Terry Tan

North York, ON | tanjiarui@hotmail.com | e-Portfolio: https://tanjiarui.dev/projects

DATA ENGINEER
specialized in data mining, data clean, and data pipeline

HIGHLIGHTS OF SKILLS

- Python3.X (senior), familiar with Numpy, Pandas, Matplotlib, and Scikit-learn
- Data warehouse/ETL (junior), familiar with dimensional modeling, schema design, and normal form modeling
- SQL and NoSQL (intermediate), familiar with MySQL, SQL Server, HiveQL, BigQuery and MongoDB
- Web framework (intermediate), familiar with Flask, FastAPI, and Socket.IO
- Cloud service (junior), familiar with AWS service and Azure service such as Web App, AD, Graph API, and Cosmos
- RPA (junior), familiar with UiPath Studio, VBA, and Airflow
- CI/CD (junior), experience in Docker, Compose, and Swarm
- Scala2.12 (junior), familiar with Spark core, Spark SQL, Structured streaming, Spark MLlib
- Troubleshooting technical issues and bugs
- Having experience working in an agile team
- Fluent in Mandarin and English

EDUCATION

Diploma, Artificial Intelligence   Sept 2020 – May 2022
Centennial College, Scarborough, ON
- **Relevant Courses:** Introduction to Artificial Intelligence, Advanced Database Concepts, Supervised Learning, Big Data Tools for Machine Learning, Unsupervised and Reinforcement Learning, Neural Networks, Natural Language Processing and Recommender Systems, Deep Learning, Cloud Machine Learning

Master of Engineering, Software Engineering   Sept 2016 – Jun 2020
University of Science and Technology of China, Suzhou, China
- **Relevant Courses:** Multimedia Signal Processing, C++ Programming, Multimedia System and Application, Design of Neural Network, IT Service Management, Big Data Basis, Information Security

Exchange Program at Clemson University, South Carolina, USA   Aug 2017 – Feb 2018
- **Relevant Courses:** Statistical Method (test hypothesis, completely randomized design, randomized block design), Software Verification, Applied Data Science (Regression, Linear/Non-linear logistic, SVD, SVM, k-means)

Bachelor of Engineering, Communication Engineering   Sept 2011 – Jul 2015
Chongqing University of Posts and Telecom, Chongqing, China
- **Relevant Courses:** C Language Programming, Signal and System, Digital Signal Processing, Database Application, Principle of Microcomputer and Interface Technology

EDUCATION PUBLICATIONS

Terry Tan. 2019. Singing Evaluation based on Deep Metric Learning. Proceedings of the 2019 3rd International Symposium on Computer Science and Intelligent Control. Association for Computing Machinery, New York, NY, USA, Article 17, 1–5.

WORK EXPERIENCE

BBD Insights   July 2022 – Present
Principal Developer
- Design and implement a model governance system for the bank's validation department, which was facing model management chaos
- According to the requirement, the system was modularized as five components: model, validation, issue, OPM and admin
- Each module experienced rounds of development and validation based on the continuous feedback. The project was delivered and deployed on Azure Kubernetes Service

- The frontend applied React and Azure Active Directory for the authentication process, while the backend applied Flask, Graph API, and Azure SQL
- The project was successfully launched and passed the OSFI E-23 audit. It also improved the work efficiency of the validation team by two times and ensured smooth data migration

WIMTACH, Centennial College   **Jan 2022 – Jun 2022**
Student Researcher
- Help a game start-up to analyze players' mental and psychological health by building a facial emotion recognition framework
- The framework was split as data pipeline and model pipeline. The data pipeline is for image annotation and transformation while the model pipeline is for training and inference
- The data pipeline adopted AWS Rekognition and S3 services. The model pipeline adopted Efficientdet for end-to-end modeling, onnx runtime for inference, and Grad-CAM for explainable AI
- Ending up with 0.74 mAP and 0.3 fps

Artefact   **Mar 2020 – Oct 2020**
Data Engineer
- Built a data clean and mining pipeline by analyzing client CRM and user behavior data, and then extracting correlated features
- Improve clienteling via an ensemble learning model to predict second purchase propensity based on extracted features, resulting in a metric with 77% accuracy and 71% F1 score
- Create RPA workflow to extrapolate product volume up to 2030 using growth trend from past decade
- Apply Airflow and VBA to automate the report generation process after the business case was determined, resulting in a detailed and complete Excel report for strategy and decision-making

Institute of Automation of the Chinese Academy of Sciences   **Apr 2018 – Mar 2019**
Intern
- Test the C++ based deep learning model quantizer using parallel test approach
- Compared the runtime and precision of quantized and not quantized models, ending up with compressed runtime by 30% with only 0.015 mAP loss

OTHER QUALIFICATIONS

Microsoft Office Specialist   **May 2015**
Online Courses:
- Achieving Advanced Insights with BigQuery (Coursera)   **Mar 2021**
- Principle and Application of Big Data (MOOC)   **Mar 2020**
  - HDFS (HA, Federation), MapReduce, YARN, NoSQL, Hive
  - Spark Core, Spark SQL, Spark streaming, Spark MLlib
- Deep Learning Specialization (deeplearning.ai)   **Mar 2018**