****

**SRM INSTITUTE OF SCIENCE AND TECHNOLOGY**

**SCHOOL OF COMPUTING**

**DEPARTMENT OF COMPUTING TECHNOLOGIES**

**VIII SEMESTER – JANUARY 2023**

**18CSP109L - PROJECT**

**Guide name: Dr. Vegesna S M Srinivasavarma**

**Designation: Research Assistant Professor  
Department: CTECH**

**Batch ID: 415**

Student 1 Reg. No: RA1911033010109

Name: Burra Sai Krishna Bharadwaj

Student 2 Reg. No: RA1911003010621

Name: Gutta Koteswararao

**PROJECT TITLE**

**A Hybrid Approach for Detecting Cyberbullying on Social Media Platform using Machine Learning Algorithms**

# ABSTRACT

Cyberbullying is becoming more common on social media platforms. Given the popularity and prevalence of social media by people of all ages, it is very important to protect social media platforms from cyberbullying.

This study focuses on how to identify, understand the comments and detect the cyberbullying. This model presents a hybrid machine learning model to detect CyberBullying on social media network.

Cyberbullying is spread across various social media platforms. It is a wrong deed in which the victim is harassed by receiving the derogatory / provocative / sensitive images or text messages by the bully. Detection of such message/post in such large platforms is very difficult and may sometimes lead to false detection. A approach is proposed to cyberbullying detection in social media platforms by using different machine learning algorithms such as SVM, logistic regression, RNN which improves over the existing results. The model is trained and evaluated on two social media datasets of which one dataset is small size and the second dataset is relatively larger size.

# INTRODUCTION

Social media applications such as Instagram, twitter, facebook, whatsapp etc have become most used and preferred platforms for interaction between people across the globe. These platforms enable people to communicate in different ways, it also leads to vicious ways.

Our proposed work provides a way in detecting cyberbullying using algorithms such as Random Forests, SVM, RNN etc and evaluate the best possible way in detection.

To develop an effective system for automatically detecting instances of cyberbullying on WhatsApp using a hybrid approach that combines rule-based and machine learning techniques. And improve the accuracy and efficiency of cyberbullying detection compared to existing state-of-the-art methods. Developing a system that is capable of handling the complexities of language and behavior used in cyberbullying and can detect new and emerging forms of such behavior.

We ensure that the system protects user privacy and security while detecting instances of cyberbullying. By explore the potential applications of the system in social media platforms, educational institutions, law enforcement agencies, and mental health professionals. We are going to contribute the growing body of research on the use of machine learning for cyberbullying detection and prevention.

# 2.LITERATURE REVIEW

This section gives an overview the study of previously proposed models in cyberbullying detection which are related to our current work.

Detection of Hate Tweets using Machine Learning and Deep Learning by Lida Ketsbaia, Biju Issac, Xiaomin Chen , explores machine learning approaches using word embeddings such as Distributed Bag of Words and Distributed Memory Mean and the performance of Word2vec Convolutional Neural Networks (CNNs) to classify online hate. This paper used several methods for two data sets linear SVC produced highest accuracy and Bernoulli naïve Bayes produced lowest. using continuous bag of words + CNN resulted them an accuracy of 94% and 98.20% of accuracy while integrating CNN + Skip gram [1].

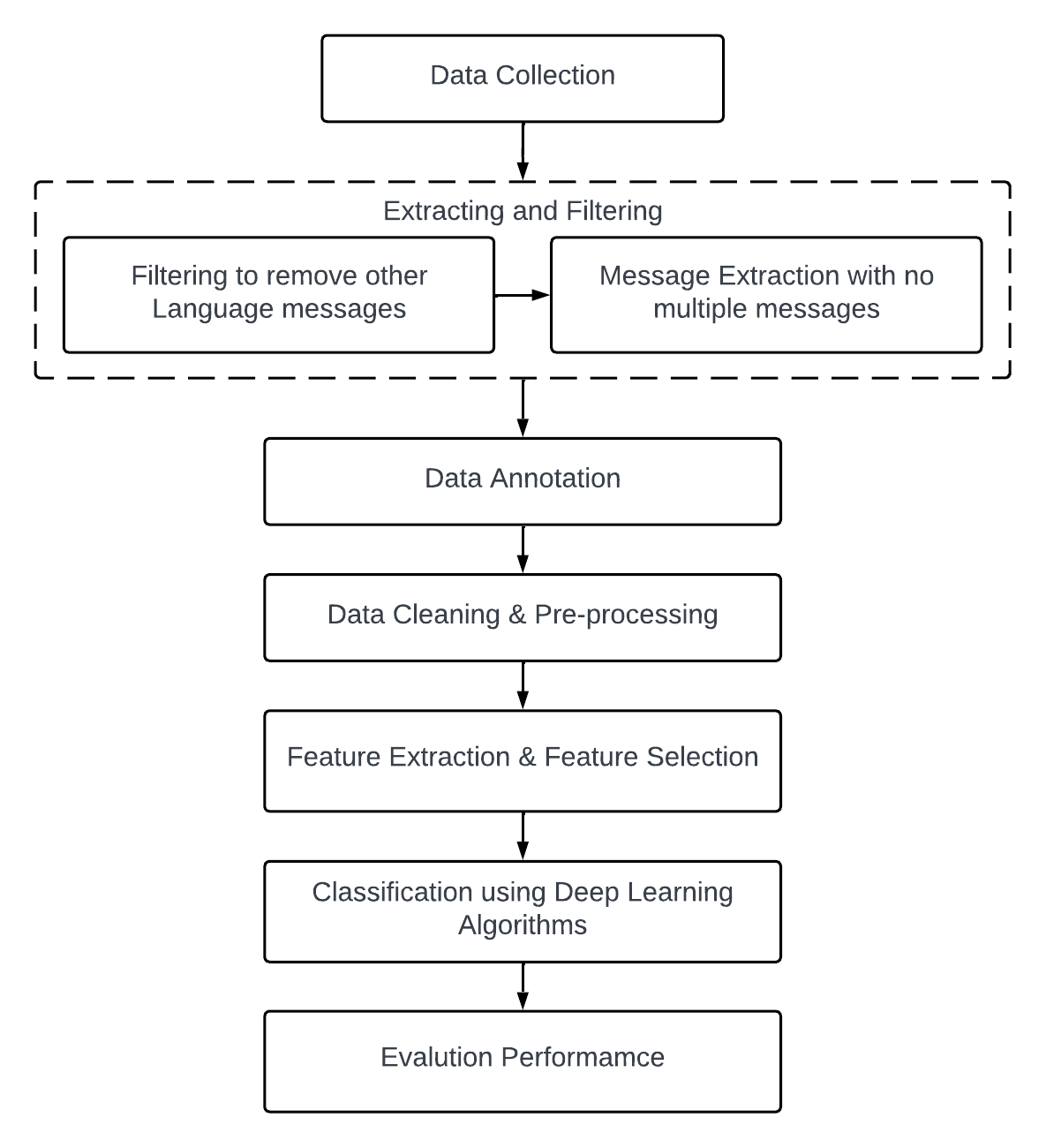
DEA-RNN: A Hybrid Deep Learning Approach for Cyberbullying Detection in Twitter Social Media Platform by Belal Abdullah Hezam Murshed, Jemal Abawajy, Suresha Mallappa, Mufeed Ahmed Naji Saif, Hasib Daowd Esmail Al-Ariki model combines Recurrent Neural Networks with Dolphin Echolocation Algorithm for classifying cyberbullying in tweets. The produced results of this paper show that DEA-RNN was having upper hand in detecting cyberbullying compared to Bi-directional long short term memory, Multinomial Naïve Bayes [2].

Optimized Twitter Cyberbullying Detection based on Deep Learning this paper they proposed an optimised twitter cyberbullying detection based on deep learning(OCDD) using CNN. They explored new detection techniques using training data by human intelligence and word embedding was for each word using Glo Ve technique. CNN was used for classifying. Which gave great results on text mining tasks however not used on cyberbullying detection context [3].

A new Approach is proposed to cyberbullying detection using pre-trained BERT model with a single linear neural network layer on top as classifier. In this approach of using pre-trained BERT model which is based on the complex and novel deep neural network using single linear linear layer of neural network gives better and stable results compared to CNN [4].

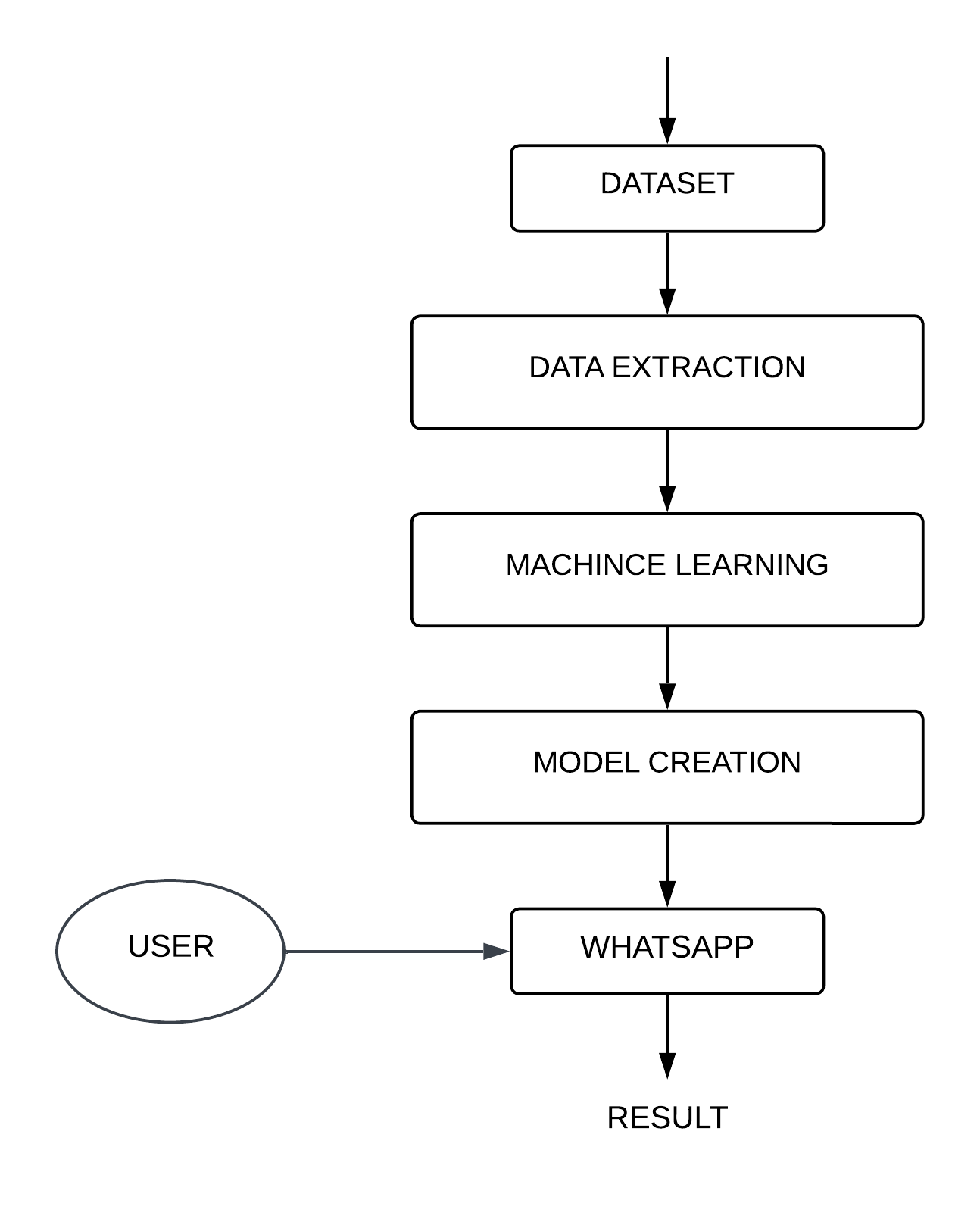
This paper proposed a prediction method of the user behaviour on his posts on social networking sites specifically twitter tensor flow have been used to predict analysis. NLP is applied to break text and CNN for classifying images. This experiment presents classifying cyberbullying and cyber trolling. Tensor flow contribution library is used to generate the prediction from data set. And converted training and test set into Numpy array.

**3. Architecture Diagram**

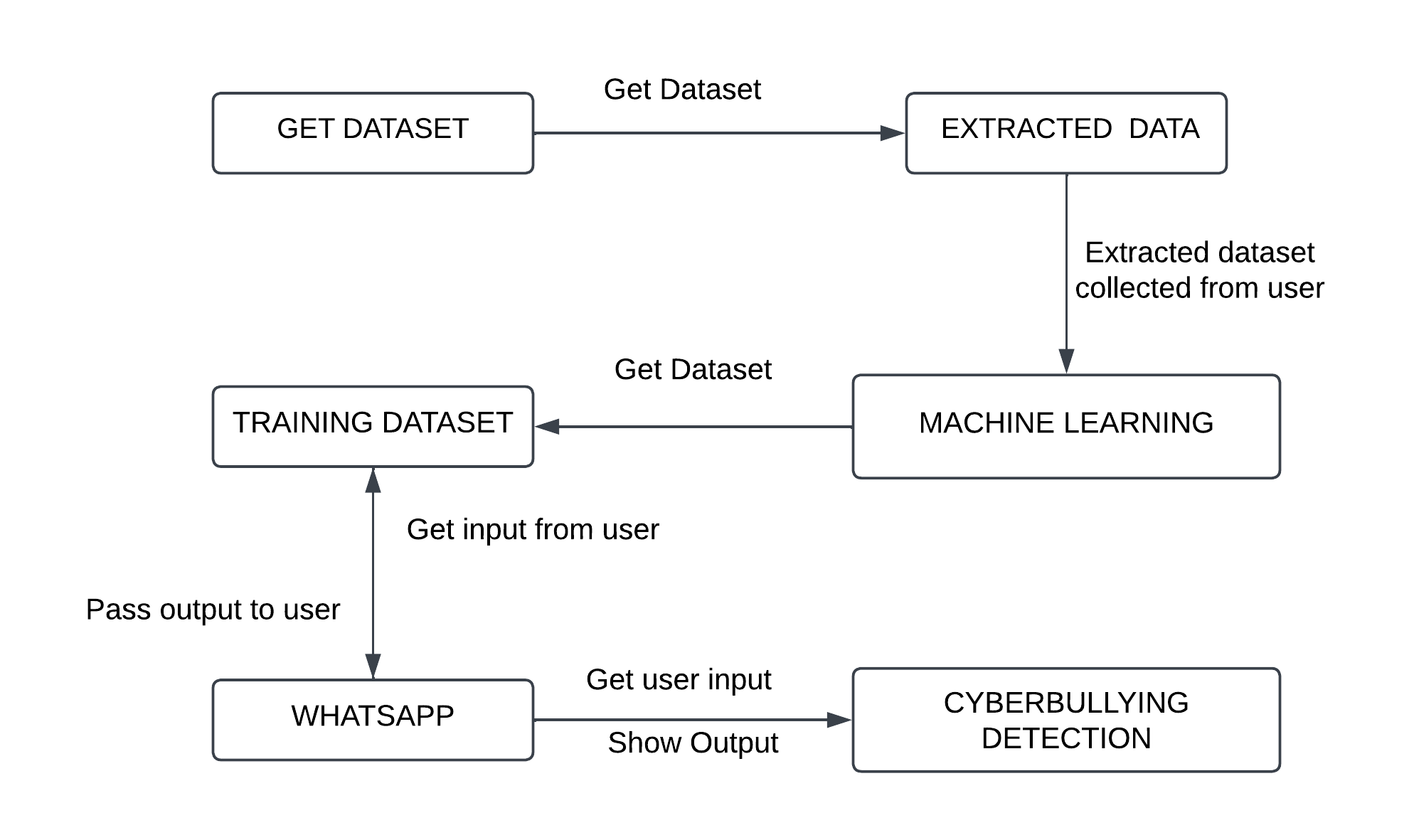


### Fig 3.1: Architecture Diagram

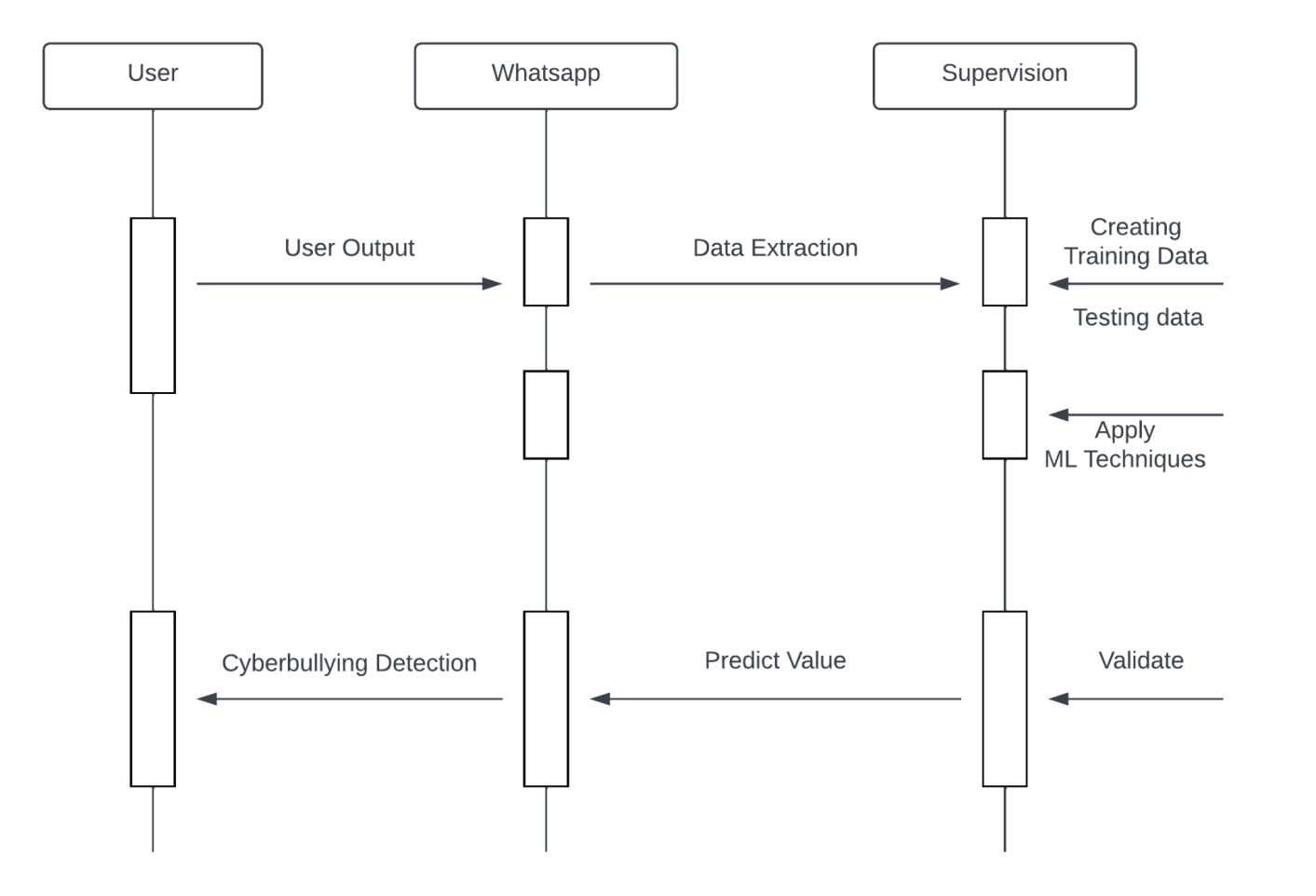
**4 .UML Diagrams**

****

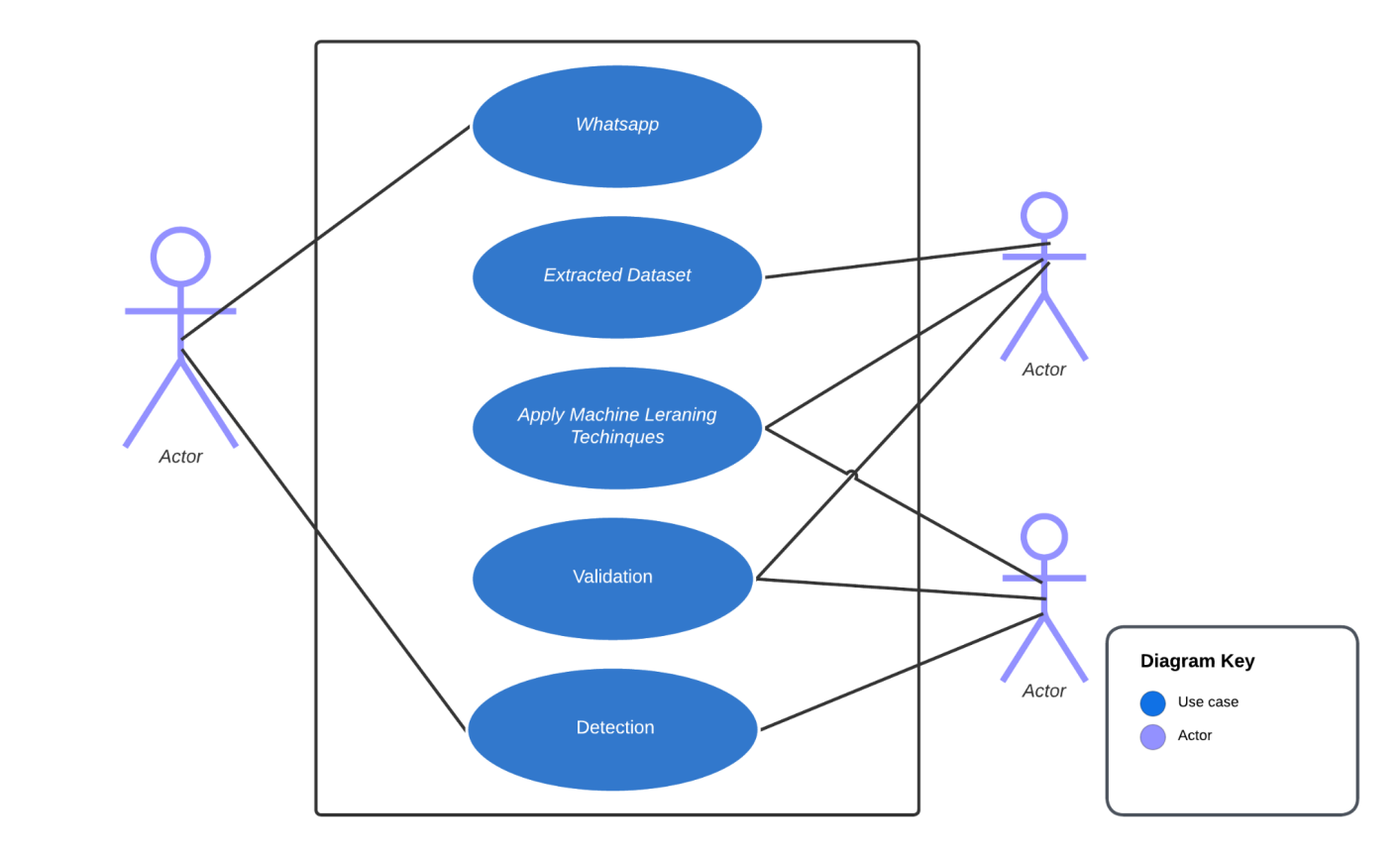
**Fig 4.1: Activity Diagram**



**Fig 4.2 : Collaboration Diagram**

****

**Fig 4.3 : Sequence Diagram**



**Fig 4.4 : Use Case Diagram**

**Modules Description**

**1. Data Collection**

Dataset required for our work is collected from Google website.

**2. Data Pre-Processing**

The data cleansing and pre-processing phase contain three sub-phases. In the first sub-phase, noise removal such as URL removal, hashtag/mentions removal, punctuation/ symbol removal, and emoticon transformation processes are performed. In the second sub-phase, Out of Vocabulary Cleansing such as spell checking, repeated Characters removal are performed. In the final sub-phase, tweet transformations such as lower-case conversion, stemming word segmentation, and stop word filtering are conducted. These subphases are performed to enhance the tweets and improve feature extraction and classification accuracy.

**3. Classification and Detection**

Random Forest is a commonly used machine learning algorithm that combines the outputs of multiple decision trees to produce a single result. Its ease of use and flexibility are driving its adoption as it can handle both classification and regression problems.

SVM is one of the most popular supervised learning algorithms used for both classification and regression problems. It is mainly used for machine learning classification problems. It creates the decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category.

Recurrent neural network (RNN) is a type of artificial neural network which uses sequential data or time series data. This deep learning algorithm are commonly used for ordinal or temporal problems, such as language translation, Natural Language Processing (NLP), speech recognition.

**4.Extracting and filtering:**

Extracting and filtering of data are two important data processing tasks that are often used in data analysis and machine learning.

Extracting data refers to the process of selecting and retrieving specific data from a larger dataset. This is typically done when we want to focus on a particular subset of the data that is relevant to our analysis. Extraction can involve selecting specific columns or rows of data based on some criteria, such as a date range, a particular category or tag, or a specific data value. Extraction can be performed using tools such as database queries, scripting languages like Python, or data visualization software.

**5. Feature Extraction and Feature Selection:**

Feature Extraction and Feature Selection are both techniques used to reduce the dimensionality of data in machine learning. Feature Extraction involves transforming raw data into a set of meaningful features that can be used to train a model, while Feature Selection involves selecting a subset of the most relevant features from the original dataset.

**6.Algorithms**

Support Vector Machine (SVM) is a type of supervised learning algorithm that aims to find the best separating boundary (hyperplane) between different classes of data. The algorithm uses a kernel function to map the input data into a high-dimensional feature space, where the boundary can be found more easily.

Logistic Regression is also a type of supervised learning algorithm that is used for classification tasks. It models the probability of the outcome variable (i.e., the class label) given the input data. Logistic regression assumes a linear relationship between the input features and the log-odds of the outcome variable.

References

1. Lida Ketsbaia, Biju Issac and Xiaomin Chen, “Detection of Hate Tweets using Machine Learning and Deep Learning,” in Proc. of the 19th IEEE International conference on Trust, Security and Privacy in Computing and Communications, Guangzhou, China, pp. 751-758, 2020.
2. Belal Abdullah Hezam Murshed, Jemal Abawajy, Suresha Mallappa, Mufeed Ahmed Naji Saif, and Hasib Daowd Esmail 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.
3. Monirah A. Al-Ajlan, and Mourad Ykhlef, “Optimized Twitter Cyberbullying Detection based on Deep Learning,” in Proc. of the 21st Saudi Computer Society National Computer Conference (NCC), Riyadh, Saudi Arabia, pp. 1-5, 2018.
4. Jaideep Yadav, Devesh Kumar, and Dheeraj Chauhan, “Cyberbullying Detection using Pre-Trained BERT Model,” in Proc. of the International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, pp. 1096-1100, 2020.
5. Nidhi Chandra, Sunil Kumar Khatri, and Subhranil Som, “Cyberbullying Detection using Recursive Neral Network through offline Repository,” in Proc. of the 7th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, pp. 748-754, 2018.