

Vinay Kiran Raju

Data Science

Email: raj.vinay2408@gmail.com | Phone: +1 (201)-492-8306 | LinkedIn: [Vinay](#)

Portfolio: vinay.github.io | Blog: [Medium - Vinay Kiran Raju](#)



PROFESSIONAL SUMMARY

Results-driven Data Analyst with a Master's in Data Science and hands-on experience in **data analysis, ETL pipelines, and business intelligence Tools**. Currently working in **capital markets**, specializing in **trade data analysis, securities transactions, and risk assessment**. Skilled in **SQL, Python, Tableau, and AWS**, with a proven ability to extract insights, optimize workflows, and enhance decision-making. Adept at **automating data processes, building interactive dashboards, and collaborating with cross-functional teams** to drive business efficiency.

TECHNICAL SKILLS

- **Programming Languages:** Python (Pandas, Beautiful Soup, NumPy, Matplotlib, Seaborn, scikit Learn), Scala, SQL.
- **Machine Learning Algorithms:** Random Forest, Fb Prophet, Naive Bayes, Logistic Regression, XG Boost, Decision-Tree.
- **Databases & Data Management:** MySQL, PostgreSQL, Snowflake, Data Modeling, Data Warehousing.
- **Visualization Tools:** Tableau, MS EXCEL, Google Sheets.
- **Cloud Services:** Amazon Web Services, Microsoft Azure.
- **Version Control:** Git, GitHub.
- **Domain:** Capital Markets

WORK EXPERIENCE

Company: Cognizant (Consultant)

Apr 2024 - Present

Role: Data Analyst

New Jersey, USA

- Analyzed large datasets using **MySQL, Excel** to identify trends, consolidate data, and provide actionable business insights.
- Automated data processing with **SQL and Tableau**, streamlining data integration and statistical computations while creating CRM and financial dashboards to enhance decision-making.
- Developed Python-based ML models for trade analytics and risk assessment, and created custom visualizations in **Python** using **Plotly, Pandas, and Matplotlib**.
- Developed and maintained BI reports and **analytical dashboards**, collaborating with cross-functional teams to deliver data-driven insights aligned with project goals and KPIs.
- Enhanced conversion metrics by 20% and developed data-driven solutions that improved business penetration by 10%, boosting profitability by 30% in Q3.
- Developed process maps and workflows in **Microsoft Visio** to support regulatory compliance and quality assurance initiatives.
- Managed high-volume data processing and reporting for securities transactions, utilizing the DTCC system to track and manage securities positions effectively.

Company: Natsoft Corporation

Feb 2023 – Apr 2024

Role: Data Engineer

New Jersey, USA

- Designed and deployed automated ETL workflows using **AWS Lambda, S3, Glue, and Redshift**, optimizing data organization and query performance.
- Implemented data transformations and aggregations with **PySpark** on EMR clusters, ensuring high data quality by reducing null values and anomalies.
- Developed and scheduled ETL jobs with **Apache Airflow**, enhancing workflow optimization with advanced features like Pool, Executors, and multi-node functionality.
- Created interactive data visualizations in **Tableau** to drive data-driven decision-making and presented actionable insights to stakeholders.
- Streamlined deployment processes by integrating Git version control with Jenkins, reducing build automation time by 10%.

Company: Natsoft Corporation

Sep 2020 – Jan 2022

Role: Data Engineer

Hyderabad, India

- Established and managed clusters on **Amazon EC2** and deployed **EMR** to establish big data environments for developing ETL pipelines and workflow.
- Increased ETL processes for ingesting data from diverse sources into HDFS and Hive using Sqoop, resulting in a 20% improvement in data ingestion speed.
- Processed web URL data using **Scala** and transformed into Spark Data Frames for analysis using **Spark SQL queries**.
- Deployed **Spark RDD** and Data Frames to rapidly process large datasets, transforming, filtering, and analyzing data, leveraging Spark's in-memory and lazy evaluation capabilities.

- Loaded and processed semi-structured data such as XML, JSON, Avro, and Parquet, optimizing Spark SQL queries to enhance data access speed by 10%.
- Orchestrated and scheduled data pipelines with **Apache Airflow**, utilizing concepts like DAGs, operators, and hooks to customize pipeline behavior.

PROJECTS

[End-to-End Automated ETL Pipeline For Zillow](#)

Jan 2024

- Built an end-to-end ETL pipeline for Zillow using **Python, Spark, AWS, and Apache Airflow**. Extracted real estate data from Zillow's API, transformed it on an EMR cluster with **PySpark**, and loaded it into S3. Automated the entire workflow with Airflow and integrated **Tableau** for data visualization, showcasing expertise in ETL processes, data transformation, and data analysis.

[Loan Default Prediction using ML](#)

Nov 2023

- Developed a Python-based **Decision Tree ML model** for loan default prediction with **95.21% accuracy** and **92.17% AUC** on 35,747 entries. Conducted extensive preprocessing to handle missing values, encode categorical features, and remove duplicates. Performed exploratory data analysis and visualized key insights using **Tableau**, improving loan risk assessment.

[Credit Card Fraud Detection using ML](#)

Dec 2023

- Developed a Python-based **Random Forest ML model** for fraud detection with **99% accuracy** and **98% cross-validation** on 550,000 entries. Conducted extensive preprocessing to handle missing values, duplicates, standardize features, and address class imbalance. Used **exploratory data analysis** and visualization to uncover patterns and correlations among 28 principal features, aiding in fraud detection.

EDUCATION

Pace University, Seidenberg School of Computer Science and Information Systems

New York, USA

MS in Data Science | **Concentration:** Data Analysis, Visualization & Machine Learning | **GPA:** 3.71/4

May 2023

Gandhi Institute of Technology and Management

Andhra Pradesh, India

Bachelor of Technology in Electronics and Communication Engineering | **GPA:** 8.38/10

Apr 2021

CERTIFICATIONS

- [Snowflake -BUILD 2023 LLM Bootcamp](#) – Snowflake Dec 2023
- [Agile Methodology Virtual Experience Program](#) – Cognizant Jun 2023
- [Data Science Virtual Experience Program](#) – British Airways Jun 2023
- [Azure Data Fundamentals](#) - Microsoft Apr 2023
- [Tableau Desktop Specialist](#)- Tableau Mar 2023
- [Google Data Analytics](#) -Coursera Nov 2022