

Mohammad Sabik Irbaz

LinkedIn: [linkedin.com/in/sabik-irbaz/](https://www.linkedin.com/in/sabik-irbaz/) **Twitter:** twitter.com/msi_bertman
Portfolio: msi1427.github.io/ **Email:** sabikirbaz@iut-dhaka.edu
GitHub: github.com/msi1427 **Phone:** +8801623848732

Last updated December 14, 2021

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, 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:

- Efficient modeling, benchmarking, evaluation and robustness in NLP [P5, P6]
- Annotating and cleaning datasets with minimal cost [P5, P8]
- Fact verification using historical data [P7]
- Explainable, Interpretable, and Reproducible AI
- Low-resource and Multilingual NLP [P3, P4]
- AI for Good (Social, Clinical & Other Disciplines) [P1, P2, P5, P6, P7]

Education

Islamic University of Technology(IUT)	Dhaka, Bangladesh
B.Sc. in Computer Science and Engineering (CSE)	Jan, 2017 - Mar, 2021
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	
Notre Dame College(NDC)	Dhaka, Bangladesh
Higher Secondary Certificate (HSC), 2016	Jul, 2014 - Dec, 2016
GPA: 5.00/5.00	
Cox’s Bazar Government High School	Cox’s Bazar, Bangladesh
Secondary Secondary Certificate (SSC), 2014	Jan, 2009 - Dec, 2013
GPA: 5.00/5.00	

Research Experience

{ ^α primary contributor; ^β secondary contributor; ^γ supervisor}	
Fordham Human Centered AI Research Lab	Mar, 2021 - Present
<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) ^α Mitigating Social Bias in NLP — <i>NLP, Explainability, Text Style Transfer</i>	
Computer Vision Lab (CVLab)	Nov, 2020 - Present
<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 Data-centric Approach for Explainable Text Classification — <i>NLP, Active Learning, Explainability</i>	
(3) ^γ Logical Reasoning QA using Standardized Test Questions — <i>NLP, QA, CSR, Data Collection, NLI</i>	
Systems and Software Lab (SSL)	Jul, 2019 - Present
<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, Aduino</i>	
(2) ^α Medical Expertise Text Style Transfer — <i>NLP, Active Learning, Text Style Transfer</i> [P5, P6]	
(3) ^α Efficient Active Learning Approach for Data Annotation — <i>ML, Active Learning, Data Annotation</i> [P8]	
(4) ^γ Query-focused Multi-Document Text Summarization for Bangla — <i>NLP, Low-resource QA, Summarization</i>	
(5) ^α Trustworthy NLP — <i>NLP, Out-of-domain Data Filtering</i>	

Publications

More updated list [here](#) or [Google Scholar](#)
{*equal contribution, ^ccorresponding author}

[P8] **Mohammad Sabik Irbaz**, Raian Rahman, Hasan Mahmud, and Md. Kamrul Hasan^c. “**Efficient Active Learning Approach for Data Annotation**”. *Manuscript in Preparation*.

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

[P6] **Mohammad Sabik Irbaz**, Abir Azad, Raian Rahman, Hasan Mahmud, and Md. Kamrul Hasan^c. “**Medical Expertise Style Transfer using Denoising Autoencoders**”. *Manuscript in Preparation*. [[Book](#)]

[P5] **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*.

[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, 2022* [[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

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](#)]

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.

Dhaka, Bangladesh

Lead Machine Learning Engineer

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) Trello Automation for Directed Premiere Video Editing — *Vision, OpenCV, dlib, Adobe Premiere, Trello* [[Code](#)]

Nascenia Ltd.

Software Engineer, Intern

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

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

Dhaka, Bangladesh

Nov, 2019 - Jan, 2020

Teaching Experience

Amar iSchool

Head Course Instructor

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 15 lectures, all self-developed
- Developed course assignments and projects

Dhaka, Bangladesh

Mar, 2021 - Dec, 2021

Selected Projects

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

- (1) **Medical Expertise Style Transfer System for Layman Patients** — *PyTorch, Fastai, Flask* [[Code](#)]
- (2) **Multimodal Multi-label Genre Classifier** — *PyTorch, Fastai, Numpy, Pandas* [[Code](#)]
- (3) **Anime vs Animation vs CGI** — *PyTorch, Fastai, Flask* [[Code](#)]
- (4) **Original Transformer for Bengali Translation** — *PyTorch* [[Code](#)]

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, Bangla, Hindi, Arabic

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**, 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.
- Part of ML Discussion Groups: Fordham DS Research Group, Turing NLP Group
- **Associate Member**, Omdena AI Global Community
- **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

References

Available upon request