Sri Hari Shankar Reddy, Mandapati

sriharishankar02.m@gmail.com | 980-370-1238 | North Carolina, USA | Linkedin | Portfolio

PROFESSIONAL SUMMARY

- Data Scientist with 3+ years of experience building machine learning models, statistical methods, and scalable data pipelines to drive business impact.
- Holds a Master's in Computer Science with a Data Science specialization from UNC Charlotte.
- Thrives on solving complex problems and translating data into actionable insights that inform business strategy.
- Skilled in supervised/unsupervised learning, time series forecasting (ARIMA, GARCH), and recommender systems, with practical experience in deep learning for NLP (LLMs, BERT, LSTM, GRU), Generative AI, and computer vision (CNNs).
- Strong foundation in statistical modeling, including ANOVA, PCA, clustering, and discriminant analysis, and experienced in Agile environments with strong communication skills.

SKILLS

- **Languages:** Python, Java, R, SAS, SQL, T-SQL
- **Machine Learning & AI:** Regression, Decision Trees, Random Forest, SVM, XGBoost, Neural Networks, Ensemble Methods, Recommender Systems, Time Series (ARIMA, GARCH), NLP, Sentiment Analysis, LLMs (BERT, GPT), Generative AI, CNNs.
- **Statistical Methods:** Hypothesis Testing, ANOVA, PCA, Clustering, Dimensionality Reduction, Cross-Validation, Time Series Analysis, Discriminant Analysis
- **Big Data & Cloud:** Hadoop, Spark, HiveQL, HDFS, PySpark, MapReduce, AWS (S3, EC2, Glue, Redshift), GCP
- **Databases:** MySQL, PostgreSQL, Oracle, HBase, Amazon Redshift, MS SQL Server, Teradata
- **ETL & Reporting:** Airflow, SSIS, SSRS, SQL Server
- **Data Visualization:** Tableau, Power BI, MatPlotLib, Seaborn, ggplot2
- **Deployment & Automation:** Docker, Apache Airflow, Git, CI/CD Pipelines
- **Operating Systems:** Linux

PROFESSIONAL EXPERIENCE

Providence Health & Services, *Data Scientist*   Apr 2024 - Present | Renton, USA
Responsibilities:

- Built and deployed ML models using Python, scikit-learn, TensorFlow, and Keras to predict patient outcomes and reduce hospital readmission rates.
- Automated data pipelines with Apache Spark, Hadoop, and Airflow to optimize data ingestion and model training workflows.
- Applied statistical models to forecast treatment success and optimize resource allocation.
- Implemented NLP methods (spaCy, NLTK, Transformers) to extract insights from unstructured EMRs and patient feedback.
- Developed fraud detection and sentiment analysis systems using XGBoost, Beautiful Soup, and Tweepy; visualized insights via Tableau, Power BI, and Matplotlib dashboards.
- Engineered and maintained data lakes and warehouses with Amazon Redshift and Snowflake; enabled real-time analytics using Kafka, Flink, and GCP Pub/Sub.

- Applied CNN-based deep learning models (PyTorch, Keras) for medical image analysis, aiding diagnostic accuracy.

The Bank of New York Mellon Corporation, *Data Scientist Intern*   Dec 2023 – Mar 2024 | Pittsburgh, USA
Key Contributions:
- Built fraud detection models using Scikit-learn, handling imbalanced data with SMOTE and cost-sensitive learning. Developed predictive models using Spark and MapReduce on Hadoop (AWS).
- Wrote Spark SQL queries and performed data cleaning, feature engineering, and model tuning to improve fraud prediction accuracy.
- Applied ML algorithms including Logistic Regression, Random Forest, SVM, and XGBoost for loan default classification.

Client: ManipalCigna Health Insurance, *Data Analyst/ Data Scientist*   Aug 2022 – Aug 2023 | Mumbai, India
Responsibilities:
- Analyzed large healthcare datasets using SQL, Excel, and Tableau to extract insights and improve claims, customer health, and network strategies.
- Built automated dashboards and conducted statistical modeling with Python and R to support KPI tracking and decision-making.
- Performed customer segmentation and A/B testing using machine learning and analytics tools (Google Analytics, Python, SQL).
- Designed ETL processes to integrate multi-source data and ensured data accuracy using Hadoop, AWS Redshift, and big data tools.
- Delivered executive-level reports and visualizations (Tableau, PowerPoint), and conducted financial modeling for pricing and revenue optimization.

Client: Johnson & Johnson, *Data Analyst*   Jun 2021 – Jul 2022 | Mumbai, India
Responsibilities:
- Built ETL packages using SSIS (Pivot, Fuzzy Lookup, Aggregate, Execute SQL) to load data into data warehouses and marts.
- Developed and optimized complex T-SQL queries, stored procedures, CTEs, and views for efficient reporting and data transformation.
- Created dynamic and parameterized reports (P&L, EBIT, ROIC) using SSRS and Excel; distributed reports via SharePoint subscriptions.
- Migrated and processed data from SAS to SQL Server; used SAS for statistical analysis and financial modeling (regression, clustering, discriminant analysis).
- Delivered data insights and ad hoc reports to support mortgage product analytics and business strategies.

EDUCATION

Master of Science | Major: Computer Science | Specialization: Data Science
University of North Carolina at Charlotte
- **Related coursework:** Artificial Intelligence, Visual Analytics, Knowledge discovery in databases, Database Systems

CERTIFICATES

- **AWS Cloud Practitioner,** Amazon Web Services. [View Certificate]