Ameer Sohail Shaik

+1 (813)-817-1935 | skameersohail148@gmail.com | LinkedIn | Portfolio | GitHub

Experience

Machine Learning Engineer August 2024 – August 2025
Proceedit – (Remote) *Barcelona, Spain*

- Architected end-to-end ML pipelines for stock price forecasting using ARIMA, SARIMAX, LSTM, and CNN-LSTM models with custom genetic optimization for hyperparameter tuning, leading research across 15+ ML/DL architectures (XGBoost, Random Forest, CNN-LSTM) that improved prediction accuracy by 6% and backtested trading signal precision by 4%.
- Engineered production-grade GraphQL API layer with Flask-based microservices architecture, building a Master AI bot orchestrating multiple deep learning model endpoints and handling 10K+ daily prediction requests with Git-based CI/CD pipelines for automated testing and deployment.
- Built automated data synchronization and ETL system integrating PostgreSQL with Google Sheets API and restructured the no-code engine for rapid application development, reducing manual data workflows by 92% (60+ min to 3 min) and accelerating model deployment cycles by 40%.

Data Science Intern January 2024 – May 2024
SBP Consulting Pvt Ltd *Hyderabad, India*

- Built predictive analytics models using Python (scikit-learn, XGBoost, pandas) and SQL to forecast client revenue trends by analyzing historical transaction records from SAP S/4HANA, performing feature engineering and model evaluation to achieve 87% forecast accuracy that supported senior consultants in quarterly business planning.
- Developed automated data extraction and cleaning scripts using Python (pandas, NumPy) and SQL to consolidate raw data from multiple SAP modules into structured formats, reducing manual data preparation time by 45% and ensuring consistent data quality for downstream analytics workflows.
- Created interactive dashboards in Power BI to visualize sales trends, revenue patterns, and key business metrics for client reporting, automating weekly report generation that replaced manual Excel-based workflows and improved reporting turnaround time from 2 days to 4 hours.

Education

Master of Science in Data Science Aug 2025 - Present
University of Maryland - College Park *College Park, MD*

Bachelor of Technology in Computer Science and Engineering 2020 - 2024
Vellore Institute of Technology, AP *AP, India*

Projects

Analysis and Prediction of Diabetes using Health Indicators | *Python, Machine Learning* Dec 2025

- Developed binary classification system predicting diabetes risk from 21 health indicators using ensemble methods (XGBoost, Random Forest, Gradient Boosting), achieving 95.9% Recall and 0.9935 ROC-AUC on 150K+ patient records from CDC BRFSS dataset.
- Engineered 18 predictive features through domain-driven feature engineering including interaction terms (BMI×Age), polynomial features, and composite health scores, improving model recall by 12% to minimize false negatives.
- Implemented SHAP framework for model interpretability, identifying GenHlth, BMI, and Age as top risk factors with quantified feature importance scores, enhancing clinical decision-making transparency.

AutoML-ify | *Python, Streamlit, Scikit-learn* August 2025

- Developed end-to-end AutoML web application enabling non-technical users to build ML models without coding, supporting 5+ algorithms with automated hyperparameter tuning using GridSearchCV.
- Engineered scalable data pipeline supporting multiple file formats with automated data validation and preprocessing, processing datasets up to 100K rows with real-time tracking.
- Implemented comprehensive model evaluation dashboard with visualizations, reducing model development time by 80% for business analysts.

Dynamic SQL Assistant | *Python, Streamlit, LangChain, Groq API, SQLite, NLP* May 2025

- Architected Text-to-SQL analytics engine using Python, Streamlit, and SQLite that translates natural language questions into executable SQL queries with 90%+ accuracy.
- Engineered LangChain-based prompt orchestration with Groq (Llama 3.3) implementing dynamic schema injection and validation logic, achieving 100% SQL syntactic validity and zero hallucination errors across complex joins and aggregations.
- Developed Pandas ETL pipeline with intelligent data sanitization, automated type inference, and SQLite integration that processes CSV uploads containing 100K+ rows into queryable databases in under 2 seconds with 99.9% data integrity.

Technical Skills

Languages & Databases: Python, SQL, Java, PostgreSQL, MySQL, SQLite, MongoDB
Data Science & ML: Pandas, NumPy, Scikit-learn, TensorFlow, EDA, ETL, Feature Engineering
NLP & GenAI: LangChain, Llama, RAG, AI
Tools & Deployment: Flask, GraphQL, Git, CI/CD, Docker, Airflow, Spark, Streamlit, Tableau, Power BI
Cloud: AWS (S3, Redshift, QuickSight, Lambda)