Yu-Chieh Chen

(626) 628-5058 | yc4015@columbia.edu | github.com/elvychen

EDUCATION

Columbia University — New York, NY
M.S. in Data Science — Expected Dec 2022

University of California San Diego — La Jolla, CA
B.S. in Data Science; Minored in Cognitive Science & Economics, GPA 3.83/4.0 — Jun 2021

SKILLS

Programming	Python, Java, SQL, JavaScript, HTML/ CSS, R, MATLAB
Packages	Numpy, Pandas, Scikit-learn, BeautifulSoup, PySpark, Dask, TensorFlow
Database	SQLite, PostgreSQL, Redis, MongoDB, Neo4J, Cassandra
Data Visualization	Matplotlib, HighCharts, Tableau

RESEARCH EXPERIENCES/INTERNSHIP

Halıcıoğlu Data Science Institute — La Jolla, CA
Teaching Assistant — Apr 2019 – Jun 2021

- Tutored for 4 different classes (data structure/advanced machine learning/data science) with class size up to 250 and led more than 15 tutors each quarter
- Held weekly office hours to answer questions, wrote AutoGrader scripts in Python for grading

Institute of Information Science, Academia Sinica — Taipei, Taiwan
Research Intern (Supervised by Prof. Meng Chang Chen)

Fake News Identification with CNN models — Jul – Sep 2020

- Optimized CNN model to process dataset (65,639 headlines) with and received 93.99% accuracy on testing data (4,462 headlines)
- Conducted exploratory data analysis to compare content farm headline with national news agency headline, found out three features (wording embedding, sentiment score, parts-of-speech) and applied bigram for more sentiment scores
- Scrapped and cleaned 33,000+ documents as training data for tfidfVectorizer and Multinomial Naïve Bayes Classifier to categorize news, achieved 0.92 F1-score
- Published paper, "Headline Diagnosis: Manipulation of Content Farm Headlines," as first author in Taiwan Academic Network Conference (TANET) 2020

Prediction of Air Quality Based on K-Nearest Neighbor (KNN) — Jul – Aug 2019

- Conducted KNN algorithm to select time and space-related monitoring stations, trained and verified KNN models to predict PM2.5 values and monitor air quality and implemented data visualization

Deloitte Touche Tohmatsu Limited — Taipei, Taiwan
Audit Department Intern — Jul 2018

- Specialized in audit software, EMS, to update financial statement
- Audited and analyzed financial statements of clients and changes over quarters

PROJECTS

Malware Detection — Oct 2020 – Mar 2021

- Designed a database (1.7GB) to store API information from 5,522 different apps (85+GB) in 3rd Normal Form
- Implemented new features and HinDroid kernels with SVM model, and received 99.17% accuracy and 0.9919 F1 score on testing data (362 apps)

Data Visualization of Suicide Rate — Nov – Dec 2019

- Analyzed highest suicide rate among different age group between 2005 to 2015
- Performed results by synchronized visualizations with HighCharts, JavaScript, HTML/CSS, and Python

Unsupervised Learning on Read and Unread Prediction — Nov 2019

- Predicted user read/unread behavior based on user book lists
- Constructed recommendation model with Popularity Ratio, Jaccard Similarity, Cosine Similarity, and Pearson Correlation to select optimal decision rule
- Accomplished top 10% ranking in class with an 74% accuracy score (about 380 undergraduate and 350 graduate students)

Real-time Face Recognition — Mar – Jun 2019

- Created local database and pre-store features for each face for reducing time consumption
- Extracted facial landmark with Dlib library and applied supervised learning to predict facial emotion
- Achieved highest validation accuracy of 89.43% with logistic regression