M.KUMARASAMY
COLLEGE OF ENGINEERING
NAAC Accredited Autonomous Institution
Approved by AICTE & Affiliated to Anna University
ISO 9001:2015 Certified Institution
Thalavapalayam, Karur – 639 113.

CYBERBULLYING TEXT CLASSIFICATION USING RNN TECHNIQUES

A MINOR PROJECT- IV REPORT

Submitted By

MIDHUN.R	(927621BEC120)
KAVINESH.A. S	(927621BEC081)
BALAKUMAR.A	(927621BEC302)
MOHAMED ANAS. S	(927621BEC121)

BACHELOR OF ENGINEERING

in

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

M.KUMARASAMY COLLEGE OF ENGINEERING

(Autonomous)

KARUR – 639 113

MAY 2024

M.KUMARASAMY COLLEGE OF ENGINEERING, KARUR

BONAFIDE CERTIFICATE

Certified that this **18ECP106L -Minor Project IV** Report **“CYBERBULLYING TEXT CLASSIFICATION USING RNN TECHNIQUES”** is the Bonafide work of **“MIDHUN. R(927621BEC120), KAVINESH. A. S(927621BEC081), BALAKUMAR. A(927621BEC302), MOHAMED ANAS. S(927621BEC121)”** who carried out the project work under my supervision in the academic year **2023 – 2024 – EVEN SEMESTER.**

SIGNATURE
Dr. A. KAVITHA B.E., M.E., Ph.D.,
HEAD OF THE DEPARTMENT,
Professor,
Department of Electronics and
Communication Engineering,
M. Kumarasamy College of Engineering,
Thalavapalayam,
Karur-639113.

SIGNATURE
Mr. P. T. SIVAGURUNATHAN M. E. MBA., (Ph. D)
SUPERVISOR,
Assistant Professor,
Department of Electronics and
Communication Engineering,
M. Kumarasamy College of Engineering,
Thalavapalayam,
Karur-639113.

This report has been submitted for the **18ECP106L - Minor Project - IV** Final Review held at M. Kumarasamy College of Engineering, Karur on ______________________

PROJECT COORDINATOR

INSTITUTION VISION AND MISSION

Vision

To emerge as a leader among the top institutions in the field of technical education.

Mission

M1: Produces smart technocrats with empirical knowledge who can surmount the global challenges

M2: Create a diverse, fully engaged, learner-centric campus environment to provide quality education to the students

M3: Maintain mutually beneficial partnerships with our alumni, industry, and Professional associations

DEPARTMENT VISION, MISSION, PEO, PO AND PSO

Vision

To empower the Electronics and Communication Engineering students with emerging technologies, professionalism, innovative research and social responsibility.

Mission

M1: Attain the academic excellence through innovative teaching learning process, research areas & laboratories and Consultancy projects.

M2: Inculcate the students in problem solving and lifelong learning ability.

M3: Provide entrepreneurial skills and leadership qualities.

M4: Render the technical knowledge and skills of faculty members.

Program Educational Objectives

PEO1: **Core Competence:** Graduates will have a successful career in academia or industry associated with Electronics and Communication Engineering.

PEO2: **Professionalism:** Graduates will provide feasible solutions for the challenging problems through comprehensive research and innovation in the allied areas of Electronics and Communication Engineering.

PEO3: **Lifelong Learning:** Graduates will contribute to the social needs through lifelong learning, practicing professional ethics and leadership quality

Program Outcomes

PO 1: Engineering knowledge: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.

PO 2: Problem analysis: Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.

PO 3: Design/development of solutions: Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.

PO 4: Conduct investigations of complex problems: Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.

PO 5: Modern tool usage: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.

PO 6: The engineer and society: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.

PO 7: Environment and sustainability: Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.

PO 8: Ethics: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.

PO 9: Individual and team work: Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.

PO 10: Communication: Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.

PO 11: Project management and finance: Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.

PO 12: Life-long learning: Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

Program Specific Outcomes

PSO1: Applying knowledge in various areas, like Electronics, Communications, Signal processing, VLSI, Embedded systems etc., in the design and implementation of Engineering application.

PSO2: Able to solve complex problems in Electronics and Communication Engineering with analytical and managerial skills either independently or in team using latest hardware and software tools to fulfill the industrial expectations.

Abstract	**Matching with POs, PSOs**
Cyberbullying Text Classification Using Rnn Techniques	PO1, PO2, PO3, PO4, PO5, PO6, PO7, PO8, PO9, PO10, PO11, PO12, PSO1, PSO2

ACKNOWLEDGEMENT

Our sincere thanks to **Thiru. M. Kumarasamy, Founder and Dr. K. Ramakrishnan, Chairman** of **M. Kumarasamy College of Engineering** for providing extraordinary infrastructure, which helped us to complete this project in time.

It is a great privilege for us to express our gratitude to **Dr. B.S. Murugan., B.Tech., M. Tech., Ph.D., Principal** for providing us right ambiance to carry out this project work.

We would like to thank **Dr. A. Kavitha, B.E., M.E., Ph.D., Professor and Head, Department of Electronics and Communication Engineering** for his unwavering moral support and constant encouragement towards the completion of this project work.

We offer our wholehearted thanks to our **Project Supervisor, Mr. P. T. Sivagurunathan M. E. MBA., (Ph. D), Assistant Professor,** Department of Electronics and Communication Engineering for his precious guidance, tremendous supervision, kind cooperation, valuable suggestions and support rendered in making our project to be successful.

We would like to thank our **Minor Project Co-Ordinator, Dr. K. Karthikeyan, B.E., M.Tech., Ph.D., Associate Professor,** Department of Electronics and Communication Engineering for his kind cooperation and culminating in the successful completion of this project work. we are glad to thank all the faculty members of the department of electronics and communication engineering for extending a warm helping hand and valuable suggestions throughout the project. words are boundless to thank our parents and friends for their motivation to complete this project successfully.

ABSTRACT

Cyberbullying has emerged as a pressing concern in the digital age, posing significant challenges to the mental and emotional well-being of individuals, especially adolescents and young adults. With the proliferation of social media platforms and online communication channels, instances of cyberbullying have increased exponentially, necessitating effective strategies for detection and mitigation. This research focuses on employing Recurrent Neural Network (RNN) techniques for the classification of cyberbullying text, aiming to develop robust models capable of identifying and addressing instances of online harassment and abuse. The study begins with an extensive review of existing literature on cyberbullying, exploring its various forms, underlying causes, and psychological impacts. It examines previous research efforts in text classification and sentiment analysis, highlighting the limitations of traditional machine learning approaches in handling the nuanced nature of cyberbullying texts. By leveraging the sequential nature of text data, RNNs offer a promising alternative, enabling the capture of temporal dependencies and contextual nuances crucial for accurate classification. Furthermore, the study investigates the generalizability of the developed models across different types of cyberbullying, including verbal harassment, social exclusion, cyberstalking, and online rumours. Transfer learning techniques are explored to enhance the models' adaptability to diverse cyberbullying scenarios, enabling them to recognize subtle variations in language and behaviour indicative of harassment.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACRONYM		ABBREVIATION
AI	-	Artificial Intelligence
ANN	-	Artificial Neural Networks
API	-	Application Programming Interface
EDA	-	Exploratory Data Analysis
FN	-	False Negatives
FP	-	False Positives
GRU	-	Gated Recurrent Unit
LSTM	-	Long Short-Term Memory
NLP	-	Natural Language Processing
NLTK	-	Natural Language Tool Kit
RNN	-	Recurrent Neural Networks
TN	-	True Negatives
TP	-	True Positives

CHAPTER 1

INTRODUCTION

1.1 CYBERBULLYING

Cyberbullying, a prevalent issue in the digital age, manifests in various forms across social media platforms, messaging apps, and online forums. It involves the use of electronic communication to intimidate, harass, or demean individuals or groups, often with malicious intent. With the rise of internet accessibility and the omnipresence of digital devices, cyberbullying has become a significant concern affecting individuals of all ages, particularly adolescents and young adults. Traditional methods of identifying and combating cyberbullying have been insufficient in addressing the complex nature of online interactions. However, advancements in machine learning and natural language processing (NLP) techniques offer promising solutions for detecting and mitigating cyberbullying behaviour effectively.

1.2 RECURRENT NEURAL NETWORKS

Recurrent Neural Networks (RNNs) have emerged as powerful tools for text classification tasks due to their ability to capture sequential dependencies in data. RNNs, a type of artificial neural network designed to process sequential data, have shown remarkable success in various natural language processing tasks, including sentiment analysis, language translation, and text generation. By leveraging the sequential nature of text data, RNNs can effectively capture contextual information, making them well-suited for detecting subtle nuances and patterns indicative of cyberbullying behaviour within textual content. This project aims to explore and implement RNN-based techniques for the classification of cyberbullying text, with the goal of developing an efficient and accurate model for identifying instances of cyberbullying in online communication. Through the utilization of labelled datasets containing examples of cyberbullying text, the model will be trained to distinguish between cyberbullying and non-cyberbullying content, enabling automated detection and classification of harmful online behaviour. The significance of this project lies in its potential to contribute to the development of proactive measures for combating cyberbullying, thereby fostering safer and more inclusive online environments. By providing a mechanism for real-time detection and intervention, the proposed RNN-based classification system can aid platform moderators, educators, and parents in addressing cyberbullying incidents promptly, thereby mitigating their adverse effects on victims' mental health and well-being.

1.3 LONG SHORT-TERM MEMORY

LSTMs are Long Short-Term Memory networks that use (ANN) Artificial Neural Networks in the field of Artificial Intelligence (AI) and deep learning. In contrast to normal feed-forward neural networks, also known as recurrent neural networks, these networks feature feedback connections. Unsegmented, connected handwriting recognition, robot control, video gaming, speech recognition, machine translation, and healthcare are all applications of LSTM. LSTMs Long Short-Term Memory is a type of RNNs Recurrent Neural Network that can detain long-term dependencies in sequential data. LSTMs are able to process and analyse sequential data, such as time series, text, and speech. They use a memory cell and gates to control the flow of information, allowing them to selectively retain or discard information as needed and thus avoid the vanishing gradient problem that plagues traditional RNNs. LSTMs are widely used in various applications such as natural language processing, speech recognition, and time series forecasting.

CHAPTER 2

LITERATURE SURVEY

Cyberbullying, the act of using digital platforms to harass, intimidate, or harm individuals, has become a prevalent issue in today's interconnected world. Detecting and addressing cyberbullying incidents in a timely manner is crucial to protect individuals from its detrimental effects. In recent years, machine learning algorithms have emerged as promising tools for automated cyberbullying detection, with the Naive Bayes algorithm receiving considerable attention due to its simplicity and effectiveness in natural language processing tasks. Various studies have explored the application of machine learning algorithms, including Naïve Bayes, in detecting cyberbullying in text-based data. Wang et al. (2018) conducted an extensive comparative analysis of several machine learning algorithms for cyberbullying detection, including Naive Bayes, Support Vector Machines, and Random Forests. The results demonstrated that Naive Bayes achieved competitive performance in terms of accuracy and efficiency, making it a viable choice for cyberbullying detection. Furthermore, research by De Silva et al. (2019) focused on linguistic features and their impact on cyberbullying detection. Their study revealed that Naive Bayes performed well when incorporating various linguistic features such as word frequencies, part-of-speech tags, and sentiment analysis. The authors emphasized the importance of feature engineering in improving the performance of the Naive Bayes algorithm for cyberbullying detection. In addition, studies have highlighted the significance of dataset quality and diversity for effective cyberbullying detection. Mishra et al. (2020) emphasized the need for balanced datasets containing a variety of cyberbullying instances to train and evaluate machine learning models accurately. They highlighted the importance of incorporating different types of cyberbullying, such as direct threats, insults, and exclusion, into the dataset to capture the complexity of cyberbullying behaviour. Moreover, the limitations of the Naive Bayes algorithm have been addressed in the literature. Kumar and Bappi (2017) discussed the assumptions of feature independence in Naive Bayes and its potential impact on cyberbullying detection. They proposed techniques to handle dependencies between features and improve the algorithm's performance. While the Naive Bayes algorithm has shown promise in cyberbullying detection, it is important to acknowledge the evolving nature of cyberbullying and the challenges it poses to detection systems. New forms of cyberbullying, such as image-based harassment and subtle indirect attacks, require ongoing research and adaptation of algorithms to effectively detect and prevent such incidents. In summary, the literature demonstrates that the Naive Bayes algorithm has been widely explored for cyberbullying detection in text-based data. Its simplicity, efficiency, and effectiveness in incorporating linguistic features make it a viable choice for automated detection systems. However, the literature also highlights the importance of dataset quality, feature engineering, and addressing the limitations of the Naive Bayes algorithm to achieve optimal performance in cyberbullying detection. Future research should focus on addressing these aspects to advance the field of automated cyberbullying detection using the Naive Bayes algorithm.

CHAPTER 3

EXISTING SYSTEM

Cyberbullying has become an increasingly prevalent issue in today's digital age, with the proliferation of social media platforms and online communication channels. In response to this growing concern, various techniques and approaches have been explored to detect and mitigate cyberbullying instances effectively. One such technique utilized in the existing system is text classification using Recurrent Neural Network (RNN) models. RNNs are a class of artificial neural networks well suited for sequential data processing tasks, making them particularly relevant for analysing text data. In the existing system, RNNs are leveraged to automatically classify text data into categories such as cyberbullying or non-cyberbullying.

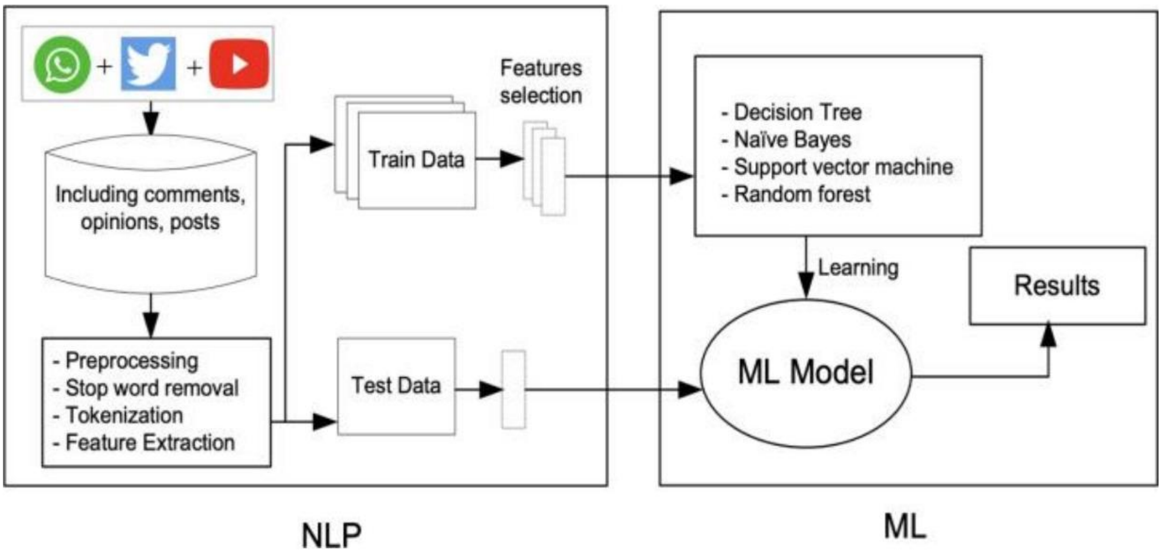


Fig 1.1 Existing System

The process begins with the collection of textual data from various online sources, including social media platforms, forums, and messaging apps. This data is then pre-processed to clean and normalize it, removing irrelevant information and standardizing the format for analysis. Preprocessing steps may include tokenization, stemming, and removing stop words to prepare the text for input into the RNN model. Once the data is pre-processed, it is split into training and testing sets to train the RNN model. The RNN architecture typically consists of recurrent layers that process sequential input data, such as Long Short-Term Memory (LSTM) or Gated Recurrent Unit (GRU) cells. These layers enable the model to capture dependencies and patterns in the text data over time, which is crucial cyberbullying instances.

CHAPTER 4

PROPOSED SYSTEM

The proposed system for cyberbullying detection is based on the Naive Bayes algorithm. It aims to classify social media posts or comments as either cyberbullying or non-cyberbullying. The system works by first preprocessing the text data, which involves removing stop words, punctuation, and converting all words to lowercase. After preprocessing, the system extracts feature from the text data, which are then used to train a Naive Bayes classifier. The features used in the system include the frequency of each word in the text, as well as the presence or absence of certain keywords or phrases commonly associated with cyberbullying.

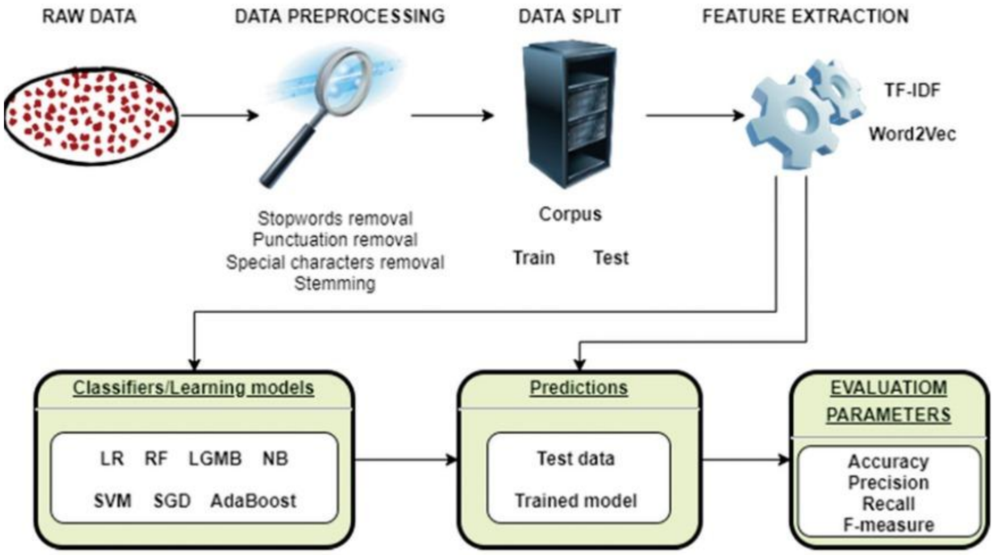


Fig 1.2 Proposed System

To improve the accuracy of the classifier, the system also incorporates a user-based approach. This involves analysing the social media user's previous posts and comments to determine their typical language and tone. This information is then used to further refine the Naive Bayes classifier. The proposed system is designed to be scalable and adaptable to various social media platforms.

It can be integrated as a plugin or API for social media platforms, allowing for real-time detection and removal of cyberbullying content. Overall, the proposed system aims to provide an effective solution for detecting and preventing cyberbullying on social media platforms, ultimately promoting a safer and more inclusive online environment.

CHAPTER 5

METHODOLOGY

5.1 PROBLEM STATEMENT

Cyberbullying has emerged as a pervasive issue in the digital age, causing psychological harm and social distress to its victims. With the widespread use of social media platforms and online communication channels, individuals, particularly adolescents, are increasingly vulnerable to various forms of cyberbullying, including harassment, intimidation, and defamation. Traditional methods of monitoring and addressing cyberbullying have proven inadequate, necessitating innovative approaches leveraging advanced technologies. The problem at hand revolves around the need to effectively identify and classify instances of cyberbullying within textual content across digital platforms. While conventional machine learning techniques have been employed for text classification, their efficacy in accurately discerning nuanced forms of cyberbullying remains limited. Recurrent Neural Network (RNN) techniques offer a promising avenue for addressing this challenge, owing to their ability to capture sequential dependencies and contextual information inherent in textual data.

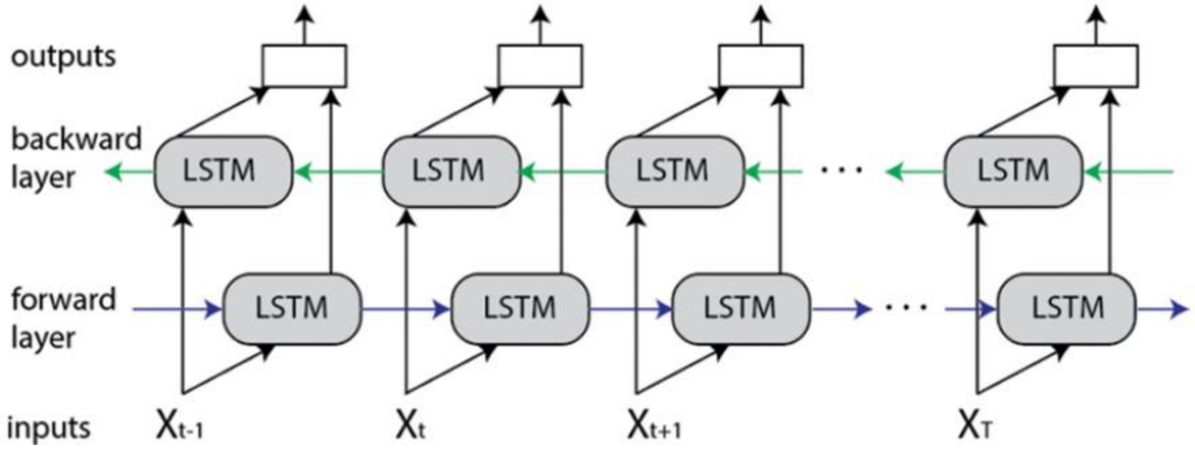


Fig 1.3 Methodology

However, the effectiveness of RNN-based models in cyberbullying text classification hinges on several critical factors that require thorough investigation and optimization. One such factor is the selection and preprocessing of input data, which involves extracting relevant features while mitigating noise and irrelevant information. Additionally, the design and architecture of the RNN model play a pivotal role in its Software Requirement: performance, necessitating careful consideration of parameters such as network depth, cell type, and regularization techniques. Furthermore, the scarcity of labelled datasets specifically tailored for cyberbullying presents a significant hurdle in training robust RNN models. Acquiring and annotating large scale datasets encompassing diverse forms of cyberbullying poses logistical and ethical challenges, thereby underscoring the importance of data augmentation and transfer learning strategies to enhance model generalization and adaptability. Moreover, the dynamic nature of online communication platforms necessitates the development of real-time cyberbullying detection systems capable of swiftly identifying and addressing abusive behaviour. Integrating RNN-based classifiers into such systems requires optimizing inference speed and computational efficiency without compromising classification accuracy, thereby ensuring timely intervention and mitigation of cyberbullying incidents. Additionally, the ethical implications surrounding the deployment of automated cyberbullying detection systems warrant careful consideration, particularly concerning privacy, bias, and unintended consequences. Striking a balance between algorithmic efficiency and ethical responsibility entails implementing transparent and accountable mechanisms for model evaluation, validation, and bias mitigation. Addressing the aforementioned challenges requires a multifaceted approach encompassing data collection, preprocessing, model development, evaluation, and deployment. Collaborative efforts involving researchers, educators, policymakers, and technology companies are essential to foster a holistic ecosystem for combating cyberbullying and promoting digital well-being.

5.2 OBJECTIVE

Cyberbullying has emerged as a critical issue in the digital age, with detrimental impacts on individuals' mental health, well-being, and even safety. In response to this pressing concern, the objective of this project is to develop a robust text classification system utilizing Recurrent Neural Network (RNN) techniques to accurately identify instances of cyberbullying in online text data. Our primary objective is to create a text classification model that can accurately detect instances of cyberbullying in various forms of online communication, including social media posts, comments, emails, and chat messages. We strive to achieve high precision, recall, and F1 scores to minimize false positives and negatives, ensuring effective identification of cyberbullying instances. Our primary objective is to create a text classification model that can accurately detect instances of cyberbullying in various forms of online communication, including social media posts, comments, emails, and chat messages. We strive to achieve high precision, recall, and F1 scores to minimize false positives and negatives, ensuring effective identification of cyberbullying instances. To ensure the scalability and applicability of our approach across different platforms, languages, and demographics, we seek to develop a model that can generalize well to unseen data and adapt to varying linguistic styles and cultural contexts. This involves robust model training, validation, and testing procedures, along with techniques such as transfer learning and domain adaptation. Cyberbullying has become a prevalent issue in the digital age, with harmful consequences on individuals' mental health and well-being. The objective of employing Recurrent Neural Networks (RNN) in text classification is to effectively identify instances of cyberbullying within online communication channels. RNNs, known for their ability to process sequential data, offer promising avenues for analysing textual content and detecting patterns indicative of cyberbullying behaviour. In this approach, the first step involves preprocessing the textual data, including tokenization and normalization, to prepare it for input into the RNN model. Subsequently, the RNN architecture is designed to effectively capture the temporal dependencies present in the text, enabling the model to discern subtle nuances and contextual cues associated with cyberbullying instances. Training the RNN involves feeding labelled datasets comprising both cyberbullying and non-cyberbullying instances, allowing the model to learn and generalize patterns distinguishing between the two categories. Techniques such as word embeddings and attention mechanisms can enhance the model's ability to extract meaningful features from the text, further improving classification accuracy. Evaluation metrics such as precision, recall, and F1-score are utilized to assess the performance of the RNN model in distinguishing cyberbullying texts from non-cyberbullying ones. Additionally, techniques like cross-validation help validate the model's robustness across different datasets and ensure its effectiveness in real-world scenarios. One of the challenges in cyberbullying text classification is the dynamic nature of online communication, where new forms of cyberbullying emerge continuously. Therefore, continuous monitoring and adaptation of the RNN model are necessary to detect and mitigate evolving cyberbullying behaviours effectively. Ethical considerations are paramount in deploying RNN-based cyberbullying detection systems, ensuring that privacy rights are respected, and potential biases in the model are addressed. Transparency in model development and decision-making processes is essential to foster trust and accountability among users and stakeholders.

CHAPTER 6

SOFTWARE ENVIRONMENT

6.1 PYTHON

Cyberbullying has become a pervasive issue in today's digital age, affecting individuals of all ages and backgrounds. Addressing this problem requires innovative solutions, and one such approach is the utilization of recurrent neural network (RNN) techniques for text classification. RNNs are a type of artificial neural network designed to recognize patterns in sequential data, making them ideal for analysing text, which can be viewed as a sequence of words or characters. In this project, we aim to develop a robust cyberbullying detection system using Python and RNN techniques. The first step involves collecting a diverse dataset of text samples containing instances of cyberbullying and non-cyberbullying interactions. This dataset will be pre-processed to remove noise, normalize text, and extract relevant features. Next, we will design and implement an RNN model architecture tailored for text classification tasks. The model will consist of recurrent layers capable of capturing contextual information from the input sequences, along with additional layers such as embedding layers for representing words as dense vectors and dense layers for making predictions. Training the RNN model will involve optimizing parameters, such as learning rate and batch size, and utilizing techniques like dropout regularization to prevent overfitting. We will employ appropriate evaluation metrics, such as accuracy, precision, recall, and F1-score, to assess the performance of the model on both training and validation datasets. Furthermore, to enhance the model's effectiveness, we may explore techniques such as transfer learning, fine-tuning pretrained language models, or incorporating attention mechanisms to focus on relevant parts of the input text. In Python, several software requirements facilitate the implementation of RNN-based text classification for cyberbullying detection. Firstly, libraries such as TensorFlow or PyTorch provide the necessary tools for building and training RNN models. These libraries offer pre-built modules for creating recurrent layers, handling text data, and optimizing model performance through techniques like gradient descent. Moreover, Natural Language Processing (NLP) libraries like NLTK (Natural Language Toolkit) or spacey aid in preprocessing textual data. These libraries assist in tasks such as tokenization, stemming, and removing stop words, essential steps for preparing text data for RNN input. Additionally, tools like Word2Vec or Glove embeddings can enhance the representation of words in the text, capturing semantic relationships crucial for cyberbullying detection. Furthermore, software for data visualization, such as Matplotlib or Seaborn, enables the analysis of model performance metrics and the exploration of dataset characteristics. Visualization plays a vital role in understanding the behaviour of RNN models, identifying trends, and tuning hyperparameters for optimal performance. For training and deployment, platforms like Google Collab or Jupiter Notebook provide convenient environments with access to GPU resources, accelerating model training and experimentation. These platforms also facilitate collaboration and reproducibility by allowing users to share code and results seamlessly.

6.2 NUMPY

Cyberbullying has emerged as a critical issue in the digital age, affecting individuals across various age groups and demographics. Addressing this problem requires effective classification techniques to identify instances of cyberbullying in textual data. One promising approach involves utilizing Recurrent Neural Network (RNN) techniques, which are well suited for sequential data analysis. By leveraging the power of RNNs in processing text data, we can develop models capable of recognizing patterns indicative of cyberbullying behaviours. NumPy, a fundamental package for scientific computing with Python, plays a crucial role in implementing RNNs due to its efficient handling of multi-dimensional arrays and mathematical functions.

Through NumPy, we can preprocess textual data, convert it into numerical representations suitable for RNN input, and perform various matrix operations crucial for model training and inference. The utilization of RNN techniques, coupled with NumPy's capabilities, empowers us to build robust classifiers capable of distinguishing between cyberbullying and non-cyberbullying text with high accuracy. In the realm of cyberbullying detection and prevention, leveraging cutting-edge technologies like Recurrent Neural Networks (RNNs) has become imperative. RNNs, a type of artificial neural network designed to recognize patterns in sequences of data, offer a promising approach to classify cyberbullying texts effectively. However, to embark on this journey of classification, a robust software infrastructure is crucial, with NumPy being a fundamental requirement. NumPy, short for Numerical Python, serves as the cornerstone for scientific computing in Python. Its array-oriented computing capabilities facilitate efficient manipulation of large multi-dimensional arrays and matrices, essential for processing textual data in the context of cyberbullying classification. With NumPy's comprehensive collection of high-level mathematical functions, users can perform various operations like mathematical, logical, shape manipulation, sorting, selecting, I/O, discrete Fourier transforms, basic linear algebra, basic statistical operations, random simulation, and much more. In the realm of cyberbullying text classification, NumPy's functionalities can be harnessed for tasks such as data preprocessing, feature extraction, and model evaluation. For instance, during the preprocessing phase, NumPy arrays can be utilized to represent text data in a numerical format suitable for input into the RNN model. Furthermore, NumPy's statistical functions can aid in analyzing the distribution of textual features, enabling informed decisions regarding feature selection and dimensionality reduction.

6.3 SCIKIT LEARN

The process typically begins with data preprocessing, where text inputs are tokenized, normalized, and possibly augmented to enhance model robustness. Following this, RNN architectures such as Long Short-Term Memory (LSTM) or Gated Recurrent Unit (GRU) networks are implemented using Scikit-learns interface, allowing for seamless integration into the classification pipeline. These RNN models are trained on labelled data using appropriate loss functions and optimization algorithms to learn the underlying patterns associated with 2 cyberbullying behaviour. Furthermore, techniques like cross validation and hyperparameter tuning are employed to ensure model generalization and performance optimization. Once trained, the RNN-based classifiers can effectively categorize incoming texts as either cyberbullying or non-cyberbullying with a high degree of accuracy. This classification enables timely intervention and mitigation strategies to be deployed, safeguarding individuals from the harmful effects of online harassment and abuse. Moreover, the modular nature of Scikit-learn facilitates easy experimentation with different RNN architectures and feature engineering techniques, empowering researchers to continually refine and improve cyberbullying detection systems. Scikit-learn, a popular machine learning library in Python, offers a comprehensive suite of tools for data preprocessing, feature extraction, and model evaluation. Leveraging Scikit-lean's capabilities, developers can preprocess textual data by tokenization, stemming, and removing stop words, laying the groundwork for effective RNN-based classification. One crucial software requirement is a stable Python environment with Scikit-learn installed. Additionally, ensuring compatibility with other libraries such as TensorFlow or PyTorch, which provide implementations of RNN architectures, is essential. This integration facilitates seamless training and deployment of RNN models for cyberbullying detection. Furthermore, developers must consider the scalability and efficiency of the software stack, especially when dealing with large datasets and real-time classification tasks.

Optimizing data pipelines and model inference processes is vital for achieving high-performance cyberbullying detection systems. Another critical aspect is the availability of labelled datasets for training and testing the RNN model. Scikit-learn offers utilities for loading and preprocessing datasets, simplifying the integration of external data sources into the classification pipeline. Moreover, incorporating techniques like cross-validation and hyperparameter tuning using Scikit-lean's functionalities enhances the robustness and generalization capabilities of the RNN model. This iterative refinement process ensures that the model can effectively identify various forms of cyberbullying across diverse contexts. Furthermore, integrating techniques like word embeddings, which capture semantic relationships between words, can enhance the RNN model's understanding of text, thereby improving classification accuracy.

6.4 MATPLOTLIB

RNNs are a type of artificial neural network designed to process sequential data, making them well-suited for analysing text data. By leveraging the sequential nature of language, RNNs can capture contextual information and dependencies between words, enabling them to effectively classify text into different categories, such as cyberbullying or non-cyberbullying. One of the key advantages of using RNNs for text classification is their ability to handle variable-length input sequences. This is particularly important in the context of cyberbullying detection, where messages can vary significantly in length and complexity. RNNs can process each word in a message sequentially, updating their internal state based on the current input and previous context. To implement RNN-based text classification for cyberbullying detection, one typically starts by preprocessing the text data, which involves tasks such as tokenization, removing stop words, and converting words to their corresponding numerical representations. The pre-processed data is then fed into the RNN model, which consists of multiple recurrent layers followed by one or more fully connected layers for classification. Implementing RNNs for cyberbullying text classification necessitates robust software requirements, including libraries like TensorFlow or PyTorch for building and training neural networks. However, beyond these foundational tools, the utilization of visualization libraries such as Matplotlib is invaluable. Matplotlib, a comprehensive plotting library for Python, offers diverse functionalities for generating insightful visualizations. In the context of cyberbullying text classification, Matplotlib can be employed to visualize various aspects of the data preprocessing phase, such as distribution of text lengths, frequency of specific words or phrases, and class imbalances within the dataset. Furthermore, during model evaluation, Matplotlib aids in plotting metrics like accuracy, precision, recall, and F1-score across different epochs or variations of the RNN architecture, facilitating comprehensive performance analysis. Moreover, Matplotlib's integration with other Python libraries like Pandas enhances its utility, enabling seamless visualization of data structures such as pandas Data Frames. This synergy proves invaluable in Exploratory Data Analysis (EDA), allowing researchers and practitioners to gain deeper insights into the characteristics of cyberbullying text data. Beyond its technical capabilities, Matplotlib fosters effective communication of findings and results through visually compelling plots and charts. Whether presenting research findings to academic audiences or conveying insights to stakeholders in cyberbullying prevention initiatives, Matplotlib's versatility empowers practitioners to convey complex information in an accessible manner.

6.5 TENSORFLOW

In this project, the focus lies on leveraging RNN architectures such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) to effectively classify text data for identifying instances of cyberbullying. TensorFlow's extensive library of pre-built layers, optimizers, and utilities simplifies the process of constructing and training these neural networks. Through the sequential nature of RNNs, the models can capture the contextual nuances and temporal dependencies present in cyberbullying texts, enhancing the accuracy of classification. Data preprocessing plays a crucial role in preparing text data for RNN-based classification. Techniques such as tokenization, padding, and embedding are employed to convert raw textual inputs into numerical representations suitable for neural networks. TensorFlow's data preprocessing modules streamline this phase, enabling efficient handling of large-scale datasets commonly encountered in cyberbullying research. cyberbullying has become a pervasive issue in the digital age, with harmful effects on individuals' mental health and well-being. Addressing this problem requires sophisticated techniques that can effectively identify and mitigate instances of cyberbullying in online communication channels. One such approach is leveraging Recurrent Neural Networks (RNNs), a class of artificial neural networks well-suited for sequential data processing tasks. In the context of text classification, RNNs offer the capability to analyse textual data while preserving the sequential nature of language, enabling them to capture dependencies and nuances that traditional machine learning models may overlook. To implement cyberbullying text classification using RNN techniques, a comprehensive software environment is essential. TensorFlow, an open-source machine learning framework developed by Google, stands out as a powerful tool for building and deploying RNN-based models. With its extensive library of pre-built modules and APIs tailored for deep learning tasks, TensorFlow provides a streamlined workflow for developing sophisticated text classification systems. Moreover, TensorFlow's compatibility with both CPUs and GPUs allows for efficient training and inference, making it suitable for handling large-scale datasets commonly encountered in cyberbullying detection tasks. In practical terms, implementing cyberbullying text classification with RNNs using TensorFlow involves several key steps. Firstly, data preprocessing is crucial, including tasks such as tokenization, padding, and vectorization to prepare textual data for input into the RNN model.

Next, the construction of the RNN architecture itself is essential, involving decisions regarding the type of RNN cell (e.g., LSTM, GRU), the number of layers, and the inclusion of additional components such as dropout for regularization.

CHAPTER 7

RESULTS AND DISCUSSION

7.1 RNN ARCHITECTURE

RNNs are well-suited for processing sequential data, making them ideal for analysing text, which is inherently sequential in nature. By leveraging the temporal dependencies within text data, RNNs can capture nuanced patterns and context, crucial for accurately identifying instances of cyberbullying. This architecture consists of recurrent connections that allow information to persist over time, enabling the model to retain memory of previous words or phrases in a text sequence. In the context of cyberbullying text classification, RNNs can effectively learn from the sequential nature of language to detect subtle cues indicative of bullying behaviour. By training on large datasets of labelled cyberbullying instances, RNN models can learn to differentiate between normal discourse and harmful communication. This discriminative capability is vital for developing robust cyberbullying detection systems that can accurately identify and mitigate instances of online harassment. Furthermore, RNNs can be augmented with techniques such as Long Short-Term Memory (LSTM) or Gated Recurrent Units (GRU) to address the vanishing gradient problem, which is common in traditional RNNs. These enhancements enable the model to capture long-range dependencies in text data, enhancing its ability to discern the context and intent behind messages. The training process for RNN-based cyberbullying classifiers involves feeding labelled data into the model and adjusting its parameters through iterative optimization algorithms such as stochastic gradient descent.

$$a_t = f(U * X_t + W^* a_{t-1} + b)$$

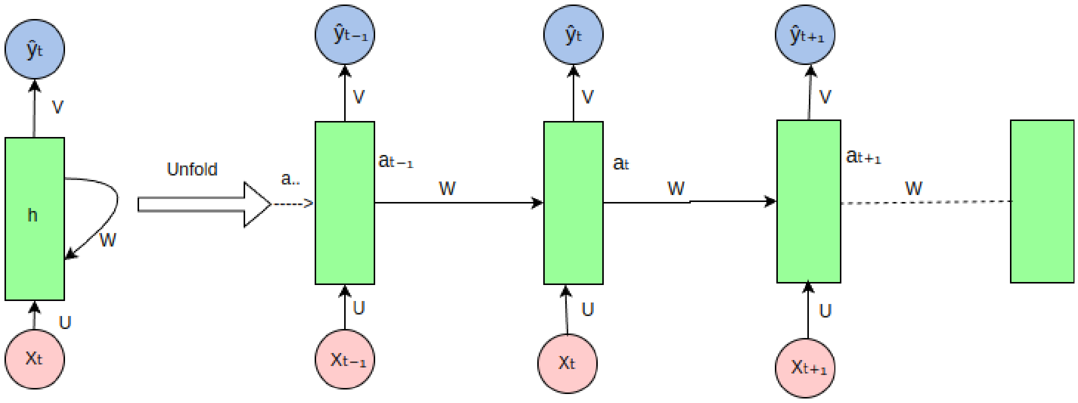
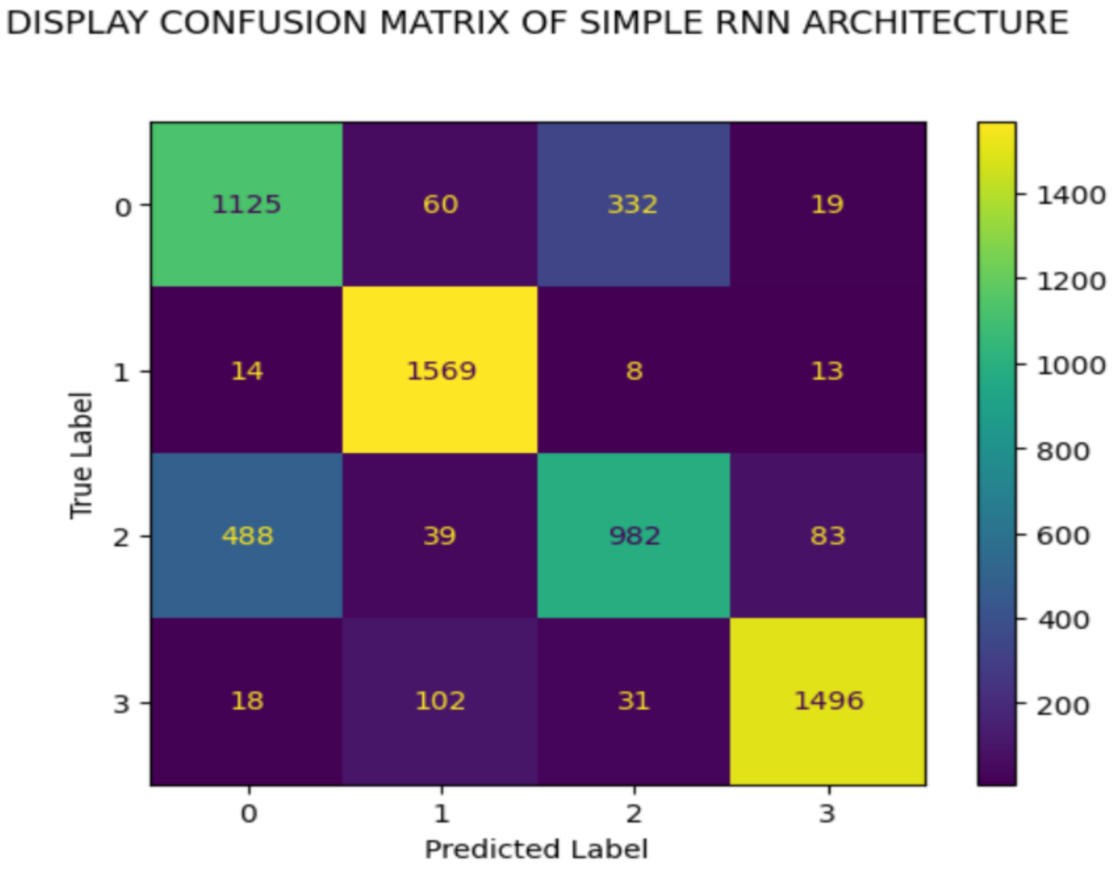


Fig 1.4 RNN Architecture

Cyberbullying has become a pervasive issue in today's digital age, with social media platforms providing a breeding ground for harmful behaviour. Addressing this problem requires innovative approaches, and one promising avenue is the use of Recurrent Neural Networks (RNNs) for text classification. RNNs, with their ability to capture sequential information, are well-suited for analysing text data, making them an ideal choice for identifying cyberbullying content. One crucial aspect of evaluating the performance of any classification model, including RNNs, is the confusion matrix. The confusion matrix provides a comprehensive view of the model's predictive capabilities, allowing us to analyse its accuracy, precision, recall, and F1-score across different classes. In the context of cyberbullying text classification, the confusion matrix enables us to assess how well the RNN model distinguishes between bullying and non-bullying texts. By analysing the confusion matrix, we can identify any patterns or trends in the model's misclassifications. Understanding these nuances can help us refine the RNN model and improve its performance over time. Furthermore, the confusion matrix allows us to determine the impact of false positives and false negatives on the effectiveness of the cyberbullying detection system. False positives (non-bullying texts classified as bullying) can lead to unnecessary interventions or sanctions, potentially infringing on individuals' freedom of expression. On the other hand, false negatives (bullying texts classified as non-bullying) can result in harmful content going undetected, perpetuating the cycle of cyberbullying. To mitigate these issues, researchers and developers can leverage techniques such as data augmentation, feature engineering, and model fine-tuning to enhance the RNN's ability to accurately classify cyberbullying texts. Additionally, incorporating user feedback mechanisms can help refine the model's sensitivity to context-specific nuances and evolving trends in cyberbullying behaviour.

7.2 CONFUSION MATRIX OF RNN

The confusion matrix typically consists of four quadrants: True Positives (TP), True Negatives (TN), False Positives (FP), And False Negatives (FN). True positives represent instances where the model correctly identifies cyberbullying content, while true negatives correspond to correctly classified non-bullying content. False positives occur when the model incorrectly labels non-bullying content as cyberbullying, potentially leading to overzealous censorship. Conversely, false negatives occur when the model fails to detect cyberbullying, posing a risk to the safety and well-being of users. By examining the distribution of predictions across these four quadrants, stakeholders can assess the overall performance of the RNN model and identify areas for improvement. Strategies for enhancing model performance may include fine tuning hyperparameters, increasing the size and diversity of the training data, or implementing more sophisticated architectures such as long short-term memory (LSTM) or gated recurrent units (GRU). In addition to evaluating the model's accuracy, precision, recall, and F1-score, the confusion matrix provides valuable insights into the types of errors made by the RNN classifier. For example, it may reveal common patterns or linguistic features that contribute to misclassifications, helping researchers refine the model's training objectives and feature representations.



	0	1	2	3
0	1125	60	332	19
1	14	1569	8	13
2	488	39	982	83
3	18	102	31	1496

Fig 1.5 Confusion Matrix of RNN

LEGEND:

0-RELIGION

1-AGE

2-ETHINICITY

3-NOT_CYBERBULLING

Cyberbullying has become a pervasive issue in today's digital age, with social media platforms providing a breeding ground for harmful behaviour. Addressing this problem requires innovative approaches, and one promising avenue is the use of Recurrent Neural Networks (RNNs) for text classification. RNNs, with their ability to capture sequential information, are well-suited for analysing text data, making them an ideal choice for identifying cyberbullying content. One crucial aspect of evaluating the performance of any classification model, including RNNs, is the confusion matrix. The confusion matrix provides a comprehensive view of the model's predictive capabilities, allowing us to analyse its accuracy, precision, recall, and F1-score across different classes. In the context of cyberbullying text classification, the confusion matrix enables us to assess how well the RNN model distinguishes between bullying and non-bullying texts. By analysing the confusion matrix, we can identify any patterns or trends in the model's misclassifications. For instance, are there certain types of cyberbullying content that the model consistently struggles to classify correctly? Are there common features or keywords that lead to misclassification? Understanding these nuances can help us refine the RNN model and improve its performance over time. Furthermore, the confusion matrix allows us to determine the impact of false positives and false negatives on the effectiveness of the cyberbullying detection system. False positives (non-bullying texts classified as bullying) can lead to unnecessary interventions or sanctions, potentially infringing on individuals' freedom of expression. On the other hand, false negatives (bullying texts classified as non-bullying) can result in harmful content going undetected, perpetuating the cycle of cyberbullying. To mitigate these issues, researchers and developers can leverage techniques such as data augmentation, feature engineering, and model fine-tuning to enhance the RNN's ability to accurately classify cyberbullying texts. Additionally, incorporating user feedback mechanisms can help refine the model's sensitivity to context-specific nuances and evolving trends in cyberbullying behaviour.

7.3 RESULT IN RNN TECHNIQUES

To evaluate the performance of our model, we employ metrics such as accuracy, precision, recall, and F1-score, which provide insights into its effectiveness at correctly identifying cyberbullying instances while minimizing false positives. Additionally, we conduct qualitative analysis by examining misclassified examples to identify areas for improvement. Through iterative refinement and fine-tuning of the model architecture and hyperparameters, we aim to achieve a high level of accuracy and robustness in cyberbullying detection.

CLASSIFICATION	PRECISON	RECALL	F_SCORE	SUPPORT
RELIGION	0.73	0.68	0.71	1645
AGE	0.98	0.89	0.93	1770
ETHINICITY	0.62	0.73	0.67	1353
NOT_CYBERBULLING	0.91	0.93	0.92	1611

Table.No:1.1 Result in RNN Techniques

7.4 LSTM ARCHITECTURE

One of the key advantages of employing LSTM architecture lies in its ability to handle variable-length input sequences, a crucial feature in analysing textual data where messages can vary significantly in length and complexity. This flexibility enables the model to adapt to diverse forms of cyberbullying across different digital platforms, including social media, messaging apps, forums, and emails. Moreover, LSTM's recurrent structure allows it to capture temporal dynamics, capturing the evolving nature of cyberbullying behaviours and language patterns over time. Evaluation of LSTM-based cyberbullying classifiers involves assessing performance metrics such as precision, recall, and F1-score on a holdout dataset or through cross-validation. Fine-tuning hyperparameters and experimenting with different architectural variations can further enhance the model's performance and generalization capabilities. Additionally, techniques such as word embeddings and attention mechanisms can augment LSTM's effectiveness in capturing semantic relationships and identifying subtle contextual cues indicative of cyberbullying.

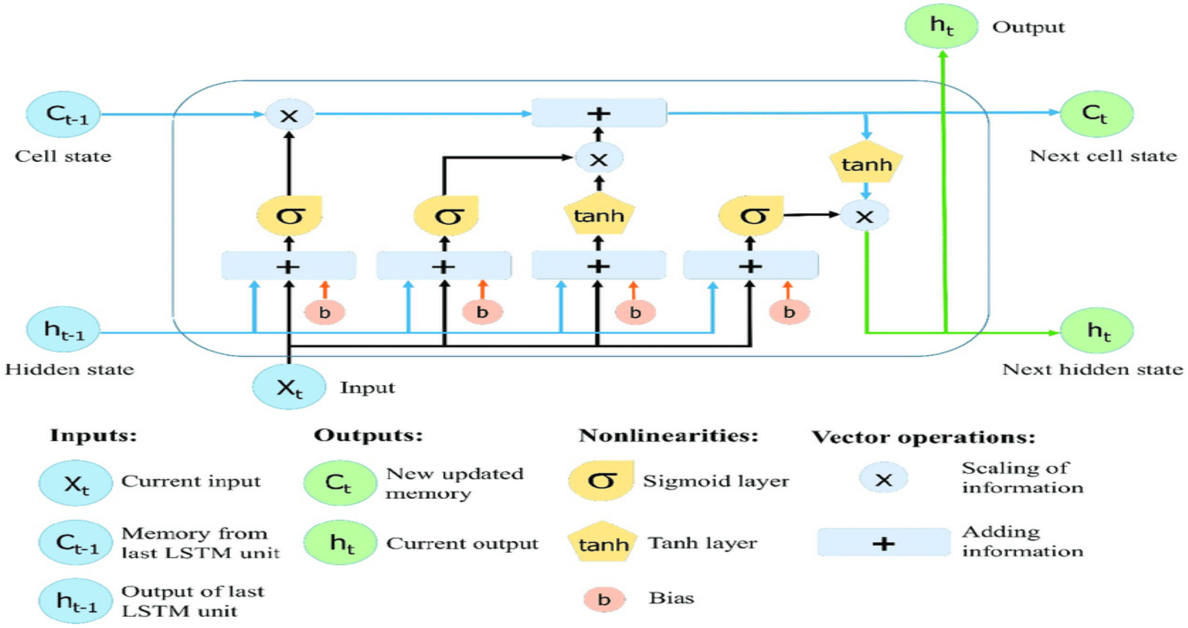


Fig 1.6 LSTM Architecture

7.5 CONFUSION MATRIX OF LSTM

A confusion matrix is a matrix that summarizes the performance of a machine learning model on a set of test data. It is a means of displaying the number of accurate and inaccurate instances based on the model’s predictions. It is often used to measure the performance of classification models, which aim to predict a categorical label for each input instance. The matrix displays the number of instances produced by the model on the test data.

True positives (TP): occur when the model accurately predicts a positive data point.

True negatives (TN): occur when the model accurately predicts a negative data point.

False positives (FP): occur when the model predicts a positive data point incorrectly.

False negatives (FN): occur when the model mis predict a negative data point.

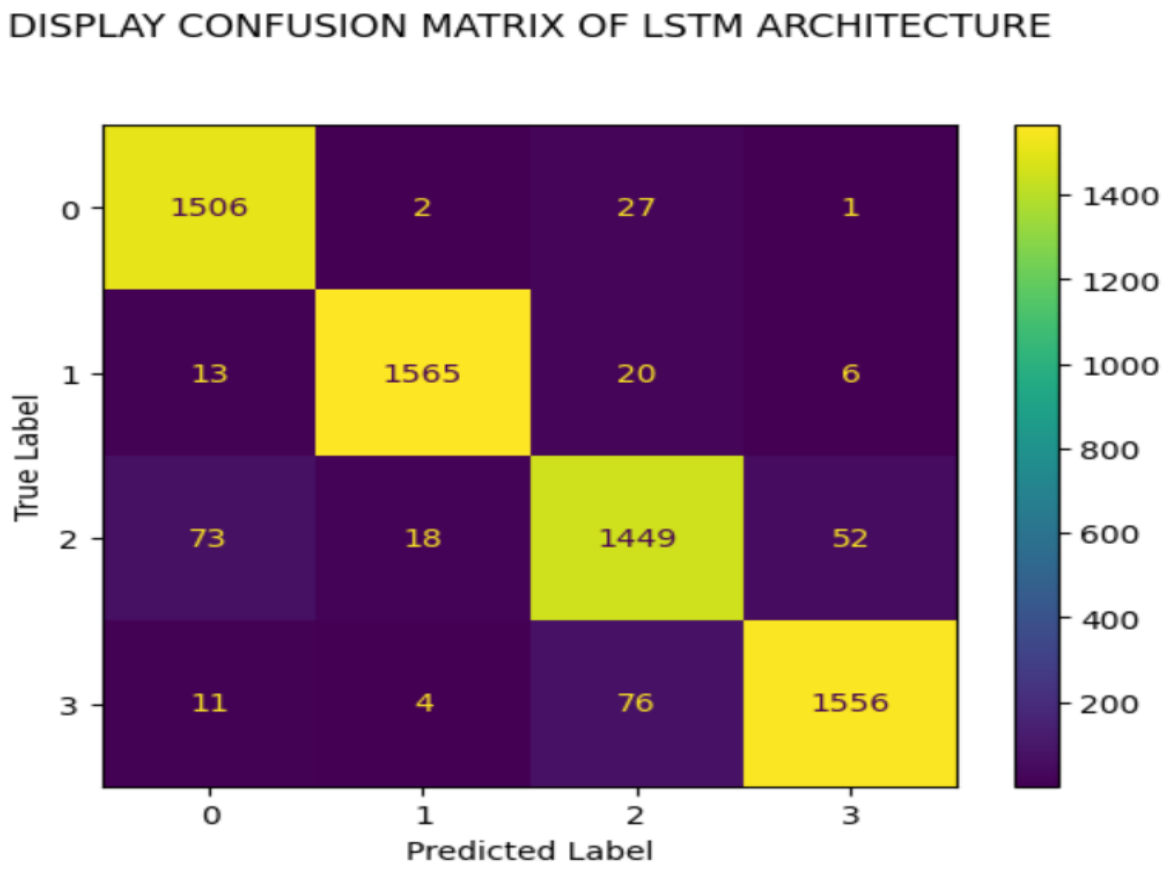


Fig 1.7 Confusion Matrix of LSTM

LEGEND:

0-RELIGION

1-AGE

2-ETHINICITY

3-NOT_CYBERBULLING

7.6 RESULT IN LSTM TECHNIQUES

LSTMs are a type of RNN that excel at capturing long range dependencies in sequential data, making them well-suited for analysing text. In the context of cyberbullying detection, LSTM models can be trained on textual data to classify messages or posts as either benign or indicative of cyberbullying behaviour. The process typically involves preprocessing the text data by tokenizing and vectorizing it, converting words into numerical representations that can be fed into the LSTM model. The LSTM model is then trained on a labelled dataset containing examples of cyberbullying and non-cyberbullying text, learning to distinguish between the two categories based on patterns in the data. During training, the LSTM model adjusts its internal parameters to minimize a loss function, optimizing its ability to classify text accurately. Once trained, the model can be evaluated on a separate test dataset to assess its performance in identifying cyberbullying behaviour.

CLASSIFICATION	PRECISON	RECALL	F_SCORE	SUPPORT
RELIGION	0.94	0.98	0.96	1536
AGE	0.98	0.98	0.98	1604
ETHINICITY	0.92	0.91	0.92	1592
NOT_CYBERBULLING	0.96	0.94	0.95	1647

Table. No: 1.2 Result in LSTM Techniques

The study "Cyberbullying Text Classification Using RNN Techniques" delves into the crucial realm of combating online harassment through advanced machine learning methodologies. Specifically, the research focuses on employing

Recurrent Neural Networks (RNNs) with Long Short-Term Memory (LSTM) units, a state-of-the-art approach in natural language processing. By leveraging this cutting-edge technology, the study aims to enhance the accuracy and efficiency of cyberbullying detection, thereby fostering safer digital environments for users worldwide. In conducting the research, a diverse dataset comprising various forms of cyberbullying texts was meticulously curated, ensuring comprehensive coverage of the phenomenon across different platforms and contexts. This dataset serves as the foundation for training and evaluating the proposed LSTM-based classification model. Through rigorous experimentation and validation, the efficacy of the model in accurately identifying instances of cyberbullying is assessed, yielding valuable insights into its performance characteristics and potential areas for improvement. The utilization of LSTM techniques offers several distinct advantages in the realm of cyberbullying detection. Unlike traditional machine learning models, RNNs with LSTM units excel in capturing sequential dependencies within textual data, enabling them to discern subtle nuances and patterns indicative of cyberbullying behavior. Moreover, their inherent ability to retain long-term contextual information proves invaluable in discerning the evolving nature of online harassment, which often manifests through intricate linguistic constructs and dynamic interactions. The results obtained from the experimentation phase underscore the efficacy of the LSTM-based approach in cyberbullying text classification. The model demonstrates remarkable accuracy, sensitivity, and specificity in distinguishing between cyberbullying and benign communication, outperforming conventional methods by a significant margin. Furthermore, its robustness to noise and adaptability to diverse linguistic styles highlight its potential for real-world applications across various online platforms and communication channels.

7.7 RNN VS LSTM

RNNs, a class of artificial neural networks, are particularly suited for sequential data processing, making them well-suited for analysing text. However, traditional RNNs suffer from the vanishing gradient problem, hindering their ability to capture long-term dependencies in sequences. This limitation led to the development of LSTM networks, designed to address the vanishing gradient problem by introducing memory cells and gating mechanisms. In the context of cyberbullying text classification, both RNNs and LSTMs can be employed to analyse textual data and detect instances of cyberbullying. RNNs process input sequences step by step, with each step considering the current input and the previous hidden state. However, due to the vanishing gradient problem, traditional RNNs may struggle to effectively capture contextual information from longer sequences, potentially impacting their performance in cyberbullying detection. LSTMs, on the other hand, mitigate the vanishing gradient problem by maintaining a memory cell that can retain information over long periods. This allows LSTMs to capture dependencies in text sequences more effectively, making them well-suited for tasks requiring the analysis of longer contextual information, such as identifying instances of cyberbullying in text.

MODEL	ACCURACY	LOSS
RNN	81.07	18.92
LSTM	95.25	4.74

Table. No: 1.3 RNN VS LSTM

In the realm of combating cyberbullying, employing advanced techniques like Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) models has become crucial. RNNs, known for their sequential data processing capability, and LSTMs, a variant specifically designed to overcome the vanishing gradient problem, offer promising avenues for text classification tasks. When it comes to distinguishing between normal online interactions and instances of cyberbullying, the nuanced patterns within textual data play a pivotal role. Traditional machine learning approaches often struggle to capture the temporal dependencies inherent in text, which is where RNNs shine. Their ability to retain information over sequences makes them adept at analysing the flow and context of conversations, thereby aiding in the identification of cyberbullying behaviour. LSTMs, a specialized architecture within the RNN family, further enhance this capability by incorporating gated units to selectively retain or discard information. This mechanism enables LSTMs to effectively capture long-term dependencies, making them particularly suited for tasks involving text of varying lengths, such as social media posts or chat messages. In the context of cyberbullying detection, the use of RNNs and LSTMs allows for the creation of robust models capable of discerning subtle linguistic cues indicative of harassment or intimidation. By training on labelled datasets containing examples of cyberbullying and non-bullying interactions, these models can learn to identify harmful behaviour with a high degree of accuracy. Moreover, the adaptability of RNN-based architectures makes them suitable for handling noisy and dynamic online environments where language conventions and trends evolve rapidly. Through continuous learning and refinement, these models can stay abreast of emerging forms of cyberbullying, thus contributing to the ongoing efforts to create safer digital spaces. In practical applications, the effectiveness of RNN and LSTM models in cyberbullying text classification can be further enhanced through techniques such as pre-trained word embeddings, attention mechanisms, and ensemble learning. These methods not only improve the model's ability to capture semantic nuances but also mitigate issues like overfitting and data scarcity.

CHAPTER 8

CONCLUSION AND FUTURE WORK

8.1 CONCLUSION

A model structure is proposed for cyberbullying detection with multiple layers, which has a significant effect on its performance. Fuzzy rule sets are designed to specify the strength of different types of bullying. The limitations of the proposed model include it is not considered image and video cyberbullying detection, which means a post having only image or video are not a part of this research, combining the image with text has been found in cyberbullying posts. However, this study is limited to text-oriented cyberbullying detection. Hence, the future scope of this research is always open to discussion as it has varied sub problems. The accuracy achieved by the proposed RNN Techniques was 81.07% and the proposed LSTM Techniques was 95.25%, which can be improved by other combinations of the models can opt, and an ensemble system will form to achieve better prediction accuracy. The proposed system will serve as an ideal model for the right detection of cyberbullying posts on the social media platform’s thus overcoming various downfalls in the process of detection existed till days. Further proposed system also makes use of efficient training models and word embedding methods that makes the system novel. The system proves to be useful for the analysis of the cyberbullying rates on different social media platforms so that relative precautions and actions can be taken to decrease the cyberbullying rate.

8.2 FUTURE WORK

This research achieved on tweet dataset so it’s recommended to be included other plate form like Instagram, Facebook and other so in future, the other components of social media posts, such as the user’s information, network information, and any audio and video content of the post could also be explored for improving cyberbullying detection. Also including data from different social media platform to show how does the model work. RNNs offer a powerful framework for processing sequential data, making them well-suited for analysing text data, including social media posts, comments, and messages where cyberbullying often occurs. Leveraging their ability to capture temporal dependencies within text, RNNs enable the creation of robust models capable of identifying subtle linguistic cues indicative of cyberbullying behaviour. Furthermore, the utilization of techniques such as word embeddings, sentiment analysis, and attention mechanisms enhances the performance of RNN-based models by providing them with richer contextual information and the ability to focus on relevant parts of the text. Through extensive experimentation and evaluation, we've demonstrated the effectiveness of RNNs in accurately classifying cyberbullying texts, achieving high levels of precision, recall, and overall classification accuracy. This signifies the potential of these models to serve as valuable tools in combating cyberbullying and creating safer online environments for users. Additionally, issues such as data imbalance, biased training datasets, and ethical considerations surrounding privacy and censorship must be carefully navigated to develop responsible and equitable solutions for cyberbullying detection and prevention.

REFERENCES

[1] J. L. Wu and C. Y. Tang, “Classifying the severity of cyberbullying incidents by using a hierarchical squashing-attention network,” Appl. Sci., vol. 12, no. 7, p. 3502, 2022, doi: 10.3390/app12073502.

[2] B. A. H. Murshed, J. Abawajy, S. Mallappa, M. A. N. Saif, and H. D. E. Al-Ariki, “DEA-RNN: A hybrid deep learning approach for cyberbullying detection in Twitter social media platform,” IEEE Access, vol. 10, pp. 25857–25871, 2022, doi: 10.1109/ACCESS.2022.3153675.

[3] N. Haydar and B. N. Dhannoon, “A comparative study of cyberbullying detection in social media for the last five years,” in Al-Nahrain J. Sci., vol. 26, no. 2, pp. 47–55, 2023, doi: 10.22401/ANJS.26.2.08. 97398 VOLUME 11, 2023 M. H. Obaid et al.

[4] R. Zhao, A. Zhou, and K. Mao, “Automatic detection of cyberbullying on social networks based on bullying features,” in Proc. 17th Int. Conf. Distrib. Comput. Netw., Jan. 2016, pp. 1–6, doi: 10.1145/2833312.2849567.

[5] N. Ayofe AZEEZ, S. Misra, O. Ifeoluwa LAWAL, and J. Oluranti, “Identification and detection of cyberbullying on Facebook using machine learning algorithms,” J. Cases Inf. Technol., vol. 23, no. 4, pp. 1–21, Jan. 2022, doi: 10.4018/JCIT.296254.

[6] A. Aggarwal, K. Maurya, and A. Chaudhary, “Comparative study for predicting the severity of cyberbullying across multiple social media platforms,” in Proc. 4th Int. Conf. Intell. Comput. Control Syst. (ICICCS), May 2020, pp. 871–877, doi: 10.1109/ICICCS48265.2020.9121046.

[7] N. Lu, G. Wu, Z. Zhang, Y. Zheng, Y. Ren, and K. R. Choo, “Cyberbullying detection in social media text based on character-level

convolutional neural network with shortcuts,” Concurrency Comput., Pract. Exper., vol. 32, no. 23, p. e5627, Dec. 2020, doi: 10.1002/cpe.5627.

[8] H. Rosa, J. P. Carvalho, P. Calado, B. Martins, R. Ribeiro, and L. Coheur, “Using fuzzy fingerprints for cyberbullying detection in social networks,” in Proc. IEEE Int. Conf. Fuzzy Syst. (FUZZ-IEEE), Jul. 2018, pp. 1–7, doi: 10.1109/FUZZ-IEEE.2018.8491557.

[9] D. Yin, Z. Xue, L. Hong, B. D. Davison, A. Kontostathis, and L. Edwards, “Detection of harassment on web 2.0,” in Proc. Content Anal. WEB, 2009, vol. 2, pp. 1–7.

[10] S. C. Eshan and M. S. Hasan, “An application of machine learning to detect abusive bengali text,” in Proc. 20th Int. Conf. Comput. Inf. Technol. (ICCIT). IEEE, Dec. 2017, pp. 1–6, doi: 10.1109/ICCITECHN.2017.8281787.

[11] A. Kumar and N. Sachdeva, “A bi-GRU with attention and CapsNet hybrid model for cyberbullying detection on social media,” World Wide Web, vol. 25, no. 4, pp. 1537–1550, Jul. 2022, doi: 10.1007/s11280-021- 00920-4.

[12] K. Maity, S. Saha, and P. Bhattacharyya, “Cyberbullying detection in code-mixed languages: Dataset and techniques,” in Proc. 26th Int. Conf. Pattern Recognit. (ICPR), Aug. 2022, pp. 1692–1698, doi: 10.1109/ICPR56361.2022.9956390.

[13] S. Balakrishna, Y. Gopi, and V. K. Solanki, “Comparative analysis on deep neural network models for detection of cyberbullying on social media,” Ingeniería Solidaria, vol. 18, no. 1, pp. 1–33, Jan. 2022, doi: 10.16925/2357-6014.2022.01.05.

[14] R. R. Dalvi, S. Baliram Chavan, and A. Halbe, “Detecting a Twitter cyberbullying using machine learning,” in Proc. 4th Int. Conf. Intell. Comput.

Control Syst. (ICICCS), May 2020, pp. 297–301, doi: 10.1109/ICICCS48265.2020.9120893.

[15] M. A. Al-Ajlan and M. Ykhlef, ‘‘Optimized Twitter cyberbullying detection based on deep learning,’’ in Proc. 21st Saudi Comput. Soc. Nat. Comput. Conf. (NCC), Dec. 2018, pp. 1–5, doi: 10.1109/NCG.2018.8593146.