

KJ MOCHROI

DATA SCIENTIST

✉ kjmochroi@gmail.com
🌐 www.kjmochroi.com
☎ 971-732-0684
📍 Beaverton, Oregon
in linkedin.com/in/kjmochroi/
📄 github.com/kjmochroi/Portfolio

Summary

I am a recent data science graduate with a strong foundation in machine learning, statistical analysis, and programming. Dedicated to leveraging advanced analytics techniques and tools to extract meaningful patterns and solve complex business problems. Eager to apply my skills and knowledge in a data scientist role to contribute to data-driven decision-making and deliver impactful insights.

Skills

DATA SCIENCE

Python
R
SQL
Data Cleaning and Preprocessing
Data Wrangling and Transformation
Statistical Analysis
Machine Learning
Big Data
Database Technologies
Communication
Data Visualization
Problem Solving
Data Management
Advanced Mathematics
Critical Thinking

Projects

Autism Classification Using Support Vector Machine Model and the AQ-10 Survey

For this project I used data collected from a ten-question survey called the Autism Quotient - 10 Questions (AQ-10) to create a binary support vector machine (SVM) classification model that can determine whether an adult should be referred to a professional clinician for the possible diagnosis of autism spectrum disorder. This investigation can be used to determine whether this short questionnaire is adequate in identifying such adults or whether a more thorough approach is needed. After cleaning the training data, I performed hyperparameter tuning using the gridsearch technique. With those optimized parameters I was able to train an SVM model with an average of 99.4% accuracy across the 5-fold cross-validation method I used to validate the model. This is significantly better than the 70% accuracy one would obtain with the null model.

A Machine Learning Approach to Cardiotocography Interpretation

•Cardiotocography (CTG) is widely used over the course of pregnancy and during labor and allows monitoring of fetal heart rate and uterine contractions to prevent fetal hypoxia, or oxygen loss in body tissues. Creating effective modeling for CTG exam data can provide additional support to those dedicated preventing harm and loss of life. The goal of this project will be to build an effective model to classify CTG data using supervised machine learning. We implemented two different models for this project, both of which performed classification on the fetal heart health target feature. The first model is a neural network and the second being a naive Bayes model. I worked in optimizing the naive Bayes model while my collaborator, Holly Figueroa, created the neural network.

•For the Naive Bayes classifier, we chose to use the Complement Naïve Bayes due to its ability to handle imbalanced classes in the target feature. We used a train-test-split ratio of 80/20 because it yielded the highest accuracy of any ratio we tested, without overfitting the model too much. We then applied a minmax scalar because the model only accepts positive values as input. After that we did a grid search to perform hyperparameter tuning on the only hyperparameter in a complement naïve Bayes model, using accuracy as our scoring metric. This resulted in a model with approximately 84% accuracy.

•For our neural network model, a multilayer perceptive classifier was chosen for our project. Given the strong imbalances between classes, data was train-test-split in a stratified manner. This offered assurance that classes with small counts were not further reduced during the split. For our MLP classifier, a grid search was conducted to find ideal model hyperparameters. Cross validation however, revealed a lower performance when predicting test data, indicating overfitting. The hyperparameters were tuned manually with cross validation to reach a higher accuracy in predicting outcomes for our test data. The resulting classifier model calculated outcomes with a .91 accuracy score.

Education

Bellevue University
M.S. Data Science 2023

Mar. 2022 to June 2023

Lewis & Clark College
B.A. Mathematics and Physics Double Major 2018

Sept. 2015 to May 2018

Employment

Lewis & Clark College
Instructor

Portland, Oregon
Jan. 2023 to May 2023, Jan. 2020 to May 2020

- Designed and delivered lectures on various computer science topics, including logic gates, programming, data science and the history of computer science.
- Developed course materials, including assignments, exams, and projects, to assess students' understanding and progress.
- Provided guidance and mentorship to students, fostering a positive and engaging learning environment.
- Mentored students in developing strong mathematical reasoning and analytical skills.

BMO Family Office
Hedge Fund Research Analyst

Portland, Oregon
Sept. 2018 to Aug. 2019

- Gathered, organized, and manipulated data from hedge fund managers to evaluate fund performance and exposures.
- Assisted senior research professionals in analyzing information and preparing investment materials.
- Maintained internal databases and programs to disseminate research reports and manager information to advisors and clients.