ANTHONY TAGLIENTE

DATA SCIENTIST

anthonyttagliente@gmail.com (845)642-7247 Astoria, NY 11105 anthony-tagliente-75735745/ atag34

SKILLS

TECHNICAL: R, Python, SQL, Git, Docker, Keras, Numpy, Pandas, sklearn, XGBoost, Tensorflow, dplyr
DATA VISUALIZATION: Shiny, Tableau, Plotly, Streamlit, Seaborn, ggplot2, Matplotlib, Leaflet
DATA WAREHOUSING: AWS, BigQuery, Spark, Salesforce

EXPERIENCE

Metis, *Data Scientist*, New York, NY Sept. 2020 - Dec. 2020

Metis is an ACCET accredited 12-week immersive data science bootcamp with focus on python, statistics, supervised and unsupervised machine learning, data analysis, data visualizations, database management, and communication. Some projects focused on building classification models for public company bankruptcy filings and clustering municipal bond issuing documents. Please see project section for more details.

The Heron Foundation, *Data Associate*, *Integrated Capitals*, New York, NY Jan. 2018 - Sept. 2020

- Developed a series of Tableau and Shiny dashboards using data from Salesforce, APIs and internal SQL database to track social and financial performance of investments.
- Dashboards were used as routine reporting materials for Heron's board (including but not limited to the Board of Directors, Investment Committee, and advisory boards.)
- Supported the Quarterly Investment Review process including automated scripts for collecting changes in portfolio holdings and social impact scores and live interactive html tables.
- Created and managed internal SQL database on AWS for storing data collected via automated scripts.

The Heron Foundation, *Associate*, New York, NY Mar. 2015 - Jan. 2018

- Managed and maintained website content, especially daily changes to the homepage. Collaborated with the K&I team to research, write, edit and disseminate engaging, informative articles, blog posts and other communications materials on topics ranging from the economy to investing to philanthropic practices.

PROJECTS

Streaming Audio Gunshot Classification

Created a convolutional neural network to identify handguns being fired based on an audio sample. Specifically focusing on the two most popular handguns in the United States, the Glock 19 and Sig Sauer P320, as a proof of concept. The dataset was built using samples from Youtube and audio augmentation techniques similar to those using in image classification. In order to make the model more generalizable, generic audio samples that did not include gunshots were included in training as a third non-gunshot class. The project included a method of classifying streaming audio in quarter second increments and displaying a running one second average for live classification.

Finding Impact in Municipal Bonds

An analysis of General Obligation bond documents using NLP techniques to identify investments that may be of interest to social impact focused clients. Several topic modeling techniques were tested to identify the unique use of proceeds in each document and cluster them together. The most effective approach proved to be using non-negative matrix factorization as a feature reduction technique and looking at outliers at the state level. I was then able to identify bonds that had socially impactful purposes such as veteran support.

Bankruptcy Prediction Using SEC EDGAR Filings

Using balance sheet data from 10-K and 10-Q filings, predicted if a company will file for bankruptcy within the next year using binary classification models. The difficulty in collecting usable data for this project offered a great opportunity to practice dealing with data in the real world. Synthetic Minority Oversampling Techniques were used to help models deal with significant class imbalances and improved model accuracy. While many models were tested, including KNN, RandomForest and Logistic Regression, the final XGBoost model offered the best results.

Warcraftlogs and Linear Regression

A deep dive into scraped data from warcraft logs to predict how long a group of players will take to defeat a boss in World of Warcraft classic. Focused on collecting data using Selenium and BeautifulSoup as well as linear modeling techniques, tuning and evaluation metrics.

MTA Turnstile Data

Explored public data provided by the New York City Metropolitan Transportation Authority (MTA) with the goal of assisting a hypothetical nonprofit with their street canvassing strategy targeting subway stations in NYC. By providing census demographics, built a weighted scoring system to identify subway stations that would best represent the organizations target audience in a way that can be adjusted.

EDUCATION

State University of New York at Albany
BA Economics and Political Science 2012