




# Sri Hari Shankar Reddy Mandapati

✉ sriharishankar02.m@gmail.com ☎ +1 (980)-946-5167  LinkedIn  Github  MyPortfolio

## PROFESSIONAL SUMMARY

---

- Data Scientist with over 3+ years' experience in Data Extraction, Data Modelling, Data Wrangling, Statistical Modeling, Data Mining, Machine Learning and Data Visualization.
- Expertise in transforming business resources and requirements into manageable data formats and analytical models, designing algorithms, building models, developing data mining and reporting solutions that scale across a massive volume of structured and unstructured data.
- Proficient in managing entire data science project life cycle and actively involved in all the phases of project life cycle including data acquisition, data cleaning, data engineering, features scaling, features engineering, statistical modeling, testing and validation and data visualization.
- Proficient in Machine Learning algorithm and Predictive Modeling including Regression Models, Decision Tree, Random Forests, Sentiment Analysis, Naïve Bayes Classifier, SVM, Ensemble Models.
- Proficient in Statistical Methodologies including Hypothetical Testing, ANOVA, Time Series, Principal Component Analysis, Factor Analysis, Cluster Analysis, Discriminant Analysis.
- Knowledge on time series analysis using AR, MA, ARIMA, GARCH and ARCH model.
- Knowledge on Natural Language Processing (NLP) algorithm and Text Mining.
- Worked in large scale database environment like Hadoop and MapReduce, with working mechanism of Hadoop clusters, nodes and Hadoop Distributed File System (HDFS).
- Strong experience with Python (2.x, 3.x) to develop analytic models and solutions.
- Proficient in Python 2.x/3.x with SciPy Stack packages including NumPy, Pandas, SciPy, Matplotlib and Python.
- Working experience in Hadoop ecosystem and Apache Spark framework such as HDFS, MapReduce, HiveQL, SparkSQL, PySpark.
- Very good experience and knowledge in provisioning virtual clusters under AWS cloud which includes services like EC2, S3, and EMR.
- Proficient in data visualization tools such as Tableau, Python Matplotlib, R Shiny to create visually powerful and actionable interactive reports and dashboards.
- Excellent Tableau Developer, expertise in building, publishing customized interactive reports and dashboards with customized parameters and user - filters using Tableau (9.x/10.x).
- Experienced in Agile methodology and SCRUM process.
- I can easily articulate business problems with my storytelling skills and solve them with my quantitative skills.

## SKILLS

---

**Databases:** MySQL, PostgreSQL, Oracle, HBase, Amazon Redshift, MS SQL Server 2016/2014/2012/2008 R2/2008, Taradata

**Statistical Methods:** Hypothetical Testing, ANOVA, Time Series, Confidence Intervals, Bayes Law, Principal Component Analysis (PCA), Dimensionality Reduction, Cross-Validation, Auto-correlation

**Machine Learning:** Regression analysis, Bayesian Method, Decision Tree, Random Forests, Support Vector Machine, Neural Network, Sentiment Analysis, K-Means Clustering, KNN and Ensemble Method, Natural Language Processing (NLP)

**Hadoop Ecosystem:** Hadoop 2.x, Spark 2.x, MapReduce, Hive, HDFS, Sqoop, Flume


**Reporting Tools:** Tableau Suite of Tools 10.x, 9.x, 8.x which includes Desktop, Server and Online, Server Reporting Services (SSRS) Data Visualization: Tableau, Matplotlib, Seaborn, ggplot2

**Languages:** Python (2.x/3.x), R, SAS, SQL, T-SQL

**Operating Systems:** PowerShell, UNIX/UNIX Shell Scripting (via PuTTY client), Linux and Windows

## PROFESSIONAL EXPERIENCE

---

**Client: Providence Health & Services, Data Scientist** 

Apr 2024 – present | Renton, USA

**Description:** Providence Health & Services is a comprehensive, nonprofit healthcare organization operates a wide range of healthcare services, emergency care, cancer treatment, cardiology, women's health, paediatric care, and mental health services. I am responsible to look at health-related information from patients to identify patterns and trends that can help improve patient care and outcomes and looking at historical data, you would help predict future trends in healthcare.

### Responsibilities:

- Developed machine learning models using Python, scikit-learn, TensorFlow, and Keras to predict patient outcomes, optimize treatment plans, and improve operational efficiency in healthcare services.
- Applied statistical analysis and regression techniques to predict patient behaviour, treatment success, and healthcare resource usage, improving patient care and reducing hospital readmission rates.
- Built and automated data pipelines with Apache Spark, Hadoop, and Airflow to streamline data collection, processing, and model training for healthcare applications and conducted healthcare risk analysis and scenario modelling using Excel, MATLAB, and R to evaluate patient risks, forecast treatment costs, and improve healthcare provider decision-making.
- Implemented natural language processing (NLP) techniques using spaCy, NLTK, and Transformers to analyse unstructured medical records, patient feedback, and clinical notes for insights into patient care and treatment outcomes.
- Developed and deployed fraud detection systems using machine learning algorithms in Python to identify anomalies in patient claims and insurance fraud, collaborating with data engineers to build data lakes and data warehouses using Amazon Redshift and Snowflake, created interactive data visualizations using Tableau, Power BI, and Matplotlib to provide actionable insights on patient care, hospital operations, and financial performance to senior management.
- Designed algorithms for patient sentiment analysis by analysing patient feedback and surveys using Beautiful Soup and Tweepy to gauge patient perceptions and improve healthcare services.
- Utilized big data technologies like Apache Kafka and Amazon Kinesis to stream and analyse real-time healthcare data for better decision-making in patient care and treatment plans.
- Integrated data from Google Analytics, Mixpanel, and SQL databases to automate reporting systems, improving the tracking of healthcare service performance and patient engagement.
- Leveraged machine learning models in Python and XGBoost to predict patient admission rates and optimize resource allocation across hospital departments, enhancing operational efficiency and Integrated real-time healthcare data streams using Apache Flink and Google Cloud Pub/Sub to monitor patient vital signs and provide immediate alerts to medical staff for critical care interventions and Implemented deep learning models with Keras and PyTorch to analyse medical imaging data, assisting radiologists in detecting anomalies and improving diagnostic accuracy.

**Environment:** Python, scikit-learn, TensorFlow, Keras, Apache Spark, Hadoop, Airflow, spaCy, NLTK, Python, Amazon Redshift, Snowflake, Tableau, Power BI, Matplotlib, Tweepy, XGBoost, Apache Flink, Google Cloud Pub/Sub, Keras, PyTorch, Power BI, SQL.

**BNY Mellon, Data Scientist** 

Dec 2023 – Mar 2024 | Pittsburgh, USA

**Description:** BNY Mellon is a global leader in investment and securities services, offering innovative solutions in asset management, wealth management, and treasury services to help clients achieve their financial goals. Developed predictive models and machine learning solutions for fraud detection and loan default prediction using Python, Spark, and AWS, applying advanced techniques like SMOTE, feature engineering, and ensemble methods for improved performance.

### Responsibilities:

- Tackled highly imbalanced Fraud dataset using under sampling, oversampling with SMOTE and cost sensitive algorithms with Python Scikit-learn.
- Wrote complex Spark SQL queries for data analysis to meet business requirement.
- Developed MapReduce/Spark Python modules for predictive analytics & machine learning in Hadoop on AWS.
- Worked on data cleaning and ensured data quality, consistency, integrity using Pandas, Numpy.

- Participated in feature engineering such as feature intersection generating, feature normalize and label encoding with Scikit-learn preprocessing.
- Improved fraud prediction performance by using random forest and gradient boosting for feature selection with PythonScikit-learn.
- Performed Naïve Bayes, KNN, Logistic Regression, Random Forest, SVM and XGboost to identify whether a loan will default or not.

**Client: ManipalCigna Health Insurance,**

Aug 2022 – Aug 2023 | Mumbai, India

*Data Analyst/ Data Scientist* ☑

**Description:** ManipalCigna Health Insurance is a prominent health insurance provider offering a range of health plans designed to meet the diverse needs of individuals, families, and businesses. Identifying patterns, trends, and relationships within the data through visualization techniques and selecting appropriate statistical and machine learning algorithms based on the project goals and continuously monitoring model performance and updating them as needed.

**Responsibilities:**

- Analysed large health-related datasets to extract actionable insights, improving operational strategies across customer health, claims, and network performance using SQL, Excel, and Tableau.
- Developed and maintained automated dashboards to track KPIs such as claim processing time, customer satisfaction, and network efficiency for various departments and applied Python and R for advanced data analysis, cleaning, and statistical modelling to inform strategic decisions and improve claims processing efficiency.
- Conducted customer segmentation analysis with machine learning algorithms to uncover patterns and optimize targeting strategies for health products and retention efforts.
- Designed and optimized ETL processes to integrate data from multiple sources, including claims data, customer demographics, and provider network performance for accurate reporting.
- Led A/B testing for new health products, services, and marketing campaigns using Google Analytics and Excel, delivering detailed performance analysis to guide decision-making and collaborated with customer service and marketing teams to create customer satisfaction surveys, analysing feedback data using SQL and Python to drive service improvements.
- Monitored and reported on operational performance metrics using Hadoop and Big Data technologies, ensuring data accuracy for real-time decision-making in healthcare operations and utilized AWS Redshift to manage and analyse large health datasets, ensuring high data quality and optimizing query performance for better insights.
- Prepared comprehensive reports and executive summaries for senior leadership, leveraging Tableau and PowerPoint to present key insights and recommendations for operational improvements.
- Analysed financial health data, including claims costs and revenue, to optimize pricing strategies and market share using Excel and Python for detailed financial modelling.

**Environment:** SQL, Excel, Tableau, KPIs, Python, R, ETL, CRM, Hadoop, Big Data technologies, AWS Redshift, PowerPoint, machine learning algorithms.

**Johnson & Johnson,** *Data Analyst* ☑

Jun 2021 – Jul 2022 | Mumbai, India

**Description:** Johnson & Johnson is a global healthcare leader known for its innovative products across pharmaceuticals, medical devices, and consumer health sectors. Involved in developing and optimizing data integration processes, creating detailed financial reports, and performing advanced statistical analyses to support business decision-making.

**Responsibilities:**

Used SSIS to create ETL packages to Validate, Extract, Transform and Load data into Data Warehouse and Data Mart.

Maintained and developed complex SQL queries, stored procedures, views, functions and reports that meet customer requirements using Microsoft SQL Server 2008 R2.

Created Views and Table-valued Functions, Common Table Expression (CTE), joins, complex subqueries to provide the reporting solutions.

Optimized the performance of queries with modification in T-SQL queries, removed the unnecessary columns and redundant data, normalized tables, established joins and created index.

Created SSIS packages using Pivot Transformation, Fuzzy Lookup, Derived Columns, Condition Split, Aggregate, Execute SQL Task, Data Flow Task and Execute Package Task.

Migrated data from SAS environment to SQL Server 2008 via SQL Integration Services (SSIS).

Developed and implemented several types of Financial Reports (Income Statement, Profit& Loss Statement, EBIT, ROIC Reports) by using SSRS.

Developed parameterized dynamic performance Reports (Gross Margin, Revenue base on geographic regions, Profitability based on web sales and smartphone app sales) and ran the reports every month and distributed them to respective departments through mailing server subscriptions and SharePoint server.

Designed and developed new reports and maintained existing reports using Microsoft SQL Reporting Services (SSRS) and Microsoft Excel to support the firm's strategy and management.

Created sub-reports, drill down reports, summary reports, parameterized reports, and ad-hoc reports using SSRS.

Used SAS/SQL to pull data out from databases and aggregate to provide detailed reporting based on the user requirements.

Used SAS for pre-processing data, SQL queries, data analysis, generating reports, graphics, and statistical analyses.

Provided statistical research analyses and data modeling support for mortgage product.

Perform analyses such as regression analysis, logistic regression, discriminant analysis, cluster analysis using SAS programming.

Environment:SQL Server 2008 R2, DB2, Oracle, SQL Server Management Studio, SAS/ BASE, SAS/SQL, SAS/Enterprise Guide, MS BI Suite (SSIS/SSRS), T-SQL, SharePoint 2010, Visual Studio 2010, Agile/SCRUM

## EDUCATION

---

**Masters in Computer Science and Specialization in 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 ☑