In Sri Lanka, the Internet and social media are platforms that are widely open to almost everyone. Usage of Internet users is caped 14.6 million and social media is increasing day by day and up to date it is 7.2 million. People write, post, comment, and share their thoughts on these platforms, which is considered freedom of speaking of humans. Free speech has opened doors for everyone to speak up and react. Hate speech, cyberbullying, and online harassment have taken place due to the freedom of speech. Darkside of freedom of expression has led to threaten, abuse, harass, offend, and defaming individuals or entities. This study addresses the negative impacts that hate speech and hate crimes have on Sri Lankans. We seek to understand the emotional, social, and psychological impact these incidents exact on individuals and communities by looking at real-life experiences and perspectives.

The study also emphasizes how important it is to have efficient hate speech detection technologies on social media platforms. Such types of solutions are essential for preventing hate speech from spreading further since internet platforms are becoming more and more like breeding grounds for it. These technologies can help mitigate the negative effects of hate speech by promptly identifying and eliminating it, providing a more comfortable and secure environment on the internet.

In this study, I want to draw attention to the critical need for proactive steps toward preventing hate speech and hate crimes in Sri Lanka, highlighting the vital role that technology plays in preserving social harmony and well-being.

The use of social media and the internet has increased significantly in Sri Lanka in recent years, following worldwide trends toward increased digital connectivity. The nation's communication and interaction dynamics have experienced an important transformation due to the growing number of people using online platforms. But in addition to this digital transformation, hate speech, hate crimes, and cyberbullying on social media are widespread problems in Sri Lanka and many other countries.

despite the fact that these instances are extremely widespread, there is a worrying tendency to minimize or ignore the importance of them. While some people consider hate speech and cyberbullying to be just forms of free speech, others blame a lack of regulations or even dismiss them as harmless internet jokes. The truth is far from that, however, those who are subjected to this kind of online abuse frequently experience severe mental suffering as well as negative social consequences such as depression and suicide.

This study aims to address the urgent need for a proactive monitoring system customized for the Sri Lankan setting in light of these critical concerns. Our goal is to give an effective solution to protect the worth and well-being of Sri Lankan internet users by creating a cutting-edge tool that can identify hate speech and cyberbullying on social media platforms. By this research, we hope to strengthen the resilience and security of people and communities in the realm of the internet and develop an inclusive and respectful online culture.

Social media offers fresh platforms for communication, information exchange, and self-expression, and it has quickly become a part of Sri Lankan day to day life. But despite its advantages, abuse has resulted in alarming problems. Misinformation, addiction, cyberbullying, and hate crimes have become important issues that affect both individuals and communities. Gaining an understanding of the motivations behind social media use in Sri Lanka is essential to appreciating its influence on society. For a variety of purposes, such as news access, your business advancement, and connections upkeep people interact with these platforms. Examining these driving forces indicates the ways in which social media shapes public opinion and behavior. Creating a safe environment for all social media users by mitigating the harmful forces that are reaching through social media.

+ Growth of the usage of social media.

Sri Lankans have been using social media with increasing frequency in recent years, with platforms such as Facebook, Instagram, WhatsApp, and X growing in popularity. This growth may be related to factors like an increased young population, more accessible internet, and the widespread availability of connection methods.

According to GWI and data.io, there are more than 6.85 million social media users who are above 18 years and 37.2 percent of them are females and 67.8 percent are males.

social media has become an essential part of daily life, working as a main platform for entertainment, communication, and information exchange.

+ Current context of social media-related negativities

Social media-related problems like hate speech, hate crimes, and cyberbullying have become major concerns in the contemporary Sri Lankan setting, having a profound effect on both individuals and communities. Unfortunately, the ease with which information can be posted on social media platforms like Facebook and Twitter has led to the dissemination of harmful content, which has resulted in instances of harassment, threats, and discrimination. Hate crime victims frequently experience severe psychological suffering, anxiety about their safety, and social exclusion. Furthermore, the increase of hate speech and cyberbullying worsens tensions among communities and threatens the harmony of society.

+ Current barriers that have provided for the safety of users of social media

A growing number of Sri Lankans can benefit from the positive effects of enhanced online experience protection provided by the Sri Lankan police force established the Cybercrime Division. In addition, community groups and social activists are essential for supporting victims and encouraging users proper online behavior. By means of awareness programs and advocacy initiatives, they enable people to securely utilize social media and foster an online environment that values inclusion and respect. When combined, these programs offer a thorough strategy for preserving the dignity and general well-being of Sri Lankan social media users.

Sri Lanka has a boundless and widespread connectivity for social media and the most popular social media for content posting, sharing, and reacting is Facebook with over 6.85 million users according to the GWI. And data.io till January 2023.

Facebook has enabled its users to engage more with its facilities and features such as stories, short videos, live streaming, and mostly the posting of both text and content. Due to the high number of users incidents that are related to Facebook have also risen. Such as hate crimes, online fraud, harassment, threats, etc. To increase the security of the local users and also to provide overall cyber protection, Sri Lankan police have implemented the Cybercrime division.

Between 2019 and 2020, 3562 cases have been reported as cybercrimes and CERT has mentioned that 97.4 percent was from social media incidents. That is clear that social media is becoming a major issue and from the above-mentioned incidents, Facebook has become the most popular platform for cybercrimes.

+ lack of ethical behavior on Facebook

User behavior on Facebook has been a big issue where users have no idea about the content that they users are posting and do not know the consequences of posted content and justify hate-spread content using the terms of free speech and criticism.

Ethical behavior should be there with the users in both real and social media environments.

+ Inability to track fake profiles on Facebook

Even with the restrictions that are given by Facebook (META) people are still able to hide their identity using fake accounts. Most people use fake accounts to commit hate crimes including hate speech to defame an individual, community, or entity while fewer make positive impacts using fake accounts to the users of Facebook or the Facebook communities.

+ inability to differentiate criticism over hate speech

Free speech is expressing the point of view of a person towards a person, situation, or entity. Criticism is making a comment or an opinion about something, or situation, or an organization with previous experience. Hate speech is defaming, threatening, harassing, or using abusive terms against an individual or a community.

Every person should educate their selves with the correct definitions and ideologies of each term while using their opinions in both public and social media platforms such as Facebook.

+ Lack of tools and implementation to mitigate defaming content on Facebook

Current tools and systematic implementations are provided by Facebook to minimize the creation of fake profiles when a user shares or uploads hateful, sexist, or content that is inappropriate for the users. However, the implemented tool has its own bounds and barriers that they are not capable of doing against the above-mentioned content therefore to close the rabbit holes in the current system should be addressed and filled for a secure and more harmonized realm of internet and social media.

+ Lack of knowledge among users about the authorities related to cybercrimes

Users should know the authorities to address if that particular user has become a victim of a cybercrime or a hate crime. Disclosing the information related to the incident will make the victim safer and the social activists to encourage and provide an upper hand to secure the victim from the negative forces of the external parties.

+ Pages and “memes” are created to defame a certain community

Pages have become the most common community addressing method. Some pages are made on Facebook to defame a certain community or to build a target on a certain community. Therefore, regularizing or monitoring the content that is posted through the page and comments that its followers are dropping to their posts should be monitored and analyzed for the further goodness of the rest of the users.

What is the difference between hate speech and free speech?

What are methods that hate speech spread over Facebook?

What are the negative effects of hate speech?

How are cyberbullying and hate crimes related to each other?

What are the positive impacts of filtering hate speech?

Why do people use hate speech? And what are their motives?

What are the current hate speech detection methodologies?

Will hate speech detection affect to the freedom of speech?

1. To study the Sri Lankan Facebook usage and socio-impacts.
   1. 6.55 million users have been recorded for 2023 January. The community of Facebook has spread all across the island and it has increased drastically in recent years. This has affected both positively and negatively the society and the person's life. Facebook has become a threat to individuals' lives because it has become an addiction where people have lost their minds and given up on their work. Some have been isolated and become cyber victims through the fraud, harassment, and online crimes that have taken place through Facebook.

Facebook is a platform that we can utilize to build a productive environment where people can use it to make their lives easier by creating online helping communities, helping and protecting the users, also to spread news and updating certain situations in real-time, business creation, and as a marketing platform, etc. Facebook has its own both negative and positive impacts based on the user and their community.

1. To study about are the hate crimes, hate speech, and cybercrimes that occur because of Facebook
   1. Facebook's extensive user base and significant influence over online discourse have led to several instances of hate crimes, hate speech, and cybercrimes worldwide. For example, the platform has come under fire for aiding in the spreading of hate speech that calls for the violence against marginalized groups. Facebook has also been used to plan and carry out hate crimes, including the live-streaming of violent assaults. Additionally, phishing schemes and identity theft have become increasingly common on the network, taking advantage of users' weaknesses with regard to their personal data.
2. To study mitigation techniques and regulation steps that are taken to control hate speech in the world
   1. The negative consequences of hate speech have been restrained globally via the use of different mitigating strategies and legal procedures in reaction to its spread online. Using artificial intelligence systems and content moderation algorithms is one popular method for quickly identifying and eliminating hate speech. For instance, abusive content is now automatically detected and removed by Twitter and YouTube using pre-established criteria. The identified content is then reviewed by human moderators. In addition, a number of nations have proposed laws to hold online platforms responsible for allowing hate speech. As an example, the Digital Services Act proposed by the European Union imposes strict guidelines mandating that internet companies promptly delete any unlawful information, including hate speech, or risk paying severe penalties. Furthermore, the goal of awareness campaigns and educational programs is to provide users with the knowledge and skills necessary to identify and properly report hate speech, promoting an attitude of good citizenship and responsible online conduct. Through a blend of technological, legal, and instructional approaches, global stakeholders strive to establish online spaces that are safer, more welcoming, and devoid of hate speech's deleterious impacts.
3. To study the Sri Lankan context of hate speech crimes and their controllability of it.
   1. Hate speech has always been a problem in Sri Lanka, frequently increasing tensions between different ethnic and religious groups. Studies by Gunaratne and Seneviratne (2020) demonstrate how common hate speech is on social media, especially when it comes to targeting minorities. Also, Sri Lankan police have made a separate division named the Cybercrime Division to address cyber-related issues. Different mitigating techniques have been put into place to address this problem. The International Covenant on Civil and Political Rights Act, for example, was passed by the Sri Lankan government and makes hate speech and incitement to violence illegal (United Nations, 2017). Social media companies have also implemented content moderation guidelines and hate speech detection and removal capabilities. However, because of Sri Lanka's complicated sociopolitical environment, difficulties continue to arise in properly implementing these policies (Jayaweera, 2019).
4. To implement a tool for detecting Facebook hate speech
   1. Developing a Facebook hate speech detection tool is crucial for creating a more respectful and secure online community. Because of the platform's extensive reach and power, the tool can quickly detect and delete offensive content, shielding users from the negative repercussions of hate speech. The tool contributes to the development of a more welcoming online community where people can express themselves without worrying about harassment or discrimination by encouraging a culture of respect and tolerance. In the end, purchasing such technology shows Facebook's dedication to maintaining user security and welfare on its network.
5. To study the pros and cons of the tool.
   1. Creating a hate speech detection tool for Facebook has benefits and drawbacks. Positively, by quickly detecting and eliminating harmful content, this technology could significantly enhance user safety and promote a more welcoming and joyful online community. Furthermore, it would show a dedication to maintaining platform integrity and community standards, earning the trust of users and stakeholders.

But it's important to take into account any possible downsides. Accuracy issues with automated detection algorithms might result in excessive censorship and the suppression of legitimate speech. Furthermore, creating and maintaining such a tool is heavy on resources and presents difficult ethical dilemmas relating to verbal freedom and censorship. Despite these difficulties, platforms looking to encourage safety and civility online should consider the potential rewards of putting in place a hate speech detection technology.

What is hate speech? originating and what free speech

Impacts of hate speech and necessity of such tool

Current tools and methods of tracking hate speech

Figure 1. depicts the methodology of the hate speech detection tool with its origination to the very end till either posting the text input or pooling it without posting it on Facebook. Each stepper includes one or more processes and each stepper needs to be completed before starting the next stepper. The user input text will pre-processed and will create a token as reference to the pre-processed text and analyzed against the dataset to identify whether the user input text contains any hateful content. Such content will eliminated and will pool in a DB with the details of the user and the content.

The research aim is to counter the above-mentioned channels which are related to the assurance of the Facebook user community due to the online hate speech on Facebook. Lay out for the above-mentioned method and the construction of the conceptual frame have been countering the neediness of the gap that has been identified in the context of Sri Lankan Facebook users.

Research Methodology

The conceptual framework has been analyzed with the solution against the population the sample of the internet users and the number of Facebook users, collecting the extract quality of the study to distinguish.

Population

The study has mainly focused on Facebook users and hate speech text content using English and Singlish language. Therefore the population of the study is all Facebook users in Facebook.

All the sectors that are using Facebook.

All the users who are 18 years old and 18-plus people

Sample

The study is about Hate speech on Facebook in Sri Lanka. The study is focused on English and Singlish language text detection, which excludes the images, sounds, and videos related to hate speech.

Sample sector 1 – English and Singlish comments.

Sample sector 2 – English and Singlish Status posts.

Data collection methods

One to one interview,

Conducting a series of one-to-one interviews to pre-define the key roles that have been

identified in the previous chapters.

• Facebook users,

• Legal background of the hate speech.

Survey / Questioner,

A survey is to be conducted about general information that needs to be sent to everyone.

Sample data sets

data set needs to be created to create the model to identify the hate speech against the preprocessed text input.

Observations

Process observation on how and what steps are taken for the hate speech detection and how the authorities and the meta are taken to mitigate the hate speech-related content.

Quantitative analysis

Descriptive statistics

This helps to get the basic statistical point of view such as the mean median mode standard deviation and variance to identify the basic tendencies between the variables, here the author of the study can use them directly and indirectly for the study.

Regression analysis

The relationship and interconnection between a dependent variable and an independent variable can be evaluated and the patterns among them to make the study more accurate and relevant.

Inferential statistics

Inferential statistical analysis is an advanced method of analyzing data and creating patterns, identifying co-relations and inter-related relationships between variables, and hypothesis analysis using the collected population and sample data

Qualitative analysis method

Thematic Analysis

The thematic analysis evaluates the qualitative data to provide/create an idea or concept based on the themes. The questionnaire has been created from both ends of quantitative and qualitative, where thematic analysis helps with the qualitative part of the survey.

Content Analysis

Analyzing documents, audio, videos, and various data types. Observations and document analysis methods are analyzed under the content analysis to identify the respective relations

Hypothesis

Hypothesis 1 – Implementing the hate speech detection tool will ensure the safety of Sri Lankan Facebook users.

Hypothesis 2 – The Hate speech detection tool will increase the number of users on Facebook due to the safe environment of the platform.

Hypothesis 3 – The Hate speech detection tool might fail to detect both languages and will not be able to achieve the study objectives.

Limitations of the study

Data collection - Since the research topic is related to hate speech on Facebook, it's not easy to get information about a personal experience and the social media life of a person is confidential.

Sample selection - Since the population is large, it is very challenging to select a perfect sample that is most suitable for the study.

Increasing engagement - Engaging and ready to help individuals are rare in the current context. Therefore, finding such individuals among the selected sample is also can be considered as a limitation.

Fewer data for modeling – Data sets for the Singlish language need to be created since such data sets haven’t been created.

In conclusion, this study addresses the issue that is underrated and not heavily spoken among the society in Sri Lanka, hate speech has become a major issue and its criticality has been life-threatening to its victims. Therefore implementing a tool to identify both English and Singlish which is a user-created language that Sri Lankan users created by combining both English and Sinhala because Singlish is widely spread across the island.

Singlish was never filtered and never been detected in the systems and this was a crucial issue that Facebook users have been facing for a long period of time.

Furthermore, the suggested tool is able to pool the hate speech content with the user details for further uses.