

# Jameson Thai

408-520-8203 | jameson.k.thai@gmail.com | <https://jamesonthai.github.io>

## EDUCATION

University of Southern California Dec 2020  
Master of Science in Computer Science - Data Science

San Jose State University May 2018  
Bachelor of Science in Computer Science, Minor in Mathematics - *Cum Laude (GPA 3.62)*

## SKILLS

**Programming Languages:** Python, R, SQL, NoSQL, Java, JavaScript

**Working Knowledge:** AWS, Hadoop, Keras, Linux, Numpy, Pandas, Plotly, PyTorch, Scikit-Learn, Spark, Tableau, TensorFlow

## EXPERIENCE

**Course Producer - University of Southern California** Jan 2020–Dec 2020

*Technologies Used: Python, Pandas, Numpy, AWS, GCP, PyTorch, TensorFlow, Keras, Git, Linux*

- Guided over 93 students to obtain practical experience in applying machine learning to video games.
- Co-managed 36 student projects over 2 semesters; secured and setup GCP and AWS environments for 18 teams.
- Surveyed students on staff objectives for research paper; 78% gained applicable practical experience from the course.

**Directed Research Developer - University of Southern California** Sep 2019–Dec 2019

*Technologies Used: Python, Numpy, MySQL, Plotly, Scikit-Learn, TensorFlow, JavaScript, Vagrant, Docker, Git, React*

- Restructured web app for analyzing and comparing multiple software systems development histories.
- Incorporated 2 Plotly features for visualizing how and to what extent developers influence software quality.
- Resolved 5 issues on Plotly displays resulting in a clearer visualization of history analysis by queried elements.

**Data Science Intern - Warner Bros. Entertainment** Jun 2019–Aug 2019

*Technologies Used: Python, Pandas, Numpy, Scikit-Learn, Plotly, Matplotlib, TensorFlow, Keras, AWS, DCU API, Git*

- Identified Batman with 80% accuracy using a Python image recognition model trained on AWS.
- Examined over 30,000 DC images by designing Python ETL pipeline for automated collection and augmenting.
- Reduced training time for a generalized model by 33% using transfer learning.
- Demoed findings in presentation to WB executives and data team by testing 2 DC characters with 74% accuracy.

**Web Developer - San Jose State University** Sep 2017–May 2018

*Technologies Used: Python, SQL, Plotly, JavaScript, Google Scripts, Git, PHP*

- Accelerated task completion in over 60 yearly projects by overhauling ticketing system in Google Scripts.
- Improved existing SQL queries leading to easier storage and faster retrieval of user data.
- Visualized attendance and subscription data in over 30 performances into actionable insights.

## PROJECTS

**Wildfires – Predicting Statistical Causes** Oct 2020–Present

Project Repository: <https://www.kaggle.com/jamesonthai/predicting-statistical-causes-of-historic-wildfire>

*Technologies Used: Python, Pandas, Plotly, Scikit-Learn, SQL, JSON, Random Forests, XGBoost, SVM, Linux*

- Predicted statistical causes of historical wildfires in the United States from 23 years of recorded SQL data.
- Achieved 80% accuracy in XGBoost model outperforming SVMs and Random Forests by 20%.
- Narrated time series analysis of 13 statistical causes of wildfires in 3221 US counties with Plotly and choropleths.

**Los Angeles Neighborhood Score** Sep 2020–Dec 2020

Project Repository: [https://github.com/JamesonThai/DSCI551/tree/master/DSCI551\\_Project\\_LA](https://github.com/JamesonThai/DSCI551/tree/master/DSCI551_Project_LA)

*Technologies Used: Python, Pandas, SQL, Plotly, MapReduce, Spark, Firebase, AWS, Flask, Rest, Linux*

- Co-led a team of 3 students to build an interactive LA neighborhood infographic.
- Aggregated 3 different time series datasets, home value, crime reports, and ACT scores, into SQL database.
- Analyzed over 70 million data points; clustered summaries by LA geospatial data into Firebase using Spark.
- Charted 97 LA neighborhoods mapped to Plotly summaries and choropleth in an interactive web dashboard.