AVIJIT JANA

Kolkata, West Bengal, India | +91 9830641205 | avijitjanascience24@gmail.com
LinkedIn: https://www.linkedin.com/in/avijit-jana/ | GitHub: https://github.com/avijit-Jana

SUMMARY

Results-driven Data Scientist with hands-on experience in machine learning, predictive modeling, and data visualization. Skilled in Python, SQL, MongoDB and AWS S3; experienced building end-to-end solutions (feature engineering, model training — including deep learning & transformers — evaluation, and deployment). Practical experience with retrieval-augmented generation (RAG) systems using Pinecone and LangChain, interactive dashboards, and production APIs to turn data into actionable business outcomes.

SKILLS

- **Languages & Databases:** Python, SQL, MySQL, MongoDB, C/C++, Java
- **Machine Learning & AI:** Supervised learning, Deep Learning, Transformers, NLP, Time Series Forecasting, Predictive Modeling, Feature Engineering, Cross-Validation, Hyperparameter Tuning, Model Evaluation (Accuracy, Precision, Recall, F1, AUC)
- **Libraries & Frameworks:** pandas, NumPy, scikit-learn, TensorFlow, PyTorch, Keras, XGBoost, LightGBM, LangChain
- **Vector DB / RAG:** Pinecone, Retrieval-Augmented Generation (RAG) workflows, LangChain integrations
- **Deployment & Tools:** FastAPI, Streamlit, Docker, Git/GitHub/GitLab, Postman, Selenium, web scraping
- **Cloud & Storage:** AWS S3
- **Visualization & BI:** Power BI, Excel, Matplotlib, Seaborn, Plotly

PROJECTS

1. **RAG Streamlit UI | GitHub**
   - De Built a Streamlit chat app for Retrieval-Augmented Generation with Pinecone vector search.
   - Enabled PDF/TXT ingestion, chunked embeddings, and semantic query retrieval.
   - Integrated Google GenAI for optional answer generation and MongoDB for chat logging.
2. **Classifying Cybersecurity Incidents using Machiachine Learning | GitHub**
   - Developed a classification model to detect cybersecurity incidents (True Positives, Borderline Positives, False Positives) using the GUIDE dataset.
   - Automated threat detection to enhance SOC efficiency and reduce response times.
   - Achieved 90% accuracy with a Macro F1 Score of 89% through advanced feature engineering and model evaluation (Precision, Recall).
3. **Used Car Price Prediction | GitHub**
   - Built a machine learning model to predict used car prices based on key features such as make, model, year, and fuel type.
   - Deployed a Streamlit-based web application for real-time price estimation to assist customers.
   - Conducted Exploratory Data Analysis (EDA) and optimized model performance to achieve an R^2 Score of 93%.
4. **RedBus Data Scraper Dashboard | GitHub**
   - Automated bus travel data extraction from Redbus using Selenium, storing results in SQL for analysis.
   - Built a Streamlit dashboard with filters (route, type, price, rating, availability) for interactive insights.
   - Enabled business use cases like travel aggregation, competitor analysis, and customer-centric service improvements.

WORK EXPERIENCE

- **AI/ML Intern – Siam Computing**
  Nov 2025 – Jan 2026 (2 Months)

EDUCATION

- **Advanced Programming Professional & Master Data Science**
  IIT-M GUVI, 2024 – 2025
- **Bachelor of Science in Computer Science (Honours)**
  University of Calcutta, 2021 – 2025

CERTIFICATIONS

- Data Analytics Using Pandas – GUVI
- Master Data Science – GUVI
- Image processing – GUVI