

# GAURI CHAUDHARI

chaudharigauri.9@gmail.com | (812) 803-9913 | [linkedin.com/in/gauri-chaudhari](https://www.linkedin.com/in/gauri-chaudhari) | [Github](https://github.com) | [Tableau](https://www.tableau.com)

## EDUCATION

**Indiana University - Bloomington, IN, United States**

May 2024

Master in Data Science

**University of Pune, India**

June 2019

Bachelor of Computer Science

## SKILLS & CERTIFICATIONS

- Programming Languages: Python, R(statistics)
- Data Science: SQL, Machine Learning, Natural Language Processing (nlp), Neural Networks, Microsoft Excel, Apache Spark, Airflow
- Libraries and Frameworks: Pandas, Numpy, Matplotlib, Seaborn, Scipy, Scikit-Learn, NLTK, Tensorflow, Keras, Hypothesis testing, PySpark
- Data Visualization Tools: Tableau
- Cloud: Amazon Web Services (AWS)(Redshift, S3, EMR, Athena, Glue), Snowflake
- Other: Git, Docker, Linux, Cloud Foundry, Jupyter
- **Google Data Analytics Professional Certificate** | Google Sept 2023
- **Associate Data Analyst Certificate** | DataCamp Feb 2024
- **Microsoft Azure AI Fundamentals** | Microsoft Apr 2024

## WORK EXPERIENCE

**Tata Consultancy Services, Mumbai, India**

*Technical Data Analyst*

Jul 2019-Jun 2022

- Designed and modernized **intuitive dashboards** and visualizations in **Splunk, Tableau** and Microsoft **Excel** for GE Digital internal teams with real-time insights into operational data trends, anomalies, and key performance indicators, facilitating data-driven decision-making.
- Exhibited robust analytical and problem-solving prowess by **analyzing and identifying anomalies and patterns** across internal dataset conducting thorough root cause analysis (RCA) on **AWS/Predix** using **SQL** and **Python** achieving a 75% improvement in system reliability and enhancing business performance.
- Performed **data modeling to predict system failures** and address potential issues, leading to a **15%** reduction in downtime.
- Collaborated with various teams and stakeholders to gather, synthesize, and present data for analysis, demonstrating strong teamwork and self-driven approach for driving a business

**Armstrong Machine Builders Pvt Limited, Nashik, India**

*Data Scientist Intern*

Sep 2018-Mar 2019

Designed and implemented a real-time analytical predictive model utilizing **back-propagation algorithm** to identify the type, timing, and root cause of conveyor belt faults with **90% accuracy**, resulting in a **98%** uptime for Propus conveyors and optimizing their operational efficiency.

## PROJECTS

### Optimizing fraud detection in credit card transactions

- Implemented and optimized a Credit Card Fraud Detection system leveraging **Random Forest** and **Decision Trees**.
- Developed a sophisticated **Deep Neural Network** model, enhanced with Synthetic Minority Over-sampling Technique (**SMOTE**), to identify fraudulent activities within a dataset of over 284,000 credit card transactions.
- Leveraged Python and machine learning libraries (pandas, numpy, keras) for data preprocessing, feature engineering, and model evaluation, leading to a 97.6% precision rate in **fraud detection**.

### Quantitative Analysis and Prediction of FAANG Stock Performance

- Conducted comprehensive **risk analysis** of FAANG stocks, utilizing **Python** and **statistical methods**, leading to the identification of a strong correlation between Microsoft and Amazon stocks, enhancing portfolio diversification strategies.
- Implemented **Monte Carlo simulations** for predicting future stock values, focusing on Google, which indicated a stable price forecast with a minimal average difference of **\$0.11**, aiding in long-term investment planning.
- **Visualized data trends** and volatility through advanced **matplotlib** and **seaborn** plots, successfully identifying Netflix's consistent stock performance over a **50-day moving average**, informing data-driven investment decisions.

### Automated Reddit Data Pipeline and Analytics using Airflow and AWS integration

- Designed and implemented an **ETL** pipeline using **Apache Airflow**, **PostgreSQL**, and AWS services (**S3**, **Glue**, **Athena**, **Redshift**) to automate the extraction, transformation, and loading of Reddit data.
- Utilized **Pandas** and **PySpark** for data manipulation and cleaning, achieving a 95% accuracy in data transformation tasks; seamlessly integrated data into **Redshift** for advanced analytics and visualization.
- Automated data extraction from Reddit APIs, storing results in CSV format and uploading to S3, facilitating real-time data querying and visualization, leading to a 30% increase in reporting efficiency.

## EXTRA-CURRICULAR & MENTORSHIP ACTIVITIES & AWARDS

- **3rd Position** in O'Neill School of Public and Environmental Affairs Data Visualization Contest (March 2024)
- Grace Hopper Conference 2023

- Awarded title of **TCS GE ISU - Young Achiever** For Q1 FY21 and Q1 FY22 by TCS (2021-2022)