Resume

Name	Wang Siyu	
Birthdate	June 3th, 1994	Male
E-MAIL	siyunb323@yahoo.co.jp	
Adress Building 13, Central University of Finance and Economics, Changping District, Beijing, China		**Postal number** 102206
State of health	Good	**Marital status** Single

Time	Education	Major	Main Courses	Grade (rank)
2013.09—2017.06	Central South University (undergraduate)	Statistics	**Required Statistics Courses：** Mathematical Analysis, Higher Algebra, Probability Theory and Mathematical Statistics, Regression Analysis, Time Series, Multivariate Statistical Analysis, Sampling Technology, etc. **Elective Mathematics Courses：** Numerical Calculation, Operations Research, Optimization Principles and Algorithms, Applied Stochastic Processes, Complex Functions, Mathematical Modeling, Analytic Geometry, etc. **Computer Courses：** C++ Language Programming, Data Structure, Database Principles (SQL, VBA), SPSS, R Language, Data Mining (Python), etc.	**90.35 (4/90)**
2017.09—2019.06	Central University of Finance and Economics (postgraduate)	Applied Statistics (big data oriented)	Computer Foundation of Big Data(LINUX\SQL\Python\R) Distributed Computing of Big Data （MapReduce、Hive、Hadoop、Spark） Statistical Basis of Big Data (Exploratory Analysis, Visualization, Data Preprocessing, etc.) Big-Data Mining and Machine Learning (Ensemble Learning, Cluster Analysis, Support Vector Machines, Naive Bayesian, Neural Networks, Recommendation Systems, etc.) Unstructured Data Analysis (Complex Networks, Natural Language Processing, Image Recognition) Statistical modeling of big data (Lasso, Ridge Regression, Elastic Network, Minimum Angle Regression, Correspondence Analysis, Time Series Analysis, etc.)	**92.20 (1/54)**

Main skills	**Foreign Languages**	English	College English Test- CET 6
			TOFEL 90
		Japanese	JLPT-N2 135， JLPT-N1 81
	Skills	Programing language	R，Python，SQL Server，SPSS，C++，Lingo，MATLAB
		Distributed processing	Hive，Hadoop，Spark，Linux
		MS Office	Word，Excel，PPT，VBA ，PowerBI

Time	Competition of scientific research
2015.06	The third prize of Mathematical Modeling Contest in Central South University
2015.09	The third prize of China Undergraduate Mathematical Modeling Contest
2016.01	The Honorable Mention Award of American Undergraduate Mathematical Modeling Contest
2018.02	The third prize of "Orient Futures Cup" China Postgraduate Statistical Modeling Contest
2018.03	The first prize of Statistical Modeling Contest in Central University of Finance and Economics
2018.06	The third prize of "Fengyun Cup" Machine-Learning Modeling Contest
2018.09	The third prize of Case Contest of National Professional Postgraduate in Applied Statistics

Time	Personal honors
2014.09	Outstanding student of Central South University in 2013
2016.09	Outstanding student of Central South University in 2015
2017.06	Outstanding graduate of Central South University in 2017
2018.10	National scholarship for postgraduate students in 2018
2019.06	Excellent Graduates of Beijing in 2019

Time	Project name	Project experience
2015.06—2016.03	Innovation and Entrepreneurship Competition of Central South University	As the team leader, I successfully applied for the project of "Gold Futures Price", which was successfully completed after the processes such as social research, data collection and establishment of vector autoregressive model. Meanwhile, based on the project, we wrote the report entitled *Empirical Analysis of the Linkage Effect between Gold Futures Price and Actuals Price*.
2017.01—2017.05	Metabolic analysis based on fingerprint of Angelica sinensis	The bio-fingerprint information of Angelica sinensis extracts in different years was determined by means of NMR, GC-MS and infrared spectroscopy, the effective bands of which were estimated by orthogonal partial least squares (OPLS). The medicinal ingredients of Angelica sinensis were identified according to the corresponding chemical characteristics of the bands. The applied research

		can guide the picking period of Angelica sinensis and the extraction of effective medicinal ingredients. The paper entitled *Metabolic Analysis Based on Fingerprint of Angelica Sinensis*, was completed in this process, which stood out among 350 papers of the whole college and was included in the Excellent Graduation Thesis Collection of 2017 Undergraduates of Central South University.
2017.12—2018.01	Classification of song styles based on spark	The crawler of Python was used to crawl the lyric information of two kinds of music styles (popular and classical) on the Internet music platform, of which the sample lyrics were used to construct corpus and were transformed into Lyric vectors with latent semantics by LSI model. Firstly, Lyric vectors with latent semantics was as input and the unknown-style songs were classified by the neural network built through Keras of python. Then, the music styles for large quantities of Lyric texts were classified by using machine learning methods such as Naive Bayesian, Support Vector Machine and Random Forest in Mllib of Spark, and the cross-validation accuracy of model was more than 95%.
2018.04—2018.06	Personal credit evaluation model based on machine learning	Based on the credit history and online transaction data of 21246 people, this paper firstly filtered and constructed the characteristics by various methods. Then taking the prediction effect of logistic model as the benchmark, the xgboost model of credit audit for lenders with high prediction accuracy was established. The accuracy of prediction in the test set can reach about 93.61%, while the recall rate is close to 90%.
2018.01-now	CluBear	CluBear（http://www.xiong99.com.cn/）, an online education platform of data science, is founded by Wang Hansheng, a well-known professor of Statistics at Peking University in China. During postgraduate student period, I published two popular science cases on this platform, one of which entitled *Talking about the Characteristics of Data Network Based on Erdos Number* is about the complex network, other of which named *Reading Big-Data Books in the Text Analysis Perspective* is about the natural language processing. Both are highly praised by teachers and students.
2018.04—2018.06	Community detection of paper cooperation based on text LDA	Through python's selenium module, I dynamically crawl the paper data about statistics and probability from apps. webofknowledge. Firstly, taking the authors as nodes, the cooperative authors were connected by edges, to construct the paper cooperation network. LDA topic model was used to extract abstract information as additional information of the nodes. Secondly, community detection with covariate was carried out. On the one hand, small marginal communities were detected by searching for connected components. On the other hand, the node similarity was constructed based on topological structure and LDA topic information, which was used to improve Label Propagation, so as to carry out community detection for large connectivity components in the paper cooperation network. Thus, the community structure of the whole network can be revealed through the above two aspects.

2019.04-now	Community detection of large paper networks	At present, I am working in the thesis project with my master supervisor. About 150,000 papers of 133 famous journals on statistics and probability from 1998 to now have been crawled from *web of science*. The four networks, the author cooperation network, the paper citation network, the keyword topic network and the magazine citation network, were constructed according to the year, then based on the dynamic evolution of which, the trend of this complex system in the field of Statistics and the trend of cooperative groups were researched at the macro level, and the growth processes of scientific research of several experts in the field of Statistics were researched at the micro level.