LINYI (LINCOLN) GUO

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

SUMMARY

Experienced data science practitioner with over 5 years of expertise in data analysis, model optimization, machine learning and pipeline development. Skilled in Python, SQL, PySpark, R and SAS with proven track record of reducing exposure by 3 billion CAD and improving efficiency by 30-50%. Strong communicator and strategic thinker, dedicated to lifelong learning and passionate about data-driven projects. Holds M.Sc. in Stat from uOttawa with GPA 9.8 and FRM certification.

PROFESSIONAL EXPERIENCE

Quantitative Analytics Manager, The Bank of Nova Scotia - Toronto, ON **November 2021 - Current**

- Conducted **quant analysis** and optimized **SACCR model**, achieving an accumulated decrease in exposure of **3 billion CAD**
- Developed and implement **25k+ Python**, PySpark and **SQL** lines, **reducing** running time and manual **workload** by **30 - 50%**
- Initiated **outlier detection** automation using **DBSCAN** and engineered **fuzzy string matching** code to improve data accuracy
- Employed **Bash** and **Docker** commands to efficiently query and inspect raw data from diverse data sources
- Orchestrated **version controls** using **Git** and collaborated with stakeholders to deploy repositories into **Oracle databases**

Statistical Programmer, Everest Clinical Research Company - Toronto, ON **May 2021 - October 2021**

- Implemented **10k+ SAS lines** to produce meticulously formatted, precise, and validated **statistical TLGs**
- Executed clinical trial **database quality** and **integrity checks**, leveraging TLGs for data review and cleansing
- Formulated SDTM **dataset specifications**, and programed and performed **quality control** of SDTM and ADaM datasets
- **Maintained documentation** of data and programming procedures in alignment with Corporate Governing Documents

Research & Teaching Assistant, University of Ottawa - Ottawa, ON **September 2018 - August 2020**

- Deployed **hidden Markov models** (HMM) and **Bayesian analysis**, reduced MSE by **25%** for the US unemployment data
- Cooperated with **Statistics Canada**, and applied HMMs to national retail sales data to better **detect and interpret trends**
- **Simulated 5000+** time series data using Monte Carlo Markov model (**MCMC**) across **R**, **Python** and **AWS cloud server**
- Optimized HMMs by **fine-tuning parameters** and adopted **parallel computation** to accelerate the training process by **80%**

PROJECTS

House Price Analysis and Prediction (Regression)

- **EDA**: Harnessed **matplotlib** and **seaborn** to analyze relationships and distribution of variables
- **Feature**: Resolved **outliers** and **missing** values based on EDA and conduct **feature enhancement** using domain knowledge
- **Model**: Trained and tested with regression (**Lasso**) and ensemble models (**XGBoost** and **LGBM**) with 10-fold CV
- **Result**: Ranked as the **top 5%** optimal solution with minimal RMSE globally

Mobile Phone Customer Behavior's Analysis and Prediction (Classification)

- **EDA**: Calculated **descriptive statistics** and geneated data **visualization** with R, SQL and Excel
- **Feature**: Employed techniques in **survival analysis** such as **cox regression** for analysis and feature selection
- **Model**: Trained and tested **logistic** regression and **decision trees** (**C5.0** and **CART**) with 10-fold CV
- **Result**: Achieved **86.71%** prediction accuracy with C5.0 tree and improved around **3%** compared to bmk

SKILLS

Coding: Python, SQL, PySpark, R, SAS, Git, Bash/Shell, Docker, Markdown, LaTeX, C++, MATLAB
Tools: Power-BI, Airflow, Eclipse, Tableau, Unix, Docker, Github, Jira, Bitbucket, Confluence
Skills: Machine Learning, Statistical Modeling, Quantitative Analytics, Data Manipulation, Data Pipeline Development, Data Visualization, Natural Language Processing, Pricing Model, Financial Modeling, Regression Testing, Risk Quant Management

EDUCATION

FRM, Global Association of Risk Professionals **2021 - 2023**

M.Sc. in Statistics (9.8/10.0), University of Ottawa – Ottawa, Canada **2018 - 2020**

B.Sc. in Mathematics (3.7/4.0), Zhengzhou University – Zhengzhou, China **2014 - 2018**