JINGJING LIN (She/Her)

isjingjing.lin@gmail.com | Arlington, VA 22202 | ⚡ github.com/JJJJJingL/portfolio 🔗 linkedin.com/in/imjingjinglin

EDUCATION

Georgetown University   08/2018 – 05/2020
- Master of Science: Data Science and Analytics, GPA 3.5/4.0   Washington D.C.
- Peer lead mentor, supervised and collaborated with other mentors to support ~70 first-year graduate students

University Of Manchester   09/2015 – 12/2016
- Master of Science: Management and Information Systems, GPA 3.3/4.0   Manchester, UK

Tianjin Polytechnic University   09/2011 – 06/2015
- B.Eng.: Software Engineering, GPA 85/100 and B.Econ.: Finance(2nd), GPA 88/100   Tianjin, China
- 2014 Presidential First-Class (top 3%), and 2013 Second-Class (top 5%) Scholarship; Outstanding Graduate (top 2%)

SKILLSET

Programming: **Python**(scikit-learn, pandas), **R**(dplyr, glmnet), **SQL**, VBA(Macros), JAVA, HTML, CSS, Command-line
Machine Learning: Regression, Decision Tree, Random Forest, Natural Language Processing, Deep Learning, Clustering
Statistics: Probability, Distribution, Sampling, Hypothesis Testing, Bayes Theorem, Correlation, A/B testing
Cloud Computing: AWS (EMR, S3, Hadoop, MapReduce, Spark, Git); Google Cloud (BigQuery, storage buckets)
Visualization & Tools: Tableau, Plotly, Matplotlib and ggplot2; MySQL, Jupyter, Google Analytics (Advanced Certificate)

EXPERIENCE

Georgetown University Graduate School | *Higher Education*   08/2020 – Present
Python Development Engineer (Data Science Researcher)   Washington D.C.
- Creating a website for tracking 100+ coronavirus vaccines development process based on scientific papers and news
- Building a Tableau dashboard to quantify and visualize textual vaccines development information, in order to explain the scientific progress in the fight against COVID-19 to a non-medical audience

Center for Security & Emerging Technology, Georgetown University | *Higher Education*   09/2019 – 12/2019
Data Science Research Assistant   Washington D.C.
- Performed exploratory data analysis (EDA) on 3 academic publication datasets (130+ million rows of 14+ GB data) to characterize tech fields in Artificial Intelligence through BigQuery, storage buckets, and virtual machines in GCP
- Increased matching rates to ~20% in non-matched records from filtered databases by conducting textual analysis (NLP), including converting bags-of-words, vectorizing tf-idf and running text similarity algorithms

Unilever - Dollar Shave Club | *Consumer Goods*   06/2019 – 08/2019
Marketing Technology Intern   Los Angeles, CA
- Sole analyst responsible for analyzing and optimizing the current manual-operated 20+ spreadsheets with 200+ tags of Urchin Tracking Module (UTM) information
- Reduced parameters setting time by 90% for marketing team, and implementing time by 33% for data systems team by establishing automated tags relationships pipeline using VBA, SQL and Macros, and re-designing the dashboard
- Delivered a new UTM solution to teams (~20 senior members) independently, designed a plan for long-term operations
- Investigated and organized existing marketing campaigns and supply chain, to create a business proposal for 'DSC X Military' to build connections with military communities

WST Consulting (Startup) | *Education Consulting*   09/2017 – 04/2018
Research Analyst   Shanghai, China
- Gathered and analyzed data on target firms' finance and development strategy; created periodic reports and guidebooks (4 chaps) on recruitment programs of 4 fields (finance, consulting, data and technology) across global markets
- Led resource management effort to create and restructure marketing materials that yielded 50% increase (from ~3000 to ~4500) in average view count of over 15 supported articles on WeChat platform
- Supervised an intern and 2 junior colleagues on document research methods and writing materials revision

ChinaSoft International Ltd. (Gartner 2019 Top 100 global IT service providers) | *Internet*   Summers, 2012 – 2015
Software Development Engineer Co-op   Tianjin, China
- Designed and built 4 types of systems (management, retail, social and hybrid) with Java, MySQL, HTML and CSS
- Conducted requirement analyses, database and UI design, conceptual prototype building and testing for all projects
- Documented feasibility analysis reports and project development plans; delivered final presentations

PROJECTS (More projects can be found on GitHub)

AWS - Data Engineering: Top Comment Identification in Reddit   04/2019 – 05/2019
- Accessed and loaded datasets of Reddit comments(~500GB) in JSON from S3 and preprocessed data, including cleaning (handling missing and inconsistent values), discretization, transformation and sampling, using PySpark in EMR
- Performed EDA with Spark SQL; created features in numeric (text-length) and categorized (scores) variables
- Conducted features encoding through MLlib; modeled a top comment identifier with logistic regression through machine learning pipeline, achieving the accuracy of 90%

Data Analytics: Where Should You Live for Your Health   09/2018 – 12/2018
- Acquired datasets through API and conducted multiple ML algorithms' analyses (clustering, association rule mining and regression) on several factors to investigate how individual choices in residence may impact their health
- Performed data wrangling and normalization(~20k rows) using pandas and posed hypotheses based on prepared data
- Architected ML models including KNN, Naïve Bayes, SVM to validate hypotheses and predicted the cancer trend