

# Kelsey Glenn

kelseyfglenn@gmail.com • (510)266-2120 • Newark, CA 94560

github.com/kelseyfglenn • linkedin.com/in/kfglenn

---

## SKILLS

---

**Programming Languages:** Python, SQL, Jupyter, HTML/CSS

**Libraries:** pandas, NumPy, SciPy, scikit-learn, NLTK, Flask

**Data Visualization:** Tableau, Matplotlib, Seaborn, Yellowbrick, Excel

**Machine Learning:**

- **Regression** (Linear, Logistic, Tree-based, Gradient Boosted)
- **Classification** (Logistic Regression, Naive Bayes, Tree/Random Forest, AdaBoost, XGBoost)
- **Dimensionality Reduction** (PCA, SVD) & **Clustering** (K-Means, DBScan)
- **NLP** (Vectorization, LSA/SVD/LDA, Sentiment analysis, word embeddings, neural language models)

## EXPERIENCE

---

### Metis

*Data Science Teaching Assistant*

Instructional aid for a 12-week immersive Data Science bootcamp.

**September, 2020 – Present**

*San Francisco, CA*

**June, 2020 - September, 2020**

*San Francisco, CA*

*Data Science Fellow*

Implemented end-to-end data pipelines from collection to productionization:

- Data wrangling, EDA, ML and modeling, visualization, production & presentation

### James Logan High School

*Lead Instructor, LD & Public Forum Debate*

▪ Lead debate coach and lecturer for a top-10 national program. Qualified multiple students to state and national championship tournaments.

▪ Implemented objective-based syllabi to educate students on advanced concepts in politics, philosophy, and economics.

▪ Provided guidance on qualitative and quantitative research skills, argumentative strategy and rhetorical techniques.

**June, 2018 - June, 2020**

*Union City, CA*

### VR Research

*Intern, Research Analysis*

▪ Performed due diligence research across domains including news and social media, government records requests, local property and tax documentation, campaign spending databases, and municipal budgets.

▪ Applied research findings to create actionable strategic dossiers for clients including public office candidates, special interest groups, and corporate entities. Composed reports with both local and Washington, D.C.-based research teams.

**January, 2016 - April, 2016**

*Oakland, CA*

## PROJECTS

---

### AniMaker: T.V. series synopsis generation with GPT-2

Novel text generator and user rating predictor using OpenAI's GPT-2 trained on Anime series plot synopses.

- Fine-tuned GPT-2 language model using TensorFlow in Google Colab
- Estimated user ratings via LSA semantic feature extraction and Random Forest Regression to curate generated samples
- Deployed web app for custom generations via Streamlit

### Clusta' Rhymes: Categorizing hip-hop styles through semantic and prosodic clustering

Clustering and recommender system for hip-hop artists based on lyrics.

- Created lyric corpus via Genius API and BeautifulSoup
- Performed semantic topic modeling via sklearn's NMF decomposition and engineered prosodic features via NLTK
- Applied K-Means clustering and generated distance-based artist recommendations
- Deployed the recommender app and visualization via Flask with embedded Tableau dashboard

### Bot or Not: Twitter bot detection

Gradient-boosted classifier to detect twitter bots through user metadata

- Achieved F1 score of .964 on test data using an XGBoost classification model
- Engineered additional user features including fictitious name detection
- Compared and optimized tree-based and gradient boosted models using sklearn's RandomizedSearchCV and GridSearchCV

## EDUCATION

---

University of California, Davis

*B.A., Linguistics*

**2013 - 2017**