Yahan Yang
yangyahan0321@gmail.com | 617-331-6941 | LinkedIn

SUMMARY

3+ years of industry experience in data analytics. 2+ years of industry exposure to machine learning and data engineering. Expert in **SQL, Python**, and developing automated data pipelines, predictive models, and dynamic visualizations using tools like **Snowflake, Redshift, Jupyter Notebook, and Tableau**. Proven ability to deliver actionable insights that drive business outcomes, with strong communication skills and a track record of collaborating effectively with both technical and non-technical teams to inform strategic decisions.

WORK EXPERIENCE

Putnam Associates Boston, MA   Apr. 2022 – May. 2024
Junior Data Scientist

- Developed **SQL** queries (**Snowflake, Redshift**) to integrate multi-source claims data and generate key insights via Tableau dashboards. Improved treatment feature identification algorithms, increasing feature capture accuracy by 5% and fully captured treatment site by 4.5%. Automated monthly analytics process, reducing manual effort by 50%.
- Implemented statistical modeling (**K-means, Bayesian networks, Random Forest**) to segment patients within various client's portfolio, to identify common patient journeys, which enables client to optimize patient opportunity.
- Utilized advanced data cleansing and imputation techniques (**K-NN, Random Forest**), and machine learning algorithms (**Random Forest, XGBoost, H2O AutoML**) to predict physician prescribing behavior, achieving 89% accuracy in predicting patient treatment eligibility.
- Spearheaded and contributed to 12+ successful patient journey analytics and market opportunity projects by analysing healthcare large-scale data in oncology, rare diseases and blood disorders, effectively supporting business decisions for biopharmaceutical clients.
- Demonstrated strong multi-tasking, communication and presentation skills, effectively addressing client needs, collaborating with cross-functional teams, and delivering timely insights.

Vee New York, NY   Dec. 2021 – Apr.2022
Data Science Intern

- Built an ETL pipeline for candidate evaluation, including multi-source data aggregation, data cleaning and transforming with Lambda, and API Gateway establishment to **Power BI** dashboards. The whole streamlined evaluation processes reduced overall 20% recruiting time.
- Collaborated with people science team to design and embed the Power Apps visuals into Power BI dashboards, and trained end-users on automated updates via API Gateway.

PathosAI. Medford, MA – Github Link   Feb. 2021 – May 2021
Data Science Intern

- Designed an LSTM-based **deep learning** model (TBANet) using **Keras** and **Google Cloud**, improving emotion detection accuracy by 16% through temporal and bodily attention mechanisms.
- Partnered with company founders to design BI dashboards showcasing TBANet's result, including experimental tracking and attention score analysis.

EDUCATION

Northeastern University *Operational Research* (GPA 4.0/4.0)		May 2021
Shandong Agricultural University *Information and Computing Science* (GPA 3.5/4.0)		June 2019

TECHNICAL SKILLS

- Languages and Scripts: Python, SQL, R
- Databases: Snowflake, Redshift, PostgreSQL
- Data Science Libraries: NumPy, Pandas, Scikit-learn, TensorFlow, Keras, SciPy, NLTK, H2O
- Data Visualization Tools: Tableau, Power BI, MS Excel, Matplotlib, Seaborn
- Cloud-based Technologies: AWS (Redshift, Glue, S3, Lambda, API Gateway, Athena), Google Cloud Platform