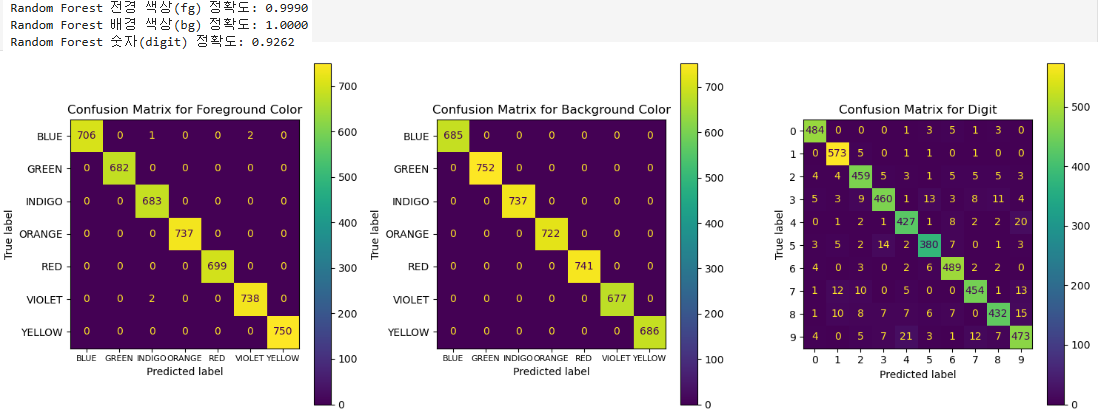
먼저 MNIST Data에 색상을 입힌 후 모델에 적용하기 위해 X\_color에 적재함.

학습을 위해 X\_color를 (60000, 28, 28, 3)에서 (60000, 2352)로 flatten.

이후 적절한 모델을 찾기 위해 다양한 모델 사용. (이때 전처리는 하지 않은 상태)



적합한 결과를 보여줌. Confusion Matrix를 통해 색상이 입혀지다보니 글씨 예측 정확도가 더 낮아진 부분을 확인할 수 있었다

어떻게 그렇게 판단할 수 있었는지 -> <https://yeoulcoding.tistory.com/102>  
= 일반 MNIST를 RF로 분류했을 때 정확도가 96% 이상이 나왔고  
추가적으로 데이터셋을 5배로 증강시켜보니 0.01%정도 정확도가 향상함.

아래는 여러가지 모델을 적용했던 기록들

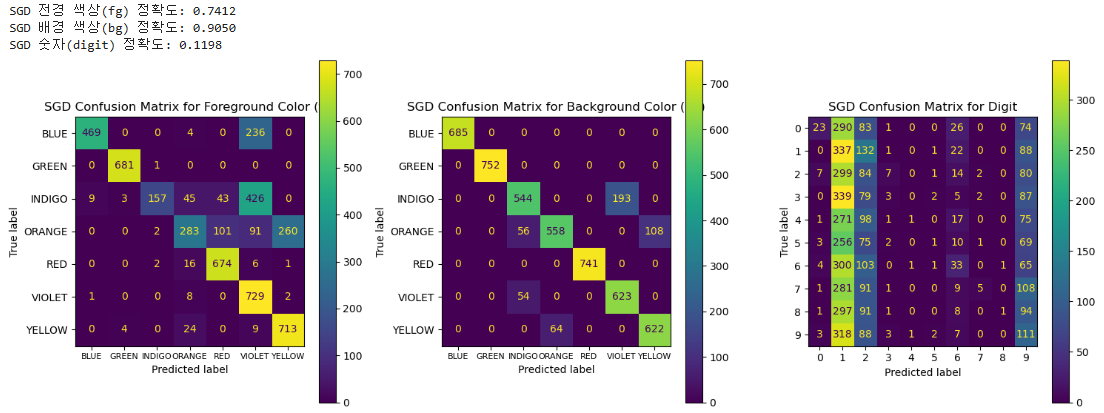
SVM

SVM 전경 색상(fg) 정확도: 0.9924  
SVM 배경 색상(bg) 정확도: 1.0000  
SVM 숫자(digit) 정확도: 0.8200  
소요시간 5분 – 이진분류를 통해 다중분류를 시도(OvO or OvR)했기 때문에 시간이 오래걸렸을 것이라고 생각.

KNN

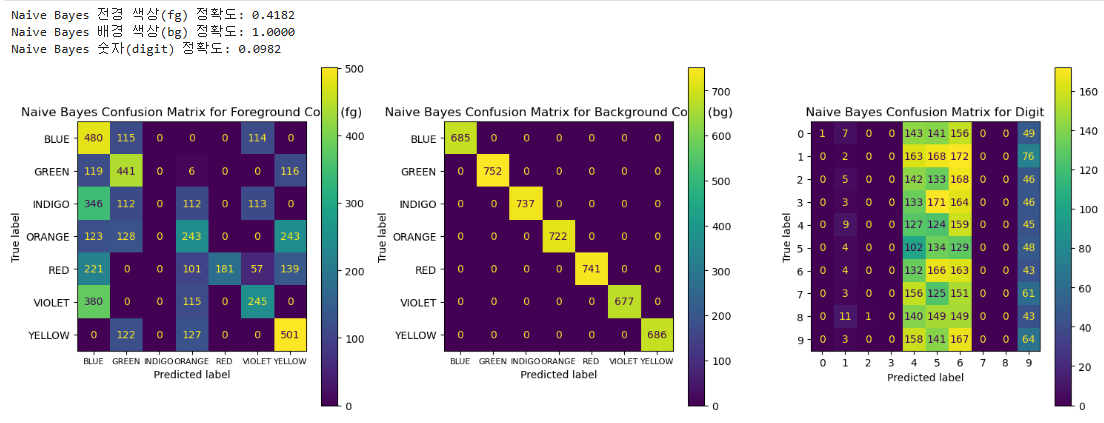
KNN 전경 색상(fg) 정확도: 0.8382  
KNN 배경 색상(bg) 정확도: 1.0000  
KNN 숫자(digit) 정확도: 0.8894  
소요시간 짧음 전경은 아무래도 색이 포함된 부분이 적다보니 정확도가 낮음. 그러나 BG 분류 시 높은 정확도를 보여줌. Label별로 모델을 다르게 한다면 고려해보아도 좋을 것 같음 (색상만 구분하는 모델이면 전처리를 색 스펙트럼으로 or 많은 색 그다음으로 많은 색 2개를 뽑아 남기고 나머지는 없애면 연산량 크게 줄일 수 있음)  
4/9 , 5/3 HW1에서 했던 오류와 비슷한 모습을 보임.

SGD



쓰레기?  
근데 모델 선정 전 예상했다시피 Violet이 Blue, Indigo 등 푸른 계열과 비슷해서 잘 맞추지 못한거같음.  
글자는 색상이 있다보니 경계가 모호해서 다 1로 판단하는거같음. 이건 전처리하면 개선될거같은데 그냥 이 모델이 Colored MNIST에 적합하지 않다고 판단함.

Naive Bayes



배경만 잘 맞추는 쓰레기 2