# SRI SAI SUBHASH KASIREDDY

Pleasanton, CA | Phone: (857) 419-2141 | E-Mail | LinkedIn | GitHub

## Professional Summary

Experienced Data Engineer with around 5 years of industrial experience specializing in big data analysis, I have honed my expertise in data mining, statistical inference, A/B testing, machine learning, and ETL data pipelines. Proficient in SQL, including MySQL and MS SQL Server, I excel in crafting stored procedures, triggers, and complex queries for insightful reporting. Leveraging Tableau, I have designed dashboards that facilitated strategic decision-making for diverse stakeholders. My proficiency in Python spans the entire data science project lifecycle, encompassing data acquisition, cleaning, exploration, and modeling using Pandas and Scikit-learn. Furthermore, I possess hands-on experience in statistical analysis, hypothesis testing, and machine learning model development, including Linear Regression, Logistic Regression, Random Forest, XGboost, K-means, KNN, and Neural Networks. Additionally, I have worked with NoSQL databases and big data tools such as Hadoop, Hive, and Spark. Cloud platform familiarity includes AWS and Google Cloud, complemented by Shell Scripting and version-control proficiency in Linux and Git. Detail-oriented and results-driven, I have a proven track record of translating business requirements into technical solutions, designing KPIs, and collaborating effectively within cross-functional teams.

## SKILLS

**Data Visualization (Business Intelligence) Tools and Packages:** PowerBI, Tableau, seaborn, ggplot2, plotly, matplotlib, Data Wrapper, pandas, NumPy, scikit-learn, TensorFlow, keras, SciPy.

**Big Data Tools:** MongoDB, Hive, Presto, Big Query, Hadoop, MapReduce, HDFS, Sqoop, PIG, HBase, Apache Airflow, Apache Spark, Flask, Apache Kafka, Oozie, Flume, NiFi, Yarn, Sparklib, Zookeeper, Mahout.

**IDE:** Jupyter Notebook, Google Collab, Visual Studio Code.

**Cloud:** GCP (Big Query, Kubernetes), AWS(EC2, S3, RDS RedShift, EMR).

**Object Oriented Programming:** Data Structures and Algorithms with Python, Java, and C++.

**Machine Learning:** Regression, Classification, Clustering, Dimensionality Reduction (PCA), Deep Neural Networks, Data Science Pipeline (Cleansing, Data Processing, Data Mining, Wrangling, Visualization, Modeling, Interpretation), Statistics, Statistical Analysis.

**Web Technologies:** HTML, CSS, JavaScript.

**Version Control:** Git, GitBucket.

**Agile:** Jira, Confluence.

**Database & Data Warehouse:** MySQL, PostgreSQL, MS SQL Server, AWS RDS, AWS Redshift, AWS Redis.

**MS Office and Tools:** Word, PowerPoint, Excel, Docker, Kubernetes.

## EXPERIENCE

**McKinsey & Company, Boston, USA** **Jan 2022 – Present**

*Data Analyst*

- Created and analyzed business requirements to compose functional and implementable technical data solutions.
- Created new data constraints and leveraged existing constraints for reuse.
- Created data dictionary, Data mapping for ETL and application support, DD, ERD, mapping documents, metadata, DDL and DML as required.
- Performed data quality issue analysis using SnowSQL by building analytical warehouses on Snowflake.
- Implemented complex SnowSQL scripts in snowflake cloud data warehouse to business analysis and reporting.
- Performed statistical analysis such as hypothesis testing casual inference and Bayesian analysis.
- Designed metrics for A/B testing, creating dashboards in Tableau to monitor test processes, analyzing test results, interpret and give recommendations to stakeholders.
- Created a long-term vision plan for future working solutions and clear documentation of present work and research.
- Ensured seamless deployment of analytical services, contributing to enhanced efficiency and reduced user-reported issues.
- Achieved an accuracy of 98.7% to predict future application downtime using an XGBoost Model with feature engineering, details from the wide and deep model embedding matrix and overlaps using an Apriori algorithm.
- Spearheaded the building and deployment of a Flask application to automate validation of metrics for insurance business, decreasing turnaround time by more than 90% and boosting accuracy by over 25%.

- Designed an end-to-end PyTorch Natural Language Processing library for text classification to finetune Hugging Face Transformer models, reducing development time by 90%.
- Implemented automation solutions for routine tasks, reducing manual errors by 30% and increasing overall efficiency.
- Developed dashboards using Tableau and present data to track KPI's and product performance using data from various sources such as MS Excel, AWS S3, AWS RDS, AWS Redshift, JSON and XML.
- Deployed Big Data tool Spark and Hive to analyze large datasets up to 2TB stored in Hadoop HDFS, including performing filtering and aggregation using SparkSQL based on Spark DataFrame.
- Analyzed discrepancies and synchronized the Staging, Development, UAT and Production DB environments with data models.
- Reviewed and revised data models for soundness of data structures and adherence to client standards.
- Wrote T-SQL queries involving multiple joins, subqueries and common table expression to perform data aggregation for reporting.
- Defined metrics to estimate impact of new features and give recommendations for business decisions based on data analysis.
- Participated in data project planning, gathering business requirements and translating them into technology requirements.

**JPB Solutions, Hyderabad, India** **Mar 2020 – Dec 2020**

*Data Analyst*

- Responsible for creating reporting dashboards, performing data mining and analysis to understand customer purchase behavior.
- Created real-time dashboards in Tableau to visualize and monitor key metrics and A/B test processing using both external and internal data.
- Analyzed large scale user log data and generated features for classification models using SparkSQL in Spark.
- Developed and implemented optimized data pipelines for ingesting, transforming and loading data into Snowflake, resulting in a 30% reduction in data processing time and improved data accuracy.
- Conducted root cause analysis on discrepancies within the WMS, identifying issues related to data entry errors and system integrations.
- Performed statistical analysis such as hypothesis testing, regression analysis, confidence interval and P-value calculation using R to find insights to increase click through rate and sales, and built web applications for ad-hoc interactive dashboard.
- Wrote scripts to store data into Hadoop HDFS from various sources including AES S3, AWS RDS and Web API and NoSQL Database MongoDB.
- Developed Python scripts to automate data validation and data cleaning processes such as deduplicating and checking data consistency using Pandas and Apache Airflow.
- Collaborated with cross-functional teams to implement inventory control strategies, resulting in a 10% reduction in carrying costs.
- Designed ETL specification documents to load the data in target using various transformations according to the business requirements.
- Performed Data profiling, Validation, and Integration. Involved in Data migration and Data distribution testing.
- Created materialized views to improve performance and tuned the database design.

**Tata Consultancy Services, Hyderabad, India** **Jul 2019 – Feb 2020**

*Assistant System Engineer*

- Conducted exploratory data analysis (EDA) on large datasets using Python, identifying trends, outlier problems, and key insights to support data-driven decision making.
- Created interactive and insightful data visualizations in Tableau, transformed the data for presentations and reports.
- Analyzed key performance indicators (KPIs) such as order fulfillment rates and inventory turnover to optimize warehouse operations.
- Used ETL tool SSIS to extract data from both external and internal sources including csv json excel and load data into database.
- Generated reports to track and report on the efficiency of WMS processes, leading to a 15% improvement in order processing time.
- Integrated Git for version control, facilitating collaboration with team members on code development.
- Designed and implemented analysis pipelines using JSON for iterative prototyping, enabling scalable insights generation and data-driven decision-making.

- Leveraged cutting edge GCP's data processing tools, including Google Cloud Dataflow and Dataproc, to clean, transform, and prepare raw data for analytics.

**Tech Mahindra, Hyderabad, India** **May 2018 – Jun 2019**

Data Analyst

- Developed various solution driven views and dashboards by developing different chart types including Pie Charts, Bar Charts, Tree Maps, Circle Views, Line Charts, Area Charts, Scatter Plots in PowerBI.
- Developed triggers on views to populate base tables.
- Rigorously tested and debugged the stored procedures and used triggers to test the validity of the data after the insert, update and delete.
- Create stored procedures and user defined functions to support efficient data storage and manipulation.
- Performed transformations like Pivoting, Unpivoting and Splitting of the columns using Power Query.
- Created different chart types and drill down reports using PowerBI desktop (SSRS).
- Published reports and visualizations to PowerBI Service and thereby created dashboards.
- Experienced in Data Modeling used to define and analyze data requirements needed to support the business processes within the scope of corresponding information systems in organizations.
- Strong Data Modeling experience using dimensional data modeling, star schema modeling, snow-flake modeling using tools like Erwin.
- Experience throughout developing OLAP cubes for data mining and generated reports from the OLAP cubes.

## Certification

**Microsoft Certified – PowerBI Data Analyst Associate** **Dec 2025**

## EDUCATION

**Northeastern University, Boston, MA**

*Masters in Data Analytics Engineering*

**GITAM University, Visakhapatnam, IN**

*Bachelor of Technology in Computer Science & Engineering*