Arth Talati

linkedin.com/in/arthtalati || github.com/arthtalati

atalati@seas.upenn.edu

(267)206-8721

Education

- **University of Pennsylvania** — Philadelphia, PA
  Master of Science in Engineering: Data Science; GPA: 3.78/4 — *May 2021*
- **National Institute of Technology,** — Surat, India
  Bachelor of Technology: Electronics Engineering with Minor in Data Science; GPA: 8.52/10 — *May 2019*

Relevant Coursework Big Data Analytics, Statistics for Data Science, Machine Learning, Internet of Things and Edge Computing, Data Science for Public policy, Database, Computational Linguistics, Computer Architecture

Skills

- **Languages**: SQL, Python, Java, R
- **Technologies and Frameworks**: PostgreSQL, MySQL, MATLAB, Neo4j, Spark, Pandas, R Shiny, Jupyter, AWS, Git, regex, XPath

Experience

Teaching Assistant — Aug 2020 - Ongoing

School of Engineering and Applied Sciences

- TA for CIS 545 Big Data Analytics course. Helped in designing home works and hosting big data on S3 instances on AWS. Developed a guide to spin up EMR cluster using AWS CLI. Mentored 8 student groups for their final projects.
- TA for GAFL 531 Statistics for Public Policy class. Helped students during Office hours and assisted in checking home works.

Data Science Intern — May 2020 - Aug 2020

CSpace

- Overhauled Customer Maturity Model for customer survey data of 300 publicly traded companies. Performed K-means clustering to segment companies into separate cohorts w.r.t their public opinion. Performed sentiment analysis on customer feedback for companies.
- Examined financial trends by analyzing different financial indicators (like Cap-Ex, Long-Term Debt, etc) to validate clustering model and provided detailed analyses for different industries (like Retail, Healthcare, etc).

Research Analyst — Jan 2020 - May 2020

Penn Data Science Group

- Assisted the Procurement Department at university to identify high-risk transactions as part of Purchasing Services High Risk Procurement Project.
- Designed custom anomaly detection algorithm from Local Outlier Factor and Isolation Forests algorithm for dataset of 1.8 million Oracle Financials invoices.

Select Projects

Football Freak

- Created a soccer app using data scraped from sofifa.com hosted on AWS Lamba 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.
- Developed the backend for an API using Express.js using couple of middleware to abstract away the need to manually connect to the client database.

Deep Learning for Authorship Identification

- Executed multi-class classification for 50 news article authors using LSTM, Bi-LSTM, GRU neural networks at sentence and article levels for corpora of articles.
- Improved the classification accuracy by 20% over the baseline LSTM model using SVM and 40 different Stylometry features.

SEPTA On-Time Performance Analysis

- Investigated SEPTA regional rail's claim of 91% of On-Time-Performance. Utilized 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 of Philadelphia to improve the prediction model. Streamlined Spark pipeline to train a random forest regressor model to predict the delays in arrival time.

US Flights Delay Analysis

- Analyzed the dataset of 5.8 Million flights delays and cancellations by US Department of Transportation to find the effects of delays on domestic flight operation over 14000 airports.
- 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.