Aravind Kumar Reddy Yeddula

+1 (551)-260-8700 | aravindyeddula2025@gmail.com | LinkedIn

SUMMARY

Data Scientist & AI Engineer with 5+ years of experience designing, deploying, and scaling machine learning pipelines and production-grade AI solutions on Microsoft Azure and AWS. Drive innovation in predictive modeling, anomaly detection, and explainable AI to ensure transparency and compliance. Experienced in integrating large-scale structured and unstructured data, automating ETL workflows, and delivering actionable insights that accelerate decision-making in healthcare and finance.

PROFESSIONAL EXPERIENCE

CitiusTech, USA **Dec 2024 – Current**

AI Engineer

Project: Predictive Healthcare Analytics Platform

- Built machine learning models using TensorFlow, XGBoost, and Keras that improved patient outcome prediction accuracy by 18%, enabling faster decisions and risk reduction through early intervention and insights.
- Automated ETL pipelines using Azure Data Factory and Apache Airflow, reducing data ingestion latency by 25%, integrating healthcare data sources for improved real-time analytics and reporting accuracy.
- Developed streaming data solutions with Apache Kafka and Databricks for anomaly detection in medical telemetry, generating instant alerts that reduced device downtime by 40% and prevented potential clinical errors.
- Secured cloud infrastructure on Microsoft Azure with RBAC, Azure AD, and encryption, ensuring 100% HIPAA compliance and protecting data for 50,000+ patient records.
- Designed interactive dashboards in Power BI and Matplotlib that transformed complex datasets into actionable insights, boosting clinician decision-making efficiency by **35%** and reducing report turnaround from days to minutes.
- Created explainable AI dashboards using SHAP and LIME to help compliance teams interpret risk model decisions.
- Integrated Large Language Models (LLMs) via Azure OpenAI Service to summarize unstructured clinical notes, extracting 20+ key risk factors per patient and improving feature engineering efficiency by 50%.

Deloitte, India **Jun 2019 – Jun 2023**

Data Scientist

Project: Financial Risk Modeling and Data Warehouse Modernization

- Engineered ETL workflows using Apache Spark and PySpark, increasing financial data processing throughput by 30%, improving availability and accuracy of datasets used in client analytics and regulatory reporting.
- Designed and Enhanced data warehousing solutions on Amazon Redshift and Snowflake, enhancing SQL query performance by 40% and reducing report generation time for business users.
- Applied machine learning techniques including regression and classification to improve client risk prediction accuracy by 15%, supporting better financial decision-making and portfolio risk management.
- Led cloud deployment on AWS S3, ensuring high availability, data durability, and disaster recovery readiness for critical financial datasets, meeting compliance and security standards.
- Automated CI/CD pipelines using GitHub Actions and Terraform, accelerating deployment speed and improving reliability of software releases across development teams.

Deloitte, India **Dec 2018 – Jun 2019**

Data Analyst Intern

- Conducted in-depth financial risk analysis using Python, Pandas, and NumPy, identifying critical credit and market risk factors, enabling mitigation strategies that reduced potential losses by 18% for clients.
- Deployed predictive risk models using regression, classification, clustering in Scikit-learn, validated via cross-validation and precision/recall, improving credit risk scoring accuracy by 22% and enhancing portfolio segmentation for 5,000+ accounts.
- Created dashboards in Power BI and Excel to visualize portfolio performance, compliance, and risk metrics, reducing reporting time by 40% and enabling executives to make investment decisions 30% faster.
- Collaborated in modernizing data warehouse infrastructure by integrating Snowflake with Azure Data Factory and AWS S3, Streamlined ETL pipelines, reducing data latency by 25%, and supporting scalable, real-time financial reporting.
- Applied ARIMA-based time-series forecasting for market trends and implemented NLP techniques on financial reports, increasing predictive accuracy by 15% and providing analysts with richer insights for risk management and investment strategies.

EDUCATION

Master of Science in Data Science **Sep 2023 – May 2025**

Pace University, Seidenberg School of Computer Science and Information Systems, New York

Bachelor of Technology in Electronics and Communication Engineering **Jul 2015 – May 2019**

SRM Institute of Science and Technology, Chennai, India

PROJECTS

Efficient Fake News Identification through Hybrid NLP Pipelines **Dec 2024**

- Leveraged NLP techniques (TF-IDF, word embeddings, transformers) to preprocess text, extract features, and generate clean training datasets from news sources.
- Trained and compared Logistic Regression, Random Forest, SVM, LSTM, and BERT models and optimized inference latency via parallel matrix computations, reducing prediction time from 120ms to 35ms per input.
- Assessed model performance using precision, recall, F1-score, and ROC-AUC, selecting the most balanced and high performing model for deployment.

Flight Delay Prediction Using RapidMiner **May 2024**

- Queried and cleaned 30+ million flight records using Hadoop Hive, then securely stored transformed data in AWS S3 for streamlined access and further machine learning processing.
- Developed a RapidMiner pipeline implementing classification algorithms to predict flight delays with 98% accuracy, supplemented by Seaborn and Matplotlib visualizations for comprehensive data insight.

TECHNICAL SKILLS

- **Programming**: Python, SQL, PySpark
- **Data Engineering & Processing**: Apache Spark, Databricks, Apache Kafka, Airflow, Azure Data Factory, dbt, Delta Lake, AWS Glue
- **Data Warehousing & Databases**: Snowflake, Azure Synapse Analytics, Amazon Redshift, Google BigQuery, MySQL, PostgreSQL
- **Cloud Platforms**: Microsoft Azure (Data Factory, Databricks, Synapse, Key Vault, Azure Monitor), AWS(S3, EC2, IAM, Glue)
- **AI / Machine Learning**: TensorFlow, Keras, Scikit-learn, XGBoost, RapidMiner, Regression, Classification, Model Tuning, Time-series forecasting models (ARIMA, Prophet, LSTM), NLP models (BERT, spaCy)
- **Data Visualization & Reporting**: Power BI, Tableau, Matplotlib, Seaborn
- **Workflow Automation & CI/CD**: Azure DevOps (YAML pipelines, ARM templates), GitHub Actions, Docker, Kubernetes, Terraform
- **Version Control & Collaboration**: Git, Jira, Confluence
- **Data Manipulation**: Pandas, NumPy
- **Security & Compliance**: HIPAA compliance, Encryption at rest and in transit, RBAC with Azure AD

CERTIFICATIONS

AWS: Certified Data Engineer – Associate
AWS: Certified Machine Learning – Specialty
IBM: Data Science Certification

ACHIVEMENTS

Data Science Scholar Award from Pace University