Lepeng Zhou

3500 Richmond Rd, Victoria, BC | 825-983-5399 | lepeng@uvic.ca | Website: lepeng1024.github.io

Education

University of Victoria
Master of Engineering - Applied Data Science — Sep 2023 ~ Present
GPA: 8.32 / 9 ≈ 3.8 / 4 — Victoria, BC

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

Skills and Competencies

Data Mining: Proficient in Python, Pytorch, MATLAB, R and C for complex data analysis, with a focus on mining time series data and forecasting the future.
Database and Software Development: SQL; Pandas; Spark; Neo4j; Java; HTML; CSS.
Data Visualization: Power BI, Tableau, fundamental design principles.
Statistics: Solid knowledge in regression analysis, hypothesis testing, statistical significance, indicator variables, probability theory, and tree-based machine learning.
Professional Communication: Team player in collaborative projects, proactive in self-learning, comprehensive in report writing, and effective in presenting.

Personal Projects

Time Series Forecasting on Supermarket Sales
Forecasted sales for a major Ecuadorian supermarket chain by category and store, analyzing data from 2013-2018 as part of a Kaggle competition.

- Employed time series forecasting techniques; modeled trends and seasonality using Fourier Transform to deconstruct components of the 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 when validating on some supermarket categories.
- https://github.com/Lepeng1024/Sales-Forecasting-XGBoost

Interactive Data Visualization on Road Accidents Involving Cyclists
Created an interactive visualization tool with Power BI to alert cyclists where accidents happen more frequently. Supports multiple interactions such as select, filter, etc.

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

Parallel Clustering on Big Graph Data with PySpark
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 the disk. Not limited by memory size. Distributed computing allows multiple computers or multiple CPU cores to do the task simultaneously, greatly reducing the overall processing time.
- https://github.com/Lepeng1024/Parallel-Structural-Clustering-PySpark

Sentiment Analysis on Movie Reviews
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 dimensionality by 80% using PCA, and finally applied linear SVM to classify the data.
- Achieved a test accuracy of 88% on the test set.
- https://github.com/Lepeng1024/Sentiment-Analysis-SVM

Work Experiences

Residence Assistant Aug 2022 ~ Oct 2022
University of Alberta Edmonton, AB
Enhanced the dormitory experience for first-year students.

- Helped students with move-in, building patrols, and providing general information.
- Improved my communication, activity programming, and teamwork skills.

Awards & Certificate

Dean's Honor Roll 2021-2022 May 2022
Faculty of Science, University of Alberta

Time Series
Course Completion Certificate, Kaggle

Interests

Cycling, Virtual Reality, Global Affairs, Large Language Models, Computer Vision.

References

Endorsements from current and past IT professionals available upon request.