SHARATH JAGADABHI

jagadabhisharath401@gmail.com | +91-8074291126 | LinkedIn | GitHub

SUMMARY

Results-driven Data Scientist with a Bachelor's of Science in Statistics and strong hands-on experience in predictive modelling, exploratory data analysis, and AI/ML solution development. Skilled in Python, SQL, and modern ML frameworks with proven ability to translate business requirements into data-driven insights and scalable prototypes. Experience in building end-to-end ML pipeline testing models, evaluating performance, and communicating findings to technical and non-technical stakeholders. Collaborative team contributor who aligns technical solutions with business goals to improve efficiency, decision-making and user impact.

EDUCATION

Bachelors of Science in Statistics (CGPA: 7.53) **Jul 2020 - Sep 2023**
Pragati Mahavidyalaya Degree College, Hyderabad, India

SKILLS

- **Programming:** Python, SQL
- **Analysis:** Pandas, NumPy, Matplotlib, Seaborn, Statistical Analysis, EDA
- **ML Algorithms:** Scikit-Learn, Linear/Logistic Regression, Decision Trees, Random Forest, SVM
- **AI Concepts:** Deep Learning, NLP, LLMs, RAG, Computer Vision, Neural Networks, TensorFlow, PyTorch, Transformers
- **Cloud & Tools:** GCP, Azure, Power BI, Docker, Git
- **Soft Skills:** Analytical Thinking, Communication Skills, Project Planning, Problem-Solving, Team Collaboration

PROFESSIONAL EXPERIENCE

Machine Learning Intern | Internship Studio **Feb 2025 - Aug 2025**

- Engineered end-to-end ML pipeline predicting student academic outcomes with 84% accuracy, directly informing retention strategies that identified at-risk students 6 weeks earlier than previous manual review processes.
- Utilized Python and SQL for initial data collection and exploration, successfully identifying data quality problems and automating preprocessing for a dataset with complex demographic attributes.
- Executed A/B tests on various modelling techniques (Linear Regression, Decision-Tree building, Random Forest) recommending Random Forest as optimal solution based on 12% higher F1-score and 40% faster inference time.
- Connected model performance to business objectives by translating complex outputs into actionable recommendations regarding student success factors, presenting insights to project stakeholders.
- Validated model performance through rigorous feature engineering and hyperparameter tuning using GridSearchCV, improving prediction reliability by 18% and reducing false negative rate by 22% to meet project accuracy requirements.

Artificial Intelligence Intern | Internship Studio **Jul 2024 - Jan 2025**

- Developed production-ready LLM-powered chatbot using GPT-3.5 and Streamlit, processing 500+ user queries during pilot testing with 89% user satisfaction rating and average response time under 2 seconds.
- Architected modular RAG system improving response accuracy by 35% and reducing hallucination rate from 28% to 8% through custom context retrieval logic tested on 200-query validation set.
- Implemented robust logging and session-state management to facilitate testing on production like data, ensuring application stability and scalability.
- Containerized the application using Docker and maintained version control via Git, aligning with industry standards for reproducible builds and deployment strategies.

PROJECTS

Cancer Image Detection | GitHub Repo

- Developed CNN-based medical image classifier achieving 92% accuracy and 0.89 AUC-ROC score on 5000-image histopathology dataset, demonstrating 15% improvement over baseline transfer learning approach.
- Optimized model architecture through systematic hyperparameter tuning using grid search, reducing training time by 33% while maintaining accuracy through mixed-precision training on GPU environments.
- Applied Grad-CAM to visualize model attention, validating that 87% of predictions focused on clinically relevant tissue structures, building potential clinical deployment and identifying 3 edge cases requiring model improvement.
- Evaluated model performance using confusion matrix analysis revealing 94 sensitivity and 90% specificity, meeting clinical thresholds for screening applications and identifying specific cancer subtypes requiring additional training data.

AI Shopping Assistant | GitHub Repo

- Designed a cloud-based NLP assistant to process vague product inquiries, leveraging LLMs to align technical inputs with business requirements for actionable search results.
- Engineered real-time query summarization pipeline using LangChain integrating SerpAPI for live product data retrieval, processing user intent with 82% accuracy and returning top 5 relevant products within 3 seconds average response time.
- Optimized query processing workflow through iterative A/B testing, reducing average response time from 4.2 to 3.0 seconds (28% improvement) and increasing user task completion rate from 68% to 83% based on user beta test.
- Managed structured project repositories using Git, demonstrating knowledge of essential tools for scalable data solutions and enabling team collaboration.

CERTIFICATIONS

Generative AI Professional | Oracle | Certificate

- Gained hands-on skills in Generative AI, LLMs, prompt engineering, RAG (Retrieval-Augmented Generation), and building AI applications using Oracle Cloud.

Python 101 for Data Science | IBM | Certificate

- Gained hands-on with core python libraries as Pandas, NumPy, and Matplotlib for data analysis and visualization.
- Developed a solid foundation in data science workflow, including data wrangling, exploratory data analysis (EDA). And basic machine learning using Scikit-learn.