

OLIVER K. PAN

 oliver.pan@mail.utoronto.ca  oliverkpan  oliverpan  oliverkpan.com  (416) 832-0873

TECHNICAL SKILLS

Languages & Frameworks

Python · SQL (Oracle, MS SQL, MySQL) · Presto · R · Flask · Ruby · Matlab · Javascript

Data Science

pandas · numpy · scikit-learn · Tensorflow · Streamlit · Spacy · PyTorch · Tableau · OpenAI · FBProphet

Developer Tools

Git · AWS (EC2, SageMaker, Lambda) · Apache Airflow · PySpark · Docker · Hadoop

PROFESSIONAL EXPERIENCE

Data Engineer Intern

May 2022 - Aug 2022

Shopify

- Designed an **extraction data pipeline** by pulling 1000+ rows of external company data from an **API** and writing files to **GCS**, constructing data schemas with Python to provide dimensional data models to sales team daily.
- Utilized **Ruby & Longboat** to engineer two crucial data extractors, streamlining processes with **Apache Airflow**.
- Modularized data flow by querying three million merchant balances & under sampling with **SMOTE**, then using **SKOPE-rules** to define ML rules with **Precision of 98%**, ultimately generate insights on high-risk merchants.

Data Science Intern

Jan 2022 - Apr 2022

Shopify

- Clustered over **100,000 companies** extracted from an external **API**, utilizing **KPrototypes** for various data types, resulting in generating insights on **Total Addressable Market (TAM)** for 10+ operations stakeholders.
- Constructed performance dashboard with **Mode**, writing **20+ queries** with **PrestoSQL** and designing visuals, showcasing pilot project initiatives to over 50 data scientists via automated daily runs.

Data Science Intern

May 2021 - Aug 2021

CIBC - Canadian Imperial Bank of Commerce

- Designed internal **NLP** web application for 30+ team members, implementing **XLNet & Bert** to summarize articles with **98% Precision**, packaging content with **Docker** and deploying app with **AWS (Amazon EC2)**.
- Customized and re-trained **NLP pipeline** using **SpaCy** to improve entity detection **accuracy up to 88%**, **scraping 600+ articles** to detect money laundering behaviour; refactored code to **optimize runtime by 70%**.

Data Management Operations Intern

Sep 2020 - Dec 2020

Ontario Teachers' Pension Plan

- 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.

Data Science Intern

Jun 2020 - Aug 2020

Royal Canin

- 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**.

Research Assistant

Sep 2021 - Present

University of Toronto

- Performs data analytics research on 4000+ undergraduate students, clustering data with **K-Means & Gaussian Mixture Modelling**, to understand student personas and student archetypes from extracurricular activities.
- Conducted literature review on 10+ academic articles provided by JEE & IJEE, to improve internal research materials.

PROJECTS

drupp.io | [Github](https://github.com/oliverkpan)

- Utilized OpenCV to develop a computer vision application that scans personal outfits and returns recommendations, designing prototype with Streamlit and deploying application via Heroku

EDUCATION

University of Toronto

Expected 2024

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