# Mingpeiyu ZHANG

Gongshu District, Hangzhou, Zhejiang Province, China

http://zhangmpy.top/ zmpyzmpy@gmail.com +86-13175033715

## SUMMARY

An aspiring data scientist & technologist with a background in pure maths & computer science. I'm driven to apply machine learning techniques to solve complex business problems.

## EDUCATION

**Eindhoven University of Technology** **Eindhoven, the Netherlands**

*Master of Science* *Sep 2016 - Mar 2018*

- Major: Data Science in Engineering (graduated 6 months earlier)
- Honors: **cum laude**
- GPA: 8.42/10.0 (top **5%**)
- Thesis: Concept Drift Investigation (grade: **9**/10)

**Beihang University** **Beijing, China**

*Bachelor of Engineering, Bachelor of Science* *Sep 2012 - Jun 2016*

- Major: Computer Science and Technology (top **3** in China)
- GPA: 3.22/4.0 (top 30%)
- Double Major: Applied Mathematics

## WORK EXPERIENCE

**Guandata** **Hanghzou, China**

*Algorithm Engineer, Project team leader* *May 2018 - Present*

- Participated in Unilever demand forecast project, from POC to project delivery.
- Became project leader since 2019, responsible for teams resources management and coordination.
- Cleaned data based on business background, predict the next 3 months' sales with ML models (GBDT, RNN, Prophet).
- Responsible for the sales forecast of 3 categories (4 in total) of Unilever mainland, and the entire project of Unilever Taiwan. Managed to achieve the accuracy KPI (10% higher than manually forecast) from Unilever.
- The main innovations: used custom loss function and a special multi-model blending strategy to make up for the defect of the tree model prediction value.

**Rabobank** **Utrecht, the Netherlands**

*Data Scientist Intern* *Jul 2017 – May 2018*

- Learned H2O, implemented a paper's algorithm, and did several experiments for fraud detection.
  1. Used PAUC to evaluate models to ensure a higher TPR with a fixed FPR.
  2. Used cluster-based methods, trained different models on different clusters.
- Worked on my master thesis from 2017-09, proposed a new concept drift detection algorithm (DDEAL) in the context of fraud detection and phishing mail detection
  1. These two contexts both can not acquire labels in real time, the algorithm proposed to use a feature as proxy label.
  2. The algorithm used an ensemble of multiple algorithms, compared with other state of art algorithms by experiments.
  3. Developed a visutalization application for domain experts to investigate the drift detection results.

**thisworld.cn photography forum (startup project)** **Eindhoven, the Netherlands**

*Algorithm partner* *Jul 2017 - Feb 2018*

- Searched and crawled photographers' blogs and photography articles.
- Built a photography vocabulary list (in Chinese) based on the crawled articles, modelled with Word2vec to implement the recommendation of words in the field of photography.
- Built a recommendation system on articles and photographers based on Alibaba Cloud services.

## PROJECTS

**Kaggle Competitions Expert(current ranking top 2%)** *2019-03 - Present*

1. Santander Customer Transaction Prediction top 6%
   - Models: LGBM and CNN.
   - Used data augmentation to enhance training set. Used Focal loss to solve imbalanced problem.
2. Gendered Pronoun Resolution top 12%
   - Model: Bert + dense layer.

**Truth finding and Rumor Analytics for iPhone 8** *Sep 2016 – Nov 2016*

- Worked on the project with other 4 teammates, responsible for Twitter crawler and tweet classification.
- Developed a crawler based on Python Scrapy and Twitter API, labeled the dataset manually.
- Responsible for data cleaning, word segmentation, used Word2vec on the dataset, text classification based on ensemble of models (SVM, KNN, NB, LR).
- Transfer learning between iPhone 7 training set and iPhone 8 testing set.
- Managed to predict that the next generation iPhone will remove the home button.

## SKILLS

| | |
|---|---|
| **Program languages** | Python, R, Java, C#, HTML, C++, C, PHP, JavaScript, Matlab, SAS, SQL |
| **Professional skills** | Multi-tasking, Quick Learner, Data analytics |
| **Language skills** | Fluent in English and elementary in Dutch |