

Pranab Islam

M: (917) 574-0680

pfi203@nyu.edu

github.com/pranabislam

linkedin.com/in/islampranab

Education

Sep 2020 – Dec 2022

New York University

New York, NY

Master of Science in Data Science | GPA: 3.88

Coursework: Deep Learning, Natural Language Understanding, Machine Learning, Big Data, Probability & Statistics, Optimization, Linear Algebra, A/B Testing

Sep 2015 – Jun 2019

The University of Chicago

Chicago, IL

Bachelor of Arts in Economics | GPA: 3.81

Skills

Languages: Python • SQL • JavaScript

ML Modeling: PyTorch • Hugging Face • scikit-learn • Pandas • Dask

Cloud / ML Ops: Airflow • Prefect • PySpark • AWS (SageMaker, Lambda) • GCP • Docker

Version Control / Other: Git • Linux / Bash • Looker • React • Next.js

Professional Experience

Sep 2022 – Dec 2022

Cash App | Machine Learning Modeler Intern

New York, NY

Search & Discovery Machine Learning Engineering Team

- Developed query intent model using XGBoost to classify user search queries
 - Created daily training / feature engineering pipeline from logging data using SQL and Airflow
 - Deployed model to run inference daily via Prefect for batch predictions and analysis. 95% ROC-AUC achieved with a precision-recall AUC of ~50% on 1.5 months of post-training data
 - Constructed two low-latency model approximators (using embeddings from matrix factorization) to deploy for real-time customer search but determined that approximators lacked sufficient performance after back-testing

Jun 2022 – Aug 2022

Roku | Data Science Intern

San Jose, CA

Core Analytics Team

- Constructed a last touch traffic source attribution model with full-service dashboard, allowing management to understand which marketing campaigns were performing on target and why
 - Executed project from end-to-end (data gathering, cleaning, model-building, data pipelining via Apache Airflow, visualizations / reporting, and future model improvement prototype creation)

Jul 2019 – Jul 2020

Mizuho Securities | Investment Banking Analyst

New York, NY

Financial Sponsors Group

Technical Projects

Sep 2022 – Present

Multimodal Contract Segmentation

In-progress research effort with goal to understand how using hierarchical document segmentations could improve state of the art ML system performance on various downstream legal NLP tasks; current (draft) open-source library repository: <https://tinyurl.com/docSegmentationDraft>

- Constructed a pipeline to programmatically label section titles in legal contracts which will be used to fine-tune an image segmentation transformer model to better segment legal contracts

Feb 2022 – May 2022

Analyzing Bagging Methods for Large Language Models

Natural language research project analyzing whether various bagged ensembles of large language models could outperform single language model baselines, holding model parameter count constant; project detail and results: <https://arxiv.org/abs/2207.09099>

- Developed an automated pipeline that fine-tuned large language models, created various bagged ensembles of them, and evaluated ensemble performance using [SuperGLUE benchmark](#)

Apr 2021 – May 2021

Music Recommender System

Implemented a recommender system with the Million Song Dataset using collaborative filtering via alternating least squares in PySpark; achieved a mean average precision of 30x the popularity baseline model (0.08 vs 0.0026); project detail and results: <https://tinyurl.com/musicRecSys>