Jack Li

Data analyst

Jack.li.2000@hotmail.com | (437) 265-0290 | https://www.linkedin.com/in/jack-li-2000 |
https://Jack-Li-2000.github.io/JackLiPortfolio/

Skills

Python: Highly proficient with pandas, numpy, matplotlib, seaborn, scikit, scipy, etc.
R: Highly proficient with ggplot, glm, scripting, visualizations.
Excel: Proficient with pivot tables, formula, vlookups, xlookup, conditional formatting, charts, data cleaning, etc.
SQL: Experienced with query calls, views, joins, conditional statements, database management, etc.
Data Visualization: Familiar with telling stories using interactive dashboards in Tableau and PowerBI.
Statistical Methods: Highly proficient with regression, classification, time series, hypothesis testing, and ML.
Cloud computing: Experienced with Google Cloud Platform (GCP) and Azure.

Work History

Data Scientist Intern, Teck Resources
(2021 – 2022, Vancouver BC)

Machine Failure Classification
- Performed data cleaning, provided **exploratory** and **ad hoc analysis** with **Python** and **SQL**.
- Retrieved relevant features for **classification** using stepwise, lasso, and elasticnet **feature selection** methods.
- Identified equipment failure using **classification models,** validated with F1-score of 0.85.
- **Reduced time taken to identify failure by 45%** and **increased accuracy of failure detection by 125%** in material processing pipeline.

Equipment Life Cycle Prediction
- Predicted equipment lifecycle, development of repair vs replace model using **quantile regression**.
- Validated equipment lifecycle **model prediction** results with acceptable RMSE, MAPE with **projected $9Mil** reduction on excessive expenditures for premature replacements and extended repairs.
- Displayed and cleaned analytical insights with **DAX** and **visualizations** in **PowerBI**.

Projects

Wordnet Similarity Recognition Analysis
Analyze NLP data structure similarities with Human semantic relationships between words. Using data from the *Semantic Priming Project* for human reaction time and accuracy based on response to stimuli, analyze **correlation** between *Wordnet* path similarity and Human word-pair recognition task performance with hypothesis and **significance testing** in **Python.**

Toronto AirBNB Popularity Insights Dashboard
Gained insight into AirBNB customer tendency towards stays in Toronto with data scraped from *InsideAirBNB* between Nov. 2022 and Nov. 2023, perform analysis on popularity in attributes like location, pricing, room type, etc. Analytical insights expressed in **Tableau** dashboard with interactive map and time filtering. **Data cleaning** in **Excel** and **Python**.

Global COVID-19 EDA & Visualization
Analysis of confirmed global and localized infection, mortality, vaccination, etc by country, continent, and worldwide of COVID-19 pandemic with data sourced from Our*WorldInData*. Datasets joined, cleaned, analyzed with **MS SQL Server** and presented as interactive dashboard in **Tableau Public** to provide insight into global reactions and results of Pandemic.

Education

Honours B.Sc. in Statistics & Computational Cognitive Science
(University of Toronto, Toronto Ontario)
Relevant skills include practical knowledge in data regression models, forecasting and predictive techniques, classification, time series analysis, dimensionality reduction, data visualization, and drawing key information to deliver insights on complex data sets. Also, in-depth knowledge of data structures and programming.