

JAY ALEXANDER TREVINO

Student and Research Intern @ Columbia University

@ jay.trevino@columbia.edu

🔗 <https://github.com/jat2211>

📍 <https://jat2211.github.io/home/>

EDUCATION

BA in Computer Science

Columbia University Class of 2023

- **Computer Science Coursework:** Machine Learning*, Natural Language Processing*, Introduction to Databases*, Applied Data Science*, Data Structures and Algorithms
- **Mathematics Coursework:** Probability Theory*, Advanced Linear Algebra*, Multivariable Calculus, Discrete Mathematics
- *indicates graduate level coursework

EXPERIENCE

NLP Research Assistant

Columbia Data Science Institute and Rights CoLab

📅 10/2021 - Ongoing

- Used **Google Big Query** to access company 10-K data and break down data by industry and manipulate data using **Pandas**
- Parsed for practice & risk terms considered to be financial materiality to find co-occurrences using **natural language processing** techniques
- Used **Excel** to validate a sample of the results and update regex terms
- Created **box plots** and graphs that assess the industries holding the most flagged companies and the change over fiscal years from 2013 to 2021

Data Engineering Intern

CrowdConnect

📅 09/2021 - Ongoing

- Designed the collection strategy of student engagement data in lectures that provides real-time feedback to professors using **SQL** and **MongoDB**
- Designed a **dashboard** of graphs that display percent engagement over time for each unique room IDs using data from company database in **Tableau**
- **Fixed bugs** that caused application to crash in network socket
- Engineered a connection between a BI connector and a database

Summer 2021 Astronomy REU Intern

University of Texas at Austin

📅 06/2021 - 08/2021

- Modeled large galactic datasets using **Pandas** and CLOUDY simulation software to optimize the most probable set of parameters for a galaxy
- Created a dictionary of 2-dimensional arrays to represent the data in order to constrain the parameters before analysis in **Python**
- Modeled input files by stepping through a range of values defined by a set of constraints provided by another research team member
- Created a **bash** script to run 108,000+ simulation files on an **HPC cluster**
- Active communication with team of researchers in order to **co-author** a published paper and present research at AAS conference in June 2022

Data Analyst Research Assistant

Columbia University Cosmology

📅 08/2020 - 02/2021

- Cleaned CMB frequency map datasets by flattening the maps and built matrices according to sum of the products of each map
- Used Lagrange multipliers to minimize variance and skewness metrics in **Python** and visualized the weighted statistical solutions with **Matplotlib**
- Used remote connection to **Linux** to process large datasets on an HPC cluster
- Gave a **45-minute talk** of the final research project with Q&A session

SKILLS

Python

Java

Linux

Tableau

SQL

Pandas

Scikit-learn

NumPy

SciPy

Tensorflow

MongoDB

PERSONAL PROJECTS

Neural Network Dependency Parsing

Implemented input representation for a feed-forward neural network, decoded its, and specified the network architecture. Trained the model using **TensorFlow** and Keras to predict the transitions of a dependency parser with 70% accuracy on labeled attachments and 75% accuracy on unlabeled attachments.

Spotify Data: KNN, Linear Regression

KNN: Pre-processed Spotify data and analyzed the results of project using sklearn.metrics, which classifies (binary, multi-class) the genre of a song.

Linear Regression: Finds ideal parameters that maximize a track's popularity in a genre. Emphasis in understanding why data quality is important.

Text Classification

Built a trigram language model that was used to classify a dataset of essays. Evaluated the model using a perplexity function (with n-gram probability functions) and smoothing techniques implemented by hand.

Student Coursework Database

Implemented an **SQL** schema that creates front-end web **app** with Flask and HTML that queries fake university database to let user access course info through secured queries. Application ran through **Google Cloud Platform**.

Convolutional Neural Network and Andy Warhol

Work in progress project that uses a CNN to analyze the style of an Andy Warhol flower painting. The model uses average pooling on an image to extract its style and apply it to a different image. The motivation is to explore various artwork while learning CNNs unconventionally.

Predicting Cancer Outcome

Data science project that uses a random forest machine learning model to predict whether a patient diagnosed with lung cancer will survive with 89% accuracy. Emphasis on preprocessing with feature expansion, one-hot encoding, and feature scaling.

AWARDS AND ACTIVITIES

Summer 2022 Intern @ Workday

Incoming operations analyst intern to work on data analytics projects that optimize Workday spending.

Columbia DSI Scholar

Use data science for social and ethical good.

Ratrock Magazine Video Producer

Producing a series of artist interview. Producer in feature-length documentary of ballet group.

CDSS Data Science Hackathon Finalist

Top 9 finalist in hackathon with 200+ participants.

LatinX in AI (LXAI) Member

Community working to further AI innovation and resources for LatinX individuals globally.