VOLHA PUZIKAVA

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

DATA SCIENTIST | MACHINE LEARNING ENGINEER

Over 5+ years of experience in criminalistics including statistical analysis, data analysis and data visualization. Also skilled in data acquisition and data modeling, artificial intelligence (AI) and machine learning (ML), deep learning, Natural Language Processing (NLP), working with large datasets. With a Bachelor's degree in forensic science, I bring strong communication, analytical and problem-solving skills to help data-driven companies extract meaningful insights from raw data and make business decisions based on the findings.

TECHNICAL SKILLS

Python (Pandas, NumPy, Scikit-Learn, Scikit-Surprise, Matplotlib, Seaborn, Statsmodels, PySpark), **SQL**, R, HTML, CSS, **Data Mining, Data Visualization** (**Tableau**), Supervised and Unsupervised **Machine Learning** (Regression, Classification, Time Series Forecasting, Recommender Systems, Clustering, Dimensionality Reduction), **Big Data**, **Neural Networks** and **Deep Learning** (**Keras**, **TensorFlow**), **Ensemble Learning**, **A/B Testing**, **Web Scraping** (BeautifulSoup), **Git**, Jupyter, Dash, Flask, AWS, Microsoft Office

TECHNICAL PROJECTS

Office Supplies Recommendation System | Supervised Machine Learning Project - Github
Recommend office supplies based on reviews and advise if a two-pack offer is valuable.
Python (Pandas, Seaborn, Matplotlib, Statsmodels, SciPy)
- Collected 5,581,313 reviews of 315,644 products from Amazon review page for features such as ratings, product IDs, reviewer IDs and product titles
- Tested various collaborative filtering methods in Surprise library (SVD model) and Spark environment (ALS model)
- Performed A/B Testing to determine a two-pack would decrease a buying rate by 6.25%

Cryptocurrency Prediction Analysis | Supervised Machine Learning Project (Time Series Modeling) - Github
Forecast price trends for the top two cryptocurrencies for a half year out starting September 2022.
Python (Pandas, Matplotlib, Scikit-Learn, NumPy, Statsmodels)
- Obtained 2,070 days of cryptocurrencies' historical data from Investing.com website for use in time series models
- 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 | Supervised Machine Learning Project (Classification) - Github
Predict if patients will develop strokes in their lifetime given clinical features of the patients.
Python (Pandas, Seaborn, Matplotlib, Scikit-Learn, Imbalanced-Learn)
- Gained data from 5,110 patients from Kaggle website for use in classification models with imbalanced data
- 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, **NYC Office of Chief Medical Examiner**, New York, NY   07/2018 - Present
- Perform detailed examination and scientific analyses of evidentiary material to generate quantitative data for the interpretation and evaluation of forensic DNA profiles
- Apply biological modeling, statistical theory, computer algorithms, and probability distributions for DNA profile interpretation using probabilistic genotyping software
- Calculate likelihood ratios for the DNA profiles by considering prosecution and defense hypotheses using allele frequency databases and DNA profile interpretation software
- Prepare administratively closed reports on the results of DNA profile interpretation in internal database

Criminalist, **NYPD Forensic Laboratory**, Queens, New York   08/2017 – 06/2018
- Performed analytical analyses of evidentiary material to generate the quantitative data for the identification of controlled and non-controlled substances
- Applied linear regression analysis on data using workstation software to identify sample with 99% accuracy
- Prepared detailed laboratory reports on the results of quantitative data analyses in internal database
- Communicated analytical results with both technical and non-technical audiences including courtroom testimony

EDUCATION

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

John Jay College of Criminal Justice, New York, NY   09/2012 – 05/2016
Bachelor of Science in Forensic Science (STEM Major)