

MAHMOUD SAAD

 908-630-7978

 mahmoud2511.saad@gmail.com

 Avenel, NJ

 [GitHub](#) | [Portfolio](#) | [LinkedIn](#)

PROFESSIONAL SUMMARY

Aspiring Data Scientist with a B.S. in Computer Science (NJIT, 2024) and 2+ years of experience applying machine learning, predictive modeling, and cloud-based analytics across healthcare and software development. Proven ability to build scalable pipelines, deploy production-grade models, and translate data into strategic decisions. Adept in Python, SQL, AWS, and Power BI, with a strong foundation in both academic research and industry-focused automation projects.

Certifications: Microsoft Power BI Data Analyst Associate (Feb 2025) | Google Advanced Data Analytics (Sep 2024) | AWS Certified Cloud Practitioner (Oct 2024)

KEY ACHIEVEMENTS

- Delivered an **Azure-based** engine at **Merck**, automating **5,000+ emails/year**, saving **4,300 hours** annually, and improving audit readiness for regulatory compliance.
- Earned **Best ML Project** at **HackNJIT** and **3rd Place** in **NJIT Capstone** for developing real-time stock prediction models and enterprise automation solutions.
- Built and deployed predictive models for hospital readmissions achieving **80% accuracy** and social media sentiment with **91% accuracy**, leveraging cloud-based pipelines in healthcare and tech.
- Designed **interactive dashboards** and model explanations that influenced strategic decisions at **Merck** and across academic research teams.

SKILLS

Languages: Python, SQL, C++, TypeScript, JavaScript | **Data Science Tools:** Pandas, NumPy, Scikit-learn, TensorFlow, PyTorch, Power BI, Tableau, Matplotlib, Seaborn, Streamlit, Excel | **Machine Learning:** Random Forest, XGBoost, TabNet, BERT, LSTM, PCA, Neural Networks, Feature Engineering, Hyperparameter Tuning | **ETL & Cloud:** AWS (S3, Lambda, ECR, EventBridge), Azure (Power Automate, Dataverse), Docker, SharePoint, API Integration | **Development:** Git, Agile Methodologies, Test-Driven Development (TDD), MLOps (Model Monitoring, Retraining), 3NF Database Design

EXPERIENCE

DATA ANALYST - Remote

Merck Pharmaceuticals, September 2024–December 2024

- Engineered an **Azure-based** notification engine using **Power Automate**, integrating **SharePoint** and **Dataverse** to process thousands of daily emails, saving **4,300 hours** annually (**\$200K+** in labor costs).
- Led weekly stakeholder meetings using **Agile methodologies**, translating business needs into technical solutions with **high-level architecture diagrams** & iterative feedback, presenting to **Merck’s COO** with **Power BI** dashboards.
- Optimized scalability by adhering to **Microsoft Power Platform** best practices, enabling seamless handling of high-volume notifications with **zero delays**.
- Developed dynamic **email templates** and **audit logs**, reducing human error by **80%** and ensuring compliance with regulatory standards for HR and clinical communications.
- Mitigated spam risks by scheduling operations and securing data in **Dataverse**, addressing stakeholder concerns and reducing email volume by **80%**.

ARTIFICIAL INTELLIGENCE RESEARCH ASSISTANT – Newark, NJ

New Jersey Institute of Technology, February 2024–July 2024

- Generated synthetic datasets using **NVIDIA Omniverse Isaac Sim**, achieving **88% accuracy** in AI segmentation tasks by converting **KITTI** to **COCO** formats.
- Conducted comparative analysis of synthetic vs. real datasets, demonstrating **90% accuracy** parity, enabling cost-effective **AI model training**.
- Performed rigorous **data quality audits** to ensure consistency, supporting advanced **computer vision** research for autonomous driving applications.

NOTABLE PROJECTS

Hospital Readmission Prediction | *Python, Scikit-learn, TabNet, PCA*

- Built an **80% accurate** model using 10 years of data from 130 US hospitals to predict 30-day readmissions for diabetic patients. Identified **A1C testing**, discharge disposition, age, and race as key predictors, potentially saving **billions in Medicare costs** (\$15K/case vs. \$70/test). Applied **Z-scores** to remove outliers, **chi-squared tests** for feature selection, **PCA**, and **upsampling** to address 10% class imbalance, improving precision and recall to 70%. Implemented **continuous monitoring** and structured feature testing cycles to ensure clinical relevance while anonymizing **patient IDs** for data privacy.

Twitter & Reddit Sentiment Analysis | *AWS, BERT, Docker, Streamlit*

- Designed a real-time NLP system (initiated as an early academic project) using a **BERT** model trained on 1.6M tweets and weekly Reddit streams, achieving **91% accuracy**. Built a production-grade **ETL pipeline** on AWS (Lambda, S3, ECR), handling data balancing with **weighted loss functions**, automating weekly retraining to maintain performance. Delivered interactive **Streamlit** dashboard ranking tech company sentiment, ensuring ethical use via **data anonymization**. Rapidly learned **AWS** via tutorials; optimized costs (**AWS Cost Calculator**) for scalable deployment.

High-Frequency Trading Model | *XGBoost, Streamlit, Time Series*

- Created a **74.63% accurate** trading model for Tesla/Ford, expanding features from 7 to 24 (e.g., **RSI**, moving averages, news sentiment) via **Grid Search** optimization. Delivered real-time **SHAP-based** predictions in an interactive **Streamlit** app, enabling data-driven trading. Won **HackNJIT Best ML Project** for innovative **feature engineering**.

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

- Designed a normalized **relational schema** for a multi-branch bank system, ensuring **3NF compliance** via functional dependency analysis. Wrote complex **SQL queries** (nested joins, aggregations) to extract loan approval and transaction insights, improving performance by **30%** through optimized design.

DATA VISUALIZATION PROJECTS ([VIEW LIVE](#))

- HR Analysis Dashboard** (*Power BI*): Linked sales outcomes with team- and manager-level budget allocations to uncover **high-ROI contributors**. Used **star schema modeling** and custom filters to enable real-time exploration of performance drivers across departments.
- Sales Performance & Workforce Risk Dashboard** (*Power BI*): Built a 3-page solution integrating employee performance, sales revenue, and HR metrics. Identified **top-performing staff** and **early retirement risks** using **KPI cards**, filters, and dynamic visuals to support strategic workforce planning.
- Algorithmic Trading Report** (*Power BI*): Visualized real-time, minute-level trading performance (profit, return rate, profit/unit) for a high-frequency bot, using **SQL** to transform data from **Google Big Query**. Recommended strategy adjustments based on trading times and **Nasdaq momentum trends**, contributing to a potential doubling of returns.

EDUCATION

Bachelor of Science: Computer Science New Jersey Institute of Technology - Newark, NJ | December 2024

- Dean's List: 4 Semesters | Key Coursework: Data Mining, Data Structures & Algorithms, Artificial Intelligence, ML