Ethan Lam

lamethan204@gmail.com · (714) 920-2105· linkedin.com/in/ethan-lam-542312204 · github.com/ethanlam101

EDUCATION

University of California, San Diego - Halıcıoğlu Data Science Institute   San Diego, CA

Bachelor of Science, Data Science | Minor in Machine Learning   Expected Graduation: 2027

Provost Honors, Data Science Student Society, Vietnamese Student Association   GPA: 3.55

PROJECT EXPERIENCE

Severe Power Outages Exploratory Data Analysis & Prediction Model   December 2024

GitHub Link | Website Link

- Conducted a comprehensive data analysis of a public data set consisting of severe power outages in the United States using **statistical analysis** and **data visualizations** to identify data patterns and relationships.
- Assessed the missingness mechanisms of several features and performed **permutation tests** to verify dependency.
- Developed **machine learning pipelines** using **scikit-learn** to implement a **RandomForestClassifie**r with optimal hyperparameters using **GridSearchCV** for a total of 9 features based on climate, electricity usage, and more.
- Tested model accuracy/precision/recall by using K-Fold cross validation, yielding a **70% precision rate**.
- Deployed the entire project and its findings on Github pages to share insights and showcase skills.

N-Gram Language Model   November 2024

- Prepared a corpus of literary text by **web scraping** a public eLibrary using **RegEx** and **BeautifulSoup**.
- Built a **Uniform** and **Unigram Language Model**, leveraging the corpus to produce random sentences.
- Designed and implemented a N-Gram Language Model incorporating conditional probabilities of word sequences to generate coherent and meaningful text.
- Trained model on *The Complete Works of Shakespeare*, successfully and efficiently generating a **100-word passage in under 30 seconds**.

Geological Dinosaur Fossil Findings Visualizer   April 2024

- Worked with a team of 4 coding beginners to **clean, analyze, and visualiz**e a raw dataset of over a million entries.
- Discovered trends using **Pandas** between variables such as diet, geological fossil location, and aggression.
- Visualized carnivorous fossil findings using **density heatmaps** to give insight into how likely you would be eaten by a dinosaur based upon your current location.

TECHNICAL SKILLS

Technical Proficiency: Python, SQL
Data Science Tools: Machine learning, EDA, EDL, Pandas, Numpy, scikit-learn, Matplotlib, Plotly

AWARDS & CERTIFICATIONS

2nd Place CyberPatriots XIV, *CyberSecurity, Linux, Cisco*   January 2023

- Achieved **2nd place** out of a pool of 1000+ teams from different schools across California
- Linux points awarded for securing a virtual operating system from cybersecurity vulnerabilities including potential malware, phishing, and system misconfigurations

Coursera Online Academy   March 2024

Certified in Supervised Machine Learning: Regression and Classification