Oliver K. Pan

github.com/oliverkpan   linkedin.com/in/oliverpan   oliverkpan.com   oliver.pan@mail.utoronto.ca

TECHNICAL SKILLS

Languages & Frameworks: Python, SQL (Oracle, MS SQL, MySQL), C, R, Flask, ReactJS

Data Science: pandas, numpy, scikit-learn, Keras, Tensorflow, nltk, SpaCy, PyTorch, Tableau, FastAI

Developer Tools: Git, Github, AWS (EC2, SageMaker, Lambda), Apache Airflow, Spark, Docker, Hadoop, SQLite

EXPERIENCE

Data Science Intern
May 2021 - August 2021

CIBC Data Studio
Toronto, ON (Remote)

- Managed internal **NLP application,** implementing **BERT** to automatically summarize articles, deploying the app with **AWS (Amazon EC2)** and packaging content with **Docker.**
- Customized **NLP pipeline** using **SpaCy,** scraping 160 article sources and implementing **sentiment analysis** to detect money laundering behaviour, optimizing runtime of company application by **70%.**
- Generated synthetic data, running various **Monte Carlo simulations** to create data for internal APIs.

Data Management Operations Intern
September 2020 - December 2020

Ontario Teachers' Pension Plan
Toronto, ON (Remote)

- Formulated **ETL** process, pulling from large databases by writing **100+ SQL queries** and constructing **Python scripts** on 32 critical datasets, capturing **100%** of data breaches during 2 months of production.
- **Automated 10+ Python scripts**, notifying business stakeholders by email notifications, which resulted in reducing daily time spent by **25%** for a team of eight members.

Data Science Intern
June 2020 - August 2020

Royal Canin
Guelph, ON (Remote)

- Developed **A/B tests** on 300,000 consumer purchases and modelled **customer churn** with **XGBoost,** using **survival analysis** and **SHAP** to present findings, which resulted in the company's investment of **$100,000.**
- Built a **Streamlit** app from a hyper-parameter optimized **CatBoost model,** to assist veterinarians in **multi-label classification** of urolith stones, achieving an accuracy of **90% F1 Score**.
- Constructed time-series models with **ARIMA** and **FBProphet** to forecast product shipments, displaying weekly results with **Flask,** ultimately improving accuracy by **30%** using machine learning methods.

Data Analyst Intern
September 2019 - December 2019

CardSwap Inc.
Toronto, ON

- Investigated **500+** transactions, reporting suspicious activity with **Excel**, preventing **$1000+** in fraud purchases.

PROJECTS

Menda (Hack Western 7) | *Python, OpenCV, nltk, Flask, Firebase*

- Developed mental health management application which utilized **OpenCV** for emotion detection and **nltk** for sentiment analysis, connecting ML Model with **Flask** and storing backend data with **Firebase.**

Iowatch | *Python, matplotlib, scikit-learn, XGBoost, CatBoost, Streamlit*

- Scraped pricing data with **BeautifulSoup**, predicting house prices with an accuracy of **<0.05 normalized RMSE**, using techniques such as **Logistic Regression** on 80+ variables to deploy model with **Streamlit**.

EDUCATION

University of Toronto
September 2018 - December 2023

B.Sc in Statistics and B.B.A in Finance, Minor in Computer Science
Toronto, ON

- **Activities:** Teaching Assistant, Management Economics Student Association, Calculus Peer Tutor