Karthik Mettu

karthikrm202002@gmail.com | +1(713)202-0004 | https://www.linkedin.com/in/karthik-mettu/ | Houston, Texas

https://karthik1636.github.io/karthik.github.io/

Professional Summary

Experienced **Data Analyst and GenAI Developer** with expertise in ETL processes, managing relational and NoSQL databases, and deploying AI/ML models into production using tools like Azure AI Studio and DataStage. Proficient in Python, R, SQL, and visualization tools like Tableau and Power BI, with a strong focus on delivering data-driven insights, automating workflows, and practicing Agile methodologies. Effective communication, collaboration & team building skills with proficiency in quickly grasping new technical concepts and practicing Agile methodologies.

Technical Skills

- **Big Data Technologies:** Apache Spark, Hadoop (HDFS, MapReduce), AWS (S3, Redshift, DynamoDB), Google Cloud Platform, Tableau, Power BI, ETL, CI/CD, SCD.
- **Big Data Analytics Frameworks**: MLlib (Spark), Hadoop Ecosystem, AWS SageMaker.
- **Programming Languages:** Python, SQL, PL/SQL, Linux shell scripts, R, C
- **RDBMS:** IBM DB2, MySQL, Microsoft SQL server, MS Access, PostgreSQL, Oracle
- **NoSQL**: Cassandra, Mongo DB
- **Cloud Technologies**: Azure AI Studio, AWS (S3, EC2, Redshift, DynamoDB), Google Cloud Platform.
- **Data Science and Machine Learning Tools**: Scikit-learn, TensorFlow, PyTorch, Statsmodels, Prophet, FactoMineR, caret.
- **Visualization and Reporting Tools**: Tableau, Power BI, Matplotlib, Seaborn, Plotly, ggplot2, Excel.
- **Methodologies**: Agile, Waterfall, Statistical Modeling, Hypothesis Testing, Dimensionality Reduction, Time-Series Analysis.
- **Tools and Utilities**: Jupyter Notebook, Google Colab, Anaconda, VS Code, Putty, MS Office.

Education Details

Master of Science, Statistics and Data Science **Aug 2023 – December 2024**

University of Houston, Texas **GPA: 3.74**

Professional Experience

Instructional Assistant, University of Houston (Houston, Texas) **Aug 2023 – December 2024**

As an **Instructional Assistant** at the University of Houston, I guided Data Science graduate students in developing and validating machine learning models, conducting statistical analyses, and optimizing data workflows using Python, R, Tableau, and advanced data science techniques as listed below.

- Programmatically acquired data using API requests (requests, httr), command-line tools, database queries (SQL, PostgreSQL, MongoDB), and web scraping tools (Beautiful Soup, Selenium), leveraging Google Colab and Jupyter Notebooks for collaborative workflows.
- Ensured data quality and analysis readiness through validation, cleansing, and enrichment using Python (Pandas, Scikit-learn), R (dplyr, tidyr), Tableau (calculated fields, prep flows), and Excel (Power Query).
- Identified trends and relationships in datasets using Python (Pandas, Matplotlib, Seaborn, Plotly) and R (ggplot2, DataExplorer) for visualizations and insights.

- Conducted hypothesis testing (t-tests, ANOVA), regression analysis (linear, LASSO, Ridge), and Bayesian inference with Python (SciPy, Statsmodels) and R (aov, MCMCpack).
- Enhanced model performance with feature creation, transformations, and dimensionality reduction techniques like PCA (Scikit-learn, FactoMineR) and Tableau (calculated and combined fields).
- Built and optimized supervised and unsupervised models, including Random Forests, SVM, k-Nearest Neighbors, and k-Means clustering, using Python (Scikit-learn, TensorFlow) and R (caret).
- Developed sentiment analysis pipelines with Bidirectional LSTM and Transformer models, leveraging Python (TensorFlow, PyTorch) for context retention and precision, achieving over 99% validation accuracy.
- Designed forecasting models for sequential data using Python (statsmodels, prophet) and R (forecast, ts), enabling accurate trend predictions.
- Enhanced accuracy and robustness with hyperparameter tuning and resampling methods like cross-validation and bootstrap, using Python (GridSearchCV) and R (caret).
- Delivered insights through interactive dashboards and visualizations in Tableau (dashboards, stories) and Python (Matplotlib, Seaborn, Plotly).
- Applied probabilistic techniques like Markov Chains and Monte Carlo simulations using Python (PyMC3, Statsmodels) and R (MCMCpack) for predictive analysis.
- Created workflow documentation and reports using Jupyter Notebooks, Google Colab, R Markdown, and Tableau Stories to support knowledge sharing and effective communication.

GenAI Developer and Data Analyst, American Red Cross (Houston, Texas) **May 2024 – Present**

As a **GenAI Developer and Data Analyst** at the American Red Cross, I leverage Microsoft Azure tools to develop and deploy AI-driven solutions, automate ETL pipelines, and provide actionable insights to enhance disaster relief and training services.

- Developed AI models for disaster relief and training services using Azure AI Studio, PromptFlow, and the RAG framework, achieving 90% accuracy in classifying purchases and driving cost savings.
- Enhanced CPR manual content by generating profession-specific GenAI images through Azure AI Studio, improving reader comprehension by 30%.
- Collaborated with stakeholders to define business objectives, translating them into actionable AI/ML workflows aligned with organizational goals.
- Extracted data from diverse sources, including databases, flat files, APIs, and external data sources, using SQL and NoSQL tools such as MySQL, PostgreSQL, MongoDB, and Cassandra.
- Conducted EDA using Tableau and Power BI to identify patterns, trends, and anomalies, creating interactive dashboards for monitoring LMS applications.
- Deployed machine learning models to production environments using Azure Machine Learning, Flask, and FastAPI, ensuring seamless integration and scalability.
- Optimized system performance by implementing advanced ETL pipelines and automation workflows using CI/CD pipelines and tools like Airflow and Azure DevOps.
- Monitored deployed models for drift, retraining them as needed to maintain performance and ensure alignment with evolving data trends.
- Documented AI/ML workflows, ETL pipelines, and process dependencies for efficient knowledge transfer and streamlined troubleshooting.
- Delivered AI/ML projects using Agile and Scrum methodologies, ensuring iterative development and timely delivery of high-impact solutions.