LOUIS CHU

(415) 806-4691 | louischu1010@gmail.com

EDUCATION

University of California, Irvine Expected June 2024

- B.S. in Data Science GPA: 3.83
- B.A. in Business Administration
- Relevant Coursework: Probability and Statistics I-III, Statistical Computing and Exploratory Data Analysis, Neural Networks and Deep Learning, Multivariate Statistical Methods

EXPERIENCE

University of California, Irvine October 2022 - Current

Research Assistant | Irvine, California

- Streamlined data management processes by consolidating MIMIC-III dataset using Python's Pandas library and Google Cloud Platform; reduced data processing time by 50% and increased efficiency of medical data analysis
- Automated medical term extraction using NLTK's POS tagging and other techniques to analyze 14K+ medical terms
- Visualized the frequency distribution of diagnosis and prescription records using WordCloud to find the most impactful disease and medications

University of California, Irvine July 2023 - Current

MUST Project Research Assistant | Irvine, California

- Conducted comprehensive analysis of student behavioral patterns within the Canvas Learning Management System, identifying and refining over 150 unique action patterns to enhance user engagement and educational outcomes
- Integrated 3 distinct schemas within AWS Redshift, extracting 80K+ rows of student behavior data categorized by day and hourly intervals; facilitated data-driven decision-making through trend analysis and behavioral pattern identification
- Implemented advanced machine learning models, including Random Forest and XGBoost, using RStudio to evaluate student continuity in a three-course sequence, achieving a ROC score greater than 0.60 for specific courses
- Engaged in a collaborative effort with fellow lab members to consistently deliver high-quality weekly presentations

Green Dot Corporation

Data Engineering Intern | Shanghai, China July 2023 – September 2023

- Monitored 20+ running jobs with an average of 10+ sub tasks to achieve a success rate of 99% via Informatica under the database engineering team
- Collaborated with 5+ Database Engineering Team members to sync up data model change from OLTP to Amazon Redshift via Jira
- Streamlined a standardized machine learning pipeline on company confluence page to analyze customer credit rating with a F1-score of 0.81; delivered a PowerPoint presentation to the executive members of the data team

PROJECTS

Sepsis Data Analysis April 2023

- Coordinated in a 4-member team to develop and deploy 7 machine learning models using Python and Jupyter Notebook to predict the likelihood of Sepsis cases; models achieved an average F1 score of 0.78 and measured losses using Binary Cross-Entropy
- Employed SQL queries in Azure Data Studio to extract valuable insights from patient data, contributing to the development of accurate predictive models for early detection of Sepsis

Yelp Review Analysis April 2023

- Collaborated with a team of 3 to code a ggplot2 correlation matrix which visualized 10+ restaurant categories in a 2-day hackathon
- Implemented a simple NLP program to classify customer reviews into time or service-related complaints into a binary variable

CIFAR-10 Image Classification December 2022

- Spearheaded a team of 2 to conduct comprehensive exploration of machine learning techniques on CIFAR-10 dataset
- Optimized the Random Forest model to use 20% fewer parameters and achieve 48% accuracy in image classification, reducing processing time by 50% and ensuring efficient usage of computational resources
- Conducted in-depth analysis of machine learning model performance and fine-tuned over 20 hyperparameters in the models' parameter space

TECHNICAL SKILLS

Programming Languages: Python (NumPy, Sklearn, Pandas, PyTorch, TensorFlow, NLTK), R (tidyverse, ggplot2, dplyr), MySql PostgreSQL, Amazon RedShift, C++, Java

Software: Anaconda, Visual Studio, VMWare, FileZilla, Excel (vlookup), Azure Data Studio

Statistical Analysis: Hypothesis testing (t-test, f-test), Multiple Linear / Non-Linear Regression, MANOVA, GLM

Machine Learning: Convolutional Neural Networks, XGBoost, Random Forest, Logistic Regression, kNN, Clustering