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: 3.73

Coursework: Deep Learning, Natural Language Processing, Machine Learning, Big Data, Probability & Statistics, Optimization & Computational Linear Algebra, A/B Testing (Google Online Course)

Sep 2015 – Jun 2019	**The University of Chicago**	*Chicago, IL*

Bachelor of Arts in Economics | GPA: 3.81

Professional Experience

Jun 2021 – Present	**Algorand,** a leading blockchain and decentralized application protocol company	*Boston, MA*

Data Science Intern, Product Team

- Created the first weekly clustering system that grouped 10,000+ decentralized apps ("dApps") on Algorand using set similarity on segments of the dApps' code, creating a systematic way for Algorand to collaborate with dApp developers (due to company names and assets now being attached to clusters of previously unanalyzed dApps)
- Constructed 10+ metrics from existing data and designed the first iteration of event tracking metrics with seven separate event types for Algorand's developer documentation website in order to make data-driven design decisions based on the discovered user behaviors
    - Tools: Google Analytics API, Google Tag Manager, Python, Pandas, SQL

Feb 2021 – May 2021	**Global Association of Risk Professionals (GARP)**	*Jersey City, NJ*

Data Analyst Intern

- Developed a pipeline, written in Python, to scrape a large volume of financial pdf files, extract the tabular data and page information, and output the results where the data visualizations team could readily access the data

Jul 2019 – Jul 2020	**Mizuho Americas,** the investment banking arm of a top 20 global bank	*New York, NY*

Investment Banking Analyst, Financial Sponsors Coverage Group

- Conducted data-driven financial analysis on companies worth up to $6 billion, which included building 7-year financial models, credit risk analysis, regression analysis, and a multitude of qualitative analyses in order to execute or evaluate 20+ leveraged buyout and debt funding transactions on behalf Mizuho's clients
- Automated Mizuho's weekly leveraged finance market report to translate raw web data into a seven-page, client-ready presentation using VBA and Excel, saving two man-hours of labor per week

Technical Projects

Nov 2020 – Dec 2020	**NBA Underdog Matchup and Betting Analysis**

Using Python, analyzed how one could create profitable betting positions by focusing on significant underdog matchups; link to the project's report detailing the process and results here: https://tinyurl.com/nbaUnderdogs

- Created two classification models (random forest and gradient boosted trees) that made positive expected value bet decisions that outperformed the baseline models (logistic regression, support vector machine, and random guessing) with each model optimized differently to simulate different types of betting behavior; final classifier was an ensemble of the two tree-based classifiers and generated a 9% return on investment
- Extracted, transformed, and loaded the data into shared repositories; scraped 200,000+ lines of data with BeautifulSoup; reconfigured a basketball-reference.com API to pull statistics; cleaned, merged, and transformed data to create a finalized set of features to analyze

April 2021 – May 2021	**Music Recommender System**

Utilized ALS implicit collaborative filtering in PySpark to create a recommender system with the Million Song Dataset; link to the project's report detailing the process and results here: https://tinyurl.com/musicRecSys

- Final model had a mean average precision of 0.08 (30x improvement from the baseline popularity-based model)
- Created dimensionality-reduced visualizations of user factor embeddings using t-SNE and UMAP techniques
- Produced data processing pipeline using PySpark & HDFS and extensively tuned hyperparameters thereafter

Skills

Python (NumPy, Pandas, scikit-learn, PySpark, PyTorch, Matplotlib), SQL, MapReduce, Hadoop, Dask, Tableau, Google Cloud Platform, Google Analytics API, Linux, Bash, Excel, Visual Basic, Git, Financial Modeling