

Cristobal Zamorano Astudillo

Software Engineer

📍 Berkeley, CA • U.S. Citizen 📞 310 980 9686

✉ cristobal.zamorano@berkeley.edu 🏠 www.cristobalza.com 🌐 in/cristobalza 🐙 cristobalza

EDUCATION

- **University of California, Berkeley** Berkeley, CA
B.A. of Computer Science, Data Science - Applied Mathematics & Modeling 2018 - 2021
 - **Major Coursework:** Advanced Machine Learning, Data Mining, Databases Design, Data Structures and Algorithms, Principles & Techniques of Data Science, Probability Theory

SKILLS

- **Languages:** Python, SQL, Java, R, C, C++, Spanish
- **Web-Development Technologies:** HTML, CSS, JavaScript, NodeJS, VueJS
- **Databases:** MongoDB, Snowflake, Postgres, Spark
- **SWE/DS Frameworks:** Flask, Django, Heroku, FastAPI, S3, Keras, Pytorch, Tensorflow, Selenium
- **Other:** Docker, Airflow, Kafka, Superset, Git

HIGHLIGHTED EXPERIENCES

- **Software Engineer Intern (Data Engineer) @ Intelligent Automation, Inc.** Rockville, MD
Signal Analysis & Control, Artificial Intelligence and Machine Learning Division Jun 2021 - Aug 2021
 - Stood up an internal Data Catalog for a medium size business's datasets to allow company wide sharing of their internal data using Amundsen, an open source Data Catalog.
 - Designed and developed a full stack web portal, Metis, using VueJS, FastAPI, Postgres, and Docker, for company's +100 engineers data ingestion and extraction.
 - Implemented and tested a real-time data pipeline to process semi-structured data by integrating 10 million raw records from Maryland highway data source using Kafka, Airflow, S3 and stored processed data in Postgres.
- **Data Scientist Research Intern @ Latinx Research Center at UCB** Berkeley, CA
Democracy + Media Lab Research Team Jan 2021 - May 2021
 - Built an end-to-end data mapping tool to visualize demographics of minority communities that were impacted by the 2020 elections with Professor Angela Marino and the rest of researchers team using econometrics data.
 - Performed topic modeling of politicians data using deep learning techniques (Speech2Vec, LDA, Tf-idf).
 - Attending weekly meetings to communicate progress, discuss challenges and troubleshooting of the project in the overall process with Professor and fellow researchers.

PROJECTS

- **Future Salary App** — <https://future-salary.herokuapp.com/> Python
 - Designed a salary predictor that only uses job descriptions in order to output salary ranges
 - Utilized SQL queries to retrieve data from more than 20,000 tech related jobs around the United States on an internal scrape
 - Implemented a word2vec Neural Network to rank text data scrapped from 20K+ different job descriptions.
 - Implemented my own Neural Network Architecture to benchmark other ML models. My Neural Network had the highest Accuracy metric of all the ML models with 85% on the test data.
- **Toxic Comments Detector App** — <https://toxic-text-detector.herokuapp.com> Python
 - Designed a multi-label classifier that reduces the inadvertent bias of a Machine Learning model to detect toxic comments.
 - Implemented an optimal feature engineering for vectorization of the text data and hypertuning parameter analysis.
 - After the model benchmark results, the final model had an overall of 90% F1 score against all toxic features.
- **Spotify Music Insights Project** — <http://www.cristobalza.com/> Python
 - Implemented the Spotify's API to obtain my music metrics to create an insight on my data.
 - Designed and implemented my own PCA algorithm using Eigenvectors from my own music data.
 - Designed and implemented my own ML recommender model that could compete against my "Discover Weekly" playlist made by Spotify.