

PATRICK DE GUZMAN

Email: patrick.de.guzman.2020@anderson.ucla.edu | Phone: (408) 775 – 4145 | Los Angeles, CA
Personal Website: patrickdeguzman.me | LinkedIn: [linkedin.com/in/patrickjdeguzman](https://www.linkedin.com/in/patrickjdeguzman) | Github: github.com/pdeguzman96

EDUCATION

UCLA, Anderson School of Management	December 2020
Master of Science in Business Analytics (MSBA)	GPA: 3.92
Santa Clara University	June 2018
Bachelor of Science in Commerce, Accounting & Information Systems	GPA: 3.86
Udacity Nanodegrees	
Intro to Machine Learning Data Scientist Data Visualization	

TECHNICAL SKILLS

Languages: Python, SQL, R, Shell, JavaScript (NodeJS), HTML, CSS

Data Science & ML: Deep Learning (Keras & PyTorch), Recommendations, Topic Modeling, Ensemble Models

Domain Knowledge: Spark, Causal Inference (Quasi-Experiments), Linear Optimization, Clustering, Hypothesis Testing

Other: Docker, API Development, CI/CD Pipelines, Distributed Processing, Web Servers, Git, Tableau, PowerBI, NoSQL

AWS Certifications: Solutions Architect Associate, Developer Associate, Certified Cloud Practitioner

PROFESSIONAL EXPERIENCE

Intuit	Mountain View, CA
<i>Marketing Analytics Intern</i>	June 2020 – Present

- Designing and developing a geo-based experimentation Python package that will be able to split geos into treatment and control groups using a matched markets approach and run pre/post experiment analyses
- Setting up automated Hive pipeline to transform display and social ad data for audience reporting and analytics
- Building a dashboard that facilitates point-in-time and time series comparisons between different types of audiences based on spend, performance, engagement, and other KPIs
- Collecting and synthesizing partner marketing agencies' audience experimentation strategies and past learnings to facilitate planning overall business' FY21 audience strategy and experimentation roadmap

Datasleek	Los Angeles, CA
<i>Backend API Developer</i>	March 2020 – June 2020

- Built Flask REST APIs that communicate with MemSQL, MySQL, and Amazon S3 to parse and process invoice image data from Amazon Textract and to facilitate CRUD operations from a web UI
- Set up a CI/CD pipeline with CircleCI and Github Actions to deploy codebase changes using Docker images to a serverless staging and production environment hosted on AWS Fargate
- Launched a MemSQL Pipeline to automatically ingest preprocessed image data into raw staging tables
- Deployed and secured APIs using SSL for encryption and Auth0 for authentication behind a single load balancer that transfers requests to stage or production endpoints using host-based routing
- Developed a templating system that enables one preprocessing worker to parse unstructured invoice image data from any supplier then store clean results using a relational schema

PwC (PricewaterhouseCoopers)	San Francisco, CA
<i>Risk Assurance Experienced Associate</i>	August 2018 – May 2019

- Developed and deployed PowerBI data pipeline and dashboard to summarize project progress data
- Tested over 100 key controls and reports containing financially relevant data through statistical sampling and code review (in SQL, Python, Bash, ABAP, and Perl) to reduce risk of financial misrepresentation
- Established automated testing procedures with UiPath to automate workload, reducing about 200 hours of work per year

PERSONAL TECHNICAL PROJECTS

PySpark: Predicting Customer Churn	Distributed Processing, Cloud Computing
---	--

- Analyzed, cleaned, and prototyped a machine learning algorithm on a truncated user log dataset using PySpark locally
- Used AWS EMR to run pipeline on a larger dataset on four machines, improving test performance by 15%

Computer Vision App: Build-Your-Own CNN	Deep Learning, Computer Vision
--	---------------------------------------

- Command-line application to retrain pretrained ImageNet Convolutional Neural Network on any set of images for prediction

Webapp for Disaster Message Classifier	Text Preprocessing, Text Classification
---	--

- Text preprocessing pipeline using nltk & Sklearn implemented in Flask webapp to classify text messages

Predicting Credit Card Fraud (Kaggle)	Class Imbalance, Feature Engineering
--	---

- XGBoost tuned with hyperopt to predict whether transaction was fraudulent with .90 testing ROC AUC
- Dask to manipulate large dataset in chunks and featurerools to perform automated feature engineering

Recommendation Engine for IBM Watson Articles	Matrix Factorization, Recommendation Systems
--	---

- Combined rank-based, collaborative, and content-based filtering to create a recommendation system for articles
- Manually created a matrix factorization algorithm (based on FunkSVD) to predict whether users will read an article

NodeJS Movie CRUD App	NodeJS API Development, NoSQL
------------------------------	--------------------------------------

- Simple MongoDB-backed movie rental CRUD API for managing movies, genres, and rentals using the Express framework