

Juan Acosta

New York, NY | 347-622-4391 | jacosta285@gmail.com | [Github](#) | [Medium](#) | [LinkedIn](#)

DATA SCIENTIST

Experience in data acquisition and data modeling, statistical analysis, machine learning, deep learning, and NLP. With a background in music and teaching, I bring strong collaboration skills, as well as highly analytical problem-solving skills.

TECHNICAL SKILLS

Python, SQL, numpy, pandas, scikit-learn, tensorflow, algorithms and data structures, AWS, Javascript, HTML, CSS.

TECHNICAL PROJECTS

RAG Chatbot

A real estate chatbot designed to provide property information and connect customers with agents.

- Utilized the Twilio API to integrate WhatsApp messaging, enabling direct communication with customers.
- Deployed a Flask server to manage interactions between customers and an LLM, using LangChain for dynamic response generation.
- Fine-tuned a BERT-based intent classifier to accurately identify when customers want to contact a real estate agent.
- Developed a vector database to efficiently store and retrieve property information, optimizing data access for the LLM.
- Implemented context management to handle multi-step interactions and maintain conversation flow.

Distance Matrix

Create distance matrices and retrieve information from Google Maps API

- Set up an API to retrieve information in a quick and precise way
- Formatted the information to properly work with Python and Pandas
- Created functions to help retrieve, format, and store the information properly
- Included functionality to use it in SQL

Urban Sound Classifier

Use machine learning techniques to classify different urban sounds

- Obtained audio files from Kaggle's UrbanSound8K
- Extracted relevant features of each audio signal using Librosa (MFCCs, Spectral Centroid, Spectral Flux, etc.)
- Tested multiple ML models for prediction (Logistic Regression, XGBoost, CNN)
- XGBoost obtained the highest accuracy at ~70%

Speech Recognition

Create a neural network that recognizes the voices of 10 different individuals

- Extracted and preprocessed Mel-Spectrograms and MFCCs from audio clips
- Performed EDA to determine the most valuable features of the model
- With a focus on MFCCs, tested different neural network architectures to determine the best approach
- Network with Bidirectional LSTM layers yielded the best results with a test accuracy of 85%

EDUCATION

Flatiron School, New York, NY Immersive Data Science Bootcamp program	11/2023 - 03/2024
SUNY Purchase, Purchase, NY Master of Music in Jazz Performance	08/2015 - 05/2017
Webster University, St. Louis, MO Bachelor of Music in Jazz Studies	08/2010 - 05/2014

EXPERIENCE

Barista, Starbucks , New York City, NY	08/2022 - Present
---	-------------------

- Customer service, and beverage preparation

Music Teacher, **Colegio Menor**, Quito, Ecuador

01/2019 - 02/2020

- Restructured music theory and audio technology curriculum
- Wrote material for teaching scales, chords, and rhythm
- Taught music technology skills like recording, mixing, sound editing, and sampling

Freelance musician and instructor, **Self-employed**, New York City, NY

08/2015 - 05/2018

- Performed in multiple styles in New York City and St. Louis
- Taught private music lessons to children and adults