SEA RAN CLEON LIEW

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

SKILLS

Programming Languages: Python, R, SQL, C, VBA, MATLAB, LaTeX

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

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

EDUCATION

University of Waterloo Waterloo, ON

Bachelor of Mathematics, Major in Computational Mathematics (Co-op), Minor in Computing Sep 2020 – Jun 2024

- Cumulative GPA: 85/100; Graduated with Distinction
- Relevant Coursework: Designing Functional Programs; Elementary Algorithm Design and Data Abstraction; Data Types & Structures; Computational Statistics and Data Analysis; Neural Networks; Data Visualization; Applied Linear Models

WORK EXPERIENCE

Data Science Research Assistant Waterloo, ON

University of Waterloo May 2023 – Feb 2024

FHIR-Formatted Healthcare Data Encoding Project under Prof. Bryan Tripp

- Implemented **data encoding methods** such as Spatial Semantic Pointers and Sinusoidal Encoding in **Python** to enhance the quality of FHIR-formatted data inputs for the **NLP Transformer Model**, improving the model's accuracy by 10.7%.
- **Cleansed, normalized, and structured FHIR-formatted datasets**, optimizing them for model input using **SQL**.
- Utilized **NumPy, SciPy, and Pandas** for data preprocessing, transformation, and analysis.

Metastatic Cancer Medical Image Analysis Project under Prof. Subha Kalyaanamoorthy

- Developed **machine learning models** for nerve cell image classification using **PyTorch**, achieving 78.28% accuracy.
- Applied **feature engineering, cross-validation, and hyperparameter tuning** to **fine-tune** and optimize **Computer Vision Models** such as **Vision Transformer, CNNs, ResNet, and Inception**.
- Visualized model performance and plotted data using **R (ggplot2)**, **Matplotlib**, and **Pandas** to support research insights.

DPI Automation Data Analyst Intern Ottawa, ON

Nokia Sep 2022 – Dec 2022

- **Automated** network traffic detection QA for 20+ applications (Web, Android, iOS) using **Python**, **Selenium**, and **Appium**.
- Analyzed traffic data captured by 60+ application filters by performing packet-level analysis with **Wireshark** to **troubleshoot issues** and **submit bug reports** for 4 filters to improve classification accuracy and enforce proper QoS policies.
- Scheduled recurring automated testing tasks using **Linux Crontab** to collect testing data.

Data Quality Assurance Intern Calgary, AB (Remote in Toronto, ON)

Ontopical Jan 2022 – Apr 2022

- Developed over 50 web scrapers to aid the ETL process for multimedia **data collection** using **Scrapy**, **Splash**, and **Lua**.
- Executed web scrapers on **AWS Batch**, monitored their performance using **AWS CloudWatch**, collected and aggregated document type counts from each scraper post-execution, and leveraged **Power BI** to visualize document distribution trends.
- Did **version control** and task management within an **Agile Scrum** framework using **Git**, **Bitbucket**, **Jira**, and **Confluence**.

Financial Auditor Intern Hong Kong

PricewaterhouseCoopers May 2021 – Aug 2021

- Analyzed financial statements, identified trends and fluctuations to inform audit conclusions using dynamic reports in **Excel (Pivot Tables, VBA, and Macros)**, and created audit reports for 4 companies.

PUBLICATIONS

Use Of Subword Tokenization For Domain Generation Algorithm Classification Cybersecurity 6, 49 (2023).

Beam - An Algorithm For Detecting Phishing Link 2022 APSIPA Annual Summit and Conference, pp. 598-604.