

Sumon Kanti Dey

<https://github.com/SumonKantiDey>

<https://www.kaggle.com/sumon23>

<https://www.linkedin.com/in/sumon-kanti-dey-96321b10b>

sumonkantidey23@gmail.com | +8801879862842

EDUCATION

NOAKHALI SCIENCE AND TECHNOLOGY UNIVERSITY | B.Sc. IN COMPUTER SCIENCE

Passed Dec 2018 | Noakhali, Bangladesh • GPA: 3.32 / 4.00

EXPERIENCE

LEADBOOK.COM | ENGINEER SCIENTIST

Apr 2020 - Present | Singapore

- Maintain data pipeline to ingest data from various sources into the data warehouse correctly, efficiently, and securely.
- Implemented Apache Kafka-based email generation, verification, and recommender pipeline to produce the best email pattern to customer.
- Designed, refactored, and developed backend APIs using FastAPI and work with containers including Docker and Kubernetes.
- Developed and deployed a transformer-based model for industry-type classification.
- Designed a heuristic-based information retrieval tool to crawl company postal addresses from company metadata.

HIFI DIGITAL LTD | PYTHON DEVELOPER

Mar 2019 - Mar 2020 | Banani, Dhaka

- Developed large-scale information Retrieval tool using Selenium and BeautifulSoup.
- Bootstrapping production-ready machine learning models.
- Present information using data visualization techniques like Tableau, Plotly.
- Processing, cleansing, and verifying the integrity of data used for analysis.
- Designed, developed, refactored, and deployed backend API in Django and AWS.

TEACHING ASSISTANT | NOAKHALI S&T UNIVERSITY

Jun 2017 - Jun 2018 | Noakhali, Dhaka

- CSTE 2202 Data Structure and Algorithms Lab (2017 & 2018 session).
- Designed Course material, lesson plans, homework assignments.
- Helped students to gain a deeper understanding of concepts introduced in lecture.

TECHNICAL SKILLS

Languages : Python, C/C++, Java, JavaScript, D3.js, SQL, MongoDB, Prolog.

Frameworks/Packages : BeautifulSoup, Selenium, Tableau, FastAPI, Flask, Django, Apache Kafka.

ML Libraries : Scikit-learn, Pytorch, Keras, NLTK, Matplotlib, Spyder, Tweepy, Pandas.

Tools : Linux, Git, AWS, Docker, Kubernetes(Learning).

PROJECTS

COVID19 FAKE NEWS DETECTION IN ENGLISH [[GITHUB](#)]

- It was Constraint@AAAI2021 shared task where the target to predict the committed tweet is real or fake.
- Used transformers-based variant of Bert and Roberta models alongside different architecture on top of Bert and Roberta models.
- Utilized Ensemble-based approach based on top three model. Our best model achieved the weighted F1-score of 0.978 on the test set

JIGSAW MULTILINGUAL TOXIC COMMENT CLASSIFICATION [[GITHUB](#)]

- Used Multi-Stage training based on different architectures on top of XLM-RoBERTa.
- The small-scale model finished at 151st place out of 1621 teams.

QUORA INSINCERE QUESTIONS CLASSIFICATION [[GITHUB](#)]

- Applied 5 folds stack attention based Bi-GRU with different embedding (Glove, Word2Vec, FastText).
- Build a web app to see real time sincere vs insincere question classification.

BENGALI.AI HANDWRITTEN GRAPHEME CLASSIFICATION [[GITHUB](#)]

- Multi-class classification challenge where need to classify three constituent elements from an image: grapheme root, vowel diacritics, and consonant diacritics.
- 64x64 pixels with one single channel image fit into a MultiOutput CNN model to classify three classes from the image. Build a Flask WebApp that will help to identify real-time Grapheme with three constituent elements.

APTOS 2019 DIABETIC RETINOPATHY DETECTION TO STOP BLINDNESS

- Used Pre-trained EfficientNet with fine-tuning for classification tasks.
- **Secured 199th Position (top 7 percent) among 2931 teams.**

TOXIC COMMENT CLASSIFICATION CHALLENGE ON WIKI DATA

- Six types of toxicity detection from wiki conversation text using Logistic Regression with TF-IDF words and char n-grams. Accuracy 0.9754.
- Implemented simple (Bi-GRU+LSTM+CNN+Pooling) with Pre-trained FastText Word Embeddings, which outperformed the previous model. Accuracy 0.9828.

SELECTED PUBLICATIONS

- Md. Rezaul Karim, **Sumon Kanti Dey** , Tanhim Islam, Bharathi Raja Chakravarthi, **Multimodal Hate Speech Detection from Bengali Memes and Texts** [[Preprint](#)]
- Md. Rezaul Karim, **Sumon Kanti Dey** , Tanhim Islam, Sagor Sarker, Mehadi Hasan Menon, Kabir Hossain, Bharathi Raja Chakravarthi, Md. Azam Hossain, Stefan Decker. **DeepHateExplainer : Explainable Hate Speech De-tection for Under-resourced Bengali Language.** IEEE (DSAA'2021)[<https://arxiv.org/pdf/2012.14353.pdf>]

ACHIEVEMENTS AND COMPETITIVE PROGRAMMING

I have joined several national and regional programming contest and solved almost 700 online judge problems in **Lightoj, UvaOJ, Codeforces, Codechef, Spoj**. I have completed Andrew NG's machine learning Course on Coursera.

- Kaggle Competition Expert (Ranked 3471 among 144,292 kagglers in the world). 2020
- Earn a **Certificate** for completing Andrew NG's machine learning course on Coursera. 2018
- 24th place at SUB Inter University Programming Contest. 2017
- 17th place at DIU ACM ICPC Dhaka Region Worlds finals Warm-up. 2016
- 1st place at inter University programming Contest. 2016
- Honorable mention at ACM ICPC Asia Dhaka Regional Contest. 2016
- 1st place at CSTE carnival'15 Programming Contest, NSTU. 2015
- Top 1% in Noakhali S&T University Entrance Exam. Ranked 54 among thousands participants. 2014

SOCIETIES

- Served as the General Secretary at departmental computing club. Help to grow up a community of 150+ people. 2017-2018
- Organized 8+ programming competitions, 1 boot-camp, 6+ workshops and 10+ seminars.
- Chief Judge at Intra NSTU PI Day Programming Contest(26 teams). 2018
- Organizer and Mentor at Junior Programming Boot-camp goicpc18 [45 participants]. 2017
- Workshop Trainer at CSTE Club Competitive Programming. 2016-2017
- Judge and Problemsetter at NSTU Intra CSTE programming Contest [30 teams]. 2017

TALKS

- Introduction to Machine learning and Kaggle at Noakhali S&T University. April 2020
- Choice of a Career on becoming a Potential Computer Science Grad at Noakhali S&T University. Oct 2019