Financial Argument Analysis in Bengali

Rima Roy, Sohom Ghosh and Sudip Kumar Naskar

November 16, 2023

Financial Argument Analysis in Bengali

Rima Roy
Jadavpur University
Kolkata, India
rima.roy1006@gmail.com

Sohom Ghosh
Jadavpur University
Kolkata, India
sohom1ghosh@gmail.com

Sudip Kumar Naskar
Jadavpur University
Kolkata, India
sudip.naskar@gmail.com

ABSTRACT

Argument mining is an emerging area of research. Argument mining in financial domain specifically for low-resources language like Bengali is in its nascent stage. There exist no datasets for argumentative financial texts mining in Bengali. In this paper, we propose two new datasets in Bengali for financial argument analysis. Subsequently, we released two transformer-based models fine-tuned on these datasets as baselines for financial argumentative unit classification and for detecting the relation between two argumentative financial texts.

CCS CONCEPTS

• **Computing methodologies → Information extraction**; **Language resources**.

KEYWORDS

Argument Mining, Language Resources in Bengali, Financial Natural Language Processing, Natural Language Processing

ACM Reference Format:
Rima Roy, Sohom Ghosh, and Sudip Kumar Naskar. 2023. Financial Argument Analysis in Bengali. In *Forum for Information Retrieval Evaluation (FIRE 2023), December 15–18, 2023, Panjim, India.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3632754.3632763

1 INTRODUCTION

In natural language processing (NLP), argument mining has of late become one of the popular areas of research. The goal of argument mining is to automatically extract and identify argumentative structures from texts. Argumentative texts are present in various places like investor-generated text, social media posts, etc. While argument mining has been a subject of discourse for several years, financial argument analysis is still in the early stage. To the best of our knowledge, no work has been done to date on financial argument analysis in low-resource languages such as Bengali.

In this paper, we focus on developing resources for mining arguments from financial texts in Bengali. First, we introduce the task of **Argument Unit Classification** to classify financial argumentative texts in Bengali into 'Premise' or 'Claim'. 'Premise', being more truthful than the 'Claim', helps investors make data-driven decisions. Secondly, we present the task of **Argument Relation Identification** to understand if two financial argumentative texts in Bengali are supporting, attacking or not related to each other. To quantify the effect of social media posts and understand an entire discussion, it is essential to examine the posts that support or attack each other.

Our contributions

•We created two datasets in Bengali for the task of Financial Argument Unit Classification and Argument Relation Identification. These datasets are released under CC BY-NC-SA 4.0 licence.[1]
•We fine-tuned two pre-trained language models to accomplish the above-mentioned tasks. We open-sourced these models so that the research community can use them as baselines.
•We developed a user-friendly tool (**Financial Argument Analysis in Bengali (FAAB)**) for demonstration[2] and hosted it in HuggingFace Spaces.[3]

2 RELATED WORKS

Argument analysis is one of the emerging research area in Natural Language Processing. Lippi and Torroni [12] surveyed existing works relating to argument mining across various domains like economic sciences, policymaking, and information technology. Cabrio and Villata [3] narrated various machine learning and deep learning algorithms to predict relationships between texts. Schaefer and Stede [19] reviewed existing works on argument mining specifically for Twitter. They discussed the approaches used for modelling the structure of arguments in the context of tweets. Furthermore, they studied the current progress in detecting arguments, and their relations in tweets. Finally, they explored the overlap between stance detection and argument mining. Lawrence and Reed [11] also surveyed recent advances in the domain of Argument Mining. Argument analysis has been widely adopted across various domains like legal [20] and Finance [5]. Xu and Ashley [20] conceptualised argument mining as a word-level classification problem instead of a sentence-level classification problem. Chen et al. [5] proposed the structures between the opinions and those between the financial instruments. They discussed how opinions from various sources can be used to extract opinion components and detect the relation between the opinions. Chen et al. [6] explored the applicability of option mining in the financial domain. They analysed the investor's opinion. In [21], Zhai et al. proposed a dataset called AntCritic which consists of 10K free-form and visually-rich financial comments and supports both argument component detection and argument relation prediction task.

Although there have been significant advances in the field of Argument Mining on financial English texts, a lot is yet to be done

FIRE 2023, December 15–18, 2023, Panjim, India

ACM ISBN 979-8-4007-1632-4/23/12.
https://doi.org/10.1145/3632754.3632763

[1] https://github.com/rima357/FAAB_Financial_Argument_Ananlysis_Bengali (accessed on 11th Aug 2023)
[2] https://youtu.be/4JwVl4mbj6Q (accessed on 9th Aug 2023)
[3] https://huggingface.co/spaces/rima357/FinArgBengali (accessed on 9th Aug 2023)

for low resource languages like Bengali. To address this knowledge gap, we release datasets, models, and, a tool for effectively analysing arguments in Bengali financial texts.

3 PROBLEM STATEMENT

We want to accomplish the following tasks.
Task-1: Given a financial argumentative text in Bengali, our aim is to classify it as 'Premise' or 'Claim'.
Task-2: Given two financial argumentative texts in Bengali, we want to classify the relationship between them. The relation can be 'support', 'attack' or 'none' (i.e. no relation).

4 DATASETS

Due to a lack of resources to create ground truth for Bengali datasets from scratch, we leveraged the resources available in English. We translated the datasets released by Chen et al. [7] from English to Bengali using the translation system released by Ramesh et al. [13]. To understand the applicability of the translation system for argumentative financial text, we evaluated it using a two-step approach. Firstly, we manually translated 100 instances from English to Bengali and calculated the average BERTScore [22] and Cross-lingual Optimised Metric for Translation Evaluation (COMET) [14] scores for machine-translated sentences and human-translated gold standard reference translations in Bengali. The scores obtained were 0.964 and 0.859, respectively. Secondly, as manual translation was not scalable, we randomly picked 8000 instances from the overall dataset. We translated the English texts into Bengali using the MT system [13]. Then, we back translated the Bengali texts to English using the same MT system (in the opposite language direction). We calculate the BERTScore between the original English text and the back translated English text. Due to absence of actual ground truth for reference, we could not calculate the COMET score. We obtained an average BERTScore of 0.944. To further ensure the validity of the machine translation system, we calculated the LaBSe [9] based cosine similarity between sentence embeddings [16] of original texts in English and texts in Bengali obtained by machine translation. The average cosine similarity score was 0.845. As the scores obtained were high in all scenarios, we conclude that the system developed by Ramesh et al.[13] to translate to and from Indian languages is applicable for our dataset.

To further ensure the quality of our dataset, we segmented the data set into several brackets based on the BERTScore and cosine similarity scores. For each of these brackets, we calculated the accuracy of translation by manually assessing how many of the 10 randomly picked instances from the corresponding bracket were translated properly. This is presented in Figure 1. Analysing this plot, we decided to retain only those translated instances having BERTScore and cosine similarity score above 0.925 and 0.800, respectively. We use the same training set, validation set, and labels as mentioned in the paper [7]. After applying the filters mentioned, the label-wise distribution of instances for Task-1 and Task-2 are mentioned in Tables 1 and 2 respectively.

For Task-1, we have text in Bengali and the corresponding label (0=Premise, 1 = Claim). For Task-2, we have two texts in Bengali and their corresponding relation (0 = No relation, 1 = Support, 2 = Attack). Few samples of Task-1 and Task-2 are shown in Tables 3 and 4 respectively.

Table 1: Data distribution of Task-1

Dataset	Premise (Label-0)	Claim (Label-1)
training	2858	2667
validation	353	334

Table 2: Data distribution of Task-2

Dataset	No relation (Label-0)	Support (Label-1)	Attack (Label-2)
Training	1970	794	28
validation	230	108	5

Figure 1: Accuracy of each bracket created using LaBSE based cosine-similarity and BERTScore

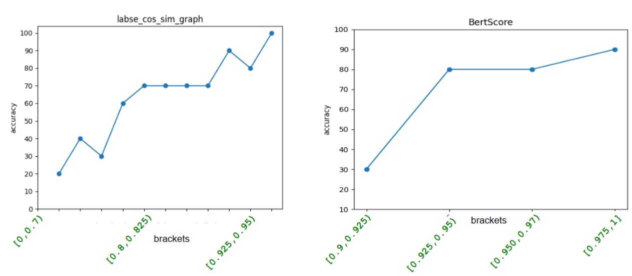

Table 3: Some instances from Task-1 dataset

Text (Bengali)	Label
এবং এই প্রেক্ষাপটে, অবশ্যই, তারা পুরোনো কিছু কাজের বোঝা তুলে নিচ্ছে এবং স্থানান্তর করছে, কিন্তু তারা পুরো ব্যবসায়িক প্রক্রিয়ার আধুনিকীকরণ করছে।	0 (Premise)
হ্যাঁ, কোয়ার্টারের জন্য, তাই এটি একটি শক্তিশালী কোয়ার্টার ছিল।	1 (Claim)

Table 4: Some instances from Task-2 dataset

Text-1 (Bengali)	Text-2 (Bengali)	Label
তাই প্রথমবার আমরা ঘোষণা করেছিলাম যেআমাদের ২০ লক্ষেরও বেশি বিজ্ঞাপনদাতা আছেন যারা ফেসবুকে বিজ্ঞাপন কিনছেন।	৮২% মানুষ যারা আমাদের সাথে বিজ্ঞাপন শুরু করেন তারা আমাদের খুব সাধারণ বিজ্ঞাপন পণ্য দিয়ে শুরু করেন।	0 (No relation)
আমাদের আরও বেশি বিজ্ঞাপনদাতা রয়েছেন যারা সঠিক ব্যক্তিকে লক্ষ্য করে তাদের বিজ্ঞাপন দেওয়ার ক্ষমতা ব্যবহার করেন।	এবং আমি মনে করি আমরা এই সমস্ত ক্ষেত্রে অভূতপূর্ব বৃদ্ধি দেখতে পাচ্ছি।	1 (Support)
আপনারা যদি চীনের মূল ভূখণ্ডের দিকে তাকান, যার ওপর আমি ব্যক্তিগতভাবে বিশেষ দৃষ্টি নিবদ্ধ করেছি, তা হলে চীনের মূল ভূখণ্ডে আমাদের সংখ্যা ১১ শতাংশ কমে গেছে।	আর তাই আমি যখন এর থেকেঘুরে দাঁড়াই এবং বৃহত্তর চিত্রের দিকে তাকাই, আমি মনে করি চীন ততটা দুর্বল নয় যতটা বলা হয়েছে।	2 (Attack)

5 EXPERIMENTS AND RESULTS

Firstly, for Task-1, we fine-tuned several variants of BERT [8] for classifying the given text into 'premise' and 'claim'. These variants are sagorsarker bangla bert base (SSB) [18], monsoon nlp bangla electra[4] (MNB), aibharat indic bert (AIB) [10], distilbert base multilingual cased (DBMC) [17], and bert base multilingual cased (BBMC) [8]. The results of these models for Task-1 are presented in Table 5. We observe that **BBMC** when fine-tuned for 5 epochs, with batch size of 8, weight decay of 0.1, and 500 warm up steps, outperformed all other variants of the BERT model. We further fine-tuned a bert-base-uncased [8] model with the English texts obtained by translating the Bengali texts. Although it performed slightly better than BBMC, the lift in performance was not statistically significant (p-value > 0.05) and deploying a machine translation system was an extra overhead.

Table 5: Results for Task-1. [A= Accuracy, P = Precision, R = Recall, F1 = F1 (binary)]

Model Name	A	P	R	F1
SSB	0.700	0.690	0.694	0.692
MNB	0.714	0.691	0.745	0.717
AIB	0.711	0.700	0.712	0.706
DBMC	0.681	0.656	0.721	0.687
BBMC	**0.719**	**0.697**	**0.745**	**0.721**

[4]https://huggingface.co/monsoon-nlp/bangla-electra (accessed on 9th August, 2023)

For Task-2, following [4], we used the cross-encoder architecture[5] [15]. We trained it for classifying the relation between two texts in Bengali into one of 3 categories ('Support', 'Attack', or 'None'). We experimented with several variants of the BERT model, as done in Task-1. The results are presented in Table 6. We observe that similar to Task-1, **BBMC** when fine-tuned for 5 epochs, with bath size of 16, and 88 warm up steps, performed the best. Subsequently, we fine-tuned a bert-base-uncased [8] model with the English texts obtained by translating the Bengali texts. Just like Task-1, it performed slightly better than BBMC for Task-2 as well. However, the performance improvement was not statistically significant (p-value > 0.05) and deploying a Bengali to English machine translation system in production was an extra overhead.

Lastly, for both the tasks, we tried to adapt **BBMC** to the financial domain using Masked Language Modelling. However, this did not improve the performance.

Table 6: Results for Task-2. [A= Accuracy, P = Precision, R = Recall, F1 = F1 score, mi = micro, ma = macro]

Model Name	A	P mi	P ma	R mi	R ma	F1 mi	F1 ma
SSB	0.708	0.708	0.442	0.708	0.417	0.708	0.419
MNB	0.705	0.705	0.235	0.705	0.333	0.705	0.275
AIB	0.682	0.682	0.476	0.682	0.348	0.682	0.303
DBMC	0.699	0.699	0.442	0.699	0.392	0.699	0.386
BBMC	**0.755**	**0.755**	**0.488**	**0.755**	**0.46**	**0.755**	**0.466**

6 LARGE LANGUAGE MODELS FOR ARGUMENT ANALYSIS

Since Large Language Models (LLMs) have been re-defining the state of the art in NLP, we experimented with llama-2[6] under zero shot and few shot setting. We further instruction fine-tuned llama-2[7] and evaluated the instruction fine-tuned version under zero shot setting. We performed these experiments with English texts obtained by translation due to the unavailability of LLMs for Bengali. The prompts which we have used are mentioned in Table 7. The results for Task-1 are presented in Table 8. Since the performance for Task-1 did not improve on using llama-2, we did not carry out experiments with LLMs for Task-2.

Table 8: Results for Task-1 using llama-2. [IF = Instruction fine-tuned, A= Accuracy, P = Precision, R = Recall, F1 = F1 (binary)]

Model	Type	A	P	R	F1
llama-2	Zero shot	0.505	0.484	0.278	0.354
llama-2	Few shot	0.522	0.500	0.186	0.271
llama-2	IF + Zero Shot	0.528	0.531	0.254	0.344

[5]https://www.sbert.net/examples/applications/cross-encoder/README.html (accessed on 9th August, 2023)
[6]https://ai.meta.com/llama/ (accessed on 18th Aug, 2023)
[7]https://huggingface.co/TinyPixel/Llama-2-7B-bf16-sharded (accessed on 18th Aug, 2023)

Table 7: The Zero Shot and Few Shot Prompts for Task-1

Prompts for LLM (TinyPixel/Llama-2-7B-bf16-sharded)
Zero-Shot: Classify the following Input text into one of the following two categories: [‘Premise’, ‘Claim’] Input : {text}
Few-Shot: Classify the text given bellow into Premise or Claim based on the meaning of the text. Choose only one Class either Premise of Claim for a text. Input : I mean, sometimes it is not that you came up with a bright strategy, it is like doing really good work continuously for a long time. Response : The class of the text is Premise. Input : See, first of all, I would like to say that the opportunity for our shareholders was never better when they thought of Microsoft. Response : The class of the text is Claim. Input : For example, we have never participated so much, I would call it a burden of all sophisticated work in the non-developed market, medium and small businesses. Response : The class of the text is Premise. Input : However, primarily, the feed for the video is going to focus on making money through advertisements. Response : The class of the text is Claim. Input : {text}

7 TOOL DESCRIPTION

For helping investors, we developed a user-friendly tool, **Financial Argument Analysis in Bengali (FAAB)**. In the back-end, we use the best performing model, i.e., bert-base-multilingual-cased [8]. This has been fine-tuned separately for Task-1 and Task-2. The front-end was developed using Gradio [1].

Figure 2: Screenshot of Tab-1 of the tool kit

Figure 3: Screenshot of Tab-2 of the tool kit

As presented in Figures 2 and 3, the tool consists of two tabs. The first tab (Tab-1) is used to classify a Bengali argumentative text into premise or claim. The other tab (Tab-2) is used to detect the relation between two Bengali argumentative texts.

In Tab-1, there is a text box where users can write the text in Bengali. Below this there is the ‘classify’ button which on clicking will classify the text into ‘Premise’ or ‘Claim’. The output will be shown below. In Tab-2, that there are two text boxes, one is for Text-1 and the other is for Text-2 where users can enter the texts in Bengali. By clicking on the ‘Detect the relation’ button, it detects the relation between these two Bengali texts and the output is shown below. We have provided a list of examples for both the tasks at the bottom.

8 CONCLUSION

In this paper, we proposed two datasets for mining argumentative financial texts in Bengali. Subsequently, we released the baseline models and open-sourced the **Financial Argument Analysis in Bengali (FAAB)** tool. The baseline models were obtained by fine-tuning bert-base-multilingual-cased. Collecting more data (specifically for Task-2), estimating profitability from these arguments, and experimenting with INDICXNLI datasets [2] are interesting directions for future work.

REFERENCES

[1] Abubakar Abid, Ali Abdalla, Ali Abid, Dawood Khan, Abdulrahman Alfozan, and James Zou. 2019. Gradio: Hassle-Free Sharing and Testing of ML Models in the Wild. arXiv:1906.02569 [cs.LG]

[2] Divyanshu Aggarwal, Vivek Gupta, and Anoop Kunchukuttan. 2022. IndicXNLI: Evaluating Multilingual Inference for Indian Languages. In Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang (Eds.). Association for Computational Linguistics, Abu Dhabi, United Arab Emirates, 10994–11006. https://doi.org/10.18653/v1/2022.emnlp-main.755

[3] Elena Cabrio and Serena Villata. 2018. Five Years of Argument Mining: a Data-driven Analysis. In Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18. International Joint Conferences on Artificial Intelligence Organization, 5427–5433. https://doi.org/10.24963/ijcai.2018/766

[4] Swagata Chakraborty, Anubhav Sarkar, Dhairya Suman, Sohom Ghosh, and Sudip Kumar Naskar. 2023. LIPI at the NTCIR-17 FinArg-1 Task: Using Pre-trained Language Models for Comprehending Financial Arguments.
[5] Chung-Chi Chen, Hen-Hsen Huang, and Hsin-Hsi Chen. 2021. From Opinion Mining to Financial Argument Mining. From Opinion Mining to Financial Argument Mining (2021). https://api.semanticscholar.org/CorpusID:234787011
[6] Chung-Chi Chen, Hen-Hsen Huang, and Hsin-Hsi Chen. 2021. A Research Agenda for Financial Opinion Mining. Proceedings of the International AAAI Conference on Web and Social Media 15, 1 (May 2021), 1059–1063. https://doi.org/10.1609/icwsm.v15i1.18130
[7] Chung-Chi Chen, Chin-Yi Lin, Chr-Jr Chiu, Hen-Hsen Huang, Alaa Alhamzeh, Yu-Lieh Huang, Hiroya Takamura, and Hsin-Hsi Chen. 2023. Overview of the NTCIR-17 FinArg-1 Task: Fine-Grained Argument Understanding in Financial Analysis. In Proceedings of the 17th NTCIR Conference on Evaluation of Information Access Technologies, Tokyo, Japan.
[8] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). Association for Computational Linguistics, Minneapolis, Minnesota, 4171–4186. https://doi.org/10.18653/v1/N19-1423
[9] Fangxiaoyu Feng, Yinfei Yang, Daniel Cer, Naveen Arivazhagan, and Wei Wang. 2022. Language-agnostic BERT Sentence Embedding. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, Dublin, Ireland, 878–891. https://doi.org/10.18653/v1/2022.acl-long.62
[10] Divyanshu Kakwani, Anoop Kunchukuttan, Satish Golla, Gokul N.C., Avik Bhattacharyya, Mitesh M. Khapra, and Pratyush Kumar. 2020. IndicNLPSuite: Monolingual Corpora, Evaluation Benchmarks and Pre-trained Multilingual Language Models for Indian Languages. In Findings of the Association for Computational Linguistics: EMNLP 2020, Trevor Cohn, Yulan He, and Yang Liu (Eds.). Association for Computational Linguistics, Online, 4948–4961. https://doi.org/10.18653/v1/2020.findings-emnlp.445
[11] John Lawrence and Chris Reed. 2020. Argument mining: A survey. Computational Linguistics 45, 4 (2020), 765–818.
[12] Marco Lippi and Paolo Torroni. 2016. Argumentation Mining: State of the Art and Emerging Trends. ACM Trans. Internet Technol. 16, 2, Article 10 (mar 2016), 25 pages. https://doi.org/10.1145/2850417
[13] Gowtham Ramesh, Sumanth Doddapaneni, Aravinth Bheemaraj, Mayank Jobanputra, Raghavan AK, Ajitesh Sharma, Sujit Sahoo, Harshita Diddee, Mahalakshmi J, Divyanshu Kakwani, Navneet Kumar, Aswin Pradeep, Srihari Nagaraj, Kumar Deepak, Vivek Raghavan, Anoop Kunchukuttan, Pratyush Kumar, and Mitesh Shantadevi Khapra. 2022. Samanantar: The Largest Publicly Available Parallel Corpora Collection for 11 Indic Languages. Transactions of the Association for Computational Linguistics 10 (02 2022), 145–162. https://doi.org/10.1162/tacl_a_00452 arXiv:https://direct.mit.edu/tacl/article-pdf/doi/10.1162/tacl_a_00452/1987010/tacl_a_00452.pdf
[14] Ricardo Rei, Craig Stewart, Ana C Farinha, and Alon Lavie. 2020. COMET: A Neural Framework for MT Evaluation. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). Association for Computational Linguistics, Online, 2685–2702. https://doi.org/10.18653/v1/2020.emnlp-main.213
[15] Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Kentaro Inui, Jing Jiang, Vincent Ng, and Xiaojun Wan (Eds.). Association for Computational Linguistics, Hong Kong, China, 3982–3992. https://doi.org/10.18653/v1/D19-1410
[16] Nils Reimers and Iryna Gurevych. 2020. Making Monolingual Sentence Embeddings Multilingual using Knowledge Distillation. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Bonnie Webber, Trevor Cohn, Yulan He, and Yang Liu (Eds.). Association for Computational Linguistics, Online, 4512–4525. https://doi.org/10.18653/v1/2020.emnlp-main.365
[17] Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2019. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. ArXiv abs/1910.01108 (2019).
[18] Sagor Sarker. 2020. BanglaBERT: Bengali Mask Language Model for Bengali Language Understanding. https://github.com/sagorbrur/bangla-bert
[19] Robin Schaefer and Manfred Stede. 2021. Argument Mining on Twitter: A survey. it - Information Technology 63, 1 (2021), 45–58. https://doi.org/10.1515/itit-2020-0053
[20] Huihui Xu and Kevin Ashley. 2022. Multi-Granularity Argument Mining in Legal Texts. Frontiers in Artificial Intelligence and Applications 362 (2022), 261–266.
[21] Tianyi Zhang*, Varsha Kishore*, Felix Wu*, Kilian Q. Weinberger, and Yoav Artzi. 2020. BERTScore: Evaluating Text Generation with BERT. In International Conference on Learning Representations. https://openreview.net/forum?id=SkeHuCVFDr
[22] Tianyi Zhang*, Varsha Kishore*, Felix Wu*, Kilian Q. Weinberger, and Yoav Artzi. 2020. BERTScore: Evaluating Text Generation with BERT. In International Conference on Learning Representations. https://openreview.net/forum?id=SkeHuCVFDr