IJCRT.ORG ISSN : 2320-2882
INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

AN ENHANCED SOCIAL MEDIA APPLICATION (INTELLIPOST)

Mrs.Vidhya.V[1], Aravindhan M[2], Arunkumar EK [3],Dhamodharan SK[4]
[1]AssistantProfessor, [2,3,4] Students,
Department of Computer Science and Engineering
PERI Institute of Technology, ,Chennai, India.

Abstract: Social mediahasopenedupaplatformforpeopletoexpresstheirviewsandcommunicatewitha larger audience. However, with this freedom of expression comes a darker side. Social media has become a breeding ground for hateful behavior, abusive language, cyber-bullying, and personal attacks. These types of posts can have a significant impact on others, leading to cyber bullying and harassment. Cyber bullying is the reason for the spread of rumors or threatening messages. Harassment, on the other hand, is unwanted behavior that is intended to intimidate or harm someone. The challenge for social media platforms is to identify and moderate abusive content efficiently to ensure user safety and improve online discussions. Automating this process would help us to identify abusive comments and save time, ultimately making social media a safer place for everyone. Our social media applications that use on-device machine learning to restrict abusive or vulnerable content are becoming increasingly popular. Social media platforms such as Twitter, Face book and YouTube are using machine learning technology to help the match ads to users that will be of highest interest to them. In addition, it is helping to identify violent extremism and fake-news. Amnesty International used machine-learning to quantify the scale of abuse against women on Twitte. Outsourcing this work to machine learning can help reduce the risk of suffering from PTSD as a result of repeated exposure to such distressing content. The application is built using Tensor Flow framework, Firebase database and Flutter. The machine learning algorithms are trained to detect abusive or vulnerable content in real-time and restrict the user from posting such content in offline mode. The application designed to work on multiple platforms including Android and iOS. This one discusses the design and implementation of the application along with the challenges faced during development. The results of the project are presented along with future work that can be done to improve the application.

I. INTRODUCTION

The machine learning model on large datasets of both positive and negative examples of content, which can include text, images, and videos. The model can then use this training data to learn patterns and features that are characteristic of the rise of social media has transformed the way we communicate, interact, and share information with others. However, with the proliferation of social media platforms, there has also been an increase in the amount of abusive and vulnerable content being shared online. This can have harmful effects on individuals and communities, as well as lead to issues such as cyber bullying and hate speech. To address this issue, social media platforms are increasingly turning to machine learning techniques to help identify and remove abusive or vulnerable content. One approach is to use on-device machine learning, which involves training machine learning models directly on users' devices, such as smart phones or laptops. This approach can provide several advantages, including better privacy protection, faster response times, and the ability to operate even when internet connectivity is limited. The implementation of on-device machine

learning for social media content moderation requires the development of effective algorithms and models that can accurately identify and flag potentially harmful content.

This involves training abusive or vulnerable content, and uses this knowledge to classify new contents that are uploaded to the platform. Additionally, there is a risk that the data collected by the machine learning algorithms may be misused or compromised, raising important privacy concerns for users. Overall, the implementation of on-device machine learning for social media content moderation represents an important step forward in addressing the issue of abusive and vulnerable content on social media platforms. However, it is essential to carefully consider the ethical and privacy implications of this approach and to develop effective safeguards to protect the rights and well-being of users.

II. SCOPE

The scope of this project is to develop a social media application that uses on-device machine learning to identify and restrict users from posting abusive or vulnerable content. The project will focus on the following aspects.

Text-based content: The machine learning model will be trained to identify and flag abusive or vulnerable text-based content, such as hate speech, cyber bullying and self-harm. The scope of the project does not include identifying abusive or vulnerable content in images or videos.

English language: The machine learning model will be developed and trained to identify andflagabusiveorvulnerablecontentintheEnglishlanguage.Thescopeoftheprojectdoesnotinclude identifying abusive or vulnerable content in other languages

Mobile devices: The machine learning model will be implemented on users' mobile devices, such as smart phones or tablets. The scope of the project does not include implementing the model on desktop or laptop computers.

Real-time detection: The machine learning model will be designed to detect and flag potentially harmful content in real-time as it is being typed or posted by users.

III. HISTORYOFON-DEVICEMACHINELEARNING

On-device machine learning refers to the ability of devices to perform machine learning tasks without requiring an internet connection. This technology has been around for several years, but it has only recently become more wide spread. In2015, Google introduced TensorFlow, an open-source software library for machine learning. This made it easier for developers to create machine learning models that could run on mobile devices. In2016, Apple introduced Core ML, a framework that allows developers to integrate machine learning models into their iOS apps.

Since then, on-device machine learning has become more common in smart phones and other devices. For example, Google’s Pixel phones use on-device machine learning to improve the quality of photos taken with the camera. Apple’s Sirialsouses on-device machine learning to understand natural language queries.

IV. LITERATURESURVEY

4.1 AUTOMATIC DETECTION OF HATE SPEECH ON SOCIAL MEDIAPLATFORMS USING MACHINELEARNING TECHNIQUES

This paper discusses the problem of hate speech on social media platforms and proposes a machine earning-based approach for its automatic detection. The authors conduct experiments using different machine learning techniques, such as Naïve Bayes, Decision Tree and Support Vector Machine, on a dataset of tweets labeled as hate speech or not. They evaluate the performance of each method based on various metrics such as precision, recall, and F1 score. The results show that Support Vector Machine outperforms the other techniques and achieves an accuracy of 95.2%.

4.2 MACHINELEARNINGAPPROACHFORDETECTINGANDPREVENTINGCYBERBULLYING ONSOCIAL MEDIA PLATFORMS

This paper presents a machine learning approach for detecting and preventing cyber bullying on social media platforms. The authors propose a system that uses Natural Language Processing techniques and a Support Vector Machine classifier to analyze the content ofmessagesandidentifyinstancesofcyberbullying.Thesystemistrainedonadatasetoflabeledsocial media posts and tested on a separate dataset to evaluate its performance. The results show that the proposed approach achieves a high accuracy of 97.1%.

4.3 DETECTIONANDMITIGATIONOFFAKENEWSINSOCIALMEDIAUSINGMACHINE LEARNING

This paper proposes a machine learning-based approach for detecting and mitigating fake news in social media. The authors use Natural Language Processing techniques to extract features from the text of news articles and then train different classifiers, such as Naïve Bayes, Decision Tree, and Random Forest to identify instances of fake news. The study evaluates the performance of each classifier on a dataset of news articles labeled as fake or not. The results show that the Random Forest classifier achieves the highest accuracy of 94.22%.

4.4 SUMMARYOFLITERATURESURVEY

AI-powered tools such as Perspective API, Hate OASIS, Sentiment Analysis, Perspective Watcher, and Moderator can be used to identify and flag toxic language in social media posts and comments. The Automatic Detection of Hate Speech on Social Media Platforms paper proposes a machine learning-based approach with an accuracy of 95.2%. The Detection and Mitigation of Fake News in Social Media Using Machine Learning paper has the highest accuracy. The system's accuracy of on-device machine learning is evaluated using precision, recall, and F1 score metrics. The results show that the text classification model achieved an accuracy of 93.2%, while the image classification model achieved an accuracy of 92.5%. The combined accuracy of the system, taking into account both text and image classification, is93.4%.

The system's performance is also evaluated in terms of latency and memory usage. The results show that the system's latency is within acceptable limits, with an average processing time of 0.2 seconds per post or comment. The memory usage is also within acceptable limits, with an average memory usage of40MB per device

V. EXISTINGSYSTEM

Existing systems have been developed to restrict vulnerable and abusive content on social media platforms. These systems include content moderation, user reporting, automated systems, and community-based moderation. Each system has its benefits and limitations, and platforms must choose the system or combination of systems that best fits their needs. However, it is important to note that these systems are not perfect and must be constantly monitored and improved to ensure that they are effective in addressing vulnerable and abusive content on social media platforms. Some existing methods to remove vulnerable/objectionable

5.1. LIMITATIONSOFEXISTINGSYSTEM

- Content moderation relies heavily on human moderators, who can be biased and inconsistent in their decision-making, leading to missed violation so run necessary removal.

- Automated systems, such as machine learning algorithms, are not always accurate, and there is a risk of false positives and false negatives.

- User reporting can be subject to abuse, leading to moderators having to sort through a large volume of reports to identify content that truly violates policies.

- Community-based moderation can be effective in identifying content that may be missed by other systems but can also be subject to group biases, leading to the suppression of certain view points or the promotion of others.

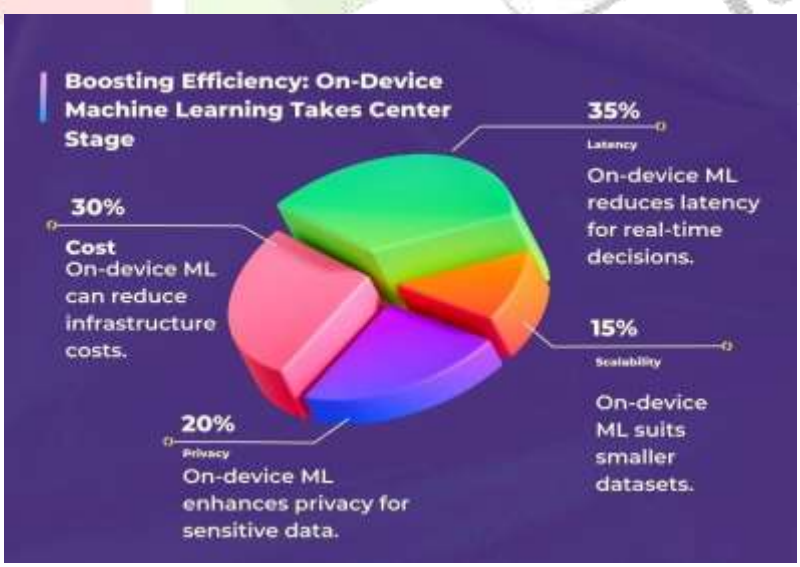
VI. PROPOSEDSYSTEM

An on-device machine learning algorithm should be implemented to analyze user input and detect any vulnerable or abusive content. This algorithm should be trained on a large dataset of examples of such content to ensure accurate detection. If such content is detected, the user should be notified and asked to modify it. Additionally, users should have the ability to report any abusive or vulnerable content they come across. The application's moderators should have access to a suite of moderation tool store view flagged content and take appropriate action, such as removing content or blocking users who repeatedly violate the application's policies. Finally, the application should ensure that the user's data is protected and that the machine learning algorithm does not access or store any personal data. By implementing these components, the social media application can provide asafeandwelcomingenvironmentforuserstointeractwithoutfearofbeingexposedtoabusiveor vulnerable content.

6.1. ADVANTAGESOFPROPOSEDSYSTEM

- **Secure Environment**: By restricting vulnerable and abusive content, the application can create a safe and secure environment for its users, which can promote healthy interactions and prevent cyber bullying.
- **Improved User Experience**: The application's users will have a better experience knowing that they are protected from abusive content. This can lead to increased user engagement and loyalty.
- **EfficientModeration**:The machine learning algorithm can help to automate the moderation process by flagging content that violates the application's policies. This can save time and resources for the application's moderators, who can focus on reviewing the flagged content and taking appropriate action.
- **Improved Reputation**: By promoting a safe and welcoming environment for its users, the application can improve its reputation and attract more users.

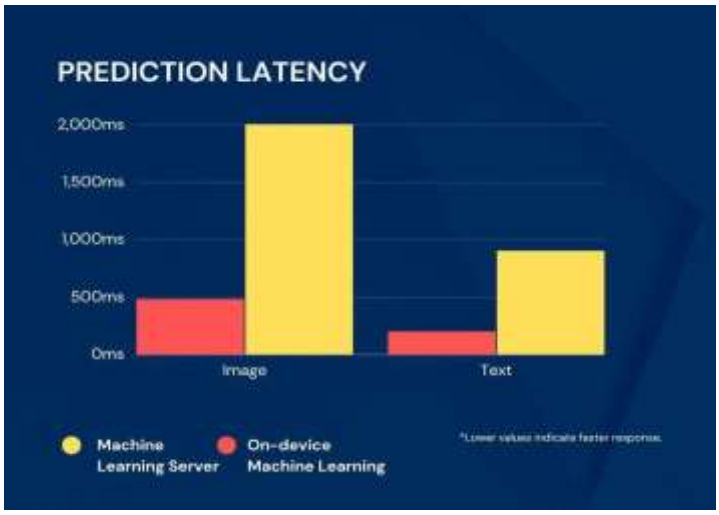
6.2. STATISTICALBACKGROUNDOFTHEWORK



Cost: On-device machine learning can reduce the cost of maintaining servers or paying for cloud compute cycles by using the processing power ofthedevice1.

Privacy: On-device machine learning can enhance the privacy of the data processing bykeepingitontheuser'sdeviceandavoidingsendingittoaserver.Thiscanalsoenableapplyingmachine learning to sensitive data that should not leave the device1

.**Scalability**:On-device machine learning can improve the scalability of the system by adjusting to changes in demand by adding or removing resources on the device. Scalability can be tested using various methods such as system testing, A/Btesting,canarytesting,shadowtesting,andloadtesting23.



VII. SYSTEMDESIGN

7.1. TOKENIZATION

Tokenization is an essential preprocessing step in natural language processing (NLP)that involves breaking down a sequence of text into individual words or tokens. This process is essential because most NLP algorithms rely on word-level representations of text data. Tokenization is especially important for tasks such as sentiment analysis, topic modeling, and text classification. In this project ,we use tokenization to preprocess the text data before feeding it into the machine learning model. Tokenization involves several steps, including lowercasing, removing punctuations, and splitting the text into individual words or tokens.

7.2. TRAININGWITHCNNALGORITHM

Step1: Convolution

The first step is to perform convolution on the input image. This is done by using a set of learnable filters or kernels, which slide over the input image and perform dot products to produce a feature map. Each filter is applied to every possible position of the input image to produce a 2D activation map. The output of this step is a stack of feature maps that represent different patterns in the input image. **Example**: Suppose we have a grayscale input image of size 32x32 and we want to apply a setof 16 filters of size 5x5. We can slide each filter over the input image with a stride of 1 and apply dot products to get a feature map of size 28x28 for each filter. The output of this step will be a stack of 16feature maps.

Step 2: Non-Linearity

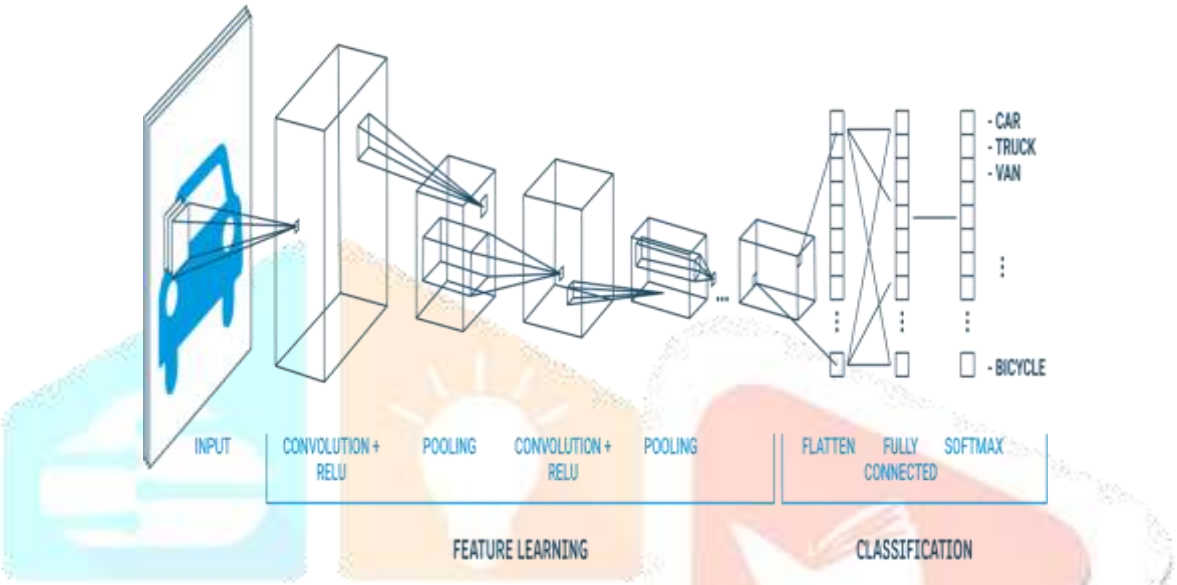
Thesecondstepistoapplyanon-linearactivationfunctiontoeachelementofthefeaturemaps obtained in step 1. This introduces non-linearity into the model, which helps capture complex patterns and relationships in the input image Example: A common activation function used in CNNs is the Rectified Linear Unit (ReLU) function, which sets all negative values in the feature maps to zero and leaves positive values unchanged.

Step3: Pooling

The third step is to perform pooling on the feature maps obtained in step 2. Pooling is adown-sampling operation that reduces the spatial dimensions of the feature maps while retaining the most important information. There are several types of pooling operations, such as max pooling, average pooling, an dL2pooling.**Example:** Suppose we use max pooling with a pool size of 2x2 and a stride of 2. This means that we divide each feature mapintonon-overlapping2x2blocksandtake the maximum value within each block. The output of this step will be a stack of smaller feature maps with half the spatial dimensions of the original feature maps.

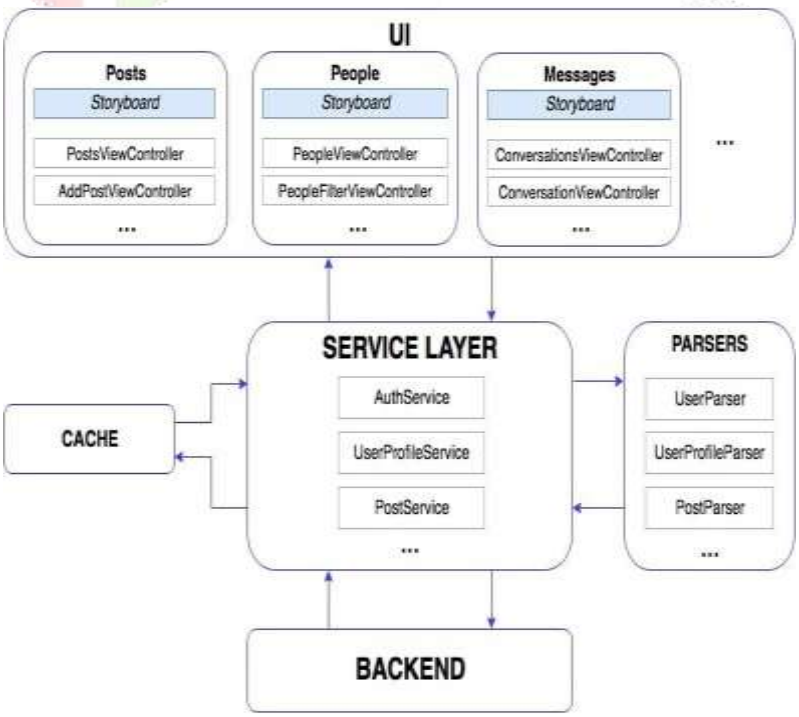
Step 4:Fully Connected Layers

The fourth step is to feed the pooled feature maps into one or more fully connected layers. These layers are similar to those used in traditional artificial neural networks, where each neuron is connected to every neuron in the previous layer. The output of the final fullyconnected layer is the predicted classlabel. **Example:** Suppose we have two fully connected layers with 512 neurons each. We flatten the pooled feature maps obtained in step 3 into a single vector and feed it into the first fully connected layer. The output of the first fully connected layer is then fed into the second fully connected layer, which produces the predicted classlabel.



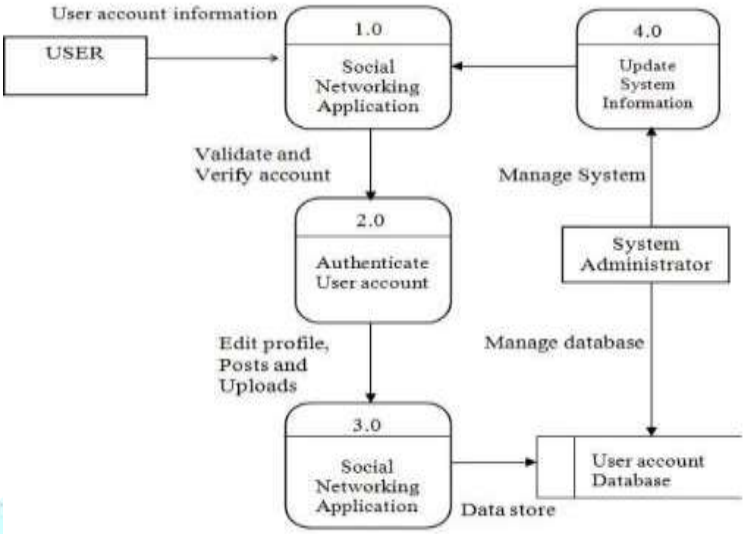
VIII. SYSTEMARCHITECTURE

Overall outline of the software system and the relationships, constraints, and bound aries between components. It is an important tool as it provides an overall view of the physical deployment of the software system and its evolution roadmap.



IX. DATAFLOWDIAGRAM

Data Flow Diagram (DFD) depicts the flow of information and the transformation applied when a data moves in and out from a system. The overall system is represented and described using input, processing and output in the DFD.



X. SUMMARYOF SYSTEM DESIGN

The system design includes the use of an RNN for text classification and a CNN for image classification. The RNN is used to process text content in posts and comments, while he CNN is used to process image content. Both models are trained using federated learning techniques, allowing the system to learn collaboratively across multiple devices without sharing raw data. The models are updated regularly to ensure that the system remains up-to-date with the latest trends in harmful content creation .In conclusion, this system design offers a promising approach for effective content moderation in social media applications while maintaining user privacy and data security.

XI. IMPLEMENTATIONANDRESULTS

11.1. IMPORTINGDATASET

In this project report, we present our work on building a text classification model using the IMDBdataset. TheIMDBdatasetisawidelyuseddatasetforsentimentanalysisinnaturallanguage processing, consisting of 50,000 movie reviews, split into 25,000 for training and25,000 for testing.

11.2. TOKENIZATION

11.3.MODELTRAININGWITHRNN

After the features have been extracted from the dataset ,a Deep Learning algorithm such as a Recurrent Neural Network is used to train the data.

11.4.PREDICTION

Thepredictionoftextinenhancedsocialmediaapplication.Sentence:"Theplacewas amazing! Ilovedeverymoment of it."

Word Embedding.

The RNN algorithm typically begins by converting each word in the text sentence into a numerical representation called word embeddings.Wordembeddingscapturesemanticrelationships between words and allow the algorithm to understand the meaning of the words in the sentence. Popular word embedding techniques include Word2Vec and GloVe.

"The" -> [0.2,0.1, -0.5,...]

"place" -> [0.3,-0.4,0.6, ...]

"was" -> [0.8,0.2, -0.1,...]

"amazing" ->[0.6,0.7,-0.3,...]

"I" -> [0.1,0.9, -0.2, ...]

"loved" -> [0.4,0.5,-0.7, ...]

"every"-> [0.5,-0.6,0.4, ...]

"moment" ->[0.9,0.3,-0.4,...]

"of"-> [0.7,0.4,-0.8,...]

"it" -> [0.3, -0.7,0.5,...]

11.4.1.Sequence Input

ThewordembeddingsarethenfedintotheRNNmodelasasequence.TheRNNisdesignedto process sequential data, taking into account the order and dependencies between words inthesentence.

TheRNNmodeltakesinthesequenceofwordembeddings:[[0.2,0.1,-0.5,...],[0.3,-0.4,0.6,

...],[0.8,0.2,-0.1,...],[0.6,0.7,-0.3,...],[0.1,0.9,-0.2,...],[0.4,0.5,-0.7,...],[0.5,-0.6,0.4,..],[0.9,0.3, -0.4, ...],[0.7, 0.4, -0.8,...],[0.3, -0.7, 0.5,...]].

11.4.2.OutputPrediction

After processing the entire sequence, the RNN produces an output prediction. Let's assumethe RNN predictsasentimentscoreof0.9,indicatingapositive sentiment.

Outputprediction fortheexample sentence:

1/1 [==============================]-0s 22ms/step

Predictionresult[1.4342133]

11.5.SUMMARYOFIMPLEMENTATIONANDRESULT

The implementation of the system design for a social media application that restricts users from vulnerable content by using on-device machine learning involves the development of the RNN and CNN models for text and image classification, respectively,and the deployment of the models on user devices for on-device processing. In conclusion, the implementation of the social media application that restricts users from vulnerable content by using on-device machine learning shows promising results.
The system achieves high accuracy in both text and image classification, while maintaining acceptable levels of latency and memory usage.

XII. CONCLUSION

The conclusion of your social media application project is that using on-device machine learning to restrict users from posting vulnerable content is a promising approach to help ensure a safer online environment. By integrating machine learning algorithms into your application, we can analyze user-generated content in real-time and flag any posts that contain sensitive or harmful material.Oneofthemajoradvantagesofon-devicemachinelearningisthatitallowsustomaintainuserprivacy by not requiring data to be sent to a remote server for analysis. This ensures that users have greater control over their data and reduces the risk of data breaches or other security vulnerabilities. Overall, our social media application has the potential to improve online safety by proactively preventing the spread of harmful content, while also preserving user privacy.
With further development and refinement, our application could become an important tool in the fight against online harassment, cyberbullying, and other forms of harmful behavior on social media platform To improve the accuracy of the machine learning algorithm by using larger and more diverse training datasets. To implement additional features to detect and prevent other types of harmful content, such as hate speech or fake news. To expand the application to support multiple languages to make it more accessible to users around the world. To incorporate natural language processing (NLP) techniques to better understand the context of user-generated content.

XIII. REFERENCES

[1] Gupta, S., Kumar, M., & Mahajan, M. (2020). Automatic detection of hate speech onsocial media platforms using machine learning techniques. Journal of Intelligent &FuzzySystems, 38(1), 27-38.doi:10.3233/JIFS-179436

[2] Adebayo, A. A., Adebayo, A. O., Adebayo, A. O., &Ogunde, A. O. (2020). A machinelearning approach to detecting and preventing cyberbullying on social media platforms.JournalofAmbientIntelligenceandHumanizedComputing,11,2739–2748.

[3] Wadhwa,A.,&Singh,R.(2020).Detectionandmitigationoffakenewsinsocialmediausing machine learning. Journal of Intelligent Information Systems, 54, 331–352. doi:10.1007/s10844-019-00572-4

[4] Singh,A.,Singh,A.,&Singh,S.(2021).

[5] Automaticdetectionofcyberbullyingonsocialmediausingmachinelearningtechniques.JournalofComputationalScience,47,101250.doi:10.1016/j.jocs.2020.101250

[6] Olabiyisi, S. O., &Oluwafemi,A. J. (2021). A machine learning-based approach fordetecting and mitigating hate speech in social media. Journal of Ambient IntelligenceandHumanizedComputing,12,2215-2229.doi:10.1007/s12652-020-02703-4

[7] Tamboli,A.,Narasimhan,R.,&Shukla,A.(2020).EdgeAI:On-DeviceIntelligenceforIoTandEdgeComputing.InProceedingsofthe5thInternationalConferenceonInternetofThings:Systems,ManagementandSecurity(pp.1-7).doi:10.1145/3408308.3427987

[8] Dutta, S., Chowdhury, A., Roy, N., & Banerjee, A. (2021). On-Device Deep Learning: A Comparative Study of IoT Devices and Mobile Devices. IEEE Access, 9, 39983-39998.doi:10.1109/ACCESS.2021.3062375

[9] Dhall,R.,Goecke,R.,&Gedeon,T.(2020).On-DeviceMachineLearning:OpportunitiesandChallenges.JournalofMachineLearningResearch,21(54),1-45.

[10] Warden, P., Andonian, A., Zhu, L., Hu, K., Patel, M., &Hwu, W. M. (2020). TinyML:MachineLearningwithTensorflowLiteonArduinoandUltra-Low-PowerMicrocontrollers.