Yuncheng Liang

E-mail: liangyuncheng2019@163.com |Telephone: +14376846660 |LinkedIn: http://linkedin.com/in/yuncheng-liang-4a1147133

Skills & Certifications

Languages: Python, SQL, PySpark, R, CSS, JavaScript

Tools / Techniques: Pandas, Scikit-learn, TensorFlow, PyTorch, Azure, AWS, Databricks, Docker, Kubernetes, Airflow, Flask, Oracle / Hadoop, Spark, CI/CD, MLOps, Github, NLP, Computer Vision, Recommendation System, A/B Testing, Prompt Engineering, LangChain

Certificates: Microsoft Azure Data scientist Associate, Microsoft Azure Data Engineer Associate, Microsoft Azure AI Engineer Associate, Microsoft Azure Power BI Analyst Associate, Microsoft Azure Developer Associate, Google Data Analytics Certificate

Work Experience

Rogers Communication, Data Scientist — May. 2022 – Present

- Developed end-to-end ML and data science projects as a core engineer of Roger's central AI team focusing on building ML solutions for NLP based customer churn and cancelation analytics, MLOps, feature store and data engineering.
- Implemented **SparkNLP** in NLP use cases as a replacement for Spacy, **reducing the ETL preprocessing time by 20+ times from 45 hours to less than 2 hours** further reducing computation and run-time cost.
- Utilized and fine-tuned **Guided LDA topic modeling** model for: churn customers' call drive analysis, spotting non-churn customers with cancel and move topic related conversations for customer service agents to follow up.
- Designed and trained **Doc2Vec and DNN** for churn and non-churn customer classification using online chat and call transcription data. Stacked output scores with other features such as embeddings and used **temporal aggregation** to create a customer level churn predictor, achieving **0.898 weighted AUC and 0.905 weighted F1-score on autoML**. Churn predictor was used by several downstream use cases and feature store to support customer level analytics, dashboarding and outbound campaigns.
- Created daily customer snapshots for model inference for above NLP churn classification pipeline and created a **feature drift detection module** to alert model retraining. The feature drift detection module then was also adopted by other projects' model inference pipelines.
- Supported **feature store** pipeline creation which serves to build an autoML model for churn classification using features across different projects including NLP, Modem Offline Prediction and Call Prediction. The final classification model **achieved 0.92 weighted AUC**.
- Consistently improved team's arsenal by developing real-time / batch processing pipelines, CI/CD pipelines, model based testing, BI dashboards, troubleshooting and debugging workflows, monitoring for models in production.

Baidu, Logistics and E-commerce Data Analyst Intern — Jun. 2021 – Sep. 2021

- Conducted **data automation using Python** to rebuild and optimize daily logistic and daily storage dashboard. Enabled new automated reports on: anomaly monitoring, delivery time monitoring, logistics cost calculation, cash on delivery estimation. Successfully **increased weekly sign rates** of major logistics provider **from 35-38% to steadily around 50% in 2 months**.
- Deployed strategies to propose, test and validate hypothesis regarding low sign rates. Through investigation, we found problematic events such as dishonest reporting of order confirmation from agents and some delivery worker charged addition money from customers.
- Communicated ETL code logic to data engineering department, together **built essential tables on-premises** for all team to consume.
- Led four interns across departments to predict signing of the orders and **achieved 67% accuracy using voting classifier**. Used **shapely value and partial dependency** plots to derive business insights and left the team with guidelines for model modification.

Personal Projects

Mask Detection End-to-End Project, https://github.com/YunchengLiang/Mask_Detection_Project

Adopted a two-stage architecture of first using pretrained SSD for face detection and then feed into 'mask on/not on face' classification model based on **mobilenet-v2 finetuned on dataset augmented by using openCV**. Served on webpage through flask framework.

Bertopic Topic Modeling Project, https://github.com/YunchengLiang/Bertopic_project

Topic modeling using Bertopic of three versions: **guided, dynamic, semi-supervised, includes side-to-side comparison with Guided LDA** using telecommunication conversation datasets, Elon Musk's tweets between 2015 to 2020 and Research articles with labels.

Education Background

University of Toronto — Sep. 2020 – Jun. 2022

Major: Chemical Engineering—Analytics Specialization GPA: 3.8/4.0 Master of Engineering

Courses: Foundations of Data Analytics and Machine Learning, Deep Learning, Data Mining in Engineering, Big Data Science, Applied Mathematics, Algorithms and Data Structures, Applications of Statistics, Process Data Analytics, Artificial Intelligence in Finance