LI-AN YANG

858-344-2107   luke831215@gmail.com   linkedin.com/in/li-an-yang   li-an.me

Skills

Areas of Interest: Personalization, Real-time Ranking System, Data Engineering, Information Retrieval

Skills: Apache Storm, Apache Beam, BigQuery, GCP Dataflow, Kafka, Redis, Kubernetes, Terraform

Miscellaneous: Distributed Streaming System, Recommender System, Model Serving Automation

Experience

SOFTWARE ENGINEER, YAHOO INC. – MAR 2020-PRESENT

A DATA INGESTION PIPELINE THAT PREPROCESSES RAW EVENTS AND PUBLISHES THEM TO APACHE KAFKA IN REAL-TIME

‣ Saved 50% of machine resources and 20% of redundant traffic in one upstream consolidation
‣ Tuned the JVM-based pipeline to scale with the volume of user instrumentation events at Yahoo
‣ Enriched data with Redis caching for downstream use cases, i.e., feature computation, user profiles, and analytics

HULKBUSTER - A RECOMMENDATION REASONING PIPELINE THAT BADGES ARTICLES FOR USERS

‣ Delivered badges such as *Most Read, Most Discussed, You Follow,* and *New Topic* based on ML features
‣ Served as the Engineering Lead in collaboration with a group of Product Managers, Designers, and Editors

LEOPARDON - MODEL SERVING AUTOMATION FOR CONTENT RECOMMENDATION SYSTEMS

‣ Raised model velocity to 100% automation by helping Data Scientists launch online experiments in one go
‣ Enabled continuous deployment of retrained models for online A/B testing

SOFTWARE ENGINEER INTERN, YAHOO INC. – JUN-SEP 2019

‣ Designed a Referrer Feature to capture content signals coming from external domains
‣ Completed a full project cycle from event processing, feature computation, to re-training models

Projects

AMAZON PURCHASE PREDICTION CHALLENGE ON KAGGLE

‣ Ranked 7th among 800 people with an accuracy of 72.96%

A VOICE-CONTROLLED STREAMING JUKEBOX BASED ON IBM BLUEMIX CLOUD SERVICE

‣ Established music streaming services on Raspberry Pi featuring personalized music recommendation

AUTOMATIC PLAYLIST CONTINUATION AND ADD-TO-PLAYLIST PREDICTION ON SPOTIFY

‣ Suggested the playlists for users to add a liked track to with logistic regression models and XGBoost

Education

University of California, San Diego – M.S. in Computer Science, Mar 2020
National Chiao Tung University – B.S. in Computer Science, Jun 2016