Xishi Chen

203 W 108th St, New York, NY10025 | 929-319-8653 | xc2455@columbia.edu | https://www.linkedin.com/in/xishi-chen

EDUCATION

Columbia University, Graduate School of Arts and Sciences   New York, NY
Master of Arts in Statistics (GPA 3.9/4.0)   Expected Dec. 2019
Relevant Courses: Linear Regression Model, Machine Learning, Data Visualization, Applied Data Science, Deep Learning

Wuhan University   Wuhan, China
Bachelor of Science in Statistics (GPA 3.7/4.0)   Jun. 2018
Relevant Courses: Data Structure and Algorithm, Time Series

SKILLS

Advanced: Python, R, Excel, SQL, Tableau, Spark, JavaScript, HTML, Machine learning, Time series, Hadoop
Intermediate: MongoDB, AWS, NLP

EXPERIENCE

Data Analyst, 72 DRAGONS, New York, NY   Mar. 2019 – Oct. 2019
- Extracted data from film industry database and analyzed how factors such as directors, distributors and budgets affected box office revenues, using Random Forest with Python.
- Visualized the result with Tableau and Rshiny to demonstrate to filmmakers the contribution of the each of factors.
- Led the team of data analysts, assigned and summarized results of the team as well as reported to my supervisor

Data Scientist, Avlino Inc., Holmdel, NJ   May 2019 – Aug. 2019
- Applied n-step SARSA Reinforcement Learning Model in Python to schedule the movement of cranes at a port in order to maximize shipping container throughput.
- Used the model's result to develop a strategy of putting containers with similar dwell times together to reduce the cranes' travel time by 30%.

Data Engineer, Shubo Technology, Wuhan, China   Mar. 2018 - Jun. 2018
- Scrapped thousands of pages to collect textual data about corporations' histories of fraud from public websites using Scrapy, extracted dates and details from texts with regular expression in order to develop databases that can be sold to credit check companies
- Transformed data into a unified format with Python so that it could be easily used by clients

Futures Researcher, Zhonghui Futures, Shanghai, China   Jun. 2017 - Sept. 2017
- Created code in VBA to assist employees in completing daily analysis of market data, saving 30+ mins per day
- Applied Markov AR model to predict prices for gold, stocks and bonds to help clients make investment decision

Researcher, Big Data and Cloud Computing Laboratory at Wuhan University, Wuhan, China   Mar. 2017 - Mar. 2018
- Applied ARIMA with R to predict volume of information on websites to decide bandwidth allocation in web scraping

PROJECTS

Recommendation System, Columbia University, New York, NY   Apr. 2019
- Collaborated with team to analyze 20 gigabyte films rating data with PySpark on AWS, apply Alternating Least Squares to make film recommendations to users and achieve Root Mean Square Error at 0.8188 for test data.

House Price Prediction, Easyrent Corporation, New York, NY   Feb. 2019
- Cleaned data and visualized the relationship between variables in order to choose the right ones for price prediction
- Predicted values of houses with XGBoost and Linear Regression and obtained log error at 0.31 for regression

Russian Inference in US Presidential Election, Columbia University, New York, NY   Jan. 2019
- Cleaned a dataset of 1,000,000 tweets with Python, including tokenization, removing stop words and stemming
- Applied LDA model using Python in order to find which topics were more commonly discussed by partisan groups

Rating on Peer-to-Peer Online Platforms, Wuhan University, Wuhan, China   Sept. 2015 - Sept. 2016
- Analyzed collected data of P2P platforms using R with a PCA model to reduce the dimensions and collinearity of factors, solving a convergence problem with the original logistic regression model
- Applied new logistic regression model to data to propose an evaluation for P2P platforms with 95% accuracy

LEADERSHIP AND VOLUNTEER EXPERIENCE

Vice President, Student Association in School of Statistics, Wuhan University, Wuhan, China   Apr. 2015 - Apr. 2016
- Promoted Math Festival to students through social media and attracted 150+ students