Minh (Halley) Dao

0426051477 | halleydao2004@gmail.com | www.linkedin.com/in/halley-dao | https://github.com/halleyd2804 | Halley's Portfolio

EDUCATION

Bachelor of Science, Computing and Software Systems Jul, 2022 - Jul, 2025
The University of Melbourne

- Recipient of the **Melbourne International Undergraduate Scholarship**, awarded for academic excellence and potential
- Awarded **Melbourne Plus: Innovation** for contributions to innovation-focused extracurriculars and projects

RELEVANT EXPERIENCE

Data Engineer Intern Jul, 2024 - Oct, 2024
Walter and Eliza Hall Institute of Medical Research, Melbourne, VIC

- Optimized website performance by implementing in-memory caching, improving loading times and reducing crashes, while utilizing **Nectar Cloud** for deploying and running R-based data analysis and machine learning models on large datasets to enhance computational efficiency and reduce processing time
- **Transformed data files from H5 format to RDS format** for enhanced compatibility and **streamlined direct data retrieval in R**, increasing efficiency in data processing and boosting website speed by **20%**, while reducing reliance on Python data transfers
- Leveraged **Excel** and **SharePoint** for effective collaboration, team communication, ensuring accurate data presentations and improving workflow integration

Data Analyst Intern Jul, 2023 - Jul, 2023
Viettel Digital, Hanoi, Vietnam

- Utilized **Tableau** (including Level of Detail (LOD) expressions, calculation fields, table calculations, parameters, and data blending) to design interactive, dynamic visualisations
- Cleaned and pre-processed **over 100,000 data records** using **Excel** to ensure accuracy and consistency, streamlining analysis and improving overall quality of visualisations and insights

TECHNICAL SKILLS

- **Programming & Development**: Experienced in Python (NumPy, Pandas, Matplotlib, Scikit-learn, NLTK, Beautiful Soup, regex), R, JavaScript, HTML, CSS, SQL, MATLAB, Git, GitHub, Jupyter Notebook, VSCode, RStudio, Nectar
- **Data Science & Databases**: Proficient in Machine Learning, Data Modelling, Data Mining, Database Querying, EDA, MS Excel (Data Cleaning, Analysis, Visualisation), Tableau, MySQL, PostgreSQL, MongoDB (NoSQL), NLP (Prompt Engineering, Text Normalization, Information Extraction), Canva
- **Collaboration & Tools**: Adept at utilizing Agile methodologies (Jira, Confluence), SharePoint for project management and collaboration

RELEVANT PROJECT WORK

Analyse 10-K reports using GENAI

- Cleaned and preprocessed raw text from 10-K filings using Python libraries such as **regex** and **Pandas** and utilized **Beautiful Soup** specifically for **HTML parsing** to extract relevant sections and remove unwanted characters.
- Engineered prompts to guide the **Gemini API** in generating structured summaries of financial data, ensuring accurate extraction of key insights for AI-related investments and trends.

Medical Pantry Admin Portal and Medical Item Identification

- Leveraged **Jira** and **Confluence** to implement **Agile methodologies**, streamlining task and deliverable management while fostering close collaboration with stakeholders to ensure seamless alignment with strategic business objectives.
- Developed a fuzzy matching indexing system in **JavaScript**, leveraging tokens derived from scanned text and barcodes with enhanced weighting for barcodes, achieving **80% accuracy** in matching data with items in our **MongoDB** database

Financial Portfolio Construction

- Developed a **Ridge regression model** using **Scikit-learn** to predict future returns, optimizing portfolio performance through Sharpe ratio maximization
- Preprocessed and normalization with **Pandas** and **NumPy**, handling missing values and computing correlations for reliable forecasting and visualized portfolio performance with **Matplotlib**

Fraud Detection

- Constructed **Logistic Regression** models in **statsmodels** and **Neural Networks** in **TensorFlow**, applying advanced feature engineering with **Pandas** to improve prediction accuracy
- Attained an **AUC score of 0.81** for the Neural Network model and optimized prediction accuracy through model evaluation