

# JINGJING LIN (She/Her)

+1 202.460.4927 | [isjingjing.lin@gmail.com](mailto:isjingjing.lin@gmail.com) | Arlington, VA 22202 | [github.com/JJJJingL/portfolio](https://github.com/JJJJingL/portfolio) [linkedin.com/in/imjingjinglin](https://www.linkedin.com/in/imjingjinglin)

## EDUCATION

<b>Georgetown University</b>	08/2018 – 05/2020
▪ <i>Master of Science: Data Science and Analytics</i> , GPA 3.5/4.0	Washington D.C.
▪ Peer lead mentor, supervised and collaborated with other mentors to support ~70 first-year graduate students	
<b>University Of Manchester</b>	09/2015 – 12/2016
▪ <i>Master of Science: Management and Information Systems</i> , GPA 3.3/4.0	Manchester, UK
<b>Tianjin Polytechnic University</b>	09/2011 – 06/2015
▪ <i>Bachelor of Engineering: Software Engineering</i> , GPA 85/100   <i>Minor: Finance</i> , 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 (Linear/Logistic), Clustering (K-Means, Hierarchical), Classification (SVM, Random Forest, Decision Tree), 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 (Advanced Certificate)

## EXPERIENCE

<b>Georgetown University Graduate School</b>   <i>Higher Education</i>	08/2020 – Present
<i>Python Development Engineer (Data Science Researcher)</i>	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	
<b>Center for Security &amp; Emerging Technology, Georgetown University</b>   <i>Higher Education</i>	09/2019 – 12/2019
<i>Data Science Research Assistant</i>	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	
<b>Unilever - Dollar Shave Club</b>   <i>Consumer Goods</i>	06/2019 – 08/2019
<i>Marketing Technology Intern</i>	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	
<b>WST Consulting (Startup)</b>   <i>Education Consulting</i>	09/2017 – 04/2018
<i>Research Analyst</i>	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	
<b>ChinaSoft International Ltd.</b> (Gartner 2019 Top 100 global IT service providers)   <i>Internet</i>	Summers, 2012 – 2015
<i>Software Development Engineer Co-op</i>	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 accessible on [GitHub](https://github.com))

<b>AWS - Data Engineering: <a href="#">Top Comment Identification in Reddit</a></b>	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%	
<b>Data Analytics: <a href="#">Where Should You Live for Your Health</a></b>	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	