MALACHI HALE

DATA SCIENTIST

949-748-0500   malachi.d.hale@gmail.com   /malachi-hale   in/malachi-hale

Data Scientist with a background in mathematics, statistics, and languages. Recently completed Codeup Data Science, a fully-immersive, project-based 22-week career accelerator that provides 670+ hours of expert instruction in data science. Excellent communicator, concise speaker, and an ethusiastic team player. My inquisitive nature and deep, wide-ranging interests make me a fast learner and a meticulous analyst. I am looking forward to using my skills to provide insights that aid in your company's growth and innovation.

TECHNICAL SKILLS

Applied Statistics - SQL - Python - Git- Pandas - Matplotlib - Data Cleaning - SciPy - NumPy -sklearn - Spark - Natural Language Processing - Tableau - Data Storytelling - Jupyter Notebooks - Machine Learning- Clustering- Classification- Anomaly Detection - Time Series Analysis - Regression - Data Visualization

EDUCATION/EXPERIENCE

CODEUP DATA SCIENCE

June 2021 - December 2021

Certificate of Completion

Developed expertise across the full data science pipeline (planning, acquisition, preparation, exploration, modeling, delivery), becoming comfortable working with data to deliver actionable insights to diverse stakeholders.

US ARMY

2018-2021

Served as an enlisted member of the Army , in the role of cryptologic language analyst, specializing in Russian. Achieved proficiency in the Russian language. Analyzed a wide variety of intelligence information.

UNIVERSITY OF ARKANSAS

2014-2018

BS Applied Mathematics
BA Spanish

Completed two degree tracks while on a full-ride Honors College Fellowship, in addition to the Arkansas Governor's Distinguished Scholarship. Studied internationally in Spain and Brazil. Successfully defended a thesis about number theory and its application in Noetherian rings.

DEVELOPMENT PROJECTS

IS AMERICA'S GLASS HALF-EMPTY?

CLASSIFICATION, DATA EXPLORATION
November 2021

Acquiring survey data from the Pew Research Panel, our team explored the drivers of pessimism in American prospective attitudes. Understanding what drives pessimistic or optimistic thinking about the future will help business leaders clarify strategies for how to best serve customers, which products and services to offer, and which investments to make.

PREDICTING PROGRAMMING LANGUAGE

NATURAL LANGUAGE PROCESSING, CLASSIFICATION
October 2021

I utilized the GitHub API to acquire the of READme files of 200 repositories. After cleaning the text data of the READme files, I performed Term Frequency Inverse Document Frequency Vectorization on the text. I then deployed classification models that used the READme contents to predict the programming language of the repository. My final model successfully used the READme text to predict the repository's programming language 28% better than baseline.

WHAT MAKES A HIT?

LINEAR REGRESSION, CLUSTERING, CLASSIFICATION
October 2021

Using a dataset acquired from Kaggle that included 500,000 tracks from the Spotify API, as well the fifteen musical metrics of these tracks, I performed a series of machine learning analyses. First, I constructed several linear regression models that predicted the popularity index of a track based on that track's auditory characteristics. My best performing model, a Fourth Degree Polynomial Regression Model, outperformed baseline by approximately 20%. Second, I used k-means clustering to sort the tracks into five distinct mood groups. Third, I used Random Forest model to predict with 77% accuracy which musical features were most predictive of a track being a hit.

PREDICTING PROPERTY VALUES

LINEAR REGRESSION
September 2021

Using the Zillow dataset of single-unit properties sold between May and August 2017 in Southern California, I employed the data science pipeline to acquire the data, clean the data, run statistical testing, create Seaborn visualizations, and utilize Select K Best and Recursive Feature Elimination to find the most relevant features for predicting cost. I constructed several linear regression models that used the features in the dataset to predict the tax value of each property. My best performing model, a Third Degree Polynomial Regression Model, outperformed baseline with a lower random mean sampling error on the test dataset.