Yuan Yuan

Manor, TX 78653 | 832-352-5218 | yyccphil@gmail.com | linkedin.com/in/philyy | github.com/yyccPhil

Education

Rice University, Houston, TX	2022/08 - 2023/12

Master of Science in Computer Science, GPA: 3.86/4.0

Beijing University of Technology, Beijing, China	2016/09 - 2020/07

Bachelor of Engineering in Electronic Information Engineering (Hons), GPA: 3.5/4.0

Experience

Tesla, Inc., Austin, TX	*Data Engineer*	2024/01 - Present

- Deployed an **Airflow** on **Kubernetes** from scratch, incorporating features like **CI/CD** workflows for automated Helm chart updates, S3 logging storage, and SMTP email alerts to enhance robustness, and set up Grafana alerts sent to Teams channel to monitor pod health; managed 50+ active DAGs, achieving a system availability of 99.97%.
- Developed multiple real-time alert systems across factories by integrating streaming data via CDC from RDBMS with **Kafka**, REST APIs, and Splunk, enabling proactive detection of critical NC and significant inspection status changes.
- Built 2 data marts from scratch on **SQL Server** and **ClickHouse**, reducing dashboard average load time by over 50%; and implemented distributed storage in ClickHouse to enhance performance and support high-volume data analytics.
- Created a sentiment analysis tool, extracted Reddit via PRAW, scraped CyberTruck Forum using Scrapy, applied NLP methods to process data and trained a BERT model to label quality-related comments with an accuracy exceeding 91%.
- Created and maintained 10+ Tableau dashboards covering multiple quality areas for daily quality issue monitoring.

Torqata Data and Analytics LLC, Huntersville, NC (Remote)	*Data Engineer (Intern)*	2023/05 - 2023/09

- Built an end-to-end **ETL** pipeline from 5 auto clients based on **GCP** to feed the data warehouse and power Looker.
- Implemented a highly efficient database schema, ensured that every transaction placed was available in the data warehouse within 15 minutes of order placement and the low latency dashboard can adjust the filter within 5s or less.
- Collaborated closely with data scientists using Vertex Workbench to implement data solutions that drive key corporate decision-making, resulting in an 1100+ expansion of the TAM for retail applications thereby affecting OKRs.
- Conducted comprehensive analysis of Pricing and Transaction datasets, extracting and comparing data at the SKU date level to uncover retail pricing patterns, resulting in clients expanding their market reach by over 200%.

Maimai, Inc. (Chinese version of LinkedIn), Beijing, China	*Data Analyst (Intern)*	2022/04 - 2022/07

- Built and maintained content models for consumption, production, retention, and experimentation, along with real-time partition data monitoring for 10M+ active users; supporting the growth of "circle" business from 0.3M to 4M+ users.
- Clarified product requirements and used ClickHouse for efficient data querying and Redash for data visualization.
- Designed, implemented, and analyzed A/B testing, enhancing product experience by optimizing business strategies; such as adjusting post-display order, resulting in 20% increase in readings per person and 30% rise in circle entry rate.

Project

Key Frame Extraction Based on Global Motion Statistics for Team-sport Videos (SCI, Q2)	2020/01 - 2021/01

Beijing Municipal Key Lab of Computation Intelligence and Intelligent Systems

- Proposed a **Computer Vision** framework named KEGMS (github.com/yyccPhil/KEGMS), which designs rules for camera resultant motion amplitudes instead of just blindly increasing network depth for improved performance.
- Automated the identification of camera translation and zooming magnitudes based on corner points of the optical flow.
- Built SportKF as a dataset for team-sport videos key frame extraction; performed validation using 4 types of sports videos collected from YouTube-8M, covering a total of 200K+ frames with 764 ground-truth key frames.
- Experimentally benchmarked the performance, showing favorable results that the KEGMS achieved the highest recall of 0.98, which is higher by at least 15.3%; achieved the highest F value of 0.44, which is higher by at least 10.8%.

Skills

Python, pandas, MySQL, SQL Server, ClickHouse, Airflow, K8s, Docker, CI/CD, REST API, Tableau, Kafka, Azure