Aakash Kunarapu

Kent, OH | +1 330-281-0912 | aakashkunarapu17@gmail.com | Aakash Kunarapu

Education

Kent State University, OH — **Kent, OH**
M.S in Computer science — *Graduation Date: May 2025*
Kakatiya University, Warangal — **Wanrangal, TG**
Bachelor's in computer applications — *Graduation Date: May 2022*

Work Experience

Genpact - Data Scientist / Platform Compliance Analyst (Meta Platforms) — **Jan 2023 – Jul 2023**

- Audited Facebook Login and Account Kit integrations for 400 web & mobile apps, parsing OAuth transaction logs and SDK telemetry with Python/SQL to surface common infringements and map violation trends by platform and region.
- Designed a scikit-learn risk-scoring model (gradient-boosted trees) that used 25 engineered features, permission mix, redirect-URI entropy, error-rate drift, user-drop-off to rank integrations by likelihood of non-compliance; the top-20 list caught 78% of true violations and cut manual review hours 25%.
- Built PySpark pipelines in Airflow to ingest daily authentication events into Redshift and power a Tableau dashboard for product, legal, and engineering stakeholders, giving near-real-time visibility into compliance status, incident spikes, and login-success KPIs.
- Worked directly with external developer teams to translate data findings into concrete remediation steps, guiding 120 apps to full guideline alignment and boosting first-time-login success rates by 12% month-over-month.

Project Experience

Skytrax Reviews Analysis for British Airways (Data Science Virtual Internship):

- Analyzed approximately 3,940 Skytrax reviews using sentiment analysis (VADER) and NLP techniques.
- Identified key sentiment trends: 56% positive, 41% negative, highlighting critical customer insights.
- Extracted major positive and negative themes (e.g., cabin crew service, premium economy experience).
- Recommended actionable insights for improving customer satisfaction, particularly on long-haul routes.

Network Analysis of Marvel Universe Characters:

- Performed large-scale network analysis of 6 000+ Marvel characters and 160 000 interactions in NetworkX, computing centrality metrics to surface the most influential nodes.
- Discovered latent communities and alliances via Greedy Modularity and Girvan–Newman algorithms, then evaluated network density, transitivity, core-periphery structure, and overall robustness.
- Created interactive Plotly and static Matplotlib visualizations that translated complex graph insights into clear, stakeholder-ready stories while honing advanced data-wrangling and graph-construction skills.

Sepsis Detection with LSTM-GBM Ensemble:

- Built a data pipeline that integrated time-series preprocessing and model stacking for 40,000+ ICU records.
- Engineered features from clinical signals and deployed ensemble models to detect anomalies, simulating fraud detection logic.
- Validated using AUC-ROC, F1-score, and precision-based metrics critical in fraud classification.

Bayesian Grid Risk Forecasting:

- Built probabilistic models for grid stability with Monte Carlo simulations and Bayesian networks, highlighting experience in complex risk modeling and statistical learning.
- Delivered insights via dashboards, emphasizing effective communication of technical finding**s** to stakeholders.

Customer Churn Prediction via Data Mining:

- Modeled behavior prediction using Random Forest (93%) and Decision Trees, with heavy emphasis on imbalanced data handling.
- Applied SHAP-based feature interpretation, replicable for explaining fraud model decisions in production.
- Optimized end-to-end flow from data wrangling to actionable insight delivery.

Skills and Interests

- **Languages & Tools:** Python, SQL, PySpark, Java, PostgreSQL, MongoDB, Redshift.
- **Big Data & Processing:** Apache Spark, Data Pipelines, Feature Engineering**.**
- **Machine Learning**: Supervised & Unsupervised Learning, Linear & Logistic Regression, Decision Trees, Random Forest, SVM, K-Means, Naive Bayes, Model Evaluation & Tuning (Cross-Validation, Grid Search), Classification, Gradient Boosting, Feature Engineering, NLP Algorithms, Data Science Toolkits, Anaconda, TensorFlow.
- **Graph Analytics:** NetworkX, Gephi, Graph Features, Connectivity Analysis.
- **ML Deployment:** Model Hosting, API Integration, Data Parsing, Operational Monitoring.

- **Statistical Analysis**: Hypothesis Testing, Regression, Distributions, PCA.
- **Cloud & SaaS**: AWS, GCP, Slack, Okta (basic familiarity), Google Workspace.
- **Libraries**: Pandas, NumPy, Scikit-learn, PyMC3.
- **Visualization**: Power BI, Tableau, Dashboards.

Software Practices: Agile, Version Control (Git), Object-Oriented Design.

Certifications

- Python for data science and machine learning essential training.
- British Airways Data Science Job Simulation.
- Big Data Analytics with Hadoop and Apache Spark.
- Introduction to prompt engineering for generative AI.