ALEXANDRA (JIAHAO) HUO

jiahao.h@columbia.edu | (424) 302-8221 | Portfolio Website | GitHub | LinkedIn

Data scientist with experience of leveraging predictive modeling, data mining algorithms, and data processing to deliver actionable insights and solve challenging business problems in professional/research settings. Well-versed in communicating out key findings to stakeholders of different technical backgrounds. Proficient knowledge in analytics, statistics, mathematics, and programming.

EDUCATION

Columbia University: Columbia Engineering | Columbia Business School Aug 2018 – Dec 2019
M.S. Business Analytics, GPA 3.84/4.00 New York, NY
- **Teaching Assistant:** Analytics in Python (for two semesters)

University of California, Los Angeles (UCLA) Sep 2014 - Mar 2018
B.S. Statistics & B.S. Financial Actuarial Mathematics, SUMMA CUM LAUDE Los Angeles, CA

WORKING & RESEARCH EXPERIENCE

Activision Blizzard May 2019 – Aug 2019
Data Scientist Intern (Player Science and Economics, Global Analytics) Los Angeles, CA
- Examined regional effects of character usage in Call of Duty (CoD) using PySpark and Presto on Qubole, and communicated out results to designers and partner data science team to increase future titles' gender and racial diversity
- Customized parametric survival models using Python Lifelines, partnered with producers to balance game Battle Royale
- Tracked real-time CoD engagement metrics and impact of new item releases using D3 visualization
- Developed performance metrics of weapons with SQL, and delivered dashboards to automatically monitor in-game ecosystem

OCP Group Nov 2018 – Present
Research Data Scientist (Text Mining and Nature Language Processing) New York, NY
- Parsed named entities from mineral reports with DBpedia Spotlight annotation, domain specific lexicon induction and n-grams
- Trained unsupervised word embeddings using Word2vec with Stanford CoreNLP on Python in parallel on google cloud
- Delivered knowledge graphs of connections between entities to stakeholders to predict changes in fertilizer prices

WiZR (Computer Vision Startup) Feb 2018 – June 2018
Data Scientist Los Angeles, CA
- Programmed AI logic rules of video analytics that boosted detection accuracy by 7.6% and reduced false alarm rate by 5.9%
- Built ML models, and created a pipeline to deliver scheduled analytical reports to clients using SQL, Python, R, and Crontab

Vale S.A. June 2017 – Aug 2017
Assistant Data Analyst Shanghai, CN
- Assessed impact of demographic trends of China on economy and steel consumption using R, Python, and SQL

UCLA Department of Statistics: NBA Players Analytics Mar 2017 – June 2017
Research Assistant (Sports Analytics) Los Angeles, CA
- Devised a ranking algorithm with self-defined KPIs and distance metrics to evaluate players' performance using R
- Visualized sweet spots and unsweet spots for the NBA Finalists with Spatial Heat Maps using Tableau

PROJECT & LEADERSHIP EXPERIENCE

The Data Open: Renewable Energy Feb 2019
Project Manager Princeton, NJ
- Applied Panel Regression, Breusch-Pagan test, the Hausman test, and the Granger Causality test to examine the relationship between non-renewable energy consumption and asthma prevalence with Python StatsModels

Game Analytics: Identify Cheaters in game PlayerUnknown's Battlegrounds (PUBG) Dec 2018
Team Lead New York, NY
- Implemented unsupervised anomaly detection algorithms (One-Class SVM and Isolation Forest) to detect cheaters with Python
- Defined metrics to evaluate performance of the unsupervised models, and calculated proportion of cheating population in game

DataFest 2018: Introduce Revenue Maximization Algorithm for *Indeed.com* Apr 2018
Best Insight Award | *Team Lead* Los Angeles, CA
- Delivered a XGBoost regression model to compute business values of job postings to optimize job posting rank and revenue

Kaggle Competition: Predict Average Response Time of the Los Angeles Fire Department Mar 2017 – June 2017
Ranked 3rd/92 teams | *Team Lead* Los Angeles, CA
- Implemented XGBoost in R, and tuned hyperparameters using random search in parallel with parallelMap

SKILLS & CERTIFICATIONS

- **Languages**: Python, SQL, Scala, R, Tableau, C++
- **Technical Skills:** Git, Apache Spark, Google Cloud, Machine Learning, Data Analytics, Data Processing, Text Mining
- **Society of Actuarial Exams**: P: Probability Theory; FM: Financial Mathematics; MFE: Models for Financial Economics