MICHELLE (XIAOYUN) WANG

Email: xiaoyun.wang00@gmail.com | Phone: 919-638-1232 | Portfolio: xwang.me
3+ Yoe | Expert in Python ETL, SQL, Machine Learning| Duke & Fudan Alumni

PROFESSIONAL EXPERIENCE

Xenon Health   Oct 2023 – Present
Data Scientist (Full-Time) | Date Engineer ETL; Machine Learning; Web App Development   *Miami, FL*
- **Building an end-to-end ETL pipeline** using **Airflow**, and **Azure PostgreSQL**, automating data integration from Google Sheets and APIs, and reducing data processing time by **40%** through efficient data transformation and storage optimization.
- **Leveraging Spark (PySpark)** on **Azure Databricks** to process and analyze **3 million** patient records, optimizing data preprocessing for machine learning models and reducing data transformation time by **50%**.
- **Constructing predictive models** (Random Forest, XGBoost, SVM) to classify claims revenue into different categories, enhancing business forecasting accuracy by achieving an **AUC-ROC score of 0.85** and improving resource allocation.
- **Designing and optimizing SQL queries** for ad-hoc analyses and BI reporting, providing real-time insights to clinical decision-makers, enabling improved KPI tracking and operational performance across multiple departments.
- **Creating 10+ Power BI dashboards** to visualize patient risk scores and KPIs, providing actionable insights to over **5** clinical departments and improving operational efficiency across the board.
- **Developing a web application** using Azure WebApp, JavaScript, and HTML5/CSS to integrate dashboards for better access.
- **Mentoring 3 interns**, providing guidance on engineering SQL and Python scripts for reporting and visualization, fostering a collaborative environment and enhancing project management skills within an **Agile** framework.

Wells Fargo   Feb 2023 – May 2023
Data Scientist | NLP Analyze ESG reports; Business Intelligence; AWS SQL DBMS   *(Capstone Project) Durham, NC*
- **Developed and presented 10+ interactive Tableau dashboards**, providing actionable insights to senior leadership on sustainability and safety metrics, guiding strategic decision-making for key business initiatives.
- **Applied NLP techniques** using **NLTK** to analyze and extract actionable insights from **ESG reports**, supporting sustainability goals and performance tracking across multiple business units.
- **Implemented a robust data pipeline** on **AWS Cloud RDS**, improving data management processes for cross-functional teams.

Triangle J Council of Governments   July 2023 – May 2024
Modeling Scientist (Part-Time) | Forecast Modeling; Map-Based Reporting   *Durham, NC*
- **Developed transportation forecasting models** for the **Durham** and **RTP Metropolitan areas**, predicting key metrics such as **travel time reliability**, enabling data-driven decision-making for urban planning and infrastructure improvements.
- **Optimized predictive models** using **Python** to process large volumes of unstructured transportation data, improving model accuracy and performance by **15%**.
- **Created and published interactive online visualizations** and **map-based reports** with ArcGIS, effectively communicating complex data insights to cross-functional teams and non-technical stakeholders, driving actionable outcomes in transportation planning.

Tencent   June 2021 – Aug 2021
Data Scientist (Internship) | Data Analytics; Publish 2021 "Cloud + AI" White Paper   *Shanghai, China*
- **Analyzed data from 335 cities** across 22 digital industries to contribute to the **2021 "Cloud + AI" White Paper**, delivering a data-driven overview of emerging business trends, and providing strategic insights for Web 3.0 discussions.
- **Facilitated cross-departmental collaboration**, summarizing key discussions and producing **80+ presentation slides** for the 2021 Tencent Global Digital Ecosystem Summit, which played a pivotal role in fostering strategic partnerships.

EDUCATION

Duke University   July 2022 – May 2023
Master of Science, Quantitative Management Business Analytics (MQM)   *Durham, NC*
- Tutor & Grader for DECISION 610W: Probability and Statistics (2022 Fall), Applied Probability & Statistics (2023 Spring)

Fudan University   Sept 2018 – Jun 2022
Bachelor of Laws, International Politics   *Shanghai, China*
- Scholarship for Outstanding Student (top 5% students, 1st prize); Graduate Scholarship (2022)

SKILLS

Programming: C++/C, Python, SQL, R, PostgreSQL, MySQL, Excel (Pivot Table, VBA), Tableau, Power BI, Looker, Shell Script, Linux, Git, Scrum / Agile, Apache Hadoop, Apache Spark (PySpark), Apache Airflow
Machine Learning and Data Mining: Statistics, Supervised/Unsupervised Learnings, Model Optimization, A/B Testing
Data Warehouse / Cloud Platform: Azure Cloud Service, AWS Cloud Service (AWS Certified Cloud Practitioner), Google BigQuery