Arth Talati

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

EDUCATION

University of Pennsylvania *Philadelphia, PA*
Master of Science in Data Science *May 2021*
- Cumulative GPA: 3.78/4.0
- Best Project Design Award – Databases and Info Sci Course
- Relevant Coursework: 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, R
Big Data & Machine Learning: Spark, Pandas, NumPy, Hadoop, Mongo DB, Neo4j, MATLAB
Data Science Technologies: ETL, Data science pipeline (cleansing, wrangling, visualization, modeling, interpretation), Statistics, Experimental design, Hypothesis testing, Git

EXPERIENCE

School of Engineering & Applied Sciences *Philadelphia, PA*
Teaching Assistant *Aug 2020 – Dec 2020*
- Collaborated with instructor to design HW assignments, answer 300+ student questions and mentored 8 project groups.
- Helped in hosting HW data in S3 instances and developed a guide to spin up EMR cluster on AWS.

CSpace *Boston, MA*
Data Science Intern *May 2020 – Aug 2020*
- Lead a team of 5 in 3 locations (Philadelphia, San Francisco and Miami) as a Tech lead, of entry-level analysts and collaborated with data science and analytics team at CSpace.
- Overhauled Customer Maturity Model for customer survey data of 300 publicly traded companies.
- Performed feedback sentiment analysis and K-means clustering to segment companies into cohorts w.r.t their public opinion.
- Designed Python dashboard to analyse trends between different financial indicators and clustered companies.

Penn Data Science Group *Philadelphia, PA*
Research Analyst *Jan 2020 – May 2020*
- Assisted the Procurement Department to identify high-risk transactions as part of Purchasing Services High Risk Project.
- Designed custom anomaly detection algorithm using Local Outlier Factor and Isolation Forests algorithms for dataset of 1.8 million Oracle Financial invoices.

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

SELECT PROJECTS

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

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

FPGA based Machine Learning for Cardiac Diseases Detection | Senior year Thesis-Project
- Designed an ASIC prototype to be embedded on handheld devices to detect diseases by ECG of a person's using machine learning algorithms. Modelled using extracted features to give disease predictions with 93% accuracy of a person having Premature Ventricular Contraction.

SEPTA On-Time Performance Analysis
- Customized TWINT API to get 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.