JAY ALEXANDER TREVINO

Computer Science Intern

713-248-3303 | jay.trevino@columbia.edu | https://github.com/jat2211

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

Software 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: Attempts to find the ideal parameters that maximize the popularity of a track in a given 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

Created an entity-relationship diagram that maps and relates unique data types from a university database. Implemented these sets as an **SQL** schema to create web app with Flask that allows a user to design a course guide for their major.

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 convolutional neural networks in an unconventional way.

Hash Table Spell-checker and K-Best Values Priority Queue

First part of project implements a spell-checker using a hash table and makes suggestions to correct the errors. The second part of project implements a priority queue that returns the k-best (i.e. largest) values in a set of data. Project done in its entirety in **Java**.

AWARDS AND ACTIVITIES

Columbia DSI Scholar

Accepted into Data Science Institute Data for Good Scholar's program. The aim of the program is use data science to promote social and ethical good.

CDSS Hackathon Finalist

Top 9 finalist in data science hackathon with 200+ participants. Analyzed EPA data and created a mock app that shows pollutant data for given zip codes.

Ratrock Magazine Video Creator

Create videos that feature local artists. Featured in publication for creating CNN model that captures an artist's style and replicates it onto another image.