Biswajit (Sumon) Banerjee

Atlanta, GA • +1-(404)203-8546 • bbanerjee32@gatech.edu • linkedin.com/in/bbaner • github.com/Biswajit-Banerjee

EDUCATION

Georgia Institute of Technology — *GPA: 4.0*   **Atlanta, GA**
Master of Science in Bio-Informatics with Machine Learning Specialization   *Aug 2023 – Dec 2024(expected)*

— **Relevant Courses:** Machine Learning, Deep Learning, Deep Learning on Text, Deep Reinforcement Learning

Asansol Engineering College — *GPA: 3.3*   **Asansol, India**
Bachelor of Technology in Computer Science & Engineering   *Aug 2015 – Aug 2019*

RESEARCH EXPERIENCE

NASA Center for the Origin Of Life, Georgia Tech   **Atlanta, GA**
Graduate Researcher — Machine Learning   *Aug 2023 - Current*

— Researching metabolic fold space search-ability through **protein fold level embedding** and function alignment.

— Co-authored a research paper on RiboVision2, a novel web server employing **Graph Machine Learning** for comprehensive ribosomal RNA structure prediction and visualization. [Publication]

— Engineered a novel pipeline to map metabolic pathways to catalyzing protein folds, integrating multiple public databases and resolving ambiguities. Presented findings at the **ExOrigins Colloquium 2024**. [Poster]

WORK EXPERIENCE

ExtraHop   **Seattle, WA**
Machine Learning Research — Summer Intern   *May 2024 - Aug 2024*

— Developed a compact LLM using network security detection data across **2271 customer networks** that reduced around **68% of false positives** and lowered user alert fatigue.

— This identified groups of related detection events having a high likelihood of being **cyber kill-chain activity**.

Stellapps   **Bangalore, India**
Machine Learning Engineer II — Computer Vision   *May 2022 - Jul 2023*

— Developed an end-to-end open-set object detection system for cattle verification utilizing custom tuned **MobileNetv2 architecture** to accurately distinguish registered cattle (in-set) from unregistered cattle (out-of-set).

— Prioritized model optimization efforts using the 80/20 rule, achieving a **2ms forward pass** while maintaining **84% accuracy** for in-set and **92% for out-of-set** cattle identification.

— Trained **YOLOv6 network** to identify the number of distinct farm animals present in the frame with an **MaP of 88% at IoU 0.95** and deployed in mobile devices.

Synopsys   **Bangalore, India**
Research & Development Engineer II — AI/ML   *Oct 2019 - May 2022*

— Created algorithm to optimize physical circuit module placement, which was **featured in Forbes** [link].

— Engineered a modular pipeline to utilize **Spark and Airflow** to generate training data with dynamic transformations for different data requirements.

— Created a pipeline to train a slack prediction (regression) model which replaced **3 weeks long interpolation method to 2-3 hours** of training and inference with **margin of error of 10 microseconds**.

— Identify which timing path needs optimization by casualty analysis which resulted in **45% less power consumption and 30% performance improvement** for ARM series chip-sets.

— Created chip congestion detection by aggregated complex flow traffic data into intuitive heat-maps.

TECHNICAL SKILLS

Machine Learning:	**Big Data:**	**Visualization:**
Pytorch, Tensorflow, Sklearn	Pyspark, Pandas, SQLAlchemy	Matplotlib, Seaborn, Plotly
Languages:	**Version Control:**	**Other:**
Python, JavaScript, C, C++, bash	Git, Perforce, CICD Pipelines	Airflow, FastAPI, MinIO, PySyft