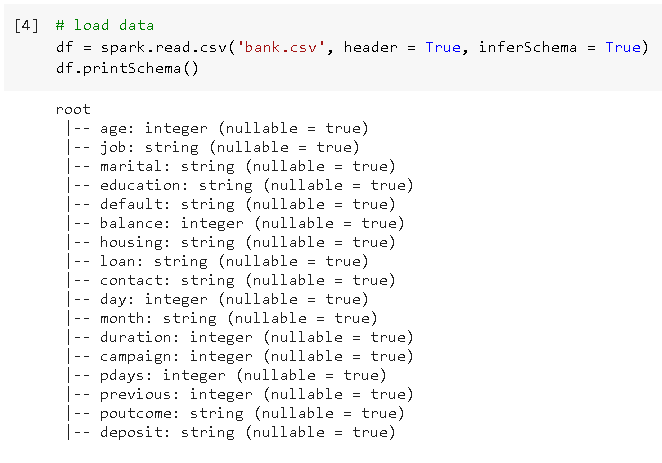
Machine Learning

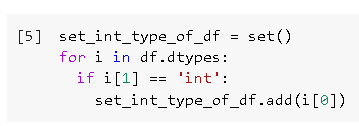
Tạo môi trường pyspark:

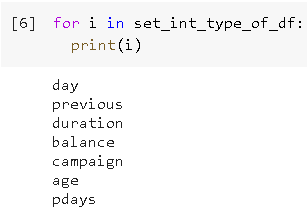


Load data:

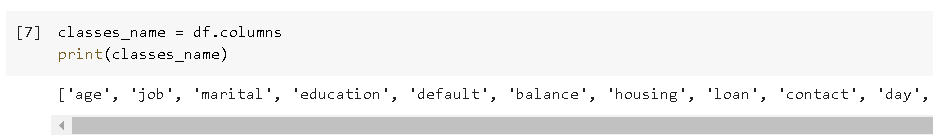


Tạo 1 set() chứa các field có dạng số:

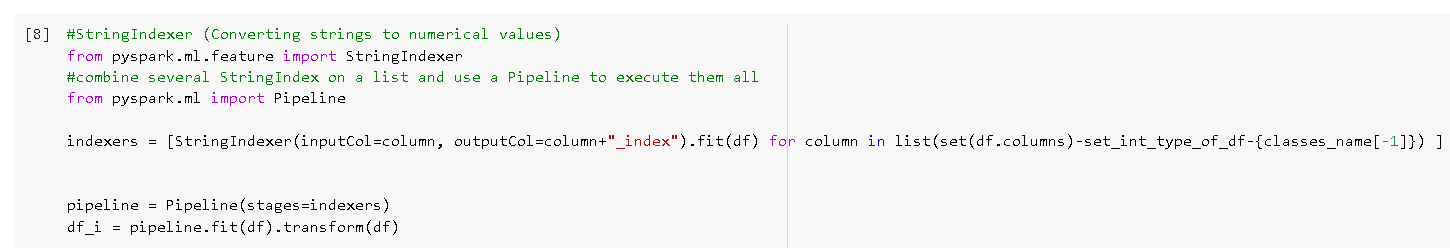


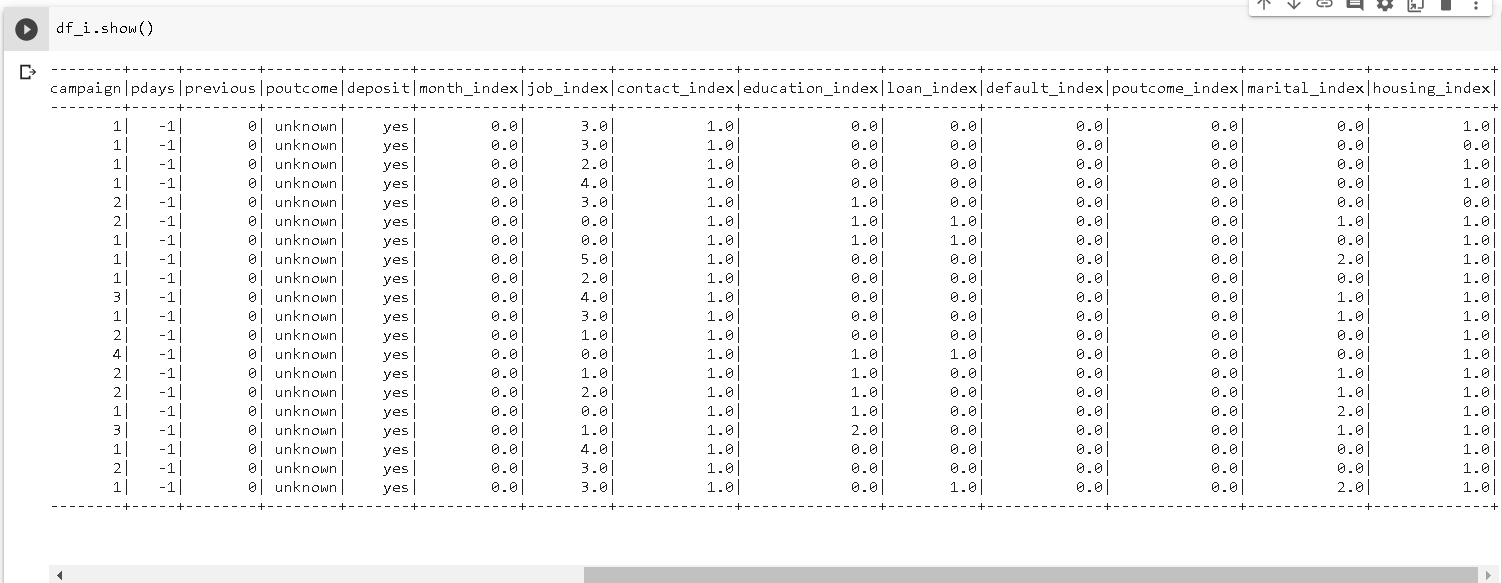


Lấy tất cả columns trong dataframe:

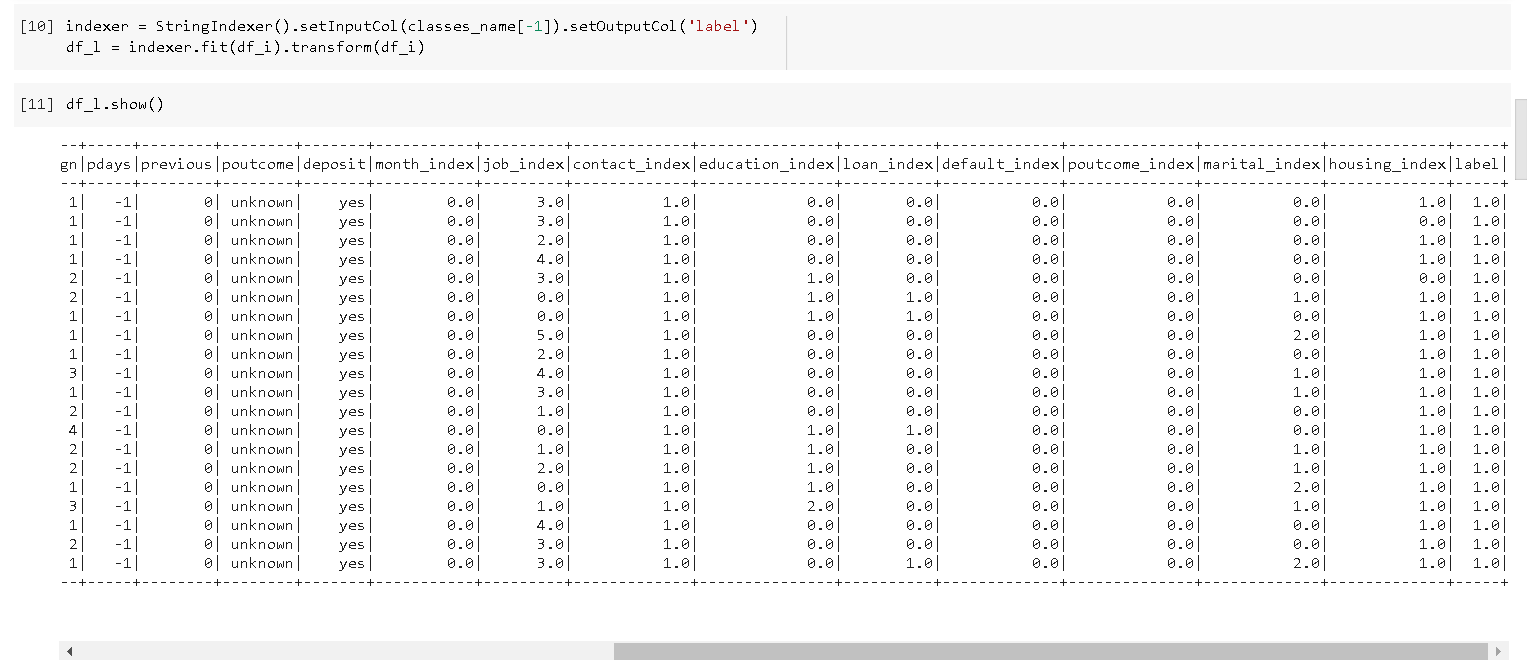


Sử dụng StringIndexer để chuyển những columns dạng string sang dạng index:

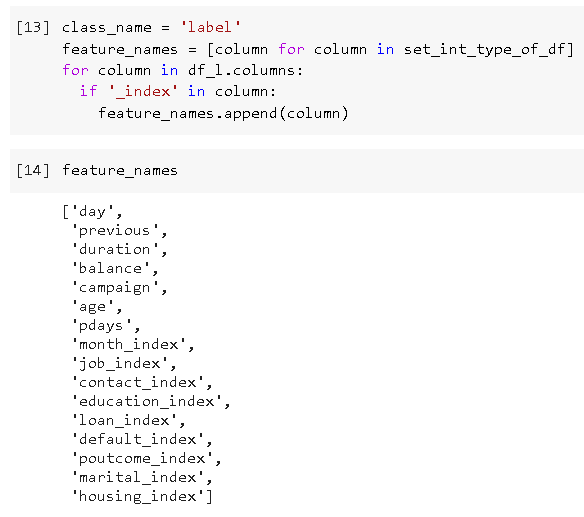




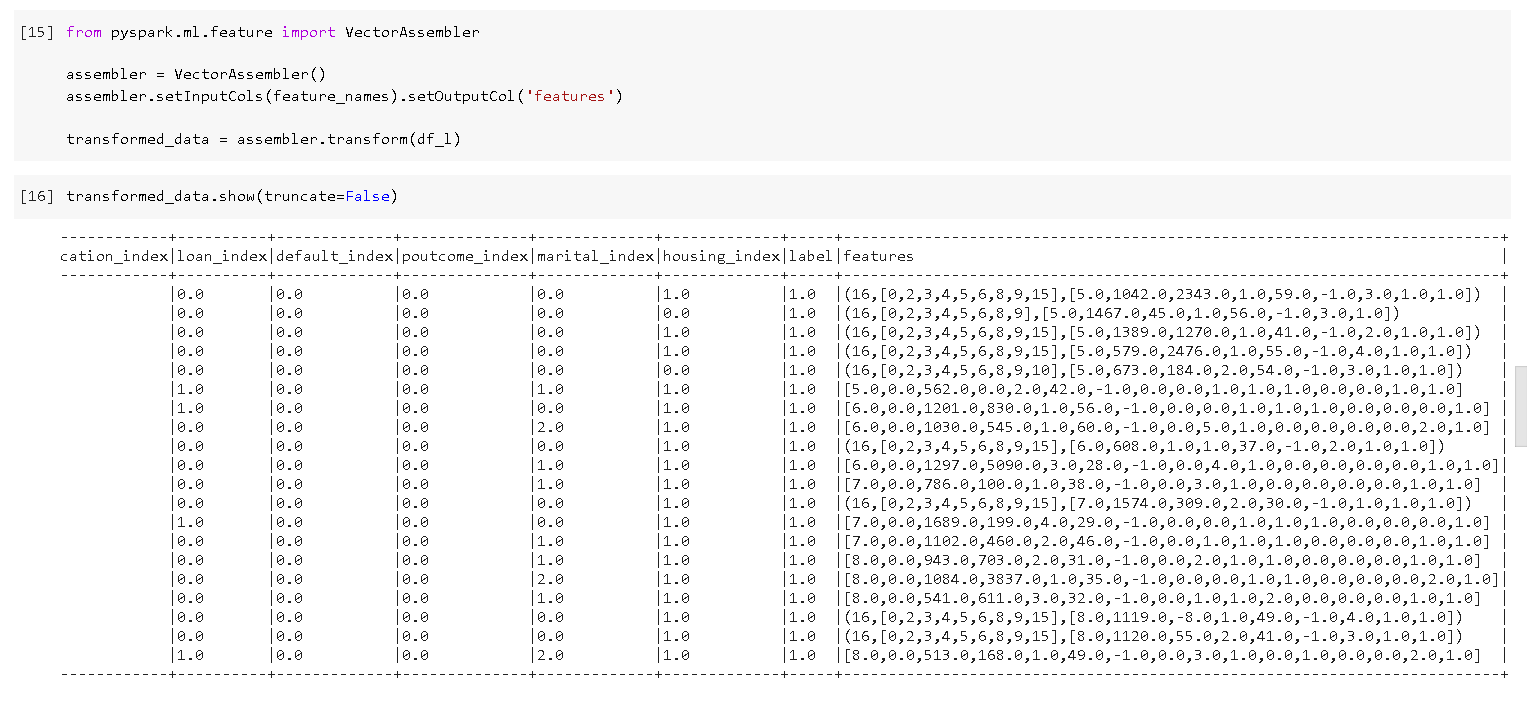
Chọn label là column cuối cùng của dataframe (ở trong ví dụ thì label là deposit)



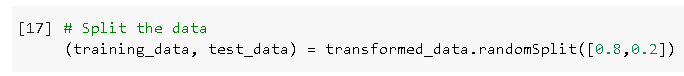
Lấy các tên features của dataframe:



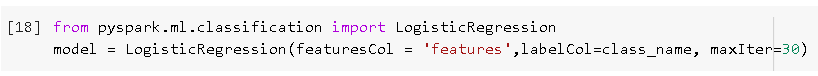
Chuyển các features sang vector



Tách dữ liệu 1 phần cho train, 1 phần cho test:



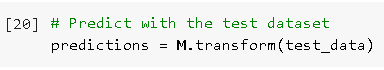
Tạo model với features column, label column, số lần lặp tối đa:



Train model:



Dự đoán của model với tập dữ liệu test:



Độ chính xác của model:

