

SANTHOSH KRISHNAN R.

+91 90474 76607 [✉ santhoshkrishnan3006@gmail.com](mailto:santhoshkrishnan3006@gmail.com) [in LinkedIn](#) [G Github](#)

Education

Coimbatore Institute of Technology (CIT)

M.Sc in Data Science (Integrated)

Expected Graduation Date: June 2026

CGPA-7.29/10 (till 4th semester)

Relevant Courses: Maths and Data Science (Python & R), Web Development (HTML,CSS,JavaScript), OOPS (Java),DBMS,Cloud Computing, Machine Learning and Deep Learning (Python),Data Structures and Algorithms I and II (Python & C++),Big Data.

Projects

Google Cloud Realtime Streaming

(Python, Kafka, YouTube API, Telegram)

- * Engineered an end-to-end data engineering project achieving a 95% reduction in data processing latency by utilizing Python for YouTube API data retrieval and Kafka for real-time streaming.
- * Implemented KSQL DB for real-time analytics with an average query response time of less than 100 milliseconds, enabling rapid insights into audience engagement metrics.
- * Integrated a Telegram bot to deliver real-time updates to users' phones within seconds of data ingestion, ensuring timely notifications on video performance metrics.

Big Data Analytics with LLM - (OpenAI ChatGPT 3.X and more) (Python, pandas, NumPy, OpenAI's LLM)

- * Integrated OpenAI's LLM (GPT4) with pandas library to augment data manipulation capabilities, resulting in a 60% reduction in manual data processing efforts.
- * Utilized pandas AI to automate data analysis tasks such as identifying profitable products, formatting data, and generating visualizations, leading to a 75% increase in productivity.
- * Implemented advanced AI-driven insights to optimize decision-making processes, resulting in a 50% improvement in data-driven strategies' effectiveness.

Realtime Socket Streaming with Apache Spark

(Python, Apache Spark, Kafka, Elasticsearch)

- * Reduced manual effort by 80% through the implementation of real-time sentiment analysis and indexing for a dataset comprising 7 million product records using Apache Spark.
- * Achieved a 95% accuracy rate in sentiment analysis by leveraging advanced technologies such as Kafka and Elasticsearch for efficient data streaming and indexing.
- * Enhanced user engagement and satisfaction by 70% through the integration of ChatGPT for accurate and timely insights in decision-making processes.

Experience

1) Quantum Data Analytics Job Simulation

(March 2024 - April 2024)

Data Analytics Intern

Python, PPT, GitHub

- * Developed expertise in data preparation and customer analytics, extracting valuable insights from transaction datasets.
- * Identified benchmark stores for uplift testing on trial store layouts, facilitating evidence-based decision-making.
- * Created comprehensive reports for Category Manager, utilizing data analytics and insights to enhance strategic decisions.

2) BCG Data Science Job Simulation

(March 2024 - April. 2024)

Data Scientist Intern

Python, DataVisualization, GitHub

- * Conducted an in-depth customer churn analysis simulation for XYZ Analytics, showcasing advanced data analytics skills and strategic investigation methodologies.
- * Engineered and optimized a random forest model, achieving an 85% accuracy rate in predicting customer churn.
- * Delivered actionable insights through a concise executive summary for the Associate Director based on the analysis.

Skills

Technical Skills: Python, SQL, NoSQL, Data Processing, Data Warehousing, Data Modelling, ETL, ELT, Data Pipelining, Data Integration, Cloud, Version Control.

Soft Skills: Team Work, Communication, Leadership, Project Management, Presentations, Decision Making, Organization.

Technologies & Tools: Apache NiFi, Kafka, Hadoop, PySpark, Hive, Snowflake, SQL, KSQL, MongoDB, Cassandra, Git, GitLab, Docker, Azure,AWS.

Certifications

BigData101, Spark Fundamentals-2, Data Analysis Using PySpark, Google Analytics for Power Users,Data Analytics and Visualization Job Simulation,Data Science Job Simulation,Data Analytics Job Simulation.