BERNARD OPOKU
MACHINE LEARNING ENGINEER

EXPERIENCE

Metis
Data Scientist — Remote — June 2021 to Aug. 2021

Completed an immersive data science and machine learning bootcamp, which included regression, web scraping, project scoping, visualization, classification, NLP, deep learning, and data engineering. **Selected work includes**:

Auto FAQ Answering System
- Utilized Natural Language Processing to build an auto FAQ answers for a bank
- Used matrix factorization methods like SVD and NMF to reduce dimensions
- With the help of a 2014 paper on IEEE, extracted 169 principal components from the data, which did not only reduce the space needed but also ensured the data does not consist of unnecessary noise.
- Built a recommendation engine using cosine similarity to return the answer to the most similar customer queries
- Designed a Flask web app and deployed it on Heroku.

Fantasy Football Predictor Web App
- Built on a previous project to create an end-to-end data pipeline to periodically update data, in order to make accurate predictions with the most recent data
- Used a SQL database to update data using the official FPL API
- Built a regression model to predict the current week's players' results; the model is then updated when the current week's round is completed
- The pipeline is put into a flask web app and deployed with Heroku

Customer Churn Classification
- Developed a classification model to predict whether or not a customer will leave a company from a given customer's info
- Used class imbalance techniques such as threshold adjustment and oversampling(SMOTE, ADASYN) to overcome the imbalance in the dataset
- Explored the data by performing both univariate and bivariate analysis to get more insights on which groups of individuals are churning and not churning
- Used multiple algorithms (XGBoost, Naive-Bayes, Random Forests, Logistic Regression) to select the ideal model for the use case.
- Focusing on f_beta(beta=2) in attempt to minimize false negatives while not sacrificing completely on the false positives, the model was able to achieve an f_2 score of 75% with a decision threshold of 0.21

Fantasy Premier League Points Prediction
- Scraped data from the dynamic official FPL website to acquire player statistics for about 600 players and over 30 weeks of Statistics for the 2020-2021 season, using BeautifulSoup and Selenium.
- Built and optimized a linear regression model using both StatsModels and SKLearn to predict a player's expected points given the previous week's data.
- Though the final model has an R^2 score of 0.34, An MAE score of 0.06 points was very much desired
- Displayed results and residuals using MatPlotLib, further explaining model performance

MTA Track Data
- Analyzed New York City MTA track data to determine optimal station placement for new underground tunnel construction
- Used the data and Exploratory Data Analysis to determine how big a station should be
- Used the data to suggest staffing and security prioritization for stations depending on the time

Zeauf Hotels
Sales Manager — Limerick, PA — 2018 to 2021

- Oversaw all room sale activities for Zeauf hotel brands including the Holiday Inn Express & Suites and Holiday Inn
- Tasked with acquiring and managing group booking; generated Corporate accounts for each hotel by visiting surrounding businesses
- Hired and trained new sales associates to efficiently provide the utmost hospitality to guests and Corporate partner

Cheyney University Department of Computer Science and Mathematics
Tutor & TA — 2015 to 2017

- Instructed students on various math courses including elementary algebra, college math, elementary functions
- Develop class structure and weekly lesson plans, and hold weekly office hours
- Instructed the required computer skills course for approximately 50 incoming freshmen

CONTACT

bernardopoku160@yahoo.com
610-931-6706
Philadelphia, PA
dranreb1660

SKILLS

PROGRAMMING
Python
SQL

LIBRARIES AND FRAMEWORKS
Numpy
Pandas
Scikit-learn
NLTK
SpaCy
Flask
Django

DATA VISUALIZATION
Seaborn
Matplotlib
Tableau

MACHINE LEARNING
Linear Regression
Classification
Model Cross Validation and Selection
Clustering
Dimensionality Reduction
Natural Language Processing
Recommendation System

DATABASES
SQLite
SQL Alchemy

CLOUD COMPUTING & HOSTING
Google Cloud Platform
Heroku
Docker

OTHER
Git
HTML
Css
MS Office Suite

EDUCATION

Cheyney University of Pennsylvania
BSc. Computer & Information Science (CIS)
2018