Yihui Fan
Data Scientist | Data Product & Insights

Postgraduate training in both machine learning and social science, together with consulting experience in public and private sectors equipped me with a unique skill set for solving business problems, generating data-driven insights, and developing data products.

pressto.fanyihui@gmail.com ✉

+44(0)7738 630073 📱

London, United Kingdom 📍

databentobox.com 🌐

linkedin.com/in/yihui-fan-datascience in

WORK EXPERIENCE

Senior Data Scientist
Decision Technology

10/2014 – Present   *London*

Consulting company specialising in experimental psychology, cognitive science and statistics

Achievements/Tasks

- Build statistical models from clients' big data to understand customers' behaviours, identify valuable trends and opportunities, and assist product developments.
- Design, execute and evaluate experiments to optimise products, improve customer journey and identify levers to help move key metrics.
- Develop data products such as predictive models, automated pricing system, commercial dashboards and visualisation services to facilitate data-driven decision making.
- Experienced in translating data to recommendations and communicating results to (non)technical stakeholders across different functions.

EDUCATION

MSc Advanced Computing Technologies (Machine Learning)
Birkbeck, University of London

09/2016 – 09/2018   *Distinction*

Courses

- Big Data Analytics using R
- Data Warehousing and Data Mining
- Cloud Computing
- Information Security
- Machine Learning
- Advanced Database Management

MSc Sociology
University of Oxford

09/2013 – 09/2014   *Distinction*

Courses

- Statistics
- Social Network Analysis
- Quantitative Method and Survey Research
- Experimental Social Science
- Research Design
- Qualitative Research Methods
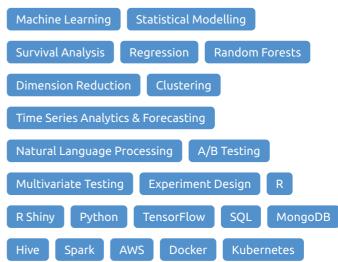
BA Linguistics and Sociology
University of Manchester

09/2011 – 09/2013   *First*

SKILLS

Machine Learning | Statistical Modelling

Survival Analysis | Regression | Random Forests

Dimension Reduction | Clustering

Time Series Analytics & Forecasting

Natural Language Processing | A/B Testing

Multivariate Testing | Experiment Design | R

R Shiny | Python | TensorFlow | SQL | MongoDB

Hive | Spark | AWS | Docker | Kubernetes

PROJECTS

Data Science blog: databentobox.com (10/2016 – Present) ↗

- databentobox.com is a blog where I showcase my personal data science projects and share thoughts on statistics and programming.
- Blog articles are frequently re-posted by R-bloggers and KDnuggets and have gained traffic in the data science community.
- Specifically, the recent project Simulating Coronavirus Outbreak with Origin-Destination Matrix and SEIR Model attracted 20k views.

MSc Machine Learning Project: Tackling Toxic Comments Using Convolution and Recurrent Neural Networks (06/2018 – 09/2018)

- Going beyond standard sentiment analysis, the goal of this project is to build an automated tool to identify online toxic comments.
- Neural network based classification models (bidirectional GRU-CNN) with pre-trained word embedding were built.
- The final model achieved state-of-the-art classification results and the MSc project was awarded distinction.

MSc Sociology Thesis: Gender Equity and Women's Fertility Intention in European Countries (06/2014 – 09/2014)

- This thesis provides an empirical test of gendered fertility theory by examining whether low gender equity at home lowers women's fertility intentions in 31 European countries.
- Using a series of multilevel logistic regressions, together with the data from the second and the fifth round of European Social Survey, the MSc thesis generated unique insights and was awarded distinction.

LANGUAGES

English
Native or Bilingual Proficiency

Mandarin Chinese
Native or Bilingual Proficiency