



ASADULLAH AL GALIB <asadullah.al.galib@g.bracu.ac.bd>

26th International Conference on Computer and Information Technology : Submission (309) has been created.

1 message

Microsoft CMT <email@msr-cmt.org>

Tue, Sep 5, 2023 at 3:21 AM

Reply-To: Microsoft CMT - Do Not Reply <noreply@msr-cmt.org>

To: asadullah.al.galib@g.bracu.ac.bd

Hello,

The following submission has been created.

Track Name: ICCITconf2023

Paper ID: 309

Paper Title: Genre Classification: A Machine Learning Based Comparative Study of Classical Bengali Literature

Abstract:

Bengali literature, specifically classical Bengali literature has been a source of inspiration, a spark for paradigm-shifting revolutions, and the sole sustaining source of cultural thirst for hundreds of millions of people over many generations. Unfortunately, very few attempts have been made to analyze this never-ending collection of literary works from the luminary figures of Bengali literature. The availability of high-quality research-ready datasets comprising all the authenticated literary works has been a key obstacle in conducting NLP research, utilizing the most recent advancements in deep learning and large language models. Identifying the genre of a given text snippet is a key step in analyzing a vast collection of works comprising different styles, themes, and motivations from classical authors. From classifying previously unexplored archival documents to identifying and suggesting similar literary works for modern recommender engines, genre classification opens the door for many downstream and specialized use cases. In this project, we initiate an ambitious goal of compiling a comprehensive dataset of literary works from classical authors and eventually extending the collection to contemporary writers as well. We explore both classical methods such as Naive Bayes as well as LSTM and recent transformer-based models to classify genre from short text snippets. We concluded that fine-tuning pre-trained BERT models produced much higher accuracy than both classical and LSTM models.

Created on: Mon, 04 Sep 2023 21:21:00 GMT

Last Modified: Mon, 04 Sep 2023 21:21:00 GMT

Authors:

- asadullah.al.galib@g.bracu.ac.bd (Primary)
- maisha.mostofa.prima@g.bracu.ac.bd
- satabdi.rani.debi@g.bracu.ac.bd
- md.muntasir.mahadi@g.bracu.ac.bd
- nayema.ahmed@g.bracu.ac.bd
- ehsanur.rahman.rhythm@g.bracu.ac.bd
- adib.muhammad.amit@g.bracu.ac.bd
- annajiat@gmail.com

Primary Subject Area: Deep Learning/Machine Learning

Secondary Subject Areas: Not Entered

Submission Files: Genre_Classification_A_Machine_Learning_Based_Comparative_Study_of_Classical_Bengali_Literature.pdf (500 Kb, Mon, 04 Sep 2023 21:20:46 GMT)

Submission Questions Response: Not Entered

Thanks,
CMT team.

Download the CMT app to access submissions and reviews on the move and receive notifications:

<https://apps.apple.com/us/app/conference-management-toolkit/id1532488001>

<https://play.google.com/store/apps/details?id=com.microsoft.research.cmt>

To stop receiving conference emails, you can check the 'Do not send me conference email' box from your User Profile.

Microsoft respects your privacy. To learn more, please read our [Privacy Statement](#).

Microsoft Corporation

One [Microsoft Way](#)

Redmond, WA 98052