

# YUAN ZHANG

✉ yuanz2299@gmail.com ☎ 514-651-2668 📧 yzhang0.github.io 🌐 yzhang0

## EDUCATION

### University of Waterloo, Canada

Sep 2015 - Aug 2019

Honours B. Math., Joint Major in Computer Science and Statistics

Coursework: Algorithms, Data Structures, Operating Systems, Object-Oriented Programming, Statistical Learning, Forecasting, Applied Linear Models.

## SKILLS

### Languages

Python (pandas, scikit-learn, flask), SQL, C++, PHP, D3.js, Java, R (dplyr, xgboost)

### Tools/OS

Hadoop (Spark, MapReduce, Oozie, Hive, Sqoop), git, bash, Linux, LaTeX

### Interests

Applied Machine Learning, Data Visualization, Time Series, Deep Learning

## EXPERIENCE

### Facebook

Jan 2019 - Apr 2019

Software Engineering Intern

Menlo Park

- Built and launched classifiers on **fact checkability** of content using **gradient boosting** and **word embedding** to optimize Fake News ranking algorithm. The model is currently used for two post types and improves misinformation models' precision by **25%**.
- Initiated and implemented the usage of new features in current misinformation models.

### Salesforce

May 2018 - Aug 2018

Software Engineering Intern

San Francisco

- Optimized platform components in **Python** and **Java** with replication and scalability features to transport large scale data from key-value store to relational databases.
- Developed **ETL pipelines** and **UDF** to efficiently store, query and analyze data in **Hadoop (Oozie, Hive)**.
- Developed new database schema with adjusted metric definitions to minimize discrepancies.
- Assisted data scientists with **time series** analysis to predict resource inadequacy events.

### Scotiabank

Sep 2017 - Dec 2017

Data Science Intern

Toronto

- Cleaned and feature engineered millions of customer browsing data points using **R** and **Hadoop Framework (HDFS, MapReduce, Spark)** to investigate the efficiency of third party data for marketing campaigns.
- Utilized **R** for data exploration and dimensionality reduction through data plots and best subset selection.
- Tuned hyperparameters through grid search and k-fold cross validation techniques.
- Predicted digital banking likelihood of new customers using **xgboost** and random forest with **73%** accuracy.

### IBM

Jan 2017 - Apr 2017

Data Analytics Developer

Ottawa

- Built an **automated** system to create scheduled data visualization using **D3.js** and **Jinja2**.
- Made recommendations on data visualizing libraries based on experiments with **Kibana**, **Plot.ly**, and **NVD3**.

## PROJECTS

### Instagram Destinations

In Progress

- Created a map of suggested locations from 10 travel instagrammers with pictures using **Mapbox**.
- Implementing real time data fetching using a **PostGres** database and allow hashtag/username queries.

### Bookclub

Jul 2018

- Created a website where users can sign in to review 5000 books from the GoodRead API.
- Connected database using **SQLAlchemy** and created web application using **Flask** and **Bootstrap**.

### Predict and Visualize Software Internships Trends at UWaterloo

Jan 2018

- Built a reactive directed graph of 86 most recurrent companies and their number of interns using **D3.js**.
- Collected company ratings, revenue, etc. from Glassdoor using **Selenium** and **Requests**.
- Trained an **ensemble model** to predict the most fitting company given a student's experience.

### ChamberCrawler3000 (CS246)

Jul 2017

- Designed and implemented 20 character, potion, and treasure classes in **C++** to interact with game engine.
- Developed random generator algorithm to place items of different frequencies on floor map.

## ACTIVITIES AND VOLUNTEER EXPERIENCE

### Future Blue Lead, IBM

Jan 2017 - Apr 2017

- Co-Organized bi-monthly career development workshops and social events for 30 interns.