

Nick Pondok

Experience

Metis · Remote

July 2021 to Sept. 2021

Data Scientist

Completed the Metis 10 week data science and engineering bootcamp focused on Python programming, machine learning, statistical modeling, data visualization, project design, and communication. See projects below.

NLP with Movie Reviews:

- Generated a topic model for movie reviews for Marvel's Shang-Chi. Web scraped over 2,000 movie reviews from google reviews using BeautifulSoup and Selenium
- With use of tools such as Regex and Spacy, all data was run through a preprocessing data pipeline. The pipeline consisted of the removal of numbers, punctuation, and stop words, the compounding of actor names and movie titles, spell checking, and lemmatization.
- A doc term matrix was generated using a CountVectorizer and was fit onto a Corex Topic model. The polarity of each document was calculated and each document was assigned as either positive, negative or neutral sentiment. Additionally a doc topic matrix was generated and each document was assigned a topic based on the highest topic correlation.
- Audience reviews talked about four main topics: Action sequences, plot, cast and Asian culture. A larger percentage of negative reviews surrounded the movies Plot when compared to positive reviews.

Spotify Genre Classifier:

- Built a music genre classification model using Spotify's audio features.
- Songs were pulled from Playlists categorized by Spotify (i.e. a song could be from a "pop" playlist or a "hip hop" playlist).
- Spotify audio features were used as the features and genre as the target. Performed a train/test split and fit the data onto a Random Forest classifier. Class weight = 1, n_estimators = 100.
- The aim for the model was to achieve the highest accuracy with interpretability being less important. Random Forest was chosen due to having an almost a 10% increase in accuracy over KNN and 5% increase compared to Logistic Regression. The model produced an accuracy score of .437

Exploratory analysis of Spotify Playlists:

- Created a functioning data pipeline to collect, store, and update data as well as deploy a web app with my findings. Project focused on identifying audio features most present in specific genres as well as audio features within specific playlists of those genres.
- Over 90k songs were collected using the Spotify python wrapper Spotipy an put through a pipeline that consisted of three parts 1) data cleaning 2) data storage and 3) updating the web app
- The data cleaning and wrangling was done with pandas and then stored into a SQL database. A cron-job was set up to check for new data once a week.
- A streamlit .py file was created and pulled the most recent data from the SQL database and displayed audio features on the web app.
- The app consisted of bar charts with the average values from each audio features (i.e. danceability, loudness, etc.). Additionally users can view specific playlists within that genre and see the averages of the audio features for songs in that playlist.

NBA Salary Predictor:

- Used linear regression to determine which features carry the most weight when trying to determine a player's salary.
- Season averages and contract information were pulled for every player over the past seven seasons (2013-2020). Data was collected via basketballreference.com through use of the web scraping tool BeautifulSoup.
- Set up a cross validation train, test split on a Ridge Regression model across five K folds in order to ensure interpretability of the coefficients. The model was able to achieve an R² of .58 and had an RMSE of ~\$35M
- Experience proved to be the most important factor in a players contract size.

MTA Traffic Analysis for softball advertising:

- Cleaned, organized and visualized MTA turnstile data to find the optimal time for a local adult softball league to advertise in the field.
- Used pandas to clear null values and updating any inconsistencies with the data
- Grouped data by time and day to find the times with the highest foot traffic at several MTA stations
- Found later afternoon was optimal time for advertising.

Human Interest · San Francisco, CA

Oct. 2017 to June 2021

Operations Lead (2020 - 2021)

- Recruit, interview, onboard, guide and supervised a team of 12 retirement operations professional and ensured operational compliance
- Drove initiatives forward, managing internal and external stakeholders to deliver positive experiences to our customers through timely financial and census data transfers
- Created and develop processes for onboarding and maintaining clients on new payroll providers
- Worked cross-functionally with account management, implementation, engineering, partnership, and sales teams to ensure customers are onboarded successfully, have requests our product can handle, and experience timely contribution processing
- Successfully worked with external auditing team to assist and complete the SOC 1 Type 1 Audit

Operations Associate (2017 - 2020)

- Communicate directly with customers to identify their needs and resolve issues
- Partner with product teams to drive platform improvements minimizing human error and improve metrics
- Validate, reconcile, and process \$100,000+ in 401(k) contributions.
- Upload and maintain census data for over 1,000 clients
- Identify bugs within the product and work with Engineering department to reconcile

Flynn, Shojinaga & Associates, Inc. · Alameda, CA

Sept. 2014 to Oct. 2017

Assistant Pension Administrator

- Analyzed pension plans to prepare and distribute loan documents to plan participants & clients
- Communicate daily with clients & participants to explain loan documentation & distributions
- Prepare and package IRS approved plan documents for 100+ pension plans annually
- Review Qualified Domestic Relations orders to ensure compliance with relevant plan guidelines
- Create and prepared materials concerning Adoption Agreements for clients & potential participants for enrollment purposes
- Reviewed weekly payroll reports and uploaded 401(k) contributions for processing

Contact

✉ nick.pondok@gmail.com

☎ 510-303-1924

📍 Alameda, CA

🌐 /in/nickjpondok/

🔄 nickpondok

Skills

PROGRAMMING LANGUAGES

SQL

Python

MACHINE LEARNING

Regression

Classification

Natural Language Processing

VISUALIZATIONS

Excel/Powerpoint

Matplotlib

Seaborn

LIBRARIES

Pandas

Scikit-Learn

NLTK

Numpy

DATABASES

MySQL

MongoDB

SQLite

WEB SCRAPING

BeautifulSoup

Selenium

APP DEPLOYMENT

Streamlit

Flask

Education

University of California, Santa Barbara

B.A. Communication 2014

Minor in Asian-American Studies