

Dylana Wang

Portfolio: <https://wangyangtot.github.io>

LinkedIn: <https://www.linkedin.com/in/dylana-wang-971982163/>

GitHub: <https://github.com/wangyangtot>

Email : wangyangtot@gmail.com

Mobile : +1-951-801-1451

Location : Seattle, WA

EDUCATION

- **University of California, Riverside** Riverside, CA
Ph.D Candidate, Computational Chemistry *Sep.2015 – Nov.2019(Expected)*
- **University of Science of Technology of China** Hefei, China
Bachelor of Science, Chemistry *Sep.2011 – July. 2015*

SKILLS AND TOOLS

- **Languages:** Python(Numpy, Pandas, Scikit-Learn, Matplotlib), C++/C, SQL, Java, Javascript, HTML, CSS
- Technologies:** Hadoop, Spark, Keras, Flask, ElasticSearch, Redis, D3, JQuery, Ajax, Bootstrap, MySQL, AWS, Git
- Data Science:** Statistics and Probability, Calculus, Linear Algebra, Differential Equation, Equations of Mathematical Physics, Machine Learning(SVM, Neural Networks, Continuous Latent Variable Models, Decision Tree, K-Nearest Neighbors, Kernel Methods, Ensemble Methods, Expectation-Maximization, PCA, Matrix Factorization, CNN, RNN, Sentiment Analysis)

EXPERIENCE

- **Insight Data Engineer Fellow** <https://github.com/wangyangtot/CustomSearch>
An Interactive Search Engine *March 2019 - April 2019*
 - Build an argument search engine to enable users search viewpoints, arguments or evidences that for or against a given topic to overcome the problem that they have to open and scan all web pages to get viewpoints and arguments surrounding a controversial topic manually.
 - Ingested **20TB+** Web ARChived format data from CommonCrawl, batch cleaned to plain text with **Spark** and **Redis**, then document-indexed into **Elasticsearch** which summed up to **10 millions** web pages.
 - Trained **biLSTM model and biLSTM with attention mechanism and Topic Similarity Features**, achieving **0.815 at ROC-AUC** in **Keras** and implementing it to identify and classify arguments as for or against query topics.
 - Deployed an interactive web-app on AWS using **Python, Flask, HTML and CSS** for interactive searching.

PROJECTS

- **EmotionalArcMovie** <https://emotionalarcmovie.herokuapp.com>
An Hybrid and Interactive Recommendation Engine *June 2018 - Sep 2018*
 - Build an interactive recommendation engine to support discovery of unknown movies with the desired sentiment arcs to go beyond the static ranked list paradigm for movie recommendation
 - Scraped **23576** movie scripts from SpringfieldSpringfield, cleaned data with **NLP(Natural Language Process)** with **Python and R**.
 - Performed **Sentiment Analysis**, clustering on text with **Scikit-learn, TuriCreate and Numpy, pandas and C++**.
 - Applied **t-SNE algorithm** of dimensionality reduction for visualization using **D3**.
 - Integrated the model into **Flask app**, recommending registered user movies based on **matrix factorization method** and unregistered users with content-based filtering method by **HTML, CSS Bootstrap jQuery, D3, Ajax and MySQL**.
- **Home Credit Default Risk** <https://www.kaggle.com/youeer>
Kaggle Competition *Mar 2018 - April 2018*
 - To evaluate bank clients' repayment abilities, used various statistical and machine learning methods to make predictions about the probability of default.
 - Made use of **2.89 GB** data from kaggle which includes a variety of data including telco and transactional information.
 - Conducted feature engineering by applying aggregate functions to create features and **tree-based feature selection** to reduce the number of features from 1000 to 400 with **Pandas, Numpy, Matplotlib and Seaborn**.
 - Implemented **logistic regression, XGBoost, CatBoost, LightGBM with Stratified KFold methods** as base models and evaluated model performance with **operating characteristic curve(ROC)** with **Scikit-learn, Pandas, Numpy**.
 - Made second-level stackers from base models and submitted one with most ROC(0.793), Ranking **top 19%** in the competition.