Kunj Chetan Mehta

New Brunswick, NJ | +1 848-437-1589 | kunjcmehta@gmail.com | kunjmehta.github.io | linkedin.com/in/kunjmehta

EDUCATION

Master of Science in Computer Science | *Rutgers University, New Jersey* | **GPA** – 3.91/4 **Expected May 2023**

Computer Science Exchange Program | Princeton University, New Jersey **Sep 2022 – Dec 2022**

Bachelor of Technology in Computer Engineering | *Mumbai University, Mumbai, India* | **GPA** – 3.41/4 **Oct 2020**

Relevant Coursework – Data Warehousing and Mining, Distributed Cloud Computing, Digital Image Processing, Database Management System, Algorithms and Data Structures, AI, ML, NLP, Advanced Computer Vision

EXPERIENCE

Data Scientist Intern| Eluvio | *Berkeley, CA* **Jun 2022 – Aug 2022**

- Part of the machine learning team building the media meta-tagging framework for *media distribution on the blockchain*
- Engineered the logo detection and classification pipeline *from supervised to zero-shot learning paradig*m. Reduced the number of *false logo detections by 8 percentage points*
- Pioneered a NFT recommender system end-to-end. Led the creation of a near real-time ETL pipeline to ingest model-ready blockchain data for training. Deployed a test MVP of the recommender handling *3000 concurrent users* efficiently

Teaching Assistant | Rutgers University | *New Brunswick, NJ* **Sep 2021 – Present**

- Taught R, SQL and Amazon Redshift and graded weekly assignments and exams for *150 students across two courses* – "Data 101" and "Database Systems for Data Science"

Business Analyst| Quantiphi, Inc**.** | *Mumbai, India* **Oct 2020 – Aug 2021**

- Researched and presented highlights of the *three US pandemic stimulus* bills to internal stakeholders that informed Quantiphi's Public Sector business strategy
- Performed *market research on 200 organizations in the US Education industry* and came up with an effective go-to market strategy that *converted four cold leads*
- Presented solution deck to four leads showcasing how machine learning can be incorporated into leads' existing processes, *converting two*

Project Intern | Fractal Analytics | *Mumbai, India* **Jun 2019 – Jul 2019**

- Implemented the object classification phase of a project that analyzed consumer behavior at stores for a Fortune 500 FMCG company. Developed a model for classifying 50 product SKUs in the product range with *80% accuracy*
- Set up a data augmentation and ingestion pipeline for the classifier. Coded a script for scrapping images of representative products from e-commerce websites to augment data

PROJECTS

Food AI| **Multi-modal Representation Learning** *(Python, PyTorch, Hugging Face, seaborn)* **May 2022**

- Performed retrieval of food recipes given food images using crossmodal techniques and beat the baseline top-10 recall for recipe retrieval in the original im2recipe *paper by 20 percentage points* by improving the feature extraction pipeline
- Improved recipe retrieval further by projecting learnt features to a shared space using triplet loss trained neural networks and attained *median retrieval rank of 1 and top-10 recall of 82.49% for 1,000 random food images* [Repo link]

Movie Recommendation from Conversational Data | NLP *(Python, PyTorch, surprise, seaborn)* **May 2022**

- Built a movie recommendation system leveraging user conversations, critics data and domain adaptation techniques, which is a re-implementation of this paper
- Tuned hyperparameters for three CF approaches: KNN, SVD and SVDpp and GBRT *to obtain 3% improvement* in results
- Experimented with neural MF and obtained comparable results of *RMSE=1.232 and MAE=0.9569* [Repo link]

New York Taxi Fare Prediction | **Big Data** *(Python. pandas, matplotlib, PySpark, AWS EMR, AWS EC2)* **Oct 2020**

- Performed feature engineering to focus on trips to and from airports and across different boroughs of NYC
- Predicted taxi fares to a *RMSE score of 4.28* by training a Random Forest on the augmented data [Repo link]

FPL Teammaker | **Data Analysis** *(Python, NumPy, pandas, streamlit, matplotlib, PuLP)* **Sep 2020**

- Developed and deployed an application that performs exploratory data analysis on the English soccer Fantasy Premier League (FPL) game data to suggest an optimal team to be entered into the game
- *50+ monthly active users. Ranked top 2% in worldwide* ranking among 8.2 million players in the year 2020 [Repo link]

PUBLICATIONS

- "Simplification with the Transformer - Its Drawbacks"(*International Journal of CS and Engineering)* [pdf]
- "Abalone Age Prediction Problem: A Review" (*International Journal of Computer Applications*) [pdf]
- *Amassed 35,000+ views* on articles on Medium publications Towards Data Science and Towards AI [link]

TECHNICAL SKILLS

- **Development:** Python, Java, SQL, PySpark, R, C++, JavaScript, Android, MS Excel
- **Frameworks and Libraries:** NumPy, pandas, matplotlib, seaborn, scipy, scikit-learn, MLlib, streamlit, Pytorch, Deep Graph Library, NLTK, Pillow, OpenCV, Flask
- **Cloud and Engineering:** AWS Sagemaker, AWS EMR, AWS Lambda, Google Big Query, Docker, Apache Airflow, Kafka
- **Databases and Visualization:** MySQL, AWS Redshift, Amazon RDS, Tableau, Power BI, Looker
- **Machine Learning:** EDA, CNN, GNN, NLP, Time Series, Ensemble & Multimodal Learning, Recommender Systems
- **Certifications:** Cloud Practitioner (*AWS*), LookML Developer (*Looker*), Machine Learning Engineer (*Udacity*), Deep Learning Specialization (*Coursera*), Applied Data Science with Python Specialization (*Coursera*) [link]