

#107

BICA*AI
2023
BICA VPS 2023

Unraveling the Elements of Effective Altruistic Appeals through Machine Learning and Natural Language Processing

Processing *Sourav Yadav¹, Sankalp Arora^{*2}, Akash Kumar, Kaveri Verma.*

¹syadav12@hawk.iit.edu ²sankalparora5@gmail.com.



First Author



Second Author



SUMMARY

- In today's digital era, online platforms enable altruistic actions without expecting reciprocation.
- This study presents a novel model combining sparse text analysis with rich features from prior research, emphasizing the significance of the first model's probability and the request's comment count in outcome prediction.

APPROACH

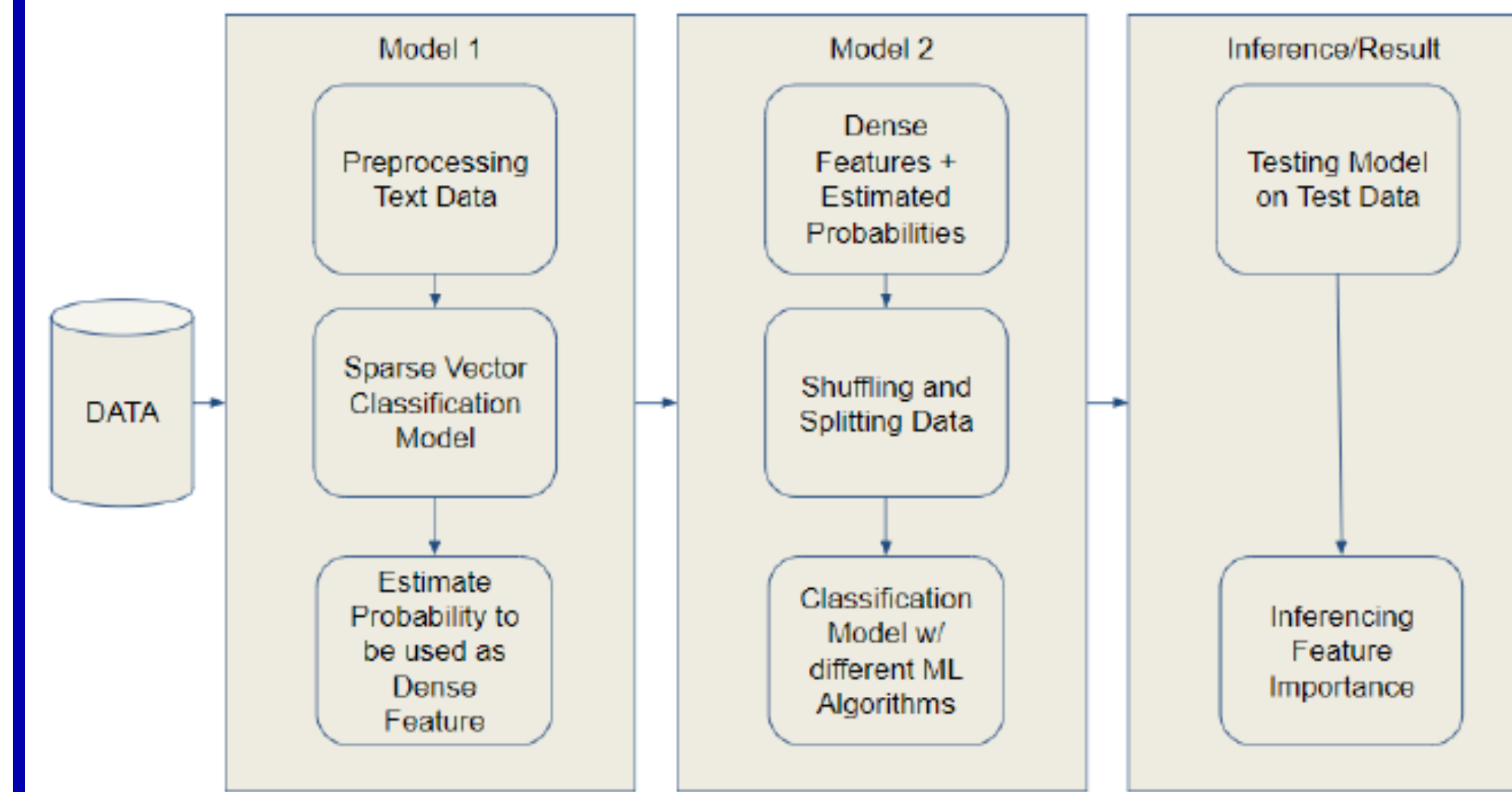


Fig 1. Model Approach

RESULTS

- To assess performance and gauge feature significance, the dataset of 5671 entries was split into 4537 training rows and 1134 testing rows.
- Out of the three machine learning algorithms used, Random Forest performs equally well for each text embedding technique, while the performance under Logistic Regression improves with the use of a more complex and advanced text embedding technique
- To evaluate the final model's performance, the test set is transformed into vectors using different embedding methods and combined with numeric data.
- The model's accuracy remains highest at around 75 when we use Random Forest to model the final data regardless of the embedding method used.

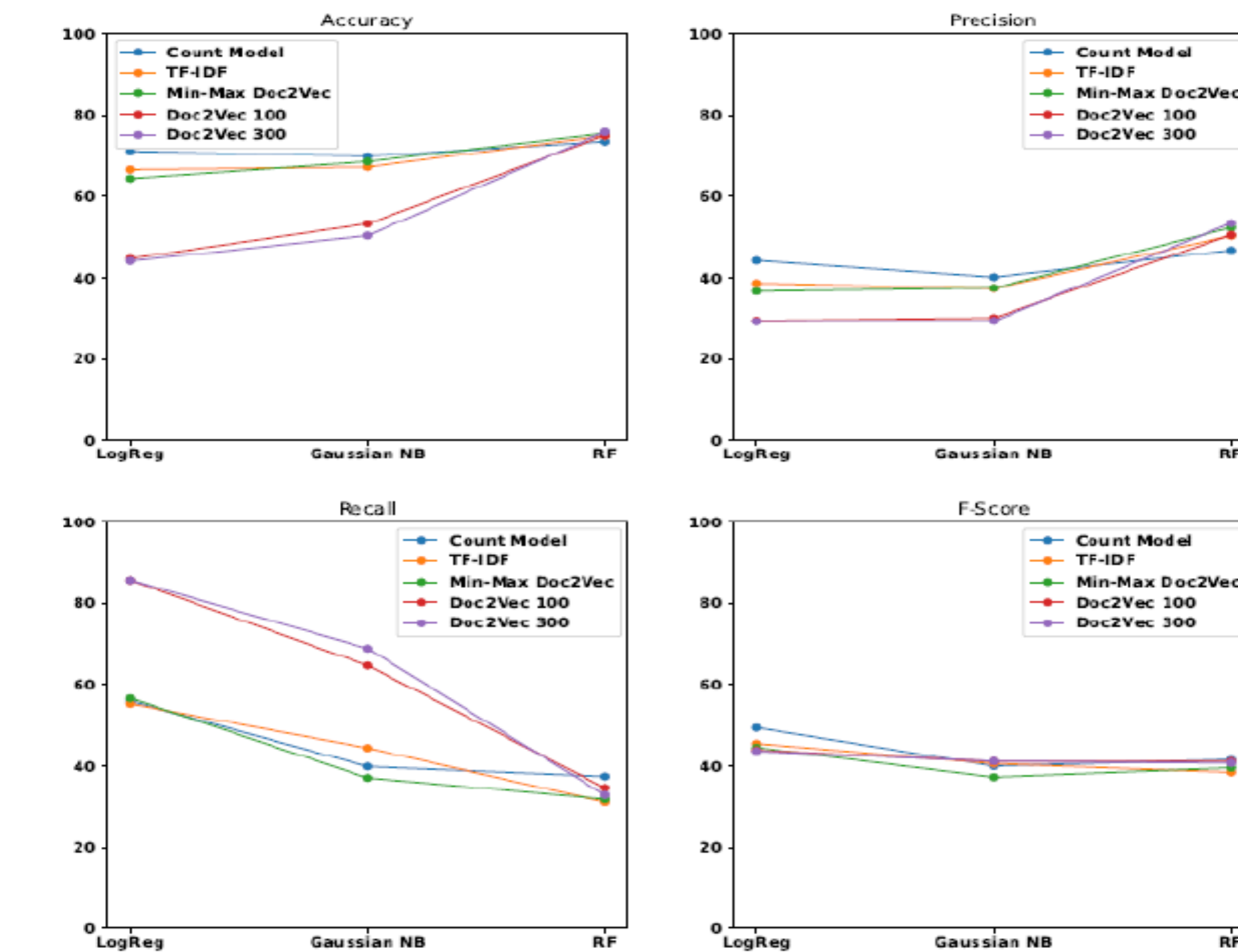


Fig 2. Performance of the Final Model in terms of Accuracy, Precision, Recall & F-Score

DISCUSSION

- The probability estimate derived from the text plays a crucial role in predicting whether a requester will receive pizza or not.
- Additionally, factors such as the number of comments on the request post, upvotes minus downvotes, and text length contribute to prediction.
- However, whether the requester has posted in the subreddit before or included an evidence link has minimal impact, as indicated by their low SHAP values in the models.

INTRODUCTION

- With the popularity of social media and Natural Language Processing in a social setting, the rise of philanthropic communities on the internet is notable.
- Lately, there has been a surge in use cases within the field of Computational Social Science that demand the utilization of advanced NLP tools for their resolution.
- The project proposes an architecture of a two-stack binary classification model to predict the success of an altruistic request.
- The dataset contains 5671 requests collected from the Reddit community between December 8, 2010, and September 29, 2013.

METHODS

- The project proposes an architecture of a two-stack binary classification model to predict the success of an altruistic request.
- The first model focuses on textual data, employing feature extraction techniques to transform raw text into vectors. It then classifies these vectors against the target variable.
- The finished model is then used to estimate the probabilities of the request being successful by the text.
- The probability estimate is incorporated as a dense feature alongside other factors such as account age of the requester, the number of likes and dislikes, and more.

ANALYSIS

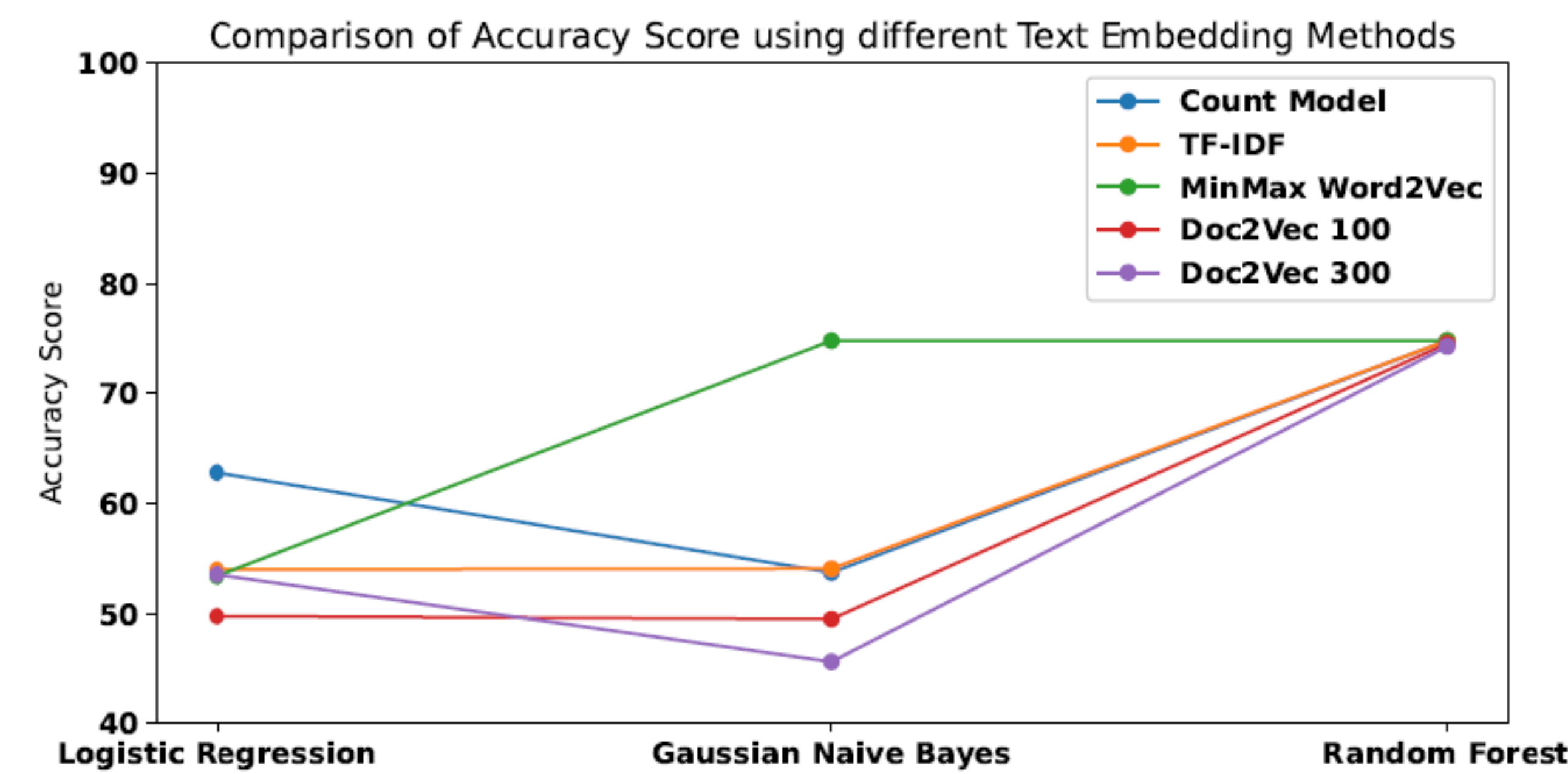


Fig 3. Comparison of Accuracy Score using different Text Embedding Methods

CONCLUSIONS

- This work introduces a two-model approach that handles text and numeric data separately. It enables the conversion of high-dimensional text vectors into a probability estimate while preserving the importance of the original dataset's features.
- In summary, our model can forecast whether a requester's post will trigger altruistic responses from fellow Redditors.
- Furthermore, this study has the potential for extension through the application of advanced machine learning and deep learning models and the inclusion of additional data to enhance model robustness.

ACKNOWLEDGMENTS

- Althoff, T., Salehi, N., Nguyen, T. (2013). Random Acts of Pizza : Success Factors of Online Requests
- Breiman, L. Random Forests. Machine Learning 45, 5–32 (2001)
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). "Efficient Estimation of Word Representations in Vector Space."