Prathik Bharath Jain

+ 1 (720) 990-4940 | prathikbafna0@gmail.com | LinkedIn | GitHub

EDUCATION

University of Colorado Boulder, United States **CGPA: 4.0/4.0**
Master of Science in Data Science August 2023 – May 2025

PES University, India **CGPA: 8.4/10.0**
Bachelor of Technology in Computer Science August 2018 – June 2022

TECHNICAL SKILLS

Programming/Scripting Languages: Python, R, Java, Scala, SQL, JavaScript.
Tools: MongoDB, MySQL, AWS, Machine Learning, Statistical Analysis, Apache Kafka, Snowflake, Visualization Redshift, Airflow, Databricks, Hadoop, Hive, DBT, Big Query, Presto, Redshift, Power BI, Apache Hudi, Apache Iceberg,

EXPERIENCE

Data Engineering Intern | Lincoln Electric, Cleveland, OH May 2024 – Present
- Built data pipelines in **Azure Data Factory** using spark, azure data bricks and medallion architecture to transform and migrate data for **Power BI**, facilitating business intelligence and creating **SQL scripts** for ad-hoc requests.
- Extracting tables from SAP tables using OData connectors and SAP table connectors to Azure Synapse environment building the Datawarehouse environment and scheduling upsert pipelines in azure synapse analytics.
- Independently improved major bottle necks in algorithm and metric processing by decreasing run time by 75%
- Transcribed Oracle stored procedures and scripts into SQL Server, migrated data flows from Alteryx to Azure
- Designed and implemented a data pipeline to extract logs from Power BI using spark, ingest them into Azure, establishing a comprehensive log history and develop a monitoring dashboard, enabling proactive tracking and actionable insights.

Machine Learning Technician | Gil's Lab, Boulder, CO September 2023 – May 2024
- Applied advanced **computer vision** techniques to tune hyperparameters of **YOLOv8** for small **object detection**.
- Implemented a **bot-sort algorithm** for fish tracking, and established a streamlined pipeline with **93.87%** accuracy, showcasing proficiency in efficient optimization and enhanced insights within the domain of computer vision.

Software Development Engineer | Zee5, Bangalore, KA July 2022 - July 2023
- Utilized Java Spring Boot framework to develop the new Juspay Payment page integration. Ensured seamless compatibility with the existing infrastructure and successfully transitioned **95%** of total payment traffic to the platform.
- Revamped **Apple payment** token verification process to in-house verification and reduced the API time by **90%**.
- Led integration efforts for Juspay Express Checkout, designing the API contracts, and optimizing the back-end **payment service**. Implemented strategic improvements that significantly increased transaction success rates, contributing to an annual revenue increase of **2.5 million dollars**.
- Designed and implemented a generic asynchronous **server-to-server notification** handling system, which can be easily adapted to any payment provider. This reduced the **cloud cost** and helped in **tracking the notifications better**.
- Addressed more than a **hundred mission-critical bugs**, garnering appreciation for the achieved performance.

Junior Developer Intern | Riskcovry, Bangalore, KA June 2021 - July 2021
- Built a discharge summary validator utilizing OCR and NLP to speed up the claims and reduce manual efforts.
- Integrated multiple third-party insurance provider APIs via JavaScript and created swagger documentation.

PROJECTS

Social Pulse – Social Media Analysis System | API, MongoDB, Airflow, NLP
- Designed and implemented a system for API information extraction with the help of airflow jobs, MongoDB storage, and advanced social presence analysis across YouTube, Reddit, Twitter(X), and Tumblr. Utilized Natural Language Processing and Topic Modeling for insightful company profile assessments

Stream Forge - Real-time Data Streaming | Python, Airflow, Spark, Kafka, Databases, AWS
- Implemented a real-time data streaming pipeline using Apache Airflow to fetch the information from an API, pass it to Kafka stream, and store it in PostgreSQL, which is further transformed by spark and stored in Apache Cassandra.
- Developed a pipeline to extract data from an AWS S3 bucket into AWS Glue Data catalog using Glue Crawlers and created an ETL job to transform the data using Spark, Lambda, and stored in the Parquet format and analyzed using Athena.

Lend Lens – Bank Loan Analytics | SQL, Tableau, Excel
- Utilized SQL and Tableau for a dynamic report on key loan KPIs like Total Applications, Funded Amount, Amount Received, Average Interest Rate, and Debt-to-Income Ratio.
- Developed charts for Monthly Trends, Regional Analysis, Loan Term, Employee Length, Loan Purpose, and Home Ownership to provide visual insights into crucial lending aspects

ACHIEVEVMENTS

Winner of Data Science hackathon conducted by University of Colorado Boulder **November 2023**
Nuero-Gaze Paper published in IEEE International Conference on Distributed Computing **April 2022**