Jenny Lee

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

Work Experience

Pacific Salmon Foundation (Capstone Project) | Vancouver, BC — 04/2024 - 06/2024

Data Science Consultant — *Python, R, PL/Python, Airflows, AWS S3*

- Developed a Hierarchical Bayesian Model; Conducted a survival analysis on Pacific salmon populations by incorporating beta regression models. Built a two-step hierarchical structure to effectively analyze the impact of various covariates on survival rates.
- Developed a robust pipeline to forecast the outmigration timing of salmon into the ocean. Implemented LASSO Regression in conjunction with XGBoost, achieving a significant reduction in RMSE by 100.

University of British Columbia, Center of Sustainability | Vancouver, BC — 08/2022 - Present

Data and Reporting Analyst GitHub — *Python, R, Git, GitHub, Docker, Plotly Dash*

- Developed an efficient ETL pipeline to extract procurement data from a PostgreSQL server for UBC Food Services, applying rigorous data wrangling techniques to ensure the data was transformed into a usable format.
- Leveraged advanced regression methodologies to evaluate the current trajectory towards achieving the milestones set forth in the Climate Action Plan 2030. In cases where meeting the goal is doubtful, the models offer recommendations to reduce annual greenhouse gas emissions.
- 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.

Teck Resources Limited | Vancouver, BC — 01/2023 - 08/2023

Data Scientist Intern — *Python, SQL, Docker, Git, GitHub, Databricks, AWS Services (S3, EC2), Microsoft Azure*

- Retrieved raw data from AWS using Apache Spark clusters. Cleaned raw data on Databricks working space for further processing.
- 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.

Ciena Corporation | Ottawa, ON — 09/2022 - 12/2022

Big Data, Modeling and Machine Learning Engineer Intern — *Python, PostgreSQL, Bitbucket, NetworkX, PyTorch, Tensorflow*

- Enhanced an outdated artificial neural network model through real-time data analysis. Conducted comprehensive testing using six different hyperparameter optimizations and generated a detailed report comparing the performance of each approach.
- Implemented a novel algorithm for feature selection, significantly improving the efficiency and accuracy of the neural network model. This streamlined the model's performance and reduced computational workload.
- 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. | Calgary, AB — 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.

Education

University of British Columbia — 08/2023 - 06/2024
Master of Data Science | Link to all courses — ***GPA: A+***

University of British Columbia — 2021 - 2023
Major in Statistics (Additional Coursework) — ***GPA: B+***

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

Technical Skills

- **Languages:** Python, SQL, R, MATLAB, Java, Dr.Racket
- **Libraries and Tools:** PyTorch, Tensorflow, Keras, Scikit-learn, Sktime, Spark, Git, PowerBI, Pandas, Numpy, Matplotlib, Seaborn, Plotly, Docker, AWS, Azure, Altair, Databricks, Jupyter Notebook, MLOps
- **Dashboards:** Plotly Dash, Streamlit, Shiny (R), Tableau, PowerBI, Superset
- **Concepts:** Machine Learning, Neural Networks, API, Cloud Computing, Data Engineering, Natural Language Processing (NLP)

Research Experience

Stabilized COre gene and Pathway Electron Algorithm 2022

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

- Utilized a series of machine learning algorithms to engineer a comprehensive statistical model. This encompassed the application of LASSO regression for feature selection, correlational analysis to uncover relationships, co-differential analysis for nuanced insights, and pathway enrichment via Over Representation Analysis, culminating in a robust analytical framework.
- Implemented the SCOPE algorithm on the melanocytic tumor transcriptome dataset from NCBI. Assisted in debugging errors.

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

Pythoshop: Python Graphics Package 2024

DSCI 525: Software Engineering Assignment GitHub | Read the Docs *Python, CI/CD, Docker, Git, RESTful APIs*

- Developed and deployed a Python graphics package inspired by the functionalities of Photoshop.
- Implemented a CI/CD pipeline using Jenkins and Docker to automate the build, test, and deployment processes of the package.
- Integrated automated testing frameworks (pytest and unittest) to ensure code quality throughout the CI/CD pipeline.

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.