Sish Ahemmed Shozol

sishahemmed102@gmail.com; +8801735746667;
linkedin.com/in/sishahemmedshozol;

Education:

Khulna University
B.Sc in Statistics *Graduation: Jan 2025*
- **CGPA:** 3.75/4.00 (Passed with distinction)

Dhaka City College *July 2017 - Jun 2019*
- Cumulative **GPA** in **H.S.C.:** 5.00/5.00

Skills:

Research Methods: Data Scraping, Surveying, Interviewing
Languages: Python, R, C/C++, SQL
Frameworks: BERT
Tools/Software: Git, PyCharm, CodeBlocks, Visual Studio Code, Jupyter Notebook, Google Collaboratory, Google Suite, LateX, SPSS
Libraries: Pandas, NumPy, Matplotlib, NLTK, BeautifulSopu4, TensorFlow, Keras,

Problem Solving:

[01] Kaggle Competitions Contributor. kaggle.com/sishahemmed

[02] 100 problems solved in URI online judge. urionlinejudge.com.br/judge/en/profile/227574

Publication:

- Biswas, A., Mir, H., Sultana, N., **Shozol, S. A.**, & Talukder, M. U. (2025, March). Framing Cancer in Media: Analyzing Risk Communication in Newspaper Reporting [Conference session]. International Crisis and Risk Communication Conference (ICRCC), Clemson University in Clemson, South Carolina, USA.

  Contribution: Developing framework, Topic Modeling, Method chapter, Result chapter, and Writing Materials.

- Tasnim, F., Hasib, M., **Shozol, S. A.**, Sultana, N., Akil, M., & Biswas, A. (2025, June). *Communicating health: Framing health-related news in developed and least developed countries during COVID-19* [Conference session]. Canadian Communication Association, George Brown College, Toronto, Ontario.

  Contribution: Data Scraping, Method chapter, Analysis, Topic modeling, Result Chapter.

- Nath, C. D., Biswas, A., Sultana, N., & **Shozal, S. A.** (2024). Framing Russia-Ukraine war in the newspapers of Bangladesh: A topic modeling approach. *South Asian Journal of Social Sciences and Humanities*, *5*(2), 252-267. **https://doi.org/10.48165/sajssh.2024.5215**

  Contribution: Writing Materials and Method chapter, News scrapping, and topic modeling analysis

- Hasib, M., & **Shozol, S. A.** (2024, April). *Exploring Online Discourse of ChatGPT through Sentiment Analysis and Topic Modelling* [Poster session]. Promoting Cross-disciplinary Research, Texas Tech University.

  Contribution: Twitter Scraping, Sentiment Analysis and Topic Modeling, Poster Designing, Writing Materials, and Method Chapter

Research Experience:

Undergraduate Thesis

Title: **Enhancing Fake News Detection Using Data Augmentation and Advanced Machine Learning Algorithms with Explainable AI (XAI)**

Supervisor: Sutapa Dey Barna(KU)

- **Objective**: Enhanced fake news detection by integrating data augmentation techniques with advanced machine learning algorithms, utilizing Explainable AI (XAI) for model transparency.
- **Developed**: A fake news detection system using machine learning models (Logistic Regression, Decision Trees, XGBoost, AdaBoost, and BERT), incorporating data augmentation (synonym substitution, numerical mapping) and LIME for interpretability.
- **Uniqueness**: Focused on both model accuracy and transparency by combining data augmentation and LIME, improving explainability in complex models like BERT.
- **Results**: Achieved exceptional results with BERT, including accuracy, precision, and recall rates up to 99%. Integrated LIME further enhanced model interpretability.
- **Tech Used**: Python, machine learning algorithms (Logistic Regression, Decision Trees, XGBoost, AdaBoost, BERT), data augmentation (Synonym Substitution, Numerical Mapping), LIME.

Research on the July Mass Uprising in Bangladesh(On-going)

Title: **Global Media Topic Modeling and Sentiment Analysis on the Student Movement Issue in Bangladesh: A Comparative Study of News Articles from India, Pakistan, and Bangladesh**

Supervisor: Dr. Mahdy Rahman Chowdhury(NSU)

- **Objective**: Conducted a comparative analysis of news coverage on the July Mass Uprising across Bangladesh, India, and Pakistan, focusing on sentiment distribution and topic modeling.
- **Sentiment Analysis**: Scraped news articles from the three countries and performed sentiment analysis using classical, ensemble, and transformer-based models. Employed LIME for model interpretability and applied few-shot, Chain of Thought (CoT), and Tree of Thought (ToT) prompting techniques with LLMs for sentiment prediction.
- **Topic Modeling**: Applied Latent Dirichlet Allocation (LDA) to extract the top 5 topics for each country and analyzed how the coverage of the movement evolved over time. Extracted the top 25 words and key phrases to understand key themes in the news.
- **Tech Used**: Python, LLM prompting techniques (Few-shot, CoT, ToT), sentiment analysis models, LDA for topic modeling, LIME for interpretability. BeautifulSoup4.

Research Assistant

Paper Title: **Framing Environmental Issues of Bangladesh in International Media: A Topic Modelling Analysis on The Guardian Newspaper.**

Supervisor: Farzana Tasnim Pinky

- Published in Sage Journals.
- Responsibilities:
    - Performed data pre-processing on textual data using **NLTK** to prepare it for further analysis.
    - Applied **Latent Dirichlet Allocation (LDA)** for topic modeling to identify key themes and topics within the articles related to environmental issues.
    - Conducted keyword and key-phrase extraction using the **KeyBERT model** to identify significant terms and phrases associated with the identified topics.

kaggle.com/code/sishahemmed/enviroment-topic-modeling

Projects:

---

[01] Twitter Scraping with SNscrape

***Goal*:** To scrape tweets from a specific country on a particular topic using SNscrape, with the aim of analyzing the sentiment and perspective on Twitter regarding that topic.

***Tools Used*:** Python, NumPy, Pandas, SNscrape, Data Processing.
kaggle.com/code/sishahemmed/twitter-scrape-snscrape

[02] News Article Scraping and Content Analysis from Washington Post Archive

***Goal*:** To scrape news articles from the Washington Post's health section using the PRISM API for content analysis and comparative research on media coverage across different regions.

***Tools Used*:** Python, BeautifulSoup4, JSON, NLTK, NumPy, Newspaper3k.
kaggle.com/code/sishahemmed/washington-post-archive-api-scrape

[03] Online Judge Problem Scrape with BeautifulSoup4 Library:

The goal of this project is to scrape programming problems from different online judge sush as: Codeforces, Timus etc. This program can scrape different problem based on their category, and store them into csv file.

Tech Used: Numpy, Pandas, BeautifulSoup4, Request Library, Parsing.
kaggle.com/code/sishahemmed/online-judge-scrape

[04] Predicting House Prices with Regression using TensorFlow:

This ML project can predict House Prices in a particular area using TensorFlow. Data cleaning, feature engineering, and data visualization are also applied here.

Tech Used: NumPy, Pandas, Matplotlib, and TensorFlow. coursera.org/verify/ZHRWPLAVXNC4

[05] Titanic - Machine Learning from Disaster:

Used machine learning to create a model that predicts which passengers survived the Titanic shipwreck. Data cleaning, feature engineering, and data visualization are also applied here.

Tech Used: Numpy, Pandas, Matplotlib, Linear-regression.
kaggle.com/sishahemmed/tabular-data-april

Training and Course Completion:

- ☐ **Applied Machine Learning/Applied Deep Learning/Artificial Inteligence (Natural Language Processing) Program.**
  Mahdy Research Academy
  Supervisor: Dr. Mahdy Rahman Chowdhury(NSU)
  - Gained in-depth knowledge of **machine learning algorithms** and their applications in real-world problems.
  - Hands-on experience with **data preprocessing**, **feature engineering**, and applying both **machine learning** and **deep learning models** in Python.
  - Gained proficiency in **research tools** such as **LaTeX**, **draw.io**, and other research methodologies.
  - Developed problem-solving skills for **Natural Language Processing (NLP)** applications and how to approach these problems efficiently.
  - Learned how to formulate methodologies in a structured and constructive manner for research projects.

  Project: **Global Media Topic Modeling and Sentiment Analysis on the Student Movement Issue in Bangladesh: A Comparative Study of News Articles from India, Pakistan, and Bangladesh.** (On-going)

Tech Used: Python, LLM prompting techniques (Few-shot, CoT, ToT), Sentiment analysis models, LDA for topic modeling, LIME for interpretability, BeautifulSoup4, Parsing, NLTK, Request Library.

- [ ] **Data Science Bootcamp Program**
  Intelligent Machines Limited
  Completed: Dec 2021 drive.google.com/file/d/1ChWQLDw5OUHedZjlESK1l_V39vifMnxW/view?usp=sharing

  - Gained expertise in data analysis, machine learning algorithms, and statistical modeling.
  - Worked with real-world datasets, applying data cleaning, feature engineering, and predictive modeling techniques.
  - **Tools Used**: Excel, Tableau

- [ ] **Python for Everybody Specialization.** coursera.org/verify/P93EDGF7BATK

- [ ] **Statistics 101.** courses.cognitiveclass.ai/certificates/6e36b50de29742cd9b953bdef7ae7237

- [ ] **Linear Algebra.** coursera.org/verify/L4YEYL9CKJS3

- [ ] **Differential Calculus through Data and Modeling.** coursera.org/verify/specialization/45Z8YUHQKWK4

Achievements:

- **Undergraduate Scholarship**: Awarded scholarship for academic excellence during my undergraduate studies based on exceptional academic performance at **Khulna University**.
- **5th BDBO-Samakal National Biology Olympiad, 2019**:
  Position: 2nd Runner Up, **Dhaka South Region**, **Category: Higher Secondary**.

Extracurricular Activities:

- **Executive Member,** কালের কণ্ঠ-শুভসংঘ, ঈশ্বরদী *March 2017 - June 2017*
  Actively participated in community-driven activities and youth empowerment initiatives.
- **Venue Manager**, **1st DCCDC National Debate Championship, 2018**
  Coordinated and managed the logistics and operations of the event, ensuring smooth execution.

Standardized Test Scores:

IELTS (International English Language Testing System)
Score: 7.5/9 · March 2024
IELTS Academic Module
Listening: 8.0 **Reading**: 7.5 **Writing**: 6.5 **Speaking**: 7.0

Interests:

- ML, LLM, AI Agent