

Mohammad Sabik Irbaz

Portfolio: msi1427.github.io/
GitHub: github.com/msi1427

Email: sabikirbaz@iut-dhaka.edu
Phone: +8801623848732

I am a Data Scientist at Leadbook Pte. Ltd. I am currently working with Dr. Md. Ruhul Amin from Fordham University on data-driven Social NLP problems. I used to work as a Lead ML Engineer at Omdena Inc. and Pioneer Alpha Ltd. where I have led several ML projects from research, and development to production. During my four years of undergraduate studies, I was immensely fortunate to be advised by Dr. Md. Kamrul Hasan and Dr. Hasan Mahmud where I have mostly worked on Text Style Transfer (TST), Clinical ML, and Data-centric AI.

Research Interests

I am primarily excited about Natural Language Processing (NLP), Artificial Intelligence (AI), Computational Social Science, and Data Science research. My research interests include:

- NLP Applications (Low-resource NLP, QA, Commonsense Reasoning, etc.) [P3, P4]
- AI for Good (Social, Clinical Other Disciplines) [P1, P2, P5, P6, P7, P8]
- Fact verification using historical data [P7]
- Data-centric AI [P6]

Education

Islamic University of Technology(IUT) B.Sc. in Computer Science and Engineering (CSE) CGPA: 3.44/4.00 Dissertation: “Medical Expertise Style Transfer using Denoising Autoencoders” [Thesis Book] [Presentation] Supervisor: Dr. Md. Kamrul Hasan and Dr. Hasan Mahmud	Dhaka, Bangladesh Jan, 2017 - Mar, 2021
Notre Dame College(NDC) Higher Secondary Certificate (HSC), 2016 GPA: 5.00/5.00	Dhaka, Bangladesh Jul, 2014 - Dec, 2016
Cox’s Bazar Government High School Secondary Secondary Certificate (SSC), 2014 GPA: 5.00/5.00	Cox’s Bazar, Bangladesh Jan, 2009 - Dec, 2013

Research Experience

{^αprimary contributor; ^βsecondary contributor; ^γsupervisor}

Fordham Human Centered AI Research Lab <i>Department of CIS, Fordham University, New York, USA</i> Advisor(s): Dr. Md. Ruhul Amin (1) ^α COVID-19 Public Policy Conformity Analysis — <i>NLP, Data Extraction, Topic Modeling, NLI, EDA</i> [P7] (2) ^β Mitigating Social Bias in NLP — <i>NLP, Explainability, Text Style Transfer</i> [P8]	Mar, 2021 - Present
Computer Vision Lab (CVLab) <i>Department of CSE, IUT, Dhaka, Bangladesh</i> Advisor(s): Sabbir Ahmed (1) ^α Automatic Radiology Report Generation — <i>NLP, Image Captioning, Image Segmentation</i> [Report] (2) ^γ Logical Reasoning QA using Standardized Test Questions — <i>NLP, QA, CSR, Data Collection, NLI</i>	Nov, 2020 - Present
Systems and Software Lab (SSL) <i>Department of CSE, IUT, Dhaka, Bangladesh</i> Advisor(s): Dr. Md. Kamrul Hasan, Dr. Hasan Mahmud (1) ^β Automated wheelchairs for semi-paralyzed patients. — <i>Signal Processing, ML, Arduino</i> (2) ^α Medical Expertise Text Style Transfer — <i>NLP, Active Learning, Text Style Transfer</i> [P6] (4) ^γ Query-focused Multi-Document Text Summarization for Bangla — <i>NLP, Low-resource QA, Summarization</i>	Jul, 2019 - Present

Reviewing Experience

Reviewed multiple conference papers for **EMNLP 2022** and **WACV 2023**

Publications

More updated list [here](#) or [Google Scholar](#)

{*equal contribution, ^ccorresponding author}

[P8] Md. Shihab Shahriar, **Mohammad Sabik Irbaz**, and Md. Ruhul Amin^c. “**Mitigating Social Bias in NLP using explainable AI methods**”. *Manuscript in Preparation*.

[P7] **Mohammad Sabik Irbaz**, Rupak Sarkar, Md. Ruhul Amin^c and Ashiqur R. KhudaBukhsh. “**COVID-19 Public Policy Conformity Analysis**”. *Manuscript in Preparation*. [Presentation]

[P6] **Mohammad Sabik Irbaz**, Abir Azad, Zarif Ahmed, Md. Ashfaq Raiyan, Raian Rahman, Hasan Mahmud, and Md. Kamrul Hasan^c. “**c-MSD: A Curated Dataset for Medical Expertise Style Transfer using Active Learning Approach**”. *Under Review*.

[P5] **Mohammad Sabik Irbaz**^c, Fardin Ahsan Sakib, and Lutfun Nahar Lota. “**Predicting User-specific Future Activities using LSTM-based Multi-label Classification**”. *Proceedings of the 4th International Conference on Activity and Behavior Computing 2022* [Presentation]

[P4] Alif Ashrafee*, Akib Mohammed Khan*, **Mohammad Sabik Irbaz**^c, and MD Abdullah Al Nasim. “**Real-time Bangla License Plate Recognition Pipeline for Low Resource Video-based Applications**”. *Real-World Surveillance: Applications and Challenges Workshop, collocated with IEEE/CVF WACV 2022* [Presentation] [Poster]

[P3] Fahim Shahriar Khan, Mueeze Al Mushabbir, **Mohammad Sabik Irbaz**^c, and MD Abdullah Al Nasim. “**End-to-End Natural Language Understanding Pipeline for Bangla Conversational Agents**”. *Proceedings of the 20th IEEE International Conference on Machine Learning and Applications, 2021* [Presentation]

[P2] **Mohammad Sabik Irbaz**^c, MD Abdullah Al Nasim, and Refat E Ferdous. “**Real-Time Face Recognition System for Remote Employee Tracking**”. *Proceedings of the International Conference on Big Data, IoT, and Machine Learning, 2022. Springer, Singapore*.

[P1] **Mohammad Sabik Irbaz**^{*}, Abir Azad^{*}, Tanjila Alam Sathi, and Lutfun Nahar Lota^c. “**Nurse Care Activity Recognition Based on Machine Learning Techniques Using Accelerometer Data**”. *Proceedings of the 8th International Workshop on Human Activity Sensing Corpus and Applications, collocated with ACM UbiComp/ISWC 2020* [DOI] [Presentation]

Industry Experience

Leadbook Pte. Ltd.

Singapore

Data Scientist

Dec, 2021 - Present

Supervising and maintaining the following projects to improve our Lead Generation process

(1) *Industry Type Classification*: According to LinkedIn, there are 148 types of industries and one company can be attributed to multiple industries. We hired 10 annotators and trained them to annotate 30K data to facilitate multi-label classification. Our final model is trained in 3 steps. a) Language Model (LM) finetuning with 1M company data, b) Knowledge distillation from large zero-shot classifier c) Using the finetuned LM and distilled classifier to finetune on the 30K annotated data.

(2) *Job Title Clustering*: We have millions of contacts in our database. An important problem is to cluster contact job titles into categories in an unsupervised or semi-supervised manner. Most of my works involves improving and maintaining this pipeline. The project involves — NLP, Data Cleaning, Clustering, Phrase Similarity [Code]

(3) Some of my regular responsibilities involve maintaining data quality and cater for customized lead generation requests utilizing different data transformation, scraping and ML methods.

Omdena Inc.

New York, USA

Lead Machine Learning Engineer

Apr, 2021 - Jun, 2021

Collaborated with 50 ML and data engineers from 20 countries on **Equilo Project** - Building Gender Equality Assessment Functionality using NLP. This large scale work was divided into 7 major tasks. I was the **task leader** of the team building the **recommender system** for the organizations with low gender equality scores. I also lead the team building the web app and deploying the whole data and ML pipeline to **GCP**.

This project involved — *Scraping, NLP, GCP, Firestore, Django, ReactJS, StreamLit, Cloud SQL* [Article]

Pioneer Alpha Ltd.

Lead Machine Learning Engineer

Dhaka, Bangladesh
Jan, 2021 - Oct, 2021

Supervised and delivered the following core projects successfully to the corresponding clients:

- (1) Multilingual Chatbot as Business Assistant — *NLP, NLU, Low-resource, Tensorflow, Rasa, Flask* [Code] [P3]
- (2) Low-resource Bangla License Plate Detection and Recognition — *OCR, Tensorflow, OCR, Flask* [Code] [P4]
- (3) Face Detection and Recognition for Remote Office — *Vision, OpenCV, dlib, Flask* [P2]
- (4) Adobe Premiere Plugin for Automated Video Editing — *Vision, OpenCV, dlib, Adobe Premiere* [Code]
- (4) Trello Automation for Directed Premiere Video Editing — *Vision, OpenCV, dlib, Adobe Premiere, Trello* [Code]

Nascenia Ltd.

Software Engineer, Intern

Dhaka, Bangladesh
Nov, 2019 - Jan, 2020

Worked on a large-scale government project on VAT automation, **Project Uddogi**

This project involved — *Ruby on Rails, HTML, CSS, Heroku* [Article]

Teaching Experience

MasterCourse

Dhaka, Bangladesh
Sep, 2022 - Present

Data Science Instructor

Dokkho Data Science Career Program — [Course URL]

The goal of this program is the following to bridge the gap between academia and industry in the field of Data Science and Machine Learning and build the next generation of data scientists and machine learning engineers. I am taking care of teaching and mentoring side of this program, MasterCourse is handling logistics and other management issues.

Amar iSchool

Head Course Instructor

Dhaka, Bangladesh
Mar, 2021 - Dec, 2021

Applied Data Science and Machine Learning — [Course Repository]

Our prime goal is to empower the local ML enthusiasts with a limited background CS and Mathematics.

- Taught data science and machine learning for production
- Gave 10 lectures, all self-developed
- Developed course assignments and projects

Skills

Programming Language: Python, C, C++, Ruby, Octave

Framework: Pytorch, fastai, Flask, Tensorflow, Keras, Ruby on Rails

Tools and Libraries: Git, LaTeX, Google Colab, HuggingFace, Google Cloud Platform

Languages: English (*fluent*), Bangla (*fluent*), Hindi (*medium*), Urdu (*medium*), Arabic (*beginner*)

Selected Awards & Achievements

An exhaustive list can be found [here](#)

- Won the OIC Scholarship that waived 75% of my undergraduate education costs (2% acceptance rate)
- **Champion**, 4th Nurse Care Activity Recognition Challenge 2022. Won 100,000 JPY. [P5]
- **Champion**, 2nd Nurse Care Activity Recognition Challenge 2020. Won a trip to Japan. [P1]
- **Finalist**, AI For Bangla Challenge 2021 (*among 147 teams*) [P3]
- **Runners Up**, IUT Computer Programming Contest 2019 (*among 77 teams*)
- **2nd Runners Up**, EDU Inter University Programming Contest 2020 (*among 50 teams*)

Extra-curricular Activities

- **Travelled** across major locations in Bangladesh.
- **Contest Judge**, 16th December Programming Contest 2019
- **Former Programming Instructor**, IUT Programming Community
- **Former Vice President**, IUT Computer Society
- **Assistant General Secretary**, IUT 10th ICT Fest
- **Volunteer Coordinator**, IUT 9th ICT Fest
- **Class and Batch Representative**, IUT CSE Batch'16