

Tahsinul Haque Dhrubo

Melbourne, VIC, Australia

+61 0422 138 819 | tahsinul.haque.dhrubo@gmail.com | [LinkedIn](#)
[GitHub](#) | [Personal Website](#)

PROFESSIONAL SUMMARY

Experienced Data Scientist with a strong foundation in machine learning and a focus on developing end-to-end ML pipelines for document classification and low-resource language tasks. Successfully transitioned logistics document detection systems from rule-based to scalable semantic classification using LLaMA models. Published in ACM and IEEE, skilled in leveraging Python, Scikit-learn, and Docker to deliver data-driven solutions. Eager to apply expertise in machine learning to drive innovation and efficiency as a Machine Learning Engineer.

TECHNICAL SKILLS

- **Programming Languages:** Python, Go, C++, C#, JavaScript, TypeScript
- **ML & AI:** PyTorch, TensorFlow/Keras, Scikit-learn, SpaCy, Hugging Face Transformers, NLP, LSTM, OpenCV, YOLOv5
- **Data & MLOps:** Pandas, NumPy, Docker, Git, CI/CD, Kubernetes
- **Search & Analytics:** Elasticsearch, Kibana, Neo4j, PostgreSQL, MongoDB
- **Backend & APIs:** FastAPI, Django, Flask, Node.js, REST & GraphQL, WebSockets
- **Frontend:** React, Next.js, Tailwind CSS, HTML/CSS

PROFESSIONAL EXPERIENCE

Inteliweave Ltd. (Umbrella Company of AIDocbuilder INC)

Jan 2025 - Jul 2025

Junior Data Scientist (Remote)

Dhaka, Bangladesh

- Led the development of an automated system for detecting logistics-related documents for global partners such as DHL, transitioning from rigid rule-based methods using spaCy and regex to a scalable semantic classification pipeline powered by LLaMA models.

AIDocbuilder Inc.

Sep 2024 - Jan 2025

Junior Data Scientist (Remote)

Toronto, Canada

- Built initial rule-based heuristics using spaCy and regex, along with OCR cleanup and preprocessing pipelines, to classify incoming logistics documents from freight and customs channels. Integrated predictions into a review interface and generated annotated training corpora that seeded the subsequent machine learning migration, improving document classification accuracy

TECHNICAL PROJECTS

Study AI App | <https://github.com/Exile404/Study-AI-App>

Jul 2025 - Aug 2025

- Developed a full-stack AI tutoring platform built with FastAPI and LangChain backend using Google Gemini API for LLM alongside YouTube transcript integration, MongoDB-backed chat memory, and dynamic content generation for summaries, quizzes, resources, and performance reports through a NextJS Shadcn UI frontend delivering personalized interactive study support.

Socrates LLM | <https://github.com/Exile404/Socrates-LLM>

Apr 2025 - May 2025

- Developed Socrates LLM, a custom large language model fine-tuned for contextual reasoning and document understanding, optimized for lightweight deployment and fast inference.

Chakri Khujo - Student Job Board | https://github.com/Exile404/Chakri_Khujo

Nov 2023 - Dec 2023

- Built a full-stack job board enabling students to post and search for internships; implemented user authentication, resume attachment uploads, and advanced filtering using React and Tailwind CSS

SoloNest - Hostel Management System | <https://github.com/Exile404/SoloNest>

Jul 2023 - Aug 2023

- Designed and developed an off-campus housing management platform with room allocation, tenant check-in/out workflows, and role-based access control using Django and PostgreSQL

Mongol Tori Rover Control | <https://github.com/Exile404/Socket-Control>

Feb 2022 - Mar 2022

- Developed a real-time teleoperation and telemetry dashboard for the Mongol Tori Rover, streaming control commands and sensor data over Python sockets with a custom GUI

RESEARCH & PUBLICATIONS

- Tahsinul Haque Dhrubo, Tanzim Reza. Bridging the Communication Gap: A Custom Transformer Encoder Stack Model for Bangladeshi Sign Language Translation. NLP Research Python, Transformer, Mediapipe Developed a Custom Transformer Encoder Stack Model trained on a curated dataset of 102 Bangladeshi Sign Language signs, achieving 99.52% accuracy on training and 98.12% on testing. Modernized assistive technologies convert sign sequences to text in real time, fostering seamless communication between deaf and hearing communities. Architected the model stack for scalability, allowing integration of additional sign languages via dataset updates. Publication: 10.13140/RG.2.2.33230.78406
- Tahsinul Haque Dhrubo, Noshin Tabassum, Md. Shaeak Ibna Salim, Md Tifur Waesh Niloy. BengaliBot: Bridging Language Barriers with AI-Driven Conversations. Deep Learning Research Python, PyTorch, NLTK Developed an offline Bengali conversational agent optimized for low-connectivity settings, combining custom NLTK-based intent detection with a PyTorch sequence-to-sequence model for response generation. Publication: 10.1145/3723178.3723216

- Tahsinul Haque Dhrubo, Noshin Tabassum, Asm Tareq Mahmood, Riead Hasan Khan, Dr. Farig Sadeque, Dr. Muhammad Iqbal Hossain. Bridging Worlds: Enhancing Low-Resources Sign Language Recognition and Seamless BdSL-to-Bengali Conversion with Sequential LSTM Innovation. NLP Research Python, LSTM, Mediapipe Created a proprietary BdSL gesture dataset using Mediapipe keypoint extraction, and trained an LSTM-based model to translate gesture sequences into Bengali text with real-time inference capabilities. Publication: 10.1145/3723178.3723214
- Md Shohanur Rahman, Md. Atikur Rahman, Tahsinul Haque Dhrubo. In-Depth Analysis and a Machine Learning Approach for Predicting Smoking Status. Epidemiology Research scikit-learn, Pandas Conducted comparative analysis of logistic regression, random forest, and gradient boosting classifiers on structured health datasets to predict smoking status; engineered features in Pandas and evaluated via cross-validation. Publication: 10.1109/BECITHCON64160.2024.10962796

EDUCATION

Deakin University <i>M.S.</i>	Jul 2025 - Jul 2027 <i>Melbourne, Australia</i>
BRAC University <i>B.Sc., Computer Science</i>	Jan 2020 - Jan 2024 <i>Dhaka, Bangladesh</i>

ACHIEVEMENTS

- **International:** Kibo Robot Programming Challenge – Crew Award (2022); Runner-up + Crew Award (2021)
- **National:** Robi Datathon 3.0 – Finalist (2024); Bangladesh Blockchain Olympiad – Finalist (2022); BRAC University Programming Contest – Champion (2021); BRAC University Hackathon – Runner-up