Sudheer Kumar Tatavalu

New Haven, CT | (203) 444-2434 | sudheerkumartatavalu@gmail.com
linkedin.com/in/sudheerkumartatavalu | github.com/sudheerkumartatavalu |
sudheerkumartatavalu.github.io/data-scientist

Professional Summary:

Insight-driven Data Scientist with proven experience deploying machine learning solutions in cloud environments. Skilled in Python, R, and SQL, with hands-on expertise in Spark, Flask, and AWS (SageMaker, EC2, DynamoDB). Successfully designed and deployed end-to-end ML pipelines for fraud detection, recommendation systems, and real-time logistics optimization at scale. Adept at combining statistical analysis, NLP, and cross-functional collaboration to drive automation, enhance decision-making, and deliver measurable business value.

Technical Skills:

- **Programming Languages:** Python, R, SQL, HTML, JavaScript
- **Data Manipulation & Analysis**: Pandas, NumPy, dplyr (R), data.table (R)
- **Data Visualization**: Matplotlib, Seaborn, Plotly, Tableau, Power BI
- **Machine Learning Libraries:** Scikit-learn, XGBoost, TensorFlow, Keras, PyTorch
- **Statistical Analysis:** Hypothesis Testing, Regression, ANOVA, Time Series Forecasting
- **Deep Learning & NLP:** Neural Networks, CNNs, RNNs, Natural Language Processing (spaCy, NLTK, BERT)
- **Big Data Technologies**: Apache Spark, Hadoop, MongoDB, Cassandra
- **Databases**: MySQL, PostgreSQL, Microsoft SQL Server
- **Cloud Platforms:** AWS (S3, EC2, SageMaker), Microsoft Azure
- **Version Control & Collaboration:** Git, GitHub
- **Tools & Platforms**: Jupyter Notebook, Google Colab, VS Code, Git, GitHub, Flask, ClearML, Docker (basic), Amazon QuickSight
- **Soft Skills:** Communication, Problem Solving, Critical Thinking, Cross-functional Collaboration

Work Experience:

Transportation Analyst August 2020 – July 2023
Amazon Development Centre (India) Pvt. Ltd, Hyderabad, India

- Enhanced operational efficiency by 25% using Python and SQL to analyze logistics data and drive optimization.
- Built and maintained interactive Power BI dashboards for real-time transportation performance tracking.
- Applied machine learning and statistical modeling to optimize route planning and resource allocation.
- Deployed a real-time logistics optimization model using Flask on AWS EC2, improving anomaly detection.
- Used ClearML to manage NLP experiment tracking, metric monitoring, and version control during internal model iterations.
- Collaborated with cross-functional teams to improve data-driven decision workflows across operations.
- Conducted exploratory data analysis (EDA) to uncover performance trends and improve planning accuracy.
- Automated periodic data reporting workflows by integrating Python scripts with AWS S3 and scheduling with CRON, reducing manual reporting time by over 40%.
- Contributed to cross-team initiatives to refactor and standardize ML model deployment practices across logistics analytics, improving model reusability and version control.

Associate Analyst October 2018– October 2019

GlobalLogic Technologies Limited | Hyderabad, India

- Conducted in-depth data analysis on client projects, employing statistical modeling to enhance decision-making processes through actionable insights.
- Built and maintained data visualization tools in Tableau and Power BI to track key performance indicators (KPIs) and provide business intelligence to stakeholders.
- Assisted in the design and implementation of machine learning models for predictive analytics, improving accuracy of forecasts and operational strategies.
- Developed and maintained efficient ETL pipelines to streamline data processing, ensuring high-quality data for analysis and reporting.
- Led efforts to enhance data management practices, ensuring data integrity and consistency across various platforms through rigorous data governance.

Education:

University of New Haven West Haven, CT

M.S. in Data Science, GPA: 3.93/4.0 August 2023 – May 2025

Projects:

Credit Card Fraud Detection (Python, Scikit-Learn) GitHub(2024)

- Developed a logistic regression model to detect fraudulent credit card transactions with 98% accuracy.
- Conducted feature engineering and model evaluation, visualized results through interactive Tableau dashboards.

Power BI Health Analytics GitHub(2024)

- Analyzed U.S. healthcare data (2019–2020) to assess hospital performance and patient outcomes.
- Built dynamic dashboards and DAX-based KPIs to support operational and strategic decision-making.

YouTube Data Analysis (AWS, PySpark, Amazon QuickSight) GitHub(2024)

- Designed a cloud-based ETL pipeline to process structured and semi-structured YouTube data using PySpark and AWS S3.
- Delivered real-time performance analytics through Amazon QuickSight dashboards, enabling content optimization.

AWS-Based Recommender System Infrastructure (S3, DynamoDB, SageMaker) Github**(**2025)

- Designed and implemented a cloud-native architecture for a content-based recommender engine using Amazon S3 for scalable data storage and DynamoDB for fast, structured metadata access.
- Trained and deployed the recommendation model on AWS SageMaker, optimizing inference time and resource utilization for real-time performance.

Team-Based ML Deployment Pipeline (Flask, AWS) Github(2025)

- Collaborated with a 3-member cross-functional team to deploy machine learning models as RESTful APIs using Flask, enabling seamless integration into existing AWS infrastructure.
- Developed and linked real-time dashboards for monitoring predictions and model metrics, enhancing visibility for stakeholders and non-technical users.

Certifications:

- Google Data Analytics – Coursera, 2025
  Gained practical experience in SQL, data wrangling, and visualization using R and Tableau
- Machine Learning on AWS – Coursera, 2025
  Trained and deployed ML models using SageMaker and explored core AWS ML workflows.