* Machine Learning -- Patient Readmission

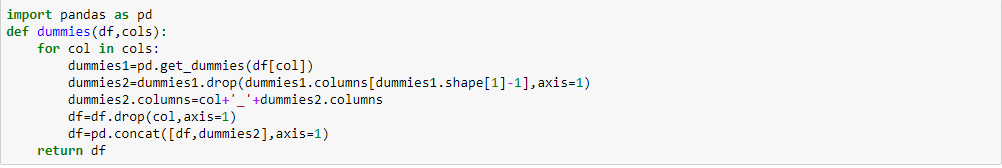
Predict which hospitalized diabetes patients will be readmitted for their condition at a later date.

Start with a dataset that contains records of diabetes patients admitted to hospitals. Write code in Python with packages matplotlib, seaborn, pandas, numpy, scikit-learn, etc. Prepare dataset for analysis. Explore features in the dataset with visualization. Fit Logistic Regression Classifier. Evaluate performance of the classification model. Do feature selection, cross validation, regularization, etc. Fit Gradient Boosting Classifier, etc. Evaluate performance of the optimized classification model.

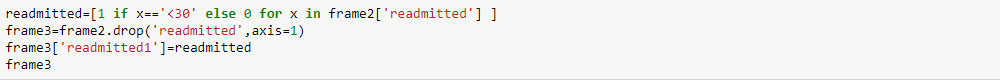
--Python--



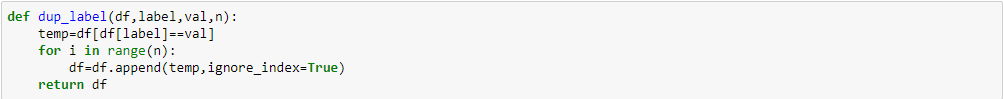






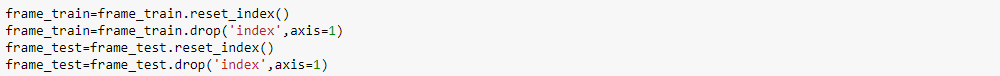


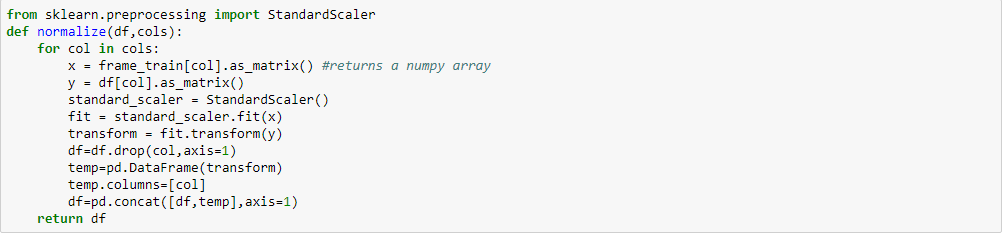


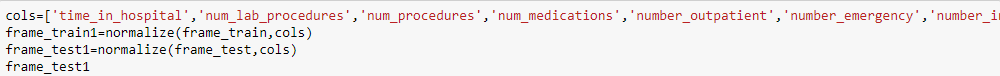


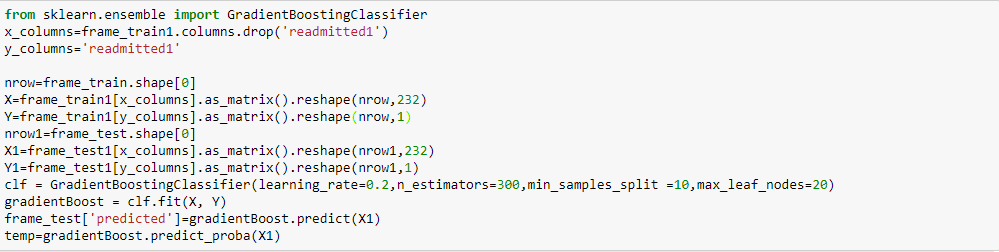




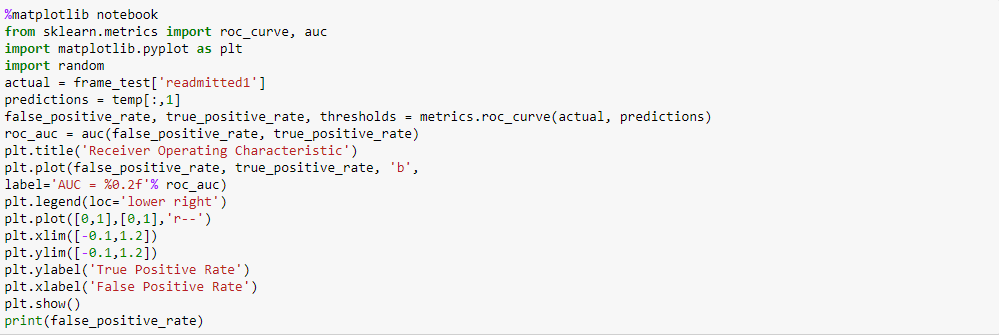












* Apache Spark -- Marketing Analysis

A banking institution ran a marketing campaign to convince potential customers to invest in bank term deposit.

Start with a dataset that contains records of customers contacted in the campaign. Do data cleansing. Write code in Spark-Shell Scala. Load data and create Spark data frame. Write SQL queries to obtain marketing success rate, explore features and do feature engineering.

--Spark--

