

MICHAEL YAN

Principal Data Scientist

📞 4164185818 @ tanbingy@gmail.com 🌐 www.thinkdatascience.com

EXPERIENCE

Principal Data Scientist

Ingram Micro

📅 01/2019 - Ongoing 📍 Toronto

- Built NLP topic models including Logistic regression, Naive Bayes and SVM using Elasticsearch and Python sklearn to identify potential customers from over 1TB unstructured text data
- Scraped and stored websites of 3M+ business websites based on Dun & Bradstreet data and built customized analyzer in Elasticsearch reflecting complex business rules
- Researched and developed various product recommendation engines including collaborative filtering and probability graphic models for IM's online shopping platform in Canada, Germany and Spain

Senior Data Scientist

Crescendo Technology (Pinnacle Sports)

📅 07/2016 - 12/2018

- 📍 Toronto
- Successfully built/tested/deployed different machine learning models for major sports such as Soccer, NBA, NFL to predict game spread and total
- Successfully built an automated trading system in R using own model predictions and achieved 3% return in 3 months
- Built different machine learning models for NBA live game predictions such as LM, Poisson Regression, XGBoost and Random Forest
- Built client facing interactive web apps using R/Shiny to exploring model predictions, conveying important insight to business stakeholders
- Developed and maintained Slack app integrations which have been using company-wide using AWS Lambda, EC2 and API Gateway

Data Scientist

Compass Digital Labs

📅 06/2015 - 07/2016 📍 Mississauga

- Built an in-app personalized campaign recommendation system which helped marketing team increase campaign performance over 20%
- Successfully built market segmentation model on customer/business unit level, helped business development team acquire two major accounts in the six months
- Mined Twitter data using twitter API to identify food trend, analyze social sentiments of different products using
- Conducted market segmentation analysis using various clustering algorithms to help business stakeholders to understand value-based customer group

Quantitative Analyst

First Paladin Inc

📅 01/2013 - 05/2015 📍 Toronto

- Developed and back-tested multiple hierarchical regression model to predict equity returns for various markets globally
- Worked with portfolio managers in the identifying of stock screening factors including fundamental factors, quantitative technical factors and macroeconomic factors
- Designed and implemented a complex ETL process using data source from FactSet, WorldBank and IMF

Revenue Analyst

Canadian Primedia

📅 11/2011 - 12/2012 📍 Toronto

- Create and managed MS SQL database and data ETL process
- Manage and support in the development, testing and implementation of sales reports for the internal and external sales team

SUMMARY

A full stack data scientist with 8+ years of experience of machine learning and predictive modeling in finance, retail and gaming industries. A person with insatiable intellectual curiosity who is passionate about finding actionable insights hidden in vast amounts of raw data. Currently leading a team of 4 Data Scientists spearheading Ingram Micro's global data science practice. Expert in research, develop and bring-to-market advanced machine learning solutions to address business problems with immediate and long-term impact.

HIGHLIGHTS



Taught data science training courses in Python and R for Environment Canada and Agriculture and Agri-Food Canada



Maintainer and contributor of many open source Python and R packages



Co-founder of an investment management startup in 2015

SKILLS

Python	R	SQL	C++	Linux
MongoDB	AWS	Azure	Spark	
Elasticsearch	Kubernetes	Docker		
VBA				

EDUCATION

MSc Statistics

Brock University

📅 2009 - 2011

GPA
3.88 / 4

BA Economics

Beijing Forestry University

📅 2005 - 2009

GPA
3.7 / 4.0