Pawan S. Bant

Data Scientist & Machine Learning Engineer

📞 +91-7676868328   ✉ bantpawan@gmail.com   GitHub   Portfolio   in LinkedIn

Professional Summary

Data Scientist & Machine Learning Engineer with expertise in designing and deploying **end-to-end machine learning solutions**. Improved **model accuracy by 45%** and reduced **data pipeline runtime by 25%** through optimized processing and scalable ML pipelines. Skilled in **Python, PySpark, Azure ML, Power BI, and Streamlit**, with experience developing analytics solutions that enhance business decision-making and product performance.

Technical Skills

- **Programming & Libraries:** Python (Pandas, NumPy, Scikit-learn, PyTorch), SQL
- **Machine Learning:** Regression, Classification, Ensembles, Feature Engineering, Model Evaluation, NLP, Transformers (Hugging Face), RAG, LangChain
- **Data & Big Data:** Pandas, PySpark, Databricks, Delta Lake, ETL Pipelines
- **Cloud & MLOps:** Azure ML, Docker, CI/CD
- **Visualization:** Power BI (DAX, data modeling), Streamlit, Matplotlib
- **Databases & Versioning:** MySQL, MongoDB, Git / GitHub

Projects

Research Paper Q&A Assistant (Python, NLP, Streamlit, LangChain)   **2024 − 2025**

- Engineered a Retrieval-Augmented Generation (RAG) pipeline with FAISS + MiniLM embeddings and 4-bit quantized LLaMA-2, cutting memory usage by **50%** and speeding up indexing by **30%**.
- Delivered structured academic-style Q&A, summaries, and quizzes via Streamlit; applied prompt engineering and BitsAndBytes optimization for efficient LLM inference.

Real Estate Price Prediction System (Python, Machine Learning, Scikit-learn)   **2023 − 2024**

- Designed an ML pipeline with advanced feature engineering, SHAP-based explainability, and ensemble models (RF, GBM), boosting R^2 from **0.62** to **0.89** (**+43%**).
- Developed a streamlined data cleaning process using Pandas, resulting in improved data accuracy and a **+15%** reduction in errors identified during the model training phase.
- Deployed via Streamlit with recommender system and analytics dashboards (heatmaps, scatter plots, word clouds), enabling interactive property insights.

Enterprise Sales Dashboard (Power BI, MySQL, Advanced Excel, Data Visualization)   **2023 − 2024**

- Built SQL-driven Power BI dashboards with calendar heatmaps, category-wise sales, and weekday/weekend segmentation, improving decision turnaround by **10%**.
- Automated KPI analysis (sales, orders, quantity), highlighting top products and location-based trends with MoM growth metrics.

Professional Experience

Data Science Intern, Besant Technologies   **Aug 2024 − Dec 2024**

Bengaluru, Karnataka

- Developed ML models (regression, deep learning) for taxi-fare prediction, achieving $R^2 = 0.84$ (**+30%** over baseline).
- Optimized ETL for 50K+ rows using Pandas & PySpark, cutting runtime by **25%**.
- Deployed Dockerized solutions and Azure ML workflows, improving scalability by **15%**.

Education

B.Tech in Engineering, KLE Institute of Technology   **2015 − 2019**

Hubballi, Karnataka

Key Achievements

- Improved real-estate forecasting model from **62%** to **89%** R^2.
- Built research-document retrieval pipeline that halved memory usage with FAISS indexing.
- Automated preprocessing pipelines that consistently reduced manual prep by **25%**.