



Sarath Babu

Data Science and AI

babusarath05@gmail.com 

+91-8148956716 

Chennai, India 

7.5+ years of experience with Data Science and 1.5 years of experience with AI in building data-driven solutions for a variety of requirements in diverse domains. Proficiency in creating, developing, and deploying highly adaptive machine learning, deep learning and generative AI models and pipelines for various business solutions. Looking forward to opportunities to build my career and apply my skills in innovative ways.

SKILLS

1 Programming Languages

- Python
- R
- PySpark
- SQL

2 Domains

- Statistics
- Machine Learning
- Deep Learning
- Image Analytics
- Text Analytics
- Generative AI
- Artificial Intelligence
- Product Affinity (Apriori)

3 Data Visualization

- Tableau
- PowerBI
- Looker

4 Machine Learning

- Descriptive Analytics
- Predictive Analytics
- Prescriptive Analytics
- Data Pre-Processing
- Data Cleaning
- Model Building
- Model Validation
- Hyper Parameter Tuning
- Model Deployment
- Regression
- Classification
- Clustering
- Time Series Analysis
- Time Series Forecasting

5 Deep Learning

- Artificial Neural Networks (ANN)
- Tensorflow
- Keras
- PyTorch

6 Image Analytics

- Convolution Neural Networks (CNN)
- Image Processing
- OpenCV
- Object Detection
- YOLO
- SSD
- Computer Vision
- Face Recognition
- Vision Transformers

| | | | |
|--|--|---|--|
| <ul style="list-style-type: none"> Text Analytics | <ul style="list-style-type: none"> Natural Language Processing Recurrent Neural Networks (RNN) | <ul style="list-style-type: none"> Text Classification BERT | <ul style="list-style-type: none"> Word Embeddings Wake Word Detection from speech |
| <ul style="list-style-type: none"> Generative AI | <ul style="list-style-type: none"> Transformers GPT Speech to Text Langchain | <ul style="list-style-type: none"> Large Language Models Llama OpenAI Whisper FAISS | <ul style="list-style-type: none"> Gemini Retrieval Augmented Generation (RAG) Chatbot HuggingFace |
| <ul style="list-style-type: none"> Cloud | <ul style="list-style-type: none"> Azure Databricks AWS S3 AWS Sagemaker | <ul style="list-style-type: none"> Azure Data Factory AWS Athena Snowflake | <ul style="list-style-type: none"> Azure Machine Learning AWS Glue |
| <ul style="list-style-type: none"> Tools | <ul style="list-style-type: none"> Airflow | <ul style="list-style-type: none"> Docker | <ul style="list-style-type: none"> Git |
| <ul style="list-style-type: none"> Back End | <ul style="list-style-type: none"> Flask | <ul style="list-style-type: none"> Streamlit | <ul style="list-style-type: none"> Flutter (Flet) |
| <ul style="list-style-type: none"> Web Automation | <ul style="list-style-type: none"> Selenium | <ul style="list-style-type: none"> Beautiful Soup | |

Work Experience

Latent View Analytics

| | |
|--|---|
| 10/2023 - Present Domain: Computer Hardware | Gen AI Chatbot application for Data Analysis and visualization for an American Multinational Technology company <i>Achievements:</i> Incorporation of insights in decision making process for quality product managers <div> <ul style="list-style-type: none"> Front End Development with Streamlit, Flask and Back End Development with Python. SQL generation, summarization and visualization with GPT-4o, RAG with Sentence Transformer, FAISS and few shot prompts. </div> |
| 07/2023 – 09/2023 Domain: Computer Hardware | SLM Chatbot application for all Laptop and Desktop users of an American Multinational Technology company with Retrieval Augmented Generation <i>Achievements:</i> Single Source of Truth packaged application deployed for all laptops and desktops <div> <ul style="list-style-type: none"> Front End Development with Streamlit, Flask and Back End Development with Python. Text preprocessing and Embedding search with Sentence Transformer embeddings on FAISS vector store. </div> |

03/2023 – 06/2023
Domain:
Subscription Analytics

Churn Analysis for Search Engine of an American Multinational Technology Company

Achievements: 3% reduction in churns on targeted demographics

- Customer profiling with Clustering and Feature importance using ensemble tree models with Azure Databricks and PySpark

12/2022 – 02/2023
Domain:
Data Engineering

End to End Data Pipeline Deployment for an American Software Company

Achievements: Successful deployment of end-to-end pipeline

- Data Pipeline development and deployment with Airflow, Data Extraction with Snowflake, Data Transformation with Python and Data Visualization with Looker

05/2021 – 11/2022
Domain:
Supply Chain

Data Warehousing Development for Demand Planners for an American Multinational Technology Company

Achievements: 50% Reduction in manual efforts for Demand Planners

- Front End Development with Excel and Tkinter and Back End Development with Python and Xlwings
- Data Extraction from Teradata, AWS S3 & Snowflake

Domain:
Supply Chain

Time Series Forecasting for Tablet Devices for an American Multinational Technology Company

Achievements: Incorporation of Forecasting model as part of decision making

- Identification of closely resembling time series of other devices with DTW (Dynamic Time Warping)
- Time Series Forecasting with Barycenter Averaged time series

Domain:
Business Process Re-Engineering

Incident Analysis with Machine Learning Models for an American Multinational Technology Company

Achievements: 16-hour manual effort reduction for every quarter

- Incidents or Tickets visualization with Tableau Dashboards
- Identifying factors which results in tickets with Feature importance and Feature selection with Machine Learning Models

Tiger Analytics

10/2020 - 02/2021
Domain: **Retail**

Pipeline Creation using Azure Data Factory for an American Pet Care Company

Achievements: Automation with QC eliminated major manual intervention

- Migration from Pandas to Pyspark and data flow using Azure Databricks.
- Assembled the pipeline for ETL & End to End deployment in Azure Data Factory.

04/2020 - 09/2020
Domain: **Retail**

Product Affinity Model Development for an American Food Corporation

Achievements: Identified Paired Products and incorporated into Marketing Strategy

- Created modules and libraries for Apriori algorithm on Big Data in a Non-PySpark environment
- Product Affinity on county, state and national level and visualization of results using Tableau

01/2020 - 03/2020
Domain:
Transportation Engineering

Accident Visualization using Dashboard and Storyboard for an American State Transportation

Achievements: Identified County locations for major resource allocation

- Dashboard with Tableau, PowerBI and Storyboard with Tableau

09/2019 - 01/2020
Domain:
Insurance

Churn Possibility Prediction of High Premium Customers for an American Insurance Company

Achievements: Identified features which got incorporated into Marketing Strategy

- Feature Selection using Embedded methods.
- Predictive modeling using Ensemble models.

TATA Consultancy Services

05/2019 - 09/2019
Domain - **Banking**

Response Classification using Topic Modelling for an American Multinational Bank

Achievements: Reduction of Manual efforts for classification

- Implemented combination of Topic modeling methods (LDA & LSA).
- Response classification using Machine Learning Models.

02/2019 - 05/2019
Domain -
Engineering & Technology

Oil Rig Pipeline Corrosion Detection for an American Technology Company

- Image acquisition and processing with OpenCV.
- Image thresholding, Pooling and Flattening using Python.
- Corrosion detection using Machine Learning models.

09/2018 - 02/2019
Domain -
Powerplant Engineering

Reduction of NOx and CO Emissions for a Japan based Thermal Power Plant

Achievements: 5% reduction in emissions

- Data visualization and Outlier elimination.
- Machine Learning Model Selection and Hyper-Parameter Tuning.
- Created a unified model for NOx and CO predictive analytics.

02/2018 - 08/2018
Domain -
Powerplant Engineering

Clustering Analysis of Coal Feed for a Japan based Thermal Power Plant

Achievements: 5% cost optimization

- Data visualization and Outlier elimination.
- Clustering to classify coal based on qualities.

06/2017 - 01/2018
Domain - **Retail**

Data Visualization at Transactional Level for a UK based Retailer

Achievements: Insights incorporated into Decision Making

- Data visualization using Matplotlib, Seaborn and Plotly.

EDUCATION

06/2014 - 09/2016
Applied Electronics

Master of Engineering

- PSG College of Technology, Coimbatore

08/2009 - 04/2013
Electronics and
Communication
Engineering

Bachelor of Engineering

- Thanthai Periyar Government Institute of Technology, Vellore

ACHIEVEMENTS

- Published LLM based paper [BIRD: Business Insight and Recommendation Developer](#) in TechRxiv (IEEE Preprint).
- Featured in Worldwide Google AI competition: [BIRD: Business Insight and Recommendation Developer](#)
- Runner up in All India Google AI Exchange competition: [ARI: Audit Report Interpreter AI](#)
- Featured in Worldwide Databricks AI competition: [Nalanda Educational AI](#)
- Fine tuned SLM on HuggingFace: [GK_10000](#)
- Cleared level 3 of Google FooBar challenge (Google Employment Eligibility)
- Ranked 42 in Machine Hack World Rankings: [MachineHackLeaderboard](#)
- Developed and published 4 PyPi packages.
 1. [multicorr](#)
 2. [metrictable](#)
 3. [largeproductaffinity](#)
 4. [cipherLV](#)