

# Nithin Sai Jalukuru

[njalukur@buffalo.edu](mailto:njalukur@buffalo.edu) | <https://www.linkedin.com/in/nithin-sai-jalukuru/> | <https://github.com/jns1406/> | +1716-400-5612

## EDUCATION

**University at Buffalo, The State University of New York**, Master's in Data Science Jan 2022 – June 2023

- Maintained a **4.0 GPA** while completing relevant coursework in **Probability and Statistics**, **Data Analysis**, Data Mining (**R** programming Language), Databases (**SQL**), **Machine Learning**, Numerical Analysis, and **Data structures and Algorithms**.

## RESEARCH PROJECTS

### Forecasting Risk Gene discovery in Autism with Genome Scale Data

**Tech Stack** : **R** (*Caret, GGplot2, RandomForest, GBM, XGBoost, AdaBoost*), *RStudio*

- Improved the analysis** of the research paper "Forecasting risk gene discovery in autism with machine learning and genome-scale data" by Brueggeman et al. (2018) through replication and identifying and addressing **methodology gaps**.
- Compared **Bagging** and **Boosting** algorithms, such as **Random Forest**, **XGBoost**, **AdaBoost**, and **Gradient Boosting**, to identify optimal approach for predicting genes, achieved a Training error of 0.0027 and an AUC-ROC score of 84% for Random Forest.

### Data Scientist Salary prediction

**Tech Stack** : **Python** (*Pandas, NumPy, Panda's profiling, Plotly, Sqlite3, Scikit-Learn*), **SQL**, *MySQL*

- Collected, cleaned, and normalized data from Kaggle for data scientist job postings, and stored it in a **SQLite3** database. Analysed data by using complex visualizations to draw conclusions.
- Implemented **Ridge regression**, **Lasso regression**, **Naïve Bayes** and **SVM** to model the relationship between job attributes and salary and estimated the wage range for **data scientist** related jobs using the best-performing model.

### Web application on NYC Collision Analysis

**Tech Stack** : **Python**, *Streamlit, PyDeck, Folium, Google Cloud Platform, BigQuery, SQL, Looker Studio, Heroku*

- Extracted NYC collision data (~ **2 Million Rows**) from NYC Open Data and analysed it using GCP **BigQuery SQL**. Visualized the results with **Looker Studio** dashboards and built a web application with **Streamlit** to share the actionable insights, Later Deployed in **Heroku**.

### Highway Traffic Data Integration and Real-time Streaming Pipeline for Toll Plaza Analysis

**Tech Stack** : *Apache Airflow, Bash, Apache Kafka, Zookeeper, Simulators, Python, DAG's*

- Developed and implemented a data pipeline using **Apache Airflow** to download, extract, transform, and consolidate data from various file formats (CSV, TSV, fixed width) with **DAG** definition, data extraction, transformation, and pipeline submission.
- Configured and managed a streaming data pipeline using **Apache Kafka**, including setting up **Zookeeper**, starting Kafka server, creating a topic, downloading, and configuring the Toll Traffic Simulator

## SKILLS

**Programming Languages**: Python, R, Java, C, SQL, HTML/CSS, Unix/Linux

**Databases**: MySQL, PostgreSQL,

**Cloud** :Google Cloud Platform(Looker, BigQuery), Azure

**Tools**: Tableau/power BI, Excel, Celonis Studio, Apache Airflow, Apache Kafka, Apache Spark, Kanban, ServiceNow, Git

**Frameworks**: Flask, Keras, Pytorch, DataBricks, Hadoop, Hive, TensorFlow, Scikit-learn, Streamlit, Folium, Pydeck, Heroku.

## EXPERIENCE

**Graduate Student Assistant**, University at Buffalo, Buffalo, New York

Jan 2023 – May 2023

- Collaborated** with professor in developing **assessments** and **projects**, while providing ongoing **guidance** and **support** to **150+** students throughout the coursework (Fundamentals of Computational Science, resulting in **enhanced learning** outcomes and academic success.
- Facilitated Weekly **Office Hours** to address student inquiries and clarify any doubts in solving Assessment's (**Python**) and Projects.

**Assistant Systems Engineer (Analyst)**, TATA Consultancy Services - Dupont, Hyderabad, India

Sept 2020 – Jan 2022

- Orchestrated the implementation of Automatic Alerting Tool in SAP by leveraging **Digitate's Ignio™ AI** platform, resulting in a significant **15-hour reduction** in weekly downtime, generating **\$30,000 in annual savings** for the company.
- Enhanced communication and **workflow** by authoring system messages in SAP ABAP Systems for maintenance windows, managing and testing Remote Function Calls (**RFCs**) per client requests, and expediting the import of transports during migration.
- Improved **data insight and visualization** by leading communication with department teams, resulting in enhanced collaboration and data-driven decision making, Thus Achieved a **25% increase** in cross-functional team communication and decision-making.
- Leveraged knowledge of **database technologies** and management systems to optimize **data storage, retrieval, and manipulation processes**.
- Streamlined SAP system performance by utilizing **advanced SAP Transactions** to Extract Insights from complex Data Sets, Collaborated cross-functionally with IT teams to identify and address potential issues.
- Mentored 4 junior** team associates in best practices for performance analysis of SAP S/4 HANA core business concepts on **LINUX** systems, resulting in increase in productivity and a reduction in **issue resolution time by 30%**.