Abhishek Rithik Origanti

College Park, MD | (984)687-6246 | abhishekoriganti@gmail.com | LinkedIn | GitHub | Portfolio

EDUCATION

UNIVERSITY OF MARYLAND *College Park, MD*
Master of Science in Data Science – GPA: 3.57/4.0 *Aug 2024 – May 2026*
Coursework: Statistics, Machine Learning, Algorithms for Data Science, Big Data Systems, Cloud Computing, Deep Learning

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY *Chennai, India*
Bachelor of Technology in Computer Science and Engineering -AI & ML – CGPA: 8.8/10.0 *Aug 2020 – May 2024*
Coursework: Data Structures, Database Management System, Computer Architecture, Data Science, Software Engineering

WORK EXPERIENCE

COUNSELING CENTER, UNIVERSITY OF MARYLAND *College Park, MD*
Research Analyst *Sep 2024 – Present*

- Consolidated 4 to 5 campus data sources into a structured **PostgreSQL** and Python data model, validating 10+ years of FERPA and **HIPAA**-compliant records to enable demand modeling and operational reporting across institutional workflows.
- Translated requirements from 40 to 50 clinicians into SQL, Excel, and **Tableau** dashboards, delivering quarterly utilization reports that improved scheduling efficiency by **25%** and informed resource allocation decisions.
- Trained Random Forest and Logistic Regression models on EHR-style appointment data to predict no-show risk, achieving **84.7%** accuracy with ROC-AUC 0.81 and cutting manual effort by **30%** across 3 automated workflows.
- Applied NLP (TF-IDF, sentiment analysis) to 4,000+ student feedback responses and ran t-tests and chi-square analyses, surfacing service quality patterns presented via **Power BI** to the VP for accreditation reporting.

INTERLINKED CORPORATION *Berkeley, CA*
Data Engineer Intern *May 2025 – Sep 2025*

- Architected a Medallion lakehouse on **AWS S3** using Kafka for ingestion, PySpark for distributed transformation, and AWS Glue for orchestration, processing 2M+ daily records and cutting ETL latency by **85%** (3 hrs to 25 min).
- Established **GitHub Actions** CI/CD pipelines for nightly data validation runs, artifact versioning, and failure alerts, cutting p95 inference latency under **350 ms** by streamlining feature pipelines and adding prediction caching.
- Benchmarked **Redshift** schema configurations including sort keys and distribution styles via A/B testing, improving dashboard query performance by **40%** and accelerating business reporting turnaround for BI consumers.
- Instrumented production pipeline monitoring via AWS **CloudWatch** and Airflow metrics across 5 DAGs, performing root cause analysis on incident signals to reduce unplanned downtime and maintain SLA compliance.

PROJECTS

Cloud-Native Travel Data Orchestration Platform using AI and Microservices | GitHub *Aug 2025 – Dec 2025*

- Developed an MCP-based travel planner in Python 3.12 using **Claude** to orchestrate **7 specialized servers** across SerpAPI, Open-Meteo, and OpenStreetMap APIs, consolidating flights, hotels, and weather into one query.
- Deployed the system on **AWS ECS Fargate** using CloudFormation IaC, DynamoDB persistence, and Secrets Manager, achieving sub-second query resolution across 7 live API integrations with CloudWatch observability.

SafeSim: Safe Medical Text Simplification with Neuro-Symbolic Verification | GitHub *Feb 2025 – May 2025*

- Designed a neuro-symbolic NLP pipeline combining **spaCy** NER and regex extraction with 4 multi-LLM backends (GPT-4, Claude, HuggingFace BART/T5) to simplify medical discharge summaries, achieving **95%+** entity preservation.
- Implemented a deterministic verification layer with exact, normalized, and fuzzy matching across 3 configurable strictness tiers (95%, 85%, 75%), containerized via **Docker** with a FastAPI REST API and test suite.

Electric Vehicle Infrastructure Analytics and Charging Demand Forecasting | GitHub *Aug 2024 – Dec 2024*

- Analyzed a 216,772-row, 17-column EV dataset spanning 1999 to 2025 using Random Forest and Logistic Regression with SMOTE to predict electric range and classify CAFV eligibility, achieving **R2=0.918** and **99.85%** classification accuracy.
- Ran t-test, chi-square, and ANOVA hypothesis tests, then mapped Washington State rural charging infrastructure gaps via **GeoPandas** and Folium, achieving 98.7% mean 5-fold cross-validation accuracy.

SKILLS AND PROFICIENCY

Programming: Python (Pandas, NumPy, Scikit-learn), SQL, R, Scala, Excel
Data Engineering: PySpark, dbt, Apache Airflow, Kafka, ETL/ELT Pipelines, Data Modeling, Data Warehousing, Databricks, Snowflake, PostgreSQL, MySQL, MongoDB
Cloud & Infrastructure: AWS (S3, Glue, Redshift, Lambda, Step Functions, EMR, Athena, CloudWatch), Docker, CI/CD
Analytics & Modeling: Tableau, Power BI, Amazon QuickSight, Data Visualization (Matplotlib, Seaborn), KPI Tracking, A/B Testing, Hypothesis Testing, Regression, Time Series Analysis
AI & ML: Feature Engineering, Anomaly Detection, NLP, PyTorch, TensorFlow, XGBoost, RAG, LangChain, Amazon Bedrock
Certifications: AWS Certified Solutions Architect – Associate, Neural Networks and Deep Learning – DeepLearning.AI

PUBLICATIONS

- Origanti, A. R. (2025). Efficient Deep Fake Image Detection Using Dense CNN Architecture. IRCCTSD 2024, CCIS vol. 2361, Springer. **[Best Paper Award]**