Abhijit Paul

NLP Engineer | Web3 Enthusiast | Leadership

https://abj-paul.github.io | +8801968399154 | bsse1201@iit.du.ac.bd | **Github:** abj-paul

EDUCATION

University of Dhaka - Institute of Information Technology, Bachelors of Science in Software Engineering(BSSE) **2020-Ongoing**
Organizations： IITSEC, DU Research Society(DURS), DU Band Society(DUBS)

Notre Dame College, Dhaka - Higher Secondary School Certification Examination **2018-2019**

Accomplishment: Secured the 70th position out of 86,000 students in the DU Admission Examination.

DEVELOPMENT SKILLSET

Languages:	Python, JavaScript, Solidity, E-Lisp, TypeScript,Java, C, C++, SQL, Bash, Php, Rust
Libraries and Frameworks:	Angular, NodeJS, OpenCV, Hardhat, Scikit-Learn, Matplotlib, Actix Framework(Rust), BNLTK
Development Tools:	Git, Docker, Azure Cloud, Linux/UNIX, Emacs, Jupyter Notebook, Emacs
Others:	Software Requirements Specification, Database Administration, Linux System Administration

PUBLICATIONS

Stemming - The Evolution and Current State with a Focus on Bangla - On Review, EACL Conference September, 2023

      With the advent of LLM and generative agents, the need for a good stemmer is at an all time high. Bangla, being a low-resource high-inflectional language, has a natural need for good stemmers. In this research, we look at existing progress and provide future directions.

Exploring Diverse Coping Mechanisms in 2023: A Comprehensive Survey Across Backgrounds and Cultures. Arxiv June, 2023

      The work on coping mechanisms has been going on for decades. However, a critical problem with the existing literature is an intense use of small-scale surveys. This does not allow us to explore diverse coping mechanisms across cultures. In this work, we collect coping mechanisms from social media and classify them into existing taxonomy of coping mechanisms and emphasize the emergence of interesting coping mechanisms..

EXPERIENCE

IIT, University of Dhaka | NLP Research under Dr. Ahmedul Kabir May 2022 – Present

- o Researched NLP from scratch level bayesian language model - normalization to word embedding, LLMs, Bangla Sign Language, Spelling Correction, Topic Modeling, Information Retrieval, stemmers, parts-of-speech tagging, and many more.
- o As a research project lead, starting from reaching out to external organizations for resources, collaborations to recruiting new members, training and guiding them, leading the team towards the goal of the group and managing international relationships.

KolpoKoushol, Workshop by Universal Machine Inc, Silicon Valley, California. Oct 10 – Oct 15, 2023
An open platform for talented individuals to work together and develop new ideas. I worked on an AI driven distributed Food Crowdfunding Solution for 4 days with food business owners and chefs. It felt similar to workin on HCI discipline where you just use any tool to solve a problem.

Monash University | Software Requirements Analysis - Dr Humphrey O. Obie Dec 2022 - Mar 2023

Analyzed their requirements for a software product that translates android app to ensure localization that respects the human value of universalism. We devised a software requirements specification and system design as part of IEEE ICSE.

Softanbees Technologies Limited, Kolkata, India | Data science Intern Nov 2021 – Dec 2021

- o Creating models like Early Prediction of Birth Weight Based on Maternal Factors.
- o Constructing documentation for research and technical content.

PROJECTS

Assembly Line Management System | Versatile Garments Ltd | NodeJS, TypeScript, Angular, FastAPI, MySQL **Machine Learning**

- o Conducting experimentation to find a NMF-based approach at detecting congestion in the assembly line.
- o Performing Software Requirements Specification after communicating with Versatile Garments Ltd to design our system.

Video Feature Based Sign Mimicking Evaluation System | Python, Angular, TypeScript, FastAPI, Keras **Machine Learning**

- o A platform where teachers can upload their videos in sign language, students can view them and take quizzes.
- o Multiple experimentation starting from data creation to testing. Additionally, deployment in AWS Cloud Server.

Automated-Static-Control-Flow-Testing | React, TypeScript, FastAPI, Python **Testing**

- o Parsing C code to generate static control flow graphs from scratch and then hoisting variables to test using heuristics.
- o Identifying independent paths and generating test cases for each of the independent paths discovered.

Representation Discrimination of Ethnic People in Bangladesh | Python   **Machine Learning**

- o Performing topic modeling to identify the representation discrimination of ethnic people in Bangladesh.
- o An ongoing effort of sentiment analysis to identify the general notion towards how people perceive the idea of ethnicity.

BNLP Knowledge Graph | Org Mode, Emacs Lisp, Firebase   **Research Artifact**

- o The project is aimed at creating a hierarchical graph for BNLP, with subject nodes representing key topics and paper nodes representing papers related to those topics. We used tools readily available to use for nothing to fast-prototype it.

DAOist | Solidity, Hardhat, Design Patterns, IPFS   **Blockchain**

- o A decentralized platform for DAO entities to communicate, vote NFTs and auction them using native DAO currency.
- o A liquidity pool to automatically adjust exchange rates between DAO currencies based on liquidity availability.
- o For each NFT exchange, Royalty is automatically transferred to the original owner of the NFT. Weighted voting scheme to select NFT.
- o Creational pattern to simplify the process of creating DAO currency with the Royalty system, along with the DAO community.

Bangla Sign Language Learning Platform | Collaborative Project | Angular, TypeScript, FastAPI, MySQL   **Collaborative Project**

- o A platform where teachers can upload their videos in sign language, students can view them and take quizzes.
- o Multiple experimentation starting from data creation to testing. Additionally, deployment in AWS Cloud Server.

Dhaka University Bus Tracking System | Collaborative Project | Azure, React, JavaScript,LeafletJS, MySQL   **Deployed**

- o Under the guidance of Dr. Naushin Nowar, leading a team of six towards creating a fast and scalable bus tracking system.
- o Designed a distributed, anonymous implementation of Bus Tracking System. along with their business model.

NLP Library | C++, Gprof, gdb, Cmake   **SPL1**

- o Implemented language model and a wide variety of smoothing and segmentation techniques to preprocess and post-process the data.
- o Showed an application of the NLP library with a noisy channel model for spelling correction tasks and optimized existing research work.

More on Github: abj-paul

Awards

2023	**1st Runners-Up in Blockchain Segment, ITVerse, University of Dhaka, Bangladesh**
2023	**2nd Runners-Up in Notre Dame Inter-University Ideathon Contest, Bangladesh**
2023	**Finalist of National Blockchain Olympiad, Bangladesh**
2023	**1st Runners-up in Deep Learning Sprint, BUET, Bangladesh**
2023	**2nd-Runners up of IEEE BUP Project Showcasing, Bangladesh**
2020	**Champion of Datathon, Coding Objects Private Limited, India**

Research Experience

My research experience covers a wide range of fields, showcasing my versatility and depth of expertise.

1. Natural Language Processing:
   - ○ Worked on word embedding, LLM, and basic tasks like parts of speech tagging, stemming, etc.
   - ○ Current focus on applying NLP in sociology, with ongoing projects on indigenous people and coping mechanisms.
2. Software Engineering:
   - ○ Automated processes, including participation in the ICSE competition to automate localization fixes in GUI.
   - ○ Participated in conferences on software security research, LLM automation in SE, and code generation tasks.
   - ○ Proficient in Software Requirements Specification (SRS) and extensive SRS documentation for large-scale systems.
3. Sociology:
   - ○ Actively collaborating with students from sociology to bridge the gap between SE and sociology.
   - ○ Recognize the need for closer collaboration between the two fields, particularly lacking in the country.
4. Machine Learning / Deep Learning:
   - ○ Implemented ML models for various tasks, with a focus on efficiency and speed.
   - ○ I also had a very interesting perspective on the data driven algorithms to solve problems from reading a paper on localisation. All machine learning approaches are really optimizing an objective function by following a curve that reduces the cost function we define - sometimes supervised, sometimes unsupervised.
5. Computer Vision:
   - ○ Most of my DL/ML tasks were on image data. So I have a long experience working with computer vision.
   - ○ I also slowly understood the necessity of old mathematics-based algorithms. Nowadays, we use data driven algorithms for almost all image processing tasks. But we need to understand that it becomes infeasible when we have to use multiple models. For example, in many Software Engineering tasks, we have a long process and each part of the process requires an algorithm. If

we use ML/DL algorithms for each of them, then the overall performance of the pipeline would significantly deteriorate. So using the whitebox algorithms in these cases result in a controlled performance of the pipeline.

6. Data Mining:
   ○ Brief experience with data mining, including a literature review on Data Cube.
   ○ Certification on data analysis tasks.
7. Other:
   ○ Conducted literature review on E-voting and contributed to the development of a blockchain-based e-voting system.
   ○ Designing a distributed and anonymous tracking system.

CERTIFICATES

2020 **"Industrial Data Science in Python", Coding Objects Private Limited, Kolkata, India.**

2023 **"KolpoKoushol", Universal Machine Inc, California, America.**