Jaideep Bommidi

jaideepbommidi@gmail.com | Linkedin | GitHub | +1 940-629-6953

- **Data Scientist & Analyst** around 7 years building production data platforms across fintech, survey analytics, and operational intelligence; expert in **PySpark for big data ETL**, **Python**, **advanced SQL**, **A/B testing**, and **Tableau/Power BI visualization**; proven ability to **communicate data-derived insights to non-technical stakeholders** and solve business problems.
- Deep expertise in **exploratory data analysis (EDA)**, **data mining**, and end-to-end **data pipeline development**; plus **Gen AI/LLM implementation**, **survey data analysis** with automated report generation, and cloud deployment (**AWS SageMaker/Lambda**, **Azure**, **GCP Vertex AI**) with **Docker/CI/CD**. Built enterprise survey analytics platform, shipped Gen AI applications.

Technical Skills

Generative AI & LLM Tools: Langchain, Langgraph, Autogen, Hugging Face, Pinecone, Vector Databases

Programming: Python, R, C/C++, SQL, JavaScript, Shell Scripting.

Machine Learning & AI: Agentic AI, **Scikit-learn**, **TensorFlow**, **PyTorch**, Keras, NLP, Classical ML (trees, boosting, clustering, PCA), time series (ARIMA, Prophet, VAR), Computer Vision, A/B testing, LLMs/agents (prompt engineering, RAG, vector search with **FAISS**).

Web & APIs: **FastAPI**, **Flask**, **Django**, **React.js**, **AngularJS**, REST.

Data Engineering & Big Data: **SQLAlchemy**, **PySpark**, **Spark**, **Hadoop/HDFS**, **Hive**, **Pig**, MapReduce, **Snowflake**, **Redshift**, **HBase**.

Databases & Warehouses: **PostgreSQL**, **MySQL**, **SQL Server**, **MongoDB**, NoSQL.

IaC, Containers & Orchestration: **Terraform**, **CloudFormation**, **Docker**, **Kubernetes** (**EKS**, **ECS**).

MLOps & CI/CD: **Git**, GitHub, GitLab, **Jenkins**, AWS **CodeBuild**/**CodeDeploy**/**CodePipeline**, GoCD.

Testing & QA: **Selenium**, **Cypress**, JUnit, Mocha, Jest, Enzyme.

Visualization & Analytics: **Tableau**, **Power BI**, Microsoft Excel, SAS Enterprise Miner.

Cloud Platforms: **AWS** (S3, EC2, Lambda, RDS, SageMaker), **Azure** (Data Factory, Data Lake), **GCP**.

Experience

Data Scientist — May 2025 – Current

Webster Bank — *Denton, TX*

- **Implemented end-to-end agentic AI system** from architecture through production deployment for Temenos Journey Manager/Maestro, featuring **multi-agent orchestration** with autonomous decision-making, multi-stage retrieval workflows, re-ranking pipelines, and semantic search (FAISS vector store).
- **Deployed production agent framework** with **LangChain-based orchestration** implementing graph-style conditional routing, stateful memory management, and tool-calling capabilities; designed multi-agent collaboration patterns consistent with **LangGraph** state machine architecture and **AutoGen**-style conversational flows for human-in-the-loop checkpoints and autonomous task delegation.
- Built **embedding-based retrieval pipeline** leveraging **AWS deployment stack** (S3 for data lakes, Lambda for serverless orchestration, SageMaker for model hosting): ingested internal properties/logs into **FAISS vector store**, implemented **hybrid search** (dense embeddings + sparse keyword filters) for context augmentation, and **fine-tuned ranking models using LoRA** to improve **NDCG/MAP by 18%**.
- **Self-hosted, bank-grade Gen AI deployment:** Ran **open-weight LLMs (LLaMA, Mistral families) and embedding models** via **Hugging Face Transformers** on company servers with zero external calls; integrated PII scrubbing, policy guardrails, automated drift detection, and full audit logging to meet enterprise security/compliance SLAs.
- Shipped **low-latency agentic inference services** using production best practices including response caching, request batching, fallback orchestration logic, and circuit breakers to maintain **<200ms p95 response times**; established **monitoring dashboards** with Prometheus/Grafana for drift detection, agent performance tracking, and guardrail compliance validation.
- **Collaborated cross-functionally** with security, compliance, and platform engineering teams to iterate on LLM/agent components (scoring strategies, retrieval plugins, tool integrations, agent memory systems); delivered custom DDS services, form validators, and RBAC features through **CI/CD pipelines** with staged rollouts and **A/B testing** for agent behavior optimization.

Data Scientist — Sep 2024 – May 2025

Sypnios — *Dallas, Texas*

- Designed and developed an **end-to-end survey data analysis platform** enabling stakeholders to collect survey responses and **automatically generate analytical reports**; built **ETL pipelines with PySpark** to process multi-source survey data (CSV, JSON, API feeds), performed data cleaning, deduplication, and validation at scale to ensure data quality and integrity.
- Implemented **advanced data analysis features** including **cross-tabulation analysis, statistical significance testing, sentiment analysis** on open-ended responses, and automatic categorization of survey feedback; ran **A/B experiments** to validate platform features and optimize user engagement metrics (response rates, completion time, satisfaction scores).
- Built **automated report generation system with Gen AI/RAG capabilities** using **retrieval-augmented generation** to synthesize survey insights into executive-ready narratives; leveraged **Google Cloud Platform (Compute Engine, Cloud Storage, Vertex AI)** for data processing and model deployment; orchestrated semantic search with **embedding models and vector retrieval** to extract relevant trends and generate PDF/Excel/Power BI reports with compliance guardrails.
- Developed full-stack analytics application (**Python, FastAPI, SQLAlchemy**) with **robust data pipelines** to source, transform, and load survey data from multiple collection platforms; implemented **advanced SQL queries** with window functions and aggregations for real-time analytics; exposed analytics via **REST APIs**, containerized with **Docker**, and deployed to **AWS/Azure** with **CI/CD automation**.
- Created **interactive dashboards (Tableau & Power BI)** with **data visualization best practices** to surface survey KPIs (response trends, demographic breakdowns, sentiment distribution, comparative analytics); **effectively communicated data-derived insights to non-technical audiences** (commercial and NGO stakeholders), enabling data-driven decision-making.

Data Scientist Mar 2022 – Jan 2023

German Railways (Deuta Werke GmbH) *Cologne, Germany*

- Developed and deployed **ML-driven fleet demand forecasting models** (ARMA/ARIMA, exponential smoothing) using **Python/PySpark** for real-time distributed processing of 147k+ rows/hour train telemetry data; optimized **marketplace signals** (fleet availability, ETA predictions, supply reliability metrics) to improve operational efficiency, reduce downtime by 22%, and inform driver allocation and incident reporting.
- Built **end-to-end ML pipelines** for risk/incident prediction using **Python (pandas, NumPy, scikit-learn)** with advanced **feature engineering**; validated with cross-validation and deployed to production serving real-time inference for operational decision-making.
- Designed and implemented **causal inference experiments via A/B testing** to measure impact of incident-reporting rule changes on false positive rates; applied **statistical rigor** to validate algorithmic improvements and documented assumptions/acceptance criteria.
- Built **scalable big data pipelines** using **Spark, Hadoop/HDFS, Hive, MapReduce** for large-scale data transformation and access; optimized **SQL** (SQL Server, PostgreSQL) for data collection, ETL design, and data quality.
- Deployed **production ML systems** with **RESTful APIs** (Django), real-time dashboards (**Tableau**), and analytics tools enabling self-serve insights for cross-functional stakeholders (operations, safety, leadership); orchestrated **CI/CD pipelines** (**GOCD, Docker, Git**) ensuring reliable, maintainable ML solutions at scale.
- Collaborated with **cross-functional teams** (operations, engineering, safety) to shape technical strategies, launch new analytical products, and continuously improve marketplace performance through data-driven optimization.

Data Scientist Sep 2020 – Dec 2021

iFactory3D GmbH *Düsseldorf, Germany*

- Developed **deep-learning defect detection system** for 3D prints using **TensorFlow/Keras with CUDA acceleration** and **transfer learning** (ResNet, VGG architectures); achieved **95% accuracy** and **20% performance improvement** over baseline computer vision methods; deployed lightweight **C++ inference module** on printer hardware, reducing processing time by **40%** and enabling real-time quality control for production environments.
- Built **end-to-end ML pipelines** for customer analytics: designed **segmentation models** (DBSCAN, K-means++, Hierarchical clustering) to identify 5 distinct customer personas; developed **churn prediction and retention recommender** using **ensemble methods** (Random Forest, Gradient Boosting) and **reinforcement learning** for personalized interventions; delivered **30% lift in retention**, **40% increase in customer satisfaction**, and via **A/B testing**, **15% boost in sales conversions**.
- Developed **time-series forecasting models** (ARIMA, VAR, NNAR) for sales and marketing campaign prediction; improved forecast accuracy by **25%**, enabling better inventory planning and resource allocation; performed **factor analysis and cluster analysis** for market targeting in investor reports, contributing to **20% increase in investment returns**.
- Architected **scalable data infrastructure**: designed **SQL/NoSQL schemas** (PostgreSQL, MongoDB) with data integrity checks reducing data errors; built **big data pipelines** using **Hive on Hadoop/HDFS** and **Redshift** for ETL at scale, reducing query times by **50%** and enabling analytics on multi-TB datasets.
- Developed **full-stack ML applications** with **RESTful APIs** (Flask), **React.js** frontend, and **AWS deployment** (EC2, S3, Lambda); operationalized ML models with **Docker containerization** and **CI/CD pipelines** (GitHub, AWS CodeBuild/CodeDeploy/CodePipeline); versioned with **Git** in **Agile/Scrum** workflow; **mentored 2 interns** on ML best practices and conducted code reviews to maintain high engineering standards.

Software Developer Jul 16 – Aug 18

Temenos India Pvt Ltd. *Chennai, India*

- Developed and provided support for modules within the Temenos product.
- Collaborated with the EB team to design and implement optimized bank transactions using JDBC, SOAP, and Tomcat, resulting in substantial transaction speed improvements.
- Conducted thorough research aimed at designing and optimizing SVM algorithms to extract valuable insights from images.

Research Internships

Graduate Teaching Assistant Aug 23 – May 24

University of North Texas *Denton, TX*

- Instructed and supported courses in **Machine Learning**, **Big Data**, and **Tableau** for **100+** students; provided technical guidance and fostered a strong learning environment.
- Built a Python-based web scraper to collect submission metadata; anonymized and analyzed data with **pandas** and **NumPy**; applied time-series methods (**ARIMA**, **Exponential Smoothing**, **Prophet**) to uncover study-behavior patterns.
- Developed and deployed deep learning models with **TensorFlow** and **PyTorch** for the UNT Swim Team: image segmentation and keypoint detection to derive stroke angles/velocities; integrated with physiological metrics to support performance analysis.

Data Scientist Intern Feb 20 – Jul 20

Solarvibes GmbH *Master Thesis, Berlin, Germany*

- Developed a deep learning plant disease classification algorithm using computer vision techniques such as image segmentation and classification.
- **European Space Imaging Project:** Developed advanced machine learning models to analyze farmland images retrieved from AWS S3, utilizing Geographic Information Systems (GIS) for comprehensive vegetation analysis. Implemented AI-enabled data-driven models to enhance predictive accuracy in agricultural monitoring.
- Engineered a robust cloud architecture on AWS to support real-time data processing and analysis, ensuring scalability and efficient resource management. Leveraged AWS services such as Amazon SageMaker for model training and deployment, and Amazon Kinesis for streaming data analytics.
- Collaborated with cross-functional teams to integrate geospatial data with machine learning insights, facilitating decision support tools for stakeholders. Engaged in strategic discussions to align project objectives with technological advancements in geospatial analytics.
- Rapidly prototyped web apps based on stakeholder requirements and scaled cloud infrastructure to meet project specifications.

Education

M.S | Advanced Data Analytics | University of North Texas May 2024 | USA

M.S | Image Processing and Computer Vision | University of Bordeaux Aug 2018 | France

B.Tech | Electronics and Communication Engineering | National Institute of Technology, Trichy May 2016 | India