

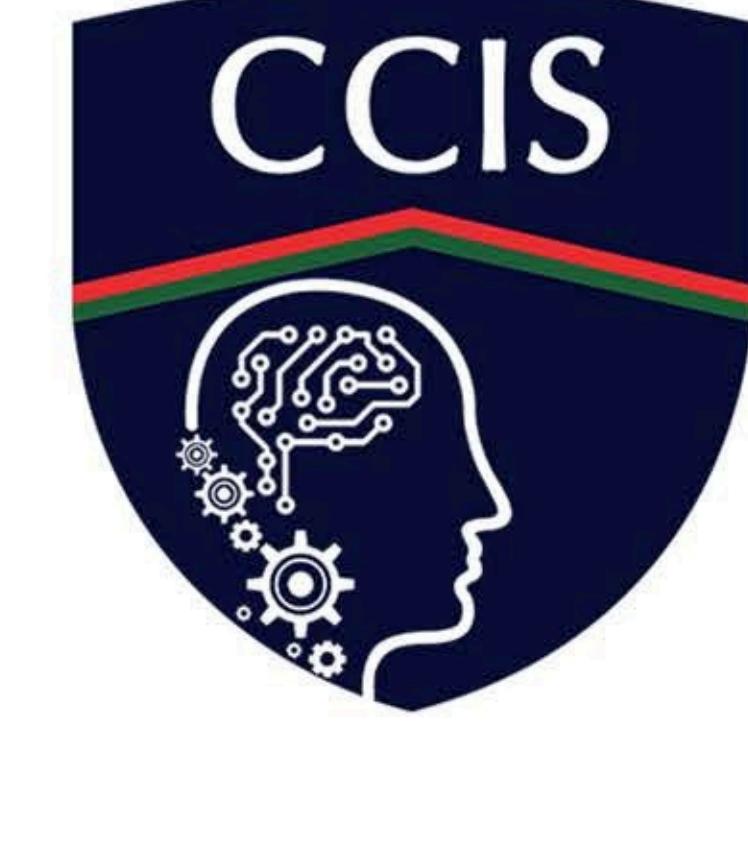
ClickSaverPH: Clickbait Detection using MobileBERT and BiGRU for Philippine News Headlines

Name of Author(s): Ang, Matthew Szenel S.¹, Abid, James Paul N.², and Tan, Andre Deuce B.³

¹ Author#1: MMCM, msAng@mcm.edu.ph

² Author#2: MMCM, jpAbid@mcm.edu.ph

³ Author#3: MMCM, adTan@mcm.edu.ph



ABSTRACT

The trend of journalism known as “clickbait” has drastically affected the credibility of online journalism, often exaggerating or misrepresenting the content of their articles or videos, with little in the way of actual substance. This paper aimed to determine the viability of using a pre-trained MobileBERT model and the BiGRU model in the training of a model capable of accurately detecting clickbait. To achieve this, a new dataset of clickbait headlines as well as a web application was developed. The research found that the model can detect clickbait headlines with an accuracy of 97.88% which affirms the validity of MobileBERT and BiGRU in the use of future models.

Keywords: Clickbait detection, Philippines, Filipino news, headlines, bilingual, English, MobileBERT, Bidirectional Gated Recurrent Unit (BiGRU), web application, clickbait dataset

[1] INTRODUCTION

Within the increasing digitalization of journalism, however, came a noticeable trend of tabloidization in news articles; often characterized by exaggerated or misleading headlines with little to no substance within the article proper.

This subset of sensationalized news articles, known as clickbait, has been used with increasing frequency by less credible news sources as a means of enticing readers to view the article, thereby increasing online traffic on the hosted website which in turn generates additional advertisement revenue (Ahmad et al., 2020; Chen et al., 2015).

In the case of the Philippines, the effects of allowing such misinformation to be easily accessible should not be understated. It is reported in 2023 according to Google's advertising statistics that an equivalent of 72.5% of the total population in the Philippines uses social media, where clickbait media is most often shared (Kemp, 2023).

Thus, there is a need for the development of tools which may be used in allowing internet users to better identify clickbait on the internet; in which this study will be verifying the viability of the MobileBERT and BiGRU algorithms in the development of such tools.

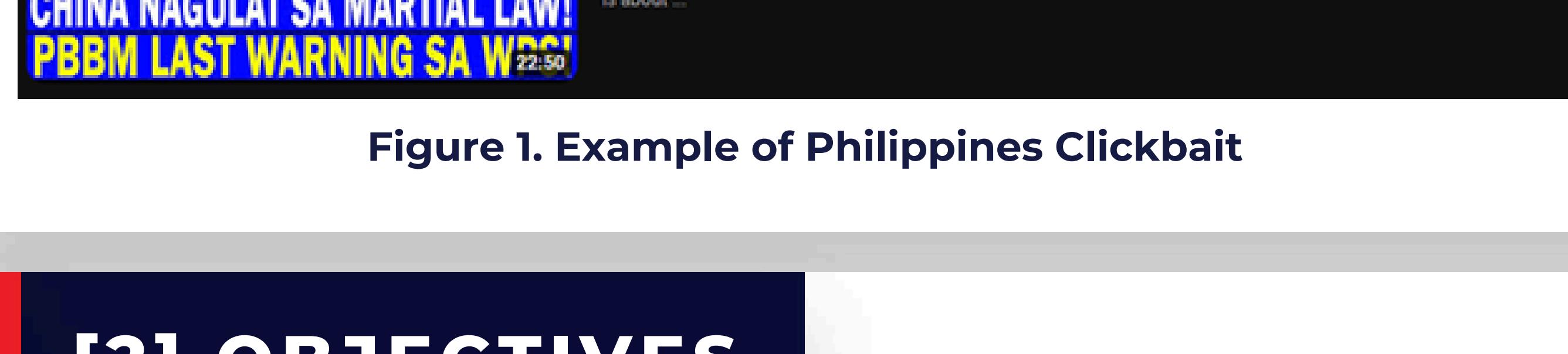


Figure 1. Example of Philippines Clickbait

[4.1] Model and Application Results

Table 1. Test Results of the MobileBERT-BiGRU model

Accuracy	Precision	Recall	F1 Score
97.88%	0.9848	0.9725	0.9786

Table 1 shows the accuracy, precision, recall, and F1 score of the model after testing it. The model was tested on 2,400 rows of data that was not used in training nor validation.

Table 2. Usability Testing Survey Results

Characteristic	Sub-characteristic	Average Rating	Interpretation
Functionality	Suitability (F1)	4.57	SA
	Accuracy (F2)	4.04	A
Usability	Understandability (U1)	4.83	SA
	Learnability (U2)	4.78	SA
	Operability (U3)	4.78	SA
	Attractiveness (U4)	4.52	SA
Efficiency	Time Behaviour (E1)	4.83	SA

Table 2 displays the results of the survey conducted on the usability of the web application with a total of thirty (30) respondents. Survey results show that, overall, the web application was met with positive response. Respondents found that the web application could effectively perform its designated functions, and that it was easy to operate the application.



Figure 1. Index Page

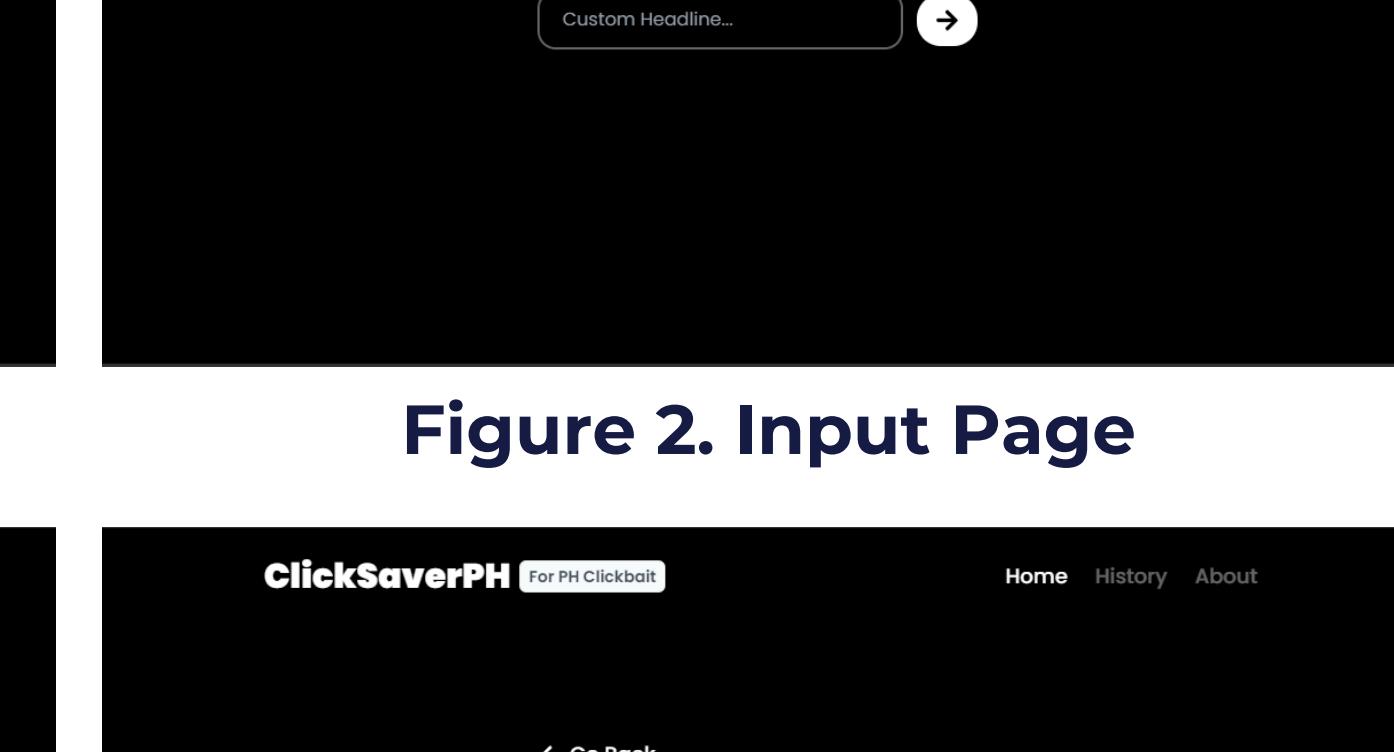


Figure 2. Input Page



Figure 3. Clickbait Result

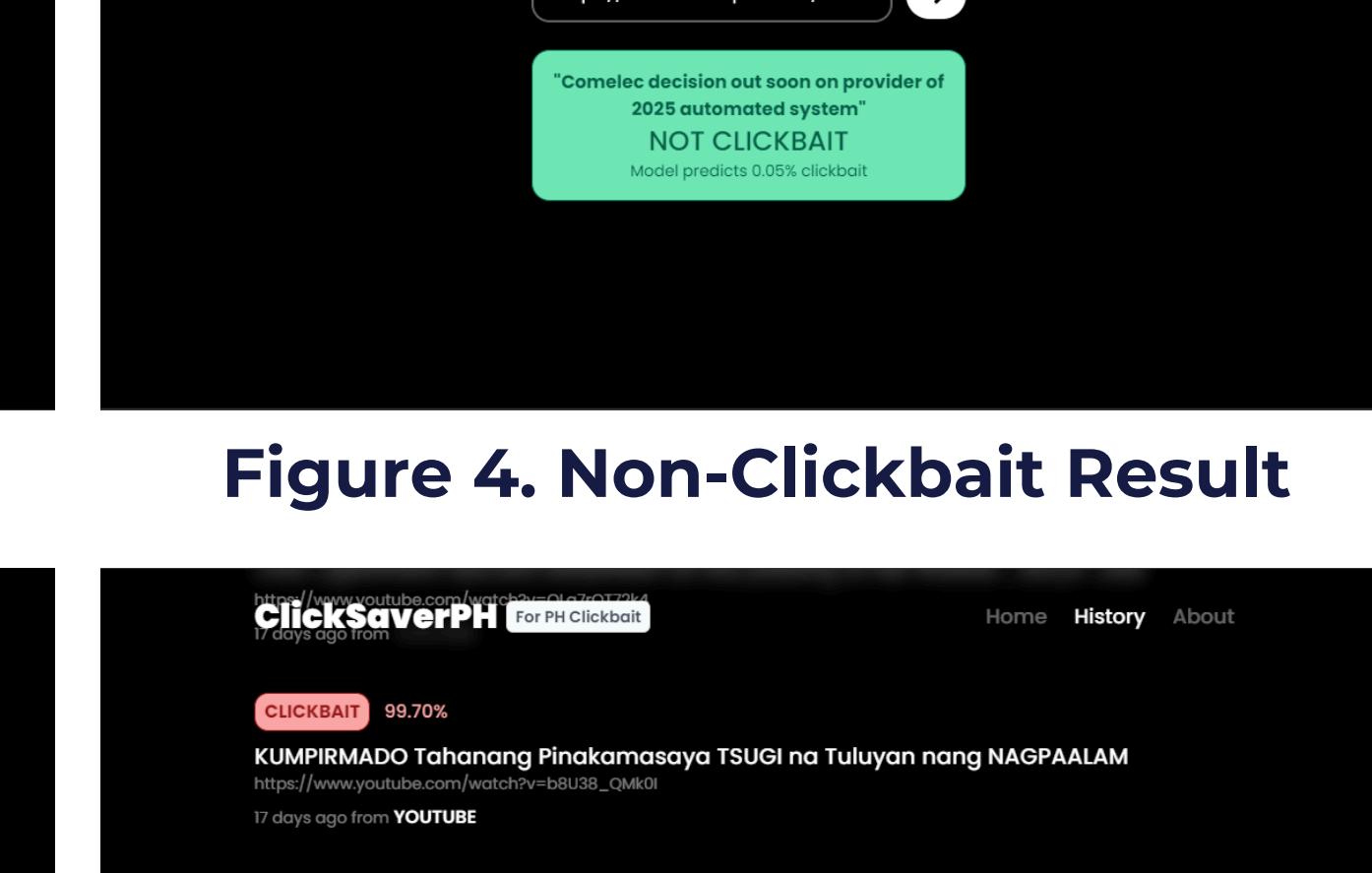


Figure 4. Non-Clickbait Result



Figure 5. About Page

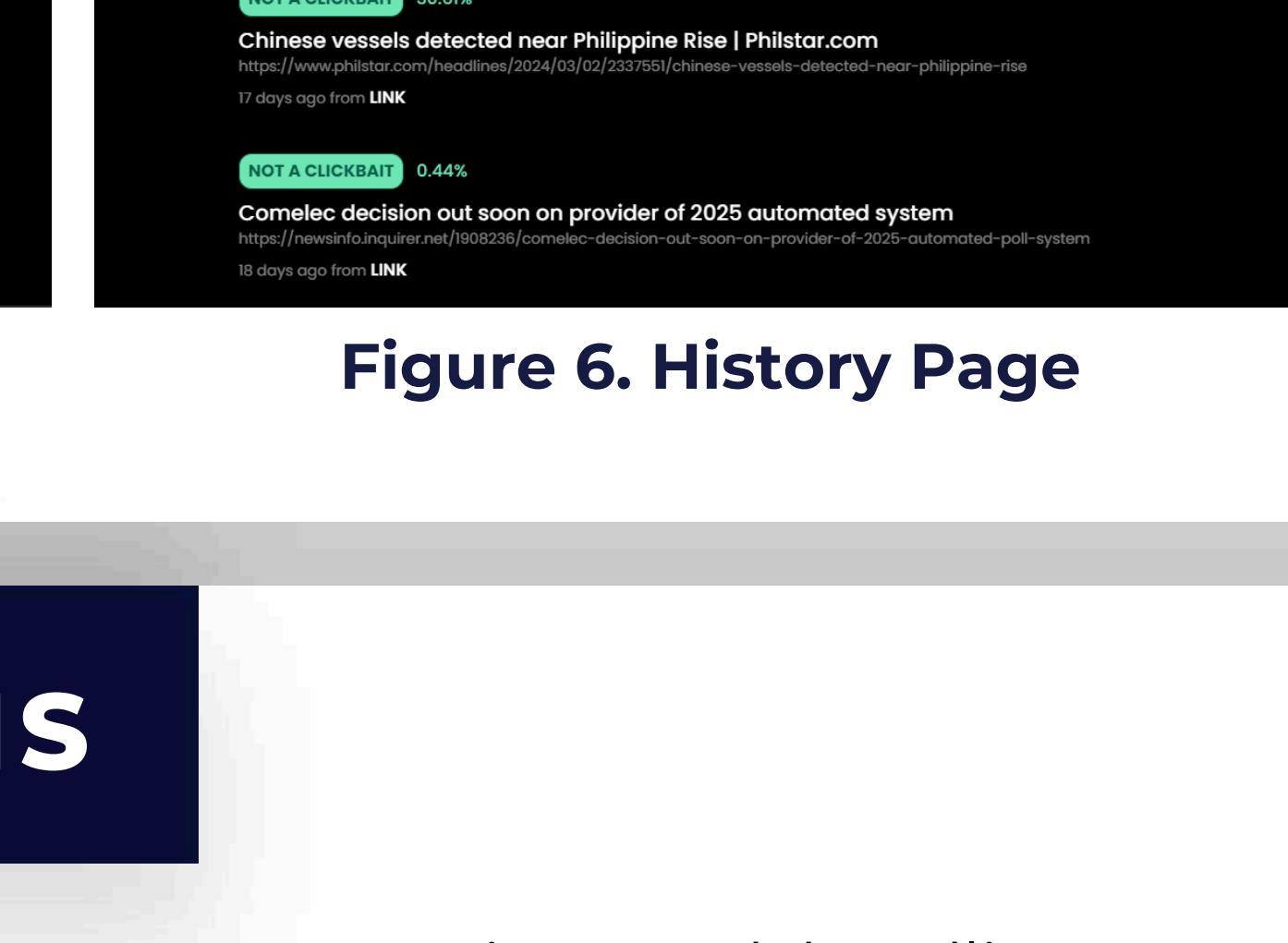


Figure 6. History Page

[5] CONCLUSIONS

While the model was generally accurate, some inputted headlines were incorrectly classified during the testing. These were generally headlines relating to international news and were not within the initial scope of the application. Furthermore, certain words were found to be incorrectly weighted towards a clickbait result such as “Russia” due to a bias caused by the political climate and general trends in clickbait news in the Philippines. It is recommended that larger datasets be used in training future models to amend these issues in future studies.

[6] ACKNOWLEDGEMENTS

We extend our deepest gratitude to several people who were pivotal to the success of this study: The institution, Mapua Malayan Colleges Mindanao, Dr. Jeanie R. Delos Arcos, Sir Neil P. Magloyuan, Dr. Patrick D. Cerna, Ma'am Genevieve A. Pilongo, Sir Martzel P. Baste, Dr. Rhodessa J. Cascaro, Mr. Ian Ray Garcia, Ms. Cristina E. Alivio, and lastly, to our family and friends, who were our greatest support throughout this entire study.

[7] REFERENCES

- Ahmad, I., Alqarni, M. A., Almazroi, A. A., & Tariq, A. (2020). Experimental evaluation of clickbait detection using machine learning models. *Intelligent Automation and Soft Computing*, 26(6), 1335–1344. <https://doi.org/10.32604/iasc.2020.013861>
- Chen, Y., Conroy, N. J., & Rubin, V. L. (2015). Misleading online content: Recognizing clickbait as “false news.” *Proceedings of the 2015 ACM on Workshop on Multimodal Deception Detection*, 15–19. <https://doi.org/10.1145/2823465.2823467>
- Kemp, S. (2023, February 8). Digital 2023: The Philippines - datareportal – global digital insights. DataReportal. <https://datareportal.com/reports/digital-2023-philippines>

[4] RESULTS AND DISCUSSIONS

A total of 27,505 English non-clickbait, 6,321 English clickbait, 18,430 Tagalog non-clickbait, and 8,813 Tagalog clickbait news headlines were collected. To prevent bias when training the model, the overall number of titles was reduced to 24,000 rows consisting of equal amounts of clickbait and non-clickbait news headlines in English and Tagalog.

A total of thirty (30) respondents of differing backgrounds were selected for testing the usability of the web application, with two (2) respondents having been professional journalists for their feedback on the reliability of the model.

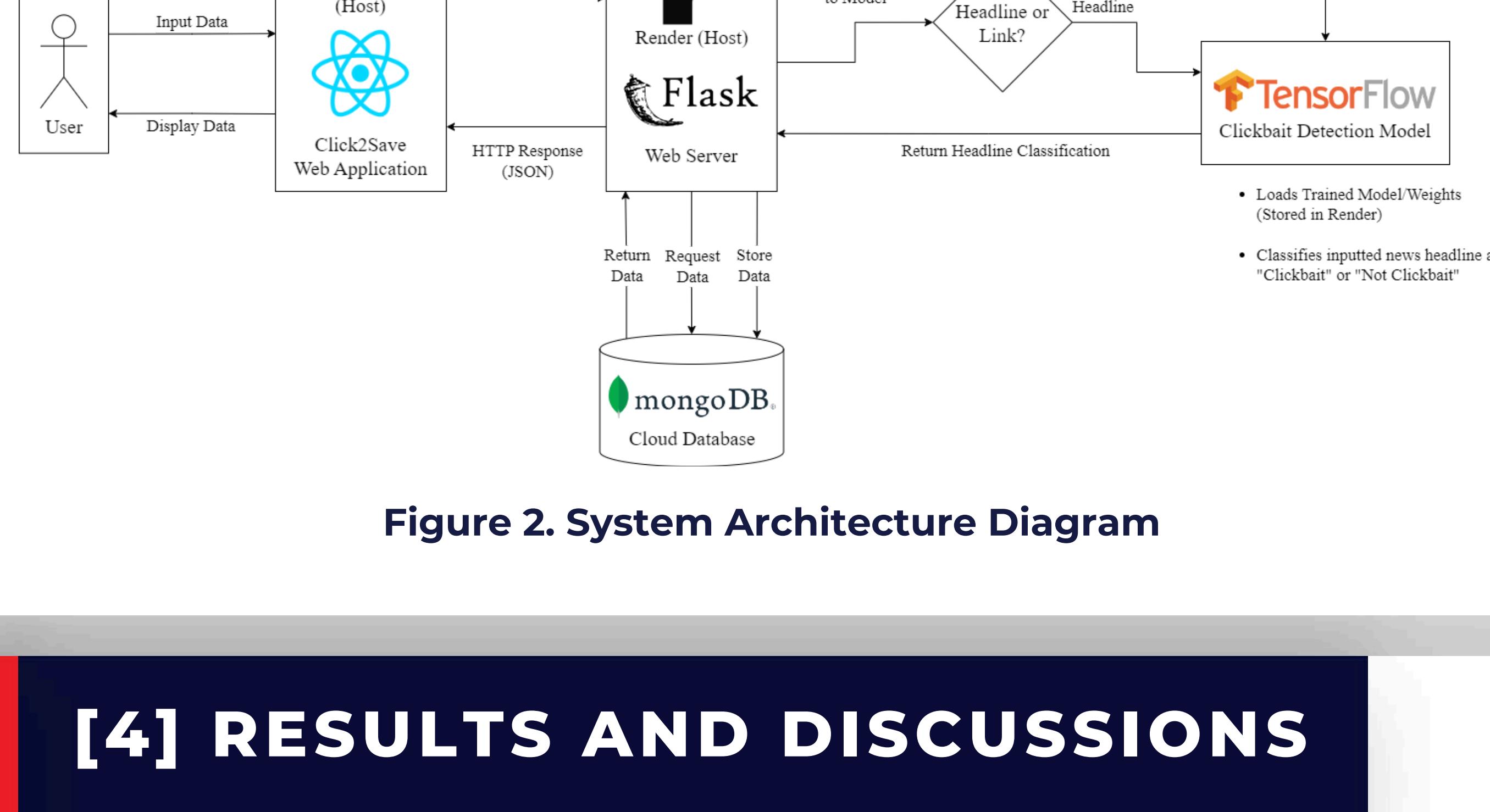


Figure 2. System Architecture Diagram