

## Summary

I have experience in all process of data science, i.e., crawling, database, exploratory data analysis(EDA), modeling and data visualization. I'm interested in position as data scientist or data analyst.

## Technical Skills

*Languages:* JavaScript, Python, HTML5, CSS3, R, shell

*Frameworks/Libraries:* jQuery, D3.js, Django, Pyramid, Mako, sklearn, pandas, matplotlib, seaborn, ggplot2, networkx, beautifulsoup.

*Database:* mysql, mongoDB, neo4j.

*Other:* Git/GitHub, Bitbucket, Jira, Markdown, Letex.

## Project Experience

### Molecular Breeding Knowledge - [mbkbase.org](http://mbkbase.org)

Dec. 2014-present

- Built a website for genotype, phenotype and germplasm search and data display of rice. Pramid is used as framework.
- Designed complex search function as data exploring tools.
- Optimized function and MongoDB query that can speed up some page load from 1min to 3-4s.
- Designed many interactive data visualization pages to enhance user's exploring of data.

### Prediction of Biological Activity of Acetylcholinesterase inhibitors - [ncbi.nlm.nih.gov/pubmed/22460031](http://ncbi.nlm.nih.gov/pubmed/22460031)

May. 2009-May. 2011

- Collected a dataset containing 714 inhibitors.
- Built MLR and SVM models to predict the bio-activity of Acetylcholinesterase inhibitors.

### Classification of Acetylcholinesterase inhibitors and decoys - [ncbi.nlm.nih.gov/pubmed/22263859](http://ncbi.nlm.nih.gov/pubmed/22263859)

May. 2009-May. 2011

- Collected a dataset containing 714 inhibitors and 3892 decoys.
- Designed different pipeline for feature selection.
- Built several SVM classification models and the accuracy is up to 96.5%.

## Work Experience

### Institute of Genetics and Developmental Biology Chinese Academy of Sciences - Assistant Engineer

Beijing, China | Jul. 2012-present

- Performed genome and transcriptome assembly.
- Developed database and website development & data visualization.

### BioDuro Co., Ltd - Intern

Beijing, China | Jun. 2012-Jan. 2014

- Developed interface for molecular properties calculation, structure search and library enumeration based on Rgroup and reaction
- Assisted the update of compound database maintained in MDL® ISIS/Base

## Education

Udacity - Data Analyst Nanodegree.

Jul. 2016

MongoDB, Inc - MongoDB for python Developers. (Training)

Feb. 2015

Beijing University of Chemical Technology - Chemical Engineering and Technology, M.S.

Jul. 2012

Beijing University of Chemical Technology - Bioengineering, B.S.

Jul.2009

## Publication

- Dejun Li, Rizhong Zeng, Yan Li, Manman Zhao, Jinquan Chao, Yu Li, **Kai Wang**, Lihuang Zhu, Wei-Min Tian, Chengzhi Liang, Gene expression analysis and SNP/InDel discovery to investigate yield heterosis of two rubber tree F1 hybrids, **Scientific Reports**, 2016. 6:24984.
- Kai Wang**, Xiaoying Hu, Zhi Wang, Aixia Yan\*, Classification of Acetylcholinesterase Inhibitors and Decoys by a Support Vector Machine, **Combinatorial Chemistry & High Throughput Screening**, 2012. 15(6), 492-502.
- Aixia Yan\*, **Kai Wang**, Quantitative structure and bioactivity relationship study on human acetylcholinesterase inhibitors, **Bioorganic & Medicinal Chemistry Letters**, 2012, 22(9), 3336-3342.
- Maolin Wang, **Kai Wang**, Aixia Yan\*, Changyuan Yu\*. Classification of HCV NS5B polymerase inhibitors using Support Vector Machine. **Int.J.Mol.Sci**, 2012, 13(4), 4033-4047.
- Aixia Yan\*, Xianglei Nie, **Kai Wang**, Maolin Wang, Classification of Aurora Kinase inhibitors by Self-Organizing Map (SOM) and Support Vector Machine (SVM), **Eur. J. Med. Chem**, 2012(61), 73-83.
- Aixia Yan\*, Yang Chong, Liyu Wang, Xiaoying Hu, **Kai Wang**, Prediction of Biological Activity of Aurora-A Kinase Inhibitors by Multilinear Regression Analysis and Support Vector Machine, **Bioorganic & Medicinal Chemistry Letters**, 2011,21(8), 2238-43.
- Kai Wang**, Aixia Yan\*, Introduction of several public databases for chemical activity on the Internet, **Computers and Applied Chemistry**, 2010,21(12): 1629-1632.