Devam Mondal

LinkedIn | GitHub | Edison, NJ | P: 908-449-7970 | devammondal123@gmail.com |

EDUCATION

Georgia Institute of Technology — Atlanta, GA
BS, Computer Science (Intelligence and Modeling & Simulation) — May 2028

Edison High School — Edison, NJ
STEM Academy — June 2024
Cumulative Unweighted GPA: 4.17/4.33
Relevant Coursework: AP Biology, AP Chemistry, AP Calculus BC, AP Physics C, AP Physics 1, Calculus 3, AP Statistics

WORK EXPERIENCE

Natural Language Processing, Data Science, and Machine Learning Intern — Hoboken, NJ
Stevens Institute of Technology, Center for Complex Systems and Enterprises — May 2023 – Present

- Authoring research paper with Professor *Carlo Lipizzi* focusing on reducing large language model bias through a novel dataset augmentation algorithm (currently awaiting publication)
- Developed two new metrics to assess bias (mb-index, db-index) using Spacy named-entity-recognition, LangChain, Pandas, HuggingFace Transformers + Datasets, and mathematical concepts (cosine similarity and lemma vectorization) with regards to datasets and large language models.
- Produced novel dataset augmentation algorithm using k-means clustering and named entity recognition to debias datasets.
- Conducted literature review with 15+ papers, presenting results on a weekly basis.

Natural Language Processing Intern — Somerset, NJ
whiz.ai — Dec 2023 – March 2024

- Developed narrative descriptions of 50+ whiz.ai data summary boards using fine-tuned large language models.
- Addressed large language model input token limitations and with creation and refinement of 5+ Data Intelligence Objects (DIOs), and chunk processing.
- Fine-tuned large language models using QLoRA adapters and validated 50+ responses with statistical testing of various metrics (perplexity, GLUE, etc.).

Machine Learning + Data Science Intern — Somerset, NJ
whiz.ai — May 2023 – Sept 2023

- Ran EDAs on three datasets (> 1 GB) using Pandas, Sweetviz, and statistical methods (Spearman rho).
- Utilized h2o.ai to train 7+ classification and regression models (stacked ensemble, DNNs, GBMs) on medical datasets.
- Carried out feature importance analysis with 50+features through multicollinearity tests (VIF), Gini impurity, and Random Forest Regressors.
- Implemented LangChain, OpenAI API, and HuggingFace fine-tuning to create automated EDA tools that reduced data exploration times by 50%.

PROJECTS

DataScribe — Dec 2023

- Designed and produced a full-stack application that automates dataset analysis and graph analysis with Flask front-end + HuggingFace, OpenAI API, PandasAI, LangChain back-end.
- Hosted app on Heroku (with custom domain) and managed version control on GitHub.

PUBLICATIONS

Mondal, D., & Lipizzi, C. (2024). **Reducing Large Language Model Bias with Emphasis on 'Restricted Industries': Automated Dataset Augmentation and Prejudice Quantification**. *arXiv [Cs.CL]*. Retrieved from http://arxiv.org/abs/2403.13925

SKILLS AND AWARDS

Technical Skills: Java, Python, Numpy, Pandas, PyTorch, TensorFlow, h2o.ai, Flask, OpenCV, Web Dev, HuggingFace, Spacy, Langchain
Certifications & Training: IBM Intro to PyTorch + Deep Learning, Intro to Cisco Packet Tracer, IBM Intro to TensorFlow
Awards: Civil Air Patrol Amelia Earhart Award, 2023 NASAT All-Star, 2022 FTC Dean's List Semifinalist