Dylana Wang

Portfolio: https://wangyangtot.github.io

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

GitHub: https://github.com/wangyangtot

EDUCATION

University of California, Riverside

Ph.D Candidate, Theoretical Chemistry; GPA: 3.70

University of Science of Technology of China

Bachelor of Chemistry; GPA:3.68 (major 3.85)

Sep. 2015 - Present

Hefei, China

Sep.2011 - July. 2015

SKILLS AND TOOLS

• Languages: Proficient: Python, Familiar: C++/C, Java, SQL, Exposed: R, Javascript, HTML/CSS, Matlab

Technologies: Hadoop, Spark, Flask, Bootstrap, D3, JQuery, Ajax, mySQL, AWS, Git

Data Science: Mathematics, Data science, Statistics and probability, Machine learning, Deep

learning, Text-mining,

Projects

 ${\bf Emotional Arc Movie}$

https://emotionalarcmovie.com

May 2018 - Sep 2018

An Interactive Recommendation Engine

- Build an interactive movie recommendation system that supports discovery of unknown movies with the desired sentiment arc to go beyond the static ranked list paradigm.
- Developed a web app as final product, recommending registered user movies based on matrix factorization method and unregistered users with content-based filtering method.
- Scraped movie scripts from SpringfieldSpringfield, Used Lexicon-Based Methods of Sentiment Analysis to analyze 9000 movie scripts and Applied t-SNE algorithm of dimensionality reduction for visualization.
- Technologies used:Flask,Scikit-learn,mySQL,HTML,Boostrap,D3,TuriCreate,Pandas,Numpy,jQuery,

Home Credit Default Risk

Kaggle Compitition

https://www.kaggle.com Mar 2018 - June 2018

- o Made use of 2.89 GB data from kaggle which includes a variety of dataincluding telco and transactional information to predict bank clients repayment abilities.
- o Conducted feature engineering by applying min, max, mean, sum and var functions to create features and used feature selection by tree-based feature selection, reducing the number of features to 400.
- Implemented logistic regression, XGBoost, CatBoost, LightGBM with Stratified KFold methods as base models and evaluated model performance with operating characteristic curve (ROC).
- Made second-level stackers from base models and submit the one with most Roc of 0.793, Ranking top %19 in the competition.
- \circ Technologiesused: Scikit-learn, Pandas, Numpy, Matplotlib, Seaborn

House Price Prediction

Kaggle Data Source

Nov 2017 - Feb 2018

- Developed and implemented Repeat Sale Method to calculate the Real Estate Price Index and Used GARCH model of Time Series to predict the Real Estate Price Index.
- Preprocessed data set by data cleaning, categorical feature transformation, normalization and feature selection. etc.
- Used ensemble methods including Random Forest and Gradient Boosting and evaluated model performance via Cross-Validation(K-fold) technique.
- Technologiesused:Scikit-learn,Pandas,Numpy,Matplotlib,Seaborn

Riverside, GA

Email: wangyangtot@gmail.com

Mobile: +1-951-801-1451