ROOCHITA IKKURTHY

+1 (850)-300-0923 | roochita27@gmail.com | https://github.com/iroochi | https://www.linkedin.com/in/roochita-ikkurthy-89b93b1b7/

EDUCATION

FLORIDA STATE UNIVERSITY Tallahassee, Fl (*August'23 – May'25*)
Master of Science – Computer Science, GPA: 3.825
Relevant coursework: Advanced Data Mining, Advanced Database systems, Artificial Intelligence, Advanced Algorithms, *Software Engineering*

SKILLS

Core Languages: Python, R, SQL
Languages: C, C++, Java, HTML5/CSS3, JavaScript
Data Engineering & Big Data: PySpark, Apache Airflow, Hadoop, Data Mining, **Microsoft Azure** (Data Factory, Data Lake Gen2, Databricks, Synapse Analytics),
Data Analytics & Visualization: Tableau, Power BI, Pandas, NumPy, Excel, BeautifulSoup
Machine Learning & Libraries: scikit-learn, TensorFlow, Keras, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Decision Tree, Logistic Regression, K-Means, Artificial Neural Networks
Database Technologies: MySQL, SQLite, Relational Databases, SQL-based ETL, PostgreSQL, Data Modeling
Development Tools & Version Control: VS Code, PyCharm, R Studio, GitHub, Bash
Core Competencies: ETL Development, Data Pipeline Automation, Data Warehousing, Data Modeling, Cloud Computing, Data Structures & Algorithms, Object-Oriented Programming (OOP), Data Cleaning & Transformation, Statistical Analysis
Awards: Awarded merit scholarships for three years in a row for excellence in academics in Undergraduate.
Certifications: Microsoft Certified – Azure AI Fundamentals, Working with BigQuery by Coursera, Tableau Training by LinkedIn Learning

EXPERIENCE

RESEARCH ASSISTANT AT FSU Tallahassee, Fl (*October'24 – May'25*)
- Engineered an end-to-end data pipeline for high-dimensional gene datasets using Python and SQL to extract, preprocess, and store structured data efficiently.
- Leveraged Azure Data Lake Gen2 and Synapse Analytics for cloud-based storage and scalable analysis, optimizing performance for hierarchical clustering workflows.
- Automated statistical analysis (Z-tests) and gene enrichment processes with Python scripts, reducing manual workload by 50% and enhancing reproducibility.

STUDENT WORKER / LEAD AT ARAMARK Tallahassee, Fl (*October'23 – December'24*)
- Led a team of student workers in managing customer service operations, ensuring adherence to Customer Relationship Management (CRM) protocol and timely resolution of service issues.
- Coordinated data collection and reporting to streamline feedback systems and support management decision-making.
- Trained new hires on data entry, CRM tools, and customer interaction guidelines, improving onboarding efficiency by 30%.
- Maintained and updated records in the Customer Relationship Management (CRM) system to track service requests and performance metrics.
- Analyzed trends in customer feedback and identified areas for service improvement.
- Supported internal reporting processes by preparing weekly status updates and performance summaries for supervisors.

RESEARCH INTERN AT BAVCH EDUTECH India(Remote) (*June'22 – August'22*)
- Collected and analyzed education tech usage metrics using Excel and Python; created reporting dashboards to monitor engagement trends.
- Designed lightweight data processing scripts for KPI tracking and supported data-driven decision-making for product development teams.

DATA SCIENCE INTERN AT VERZEO India(Remote) (*May'21 – August'21*)
- Built ETL pipelines using Pandas and SQL to clean, merge, and prepare large-scale datasets, achieving a 35% reduction in processing time.
- Integrated machine learning workflows into the data pipeline, using SVM and K-NN, and output results to Tableau dashboards for business stakeholders.
- Designed reproducible project structure using version control (GitHub) and modular Python scripts for scalable deployment.

PROJECTS

Tokyo Olympics 2020 Data Engineering Pipeline **May'25**
Technologies: Python, Azure Data Lake Gen2, Databricks, PySpark, SQL, Tableau, Apache Airflow
- Designed and implemented a scalable end-to-end data pipeline to analyze athlete performance, event metrics, and country-wise medal trends from the Tokyo 2020 Olympics.
- Ingested and processed multi-format data (CSV/JSON) using Azure Data Factory and stored it in Azure Data Lake Gen2 with sport and country-based partitioning.
- Utilized PySpark on Databricks for data cleansing, transformation, and aggregation, generating insights such as medal tallies, gender participation ratios, and event performance trends.
- Created analytical models and visualized key metrics via Tableau dashboards, enabling real-time monitoring of medal standings and sport-wise comparisons.
- Orchestrated daily data updates and pipeline execution using Apache Airflow, ensuring consistent refresh and reporting automation.

Retail Sales Analytics Pipeline **April'25**
Technologies: Python, SQL, SQLite, Tableau, Pandas
- Developed an analytics pipeline to clean, transform, and analyze retail sales data from Superstore dataset.
- Loaded data into a relational database (SQLite) and created SQL queries to extract KPIs such as sales trends, profit margins, and customer segments.
- Built a dynamic Tableau dashboard enabling business stakeholders to make data-driven inventory and marketing decisions.

Real-Time Outlier Detection System (NETS Algorithm) **April'25**
Technologies: C++, Grid-based Indexing, Streaming Simulation
- Implemented a high-speed outlier detection engine based on the NETS research algorithm for streaming data environments.
- Designed grid-based indexing and efficient dimensional filtering to handle large volumes of real-time input.

- Benchmarked system performance across multiple datasets, achieving 100x speedup compared to baseline detection methods.

Crime Forecasting Using Spatiotemporal Models   **September'24**

Technologies: Python, Graph Neural Networks (GNN), LSTM/GRU

- Engineered a data pipeline to integrate socioeconomic and crime data for the city of Chicago.
- Built predictive models using GNNs and temporal networks (LSTM/GRU) to forecast crime likelihood at regional levels.
- Highlighted how spatiotemporal modeling enhances real-time adaptability in urban safety planning.