Rishabh Gundavarapu

Treyburn Circle, San Ramon, CA | 608-960-3796 | rishgun329@gmail.com | LinkedIn | Github

Education

University of Wisconsin-Madison Madison, WI

Bachelor of Science in Data Science, Computer Science Minor, 3.7 GPA, Dean's List *May '23*

Coursework: Object Oriented Programming, Data Structures, Software Programming, Operating Systems, Big Data Systems, Database Systems, Linear Algebra, Data Modeling, Supervised Machine Learning, Data Ethics and Policy

Technical Skills

Languages: Python, R, SQL, C++ , Java, JavaScript, HTML/CSS
Developer Tools/Frameworks: Tableau, Looker Studio, Apache Kafka / Spark / Cassandra, Docker, Google Cloud Platform (BigQuery,Compute Engine,BigTable), AWS(Lambda,S3,EC2) , Selenium, Snowflake
Libraries: PyTorch, Pandas, NumPy, Matplotlib, sklearn, Tensorflow, seaborn

Experience

Data Analytics Intern December '23 - February '24

Virtical (Healthcare AI Startup)

- Utilized **Snowflake**, **Python** and **SQL** verify data integrity, generate aggregation tables, and derive analytical insights from 10M+ healthcare records. Developed a Comparative system using **Python** and statistical methods for healthcare companies to analyze their payer distribution across multiple facets. Utilized Pandas to analyze price transparency data and collaborated with stakeholders to create healthcare KPIs and dashboards using Superset.
- Executed a proof of concept utilizing **PySpark**, **SQL**, and **AWS S3** buckets to automate the integration of 100+ healthcare parquet files as part of a data pipeline comprising 20M+ records in total and more.

Undergraduate Machine Learning Research Assistant February '23 – May '23

Skunkwork Informatics, UW-Madison Department of Materials Science and Engineering

- Adapted a GBM-LASSO model using **Mast-ML** to predict the density and elasticity of SiO2 glassware and contributed to the Foundry Project enabling material science researchers to utilize the model for future research.
- Implemented GridSearch with various parameters significantly improving model accuracy to **93%** .

Projects

2023 MinneMUDAC Minnesota Data Analytics Hackathon | *Python, R , ML, Business Insights*

- Led a team of 4 individuals to analyze past data, suggest business strategies, and extract actionable insights w.r.t team rivalries, home game advantages, demographics, and giveaway items to predict and improve game attendance for the Minnesota Twins Baseball team. Recommended **business strategies** focusing on impact sponsorships, promotions, and college fan bases to stakeholders. Secured 2nd position across 40+ teams in the Insights category.
- Employed **R** for data analysis, Python's Pandas, Numpy and **sklearn** libraries to construct a multiple regression model with 0.94 R^2 score. Incorporated promotional data into the model which reduced the MSE by 12%. .

Telecom Churn Analysis Tableau Dashboard | *Tableau*

- Developed a comprehensive interactive **Tableau Story** delving into the drivers of customer churn within a telecom company. Crafted dashboards, and conducted an inferential analysis to extract insights for factors impacting high churn rates, including demographics, data usage, payments, plan types, and subscriptions.
- Enabled users to filter by factors such as contract type, churn reason, and international data plan, enhancing usability and interactivity within the visualizations.

Uber Taxi ETL Cloud Pipeline and Analytics Dashboard | *Mage AI, Python, SQL, Looker Studio, GCP*

- Analyzed NYC Taxi trips data of 120K records by leveraging Mage AI to deploy an **ETL pipeline** on GCP's **Compute Engine**. Employed a combination of BigQuery and Python(pandas,numpy) for data analysis and transformation, discovering actionable trends. Engineered an interactive dashboard using **Looker Studio**, featuring multiple filters relating to Payments,Fares, and Passenger info while directly ingesting data from BigQuery.

Wisconsin Bank Loans Analysis | *Spark, Hive Tables, Spark-SQL,HDFS, REST API, Docker Images*

- Utilized **Apache Spark** to analyze a parquet dataset of 100,000 real-world loan applications of Wisconsin deployed entirely using **docker** images. Revised **query execution plan** to improve performance by **13%**.
- Conducted intricate joins using advanced grouping, filtering, and windowing functions for efficient data analysis and insights via **SQL**. Reduced the impact on load balance's performance by using single load partition caching policy and improved query times by **40%** through bucketing via **Spark's REST API**.