Vishwesh Srinivasan

vishweshsrini@gmail.com | 781-921-4385 | LinkedIn | GitHub | Tableau

EDUCATION

Tufts University, MA, USA | M.S. in Data Analytics | GPA: 3.96 **09/2022 – 05/2024**

Coursework: Machine Learning, Big Data, Data Science for Urban Sustainability, Database Design & SQL, Adv. Statistics I & II
Teaching Assistant: Foundation of Data Analytics (DATA200) & Introduction to Data Analytics (DATA100)

National Institute of Technology Warangal, India | B.Tech in Mechanical Engineering **08/2016 – 08/2020**

EXPERIENCE

Paragon Corporation, Remote, USA | Data Analyst **01/2025 – Present**

- Optimized the ETL process for payroll data across 25+ clients by streamlining Python and SQL scripts to extract data from Excel/PDF files, standardize formats, and load into PostgreSQL databases.

Success Academy Charter Schools, NY, USA | Data Analyst **07/2024 – 01/2025**

- Reduced feedback turnaround time by 85%, enabling faster data-driven teaching adjustments and improving student outcomes by automating academic reporting workflows that integrated data from multiple sources in Google Sheets.

Data Intensive Studies Center, Tufts University, MA, USA | Research Intern (Poster) **01/2024 – 05/2024**

- Developed a novel approach to detect rare tumor cell clusters in blood by training a random forest classifier using a correlation-based feature-engineered light scattering data (Accuracy, Precision, TPR, TNR: ~90%).

JPMorganChase, DE, USA | Data Science Intern, Credit Risk Modeling **06/2023 – 09/2023**

- Developed two XGBoost classifiers (with and without additional guarantors' indicators) using a dataset of 1M+ records for small business credit decision-making and analyzed the models' performance across different data segments.
- Leveraged the analysis to propose a strategy reducing the documentation and approval time by 30% (est.) for applications with less than $250K exposure by eliminating the need for additional guarantors' information.
- Facilitated the migration of PySpark-based model development from on-premise infrastructure to the AWS ecosystem to reduce costs by collaborating with the infrastructure team and establishing best practices.

Citigroup, India | Software Developer **08/2020 – 07/2022**

- Simplified the personal loan data processing workflows serving 100K+ customers by re-engineering statement generation and payment application logic, enhancing processing accuracy and efficiency.
- Automated report processing (saving 10+ hrs/week of manual effort) by developing a VBA system that converts text reports into dashboards, delivering insights on product metrics like new account activity, balances, and transaction volumes.

Language Technologies Research Center, IIIT Hyderabad, India | Research Intern (Repository) **05/2019 – 07/2019**

- Reduced the perplexity by 16% compared to the SOTA sentence simplification model by training a Seq2Seq model using the WIKISPLIT dataset with reward (BLEU score) augmented maximum likelihood objective function.

Reliance Jio Infocomm Limited, India | Data Science Intern **05/2018 – 07/2018**

- Boosted the search efficiency of an internal platform used for resolving network coverage issues by building related search (NCD algorithm), auto-complete (N-Grams), and recommendation (TF-IDF & cosine similarity) features.

SPI Cinemas Private Limited, India | Data Science Intern **11/2017 – 12/2017**

- Decreased the turnover rate by 30% (est.) by pioneering a data-driven hiring strategy and developing a predictive model (87% accuracy) to predict the likelihood of frontline staff leaving within the first three months.

PROJECTS

Spatial analysis of air pollution exposure and walkability index in Greater Boston (Repository)

- Identified a positive correlation between exposure to air pollution and the walkability index by analyzing the visualizations (using GeoPandas) and the results of spatial regression models implemented using PySAL.

Gentrification study of New York and Los Angeles metropolitan areas (Repository)

- Implemented classification models with SMOTE techniques to predict the likelihood of census tract gentrification (2000 – 2010) using the US census data, identifying key predictors across demographic, housing, and economic factors.

TECHNICAL SKILLS

Programming Languages: Python, R, SQL, MATLAB, C++

Competencies: Linear & Logistic Regression, Tree-Based Methods, Ensemble Learning, Neural Networks, CNNs, Text & XML Processing, Word Embeddings, LSTM, Hypothesis Testing, A/B Testing, SHAP, Partial Dependence Plots

Tools/Libraries: GCP (BigQuery), AWS (S3, SageMaker, EMR), Tableau, Power BI, R Shiny, NumPy, Pandas, Matplotlib, Scikit-Learn, PySpark, GeoPandas, PySAL, PyTorch, TensorFlow, ggplot2, Linux, Git, Jira