Cristobal Zamorano Astudillo

📞 310-980-9686 ✉ cristobalzamas@gmail.com 🔗 cristobalza 🐙 cristobalza 🌐 cristobalza.com 👤 US. Citizen

Education

University of California, Berkeley   Berkeley, CA | Dec 2021
B.A. Data Science, D.E.: Computer Science Modeling and Applied Mathematics
Coursework: Data Structures & Algorithms; Databases; Operative Systems; Data Mining; Advanced Machine Learning

Highlighted Experience

Intelligent Automation, Inc. | Data Engineer Intern   Rockville, MD | Jun 2021–Aug 2021

- Developed a full-stack intranet portal to optimize processes that streamline and make internal operations of +100 engineers more efficiently by reducing loading times by 20% between apps, using **VueJS**, **Python**, **FastAPI**, **Postgres**, and **Docker**.
- Setup a CI/CD pipeline, using **GitLab**, and exceeded requirements in each stage, which led to the intranet application being deployed to their internal network, using **Docker**.
- Implemented and tested a real-time data pipeline to process semi-structured and structured data by integrating +9,000 labeled video records from a highway data source using **Kafka**, and **Airflow**. Stored processed data using **AWS S3** to store video files and use **MongoDB** to store metadata.

Projects

ETL and Dashboard using Youtube Videos Project ↗   March 2022 - May 2022

- Implemented an AWS cloud-based safe Data Lake solution to aid in the development of rich data analytics while classifying data into several storage phases, such as raw, cleansed, and analytical.
- Built ETL data pipeline to acquire +200 video data using various methods to store raw data, clean, validate, and transform data into a query-worthy format and display KPIs using **AWS S3, Lambda** and **Spark**.
- Pushed transformed data into a dashboard using **AWS Quicksight** to allow Data Scientists and Data Analysts to run efficient queries, visualize, and build Machine Learning models or analyze data.

Future Salary Predictor App ↗   Jan 2021 - March 2021

- Designed a salary predictor to predict the salary, and give job seekers more information on location job choices. The final model had 71% accuracy and was deployed on **Heroku**.
- Implemented an automatized data pipeline to extract +20,000 tech-related jobs in major U.S. cities and information such as job title, company, city, job description, and salary range, using **Python**, **Pandas** and **Selenium**.
- Implemented models that combine text features and word embeddings with features such as Cost of living Index, Purchasing Power Index, Location to add more complexity to the models, using **Keras**, **Scikit-learn**, and **NLTK**.

Toxic Comments Detector App ↗   Oct 2020 - Jan 2021

- Designed a multi-label classifier that reduces the inadvertent bias of a Machine Learning model to detect toxic comments, using **Keras** and **Scikit-learn**.
- Created a web application on **VueJS** and **Flask**, converted the ML model into a byte stream to store it in a file/database, and maintain software state across sessions. Use **Chrome Dev Tools** and **Postman** for app testing.
- Achieved an overall 90% F1 score in multi-label prediction results in the testing series. Now, web app lives in the **Heroku** cloud platform runs on every desktop and phone/tablet with an Internet connection for free usage.

Related Experience

UC, Berkeley - Latinx Research Center | Data Scientist Intern   Berkeley, CA | 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 according to econometric data, using **Pandas**, **Angular**, and **Postgres**.

Skills

Languages: Python, JavaScript, Java, SQL, HTML, CSS
Other: Postrges, Mongo, MySQL, VueJS, Docker, Airflow, Kafka, Git, GitLab, Flask, FastAPI