Toufiq Hossain

thoss0513@berkeley.edu | https://www.linkedin.com/in/toufiqhossain/ | https://github.com/thossain0513 | (510) 705-3062 | Berkeley, CA

EDUCATION

University of California, Berkeley **Expected Graduation: December 2024**
Master of Information and Data Science (MIDS) **GPA: 3.84**

University of California, Berkeley **August 2020 - May 2023**
Bachelor of Arts, Economics and Data Science, Concentration in Business and Industrial Analytics

SKILLS

- **Programming:** Python, Java, R, SQL, pandas, Keras, Tensorflow, NumPy, Javascript, PyTorch, Gensim, spaCy, NLTK, sci-kit learn, LLMs
- **Tools and Skills:** Git, Databricks, AWS Sagemaker, Tableau, Amazon Web Services (AWS), Google Cloud Platform (GCP), AWS S3, Snowflake, Flask, Spark, Heroku, FastAPI, OpenAI API, dbt, A/B Testing, MySQL, PostgreSQL, ETL, LLAMA, Docker, Kafka, SQL, NoSQL, MongoDB, Redis, Linux, Looker, Grafana, Kubernetes, BigQuery, MATLAB, Cloud Computing, DevOps, MLOps, CI/CD, Statistics, Agile methodologies, AWS Bedrock

WORK EXPERIENCE

Amaze **November 2022 – May 2023**
Data Scientist

- Trained a deep learning classification model using **Keras** and **Tensorflow** to gauge earning potential for onboarding sellers using data from **1M+** existing users, model increased sample product GMV by **5%** by allowing seller success team to enhance platform experience for predicted high-value sellers on signup, model achieved **85%** accuracy on testing set for revenue predictions, set up **Grafana** dashboard to monitor predictions
- Generated end-to-end data pipelines using **Apache Airflow** to design and aggregate key revenue data, sales data, and social media metrics on all users, accounting for over **$100M** in GMV
- Led team of four to develop, test, and deploy SpringScore, a machine learning driven prediction tool using the **scikit-learn** and **pandas** libraries, used **SQL** to aggregate user sales and social media data for training prototype model, served model predictions in real time using **AWS Sagemaker**
- Used **SQL** and **Tableau** to aggregate e-commerce related Key Performance Indicators (KPIs) for the launch of Youtube Merch-Shelf and TikTok Jump partnerships, created data visualizations to identify early successful sellers on social media platforms based on social media and merchandise sales metrics, which led to focused digital marketing initiatives on platform users

Spring **June 2022 – November 2022**
Data Science Intern

- Performed in-depth analysis of user journey, discovering critical points of dropoff in the signup, onboarding, and management funnels, reported findings directly to the executive team, saving the company **$1.44M** in annual costs
- Collaborated with product managers to redesign user-facing dashboard using a data-driven approach, used **A/B testing** methods, customer segmentation, and user behavior analysis to inform UI rebranding initiative on the seller side of the platform, redesigned dashboard generated over **60,000** social media integrations post-launch
- Built a robust reporting system on **Tableau**, leveraging statistical packages, optimized queries, and an interactive dashboard for sales, signup, and integration data, that has become a go-to resource for the executive team's day-to-day operations and long-term planning

Karbon **August 2021 – May 2022**
Data Consultant

- Used **dbt** command line and **Google BigQuery** on raw data tables containing **30M+** rows to build a data pipeline and automate **ETL** processes for the whole company, reducing manual workload for onboarding data scientists by **15%**
- Influenced business analytics, strategy, and product expansion through model ensembling and feature importance methods to draw product development insights in feedback and frequency of usage for performance analytics tools

PROJECTS

Ulta Beauty Customer Profiling Project | *Clustering, Machine Learning, Product Analytics*

- Developed and implemented a customer profiling model using **DBSCAN**, successfully clustering customers in a dataset for targeted business strategies, model received a silhouette score of **0.6**
- Communicated model and analytical findings and recommendations to senior leadership, enabling informed decision-making and personalized marketing efforts, empowering Ulta to develop targeted marketing strategies, improving customer online engagement by **10%**

Supervised GAN-BERT Model Paper on Ancient Greek Poetry | *PyTorch, Keras, Tensorflow, NLP*

- Co-authored a paper focused on fine-tuning a **BERT** model on Ancient Greek and extracted CLS tokens for GAN imitation
- Implemented a novel **BERT-GAN** model mimicking a generative model for authorship attribution in Archaic Greek texts, addressing challenges in small text corpora and linguistic obscurity, achieving improved performance over baseline models through sampling techniques for data augmentation, attained a **68%** accuracy rate on test set for authorship attribution on a 4-class classification task