# Dongcheng (Sarah) Ge

McLean VA | (917) 929-2567 | sarahgedc@gmail.com
www.linkedin.com/in/gdcsarah

## EDUCATION

**Georgia Institute of Technology** | ***MS in Analytics*** | Atlanta, GA **Dec 2022**
Coursework: Machine Learning, Deep Learning, Analytics Modeling, Computing for Data Analytics, Data and Visual Analysis, Simulation
**GPA: 4.0**

**Baruch College** | ***MS in Computer Information Systems*** | New York, NY **May 2017**

**Johns Hopkins University** | ***MS in Marketing*** | Baltimore, MD **Aug 2014**

**Nanjing Normal University** | ***BA in Business English*** | Nanjing, China **July 2011**

## SKILLS & CERTIFICATIONS

**Programming & System:** Python, SQL, R, Scala, JavaScript, CSS, HTML, VBA
**Modeling Techniques:** Regression, Classification, Clustering, Simulation
**Framework & Tools:** Spark, Pandas, NumPy, Scikit-Learn, Polars, dbt, MODE, AWS, Azure, Plotly/Dash, Simio, Tableau, Power BI
**Languages:** English, Chinese Mandarin

## PROJECTS

**Predict likelihood of members having housing insecurity issues** | *Sponsor:* Humana **Fall 2022**

- Analyzing a data set of 900 features, identified Medicare members most likely to be struggling with housing insecurity issues by applying decision tree-based ensemble algorithms, and finetuning with Optuna.
- ***Language:*** *Python;* ***Model:*** *XGBoost, LightGBM;* ***Framework:*** *Optuna*

**Retail Service Support Problem Management** | *Sponsor:* NCR **Spring 2022**

- Cleaned and analyzed over 18 million records of device incident data and service notes.
- Created and compared multiple models to explain the variation in resolution times of different incidents.
- Applied topic modeling techniques to find meaningful incident clusters.
- ***Language:*** *Python, SQLite;* ***Model:*** *Regression, Random Forest, Survival Models, NLP*

**Detection and Clustering of Digital Signal Impairments (Top 5)** | *Sponsor:* Cox Communications **Spring 2022**

- Created a sliding window algorithm using python, combining with HDBSCAN to perform clustering on full band capture data, and to detect different types of wave impairments.
- ***Language:*** *Python;* ***Model:*** *HDBSCAN*

**Stock Behavior Predicting** **Fall 2021**

- Combined the outputs of sentiment models (VADER) with time series data and trained the new dataset with multiple tree-based and regression models, to predict stock opening price trends.
- Visualized predictions and other stock or sentiment related information using Plotly (Dash).
- ***Language:*** *Python;* ***Model:*** *VADER, XGBoost, Exponential Smoothing, Auto Linear Regression, Auto Logistic Regression*

## EXPERIENCE

**Shopify (USA) Inc.** | ***Data Science Intern*** | Atlanta, GA **May 2022 – Aug 2022**

- Designed and created weekly reports using an internally built SQL alerting framework.
- Built and deployed a metric model using internally developed data transformation tools and metric definition framework.
- Created multiple ad-hoc reports for cross-functional teams such as returns, refunds and marketing, using MODE.
- Documented multiple MySQL raw tables using an internally developed Ruby Gem.
- ***Language****: SQL, Python****; Tool:*** *MODE, dbt, and other tools/frameworks that are developed in-house*

**China Overseas America, Inc.** | ***Marketing Manager*** | New York, NY **Feb 2017 – Mar 2021**

- Achieved 100% sales goals and reduced advertising cost to under 50% of the market average.
- Built and optimized pricing models using regression in Excel.
- Automated marketing and sales reports by programming in Excel, which significantly improved team efficiency.
- Led a team of 6, along with external collaborators, to conduct thorough research on New York condo market.
- Improved cross-departmental communications by promoting company-wide learning sessions, and re-engineering cross-functional reporting systems.

**United Nations** | ***Digital Marketing Analyst (Intern)*** | New York, NY **Feb – Aug 2016**

- Monitored and analyzed website/social media performance data to provide insights to supervisor.