

Adrian Lam

ayplam@gmail.com • <https://github.com/ayplam> • 650-397-1526

PROFESSIONAL EXPERIENCE

Change Healthcare

Data Scientist, AI Project Lead

Data Scientist

Emeryville, CA

Jan 2018 - Current

May 2017 – Jan 2018

Responsible for leading a team of data scientists to build end-to-end products from conception to deployment that have direct financial impact. Create and manage product roadmap for features and improvements. Work closely with members to guide both modeling and engineering efforts. Coach individuals to accelerate personal growth.

Team Responsibilities

- Productionized data science pipeline to generate audit predictions from medical record PDFs, estimated > \$700k annual impact
- Work with business units to collect requirements, evaluate feasibility of proposals and determine financial impact of models
- Plan weekly data science and engineering stories, manage progress to remove any blockers and to ensure timely delivery of features
- Pioneered development of code repositories used across the team for large scale processing of text, deep learning pipelines, and cloud utilities
- Heavily involved with standardizing DevOps practices across the AI team for code quality, CI/CD, testing, and deployment of models and pipelines

Medical Record Audit Automation

- Deployed language model to automate categorization of medical record pages for auditors
- Trained custom NER models for clinical lab value extraction to validate diagnosis codes
- Built LSTM model to assess sepsis claims by isolating relevant medical text using external datasets and hyperparameter optimization of network architecture

SAP Labs

Data Scientist, Developer

Data Scientist, Intern

Palo Alto, CA

April 2016 – April 2017

Feb 2016 – April 2016

Embed machine learning capabilities into IoT proof of concept projects, including chatbot intent classification, entity extraction, and facial detection. Mentored data science interns in machine learning, model evaluation, and good practices for software engineering.

Digital Assistant

- Successful Beta and Beta+ delivery of chatbot application with 20+ customer use cases to simplify access to CRM/HR information
- Developed complete NLP intent recognition and parameter extraction chatbot system using NLP4J by novel dependency tree comparison algorithm
- Integrated scoring system to rate intent recognition by passing engineered features into LibSVM
- Driving force behind innovation of multiple secondary features by guiding interns to successfully create and integrate TF-IDF recommendations, Naïve-Bayes autocorrect, and fuzzy matched entity suggestions

Intelevator as a Service/Leapfrog for IoT

- Worked with international team to build multiple end-to-end scenarios to support high volume IoT data by combining data generation, Flume, Kafka, Spark Streaming, SAP HANA Vora, and SAP HANA
- Major contributor to all phases of development including architectural design, data modelling, application design, testing and deployment
- Responsible for creating spark streaming applications, hourly batch aggregation jobs, incorporating predictive analytics and integrating all components into single application
- Optimized and benchmarked SAP HANA Vora aggregations to determine bottlenecks and maximum throughput of end-to-end system (~4M records/min on 7 node cluster)

The Data Incubator

Data Science Fellow

San Mateo, CA

June 2015 – Aug 2015

Generic Drug Release Modeling

- Munged FDA drug database into SQL to match brand name drugs to generic name counterparts
- Regression performed to predict generic drug launch date from vectorized pharmaceutical class and brand name release date
- Time-series analysis and matrix factorization used to find seasonality trends
- Results can improve monthly sales forecast and predict upcoming product introductions

Predictive Yelp Modeling

- Developed web scraper to automatically harvest Yelp reviews from user-specified parameters
- Bag-of-words model allowed clear differentiation of five-star from one-star reviews illuminated unique characteristics only found in quality restaurants
- Restaurant attributes to build K-Nearest Neighbors model and predict restaurant ratings with cross-validated ridge regression

EDUCATION

Nanodegree, Deep Reinforcement Learning

Coursera

December 2018

Ph.D., Bioengineering

Georgia Institute of Technology

Minor: Digital Signal and Image Processing

Atlanta, GA

December 2015

B.S., Biomedical Engineering

Dept of Biomedical Engineering, University of California, Davis

Davis, CA

June 2010

SKILLS

Code: Python, Java, Scala, SQL/NoSQL, Bash, C++, MATLAB, Lua

ML/NLP: spaCy/Prodigy, NLTK, TensorFlow, Keras, PyTorch, H₂O, SparkML

DevOps: Git/Gitlab, Docker, Jenkins, CI/CD (pytest, JUnit), Artifactory, Nexus

Apache: Spark, Airflow, Flume, Hive, Hadoop, Kafka, Zeppelin

Web: HTML, CSS, Javascript, Mako, Node.js, MongoDB, jQuery, D3