

# Edwin Huang San Diego, California

858-900-5207 | [Email](#) | [LinkedIn](#) | [GitHub](#)

## EDUCATION

---

### Bachelors of Science, Data Science

Sep 2017- June 2021

University of California, San Diego / Halicioglu Data Science Institute

- Relevant Coursework: Linear Algebra, Optimization, Statistics, Data Management, Web Mining, Deep Learning, PDE
- Minor in Mathematics, Provost Honors

## WORK EXPERIENCE

---

### Business Analyst Intern @ Applied Materials – Taiwan (<https://cutt.ly/zkEd7Wv>)

July 2020 – August 2020

- **File Parsing:** Parse raw data files (xlsx, txt, pdf) and transform them into full reports using PowerPoint.
- **RegEx:** Used Regular Expressions to clean and parse large text files distributed every week and extract information needed.
- **Automation:** Automated manual Excel tasks and improved efficiency of my department by around 4 hours combined.
- **Data Structures:** Created a program to group together similar company names in a data file of >100K rows using a generalized suffix tree. Microsoft's built in dictionary is used to check for typos and mark those words.
- Worked with people from different departments to host an event of >200 people. Helped with planning and logistics.
- Presented all my work to supervisors and helped explain my work to the IT department for maintainability.
- Volunteered as Intern leader. Helped HR with organizing small intern gatherings and check-ins with interns.

### Financial Analyst Intern @ Kingston Technologies – Fountain Valley, CA

July 2019 – September 2019

- **Web Scraping:** Scraped Marketwatch for financial information to determine competition's valuations.
- **Data Visualization:** Automated data to presentation pipeline to make sure reports include most recent data.
- **Design:** Developed a sales tool to track shipments and quantities required for customers.
- **SQL:** Wrote custom SQL to query data from IBM Cognos.
- **Automation:** Automated several Excel tasks to improve efficiency of the department.

## PROJECTS

---

### Database system for valuation of competitors

- Created a system based in Excel that automatically pulls data from Marketwatch containing companies' quarterly financial reports. Cleans and aggregates them, and automatically exports visualizations to PowerPoint.

### [Senior Capstone Project: CocoDroid Malware Detection Model](#)

- Created a model to predict malicious APKs by representing APKs using Control Flow Graphs (CFG). Did extensive research on graphs, their representations, and some ways to use them to classify such as Graph Neural Networks and common graphs.
- Used metapath2vec, doc2vec, and word2vec to build feature vectors from CFGs which are used in a model pipeline with RF, KNN, and SVM for classification.
- Achieved F1-Scores, Precision, and Recall up to 0.99, training on 8435 malicious apps and 802 benign apps.

### [Deep Convolutional Networks for Image Classification](#)

- Implemented deep convolutional networks for image classification. Used PyTorch for our framework. Used a dataset with images of birds.
- Implemented custom dataset loaders, CNN models, and experimented with transfer learning.
- Visualized weight and feature maps of trained models.
- When all layers are trainable, training accuracies was 93%.

### [Multi-Layer Neural Networks](#)

- Implemented a multi-layer neural network including back-propagation, momentum, regularization, activation functions, and network topologies. Trained model on the multi-classification SVHN dataset problem.
- Examined how learning rate, momentum, early stopping and L2 regularization impact the performance of our model.
- Achieved a test accuracy of 75% at first, and after hyperparameter tuning, achieved test accuracy of 81%

### [Classic Machine Learning Algorithms](#)

- Experimented with and implemented Logistic Regression and Softmax Regression algorithms. Performed classification of fashion items from the Fashion-MNIST data set.
- Applied Principle Component Analysis on the feature vectors, along with hyperparameter tuning on batch sizes and learning rates
- For logistic regression, we achieved a 99.9% accuracy on the test set. For Softmax Regression: 82.1% accuracy on the test set.

## Skills

---

- Programming: Python, R, SQL, MatLab, Java, HTML, CSS, VBA
- Libraries and Tools: PyTorch, TensorFlow, Keras, NumPy, Pandas, Scikit-learn, Matplotlib, Seaborn, MySQL, Spark, AWS, Git, Docker, Jupyter Notebook, Tableau, Latex, SSH, Microsoft Office, Networkx, Stellargraph, PowerBI.

