

LINYI (LINCOLN) GUO

Toronto, ON | (+1)613-630-0521 | lincoln.gly@gmail.com | [Website](#) | [LinkedIn](#) | [GitHub](#)

SUMMARY

- Over 2 years of experience in data cleaning, data visualization, data analysis, statistical modelling and machine learning
- Outstanding analytical skills with ability to tell the story behind data clearly and succinctly
- Demonstrated record of facilitating analysis/predictions and performing ad-hoc analysis with statistical methods
- Proficient in statistical programming languages/tools such as Python, SQL, SAS, R and Tableau
- Excellent team player and self-starter with great communication and presentation skills
- A motivated M.Sc. in Statistics (9.8/10.0) with a passionate interest in data science and machine learning

PROFESSIONAL EXPERIENCE

Statistical Programmer, Everest Clinical Research Company

May 2021 – Current

- Perform clinical trial database quality/integrity checks by generating **SAS** data TLGs for data reviewing and cleaning
- Develop SAS programs and generate complete, accurate and validated statistical TLGs in well-defined formats
- Program SDTM and ADaM datasets according to the dataset specifications

Research & Teaching Assistant, University of Ottawa - Ottawa, ON

September 2018 – October 2020

- Collaborated with Statistics Canada and my supervisor upon the seasonal adjustment problem
- Brought down the prediction error (MSE) by 25% on average for the US unemployment dataset by applying hidden Markov models (aka state space models) and Bayesian analysis
- Removed the seasonal influence successfully and promoted the business and data analysis process
- Collected and cleaned 30+ monthly retail sales datasets from Statistics Canada and third-party platforms by **Excel**
- Constructed HMMs to break down the retail sales dataset into trend, seasonal and noise components, and achieved better visualization results using **R** and **Python**
- Customized HMMs under specific parameters and adopted parallel computation to accelerate the training
- Delivered tutorial courses and graded assignments and test paper in Math and Stat for over 700 undergraduates
- Assisted undergraduates in dealing with learning problems, and provided regular, accurate and timely feedback

Data Analyst, China Life Insurance Company - Henan, China

July 2016 – September 2016

- Maintained the local branch's database with **SQL** daily, including updating, inserting and removing data
- Collected, input and analyzed clients' data from salesmen, marketing channels and other sources
- Utilized **Excel** to compute statistical data and draw diagrams to provide actionable insights upon clients' data
- Compiled visualised analysis reports and presented the results for decision making

PROJECTS

Hand-written Digit Recognition and Overview of Graph Neural Network

- Built a perceptron (1-layer NN) from scratch using **Python** and achieved 91% accuracy
- Utilized library **keras** to build a 2-layer NN and elevated the accuracy to 98% from 91%
- Self-studied the GNN, and wrote an overview to introduce GNN and how it works

CANDEV Data Challenge: Government Funding

- Collected the raw funding data over 400M from Canada government websites and cleaned with **SQL**
- Validated data and filtered variables by frequency tables and factor analysis
- Provided bar plots, frequency tables and other advanced diagrams such as dynamic bubble plots, geographical maps and their combination to present and illustrate the cleaned funding data using **R**
- Published the whole project as a web-based application through **R Shiny** Server

Mobile Phone Customer Behavior's Analysis and Prediction

- Drew frequency table, bar chart, fan chart and computed descriptive statistics to explore data with **R**, **SQL** and **Excel**
- Utilized techniques in survival analysis such as cox regression to analyze and select variables
- Built and trained logistic models and decision trees upon training set for prediction
- Tested and compared the performance on test set and made ad-hoc analysis
- Achieved 87% prediction accuracy on average for both models in the end

EDUCATION

M.Sc. in Statistics, University of Ottawa – Ottawa, Canada

September 2018 – October 2020

GPA: 9.8/10.0

B.Sc. in Mathematics, Zhengzhou University – Zhengzhou, China

September 2014 – June 2018

GPA: 3.7/4.0

PROGRAMMING & CERTIFICATES

Programming: Python, SQL, SAS, R, Microsoft Office, Markdown, LaTeX, Git(Hub), Linux, MATLAB

Certificates: IBM Data Science Professional Certificate Candidate (Coursera & IBM), Deep Learning (Coursera & Deeplearning.ai), SAS 9.4 Advanced & Based Programming (SAS), SAS Advanced & Based Programmer (Coursera & SAS)