
SEMESTER 1 – MACHINE LEARNING

# FASHION – MNIST:

# Pixels speak style's language

Myrsini Karakasoglou

Antigoni Moira

Dimitris Tselentis

# The Fashion-MNIST dataset

zalando


0 - T-shirt/top
1 - Trouser
2 - Pullover
3 - Dress
4 - Coat
5 - Sandal
6 - Shirt
7 - Sneaker
8 - Bag
9 - Ankle boot

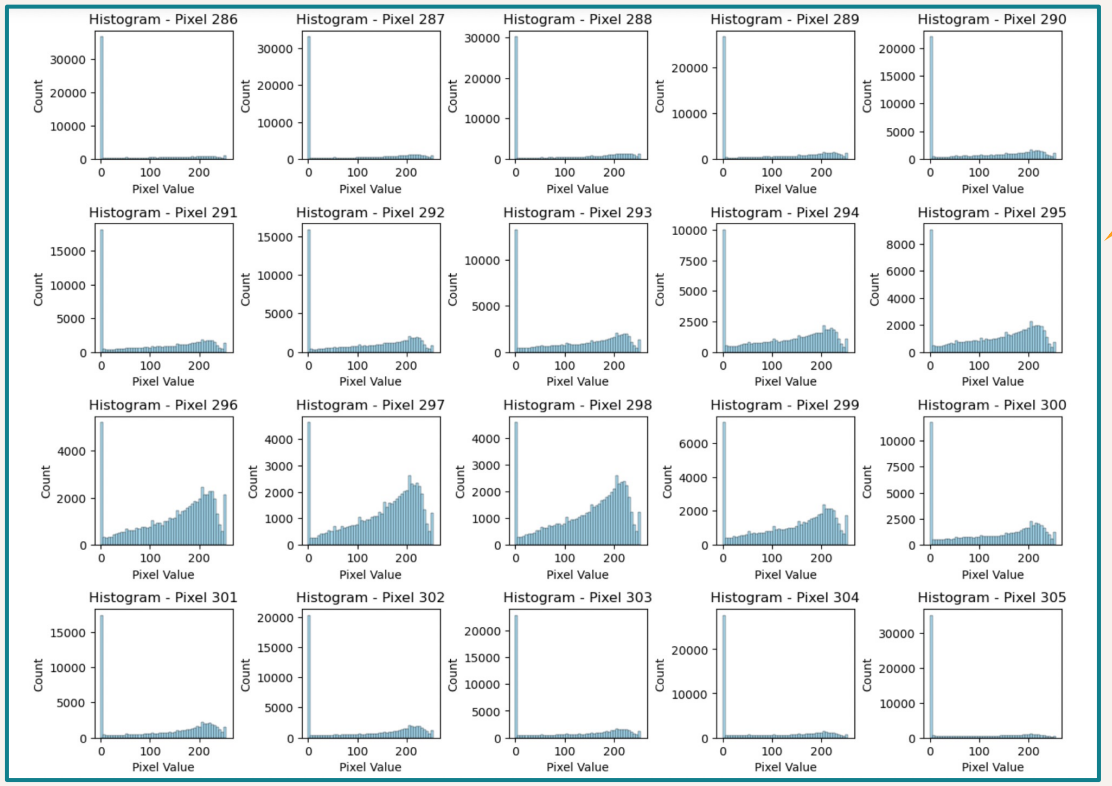
28 x 28 pixels

784 features

pixel values: 0 - 255

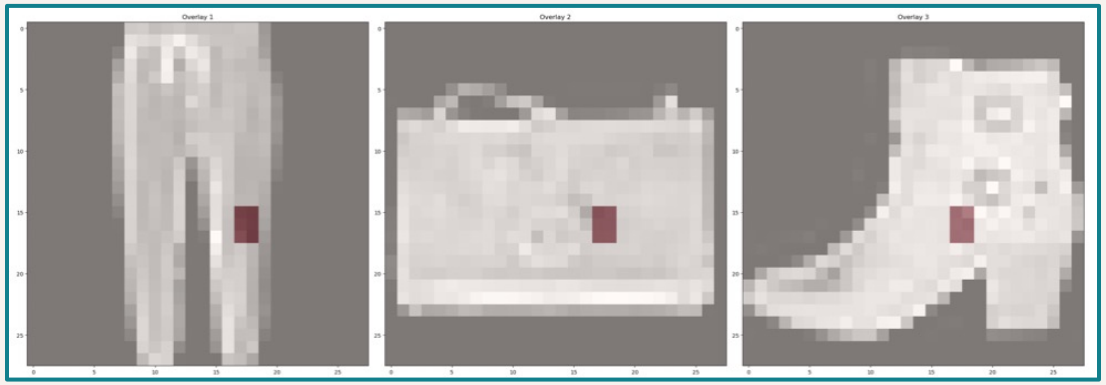
60,000 training images
10,000 test images

# Locating some “informative” pixels

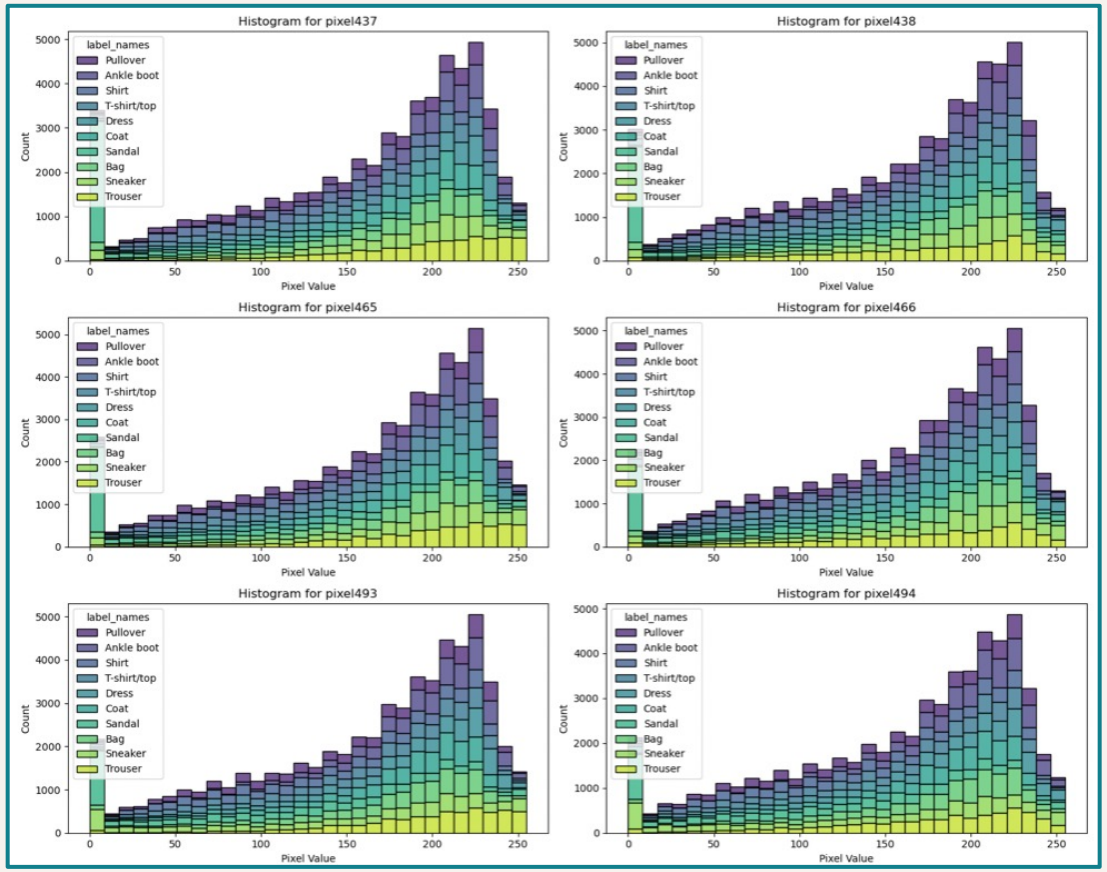


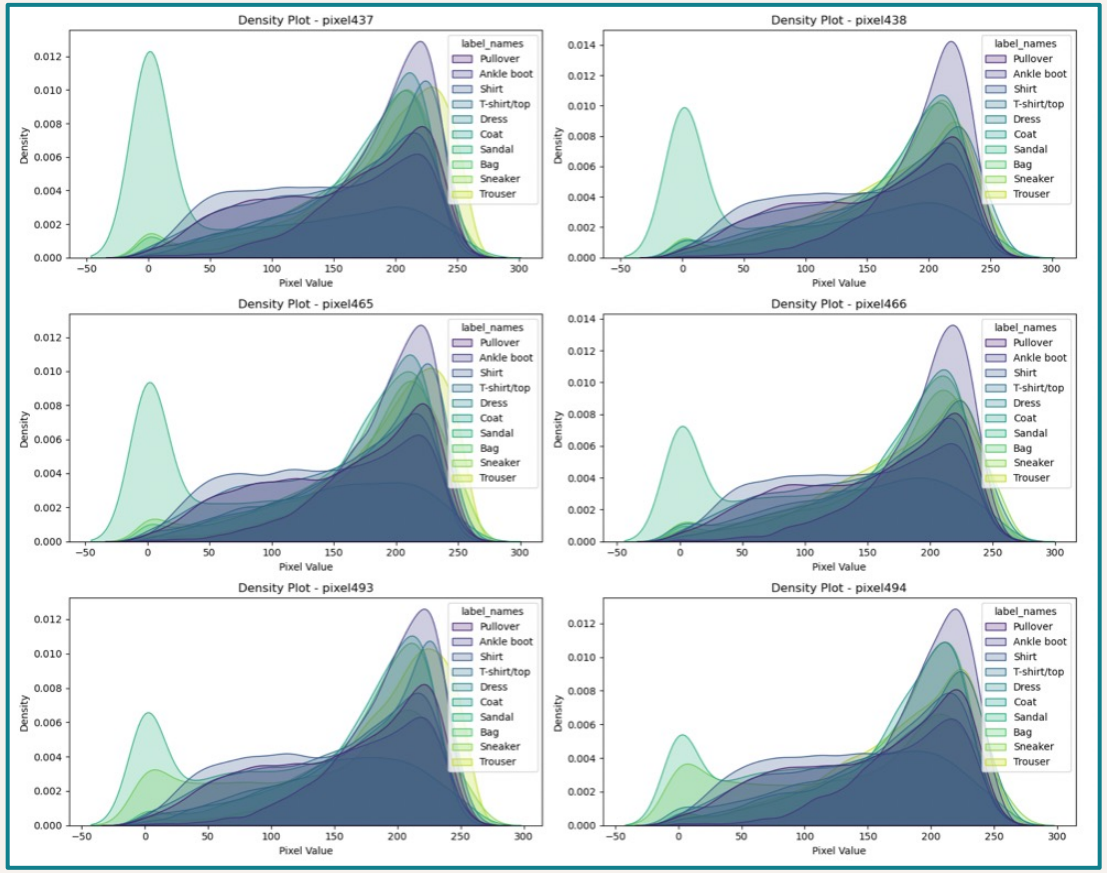

Uniform class distribution

pixel 437 | pixel 438 | pixel 465
pixel 466 | pixel 493 | pixel 494

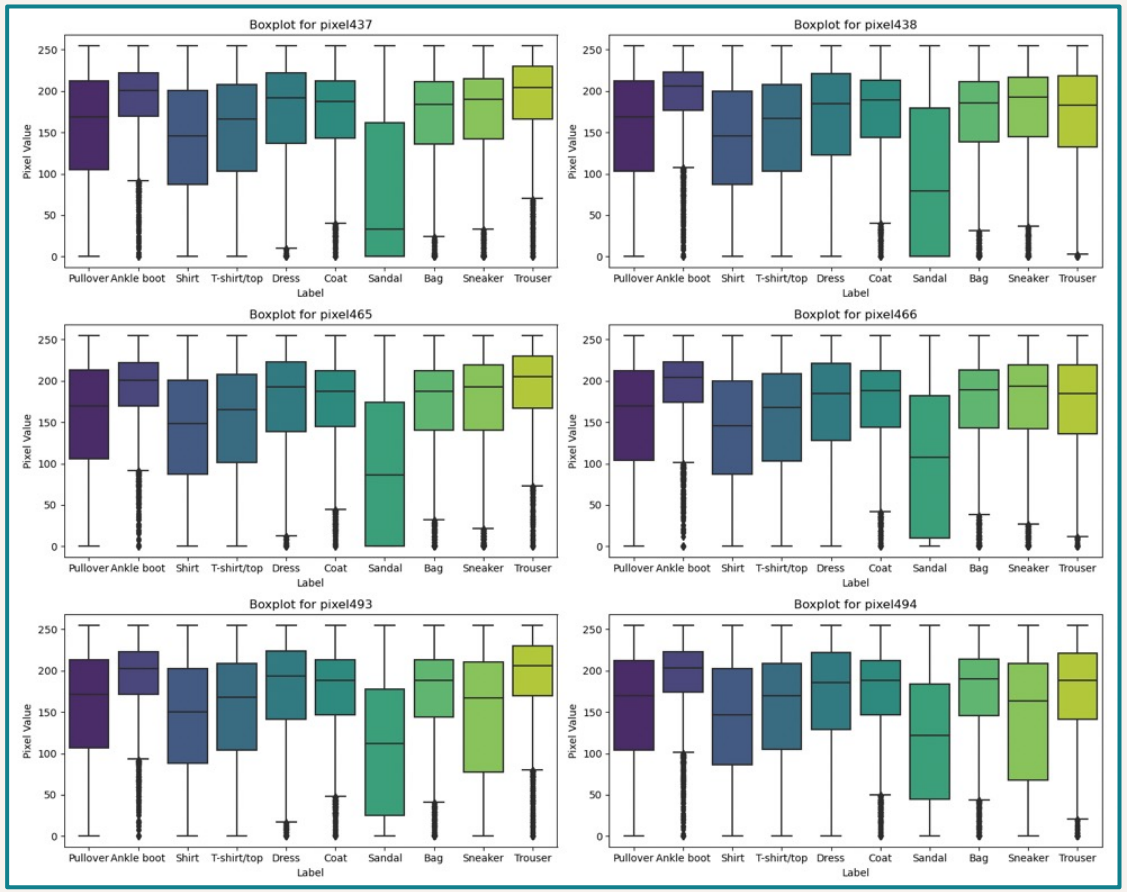


# A peak into the “informative” features

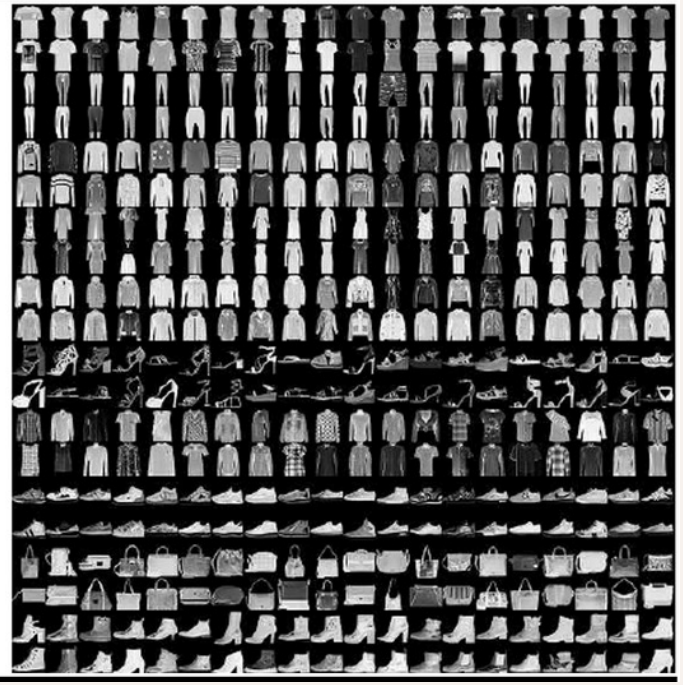


# Further insights into the data...

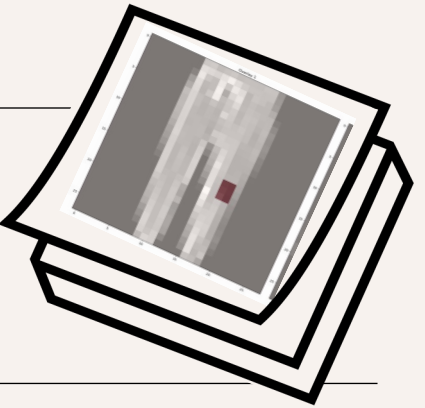


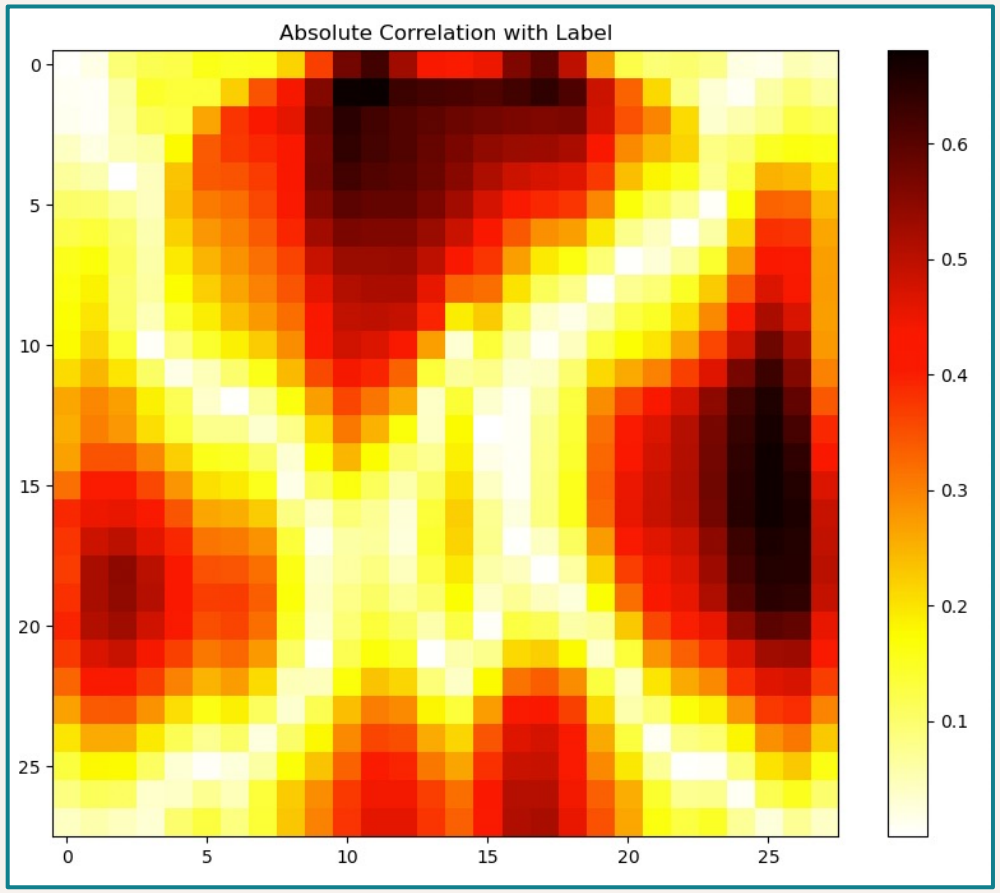

| Label | Description | Examples |
|---|---|---|
| 0 | T-Shirt/Top | |
| 1 | Trouser | |
| 2 | Pullover | |
| 3 | Dress | |
| 4 | Coat | |
| 5 | Sandals | |
| 6 | Shirt | |
| 7 | Sneaker | |
| 8 | Bag | |
| 9 | Ankle boots | |



→ *Zhao, Jin & Jiao, Licheng. (2019). Sparse Deep Tensor Extreme Learning Machine for Pattern Classification. IEEE Access. PP. 1-1. 10.1109/ACCESS.2019.2924647.*

# Locating important pixels...



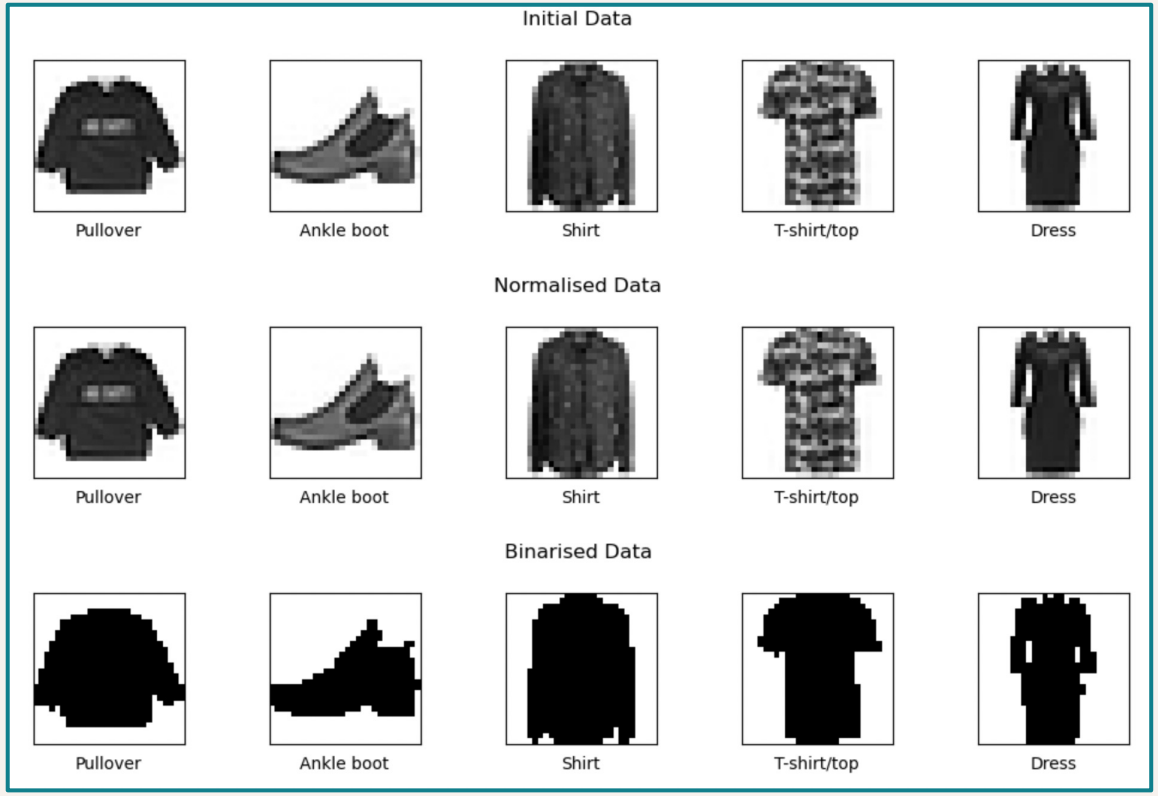



- Empty pixels?
- Collectively correlated pixels?
- Redundant pixels?

# Get the data ready for the algorithms

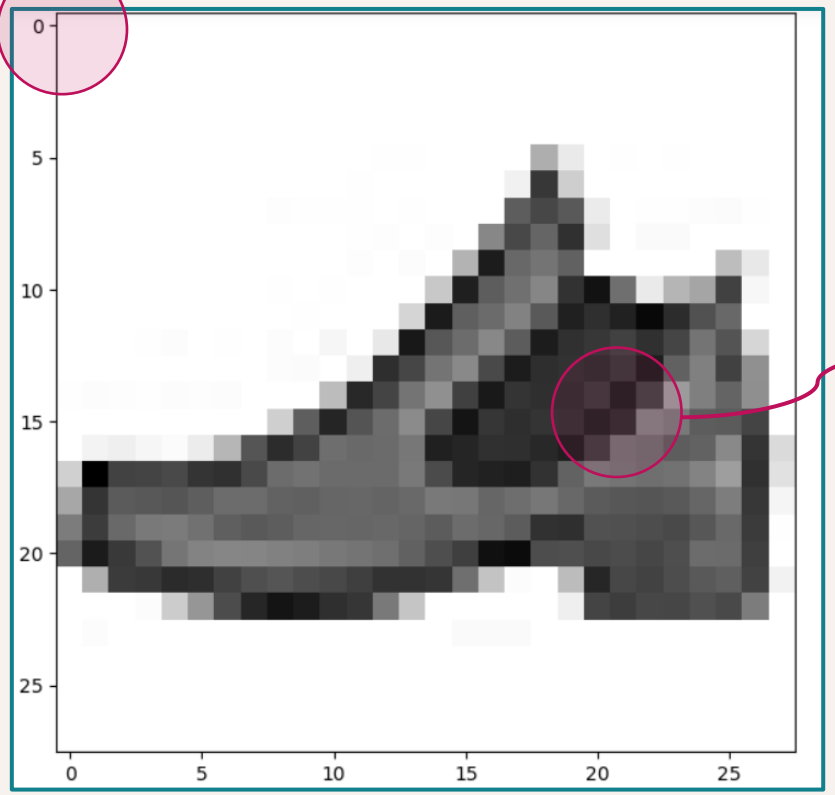

Normalisation → [0,1]

Binarisation → 0 or 1



# Locating “good” features

Irrelevant Pixels

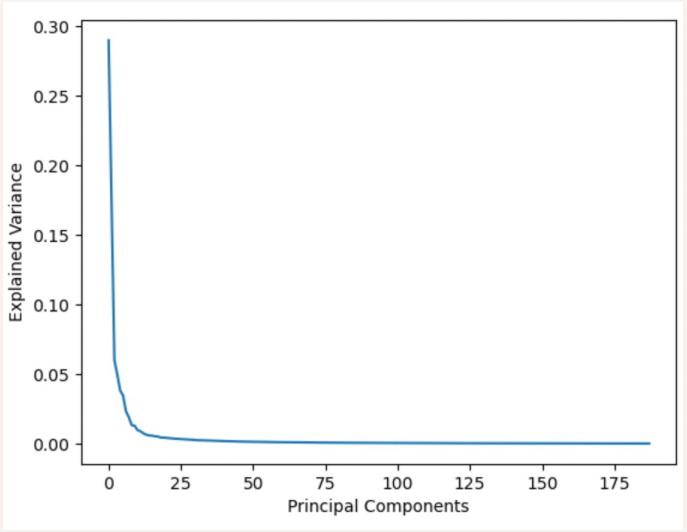



Redundant Pixels

# Feature Engineering – PCA [1]

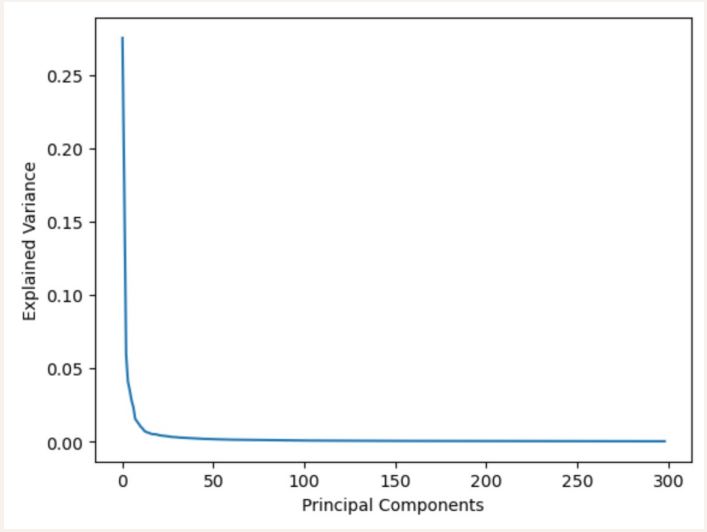
95% explained variance

## Normalised Data




188 principal components

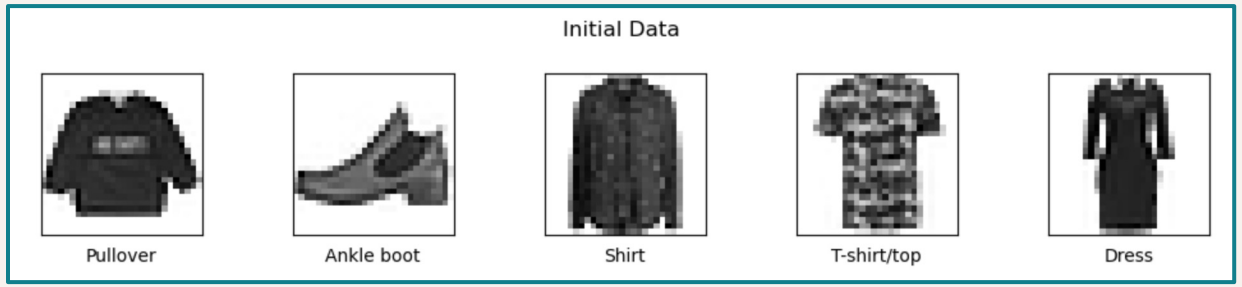
## Binarised Data




299 principal components

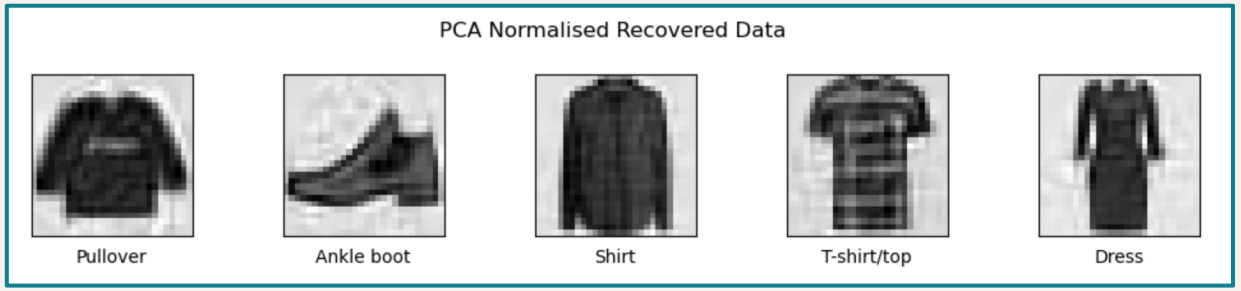
# Feature Engineering – PCA [2]

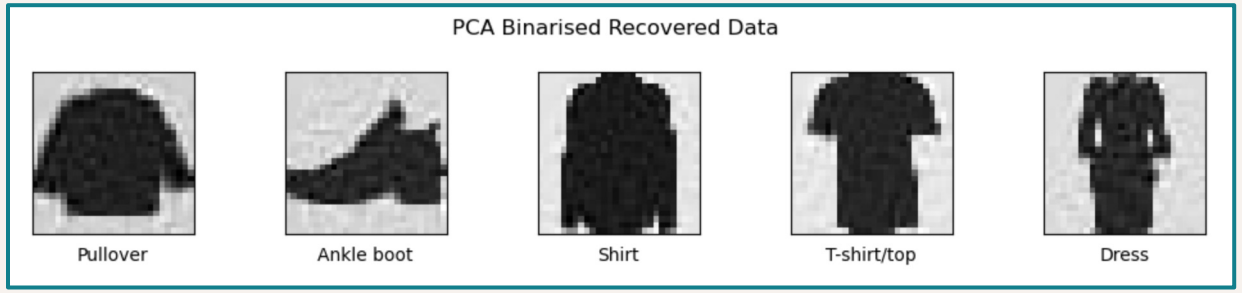
95% explained variance




784 features

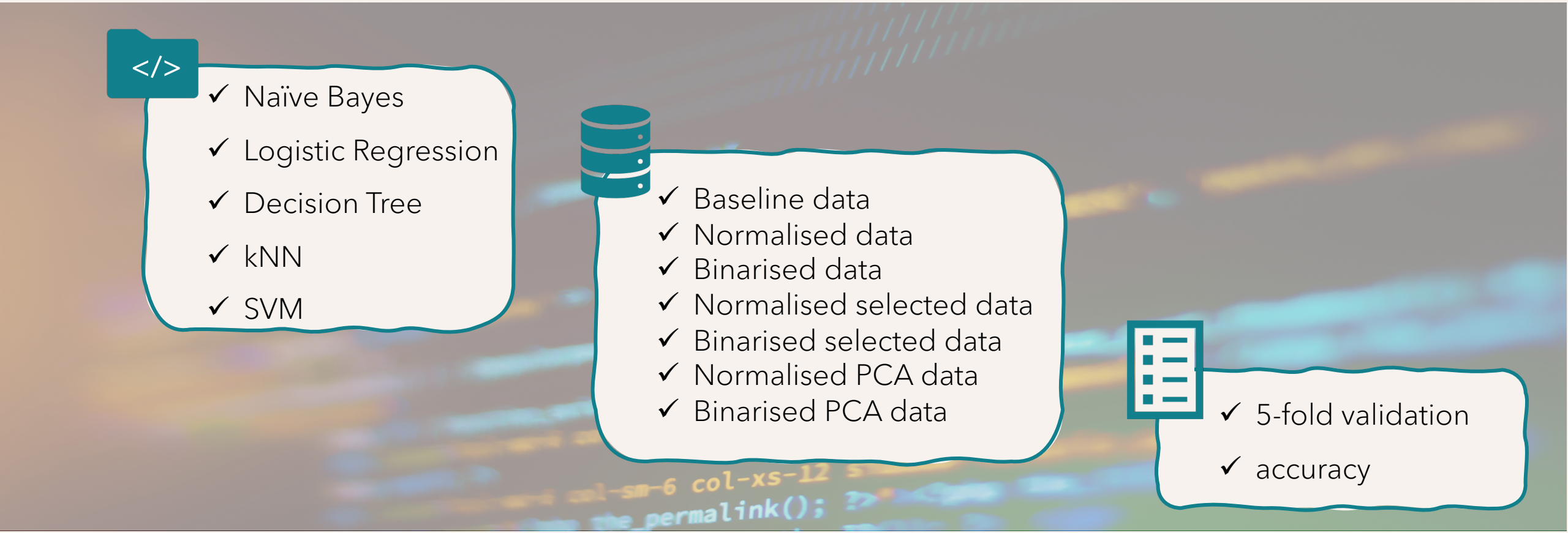
188
principal components




299
principal components

# The algorithm selection phase



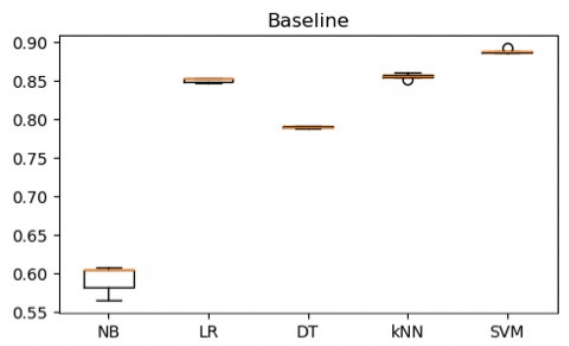
- ✓ Naïve Bayes
- ✓ Logistic Regression
- ✓ Decision Tree
- ✓ kNN
- ✓ SVM

- ✓ Baseline data
- ✓ Normalised data
- ✓ Binarised data
- ✓ Normalised selected data
- ✓ Binarised selected data
- ✓ Normalised PCA data
- ✓ Binarised PCA data

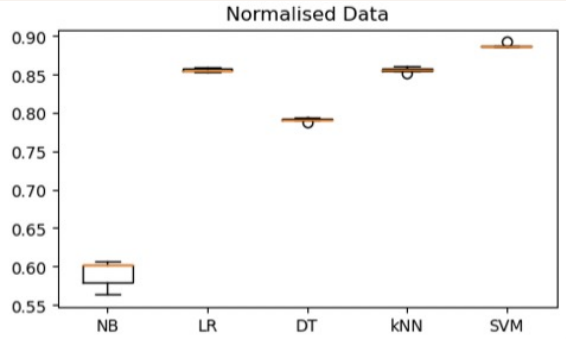
- ✓ 5-fold validation
- ✓ accuracy

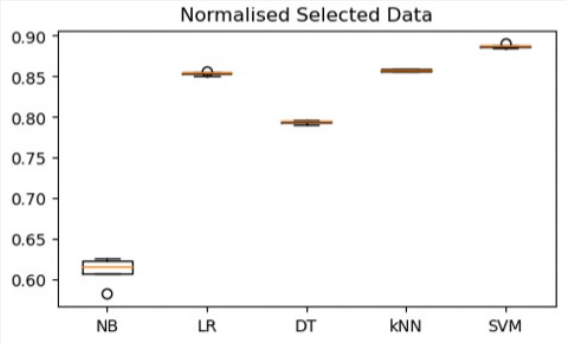
# Comparing the algorithms

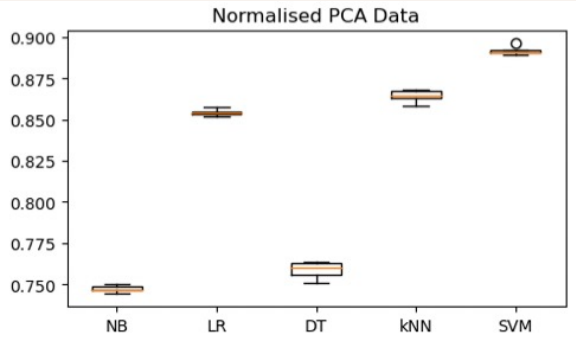
CONSTRAINTS
- time
- resources

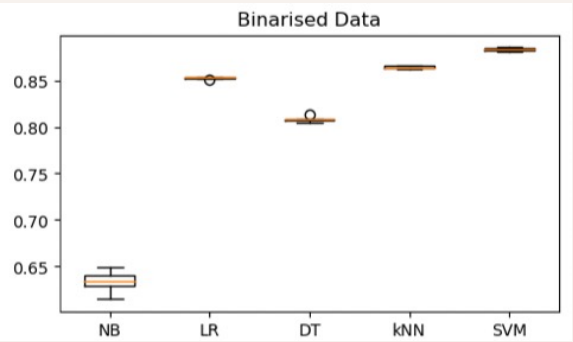


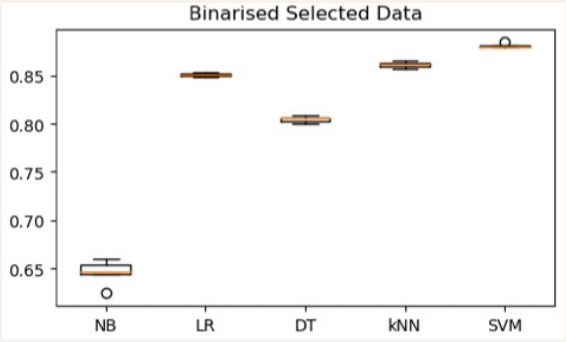

✓ SVM
✓ normalised
✓ PCA

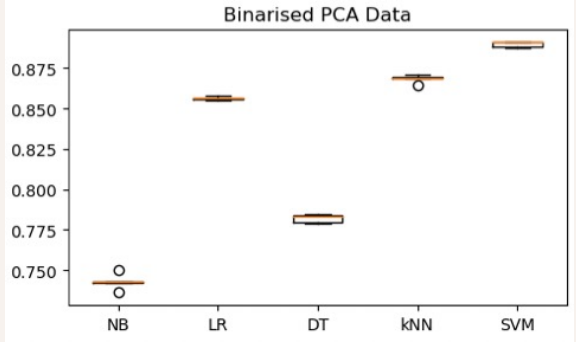


# Optimisation of SVM & Ensembles

Initial performance Acc ≈0.892

## Randomised Search

- ✓ kernel: rbf
- ✓ C: 10

Accuracy: ≈0.905

## Random Forest

- ✓ n_estimators: 200
- ✓ min_samples_split: 10
- ✓ min_samples_leaf: 2
- ✓ max_features: sqrt
- ✓ max_dept: 30
- ✓ Criterion: gini
- ✓ Bootstrap: false

Accuracy: ≈0.87

## XGBoost

- ✓ max_depth: 7
- ✓ learning_rate: 0.3
- ✓ lamda: 1
- ✓ gamma: 0
- ✓ colsample_bytree: 0.8

Accuracy: ≈0.89

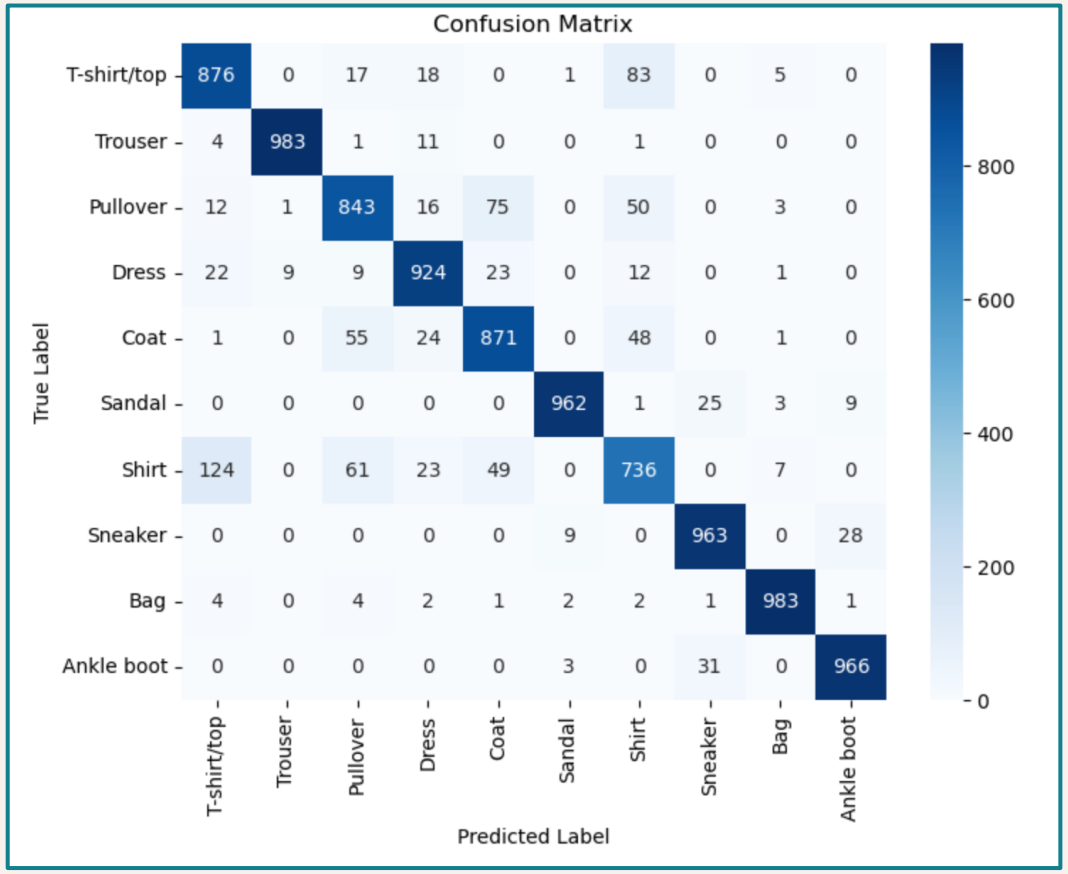
# Testing the performance of tuned SVM

Normalised data
PCA (kernel: rbf, C: 10)

Classification Report for SVM:

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.84 | 0.88 | 0.86 | 1000 |
| 1 | 0.99 | 0.98 | 0.99 | 1000 |
| 2 | 0.85 | 0.84 | 0.85 | 1000 |
| 3 | 0.91 | 0.92 | 0.92 | 1000 |
| 4 | 0.85 | 0.87 | 0.86 | 1000 |
| 5 | 0.98 | 0.96 | 0.97 | 1000 |
| 6 | 0.79 | 0.74 | 0.76 | 1000 |
| 7 | 0.94 | 0.96 | 0.95 | 1000 |
| 8 | 0.98 | 0.98 | 0.98 | 1000 |
| 9 | 0.96 | 0.97 | 0.96 | 1000 |
| accuracy | | | 0.91 | 10000 |
| macro avg | 0.91 | 0.91 | 0.91 | 10000 |
| weighted avg | 0.91 | 0.91 | 0.91 | 10000 |

60,000 training images
10,000 test images



"Elegance is elimination."
Cristóbal Balenciaga