TONMOY TALUKDER

✉ tonmoytalukder.cs@gmail.com   ☎ (+880) 1880 788228

🌐 https://tonmoytalukder.github.io/   in https://www.linkedin.com/in/tonmoytalukder/

⌂ 23/2-C, Shah Ali Bag, Mirpur 1, Dhaka 1216

Research Interests

My research focuses on exploring the utilization of *Natural Language Processing* in how to utilize these technologies to improve the understanding and use of *low-resource languages*. Alongside, I am particularly interested in exploring *Multimodal Deep Learning*, *Computer Vision*, and *Human-Computer Interaction*. Additionally, I aim to investigate the interaction between machine learning models and data representations during training, with the goal of improving their performance.

Education

Ahsanullah University of Science Technology   Dhaka, Bangladesh
Discipline: Computer Science and Engineering
Degree: Bachelor of Science   May, 2023

Research Experience

Research Assistant of Undergrad Supervisor   (2022 − Present)

- Exploring multimodal research scopes using *LLMs (Large Language Models)* in NLG, Computer Vision area for the Bangla language.

- Developed *keyword-extractor* and *sentence generation from keywords technique* for a low-resource language, Bangla. Also developed a dataset for this.

- Explored ranking-based Bengali text summarization.

Publication

Under Review

Bengali Key-to-Text: Text Generation from Keywords in Bengali
Authors: G. M. Shahariar*, and **Tonmoy Talukder***
(* denotes equal contribution, names are listed in alphabetical order).
Conference: The 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP 2023)

Ongoing Research

1. **Image2Cap: Bengali Caption Generation from Images using Pre-trained Transformers**

   - The ongoing work involves the development of a pre-trained LLM transformer-based model designed to perform both object detection and image captions generation in Bengali.

2. **Cross Lingual Key2Text: Bengali Text Generation from Cross-Lingual Keywords**

   - The previously developed **Bangla Key2Text** model has the capability to generate Bengali sentences from unordered Bengali keywords. The current focus is on enhancing the model's performance to enable it to generate meaningful sentences even from cross-lingual keywords.

3. **Multimodal Image and Text classNameification using Pre-trained Transformers**

   - This research focuses on performing multi-modal image and text classification using the LAION 400 Million Open Dataset. A portion of the multimodal dataset, comprising 400 million pairs, has been preprocessed for this purpose.

Additional research insights can be found at: https://tonmoytalukder.github.io/research.

Published

Rank Your Summaries: Enhancing Bengali Text Summarization via Ranking-based Approach Authors: G. M. Shahariar*, **Tonmoy Talukder***, Rafin Alam Khan Sotez, and Md. Tanvir Rouf Shawon
(* denotes equal contribution, names are listed in alphabetical order).
Conference: International Conference on Big Data, IoT and Machine Learning (BIM 2023)

Undergraduate Thesis

Bengali Key-to-Text: Text Generation from Keywords in Bengali
Supervisor: G. M. Shahariar, Senior Lecturer, Department of Computer Science and Engineering, Ahsanullah University of Science and Technology.

The thesis introduces Bangla Key2Text, a model designed to generate Bengali texts from keywords. A dataset of 2.01 million keyword-text pairs was collected, and the model underwent fine-tuning using two million Bengali pairs. Additionally, a keyword extractor was developed utilizing BERT. The research showcases the model's utility in various tasks, including dialogue systems, language learning, and machine translation for the Bengali language.

Projects

- **Bangla-KeywordExtractor**
  Description: BanglaBERT-based keywords extractor of Bangla Sentences. This extractor works by calculating the cosine similarity of each word embedding with the mean embedding.
  GitHub: https://github.com/TonmoyTalukder/Bangla-KeywordExtractor

- **Bengali Hate Speech Classification**
  Description: NLP classification with different machine learning algorithms with hyperparameter fine-tuning, Ensemble approaches.
  GitHub: https://github.com/TonmoyTalukder/Analysis-on-ML-Model-s-Classification-of-Bengali-Hate-Speech-in-Different-Social-Contexts

- **Diabetes Prediction Analysis**
  Description: Analysis on Pima Indian Diabetes dataset. Applied seven algorithms with hyper-parameter fine-tuning, Ensemble approaches.
  GitHub: https://github.com/TonmoyTalukder/An-Empirical-Study-of-the-Efficacy-among-multiple-MachineLearning-Algorithms-for-Diabetes-Prediction

- Additional project insights can be found at: https://tonmoytalukder.github.io/projects.

Skills

Programming Languages: Python, JavaScript, C/C++, and Markup Language (HTML, CSS)

Frameworks: Pytorch, Tensorflow, React JS, Node JS, Next JS, Material UI, Tailwind CSS

Database: NoSQL (MongoDB), msSQL

Tools: Jupyter Notebook (Anaconda 3), Colab, VS Code, Git, LaTeX

Additional Skills: Exploratory Data Analysis, Data Preprocessing, Full Stack Web Development (MERN)

Volunteering Experience

Mentor (Senior Section)   December, 2021 − Present
Technovation Girls Challenge   California, USA
Technovation is a program of Iridescent, which is a non-profit organization in the United States. It is the world's largest and longest-running tech competition for girls. My responsibility is mentoring students for their AI projects. Responsible for...

- Mentoring students to learn App Development, AI & Machine Learning Tools

- Helping students to develop projects for the Technovation Global Competition

Achievements

- Certificate for completing the course "**Machine Learning with Python**" in EMK Center

- Certificate for completing the course "**Data Analysis with Python**" in EMK Center

- Certificate for completing the course "**Data Visualization with Python**" in EMK Center

- "**Complete Web Development (MERN Stack)**" certification by Programming Hero

Activities

Mentoring

- **Team**: The Peace Ambassadors   Season Three, 2022 − 2023
  Project Name: SERENE
  Depression-based app. Implemented AI Diary, Firebase Authentication, AI Chatbot, AI Quiz along with other features.

- **Team**: Createch   Season Two, 2021 − 2022
  Project Name: ROOT
  Machine learning-based plant disease detection and blog app.

- **Team**: AlphaBit's AI   Season Two, $2021 - 2022$
  Project Name: FITIFY
  Blog app on BMI and BMR with AI calculator to find BMI and BMR scores.

Programming Contest

- Intra Aust Programming Contest   February, 2019
  7th Position, Junior Division
  Ahsanullah University of Science and Technology, Dhaka

- Makers' Mania − Programming Contest   September, 2018
  4th Position, Junior Division
  Independent University Bangladesh, Dhaka

Reference

Dr. Md. Shahriar Mahbub, Professor and Head
Department of Computer Science and Engineering
Ahsanullah University of Science and Technology, Dhaka, Bangladesh
Email: shahriar.cse@aust.edu
Web: www.aust.edu/cse/faculty_member/dr_md_shahriar_mahbub

G. M. Shahariar Shibli, Senior Lecturer
Department of Computer Science and Engineering
Ahsanullah University of Science and Technology, Dhaka, Bangladesh
Email: shahariar_shibli.cse@aust.edu
Web: https://shahariar-shibli.github.io/

Sidratul Moontaha, Founder and CEO
InnoVerse Foundation, Dhaka, Bangladesh
The Diana Award Holder 2021
Email: sidratulmoontaha143@gmail.com

I solemnly declare that all the statements mentioned in this curriculum vitae are accurate.

Tonmoy Talukder
Graduate, Department of Computer Science and Engineering
Ahsanullah University of Science and Technology
Contact No: (+880) 1880 788228
Email: tonmoytalukder.cs@gmail.com
Web: https://tonmoytalukder.github.io/