

FORREST MCCROSKY

DATA SCIENTIST



C: 727-612-1082



forrest.mccrosky@gmail.com



/ForrestMccrosky



/in/forrestmccrosky

TECHNICAL SKILLS

- SQL
- Python
- Machine Learning
- Natural Language Processing
- Applied Statistics
- Data Analysis
- Seaborn
- Tableau
- Git
- Matplotlib
- Numpy
- Pandas
- Jupyter

EDUCATION

CODEUP

Fully immersive project based 22-week career accelerator. Codeup provides a strong curriculum focused on machine learning projects after teaching core data science fundamentals. The program focuses on the full data science pipeline while working with real, messy data to deliver actionable insights to diverse stakeholders.

WORK EXPERIENCE

ICON MANAGEMENT

Guest Services Team Member
July 2019 - September 2020
Ensure all guests and members are fully accommodated and serviced in the golf facility. Responsible for complaint resolution using HEART method (Hear, Empathize, Apologize, Respond, and Take Action).

CELLULAR SALES

Business Solutions Consultant
July 2016 - July 2019
As a Business Solutions Consultant for Cellular Sales, Verizon's largest authorized retailer I ranked in the top five Account Managers in East Florida (South Florida, Miami, Ft Lauderdale etc) monthly for new sales. While in January 2018 and November 2017 was also ranked top five nationwide and was awarded two "Certificates of Excellence".

SIMPLE WEDDINGS

Event Organizer
August 2015-September 2016
As an event organizer I was responsible for onsite event interaction with attendees as well as working directly with wedding parties to ensure successful events. As well as event loading, setup, configuration, and tear down for Pinellas County Events.

ABOUT

I am a determined and results driven data scientist seeking a role that will allow me to leverage my strong work ethic, technical skills, and continue to develop my data science experience. I am looking to work with an organization that is expanding and advancing their use of data science to drive actionable insights to grow their business and marketshare.

CURRICULUM PROJECTS

HOUSTON, WE HAVE A PAY GAP

Capstone: Regression Machine Learning/Modeling September 2021

This project created regression models that predicted the annual salary of current Texas Government employees. The goal of this project was to find key drivers that influenced the annual salary of these employees. While we explored the data using seaborn and scipy stats we will also answered important questions such as: are there discrepancies for annual salary among races or genders? After we used scikit learn for regression modeling. Our best performing model was a 2nd degree polynomial with an r-squared value of 0.38 and an rmse of 20108.9 dollars.

NLP: PREDICTING REPOSITORY LANGUAGES

Multi-Classification Machine Learning/Modeling August 2021

This project is a supervised machine learning project that creates multi-class classification models with the assistance of Natural Language Processing. The goal of these models is to predict the programming language of github.com's most starred repositories readme files. Several self-written functions aided by python's beautiful soup library were used to help scrape github for a repository list and their readme's contents and we filtered for the four most common languages (Python, Javascript, Java, & Go). The best performing classification model that was used for the out-of-sample data test came back with 82 percent accuracy in predicting the correct programming language of each repository.

ZILLOW CLUSTERING PROJECT

Regression Machine Learning/Modeling June 2021

This project creates regression models aided by K-means clusters. The models goals are to predict the Zestimate error of Zillow homes in southern California. Using seaborn and matplotlib to explore the dataset I was able to select longitude, age of the property, and a cluster made from the taxamount and longitude as key drivers of logerror. My best performing model was a Tweedie Regressor Model that evaluated with an RMSE of 0.16975 and an R squared value of 0.0017.

PASSION PROJECTS

CREDIT CARD FRAUD CLASSIFICATION PROJECT

Classification Machine Learning/Modeling July 2021

For this Project I acquired Credit Card Transaction data from Kaggle.com. I did extensive feature engineering and exploration on the data, using pandas, numpy and matplotlib. Using Scikit learn and imblearn, I created a successfully performing Random Forest Classifier model. The Random Forest model is 91 percent accurate in predicting our positive case (fraud) and 91 percent accurate in predicting our negative case (not fraud).