DARIEN MITCHELL-TONTAR DATA SCIENTIST

darien.pmt@gmail.com medium.com/@darien.pmt 347-506-8872 Fort Collins, CO 80524 darienpmt darienpmt

SKILLS

LANGUAGES: Python, SQL
LIBRARIES: Pandas, Numpy, ScikitLearn, Matplotlib, Seaborn, NLTK, vaderSentiment, Keras, Tensorflow, PIL
MODELING: Linear Regression Models, Random Forest Models, Latent Semantic Analysis, Non-Negative Matrix Factorization, Latent Dirichlet Allocation, Convolutional Neural Networks, Clustering Algorithms, Image Processing
VISUALIZATIONS: Tableau, Plotly, Bokeh, OpenCV
TOOLS: Postgres SQL, Flask, AWS, Google Cloud, MongoDB

EXPERIENCE

Metis, *Data Scientist*, Online Mar. 2020 - June 2020
Metis is an accredited 12-week data science bootcamp which focuses on Python, statistics, machine learning, data analysis, data visualization and database management. The curriculum is built around five open ended projects in which students design their own project scope, obtain data and use certain specified algorithms and tools to answer their question. Each project culminates in a 5-minute presentation intended for both technical and non-technical audiences.

United Nations International School, *Math Instructor*, New York City Aug. 2018 - June 2019
Served on the curriculum committee and helped restructure grades 6-10 curriculum to be problem and standards based.

International School Nido de Aguilas, *Math Instructor*, Santiago, Chile July 2014 - June 2018
Redesigned the 9-10 grade math curriculum from scratch to be entirely student led. Developed a technology strand in which students had to meet bench mark proficiency in the use of graphing calculators, Scratch, Google Sheets and basic programming skills. Led multiple student trips abroad for athletic events.

King's Academy, *Math Instructor*, Amman, Jordan Aug. 2011 - June 2014
In addition to teaching, was an advisor to 5+ high school students each year, a dorm parent, basketball coach and led multiple student trips abroad to Lebanon for athletic events.

PROJECTS

Selfie Popularity Predictor May 2020 - June 2020
Using a dataset containing over 40,000 selfies, used Keras to train own convolutional neural network to predict whether an image would be popular or not on Instagram (based on the number of clicks it will receive). The final model achieved 72% accuracy on a balanced dataset. In order to visualize the results, created a Flask App which takes in a photo of any size and returns the corresponding probabilities of it being popular on Instagram according to my model. In addition, OpenCv and Facial Emotion Recognition (FER) to give the user information about the emotions expressed in their photo with the intention of giving them some ideas of how to increase their popularity.

Exploring Topics and Sentiment in my Grandfather's WW2 Letters May 2020 - May 2020
Explored the sentiment and topics in over 300 letters my grandfather wrote to my grandmother during WW2 using Natural Language Processing and unsupervised machine learning modeling (Topic Modeling and clustering).
Methods: Python, Pandas, Numpy, Matplotlib, NLTK, vaderSentiment, Topic Modeling, Unsupervised clustering models

Predicting U.S. Counties At Risk for Covid-19 Outbreaks Apr. 2020 - May 2020
Predicted whether a U.S. county was at risk for seeing an increase in covid-19 cases using data from the CDC, census and county hospitals. Had to create my own target since there was no "at risk" label in my dataset in addition to joining 5 different datasets together in SQL database. End product a map visualizing of my predictive model highlighting the counties which were most at risk. Found that although cases were declining in the entire country at the time of completing, they were rising in the predicted at risk areas outside of previous "hotpots."
Methods: PostgreSQL, Python, Pandas, Numpy, Matplotlib, Seaborn. ScikitLearn, Regression Models, Random Forest Models, Tableau, AWS

Predicting NFL Running Back Performance using NCAA Data Apr. 2020 - Apr. 2020
Used linear regression models to predict a future NFL running back's yards per game performance based on their college stats. Used data scraped from sports-reference.com. The data contained numerous players with the same name requiring me to give assign each player a unique ID so I could retain all of my data. Found that yard per carry in college was the best predictor for success in the NFL.
Methods: Python, BeautifulSoup,, ScikitLearn, Pandas, Numpy, Matplotlib, Seaborn, Linear Regression Models (including Ridge and Lasso)

EDUCATION

Syracuse University Aug. 2009 - May 2011
M.S. Mathematics 2011

Syracuse University Aug. 2005 - May 2009
B.S. Mathematics 2009
B.A. Physics 2009

ACTIVITIES

Trail Running & Coaching July 2014 - Current
Less importantly, I am also a semi-accomplished ultra runner. I'm extremely passionate about running, have coaching experience, organized community races and would love to incorporate my love for running into my work.