DHATHRI CHOWDARY

eluri1d@cmich.edu | dhathrieluri6@gmail.com| 479.402.1097

PROFESSIONAL SUMMARY

Data Scientist with strong experience in statistical modeling, A/B testing, quantitative analysis, and machine learning across cloud platforms (Azure, AWS, GCP). Proven track record in transforming complex data into actionable insights using SQL, Python, and advanced visualization tools. Demonstrated ability to lead strategic analytics initiatives and build scalable, statistically rigorous solutions aligned with business goals. Adept at building end-to-end analytics pipelines, collaborating cross-functionally, and applying experimentation frameworks and data storytelling to influence decision-making at scale. Passionate about innovation and delivering impact in tech-driven environments.

Dynamic and adaptable Data Engineer, Analyst, and Scientist with strong expertise in Python, SQL, Java, and full-stack development. Proven ability to build robust ETL pipelines, implement predictive models, and manage scalable data architecture across Azure, AWS, and GCP. Hands-on experience with big data tools (Spark, Kafka, EMR), data warehousing (Snowflake, Redshift), and DevOps (Jenkins, GitHub Actions). Proficient in deploying end-to-end solutions from data ingestion to visualization using Tableau, Power BI, and Python-based dashboards. Committed to data quality, governance, and leveraging modern technologies to extract business value from complex datasets.

Master of Science, Information Systems,
Central Michigan University, Mount Pleasant, MI, USA | ***January 2023 – December 2024***
Major: Information System (Business Data Analytics) | GPA:3.4

Bachelor of Technology, Electronics and Communication,
KL University, Guntur, India | ***June 2016 – August 2020***
Major: Electronics and Communication Engineering (VLSI) | Overall GPA:3.7
Relevant Coursework: Quantitative Analysis, A/B Testing, Statistical Modeling, Data Mining, Data Visualization, Python, SQL

PROFESSIONAL EXPERIENCE

Data Scientist/ Data Analyst | Toyota ***June 2024 – Present***

- Conducted **end-to-end data analysis** using **Python**, **SQL**, and **R**, including data cleaning, exploratory data analysis (EDA), statistical modeling, and visualization to support data-driven business decisions.
- Designed and implemented **predictive models** and **machine learning algorithms** (e.g., **Linear Regression**, **Logistic Regression**, **Random Forest**, **XGBoost**, **Clustering**, **Time Series Forecasting**) to solve business problems and improve key performance indicators (KPIs).
- Designed and executed **A/B testing experiments** to evaluate optimization strategies for logistics planning and customer segmentation.

- Applied **quantitative analysis and statistical modeling** (e.g., logistic regression, decision trees) to solve high-impact business problems.
- Led **data storytelling** sessions with stakeholders by developing executive dashboards using Tableau and Power BI.
- Conducted **data mining and anomaly detection** on real-time supply chain data using PySpark and Scikit-learn.
- Defined project roadmaps and collaborated across teams to align long-term analytics initiatives with business goals.
- Delivered **SQL-based metrics and KPIs** to support strategic decision-making and model monitoring.
- Developed prescriptive analytics tools to support scenario planning and transportation optimization across warehouses.
- Used **Azure Databricks and PySpark** to process and analyze large-scale logistics datasets, enabling better placement decisions.
- Automated **data extraction**, **transformation**, and **loading (ETL)** pipelines from various sources (APIs, flat files, databases, cloud storage) using **Pandas**, **PySpark**, and **Azure Data Factory**, improving data freshness and reducing manual effort.
- Collaborated with product, marketing, finance, and operations teams to gather requirements, formulate hypotheses, perform A/B testing, and deliver actionable insights based on **statistical analysis** and **business intelligence** reports.
- Led **feature engineering** efforts and applied **model selection**, **hyperparameter tuning**, **cross-validation**, and **model evaluation** techniques to ensure high performance and generalization of machine learning models.
- Built **segmentation models** using **K-means**, **Hierarchical Clustering**, and **DBSCAN** to profile customers and optimize marketing strategies.
- Developed and maintained **data models** (logical, conceptual, and physical) and supported data warehouse initiatives on platforms like **Snowflake**, **Google BigQuery**, or **Azure Synapse Analytics**.
- Utilized **version control systems** like **Git/GitHub** for codebase management and collaborated via **Jupyter Notebooks**, **VS Code**, or **Databricks Notebooks** for shared analytics workflows.
- Maintained compliance with **data governance**, **data privacy**, and **regulatory standards** such as **GDPR** and **CCPA**, ensuring secure handling and ethical usage of customer and business data.
- Performed **root cause analysis**, **anomaly detection**, and **trend analysis** to support business continuity planning and operational performance improvement.

Environment: Python, R, SQL, PySpark, Scikit-learn, XGBoost, Pandas, NumPy, Power BI, Tableau, Plotly, Azure Synapse, Azure Data Factory, Databricks, Snowflake, BigQuery, DBT, Jupyter, Git, GitHub, Azure DevOps, PostgreSQL, SQL Server, Oracle, MongoDB, Cassandra, GDPR, HIPAA, Agile, CI/CD

Data Engineer Intern |Wells Fargo, ***April 2024 – August 2024***

- Collaborated with senior data engineers and analysts to design and implement **ETL pipelines** for ingesting and transforming large-scale financial data using **Python**, **SQL**, and **Apache Spark**.
- Supported development and optimization of **data pipelines** in **Azure Databricks**, integrating **structured and semi-structured data** from internal banking systems and external APIs.
- Assisted in performing **data profiling**, **data cleansing**, and **validation** to ensure **data quality**, consistency, and compliance with regulatory standards (e.g., **SOX**, **FINRA**).
- Created and maintained **data dictionaries**, **lineage documentation**, and **metadata** for improved discoverability and governance across the enterprise data lake.
- Conducted root cause analysis for **data anomalies** and collaborated with cross-functional teams to resolve upstream data integrity issues.

- Developed **SQL queries**, stored procedures, and views to support ad-hoc reporting, exploratory analysis, and downstream data consumer needs.
- Supported business intelligence teams with SQL queries and ad hoc data mining, improving data access for analytics.
- Participated in experimentation setup and contributed to A/B test frameworks in Python and SQL.
- Created and validated ETL pipelines supporting dashboards and forecasting models, ensuring data integrity and reproducibility.
- Contributed to the development of scalable data marts for financial planning and forecasting use cases using Azure Synapse and SQL Server.
- Participated in daily stand-ups, sprint planning, and code reviews in an **Agile environment**, gaining exposure to **DevOps**, **CI/CD pipelines**, and **version control (Git)**.
- Learned and adhered to **data security**, **confidentiality**, and **compliance protocols** in a highly regulated banking environment.
- Gained hands-on experience with enterprise tools and platforms such as **Azure Data Factory**, **SQL Server**, and **Confluence/JIRA** for documentation and task management.

Environment: Python, SQL, Apache Spark, Azure Databricks, Azure Data Factory, SQL Server, Azure Blob Storage, Power BI, Git, JIRA, Confluence, SOX, FINRA, Agile

Student Services Coordinator | CMU ***June 2023 – April 2024***

- Coordinated and delivered comprehensive **student support services** across academic advising, enrollment, career development, and international student engagement to enhance student success and retention.
- Served as the **first point of contact** for student inquiries, resolving issues related to admissions, financial aid, registration, degree requirements, and campus resources.
- Managed student records using **Student Information Systems (SIS)** and maintained compliance with **FERPA** guidelines and institutional policies.
- Collected, analyzed, and reported **student performance data** to identify academic risk indicators and recommend interventions for at-risk students.
- Collaborated with academic departments and administrative units to streamline **advising processes**, organize **orientation sessions**, and support **course scheduling and catalog updates**.
- Facilitated workshops and presentations on **academic planning**, **time management**, and **career readiness**, empowering students to achieve educational and personal goals.
- Supported **data-driven decision-making** by preparing reports on student engagement, retention, graduation rates, and satisfaction surveys using **Excel**, **Power BI**, and **survey platforms**.
- Oversaw coordination of **student events**, including graduation, mentoring programs, cultural celebrations, and wellness initiatives, promoting diversity, equity, and inclusion.
- Contributed to the development and continuous improvement of **student services workflows**, utilizing feedback and analytics to implement process enhancements.
- Provided support to faculty and staff with student-related matters, course enrollment issues, and policy interpretation, ensuring a student-centered approach.
- Maintained a professional and empathetic approach while handling confidential information, student grievances, and escalations, escalating complex cases to senior leadership as needed.

Environment: PeopleSoft, Excel, Power BI, Google Workspace, Microsoft Teams, Zoom, Trello, FERPA Compliance

Data Engineer| Data Scientist | NIIT Technologies ***April 2020 – March 2023***

- Designed and implemented scalable data pipelines and advanced analytics solutions supporting strategic planning and operational optimization across retail and logistics clients.
- Built end-to-end **inventory allocation models** using **integer programming**, **metaheuristics**, and **linear optimization** to improve product placement efficiency and warehouse utilization.
- Collaborated with cross-functional teams to define business requirements and translate them into **production-grade optimization models** integrated with existing ETL workflows.
- Developed and deployed **demand forecasting models** using **time series analysis** (ARIMA, Prophet) and **XGBoost**, driving more accurate procurement and allocation strategies.
- Engineered data ingestion and transformation workflows using **Azure Data Factory**, **Databricks**, and **PySpark**, integrating structured and semi-structured data from APIs, SQL, and NoSQL sources.
- Led data profiling, validation, and feature engineering initiatives to ensure data integrity and support model robustness.
- Worked on real-time decision-making tools by integrating modeling outputs into **Power BI dashboards** and alert systems, empowering leadership with actionable insights.
- Applied machine learning algorithms (regression, decision trees, clustering) and optimization techniques to solve business problems across fulfillment, routing, and customer segmentation use cases.
- Managed model lifecycle using **CI/CD pipelines** in **Azure DevOps** and **GitHub Actions**, ensuring scalable deployment and monitoring.
- Maintained compliance with **data governance frameworks** (GDPR, HIPAA, SOX) and integrated metadata management tools to support data cataloging and discoverability.
- Mentored junior engineers and analysts on statistical modeling, optimization frameworks, and productionization best practices.
- Applied **data modeling techniques** to develop **logical**, **physical**, and **conceptual data models** using tools like **Erwin** and **Enterprise Architect**; created and maintained **data marts**, **dimensional models**, and **star/snowflake schemas** for BI and analytics.
- Conducted **technical problem formulation** by translating complex business scenarios into data science and analytics workflows; evaluated use cases to apply **machine learning methods** like regression, random forests, and decision trees for actionable insights.
- Implemented **data governance practices**, aligned with **enterprise data policies** and compliance frameworks (**GDPR**, **HIPAA**), ensuring robust **data privacy**, **security**, **lineage**, and **retention** across all data assets and systems.
- Supported **metadata management** and utilized **AI-enabled data cataloging** tools to enhance data discoverability, reuse, and literacy across technical and non-technical teams.
- Led the development of a generative recommendation engine using **OpenAI APIs** to personalize product placements in retail dashboards, boosting cross-sell conversions.
- Acted as a **servant leader** and **culture champion** by modeling values of **integrity**, **innovation**, **diversity**, and **continuous improvement**, fostering collaboration, transparency, and inclusivity.
- Mentored junior engineers, led code reviews, and contributed to the design of reusable, modular, and production-grade code libraries for **data pipeline optimization** and **metadata-driven workflows**.

Environment: Python, SQL, PySpark, Azure Data Factory, Databricks, Synapse, Snowflake, Gurobi, OR-Tools, XGBoost, Prophet, Delta Lake, BigQuery, Tableau, Power BI, Jenkins, GitHub Actions, Event Hubs, Agile, CI/CD, GDPR, HIPAA, SOX , Presto, Kafka, Delta Lake, Event Hubs, Erwin, Dimensional Modeling, Star/Snowflake Schema, Git, Jenkins, Azure DevOps, Data Governance Tools, GDPR, SOX, Agile

Student Program Coordinator | K L University ***September 2019 – April 2020***

- Coordinated and executed **academic programs**, **student engagement activities**, and **career development**

initiatives to enhance the student experience and align with institutional goals.

- Acted as a liaison between faculty, administration, and student bodies to ensure timely communication and resolution of student concerns related to **admissions**, **course registration**, **examinations**, and **event planning**.
- Facilitated onboarding and orientation programs for new students, including **academic advising sessions**, **curriculum briefings**, and **campus resource workshops**.
- Collected, analyzed, and maintained student-related data such as attendance, academic progress, and participation rates using **Excel** and university ERP systems to support **data-driven reporting** and performance tracking.
- Supported planning and organization of **seminars**, **technical workshops**, **hackathons**, **career fairs**, and **cultural events**, increasing student involvement and interdisciplinary collaboration.
- Coordinated with the **placement cell**, **industry partners**, and **alumni network** to organize **internships**, **campus drives**, and **resume-building sessions** to support students' career preparedness.
- Created reports and dashboards for faculty and management to assess program outcomes, student satisfaction, and areas of improvement using **Microsoft Office Suite** and **Google Workspace tools**.
- Promoted diversity and inclusion by engaging students from various backgrounds in leadership and mentoring programs to build a collaborative learning environment.
- Ensured adherence to **university policies**, **academic regulations**, and **student conduct guidelines**, while addressing grievances and maintaining a positive campus culture.

Environment: Microsoft Excel, Google Sheets, ERP Systems, LinkedIn, Zoom, Google Meet, MS Outlook, Excel Dashboards, TCS Ion

Communications Engineer Intern| South Central Railway ***April 2019 – August 2019***

- Assisted in the maintenance and optimization of **railway signaling and telecommunications systems**, including **optical fiber networks**, **microwave communication**, and **VHF/UHF radio systems**.
- Supported senior engineers in the **installation, testing, and troubleshooting** of electronic interlocking systems and control panels, ensuring safety and operational efficiency across train routes.
- Participated in the inspection and documentation of **communication circuits**, analyzing transmission quality, identifying bottlenecks, and recommending improvements using diagnostic tools.
- Recorded and analyzed **technical data** from real-time communication systems to track performance indicators and maintain compliance with safety and quality standards.
- Assisted in preparing **technical reports**, **schematics**, and **wiring diagrams**, enhancing understanding of telecommunication infrastructure within large-scale public transport systems.
- Gained hands-on experience in **network topology**, **signal processing**, and **data transmission protocols**, including familiarity with SCADA, OFC systems, and GSM-R technologies.
- Collaborated with cross-functional teams during system upgrades and contributed to the development of maintenance schedules and standard operating procedures (SOPs).
- Ensured alignment with **government railway safety protocols** and **Department of Telecom (DoT)** regulations for secure communication practices.

Environment: Optical Fiber Systems, VHF/UHF Radios, GSM-R, SCADA, Interlocking Systems, AutoCAD, MS Excel, MS Word, Signal Testing Tools

---

TECHNICAL SKILLS

Quantitative Analysis & Statistical Modeling: A/B Testing, Hypothesis Testing, Regression, Clustering, Time Series Forecasting

Experimentation Design: Controlled Experiments, T-tests, Chi-square, Logistic Regression
Tools & Languages: Python, SQL, R, Jupyter Notebooks, Scikit-learn, XGBoost
Data Infrastructure: Azure, AWS, GCP, Databricks, BigQuery, Snowflake
Data Visualization: Tableau, Power BI, Plotly, Seaborn, Matplotlib
Data Mining & Storytelling: EDA, Feature Engineering, Model Interpretation, Executive Dashboards
Soft Skills: Stakeholder Communication, Cross-functional Collaboration, Agile Methodologies
Data Visualization & Reporting: Power BI, Tableau, Plotly, Excel Dashboards, Matplotlib, Seaborn
Databases & Storage: SQL Server, PostgreSQL, Oracle, MySQL, MongoDB, Cassandra, NoSQL, Azure Blob Storage, Azure Data Lake Storage Gen2
Machine Learning & Analytics: Scikit-learn, XGBoost, Time Series Forecasting, Regression Models, Clustering (K-Means, DBSCAN), Feature Engineering, Model Evaluation, A/B Testing
Data Modeling & Warehousing: Star Schema, Snowflake Schema, Dimensional Modeling, Erwin, DBT, Data Marts, Data Lakes, Metadata Management
Tools & Platforms: Jupyter Notebooks, VS Code, Git, GitHub, Bitbucket, Azure DevOps, Jenkins, Confluence, JIRA, Trello, PeopleSoft, TCS iON, ERP Systems
Data Governance & Compliance: GDPR, HIPAA, FERPA, SOX, Data Quality Frameworks, Data Lineage, Security & Access Controls
Soft Skills & Methodologies: Agile/Scrum, Cross-functional Collaboration, Sprint Planning, Business Requirements Gathering, Technical Documentation, Communication & Stakeholder Engagement
Telecom & Networking (Internship): SCADA, GSM-R, VHF/UHF Radios, Signal Testing Tools, Optical Fiber Systems, AutoCAD (basic), Railway Interlocking Systems

---

PROFESSIONAL DEVELOPMENT

Intelligent Vehicle Maintenance Tracker (IVMT) ***February 2024 – July 2024***

- Developed a cloud-native backend system for predictive maintenance and real-time diagnostics for connected vehicles (Toyota project).
- Built RESTful microservices with Spring Boot for managing service schedules, fault codes, and alerts.
- Processed vehicle telemetry via Apache Kafka, enabling real-time updates with low latency.
- Secured APIs using JWT and OAuth2 for different user roles (drivers, service centers).
- Stored structured and semi-structured data using PostgreSQL and MongoDB.
- Deployed services using Docker on AWS EC2 and automated deployments with Jenkins.
- Authored comprehensive API documentation with Swagger for integration and QA testing.
- Achieved 35% improvement in ingestion speed and 28% reduction in latency.
- Delivered business value by reducing maintenance costs and enhancing driver experience with timely alerts.

Tech Stack: Java 11, Spring Boot, Spring Cloud, PostgreSQL, Kafka, Docker, Jenkins, AWS EC2, Swagger, React

Energy Forecasting Tool (Python + AWS) ***March 2024 – May 2024***

- Enhanced an existing energy demand forecasting model using machine learning (Scikit-learn, XGBoost) for improved accuracy.
- Developed a stand-alone forecasting tool in Python with data ingestion, cleaning, model training, and deployment modules.
- Deployed solution on AWS (EC2, S3) for daily operational use by internal teams.
- Documented all steps from model evaluation to deployment, improving reproducibility and scalability.
- Presented results and future roadmap to peers and mentors, simulating a real-world consulting engagement.
- Packaged solution with CLI-based interface, stored intermediate states, and deployed via EC2 for daily operational use.

Hospital Management Analytics (Azure + ADF) ***August 2023 – December 2023***

- Conducted data analysis on patient records and hospital operations to improve workflow efficiency.
- Integrated Azure Data Factory to automate data ingestion, transformation, and reporting.
- Developed an interactive dashboard for monitoring patient visits, treatment effectiveness, and staff performance.
- Implemented data governance protocols to ensure compliance with healthcare regulations.
- Connected and transformed patient records using Azure Data Factory; used Matplotlib to track KPIs.
- Built compliance checks and integrated alerts for SLA violations on patient service timelines.

Sales Performance and Market Analysis Dashboard (SQL + Tableau) ***January 2024 – July 2024***

- Developed a real-time analytics dashboard using Tableau and SQL to analyze sales trends.
- Optimized SQL queries to enhance dashboard responsiveness, reducing data refresh time by 40%.
- Integrated heatmaps and predictive analytics to forecast market demand trends.
- Presented insights to senior management, leading to a 12% increase in profitability.
- Performed time-series clustering and linear regression to anticipate market demand.
- Optimized queries in SQL Server to power interactive Tableau dashboards used by sales leaders.

A/B Testing for Conversion Rate Optimization ***Mar 2024 – Apr 2024***

- Designed and executed controlled experiments to optimize product recommendation modules.
- Designed and executed controlled A/B experiments to improve user engagement in recommendation systems.
- Validated statistical significance using T-test and Chi-square tests; modeled outcomes using logistic regression.
- Applied hypothesis testing (T-test, Chi-square) to validate uplift in conversion using control vs. test groups.
- Modeled impact using logistic regression and visualized findings using Plotly.

Customer Segmentation Using K-Means and PCA ***Jan 2023 – Mar 2023***

- Performed dimensionality reduction and clustering on customer behavioral data using PCA and K-Means.
- Identified 5 distinct customer personas based on purchasing habits and demographics.
- Enabled targeted marketing strategies, improving click-through rates by 18%.

Model Deployment with Flask and Docker (Mini Project) ***Dec 2022 – Jan 2023***

- Containerized a trained fraud detection model using Docker.
- Deployed as RESTful API using Flask for real-time scoring of transactions.
- Used Jenkins for CI/CD and monitored performance with logging dashboard.

Design and Implementation of IoT based Street Lightening Project

IEEE Paper Published, *Project* ***April 2019- April 2020***

- Designed an IoT-based system for remote monitoring and automated control of streetlights.
- Implemented machine learning models to analyze energy consumption and optimize usage.
- Established a centralized cloud-based dashboard for real-time monitoring and maintenance alerts.
- Implemented energy-saving ML algorithm; centralized real-time data with cloud dashboard.
- Conducted A/B testing to verify cost savings in live environments.

Library Management System

K L University**,** *Project* ***December 2018 -April 2019***

- Compiled a Library management system for our college library using PHP.
- Created a User-friendly web application using Flask.

Certifications

- Microsoft Certified: Azure Data Engineer Associate
- Microsoft Certified: Azure Data Fundamentals
- Pega Certified Senior System Architect (CSSA)
- Pega Certified System Architect (CSA)

- Data Science with Python – Simplilearn
- Java, Python, Problem Solving

ADDITIONAL INFORMATION

- Team leadership experience as a lead for the Library Management project.
- Published IEEE research paper on IoT-based street lighting.
- Open to relocation and eligible to work under OPT status in the USA.

ADDITIONAL PROJECT

Cloud-Native ETL and Governance Framework ***June 2023 – August 2023***

- Designed and implemented distributed ETL pipelines using Apache Spark, Kafka, and AWS EMR.
- Automated batch and streaming data loads from internal APIs and GCP DataProc to Snowflake and Azure Synapse.
- Applied feature extraction, normalization, and decision tree algorithms to build ML-ready datasets.
- Built Airflow DAGs for orchestrating end-to-end data ingestion and transformation workflows.
- Enforced data governance using metadata cataloging, schema validation, RBAC, and automated documentation.
- Developed dashboards in Tableau and Power BI to visualize operational KPIs, anomaly trends, and SLAs.
- Integrated CI/CD workflows using Jenkins and GitHub Actions for continuous integration and deployment.