

SANA AMBREEN

sanataj0409@gmail.com | +1 5127779233 | LinkedIn: [sana-ambreen-383433356/](https://www.linkedin.com/in/sana-ambreen-383433356/) | GitHub: sana0409-ai | Portfolio:

Experience

Research Intern | Verity Advisors LLC, Remote

March 2025-Present

Verity Quant AI Algo trading for Options

- Built a **real-time options trading bot** that reads big option sweeps, runs ML, and places orders through IBKR in <500 ms.
- Designed a production ready **end-to-end ML trading pipeline**- pull data → build features → train → test → deploy → monitor → retrain.
- Set up MLOps and used MLflow to track experiments and register models, kept everything versioned in Git.
- Trained an **XGBoost model** on sweep features IV rank VWAP, RSI and a **LSTM model** on 48-hour OHLCV + news.
- Set up GitHub Actions **CI/CD** to automatically run tests, build **Docker images**, and deploy the model/services on every merge.
- Calibrated the models with **Platt scaling / Isotonic regression**, then blended them with a 60/40 weight.
- Kept S&P 500 sentiment scores fresh every minute and cached them in **Redis** for instant use.
- Added risk rules (stop-loss, take-profit, time stop, position limits, cooldowns) plus a kill switch.
- Log sweeps, features, predictions, and fills to **S3/Athena**.

Tech: Python, FastAPI, XGBoost, PyTorch (LSTM), SBERT/FinBERT, MLflow, Git/GitHub Actions, Redis, MQTT, IBKR TWS, AWS (SQS, IoT Core, Fargate, S3, Athena, CloudWatch), Docker, Grafana.

Oracle Database Administrator | Wipro technologies, India

Oct-Dec 2022

- Configured and maintained **Oracle databases** for a Hospital Management System, ensuring performance, uptime, and security of patient data.
- Implemented backup strategies, performed **upgrades/patching**, and **optimized performance** through indexing, compression, and query tuning.
- Integrated databases with external vendor systems via **JDBC connections** and documented configurations/troubleshooting for team use.
- Diagnosed and resolved long-running queries using trace and **RCA**, collaborating with technical teams to ensure smooth operations

Projects

Multi modal Depression analysis (<https://multimodal-depression-analysis.vercel.app/>)

May 2025

- Built an end-to-end ML pipeline integrating audio (**OpenSmile eGeMAPS, Parselmouth**), video (**OpenFace facial landmarks**), and text (**HuggingFace MiniLM embeddings**) features for clinical depression detection.
- Developed and 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**.

Stock Sentiment Analysis

Mar-May 2025

- Designed and implemented an **LSTM-based deep learning model** to predict short-term stock direction by combining OHLCV market data with multi-source sentiment scores (Alpha Vantage, Polygon IO, IBKR news).
- Engineered features including future returns, windowed sequences, and **sentiment aggregation**, applying techniques like label shifting, sequence padding, and threshold tuning (F1, Youden's J, AUC) for robust model evaluation.
- Achieved **~69% accuracy** and 0.764 AUC, optimizing trade-off between recall and precision through custom decision thresholds.
- Built an **end-to-end pipeline** with data preprocessing, feature engineering, model training, and evaluation, supporting balanced accuracy metrics for imbalanced datasets.

Bank Account Fraud Detection

May 2025

- Developed an **XGBoost-based fraud detection model** on 1M+ bank account applications, addressing extreme class imbalance (fraud rate 1.1%) with **scale_pos_weight** tuning and **cross-validation**.
- Performed data cleaning, preprocessing, and **feature engineering**, including high-impact features such as credit-to-income ratio, distinct email count, and long-address flag to capture fraud patterns.
- Optimized model with **RandomizedSearchCV** hyperparameter tuning (600+ estimators, learning rate 0.01), achieving **~80% recall and 81% accuracy**, minimizing false negatives in high-risk cases.
- Conducted feature importance and correlational analysis, identifying **key fraud indicators** and translating insights into tiered fraud prevention strategies.

Donor Prediction and Fund-Raising Analytics

Feb- Mar 2025

- 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 Linear Regression) to predict expected donation amounts, with the 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.

Hospital Financial Dashboard

Mar- Apr 2025

- Built an **end-to-end Big Query data pipeline** processing 36,000+ records across 117 financial metrics (HCRIS, CMS, ERP, surveys), enabling structured OLAP analysis.
- Designed **15+ Big Query views** to calculate KPIs (net income, cost-per-discharge, uncompensated-care ratios), achieving 50% faster reporting compared to manual Excel workflows.
- Developed an **interactive Tableau dashboard** visualizing rural vs. urban disparities (50% lower net income, 21% higher cost-per-discharge, 3× staffing ratios).
- Applied **Big Query ML linear regression** to forecast hospital closures and financial stress, identifying Medicaid dependency and uncompensated care as key negative profit drivers.

Technical Skills

Programming Languages: Python, Java, SQL, JavaScript, TypeScript, HTML5/CSS3

Web & API Development: React, Next.js, Tailwind CSS, Node.js, Flask, Express.js, RESTful APIs

AI & ML Frameworks: PyTorch, TensorFlow, Scikit-Learn, Hugging Face Transformers, OpenAI API, BigQuery ML

Data & Visualization: Pandas, NumPy, Matplotlib, Plotly, Recharts, Tableau, Power BI

Cloud & DevOps: GCP (BigQuery, Cloud Run, Compute Engine), AWS (EC2, S3, Lambda), Vercel, Docker, GitHub Actions (CI/CD)

Databases & Vector Stores: PostgreSQL, MongoDB, Firebase, Pinecone, MySQL

Education

GPA: 4.0/4.0 — MS in Advanced Data Analytics (AI/ML), University of North Texas, Denton, TX May 2026

GPA: 3.7/4.0 — B.E in Electronics and Instrumentation, Osmania University, Hyderabad, India

