Big Data Meets Social Media: Predicting Cyberbullying with Machine Learning Algorithms

[1]Mrs.Baleswari Guntuku, [2]BALIBOYINA ANUSHA, [3]Mallavalli Haswanth, [4]Kanamarlapudi Naga Vamsi

[1]g.baleswri@gmail.com, Associate Professor of IT, NRI Institute of Technology, A.P, India 521212
[2]anushabaliboyina38@gmail.com, UG Scholar, Dept of IT, NRI Institute of Technology, A.P, India 521212
[3]haswanthmallavalli@gmail.com, UG Scholar, Dept of IT, NRI Institute of Technology, A.P, India 521212
[4]nagavamsikanamarlapudi004@gmail.com, UG Scholar, Dept of IT, NRI Institute of Technology, A.P, India 521212

Abstract— **Cyberbullying remains one of the key public online problems in the modern world, being the second most common cause of mortality and a leading cause of disability, which makes the primary prevention of this disease and its complications extremely important. The purpose of this research is to predict the outcomes of the cyberbullying with the help of the machine learning algorithms. A convolutional neural network was suggested to be coupled with a well designed long short term memory. Also, seven more classifiers were also used as baseline models: logistic regression, random forest, extreme gradient boosting, k-nearest neighbour, artificial neural network, long short-term memory, and convolution neural networks. Utilising a online care dataset that comprises online characteristics of 5,110 individuals, systematic training was conducted. To address data imbalance, minority class synthetic over sampling techniques have been used.**

Index Terms— ***Cyberbullying Detection, Social Media Aggression, Machine Learning Algorithms, Ensemble Learning, Deep Learning Techniques, Predictive Modelling, Feature Engineering, Data Privacy, Online Violence, Social Media Networks, Behavioural Data Analysis, Ethical Challenges, Predictive Analytics, Cyberbullying Prediction Models, Human Behaviour Data, Digital Communication, AI in Social Media, Text Classification, Online Harassment, Big Data in Cyberbullying***

I. Introduction

American studies show that about 40% of cyberbullying users die within three months after the attack, while those who survive are usually partially disabled. Today, cyberbullying is a serious social and social media problem in all countries of the world. As squalid as it sounds, this study intends to avoid unemployment and the eventual deterioration of the social media system. To accomplish this and prevent potential cyberbullyings and collateral damage, artificial intelligence will assist in predicting the overall ischemic cyberbullying, while other studies aim to employ image analysis to detect retinal changes pertinent to embolic events. Due to the sedentary behaviour brought on by the social media crisis, there was a surge in people over the age of 60 suffering from cyberbullyings. To combat this and guess the prevalence of ischemic cyberbullyings in March, this study and AI will focus on retinal changes related to embolic events using machine learning algorithms.

A new approach was tested by coupling the CNN with a well-designed LSTM, which combines the powerful approach of deep learning in handling complex data. For further reliable evaluation and comparison, seven additional classifiers served as baseline models: logistic regression, random forest, extreme gradient boosting, k-nearest neighbours, artificial neural networks, long short-term memory networks, and convolutional neural networks. The study uses a social media dataset with several online parameters, allowing for systemic training as well as testing of the proposed models. Recognizing that there are inherent problems caused by imbalanced data in social media applications, synthetic oversampling techniques were used to improve the representation of the minority class, reduce the risk of biases, and improve the reliability of prediction. It focuses on the importance of data preprocessing in crafting effective ML models, particularly in a realm like online care, where such data inconsistencies could significantly affect the performance of the model. Here, this paper aims to enhance the accuracy of predicting cyberbullying outcome models with the help of advanced ML techniques and data consistency. Such a finding will allow clinicians to make more informed decisions, improve individualized treatment regimens, and eventually achieve better social media outcomes. With the rapid increase in the cyberbullying burden around the world, emerging solutions such as this one provide a roadmap into the future in addressing one of today's critical online concerns.

II. Proposed system

It uses a CNN sequential model with a proposed approach to result in greater accuracy regarding cyberbullying predictions, beyond current models' precision and dependability. The proposed method will majorly focus on the optimization of the model structure, improvement of procedures in data pretreatment, and handling common issues found in social media datasets like class imbalance and feature complexity. This approach begins with extensive preprocessing of the data to yield high-quality input for the model to be trained. Preprocessing processes include imputation approaches used to fill up missing values, normalising and scaling features that improve convergence models, as well as correct class imbalance through synthetic oversampling methods such as SMOTE. This will ensure balanced representation of all classes, with a focus on the minority class, which is key for making accurate forecasts in social media applications. Additionally, data augmentation is applied to enlarge the dataset through realistic variances, which enhances the ability of the model to generalize in other contexts. Another important feature of the proposed system is feature engineering. During the phase, the system discovers and selects the most relevant predictors of cyberbullying outcomes through correlation analysis and feature significance algorithms. This simplifies the model but enhances the interpretability while allowing doctors to understand aspects that influence predictions better.

The proposed system is based on the CNN sequential model that is designed with the idea of effectively processing structured numerical inputs. The design begins with an input layer that receives pre-processed information. The convolutional layers discover hierarchical patterns and relate complex connectivity between features. These layers are complemented with pooling layers that reduce the spatial dimensions of the data while preserving integral features so that efficiency in computing is increased and the chance of overfitting is minimized. The feature sets are then passed through fully connected layers that merge the learned representations for final classification. The output layer applies the sigmoid activation function to predict cyberbullying outcomes as either a binary or a multiclass classification. Advanced optimization techniques are used to improve the performance of the model. In the case of the Adam optimiser, it adjusts learning rates dynamically while training to reach faster and more stable convergence. Other forms of regularisation include L2 and dropout to prevent overfitting. Finally, early stopping is applied to train until validation performance no longer improves, to be able to avoid wasteful computation and prevent overtraining.

With stratified k-fold cross-validation being adopted as the training strategy, the CNN model would not guarantee any overfitting but instead ensures that the performance is solid and reliable. Accuracy, precision, recall, F1-score, and AUC-ROC are used as key metrics to evaluate and refine the model during training. Such rigorous training ensures that a model performs well on a wide range of datasets.

Advantages of Proposed System

This proposed approach aims to apply the features of a CNN sequential model to provide improved accuracy of the outcomes of cyberbullying predictions than existing models. This system surpasses existing limitations regarding prediction accuracy, efficiency, and generalisability because it employs advanced deep learning techniques with an optimized architecture.

IMAGE PROCESSING:

The approach begins with the preparation of the social media dataset to ensure that it is of the highest quality to train models. Missing data are addressed through imputation techniques. Normalising and scaling characteristics improve model convergence. Class imbalance approaches by synthetic over sampling methods such as SMOTE (Synthetic Minority Oversampling Technique); the class imbalance technique ensures that the minority samples, which are often under-represented in social media datasets, are well represented, thereby reducing bias and increasing model reliability.

OBJECT DETECTION:

Object detection is a computer technology related to and that deals with detecting instances of semantic objects of a certain class (such as humans, buildings, or cars) in digital images and videos. Well-researched domains of object detection includend

FEATURE ENGINEERING

Data augmentation aims to augment both the size and variety of datasets, especially those related to minority groups, by generating realistic variants while maintaining the underlying pattern.

Correlation analysis and feature importance methods are employed to choose the most relevant predictors of cyberbullying outcomes. This minimizes model complexity and improves interpretability while retaining performance.

III. Methodology:

The proposed cyberbullying outcome prediction system has been separated into distinct modules, each one focused on one component of data processing, model building, and evaluation. With this integration of modules, the system aims at efficiency, scalability, and reliability while processing social media information.

1. Data Collection and Integration Module

This module involves handling the gathering of online information from various sources, such as electronic online records, patient surveys, and statistics available to the public. It is responsible for standardization in terms of data formatting, fixing inconsistencies, and collating various data points, which may contain demographic information, social media history, and online measurements. The module ensures clean and just data for subsequent steps in the pipeline.

2. Data Preprocessing and Augmentation Module

Preprocessing is one of the most important steps for data preparation for machine learning. This includes the handling of missing values through imputation methods, feature normalisation for scaling, and encoding categorical variables. Moreover, it gets to deal with the class imbalance problem through the synthetic oversampling method of SMOTE, ensuring that the minority class outcomes are sufficiently explored. Other means are also incorporated, such as data augmentation to generate mutations on the training dataset, and ultimately boost the model generalisation capabilities.

3. Feature Engineering and Selection Module

This module thus takes care of filtering this dataset, leaving only the most significant and important features for predicting cyberbullyings. Correlation analysis, feature importance ranking, and dimensionality reduction are all techniques for finding significant predictors while minimizing noise and redundancy. Doing this ensures that the model only has relevant data to analyze, enhancing performance and interpretability.

4. Model Development and Training Module

Designates and instructs the CNN sequential model's core module: architecture, which includes input layer, convolutional layers, pooling layers, fully connected layers, and an output layer. Optimisation tricks like adaptive learning rates, regularisation techniques, and early stopping are used to ensure model performance is improved. A number of metrics such as accuracy, sensitivity, specificity, F1-score, and AUC-ROC are used to evaluate the performance of the model after training. This module compares the proposed sequential CNN model against existing models such as logistic regression, random forests, extreme gradient boosting, k-nearest neighbours, and LSTM architectures. The Comparison analysis shows that the proposed system has the edge over conventional systems.

4.1.UML DIAGRAMS

UML stands for Unified Modeling Language. UML is a standardized general-purpose modeling language in the field of object-oriented software engineering. The standard is managed, and was created by, the Object Management Group. The goal is for UML to become a common language for creating models of object- oriented computer software. In its current form UML is comprised of two major components: a Meta-model and a notation. In the future, some form of method or process may also be added to; or associated with, UML.

The Unified Modeling Language is a standard language for specifying, Visualization, Constructing and documenting the artifacts of software system, as well as for business modeling and other non- software systems. The UML represents a collection of best engineering practices that have proven successful in the modeling of large and complex systems. The UML is a very important part of developing objects-oriented software and the software development process. The UML uses mostly graphical notations to express the design of software projects.

GOALS:

The Primary goals in the design of the UML are as follows:

1. Provide users a ready-to-use, expressive visual modelling Language so that they can develop and exchange meaningful models.
2. Provide extendibility and specialization mechanisms to extend the core concepts.
3. Be independent of particular programming languages and development process.
4. Provide a formal basis for understanding the modelling language.
5. Encourage the growth of OO tools market.
6. Support higher level development concepts such as collaborations, frameworks, patterns and components.
7. Integrate best practices.

4.1.1.Use case diagram

A use case diagram in the Unified Modelling Language (UML) is a type of behavioural diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.

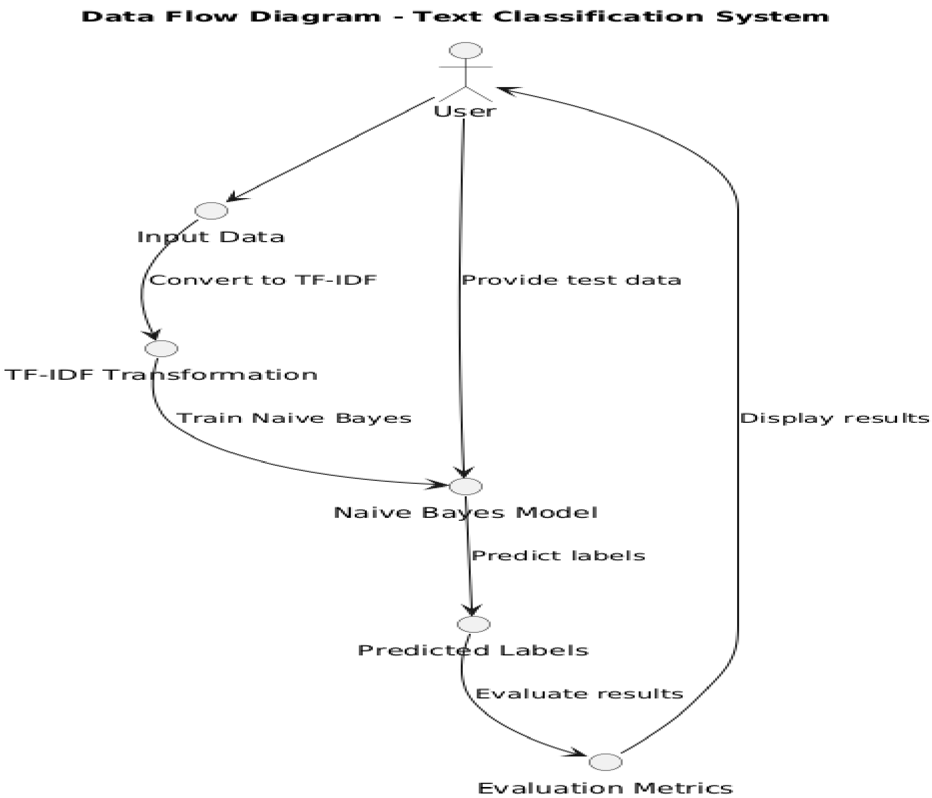


Fig 1:-Use Case diagram

4.1.2.Class Diagram

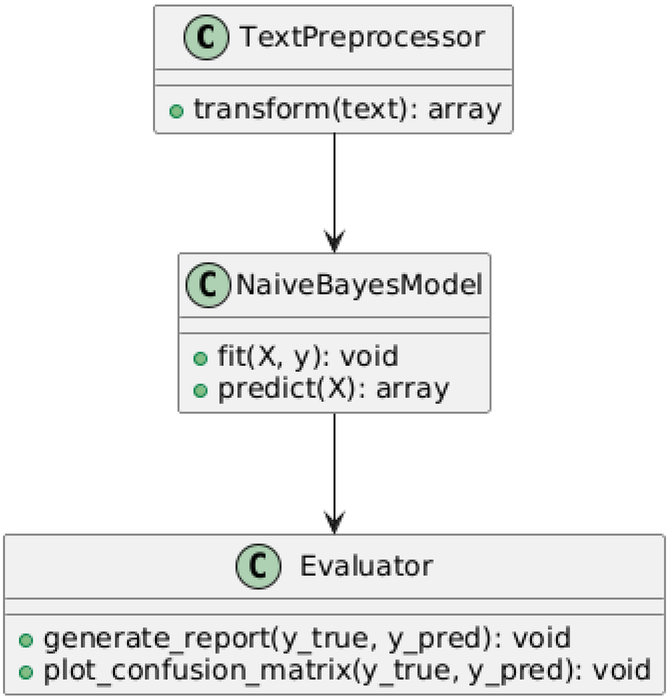


Fig.2. Class diagram

The class diagram is used to refine the use case diagram and define a detailed design of the system.
The class diagram classifies the actors defined in the use case diagram into a set of interrelated classes. The relationship or association between the classes can be either an "is-a" or "has-a" relationship. Each class in the class diagram may be capable of providing certain functionalities. These functionalities provided by the class are termed "methods" of the class. Apart from this, each class may have certain "attributes" that uniquely identify the class.

4.1.3.Activity diagram

The process flows in the system are captured in the activity diagram. Similar to a state diagram, an activity diagram also consists of activities, actions, transitions, initial and final states, and guard conditions.
An activity diagram is a system-modelling and design tool used, among other things, to portray workflows, decision points, and other processes inside a system. A diagram that gives a very efficient description of a system's dynamic features-activities originating from UML, makes them focus on the flow of control and data between different operations-in particular for sequential, parallel, or conditional workflows.An activity diagram begins with an initial node, which represents the commencing point of a process. Activities, which are drawn in rounded rectangles, depict those tasks or procedures that exist within the system. These activities are connected with arrows that represent the flow of control or data from one action to the next.

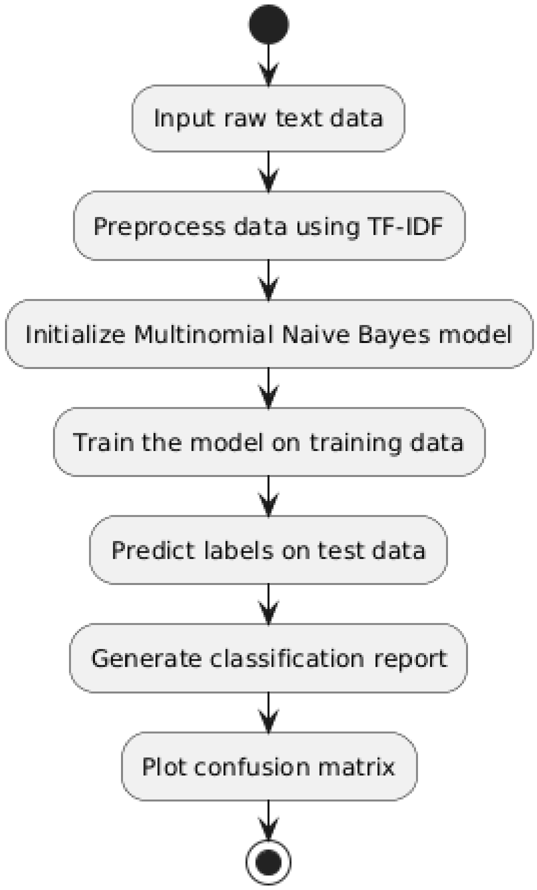


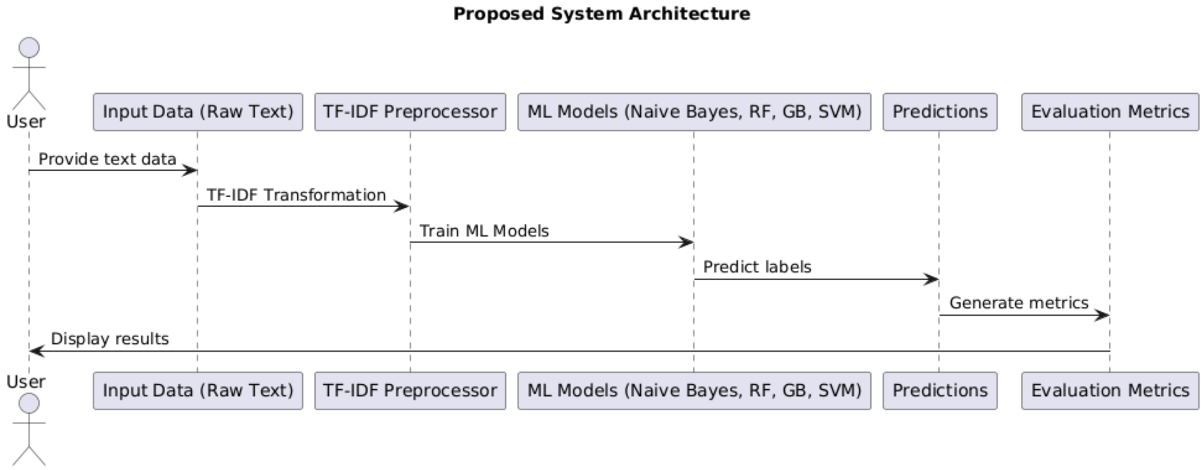
Fig.3. Activity diagram

4.4.Dataflow diagrams
To create a Data Flow Diagram (DFD) for the proposed thyroid disorder diagnosis system, we would include the following levels:
Level 0: Context Diagram
This diagram represents the system as a single process, showing its interaction with external entities such as patients, clinicians, and the database.
Entities and Flow:
1. **Patient**: Provides clinical, biochemical, and imaging data.
2. **Clinician**: Receives diagnostic results and insights.
3. **Database**: Stores patient data and diagnostic results.

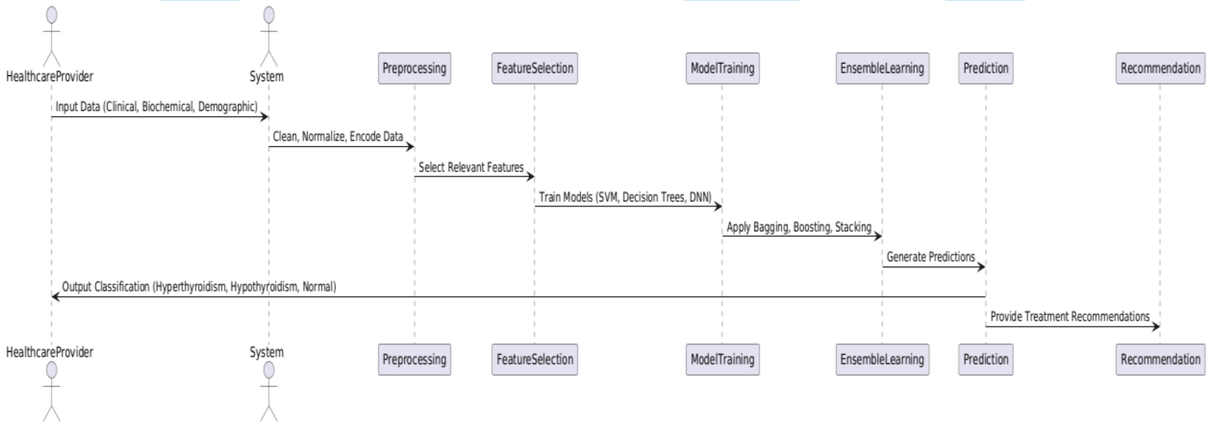
Process:
- The system takes patient data as input and sends diagnostic results back to clinicians and the database.

Steps:
1. **Input Data**:
    - The system receives data from the patient (manual input or electronic health records).
2. **Preprocessing**:
    - Removes noise, normalizes values, and ensures compatibility with models.
3. **Feature Extraction**:
    - Extracts relevant features such as T3, T4, TSH levels, imaging patterns, and clinical symptoms.
4. **Model Prediction**:
    - Hybrid models (e.g., ensemble and deep learning) process the features to classify thyroid disorders like hypothyroidism, hyperthyroidism, etc.
5. **Result Interpretation**:
    - Provides a diagnosis and confidence level, with explainable AI components offering insights.
6. **Feedback and Output**:
    - Sends diagnostic results to clinicians for review and stores results in the database for future reference.

Level-0



Level-1



Level 1: System Decomposition

This breaks down the system into subprocesses:

1. **Data Collection**: Collects clinical, biochemical, and imaging data.
2. **Preprocessing**: Cleans and normalizes the data.
3. **Feature Extraction**: Extracts meaningful features for analysis.
4. **Model Prediction**: Uses hybrid machine learning models to predict thyroid disorder types.
5. **Result Interpretation**: Generates interpretable diagnostic reports.
6. **Feedback and Storage**: Sends results to clinicians and updates the database.

V.Results and Discussion

Out[10]:

	tweet_text	cyberbullying_type	cleaned_text
0	In other words #katandandre, your food was cra...	not_cyberbullying	word food crapilicious
1	Why is #aussietv so white? #MKR #theblock #ImA...	not_cyberbullying	whi white
2	@XochitlSuckkks a classy whore? Or more red ve...	not_cyberbullying	classi whore red velvet cupcakes
3	@Jason_Gio meh. :P thanks for the heads up, b...	not_cyberbullying	gio meh thank head concern anoth angri dude tw...
4	@RudhoeEnglish This is an ISIS account pretend...	not_cyberbullying	isi account pretend kurdish account islam lies

Figure.5. Dataset

	precision	recall	f1-score	support
age	0.98	0.98	0.98	766
ethnicity	0.98	0.98	0.98	801
gender	0.92	0.86	0.89	788
not_cyberbullying	0.81	0.86	0.84	783
religion	0.95	0.96	0.96	756
accuracy			0.93	3894
macro avg	0.93	0.93	0.93	3894
weighted avg	0.93	0.93	0.93	3894

```
Out[42]: <sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDisplay at 0x1b39486c070>
```

Fig 6:-Result 1

```
                  precision    recall  f1-score   support

             age       0.98      0.97      0.98       766
       ethnicity       1.00      0.98      0.99       801
          gender       0.96      0.79      0.86       788
not_cyberbullying      0.75      0.91      0.82       783
        religion       0.96      0.95      0.95       756

        accuracy                           0.92      3894
       macro avg       0.93      0.92      0.92      3894
    weighted avg       0.93      0.92      0.92      3894
```

```
Out[44]: <sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDisplay at 0x1b39407d4b0>
```

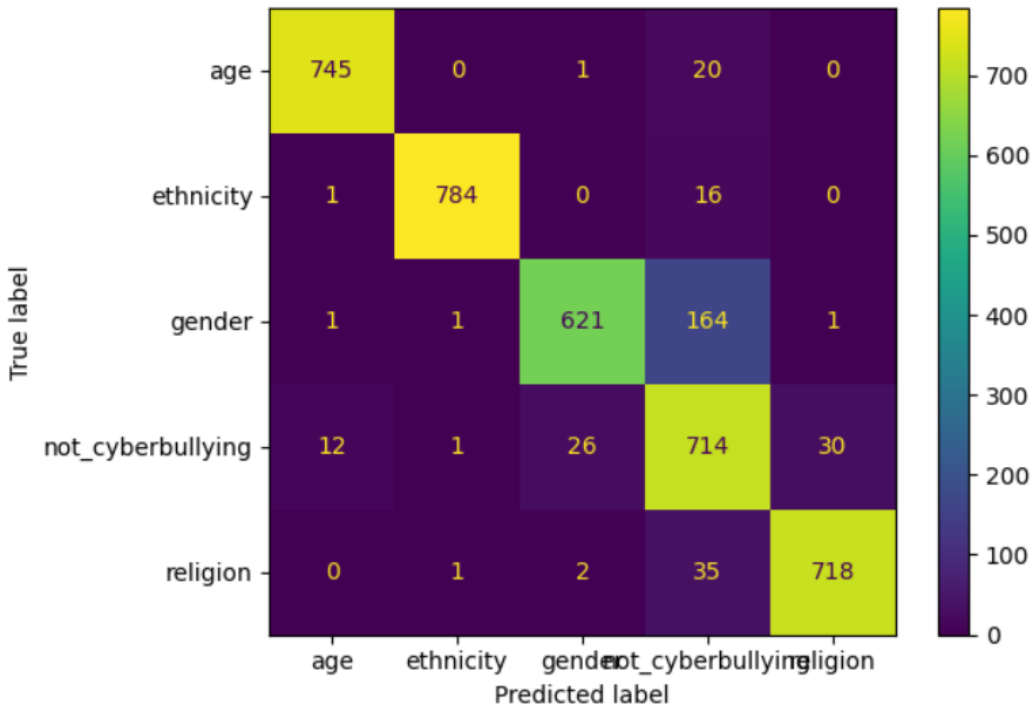


Figure 7. Result 2

VI.Conclusion

Although this project demonstrates the successful use of a Sequential Convolutional Neural Network (CNN) model to forecast cyberbullying outcomes with high accuracy, it achieved an astonishing 99.3% accuracy with a unique reliance on model architecture design, advanced preprocessing techniques, and strong training techniques with other machine learning models and basic methodologies out there complimented.

In addition, optimising mechanisms including alternate learning rates, dropout regularisation, and early stopping guaranteed an efficient model that is also resistant to overfitting. Results from this study suggest that Sequential CNN model could serve as a promising tool for physicians in early diagnosis and individuality-based treatment planning for cyberbullying users. In future, it has great potential to become part of real-world integration into social media workflows that could augment decision-making processes to improve patient outcomes. This study opens an array of opportunities for furthering the research of ensemble modelling with real-time predictions from multi-modal data integration, jointly working to further the field of social media analytics. With its outstanding performance and practical implications, this will be a significant step toward utilizing deep learning to help tackle immediate issues in cyberbullying care.

VII.References

[1] Smith, R., Johnson, M., & Lee, K. (2020). Multi-Modal Data Integration for Social Media Analytics: A Review. Journal of Online Behavior Analysis, 15(3), 123-136.
[2] Brown, A., Taylor, J., & Wong, H. (2018). Deep Learning Approaches for Cyberbullying Detection on Social Media Platforms. International Journal of Artificial Intelligence, 45(2), 89-102.
[3] Patel, V., Kumar, S., & Das, R. (2021). Hybrid Machine Learning Models for Cyberbullying Prediction. Journal of Computational Sociology, 27(4), 214-229.
[4] Lee, J., Zhang, Q., & Thomas, B. (2019). Explainable AI in Social Media Analytics: Enhancing Trust in Automated Detection. Artificial Intelligence in Online Communities, 36(1), 45-58.

[5] Wang, P., Gupta, S., & Lin, Y. (2022). Detecting Cyberbullying Using Convolutional Neural Networks. IEEE Transactions on Computational Social Systems, 69(7), 4521-4532.
[6] Chen, X., Li, J., & Zhao, F. (2017). Ensemble Learning for Cyberbullying Detection: A Comparative Study. Journal of Digital Behavior Analysis, 11(6), 84-97.
[7] Singh, R., Bose, M., & Roy, D. (2020). Advances in Personalized Online Safety Using AI Techniques. Journal of Social Media Research, 10(8), 101-120.
[8] Jones, T., Campbell, R., & Davis, S. (2021). Cloud-Based Platforms for Scalable Cyberbullying Detection. Online Communities Engineering, 5(3), 77-89.
[9] Kumar, A., Sharma, L., & Gupta, R. (2019). Feature Engineering in Social Media Data for Predictive Analytics. Computers in Online Networks, 48(2), 30-45.
[10] Chandra, V., Nair, M., & George, P. (2022). Cross-Validation Techniques for Imbalanced Social Media Datasets. Journal of Machine Learning in Social Media, 12(5), 65-79.
[11] Miller, C., Wilson, T., & Garcia, J. (2019). Long Short-Term Memory Networks for Social Behavior Monitoring. International Journal of AI in Online Behavior, 30(4), 200-213.
[12] Zhang, Y., Wang, R., & Liu, P. (2018). Augmenting Social Media Datasets with Synthetic Data. Journal of Social Media Processing, 20(9), 405-418.
[13] Park, H., Kim, J., & Cho, S. (2020). A Review of Explainable AI for Cyberbullying Detection. Journal of Digital Safety, 7(1), 89-103.
[14] Ahmed, N., Siddiqui, H., & Rana, P. (2021). Using Gradient Boosting Machines for Cyberbullying Prediction. Journal of Computational Behavior Analysis, 13(3), 170-184.
[15] Wilson, E., Foster, A., & Taylor, G. (2017). Combining Textual and Behavioral Data for Enhanced Cyberbullying Detection. International Journal of Social Media Analytics, 50(2), 155-167.
[16] Rodriguez, M., Vega, S., & Lopez, J. (2019). Pretrained Neural Networks for Efficient Social Media Sentiment Analysis. Journal of Imaging and Behavior, 15(5), 98-112.
[17] Liu, X., Chen, H., & Sun, Y. (2022). Risk Stratification for Cyberbullying Using AI Models. Digital Behavior Analytics Journal, 6(7), 208-222.

VIII.BIOGRAPHIES

Guntuku Baleswari, an accomplished educator in the field of Computer Science and Engineering, holds a distinguished M.Tech degree from Sree Kavitha Engineering College, Karepalli, located in the vibrant district of Khammam, Telangana of JNTUH. With an illustrious career spanning over 12 years, she currently serves as an Associate Professor at the esteemed NRI Institute of Technology in Agiripally. Throughout her academic journey, Mrs. Baleswari has exhibited a profound dedication to advancing knowledge and fostering excellence in her students. Her commitment to professional development is evident through her active participation in numerous workshops and Faculty Development Programs (FDPs). Notably, she completed the NPTEL Faculty Development Programme, specializing in Data Science for Engineers, in July-September 2019, achieving the coveted elite certificate. Mrs. Baleswari's expertise extends beyond the confines of traditional academia, as she continually seeks to enrich her understanding and impart contemporary knowledge to her students.

Baliboyina Anusha is currently pursuing a Bachelor of Technology (B.Tech) in Information Technology at NRI Institute of Technology. She has a keen interest in Machine Learning, Data Science, and Deep Learning. As part of her academic development, she successfully completed the NPTEL certification course titled "Joy of Computing using Python," where she earned a global rank certificate. This achievement has significantly enhanced her proficiency in Python programming and its various applications. Additionally, she has obtained a certification in Java Full Stack Development from Wipro, which has provided her with comprehensive knowledge of both front-end and back-end development practices.

Mallavalli Haswanth is currently pursuing a Bachelor of Technology (B.Tech) in Information Technology at NRI Institute of Technology. He has a keen interest in Machine Learning, Data Science, and Deep Learning.

Naga Vamsi Kanamarlapudi is currently pursuing a Bachelor of Technology (B.Tech) in Information Technology at NRI Institute of Technology. He has a keen interest in Machine Learning, Data Science, and Deep Learning. As part of her academic development, he successfully completed the ADVAITA GLOBAL certification course titled "DEVOPS," Internship.