

SKILLS**LANGUAGES**

Python
SQL
HTML
git

LIBRARIES

scikit-learn
imb-learn
pandas
pyspark
numpy
streamlit

NATURAL LANGUAGE PROCESSING

Gensim
CorEx
NLTK
SpaCy

DATA VISUALIZATION

Tableau
Seaborn
Matplotlib
Powerpoint
Google Suite

WEBSCRAPING

Selenium
BeautifulSoup

MACHINE LEARNING TECHNIQUES

Regression
Classification
Clustering
Neural Networks
Recommender Systems

EDUCATION

Rutgers University May 2016
B.A. Physics 2015
Ed. M. Science Education
2016

SUMMARY

Physics Teacher turned Data Scientist with an expansive skillset ranging from regression and classification to deep learning and language processing.

PROJECTS

Rain Prediction with Classification Fall 2020

- Deployed rain/no-rain classification model on Streamlit that interacts with 'OpenWeatherMap' API to make real-time weather predictions.
- Web-scraped 20 years of daily weather updates using BeautifulSoup/Selenium.
- Trained four unique classification models (LogReg, RandomForest, KNN, XGBoost) to predict rainy days and ensembled with a soft-voting classifier, prioritizing recall.
- Cross-validated training pipeline with imb-learn library for imbalanced classes.

Analysis of Box Office Domestic Gross Fall 2020

- Predicted movie revenue using cast & crew as features on a linear regression model.
- Web-scraped 10K+ movies from Boxofficemojo.com and IMDB.com.
- Performed LASSO regularization to optimize bias-variance tradeoff and select important features engineered to quantify past performance of cast & crew.

What Do Data Scientists Talk About? Fall 2020

- Built a Tableau dashboard to explore various styles of writing, trends in the TowardsDataScience (TDS) community, and factors affecting article popularity.
- Topic Modeled using NMF algorithm on 35000 blog posts scraped from TDS.
- Created document vectors using Gensim Doc2Vec and tested multiple clustering algorithms including DBSCAN, Hierarchical Agglomerative, and KMeans.

Hiking Trail Recommendation System Fall 2020

- Built an app to recommend hiking trails using user preferences and trail features.
- Web-scraped 400K+ hiking trail reviews and 14K+ hiking trail descriptions.
- Utilized SpaCy, CoEx and Sklearn packages to preform NLP preprocessing and topic modeling to generate the aforementioned trail features.

EXPERIENCE

Metis Data Science Bootcamp New York, New York
Sept. 2020 to Dec. 2020
Data Scientist

Metis is an ACCET accredited project oriented 12 week immersive data science bootcamp.

- Developed 4 end to end DS projects dealing with regression, classification, NLP, and clustering throughout the course accompanied by presentations to a panel of peers.
- Skills learned include: data acquisition, data cleaning, data visualization, feature engineering, and hyper parameter tuning.
- Data acquired was cleaned, stored, and queried via a local PostgreSQL database.

Fair Lawn High School Fair Lawn, New Jersey
Sept. 2016 to June 2020
Teacher of Physics

Courses: AP Physics C Mechanics, AP Physics C Electricity and Magnetism, AP Physics 1, College Prep Physics

- Simplified complex concepts for 500+ students and showed interdependence of physics and related technical subjects such as calculus and chemistry.
- Liaised between the science and IT departments to facilitate the transition to personal devices for all 1400+ high school students.
- Developed standards and benchmarks on data literacy targeting 5000+ students each year (across all 12 grades) within the Fair Lawn district.