Nazia Tasnim

 Github   nimzia.blu@gmail.com
 +880 1633647074   in Nazia Tasnim   Google Scholar
 Chittagong, Bangladesh.

EDUCATION

B.Sc in Computer Science and Engineering, Shahjalal University of Science and Technology, Sylhet.
CGPA: **3.70/ 4.00**
Graduation : September, 2021.

PROJECTS

mwparserfromhtml
- Python library for parsing and mining metadata from the Enterprise HTML Dumps
- Built the entire module from scratch using **Python**
- Efficiently extracts artticle metadata like Category, Templates, Wikilinks, External Links, Media, References etc. with their subcategories
- Large quantity of plaintext can be extracted with little resource-overhead

Shaako
- A mobile application with the rural people having little to no access to proper medical health care, to their nearest medical facilities instantly through their neighbourhood community health workers, empowered by low-cost technology.
- Built the frontend module in **Android Framework**
- Backend portal is built using **Django**, and uses both **Firebase** and **Azure CMS** for database management.
- The project secure the second runners-up place in the Microsoft Imagine Cup, 2022 - SEANM region.

SUSTCast
- Fully automated online radio platform servicing over 2000 students of Shahjalal University of Science and Technology
- Created the main app using **Firebase real-time database, Android**
- Helped develop the data collection and processing backend with **python3**

Peer to Peer Chat
A chat and file-sharing Android app, that connects two devices on the same network using their IP & port numbers. It uses **Java Sockets** for connecting users.

ACHIEVEMENTS

Outreachy Internship at the Wikimedia Foundation (2022)

Selected as **one of 66 interns from a pool of 4500+ applicants**. Under the supervision of *Dr. Isaac Johnson* and *Dr. Martin Gerlach*, I worked with the Wikimedia Research team for three months to build a parser for Wikimedia Enterprise HTML dumps.

3rd place in CVPR AI City Challenge Track-4 (2022)
- Performed Multi-class product counting and recognition from real-time videos for automated retail checkout with an **F1-score over 45%**
- Built a cascading pipeline consisting of a module foreign object identifier, a hand segmentation module using **U-Net**, an entropy based domain bias detector and **Vision Transfomer** based classifier.
- Developed a **heuristic-based filtration metric** to discard empty frames and improve precision.

Grace Hopper Scholarship Recipient (2021)

I joined the 2021's GHC event, as a student scholar. **Less than 5% applicants** are chosen as scholars each year and I was fortunate to be one of them in my first participation.

4th place in DhakaAI - AI-Based Dhaka Traffic Detection Challenge (2020)

- Secured 4th place out of 250+ teams in this international Computer Vision challenge organized by DhakaAI.
- Cleaned and pre-processed the noisy, poor quality dataset to fit with different light and weather conditions through **augmentation** and manual data collection.
- Tuned the hyperparameters of 5-8 SOTA **Image classification models (EffDet, Yolo, Cascaded-RCNN, Faster-RCNN)** and applied ensemble technique with **WBF**.

BRACATHON 3.0 - Healthcare (2019)

- Became champion competing with over 300 teams.
- Built a system that contained an app and web dashboard of learning & assessment modules for BRAC's **6000+ Health workers**.
- Developed the solution of the presented problem and workflow

EXPERIENCES

Contractual Research Scientist at the Wikimedia Foundationv(2022, September - Present)

- Collaborate with the internal Wikimedia research teams to build NLP tools usable across the 300+ languages supported by Wikipedia
- Generate paper summaries and analysis reports relevant to the computational linguistics domain

Research Assistant, BRAC University (2022, June - Present)

- Work under the supervision of Dr. Nur Yanhaona to develop an Automated Media Analysis Framework with Bias and Stance detection of News outlets
- Built an end-to-end pipeline that extensively mines entity information from news reports and produces a knowledge graph, while establishing geopolitical affiliations

Machine Learning Engineer at Giga Tech Limited (2021, August - 2022, June)

I worked on the National corpus project aiming to develop a complete NLP ecosystem for Bangla. As part of this mammoth task, I have been performing the following tasks :

- Collaboration with the linguistic team and NLP researchers in creating a knowledge base
- Create presentations and demonstrations, for workshops and intra-team knowledge sharing.
- Choose, analyze and establish dataset development and representation guidelines.
- Develop prototypes of NLP pipelines and tools.

Part-time coding instructor at KATclub.ca (2020, October - 2021, July)

- Helped Canadian middle to high school students to start programming through block programming
- Taught **web development, game development and interactive problem solving**

SKILLS

Python, Java, Android Framework, Git, Shell, Pytorch and associated ML frameworks.

Publications

Workshp, Conference and Journal Publications

1. Co-authored a journal paper **'Choice of assemblers has a critical impact on denovo assembly of SARS-CoV-2 genome and characterizing variants'**,published on *Briefings in Bioinformatics*. My tasks included analysis of the **genomic variant call patterns** and visualizing the trends.

2. Authored one conference paper in *The 34th International FLAIRS Conference*, titled : **'Observing the Unobserved: A Newspaper Based Dengue Surveillance System for the Low-Income Regions of Bangladesh'** . The task required unstructured newspaper data scraping, pre-processing, topic modeling with **Guided LDA**, classifying with **SVM** and developing insights from data visualizations.

3. Authored a workshop paper accepted for spotlight presentation in the *Joint KDD 2021 Health Day and 2021 KDD Workshop on Applied Data Science for Healthcare workshop* titled **Exploring the Scope and Potential of Local Newspaper-based Dengue Surveillance in Bangladesh**. Main contribution includes technique of developing a novel database, using **human-in-the-loop** in conjunction with common data mining techniques and methods of classification of dengue disease and its intervention-related news.

4. Authored a workshop paper for the *NAACL 2022 SemEval Workshop - Task 11*, where we demonstrate our approach in tackling the challenge of **recognizing complex Named Entities in Bangla, through data augmentation and majority-voting based ensembling efforts**. The mode performed competitively in the task and placed *8th* in the overall global ranking.

5. Co-authored a workshop paper for the *CVPR 2022 AI City Challenge Workshop - Track 4*, where a built a pipeline to **identify multi-class retail products from real-time videos for automatic retail checkout** through a combination of segmentation, classification and heuristic-based modules to have a model with over 45% F-score and globally secured the 3rd place.

6. Authored a full-length paper for the *WEB3 Conference, 2022 - Research track*, where we build an end-to-end system that extensively **mines information from English newspaper data, resolves the geopolitical affiliations and produces a knowledge-graph of entity relationships**.

On-going Research

7. Interpreting the usability of transformer based models in identification of IGHV β - cells. The task focuses on using **pre-trained deep learning models (BERT, DistilBERT)** to find linguistic patterns in RNA sequences and validating their biological significance through language modeling.

Research Interests

NLP , Computer Vision, Probabilistic ML algorithms, Data Science, Explainable AI