

JINGJING LIN (She/Her)

+1 202.460.4927 | isjingjing.lin@gmail.com | Arlington, VA 22202 | github.com/JJJJingL [LinkedIn-imjingjinglin](#)

EDUCATION

Georgetown University <ul style="list-style-type: none">Master of Science: Data Science and Analytics, GPA 3.5/4.0Peer lead mentor, advised mentors and ~70 graduate students	08/2018 – 05/2020 Washington D.C.
University Of Manchester <ul style="list-style-type: none">Master of Science: Management and Information Systems, GPA 3.3/4.0	09/2015 – 12/2016 Manchester, UK
Tianjin Polytechnic University <ul style="list-style-type: none">Bachelor of Engineering: Software Engineering, GPA 85/100 Minor: Finance, GPA 88/1002014 Presidential First-Class (top 3%), 2013 Second-Class (top 5%) Scholarship; Outstanding Graduate (top 2%)	09/2011 – 06/2015 Tianjin, China

SKILLSET

Programming: Python (Scikit-learn, pandas), R (dplyr, glmnet), SQL, VBA – [Demo](#), Java (JDBC), HTML, CSS, Shell
Machine Learning: Regression (Linear/Logistic, Lasso), Clustering (K-Means, Hierarchical), Classification (Boosting, Random Forest, SVM), Deep Learning (CNN, RNN), Natural Language Processing (NLP), Dimension Reduction (PCA)
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 (Certificated)

EXPERIENCE

Georgetown University Graduate School <i>Higher Education</i> <i>Data Science Researcher</i> <ul style="list-style-type: none">Collected 100+ vaccines development progresses from scientific reports, merged data using pandas and NLTKBuilt a map-based Tableau dashboard to quantify and visualize textual processes with geological informationGenerated weekly reports and cleaned datasets, embedded results in the website using HTML and CSS	08/2020 – Present Washington D.C.
Center for Security & Emerging Technology, Georgetown University <i>Higher Education</i> <i>Data Science Research Assistant</i> <ul style="list-style-type: none">Accessed data from 3 AI publication databases (130+ million rows of 14+ GB) using BigQuery in GCPMerged database A, B by DOIs and metadata (e.g., keywords), built a dictionary from metadata by bag-of-wordsFound the top 50 matches in B for each article in A by matching TF-IDF distance and the year of publication, created the distribution of the longest substring sizes in each match of the top 50 articlesIncreased matching rates to ~20% in articles without DOIs by running text similarity algorithms using the highest similarity score, based on the threshold (95%) from the scores distribution in matched articles with DOIs	09/2019 – 12/2019 Washington D.C.
Unilever - Dollar Shave Club <i>Consumer Goods</i> <i>Marketing Technology Intern</i> <ul style="list-style-type: none">Sole analyst responsible for analyzing and optimizing the current manual-operated 20+ spreadsheets with 200+ tags of Urchin Tracking Module (UTM) information managementReduced parameters setting time by 90% for marketing team, and implementing time by 33% for data systems team by developing an automated tags generator and re-designing the dashboard using macros, pivots and SQLDelivered a new UTM solution to teams of ~20 colleagues independently, designed a long-term operation planInvestigated and organized existing marketing campaigns and supply chain, to create a business proposal for 'DSC X Military' to build connections with military communities	06/2019 – 08/2019 Los Angeles, CA
WST Consulting (Startup) <i>Education Consulting</i> <i>Research Analyst</i> <ul style="list-style-type: none">Gathered information on target firms' finance and development strategy to generate biweekly reports, wrote 4 chapters of guidebook on recruitment programs of finance, consulting, data and tech fields across global marketsLed resource management efforts 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 platformSupervised an intern and 2 junior colleagues on document research methods and writing materials revision	09/2017 – 04/2018 Shanghai, China

PROJECTS (More projects accessible on [GitHub](#))

Data Science: Where Should You Live for Your Health (Demo , Code) <ul style="list-style-type: none">Acquired ~20k rows of data through API after scoping the project and translating it into an analytical problemAssessed data by performing data wrangling and normalization with pandas, posed 3 hypothesesArchitected multiple ML models of clustering, association rule mining and regression on factors (air and water quality) to validate hypotheses, investigated the cancer trend based on results in Tableau and network analysis
AWS - Data Engineering: Top Comment Identification in Reddit (Code) <ul style="list-style-type: none">Accessed and loaded ~500GB datasets of Reddit comments in JSON from S3 and preprocessed data, i.e. cleaning (handling missing and inconsistent values), discretization, transformation and sampling, using PySpark in EMRPerformed EDA with Spark SQL; created features in numeric (text-length) and categorized (scores) variablesConducted features encoding through MLlib; modeled a top comment identifier with logistic regression through machine learning pipeline, achieving the accuracy of 90%