

Raj Shah

<https://www.linkedin.com/in/raj-v-shah> | 904-521-5744 | rajshah@andrew.cmu.edu

WORK AND INTERNSHIP EXPERIENCE

Allianz Life – Minneapolis, MN

May 2024 – Present

Data Scientist Intern

- Reduced compliance expenses \$20,000 annually by automating legal doc prioritizations using ML models.
- Developed tooling comparing ML (BERT, LGBM, SkLearn) across 100+ permutations to achieve 84% AUC-PR.
- Automated ETL Pipeline to consume data, run versioned models on altered rows, and save results to SQL.
- Migrated fraud projects worth \$4,000,000+ quarterly from legacy databases using Python, SQL, and RStudio.
- Identified potential lapses worth \$1,000,000+ using Cox's Penalized Proportional Hazard Model in SkSurv.

Ray Analytics (Startup) – Washington, DC

Dec 2022 – July 2023

Founder/Chief Technology Officer

- Developed SaaS Platform capable of on-demand political discourse analysis and influence identification.
- Trained and deployed Random Forest (94% ROC-AUC) to classify users based on tweet patterns and metrics.
- Utilized AWS API Gateway, Lambda, and React to develop automated backend system with live analysis.
- Developed interactive sentiment analysis dashboard using D3.JS and React used by 20+ customers.

Plainspoken Digital (Startup) – Washington, DC

Mar 2022 – Dec 2022

Senior Software Engineer/Project Lead

- Implemented digital transformation from legacy system to Palantir Foundry, increasing client op efficiency 2x.
- Lead Agile processes for a development team of 7 and managed intra team project management.
- Developed automated ETL processes using cron PySpark jobs, merging client cloud and local data In Foundry.

Amazon Web Services (AWS) – Crystal City, VA

Oct 2020 – Mar 2022

Software Engineer II – L5

- Lead development and design on S3 Encryption by Default, bolstering security for billions of daily requests.
- Saved customer segment 98%+ of daily costs through implementation and testing of S3 Bucket Keys
- Utilized design patterns and unit testing when deploying Java on S3's Apache Tomcat Webserver.
- Managed DevOps pipelines for AWS S3 (code with 50+ teams), reducing feature deployment by several days.

Data Science for Social Good – Jacksonville, FL

May 2020 – Aug 2022

Data Science Fellow

- Lead a team of three to analyze data for 6000+ adult learners being helped by literacy services in Florida.
- Utilized Z Scoring for score comparisons, Gower Method for clustering demographics, and causal analysis of feature importance to identify the leading cause of attrition rates for adult learners in Northeast Florida.

Booz Allen Hamilton – Washington, DC

June 2019 – Aug 2019

Data Science Consultant Intern

- Developed with a team of 5 an application capable of discerning between AI GAN media and real images.
- Programmed a Convolutional Neural Network with TensorFlow with 88% Precision, utilizing transfer learning.

EDUCATION

Carnegie Mellon University – Pittsburgh, PA

Master of Science, Data Science and Public Policy

Graduating May 2025

- **GPA:** 3.8
- **Relevant Courses:** Immediate Econ Analysis, Data Analysis for Organizational Impact, Machine Learning with Python, Applied Econometrics, Telling Stories with Data, Optimization, Introduction to Deep Learning, Voting Disenfranchisement, Decision Analytics in Business Policy, Geographic Information Systems

University of Florida – Gainesville, FL

Bachelor of Science, Computer Science, Minor in Mathematics

Graduated May 2020

- **GPA:** 3.2
- **Relevant Courses:** Differential Equations, Calculus 3, Linear Algebra, Digital Logic, Algorithmic Design, Numerical Analysis, Data Structures, Digital Logic, Introduction to Data Science, Database Systems

PROJECTS

Introduction to Deep Learning

Aug 2024 – Dec 2024

Achieving Group and Individual Fairness in Federated Learning

- Implemented novel approach of improving Individual Fairness in Federated Instances using Flower framework.
- Trained Neural Network to fairly classify recidivism cases from COMPAS dataset without reduction in Accuracy.

MyTorch

- Utilized NumPy and base Python to implement a custom Neural Network package with Layers and Activations.
- Derived code and derivatives to implement Pooling, Loss Functions, RNN Cells, and self-attention.

Decision Analytics and Business Policy

Aug 2024 – Dec 2024

Fund the Vote: Reducing Voter Wait Times using Predict-then-Optimize

- Developed framework to predict local voting turnout and then optimize precinct resources to reduce wait time.
- Utilized Census data, GeoPandas, and LassoCV with a hierarchical Regression model to predict Voter Turnout.
- Developed Optimization Model using Gurobi to calculate voter wait times and effect on marginalized voters.

Geographic Information Systems

Aug 2024 – Dec 2024

Changes in Industry in Pittsburgh

- Created an ArcGIS Dashboard showing changes in occupation makeup across several years using Selectors.
- Performed Multivariate Clustering across all years to identify + visualize patterns across Pittsburgh precincts.

Telling Stories with Data

Mar 2024 – May 2024

Disproving the NBA Contract Year Phenomenon

- Utilized Scrollama, D3.JS, and React to compare player performance metrics in relation to free agency status.
- Scraped several datasets and feature engineered data using statsmodel to compare player distributions.

LEADERSHIP AND INVOLVEMENT

SwampHacks (QuikVote) – University of Florida

May 2020

- Developed a website utilizing Bayesian predictions through an ML model in the backend to provide a “fast pass” voting system to users to reduce the disenfranchisement of minority voters.
- Awarded InfoTech’s Best Hack using a Public Dataset.

Solar Gators – University of Florida

Dec 2016 – June 2018

- Spearheaded team of 40 novice Engineers to create a competitive solar car with limited time before competition.
- Represented the organization when communicating with the University for 10,000+ dollars of funding.
- Machined and fabricated several core parts of the car including the C-Drivetrain and the Chassis Hull.

Gator Motorsports – University of Florida

July 2016 – June 2017

- Lead a team of 6 to envision marketing plan and financial investment strategy to present Formula racecar to judges.
- Created presentation in PowerPoint and Photoshop, winning 6th place in international competition.

SKILLS

- Proficient in AWS, experience with Azure and GCP. Knowledge in Python, Java, React, C++, SQL, and GraphQL.
- Experience with RESTful API, PostMan, OpenGL, JavaScript, OpenCV, JSON, and MIPS
- Utilized technologies such as JIRA, Excel Scripting, Google Apps Scripts, Microsoft Office, and Mac iWork Suite.