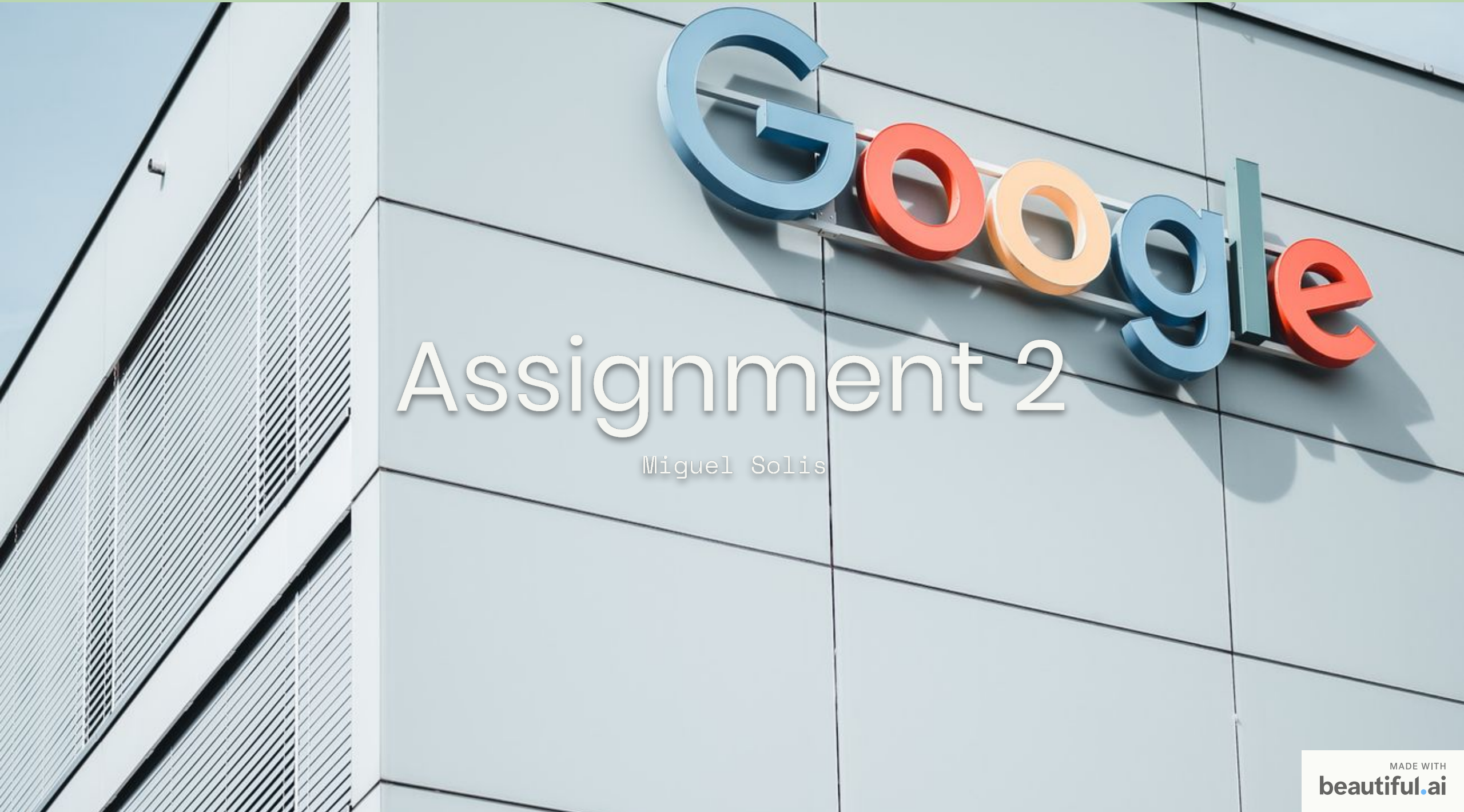
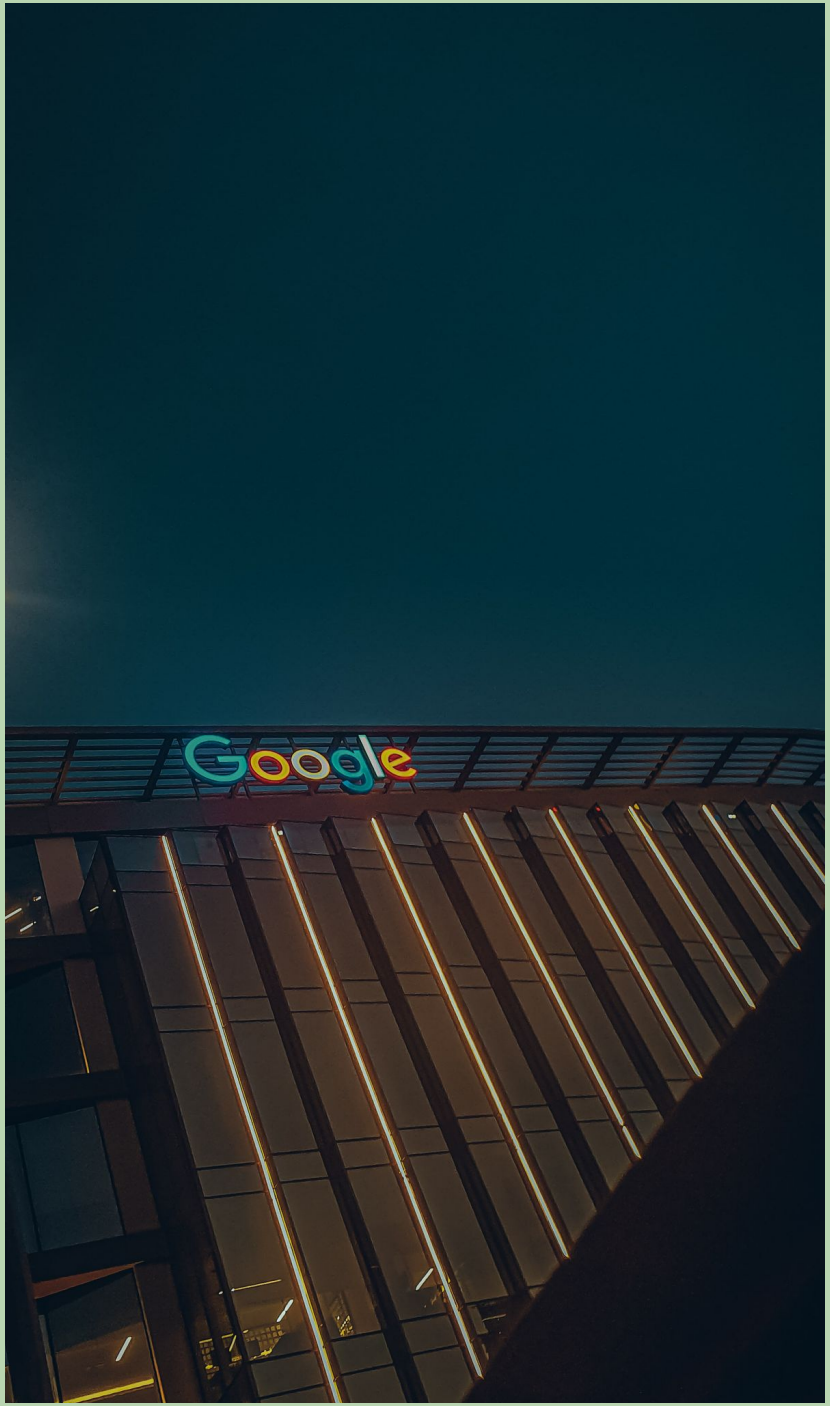


Assignment 2

Miguel Solis

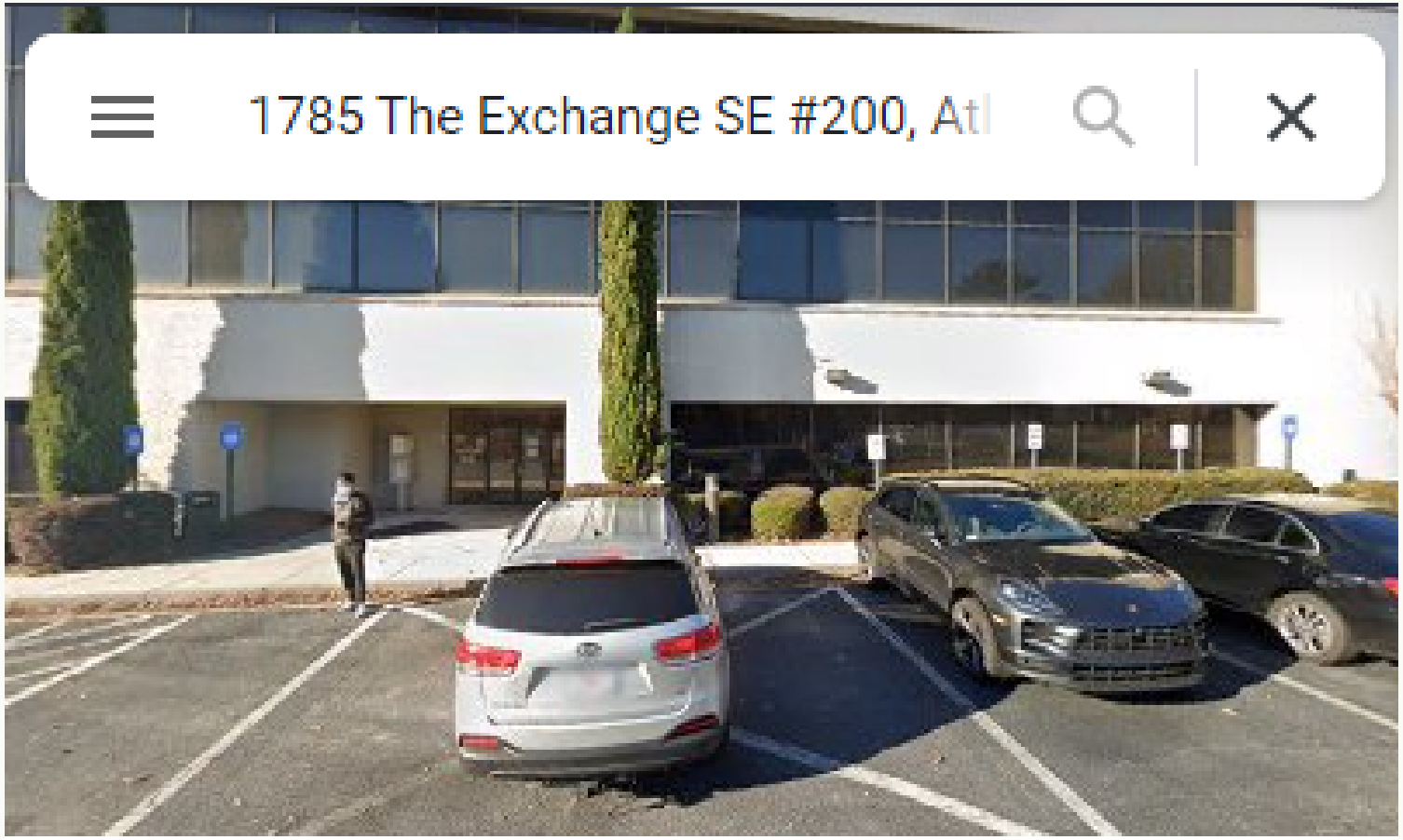


It's 2013 and we are all Google employees

CONTACT US >

Google Maps

Adress input



1785 The Exchange SE #200

Address numbers

98%

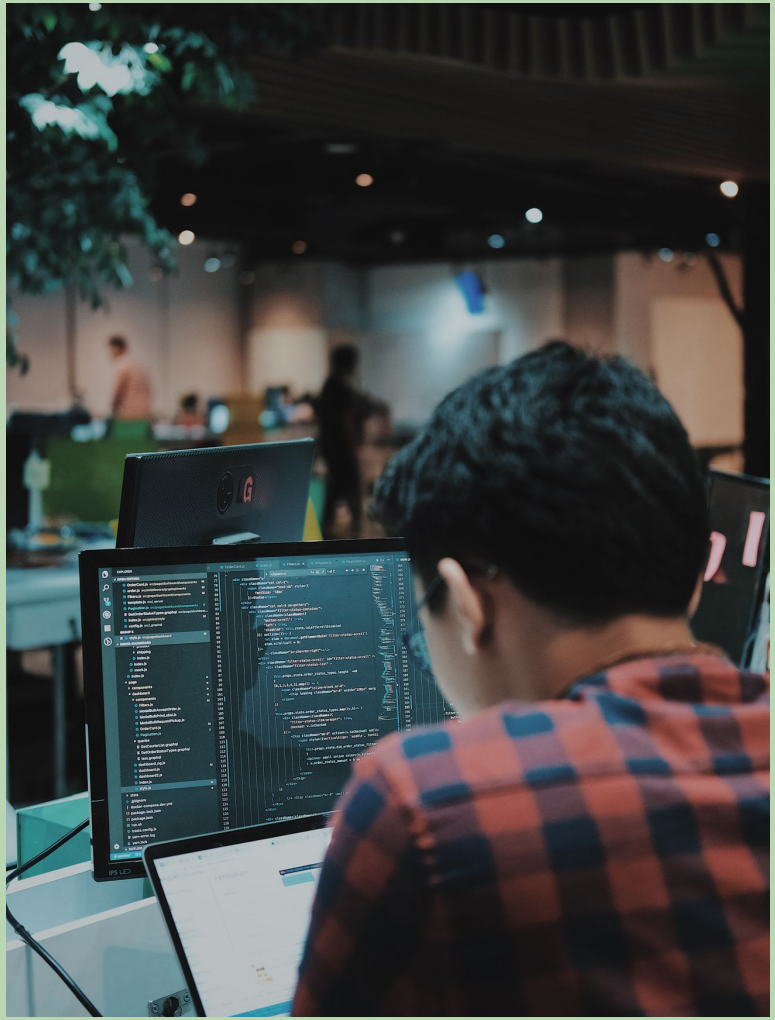
Human input

Can we get at least the same accuracy with a machine learning model?

The data set

- 2 datasets (Train and Test)
- Train = 60,000 observations
- Test = 10,000 observations
- Unnamed:0, index, labels, + 784 pixels with values from 0 to 256.

Machine learning models used

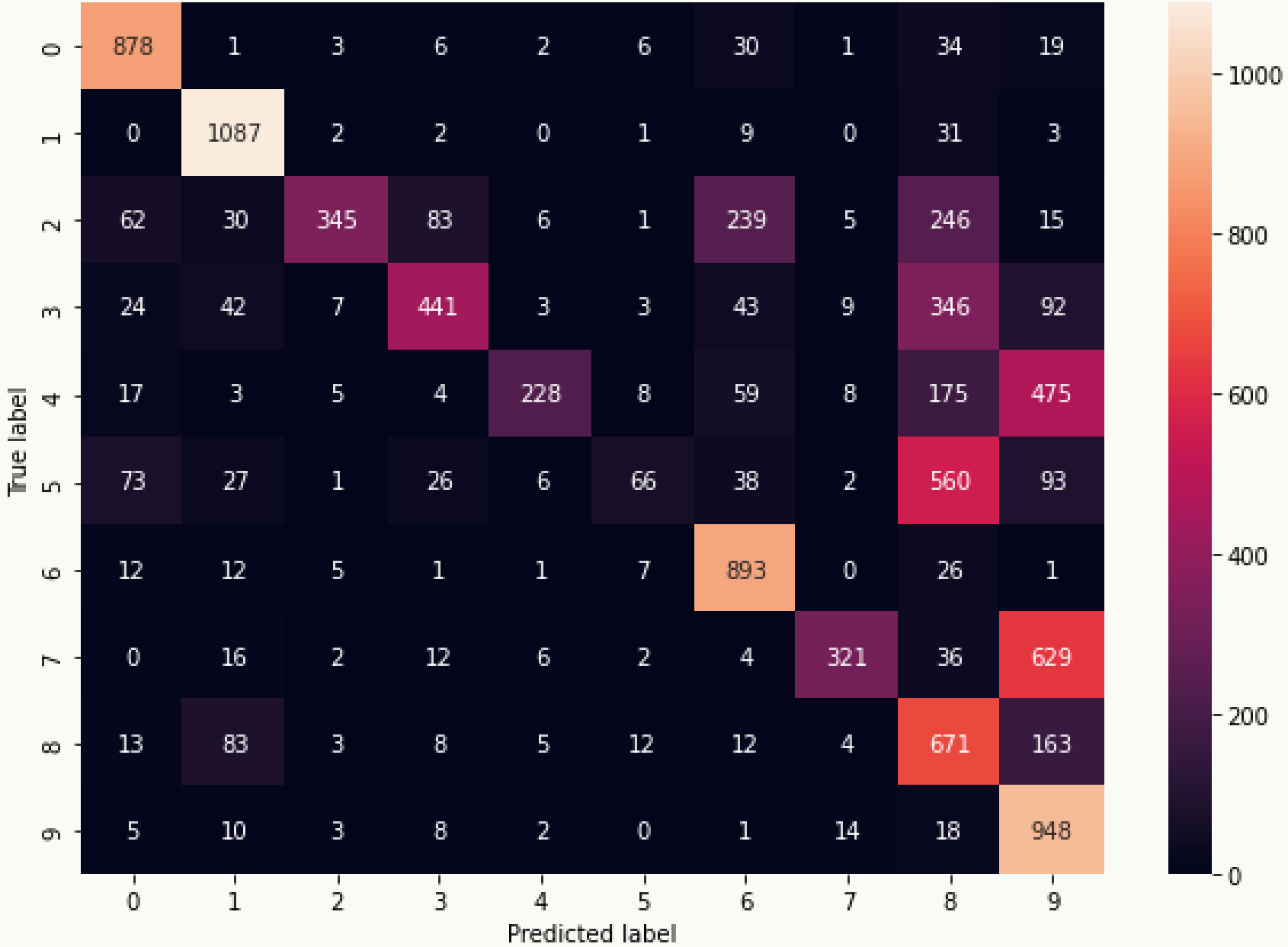


- Naive Bayes
- Gauss Bayes
- KNN

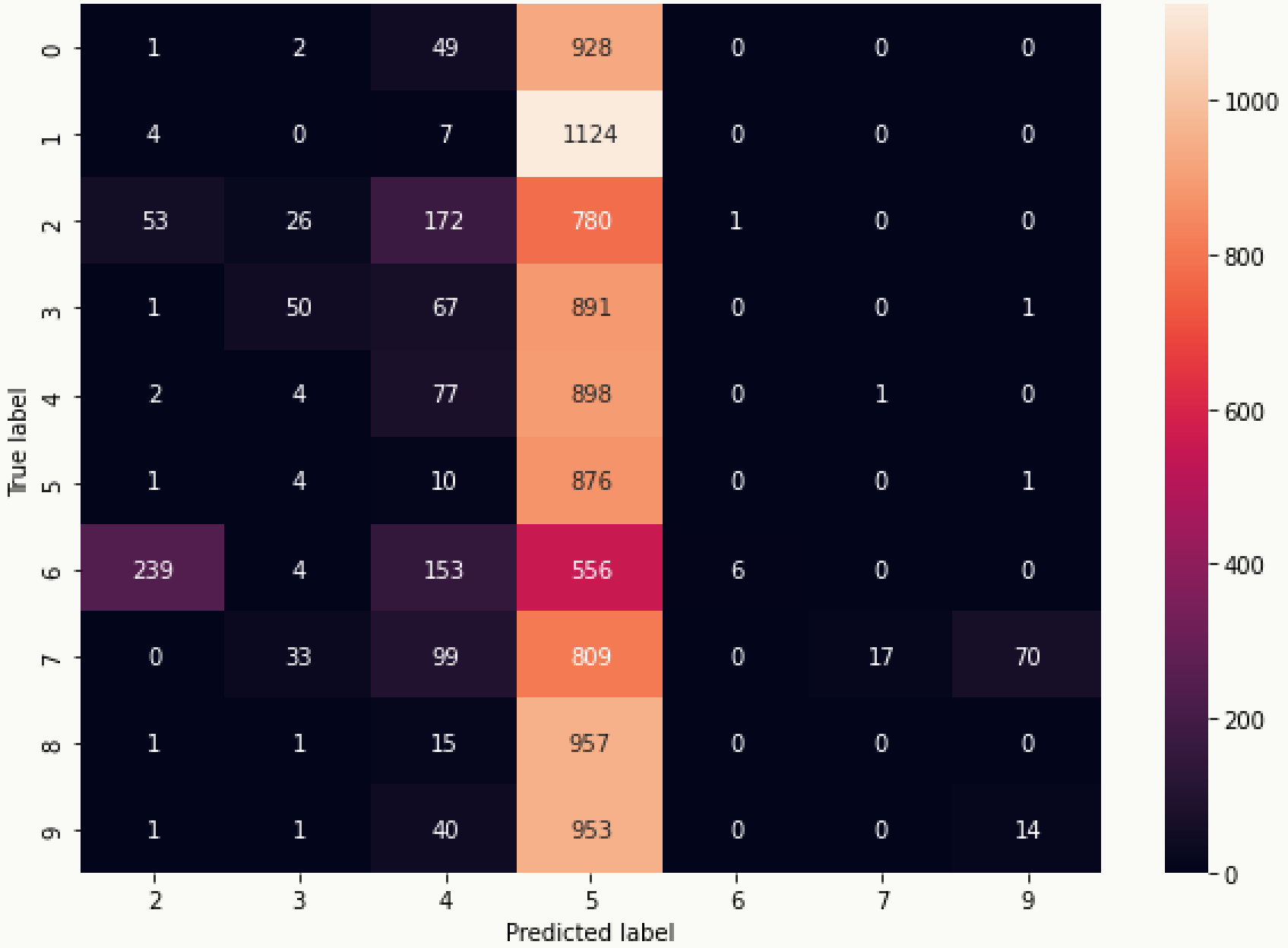
The results

Model	Train	Test
Naive Bayes	0.5937666666666667	0.5878
Gauss-Bayes	0.1165166666666667	0.1093
KNN	1.0	0.9678

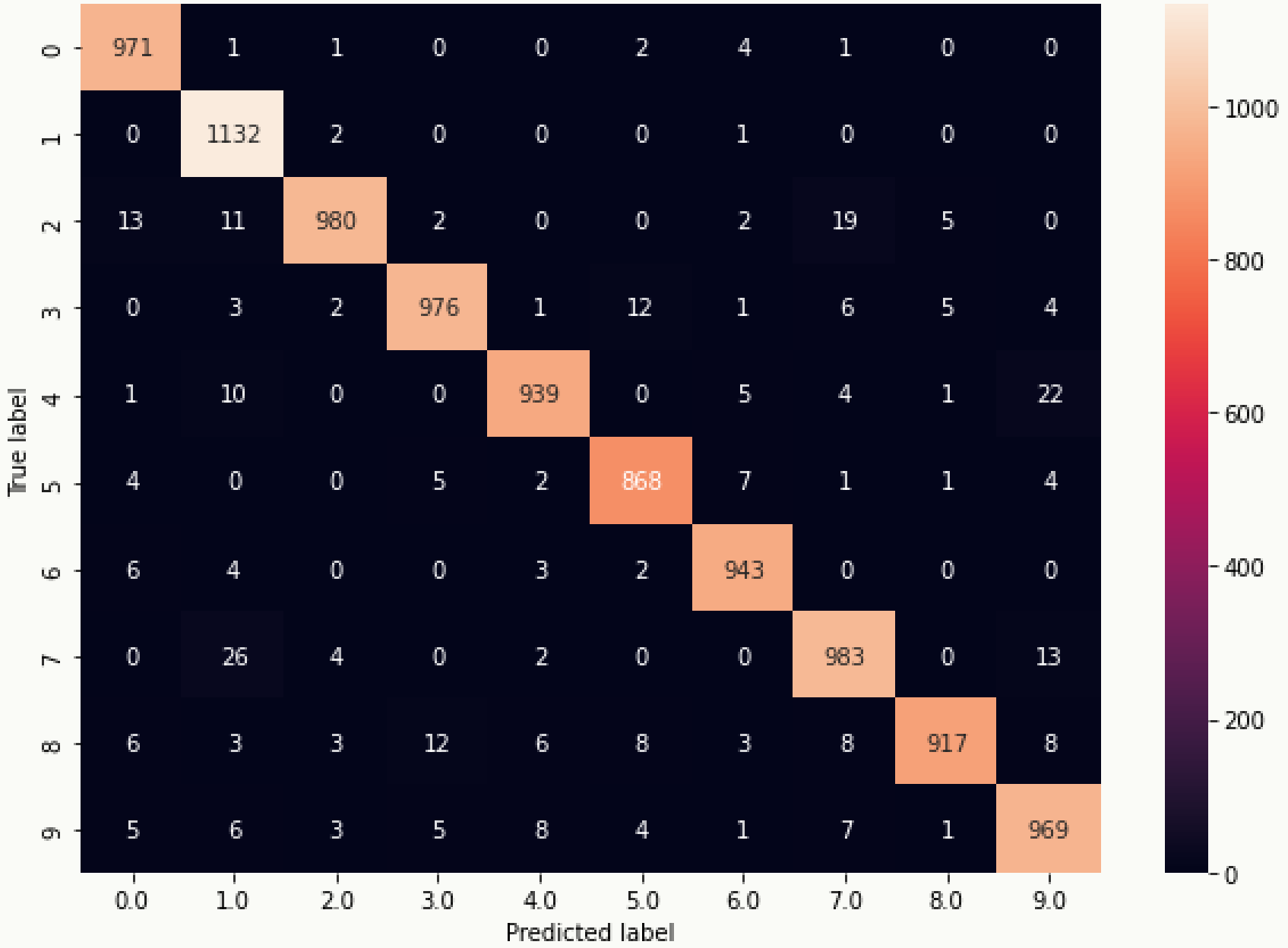
Naive Bayes


True label
Predicted label
0 1 2 3 4 5 6 7 8 9
878 1 3 6 2 6 30 1 34 19
0 1087 2 2 0 1 9 0 31 3
62 30 345 83 6 1 239 5 246 15
24 42 7 441 3 3 43 9 346 92
17 3 5 4 228 8 59 8 175 475
73 27 1 26 6 66 38 2 560 93
12 12 5 1 1 7 893 0 26 1
0 16 2 12 6 2 4 321 36 629
13 83 3 8 5 12 12 4 671 163
5 10 3 8 2 0 1 14 18 948
1000
800
600
400
200
0

Gauss-Bayes

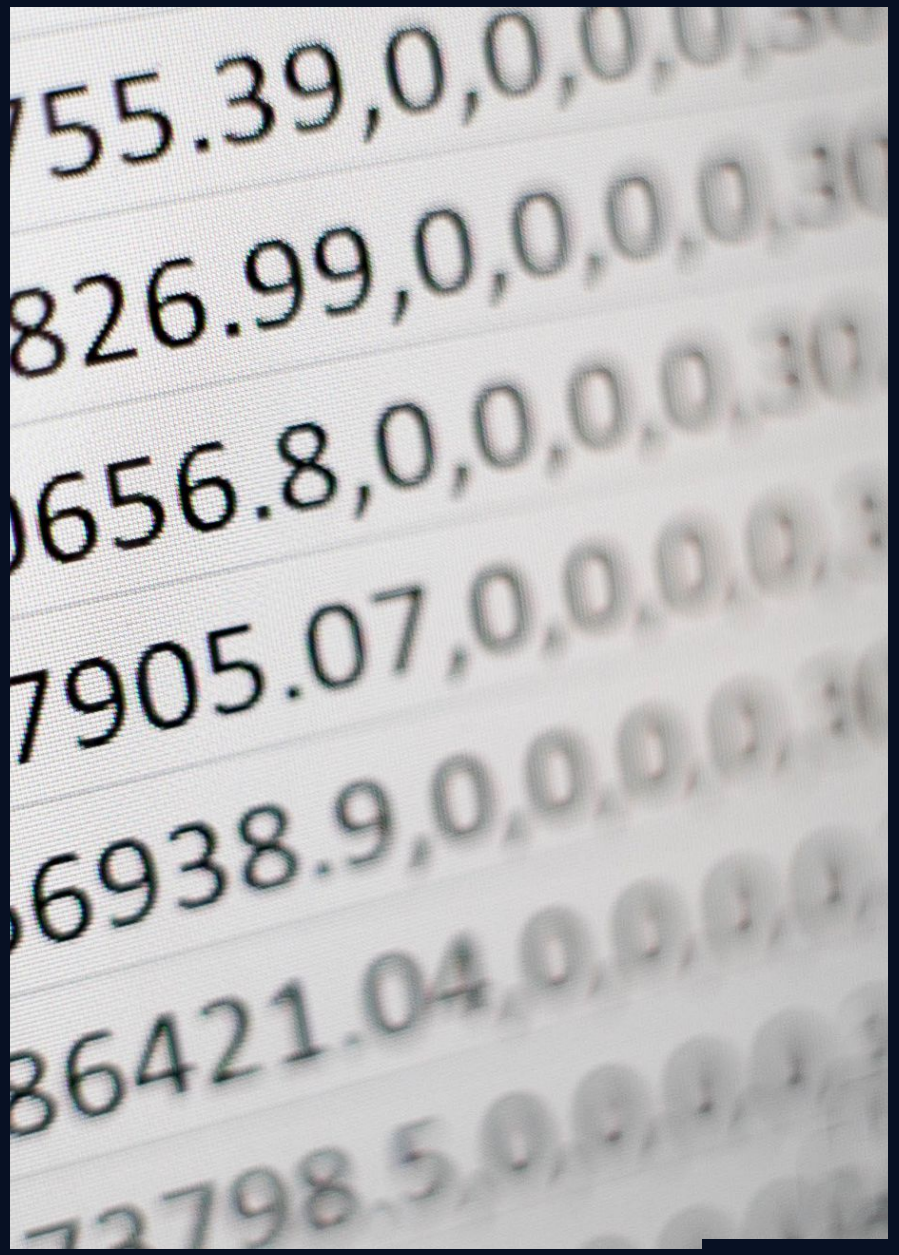

True label
Predicted label
0
1
2
3
4
5
6
7
8
9
2
3
4
5
6
7
9
1 2 49 928 0 0 0
4 0 7 1124 0 0 0
53 26 172 780 1 0 0
1 50 67 891 0 0 1
2 4 77 898 0 1 0
1 4 10 876 0 0 1
239 4 153 556 6 0 0
0 33 99 809 0 17 70
1 1 15 957 0 0 0
1 1 40 953 0 0 14
1000
800
600
400
200
0

KNN


True label
Predicted label
0 971 1 1 0 0 2 4 1 0 0
1 0 1132 2 0 0 0 1 0 0 0
2 13 11 980 2 0 0 2 19 5 0
3 0 3 2 976 1 12 1 6 5 4
4 1 10 0 0 939 0 5 4 1 22
5 4 0 0 5 2 868 7 1 1 4
6 6 4 0 0 3 2 943 0 0 0
7 0 26 4 0 2 0 0 983 0 13
8 6 3 3 12 6 8 3 8 917 8
9 5 6 3 5 8 4 1 7 1 969
0.0 1.0 2.0 3.0 4.0 5.0 6.0 7.0 8.0 9.0
1000
800
600
400
200
0

Lets normalize

Min - Max



$$v' = \frac{v - \text{minF}}{\text{maxF} - \text{minF}} (new\_max_F - new\_min_F) + new\_min_F ,$$

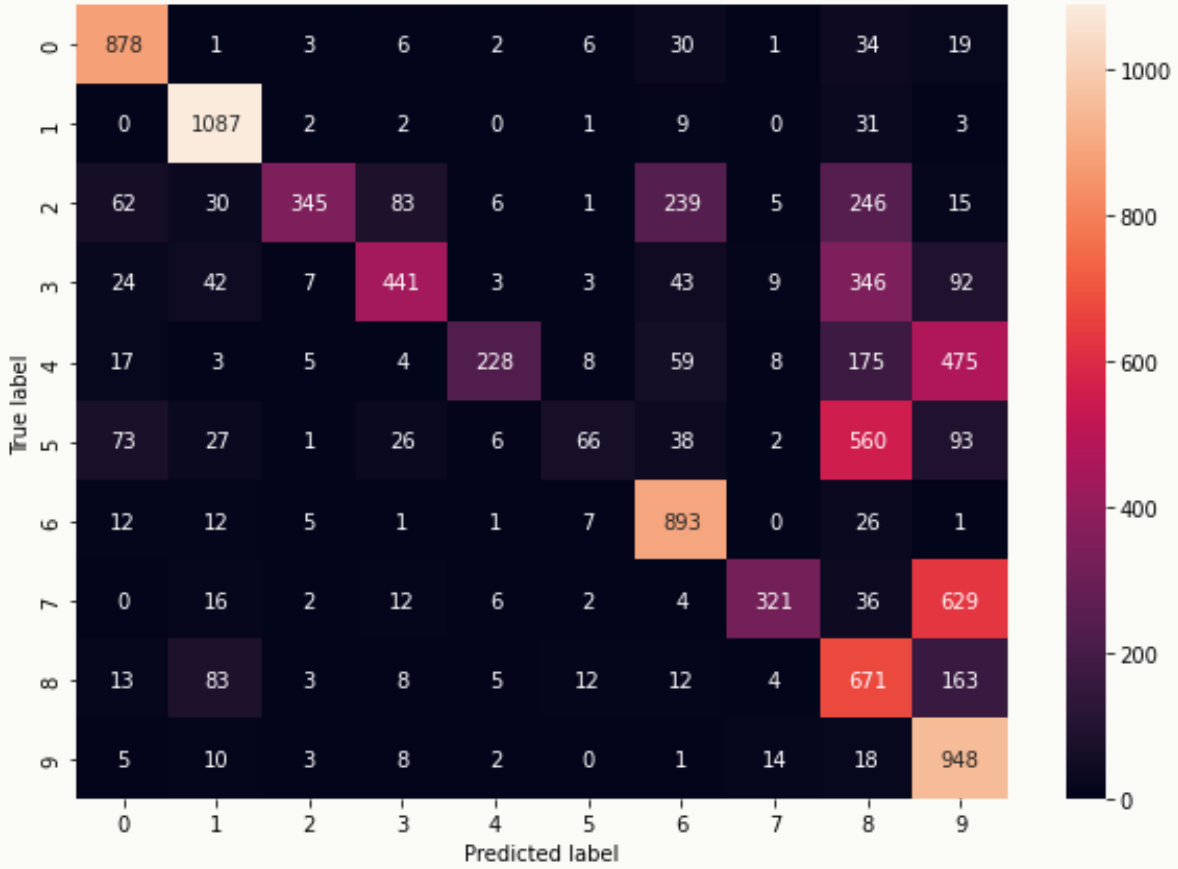
The results

After normalization

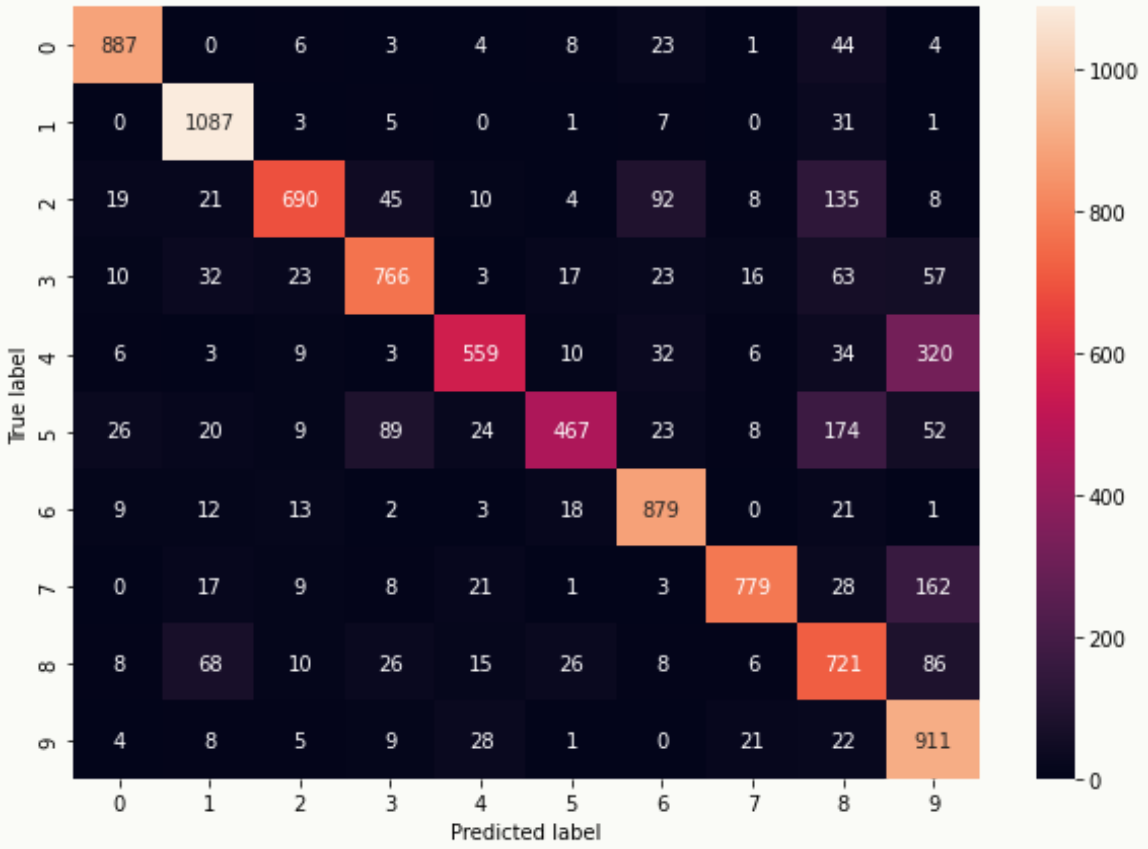
Model	Train	Test
Naive Bayes	0.5937666666666667	0.5878
Gauss-Bayes	0.1165166666666667	0.1093
KNN	1.0	0.9678

Naive Bayes

Not normalized

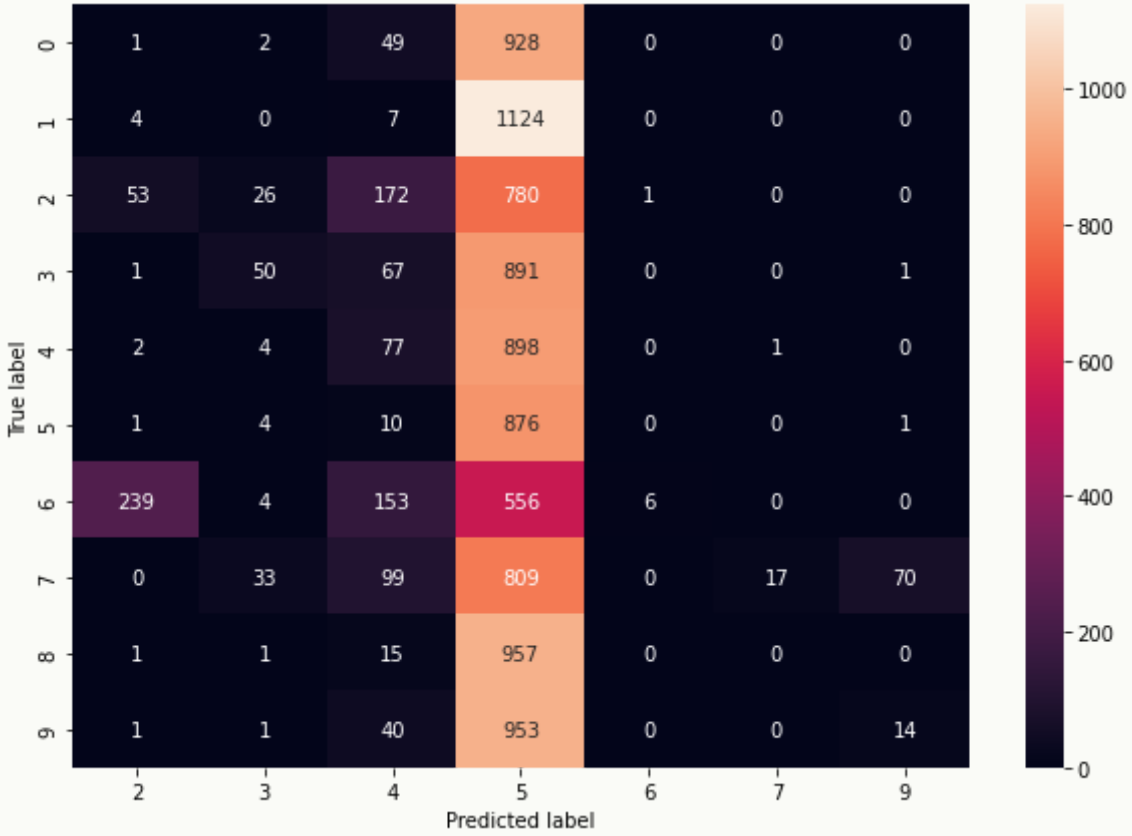


Normalized

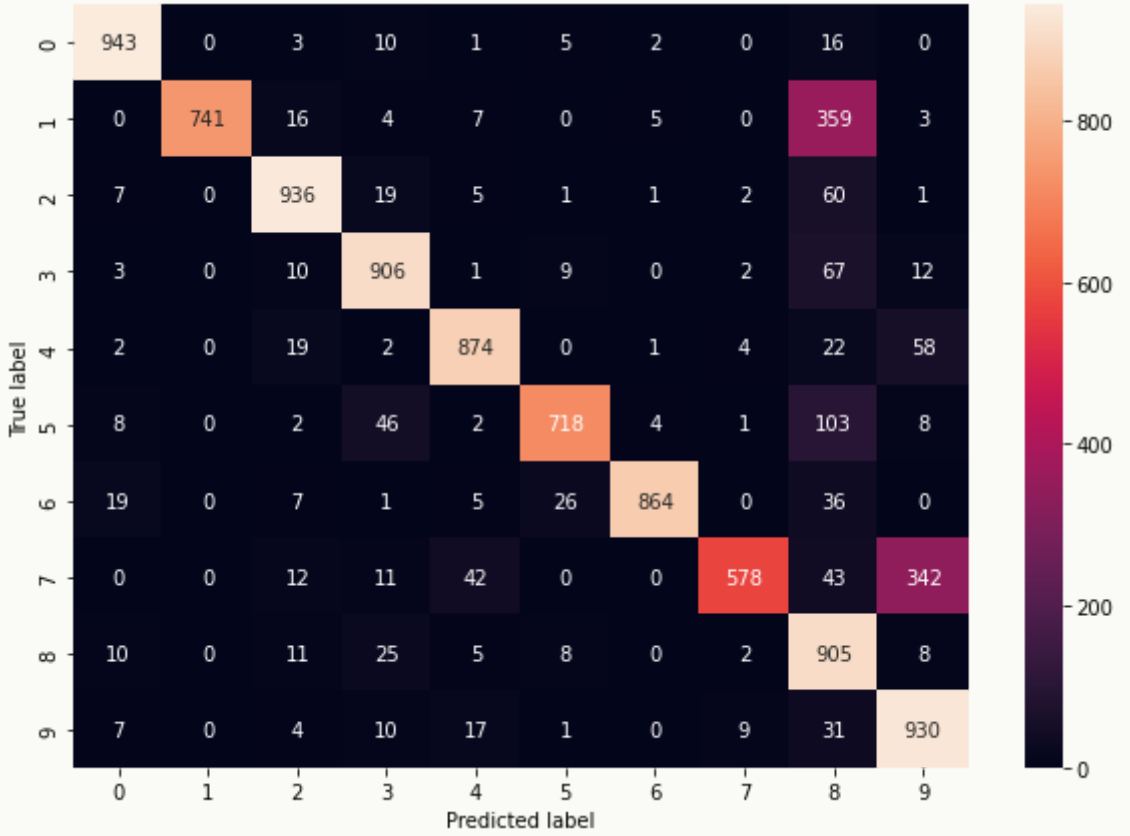


Gauss-Bayes

Normalized

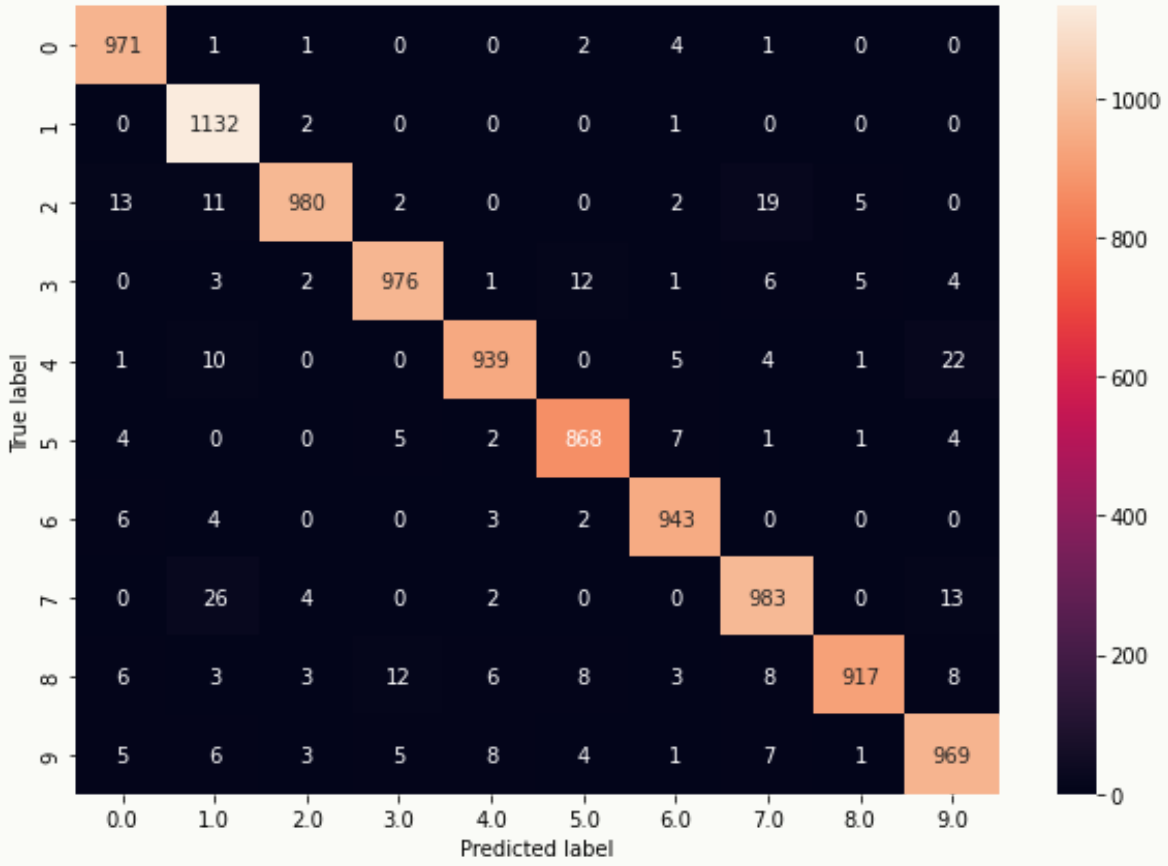


Not normalized

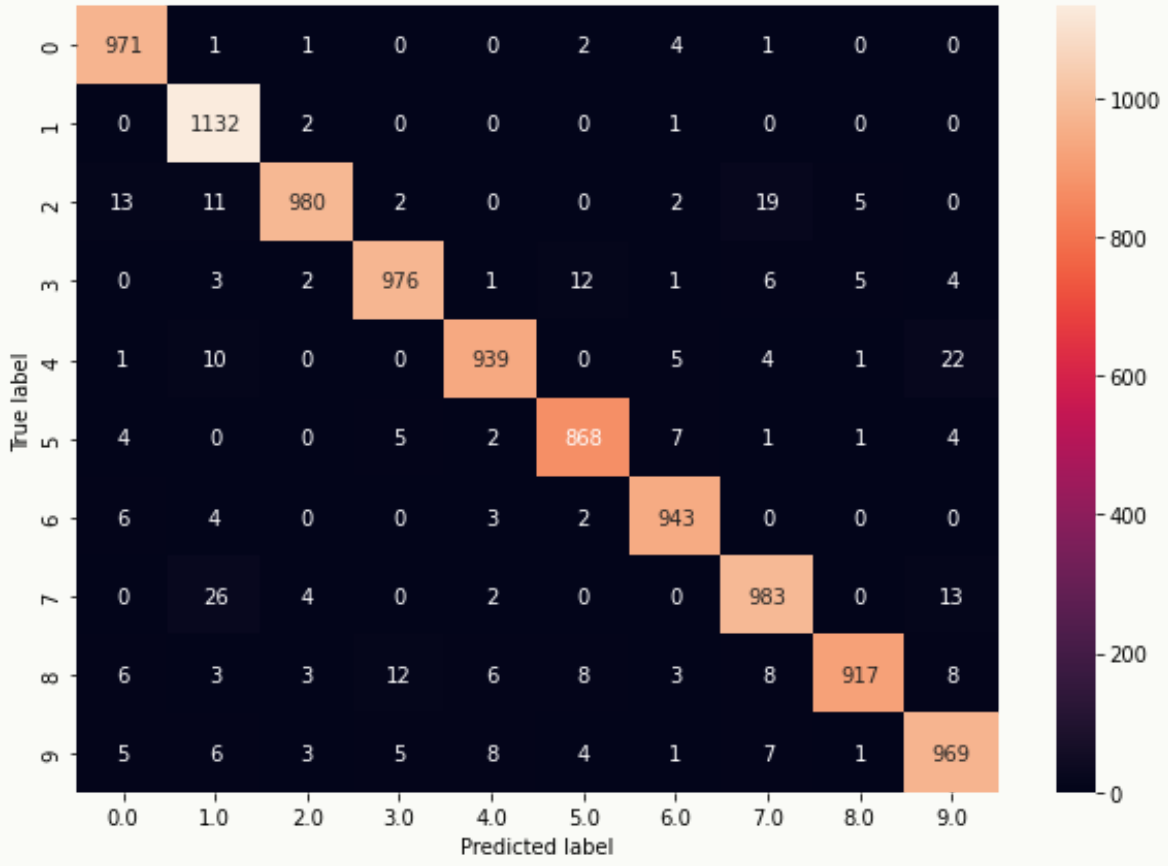


KNN

Normalized



Not normalized



Did we reached at least the same accuracy as human beings?

NOPE

Further research may be required

Conclusions

- We couldn't meet at least 98%.
- The model with the best accuracy was KNN with .9678
- Naive Bayes and Gauss-Bayes got a great enhancement after normalization