IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 6, June 2022

DOI: 10.17148/IJARCCE.2022.116104

ABUSIVE CONTENT DETECTION (Using Sentimental Analysis)

Dr. D. VIJAYA LAKSHMI[1], MURIKI SIRICHANDANA [2], ENDLA PAVANI BHAVYA [3], CH. SHASHIREETHAM [4]

[1]Head of the Dept of Information Technology, MGIT, Hyderabad,500075, India

[2,3,4]UG Student, Dept of Information Technology, MGIT, Hyderabad, 500075, India

Abstract: The increase in online services like medical consultancy on internet, business trades taking place in the web so on and so forth, in many means have proliferated the need for abusive content detection and its prevention because of the vast data flowing in the web. In addition to that, the multiplied cyber crimes have been concerning the communities that are looking to enlarge their business by making it accessible to a wide range of communities by moving to the global networks.

Abusive content is against humanity and might lead to mental disabilities like depression and many more severe issues amongst the victims, so it is a basic need to prevent the abuse and make sure the data is secure that is provided by clients belonging to various sections of the society.

This project involves detecting multiple cases of abuse in the comments, tweets, and messages. It detects positive and negative sentences.

INTRODUCTION

The word Abuse means improper usage or treatment of a thing to gain benefit improperly. It is an action that intentionally harms another person. Abuse can come in many forms, such as physical or verbal maltreatment, injury, assault, violation, rape, unjust practices, crimes, or other types of aggression. Digital Abuse is using technologies such as texting and social networking to bully, harass, stalk or intimidate a partner. Often this behavior is a form of verbal and emotional abuse perpetrated online.

LITERATURE SURVEY

S.No.	TITLE	AUTHORS	YEAR	METHODOLOGY	REMARKS
1	Robust Internet abuse detection method	Zhou Fa, Guang-Gang Geng, Zhi-Wei Yan, Xiao-Dong Lee	2018	Random Forest Naive Baye's algorithm	Used to recognize illegal websites and is highly efficient
2	Automatic Detection of Online Abuse and Analysis of Problematic Users in Wikipedia	Charu Rawat, Arnab Sarkar, Sameer Singh, Rafael Alvarado, and Lane Rasberry	2019	Logistic regression XGBoost	Used to build a model to detect blocked users
3	Offensive Language Detection using Artificial Neural Network	Meredita Susanty, Sahrul, Ahmad Fauzan Rahman, Muhammad Dzaky Normansyah, Ade Irawan	2019	ANN modeling, sigmoid function,	The model can understand the context and detect offensive content.

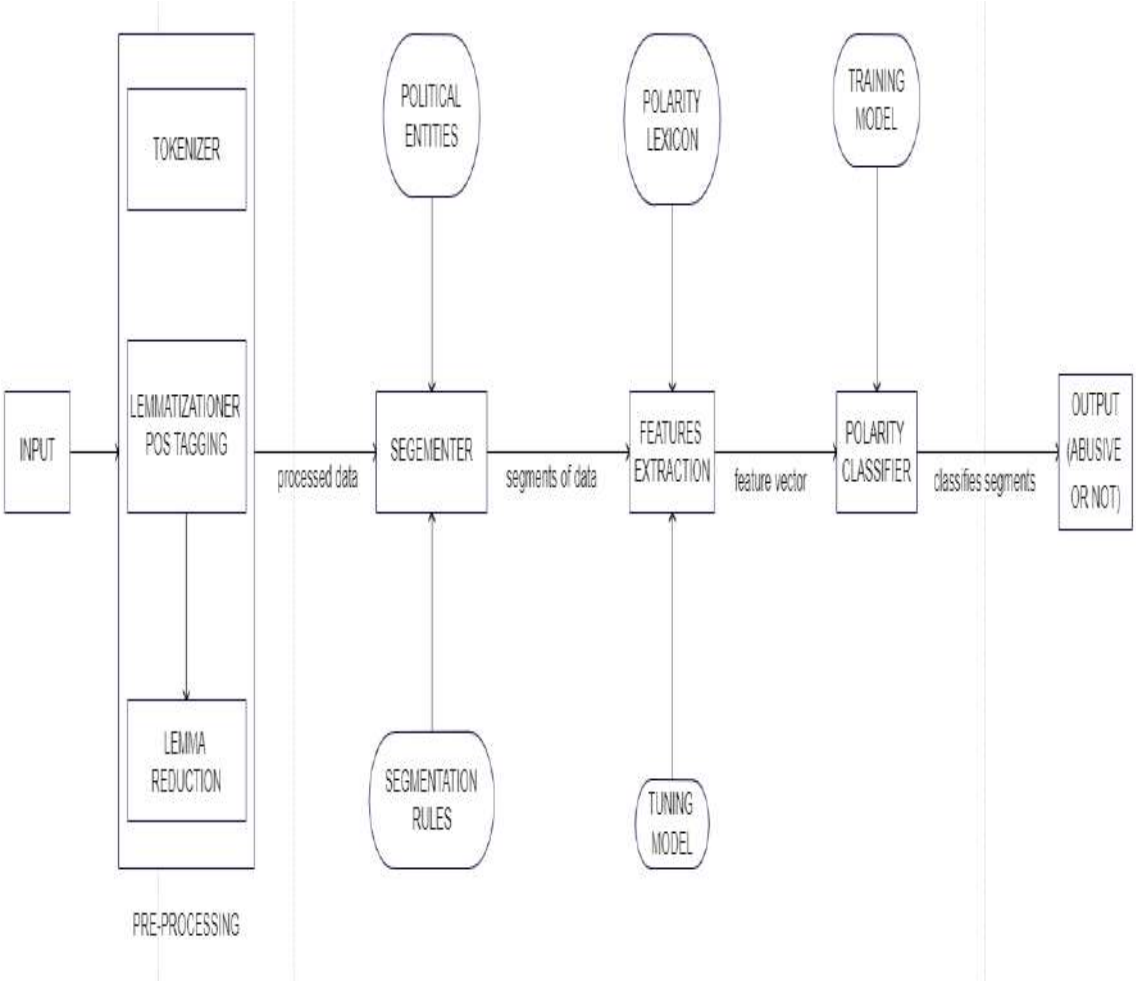
IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 6, June 2022

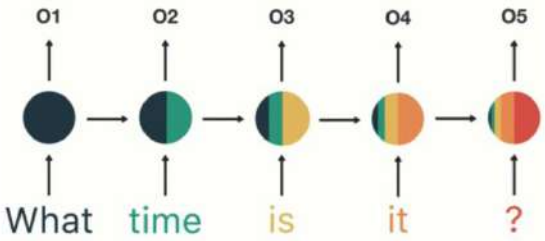
DOI: 10.17148/IJARCCE.2022.116104

SYSTEM DESIGN



Pre-processing: The given Input is pre-processed in 3 different phases

- **Tokenizer:** Tokenization is a technique to separate a piece of text into smaller units called tokens. Tokens can be either words, characters, or subwords.



In the Recurrent neural network architecture of Natural Language Processing, the tokens are received and processed at a particular time step. Here word tokenization is done for abusive content detection.

- **Lemmatization parts of speech tagging:**

The process of lemmatization involves converting the words in a sentence to dictionary forms. The parts of speech tagging are the process of identifying the parts of speech in a sentence. The POS tagging is a supervised learning approach in which it checks for previous and next words in a sentence to identify the parts of speech. nltk Library is used in this process. It extracts the stop words and they are removed to identify the words which are not frequently or commonly used.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 6, June 2022

DOI: 10.17148/IJARCCE.2022.116104

Segmenter: It is the process of segmentation that involves dividing the sentence into segments for further processing.

- **Political Entities:** Political entities are identified in POS tagging.
- **Segmentation rules:** According to the identified segmentation the segmentation rules are applied and segmentation of sentences is done. These segmented sentences are further sent for the feature extraction process.

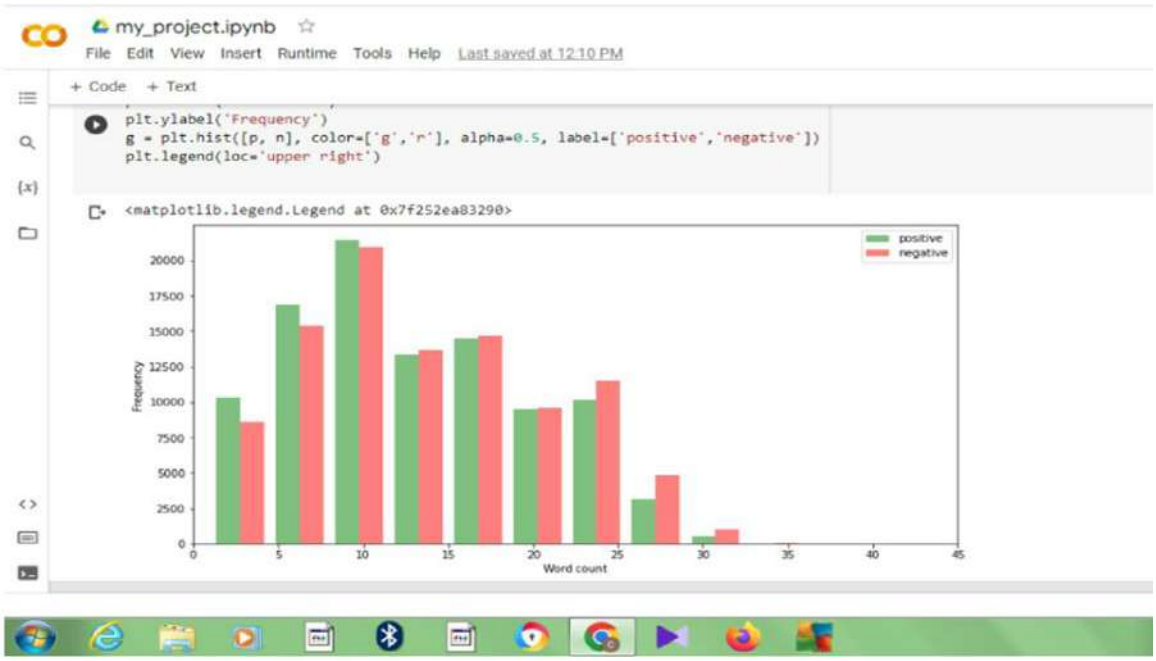
Feature Extraction: It is the process of extracting features from input data and generating distinct properties (feature vectors) which are informative and not redundant for improving the learning process of the system.

- **Polarity Lexicon:** Polarity Lexicons are associated with 3 numerical scores – positivity, negativity, and neutrality. SVM is used for assigning the polarities.
- **Tuning model:** After extracting the features, a tuning model is made to control the overall behavior of the system.

Polarity Classifier: For Polarity classification, all the object expressions are removed. The **Training model** consists of all the subjective datasets which are re-labeled. The expressions are further classified into 3 categories - namely, positive, neutral, and negative.
The complete data set consists of 24939 tweet expressions with around 15000 objective expressions and around 9000 subjective expressions. out of subjective expressions around 5000 are positive, around 3000 are negative and around 400 are neutral expressions.

Classified Outputs: The classified segments are given as outputs. The abusive content is returned with a negative message.

RESULTS

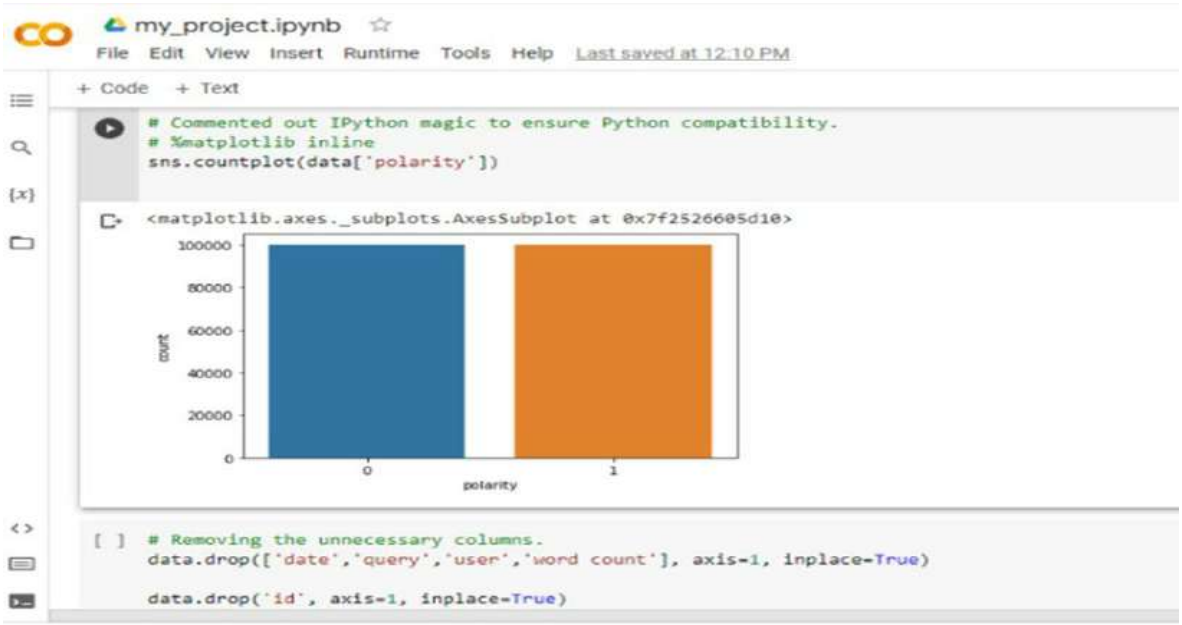


IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

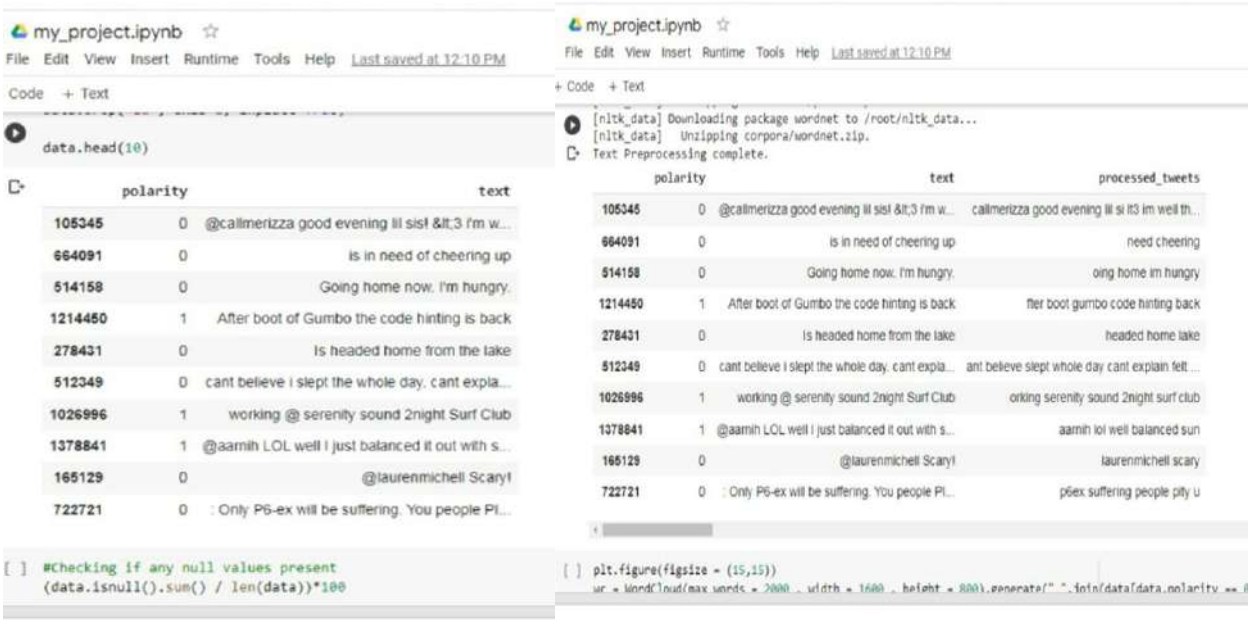
International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 6, June 2022

DOI: 10.17148/IJARCCE.2022.116104



Outputs of the analysis done using matplotlib.



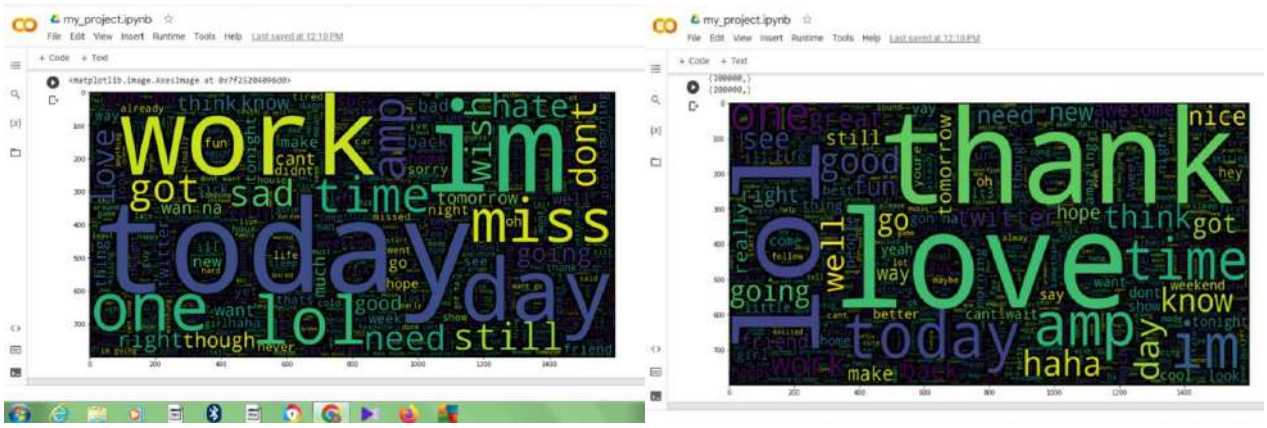
Polarities of the dataset

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

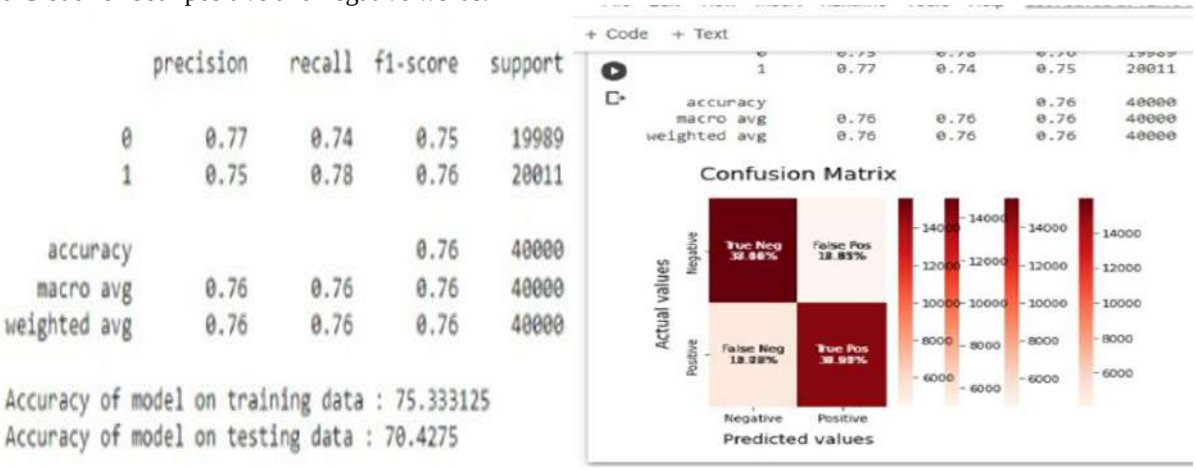
International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 6, June 2022

DOI: 10.17148/IJARCCE.2022.116104



Word Cloud for both positive and negative words.



Accuracy of the model by SVM algorithm

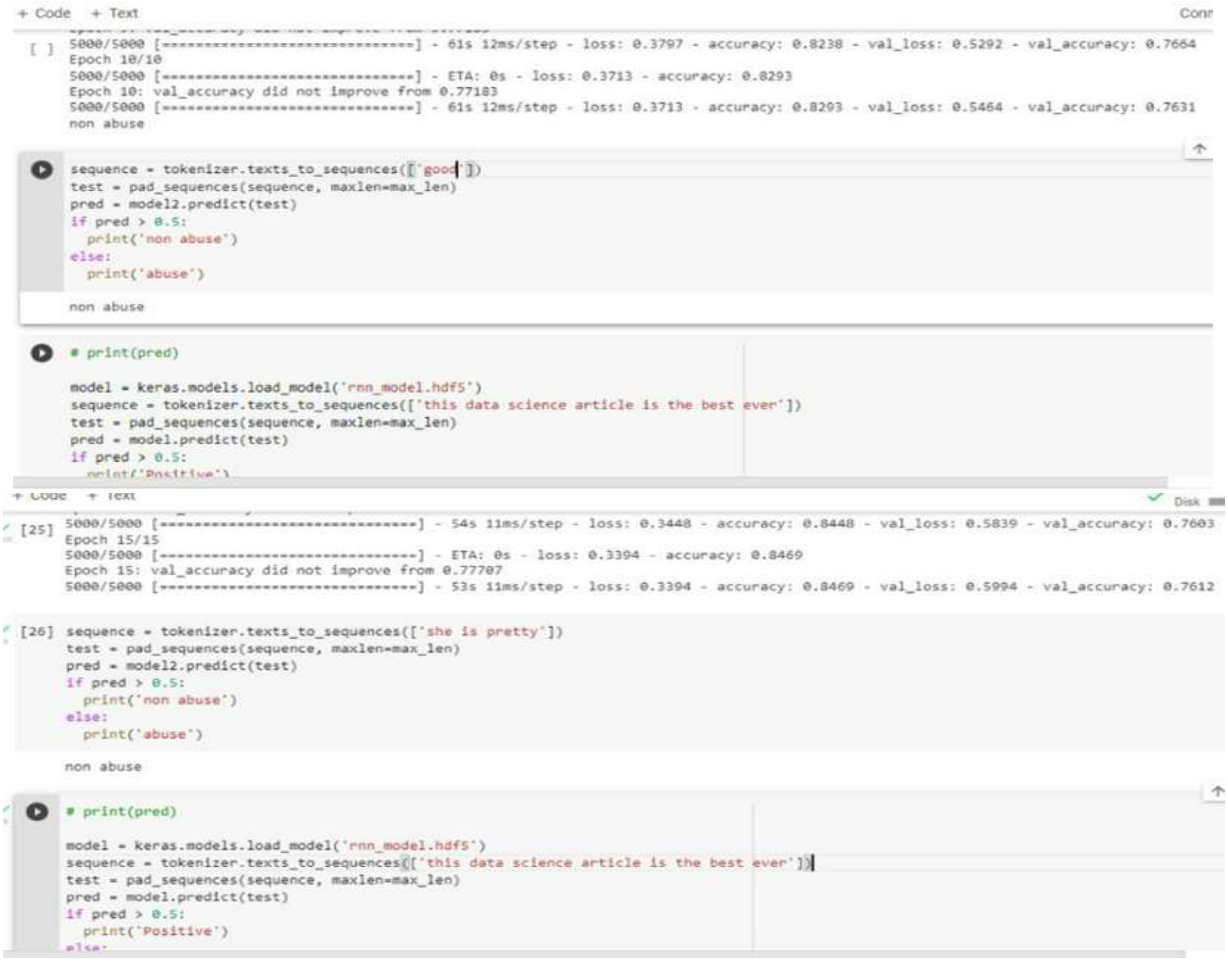
IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 6, June 2022

DOI: 10.17148/IJARCCE.2022.116104

The training of the model is done and the epoch is 15.



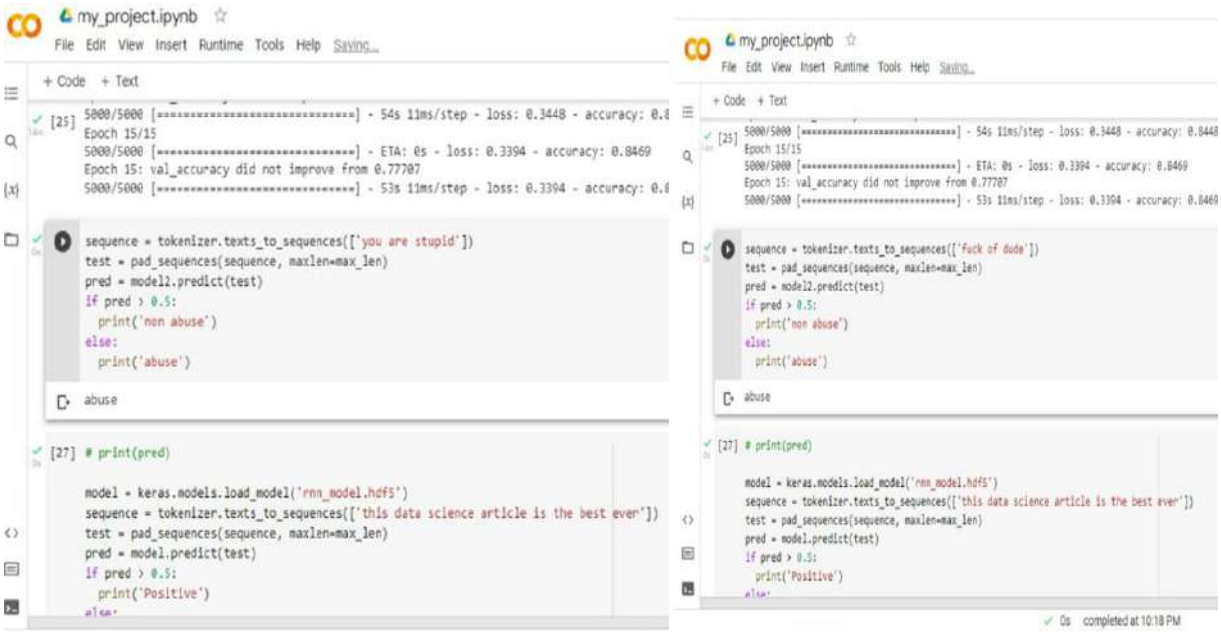
“Good”, “She’s pretty” are appreciations, so the output is non-abuse.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 6, June 2022

DOI: 10.17148/IJARCCE.2022.116104



"Fuck off dude", "You're stupid" are abusive, so the output is abuse

CONCLUSION AND FUTURE SCOPE

Abusive content detection problem is more complicated than it seems due to its unseemly, unstructured noisy data and unpredictable context. The learning performance of neural networks attracts researchers to get the highest performing output. Still there are some limitations to noisy data while training for a neural network. In our work, we have proposed an approach that considers the assets of both machine learning and neural network to get the most optimum result. Our approach performs with an F1 score of 93. In the future, this can be implemented on social media sites and block the abusers. This way cyber crimes also can be reduced.

REFERENCES

[1] Z. Fa, G. -G. Geng, Z. -W. Yan and X. -D. Lee, "A robust internet abuse detection method," 2017 IEEE International Conference on Big Data (Big Data), 2017, pp. 1712-1715, DOI: 10.1109/BigData.2017.8258113.

[2] C. Rawat, A. Sarkar, S. Singh, R. Alvarado, and L. Rasberry, "Automatic Detection of Online Abuse and Analysis of Problematic Users in Wikipedia," 2019 Systems and Information Engineering Design Symposium (SIEDS), 2019, pp. 1-6, DOI: 10.1109/SIEDS.2019.8735592.

[3] M. Susanty, Sahrul, A. F. Rahman, M. D. Normansyah and A. Irawan, "Offensive Language Detection using Artificial Neural Network," 2019 International Conference of Artificial Intelligence and Information Technology (ICAIIT), 2019, pp. 350-353, DOI: 10.1109/ICAIIT.2019.8834452.

[4] https://doi.org/10.13053/cys-24-3-3476

[5]Human aggressiveness and reactions towards uncertain decisions F Bashir, N Ashraf, A Yaqoob, A Rafiq, RU Mustafa

[6]Breiman, L. Bagging predictors. Mach Learn 24, 123–140 (1996). https://doi.org/10.1007/BF00058655

[7] https://doi.org/10.1145/3274301

[8]https://www.pewresearch.org/internet/2017/07/11/online-harassment-2017/

[9] L. Khan, A. Amjad, N. Ashraf, H. -T. Chang and A. Gelbukh, "Urdu Sentiment Analysis With Deep Learning Methods," in IEEE Access, vol. 9, pp. 97803-97812, 2021, doi: 10.1109/ACCESS.2021.3093078.