Juan Andreas

joowahn97@gmail.com
https://github.com/juanandreas
415-637-8338

SF Bay Area & Silicon Valley
https://www.linkedin.com/in/jandreasucsc/

Education

Bachelor of Science, Honors in Computer Science
University of California, Santa Cruz 2015 - 2019

Certification

Google Cloud Certified Professional Data Engineer (In Training)

Work Experience

Accolade, Inc. Jan. 2019 - Jun. 2019
Student Data Scientist - Modular Data Pipeline

- Extracted raw data stored on **Amazon S3** to transform and aggregate data for Machine Learning
- Efficiently processed large datasets by leveraging **PySpark**'s distributed functions for ETL jobs
- Trained various ML models using **Sci-Kit Learn** on historical client data, to predict Accolade's client traffic
- Established and monitored sprint requirements by coordinating team and communicated closely with Accolade
- Provided Accolade with better predictions for their staffing to better match with their client traffic in the future
- Built an automated tool for Accolade's Data Scientists to reduce the overhead of processing large raw datasets

Relevant Experience

Security Point Jan. 2019
Grand Prize winner for Microsoft's sponsored project at Cruzhacks 2019 (2-day Hackathon at UCSC)

- Extracted relevant information about security breaches using **Natural Language Processing**
- Enabled cloud function API queries by storing data on **GCP** cloud database
- Delivered security breach data using Microsoft's **Azure Maps** to visualize data from JSON object
- Provided a cluster map data visualization to gain insights on areas and businesses that are prone to breaches

Distributed Systems Project Sep. 2018 - Dec. 2018
Developed and studied distributed systems

- Achieved a scalable and fault tolerant distributed system by implementing data recovery and automatically-rebalancing shards.
- Simulated multiple servers in a distributed system by launching **docker** instances running in parallel
- Deployed **Available and Partition Tolerant distributed system** in Python by implementing a **KVS** and **RESTful API**
- Improved clarity between server-communication to maintain causal consistency using **Object-oriented programming** to organize data payloads

Natural Language Processing Mar. 2018 - Jun. 2018
Practiced various Natural Language Processing technique

- Achieved around 50-54% accuracy on a Q&A system using wordnet and dependency parse trees
- Utilized the **NLTK** library to process text from various corpora to discover relevant features needed to answer various types of questions

Additional Skills

- Scrum, Agile, Git, Windows, MAC OS, Linux, Unix
- SQL, PSQL, Jupyter Notebooks, Java
- Speaker of Indonesian and Japanese

Relevant Coursework: Distributed Systems, Software Engineering, Introduction to Algorithm Analysis, Introduction to Artificial Intelligence, Machine Learning, Natural Language Processing