Rohit Venkat

Data Scientist + Engineer

As a biomedical researcher turned data scientist, I leverage my quantitative research background to tackle complex problems and deliver novel insights to stakeholders. I bring a 7+ year proven track record of being a detail-oriented problem solver and a team player.

Nashville, TN
615-403-7340
rohitrvenkat.github.io
rohitrvenkat@gmail.com
rohitrvenkat
rohit-venkat

SKILLS

Python
pandas, numpy, statsmodels, scikit-learn, imbalanced-learn, umap-learn, PySpark, SQLAlchemy, geopandas, Folium, pydeck, matplotlib, seaborn

R
tidyverse, Shiny, sf, leaflet, mapdeck, tidycensus, Seurat, Bioconductor

Databases
PostgreSQL, SQLite, Microsoft SQL Server, Neo4j, MongoDB

Machine Learning
Linear/logistic regression, decision trees, neural networks, ensemble models, dimensionality reduction, clustering algorithms

Natural Language Processing
NLTK, spaCy, SentenceTransformers

Web Scraping
Beautiful Soup

Cloud / Distributed Computing
AWS, Azure, Hadoop, Spark, Slurm

EDUCATION

Nashville Software School
Data Science Apprentice

Vanderbilt University
Master of Science Computational Immunology

Washington University in St. Louis
Bachelor of Arts Biology

DATA SCIENCE EXPERIENCE

Nashville Software School
Data Science Apprentice — 2021 – 2022

Intensive bootcamp focusing on data science fundamentals and problem solving. Leveraged real-world datasets to develop analyses, data tools, and models to inform community business partners' strategic decision-making.

- **Bernstein Private Wealth Complexity Analysis:** Performed predictive modeling of client investment behavior for AllianceBernstein's Private Wealth Management business and identified underperforming investment products and services
- **U.S. Congress Twitter Analysis:** Analyzed 3.5 million tweets by U.S. Congress members using PySpark and NLP techniques to explore polarizing events and differences in topic engagement between Democrats and Republicans
- **SEC Filings Automated Data Extraction:** Applied NLP techniques using Python to automatically extract relevant share repurchase information from 10-K SEC filings
- **Truck Derate Prediction Model:** Trained a machine learning algorithm on vehicle on-board diagnostic data using Python to predict an upcoming truck derate in order to minimize fleet disruptions and affiliated costs
- **Rocket League Kaggle Competition:** Designed the highest-scoring machine learning classification algorithm in our class Kaggle competition for predicting Rocket League player rank based on in-game match statistics
- **Wordle SQL Algorithm Solver:** Developed a SQL algorithm for Wordle that solves all 2,315 words in under 16 seconds with an average of 3.76 guesses per word
- **VUMC Referral Network Analysis:** Analyzed Medicare data using SQL, Python, and R to identify provider referral patterns and made recommendations on how to increase Vanderbilt University Medical Center's patient volume
- **Evaluating Traffic Accident Risk for Nashville Roadways:** Created a Shiny app that probes Nashville traffic trends and visualizes the highest-risk Nashville roadways based on 234,640 reported traffic accidents between 2014–2021
- **Hauser Jones & Sas Bank Auditing Tool:** Created a bank auditing dashboard using R Shiny to help auditors compare a loan institution's lending practices to peer institutions and ensure regulatory compliance

Vanderbilt University
Graduate Research Assistant — 2015 – 2020

- Wrote a novel genomic data pipeline using R to process, analyze, and interpret high-dimensional single-cell sequencing data
- Identified epitope-specific broadly neutralizing HIV antibodies from an HIV-infected donor sample using the pipeline
- Utilized statistical and machine learning techniques including dimensionality reduction, clustering, and hypothesis testing to determine the cellular identities of individual cells and genes that are differentially expressed between clusters of cells
- Adapted data analysis pipelines to high-performance computing environments, enabling parallel processing of hundreds of genomic datasets

Research Assistant — 2013 – 2015

- Curated and performed statistical analyses on gene expression and genetic variation data using R, Bash scripts, and Linux command line tools
- Wrote Excel user-defined functions using VBA in order to analyze data generated by experiments