

BÁO CÁO BÀI TẬP THỰC HÀNH MÔN HỌC MÁY THỐNG KÊ (DS102.K21)

Giảng viên phụ trách môn : Thầy NGUYỄN TẤN TRẦN MINH KHANG – VÕ DUY NGUYỄN

Họ và tên sinh viên: Võ Hoàng Thông

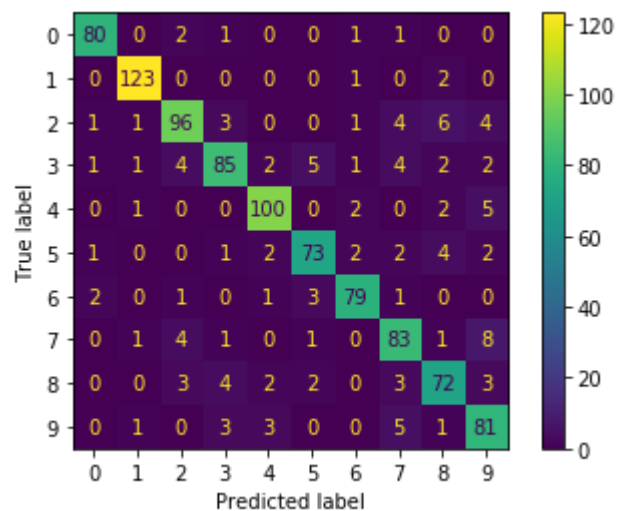
MSSV: 18521462

Bài toán: Nhận dạng chữ số viết tay.

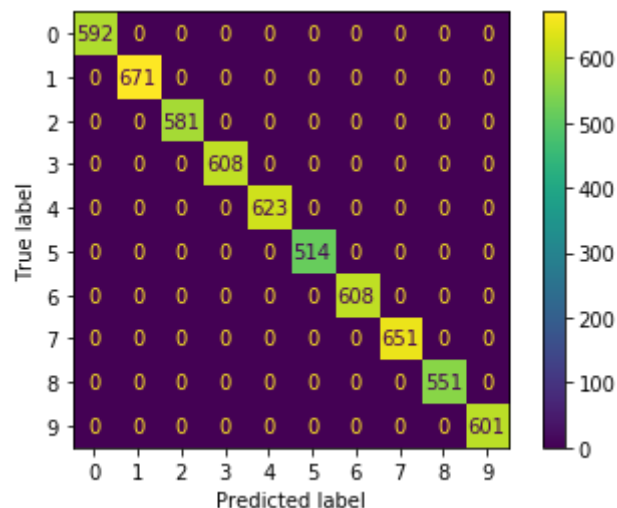
Phương pháp: Sử dụng thuật toán phân lớp Logistic Regression.

Source File: File đính kèm (Handwritten-digit-recognition.ipynb)

Confusion matrix của tập test



Confusion matrix của tập training



Đánh giá model

+ R^2 score training set = 1.0

+ R^2 score testing set = 0.872

Nhận xét:

Đánh giá độ đo trên tập training cho kết quả gần như tuyệt đối, tương đồng với confusion matrix.

→ Điều này cho thấy bộ dữ liệu huấn luyện cực kỳ tốt.

Sau khi kiểm tra trên tập testing, đánh giá độ đo cho kết quả khá tốt với tỉ lệ trên 85%, cho thấy không có nhiều sự khác nhau về độ khó trên tập train và tập test khi bộ dữ liệu phân chia độ khó khá đồng đều, chất lượng cả về số lượng và dữ liệu.