MAHMOUD SAAD

908-630-7978 | Mahmoud2511.saad@gmail.com | Linkedin | GitHub

PROFESSIONAL SUMMARY

- Data-driven professional with a B.S. in Computer Science and a proven track record in building scalable data solutions across healthcare, finance, and tech.
- Over 2 years of hands-on experience developing **end-to-end** data pipelines, predictive models, and interactive dashboards.
- Expertise in **Python**, **SQL**, and **Power BI**—backed by industry certifications and real-world project success.
- Adept at integrating data from diverse sources, performing rigorous **data cleaning** and analysis, and delivering actionable insights that improve outcomes.
- Recognized for excellence: achieved 3rd place in a capstone project at NJIT and presented directly to senior management at Merck.

TECHNICAL SKILLS

Languages: **Python**, **SQL**, C++, TypeScript, JavaScript
Data Science Tools: Pandas, NumPy, **Power BI**, Tableau, TensorFlow, Scikit-learn, PyTorch, Excel
ETL & Data Pipelines: Automated data ETL pipelines; experience with AWS, Azure, SharePoint integration
Development & Cloud Tools: Docker, Git, Agile, Test-Driven Development (TDD), SAM model, AWS

EXPERIENCE

Data Analyst — Sep 2024 – Dec 2024

Merck Pharmaceuticals — *Remote*

- Implemented a scalable automated notification engine using **Power Automate** to aggregate data from multiple sources—including a **SharePoint** database—while executing thorough **data cleaning** routines.
- Achieved a 4,300-hour annual reduction in manual effort, secured 3rd place among NJIT capstone projects, and earned a presentation to Merck's management team.
- Implemented **dynamic email templates** in Dataverse, enabling real-time data population and ensuring consistent messaging while allowing flexible updates for new business scenarios.
- Collaborated cross-functionally using **agile methodologies** and created a comprehensive audit log to ensure regulatory compliance and data consistency across HR and regulatory communications.
- Facilitated cross-functional collaboration by leading weekly meetings with stakeholders, gathering feedback to refine automation processes and resolve integration challenges.

Artificial Intelligence Research Assistant — Feb 2024 – July 2024

New Jersey Institute of Technology — *Newark, NJ*

- Developed synthetic datasets with NVIDIA Omniverse Isaac Sim by applying the **SAM model** and performed a comparative analysis between synthetic and real datasets, using conversion techniques (KITTI to COCO) to standardize data formats.
- Demonstrated that synthetic data achieved 90% of the accuracy of real data, enabling cost-effective and scalable AI research.
- Conducted rigorous data quality audits to ensure dataset consistency, supporting advanced AI model development and evaluation.

PROJECTS

Hospital Readmission Prediction | *Python, Scikit-learn, Ensemble models, TabNet, Seaborn* — GitHub

- Developed a predictive model to forecast 30-day readmission using real-world data aggregated from **130 hospitals** and **100,000 patient records**.
- Performed intensive **exploratory data analysis (EDA)** to handle outliers and missing values, applying chi-squared tests, Z-tests, and A/B testing for feature selection.
- Achieved a **0.75 AUC**, reducing false positives/negatives and enhancing patient outcomes while cutting healthcare costs; identified ACI results as a critical predictor.
- remove Implemented an end-to-end pipeline that includes continuous model monitoring and periodic retraining strategies, ensuring the model remains clinically relevant and delivers significant cost savings by reducing unnecessary readmissions.

Twitter & Reddit Sentiment Analysisl | *Python, AWS, Docker, Streamlit* GitHub

- Designed and deployed a **real-time sentiment analysis model** by training a **BERT model** on **1.6M tweets** and continuously ingesting weekly Reddit data.
- Aggregated and normalized data from multiple datasets; used upsampling to balance the neutral class and built an **end-to-end ETL pipeline** on AWS.
- Achieved **91% model accuracy** (94% precision for key classes) and developed an interactive **Streamlit dashboard** to rank tech companies by sentiment.

Bank Database Design | *Oracle SQL, Relational Database Design, 3NF Normalization* GitHub

- Designed and implemented a comprehensive relational database from scratch to support a multi-branch banking system, ensuring data integrity through rigorous testing and adherence to **3NF**.
- Developed complex **SQL queries** (aggregations, nested queries) and performed functional dependency analysis.
- Delivered a robust solution that provided actionable insights for loan approvals and inter-branch transaction volumes.

High-Frequency Trading Model | *Python, Streamlit, Time Series Modeling* GitHub

- Developed a **trading model** by collecting **5+ years** of multi-source financial data and synchronizing sentiment data for **Tesla** and **Ford** stocks.
- Feature-engineered **technical indicators** (including a custom **RSI**), applied **Random Forest** for feature selection, and conducted hyperparameter tuning using **Random Search** and **Grid Search**.
- **XGBoost** outperformed other models with **70% accuracy** in predicting stock movement.
- Deployed an interactive **backtesting environment** in **Streamlit** with **SHAP graphs** to elucidate model decisions.

EDUCATION

New Jersey Institute of Technology Newark, NJ

Bachelor of Science in Computer Science *December 2024*

- **Relevant Coursework**: Data Mining, Data Structures and Algorithms, Artificial Intelligence, Machine Learning

DATA VISUALIZATIONS (VIEW LIVE)

HR Analysis Dashboard

- Developed a **Power BI dashboard** that links sales data with budgets and performance metrics to show which teams and managers drive the most revenue.
- Employed easy-to-read visuals and filters to quickly highlight performance trends and key revenue drivers.

Sales Dashboard

- Created a **Power BI report** displaying minute-by-minute trading data, including profit, return rates, and profit per unit.
- Presented clear trends and simple recommendations to improve trading strategies, such as adjusting trading times and strategies.

Algorithmic Trading Report

- Built a **3-page interactive Power BI dashboard** that combined employee performance data from different sources to spot **top talent** and **early retirement risks**.
- Used clear charts and key indicators to provide simple, actionable insights for smarter workforce planning.

CERTIFICATIONS

Microsoft Certified: Power BI Data Analyst Associate (PL-300) *(Feb 2025)*

Google Advanced Data Analytics *(Sep 2024)*

AWS Certified Cloud Practitioner *(Oct 2024)*