



FEDERATED-BASED APPROACH TO TWITTER SENTIMENT ANALYSIS

Vibhor Bhatt¹, Yogendra Sharma², Tisha Pareek³

Btech student¹, IT Department, JECRC college, jaipur,India

ABSTRACT

In this study, there is an investigation on how federated learning can be used to protect data privacy in Twitter sentiment analysis and detect hate speech in tweets. The drawbacks of conventional, centralized learning approaches suggest a federated learning strategy to resolve these problems. Here, it is also a case study that illustrates how the suggested approach works to protect user privacy while maintaining the precision of sentiment analysis.

Keywords -Data Analysis, Text Mining, Machine Learning Algorithms, NLP

[A] INTRODUCTION

The common procedure for analysing the data is the centralized method of machine learning. In this technique the data of the clients from there devices is transferred to the centralized data centre using the communication medium and then the data is processed using machine learning to create a generic model and this new set of model is than applied to the users devices but in this the data of the user is no more private the sensitive data of the user might get leaked form the centralized data centre or in between the communication ,and this type of models also increase the latency and many more issues like bandwidth is very often limited and so this type of issue can be solved by using a machine learning technique known as federated learning[1].

The sentiment analysis is based on the NLP (natural language processing) .The ability of a machine to comprehend human language is known as NLP, it is part of AI .The machine learning models used in the sentiment analysis are text analysis and mining, data analysis and visualization[2].

Here is the idea which will go over the foundational aspects of federated learning and how they relate to data privacy in sentimental analysis, what are the aspects of the federated learning, federated averaging, model training and aggregation and the steps of implementing it on the

twitter sentiment analysis. It will also go over the difficulties and potential solutions in putting Federated Learning for sentiment analysis on Twitter into practice.

[B] RELATED WORK

Sentiment Analysis on Twitter:

The practice of figuring out the sentiment or feeling behind a piece of writing is called sentiment analysis. Businesses and scholars can use sentiment analysis to better understand how people feel about certain issues, products, or services on Twitter. Basically it's an analysers which is used to analyse the sentiment of the person who tweets by reading his/her tweets .It identifies whether the tweet is positive, neutral or negative and generates the output according.

Data Privacy Concerns:

In order to apply centralized learning methods, user data must be gathered and stored centrally. This might result in privacy violations and unauthorized access to private data. In the centralized data system all the data might get erased or stolen but whereas in the decentralized data system if the data is erased or stolen from the one system it will be available in the other servers which can be retrieved from it .

Federated Learning:

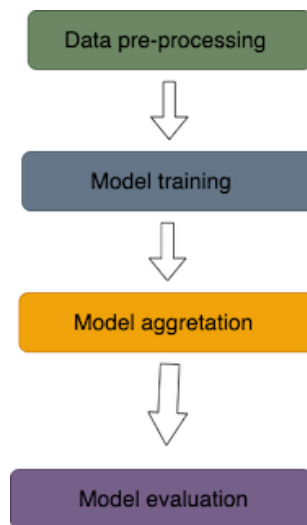
Recurrent neural networks (RNNs), such as the Long Short-Term Memory (LSTM), can acquire effective feature representations of sequences when provided with sufficient training data [4 - 6].

[C] Analysis Federated Learning for Twitter Sentiment Methodology:

Following are the steps of the Federated-based approach to twitter sentiment analysis:

- Data pre-processing: On each user's device, clean and tokenize tweets.
- Model training: Using the pre-processed data on each device, develop a sentiment analysis model.
- Model aggregation: Create a global model by combining the models from all devices.

- Model evaluation: Assess the effectiveness of the global model using a test dataset.



[5] Case Analysis:

To assess how well the federated learning technique works in protecting data privacy while keeping the precision of sentiment analysis. The findings demonstrated that this strategy Produced accuracy on par with conventional centralized solutions while protecting user privacy. The goal will be to identify hate speech in tweets. So we define hate speech as any tweet that incorporates a racist or sexist emotion. So to distinguish distinctive (sexist or supremacist) tweets from ordinary tweets. And to complete this challenge, divide it has that node '0' indicates that a tweet is not supremacist or sexist and node '1' indicates that it is.

Model Training and Aggregation:

Train local models on user devices using their respective data without sharing raw data. Update the global model through the aggregation of local model gradients or parameters. Implement techniques such as federated averaging or secure aggregation to ensure accurate model aggregation.

Framework for Federated Learning:

Create local models for the sentiment analysis task that were learned on specific devices or regional servers. Utilize a Federated Learning framework to combine the regional models while maintaining the decentralized nature of the data. To preserve privacy during model updates, use encryption and secure communication methods.

Federated Learning Framework:

Divide the sentiment analysis task into local models trained on individual devices or local servers. Utilize a Federated Learning framework to aggregate the local models while keeping

data decentralized. Employ encryption and secure communication protocols to ensure privacy during model updates.

Performance Assessment:

Measure the Federated Learning-based sentiment analysis approach's precision, effectiveness, and protection of privacy. To demonstrate the benefits of the federated learning technique, compare the outcomes with those of conventional centralized methods.

Gathering and Pre-processing of Data:

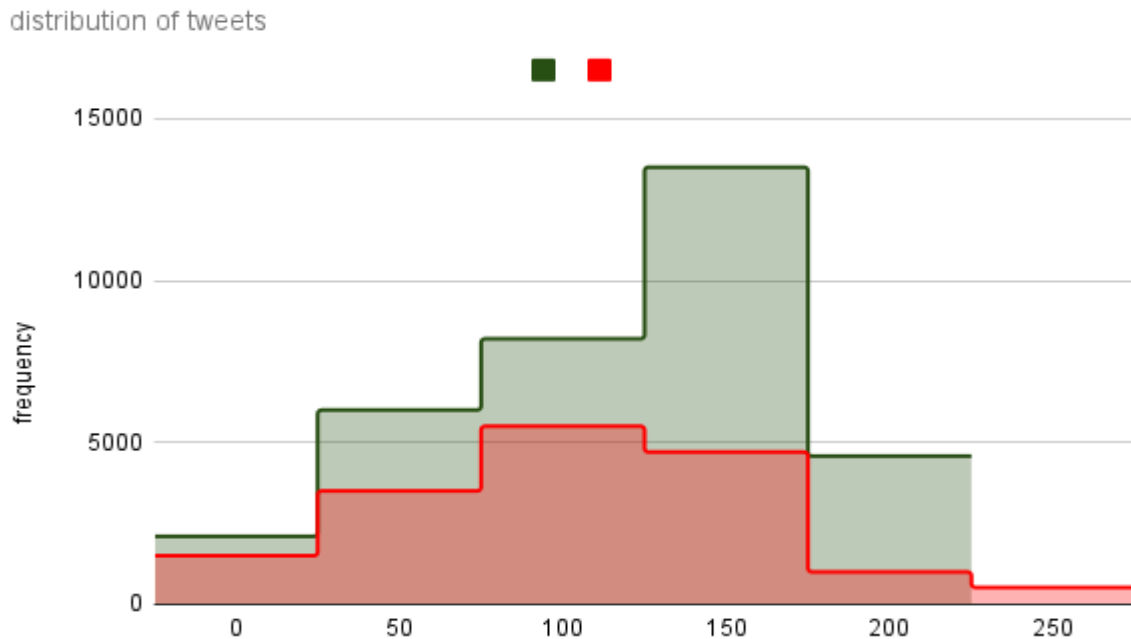
Implement a user-friendly distributed data collection process. Use pre-processing procedures to handle noise, eliminate extraneous data, and standardize the format.

Table: sample data for the sentimental analysis [7]

	ID	LABEL	TWEET
0	1	0	@user my mother is dysfunctional and is s...
1	2	0	@user @user thanks for #lyft credit i can't use.....
2	3	0	bihday your majesty
3	4	0	#model i love u take with u all the time in...
4	5	0	factsguide: society now #motivation
5	6	0	[2/2] huge fan fare and big talking before the...
6	7	0	@user camping tomorrow @user @user @user @use.
7	8	0	the next school year is the year or exams.Ÿđ....
8	9	0	we won!!!love the land!!! #allin #cavs #champ...
9	10	0	@user @user welcome here ! i'm it's so #gr...

Twitter sentiment analysis's flaw:

Federated learning can also aid in addressing the imbalanced data problem in sentiment analysis. The dataset used to train the model in conventional centralized learning techniques is



frequently skewed towards a specific user or demographic group. This could produce erroneous sentiment analysis findings and confirm current prejudices.

As each user's device contributes its own data, federated learning, on the other hand, enables a more varied dataset to be used in model training. As a result, the results of sentiment analysis may be more accurate, and bias may be lessened. [8]

In our project, we examined a Kaggle dataset CSV file that had 15,000 tweets. 90% of tweets, or 13,500 tweets, in the sample contained no hate labels, while 10%, or 1500 tweets, contained hate labels. As a result, the dataset was highly skewed.

- The number of hateful and non-hateful tweets needed to be balanced as the initial stage in developing our model.
- Run the following five classification algorithms: Gradient Boosting, Logistic Regression, Naive Bayes, Decision Tree, and Random Forest. After completing the task, these five algorithms are run repeatedly.
- Next, we employ NLP and networks (RNNs) such the Long Short-Term Memory (LSTM) for the pre-processing stage.
- Finally the working file contains the code for model building and visualisations.

Fig 1. Distribution of tweets

Federated learning can also help sentiment analysis models become more scalable. In order to store massive data sets using traditional centralized learning approaches, datasets. Federated learning, on the other hand, is more scalable and effective because the computational load is spread across several machines.[9]Overall, the use of federated learning in Twitter sentiment analysis presents a viable answer to the issues with data privacy raised by conventional centralised learning techniques. Businesses and researchers may protect user privacy while still learning important things from social media data by using a federated learning approach.

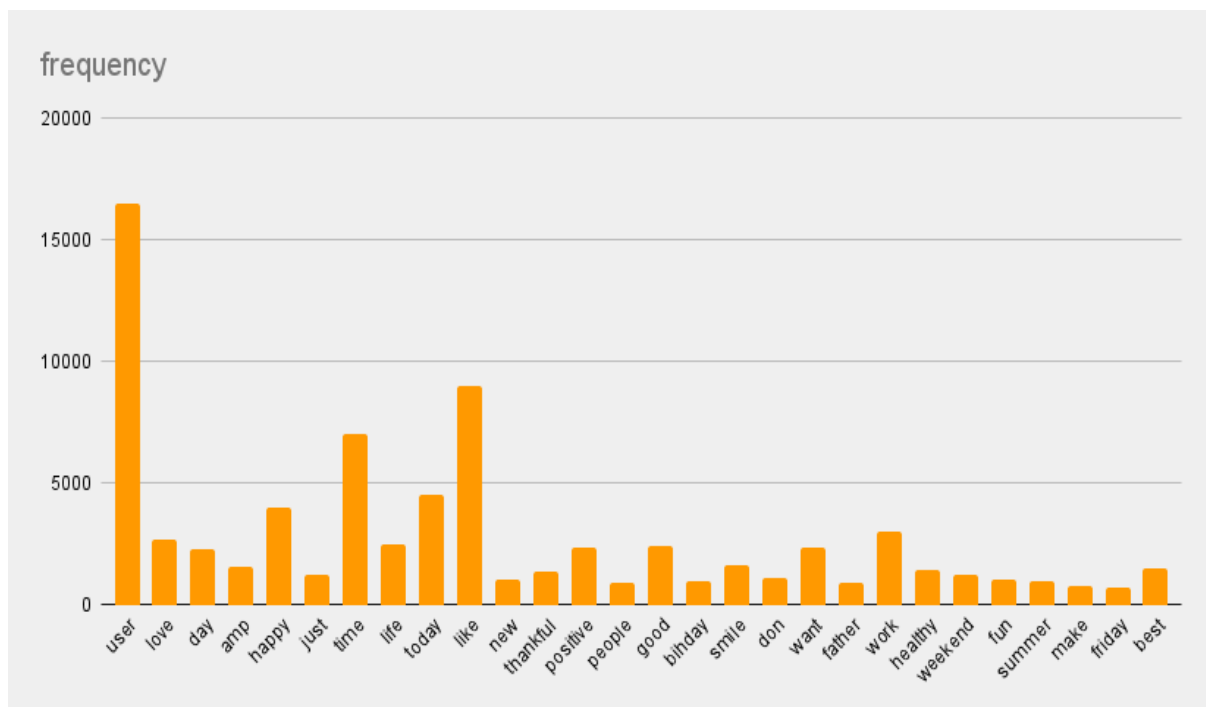


Fig 2: Most frequently occurring words

[E] SUMMARY

Comprehension of trends and public opinion requires a thorough comprehension of Twitter sentiment. The typical centralized approach, however, poses data security and privacy issues. A Federated Learning-based strategy has come to light as a potential resolution to these difficulties. This case study investigates how sentiment analysis on Twitter data might be applied while maintaining data privacy. Beginning with data collecting and pre-processing,

The study adopts a dispersed methodology to protect user privacy. The sentiment analysis task is then split into local models trained on specific devices or servers using the Federated Learning framework. This enables decentralized data processing and guarantees the confidentiality and security of data.

Training local models on user devices without transferring raw data is part of the model aggregation and training process. The global model is then updated using aggregation methods like secure aggregation or federated averaging. To safeguard sensitive user data, privacy preservation mechanisms including client-side encryption and differential privacy are put in place. The Federated Learning-based approach's performance evaluation shows precise sentiment analysis while upholding user privacy. Comparisons with established centralized methods show the benefits. Of Federated Learning in terms of accuracy, efficiency, and privacy preservation. This method has ramifications for many other fields, such as social media analytics, market research, and public opinion studies. Organizations can benefit from insightful sentiment analysis by utilizing federated learning without jeopardizing user privacy. In the case study, the significance of privacy-preserving methods in sentiment analysis is emphasized, and more study in this area is encouraged. The Federated Learning-based strategy, which prioritizes data security and privacy, provides an efficient remedy for sentiment analysis on Twitter data. Accurate sentiment analysis can be done while upholding user privacy by dispersing the learning process and combining local models. This case study highlights Federated Learning's potential for sentiment analysis on Twitter as well as its broader implications for data privacy in social media analysis.

REFERENCES

- [1] Konečný, Jakub; McMahan, Brendan; Ramage, Daniel (2015). "Federated Optimization: Distributed Optimization Beyond the Datacenter". ArXiv: 1511.03575
- [2] Goldberg, Yoav (2016). "A Primer on Neural Network Models for Natural Language Processing". *Journal of Artificial Intelligence Research*. **57**: 345–420. ArXiv: 1807.10854. doi:10.1613/jair.4992. S2CID 8273530.
- [3] Dr T Subbulakshmi, S. Mercy Shalinie "Real Time Classification and Clustering Of IDS Alerts Using Machine Learning Algorithms", International Journal of Artificial Intelligence & Applications (IJAIA), Vol. 1 **change**, No.1, January 2010
- [4] Abiodun, Oludare Isaac; Jantan, Aman; Omolara, Abiodun Esther; Dada, Kemi Victoria; Mohamed, Nachaat Abdelatif; Arshad, Humaira (2018-11-01). "State-of-the-art in artificial neural network applications: A survey". *Heliyon*. **4** (11): e00938. doi:10.1016/j.heliyon.2018.e00938. ISSN 2405-8440. PMC 6260436. PMID 30519653
- [5] Li, Xiangang; Wu, Xihong (2014-10-15). "Constructing Long Short-Term Memory based Deep Recurrent Neural Networks for Large Vocabulary Speech Recognition". arXiv:1410.4281
- [6] Mayer, H.; Gomez, F.; Wierstra, D.; Nagy, I.; Knoll, A.; Schmidhuber, J. (October 2006). *A System for Robotic Heart Surgery that Learns to Tie Knots Using Recurrent Neural Networks*. 2006 *IEEE/RSJ International Conference on Intelligent Robots and Systems*. pp. 543–548. CiteSeerX 10.1.1.218.3399. doi:10.1109/IROS.2006.282190. ISBN 978-1-4244-0258-8. S2CID 12284900.
- [7] <https://www.kaggle.com/datasets/arkhoshghalb/twitter-sentiment-analysis-hatred-speech> [online]
- [8] ncbi.nlm.nih.gov/pmc/articles/PMC6701757/ [Online]
- [9] <https://www.dailyprincetonian.com/article/2018/10/hate-speech-deserves-a-second-look>

- [10]. Xiong LIU and Haiqing LIU, (2021), “Data Publication Based On Differential Privacy In V2G Network” Int. J. of Electronics Engineering and Applications, Vol. 9, No. 2, pp. 34-44, DOI 10.30696/IJEEA.IX.I.2021.45-53.
- [11]. Mandava Siva Sai Vighnesh, MD Shakir Alam and Vinita.S, (2021), “Leaf Diseases Detection and Medication” Int. J. of Electronics Engineering and Applications, Vol. 9, No. 1, pp. 01-07, doi 10.30696/IJEEA.IX.I.2021.01-07
- [12]. Pradeep M, Ragul K and Varalakshmi K,(2021), “Voice and Gesture Based Home Automation System” Int. J. of Electronics Engineering and Applications, Vol. 9, No. 1, pp. 08-18, doi 10.30696/IJEEA.IX.I.2021.08-18
- [13]. Jagan K, Parthiban E Manikandan B,(2021), “Engrossment of Streaming Data with Agglomeration of Data in Ant Colony” Int. J. of Electronics Engineering and Applications, Vol. 9, No. 1, pp. 19-27, doi 10.30696/IJEEA.IX.I.2021.19-27
- [14]. M. Khadar, V. Ranjith, K Varalakshmi (2021), “Iot Integrated Forest Fire Detection and Prediction using NodeMCU” Int. J. of Electronics Engineering and Applications, Vol. 9, No. 1, pp. 28—35, doi 10.30696/IJEEA.IX.I.2021.28-35
- [15]. Gayathri. M, Poorviga. A and Mr. Vasantha Raja S.S, (2021), “Prediction Of Breast Cancer Stages Using Machine Learning” Int. J. of Electronics Engineering and Applications, Vol. 7, No. 1, pp. 36-42, doi 10.30696/IJEEA.IX.I.2021.36-42
- [16]. Karthikeyen, N. Ramya, M. Sai Priya and C. Yuvalakshmi, (2021), “Novel Method Of Real Time Fire Detection And Video Alerting System Using Open-CV Techniques” Int. J. of Electronics Engineering and Applications, Vol. 9, No. 1, pp. 43-50, doi 10.30696/IJEEA.IX.I.2021.43-50
- [17]. L.Prinslin, M.A.Srenivasan and R.Naveen (2021), “Secure Online Transaction With User Authentication” Int. J. of Electronics Engineering and Applications, Vol. 9, No. 1, pp. 51-57, doi 10.30696/IJEEA.IX.I.2021.51-57
- [18]. S Lokewar, A Hemaranjane and V. Narayane (2021), “Edge Based Ecosystem For Internet Of Things (EBEFIOT)” Int. J. of Electronics Engineering and Applications, Vol. 9, No. 1, pp. 58-67, doi 10.30696/IJEEA.IX.I.2021.58-67
- [19]. Prof. K. Phani Srinivas and Dr. P. S. Aithal, (2000). "Practical Oriented Analysis On The Signal Processing Using FFT Algorithm", Int. J. of Electronics Engineering and Applications, Vol. 8, Issue II, July-Dec. 2020. pp 01-10, doi 10.30696/IJEEA.VIII.II.2020.01-10
- [20]. Onintra Poobrasert, Sirilak Luxsameevanich, Sarinya Chompoobutr, Natcha Satsutthi, Sakda Phaykrew and Paweena Meekanon, (2000), “Heuristic-based Usability Evaluation on Mobile Application for Reading Disability “, Int. J. of Electronics Engineering and Applications, Vol. 8, Issue II, July- Dec. 2020, PP- 11-21, doi 10.30696/IJEEA.VIII.II.2020.11-21
- [21]. Rajeev Ranjan Kumar and S. P. Singh, (2020), “Variation Of Capacitive Reactance Of Coupled Microstrip Line Structure With Width Of The Similar Metal Strips” Int. J. of Electronics Engineering and Applications, Vol. 8, No. 2, pp. 22-28, DOI- 10.30696/IJEEA.VIII.II.2020.22.28
- [22]. Sunita Swain and Rajesh Kumar Tiwari, (2020), “Cloud Security Research- A Comprehensive Survey” Int. J. of Electronics Engineering and Applications, Vol. 8, No. 2, pp. 29-39, DOI- 10.30696/IJEEA.VIII.II.2020.29.39
- [23]. Ritesh Kumar Thakur, Rajesh Kumar Tiwari (2020), ‘Security On IoT: A Review’, Int. J. of Electronics Engineering and Applications, Vol. 8, No.2, July-Dec 2020, pp-40-48, DOI- 10.30696/IJEEA.VIII.II.2020.40.48.
- [24]. Rajesh Kumar Tiwari (2020). ‘Human age estimation Using Machine Learning Techniques’, International Journal of Electronics Engineering and Applications, Vol. 8, No. 1, pp.01-09, DOI- 10.30696/IJEEA.VIII.I.2020.01-09

[25]. Chinmayee Sahoo. 'CLOUD COMPUTING AND ITS SECURITY MEASURES', International Journal of Electronics Engineering and Applications, Volume 8, Issue I, Jan-June 2020, pp 10-19, doi 10.30696/IJEEA.VIII.I.2020.10-19