SANA AMBREEN

⊙ **Sanataj0409@gmail.com**   ⊙ *Ph: 5127779233*   ⊙ *LinkedIn*   ⊙ *GitHub*   ⊙ *Portfolio*

Experience Summary

AI/ML Engineer with **3+ years** of experience building production-ready **machine learning, deep learning, NLP, and generative AI systems**, including generative AI assistants deployed in real-world environments. Proven ability to deliver end-to-end ML pipelines covering **data ingestion, data cleaning, feature engineering, model training,** deployment, and monitoring, with exposure to **MLOps practices**. Experienced in developing solutions using **PyTorch** and **Hugging Face**, and deploying cloud-native, scalable services across **AWS, Azure, and GCP**. Adept at translating complex models and analytics into **actionable insights** and business value, driving measurable improvements in efficiency, accuracy, and user engagement.

Technical Skills

Programming & Scripting-Python, SQL, Bash/Shell, TypeScript (basic), JavaScript (basic).

Machine Learning-Supervised Learning, Unsupervised Learning, Hyperparameter Tuning, Cross-Validation, Statistical Modeling, Time-Series Forecasting, A/B Testing, Anomaly Detection, Recommendation Systems. Git, Pinecone, Vector Databases, Langchain, Fast API,

Artificial Intelligence-Deep Learning (CNN, RNN, LSTM, Transformers), Generative AI(LLMs, Prompt engineering), Text Generation and Summarization.

Natural Language Processing (NLP)-NLP Libraries(spacy, Hugging Face), Named Entity Recognition(NER), Intent Classification, Sentiment Analysis, Transformers-BERT, GPT

Data Engineering-Apache Spark, Big Query, AmazonS3, Azure Data Lake, Google Cloud Storage, PostgreSQL, Redis

Data Analysis & Visualization-Pandas, NumPy, Matplotlib, Seaborn, Tableau, Power BI, Looker, A/B Testing, Exploratory Data Analysis (EDA), Statistical Modeling

Cloud & Infrastructure-AWS (S3, SageMaker, Lambda), GCP (Big Query, VertexAI), Azure (Synapse, ML Studio), Docker, Kubernetes, CI/CD Pipelines

Career Summary

AI Intern | UpLifty AI, Remote (USA)   Oct 2025 – Present

- Led the design and pilot deployment of a **Personal AI Agent (PAI)** architecture combining **supervised learning** and **reinforcement-learning** inspired feedback loops to deliver context-aware, goal-driven recommendations, improving personalization and early user engagement during pilot rollouts.
- Built and productionized **LLM-based suggestion** and **wellness classification models**, including **fine-tuning Llama 3.2 (1B) and BERTweet**, enabling low-latency, CPU-only inference on **Google Cloud Run** and supporting scalable pilot deployments.
- Designed and implemented a **Retrieval-Augmented Generation (RAG) pipeline** using **LangChain orchestration** and a **vector database (Pinecone)** to ground LLM responses in curated internal knowledge and user context, **reducing hallucinations** and improving response relevance by 20–25%.
- Developed the **Human Growth Index (HGI) pipeline** by transforming multi-source user activity and reflection data into 13 explainable wellness dimensions using **regression and classification models,** delivering **transparent, user-visible scoring** that improved insight clarity and usability by **15–20%**
- **Generated 50K+ synthetic and weakly labeled samples** through **LLM-assisted labeling and validation workflows**, accelerating model development by **~30%** and reducing manual annotation effort **by ~35–40%**, while owning **end-to-end ML deployment pipelines** (Docker, Ollama, Cloud Run, Supabase triggers) with monitoring and retry logic.

AI Intern | Verity Advisors LLC, Remote (USA)   March 2025 – Oct 2025

- Built a real-time, **event-driven ML trading system** using **FastAPI-based microservices,** ingesting live options sweep data via **WebSockets** and processing it through **AWS SQS,** reducing signal-to-decision latency by **~40%** compared to a synchronous pipeline.
- Developed an **XGBoost classification model** with extensive **feature engineering** on options sweep data, including IV Rank, VWAP, RSI, volume, and open interest, improving signal precision and reducing low-confidence trade intents by approximately **30%** during controlled evaluation runs.
- Trained an **LSTM model using TensorFlow** on 48-hour OHLCV **time-series data** combined with **sentiment features**, where sentiment was derived using **FinBERT/SBERT** models from **Hugging Face**, enabling improved trend-aware trade scoring.
- Implemented **model ensembling,** calibration, and risk-gated execution logic, enforcing exposure limits, cooldowns, and market-hours checks before placing orders through IBKR APIs, **reducing invalid or rejected trades by ~25%** in dry-run and early live testing.
- Established **end-to-end MLOps workflows**, training models tracking experiments with **MLflow**, **versioning artifacts**, and deploying **Dockized inference services** on **AWS ECS Fargate** via **GitHub Actions CI/CD**, contributing to consistent signal accuracy and positive trading performance achieving a **70% trading accuracy**.

Data Scientist | Wipro Technologies, India   June 2023 – Dec 2024

- Supported the development of an **end-to-end A/B testing and analytics solution** by building **ETL pipelines** in **BigQuery** to aggregate and clean session-level event data, defining KPIs such as conversion rate and AOV, and performing **statistical analysis** to evaluate experiment outcomes.
- Developed and validated a **logistic regression conversion probability** model using **BigQuery ML**, applying **feature engineering** on user behavior data to predict purchase likelihood and provide early intent signals during active experiments.
- Collaborated with **cross-functional teams** including product managers, engineers, and analysts to automate data workflows and **CI/CD-style model pipelines** using **SQL, BigQuery scheduled queries, and Cloud Run,** supporting regular retraining, batch scoring, and monitoring.
- Created **Power BI dashboards** & utilized **data storytelling** to effectively communicate A/B test results, uplift, **statistical significance**, helping propose and implement solutions that contributed to an estimated **increase in annual revenue through an 15% improvement in conversion rate** while maintaining stable guardrail metrics.

Data Analyst | Wipro Technologies, India   Oct 2021 – May 2023

- Supported **hospital operations analytics** by building **ETL pipelines** and performing **data cleaning, exploratory data analysis (EDA),** and seamless data integration across patient admissions, bed and ICU occupancy, OT usage, and workforce schedules using **Python, Pandas, NumPy, SQL queries, and relational databases.**
- Performed **feature engineering** on **time-series operational data** and developed interpretable **predictive models** using **linear regression and random forest** to estimate patient inflow, ICU capacity risk, and staffing demand.

SANA AMBREEN

- Created **data visualizations and Tableau Server dashboards** using datasets and model outputs, enabling stakeholders to monitor key KPIs and **forecasts in real time.**
- Generated **data-driven and actionable insights** from operational and **predictive analyses** to support **data-driven decision making**, improving visibility into peak demand periods, staffing imbalances, and capacity risks by approximately **10–15%**

Projects

Multi modal Depression analysis | Link

- Built an **end-to-end ML pipeline** integrating **audio (OpenSmile eGeMAPS, Parselmouth), video (OpenFace facial landmarks),** and **text (HuggingFace MiniLlM embeddings) features** for **clinical depression detection.**
- Trained a **PyTorch deep neural network** with top-100 selected features, leveraging **scikit-learn StandardScaler** for preprocessing and **feature normalization.**
- Deployed the solution with **Flask + Gunicorn backend and Vercel frontend**, enabling real-time video/audio capture, feature extraction, and live probability predictions.
- Implemented MLOps best practices with **Dockerized environments**, **API endpoints** (/segment, /finalize, /predict), and scalable deployment on **Render/Railway**.

Trip Sage AI travel assistant | Link

- Built a **conversational AI travel assistant** using **Dialogflow CX,** implementing **intent classification and multi-turn webhook workflows** to support flight, hotel, and car rental search with personalized itineraries.
- **Engineered prompt-based travel suggestions** using **Gemini 2.5 Flash** (Vertex AI, Dialogflow CX), generating contextual travel suggestions & summaries during testing.
- Developed **Flask-based webhook services with REST endpoints** (/webhook, /chat), performing **data cleaning, normalization, and feature engineering** on third-party travel API responses to return ranked option cards.
- Enabled **real-time Dialogflow CX integration using ngrok-based secure tunneling**, supporting low-latency responses and live conversational flows.

Fraud Monitoring Platform | Link

- Developed and optimized an **XGBoost-based fraud detection model on 1M+ bank account applications**, addressing extreme class imbalance (~1.1% fraud rate) using scale_pos_weight tuning, **cross-validation**, **data cleaning, preprocessing, and feature engineering** (credit-to-income ratio, distinct email count, long-address flags), **achieving ~80% recall and improving high-risk fraud detection by ~25%.**
- Built a **RAG-powered fraud explanation layer using LangChain and prompt-engineered OpenAI GPT-4.1 Mini** to act as a **virtual fraud analyst**, with **Azure Cognitive Search (semantic retrieval)** with historical fraud cases and model outputs, **reducing manual analyst investigation time by ~30–35%** while improving explainability.
- Deployed a **cloud-native FastAPI microservice** exposing clean **REST APIs, fully containerized with Docker and deployed on Azure Container Apps**, integrating **Azure OpenAI (chat + embeddings)** for low-latency, scalable **real-time fraud intelligence**, achieving reliable production performance during pilot deployments.

Real-Time Electricity Load Forecasting System | Link

- Developed and **evaluated deep learning time-series models (Vanilla RNN, LSTM, GRU, Bidirectional LSTM)** on **2020–2025 ERCOT load data**, applying data cleaning, normalization, feature engineering, and **sliding-window forecasting**, achieving lowest RMSE and MAPE with a multivariate BiLSTM for next-2-hour demand prediction.
- Optimized model performance using **Optuna hyperparameter tuning** and multivariate inputs (regional loads, time-of-day, seasonality), reducing short-term **forecasting error by ~18–22% compared to univariate and baseline RNN models.**
- Deployed a production-style real-time **ML pipeline** with a **FastAPI inference service** fetching **live ERCOT data** and **serving forecasts to Tableau via Web Data Connector,** enabling continuously refreshed dashboards for **grid monitoring** and achieving reliable, low-latency operational forecasts.

COVID-19 Policy Impact on U.S. Housing Markets | Link

- Built a **state-level predictive econometric analysis** by integrating multi-source datasets (Zillow, OxCGRT, FRED, OWID) and applying **data cleaning, normalization**, & **feature engineering** on COVID-19 policy stringency, unemployment, GDP growth, interest rates, & mortality, enabling robust modeling of **U.S. housing price growth.**
- Applied **Multiple Linear Regression (OLS)** with **exploratory data analysis,** outlier handling, and model evaluation (R^2, RMSE, MAE), and delivered an **interactive Tableau dashboard** that generated **data-driven insights**, demonstrating how **lower policy stringency and low interest rates accelerated housing growth**, while rising unemployment and interest rates slowed markets in later phases.

Donor Prediction and Fund-Raising Analytics

- Built **classification models (Logistic Regression, Random Forest, Neural Network) in SAS Enterprise Miner** to predict donor likelihood, **achieving >91% accuracy and ~92% sensitivity** on validation data.
- **Developed regression models (Decision Tree, KNN, Multiple LR)** to predict donation amounts, by Decision Tree model yielding the lowest validation ASE (~23.1).
- Conducted **EDA, feature selection**, and outlier analysis on a dataset of 6,000+ records, identifying wealth, home value, and region as key predictors of donor behavior.
- Delivered a **data-driven mailing strategy** that improved expected profit per mailer from a **loss of $0.55 to a profit of $11.12,** significantly increasing campaign efficiency and donor targeting effectiveness.

Education

- M.S | **Advanced Data Analytics With AI/ML** | University of North Texas | Denton, TX | May 2026 | GPA: 4.0/4.0
- B.E. | **Electronics and Instrumentation** | Osmania University | Hyderabad, IN | June 2021 | GPA: 3.7/4.0

Certification

- DeepLearning.AI – Generative AI specialization