

Kankshat Patel

kankshatp@gmail.com < > [kankshat.github.io](https://github.com/kankshat) < > www.linkedin.com/in/kankshatpatel < > (510)-599-8727

EDUCATION

University of California Santa Barbara • B.S. in *Data Science and Statistics*.

Organizations: Data Science Club *Vice President*.

SKILLS

- *Languages:* Python, R, SQL, C++, Java, SAS.
 - *Machine Learning Models:* KNN, Logistic/Linear Regression, PCA, Decision Tree, Random Forests, TensorFlow.
 - *Elasticsearch:* Elasticsearch, Kibana, Logstash, ECK on Kubernetes.
 - *DevOps:* Ansible, Git, Jenkins.
-

PROFESSIONAL EXPERIENCE

HealthStream

Dec 2022 - Present

Data Analyst

- Evaluated and Cleaned Healthcare data to determine suitability and procedure to migrate to specific Verity Streams system.
- Performed complex Medicare data imports via bulk T-SQL query and window functions from various multi-source database systems while utilizing stored procedures.
- Monitored query performance and unoptimized indexes and created a guideline for acceptable query performance.
- Developed Python script using fuzzy logic to automate string matching for various healthcare client data.

UCSB Analytics Department

Jul 2021 – Jun 2022

Data Analytics Intern

- Integrated KNN, PCA, Decision Tree, and Random Forest models in R and Python to precisely predict the outcome of a Division-I batter's plate appearance with 80% success rate given various parameters.
- Engineered a data scraper in python using the BeautifulSoup module and regular expressions to extract and clean the data from the NCAA site transforming it into csv format, optimizing the time retrieval by 75%.
- Developed dashboards through Kibana for various metrics and communicated findings with supervisor.

JP Technology

Jun 2021 – Sep 2021

Software Developer Intern

- Installed and Configured Elasticsearch cluster through Ansible script and re-indexed data into new Elasticsearch cluster with zero downtime.
 - Designed and implemented component templates and index templates, simplifying the process of making changes to index settings and mappings for future index adjustments.
 - Updated and installed a synonyms extension used by multiple index analyzers on Elastic Cloud Environment (ECE).
-

PROJECTS

Relationship Compatibility Data Science Model

Aug 2021

- Integrated ML algorithms with normalized and interpolated data to predict matches for speed dating with 85% accuracy.
- Created a project timeline regarding code for all team members to follow and stay organized.
- Built a classification model in Python using a k-means clustering method to predict finding a successful match.

COVID-19 Case Prediction Model [Formal Write Up](#)

Dec 2021

- Extracted relevant COVID-19 data from CovidCast API and tidied dataset to apply ML algorithms on time series analysis.
 - Integrated Linear Regression and PCA and found a 77% success rate in predicting the number of new COVID-19 cases in any given California county 7 days in advance.
-

RELEVANT COURSEWORK

Machine Learning and Algorithms || Statistical Machine Learning (Graduate) || Linear/Logistic Regression || Probability and Statistics || SAS Base Programming || Stochastic Processing || OOP (C++, Java) || Discrete Structures || Linear Algebra ||
