Sanket Mishra

+91-7236891123 | sanketmishra.real@gmail.com | linkedin.com/in/sankettmishra | github.com/SanketxMishra

Summary

Data Science undergraduate at IIT Madras with strong expertise in Generative AI, Machine Learning, and Full-Stack deployment. Proven ability to architect RAG systems and predictive models that solve tangible business problems. Passionate about bridging the gap between statistical analysis and production-grade software engineering.

Education

Indian Institute of Technology Madras — Chennai, Tamil Nadu
BS in Data Science and Applications, Minor in Economics and Finance — *Sept. 2022 – Aug 2026*

Projects

AI-Powered Data Analysis Automation | *Python, FastAPI, OpenAI GPT-4* — GitHub

- Engineered an autonomous data analysis agent using GPT-4.1 to generate and execute Python code, delivering statistical insights in **<60 seconds** for 8+ file formats (CSV, JSON, Parquet).
- Implemented an intelligent **self-healing mechanism** with 3-retry logic and error parsing, achieving a **95% success rate** in unsupervised code execution.
- Architected a scalable REST API (FastAPI) orchestrating 27+ data libraries (Pandas, SciPy, GeoPandas) with multipart handling for real-time processing.
- Developed an automated visualization pipeline generating optimized (<100KB) graphs using Matplotlib and NetworkX, directly embedded into client responses.

Revenue Prediction Engine (XGBoost) | *Python, Scikit-Learn, Pandas* — GitHub

- Developed a high-performance regression pipeline on **116,000+** web traffic records to forecast revenue, achieving a top R^2 score of **0.74**.
- Outperformed baseline KNN models by **88%** by implementing an optimized **XGBoost Regressor** with Histogram-based tree methods for faster convergence.
- Conducted advanced **Feature Engineering** to create 6 interaction terms (e.g., session seasonality), significantly increasing model sensitivity to user behavior.
- Reduced dimensionality by 40% via automated null-value filtering and variance thresholding, streamlining the feature space for production deployment.

GenAI Personal Stylist (RAG Agent) | *React, LangChain, ChromaDB* — GitHub

- Built a stateful conversational AI agent using **LangGraph** and Groq LLM to provide hyper-personalized fashion recommendations based on natural language queries.
- Engineered a **Semantic Search pipeline** using **ChromaDB** (Vector Store) and HuggingFace embeddings, achieving **90% retrieval relevance** for product queries.
- Implemented **Session Memory** to maintain context across 3+ conversational turns, tracking complex constraints (price, brand, size) for precise filtering.
- Deployed a secure full-stack architecture with React and FastAPI, utilizing JWT authentication to manage user sessions and order history.

Technical Skills

Languages: Python, SQL, JavaScript (ES6+), C++
AI/ML: TensorFlow, PyTorch, Scikit-Learn, XGBoost, Statsmodels, Pandas, NumPy
GenAI & NLP: LangChain, LangGraph, RAG, HuggingFace Transformers, ChromaDB, OpenAI API, Groq
Web & DevOps: React.js, FastAPI, Flask, Node.js, Docker, Git, Google Cloud Platform (GCP)

Certifications

Dataiku Academy: Core Designer Certification, Core Developer Certification

Achievements

Winner, Trade Quest (IIT Madras): Secured 1st Position in a high-stakes live market simulation by deploying quantitative trading strategies and optimizing risk-reward ratios based on real-time technical analysis.