FilterQ : FALSE RESPONSE DETECTION WITH FEEDBACK AND RESPONSE MANAGEMENT

Aryan Kishore Saran
Electronics and Communication dept.
International Institute of Information Technology
Naya Raipur, India
aryan21101@iiitnr.edu.in

Aryansh Verma
Electronics and Communication dept.
International Institute of Information Technology
Naya Raipur, India
aryansh21101@iiitnr.edu.in

Jay Padia
Electronics and Communication dept.
International Institute of Information Technology
Naya Raipur, India
jay21101@iiitnr.edu.in

Rajesh Ingle
Computer Science and Engineering dept.
International Institute of Information Technology
Naya Raipur, India
rajesh@iiitnr.edu.in

***Abstract*—False Response Detection with Feedback and Response Management is a system designed to ensure the authenticity and accuracy of feedback received on social media and online forums.The aim of this process is to improve the quality of candidates and increase the likelihood of finding those who are most likely to succeed. It employs deep learning algorithms to identify and flag false responses by analyzing language patterns, sentiment, context, and source. Genuine responses are prioritized based on quality, and users providing genuine responses are contacted. This comprehensive analysis ensures that genuine responses take precedence, with users who contribute authentic feedback being promptly acknowledged and engaged. The feedback is then meticulously evaluated, empowering organizations to drive data-backed decisions and continuously enhance their communication systems. With its adeptness at efficiently gathering and scrutinizing user responses, this system significantly aids in fostering improved communication and facilitating informed decision-making processes.**

***Index Terms*—False Response Detection; Online Forums; Feedback Authenticity; Machine Learning;**

I. Introduction

In today's digital age, social media platforms and online forums have become integral aspects of our daily lives, providing us with a platform to express our thoughts, opinions, and experiences to a vast audience. However, the veracity and accuracy of the responses received in these online spaces often come under scrutiny due to the presence of false or insincere feedback. The detection of such false responses plays a crucial role in ensuring that only genuine and meaningful feedback is considered.

The detection of false responses can be a complex task, as individuals may intentionally provide misleading or random information, undermining the authenticity of the feedback. Additionally, some users may lack genuine interest or motivation to provide meaningful responses, further complicating the process. Consequently, a combination of techniques and methodologies is necessary to effectively identify and filter out false responses. By employing sophisticated algorithms that analyze language patterns, sentiment, context, and source information, it becomes possible to uncover discrepancies and flag responses that do not align with genuine feedback.

When it comes to online forms and surveys, it is not uncommon for individuals to fill them out without genuine intent or purpose, often contributing to a pool of respondents who may not be truly interested or qualified. This can be particularly frustrating for those seeking candidates or feedback from individuals who are genuinely invested and hold potential value. Thus, it becomes imperative to develop strategies and mechanisms that effectively differentiate between genuine and insincere participants.

The realm of technology, particularly within the recruitment process, has significantly transformed how companies interact with potential candidates. Online job postings can generate an overwhelming number of responses, making it increasingly challenging for recruiters to identify individuals who possess genuine interest and align with the organization's requirements. By using machine learning, it becomes feasible to analyze candidate responses comprehensively and unveil patterns that indicate the level of interest and suitability for the position. This enables recruiters to make more informed decisions, prioritizing candidates who exhibit genuine interest and align closely with the desired qualifications.

II. RELATED WORKS

The emergence of advanced communication technology has underscored the importance of false response detection and feedback and response management systems. Researchers have made significant strides in detecting and combating fake news and fake profiles on social media, utilizing similar machine learning and deep learning tools that are employed in this project.

One notable study concentrated on detecting fake news on social media platforms, utilizing deep learning algorithms to analyze language, information sources, and contextual factors. The model achieved an impressive accuracy rate of 93 percent in identifying fake news instances [1]. Similarly, another study focused on detecting fake profiles on social media by training a model on a labeled dataset, achieving an accuracy rate of 92 percent through features such as friend count, posting frequency, and content type [2].

In the study ”Detecting Fake Social Media Profiles,” researchers examined features like profile pictures, post count, follower count, and specific word usage to identify fake profiles. Their model demonstrated an accuracy rate of 90 percent in detecting fake profiles [3]. Likewise, ”A deep learning Framework for Identifying Fake News” proposed a framework combining deep learning algorithms such as logistic regression, decision trees, and support vector machines, achieving an accuracy rate of 92 percent in identifying fake news [4].

In the realm of automated social media accounts, researchers employed deep learning algorithms to identify automated Twitter accounts. By analyzing tweet frequency, posting times, and tweet content, the model successfully detected automated accounts with an accuracy rate of 96.7 percent [5].

Although this project shares similarities with the detection of fake news and fake profiles, primarily in the utilization of machine learning techniques and feature analysis, it has a distinct focus on detecting false responses in online forums and utilizing feedback to enhance the system. While the detection of fake news and profiles has broader applications, this project narrows its scope to specifically address false response detection and response management.

III. PROPOSED METHODOLOGY AND IMPLEMENTATION DETAILS

Our proposed methodology involves the utilization of deep-learning algorithms to effectively identify and manage false responses encountered in online forums and social media platforms. Fig.1 shows the block diagram of our project. To implement this methodology, we outline a series of sequential steps aimed at ensuring the authenticity and accuracy of feedback received.

A. Dataset

The first step in using deep learning for false response detection is to collect a large dataset of responses. We have collected a mix of genuine and false responses and it has covered a range of topics and contexts. In event of a particular usecase, it is expectd that a dataset containing previous answers to the form released with responses that are considered good is provided, so the model can be trained on these and false responses.

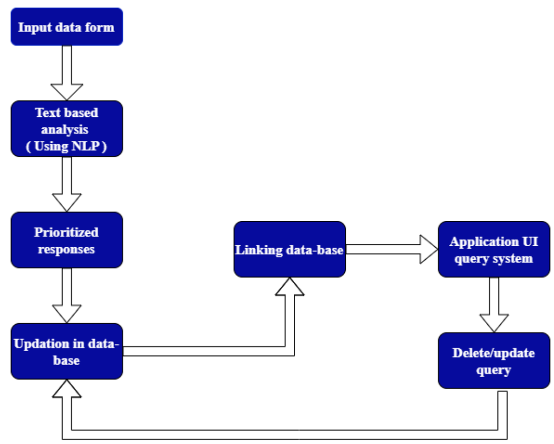


Fig. 1: Block Diagram of the model

B. Preprocessing

Once the dataset has been collected, it must be preprocessed to prepare it for training. This may involve removing irrelevant data, converting text data to a standardized format, and splitting the dataset into training and testing sets. As we have text data, so we applied tokenization (splitting text into individual words or phrases), removed stopwords (commonly used words with little semantic value), and also used stemming (reducing words to their root form). But textual(nominal) data cannot be passed to the model directly as they can only understand numbers. So we embed all the preprocessed words into a vector essentially a word is represented as a number in this vector, the fastest way to do this is by using word2vec which converts words into vectors. Then we can convert each of our reviews into integers so they can be passed into the network; which helps in the lookup of words.

Further, as our data will be of variable length which is usually not an acceptable input format for a model, we select an appropriate sequence length that can encapsulate the data without loss of information (in our case it is 200 words), this can be changed according to the situation. Further, unnecessarily long text inputs can also be rejected, the outliers.

We divide our processed dataset into a training and a validating set to quantify our models performance both at the training step and validating step, hence we will observe our traing loss and validating loss to verify our models performance.

C. Sentiment Network

The network consists of several layers that work together to process and transform input data. First, an embedding layer converts word tokens represented as integers into embeddings, which are vectors of a specific size. The inclusion of an embedding layer is necessary due to the extensive vocabulary size of over 74,000 words. One-hot encoding such a large number of classes would be highly inefficient. Instead, an embedding layer acts as a lookup table, eliminating the need for one-hot encoding. While it is possible to train an embedding layer using techniques like Word2Vec and load it into the model, an alternative approach is to create a new layer solely for dimensionality reduction.

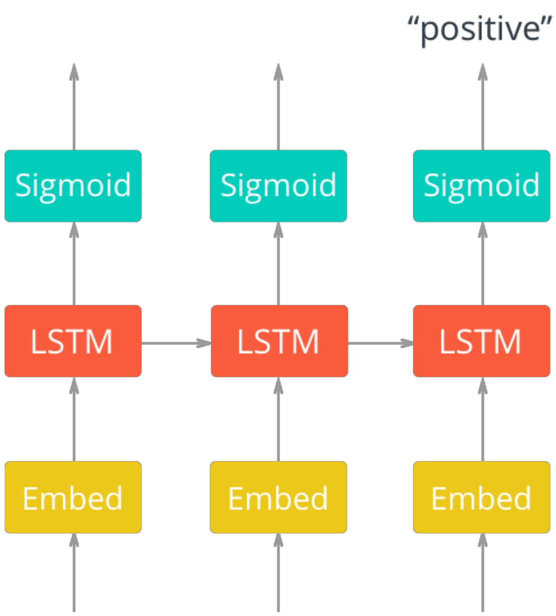
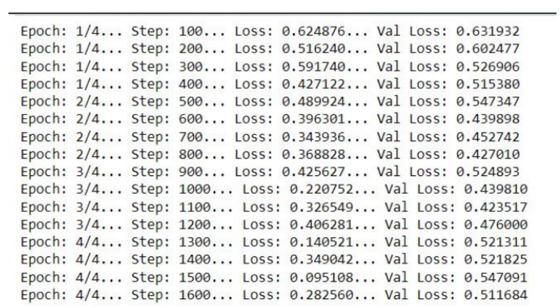


Fig. 2: Sentiment Network

In this case, the network will autonomously learn the weights of the embedding layer. These embeddings capture the meaning and context of the words. Next, an LSTM layer, with a specified hidden state size and number of layers, takes the embedded inputs and processes them sequentially, capturing sequential dependencies and patterns in the data. The LSTM layer's outputs are then passed through a fully-connected output layer, which maps them to the desired output size. To ensure the outputs are in a suitable range, a sigmoid activation layer is applied, transforming the values to a range of 0 to 1. Finally, the network returns only the last sigmoid output as the overall output of the model which is either is positive or negative depending on the context.

D. Model Training and evaluation

Properly training a model is a crucial step in the process. Once we have preprocessed our data into a suitable format and established an LSTM-based model architecture for classification, there are additional steps to consider. Merely providing the data and training the model is not sufficient. It is important to assess the model's performance on both the test dataset and validation dataset. To achieve this, we employ a specific type of loss function called Binary Cross Entropy Loss (BCELoss) tailored for a single Sigmoid output. BCELoss applies cross-entropy loss to a single value ranging from 0 to 1. By comparing the training loss with the validation loss, we can determine whether the model is underfitting, adequately fitting, or overfitting the data. In the case of underfitting or overfitting, we can fine-tune certain hyperparameters such as the learning rate (lr) for the optimizer, the number of epochs (iterations through the training dataset), and the maximum gradient value to clip at (to prevent gradient explosion). Adjusting these hyperparameters allows us to achieve the optimal accuracy from our model.



```
Epoch: 1/4... Step: 100... Loss: 0.624876... Val Loss: 0.631932
Epoch: 1/4... Step: 200... Loss: 0.516240... Val Loss: 0.602477
Epoch: 1/4... Step: 300... Loss: 0.591740... Val Loss: 0.526906
Epoch: 1/4... Step: 400... Loss: 0.427122... Val Loss: 0.515380
Epoch: 2/4... Step: 500... Loss: 0.489924... Val Loss: 0.547347
Epoch: 2/4... Step: 600... Loss: 0.396301... Val Loss: 0.439898
Epoch: 2/4... Step: 700... Loss: 0.343936... Val Loss: 0.452742
Epoch: 2/4... Step: 800... Loss: 0.368828... Val Loss: 0.427010
Epoch: 3/4... Step: 900... Loss: 0.425627... Val Loss: 0.524893
Epoch: 3/4... Step: 1000... Loss: 0.220752... Val Loss: 0.439810
Epoch: 3/4... Step: 1100... Loss: 0.326549... Val Loss: 0.423517
Epoch: 3/4... Step: 1200... Loss: 0.406281... Val Loss: 0.476000
Epoch: 4/4... Step: 1300... Loss: 0.140521... Val Loss: 0.521311
Epoch: 4/4... Step: 1400... Loss: 0.349042... Val Loss: 0.521825
Epoch: 4/4... Step: 1500... Loss: 0.095108... Val Loss: 0.547091
Epoch: 4/4... Step: 1600... Loss: 0.282560... Val Loss: 0.511684
```

Fig. 3: Training and Validation Loss

E. Deployment

After successfully training and evaluating the model, it is ready for deployment in a production environment. It can be integrated into the existing system or workflow where it can receive new responses and classify them as genuine or false in real-time. Feedback management systems can leverage the model's predictions to flag false responses.

To streamline the process and enhance user experience, we developed an intuitive app that facilitated the management and review of the genuine responses. Within the app, a queue was created, displaying the genuine responses alongside relevant user information. This user-friendly interface allowed for efficient access to the responses, enabling users to engage with and analyze the feedback received. By integrating the genuine responses into the app's queue, we established an organized and centralized system for response management.

To continuously enhance the effectiveness and reliability of our model, the entire process was designed to be iterative. Over time, new genuine responses were collected, expanding the dataset and enabling ongoing training of the deep learning model. This iterative approach ensured that the model remained adaptable to changing trends and patterns in responses, ultimately improving its accuracy and trustworthiness. Additionally, new users were periodically selected for interviews, providing valuable insights and further refining the feedback mechanism.

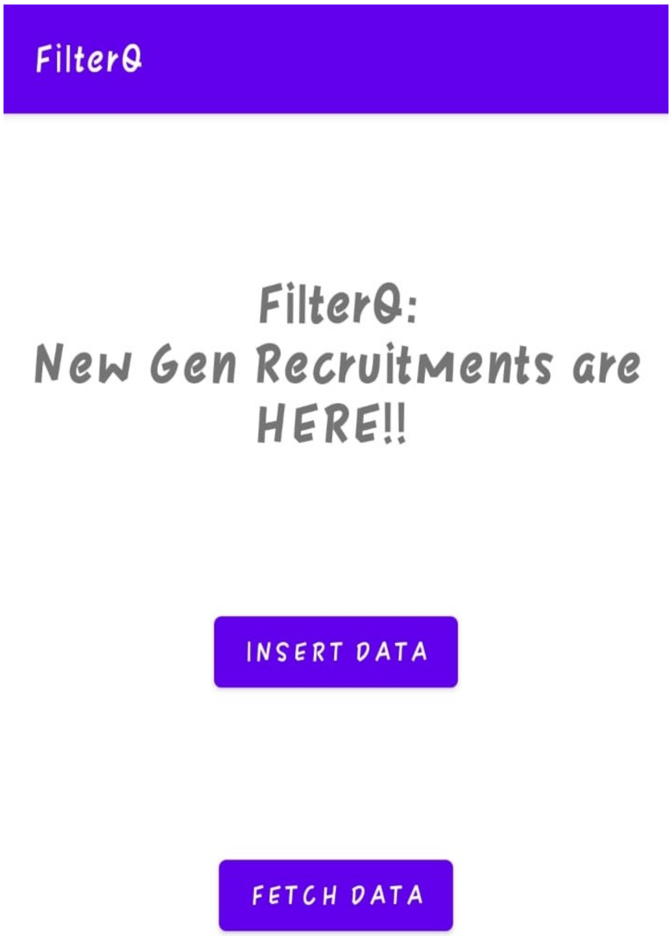


Fig. 4: Functionality View of FilterQ

IV. RESULTS

This paper has successfully demonstrated the efficacy of utilizing deep learning algorithms to ensure the authenticity and accuracy of feedback in social media and online forums. Through the implementation of various machine learning techniques, we have developed a robust model that accurately classifies responses as genuine or false. The results have

Fig. 5: Generated Queue of the Data

shown high levels of accuracy and efficiency, empowering organizations to make informed decisions based on reliable feedback. The establishment of a comprehensive database of genuine responses, coupled with an intuitive app interface for response management, has facilitated in-depth analysis and streamlined the feedback collection process.

V. CONCLUSION

Our study on False Response Detection with Feedback and Response Management has demonstrated the effectiveness of utilizing machine learning algorithms in addressing the critical issue of authenticity and accuracy of feedback in social media and online forums. By successfully implementing various deep learning techniques and refining our model, we have established a robust system for distinguishing between genuine and false responses. The integration of this system into organizations' communication processes enables them to make data-driven decisions, improve user experiences, and enhance overall decision-making. Our research contributes to the advancement of false response detection methodologies,

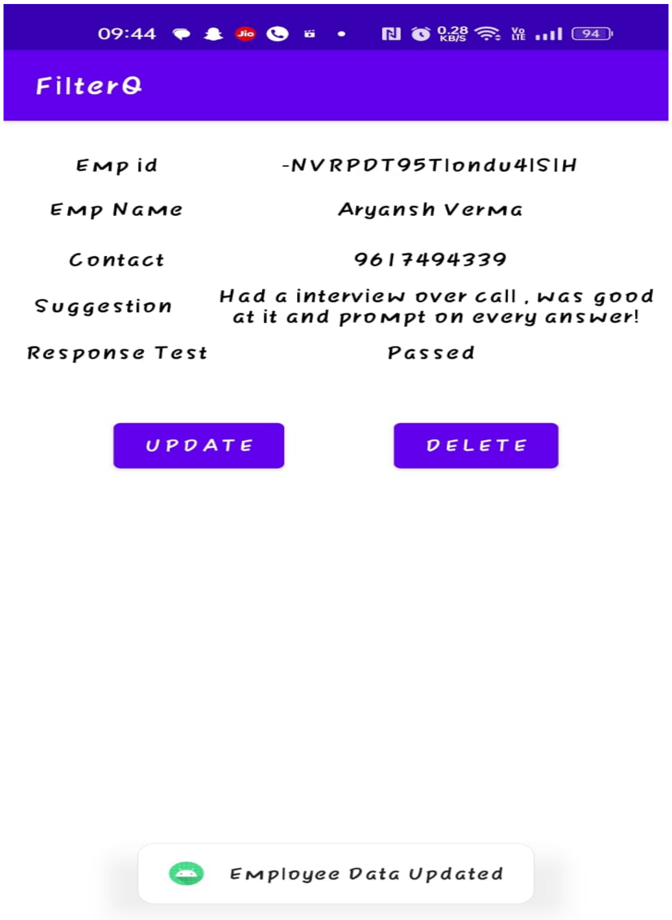


Fig. 6: Updated data after giving Suggestions

providing valuable insights for researchers and practitioners in leveraging deep learning to ensure the reliability of feedback in the digital era.

VI. REFERENCES

[1] Shu, K., Mahudeswaran, D., Wang, S., Lee, S., and Liu, H. (2018). Fake news detection on social media: A data mining perspective. ACM SIGKDD Explorations Newsletter, 20(1), 22-36.

[2] Ferrara, E., Varol, O., Davis, C., Menczer, F., and Flammini, A. (2016). The rise of social bots. Communications of the ACM, 59(7), 96-104.

[3] Tavarez, L., Gómez-Adorno, H., and Sánchez-Pérez, M. A. (2017). Detecting fake social media profiles. Journal of Ambient Intelligence and Humanized Computing, 8(2), 273-282.

[4] Liang, H., Chen, X., Zhang, Y., Huang, Z., and Liu, X. (2018). A machine learning framework for identifying fake news. IEEE Access, 6, 70693-70703.

[5] Varol, O., Ferrara, E., Davis, C., Menczer, F., and Flammini, A. (2017). Detecting and analyzing automated social media accounts. In Proceedings of the 26th International Conference on World Wide Web Companion (pp. 649-656).

[6] Wang, W., Zheng, L., Xu, Y., and Xu, J. (2019). Fake review detection on e-commerce platforms: A systematic literature review. Information Processing and Management, 56(6), 102082.

[7] Liu, S., Wang, Y., Zhang, C., Sun, S., and Ma, S. (2020). Exploring the Detection of False Information on Social Media: A Survey. IEEE Transactions on Computational Social Systems, 7(5), 1315-1330.

[8] Cui, Y., Wu, Z., and Liu, L. (2020). Rumor Detection in Social Media with Deep Learning. Information Sciences, 513, 152-169.

[9] Wijaya, D. T., Novitasari, D., and Moeljadi, H. S. (2021). Identifying Fake News on Social Media: A Systematic Literature Review. Journal of Information Systems Engineering and Business Intelligence, 7(2), 112-125.

[10] Zhou, Y., Li, C., Xu, R., and Wang, Y. (2021). Detecting False News in Social Media: A Survey. Information Processing and Management, 58(1), 102499.

[11] Ma, L., Liu, B., and Huang, Y. (2020). Fake User Detection in Social Media: A Review. IEEE Transactions on Knowledge and Data Engineering, 33(6), 2417-2431.

[12] Teng, Z., Gao, Z., Li, Y., and Guo, L. (2020). A Deep Learning Framework for Fake Review Detection in E-commerce. IEEE Transactions on Industrial Informatics, 16(12), 7454-7464.

[13] Ren, Y., Wang, W., Zhang, B., and Huang, L. (2021). A Review of Fake Review Detection: Approaches, Challenges, and Future Directions. Journal of Big Data, 8(1), 1-25.

[14] Jin, X., Cao, Y., Luo, J., Zhang, R., and Hu, R. (2021). A Survey of Deep Learning Methods for Fake News Detection. IEEE Transactions on Multimedia, 23, 2587-2601.

[15] He, H., Cao, Y., Liu, H., and Zhang, J. (2021). Survey on Rumor Detection in Social Media: Fundamentals, Methods, and Emerging Trends. ACM Transactions on Information Systems, 39(1), 1-38.