

# Steve Shim

700 W 9<sup>th</sup> St. #1506

Los Angeles, CA 90015

[steveyshim@gmail.com](mailto:steveyshim@gmail.com) | (818)565-9182

## Work Experience

---

### PatientPop

*Data Scientist*

April 2019 – Present

- Create review topics from text posted on reputation platforms (i.e. Yelp, Google) using LDA and automate the process of topic tuning with reducing feature space and optimizing for coherence and perplexity.
- Build dashboards from disparate data sources (Salesforce, Google Sheets, Pendo) to show impact of market changes on sales metrics and how they translate into product performance metrics.
- Design and implement company's first A/B test experiment based on sensitivity testing stemming from the reputation drivers' analysis, resulting in saving \$1.3 million in ARR.
- Find the relationship between online reputation and appointment requests using an XGBoost regression model and SHAP to perform a drivers' analysis.
- Perform k-means clustering to create customer segments to apply generalization of customer value to all customers.
- Utilize asynchronous API calls to systems of records to ingest data to create value story and give management its first realistic view into value provided by PatientPop products.

### NEOGOV

*Data Scientist*

January 2018 – April 2019

- Combined NLP and machine learning techniques to classify existing job data into standardized job classifications by parsing and cleaning job descriptions and minimum qualifications.
- Collaborated with product team to provide data driven insights about customer behavior and feature adoption while communicating analysis to stakeholders.
- Generated dashboards and reports in Tableau and Microsoft Excel for product management and customer service teams used for viewing of customer data and key product metrics.
- Analyzed and produced data regarding hiring trends within the public sector to be used in company's publication.
- Built automation scripts in Python with a simple user interface that provide the marketing and sales teams scalable access to product data in order to reduce the influx of ad-hoc requests.
- Defined key metrics for product suite and use metrics to identify healthy usage of products while enhancing existing tracking to gather time series data to analyze customer behavior over time.
- Scrubbed existing product usage data in MSSQL and create link with other sources such as Mixpanel and Salesforce.
- Developed tool to assist marketing and sales with creating an ROI calculator to present at customer meetings.

*Data Science Intern*

July 2017 – December 2017

- Defined and extracted product feature usage, transformed into time series usage data, and loaded into separate product analytics usage SQL tables for future use.
- Developed customer churn model by creating clusters using k-means and supervised learning models based on product feature usage to predict customers at risk of cancellation to maximize customer retention.

## Education

---

California State University, Los Angeles

*M.S., Computer Science*

August 2017

## Technical Skills

---

- Languages: Python, SQL, Java, JavaScript, HTML, C, C++
- Tools: Sklearn, Pandas, Matplotlib, Tableau, NLTK, Mixpanel, Google Analytics, Salesforce, Git, Sagemaker
- Databases: MSSQL, Redshift, Postgres, MongoDB, MySQL