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** – 3.95/4 **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**
Courses – Mac hine Learning, Computer Vision, AI, NLP, Data Mining, Cloud Computing, Database Management, Data Structures

TECHNICAL SKILLS

- **Development:** Python, Java, SQL, PySpark, R, JavaScript, Android, MS Excel
- **Frameworks and Libraries:** Pytorch, NumPy, pandas, matplotlib, seaborn, scikit-learn, streamlit, NLTK, OpenCV, Flask
- **Cloud and Engineering:** AWS Sagemaker, AWS EMR, Docker, Apache Airflow, Kafka
- **Databases and Visualization:** MySQL, AWS Redshift, Amazon RDS, Tableau, Looker
- **Certifications:** Cloud Practitioner (*AWS*), LookML Developer (*Looker*), Machine Learning Engineer (*Udacity*) [link]

EXPERIENCE

Software Engineer| Barclays | *New York, NY* **Aug 2023 – Present**
- Building a pipeline to ingest trades data for processing from MSSQL as part of the FRTB RFET regulatory project
- Increased the test coverage by writing unit tests using XUnit and Pytest from 20% to 80%

Data Scientist Intern| Eluvio | *Berkeley, CA* **Jun 2022 – Aug 2022**
- Worked on improving the media meta-tagging framework for movie distribution on the blockchain
- **Logo Classification:** Designed and restructured the logo classification pipeline in PyTorch *from supervised to zero-shot learning paradig*m, elevating the model *classification capacity five times to 500 classes and achieving 67% accuracy*
- Conducted 25+ A/B testing experiments using diverse base classification models and super-resolution blocks to select a champion model that reduced the *false positive rate of zero-shot classification from 10% to 2%*
- **NFT Recommender:** Pioneered and developed an *end-to-end ML pipeline for an NFT recommender system* and deployed an alpha version of the system using Docker *capable of handling 3,000 concurrent requests*
- Constructed a near-real-time cron automated ETL pipeline by extracting blockchain and NFT metadata from Cockroach DB using SQL and collaborated with SMEs to manually formulate rules for transforming the extracted data

Teaching Assistant | Rutgers University | *New Brunswick, NJ* **Sep 2021 – May 2023**
- Taught R, SQL and Amazon Redshift, and guided understanding of statistical concepts for *200 students across two courses* – "Data Science 101" and "Database Systems for Data Science"

Business Data Analyst| Quantiphi, Inc**.** | *Mumbai, India* **Oct 2020 – Aug 2021**
- **Data Analyst:** Analyzed, reported and presented the US Public Sector vertical's quarterly revenue figures using Looker
- **Business Analyst:** Researched, consolidated 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 formulated a go-to market strategy that *generated two Sales Qualified Leads*
- Spearheaded singlehandedly the creation of an in-house repository for monitoring research in ML which was utilized *by 230 people in the organization including founders*

Project Intern | Fractal Analytics | *Mumbai, India* **Jun 2019 – Jul 2019**
- Built a product classification model for a Fortune 500 FMCG company that classified 50 *SKUs with 80% accuracy*
- Coded a script for scraping 25,000 representative images of SKUs using Selenium from e-commerce websites to augment training data

PROJECTS

Ensembling YOLO for Detection| **Computer Vision** (*Python, PyTorch*) **Dec 2022**
- Developed algorithms for building YOLO bagging and boosting ensembles
- *Achieved an average precision of 87.5% on the Flickr-32* dataset using a generic logo detection system comprising *two instances of boosted YOLO models* [Report link]

Food AI| **Multi-modal Representation Learning** (*Python, PyTorch, Hugging Face, seaborn*) **May 2022**
- Built a system for retrieval of food recipes using images of corresponding food items
- Surpassed the CCA baseline top-10 recall for recipe retrieval in the original im2recipe *paper by 20 percentage points* using ResNet and BERT feature extractors and introducing cross-modality through a shared embedding layer [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 trip fares to and from airports and across different boroughs of NYC
- Predicted taxi fares to a *margin of +/- 2 (RMSE 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**
- Designed, and deployed on Streamlit, an algorithm 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 each week
- Consistently observing *50+ monthly active users. Recommended team 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 40,000+ views* on articles on Medium publications Towards Data Science and Towards AI [link]