SRI SAI KRISHNA KATEGARU
[LinkedIn](#) || [Github](#)

+91-6305816104
krishnakategaru@gmail.com
msdsm2204107027@alum.iiti.ac.in

Profile (Introduction)

As a recent graduate with a MASTERS IN DATA SCIENCE AND MANAGEMENT from IIT – IIM Indore, I have a solid foundation in data analysis and project management. Certified Microsoft Associate Data Scientist with 7 years of experience driving data-driven solutions across supply chain, healthcare, retail, and various sectors. Proficient in the full stack data science cycle on cloud platforms such as Azure, encompassing data structuring, mining, ETL pipelines, statistical modeling, and predictive analytics, including expertise in LLM and multi-modal algorithms in GenAI. Proven ability to collaborate with diverse stakeholders to identify business problems and independently implement analytical solutions with thought leadership. Demonstrates strong problem-solving abilities, debugging skills, and a passion for continuous learning. Eager to contribute expertise to challenging roles in data analytics, with a focus on advancing skills in machine learning, deep learning, NLP, optimization, and Artificial Intelligence

Education

M.S in Data Science & Management (Regular)	**IIT Indore - IIM Indore**	**3.63/4.33**	**2024**
B. Tech (Electrical and Electronics Engineering)	**Sree Vidyanikethan Engineering College**	**7.66/10**	**2018**

Academic & Co-Curricular Activities

- Microsoft Certified for Associate Azure Data Scientist (DP-100) VALID and Associate Azure Data Engineer (DP-203) VALID
- GATE qualified in 2018 in EE and Infosys certified Scala Developer and Python Developer
- **Secured 1400 rank in Gate for Data science& AI**
- Annual Day function Organizer in College

Experience

Data Science and Gen AI **Senior Consultant, Capgemini** **Jan'23 – Present**

- For an automotive company (German client) , individually implemented data catalogue using **Agentic** Architecture basically a multi agent workflow on around **538 billion** records, which later developed as AWS Athena Code Generator and executor also firsthand reference for analysis to user who are responsible for KPI building day to day., deployed in AWS.
- Created and deployed MVP of **Multi Agentic Model capable of multimodal rag** for supply chain and presented to various clients in process of consulting for a german client
- Acted as techno-managerial lead for a team of 2 to 3 members in POV's for the clients achieving around **90 to 100%** success rate in project delivery
- Created and deployed an MVP for **KPI Automation (Dashboard automation) using Plotly and GenAI. Impact is expected to reduce work and delivery load from months to weeks (around 70%) for German health care client.**
- To increase productivity and 360-degree solution, Created and deployed a POV to automate the process of **Project Documentation automation and Test cases automation**
- Developed and deployed a **Retrieval-Augmented Generation (RAG)** pipeline based on GPT-4 and AI Search (**semantic**) into Azure App Services independently for a client with more than 80% metrics using RAGAS. Acted as Individual Contributor.
- Designed a **Mixed Integer Linear Programming (MILP) model** to automate route optimization in transportation and logistics by forecasting demand for each outlet, deployed via **FastAPI** and a React app on Azure.
- Pitched project to client on Procurement RFP Agent Proof of Concept (POC) combining **Intent Recognition, Conversational Agents, and Document Generation** for a client.
- Lead Time Forecasting: Forecasted lead times for automotive products and projected KPIs using dashboards in Grafana, deployed via FastAPI and a React app, improving forecasting accuracy by 12%.
- Produced reports, presentations, and other documentation to communicate findings with the management team.
- Tools**: Langchain, Streamlit, FastAPI, Python, Azure VM, Azure API Management, Azure Web Apps, ReactJS, MILP, Gurobi, PuLP, ML, NLP , OpenAI Models, Llama, Mistral.**

Data Scientist **Azure Data Scientist I, SKYPOINT Cloud Pvt Ltd** **Dec'21- Nov'22**

- **Azure Data Factory (ADF) Pipelines**:Created Azure Data Factory (ADF) pipelines to manage data flows using Azure Databricks, including data wrangling, cleaning, and exploration, with Durable Functions for ingesting data into Data Lake and Delta Lake.
- **Predictive Modeling Techniques**: Implemented predictive modeling techniques for **RFM (+10% more than base models' precision for more than 23.98% profiles), Customer Lifetime Value for more than 90+ profiles tested with lesser MSE, Churn, Sentiment Analysis for around 5800+ profiles with 82%+ metrics, and Product Recommendation Models,** built and deployed using **MLOps** techniques.
- Technologies: **Azure Machine Learning, Spark, ADF, Databricks, Durable Function Apps, Azure Cosmos DB, Azure DevOps CI/CD, Python, SQL, Pandas, ML Models, PySpark, MLflow, SparkNLP, AzureAPIM.**

Data Analyst & Data Scientist **Senior Systems Engineer, Infosys** **Jun'18 - Dec'21**

Project 1: Data Scientist, US-based Supply Chain and Logistics Client

- **Data Analysis and ETL**: Collaborated across three verticals (Energy, Chemical, Retail) to extract and analyze 30GB of source data, applying ETL processes using Azure Data Factory and MS SQL. This resulted in a 25% reduction in data processing time.
- **Exploratory Data Analysis (EDA):** Led a team of 3 members in EDA, including univariate, multivariate analysis, correlation, anomaly detection, and feature engineering. Achieved a 90% accuracy rate in identifying key data trends.
- **Machine Learning Model Development:** Modelled and deployed a range of machine learning algorithms, including Logistic Regression, Decision Trees, Boosting Techniques, Naive Bayes, Gradient Boosting, KNN, SVC, Random Forest, and LightGBM Classifier. Developed and deployed deep learning models, achieving a 12% increase in predictive accuracy.
- **Business Impact:** Reduced stockouts by up to 10%, increasing sales, and decreased holding costs by 5-20%. AI reduced human errors by 50-70% in testing.
- **Time Series Analysis**: Conducted demand forecasting using SARIMAX, improving forecast accuracy by 8%.

Project 2: Data Analyst, US-based Brewery Client

- **Power BI Reporting**: Developed Power BI reports in collaboration with senior managers, enhancing data visualization by 40%.
- **SQL Queries and KPI Dashboards:** Created SQL queries for KPI dashboards using Snowflake SQL, improving data retrieval efficiency by 20%.
- **Regression Modelling:** Validated and tested KPI dashboards and performed regression modelling to predict dollar per barrel, achieving a 5% increase in predictive

accuracy.

Project 3: Developer, Germany Client

- **Data Ingestion and Transformation:** Managed data ingestion from DB2 to HDFS using SQOOP, transforming data into various formats (parquet, Avro, CSV), reducing data processing time
- **Royalty Calculations and Predictions**: Analyzed and computed data for royalty calculations using Spark SQL, Hive, Spark, and Scala. Assisted in predicting royalties using regression models, achieving a 15% improvement in calculation accuracy.

Technical Skills

PROGRAMMING LANGUAGE	Python, R, Pyspark, C, Scala
DATA ANALYSIS AND MANIPULATION	NumPy, Pandas, SQL, Plotly
DATA VISUALIZATION	Matplotlib, Seaborn, Plotly, Tableau
MACHINE LEARNING FRAMEWORKS	Scikit Learn, TensorFlow, Keras, pyTorch
DEEP LEARNING	CNN, RNN, LSTM, GRU, Transformers
NATURAL LANGUAGE PROCESSING (NLP)	NLTK, SpaCy,Genism, Transfer Learning (BERT, Llama), NER, POS, Topic Modelling, Text preprocessing
COLLABORATION TOOLS	Jupyter Notebooks, Google Colab, Git , VsCode
DATABASE MANAGEMENT	MySQL, PostgresSql,Pinecone, Faiss, Qdrant, AWS Athena, Azure SQL
SOFT SKILLS	Communication, Event Management, Problem-Solving, Critical Thinking, Team Management
CLOUD	Azure Data factory, Azure ML, Azure AI Studio, Azure SQL, Azure APIM, Azure Functions, Azure Databricks, Azure AI Search, Azure Cognitive services
ADDITIONAL PYTHON MODULES	Langchain, PandasAI, Cognite. Llama index

Projects (In college / Personal)

- **Created ALGOTRADING ALOGORITHM for back testing Masters Thesis**
- Risk Prediction on Diabetic dataset, Pattern Recognition on MIST dataset, Sentiment Analysis on IMDB dataset**, Arthritis prediction using CV + DN**, **Algo Trading** Dashboard for price indicators, monte Carlo simulations.
- Have Basic level of knowledge in C language, Algorithms, sorting techniques, oops, Java, flask, react-native, Hibernate, Spring, Html

Extra Circular Activities

Class Representative	Infosys Training	2018
Volleyball player in Zonal	Keshava Reddy school	2011

Hobbies

Gym, Swimming, Snooker .