Arth Talati

Philadelphia, PA | P: 267206-8721 | svnit.arth@gmail.com| https://arthtalati.github.io

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
<u>**Relevant Coursework:**</u> Big Data Analytics, Machine Learning, Internet & Web Systems, Computational Linguistics, Databases, Statistics for Data Science, Data Science for Public Policy

SKILLS

Programming Languages: Python, Java, SQL, HTML, JavaScript, Tableau, SparkSQL, R, Excel, Google Sheets
Big Data & Machine Learning: Apache Spark, Scikit-learn, Pandas, NumPy, TensorFlow, Hadoop, Pytorch, PostgreSQL, Redshift, Neo4j, Matplotlib, NLTK, Seaborn, Amazon Web Services (AWS), Git, Looker, ActiveMQ
Data Science Technologies: ETL, Database Management, Statistics, Anomaly Detection, Clustering, Time Series Analysis, Natural Language Processing, Reinforcement Learning, Hypothesis testing, A/B Testing, Google Analytics

WORK EXPERIENCE

Merck & Co., Inc | Sr. Specialist, Data Science   *Philadelphia | July 2021 – Present*
- Built a XGBoost regression model to identify drivers for HPV Vaccine Hesitancy in France and EU Region.
- Brainstormed with Sentiment tracker team to improve their dashboard and also worked on NASH use case ideation.
- Designed the Code Versioning template for Bitbucket to help formalise structure of upcoming data science projects.
- Analysing markets to prioritize for V116 candidate with the Pneumococcal marketing team.

Wharton Consumer Analytics & FOX Broadcasting Company | Technical Consultant   *Philadelphia | Jan 2021 – May 2021*
- Conducted in-depth evaluation of marketing campaign thru Market Mix Modelling using viewership 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 (CIS 545 Big Data Analytics)   *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 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.

Penn Medicine | Research Intern (Anaesthesiology and Critical Care Department)   *Philadelphia | Nov 2019 – Jan 2020*
- Analyzed EEG signal data from anesthetized subjects, to know more about their responsivy and side-effects of drugs.
- Extracted quantifiable information form raw data using MATLAB after doing preliminary signal processing and spectrum analysis.

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.

Deep Learning for Authorship Identification (Deep Learning | NLP | Python)
- Executed multi-class classification for 50 news article authors using LSTM, Bi-LSTM, GRU (RNNs) neural networks, improved the classification accuracy by 20% over the baseline LSTM model using SVM and 40 different Stylometry features.

Philadelphia Adverse Childhood Experiences Dashboard (RShiny | SQL | ggplot2)
- Performed survey statistical analysis using data from the 2013 Philadelphia Expanded Adverse Childhood Experience (ACE) Survey to test impacts of perceived neighbourhood 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.

US Flight Delay Analysis (RShiny | Leaflet | Plotly | R)
- Analysed the dataset of 5.8 Million flights delays and cancellations by US DOT to find the effects of delays on domestic flight operation.
- Developed an interactive dashboard (R Shiny) 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