KARTHIK M

+1(737) 239-1784 | mvkarthik2522@gmail.com | linkedin | Austin | TX

PROFESSIONAL SUMMARY

Results-driven Data Scientist with 3+ years of experience developing predictive models, NLP pipelines, and real-time analytics to optimize healthcare and financial operations. Skilled in Python, SQL, and cloud-native workflows (AWS, GCP), with expertise in machine learning (Scikit-learn, TensorFlow, PyTorch) and data engineering (Spark, Airflow, Kafka). Proven ability to translate complex datasets into actionable insights that improve decision-making, forecasting accuracy, and business performance.

EDUCATION

Purdue University, Fort Wayne   Aug 2023 – May 2025
Master of Science, Computer Science

TECHNICAL SKILLS

Languages: Python, R, SQL, SAS

Machine Learning & AI: Supervised/Unsupervised Learning, Reinforcement Learning, Neural Networks (CNN, RNN, LSTM, GAN), Random Forest, Gradient Boosting, XGBoost, SVM, K-Means, KNN, Scikit-learn, TensorFlow, PyTorch, NLP (spaCy, BERT)

Data Engineering: Apache Spark, Hadoop, Kafka, Hive, ETL/ELT, Airflow

Databases: MySQL, PostgreSQL, MongoDB, MS SQL Server, Snowflake, Amazon Redshift, BigQuery

Data Visualization: Tableau, Power BI, Excel (Advanced), Chart.js, Matplotlib, Seaborn, Plotly, ggplot2

Cloud & DevOps: AWS, GCP, Docker, Kubernetes, GitHub Actions, GitLab CI/CD, Jenkins

Other Tools & Skills: Pandas, NumPy, dplyr (R), OpenCV, SSIS/SSRS/SSAS, Informatica, Talend, MLFlow, Advanced Analytics, A/B Testing, Hypothesis & Statistical Testing, Regression, Time Series, KPIs, Accuracy/Precision/Recall/F1, ROC Curve, AUC, Data Quality & Governance, Data Warehousing, Data Transformation & Cleansing, SDLC, Agile Development, QA Automation, Testing & Deployment, Problem-Solving

Version Control: Git, GitHub, SVN

Operating Systems: Windows, Linux, MacOS

PROFESSIONAL EXPERIENCE

Data Scientist, UnitedHealth Group, TX, USA   Aug 2024 – Present

- Improved disease detection by 24% with Scikit-learn models using classification, feature engineering, and cross-validation; enabled personalized care and better resource allocation.
- Forecasted ICU bed demand and patient readmission with LSTM models in PyTorch, improving hospital planning by 21% and reducing wait times through temporal data modeling.
- Reduced ER overload by 18% via a reinforcement learning triage engine (OpenAI Gym + TensorFlow) that prioritized high-risk patients using real-time and historical clinical data.
- Increased clinical outcome prediction accuracy by 19% using Gradient Boosting with SHAP interpretability, hyperparameter tuning, and early stopping.
- Optimized large-scale clinical datasets in Amazon Redshift for HIPAA-compliant access; powered real-time dashboards for KPI tracking and care quality benchmarking.
- Boosted early intervention accuracy by 22% through advanced data mining on longitudinal EHR datasets using clustering and association rule mining.
- Built an NLP pipeline with spaCy + BERT embeddings to extract medical entities from clinical notes, improving classification accuracy and accelerating case review workflows.

Data Scientist, Cognizant, India   May 2021 – Jul 2023

- Engineered data pipelines in Snowflake integrating financial transactions with APIs and market feeds; boosted query performance by 40% with SQL transformations and secure access.
- Built Power BI dashboards for revenue trends, loan delinquency, and churn, enabling business leaders to act on KPIs across dynamic financial portfolios.
- Predicted loan default risk with optimized XGBoost models (Bayesian tuning), increasing accuracy by 14% and reducing false positives in underwriting workflows.
- Forecasted cash flows and expenses with ARIMA and Prophet, improving budgeting accuracy by 23% via seasonality-adjusted projections and diagnostics.
- Orchestrated ETL workflows in Apache Airflow to process structured data from AWS S3, ensuring task dependencies, high availability, and reliable ML pipeline delivery.
- Improved fraud detection with decision trees, logistic regression, and ensembles; classified high-risk customers through cross-validated supervised learning.
- Designed Talend ETL pipelines integrating legacy and cloud data, enhancing data consistency by 30% and accelerating financial reporting with transformation logic and error handling.

PROJECTS

EventEase - Event Management Platform   2024 – 2025

- Built an event management platform that applied NLP via OpenAI APIs to personalize onboarding and recommendations, and developed predictive models on booking data to forecast demand.
- Designed real-time analytics dashboards with Chart.js and Python (Pandas, Scikit-learn) to track KPIs such as revenue, attendance, and churn, enabling data-driven event planning.