Jenny Lee

3333 Wesbrook Mall, Vancouver, BC | V6S 0E3
jennyjeeun@gmail.com | +1 (403) 969-0567 | github.com/jlee2843 | linkedin.com/in/jennyjeeun

Education

University of British Columbia — Expected graduation date: 06/2024
Master of Data Science | Link to all courses — ***GPA: A+***

University of British Columbia — 2021 - 2023
Bachelor of Science, Major in Statistics (Partial credit transfer from RDP) — ***GPA: B***

Red Deer Polytechnic — 2020 - 2021
Bachelor of Science, Major in Statistics — ***GPA: A+***

University of Western Ontario — 2014 - 2018
Bachelor of Science, Honors Specialization in Neuroscience — ***GPA: A-***

Work Experience

University of British Columbia — 09/2023 - Present
Food Sustainability Data and Reporting Analyst GitHub — *Python, R, Git, GitHub, Docker*

- Engineered an automated workflow to evaluate the cumulative greenhouse gas emissions resulting from all procurement items acquired and utilized by UBC Food Services and AMS.
- Implemented machine learning models to gauge UBC's ability to achieve the Climate Action Plan 2030. In cases where meeting the goal is doubtful, the models offer recommendations to reduce annual greenhouse gas emissions.

University of British Columbia, Department of Statistics — 09/2023 - Present
Graduate Teaching Assistant — *R, Git, GitHub*

- Member of the teaching team for STAT 200 (Elementary Statistics for Applications) and DSCI 100 (Introduction to Data Science) at UBC.
- Attended lectures to address students' questions in real-time. Facilitated a lab section, guiding students through weekly lab materials. Assessed and graded student exams and assignments.

Teck Resources Limited — 01/2023 - 08/2023
Data Scientist Intern — *Python, Docker, Git, GitHub, Databricks, AWS*

- Developed statistical solutions to address machine learning model performance degradation resulting from data drift. Automated the process of re-scaling incoming data values upon detecting data drift.
- Explored and implemented alternative machine learning models aimed at replacing the existing high-performance but costly BERT model. Investigated models with enhanced interpretability, such as K-means clustering, logistic regression, and DBSCAN clustering.

University of British Columbia — 08/2022 - 04/2023
Climate-Friendly Food Sustainability Data Analyst Report1 | Report2 | GitHub — *Python, R, Git, GitHub*

- Engineered an automated workflow to evaluate the overall impact of newly added food items to the UBC Food Services database.
- Categorized all food items offered by UBC Food Services into green, yellow, and red based on the total greenhouse gas emissions produced by 100g of each menu item.
- Labeled newly added food ingredients in the UBC database.
- Developed a web-based application for Food Services staff to efficiently search for climate-friendly labels. The application incorporates dynamic graphs that change based on the selection of particular residence halls or vendors.

Ciena — 09/2023 - 12/2023
Modeling and Machine Learning Intern — *Python, PostgreSQL, Bigbucket, NetworkX*

- Enhanced an outdated artificial neural network model through real-time data analysis. Conducted comprehensive testing using six different methods and generated a detailed report comparing the performance of each approach.
- Collaborated with colleagues to develop an interactive app using Plotly Dash. This app enables clients to effortlessly monitor their current network traffic and explore future predictions.

Cybera Inc. 04/2022 - 03/2024

Data Science Intern *Python, R, Git, GitHub*

- Analyzed, wrangled, and visualized data sourced from an open-source database on trending topics. Gained exposure to various Python and R packages for visualization, including Matplotlib, Seaborn, Plotly, Plotly Dash, and Ggplot.
- Collaborated with colleagues to create hackathon content for Albertans. Prepared and processed open-source datasets to serve as valuable resources for a data science hackathon.

Research Experience

Stabilized COre gene and Pathway Electron Algorithm 2022

University of Calgary | Alberta Innovates Studentship 2022 Awardee GitHub *Python, R*

- Utilized ML algorithms including SCOPE-stabilized LASSO regression, correlational analysis, co-differential analysis, and pathway enrichment with Over Representation Analysis.
- Implemented the SCOPE algorithm on the melanocytic tumor transcriptome dataset from NCBI. Assisted in debugging errors within the algorithm and generated visualizations for publication.

Assessing the Impact of Peer Review: A Comprehensive Analysis of Scientific Article Effectiveness 2022

University of Calgary GitHub *Python, R*

- Conducted natural language processing tasks to extract contextual information from online articles in PDF format.
- Examined the total number of article citations and assessed the extent of changes made during the peer-review process.
- Established an automatic workflow for the process above.

Projects

Alternative Recipe 2023

The Climate Change-Makers Challenge 2023: Best of Rest Winner GitHub | Web-App *Python, Streamlit*

- Calculated the total greenhouse gas emissions produced from ingredients used in a recipe and assessed whether the emissions exceeded a baseline threshold.
- In cases of exceedance, recommended the closest alternatives to ingredients with high emission factors. Alternatives were suggested based on food category and nutrition distribution.
- Developed a web-based application using Streamlit for user interaction. Implemented email functionality to send the modified recipe to users' emails upon request.

Credit Card Fraud Detection 2023

DSCI 522: Data Science Workflows Assignment GitHub | Report *Python, Jupyterbook, Docker*

- Conducted credit card fraud detection using synthetic data generated by Capital One.
- Explored and implemented diverse machine learning models, such as logistic regression, random forest classifier, and gradient boost classifier. Performed hyperparameter optimization to enhance accuracy.
- Compiled a comprehensive report on our findings in a Jupyter book.
- Containerized the code using Docker, enabling new users to execute all the code in their terminal.

Misinformation Analysis 2023

CANIS Hackathon at the University of Calgary: Special Awards Winner GitHub *Python, Plotly Dash, Heroku*

- Conducted an analysis to classify fake news and non-fake news. Explored and implemented various machine learning models, including logistic regression, support vector machine, and Naive Bayes classifier.
- Created a web-based comprehensive report on our findings using Plotly Dash, incorporating dynamic visualizations that change upon selection. Deployed it on Heroku for accessibility.

Air and Water Quality Across Canada Visualizations 2023

Visualize air and water quality across Canadian provinces GitHub *Python, Plotly*

- Analyzed trends in air and water quality changes across Canadian provinces.
- Created interactive visualizations, including line graphs, bar graphs, pie charts, and geographical heatmaps, facilitating easy comparison over time.

Leadership and Volunteering

- *Statistics Liaison Officer*, University of British Columbia 2021 - 2022
- *External Events Director*, Undergraduate Statistics Society at UBC 2021 - 2022
- *External Director*, Women in Data Science at UBC 2022 - 2023