

# Shuang Du

984-810-9076 | isreneds@gmail.com | <https://www.linkedin.com/in/shuangdu08/>  
<https://reneedu320.github.io> | Chapel Hill, NC

## EDUCATION

### University of North Carolina at Chapel Hill, United States

May 2023

Master of Science in Information Science

Courses: Text Mining (NLP), Applied Statistics (R), Statistical Computing (C++), Algorithm (Python, Java)

### Sichuan University, China

June 2020

Bachelor of Economics in Finance

Exchange Program-University of New Mexico, United States (2018-2019, GPA 3.96/4)

Courses: Probability Statistics, Econometrics (Time series), Derivatives

## SKILLS

Programming	Python (scikit-learn, Pandas, NumPy, SciPy, seaborn, matplotlib), R (tidyverse, Shiny, ggplot2, rcpp), Hadoop, Spark, SQL, JavaScript, VBA
Tools	Docker, Git, Tableau, Power BI, Microsoft Office, Stata
Concepts	Regression, Clustering, Text Mining, NLP, Classification, KNN, Naïve Bayes, SVM, word2vec, Retrieval system

## PROFESSIONAL EXPERIENCE

### Data Analyst Intern | Bank of Montreal Financial Group

May - Jul 2021

- **Automation:** Independently consolidated 12 months routine data assessment worksheets from other departments, utilized **SQL Server** in assembling complex, multivariate datasets from both structured and unstructured sources, help built data pipeline to facilitated the process of data cleaning and analysis, increased 67% operational effectiveness and efficiency;
- **Data Visualization:** Produced informative and interactive visualization data governance reports using **R, Tableau** and pivot table;
- **Cross-functional teamwork:** Participated in alignment between analytics, business, and technical sides to identify and resolve issues related to AML regulatory compliance and potential credit risk exposures

### Data Analyst Intern | Deloitte

Jan - Apr 2021

- **Statistical Modeling:** Quantified potential risks and advise clients on complex business problems by interpreting large data sets using statistical methods and technical tools (**regression, logistic model, correlation analysis**);
- **Data mining:** Extracted data from massive client databases (**Hive**) and worksheets using query, combined data using **SQL** and **VBA** scripts; interpreted and analyzed data using **R** and **Python** to identify key metrics and transform raw data into actionable information;
- **Presentation and teamwork:** Actively engaged with the audit team and clients to create visually impactful dashboards in **Tableau**, presented the insight report to both the technical and non-technical data users

### Product Growth Analyst | Hiretual

Aug - Dec 2020

- **User behavior analysis:** Analyzed customer trends in CRM, collected data from marketing campaigns to help drive the future adoption of products; collaborated with key stakeholders to develop marketing strategies, customer retention rate improved from 10% to 38% in 1 month;
- **A/B test:** Designed A/B Test experiments on new released product to monitor performance such as CTR and retention rate, conducted significance analysis on results, summarized analyses from feedbacks and presented findings to stakeholders;

### Insurance Analytics Intern (part-time) | China Pacific Insurance Company

May – Nov 2019

- **Data processing:** Managed large, noisy, seasonal sales datasets using **SQL** and **R**, contributed to building a data pipeline to automate sales data transform and analysis; prepared forecasts using linear regression models, visualized the statistical results through **PowerBI**; presented reports on weekly meetings; Conducted on data query, cleaning, visualization, and provided technical support on the team' s usage of Python, SQL, and Excel
- **Product analytics:** Pulled data from datasets by writing SQL queries and used statistics to provide business insights; monitored product metrics and translated analytical results into business recommendations

## PROJECT

### Machine Learning Project – Fake Reviews Detection with Yelp Dataset

Aug –Sep 2021

- Utilized **Python** for web scraping, Dataset cleaning; Used **R** programming with overfitting problems and model testing
- Performed feature engineering using Unigram and Bigram models, specified the fake review words pattern with Bag of Words Model and Naïve Bayes Classification
- Applied **learning algorithms** including Logistic Regression, Linear Discriminant Analysis, Multinomial Naïve Bayes, Support Vector Machines (SVM), Neural Network model to improve prediction performance, NN model worked best, AUC(83%), F1(82%)
- (**Python**) **scikit-learn, beautiful soup, pandas, seaborn**