Arth Talati

Philadelphia, PA | P:267206-8721 | atalati@seas.upenn.edu | Portfolio arthtalati.github.io | Chef Instagram @asli_penn_masala

EDUCATION

University of Pennsylvania *Philadelphia, PA*

Master of Science in Data Science - GPA: 3.78/4.0 *May 2021*

Best Project Design **Award** for Football Freak project – Databases and Info Sci Course

Relevant Coursework: Big Data Analytics, Machine Learning, Databases and Info Systems, Internet & Web Systems, Computational Linguistics, Statistics for Data Science, Data Science for Public Policy

SKILLS

Programming Languages: Python, Java, SQL, SparkSQL, PySpark, R, Scala, MATLAB, SAS, Unix, Tableau

Big Data & Machine Learning: Tableau, Looker Apache Spark, Scikit-learn, Pandas, NumPy, Cassandra, Hadoop, TensorFlow, Pytorch, PostgreSQL, Mongo DB, Neo4j, Matplotlib, NLTK, Seaborn, Amazon Web Services(EC2, EMR, S3, DynamoDB), Git

Data Science Technologies: ETL, Database Management, Statistics, Anomaly Detection, Factor Analysis, Clustering, Time Series Analysis, Data Structure and Algorithms, Natural Language Processing, Reinforcement Learning, Hypothesis testing, A/B testing, Google Analytics

WORK EXPERIENCE

FOX Entertainment | Technical Consultant *Philadelphia | Jan 2021 – Present*

- Conducted in-depth evaluation of marketing campaign data to help derive insights to improve audience reach, increase ROI of future promo campaigns and suggested recommendations for future initiatives.
- Developed and refined customer segmentation models to analyse effects of social media campaigns supporting retention and long term promotional campaign profitability.

University of Pennsylvania | Teaching Assistant *Philadelphia | Aug 2020 – Dec 2020*

- Collaborated with instructor to design HW assignments, answer 300+ student questions and mentored 8 student project groups.
- Hosted HW datasets in S3 buckets and developed a guide to spin up an EMR cluster on AWS for students to work with Spark ML.

CSpace | Data Science Intern *Boston | May 2020 – Aug 2020*

- Derived a quantitative approach to Customer Segmentation based on levels of customer centricity, measured with CSpace survey features.
- Performed feedback sentiment analysis and K-means and hierarchical clustering to segment companies (300+) into cohorts w.r.t their public opinion. Co-Authored a white-paper soon to be published in *Harvard Business Review*.
- Modelled correlation of financial performance of companies with their perceived image in different consumer segments.
- Improved the built inhouse analytics capabilities to enhance consulting framework and developed dashboards for CSpace Customer demos

Wharton Consumer Analytics | Fraud Research Analyst *Philadelphia | Jan 2020 – May 2020*

- Assisted the Procurement Department to identify high-risk transactions as part of Purchasing Services High Risk Project.
- Designed anomaly detection algo using Local Outlier Factor and Isolation Forests algos for dataset of 1.8 million Oracle Financial invoices.
- Automated the manual transaction flagging process by anomaly detection algorithm resulting in 20% increase in fraud detection accuracy.
- Presented model analysis to both technical and non-technical (Procurement Department) teams through high-level executive summaries.

NIT Surat, Electronics design & Machine Learning Lab | Machine Learning Researcher *Surat, India | Aug 2018 – May 2019*

- Designed an ASIC (Application Specific IC) prototype to be embedded on handheld devices to detect heart diseases by ECG machine learning algorithm (SVM and Naive Bayes).
- Modelled using extracted features to give disease predictions with 93% accuracy of a person having Premature Ventricular Contraction.

SELECT PROJECTS

Football Freak (MySQL DB | AWS | Web Scraping | JavaScript)

- Created an app using data scraped from *sofifa.com and* hosted on AWS Lambda serverless instance with a relational model in 3NF format
- Deployed on AWS RDS instance with appropriate indexes and an optimal query plan to decrease the query execution time by 10x.

SEPTA On-Time Performance Analysis (Spark MLlib | Web Crawling | Python)

- Customized TWINT API to crawl tweets from @SEPTA Twitter handle and validated the claim by performing regression between actual delays and the ones claimed in tweets.
- Scrapped the weather data and streamlined Spark pipeline to train a random forest regressor model to predict the delays in arrival time.

US Flight Delay Analysis (RShiny | Leaflet | Plotly | R)

- Developed an interactive dashboard (R Shiny) by analyzing 5.8 Million flight operation details to demonstrate the key functionality of comparing different airlines departure/ arrival timings for given airport based on the day-of-week, time-of-day, taxi times and other features.
- Performed statistical analysis for US flight delay data in R to build predictive models for flight delay patterns.

Philadelphia Adverse Childhood Experiences Dashboard (RShiny | SQL)

- Performed survey statistical analysis using data from the 2013 Philadelphia Expanded Adverse Childhood Experience (ACE) Survey to test impacts of perceived neighborhood trust and safety during childhood.
- Wrangled survey log data and built pipeline for front-end dashboard (R Shiny) using SQL and hosted the webapp on RShiny server.

Search Engine (Java | AWS)

- Built a distributed, scalable, cloud-based search engine which consists of a web crawler, as indexer, a ranking engine and a web front end. Hosted it on AWS EMR as part of a course final project.