Roland Chin

Cal ‘21 - Data Science / Computer Science

rond24933chn@berkeley.edu - (408) 908 9187 - rchinnn.github.io

Senior at UC Berkeley seeking an internship / entry level job to build & apply my skills.

COURSES

Data Science

- The Foundations of Data Science
- Statistical Methods for Data Science
- Principles & Techniques of Data Science
- Probability for Data Science
- Data, Inference, and Decisions
- Intro to Artificial Intelligence
- Intro to Time Series *(next semester)*
- Introductory Applied Econometrics *(next semester)*
- *currently self studying MySQL & Tableau*

Computer Science

- Structure & Interpretation of Computer Programs
- Discrete Mathematics & Probability Theory
- Data Structures
- Machine Structures
- Efficient Algorithms and Intractable Problems

EXPERIENCE

zHealthEHR Data Science Intern

May 2020 - Aug 2020

- Cleaned and performed EDA on healthcare data from chiropractors with MySQL
- Created a prototype model that provided guidance to providers on how to increase their building based on similar overlapping features between providers using seaborn

Applied Statistics: Data Consulting Committee

Sept 2019 - Dec 2019

- Worked with Grandmark International, an auto parts company in South Africa, to create a new stocking policy based on demand forecasting and customer segmentation
- Utilized a loss function, moving window, and confidence levels to predict optimal turnover quantities for future inventory for each stock code

Project ATMA, AI Startup Internship

Aug 2018 - Feb 2019

- Worked with a team of Berkeley graduates to help develop an AI platform to can automatically call 911 through real-time analysis of security camera footage, using both the ResNet50 and Xception models
- Collected videos of emergency footage to train ATMA to recognize such situations

TEACHING

Statistics Undergrad Student Association: Education Committee

Sept 2018 - May 2020

- Created material for a DeCal (student run courses at Berkeley) on data analysis, specifically for Kaggle projects
- Hosted and taught weekly programming workshops for fellow college students, alongwith weekly programming assignments

Academic Intern for Data 8: Foundations of Data Science

Jan 2018 - Dec 2019

- Guided students with questions on labs, homework, and projects
- Created comprehensive course study material for all students
- Assisted Berkeley MOOCLab with EdX online course launch as beta tester

PROJECTS

Feature Selection for Neural Networks
- Compared losses for random, engineered, and learned features on a linear model
- Utilized statsmodels module, principal component analysis, and PyTorch

Neural Network Digit Classification
- Implemented a neural network to classify handwritten digits by altering hidden layer sizes, batch sizes, learning rate, and number of hidden layers
- Used the ReLU operation for non-linearity and trained models with loss for gradient-based updates

Pacman Game
- Implemented DFS, BFS, and A* Search agents for multiagent Pacman game using Python
- Applied minimax and Alpha-Beta pruning techniques

Decryption with Markov Chain Monte Carlo
- Created a decoder with transition matrices/bigrams to decrypt text encrypted by substitution
- Implemented the Metropolis algorithm in Python

Spam vs Ham Email Classification
- With Scikit-learn, used multiple linear regression, feature engineering, logistic regression, and cross validation to classify emails as spam or not spam
- accuracy of 95%

Genre Classification
- Predicted genre of movie from screenplay text, books from literary text, and songs from lyrics
- Utilized k-nearest neighbors algorithm with Euclidean distance of frequency of word features

NYC Taxi Ride Predictions
- Used SQL to explore NYC taxi dataset, visualizing trip patterns and collision frequencies
- Fit regression model with regularization to predict duration of taxi rides - within 3 minutes on all trips

Corona Borealis Galaxy Investigation
- Bootstrapped to estimate distribution of galaxy velocities in a region of the Corona Borealis
- Tested hypothesis of whether velocities of galaxies followed a multimodal distribution for the existence of voids & superclusters

Trump’s Tweets Evaluation
- Performed EDA and analyzed Trump’s tweets using the Twitter API
- Discovered favorite times to tweet, phone preferences, as well as the overall sentiment trends over time in relation to election season

USPS Sorting Optimization
- Worked with the Oakland Processing and Distribution Center, one of USPS’s package depots in the Bay Area, on processing and sorting of packages
- Increased package distribution efficiency by optimizing number of workers per terminal using SIMIO simulation software