VOLHA PUZIKAVA

Middle Village, NY | 929-317-9590 | helga.mikel@gmail.com | GitHub | Blog | LinkedIn

DATA SCIENTIST | MACHINE LEARNING ENGINEER

I have experience in data acquisition and data modeling, statistical analysis, machine learning, deep learning, and NLP. With a Bachelor's degree and a background in forensic science, I bring strong analytical and problem-solving skills to help data-driven companies to extract meaningful insights from the raw data and make business decisions based on the data findings.

TECHNICAL SKILLS

Python, SQL, R, Tableau, HTML, Machine Learning (Pandas, NumPy, Scikit-Learn, Scikit-Surprise, Matplotlib, Seaborn, Statsmodels, Apache Spark, Keras), Data Mining, Data Visualization, Predictive Modeling, Time Series Forecasting, Recommender Systems, Deep Learning, A/B Testing, Web Scraping, Git, Jupyter, Dash, Flask, AWS

TECHNICAL PROJECTS

Office Supplies Recommendation System - Github

Recommend office supplies based on reviews of purchased products and advise if it is valuable to offer products as a two-pack.

- Collected 5,581,313 reviews of 315,644 products from Amazon review page for features such as ratings, product ids, reviewers' ids and product titles
- Tested various collaborative filtering methods in surprise library and Spark programming environment
- Employed SVD and ALS models to build recommendation systems
- Performed A/B Testing to determine a two-pack would decrease a buying rate by 6.25%

Cryptocurrency Prediction Analysis - Github

Forecast price trends for the top two cryptocurrencies for a half year out starting September 2022.

- Obtained 2,070 days of cryptocurrencies' historical data from Investing.com website for use in time series models
- Implemented the ADF test and detrended the data through subtracting the EWMA and differencing transformations
- Modeled the data using various strategies including different orders of AR, MA, ARIMA, and SARIMA
- Forecasted the price trends of cryptocurrencies using the models with the lowest AIC of 349.75 and 328.48

Stroke Prediction Analysis - Github

Predict if patients will develop stroke in their lifetime given clinical features of the patients.

- Gained data about 5,110 patients from Kaggle website for use in classification models
- Applied one hot encoding and SMOTE-NC to handle categorical features and an imbalanced dataset
- Modeled the data using different strategies including baseline and tuned Logistic Regression, Decision Tree, Bagged Trees, Random Forest, AdaBoost, Gradient Boosting, XGBoost, Naïve Bayes, KNN, and SVM
- Achieved the recall score of 97% and identified the key factors leading to stroke

EMPLOYMENT HISTORY

Criminalist 1B, **NYC Office of Chief Medical Examiner**, New York, NY 07/2018 - Present

- Perform detailed examination and scientific analyses of evidentiary material to identify potential suspects
- Interpret STR data using combined probability of inclusion and likelihood ratios to determine genetic correlation and identity
- Prepare administratively closed reports on the results of scientific analyses in LIMS
- Determine the cause of, and propose the solution to, instrumental problems and troubleshooting events
- Endorse the inventory system in LIMS to maintain the updated records of reagents and supplies

Criminalist 1B, **NYPD Police Crime Laboratory**, Queens, New York 08/2017 – 06/2018

- Performed GC/MS analyses of evidentiary material for the presence of controlled/ non-controlled substances
- Applied linear regression analysis on data to identify the components of unknown samples with 99% accuracy
- Prepared detailed laboratory reports on the results of scientific analyses in LIMS
- Provided accurate and objective expert witness testimony for the presence of controlled substances in samples

EDUCATION

Flatiron School, New York, NY 10/2022
Immersive Data Science Bootcamp program

John Jay College of Criminal Justice, New York, NY 05/2016
Bachelor of Science in Forensic Science with Concentration in Criminalistics