Rahul Pande

github.com/rahul-pande | linkedin.com/in/pande-rahul

Worcester, MA 01609
rspande@wpi.edu | 508-333-9986

SUMMARY

Data Scientist with in-depth understanding of machine learning and statistics, bringing 3+ years of data science, analytics work experience and strong programming skills

EXPERIENCE

- **Genomic Data Scientist - Sanofi** (Contingent) Framingham, MA | August 2020 - Present
  - Developed single-cell RNA-Seq analytic pipelines using R, Python and AWS for rapid prototyping and easy collaboration with other functional groups
  - Integrated and analyzed omics data from postmortem human subjects to understand underlying mechanism of neurodegenerative diseases
- **Data Scientist Intern - National Fire Protection Association** Quincy, MA | June 2019 - December 2019
  - **Software Engineering:** Built and deployed in-house geocoding/reverse-geocoding APIs using Flask, Docker, Lucene and libpostal- average throughput of 20 requests/second
  - **Data Engineering:** Operationalized batch geocoding with Apache Spark to process 150+ million addresses, while achieving 70% decrease in execution time and median distance <1 mile from Google's Geocoding API
- **Product Analyst - Vuclip India Pvt. Ltd.** Pune, India | April 2016 - June 2018
  - **Complaints Prediction:** Developed a Random Forest model to predict if a user will complain with cross validation recall of 0.79; preemptively decreased complaints from 11% to 7%, avoiding SLA violations and penalties
  - **Data Engineering:** Authored numerous Apache Airflow data pipelines to replace cron scripts which significantly reduced database locks, failures and execution time; Improved workload management on multiple servers

RESEARCH

- Pande*, Teeple*, et al., bioRxiv 2021, "Lysosomal-Immune Axis Is Associated with COVID 19 Disease Severity: Insights from Patient Single Cell Data" https://doi.org/10.1101/2021.01.27.428394

SKILLS

Programming: Python, R, SQL, Java
Visualization: Tableau, Power BI, Spotfire, Datastudio
Data: Hadoop, Spark, Redshift, MongoDB, Hive
Developer Tools: AWS, GCP, Github, Docker, JIRA
Frameworks: scikit-learn, pandas, Tensorflow, Keras, PyTorch, Seurat, Terra, Flask, Bioconductor, Bioconda, slurm
Algorithms: Naive Bayes, Regression, SVM, Decision Trees, KNN, Survival Analysis, ETL, A/B Testing

EDUCATION

- **Worcester Polytechnic Institute (WPI)** | Worcester, MA May 2020
  Master of Science, Data Science GPA: 3.9/4.0

PROJECTS

- **Graduate Research - Takeda Pharmaceutical** Cambridge, MA | January 2019 - May 2019
  - Large scale deep learning on UK Biobank actigraphy and phenotype dataset (500K subjects) using Tensorflow
  - Achieved f1 score of .78 and 34.8 rmse in predicting insomnia and age respectively from the actigraphy data using multitask learning with 1D Convolutional Neural Network (CNN) and Long-Short Term Memory (LSTM)
- **Music Generation with Neural Networks - WPI** Worcester, MA | April 2019
  - Synthesized music sequences (MIDI notes) with deep learning, which outperformed $\sqrt{n}$ randomized real sequences, through use of GAN; Implemented using Tensorflow and Keras

SOCIAL PARTICIPATION AND ACHIEVEMENTS

- Helped teach a MATLAB introductory course (Fractals, Conway's Game of Life, etc) to high school students during Frontiers Summer Outreach program at WPI in 2019
- Developed the official Android app for VIT college, Pune; Second rank in on-spot android app development hackathon