

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 K Nearest Neighbors.

Độc dữ liệu: File đính kèm (Handwritten-digit-recognition.ipynb)

Thống kê dữ liệu:

+ Tập train: 60000 mẫu.

+ Tập test: 10000 mẫu.

STT	Tên lớp	Số mẫu
1	0	5923
2	1	6742
3	2	5958
4	3	6131
5	4	5842
6	5	5421
7	6	5918
8	7	6265
9	8	5851
10	9	5949

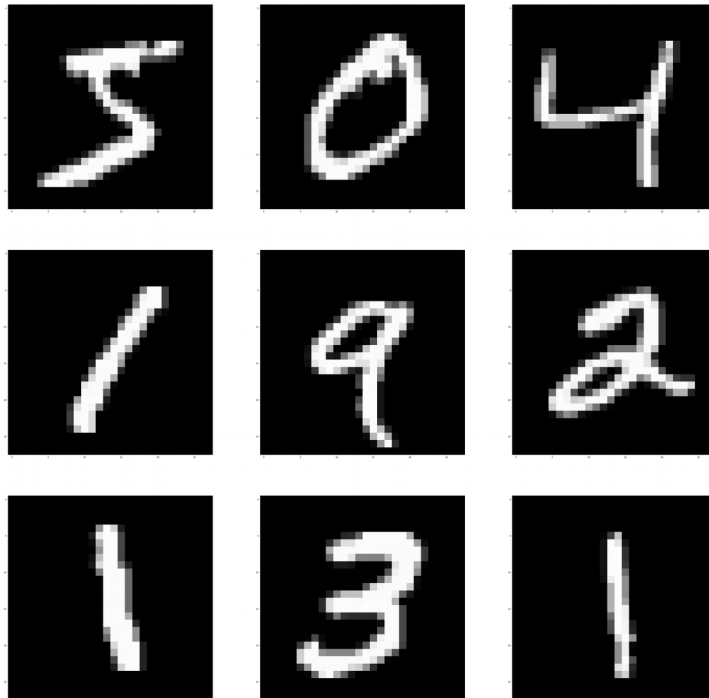
Số class và số ảnh trên mỗi class tập Training

STT	Tên lớp	Số mẫu
1	0	980
2	1	1135
3	2	1032
4	3	1010
5	4	982
6	5	892
7	6	958
8	7	1028
9	8	974
10	9	1009

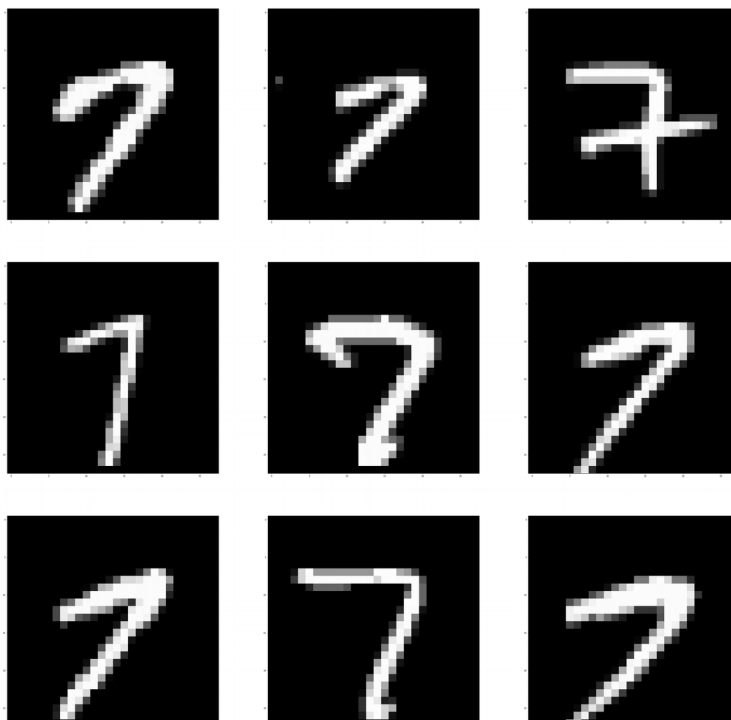
Số class và số ảnh trên mỗi class tập Testing

Visualize ảnh mẫu:

❖ Nhận dạng 9 chữ số viết tay đầu tiên của dataset:

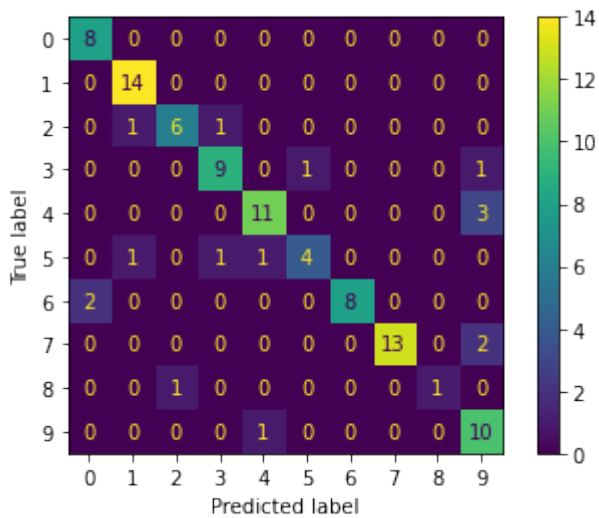


❖ Nhận dạng 9 chữ số 7 viết tay

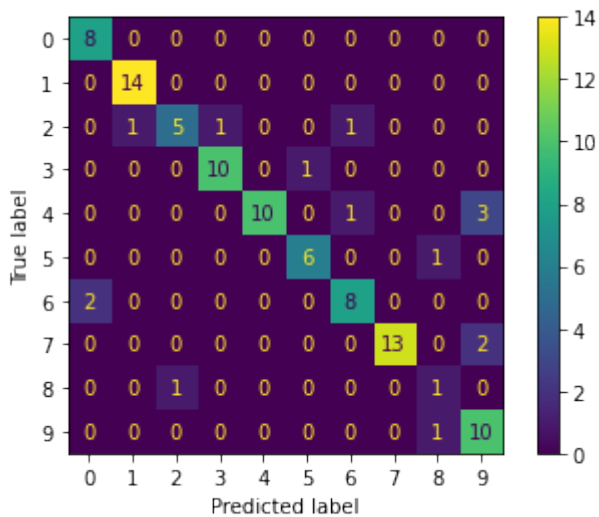


Kết quả thực nghiệm confusion matrix với các độ đo khoảng cách (minkowski, cosine, euclid) trong trường hợp $k = 1, 3, 5, 7$.

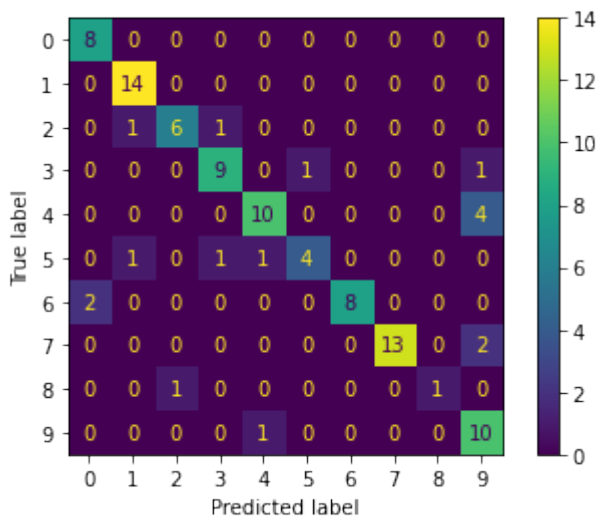
Confusion matrix with $k = 1$ and minkowski distance



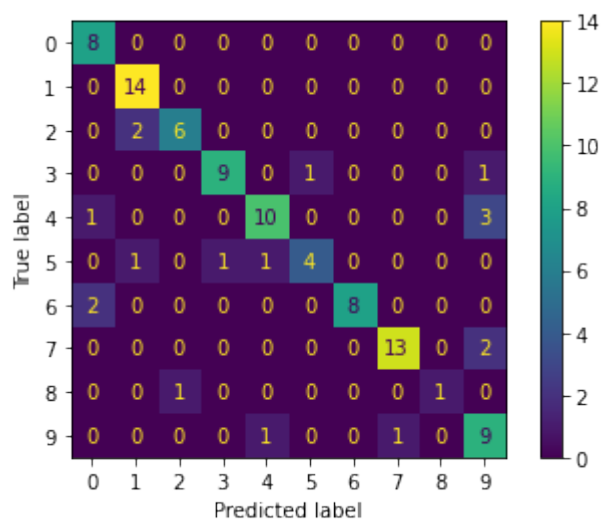
Confusion matrix with $k = 1$ and cosine distance



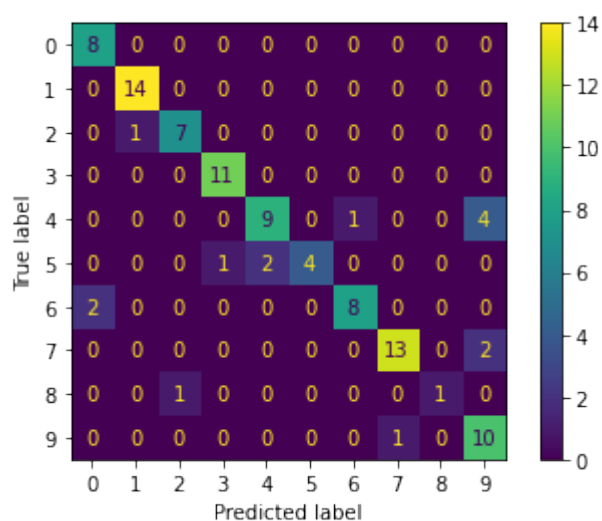
Confusion matrix with $k = 1$ and euclidean distance



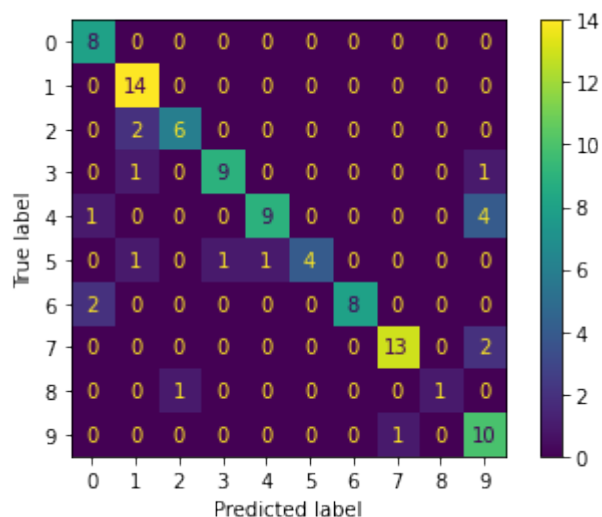
Confusion matrix with $k = 3$ and minkowski distance



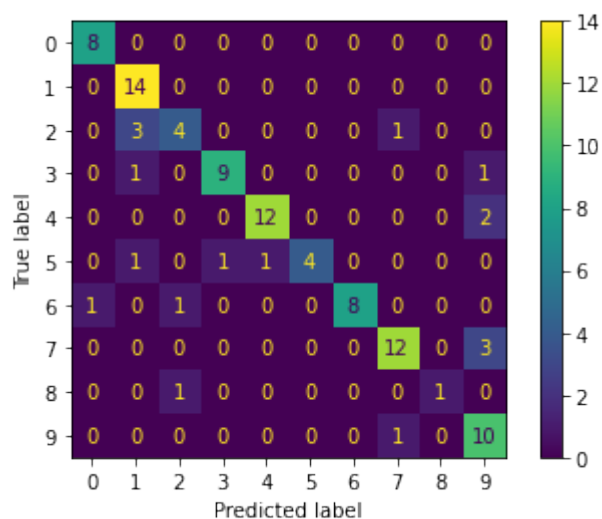
Confusion matrix with $k = 3$ and cosine distance



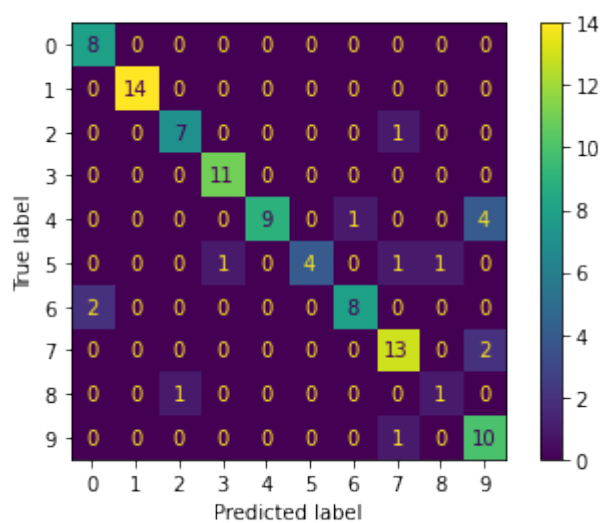
Confusion matrix with $k = 3$ and euclidean distance



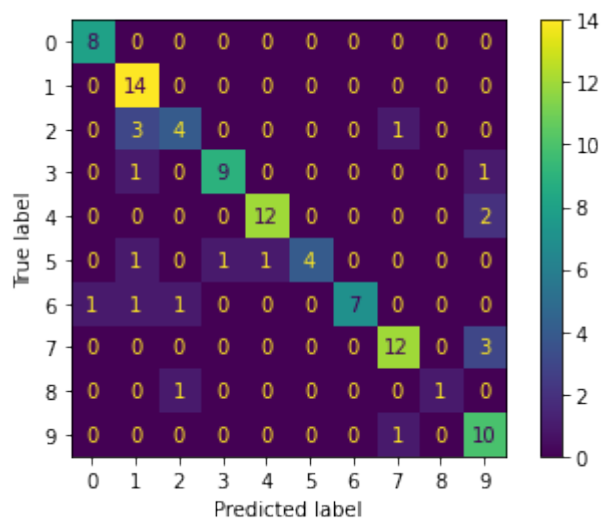
Confusion matrix with $k = 5$ and minkowski distance



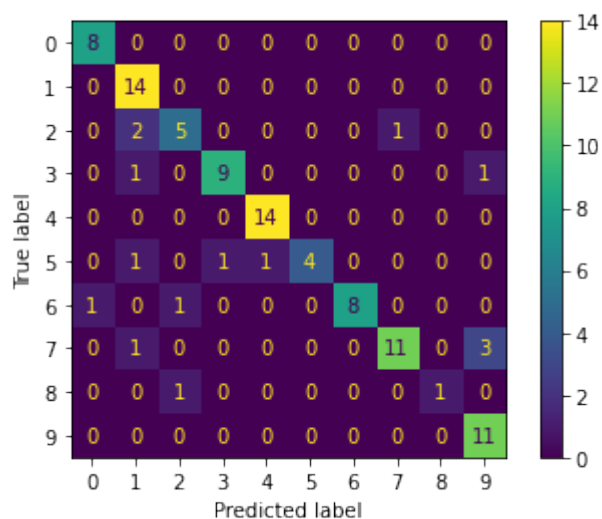
Confusion matrix with $k = 5$ and cosine distance



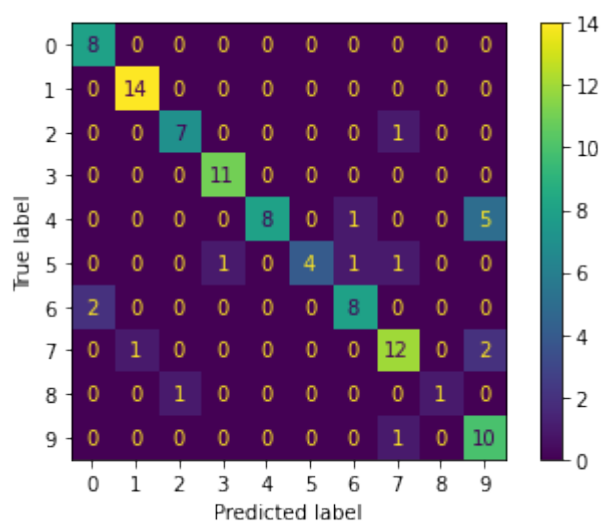
Confusion matrix with $k = 5$ and euclidean distance



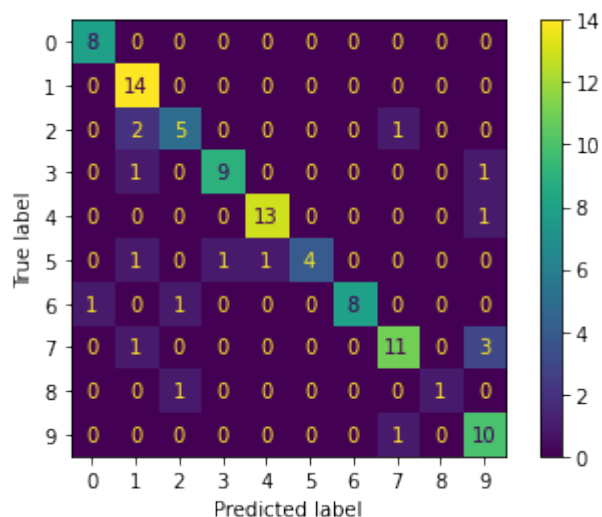
Confusion matrix with $k = 7$ and minkowski distance



Confusion matrix with $k = 7$ and cosine distance



Confusion matrix with $k = 7$ and euclidean distance



Bảng đánh giá kết quả

	k	Distance	Accuracy of values
0		1minkowski	0.84
1	1	cosine	0.85
2	1	euclidean	0.83
3		3minkowski	0.82
4	3	cosine	0.85
5	3	euclidean	0.82
6		5minkowski	0.82
7	5	cosine	0.85
8	5	euclidean	0.81
9		7minkowski	0.85
10	7	cosine	0.83
11	7	euclidean	0.83

Nhận xét:

+ Cả 3 độ đo khoảng cách đều cho kết quả tốt với độ chính xác trên 80%.

+ Với độ đo minkowski cho kết quả tốt nhất khi $k = 7$.

Với độ đo cosine cho kết quả tốt nhất khi $k = 1, 3, 5$.

Độ đo euclidean cho kết quả tốt nhất khi $k = 1$ và 7 .

+ Với 3 độ đo cho kết quả xấp xỉ khá tương đương nhau, ta nhận ra việc training model và dataset của tập training khá tốt. Bên cạnh đó, thuật toán được chọn hoạt động khá tốt.