SEA RAN CLEON LIEW

linkedin.com/in/cleon-liew | srcliew@uwaterloo.ca (226-201-3572) | cleon01.github.io (My Portfolio)

SKILLS

Programming Languages: Python, R, SQL, Java, C, Bash, Lua

Libraries: scikit-learn, Keras, PyTorch, TensorFlow, NumPy, SciPy, Pandas, Matplotlib, ggplot2, Cuda, Hugging Face Transformers, BeautifulSoup, Scrapy, Splash, Selenium, Appium

Technologies: AWS Cloud (Batch, CloudWatch), Git, Bitbucket, Jira, Confluence, Power BI, Jupyter Notebook, Linux, Wireshark, Microsoft Office Suite, Excel (Pivot Tables, VBA, Macros)

WORK EXPERIENCE

Data Science Research Assistant – University of Waterloo   Waterloo, ON (May 2023 – Feb 2024)
1) **Metastatic Cancer Medical Image Analysis Project** under Prof. Subha Kalyaanamoorthy
- Achieved 78.28% classification accuracy in metastatic nerve cell image classification using **PyTorch** by implementing CNN, ResNet, and ViT models, improving baseline by 6% through feature engineering and hyperparameter tuning.
- Presented findings on model performance to stakeholders using **PowerPoint**, **R (ggplot2)**, **Matplotlib**, and **Pandas.**
2) **FHIR-Formatted Healthcare Data Encoding Project** under Prof. Bryan Tripp
- Applied data encoding methods such as Sinusoidal Encoding in **Python** to enhance data inputs to the transformer.
- Used **SQL** to **cleanse**, **normalize**, and **structure** FHIR-formatted datasets, optimizing them for model input.

Deep Packet Inspection QA Engineering Intern – Nokia   Ottawa, ON (Sep 2022 – Dec 2022)
- Automated traffic capture tasks for 20+ apps (Web, Android, iOS) using **Python**, **Selenium**, **Appium**, and **Linux Crontab**.
- Performed packet-level analysis using **Wireshark** on traffic data from 60+ application filters, identifying issues and submitting bug reports for 4 filters to improve classification accuracy and ensure enforcement of QoS policies.

Data Validation Intern – Ontopical   Calgary, AB (Remote) (Jan 2022 – Apr 2022)
- Built 50+ **web scrapers** to contribute to the ETL process for multimedia data collection using **Scrapy**, **Splash**, and **Lua**.
- Deployed scrapers on **AWS Batch** and monitored performance using **AWS CloudWatch**, aggregating document-type counts and visualizing distribution trends with **Power BI**.
- Worked within an Agile Scrum framework, managing tasks and version control using **Git**, **Bitbucket**, **Jira**, and **Confluence**.

Assurance Practice Intern – PricewaterhouseCoopers (PwC)   Hong Kong (May 2021 – Jul 2021)
- Analyzed **financial data** and **trends** using **Excel**, identifying inconsistencies in reports to support **audit planning**.
- Performed **detailed testing** and risk assessments to enhance compliance and reporting accuracy.

FREELANCE

Backend SQL Database Development – Nanny & Eldercare Immigration Inc.   Canada (March 2025 – May 2025)
- Collaborated with a remote team to design and validate a relational database for a client management system using **MySQL**.
- Configured **Hamachi VPN** and **MySQL** Workbench for remote access; developed SQL scripts and contributed to a Python-based ETL pipeline to automate data import while maintaining schema integrity.

PROJECTS/ PUBLICATION

Beam - An Algorithm For Detecting Phishing Link 2022 APSIPA Annual Summit and Conference (**Github code**)
- Applied six different NLP AI models, including fine-tuned BERT, CNN, and LSTM networks, for phishing URL detection.
- Utilized various **NLP-based tokenizations** (character & subword) and evaluated their performance on processed datasets.

Use Of Subword Tokenization For Domain Generation Algorithm (DGA) Classification Cybersecurity 6, 49 (2023).
- Developed an integrated scheme consisting of various NLP models for better classification of domains generated by DGAs.

EDUCATION

University of Waterloo   Waterloo, ON (Sep 2020 – Jun 2024)
Bachelor of Mathematics, Major in Computational Mathematics (Co-op), Minor in Computing
- Cumulative GPA: 85/100; Graduated with Distinction