**Nealson Setiawan**

Los Angeles, CA | 747-272-5097 | [nealsons@usc.edu](mailto:nealsons@usc.edu) | [www.linkedin.com/in/nealson‑setiawan‑4a1700252](https://www.linkedin.com/in/nealson-setiawan-4a1700252) | <https://github.com/nealsonS>

**education**

**University of Southern California** Los Angeles, CA

**M.S, Masters of Science, Applied Data Science** August 2023 - May 2025

* Relevant course work: Apache Spark, Hadoop Mapreduce, Machine Learning, Keras/Tensorflow, SQL, MongoDB, DynamoDB, Unix, AWS EC2, Data Mining with Spark (ongoing), LLM Modelling (ongoing), Recommender Systems (ongoing), Scala (ongoing)

**University of California Santa Barbara** Santa Barbara, CA

**B.S, Bachelors of Science, Statistics & Data Science** September 2021 - June 2023

* Relevant course work: Statistical Machine Learning, Statistical Data Science, Time Series Analysis, Time Series Forecasting, Data Visualization, Algorithms and Data Structures, Stochastic Processes, Probability Theory, Regression Analysis, Design of Statistical Experiments, Intermediate Python, R, SQL, SAS

**experience**

**Evidation Health | University of California, Santa Barbara** Santa Barbara, CA

**Data Science Researcher** January 2023 - June 2023

* Analyzed patients' behavior to respiratory illness contraction by performing dimension reduction on health wearables data with 10+ features into 2 dimensions leveraging the t-SNE algorithm, and discovered 2-3 distinct patterns in participants' responses
* Scaled, centered, and binned time-series quantitative data; analyzed dimension reduced processed data by employing seaborn and plotly graphs and animations to perform explanatory data analysis
* Managed logs and records of all of 20+ weekly meetings with Evidation data science team to lead team of 5 members with action items and issues to be addressed for upcoming meetings

**projects**

**Developing a Recommender System | Spark, MapReduce** January 2023 - Present

* Implemented efficient SON algorithm in PySpark and applied to Kaggle dataset to find all combinations of frequent item sets
* Explored and implemented custom partitioning method to facilitate shuffling process in large dataset for MapReduce job
* Learning foundations of Recommender Systems like Locality-Sensitive Functions, content-based recommendations, and collaborative filtering to be implemented by May 2024

**CatDB - NoSQL Database System with query language | Python** August 2023 - December 2023

* Implemented a database management system manipulated chunks of 200+ MB of data with functions for scanning, external sorting, joining, filtering, grouping, aggregation, and CRUD operations
* Created a custom cat-like SQL and MongoDB inspired query language and a query execution engine to manipulate 200+ MB of data with user-defined limited memory usage to be used through a command line-interface

**Classification of Frost in NASA images | Tensorflow** November 2023 - December 2023

* Achieved a test accuracy of 93% and test F1 of 94.8% after creating data pipeline to perform transfer learning on EfficientNetB0, ResNet50, VGG16 and CNN on satellite images supplied by NASA for class final deep learning project
* Minimized effects of overfitting of CNN by applying 6 methods: L1 regularization, dropout layers, early stopping, batch normalization, data augmentation, and minimizing learning rate
* Augmented 2GB+ images on Keras to induce variability by randomly cropping, flipping, rotating, translation, brightening images

**Sentimental Analysis on Amazon Product Reviews | NLTK, Scikit-learn** January 2023 - March 2023

* Achieved a cross-validated testing ROC of 88.69% and test ROC of 80.37% after fitting 4 classification models such as random forest and support vector machines on 4000 Amazon product reviews
* Deployed 3 word embedding techniques using NLTK such as Word2Vec and TF-IDF after text preprocessing steps such as lemmatization and removal of stopwords Amazon product reviews to prepare textual data for classification
* Illustrated underlying results of a Word2Vec model by visualizing 200+ 2-dimensional word vectors leveraging the t-SNE algorithm

**Monthly Forecasting on Indonesian Earthquake Data | R** January 2023 - March 2023

* Deployed SARIMA and State Space models on monthly Indonesian earthquake data from 2008-2022 to forecast magnitude of earthquakes 5 months ahead
* Estimated SARIMA and State Space model's parameters by evaluating 3+ SARIMA models' autocorrelation function, normality and standardized residuals, and kalman filtering and smoothing of a State Space model