OLIVIA SHI

Homepage: http://olivia-shi.github.io
Mobile: +86 184-0288-1015 ⋄ Email: Olivia.C.Shi@gmail.com

RESEARCH INTERESTS

- ▷ Computational Social Science, Data Science, Machine Learning and related area.
- ▷ Natural Language Processing for social science and social network.

EDUCATION

Southwest Jiaotong University, Chengdu, China Sep. 2015 - Present
Computer Science and Technology, Senior, Major GPA: 85.34. Sep. 2016 - Present
Pharmaceutics Engineering. (change major) Sep. 2015 - July. 2016
Courses: Data Mining, Machine Learning, Internet Searching Engine, Computer Vision.
Books: *Networks, Crowds, and Markets*, *Introduction to Statistical Learning R*,*Speech and Language processing, Machine Learning.*

PROFESSIONAL EXPERIENCE

Kaggle Featured - Home Credit Default Risk (project webpage) July 2018 - Aug. 2018
Brown Medel: Top 8%(516/7198), Team Work, Group Leader

- ▷ Performed data exploratory analysis to get a deep understanding of the big data.
- ▷ Utilized feature engineering and domain knowledge to add more than 1000 time-series related features to boost cross-validation AUC score, and it turned out those features are **'magic'** ones.
- ▷ Implemented and evaluated different models: Lightgbm, Xgboost, Catboost, Neural Network to accurately predict the capability of an applicant of repaying a loan.
- ▷ Performed Bayesian Optimization to find the optimal combination of hyper-parameters, and developed more advanced ensemble and stacking strategies to reduce variance and enhance AUC score.
- ▷ Achieved a bronze medel with ranking **top 8%** by hard-working, effective communication and great leadership within nearly 2 month.

Internet Searching Engine project (project webpage) Apr. 2018
Supervisor: Prof. Wu, Natural Language Processing, Machine Learning

- ▷ Programmed web spider to filter and save 1000 websites as text,including English and Chinese.
- ▷ Accomplished words cutting and Stemming using NLTK and Jieba.
- ▷ Accomplished computing similarity between every two documents with Cosine Distance.
- ▷ Performed a k-means model to cluster 500 English websites to 16 classes and display represented documents in the biggest class.

Overlapping Social Network Detection Algorithms (webpage) May 2017 - Mar. 2018
Supervisor: Prof. Chen, Graph Data Mining

- ▷ Reproduced Clique Percolation Method and Weak Clique Percolation Method algorithms using java and python.
- ▷ Visualized and analyzed the community structure with Gephi software, feeding data from Stanford Large Network Dataset.
- ▷ Evaluated football dataset with high accuracy.
- ▷ Paralleled the algorithms using scala.

HONORS AND AWARDS

Students Scholarship for five times.	2016 - 2018
Excellent volunteer in Life Mystery Museum at Chengdu city.	Mar. 2016
Third prize, China Undergraduate Mathematical Contest in Modeling.	May. 2016

SKILLS

Programming Languages	Python, R, Java, C, C++, Scala.
Python Packages	Pandas, Numpy, Matplotlib, Scipy, Sklearn, Beautifulsoup, Seaborn NLTK, Tensorflow, Keras.
Software & Tools	Jupyter notebook, Gephi, LaTeX, Spark.

EXTRACURRICULAR ACTIVITIES

- A member of debating team in school of Life Science.
- Volunteer in Life Mystery Museum at Chengdu city.
- A member of Mathematical Modeling Institute in Southwest Jiaotong University.