

- Called the following libraries: NumPy, pandas, sklearn.model_selection, sklearn.preprocessing, sklearn.linear_model, and sklearn.metrics.
- Identified missing values and/or placeholders for missing values and replaced them with the most frequently occurring value for columns of datatype object and with the mean for columns of datatype numeric.
- Used one-hot encoding (pd.get_dummies) to convert categorical variables to numerical placeholders for analysis
- Split the dataset into a features table and a target column, with X and y labels respectively.
- Split the data into training data and testing data using train_test_split from sklearn.model_selection
- Standardized all data to ensure proper modelling and testing
- Used KFold from sklearn.model_selection to randomly perform the split on the dataset 5 times.
- Utilized GridSearchCV with cross-validation to tune hyperparameters of KNN (neighborhood size) and Logistic Regression, identifying the best-performing estimators.
- Developed and evaluated predictive models using optimized KNN, Logistic Regression, and Decision Tree classifiers.
- Applied testing of the aforementioned models on the standardized testing dataset and compared outcomes using a confusion matrix, accuracy, sensitivity (recall), specificity, and precision.