

# Cristobal Zamorano Astudillo

## Software Engineer

• U.S. Citizen ☎ +1 310 980 9686

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

### EDUCATION

---

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

### SKILLS

---

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

### HIGHLIGHTED EXPERIENCE

---

- **Software Engineer Intern (Data Engineer) @ Intelligent Automation, Inc.** Rockville, MD  
*Signal Analysis & Control, Artificial Intelligence and Machine Learning Division* Jun 2021 - Aug 2021
  - Implemented 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*, to optimize processes that streamline and make internal operations easier for company's +100 engineers, using VueJS, FastAPI, Postgres, and Docker.
  - Implemented and tested a real-time data pipeline to process semi-structured data by integrating 10 million raw video records from a highway data source using Kafka, Airflow, S3 and stored processed data in Postgres.
- **Data Scientist Research Intern @ Latinx Research Center at UC Berkeley** 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 using econometric data, using JavaScript, HTML, Pandas.
  - Performed topic modeling of candidate data using deep learning techniques (Speech2Vec, Tf-idf), using Scikit-learn, PyTorch.
  - Attended weekly meetings to communicate progress and troubleshoot challenges in the overall process with faculty lead and fellow researchers.

### PROJECTS

---

- **Future Salary App** Click here: <https://future-salary.herokuapp.com/>
  - Designed a salary predictor that only uses job descriptions in order to output salary ranges using context-independent word embedding (Word2Vec, Sentence2Vec).
  - Implemented an automatized data pipeline, using Selenium and BeautifulSoup, to extract +20,000 tech related jobs around the United States from Indeed.com
  - Implemented and save a regression model with a 85% of accuracy to be deployed into a web app using Heroku.
- **Toxic Comments Detector App** Click here: <https://toxic-text-detector.herokuapp.com>
  - Designed a multi-label classifier that reduces the inadvertent bias of a Machine Learning model to detect toxic comments, using Keras and Scikit-learn, create a web app, using NodeJS, VueJS, and deploy using Heroku.
  - Implemented an optimal feature engineering for vectorization of the text data and hyper-tunning parameter analysis.
  - After benchmark results, the final model had an overall of 90% F1 score in multi label prediction results.
- **Spotify Music Insights Project** Click here: <http://www.cristobalza.com/>
  - Implemented a data pipeline using Spotify's API in Python (Spotipy) to explore, visualize, and run ML experiments using personal music data to understand mood during the pandemic quarantines.
  - Designed and implemented custom PCA algorithm using personal music data to create 3D plots of most important music metric features, using Numpy.
  - Designed and implemented an Machine Learning recommender that could compete against my "Discover Weekly" playlist made by Spotify as testing data, using Keras, Scikit-learn.