Xuan Wang

Austin, Texas | 512-200-6763 | xxxuan0213@gmail.com | linkedin.com/in/xxuan-wang/ | xuanwangg.github.io/

EDUCATION

MS In Information Studies - The University of Texas at Austin | GPA: 3.45/4.00   Aug 2019 - May 2021

BS In Electronic Commerce - Dalian University of Technology | GPA: 3.40/4.00   Sep 2015 - Jun 2019

PROFESSIONAL EXPERIENCE

Data Analyst - Angle Health Inc.   Apr 2023 - Sep 2023

- Employed regression models to evaluate risk scores, setting up metrics to identify & monitor high-cost cases. Collaborated with actuaries to redefine insurance rate strategies, boosting customer satisfaction while maintaining manageable loss ratios.
- Designed ETL automation connecting AWS Lambda/S3 and PostgreSQL databases with optimized queries to integrate internal & external data resources, improving data retrieval speed by 60% and boosting claims processing efficiency by 30%.
- Cultivated strong stakeholder-tech cohesion, prioritized and converted PM's business needs into executable decisions; Setup customized APIs for smooth data transmission with engineers and delivered tailored Retool views to clients.

Data Analyst (AI/ML) - Movable Ink   Mar 2022 - Jan 2023

- Developed ML models and established file health metrics to identify top-performing campaigns and Ads; Applied A/B testing and experimentation to diagnose and overcome campaign conversion rate plateaus, achieving an 18% increase in conversions.
- Collaborated closely with eng teams to customize GitHub Actions for deploying models through CI/CD pipelines, prepared data for downstream use, contributed to troubleshooting, resolution, and data validation to enhance data quality.
- Spearheaded effective SQL logics on BigQuery to process TB of inbound data, aligning precisely with data patterns & trends to support the modeling team with the feature selection phase.
- Built Tableau dashboards and proactively communicated with stakeholders for effective product improvement strategies, leading to a successful 20% boost in retention rate.
- Championed the development and implementation of a Python-based reporting automation tool on Cloud Run, realizing a 7X speed enhancement, significantly reducing manual input and minimizing human errors.

Data Science & Analysis Intern - *Renzoe Box, Inc.*   Sep 2021 - Mar 2022

- Leveraged NLP for multi-label classification of text data, accurately predicting 64 tags over 28k beauty products.
- Pioneered a content-based ML solution in Spark by integrating structured & unstructured data to enhance user engagement.
- Deployed machine learning models with Flask on AWS EC2 instances to enable result visualization, resulting in personalized makeup recommendations powered by neural networks and exhaustive search algorithms.

Data Research Assistant - Texas Department of Information Resources   Dec 2020 - June 2021

- Automated a multi-threaded web scraping Python app using requests, BeautifulSoup, and selenium to fetch 12k+ vendor records through Salesforce API and feed as inputs to 15 websites for targeted file collections.
- Impact: This app leads to a time efficiency >10X, and significantly reduced oversight by contract managers by 90%.

Cloud Computing TA - UT Austin, Department of Computer Science   Jan 2021 - May 2021

- Designed Git-based automated Python grading scripts with seamless integration, reduced human supervision by 70%.
- Developed S3 & DynamoDB handlers to facilitate distributed data manipulation with AWS Python SDK in virtual envs.
- Led a group of 60+ students in container exploration and successfully deployed of Helm Charts on a GKE single-node cluster.

Big Data & Distributed Programming TA - UT Austin, McCombs School of Business   Aug 2020 - Jan 2021

- Guided students in leveraging EC2 GPUs to experience quicker execution capabilities of TensorFlow/Keras models.
- Illustrated the essential operations of RDDs on distributed datasets exceeding 1TB in size, employing Apache Spark to implement a collaborative-filtering movie recommendation system.

SKILLS

Programming Languages: Python, SQL, PostgreSQL, Apache Spark, Git, R; Scikit-learn, Pandas, NumPy, Tensor-flow, Keras

Tools: Tableau, AWS (S3, DynamoDB, EC2), GCP BigQuery, Salesforce, Jupyter, Kubernetes, Azure, Docker, Excel