Sri Hari Shankar Reddy Mandapati

✉ sriharishankar02.m@gmail.com 📞 +1 (980)-946-5167 🔗 Linkedin ⬙ Github 🔗 MyPortfolio

PROFESSIONAL SUMMARY

- Data Scientist with 3+ years of experience designing and deploying machine learning models, statistical methods, and scalable data pipelines to drive business impact. Holds a Master's in Computer Science with a focus in Data Science from UNC Charlotte.
- Thrives on solving complex problems and translating data into actionable strategy.
- Skilled in a wide range of ML techniques including 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, 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, Postgre SQL, Oracle, HBase, Amazon Redshift, MS SQL Server, Taradata
ETL & Reporting: Airflow, SSIS, SSRS, SQL Server, Excel
Data Visualization: Tableau, MatPlotLib, Seaborn, ggplot2
Deployment & Automation: Docker, Apache Airflow, Git, CI/CD Pipelines
Operating Systems: Linux, Windows, UNIX Shell (PuTTY), PowerShell

PROFESSIONAL EXPERIENCE

Providence Health & Services, *Data Scientist* ⬀   Apr 2024 – present | Renton, USA
Responsibilities:

- Built and deployed ML models (Python, scikit-learn, TensorFlow, Keras) to predict patient outcomes and reduce readmission rates.
- Automated data pipelines using Apache Spark, Hadoop, and Airflow to streamline healthcare data ingestion and model training.
- Applied statistical and regression analysis to forecast treatment success and optimize resource utilization.
- Implemented NLP techniques (spaCy, NLTK, Transformers) to analyze unstructured EMR and patient feedback for clinical insights.
- Developed fraud detection systems for insurance claims using XGBoost; collaborated on data lake/warehouse design (Redshift, Snowflake).
- Created interactive dashboards using Tableau, Power BI, and Matplotlib for operational, financial, and care-performance reporting.
- Integrated real-time health data streams with Kafka, Kinesis, and Flink to support early alerts and intervention.
- Conducted sentiment analysis and patient engagement studies using Tweepy, BeautifulSoup, and integrated analytics from Google Analytics and Mixpanel.

The Bank of New York Mellon Corporation,   Dec 2023 – Mar 2024 | Pittsburgh, USA
Data Scientist Intern ⊠
- Processed highly imbalanced fraud datasets using SMOTE, undersampling, and cost-sensitive algorithms in Scikit-learn.
- Wrote optimized Spark SQL queries for fraud analytics and business reporting.
- Built predictive ML modules using MapReduce/Spark in Hadoop (AWS environment).
- Ensured data quality through cleaning and transformation using Pandas and NumPy.
- Engineered features via normalization, encoding, and intersection techniques with Scikit-learn.
- Boosted fraud prediction accuracy through feature selection using Random Forest and Gradient Boosting.
- Applied ML models including Logistic Regression, Naïve Bayes, KNN, SVM, Random Forest, and XGBoost to predict loan defaults.

Client: ManipalCigna Health Insurance,   Aug 2022 – Aug 2023 | Mumbai, India
Data Analyst/ Data Scientist ⊠
- Analyzed large healthcare datasets using SQL, Excel, and Tableau to improve operational strategies across claims, customer health, and network performance.
- Built automated dashboards to track KPIs (e.g., claim processing time, satisfaction scores) and supported strategic decisions with Python and R-based statistical modeling.
- Conducted customer segmentation and retention analysis using ML algorithms to optimize product targeting and marketing effectiveness.
- Designed ETL pipelines to unify claims, demographic, and provider data for accurate, timely reporting.
- Led A/B testing for new health services and campaigns using Google Analytics and Excel; analyzed survey data with Python to improve customer experience.
- Utilized Hadoop and AWS Redshift to manage and query large-scale datasets, ensuring data accuracy and performance optimization.
- Delivered executive-level insights and reports using Tableau and PowerPoint to inform pricing strategy, resource allocation, and service delivery improvements.
- Performed financial modeling to assess claims cost trends, revenue performance, and pricing strategy outcomes.

Client: Johnson & Johnson, *Data Analyst* ⊠   Jun 2021 – Jul 2022 | Mumbai, India
- Developed ETL pipelines using SSIS to validate, extract, transform, and load data into Data Warehouses and Data Marts.
- Wrote and optimized complex T-SQL queries, stored procedures, CTEs, and functions to support reporting and performance tuning.
- Migrated data from SAS to SQL Server; used SAS/SQL for detailed data extraction and statistical analysis.
- Designed and delivered dynamic reports and dashboards (SSRS, Excel) including P&L, EBIT, and ROI metrics for leadership decision-making.
- Automated monthly performance reports with drilldowns and parameterized views, distributed via SharePoint subscriptions.
- Conducted advanced statistical analysis (regression, clustering, discriminant analysis) using SAS for financial and mortgage product modeling.

EDUCATION

Master of Science in Computer Science, *Specialization: Data*   Aug 2023 – Dec 2024
Science, University of North Carolina at Charlotte ⊠
- Related coursework: Artificial Intelligence, Visual Analytics, Knowledge discovery in databases, Database Systems

CERTIFICATES

AWS Cloud Practitioner ⊠   **Artificial Intelligence**
Amazon Web Services   **Foundation** ⊠
                                           NASSCOM (IT-ITeS Sector Skills
                                           Council)