Lepeng Zhou

Contact Information

3500 Richmond Road, Victoria, BC V8P 4P8 | 825-983-5399 | lepeng@uvic.ca | My Website

Objective

Master's student seeking a Co-op job in Data Science and Analytics related field to apply skills and knowledge in real-world scenarios. Helping businesses derive better decisions from data.

Competencies Highlights

- Currently pursuing a Master of Engineering in Applied Data Science at the University of Victoria, with a Bachelor's degree in Computing Science, providing a robust foundation in machine learning, computer science, and statistics
- Extensive experience in building, documenting machine learning pipelines using Python and R, deriving insights from complex, unstructured datasets such as languages and time series. Designing regression models and interpreting statistical results
- Proficient in processing data using SQL and Pandas. Managing massive databases using PySpark. Skilled in data visualization and reporting results using PowerBI
- Demonstrated ability to manage time effectively between learning new knowledge, making projects, and maintaining a high GPA throughout rigorous master's and bachelor's degrees
- Seven years of Canadian education, ensuring strong written and verbal communication skills. Experienced in collaborating on team projects and coursework

Education

Master of Engineering - Applied Data Science September 2023 ~ September 2025
University of Victoria, Department of Electrical and Computer Engineering Victoria, BC
GPA: 8.32 / 9 ≈ 3.8 / 4

- Notable Coursework: Data Analysis and Pattern Recognition (A); IOT: Analytics and Security (A+); Systems for Massive Datasets (A+)

Bachelor of Science - Honors in Computing Science September 2019 ~ June 2023
University of Alberta, Department of Computing Science Edmonton, AB
GPA: First-Class Honors

- Notable Coursework: Intelligent System (A+); Statistics II (A); Machine Learning (A-)

Skills

Python; Pytorch; Matlab; Statistics; R; C; SQL; Pandas; Spark; Neo4j; Java; Power BI; Visualization Design; Report Writing; Public Speaking; HTML; GitHub; Microsoft Office.

Projects

Time Series Forecasting on Supermarket Sales (Personal Project) May 2024
Forecasted sales for a major Ecuadorian supermarket chain by category and store, analyzing data between 2013-2018 as part of a Kaggle competition.

- Employed time series forecasting techniques; modeled trends and seasonality using Fourier Transform to deconstruct components of data
- Utilized XGBoost to model residuals, enhancing model accuracy with additional time-related features such as daily oil prices and holiday schedules
- Achieved 10% to 20% MAPE (lower is better) when validating on certain supermarket categories
- https://github.com/Lepeng1024/Sales-Forecasting-XGBoost

Parallel Clustering on Big Graph Data Set (Capstone Project) March 2024

Implemented a PySpark-based, Structural Clustering Algorithm to put people with similar interpersonal relationships into groups.

- Algorithms are built upon the famous MapReduce Model, handling the Resilient Distributed Dataset. Tested on a small, simulated toy data for validation
- Capable of handling massive datasets directly on disk. Not limited by memory size
- Distributed computing allows multiple computers or multiple CPU cores to do tasks simultaneously, greatly reducing overall processing time
- https://github.com/Lepeng1024/Parallel-Structural-Clustering-PySpark

Interactive Data Visualization on Cycling Accidents (Capstone Project) March 2024

Created an interactive visualization tool with Power BI to alert cyclists where accidents happen more frequently. Supports multiple interactions such as select, filter.

- Designed with visualization theories, discussed strength, alternatives, and drawbacks
- https://github.com/Lepeng1024/Data-Visualization-Power-BI

Sentiment Analysis on Movie Reviews (Capstone Project) November 2023

Engineered a Python-based machine learning model to classify text reviews as positive or negative, assisting in identifying high-quality Movies.

- Embedded each sentence as a 3000-dimensional vector using TF-IDF, then reduced its dimension by 80% using PCA, and finally applied linear SVM to classify data
- Achieved a test accuracy of 88% on the test set
- https://github.com/Lepeng1024/Sentiment-Analysis-SVM

Work Experiences

Residence Assistant August ~ September 2022

University of Alberta Edmonton, AB

Enhanced dormitory experience for first-year students.

- Assisted students with move-in, building patrols, and providing general information
- Improved communication, activity programming, and teamwork skill

Awards and Certificates

Dean's Honor Roll 2021-2022 *University of Alberta*

Feature Engineering Course Completion Certificate, *Kaggle*

Time Series Course Completion Certificate, *Kaggle*

Interests

Sustainable Energy; Electric Vehicles; Virtual Reality; Global Affairs; Large Language Models.

References

Endorsements from current and past IT professionals available upon request.