

# Mohammad Sabik Irbaz

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

*Last updated December 7, 2021*

I am a Data Scientist (NLP) 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 in Natural Language Processing (NLP), Artificial Intelligence (AI) and Data Science research. My research interests include:

- Efficient modeling, benchmarking, evaluation and robustness in NLP
- Annotating and cleaning datasets with minimal cost
- Fact verification using historical data
- Explainable, Interpretable, and Reproducible AI
- Low-resource and Multilingual NLP
- AI for Good (Social, Clinical & Other Disciplines)

## Education

<b>Islamic University of Technology(IUT)</b>	Dhaka, Bangladesh
B.Sc. in Computer Science and Engineering (CSE)	Jan, 2017 - Mar, 2021
<b>CGPA:</b> 3.44/4.00 [ <b>3.90</b> in the final semester]	
<b>Dissertation:</b> “Medical Expertise Style Transfer using Denoising Autoencoders” [ <a href="#">Report</a> ] [ <a href="#">Presentation</a> ]	
<b>Supervisor:</b> Dr. Md. Kamrul Hasan, Professor, Department of CSE, IUT	
<b>Notre Dame College(NDC)</b>	Dhaka, Bangladesh
Higher Secondary Certificate (HSC), 2016	Jul, 2014 - Dec, 2016
<b>Cox’s Bazar Government High School</b>	Cox’s Bazar, Bangladesh
Secondary Secondary Certificate (SSC), 2014	Jan, 2009 - Dec, 2013

## Research Experience

{ <sup>α</sup> primary contributor; <sup>β</sup> secondary contributor; <sup>γ</sup> supervisor}	
<b>Fordham Human Centered AI Research Lab</b>	Mar, 2021 - Present
<i>Department of CIS, Fordham University, New York, USA</i>	
<b>Advisor(s):</b> Dr. Md. Ruhul Amin	
(1) <sup>α</sup> COVID-19 Public Policy Conformity Analysis — <i>NLP, Data Extraction, Topic Modeling, NLI, EDA</i> [ <b>P6</b> ]	
(2) <sup>α</sup> Mitigating Social Bias in NLP — <i>NLP, Explainability, Text Style Transfer</i>	
<b>Computer Vision Lab (CVLab)</b>	Nov, 2020 - Present
<i>Department of CSE, IUT, Dhaka, Bangladesh</i>	
<b>Advisor(s):</b> Sabbir Ahmed	
(1) <sup>α</sup> Automatic Radiology Report Generation — <i>NLP, Image Captioning, Image Segmentation</i> [ <a href="#">Report</a> ]	
(2) <sup>γ</sup> Data-centric Approach for Explainable Text Classification — <i>NLP, Active Learning, Explainability</i>	
(3) <sup>γ</sup> Logical Reasoning QA using Standardized Test Questions — <i>NLP, QA, CSR, Data Collection, NLI</i>	
<b>Systems and Software Lab (SSL)</b>	Jul, 2019 - Present
<i>Department of CSE, IUT, Dhaka, Bangladesh</i>	
<b>Advisor(s):</b> Dr. Md. Kamrul Hasan, Dr. Hasan Mahmud	
(1) <sup>β</sup> Automated wheelchairs for semi-paralyzed patients. — <i>Signal Processing, ML, Aduino</i>	
(2) <sup>α</sup> Medical Expertise Text Style Transfer — <i>NLP, Active Learning, Text Style Transfer</i> [ <b>P5, P7</b> ]	
(3) <sup>α</sup> Efficient Active Learning Approach for Data Annotation — <i>ML, Active Learning, Data Annotation</i> [ <b>P8</b> ]	
(4) <sup>γ</sup> Query-focused Multi-Document Text Summarization for Bangla — <i>NLP, Low-resource QA, Summarization</i>	
(5) <sup>α</sup> Trustworthy NLP — <i>NLP, Out-of-domain Data Filtering</i>	

## Publications

More updated list [here](#) or [Google Scholar](#)  
{\*equal contribution, <sup>c</sup>corresponding author}

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

[P7] **Mohammad Sabik Irbaz**, Abir Azad, Raian Rahman, Dr. Hasan Mahmud, and Dr. Md. Kamrul Hasan<sup>c</sup>. “**Expertise Text Style Transfer using Denoising Autoencoders**”. *Manuscript in Preparation*.

[P6] **Mohammad Sabik Irbaz**, Dr. Md. Ruhul Amin<sup>c</sup>, Rupak Sarkar, and Dr. Ashique KhudaBukhsh. “**COVID-19 Public Policy Conformity Analysis**”. *Manuscript in Preparation*.

[P5] **Mohammad Sabik Irbaz**, Abir Azad, Zarif Ahmed, Md. Ashfaq Raiyan, Raian Rahman, Dr. Hasan Mahmud, and Dr. Md. Kamrul Hasan<sup>c</sup>. “**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**<sup>c</sup>, 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**<sup>c</sup>, and MD Abdullah Al Nasim. “**End-to-End Natural Language Understanding Pipeline for Bangla Conversational Agents**”. *Proceedings of the 20<sup>th</sup> IEEE International Conference on Machine Learning and Applications, 2021*

[P2] **Mohammad Sabik Irbaz**<sup>c</sup>, 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, 2021*

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

## Industrial Experience

**Leadbook Pte Ltd.** Singapore  
**Data Scientist (NLP)** 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 involve solving and maintaining this pipeline. The project involves — NLP, 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* [P3]
- (2) Low-resource Bangla License Plate Detection and Recognition — *OCR, Tensorflow, OCR, Flask* [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]
- (5) Trello Automation for Directed Video Editing — *Vision, OpenCV, dlib, Adobe Premiere, Trello*

**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 - Present

**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](#)

- **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*)
- **Runners Up**, INWED Tech Fest IUPC 2019 (*among 30 teams*)
- **2nd Runners Up**, EDU Inter University Programming Contest 2020 (*among 50 teams*)
- **2nd Runners Up**, Britannia University Inter University Programming Contest 2020 (*among 25 teams*)
- **OIC Scholarship**, 75% expense coverage for undergraduate studies at IUT. [**2%** acceptance rate]

**Competitive Programming Experience**

- Solved **2000+** programming problems in different online judges using C++, Python.
- Participated in **20+** onsite Inter University Programming Contests and **200+** online programming contests.
- Participated in **ACM ICPC Dhaka Regionals** 3 times.
- Attended BACS - BUBT **National Programming Camp** 2017.
- Achieved 1632 max rating (**Expert**) in [Codeforces](#) and 1844 max rating (**Violet**) in [Codechef](#).
- **Contest Judge**, [16th December Programming Contest 2019](#)
- **Former Programming Instructor**, IUT Programming Community

**Extra-curricular Activities**

- **Associate Member**, Omdena Global 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'16
- [Travelled](#) across major locations in Bangladesh.
- Part of ML Discussion Groups: Fordham DS Research Group, Turing NLP Group

## References

*Available upon request*

---