

# Jason Chan Jin An

DATA SCIENTIST · MACHINE LEARNING ENTHUSIAST · CERTIFIED ALTERYX CORE DESIGNER · GUITARIST · MIXED MARTIAL ARTS FAN

Tel: (852) 68094173 | Email: [jasonchanhku@gmail.com](mailto:jasonchanhku@gmail.com) | Website: [Predicting the Unpredictable](http://Predicting the Unpredictable) | GitHub: [jasonchanhku](https://github.com/jasonchanhku)

## Education

### The University of Hong Kong (HKU)

Hong Kong

B.SC IN QUANTITATIVE FINANCE, GRADUATION DATE: AUG 2016, GPA: 3.40

Sep 2012 – Aug 2016

- Notable Coursework: Financial Engineering, Quantitative Risk Management, Mathematical Finance, Multivariable Calculus, Linear Algebra, Probability Modelling, Financial Modelling, Computer Programming
- Programming Languages: R, Python, SQL, UNIX

## Relevant Work Experience

### Fidelity International

Hong Kong

DATA SCIENTIST

Dec 2016 - Present

- Performed survival analysis and cox regression on Mandatory Provident Fund (MPF) data to study survival rates of different groups of customers and variables impacting them.
- Conducted segmentation using machine learning on MPF customers' data for marketing team to conduct targeted email campaigns and improve open rates.
- Developed an early warning indicator email trigger using daily historical fund redemptions data to predict redemption anomalies using Extreme Studentized Deviate (ESD) test.
- Integrated Tableau and R to create a dynamic machine learning dashboard that predicts incremental net sales of funds from marketing campaigns.

### FDM Academy

Hong Kong

DATA ANALYST TRAINEE

Aug 2016 - Dec 2016

- Undergone 4 months of technical training in the following modules: SQL, UNIX, Python, Tableau.
- Wrote a data preprocessing pipeline of trainee placement data using a Python script.
- Performed exploratory data analysis (EDA) using Tableau and built dashboards to provide insights.
- Awarded the FDM Star Award for outstanding performance.

### GROUPON

Malaysia

DATA ANALYST INTERN

Jun 2015 - Aug 2015

- Analyzed web analytics data from Google Analytics such as sessions, views, and bounce rates to report daily findings to city planners and managers.
- Data blending of data from Google Analytics and Salesforce data to engineer new features such as coupon conversion rate.
- Helped a city planner construct a sales team tracking dashboard using Excel VBA and useful sales metrics.

## Projects ([View website](#) for full list)

### Mixed Martial Arts Fight Predictor

WRITTEN IN PYTHON ([View Web App](#))

Jul 2017 - Dec 2017

- Trained a machine learning model to output the win probability of two fighters for an upcoming fight and built a web app using Python Dash and deployed on Heroku.
- Built scheduled web crawlers to scrape and cleanse fighters' and fights data from Fightmetric and push data to Amazon S3 Bucket.
- Performed exploratory data analysis (EDA) and statistical analysis before applying feature and model selection to predict winners of fights with 70.4% accuracy (benchmark 60%).
- Featured on [Data Science Hong Kong](#) after hackathon pitch.

### Financial Engineering Capstone Project

WRITTEN IN R ([GitHub Repository](#))

Feb 2016 - Jun 2017

- Designed and structured a complex Principle Guarantee Note betting on drop in oil price and MYR/USD using vanilla and lookback options.
- Implemented the Heston Model to model underlying FX rate and volatility to overcome fat tail distribution.
- Wrote visualization, Monte Carlo option pricing, optimization codes and evaluated product performance.
- Project was graded A after presentation of project.

## Technical Skills

---

### Data Preparation and Preprocessing

- R libraries used: data.table, reshape2, dplyr, lubridate, stringr, MICE, ggplot
- Python libraries used: pandas, numpy, matplotlib, seaborn, plotly

### Machine Learning

- R libraries used: H2O.ai, rpart, RandomForest, caretEnsemble, e1071
- Python libraries used: scikitlearn, tensorflow, keras

### Web Scraping

- R libraries used: rvest
- Python libraries used: scrapy, selenium

### Alteryx

- Build efficient large-scale workflows to reduce time spent on data cleansing.
- Use of advanced batch macro to efficiently iterate R and Python Scripts across grouped datasets.

### Tableau and Tableau Server

- Build advanced dashboards using parameter control, calculated fields, and level of detail expressions.
- Integrate R and Tableau to produce dashboards with dynamic machine learning visualization.

### Big Data and Cloud

- Familiar with Google Cloud solutions such as Google Compute Engine and Google Storage
- Creating an instance on Google Compute Engine to perform data cleansing, analysis, and training ensemble models on large datasets

## Communication

---

### Website

#### [PREDICTING THE UNPREDICTABLE](#)

- Used Jekyll Now to build my website hosted on GitHub where I showcase my my projects and write about new ideas as potential projects.

### Alteryx Q12018 User Group Meeting

GUEST SPEAKER

Mar 2018

- Presented my best use cases of using Alteryx at work
- Shared some tips and tricks to building more efficient Alteryx workflows

### Data Science Hong Kong Hackathon

PROJECT PITCHER

Dec 2017

- Presented my UFC MMA Predictor Web App to the Data Science Hong Kong community
- Collaborated with hackathon participants to resolve outstanding issues in projects

## Certificates

---

### Deeplearning.ai

CERTIFICATE OF COMPLETION

Feb 2018

- Neural networks and deep learning
- Improving deep neural networks
- Structured machine learning projects
- Convolutional neural networks

### Alteryx

CERTIFIED CORE DESIGNER

Jan 2018

- Demonstrated proficiency with the fundamentals of using Alteryx Designer
- Tested concepts: Data Types, Preparation, Blending, Descriptive Analysis, Basic Macros & Analytic Apps