is passed through machine learning classifiers such as Random Forest, Logistic Regression, and others (Sanchez, 2019).

Furthermore, a brief summary of (Sreelakshmi, et al., 2020)s' Detection of hetaerism and toxicity detection in Hindi-English Code-mixed Text is provided. Data work, their technique was quite unique compared to the rest of the others because they were using a combination of two languages to classify terrorism, hetaerism and toxicity, and to pull this off, the researchers had first attempted to understand the entire semantic of the sentence as it makes the whole task easier, so this made them trial test with three aspects of algorithms and classification methods as follows:

Trial 1: For categorization, this uses the doc2vec feature (an NLP tool for modelling documents as vectors), which converts the complete sentence into a vector illustration and can record both the phrase meaning and the word as well as word ordering, further it uses Continuous Bag of Words Model to supervise the vectors of doc2vec with a tensor length of 300, a scale factor of 5, and a least measure of 1 before passing the vector through SVM and Random Forest. However, this method only produced an average accuracy of 64%, so they went on to the next phase.

Trial 2: Because the retrieved features and vectors created by doc2vec were insufficient, the researchers used phrase fragments as attributes to achieve syllable-level tokenization. For character extraction, the researchers used sector tailored vectors, and to identify these features, the entire dataset with snippers was passed through the word2vec technique, which also identified distinctive words and created word vectors of them of a predefined length, and finally the processed dataset was fed into Classification using classifier models such as Support vector machines and Random Forest, with an accuracy of 75%. As a consequence, the researchers discovered that word2vec outperformed doc2vec, and they opted to apply the vector representation of the features to a specified length and segment of the phrase since word2vec works well.

Trial 3: Because the preceding two tests failed to recognize and handle out of vocabulary terms (Hindi), the bulk of NLP investigations use pre-trained models such as FastText, which were learnt through rigorous training over a huge corpus. With the exception of word2vec, FastText evaluates each text as a piece of plot points to the lowest n-gram units, offering the model with a deeper understanding of the word. As a result, the researchers ran the dataset through FastText, which converted the labels and text formats to aggregated word vectors, parsed each text and appended the tensor representation for each word in a sentence, and then fed the data to Support Vector Machine and Random Forest for classification, yielding an accuracy of 85 percent for Hindi and English (Sreelakshmi, et al., 2020). Another example is the VAIS Hetaerism and Toxicity Detection System, in which the researchers created the Models structure by combining five various model architectures and multiple different types of CNN and RNN, as shown below.

- TextCNN: It is entirely made out of CNN blocks, with some Congested levels tossed in for good measure. The output of several CNN nodes with varied kernel size is linked to one another (Gong & Ji, 2018).
- VDCNN: Similar to the TextCNN design, it contains many Convolutional Neural Network units. This model's convolution layers are a distinguishing feature (permits gradients to pass straight across a network, bypassing non-linear convolution layers).
- Bi-LSTM: In this model, many LSTM bidirectional blocks are stacked on top of each other.
- LSTMCNN: The Bi-LSTM block will modify a sequence of syllable vectors before running them through CNN nodes.

- SARNN: Rather of storing the data series in a single conventional context tensor, it creates a vector representation that is precisely adjusted for each output sampling period.

To get the greatest overall result, each model leverages different word embedding approaches, such as CBOW, RoBERTa architecture, and FastText, or understands sentence vectorization directly (Nguyen, et al., 2019). Finally, according to the BERT model, hetaerism, toxicity, and race prejudice can be reduced through social media. The researchers built two BERT-based modules to discover hetaerism and toxicity identification and bias mitigation, which was made possible by their Model architecture and the techniques used, based on the pre-processed datasets weights ("lesbian", "gay", "Islam", "feminist", etc. weight words are passed through the mitigation module) and attributes the data is either pass through bias mitigation module or AI module. Furthermore, in order for the model to identify whether the word belongs to the mitigation model or the terrorism, hetaerism and toxicity model, various encoders were utilized to select and create the dataset's input sequences, and the BERT Framework was used to train the model. The researchers used tools available in the BERT open repository, such as text tokenizers and pre-trained Word Piece, to eliminate invalid and special characters from words and convert them to lowercased strings as the first step, and then as the second step, the researchers used the Word Piece tool to break down and subsequent words into sub words. The maximum series length was set to 64 throughout this breakdown process, and any greater than that will be padded with null data or truncated accordingly. Last but not least, researchers used the Google Collaboratory tool to fine-tune their classifiers by passing the categorized and pre-processed data via BERT + Non-Linear layer, BERT + CNN, and BERT+LSTM Models to train, test, and evaluate (Guo, et al., 2020).

2.7.Problem definition

Despite this, the solution has covered numerous techniques and ways to tackling the problem, as well as comparable solutions presented. The time constraint necessitated more detailed evaluation than what was provided in the literature study. This is due to the fact that slang and literal meanings of sentences do not always have an aggressive connotation. This might easily seem elegant and not offensive enough. However, the overall tone of such a book would almost certainly be harsh. Existing deep learning systems, such as LSTM, SVM, FastText, and numerous other algorithms described above, do not provide as much contextual weighting as the BERT algorithm. The BERT algorithm is powerful enough to recognize the context of a text even when a word has many interpretations depending on the context. However, this does not imply that the algorithms detect less correctly. Because each of these algorithms has advantages and disadvantages. However, our primary purpose is to combat hetaerism, toxicity, and terrorism by measuring text aggression in the context of the entire text.

2.8.The Multi-Classification Model

2.8.1. Algorithm

BERT is the algorithm we use in our solution. BERT (Bidirectional Encoder Representations from Transformers) is a pre-training strategy that creates deep bidirectional representations from unlabelled text by using both left and right context conditioning in all layers. As a consequence, the pre-trained BERT model may be finetuned with one or more additional output layers to build state-of-the-art models for a variety of tasks, including question answering, language inference, and text categorization, without needing major task-specific architectural changes. BERT was trained on the English Wikipedia (which has 2,500 million words) and the wordsBooksCorpus (with 800 million words) (Devlin, et al., 2019) (Guo, et al., 2020).

2.8.2. Model Architecture

BERT's model architecture is a multi-layer bidirectional Transformer encoder based on the original implementation of (Vaswani, et al., 2017), which is accessible in the tensor2tensor library.BERT$_{BASE}$ and BERT$_{LARGE}$ were the original models for the English model. There are 12 encoder layers, 768 feedforward networks, and 12 attention heads in BERT$_{BASE}$. There are 24 encoder layers, 1024 feedforward networks, and 16 attention heads in BERT$_{LARGE}$ (Devlin, et al., 2019).

2.8.3. Why BERT?

Unlike directional models, which read the text input sequentially, the Transformer encoder reads the entire sequence of words at once (left-to-right or right-to-left). As a result, it is classed as bidirectional, however it is more accurate to call it non-directional. This characteristic allows the model to determine the context of a word from its surroundings (the word's left and right). (Horev, 2018)

Chapter 03 - Methodology

This chapter explains the different techniques and the systematic procedure followed by the researcher. As the core research area is relevant to AI there are several statistical techniques and some critical technical aspects. The statistics is a key branch of mathematic, which considered as the science of data. Hence data is crucial to this study. Types of data, data sets and data collections techniques and all the statistical techniques are included to this chapter. Following figure illustrate the key steps within the research effort and following section briefly explain the relevant key contents.

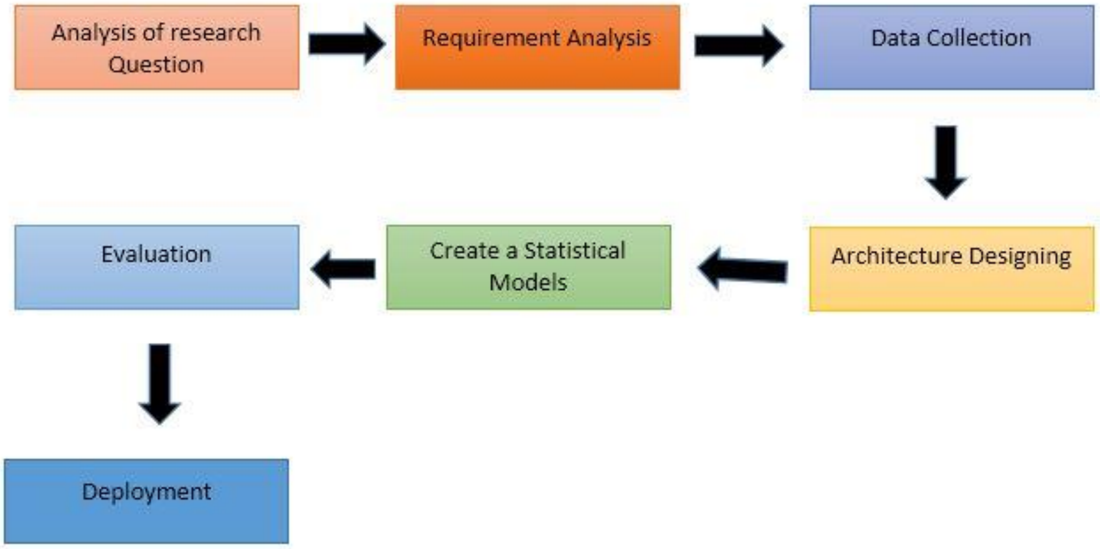


Figure 1:-Steps of Research Methodology

3.1.Design Phases of Research Methodology

3.1.1. Analysis of Research Question

This phase is used to do in-depth research on the issue statement stated in the preceding chapters. After doing an accurate analysis of the issue statement, how the Artificial Intelligence approaches should be prepared to provide the answer should be examined. It is necessary to debate which principles should be employed and what type of architecture should be built for the research.

3.1.2. Research Philosophy

The methods through which research will be done are referred to as research philosophy. It is the initial phase in the study procedure. A research philosophy is a collection of ideas and assumptions that will aid the researcher in developing knowledge about a certain issue. (Kuada, 2012). The presence of numerous phenomena that are real or thought to be true centers on the concept of research philosophies. Using a research philosophy, a researcher specifies the tactics that will be employed in the research process. The researcher's perspective and reasoning are defined by his or her research philosophy. (Kothari, 2004). It assists the researcher in understanding the various methods for collecting data. The main aspects of research philosophy are the selection of a research technique, the analysis of data, and the application of the information obtained. There are several kinds of research philosophies such as Positivism, Realism, Interpretivism and Pragmatism. The research philosophy chosen is determined by the research issue or hypothesis being developed (Welman, et al., 2006).

The researcher adopted positivist philosophy for this investigation. Because this study focuses on artificial intelligence utilized for social media analysis, as well as recognized hetaerism, toxicity, and terrorist behaviour on social media in Sri Lanka. (Chetty & Priya, 2016). Because the information linked with this issue is sometimes contradictory to the recognized knowledge of ancient social media analysis, the researcher prioritized positivist philosophy. During this study, the researcher's activity is confined to data collection and interpretation in an objective manner, and research outcomes are visible and quantifiable, which are also essential possibilities for positivist investigations. This research study is entirely dependent on the facts of social media users from various online platforms; this fact makes it clear that positivism is the most effective philosophy for this investigation. (Badke , 2011).

3.1.3. Research Approach and Technique

The analysis of the research issue provides a broad arena for understanding the research strategy. The researcher used a deductive research strategy for the study based on the circumstances of the topic. Because this study takes a top-down method, where the research was important to hypothesis testing and subsequently decision making. In the deductive technique, it is also possible to get a study result by designing a framework based on factors found from previous research studies. As a result, researchers employ a deductive research technique, progressing from existing ideas through study framework and hypothesis testing. As a result, the premise for this study, which is based on artificial intelligence analysis of online spaces in Sri Lanka, is based on current ideas and findings. The goal of using a deductive technique is to reach conclusions about questions inside the study that the researcher feels can be answered. Furthermore, the deductive approach's thought process progresses from theory to research question, data collection, findings, and rejection or confirmation of the research question. Another research method is known as the inductive approach. The inductive method is a bottom-up strategy. Thus, in the particular technique, the researcher must conduct observations and a detailed examination of the data set before identifying the hypothesis based on the patterns in the data. Finally, the hypothesis may be tested and the findings generalized into a theory. This method is more difficult and time-consuming. Due to time constraints and the availability of excellent literature sources, this study employs a deductive research technique.

Another stage in research technique is research strategy. This phase includes the technique for guiding the research investigation. There are a few approaches that the investigators can take. These are survey methods, case studies, experimental methods, action research, and grounded hypotheses. The strategy to leading the expedition is determined by the selection of these tactics. This investigation tried the overview of obtaining information by survey technique, as this research is produced using positivist philosophy, logical methodology, and included of quantitative data obtained from social media users of Kandy Sri Lanka. A survey was carried out in order to collect relevant information from the accessible data sources. The surveys generate quantitative data that may be analyzed empirically. Surveys are most commonly used to investigate causal factors between different types of data. Particularly from the statistics reports issued by numerous institutions and research groups. Thus, the study was based on secondary data, because the primary data required the researcher to obtain new data that would be used for the first time in his study. This information obtained from the study secondary data collection provides the end-product of the study to achieve the study's pre-specified generic and particular objectives.

3.1.4. Research Choice

Choices of a research comprises of three fundamental sorts like mono technique, multi strategy and the blended technique. Mono strategy for the examination the utilizing of just subjective/ qualitative or the quantitative data and data analysis techniques. The blended strategy is a technique where the utilization of both the techniques for the investigation as well as the translation of the outcomes. In the multi technique the utilization of both the sorts of strategies subjective and the quantitative should be visible. In this research the analyst has utilized the mono technique since the examination is falling under the quantitative strategy and information analyses are performed using mono methods of assessment, so the mono method was the most appropriate option for this study. But the study consists with different statistical techniques which comes under quantitative research.

3.1.5. Time horizon

Time horizon is important for the data collection process as well as for the data analysis techniques. There are two basic categories in time horizon as, cross sectional and the longitudinal. In the cross-sectional examination the exploration is directing in the short-term time frame span or single point of time while in the longitudinal investigates it very well may be done in the long terms of the review. Both the quantitative and subjective information can be utilized in the cross-sectional technique and can notice the way of behaving of the data items. The longitudinal exploration likewise permits to utilize both the subjective and the quantitative information which permits to concentrate on an engaged occasion for a long haul. Because there are just a few months to complete this investigation. As a result, the cross-sectional time horizon over a short period of time is the optimum choice for this research project. Further temporal horizon study allows the researcher to identify multiple populations in the research at a single point in time, allowing the researcher to identify distinct factors at the same time. These research surveys are also less expensive. All of the data is gathered in a short period of time.

3.1.6. Conceptual Framework

A conceptual framework is a graphical representation that aids in the illustration of the predicted link between cause and effect in a financial environment. It is often referred to as a Conceptual Model or a research model. The model includes many variables and the expected connections between those variables, which reflect the expectations.

All data was obtained using a basic random sample procedure, and a questionnaire was employed as a data collecting tool. The questionnaire includes demographic information as well as questions on independent and dependent factors.

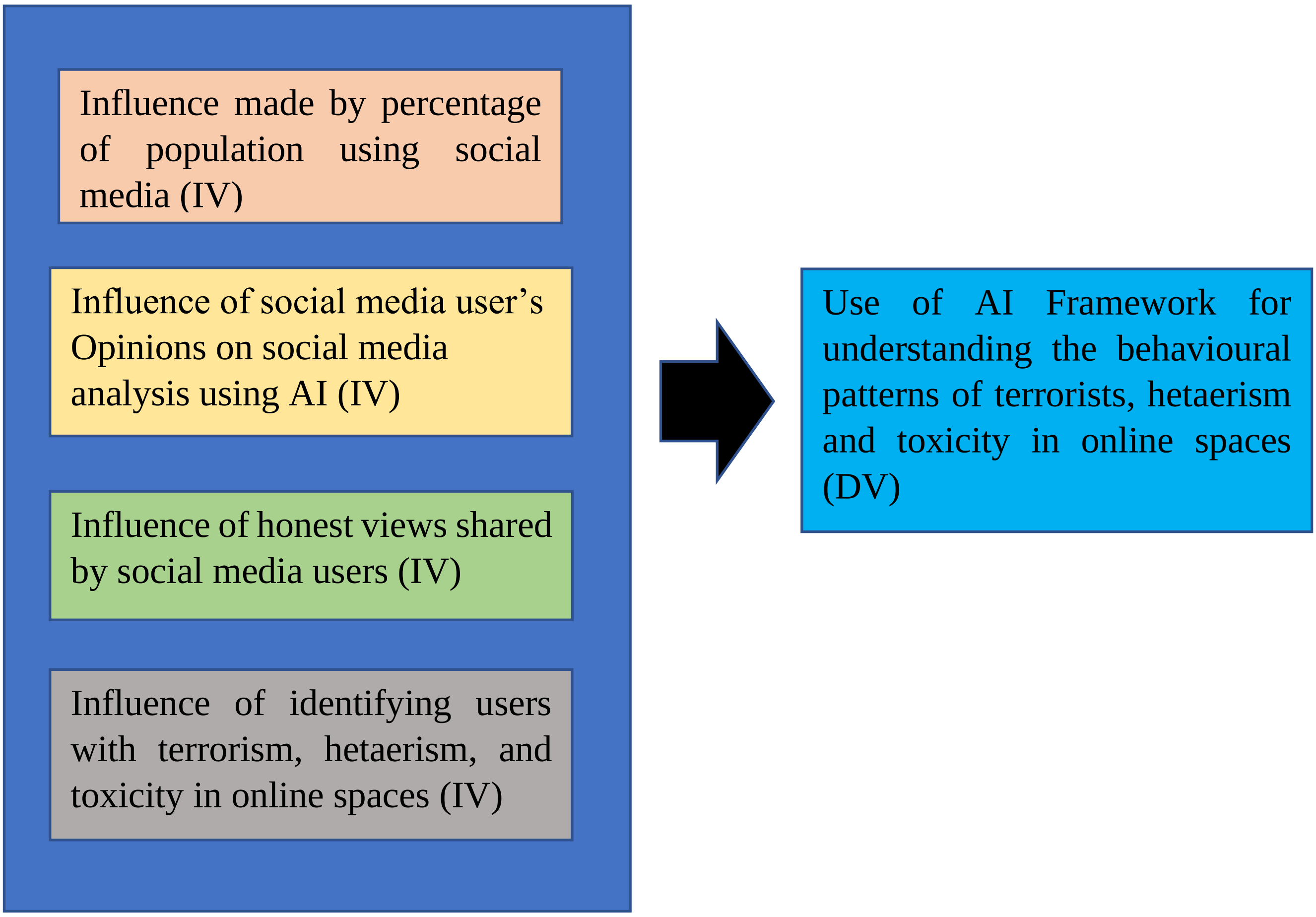

Influence made by percentage of population using social media (IV)
Influence of social media user's Opinions on social media analysis using AI (IV)
Influence of honest views shared by social media users (IV)
Influence of identifying users with terrorism, hetaerism, and toxicity in online spaces (IV)
Use of AI Framework for understanding the behavioural patterns of terrorists, hetaerism and toxicity in online spaces (DV)

Figure 2 :- Conceptual Framework of the research

3.2.Hypothesis Development

A hypothesis is an tentative statement regarding the link between two or more variables. It is a particular, testable prediction about what researchers anticipate will occur in a study (Shuttleworth & Martyn, 2019). A hypothesis in the scientific method indicates what the researchers believe will happen in an experiment, whether it is in psychology, biology, or another field (Cherry & Kendra, 2019)

3.2.1. Hypothesis of the research

- Hypothesis 1

  H_1 – AI based framework understand the behavioral patterns of terrorists, hetaerism and toxicity in online spaces

  H_0 - AI based framework does not understand the behavioral patterns of terrorists, hetaerism and toxicity in online spaces

- Hypothesis 2

  H_1 - social media is used by a significant percentage of the population.

  H_0 - social media is not used by a significant percentage of the population.

- Hypothesis 3

  H_1 - Social media users do share their honest opinions.

  H_0 - Social media users do not share their honest opinions

- Hypothesis 4

  H_1 - Identifying and filtering terrorism, toxicity and hetaerism views and opinions based on an AI tool is possible.

  H_0 - Identifying and filtering terrorism, toxicity and hetaerism views and opinions based on an AI tool is not possible.

- Hypothesis 5

  H_1 - General Public is comfortable of using an AI model to analyse social media.

H_0 - General Public is not comfortable of using an AI model to analyse social media.

3.3.Requirement Analysis

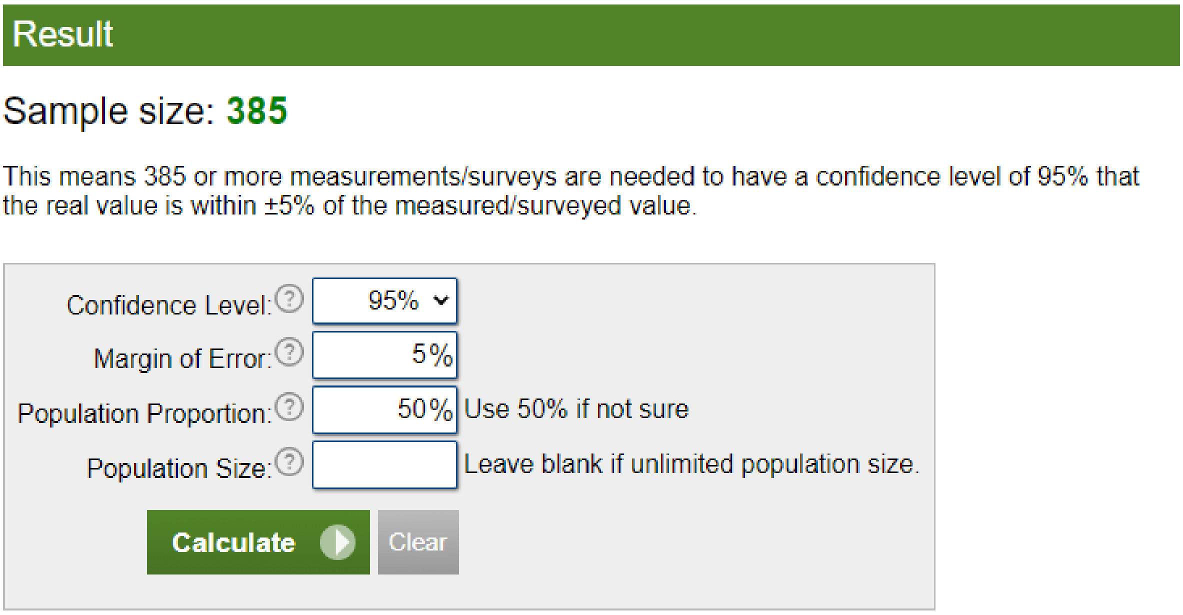
Following the analysis of the research questions, the appropriate data source or data collection should be identified within the project scope. This research should be conducted in a Sri Lankan context focusing on the Kandy area, hence there should be an exact dataset accessible that includes Sri Lankan slang terminology as well as Sinhala language utilized in English (Singlish). This can be more complicated at times since there is no pre-created data set for Slang Language and Singlish Processing in Sri Lankan. Soft copies of data relating to the scoped are extremely rare. As a result, additional effort should be made to locate an accurate and reliable dataset for the project.

3.4.Data Collection

The study design is carried farther into the procedures of data collecting and analysis with this research layer. Researchers select data gathering and analysis procedures appropriate for this study based on previous decisions made at various phases of research. Inorder to satisfy the research objective before the implementation primary data was collected based on survey method by distributing questionnaires among the selected population who uses social media in Kandy region. The above findings will support the addressing research objectives n the process of determining the solution for the problem domain.

3.4.1. Population and Sampling

According to (Kemp, 2022) the 52.6% percent of Sri Lankan population are internet users. The general population in Kandy region is around 1,501,000 (Brinkhoff, 2021). Thus, based on the Kandy population internet users the sample population generated based on a sample size calculator as per below figure.



Result

Sample size: 385

This means 385 or more measurements/surveys are needed to have a confidence level of 95% that the real value is within ±5% of the measured/surveyed value.

Confidence Level: 95%

Margin of Error: 5%

Population Proportion: 50% Use 50% if not sure

Population Size: Leave blank if unlimited population size.

Figure 3 :- Sample Population Generated

Despite figure 2 population size is to be considered as unlimited based on internet user in Kandy Region. As a percentage, to achieve 95% confidence and 5% error, 385 samples should be gained as per the equation for unknown population.

3.4.2. Sampling Technique

Using the random sample method, each member of the population has an equal chance of being chosen as a subject. The whole sampling procedure is completed in a single step, with each subject chosen independently of the other members of the population. There are several approaches to using basic random sampling. The lottery approach is the most rudimentary and mechanical. Each person in the population is given a unique number. Each number is placed in a bowl or hat and properly mixed. The researcher then draws numbered tags from a hat while blindfolded. The subjects for the study are all of the people who have the numbers chosen by the researcher. Another option is to use a computer to make a random pick from your population. The first approach is recommended for populations with a limited number of members, however if the population has a large number of members, computer aided random selection is favoured.

3.4.3. Data Collection and Analysis

This mainly focuses on the analysis of collected data via the questionnaire or the survey. The data is collected from the participants via google forms which is an online survey method. The collected data were entered into the Microsoft Excel and Statistical Package for Social Sciences (SPSS) software based on each variable of the study for the purpose of analysis. This Particular study also use a descriptive statistic which involves a process of transforming a collection of raw data into tables, charts, with frequency distribution and percentages, which is an important sector of data analysis. The Demographic information provides data regarding research participants and for the purpose of the determination whether the individuals in the particular study is representing the sample target population for generalization purposes and analysis of broad characteristics about groups of people and populations. Correlation analysis is a method used for the purpose of statistical evaluation used to study the strength of a relationship between two, numerically measured, continuous variables which is useful to establish relationships between variables. Regression analysis is also inherited to modelling and analysing several variables, where the relationship includes a dependent variable and one or more independent variables.

3.4.4. Reliability

The data from the questionnaire is precise and it includes the respondents' perception. The information gathered is therefore reliable. To derive the data, the researcher collected data from Online space users in the Kandy region to gather views of them on identifying the hetaerism, toxicity and terrorism in Sri Lanka by analysis of content posted over social media using AI. In order to preserve the reliability of the data collections, the research did not involve or interfere with the respondents to get any desired outcomes to the questionnaires from the respondents, instead the respondents were given the freedom and time to freely and voluntarily answer the questionnaires.

3.4.5. Validity

The research findings, Surveys and analysis of data are evaluated to reduce the likelihood if the researcher retains the ability to repeat the same response over and over again which shows the existence of adequate interrelationships and factor analysis in the results of the information. Therefore, if someone carries out the same study on researching the use of AI Framework for understanding the behavioural patterns of terrorists, hetaerism and toxicity in online spaces the similar data findings would be obtained.

3.4.6. Generalizability

The Generalizability is the utility or the service of the studying the results and interpretations in order to gain greater array of individuals or the scenario. Generalizability needs information on big populations, experimental quantitative research. The bigger the sample population the wider the findings and the perception can be. In this research, generalizability relates to the fact that if someone conducts research on the same context of, if someone conduct the same research on researching the use of AI Framework for understanding the behavioural patterns of terrorists, hetaerism and toxicity in online spaces the probability of getting similar results is likely possible.

3.4.7. Ethical issues of the research study

The researcher had to maintain the confidentiality and the integrity of the research by using valid data for the analysing process while the researcher has to seek informed consent in order to grant full knowledge about the research to the consent. The respondents’ privacy has been highly preserved to respect their confidentiality and anonymity. The participants aren’t pressured in any circumstance to gain their participation in the research and all the respondents participate voluntarily. All the Collected Data and Information will be secured within the researchers' devices in the format of Word Files, Excel Files, Google Forms, Images. The data and information collected will be analysed by using "IBM SPSS Software". This research is related the use of AI Framework for understanding the behavioural patterns of terrorists, hetaerism and toxicity in online spaces. The collected data will be stored for a time period of Six months. None of the data and Information will be shared among any external party. So, there won't be any ethical issues in proposed work since all the data and information to be gathered is subjected to equitable selection of participants, respect for the privacy, lack of unwanted pressure, unbiased presentation and avoided conflicting concerns. Thus, in order to get rid of the malfunctions and the defects aroused with the data and information analysis all the queries included in the survey are mandatory for the respondents. The gathered data and information will be stored and processed in compliance under the Data Protection Act 1998 and all other Data Protection Acts in Sri Lanka and UK.

3.5.Data Analysis and Presentation

3.5.1. Introduction to the Data Analysis and Presentation

When collecting data following demographic questions were considered

- Gender
- Age
- Occupations
- Education levels
- Marital status

Four independent variable questions from the topic domain were chosen for this study. "Percentage of population who uses social media" was chosen as the first independent variable to demonstrate the first specified purpose. To determine the second particular aim, the question "Do all social media users express honest ideas or do not?" was posed to the participants. To demonstrate the third specified goal, "Identifying users with hetaerism, terrorism and toxicity in online spaces" were requested. Finally, as the fourth independent variable, "General opinion of Social media user's view on utilizing an AI Based tool to analyse social media" were chosen to demonstrate the final particular aim. The dependent variable was designed to make the outcome of the independent variable true and to prove the research question.

The tools Bar chart and Pie chart were used to analyse demographic questions. Flowing analysis was then used to demonstrate both the dependent and independent variables.

- Descriptive analysis
- Frequencies analysis
- Histogram Chart
- Correlation (Bivariate)
- Linear Regression Analysis

3.5.2. Field Survey

Questionnaires were distributed to social media users in the Kandy region to collect data as per the sample population 385. Users of the target group were informed about the necessity of this study and the importance of information received from them on conducting this survey. In order to achieve maximum answers 400 questionnaires were delivered to users in accordance with the approach. Out of 400 questionnaires, 392 were returned, and 370 valid replies were examined following data purification, reaching a response rate of more than 100% in comparison to the sample size. The survey sheets that were collected were checked to ensure that all subpopulations were represented at or above their goal numbers.

3.5.3. Demographic Analysis

Demographic analysis is a strategy for gaining a knowledge of a population's age, gender, and racial composition and how it has evolved over time via the basic demographic processes of birth, death, and migration. Demographic Analysis (abbreviated as DA) also refers to a specific set of procedures for creating national population estimates by age, gender, and race from administrative information, which will be used to evaluate the quality of the decennial census.

3.5.4. Participants Gender Analysis

Gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	male	137	34.9	37.0	37.0
	female	233	59.4	63.0	100.0
	Total	370	94.4	100.0	
Missing	System	22	5.6		
Total		392	100.0		

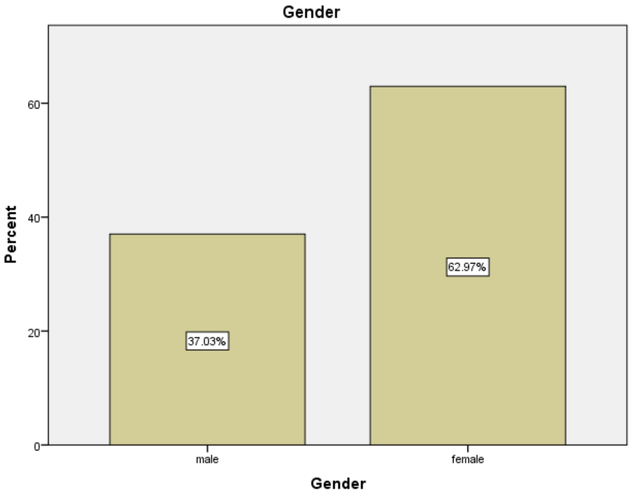


Figure 4:- Participants Gender Analysis using bar Chart

The above bar chart which shows the percentages of the gender of people who used social media as a habit. According to the chart 62.97% were female and 37.03% were Male among total of 389 participants. Hence it confirms that majority of people who used the social media are females.

3.5.5. Participants Age Analysis

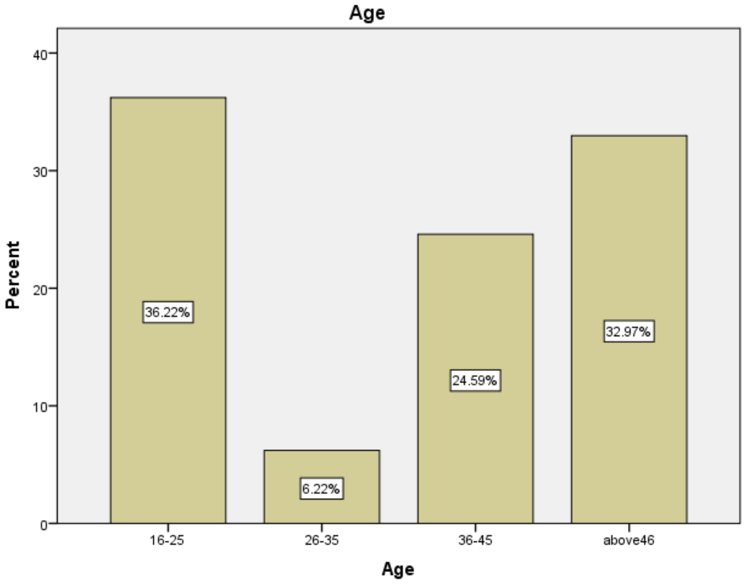


Figure 5:- Participants Age Analysis using Bar Chart

According to the above bar chart majority with 36.22% participants were in between 16 -25 age, 32.97% participants were above age 46, whereas 24.59% were between the age range 36 – 45 and the lowest as per the conduct research was 6.22% in the age range 26 - 35. When considering the overall results highest percentage of the users were in the age range 16 -25

3.5.6. Participants Occupation Analysis

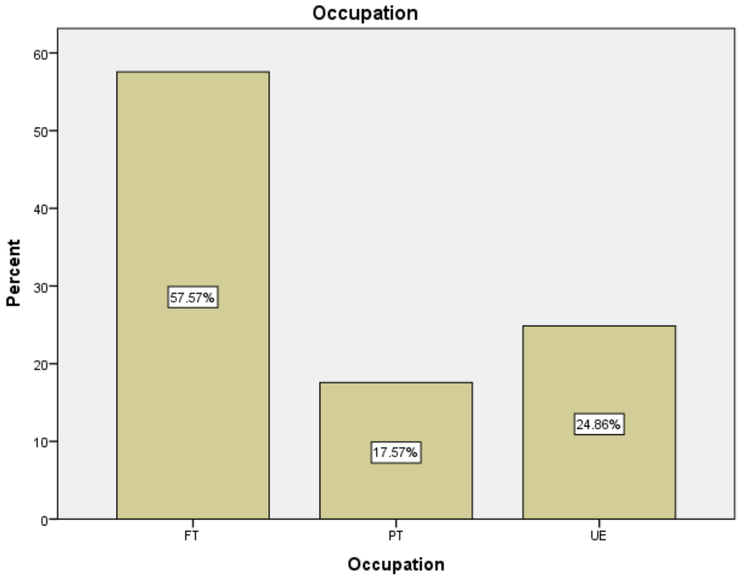


Figure 6:- Participants Occupation Analysis using bar chart

The above bar chart clearly highlights that among the people who participated into this research majority of 57.57% of people are full-time workers as well as 17.57% of people are part time workers and most importantly unemployed percentage is 24.86%.

3.5.7. Participants Education Level Analysis

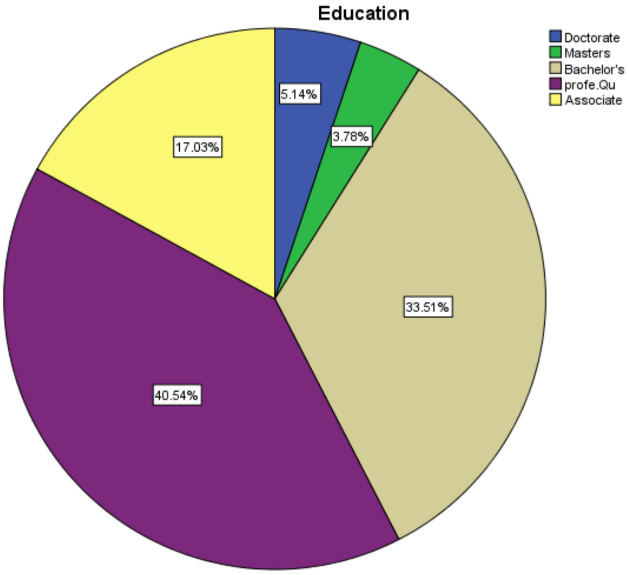


Figure 7:- Participants Education Level Analysis using Pie Chart

The figure 7 depicts the education level of the participants where the majority of 40.54% were possessing professional qualifications, 33.51% with undergraduate – bachelor's degree, 17.03% with associate degree, 5.14% with doctorates and the minimum percentage of 3.78% are with graduate- master's degree category.

3.5.8. Participants Marriage State Analysis

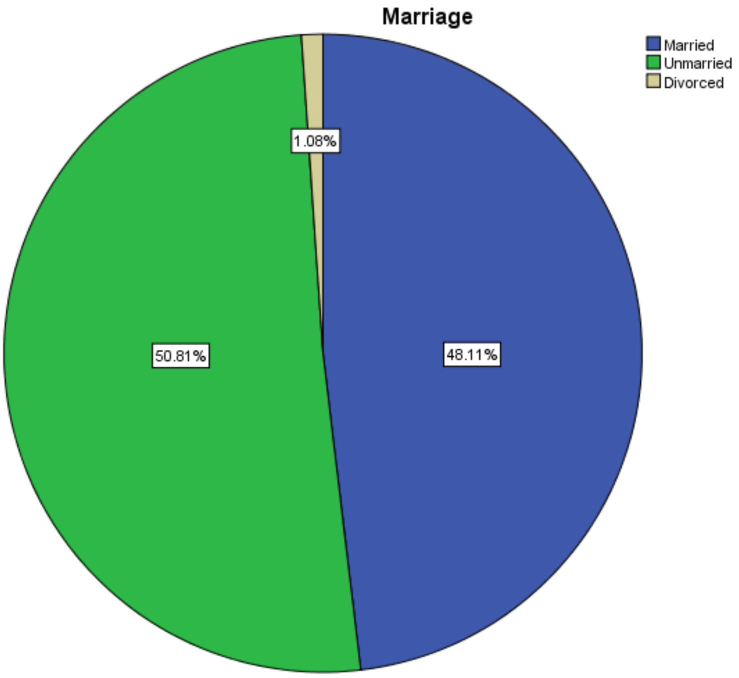


Figure 8:- Participants Marriage State Analysis using Pie Chart

The above pie chart deals with the marriage state of the participants. Mainly chart is divided into 3 categories as unmarried, married and divorced. The majority of the participant were found from unmarried state with 50.81% and 48.11% were in the married state whereas the minimum percentage was 1.08% related to divorced.

3.5.9. Participants Internet Usage Analysis

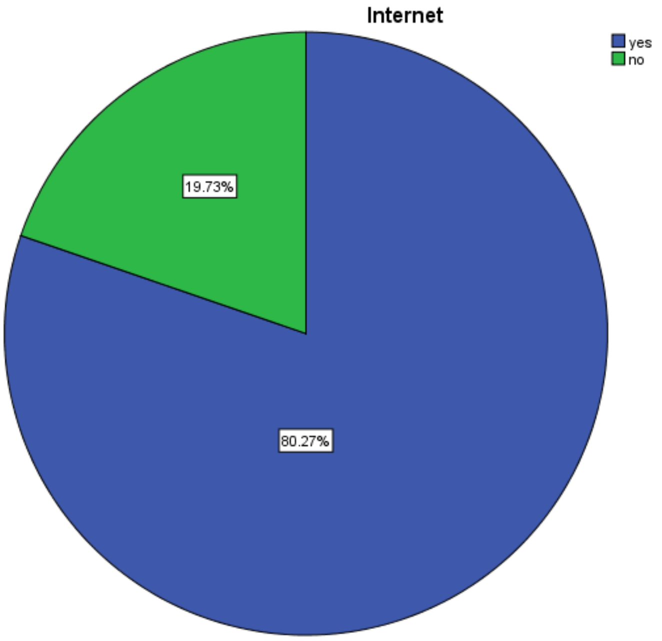


Figure 9:-Participants Internet Usage Analysis using Pie Chart

The above figure 9 depicts the internet usage percentage of the research participants. 80.27% of users tend to use an Internet Connection whereas 19.73% of users aren’t using internet.

3.5.10. Participants Device Usage Analysis for accessing Internet

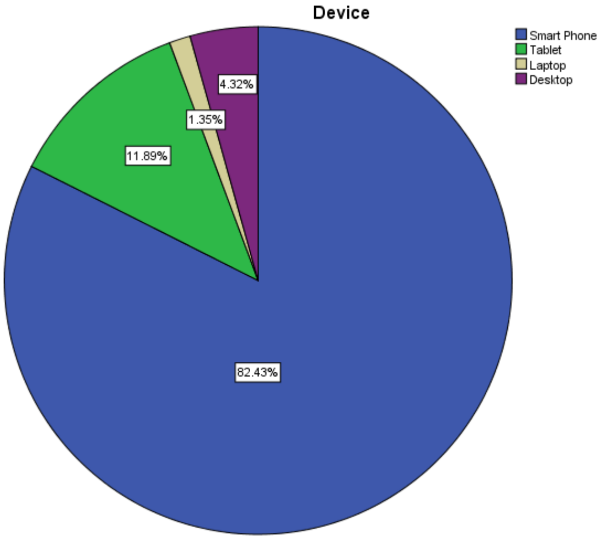


Figure 10:-Participants device (s) usage to browse internet

The above figure 10 depicts the usage of devices to browse internet. Majority 82.43% of the participants use a smart phone to access internet whereas 11.89% is designated to use a tablet.

Among all 4.32% are using desktops and 1.35% issuing Laptops.

3.6.Independent and Dependent Variable Analysis

Variable analysis is a technique that uses information analysis to reduce the features of a fundamental probability distribution. Variable evaluation concludes by reducing the population hypothesis elements of the target population test. Variable analysis is based on the statistical model used to exclude ideas from a certain statistical model.

3.6.1. Defining Variable

The study is based on four independent factors and one dependent variable, as well as demographic information. Gender, age, marriage, employment, education, and other demographic information of individuals are provided. These independent factors, as well as gender characteristics, have an impact on the usage of artificial intelligence in analyzing and filtering terrorism, toxicity and hetaerism in online spaces. Variable definitions are provided below.

Table 1 :- Variable Table

Variable Name	Description	Expected sign
Demographics	Influence of gender, age, marriage, occupation, education, usage of internet of participants	+
IV01	Influence made by percentage of population using social media	+/-
IV02	Influence of honest views shared by social media users	+/-
IV03	Influence of identifying and filtering terrorism, toxicity and hetaerism views and opinions on online spaces.	+/-
IV04	Influence or agreement of users to allow an AI Model to analyse online spaces	+/-

DV	Use of AI Framework for understanding the behavioural patterns of terrorists, hetaerism and toxicity in online spaces	+

3.6.2. Reliability analysis

Table 2:- Standard scale for reliability

Cronbach’s Alpha	Reliability Level
α ≥ 0.9	Excellent
0.9 α ≥ 0.8	Good
0.8 α ≥ 0.7	Accepted
0.7 α ≥ 0.6	Questionable
0.6 α ≥ 0.5	Poor
0.5 > α	Unacceptable

Table 3:- Reliability Statistics of the collected data set

Reliability Statistics

Variable	Cronbach's Alpha	N of Items	Status
IN.Va one	.602	4	Questionable
IN.Va Two	.712	6	Accepted
IN.Va Three	.678	6	Questionable
IN.Va four	.719	3	Accepted
De.Va	.789	3	Accepted

According to the above table 3, all four independent variables as well as the dependent variable are internally reliable which is or above the reliability level of questionable so that the analysis can be processed further. Thus, the IV (1) has a Cronbach’s value of .602 which the status is Questionable whereas the IV (2) has a value of .712 which the status is accepted. Further, the IV (3) has a value of .678 that possess the status of questionable and finally the IV (4) has a value of .719 which is accepted. Ultimate key attribute which is defined as dependent variable has a value of .789. Therefore, DV is also in the status if accepted.

3.6.3. Data representation of independent and dependent variables

Table 4 :- Independent variable one

Statistics

		IV1.1	IV1.2	IV1.3	IV1.4
N	Valid	370	370	370	370
	Missing	0	0	0	0
Mean		3.69	3.49	3.73	3.99
Mode		4	3	3	4
Std. Deviation		.788	.850	.919	.855
Minimum		2	1	1	1
Maximum		5	5	5	5

As per the table 4, all 4 questions related to the Independent Variable 1 shares a common level of mean in the range of 3. When considering the mode, the highest mode is acquired by the IV1.1 and IV1.4 with a value of 4.

Table 5:- Independent variable 2

Statistics

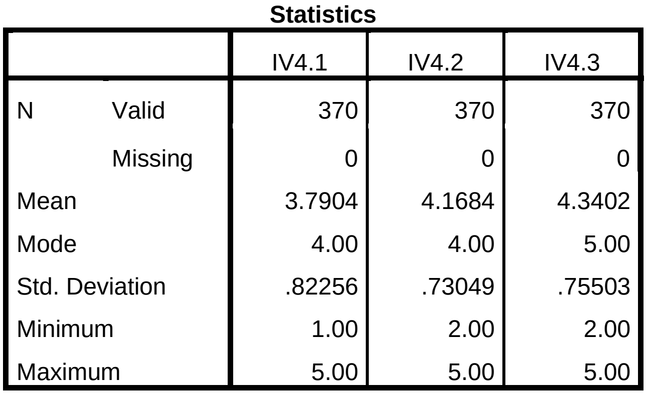
		IV2.1	IV2.6	IV2.2	IV2.3	IV2.4	IV2.5
N	Valid	370	370	370	370	370	370
	Missing	0	0	0	0	0	0
Mean		3.72	3.5811	3.69	3.91	4.1811	3.9757
Mode		4	4.00	4	3	4.00	4.00
Std. Deviation		.734	.96281	.825	.907	.78755	.89984
Minimum		2	1.00	1	1	1.00	1.00
Maximum		5	5.00	5	5	5.00	5.00

Table 6:- Independent variable 3

Statistics

		IV3.1	IV3.2	IV3.3	IV3.4	IV3.5	IV3.6
N	Valid	370	370	370	370	370	370
	Missing	0	0	0	0	0	0
Mean		3.3127	3.8832	3.7423	3.4467	3.1993	3.9244
Mode		4.00	4.00	4.00	3.00	3.00	4.00
Std. Deviation		.92948	0.05704	.74665	.82613	0.03464	.84351
Minimum		1.00	1.00	1.00	1.00	1.00	1.00
Maximum		5.00	5.00	5.00	5.00	5.00	5.00

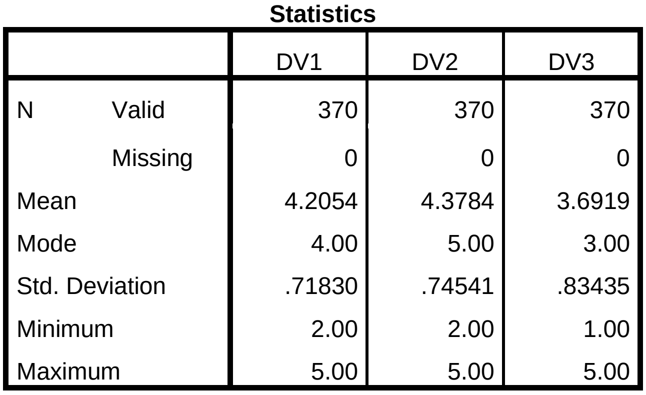
Table 7:- Independent variable 4



Statistics

		IV4.1	IV4.2	IV4.3
N	Valid	370	370	370
	Missing	0	0	0
Mean		3.7904	4.1684	4.3402
Mode		4.00	4.00	5.00
Std. Deviation		.82256	.73049	.75503
Minimum		1.00	2.00	2.00
Maximum		5.00	5.00	5.00

Table 8:- Dependent variable



Statistics

		DV1	DV2	DV3
N	Valid	370	370	370
	Missing	0	0	0
Mean		4.2054	4.3784	3.6919
Mode		4.00	5.00	3.00
Std. Deviation		.71830	.74541	.83435
Minimum		2.00	2.00	1.00
Maximum		5.00	5.00	5.00

3.6.4. Histograms for variables

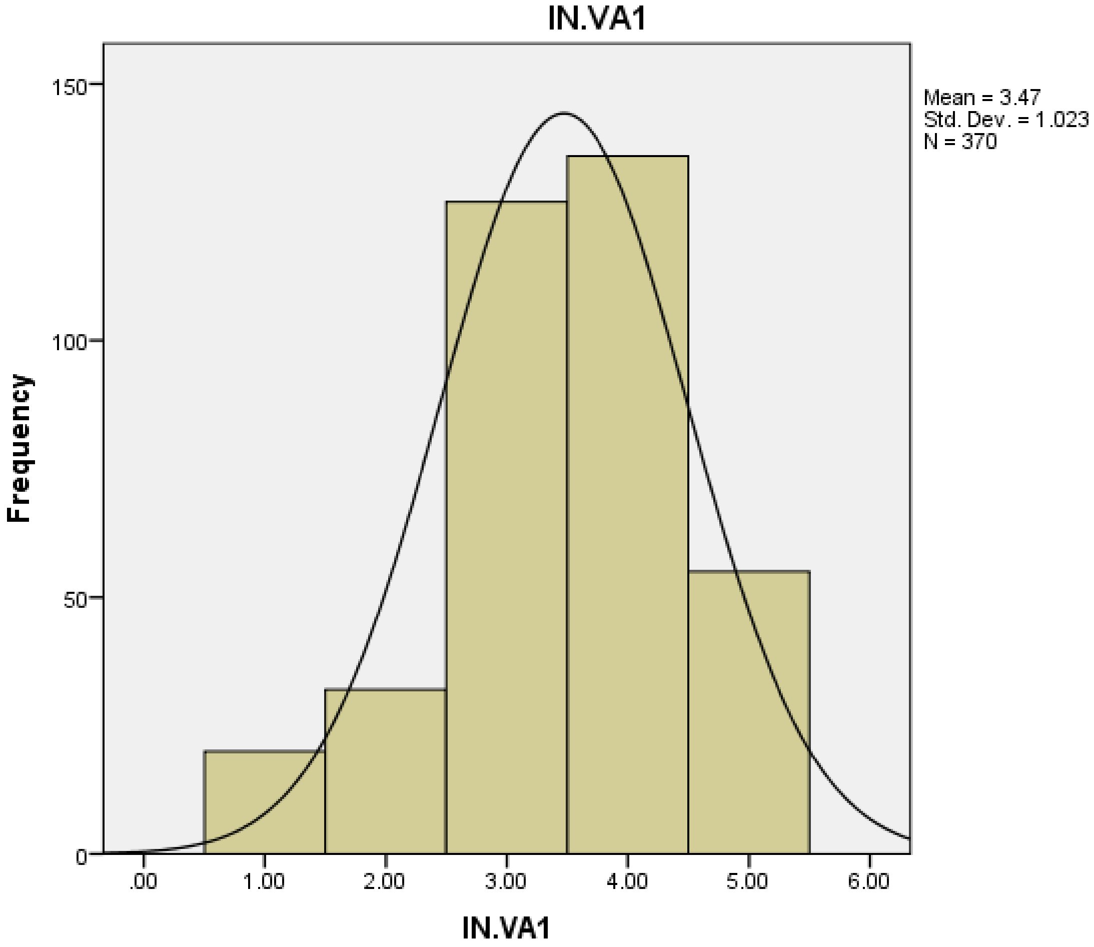


Figure 11:- Histogram for Independent Variable 1

According to the figure 11, the histogram depicts that the highest number of responses and the peak is marked at the response 4. Where this confirms that the participants in general "agree" that an "Influence is made by percentage of population using social media".

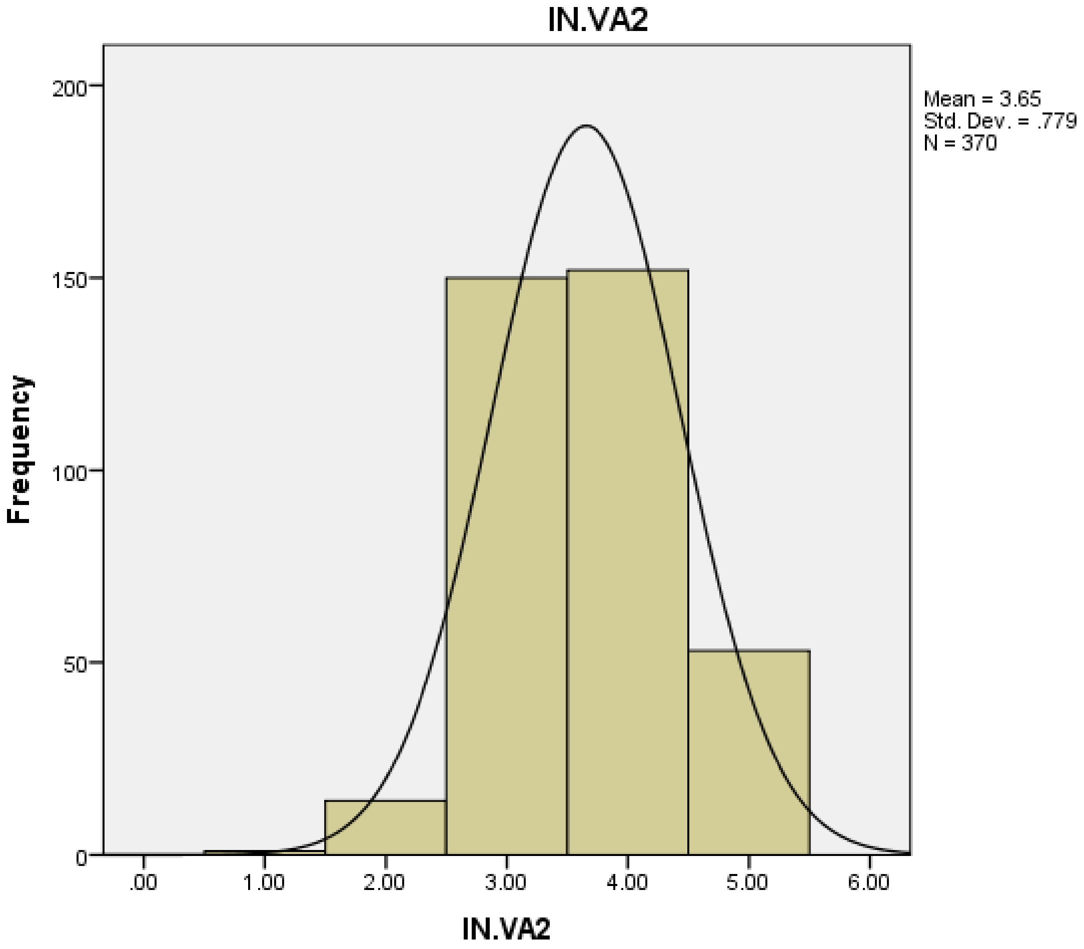


Figure 12:- Histogram for Independent Variable 2

According to the figure 12, the histogram depicts that the highest number of responses and the peak is marked at the response 4. Where this confirms that the participants in general “agree” that there is an “Influence of honest views shared by social media users”. Further, when considering the response 3 it possesses a comparably higher response rate similar to response for which representants that there is a neutral impression for the second largest number of responses.

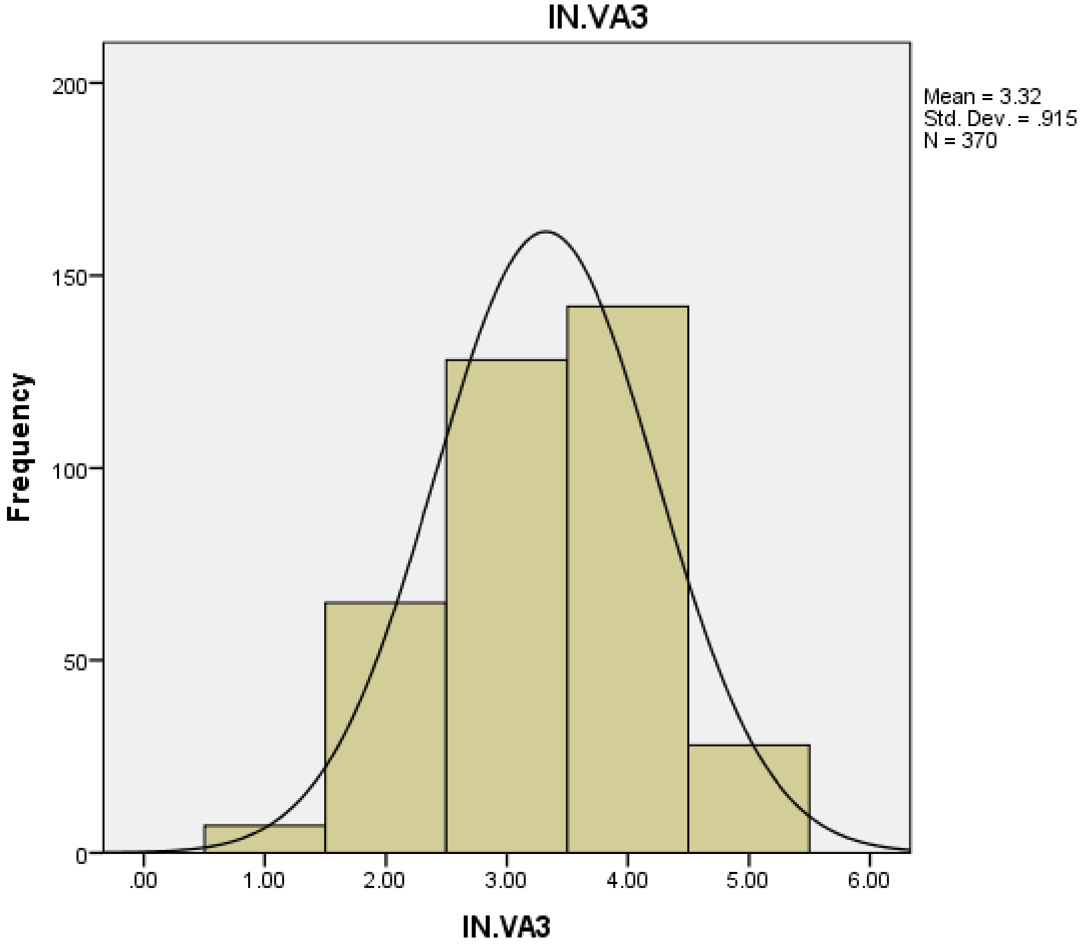


Figure 13 :- Histogram for Independent Variable 3

According to the figure 13, the histogram depicts that the highest number of responses and the peak is marked at the response 4. Where this confirms that the participants in general “agree” and has as “neutral” impression that there is “Influence of identifying and filtering terrorism, toxicity and hetaerism views and opinions on online spaces” is feasible.

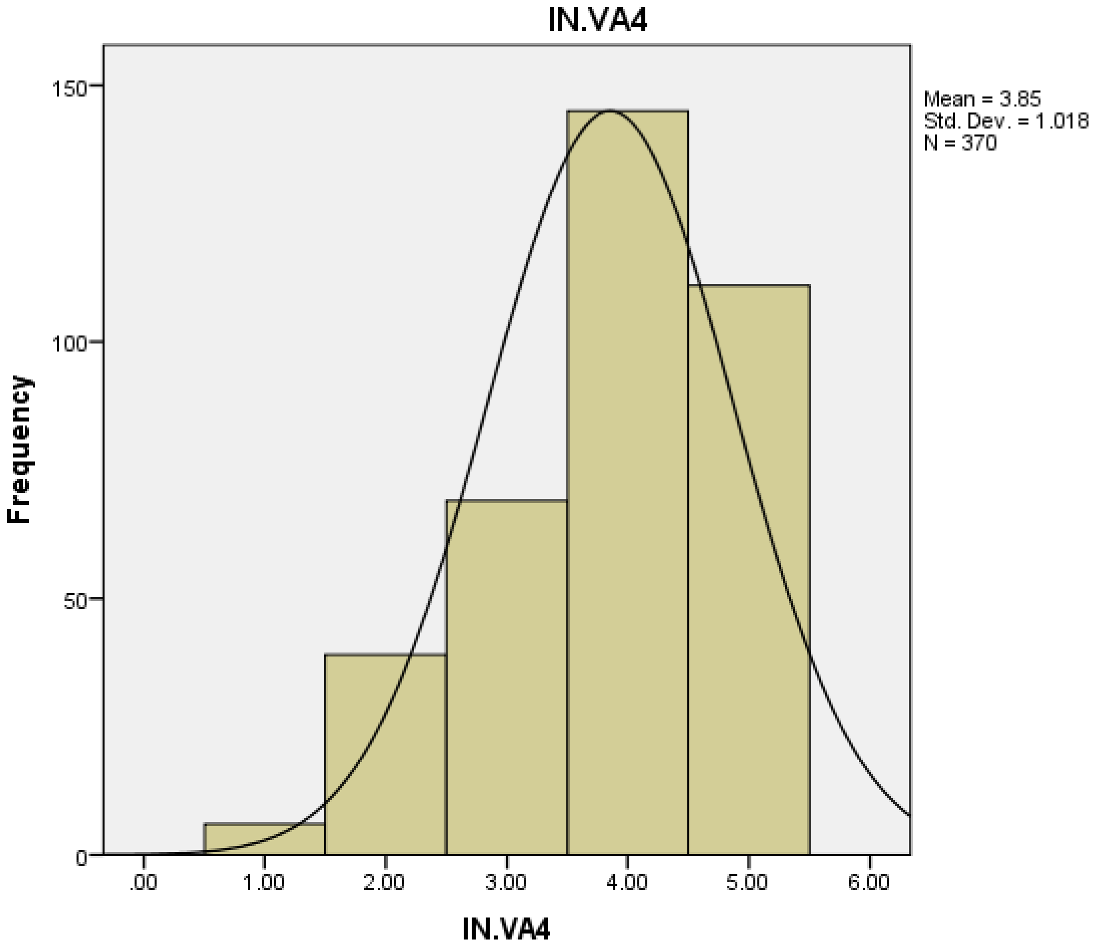


Figure 14:- Histogram for Independent Variable 4

According to the figure 14, the histogram depicts that the highest number of responses and the peak is marked at the response 4. Where this confirms that the participants in general “agree” and has a impression that the “Influence or agreement of users to allow an AI Model to analyse online spaces” overall acceptable. Moreover, than previous IVs the IV 4 has a certain higher rate of response to number 5.

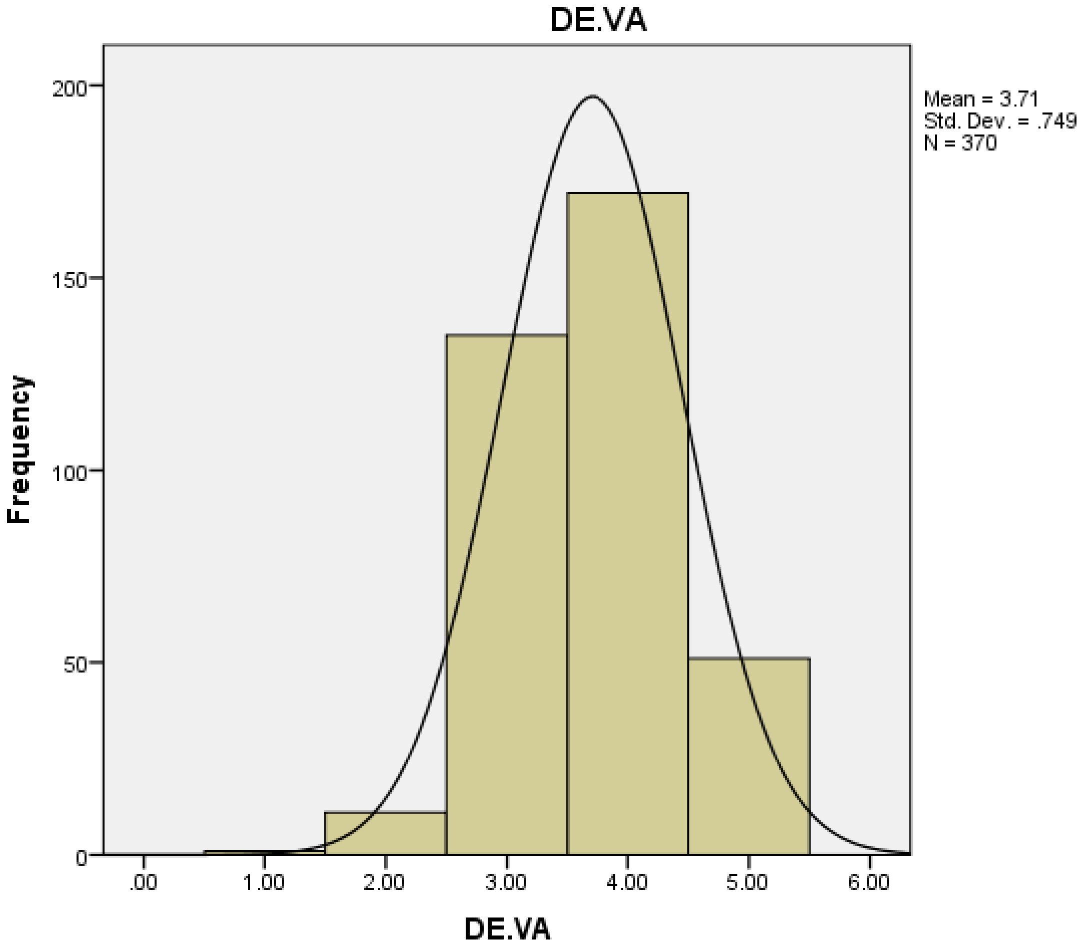


Figure 15:- Histogram for Dependent Variable

According to the figure 15, the histogram depicts that the highest number of responses and the peak is marked at the response 4 and 2nd highest response rate is at response 3 and 5 respectively. Where this confirms that the participants in general “agree” and has as “neutral” as well as “Strongly Agree” impression that there is “Use of AI Framework for understanding the behavioural patterns of terrorists, hetaerism and toxicity in online spaces” is practically feasible.

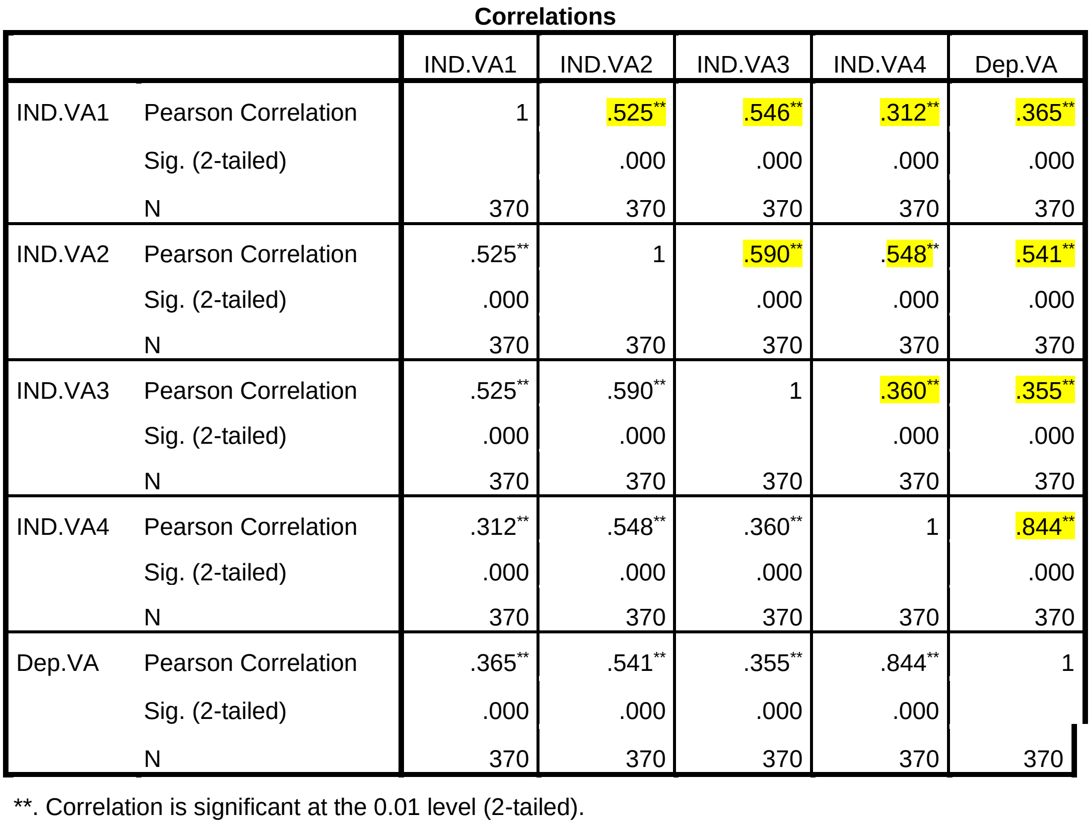
3.6.5. Correlation Analysis

Table 9:- Correlation Criterion

Value	Relationship
0.5 – 1	Strong Positive Relationship
0 – 0.5	Moderate Positive Relationship
0	No Relationship
0 – (-0.5)	Moderate Negative Relationship
(-0.5) – (-1)	Strong Negative Relationship

Correlation analysis is a methodology of statistical evaluation used to study the strength of a relationship between two, numerically measured, continuous variables. This particular type of analysis is useful when a researcher wants to establish if there are possible connections between variables. It is often misunderstood that correlation analysis determines cause and effect; however, this is not the case because other variables that are not present in the research may have impacted on the results. If correlation is found between two variables it means that when there is a systematic change in one variable, there is also a systematic change in the other the variables alter together over a certain period of time. If there is correlation found, depending upon the numerical values measured, this can be either positive or negative. Although the above-mentioned correlation is fairly obvious but when considering the gathered data, it could contain unsuspected correlations. To identify which correlations are the strongest an intelligence correlations analysis can lead to better understating of the data in hand. (The Survey System, 2018)

Table 10:- Correlations



Correlations

		IND.VA1	IND.VA2	IND.VA3	IND.VA4	Dep.VA
IND.VA1	Pearson Correlation	1	.525**	.546**	.312**	.365**
	Sig. (2-tailed)		.000	.000	.000	.000
	N	370	370	370	370	370
IND.VA2	Pearson Correlation	.525**	1	.590**	.548**	.541**
	Sig. (2-tailed)	.000		.000	.000	.000
	N	370	370	370	370	370
IND.VA3	Pearson Correlation	.525**	.590**	1	.360**	.355**
	Sig. (2-tailed)	.000	.000		.000	.000
	N	370	370	370	370	370
IND.VA4	Pearson Correlation	.312**	.548**	.360**	1	.844**
	Sig. (2-tailed)	.000	.000	.000		.000
	N	370	370	370	370	370
Dep.VA	Pearson Correlation	.365**	.541**	.355**	.844**	1
	Sig. (2-tailed)	.000	.000	.000	.000	
	N	370	370	370	370	370

**. Correlation is significant at the 0.01 level (2-tailed).

3.6.5.1. Correlation between Independent Variable 01 (IV01) and Dependent Variable (DV)

The percentage of people say that the use of AI in order to understand the behavioural patterns of terrorism, hetaerism and toxicity on social media based on AI represents a positive correlation of .365 which is greater than 0 and less than 0.5 to the dependent variable. Therefore, according to correlation criteria Influence is made by percentage of population using social media (*IV1*) and the Use of AI Framework for understanding the behavioural patterns of terrorists, hetaerism and toxicity in online spaces (DV) has a moderate positive relationship.

3.6.5.2. Correlation between Independent Variable 02 (IV02) and Dependent Variable (DV)

The percentage of participants mentions that the all the social media users do share their honest opinions represents a positive correlation of .541 which is greater than 0.5 and less than 1 to the dependent variable. Therefore, according to correlation criteria Influence of honest views shared by social media users (*IV2*) and the Use of AI Framework for understanding the behavioural patterns of terrorists, hetaerism and toxicity in online spaces (DV) has a Strong positive relationship.

3.6.5.3. Correlation between Independent Variable 03 (IV03) and Dependent Variable (DV)

The percentage of participants mentions that the identifying and filtering terrorism, toxicity and hetaerism views and opinions represents a positive correlation of .355 which is greater than 0 and less than 0.5 to the dependent variable. Therefore, according to correlation criteria Influence of identifying and filtering terrorism, toxicity and hetaerism views and opinions on online spaces (*IV3)* and the Use of AI Framework for understanding the behavioural patterns of terrorists, hetaerism and toxicity in online spaces (DV) has a Moderate positive relationship.

3.6.5.4. Correlation between Independent Variable 04 (IV04) and Dependent Variable (DV)

The percentage of participants mentions that the general public is comfortable of using an AI model to analyse social media represent a positive correlation of .844 which is greater than 0.5 and less than 1 to the dependent variable. Therefore, according to correlation criteria the Influence or agreement of users to allow an AI Model to analyse online spaces (IV3) and the Use of AI Framework for understanding the behavioural patterns of terrorists, hetaerism and toxicity in online spaces (DV) has a Strong positive relationship.

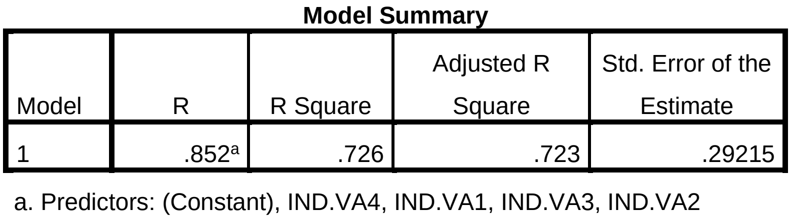
3.6.5.5. Correlation between independent variables and independent variables

There is a significant, positive correlation between IV 01 and IV02 and IV03 and IV 04, where the correlation between variables is achieved as 0. 525, 0. 546, and 0. 312 respectively. The significance values are under significance level 0.5. The correlation between variables IV02 and IV03 and IV 04 are 0.590 and 0.548, whereas significance values are under significance level 0.5. This states, there is positive significant correlation between IV02 and IV03 and IV04 as well as there is a positive correlation between IV03 and IV 04, where the correlation value is 0.3.60 and significance value is under significance level 0.05.

3.6.6. Regression Analysis

Linear regression is a useful technique in analytics that aims to evaluate the strength of the link between one dependent variable and a series of other changing variables (Independent variable). There are several methods for calculating linear regression. The ordinary least-squares approach is one of the most prevalent, and it estimates unknown variables in data by summing the vertical distances between data points and the trend line. Linear regression employs a single independent variable to explain or predict the outcome of the dependent variable, whereas multiple linear regression employs numerous independent variables to predict the outcome or final results. In this study, regression would help the researcher find the link between the dependent and independent variables (businessdictionary, 2019).

Table 11:-Model Summary



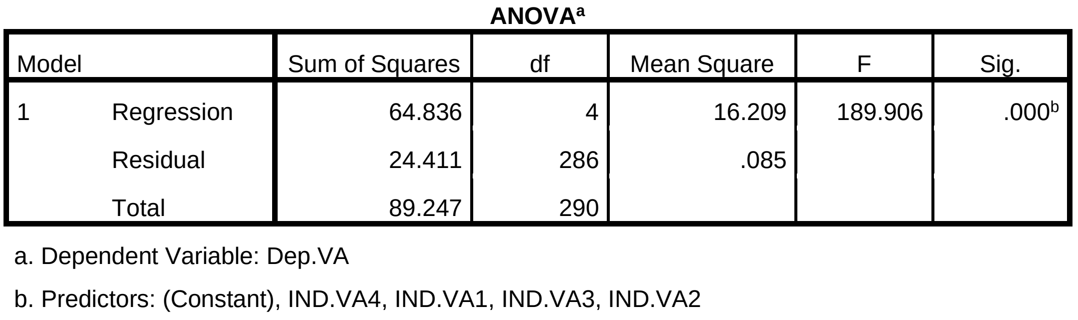
Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.852[a]	.726	.723	.29215

a. Predictors: (Constant), IND.VA4, IND.VA1, IND.VA3, IND.VA2

In above Model Summary, The R value represents the simple correlation as 0.852 (the “R” column), which indicates a high degree of correlation or a positive correlation. The R^2 (the “R Square” column) value is 0.726. R Square is the percentage of independent variables (IV1, IV2 IV3 and IV4) explained in the dependent variable (DV). Therefore, it is possible to declare that nearly 72.6% of the dependent variable (Use of AI Framework for understanding the behavioral patterns of terrorists, hetaerism and toxicity in online spaces) has been explained by the independent variables (IV1, IV2 IV3 and IV4).

Table 12:- ANOVA



ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	64.836	4	16.209	189.906	.000[b]
	Residual	24.411	286	.085		
	Total	89.247	290			

a. Dependent Variable: Dep.VA

b. Predictors: (Constant), IND.VA4, IND.VA1, IND.VA3, IND.VA2

The ANOVA represents the Significant (Sig.) relationship of independent variables (IV1, IV2 IV3 and IV4). and dependent variable (to understand the behavioral patterns of terrorism, hetaerism and toxicity in online spaces) which is in the scale of 0.000[b], when the Sig. is lesser than 0.05 it is observed that there is a significance relationship. Therefore:

- H_0 (Social media is not used by a significant percentage of the population) of hypothesis is rejected and H_1 (Social media is used by a significant percentage of the population) is accepted.
- H_0 (Social media users do not share their honest opinions) of hypothesis is rejected and H_1 (Social media users to share their honest opinions) is accepted.
- H_0 (Identifying and filtering terrorism, toxicity and hetaerism views and opinions based on an AI tool is not possible) of hypothesis is rejected and H_1 (Identifying and filtering terrorism, toxicity and hetaerism views and opinions based on an AI tool is possible) is accepted.
- H_0 (General Public is not comfortable of using an AI model to analyse social media) of hypothesis is rejected and H_1 (General Public is comfortable of using an AI model to analyse social media) is accepted.

It is can also be explained as, F (2, 286) = 189.906, P= 0.000.

Table 13 :- Coefficients

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.245	.158		1.547	.123
	IND.VA1	.093	.039	.091	2.367	.019
	IND.VA2	.086	.050	.077	1.737	.024
	IND.VA3	.018	.039	.019	.462	.044
	IND.VA4	.780	.037	.780	21.064	.000

a. Dependent Variable: Dep.VA

3.7.Conclusions and Recommendations for the research

This chapter contains interpretations generated from the study's research, including demographic interpretations and interpretations of the connection between variables using correlation and regression analysis. There are many perspectives on the potential histogram analysis of dependent variable was used to identify and filter terrorism, toxicity, and hetaerism thoughts and attitudes based on information posted on social media using an AI tool. Furthermore, this discusses the extension of knowledge in this context from the current knowledge pool, the researcher's perspective, and how this study might be improved for future research.

3.7.1. Demographic Conclusions

This section of study is comprised of conclusions derived from the outcomes of the demographic analysis.

- Conclusions on gender

  The results of demographic analysis on gender of participants in this study, who are social media users in the Kandy region, show that there is a considerable disparity in the number of men and females. This relevance may be attributed to females' significant degree of influence over males on social media platforms such as Facebook, TikTok, Instagram, and so on.

- Conclusions on Age

  According to the findings of the study's age analysis, more than 36.22% of participants are under the age of 25, while 32.97% are above the age of 46, and 24.59% are between the ages of 36 and 45, with a very low percentage in the age range 2 to 35 in the selected population. As a result, the researcher could conclude that users of social media platforms are largely young individuals under the age of 26, with a tendency of matured adults over the age of 46, while usage of young middle aged and middle-aged participants is relatively low.

- Conclusions on Occupation

  Along with the findings of the study analysis on participant occupation, it is clear that the majority of participants are full-time employees (57.5%), followed by part-time (17.57%) and unemployed (24.86%). This suggests that the majority of social media users are either full-time or part-time employees, which is consistent with the fact that employed persons in Kandy, Sri Lanka are more interested in social media.

- Conclusion on Education

  According to the study's education analysis, a bigger percentage of the population has professional credentials, whereas a lesser number of participants have Masters or Doctorates. This means that individuals with a higher education level use social media less frequently, whereas people with a lower educational background use social media more frequently.

- Conclusion on Marriage

  According to the study's Marriage analysis, both married and unmarried participants utilize social media in their daily lives to communicate their opinions to society, as a tool for entertainment, or for any other informative purpose, rather than television, radio, or any other medium.

- Conclusion on Usage of Internet

  According to the study's Internet Usage research, about 80.27% of participants utilize the internet, implying that people in society are utilizing the internet more than ever before in this rapidly evolving world.

- Conclusions on the Device Usage

  As per the outcomes of Device Usage analysis of the study, almost 82.43% of participants uses Smart Phone to access internet which interprets that the majority access the social media and online platforms based on smart phone rather than any other electronic device.

3.7.2. Revisiting research objectives

The pioneer goal of this study is to seek the possibility of identifying and filtering terrorism, toxicity and hetaerism views and opinions on online spaces based on an AI tool. This objective of the study is achieved in case of the achievement of specific objectives developed for the study by the researcher. The specific objective is included of finding out the percentage of population using social media in Sri Lanka, checking whether all the users share their honest views and opinions on social media, while identifying and filtering terrorism, toxicity and hetaerism views and general public is comfortable of using an AI model to analyze social media. This section involves interpretations derived from independent variables (specific objectives) and dependent variable (main objective). These interpretations are outcomes of analysis of data collected from the survey based on social media users in Kandy region.

3.7.3. Contextualization of research findings

Along with the results of correlation and regression analysis, it is feasible to conclude that there is a high likelihood of success in utilizing an AI Framework to comprehend the behavioural patterns of terrorism, hetaerism, and toxicity in online spaces. The interpretations of outcomes help to determine the percentage of the population in Sri Lanka that uses social media, whether all users share their honest views and opinions on social media, while identifying and filtering terrorism, toxicity, and hetaerism views, and whether the general public is comfortable using an AI model to analyse social media. Further interpretations are made on the success of employing AI in spotting terrorism, hetaerism, and toxicity in Sri Lanka by analysis of social media material using an AI technology.

3.7.4. Conclusion of dependent and independent variable relationship

- RO1: To determine the percentage of the Sri Lankan population that uses social media.

The researcher deduced from the correlation results that there is a considerably positive and strong link between IV01 and DV. This fact is supported by the study's regression analysis. This study indicates that the percentage of the Sri Lankan population that uses social media increases the likelihood of recognizing terrorism, hetaerism, and toxicity in Sri Lanka through analysis of information uploaded on social media using an AI tool. The hypotheses developed for this specific goal are H_1: A major portion of the population utilizes social media and H_0: A significant portion of the population does not use social media. According to the interpretations of RO1 study, H_1 might be generated as a precise hypothesis from this hypothesis pair because there is a strongly positive association.

- RO2: Analyse whether all the social media users share their honest views and opinions on social media.

According to the findings of the correlation and regression analyses, the researcher concluded that there is a statistically extremely strong and positive association between IV02 and DV. According to this study, social media analysis on whether social media users communicate honest ideas and opinions on social media increases the probability of spotting terrorism, hetaerism, and toxicity in Sri Lanka by analysing information posted on social media using an AI tool. The hypotheses developed for this specific goal are H_1: All social media users share their ideas and H_0: All social media users do not share their opinions. According to the interpretations of RO2 analysis, H_1: might be determined as a correct hypothesis from this hypothesis pair because there is a considerably positive association.

- RO3: identifying and filtering terrorism, toxicity and hetaerism views and opinions on online spaces is possible based on an AI tool.

The researcher deduced from the correlation results that there is a considerably positive and strong link between IV03 and DV. According to the study's regression analysis, it is feasible to detect terrorism, hetaerism, and poisonous beliefs and ideas in Sri Lanka by analysing information uploaded on social media with an AI tool. H_1: Identifying and filtering terrorism, toxicity, and hetaerism views and opinions is feasible; H_0: Identifying and filtering terrorism, toxicity, and hetaerism views and opinions is not possible. According to the interpretations of RO1 study, H_1 could be generated as a precise hypothesis from this hypothesis pair because there is a considerably positive association.

- RO4: agreement of users to allow an AI Model to analyse online spaces.

In accordance with the findings of the correlation and regression analyses, the researcher concluded that there is a substantial and very strong positive association between IV04 and DV. The study finds that social media users' willingness to enable AI analysis effects the likelihood of spotting terrorism, toxicity, and hetaerism in Sri Lanka through analysis of information uploaded on social media using an AI tool. H_1: The general public is comfortable using an AI model to analyse social media and H_0: The general public is not comfortable using an AI model to analyse social media are the hypotheses developed for this specific purpose.

According to the interpretations of RO4 study, H_1 might be deduced as the correct hypothesis from this hypothesis pair because there is a considerably positive association.

3.7.5. Conclusions on relationships between independent variables

As per the results of the correlation study between independent variables, the researcher might draw additional inferences. According to the correlation analysis, each of the independent variables IV01, IV02, IV03, and IV04 has a substantial, positive connection. This correlation indicates that the higher the percentage of the population who uses social media, the more honest the opinions shared by social media users, identifying and filtering terrorism, toxicity, and hetaerism views and opinions on online spaces, and the willingness of social media users to analyse their opinions using an AI tool. The capacity to recognize user opinions improves as the number of people who use social media grows. This might be due to Big Data analysis of patterns becoming simpler with the availability of a bigger quantity of data. Further interpretations might be made that as the number of honest opinions given by users improves, so does the capacity to recognize user viewpoints and their desire to engage in social media analysis. This might be because someone who expresses their true feelings will not be concerned about their thoughts being analysed.

3.8. Recommendations

Introduction

This component comprises of suggestions for the Sri Lankan Cyber Security Unit, the Sri Lankan Defence Ministry, and other relevant law enforcement bodies in Sri Lanka based on the knowledge gleaned from the literature research and survey findings. The proposals would help Sri Lanka's Cyber Security Unit and the Defence Ministry to reformulate their techniques for detecting terrorists. These suggestions are based on the study's four proven hypotheses.

3.8.1. Recommendation on IV1

According to the study's interpretations, terrorism, toxicity, and hetaerism behaviours can be discovered by AI analysis of information shared by social media users. With the increased number of social media users, terrorists may be readily spotted, according to this study. Because the majority of the population uses social media platforms, the Sri Lankan cybercrime unit could easily track down any terrorism, toxicity, and hetaerism behaviour, and thus any terrorism, toxicity, and hetaerism threat, such as the unfortunate situation on April 21st, 2019, on different churches and hotels across Sri Lanka, could have been curtailed before consequences. As a result, the cybercrime unit, in collaboration with the government of Sri Lanka, may intensify their attention on growing the population's use of social media platforms. This could be accomplished by promoting the benefits of using social media among the population and emphasizing the importance of social media usage in analysing terrorism, toxicity, and hetaerism among the population in order to achieve success through social media analysis to identify terrorism behaviours. Similarly, the usage of social media by the majority of the population would facilitate the study of their qualities that may be easily observable and track any crimes committed by them, allowing law enforcement agencies to quickly lessen numerous terrible crimes.

3.8.2. Recommendation on IV2

According to the interpretations of the study's findings, honest views of social media users are crucial in analysing terrorism, hetaerism, and toxicity using social media opinions and perspectives. Although there was terrorism, hetaerism, and toxicity in Sri Lanka prior to the 21st April blasts, none of those terrorism, hetaerism, and toxicity related induvial or groups were previously tipped off, either directly to law enforcement authorities or as views expressed through social media platforms. This could be due to people's fear of getting involved in any potential problems, so the researcher suggests that the government of Sri Lanka and relevant law enforcement authorities actively and passively change the perception of people in Sri Lanka to fearlessly post their views and opinions on social media, and that they be rewarded for recognizing such terrorism, hetaerism, and toxicity activities.

3.8.3. Recommendation on IV3

According to the findings of this study, terrorism, hetaerism, and toxicity may be filtered via their social media ideas and attitudes. As a result, the researcher advises law enforcement to use artificial intelligence-based social media analysis to identify terrorism, hetaerism, and toxicity-related individuals or groups, which might aid in protecting the country from future threats.

3.8.4. Recommendation on IV4

The study says most of the population prefer the use of artificial intelligence in analysing social media behaviours and identifying terrorism, hetaerism and toxicity related individuals or groups. As most of the people are pleased with the social media analysis of AI in identifying terrorism, hetaerism and toxicity, implementation of such social media analysis are recommended for relevant law authorities. This implementation would ease the process of identifying and filtering terrorism, hetaerism, and toxicity related individuals or groups as most of the population would provide their maximum support.

Chapter 04 - Design, Implementation and Evaluation

4.1.Development Methodology

The Agile methodology is the most suitable development technology which would fit the solution planned to provide, because both the web application and the AI Detection multi-classification model can be developed in sprints, and the AI model, which uses the Artificial Neural Networks (ANN) methodology, would benefit from the Agile methodology's flexibility. Therefore, as a result the ANN methodology can adapt to a continuous development phase where the model learns and adapts from training with the relevant training dataset comparing the expected to actual results, and then adjusting weights based on the error. As a result, the linkages that provide correct answers are strengthened, while those that provide incorrect predictions are weakened, allowing the model to be accurate (Crabb, et al., 2019). The Agile Methodology would be put into practice inorder to fulfil the software development life cycles, iteration and sprints are performed according to the agile methodology structure, and it allows to complete the proposed model in phases where it's easier to manage and debug any bugs during the development process and the whole project is planned on executing in 4 sprints as per the Gantt Chart in the appendix, During the first sprint, the progress of the model is to analyses and finalize the concepts and algorithms to use for the development and report in progress reports whereas for the second and third sprint the development of the model is to be started along with the development of the web app such as creating the UI of the web app, and designing the prototype, creating the structure of the model, creating the AI model prototype and training.

4.2.Feasibility Study

4.2.1. Time feasibility

The most critical component in project success is time feasibility. A project will fail if it is not completed on time. The recommended project development technique utilized on the project is agile methodology, which takes an iterative approach to project completion and helps the project to produce value faster, with improved quality and predictability, as described above. This is where scrum, an agile project management framework that offers a structure for fast-paced Agile development to prioritize, organize, and execute the project in sprints with a set of dynamic objectives and milestones that would allow the project to conclude on time, comes in.

A Gantt chart is used to carry out this task in a methodical manner, and it is placed below the Project work plan section.

4.2.2. Cost feasibility

Before allocating financial resources, organizations may use this tool to examine the viability, cost, and advantages of initiatives.

The following factors covers the cost feasibility of the study:

- Training Data: Data for training that must be gathered If that is the case, how much time and money will it require, and how much training data can be obtained for free through Kaggle
- Predictive Features: The components that are likely to forecast the target variable, as well as whether or not the data is accessible. The model and other major materials utilized are completely free and open source.
- Data Sources: What data sources would you require access to? Do you have internal help from data engineers? How much will vendor data cost if it is external? Web scraping over Reddit forums to obtain bespoke data, in addition to the original dataset that collected data from Twitter.
- Production: the cost and effort it will take to design, implement, and manage your production model. In terms of production, there were charges associated with the Google Cloud Platform, where the model is hosted and deployed as a microservice, and the project uses the free tier of Docker containers for maintenance.

4.2.3. Scope feasibility

Feasibility Scope Studies are in-depth technical assessments of your goals.

As discussed in the literature review, whatever model is to be developed, the right dataset and pre-processing method play a huge role in the accuracy and performance of the model because in some cases, as shown in the literature review, even though some researchers had the right classifying algorithms. However, their model did not perform well due to poorly pre-processed datasets, and the second factor that has the greatest impact is that even after pre-processing the datasets, the use of the right algorithms and models to classify the data into their relevant labels completes the other half of the model, so with the literature survey completed, the concluded decision for the proposed module is to use BERT algorithm and manual regex functions to manually clean the data for the construction of the final model which makes the scope of the project viable.

4.2.4. Technical feasibility

The suggested system has a good chance of being implemented.

Once a certain number of solutions have been explored during optimization, artificial neural networks (ANNs) can be trained and used to estimate the objective function. The type and topology of the goal function influence the number of training points, but the dimensionality of the goal function is the most important factor. The number of training points needed to achieve the same precision of ANN approximation throughout n-dimensional space rises exponentially with dimension. In terms of the number of optimization variables, this section emphasizes the technical feasibility of using ANNs for objective function approximation in structural optimization issues. According to the BERT algorithm, which works on the weight of context in a sentence using a method called tokenization, in which strings are converted into integers, and it passes the tokenized inputs through several layers of classifiers rather than relying on single word meaning, and to regulate this, python is used alongside other libraries such as numpy, pandas, keras, and transformers.

4.2.5. Economic feasibility

An examination of a project's expenses and revenues to see whether it is sensible and possible to carry through. The costs and benefits of the suggested plan are estimated as part of this, and the proposal is only economically feasible if the tangible or intangible benefits outweigh the expenses. The cost of implementing the model and the social web app would be high. The model and social web app on offer are both affordable and viable.

4.3.Architectural Design

4.4. An Overview of the Whole Solution

Prediction API is where the main component is based on. It has the saved classification model packaged into a dockerized image along with the FASTAPI environment. This runs in a Google Cloud Run Container. Social Media Web App and it's RESTful API will depict the use cases of a typical social media application. This will allow our Research Component which is the classification model to be demonstrated for a better understanding. Cloudinary is used for media uploads and downloads. This is used to make the RESTful API very minimal. MongoDB is used as the DBMS for the social media app. Along with its uses, this can also be used for further enhancement of the classification model in the future.

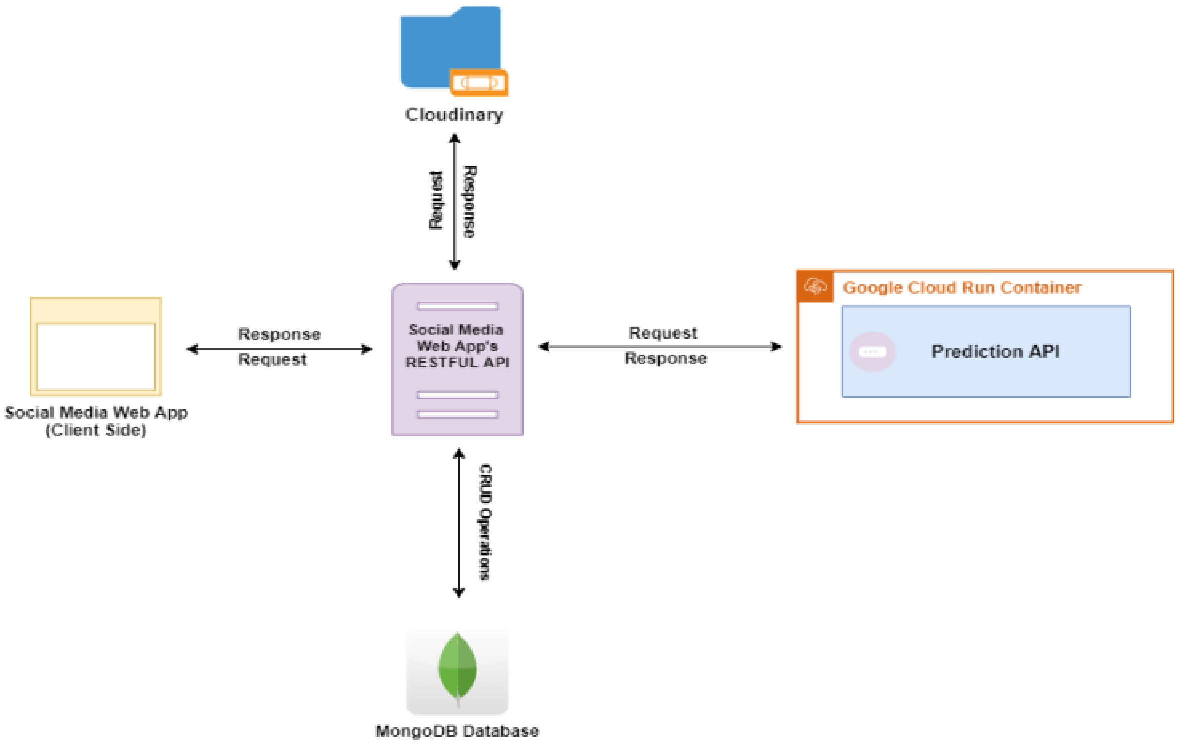


Figure 16 :- Architectural Design Diagram

4.5. Functionality of the Social Media Web App and AI Model

The below diagram depicts the interaction of the with the AI Model based on social media web application.

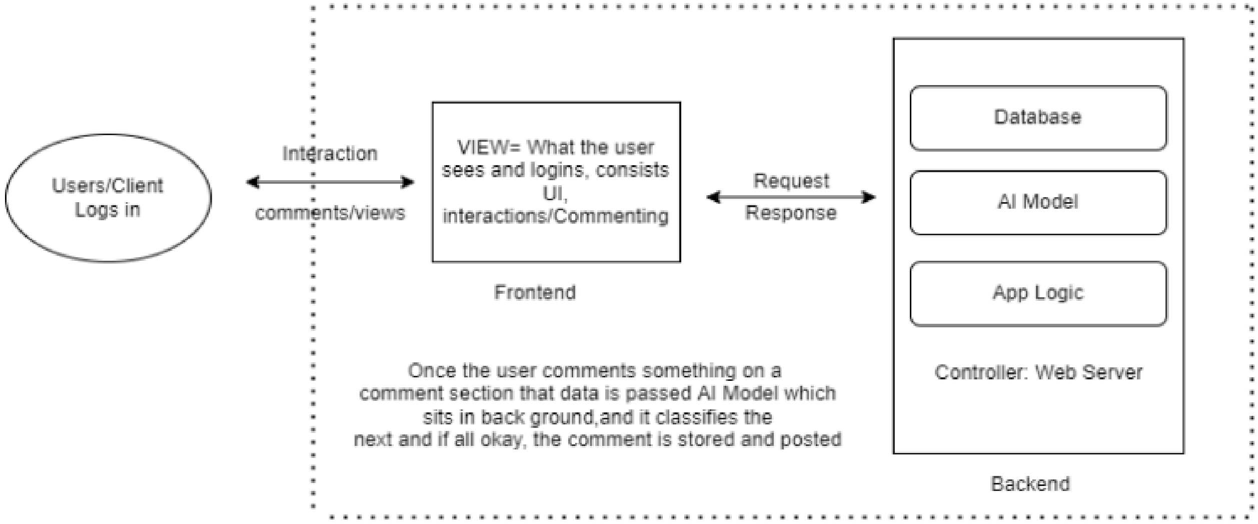

Users/Client
Logs in
Interaction
comments/views
VIEW= What the user sees and logins, consists UI, interactions/Commenting
Frontend
Request
Response
Database
AI Model
App Logic
Controller: Web Server
Backend
Once the user comments something on a comment section that data is passed AI Model which sits in back ground,and it classifies the next and if all okay, the comment is stored and posted

Figure 17 :Functionality of the Social Media Web App and AI Model

4.6.Product Flow Diagram for the AI/Detection Model

Below figure depicts the data flow of the solution based on how detection model functions along with the web application.

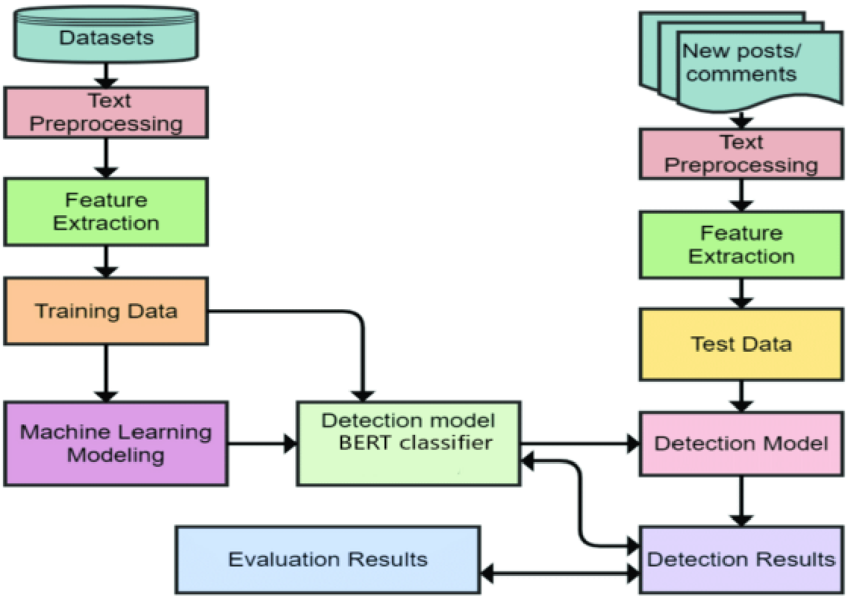


Figure 18:- Data Flow Diagram

4.7.Hardware and Software requirements

4.7.1. Hardware Requirements

In order to execute the proposed AI Model consideration of the hardware requirements in important. The server-side hardware should be somewhat powerful to more concurrent requests and data traffic, as well as in order to enable availability. It is a must to handle the system with a smooth data flow without any interruptions or distractions. Thus, if the system does not possess the ability to response and provide services in an efficient way, it grants idea that the system is overall failed. The below table consist of minimum hardware requirements.

Table 14:- Minimum Hardware Requirements

Device	Hardware Component	Minimum Requirement
	Processor Type	Core 2 Duo or higher
	Processor Speed	2.0 GHz or higher
	RAM Capacity	8 GB or higher

Server PC	Disk Space	500 GB or higher
	Bandwidth (Network Connection)	25 Mbps or higher
Client PC	Processor Type	Dual Core or higher
	Processor Speed	1.6 Hz or higher
	RAM Capacity	512 or higher
	Disk Space	500 Gb or higher
	Internet Connection	Dialog, SLT etc..
	Bandwidth (Network Connection)	10Mbps or higher

4.7.2. Software Requirements

Software requirements is a major factor in order to provide an efficient and effective service to the client as well as to execute the AI Model smoothly. This particular system is platform independent, therefore the system can be executing on almost all the platforms as any user who has internet can visit the social media website. The below table includes further details about the Minimum software requirements required.

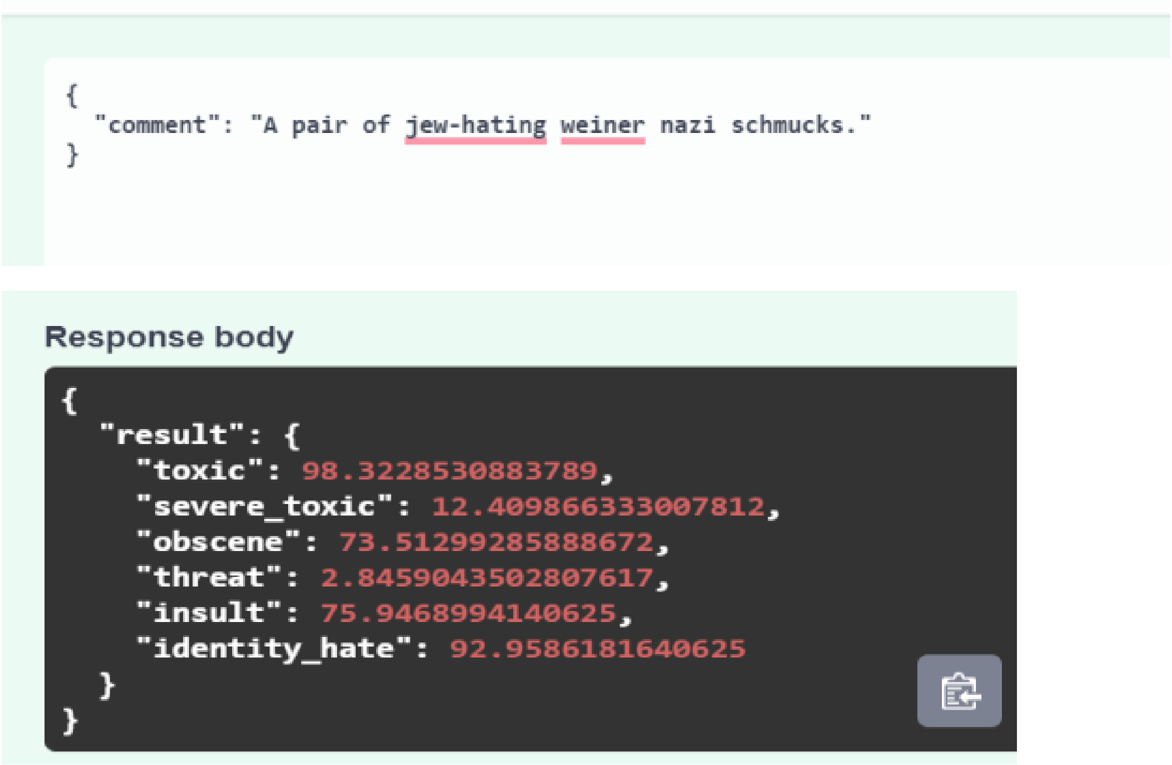
Table 15:-Minimum Software Requirements

Device	**Software**	**Minimum Requirement /Description**
Server PC	Operating System	Windows 7 or higher
	Data Base Management System	MySQL
	Browser	In order to test and implement, and regulate the workflow
	POSTMAN	To test server responses
	MongoDB Compass	Database management tool
	Visual Code	Text Editor
	Kaggle Notebook	GPU enabled notebooks for the training and development of the model

	Cloudinary	is an end-to-end image- and video-management solution covering everything from image and video uploads, storage, manipulations, optimizations to delivery
	FAST api	FastAPI is a Python framework and set of tools that enables us to use a REST interface to call commonly used functions to implement Models
	Node js	a single-threaded, open-source, cross-platform runtime environment for building fast and scalable server-side and networking applications. It runs on the V8 JavaScript runtime engine
	Google Cloud	managed compute platform that enables to run containers that are invocable via requests or events
Client PC	Operating System	Windows 7 or higher
	Browser	Any Web Browser (Recommended Google Chrome)

4.8. Evaluating of Solutions

The multi-classification detection model is the central component of the suggested approach. Which was effectively educated and tested to be open to feedback testing (beta testing). After that, it was packed, dockerized, and deployed as a microservice. It is now available to the general public. This may be implemented as an API into any system for detecting terrorism, toxicity, and hetaerism. This classification model has exactly 6 labels which are toxic, identity, hate, insult, obscene, severe, toxic and threat. If any text is not detected as one of the labels listed above, an empty result is returned. The outcome would be a percentage of the above-mentioned labels. The text's aggression is decided by the label/labels that yields the highest percentage.



```
{
  "comment": "A pair of jew-hating weiner nazi schmucks."
}
```

Response body

```
{
  "result": {
    "toxic": 98.3228530883789,
    "severe_toxic": 12.409866333007812,
    "obscene": 73.51299285888672,
    "threat": 2.8459043502807617,
    "insult": 75.9468994140625,
    "identity_hate": 92.9586181640625
  }
}
```

Figure 19:- Label/labels' which returns the most percentage based on text's aggressiveness is determined

4.9.Project Work Plan

4.9.1. Project Plan

Initially, before doing any implementation, researching the topic thoroughly is much important and key to success. Therefore, this would take the first few tasks as per the initial project plan. It was possible to do a thorough literature review on how the AI detection model is made possible by the inheritance of AI techniques and technologies. This is why the next steps involve data collection and pre-processing. Therefore, being able to understand what needs to be done to move forward with the application, the most crucial part of the whole journey is to train and test a model to perfection. Thus, if the model isn't detecting or functioning perfectly, then the web application which would be developed will fail to serve its purpose as mentioned above in the 'Project Aim'. Below figure depicts the project delivery of the project.

Table 16:-Project Deliverables

Deliverable	Recipients	Delivery Method
Project Registration Form	Canvas Portal	Soft Copy
Project Proposal	Canvas Portal	Soft Copy
Interim Report	Canvas Portal	Soft Copy
Final Report	Canvas Portal	Soft Copy
Final System	Canvas Portal	Soft Copy

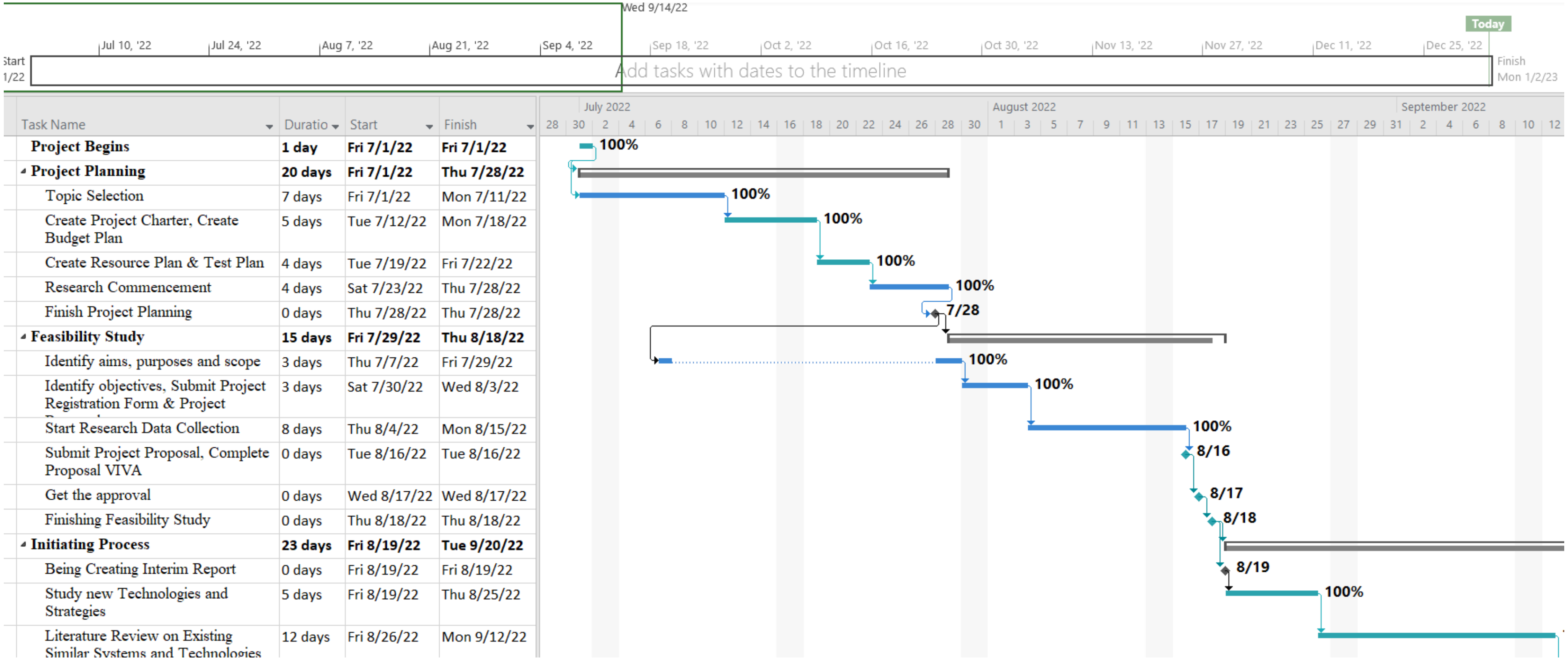
The Gantt chart is the graphical or the visual aspect of representing the task within the work break down structure of Project. This particular tool or the mechanism comfort and aids in the arranging, scheduling and proceeding the project. The Project management timelines and tasks are transformed into a horizontal bar chart, which has the beginning and end dates, which occupies inter-dependencies with the schedules and the deadlines, compromising the percentage of the task which is being completed or accomplished per levels according to the Initial Project plan done for the AI model & Web Application. Therefore, the Gantt is very much essential and a useful document which conduct tasks and track them using Milestone lists within the project. The different phases within the project have been displayed summarized which anyone could recognize and understand how the project management and the tasks are being achieved or accomplished within the project. Further, this particular graphical representation consists of the tasks or the project schedules such as the start and finish dates of the tasks and the activities which possess the summary of the resources, milestones, tasks and dependencies. This is being used to track and monitor the success criteria and the tasks of the project. Accordingly, this project is being started on 1st of July 2022 where the main tasks as dated as,

Table 17 :- Scheduled Main Tasks, Dates and Durations

Main Tasks	Start Date	End Date	Duration
Project Begins	Fri 7/1/22	Fri 7/1/22	0 Days
Project Planning	Fri 7/1/22	Thu 7/28/22	20 Days
Feasibility Study	Fri 7/29/22	Thu 8/18/22	15 Days
Initiating Process	Fri 8/19/22	Tue 9/20/22	23 Days
Designing and Development	Wed 9/21/22	Tue 11/29/22	50 Days
Testing	Wed 11/30/22	Tue 12/20/22	15 Days
Finalizing the Project	Wed 12/21/22	Mon 1/2/23	9 Days

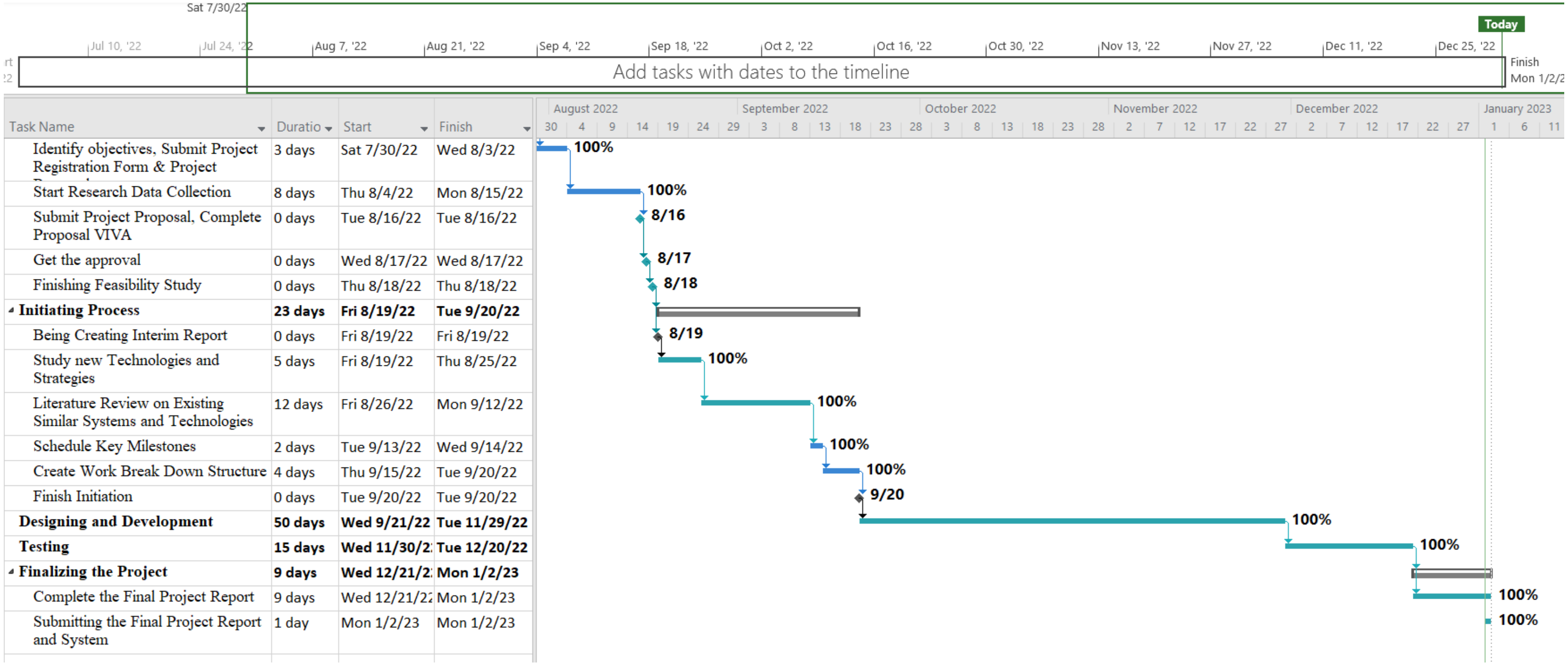
The planned time duration for the project is Approximately 6 months of time from 1st of July 2022 to 2nd January 2022 excluding public holidays, and Full Moon Poya Days. Below figures represents the initial Gantt chart with the Work Break Down Structure and Scheduled Dates.

4.9.2. Gantt Chart



Task Name	Duratio	Start	Finish
Project Begins	**1 day**	**Fri 7/1/22**	**Fri 7/1/22**
Project Planning	**20 days**	**Fri 7/1/22**	**Thu 7/28/22**
Topic Selection	7 days	Fri 7/1/22	Mon 7/11/22
Create Project Charter, Create Budget Plan	5 days	Tue 7/12/22	Mon 7/18/22
Create Resource Plan & Test Plan	4 days	Tue 7/19/22	Fri 7/22/22
Research Commencement	4 days	Sat 7/23/22	Thu 7/28/22
Finish Project Planning	0 days	Thu 7/28/22	Thu 7/28/22
Feasibility Study	**15 days**	**Fri 7/29/22**	**Thu 8/18/22**
Identify aims, purposes and scope	3 days	Thu 7/7/22	Fri 7/29/22
Identify objectives, Submit Project Registration Form & Project	3 days	Sat 7/30/22	Wed 8/3/22
Start Research Data Collection	8 days	Thu 8/4/22	Mon 8/15/22
Submit Project Proposal, Complete Proposal VIVA	0 days	Tue 8/16/22	Tue 8/16/22
Get the approval	0 days	Wed 8/17/22	Wed 8/17/22
Finishing Feasibility Study	0 days	Thu 8/18/22	Thu 8/18/22
Initiating Process	**23 days**	**Fri 8/19/22**	**Tue 9/20/22**
Being Creating Interim Report	0 days	Fri 8/19/22	Fri 8/19/22
Study new Technologies and Strategies	5 days	Fri 8/19/22	Thu 8/25/22
Literature Review on Existing Similar Systems and Technologies	12 days	Fri 8/26/22	Mon 9/12/22

Figure 20:-Initial Project Plan, Gantt Chart (Image No - 01) (Author Developed)



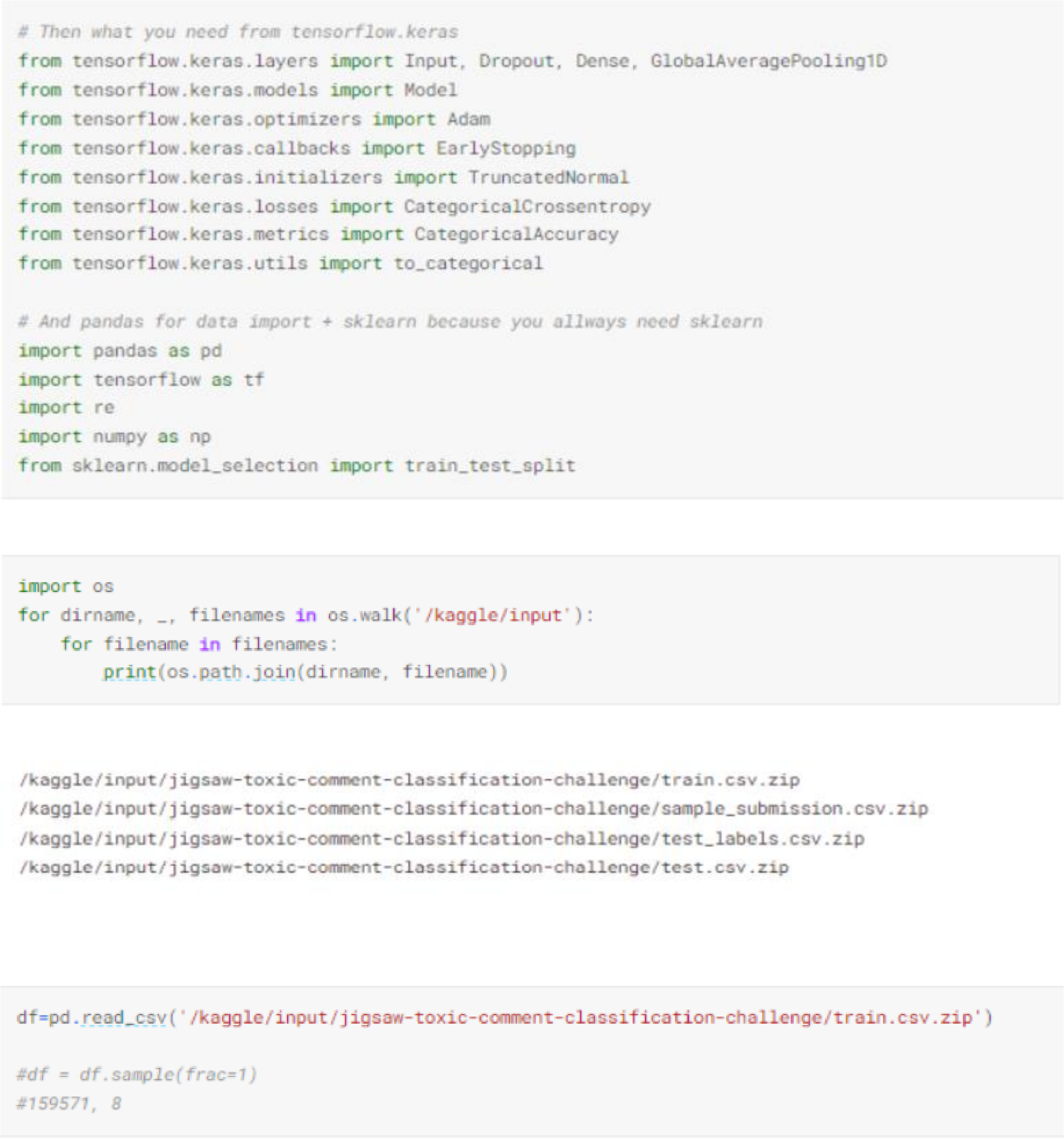
Task Name	Duratio	Start	Finish
Identify objectives, Submit Project Registration Form & Project	3 days	Sat 7/30/22	Wed 8/3/22
Start Research Data Collection	8 days	Thu 8/4/22	Mon 8/15/22
Submit Project Proposal, Complete Proposal VIVA	0 days	Tue 8/16/22	Tue 8/16/22
Get the approval	0 days	Wed 8/17/22	Wed 8/17/22
Finishing Feasibility Study	0 days	Thu 8/18/22	Thu 8/18/22
Initiating Process	**23 days**	**Fri 8/19/22**	**Tue 9/20/22**
Being Creating Interim Report	0 days	Fri 8/19/22	Fri 8/19/22
Study new Technologies and Strategies	5 days	Fri 8/19/22	Thu 8/25/22
Literature Review on Existing Similar Systems and Technologies	12 days	Fri 8/26/22	Mon 9/12/22
Schedule Key Milestones	2 days	Tue 9/13/22	Wed 9/14/22
Create Work Break Down Structure	4 days	Thu 9/15/22	Tue 9/20/22
Finish Initiation	0 days	Tue 9/20/22	Tue 9/20/22
Designing and Development	**50 days**	**Wed 9/21/22**	**Tue 11/29/22**
Testing	**15 days**	**Wed 11/30/2:**	**Tue 12/20/22**
Finalizing the Project	**9 days**	**Wed 12/21/2:**	**Mon 1/2/23**
Complete the Final Project Report	9 days	Wed 12/21/22	Mon 1/2/23
Submitting the Final Project Report and System	1 day	Mon 1/2/23	Mon 1/2/23

Figure 21:- Initial Project Plan, Gantt Chart (Image No - 02) (Author Developed)

4.10. Implementation

The necessary imports, together with the prepared datasets, are imported as the initial part of the classification model's execution stage.

External Dataset link: [Toxic Comment Classification Challenge | Kaggle]



```
# Then what you need from tensorflow.keras
from tensorflow.keras.layers import Input, Dropout, Dense, GlobalAveragePooling1D
from tensorflow.keras.models import Model
from tensorflow.keras.optimizers import Adam
from tensorflow.keras.callbacks import EarlyStopping
from tensorflow.keras.initializers import TruncatedNormal
from tensorflow.keras.losses import CategoricalCrossentropy
from tensorflow.keras.metrics import CategoricalAccuracy
from tensorflow.keras.utils import to_categorical

# And pandas for data import + sklearn because you allways need sklearn
import pandas as pd
import tensorflow as tf
import re
import numpy as np
from sklearn.model_selection import train_test_split
```

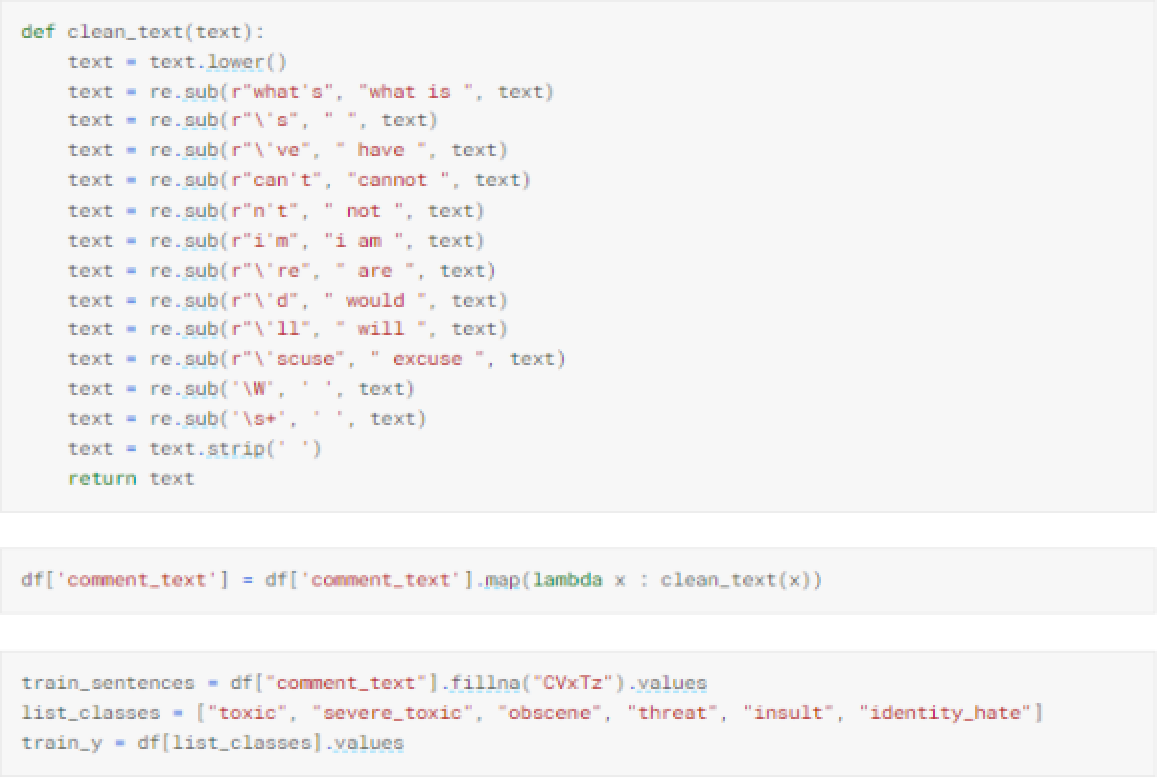
```
import os
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))
```

```
/kaggle/input/jigsaw-toxic-comment-classification-challenge/train.csv.zip
/kaggle/input/jigsaw-toxic-comment-classification-challenge/sample_submission.csv.zip
/kaggle/input/jigsaw-toxic-comment-classification-challenge/test_labels.csv.zip
/kaggle/input/jigsaw-toxic-comment-classification-challenge/test.csv.zip
```

```
df=pd.read_csv('/kaggle/input/jigsaw-toxic-comment-classification-challenge/train.csv.zip')

#df = df.sample(frac=1)
#159571, 8
```

Figure 22 :- Importing the Dataset and relevant Libraries

After the dataset is imported, it is preprocessed using regex functions. As we concluded in the literature review, the dataset plays an important role in obtaining a highly accurate model, so what preprocess does is lowercase all the sentences first and then take up the contractions words and extends it to the original form, resulting in the data carrying more context and thus the model having a better understanding.



```
def clean_text(text):
    text = text.lower()
    text = re.sub(r"what's", "what is ", text)
    text = re.sub(r"\'s", " ", text)
    text = re.sub(r"\'ve", " have ", text)
    text = re.sub(r"can't", "cannot ", text)
    text = re.sub(r"n't", " not ", text)
    text = re.sub(r"i'm", "i am ", text)
    text = re.sub(r"\'re", " are ", text)
    text = re.sub(r"\'d", " would ", text)
    text = re.sub(r"\'ll", " will ", text)
    text = re.sub(r"\'scuse", " excuse ", text)
    text = re.sub('\W', ' ', text)
    text = re.sub('\s+', ' ', text)
    text = text.strip(' ')
    return text
```

```
df['comment_text'] = df['comment_text'].map(lambda x : clean_text(x))
```

```
train_sentences = df["comment_text"].fillna("CVxTz").values
list_classes = ["toxic", "severe_toxic", "obscene", "threat", "insult", "identity_hate"]
train_y = df[list_classes].values
```

Figure 23 :- Data preprocessed using regex functions

Raw: Explanation/n Why the edits made under my username Hardcore Metallica Fan were reverted? They weren't

Raw: **e**xplanation Why the edits made under my username hardcore **m**etallica **f**an were reverted? They **were not**

Figure 24 :- Preprocessed Sentence Example

The pre-processed data set is then put via the Tokenizer for WordPiece, which works by splitting words into complete forms (e.g., one word becomes one token) or word pieces (e.g., one word can be broken into many tokens). We already know that the terms "surfboard" and "snowboard" share meaning through the wordpiece "##board" since we divide words into word pieces. There are a few key parameters to remember during startup, which are as follows:

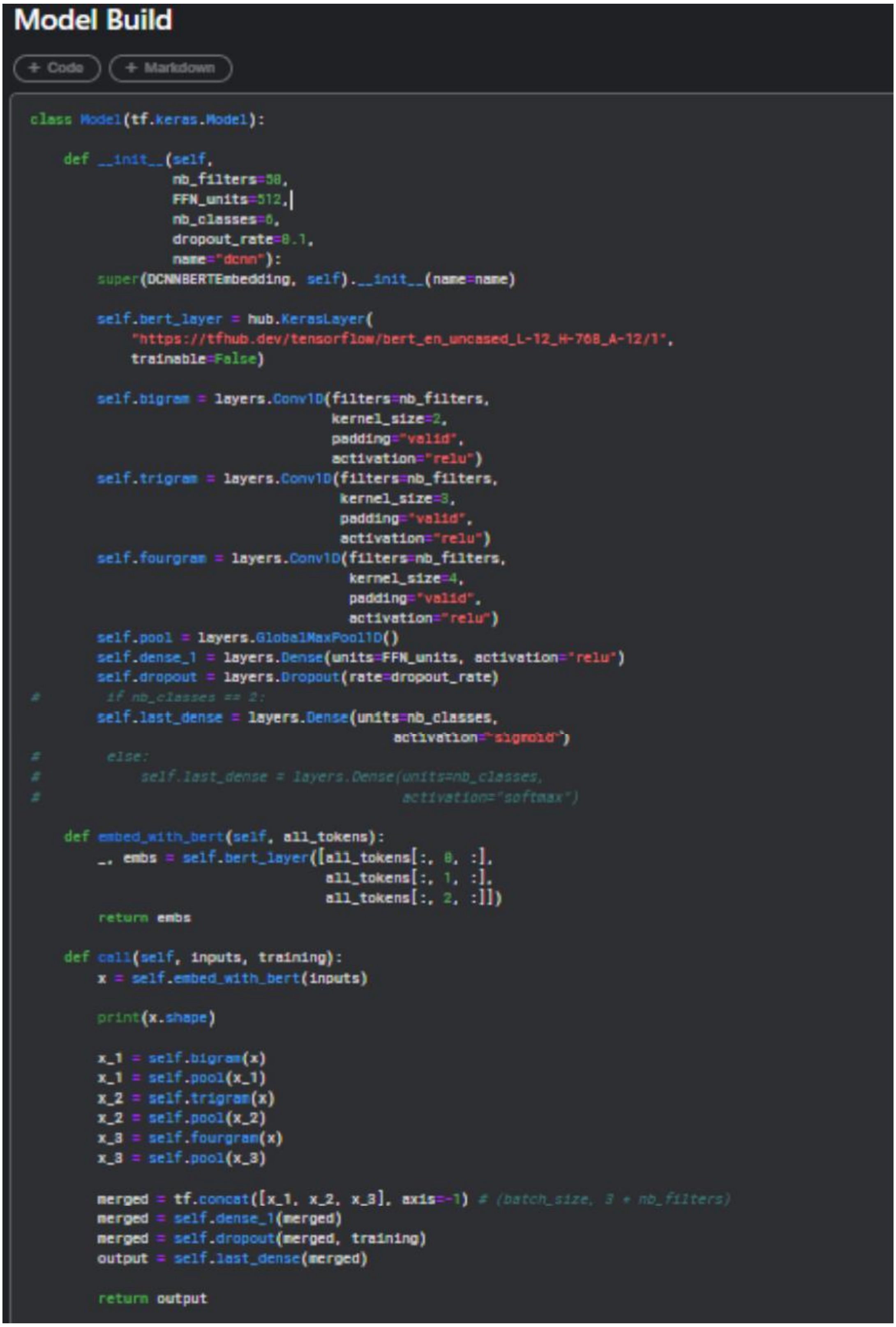
- Input IDs: The input ids are frequently the sole parameters that must be supplied to the model as input. They are token indices, which are numerical representations of the tokens that make up the sequences that the model will utilize as input.

- The attention mask: The attention mask is a binary tensor that indicates where the padded indices should be placed so that the model does not pay to them. For the BertTokenizer, a value of 1 suggests that it should be attended to, whereas a value of 0 indicates that it should be padded. This attention mask is found in the tokenizer's dictionary under the key "attention mask":

- The token type IDs: A single vector containing the whole input sentence must be supplied to a classifier for the classification operation. In BERT, the judgment is made that the concealed state of the initial token represents the entire phrase. To do this, an extra token must be manually inserted to the input phrase. The token [CLS] is used for this purpose in the original implementation. We need a mechanism to tell the model where the first sentence ends and the second sentence begins in the "next sentence prediction" task. As a result, another fictitious token, [SEP], is introduced. If we want to train a classifier, each input sample will only include one sentence (or a single text input). The [SEP] token will be appended to the end of the input text in that case. To summarize, the first step in pre-processing the input text data is to add the [CLS] token at the beginning and the [SEP] token at the end of each input text.

- Padding Tokens: As input, the BERT model is given a fixed length phrase. The maximum length of a phrase is usually determined by the data we are dealing with. To compensate for sentences that are less than this maximum length, paddings (empty tokens) will be added to the sentences. In the original version, the token [PAD] is used to represent sentence paddings.

Tokenization is thus required since it plays an important function in NLP by converting text to numbers that deep learning models may employ for processing. To put it another way, tokenizers aid in comprehending the context or constructing the model for NLP. Tokenization aids in determining the meaning of the text by evaluating the word sequence. Finally, we must setup the model to our specifications and feed the tokenized dataset through the model.

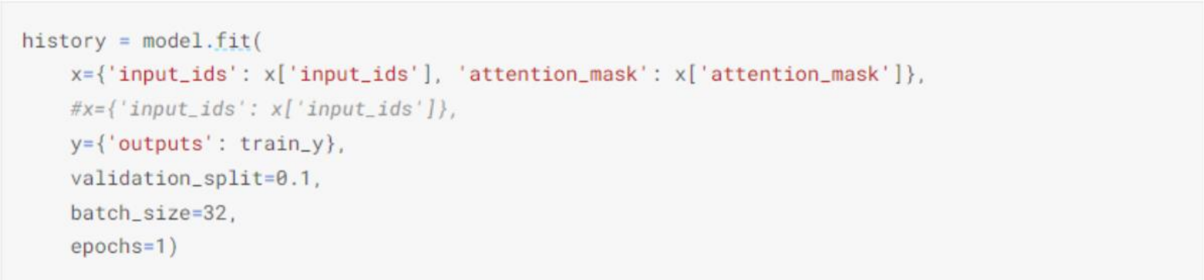


Model Build

```
class Model(tf.keras.Model):

    def __init__(self,
                 nb_filters=50,
                 FFN_units=512,
                 nb_classes=6,
                 dropout_rate=0.1,
                 name="dcnn"):
        super(DCNNBERTEmbedding, self).__init__(name=name)

        self.bert_layer = hub.KerasLayer(
            "https://tfhub.dev/tensorflow/bert_en_uncased_L-12_H-768_A-12/1",
            trainable=False)

        self.bigram = layers.Conv1D(filters=nb_filters,
                                    kernel_size=2,
                                    padding="valid",
                                    activation="relu")
        self.trigram = layers.Conv1D(filters=nb_filters,
                                     kernel_size=3,
                                     padding="valid",
                                     activation="relu")
        self.fourgram = layers.Conv1D(filters=nb_filters,
                                      kernel_size=4,
                                      padding="valid",
                                      activation="relu")
        self.pool = layers.GlobalMaxPool1D()
        self.dense_1 = layers.Dense(units=FFN_units, activation="relu")
        self.dropout = layers.Dropout(rate=dropout_rate)
#        if nb_classes == 2:
        self.last_dense = layers.Dense(units=nb_classes,
                                            activation="sigmoid")
#        else:
#            self.last_dense = layers.Dense(units=nb_classes,
#                                            activation="softmax")

    def embed_with_bert(self, all_tokens):
        _, embs = self.bert_layer([all_tokens[:, 0, :],
                                   all_tokens[:, 1, :],
                                   all_tokens[:, 2, :]])

        return embs

    def call(self, inputs, training):
        x = self.embed_with_bert(inputs)

        print(x.shape)

        x_1 = self.bigram(x)
        x_1 = self.pool(x_1)
        x_2 = self.trigram(x)
        x_2 = self.pool(x_2)
        x_3 = self.fourgram(x)
        x_3 = self.pool(x_3)

        merged = tf.concat([x_1, x_2, x_3], axis=-1) # (batch_size, 3 * nb_filters)
        merged = self.dense_1(merged)
        merged = self.dropout(merged, training)
        output = self.last_dense(merged)

        return output
```

Figure 25:- Implementation I

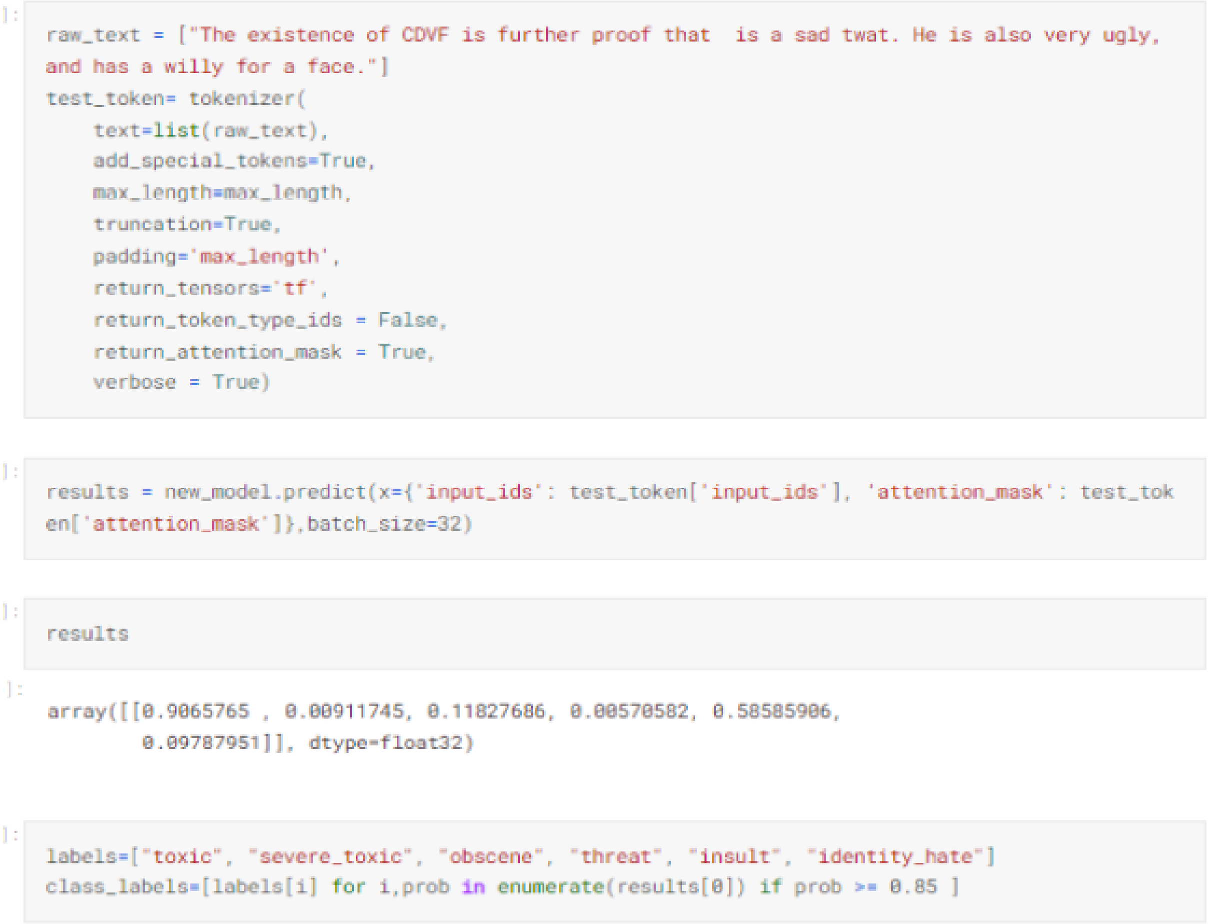
As indicated in the diagram above, our model already has 12 layers of classification, but in order to achieve high accuracy, we added four more layers of classification using text classification algorithms like as relu and sigmoid for in-depth classification. After saving the model configurations, the dataset is sent through the model and trained, and the model's accuracy is 93%.



```
history = model.fit(
    x={'input_ids': x['input_ids'], 'attention_mask': x['attention_mask']},
    #x={'input_ids': x['input_ids']},
    y={'outputs': train_y},
    validation_split=0.1,
    batch_size=32,
    epochs=1)
```

Figure 26:- Implementation II

Finally, once the training process is complete, a raw text is fed through the model to predict the result, as illustrated below:



```
]:  raw_text = ["The existence of CDVF is further proof that  is a sad twat. He is also very ugly,
    and has a willy for a face."]
    test_token= tokenizer(
        text=list(raw_text),
        add_special_tokens=True,
        max_length=max_length,
        truncation=True,
        padding='max_length',
        return_tensors='tf',
        return_token_type_ids = False,
        return_attention_mask = True,
        verbose = True)
```

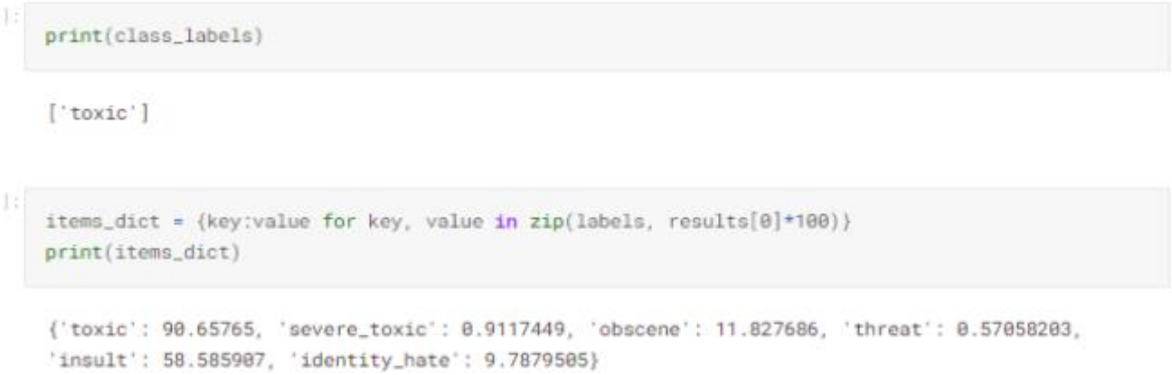
```
]:  results = new_model.predict(x={'input_ids': test_token['input_ids'], 'attention_mask': test_tok
    en['attention_mask']},batch_size=32)
```

```
]:  results
```

```
]:  array([[0.9065765 , 0.00911745, 0.11827686, 0.00570582, 0.58585906,
            0.09787951]], dtype=float32)
```

```
]:  labels=["toxic", "severe_toxic", "obscene", "threat", "insult", "identity_hate"]
    class_labels=[labels[i] for i,prob in enumerate(results[0]) if prob >= 0.85 ]
```

Figure 27:- Implementation III



```
print(class_labels)
```

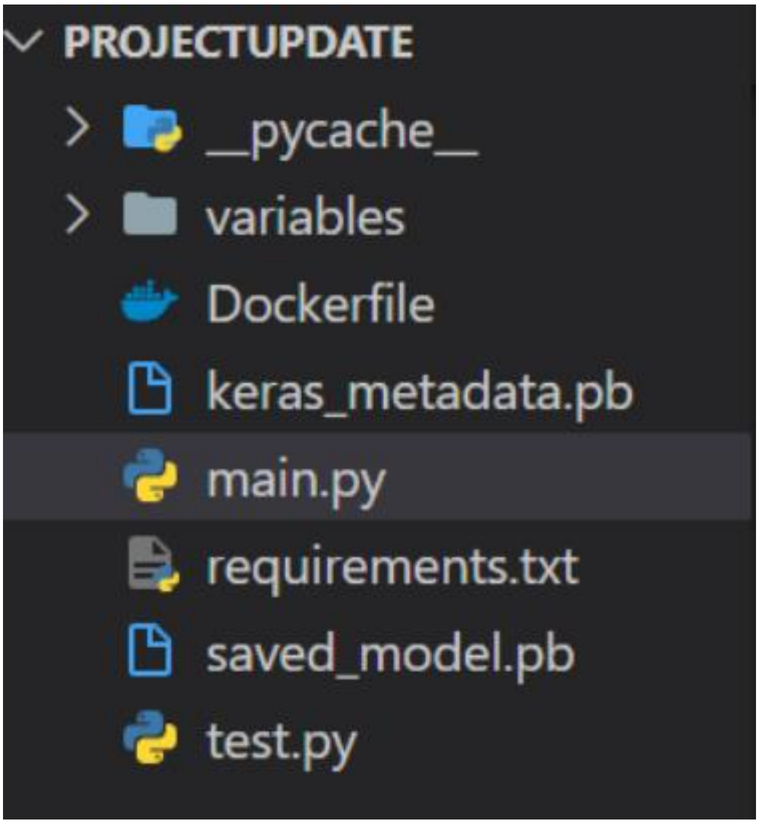
```
['toxic']
```

```
items_dict = {key:value for key, value in zip(labels, results[0]*100)}
print(items_dict)
```

```
{'toxic': 90.65765, 'severe_toxic': 0.9117449, 'obscene': 11.827686, 'threat': 0.57058203,
'insult': 58.585907, 'identity_hate': 9.7879505}
```

Figure 28:- Implementation IV

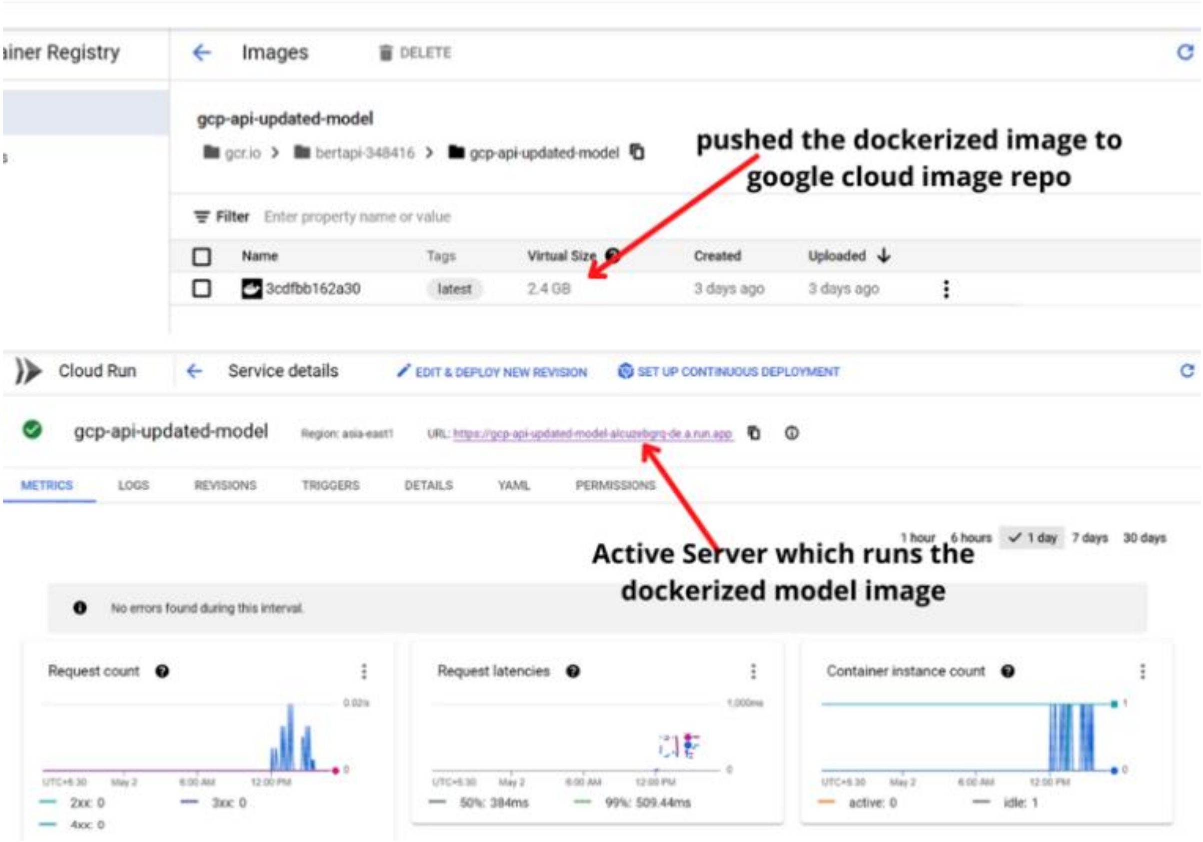
The raw text is tokenized and fed into the model, which correctly predicts the input as "hazardous." Now that the model is complete, it is provided on a fastapi server by retrieving the stored model file and loading it to the server in the correct sequence. Afterwards, the entire package is dockerized in order to build a virtual instance with all of the model's imports, dependencies, and OS needs, and these are the files in the dockerized image as follows:



```
PROJECTUPDATE
  > __pycache__
  > variables
    Dockerfile
    keras_metadata.pb
    main.py
    requirements.txt
    saved_model.pb
    test.py
```

Figure 29 :- Implementation IV

Finally, a dockerized image is deployed as a microservice in the Google Cloud Platform.



gcp-api-updated-model

gcr.io > bertapi-348416 > gcp-api-updated-model

□	Name	Tags	Virtual Size	Created	Uploaded ↓
□	3cdfbb162a30	latest	2.4 GB	3 days ago	3 days ago

pushed the dockerized image to google cloud image repo

gcp-api-updated-model Region: asia-east1 URL: https://gcp-api-updated-model-alcuzebgrq-de.a.run.app

Active Server which runs the dockerized model image

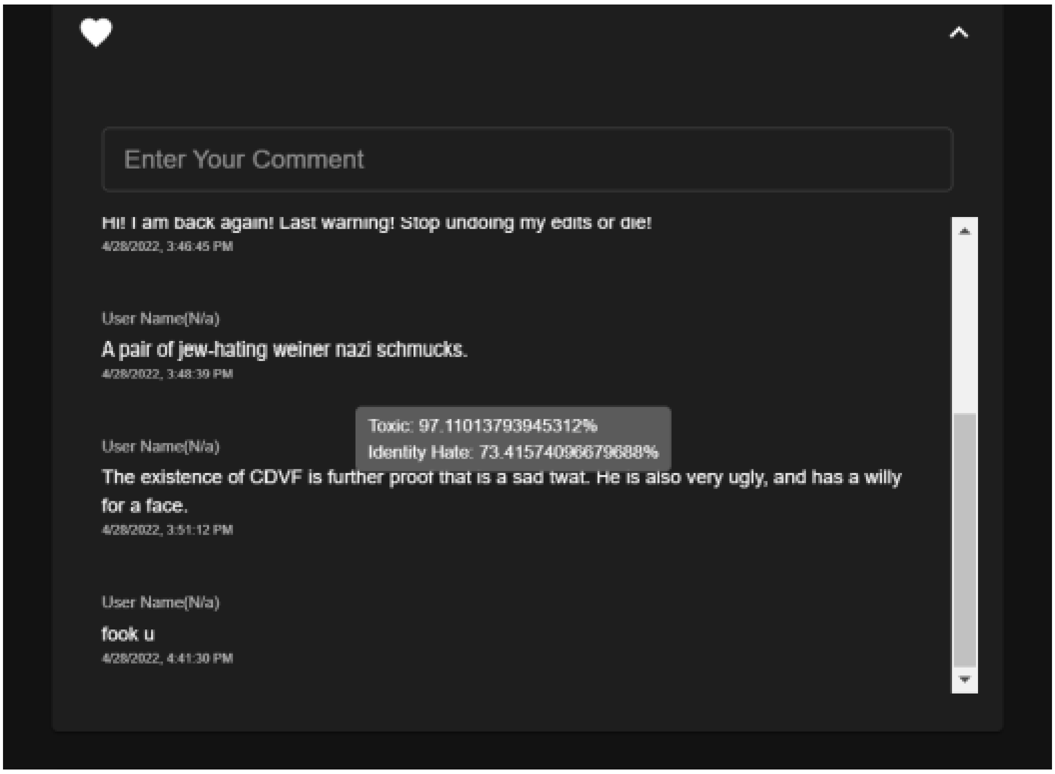
No errors found during this interval.

2xx: 0 3xx: 0 4xx: 0

50%: 384ms 99%: 509.44ms

active: 0 idle: 1

Output shown in the application for demonstration,



Enter Your Comment

Hi! I am back again! Last warning! Stop undoing my edits or die!
4/28/2022, 3:46:45 PM

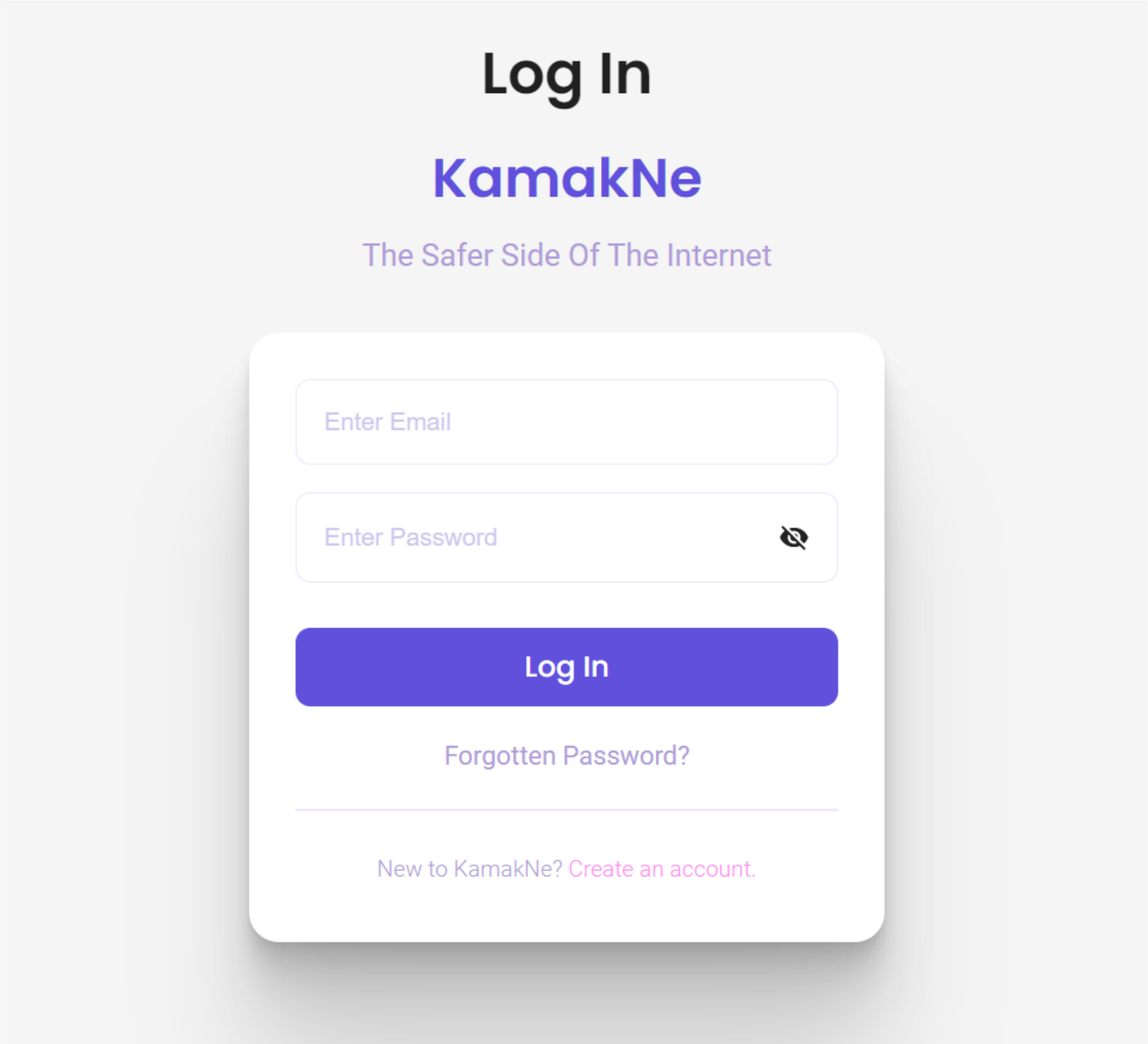
User Name(N/a)
A pair of jew-hating weiner nazi schmucks.
4/28/2022, 3:48:39 PM

Toxic: 97.11013793945312%
Identity Hate: 73.41574096679688%

User Name(N/a)
The existence of CDVF is further proof that is a sad twat. He is also very ugly, and has a willy for a face.
4/28/2022, 3:51:12 PM

User Name(N/a)
fook u
4/28/2022, 4:41:30 PM

Figure 30:- Implementation VI

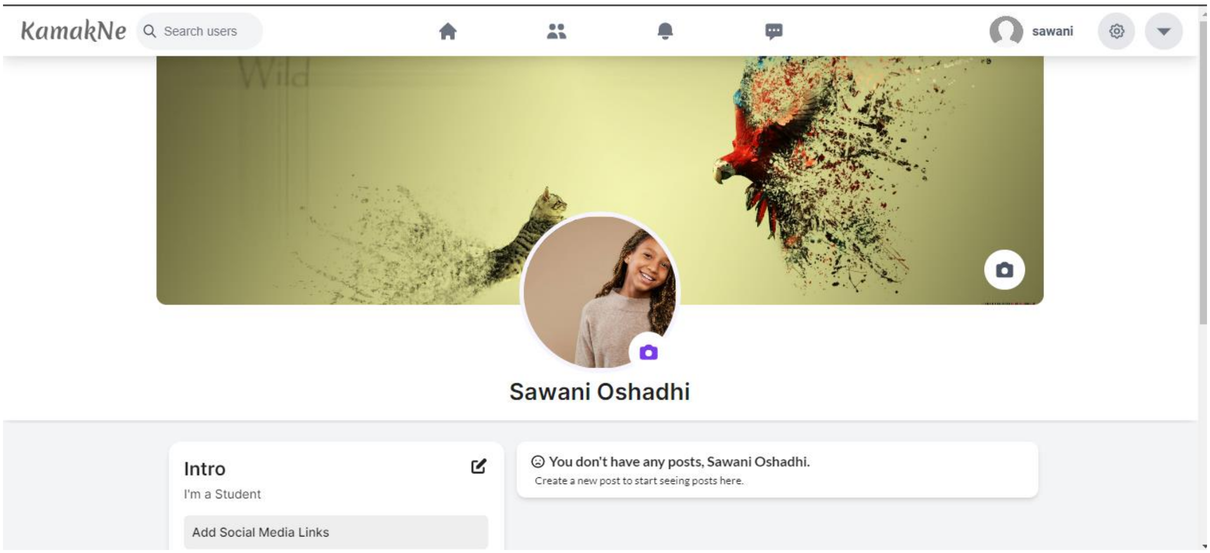

Log In

KamakNe

The Safer Side Of The Internet

New to KamakNe? Create an account.

Figure 31 :- Login for Social Media App Model



Sawani Oshadhi

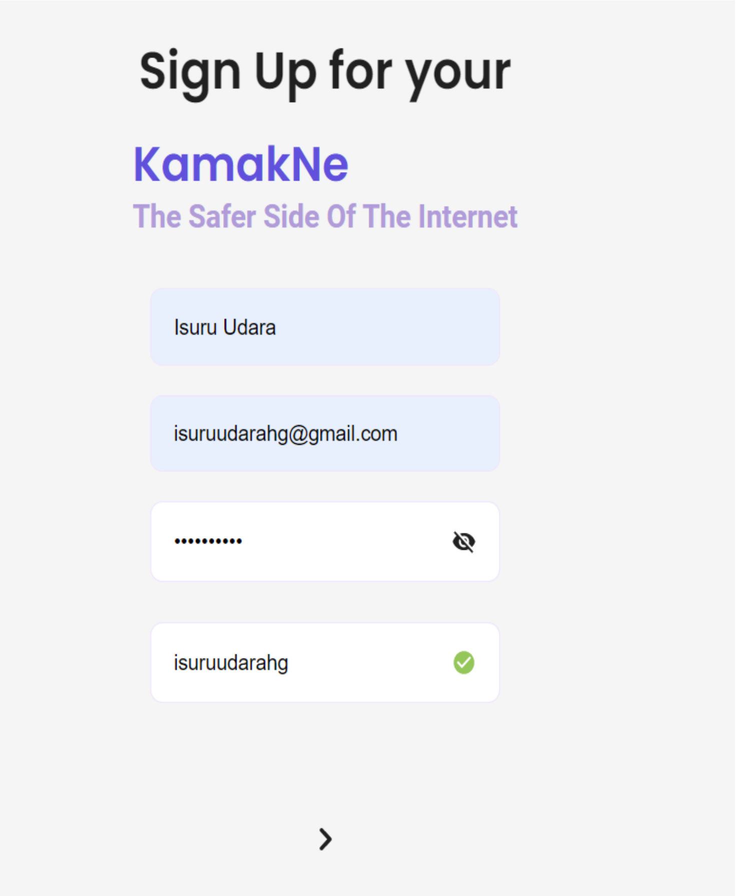
Intro

I'm a Student

You don't have any posts, Sawani Oshadhi.

Create a new post to start seeing posts here.

Figure 32:- Account Profile in Social Media App Model



Sign Up for your

KamakNe

The Safer Side Of The Internet

Isuru Udara

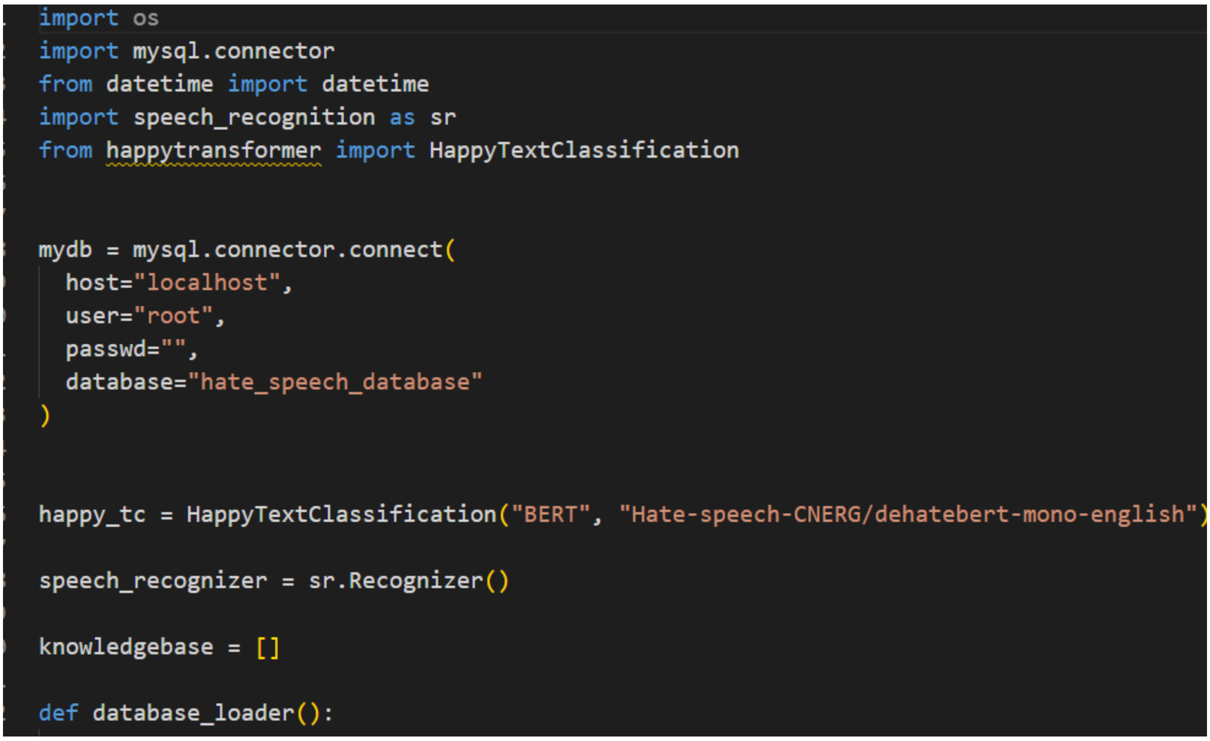
isuruudarahg@gmail.com

••••••••••

isuruudarahg ✓

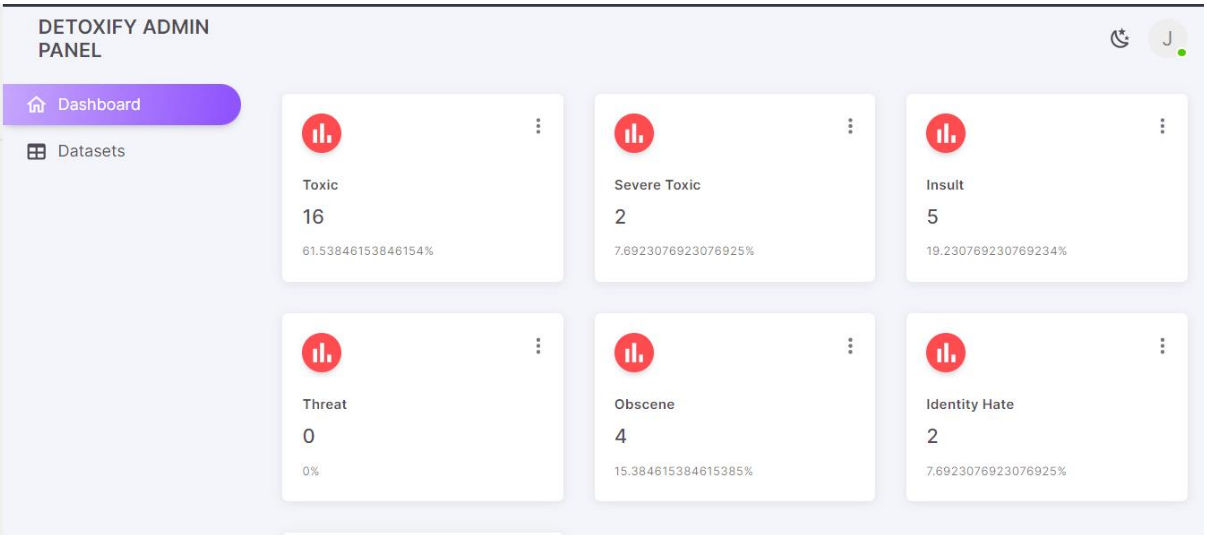
>

Figure 33:- Signup for Social Media App



```
import os
import mysql.connector
from datetime import datetime
import speech_recognition as sr
from happytransformer import HappyTextClassification


mydb = mysql.connector.connect(
  host="localhost",
  user="root",
  passwd="",
  database="hate_speech_database"
)


happy_tc = HappyTextClassification("BERT", "Hate-speech-CNERG/dehatebert-mono-english")

speech_recognizer = sr.Recognizer()

knowledgebase = []

def database_loader():
```

Figure 34:- Audio Processing using Happy transform.



Toxic
16
61.53846153846154%

Severe Toxic
2
7.6923076923076925%

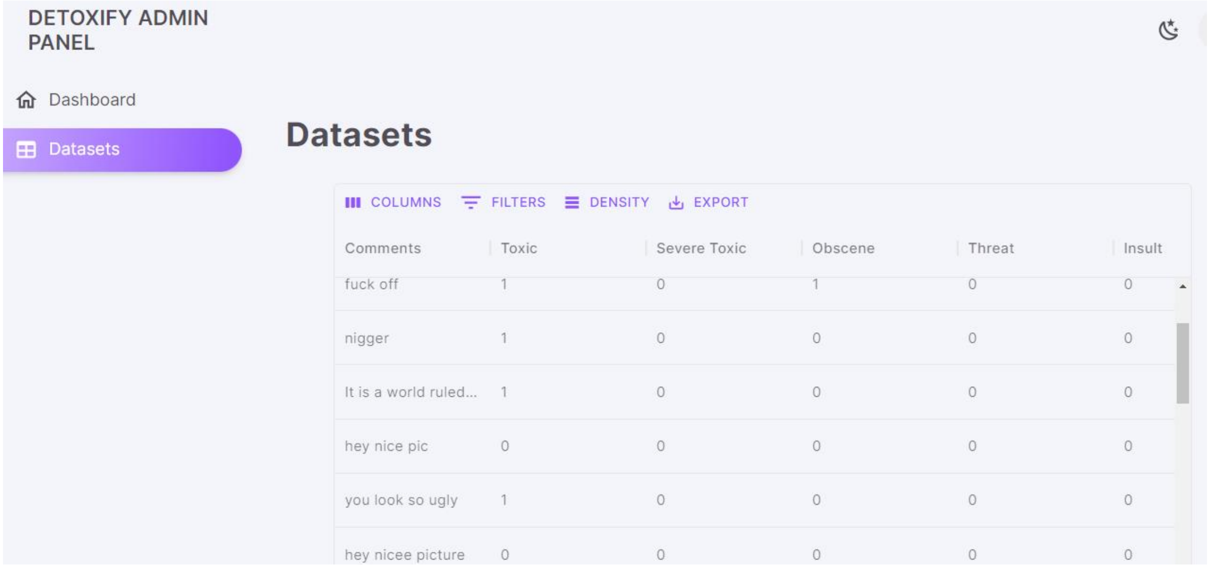
Insult
5
19.230769230769234%

Threat
0
0%

Obscene
4
15.384615384615385%

Identity Hate
2
7.6923076923076925%

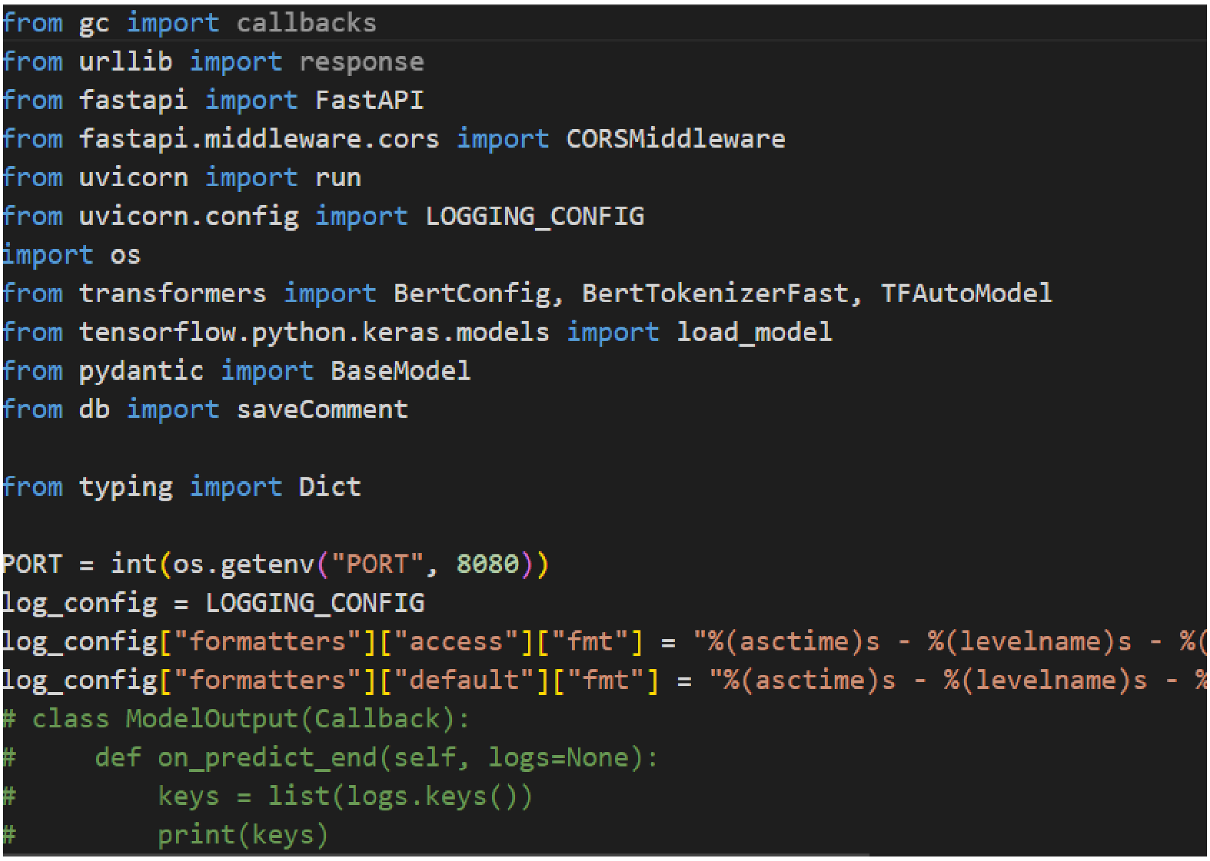
Figure 35:- Detoxify Admin Panel



Datasets

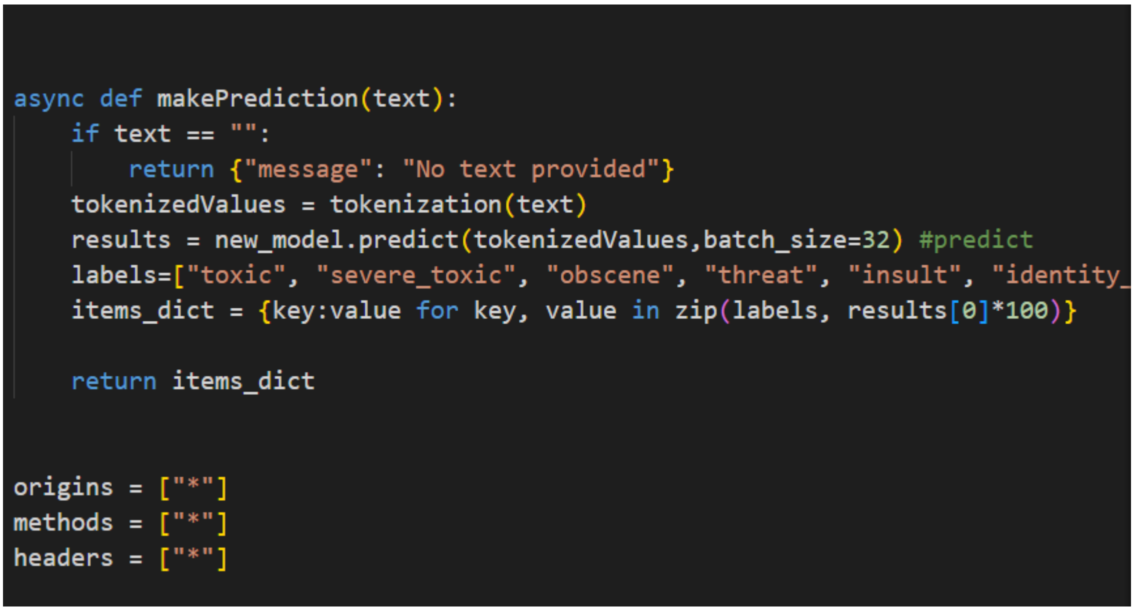
Comments	Toxic	Severe Toxic	Obscene	Threat	Insult
fuck off	1	0	1	0	0
nigger	1	0	0	0	0
It is a world ruled...	1	0	0	0	0
hey nice pic	0	0	0	0	0
you look so ugly	1	0	0	0	0
hey nicee picture	0	0	0	0	0

Figure 36:- Analysis of Comments



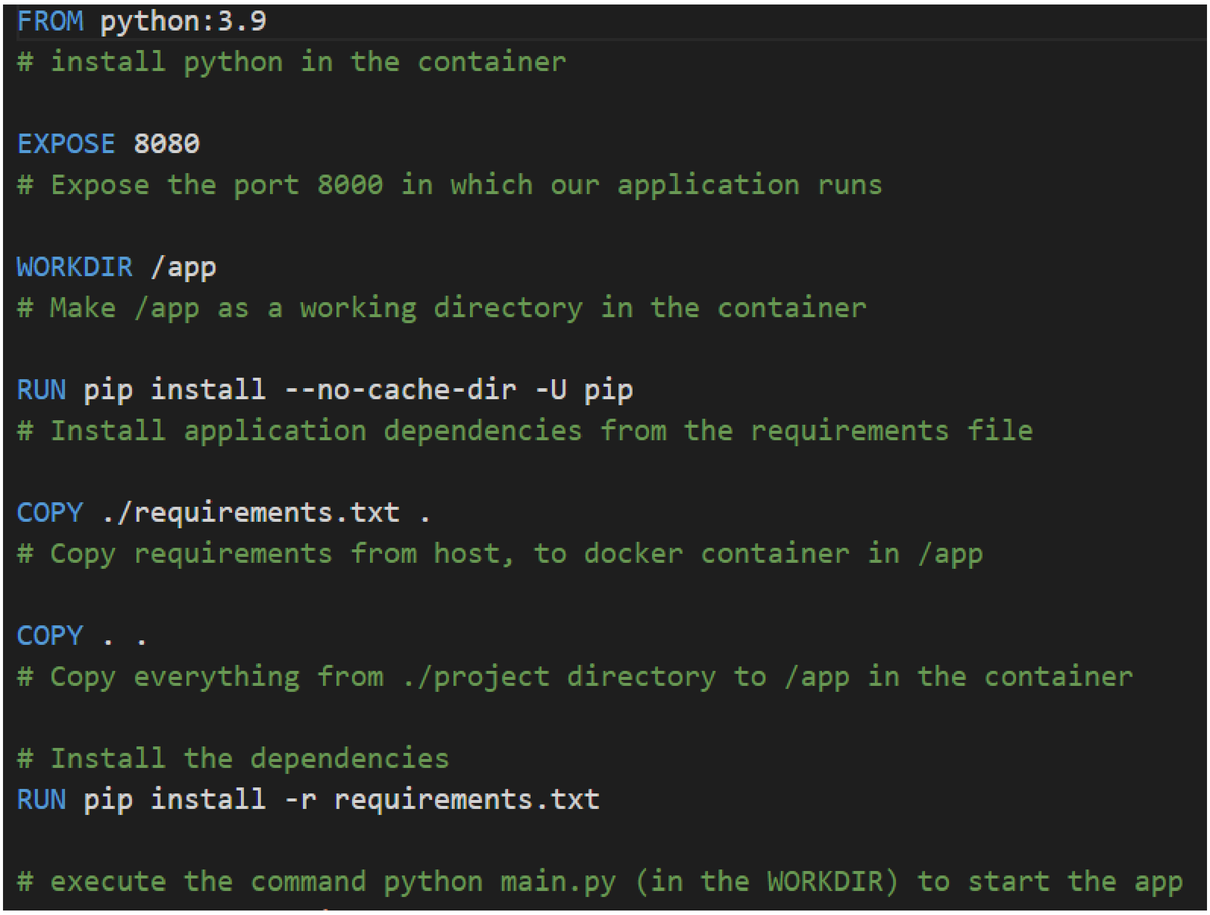
```
from gc import callbacks
from urllib import response
from fastapi import FastAPI
from fastapi.middleware.cors import CORSMiddleware
from uvicorn import run
from uvicorn.config import LOGGING_CONFIG
import os
from transformers import BertConfig, BertTokenizerFast, TFAutoModel
from tensorflow.python.keras.models import load_model
from pydantic import BaseModel
from db import saveComment

from typing import Dict

PORT = int(os.getenv("PORT", 8080))
log_config = LOGGING_CONFIG
log_config["formatters"]["access"]["fmt"] = "%(asctime)s - %(levelname)s - %(
log_config["formatters"]["default"]["fmt"] = "%(asctime)s - %(levelname)s - %
# class ModelOutput(Callback):
#     def on_predict_end(self, logs=None):
#         keys = list(logs.keys())
#         print(keys)
```

Figure 37 :- BERT Config



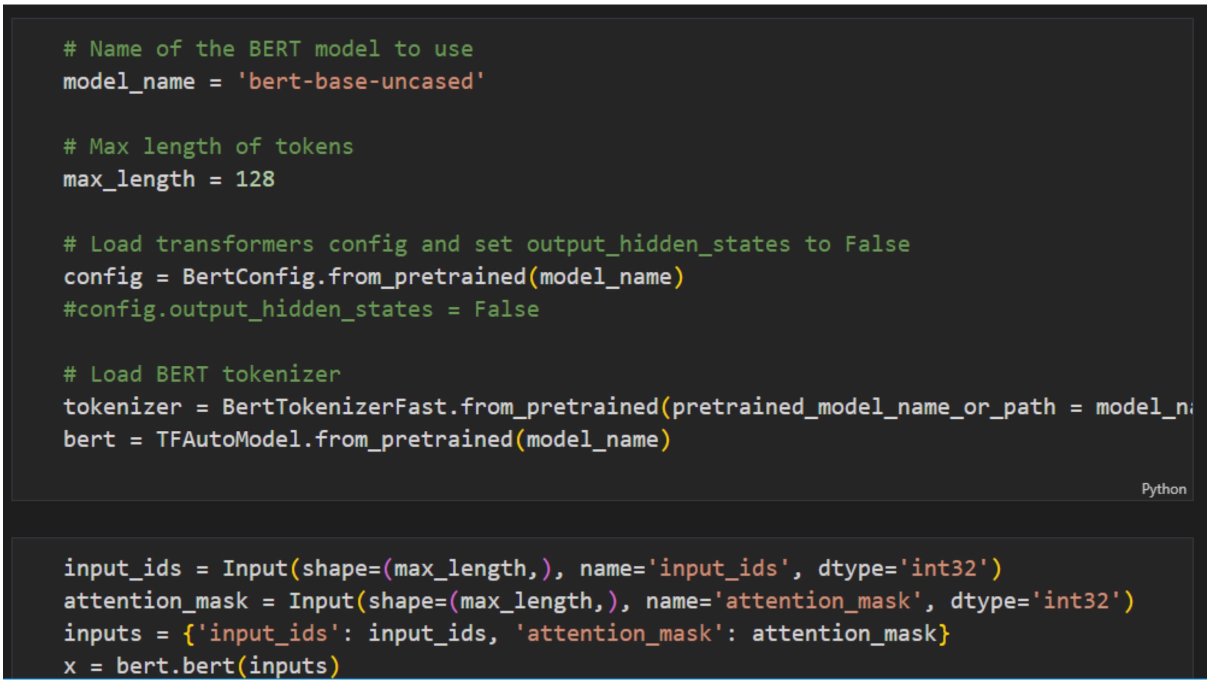
```
async def makePrediction(text):
    if text == "":
        return {"message": "No text provided"}
    tokenizedValues = tokenization(text)
    results = new_model.predict(tokenizedValues,batch_size=32) #predict
    labels=["toxic", "severe_toxic", "obscene", "threat", "insult", "identity_
    items_dict = {key:value for key, value in zip(labels, results[0]*100)}

    return items_dict


origins = ["*"]
methods = ["*"]
headers = ["*"]
```

Figure 38 :- :- BERT Config 2



```
FROM python:3.9
# install python in the container

EXPOSE 8080
# Expose the port 8000 in which our application runs

WORKDIR /app
# Make /app as a working directory in the container

RUN pip install --no-cache-dir -U pip
# Install application dependencies from the requirements file

COPY ./requirements.txt .
# Copy requirements from host, to docker container in /app

COPY . .
# Copy everything from ./project directory to /app in the container

# Install the dependencies
RUN pip install -r requirements.txt

# execute the command python main.py (in the WORKDIR) to start the app
```

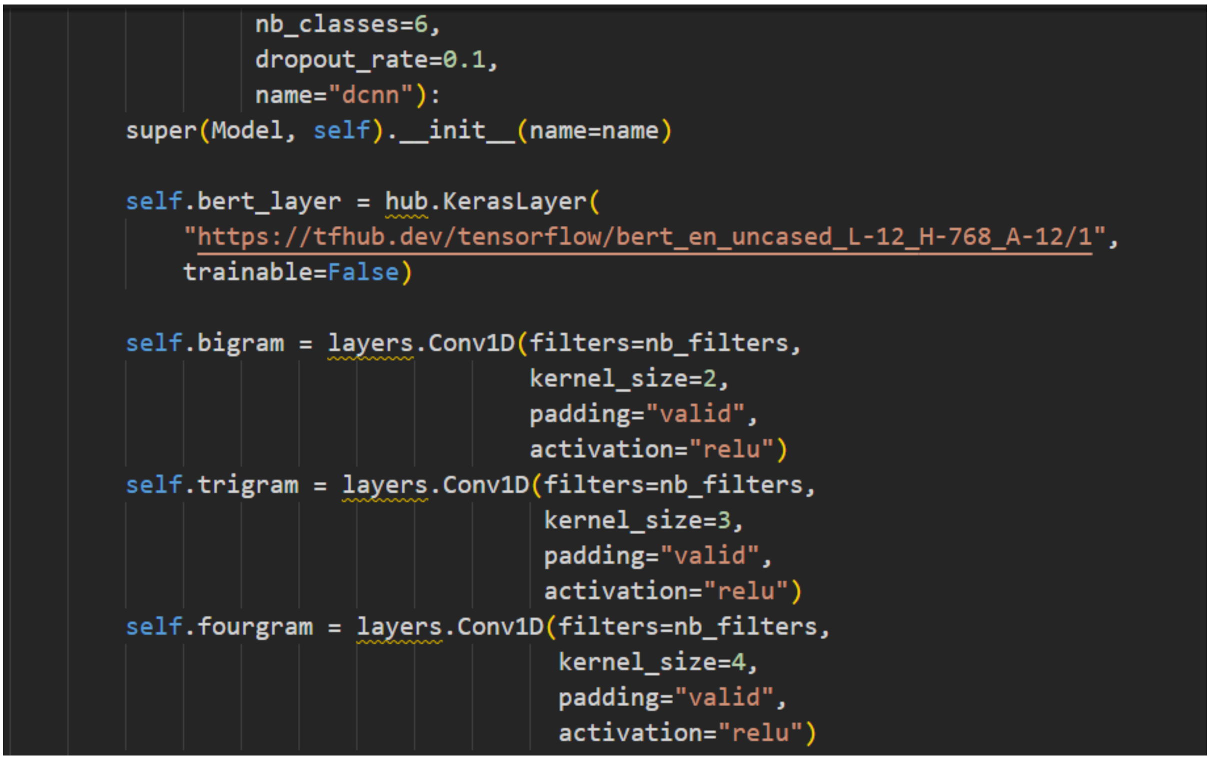
Figure 39:- Docker File



```
# Name of the BERT model to use
model_name = 'bert-base-uncased'

# Max length of tokens
max_length = 128

# Load transformers config and set output_hidden_states to False
config = BertConfig.from_pretrained(model_name)
#config.output_hidden_states = False

# Load BERT tokenizer
tokenizer = BertTokenizerFast.from_pretrained(pretrained_model_name_or_path = model_na
bert = TFAutoModel.from_pretrained(model_name)
```

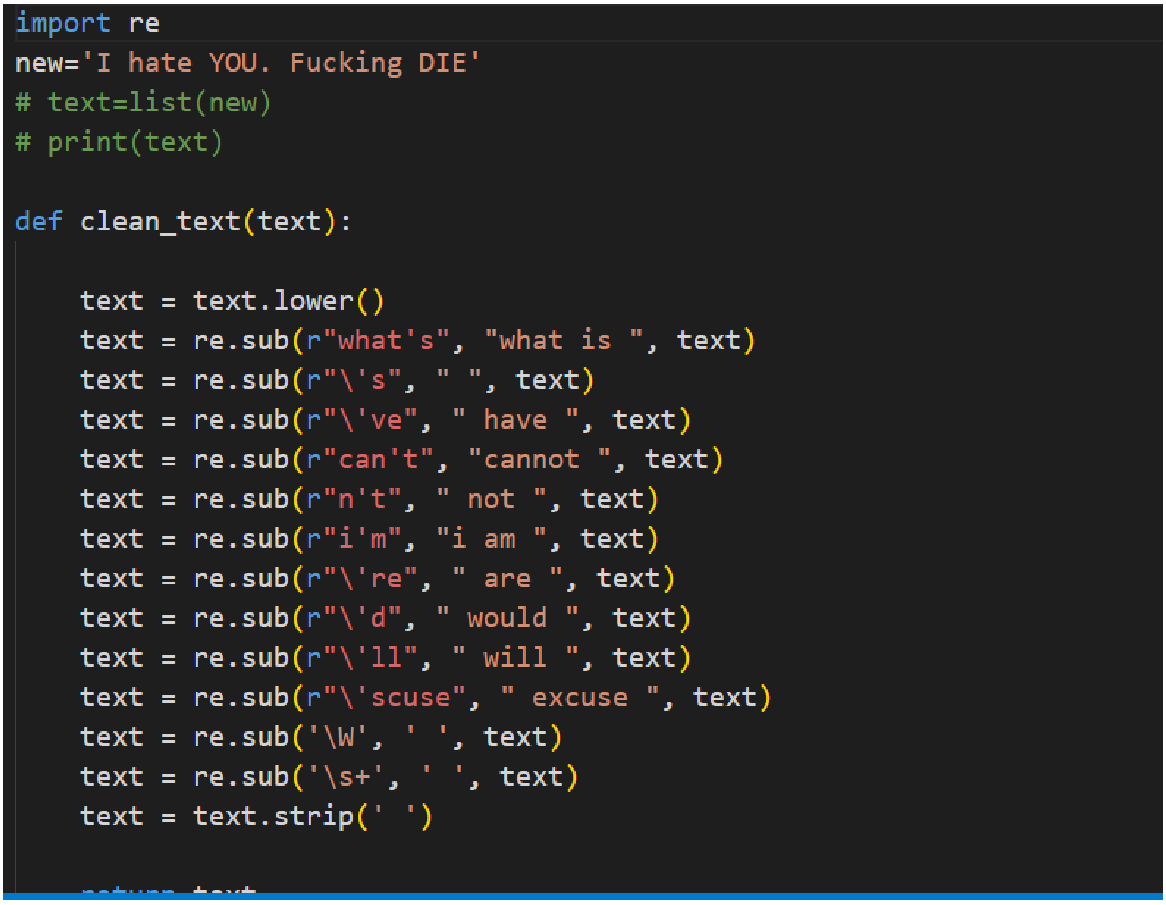
```
input_ids = Input(shape=(max_length,), name='input_ids', dtype='int32')
attention_mask = Input(shape=(max_length,), name='attention_mask', dtype='int32')
inputs = {'input_ids': input_ids, 'attention_mask': attention_mask}
x = bert.bert(inputs)
```

Figure 40:- Toxic Comments Config BERT



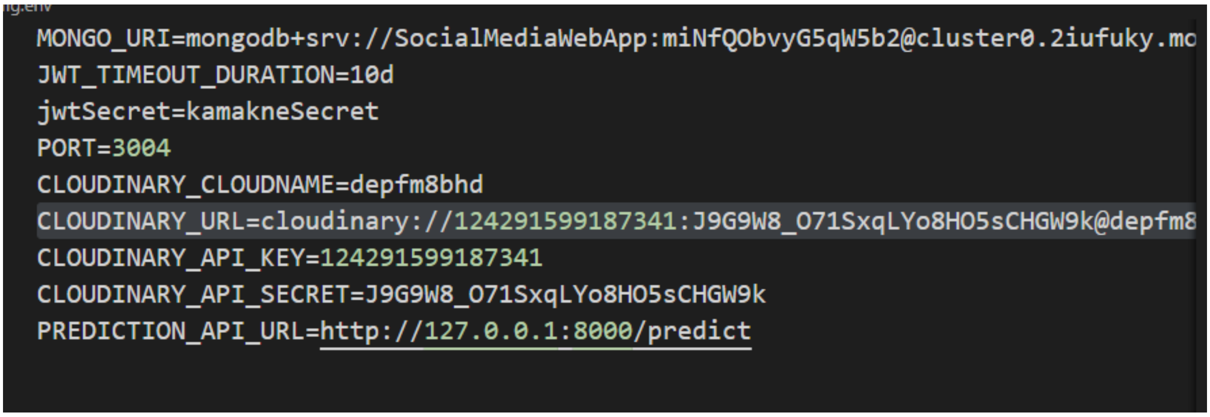
```
                 nb_classes=6,
                 dropout_rate=0.1,
                 name="dcnn"):
        super(Model, self).__init__(name=name)

        self.bert_layer = hub.KerasLayer(
            "https://tfhub.dev/tensorflow/bert_en_uncased_L-12_H-768_A-12/1",
            trainable=False)

        self.bigram = layers.Conv1D(filters=nb_filters,
                                    kernel_size=2,
                                    padding="valid",
                                    activation="relu")
        self.trigram = layers.Conv1D(filters=nb_filters,
                                     kernel_size=3,
                                     padding="valid",
                                     activation="relu")
        self.fourgram = layers.Conv1D(filters=nb_filters,
                                      kernel_size=4,
                                      padding="valid",
                                      activation="relu")
```

Figure 41 :- Toxic Comments Config BERT 2



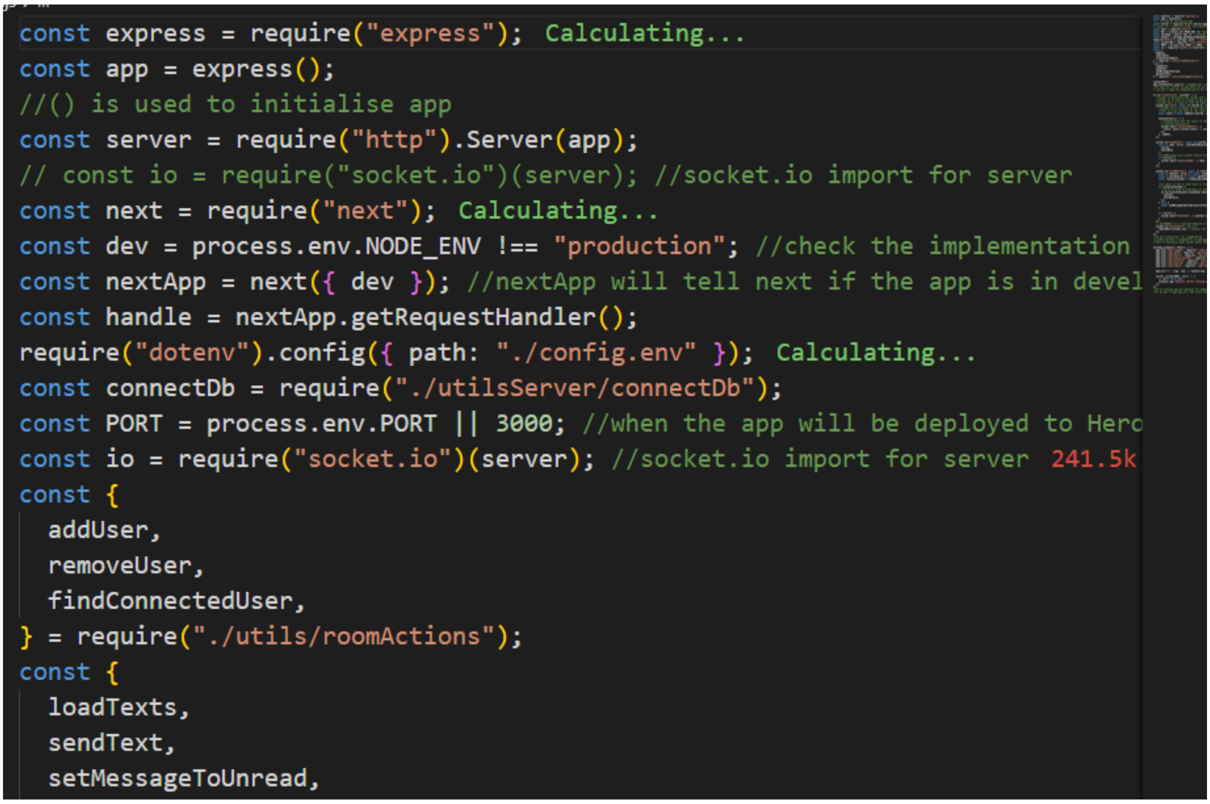
```
import re
new='I hate YOU. Fucking DIE'
# text=list(new)
# print(text)

def clean_text(text):

    text = text.lower()
    text = re.sub(r"what's", "what is ", text)
    text = re.sub(r"\'s", " ", text)
    text = re.sub(r"\'ve", " have ", text)
    text = re.sub(r"can't", "cannot ", text)
    text = re.sub(r"n't", " not ", text)
    text = re.sub(r"i'm", "i am ", text)
    text = re.sub(r"\'re", " are ", text)
    text = re.sub(r"\'d", " would ", text)
    text = re.sub(r"\'ll", " will ", text)
    text = re.sub(r"\'scuse", " excuse ", text)
    text = re.sub('\W', ' ', text)
    text = re.sub('\s+', ' ', text)
    text = text.strip(' ')

    return text
```

Figure 42:- Test Prediction



```
MONGO_URI=mongodb+srv://SocialMediaWebApp:miNfQObvyG5qW5b2@cluster0.2iufuky.mo
JWT_TIMEOUT_DURATION=10d
jwtSecret=kamakneSecret
PORT=3004
CLOUDINARY_CLOUDNAME=depfm8bhd
CLOUDINARY_URL=cloudinary://124291599187341:J9G9W8_O71SxqLYo8HO5sCHGW9k@depfm8
CLOUDINARY_API_KEY=124291599187341
CLOUDINARY_API_SECRET=J9G9W8_O71SxqLYo8HO5sCHGW9k
PREDICTION_API_URL=http://127.0.0.1:8000/predict
```

Figure 43:- Config file for Social Media Application



```
const express = require("express");  Calculating...
const app = express();
//() is used to initialise app
const server = require("http").Server(app);
// const io = require("socket.io")(server); //socket.io import for server
const next = require("next");  Calculating...
const dev = process.env.NODE_ENV !== "production"; //check the implementation
const nextApp = next({ dev }); //nextApp will tell next if the app is in devel
const handle = nextApp.getRequestHandler();
require("dotenv").config({ path: "./config.env" });  Calculating...
const connectDb = require("./utilsServer/connectDb");
const PORT = process.env.PORT || 3000; //when the app will be deployed to Hero
const io = require("socket.io")(server); //socket.io import for server  241.5k
const {
  addUser,
  removeUser,
  findConnectedUser,
} = require("./utils/roomActions");
const {
  loadTexts,
  sendText,
  setMessageToUnread,
```

Figure 44:- Server.js File

4.11. Testing and Evaluation

4.11.1. Test Plan

The following approaches are done in order to test the model and fine tune it is depending on the outcome of the tests

- Divide Test Cases by components
- Test Predictions by label
- Retrain model depending on the accuracy until the expected results are obtained
- Follow normal testing procedures to test the secondary component

The following testing's were performed after the implementation of the AI Model and the Social Media Web Application.

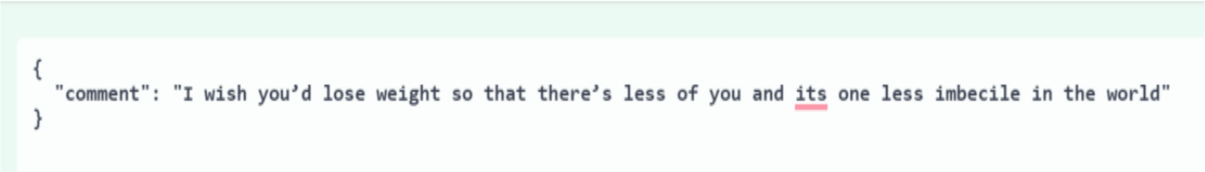
AI Model

The test inputs that are used to test the model are scrapped thorough reddit forums in order to get more original data which is closer to natural language.

The testing's performed below are to find out if the predictions were accurate enough under every label provided,

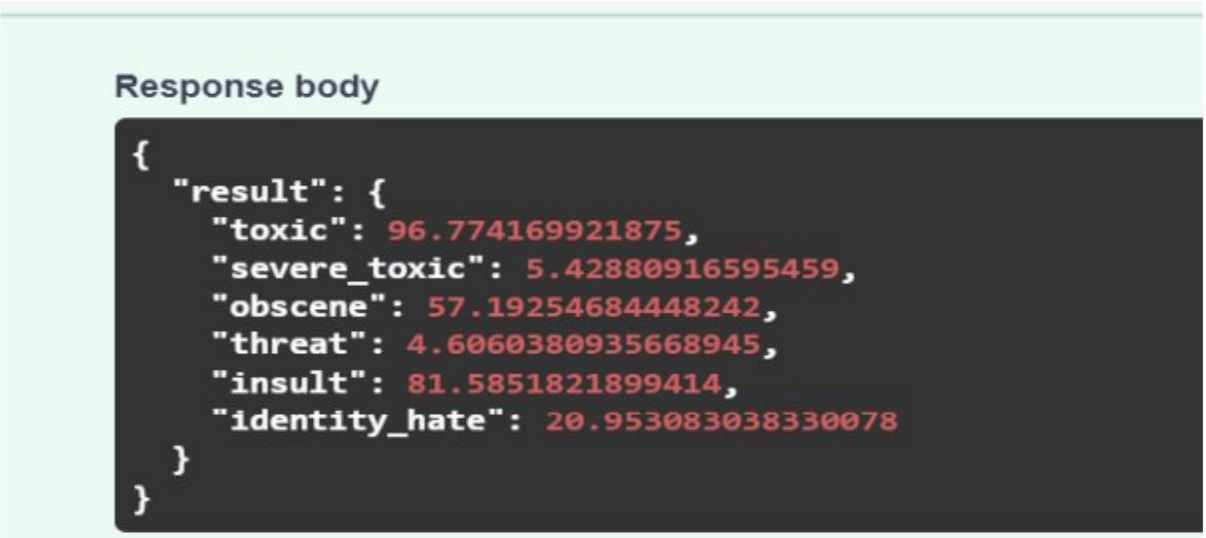
1. Label - Toxic

Comment



```
{
  "comment": "I wish you'd lose weight so that there's less of you and its one less imbecile in the world"
}
```

Prediction



Response body

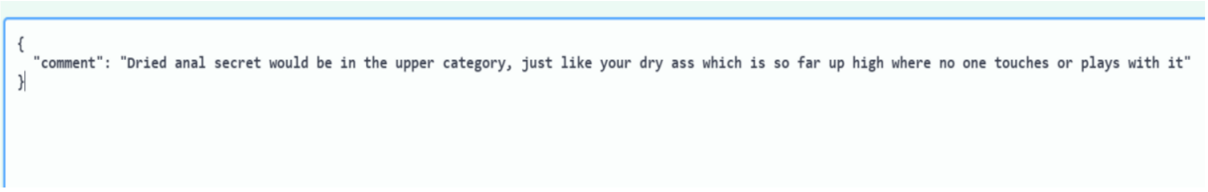
```
{
  "result": {
    "toxic": 96.774169921875,
    "severe_toxic": 5.42880916595459,
    "obscene": 57.19254684448242,
    "threat": 4.6060380935668945,
    "insult": 81.5851821899414,
    "identity_hate": 20.953083038330078
  }
}
```

Figure 45:- Toxic Prediction

Despite the figure 28, the approximate detected toxicity level is 97%.

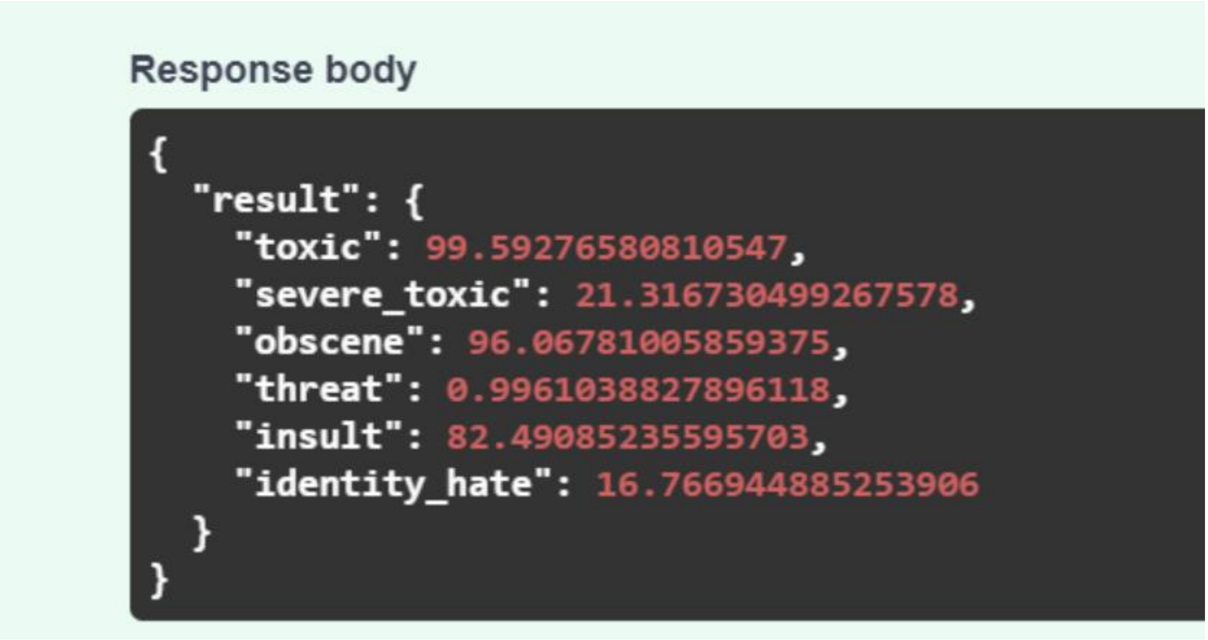
2. Label - Obscene

Comment



```
{
  "comment": "Dried anal secret would be in the upper category, just like your dry ass which is so far up high where no one touches or plays with it"
}
```

Prediction



Response body

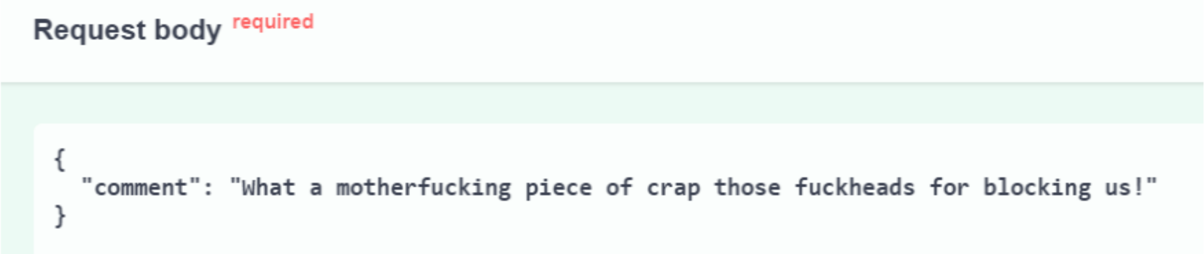
```
{
  "result": {
    "toxic": 99.59276580810547,
    "severe_toxic": 21.316730499267578,
    "obscene": 96.06781005859375,
    "threat": 0.9961038827896118,
    "insult": 82.49085235595703,
    "identity_hate": 16.766944885253906
  }
}
```

Figure 46:- Obscene Prediction

Despite the figure 29, the approximate obscene level is 96%.

3. Label - Severe Toxic

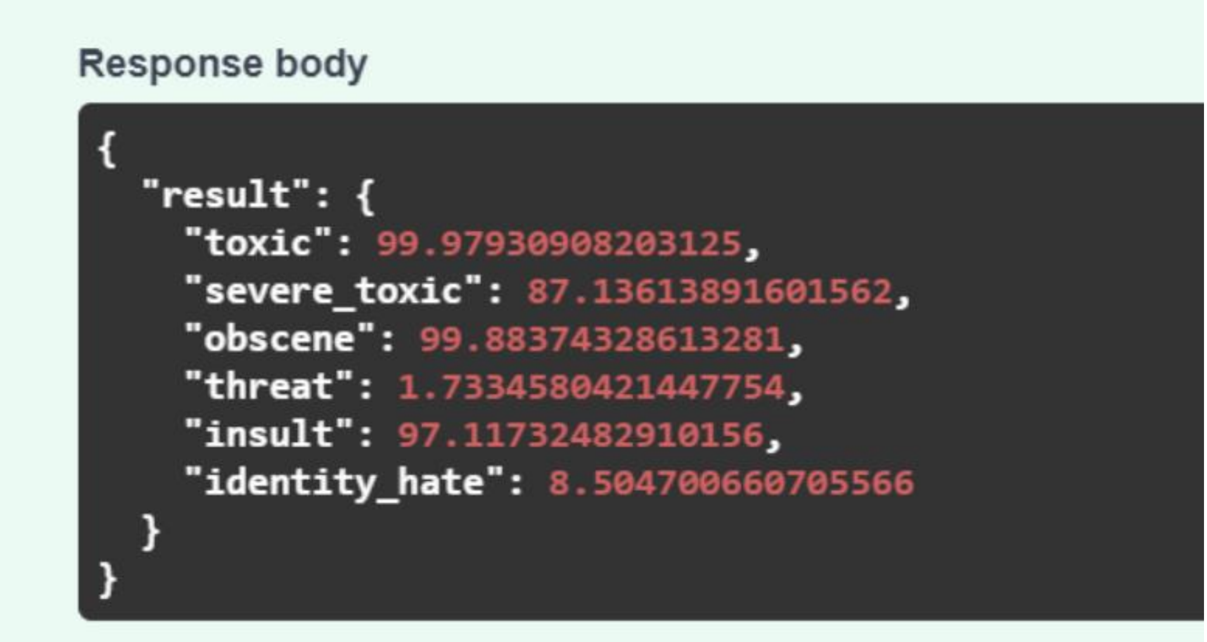
Comment



Request body required

```
{
  "comment": "What a motherfucking piece of crap those fuckheads for blocking us!"
}
```

Prediction



Response body

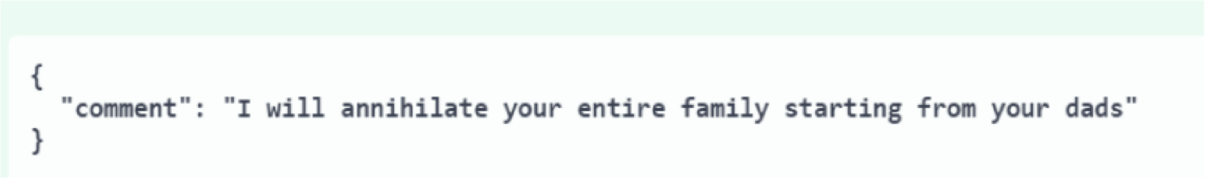
```
{
  "result": {
    "toxic": 99.97930908203125,
    "severe_toxic": 87.13613891601562,
    "obscene": 99.88374328613281,
    "threat": 1.7334580421447754,
    "insult": 97.11732482910156,
    "identity_hate": 8.504700660705566
  }
}
```

Figure 47 :- Severe Toxic Prediction

Despite the figure 30, the approximate Severe Toxic level is 87%.

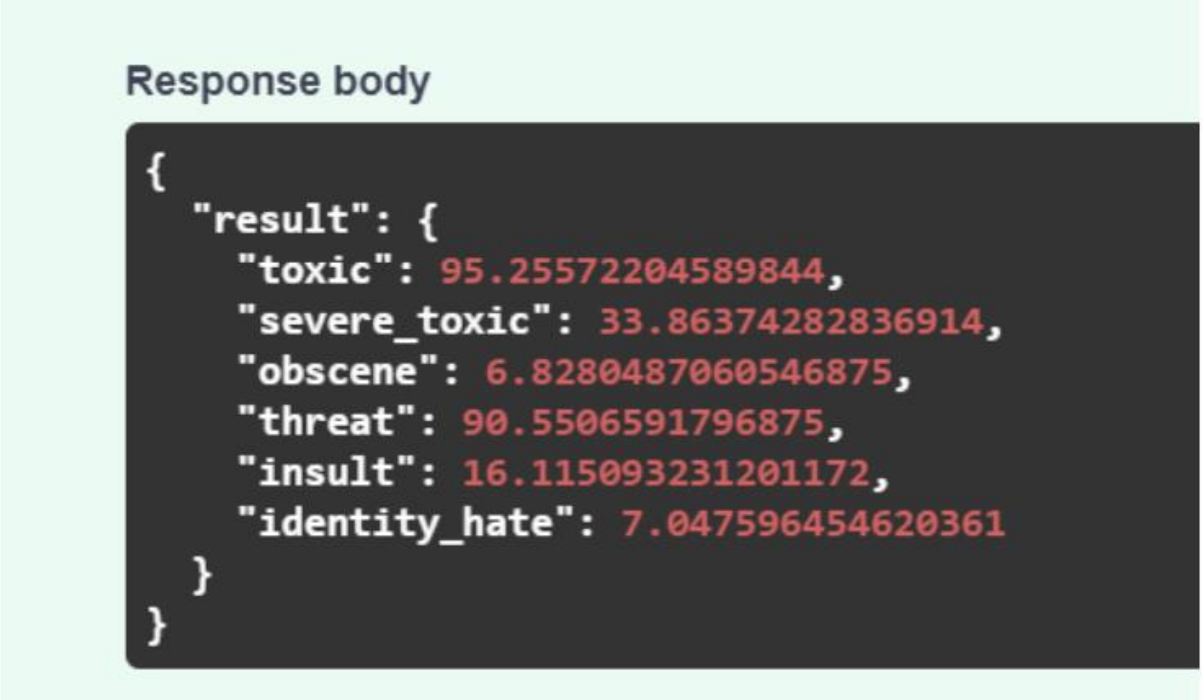
4. Label - Threat

Comment



```
{
  "comment": "I will annihilate your entire family starting from your dads"
}
```

Prediction



Response body

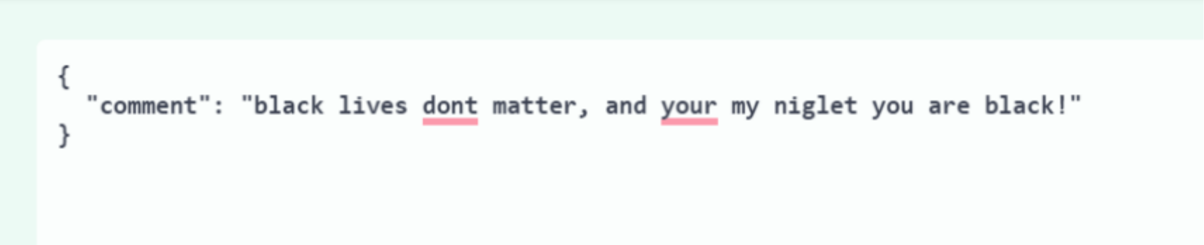
```
{
  "result": {
    "toxic": 95.25572204589844,
    "severe_toxic": 33.86374282836914,
    "obscene": 6.828048706054 6875,
    "threat": 90.5506591796875,
    "insult": 16.115093231201172,
    "identity_hate": 7.047596454620361
  }
}
```

Figure 48:- Threat Prediction

Despite the figure 31, the approximate Threat level is 91%.

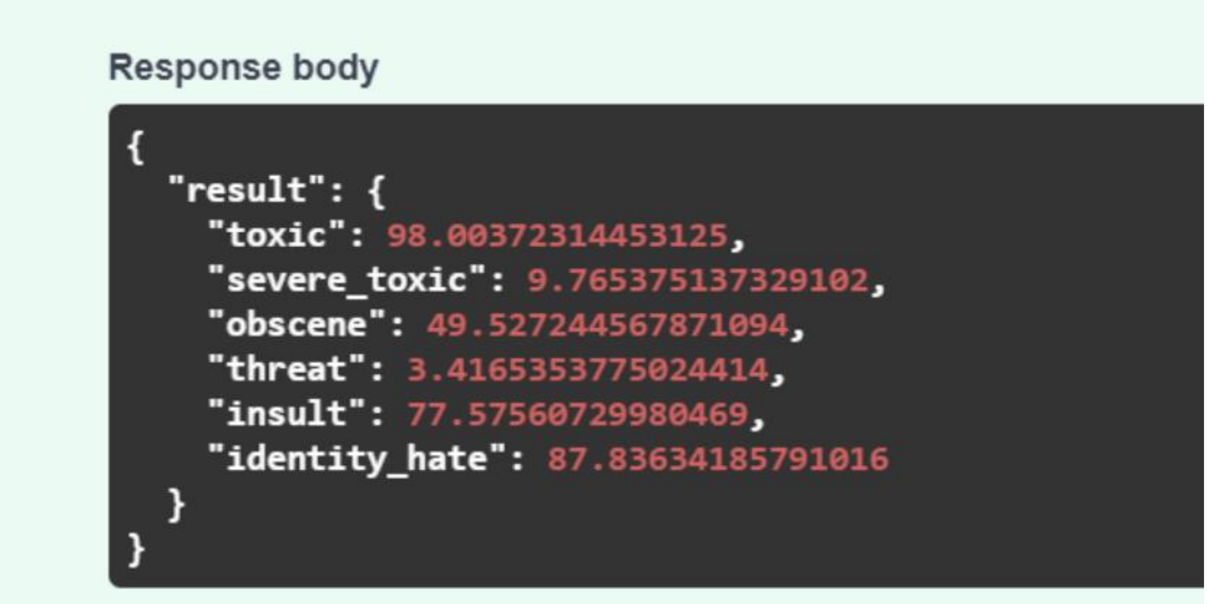
5. Label - Identity Hate

Comment



```
{
  "comment": "black lives dont matter, and your my niglet you are black!"
}
```

Prediction



Response body

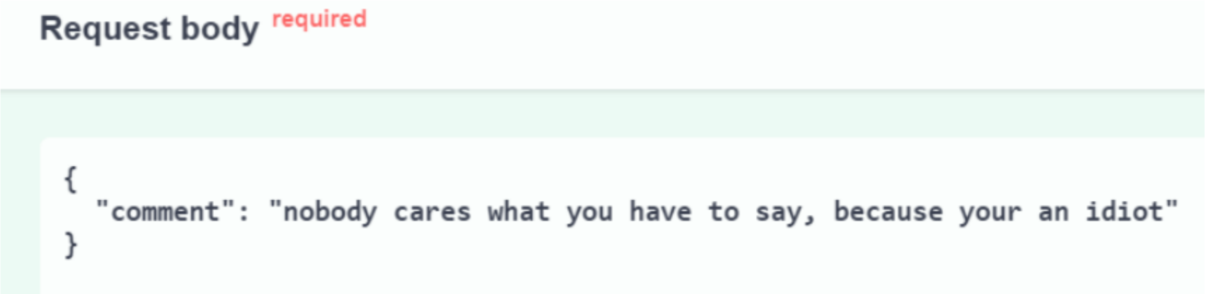
```
{
  "result": {
    "toxic": 98.00372314453125,
    "severe_toxic": 9.765375137329102,
    "obscene": 49.527244567871094,
    "threat": 3.4165353775024414,
    "insult": 77.57560729980469,
    "identity_hate": 87.83634185791016
  }
}
```

Figure 49:- Identity Hate Prediction

Despite the figure 32, the approximate Identity Hate level is 88%.

6. Label - Insult

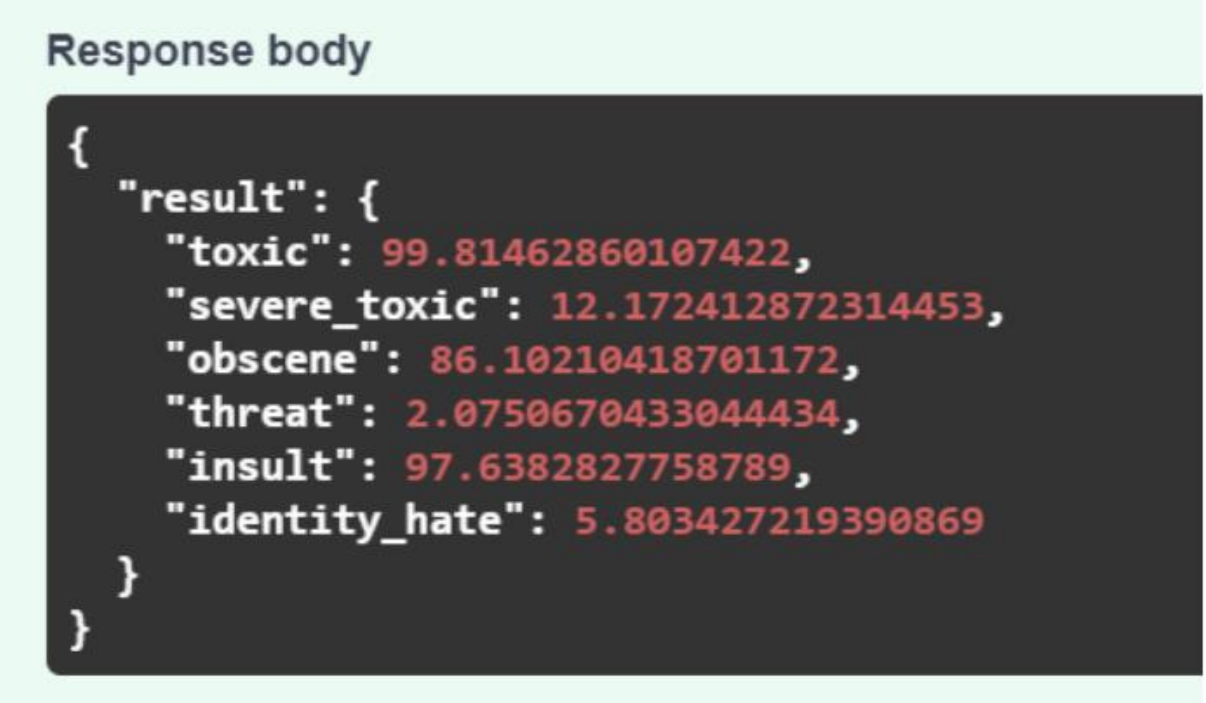
Comment



Request body required

```
{
  "comment": "nobody cares what you have to say, because your an idiot"
}
```

Prediction



Response body

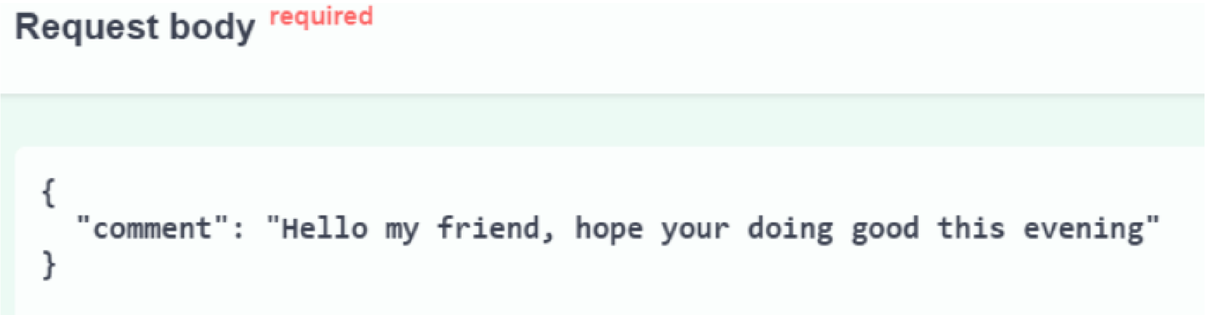
```
{
  "result": {
    "toxic": 99.81462860107422,
    "severe_toxic": 12.172412872314453,
    "obscene": 86.10210418701172,
    "threat": 2.075067043304434,
    "insult": 97.6382827758789,
    "identity_hate": 5.803427219390869
  }
}
```

Figure 50:- Insult Prediction

As per the figure 33, the approximate Insult prediction Hate level is 98%.

7. Normal Comment

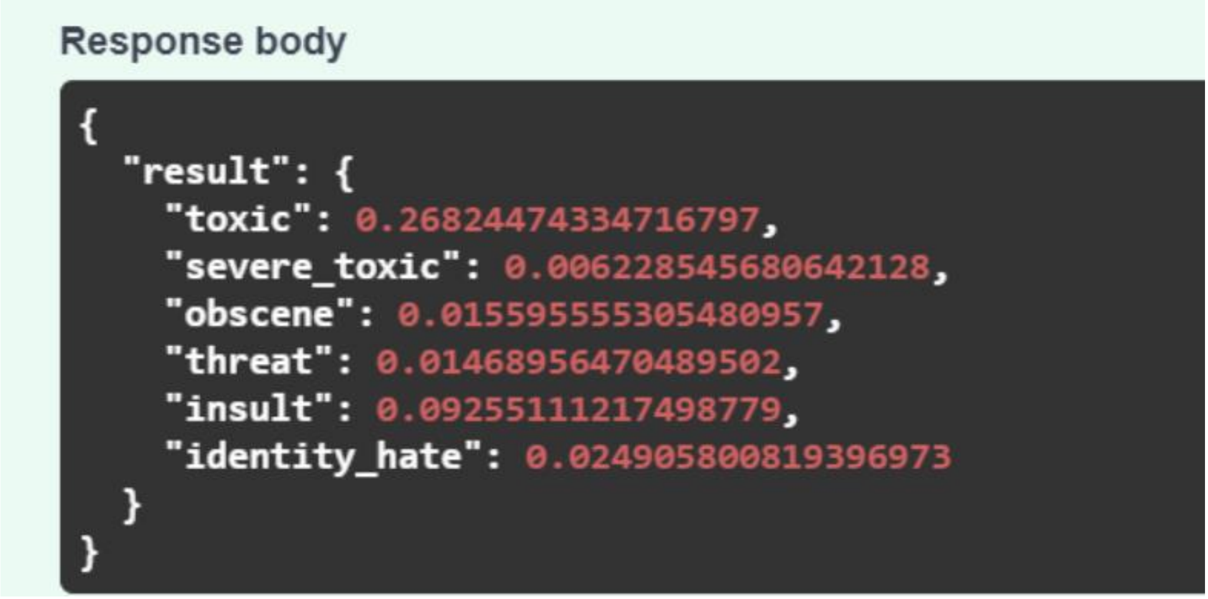
Comment



Request body required

```
{
  "comment": "Hello my friend, hope your doing good this evening"
}
```

Prediction



Response body

```
{
  "result": {
    "toxic": 0.26824474334716797,
    "severe_toxic": 0.006228545680642128,
    "obscene": 0.01559555305480957,
    "threat": 0.01468956470489502,
    "insult": 0.09255111217498779,
    "identity_hate": 0.024905800819396973
  }
}
```

Figure 51 :- Normal Comment

Percentage values are significantly lower

4.11.2. Test Cases of AI Model

FAST API

Table 18 :- Test Cases of FAST API

ID	Case	Description	Output
01	Execute the model locally	Download the saved model output from the notebook and run it locally	Pass
02	Send http request to the model api in order to predict	Send an input as a POST request to the API and obtain the results as a response	Pass

Docker

Table 19:- Test Case of Docker

ID	Case	Description	Output
03	Dockerized the Detection Model Microservice	Creates an image for developing, shipping, and running applications. Docker enables you to separate your applications from your infrastructure so you can deliver software quickly	Pass

Deployment

Table 20 :- Test Cases of Deployment

ID	Case	Description	Output
04	Push the dockerized image to google cloud image container registry	The google cloud image container registry provides a service for storing space container images and a subset of features provided by Artefact Registry, a universal repository manager and the recommended service for managing container images and other artifacts in Google Cloud	Pass
05	Push the image to google cloud run	Pushes the image to a managed compute platform that enables you to run containers	Pass

		that are invocable via requests or events. Cloud Run is serverless: it abstracts away all infrastructure management	
06	Access the microservice using the public URL	Use the deployed microservice and test it's functionality	Pass

4.11.3. Test Cases of Social Media Web Application

Integration

Table 21:- Test Cases of Integration

ID	Case	Description	Output
01	HTTP request to the Detection Model API	Send an input as a POST request to the Detection API and obtain the results as a response	Pass
02	HTTP request via the social media web API to the Detection Model API	Send an input as a POST request to the Detection API via the server-side of the web application and obtain the results as a response	Pass

Posts

Table 22:- Test Cases of Posts

ID	Case	Description	Output
03	Add Post Without Media	Uploading a post with only a caption	Pass
04	Add Post With An Image	Uploading a post with only an image	Pass
05	Add Post With A Video	Uploading a post with only a video	Pass

Comments

Table 23:- Test Cases of Comments

ID	Case	Description	Output
07	Commenting on a post	Adding a text input under a post as a comment	Pass

08	Prediction Tooltip on a comment	Text aggressiveness details shown on a comment over a tooltip	Pass

Chapter 05 – Conclusion, Personal Evaluation & Future Improvements

This Research and the Implementation is mainly focusing to analyse the use of AI Framework for understanding the behavioural patterns of terrorism, hetaerism and toxicity in online spaces. The researcher had to find and review previous literatures regarding the subject to seek the possibility of identifying the terrorism, hetaerism and toxicity behaviours in online spaces using an AI Framework related to the users Kandy, Sri Lanka. The research carried on seeking the possibility of identifying the terrorism, hetaerism and toxicity in Sri Lanka by analysis of content posted over social media using an AI model which significantly contributes to many aspects. The online spaces in Sri Lanka tend identify and filter terrorism, toxicity and hetaerism views and opinions on online spaces. Thus, when the social media users share their honest views and opinions on social media the probability to identify the terrorism, hetaerism and toxicity from the behaviours increases. The filtering of terrorism, hetaerism and toxicity views and opinions becomes possible and the ability to identify terrorism, hetaerism and toxicity using social media platforms increases. If the social media users to allow the analysis using AI it would become very handy to analyse and identify the terrorism, hetaerism and toxicity behaviour, as the population has given their consent. By providing the above-mentioned benefits to aid in increasing the social security in Sri Lanka, by establishing a AI Model to analyse terrorism, hetaerism and toxicity based on online spaces so that the government can prevent future threats by identify the offensive groups or individuals who spread terrorism, hetaerism and toxicity. It can be concluded that this research contributes significantly to increase the existing knowledge of the society.

As discussed, Sri Lanka have not been subjected to research about Artificial intelligence in identifying the terrorism, toxicity and hetaerism, therefore this study was taken into consideration to provide a fruitful solution to address the problem domain. Delimitations of this study could be amended, and much future research could be conducted. Delimitations of the geographical location in the study on Kandy Region could be altered in a way to focus the above research study towards many other private sector higher education institutes with entire population in Sri Lanka. This study is conducted based on the outcomes of data collected from social media users, where else future research could extend its samples by including all the social media users in Sri Lanka without a specific geographical limitation or restriction, whereas the research could be enhanced in the perception on using AI to identify terrorism, toxicity and hetaerism using social media platforms. As per the conclusion of this investigation, the researcher was able to gain a better knowledge and increase his competence and belief as an academic. With the production of this study, the researcher might gain extensive information on research procedures, various data analytic approaches, and statistical models. The researcher was able to gather extensive information on artificial intelligence, recognizing behaviour patterns utilizing artificial intelligence in social media, and big data analysis via the examination of several researches to assist this work. The researcher was more successful in management due to the limited time allocation and resource availability, and the experience and information gained would be valuable in future investigations.

The research approaches employed to complete this study were critical in reaching the study's aims. As a result of the researcher's application of positivism, the researcher has learned not to allow sentiments and emotions cloud his judgment, resulting in more accurate and impartial study results. Using a deductive strategy for this research, the researcher was able to explain correlations between ideas and variables, measure and quantify data, and generalize research findings to some level, so assisting the researcher in proving the study's premise. The use of the survey approach for data collecting allowed the researcher to obtain more reliable sample findings from which he could draw inferences and make critical suggestions. The researcher was able to gather accurate, trustworthy, and generalizable data collecting by using quantitative research methods, which improved the dependability of the study results. The use of a cross-sectional time horizon aided the researcher in completing this study in three months for data collecting. The use of random sampling for the population of this study decreased the bias of the results. The researcher's usage of IBM's SPSS program allowed him to readily analyse the obtained data, draw conclusions, and make suggestions to the target audience.

After working through the implementation, it is possible to notice how strongly the context of a statement may give it significance. It makes no difference if the sentence does not contain any specific harmful terms. However, if the entire meaning of a sentence might literally signify anything insulting, it must be considered hostile writing. To achieve the project's goal To on researching the usage of AI Framework for comprehending the online toxicity, hetaerism, and terrorism behaviour patterns, and to design an AI Prototype to create a peaceful, inclusive, and decent community. Contextual meaning is quite important. As a result, the BERT algorithm has a significant influence on weighting words based on context. According to the offered solution, it allows us to come one step closer to the project's goal. This is because, as stated in the problem description, it is conceivable to conclude that the project's solution addresses the issue as needed. However, there are several obstacles to overcome in order to fulfill the project's goal. However, with what we have, the following steps might have a significant influence on accomplishing the goal. Moving forward with the current solution, there are several solutions available to boost the odds of success. The following technological areas may help us boost the chances.

- Engineering a continuous learning pipeline for the classification model to consistently or periodically learn and adapt to the latest trends.
- Multiple Detection models to detect in every other language out there.
- Enhancing the Detection API to fit in many other detection models.
- Inclusion of other AI techniques such as image detection to counter terrorism, toxicity and hetaerism in any type of a media.

Gannt Chart for Report

STAGE	Aug				Sep				Oct				Nov				Dec				Jan
	W 1	W 2	W 3	W 4	W 1	W 2	W 3	W 4	W 1	W 2	W 3	W 4	W 1	W 2	W 3	W 4	W 1	W 2	W 3	W 4	W 1
Commencing Report	■	■	■																		
Chapter 1			■	■	■	■			■	■											
Chapter 2				■	■	■	■	■		■	■										
Chapter 3							■	■	■	■	■	■			■	■	■	■			
Chapter 4									■	■	■	■	■	■	■	■		■	■		
Chapter 5																				■	■

References

Afutu-Kotey, R. L., Gough, L. W. & Owusu, G., 2017. *Young Entrepreneurs in the Mobile Telephony Sector in Ghana: From Necessities to Aspirations.* s.l.:Journal of African Business.

Aida, A. & Behrang, S., 2014. *A Study on the Negative Effects of Social Networking Sites. A Study on the Negative Effects of Social Networking Sites.* 5 ed. s.l.:s.n.

Alrawashdeh & Ahmad, 2019. *What are the advantages and disadvantages of artificial intelligence?.* [Online]
Available at:
https://www.researchgate.net/post/What_are_the_advantages_and_disadvantages_of_artif
[Accessed 9 6 2022].

Australian Human Rights Commission., 2021. *Examples of Racist Material on the Internet,* s.l.: Humanrights.gov.au.

Backstrom, L., Sun, E. & Marlow, C., 2010. *Find me if you can: Improving geographical prediction with social and spatial proximity,.* California: International Conference on World Wide Web.

Badke , W., 2011. *Research Strategies: Finding Your Way Through the Information Fog.* 4th ed. s.l.:Universe.

Bajada & Josef, 2019. *Artificial Intelligence Demystified..* [Online]
Available at: https://towardsdatascience.com/https-medium-com-josef-bajada-demystifying-artificial-intelligence-6f5f7a8dd1b0
[Accessed 14 6 2022].

Bansal & Gauri, 2018. *Big Data and AI for Psychology: Predicting Human Behaviour.* [Online]
Available at: https://www.analyticsinsight.net/big-data-and-ai-for-psychological-science-predicting-human-behaviour/
[Accessed 13 06 2022].

Benevenuto, F. & Rodrigues, T., 2009. *Characterizing user behavior in online social networks.* Chicago: SIGCOMM Conference.

Benson, V., Saridakis, G. & Tennakoon, H., 2015. *Information disclosure of social media users.* Bingley: Emerald Group Publishing Limited..

Billon, M., Lera-Lopez, F. & Marco, R., 2017. *Patterns of Combined ICT Use and Innovation in the European Regions.* s.l.:Journal of Global Information Technology Management.

Blumenstock & Gillick, 2010. *Who's calling? Demographics of mobile phone use in Rwanda.* Rwanda: s.n.

Bongomin, G. .. O. .. C., Ntayi, J. M., Munene, J. C. & Malinga, C. A., 2018. *Mobile Money and Financial Inclusion in Sub-Saharan Africa: the Moderating Role of Social Networks.* s.l.:Journal of African Business.

Boyd, D. M., 2007. *Journal of Computer-Mediated Communication.* England: Oxford University Press.

Brinkhoff, T., 2021. *Kandy, District in Sri Lanka.* [Online]
Available at: https://www.citypopulation.de/en/srilanka/prov/admin/central/21__kandy/
[Accessed 16 6 2022].

Burgess, M., 2003. *A Brief History of Terrorism.* Washington: Center for Defense Information (CDI).

businessdictionary, 2019. *businessdictionary.* [Online]
Available at: http://www.businessdictionary.com/definition/regression-analysis-RA.html
[Accessed 1 11 2022].

Cervone, G. et al., 2016. *Using Social Media and Satellite Data for Damage Assessment in Urban Areas During Emergencies.* Switzerland: Springer.

Cherry & Kendra, 2019. *Forming a Good Hypothesis for Scientific Research.* [Online]
Available at: https://www.verywellmind.com/what-is-a-hypothesis-2795239
[Accessed 11 9 2022].

Chetty & Priya, 2016. *Project Guru.* [Online]
Available at: https://www.projectguru.in/publications/selecting-research-approach-business-studies/
[Accessed 15 7 2022].

Chorley, M. J., Whitaker, R. M. & Stuart , A., 2015. *Personality and location-based social networks.* Amsterdam: Elsevier Ltd.

Chun, M., 2012. *The affective/cognitive involvement and satisfaction according to the usage motivations of social network services.* London: Researchgate.

Comito, C. & Falcone, D., 2016. *Mining human mobility patterns from social geo-tagged data. Pervasive and Mobile Computing.* s.l.:33rd IEEE International Conference.

Crabb, J., Yang, S. & Zubova, A., 2019. *Classifying Hate Speech: an overview.* [Online]
Available at: https://towardsdatascience.com/classifying-hate-speech-an-overview-d307356b9eba
[Accessed 20 08 2022].

Crenshaw, M., 1995. *Terrorism in Context.* University Park: Pennsylvania State University Press.

Cui, L. & Shi, J., 2012. *Urbanization and its environmental effects in Shanghai,.* New York: Elsevier Ltd.

Devlin, J., Chang, M. W., Lee, K. & Toutonova, C., 2019. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.* [Online]
Available at: https://arxiv.org/pdf/1810.04805.pdf
[Accessed 15 7 2022].

Dorchai, S. & Meulders, D., 2009. *Statistics and Indicators on Gender Equality in Science.* Luxembourg: European Union.

Founta, A. et al., 2018. *Large Scale Crowdsourcing and Characterization of Twitter Abusive Behavior,* London: s.n.

Gates, S. & Podder, S., 2015. *Social Media, Recruitment, Allegiance and the Islamic State.* s.l.:Perspectives on Terrorism.

Geet d'Sa, A., Illina, I. & Fohr, D., 2021. *Classification of Hate Speech Using Deep Neural Networks,* s.l.: Centre de Recherche sur l'Information Scientifique et Technique (CERIST).

Gong, L. & Ji, R., 2018. *What does a TextCNN learn?.* s.l.:arXiv preprint .

Guardian News & Media Limited, 2018. *Teenagers and social networking it might actually be good for them.* [Online]

Available at: https://www.theguardian.com/lifeandstyle/2013/oct/05/teens-social-networking-good-for-them
[Accessed 4 6 2022].

Guo, W. et al., 2020. *Detext: A deep text ranking framework with bert.* s.l.:In Proceedings of the 29th ACM international conference on information & knowledge management .

Gu, Z. & Zhang, Y., 2016. *Analysis of attraction features of tourism destinations in a mega city based on check-in data mining.* China: ISPRS.

Gyarmati, L. & Trinh, T. A., 2010. *Measuring user behavior in online social networks.* York City: IEEE Network.

Hate Speech Transparency Center Meta, 2021. *Hate speech.* [Online]
Available at: https://transparency.fb.com/en-gb/policies/community-standards/hate-speech/
[Accessed 15 8 2022].

Heath & Nick, 2016. *What is AI? Everything you need to know about Artificial Intelligence.* [Online]
Available at: https://www.zdnet.com/article/what-is-ai-everything-you-need-to-know-about-artificial-intelligence/
[Accessed 12 6 2022].

Hesse, B. W., Moser, R. P. & Riley, W., 2015. *From Big Data to Knowledge in the Social Sciences.* Philadelphia: s.n.

Hoffman, B., 2006. *Inside Terrorism.* 2nd ed. New York: Columbia University Press.

Hong, I., 2015. *Spatial analysis of location-based social networks in seoul,.* Cheonansi: Department of GIS, Namseoul University.

Horev, R., 2018. *BERT Explained: State of the art language model for NLP.* [Online]
Available at: https://towardsdatascience.com/bert-explained-state-of-the-art-language-model-for-nlp-f8b21a9b6270
[Accessed 21 8 2022].

International Bar Association, 2003. *International Terrorism: Legal Challenges and Responses.* International Terrorism: Legal Challenges and Responses: Transnational Publishers.

Isaac, M., 2021. *Frances Haugen, Facebook Whistle-Blower, Testifies to.* [Online]
Available at: https://www.nytimes.com/2021/10/25/business/frances-haugen-facebook.html
[Accessed 25 8 2022].

Jin, L., Chen, Y., Wang, T. & Hui, P., 2013. *Understanding user behavior in online social networks.* s.l.:IEEE Communication.

Joint Chiefs of Staff Department of Defense - USA, 2008. *Department of Defense Dictionary of Military and Associated Terms,* Washington: Department of Defense USA.

Kemp, S., 2022. *DIGITAL 2022: SRI LANKA.* [Online]
Available at: https://datareportal.com/reports/digital-2022-sri-lanka
[Accessed 10 8 2022].

Khan, I. & Dongping, . H., 2017. *Variations in the diffusion of social media content across different cultures: A communicative ecology perspective.* s.l.:Journal of Global Information Technology Management.

Khurana, J., 2015. *The Impact of Social Networking Sites on the Youth.* Mumbai: Mass .

Kothari, C., 2004. *Research Methodology: Methods and Techniques..* s.l.:New Age International.

Krishnan, S., Ahmed, M. & AlSudiary, T., 2016. *Cultural Practices and Virtual Social Networks Diffusion: An International Analysis Using GLOBE Scores.* s.l.:Journal of Global Information Technology Management.

Kuada, J., 2012. *Research Methodology: A Project Guide for University Students.* s.l.:Samfunds Literature.

Kulandairaj, A. J., 2014. *Impact of social media on the lifestyle of youth.* Chennai: International Journal of Technical Research and Applications.

Kung, K. S., 2014. *Exploring universal patterns in human home-work commuting from mobile phone data..* China : China Plus One.

Kursuncu, U. et al., 2019. Predictive analysis on twitter: Techniques and applications, in Emerging research challenges and opportunities in computational social network analysis and mining. In: s.l.:Springer, p. 67–104.

Kursuncu, U., Mejova, Y., Blackburn, J. & Sheth, A., 2020. *Cyber social threats 2020 workshop meta-report: Covid-19, challenges, methodological and ethical considerations,.* s.l.:Workshop Proceedings of the 14th Conference on Web and Social Media.

Laqueur , W., 1987. *The Age of Terrorism.* 2nd ed. Boston: Little & Brown.

League Convention, 1937. *Convention for the Prevention and Punishment of Terrorism,* s.l.: League Convention.

Lee, K. & Joshi, K., 2016. *Importance of Globalization in the Information Technology,.* s.l.:Journal of Global Information Technology Management.

Lei, C. & Zhang, A., 2018. *Spatial-Temporal Analysis of Human Dynamics on Urban Land Use Patterns Using Social Media Data by Gender.* Coruna: ISPRS.

Lin & Karen, 2018. *Role of Data Science in Artificial Intelligence.* [Online]
Available at: https://towardsdatascience.com/role-of-data-science-in-artificial-intelligence-950efedd2579
[Accessed 21 6 2022].

Luarn , P. & Yang, J., 2015. *Why people check in to social network sites.* London: Mobile Commerce.

Maia, M. & Almeida, J., 2008. *Identifying user behavior in online social networks,.* Scotland: Workshop on Social Network Systems.

Mathur, P., Sawhney, R., Ayyar, M. & Shah, R. R., 2018. *Did you offend me? Classification of Offensive Tweets in Hinglish Language,* India: s.n.

Melodia, T., Iera, A., Rudin, H. & Stiller, B., 2014. *Estimating human trajectories and hotspots through mobile phone data.* Berlin: Springer.

Moeller, S. D., 2002. *A Hierarchy of Innocence: The Media's Use of Children in the Telling of International News.* s.l.:The International Journal of Press/Politics.

Mozofari, M., Farahbakhsh, R. & Crespi, N., 2020. *Hate speech detection and racial bias mitigation in social media based on BERT model.* [Online]
Available at: https://doi.org/10.1371/journal.pone.0237861
[Accessed 20 08 2022].

Muscanell, N. L. & Guadagno, R. E., 2012. *Make new friends or keep the old: Gender and personality differences in social networking use.* London: Elsevier Ltd.

Nguyen, T. B. et al., 2019. *VAIS Hate Speech Detection System: A Deep Learning based Approach for System Combination,* Vietnam: s.n.

Noulas, A. & Scellato, S., 2011. *An empirical study of geographic user activity patterns in foursquare.* Spain: Fifth International AAAI Conference on Weblogs and Social Media.

Parkyn, R., 2017. *The role of social media in development, The World Bank.* [Online]
Available at: https://blogs.worldbank.org/publicsphere/role-social-media-development
[Accessed 4 7 2022].

Patrikarakos, D., 2017. *War in 140 Characters: How Social Media is reshaping conflict in the Twenty-first Century.* [Online]
Available at: https://www.hachettebookgroup.com/titles/david-patrikarakos/war-in-140-characters/9780465096152/
[Accessed 6 7 2022].

Pentina, I., Basmanova, O. & Zhang, A., 2016. *A cross-national study of Twitter users motivations and continuance intentions.* London: J. Marketing Communications.

Pereira-Kohatsu, J. C., Quijano-Sánchez, L., Liberatore, F. & Camacho-Collados, M., 2019. Detecting and Monitoring Hate Speech in Twitter. *Sensors,* 26 October, p. 37.

Preoţiuc, D. & Cohn, T., 2013. *Mining user behaviours,.* Paris: Annual ACM Web Science Conference.

Pucci, P., Manfredini, F. & Tagliolato, P., 2015. *Mapping Urban Practices through Mobile Phone Data.* New York: Springer.

Purohit, H. et al., 2011. *Understanding user-community engagement by multi-faceted features: A case study on twitter.* s.l.:WWW 2011 Workshop on Social Media Engagement (SoME).

Reed., P. J. & Khan, M. R., 2016. *Observing gender dynamics and disparities with mobile phone metadata.* Ann Arbor: Ann Arbor: Technologies and Development.

Rithika, M. & Selvaraj, S., 2013. *Impact of social media on student's academic performance.* Volume 2 ed. India: Pezzottaite Journals.

Rizwan, M. & Mahmood, S., 2017. *Location based social media data analysis for observing check-in behavior and city rhythm in Shanghai.* Shanghai: International Conference on Smart and Sustainable City.

Rizwan, M., Wanggen, W. & Wanggen, O., 2018. *Using Location-Based Social Media Data to Observe Check-In Behavior and Gender Difference.* France: Université Grenoble Alpes.

Roche, S., 2014. *Geographic Information Science I : I: Why does a smart city need to be spatially enabled?.* Canada: Université Laval.

Ruggles, S., 2014. *Big microdata for population research.* India: India : Demography.

Sanchez, D. B., 2019. *Designing and Development of a Hate Speech Detector in Social Networks based on Deep Learning Technologies,* Madrid: Universidad Politecnica De Madrid.

Scellato, S., Noulas, A. & Lambiotte, R., 2011. *Socio-spatial properties of online location based social networks.* Barcelona: International Conference on Weblogs and Social Media.

Scellato, S., Noulas, A. & Lambiotte, R., 2011. *Socio-spatial properties of online location based social networks,.* Barcelona: International Conference on Weblogs and Social Media.

Schmid, A. & Jongman, A., 1988. Political Terrorism: A New Guide to Actors, Authors, Concepts Data Bases, Theories, and Literature.. In: Amsterdam: Transaction Books, p. 28.

Seda, L. & Altin, M., 2019. *Doctoral Symposium at SEPLN.* s.l.:s.n.

Sentance & Rebecca, 2017. *Artificial intelligence in marketing.* [Online]
Available at: https://econsultancy.com/15-examples-of-artificial-intelligence-in-marketing/
[Accessed 12 6 2022].

Seven Media Group, 2018. *SriLanka Social Media Audience Survey.* [Online]
Available at: https://sevenmediagroup.com/sri-lanka-social-media-audience-survey/
[Accessed 1 4 2022].

Shen, Y. & Karimi, Y., 2016. *Urban function connectivity: Characterisation of functional urban streets with social media check-in data.* London: Elsevier Ltd.

Shuttleworth & Martyn, 2019. *Research Hypothesis.* [Online]
Available at: https://explorable.com/research-hypothesis
[Accessed 10 9 2022].

Simon, J. D., 1994. *The Terrorist Trap. Bloomington.* Indiana : Indiana University Press.

Smith, A., 2018. *Why Americans Use Social Media..* [Online]
Available at: https://www.pewinternet.org/2011/11/15/why-americans-use-social-media/
[Accessed 4 8 2022].

Sreelakshmi, . K., Premjith, B. & Soman, K. P., 2020. *Detection of Hate Speech Text in Hindi-English Code-mixed Data,* Coimbatore, Amrita Vishwa Vidyapeetham, India: Center for Computational Engineering & Networking (CEN) Amrita School of Engineering.

Stefanone, M. A. & Huang, Y. C., 2011. *Negotiating Social Belonging: Online, Offline, and In-Between.* Kauai: Hawaii International Conference.

Strater, K. & Lipford, H., 2007. *Examining privacy and disclosure in a social networking community,.* Pittsburgh: Symposium On Usable Privacy and Security.

The Survey System, 2018. *Creative Research Systems.* [Online]
Available at: https://www.surveysystem.com/correlation.htm
[Accessed 19 11 2022].

Thelwall & Mike, 2008. *Social Networks, Gender, and Friending.* Wolverhampton: University of Wolverhampton.

Tuman, J. S., 2009. *Communicating Terror: The Rhetorical Dimensions of Terrorism.* Sage: Thousand Oaks.

Twitter, 2021. *Abusive behavior.* [Online]
Available at: https://help.twitter.com/en/rules-and-policies/abusive-behavior
[Accessed 20 08 2022].

U.S. Department of State, 1996. *Patterns of Global Terrorism,* Washington: U.S. Department of State.

Vairavan & Arvind, 2019. *What is AI.* [Online]
Available at: https://medium.com/datadriveninvestor/1-what-is-ai-6f8aff4e15d
[Accessed 7 06 2022].

Vaswani, A. et al., 2017. *Attention is All you Need: NeurIPS Proceedings.* [Online]
Available at:

https://www.researchgate.net/publication/317558625_Attention_Is_All_You_Need
[Accessed 15 7 2022].

Wang, S. S. & Stefanone, M. A., 2013. *Showing Off? Human Mobility and the Interplay of Traits, Self-Disclosure, and Facebook.* Kaosiung: National Sun Yat-sen University.

Waseem, Z. & Hovy, D., 2016. *Hateful Symbols or Hateful People? Predictive Features for Hate Speech Detection on Twitter,* San Diego, California: Association for Computational Linguistics.

Welman, C., Kruger, F. & Mitchell, B., 2006. *Research Methodology.* 3rd ed. s.l.:OUP Southern Africa.

White, J. . R., 2011. *Terrorism & Homeland Security.* 7th ed. Belmont: Wadsworth.

Witold , W. et al., 2020. The impact of social media on the lifestyle of young people. *Polish Journal of Public Health,* 130(Sciendo), pp. 22-28.

World Bank, 2016. *World Development Report 2016: Digital Dividends.* [Online]
Available at: http://www.worldbank.org/en/publication/wdr2016
[Accessed 6 7 2022].

Wu, C., Ye, X. & Ren, F., 2016. *Spatial and social media data analytics of housing prices in Shenzhen.* Bethesda: National Center for Biotechnology Information.

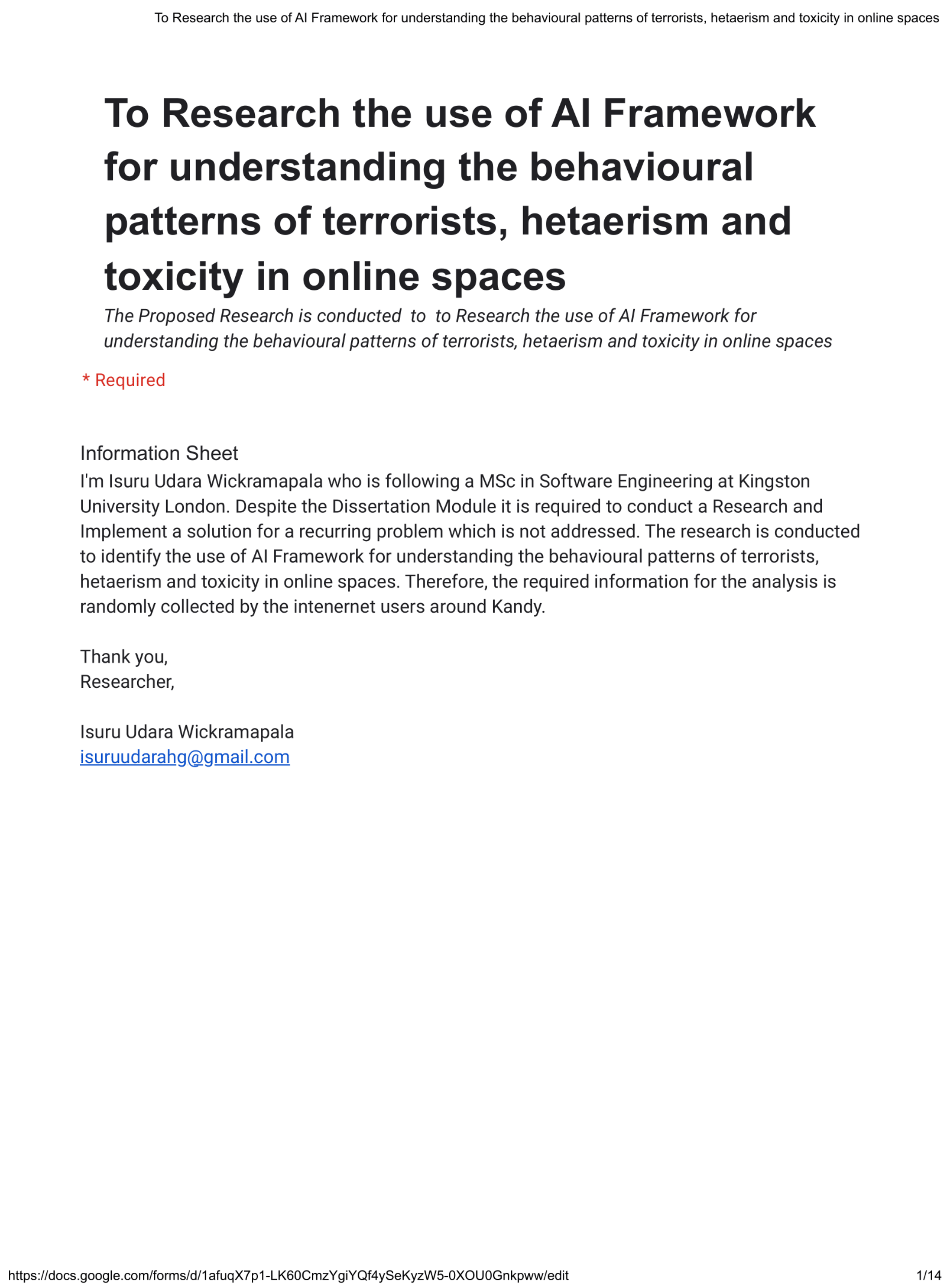
Zahi, . A., Mansoor, A. & Hashmat, S., 2016. *International Journal of Business and Management. International Journal of Business and Management.* Volume 5 ed. s.l.:s.n.

Zhang, Z. & Luo, L., 2018. *Hate Speech Detection: A Solved Problem? The Challenging Case of Long Tail on Twitter,* s.l.: s.n.

Zheng, Y. & Zhang, L., 2009. *Mining correlation between locations using human location history.* Washington: ACM SIGSPATIAL International Conference.

Appendices

Survey Form



To Research the use of AI Framework for understanding the behavioural patterns of terrorists, hetaerism and toxicity in online spaces

To Research the use of AI Framework for understanding the behavioural patterns of terrorists, hetaerism and toxicity in online spaces

The Proposed Research is conducted to to Research the use of AI Framework for understanding the behavioural patterns of terrorists, hetaerism and toxicity in online spaces

* Required

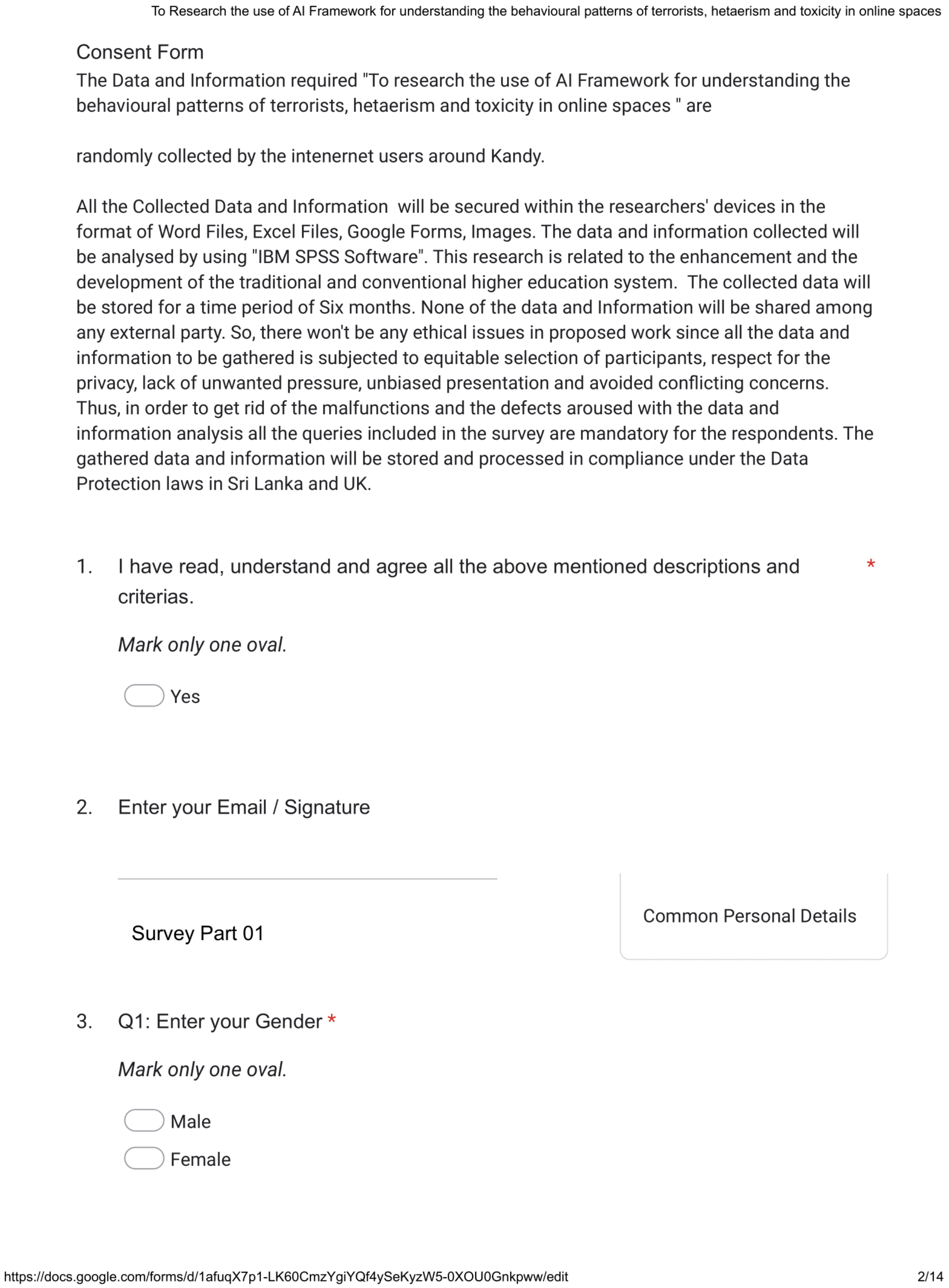
Information Sheet

I'm Isuru Udara Wickramapala who is following a MSc in Software Engineering at Kingston University London. Despite the Dissertation Module it is required to conduct a Research and Implement a solution for a recurring problem which is not addressed. The research is conducted to identify the use of AI Framework for understanding the behavioural patterns of terrorists, hetaerism and toxicity in online spaces. Therefore, the required information for the analysis is randomly collected by the intenernet users around Kandy.

Thank you,
Researcher,

Isuru Udara Wickramapala
isuruudarahg@gmail.com

https://docs.google.com/forms/d/1afuqX7p1-LK60CmzYgiYQf4ySeKyzW5-0XOU0Gnkpww/edit 1/14



To Research the use of AI Framework for understanding the behavioural patterns of terrorists, hetaerism and toxicity in online spaces

Consent Form

The Data and Information required "To research the use of AI Framework for understanding the behavioural patterns of terrorists, hetaerism and toxicity in online spaces " are

randomly collected by the intenernet users around Kandy.

All the Collected Data and Information will be secured within the researchers' devices in the format of Word Files, Excel Files, Google Forms, Images. The data and information collected will be analysed by using "IBM SPSS Software". This research is related to the enhancement and the development of the traditional and conventional higher education system. The collected data will be stored for a time period of Six months. None of the data and Information will be shared among any external party. So, there won't be any ethical issues in proposed work since all the data and information to be gathered is subjected to equitable selection of participants, respect for the privacy, lack of unwanted pressure, unbiased presentation and avoided conflicting concerns. Thus, in order to get rid of the malfunctions and the defects aroused with the data and information analysis all the queries included in the survey are mandatory for the respondents. The gathered data and information will be stored and processed in compliance under the Data Protection laws in Sri Lanka and UK.

1. I have read, understand and agree all the above mentioned descriptions and criterias. *

   Mark only one oval.

   ○ Yes

2. Enter your Email / Signature

   ____

Survey Part 01

Common Personal Details

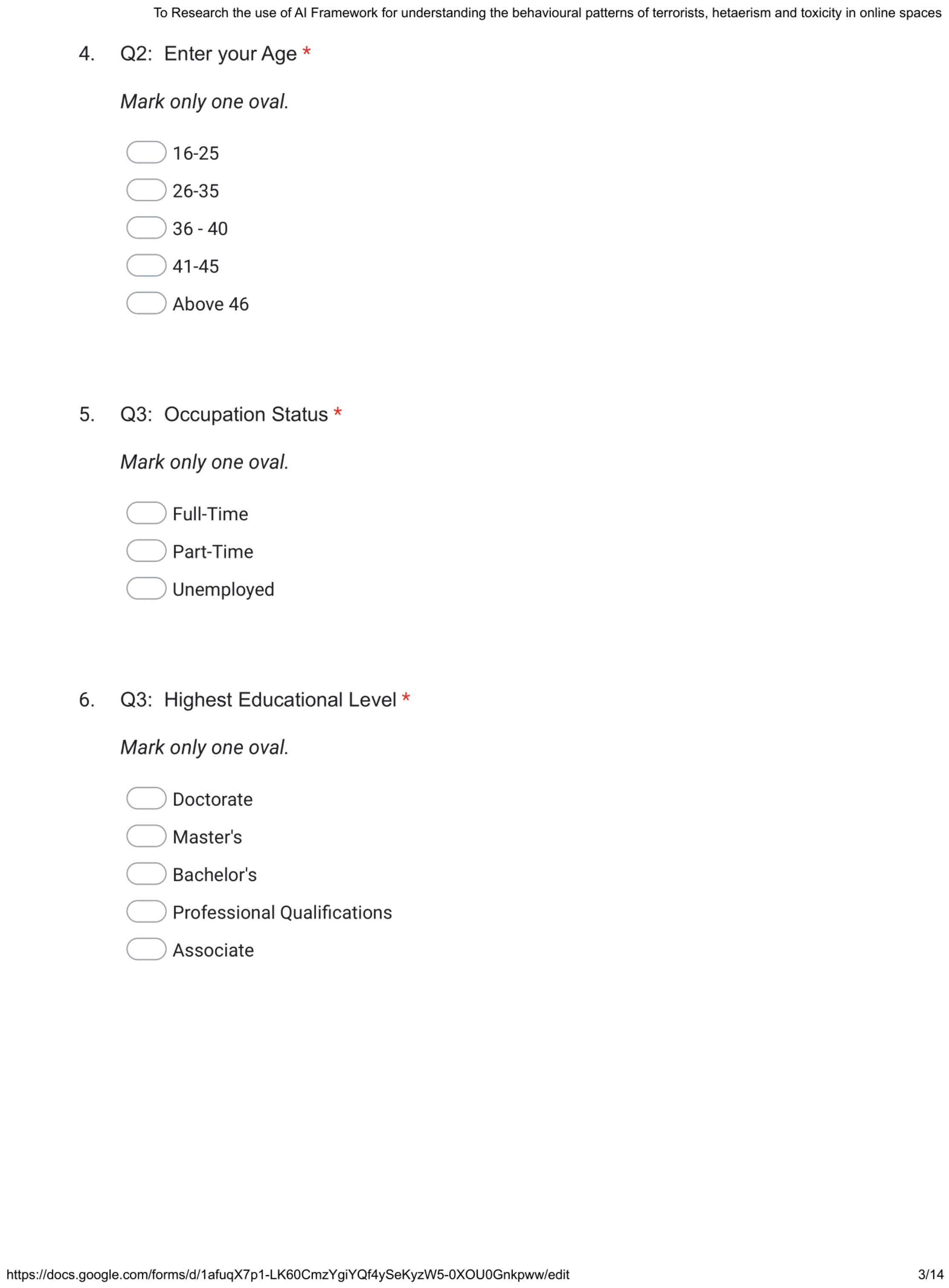
3. Q1: Enter your Gender *

   Mark only one oval.

   ○ Male

   ○ Female

https://docs.google.com/forms/d/1afuqX7p1-LK60CmzYgiYQf4ySeKyzW5-0XOU0Gnkpww/edit 2/14



4. Q2: Enter your Age *

Mark only one oval.

○ 16-25
○ 26-35
○ 36 - 40
○ 41-45
○ Above 46

5. Q3: Occupation Status *

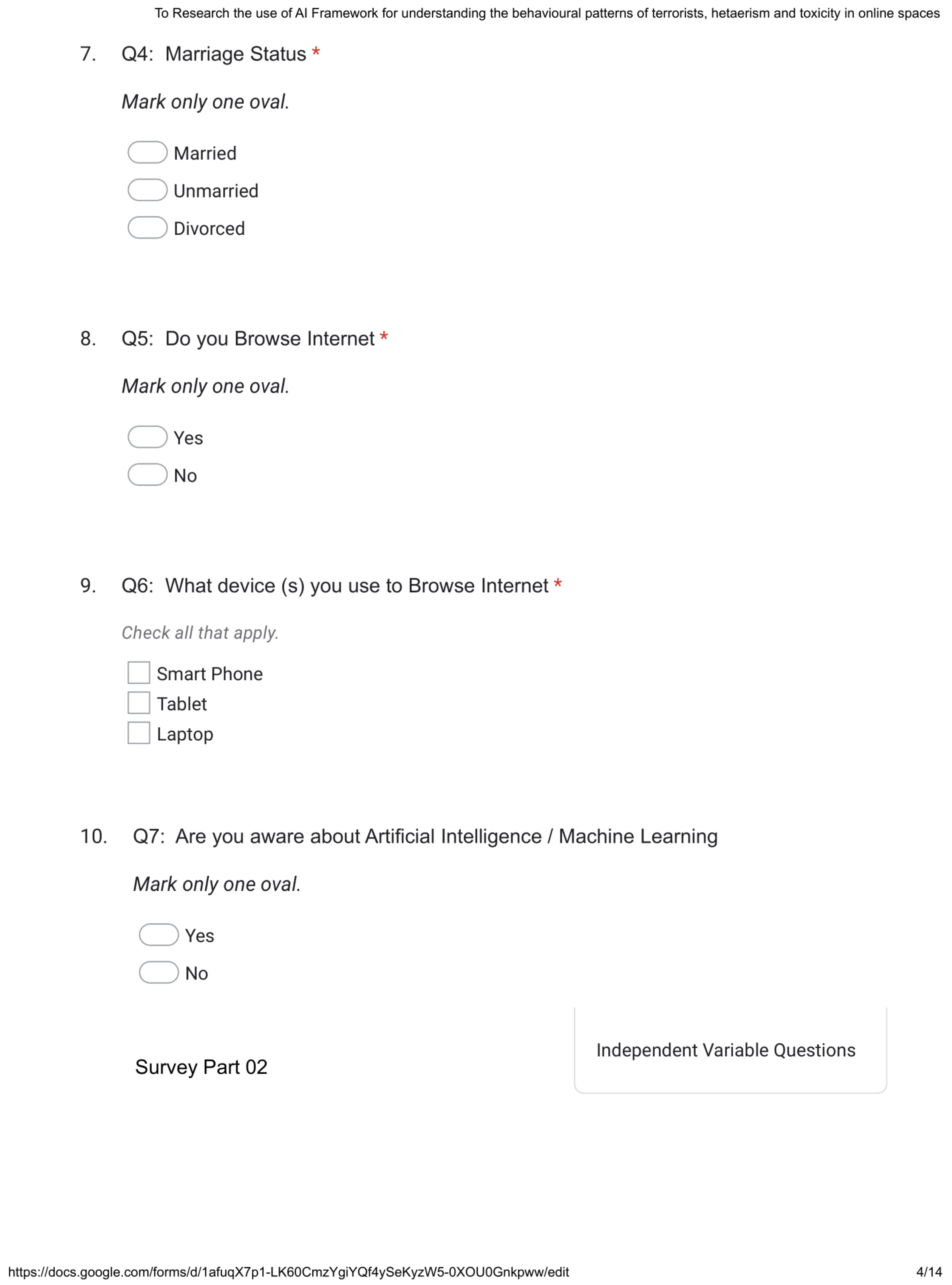
Mark only one oval.

○ Full-Time
○ Part-Time
○ Unemployed

6. Q3: Highest Educational Level *

Mark only one oval.

○ Doctorate
○ Master's
○ Bachelor's
○ Professional Qualifications
○ Associate



To Research the use of AI Framework for understanding the behavioural patterns of terrorists, hetaerism and toxicity in online spaces

7. Q4: Marriage Status *

Mark only one oval.

○ Married
○ Unmarried
○ Divorced

8. Q5: Do you Browse Internet *

Mark only one oval.

○ Yes
○ No

9. Q6: What device (s) you use to Browse Internet *

Check all that apply.

□ Smart Phone
□ Tablet
□ Laptop

10. Q7: Are you aware about Artificial Intelligence / Machine Learning

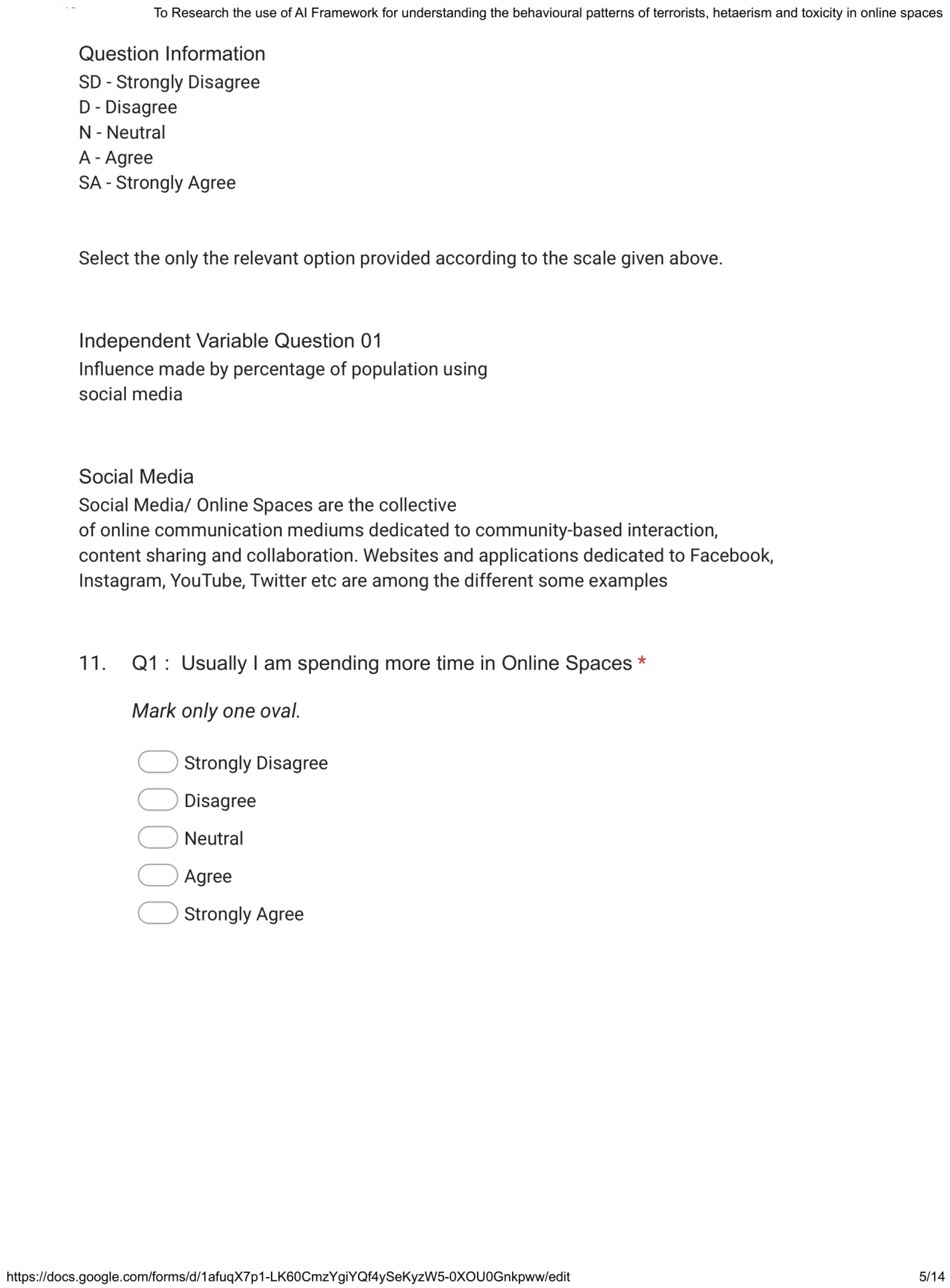
Mark only one oval.

○ Yes
○ No

Survey Part 02

Independent Variable Questions

https://docs.google.com/forms/d/1afuqX7p1-LK60CmzYgiYQf4ySeKyzW5-0XOU0Gnkpww/edit 4/14



To Research the use of AI Framework for understanding the behavioural patterns of terrorists, hetaerism and toxicity in online spaces

Question Information

SD - Strongly Disagree
D - Disagree
N - Neutral
A - Agree
SA - Strongly Agree

Select the only the relevant option provided according to the scale given above.

Independent Variable Question 01

Influence made by percentage of population using social media

Social Media

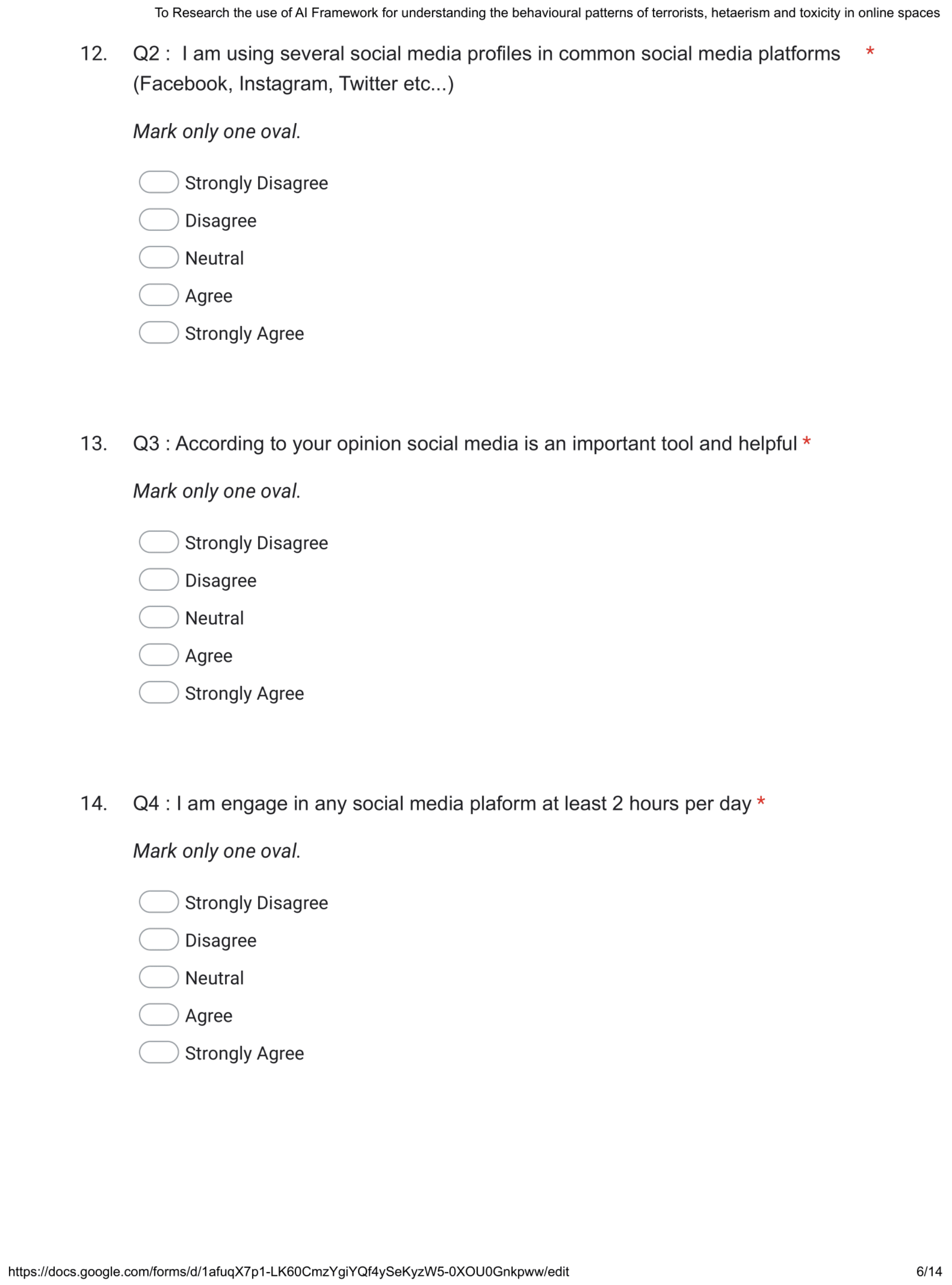
Social Media/ Online Spaces are the collective of online communication mediums dedicated to community-based interaction, content sharing and collaboration. Websites and applications dedicated to Facebook, Instagram, YouTube, Twitter etc are among the different some examples

11. Q1 : Usually I am spending more time in Online Spaces *

Mark only one oval.

○ Strongly Disagree
○ Disagree
○ Neutral
○ Agree
○ Strongly Agree

https://docs.google.com/forms/d/1afuqX7p1-LK60CmzYgiYQf4ySeKyzW5-0XOU0Gnkpww/edit 5/14



12. Q2 : I am using several social media profiles in common social media platforms (Facebook, Instagram, Twitter etc...) *

Mark only one oval.

○ Strongly Disagree
○ Disagree
○ Neutral
○ Agree
○ Strongly Agree

13. Q3 : According to your opinion social media is an important tool and helpful *

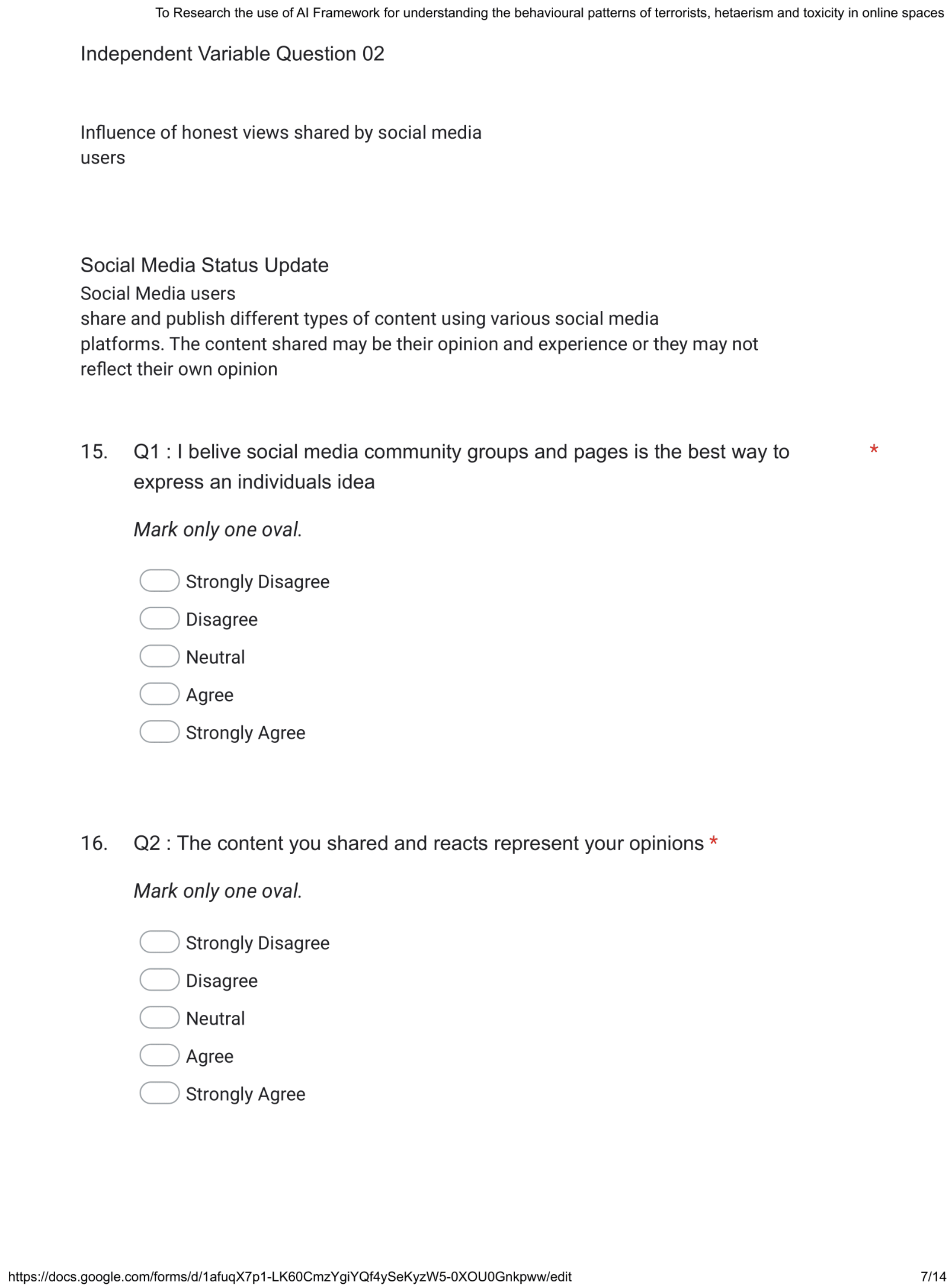
Mark only one oval.

○ Strongly Disagree
○ Disagree
○ Neutral
○ Agree
○ Strongly Agree

14. Q4 : I am engage in any social media plaform at least 2 hours per day *

Mark only one oval.

○ Strongly Disagree
○ Disagree
○ Neutral
○ Agree
○ Strongly Agree



To Research the use of AI Framework for understanding the behavioural patterns of terrorists, hetaerism and toxicity in online spaces

Independent Variable Question 02

Influence of honest views shared by social media users

Social Media Status Update

Social Media users share and publish different types of content using various social media platforms. The content shared may be their opinion and experience or they may not reflect their own opinion

15. Q1 : I belive social media community groups and pages is the best way to express an individuals idea *

Mark only one oval.

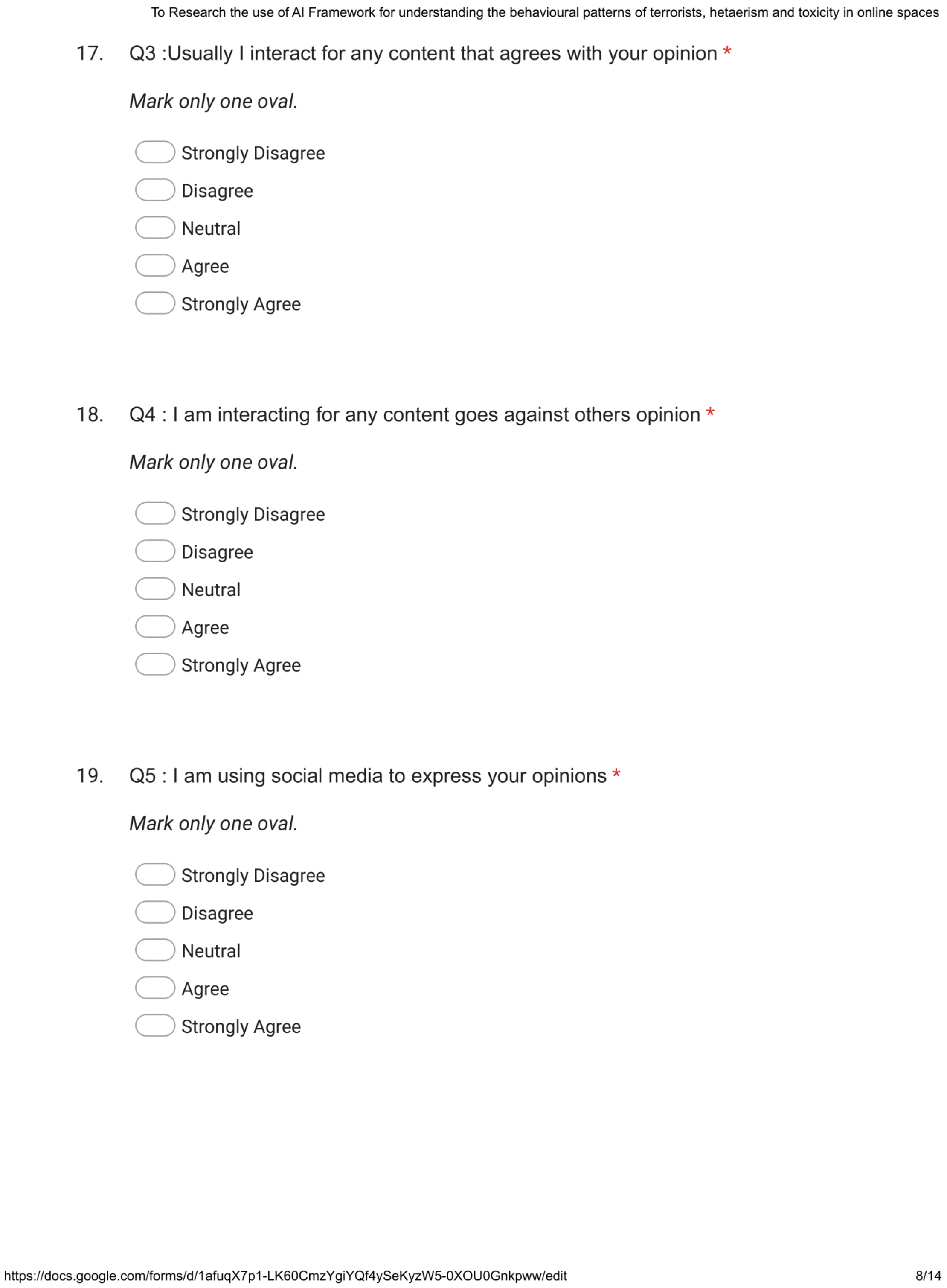
○ Strongly Disagree
○ Disagree
○ Neutral
○ Agree
○ Strongly Agree

16. Q2 : The content you shared and reacts represent your opinions *

Mark only one oval.

○ Strongly Disagree
○ Disagree
○ Neutral
○ Agree
○ Strongly Agree

https://docs.google.com/forms/d/1afuqX7p1-LK60CmzYgiYQf4ySeKyzW5-0XOU0Gnkpww/edit 7/14



17. Q3 :Usually I interact for any content that agrees with your opinion *

Mark only one oval.

○ Strongly Disagree
○ Disagree
○ Neutral
○ Agree
○ Strongly Agree

18. Q4 : I am interacting for any content goes against others opinion *

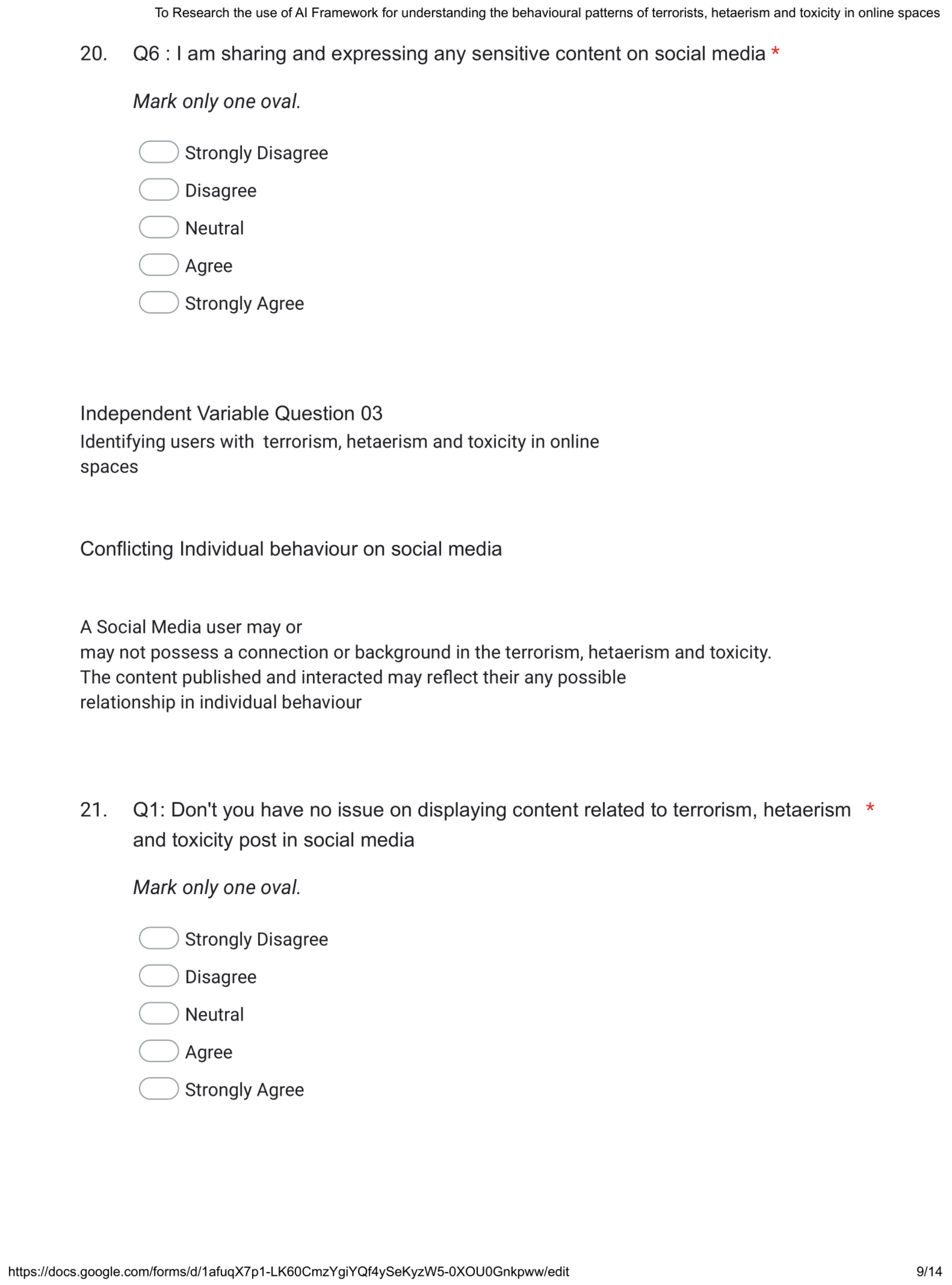
Mark only one oval.

○ Strongly Disagree
○ Disagree
○ Neutral
○ Agree
○ Strongly Agree

19. Q5 : I am using social media to express your opinions *

Mark only one oval.

○ Strongly Disagree
○ Disagree
○ Neutral
○ Agree
○ Strongly Agree



20. Q6 : I am sharing and expressing any sensitive content on social media *

Mark only one oval.

○ Strongly Disagree
○ Disagree
○ Neutral
○ Agree
○ Strongly Agree

Independent Variable Question 03
Identifying users with terrorism, hetaerism and toxicity in online spaces

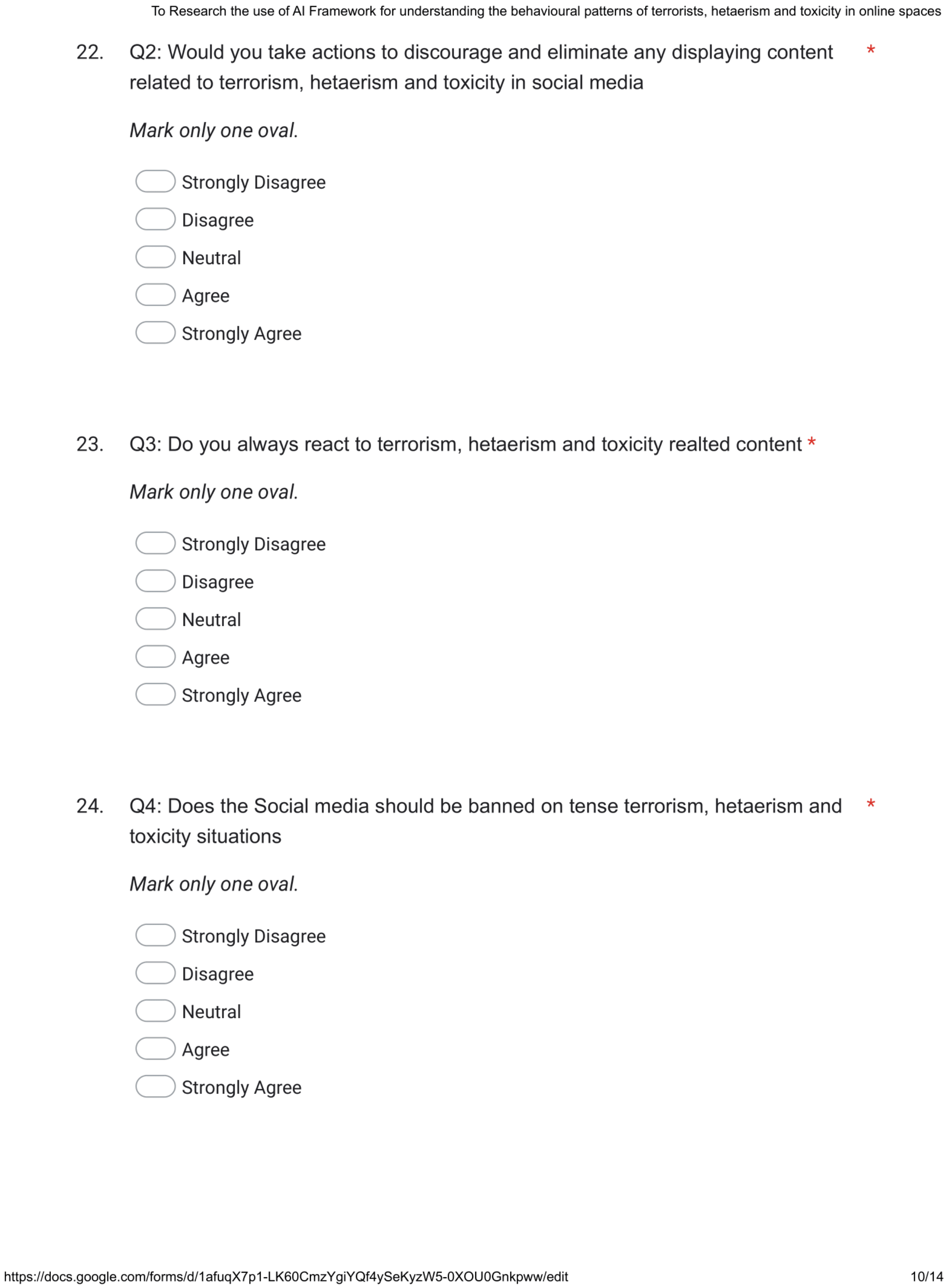
Conflicting Individual behaviour on social media

A Social Media user may or
may not possess a connection or background in the terrorism, hetaerism and toxicity.
The content published and interacted may reflect their any possible
relationship in individual behaviour

21. Q1: Don't you have no issue on displaying content related to terrorism, hetaerism and toxicity post in social media *

Mark only one oval.

○ Strongly Disagree
○ Disagree
○ Neutral
○ Agree
○ Strongly Agree



To Research the use of AI Framework for understanding the behavioural patterns of terrorists, hetaerism and toxicity in online spaces

22. Q2: Would you take actions to discourage and eliminate any displaying content related to terrorism, hetaerism and toxicity in social media *

Mark only one oval.

○ Strongly Disagree
○ Disagree
○ Neutral
○ Agree
○ Strongly Agree

23. Q3: Do you always react to terrorism, hetaerism and toxicity realted content *

Mark only one oval.

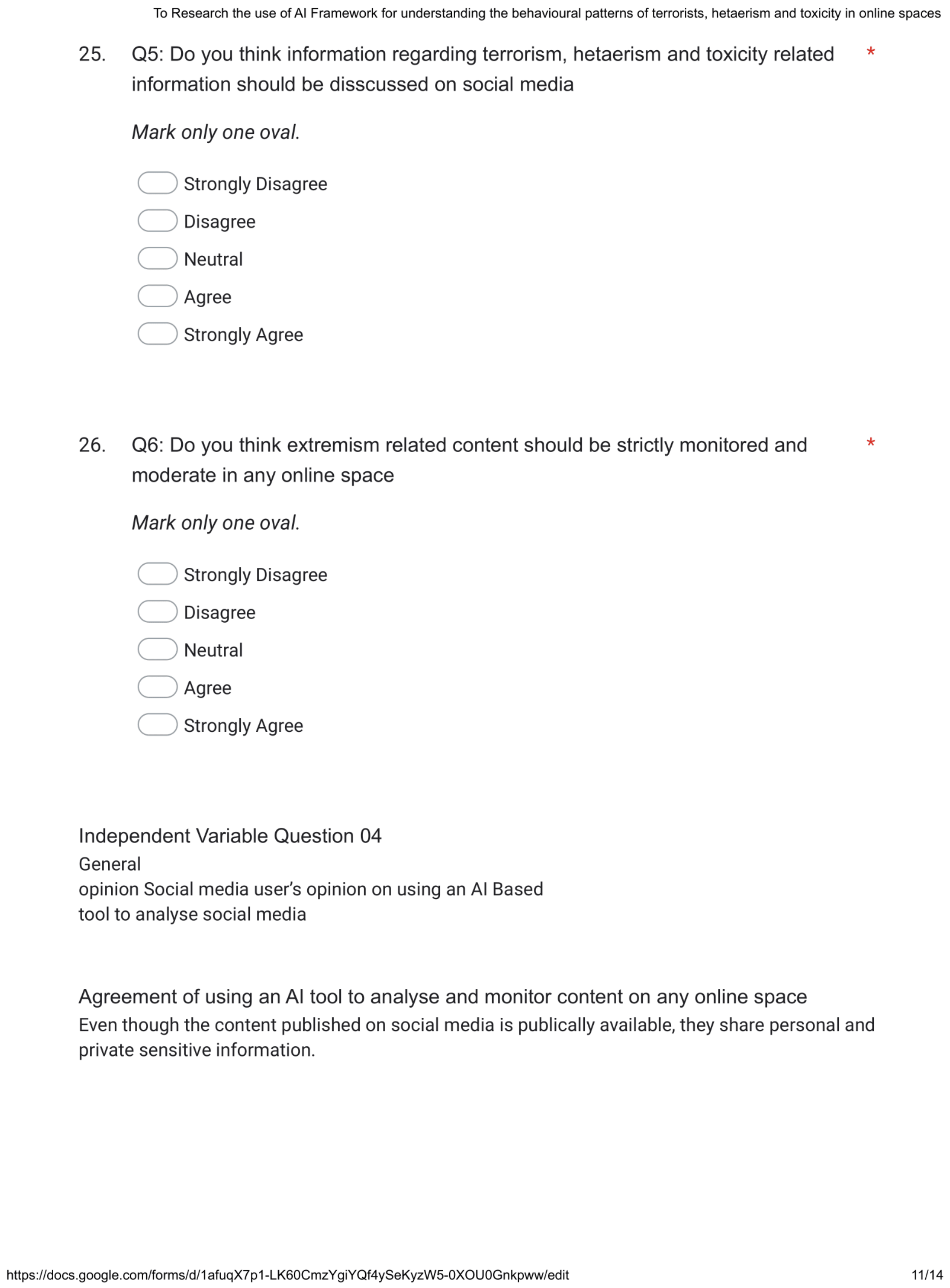
○ Strongly Disagree
○ Disagree
○ Neutral
○ Agree
○ Strongly Agree

24. Q4: Does the Social media should be banned on tense terrorism, hetaerism and toxicity situations *

Mark only one oval.

○ Strongly Disagree
○ Disagree
○ Neutral
○ Agree
○ Strongly Agree

https://docs.google.com/forms/d/1afuqX7p1-LK60CmzYgiYQf4ySeKyzW5-0XOU0Gnkpww/edit 10/14



25. Q5: Do you think information regarding terrorism, hetaerism and toxicity related information should be disscussed on social media *

Mark only one oval.

○ Strongly Disagree

○ Disagree

○ Neutral

○ Agree

○ Strongly Agree

26. Q6: Do you think extremism related content should be strictly monitored and moderate in any online space *

Mark only one oval.

○ Strongly Disagree

○ Disagree

○ Neutral

○ Agree

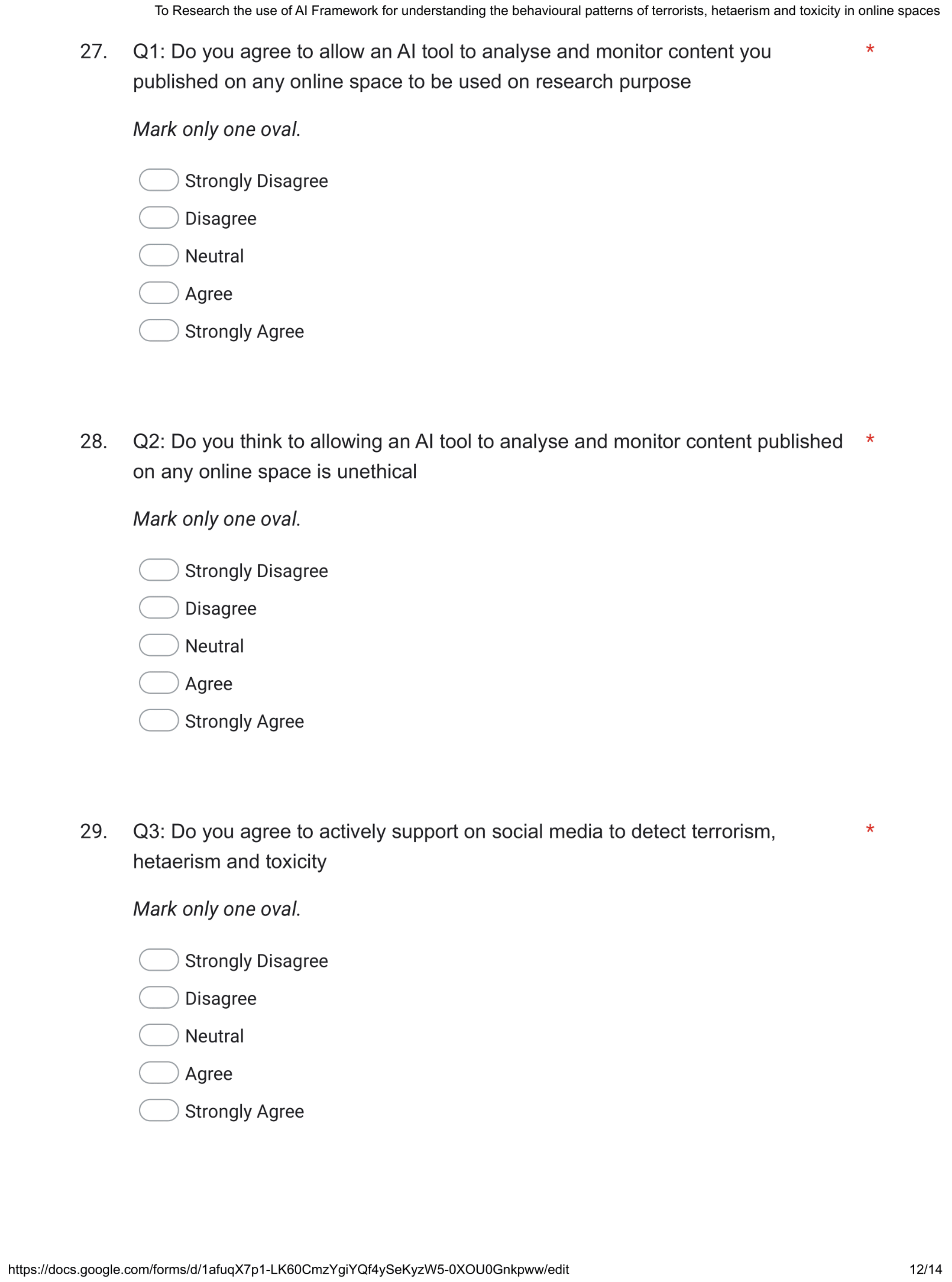
○ Strongly Agree

Independent Variable Question 04

General
opinion Social media user’s opinion on using an AI Based
tool to analyse social media

Agreement of using an AI tool to analyse and monitor content on any online space

Even though the content published on social media is publically available, they share personal and private sensitive information.



To Research the use of AI Framework for understanding the behavioural patterns of terrorists, hetaerism and toxicity in online spaces

27. Q1: Do you agree to allow an AI tool to analyse and monitor content you published on any online space to be used on research purpose *

Mark only one oval.

○ Strongly Disagree
○ Disagree
○ Neutral
○ Agree
○ Strongly Agree

28. Q2: Do you think to allowing an AI tool to analyse and monitor content published on any online space is unethical *

Mark only one oval.

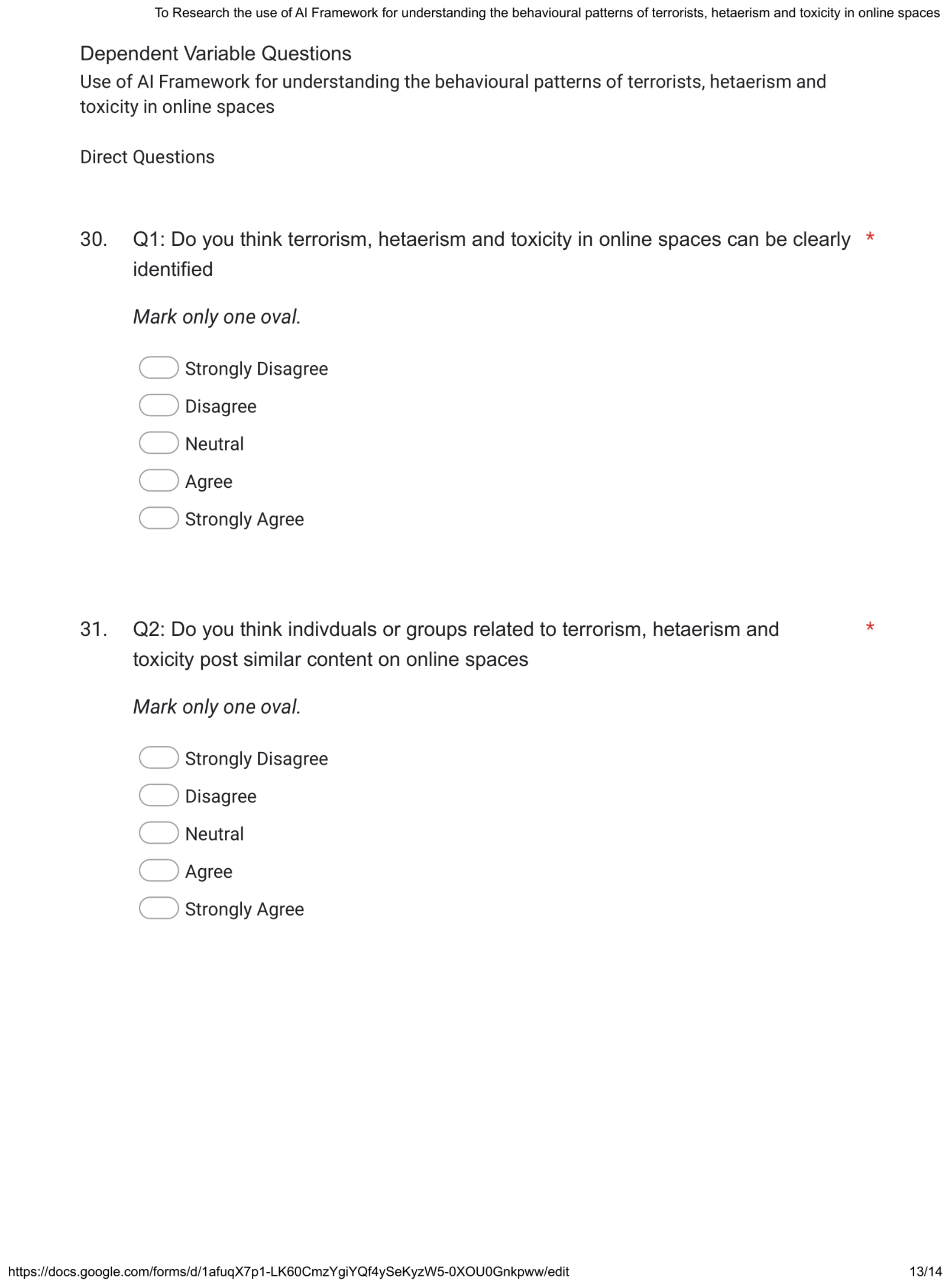
○ Strongly Disagree
○ Disagree
○ Neutral
○ Agree
○ Strongly Agree

29. Q3: Do you agree to actively support on social media to detect terrorism, hetaerism and toxicity *

Mark only one oval.

○ Strongly Disagree
○ Disagree
○ Neutral
○ Agree
○ Strongly Agree

https://docs.google.com/forms/d/1afuqX7p1-LK60CmzYgiYQf4ySeKyzW5-0XOU0Gnkpww/edit 12/14



Dependent Variable Questions

Use of AI Framework for understanding the behavioural patterns of terrorists, hetaerism and toxicity in online spaces

Direct Questions

30. Q1: Do you think terrorism, hetaerism and toxicity in online spaces can be clearly identified *

Mark only one oval.

○ Strongly Disagree

○ Disagree

○ Neutral

○ Agree

○ Strongly Agree

31. Q2: Do you think indivduals or groups related to terrorism, hetaerism and toxicity post similar content on online spaces *

Mark only one oval.

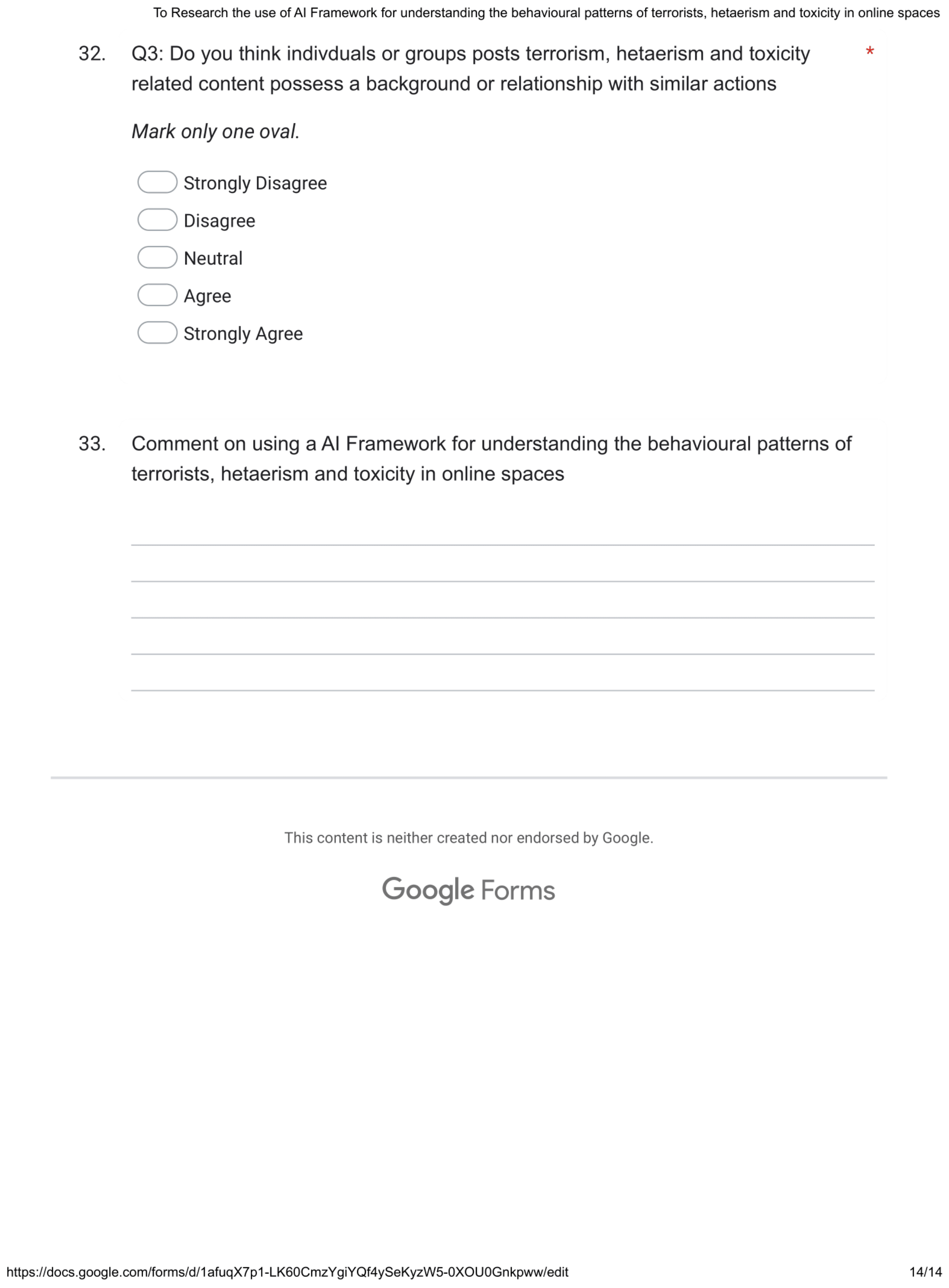
○ Strongly Disagree

○ Disagree

○ Neutral

○ Agree

○ Strongly Agree



To Research the use of AI Framework for understanding the behavioural patterns of terrorists, hetaerism and toxicity in online spaces

32. Q3: Do you think indivduals or groups posts terrorism, hetaerism and toxicity related content possess a background or relationship with similar actions *

Mark only one oval.

○ Strongly Disagree

○ Disagree

○ Neutral

○ Agree

○ Strongly Agree

33. Comment on using a AI Framework for understanding the behavioural patterns of terrorists, hetaerism and toxicity in online spaces

This content is neither created nor endorsed by Google.

Google Forms

https://docs.google.com/forms/d/1afuqX7p1-LK60CmzYgiYQf4ySeKyzW5-0XOU0Gnkpww/edit 14/14