



Digits Recognizer

Разпознаване на ръчно написани цифри

Съдържание

01

Увод

02

Преглед на
областта

03

Реализация

04

Резултати

Поколението на дигитализацията





Важни сфери



Здравеопазване
Медицински
записки

Образование
Автоматично
оценяване



Право
Обработка на
информация

Финанси
Банкови
формуляри



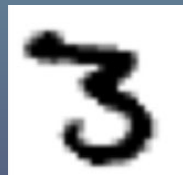
Проблем



→ 1



→ 1



→ 3




→ 7



Преглед на областта

Конволюционни
невронни мрежи
(CNN)
Рекурентните
невронни мрежи
(RNN)



Нашето решение

PCA

K-Nearest Neighbors

Logistic Regression

SVM

98%



ДАННИ

[7, 0, 0, 0, ..., 23, 45, 254, 255, 200, 199, 23, 7, 0, 5, 45, 0, 2, 34 ..., 0]

Label
[0, 9]

Пиксел 1
[0, 255]



28 px

28 px

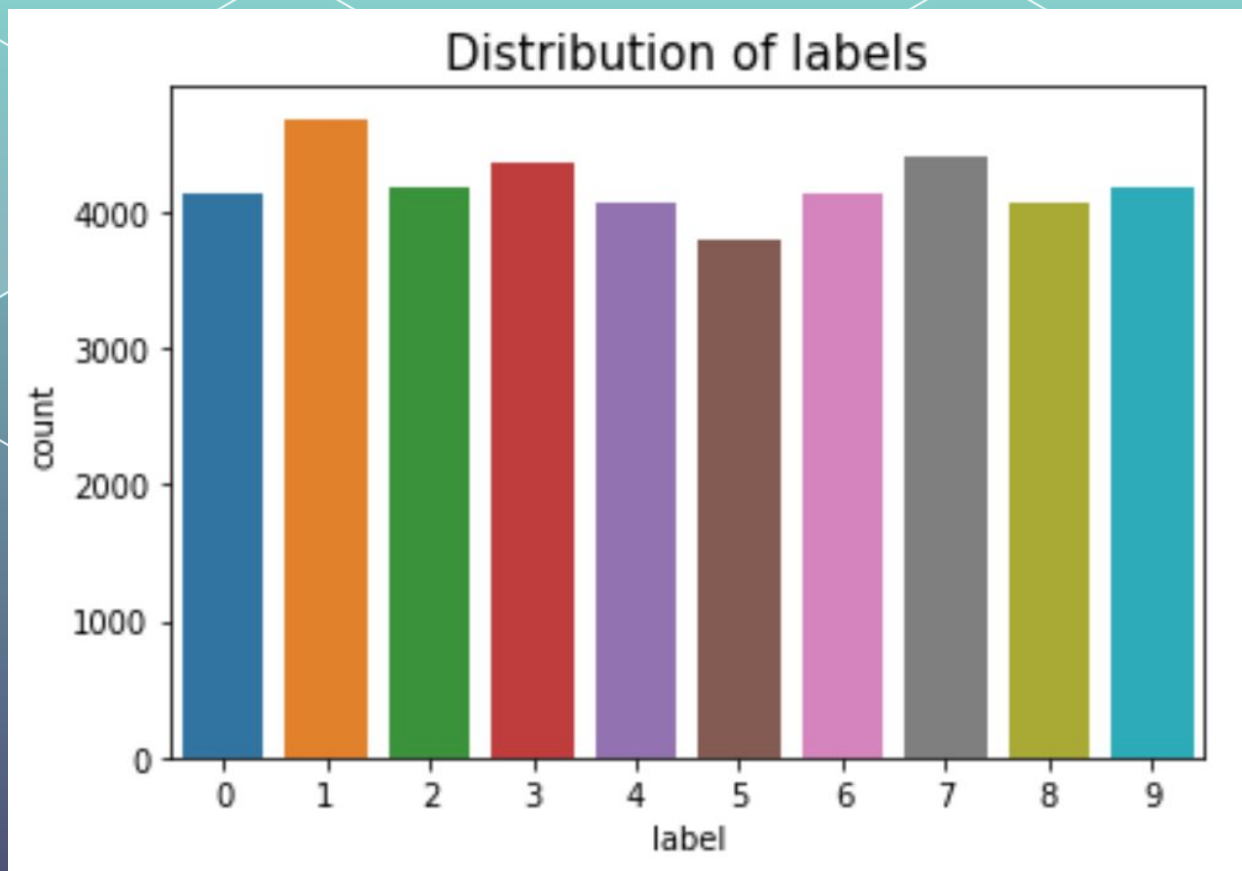
Пиксел 783
[0, 255]

42 000 записа



	label	pixel0	pixel1	pixel2	pixel3	pixel4	pixel5	pixel6	pixel7	pixel8	...	pixel774	pixel775	pixel776	pixel777	pixel778	pixel779	pixel780	pixel781	pixel782	pixel783
41990	3	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
41991	1	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
41992	9	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
41993	6	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
41994	4	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
41995	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
41996	1	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
41997	7	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
41998	6	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
41999	9	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0

	label	pixel0	pixel1	pixel2	pixel3	pixel4	pixel5	pixel6	pixel7	pixel8	...	pixel774	pixel775	pixel776	pixel777	pixel778	pixel779
count	42000.000000	42000.0	42000.0	42000.0	42000.0	42000.0	42000.0	42000.0	42000.0	42000.0	...	42000.000000	42000.000000	42000.000000	42000.000000	42000.000000	42000.000000
mean	4.456643	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.219286	0.117095	0.059024	0.02019	0.017238	0.002857
std	2.887730	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	6.312890	4.633819	3.274488	1.75987	1.894498	0.414264
min	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	2.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
50%	4.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
75%	7.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
max	9.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	254.000000	254.000000	253.000000	253.000000	254.000000	62.000000




Равномерно \Rightarrow accuracy

A decorative pattern of overlapping hexagons in various shades of teal and light blue, located at the top of the slide.

Обработка на данните

MinMaxScaler

$[0, 255] \rightarrow [0, 1]$

A decorative pattern of overlapping hexagons in various shades of purple and dark blue, located at the bottom of the slide.



Намаляване на размерността

PCA

90%

784 -> 87





Разделяне на данните

30% test data -> 29400

70 % train data -> 12 600

```
X_tr, X_ts, y_tr, y_ts = train_test_split(features_scaled_pca, labels, test_size = 0.30)
```



Избиране на алгоритми cross validation (5 folds)

Logistic Regression -> 91%

K-Nearest Neighbor -> 96%

Decision Tree -> 80%

AdaBoost -> 64%

Random Forest -> 93%

Linear SVC -> 90%

Gaussian SVC -> 97%



Logistic Regression

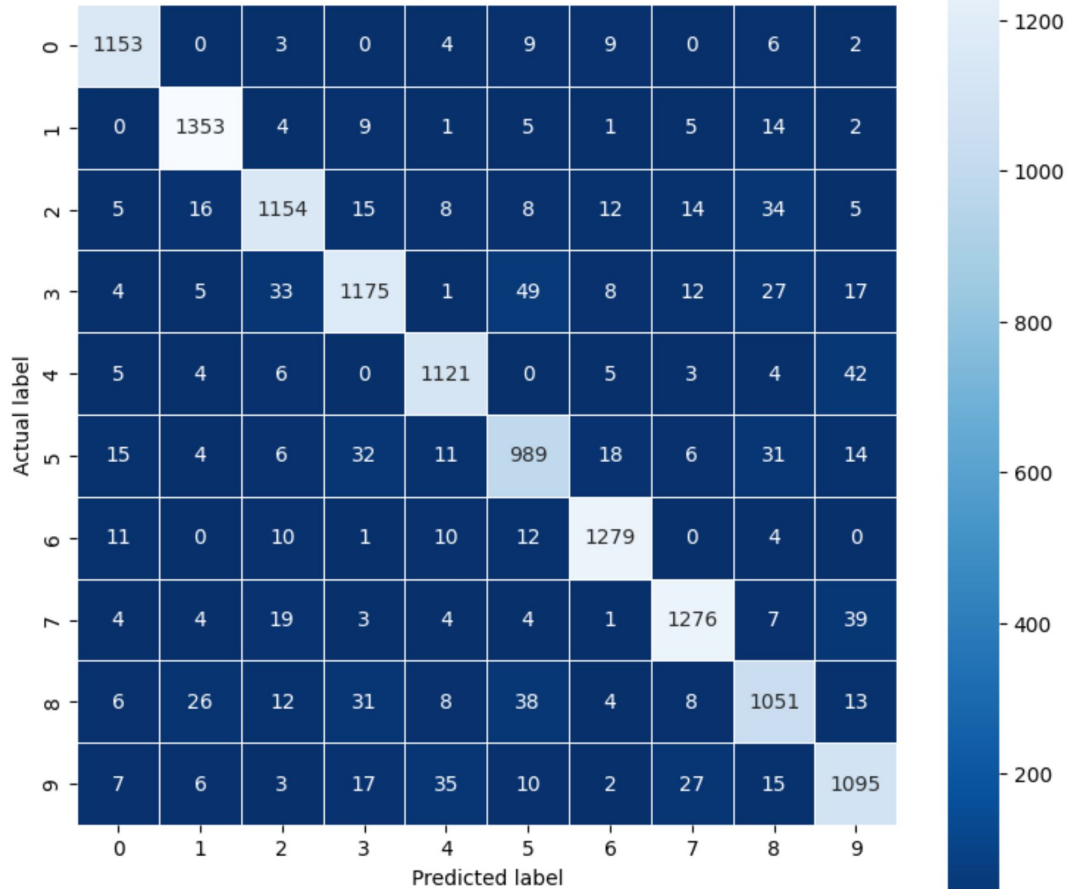
GRID Search

`LogisticRegression(C=100, solver='lbfgs')`

92.429%



Accuracy Score: 92.429%





Gaussian SVM

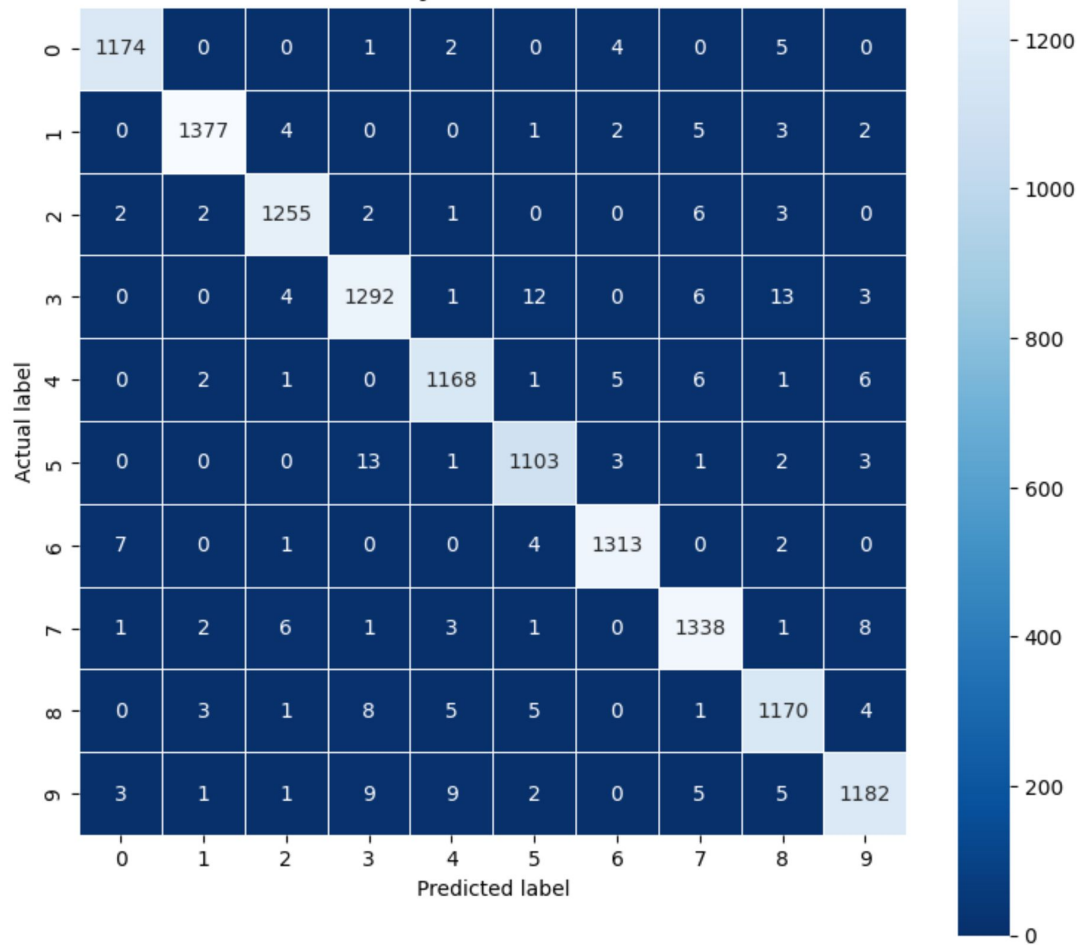
GRID Search

$SVC(C=5, \text{cache_size}=1000, \text{gamma}=0.05)$

98.190%



Accuracy Score: 98.190%





K-Nearest Neighbors

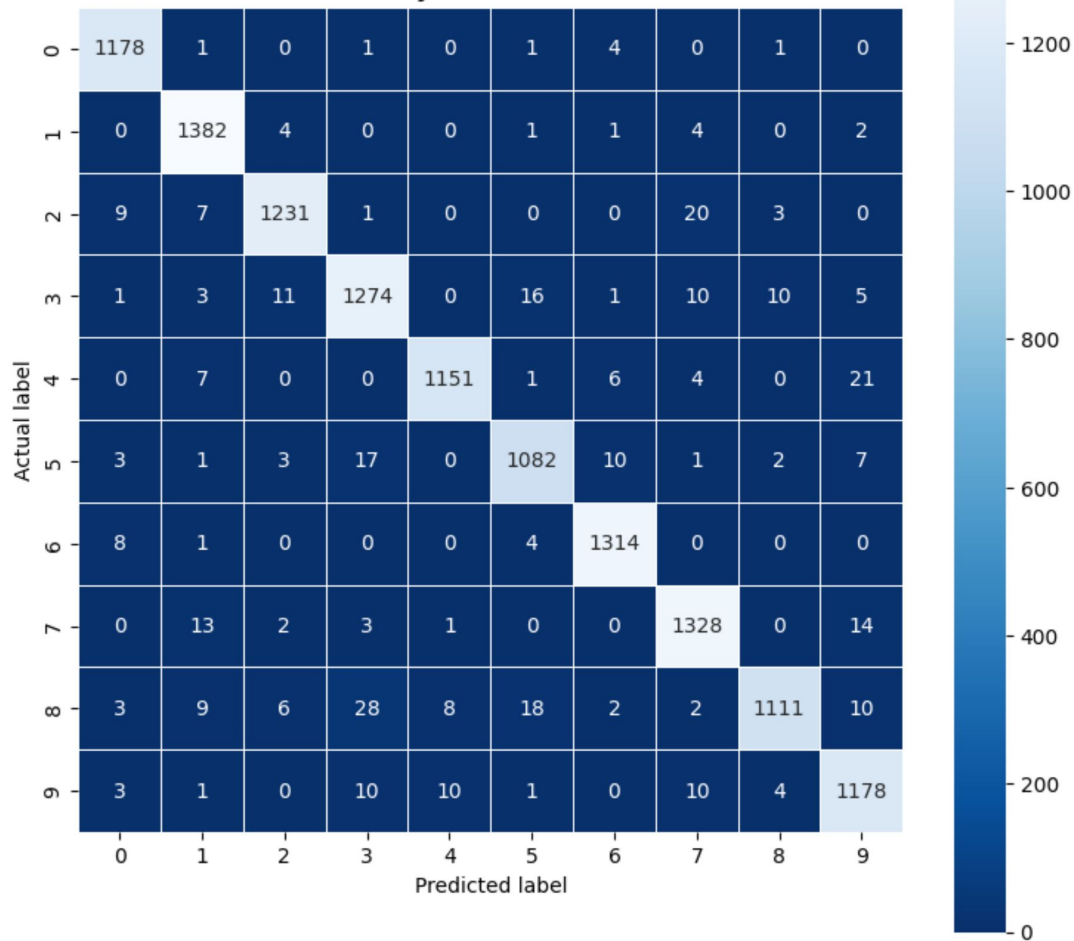
GRID Search

`KNeighborsClassifier(n_neighbors=3)`

97.056%



Accuracy Score: 97.056%





Feature selection вурху KNN

60% -> 17 компонента -> 96.13%

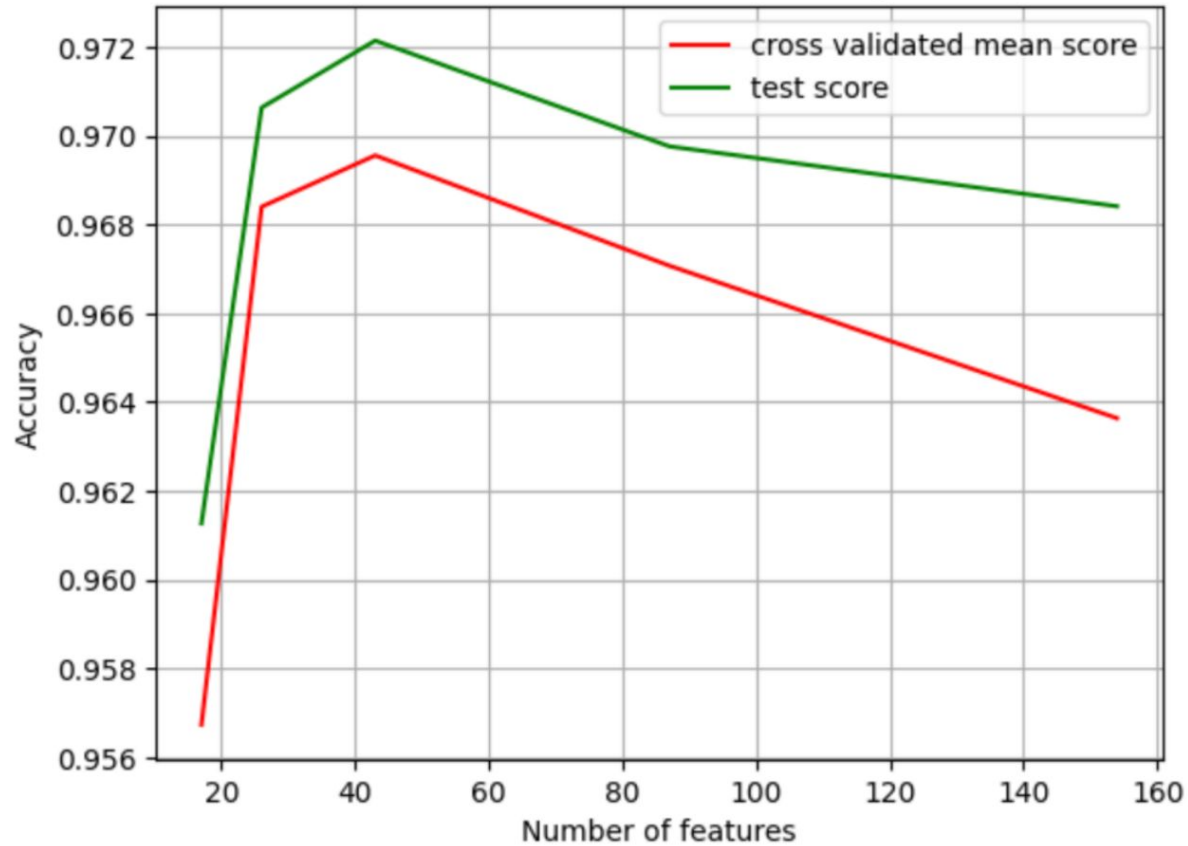
70% -> 26 компонента -> 97.06%

80% -> 43 компонента -> 97.21%

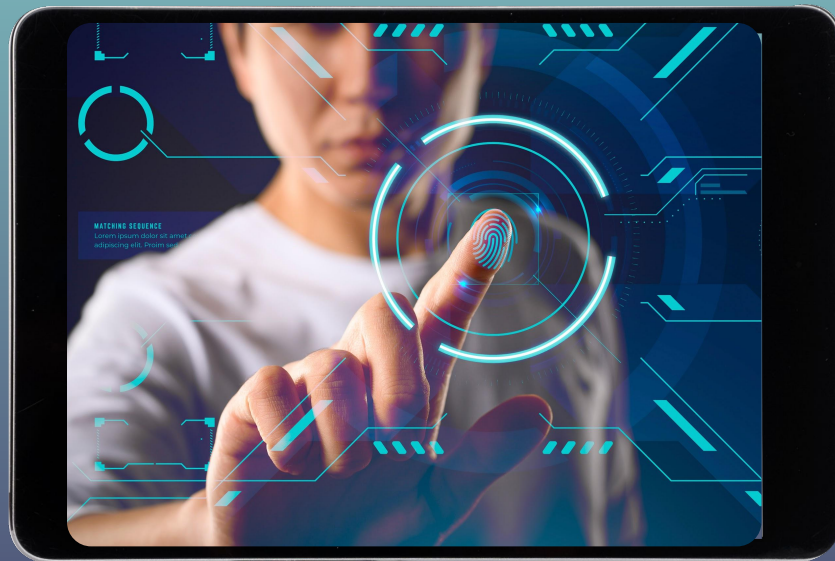
95% -> 154 компонента -> 96.84%



Mean cross validated and test accuracy vs features



Демонстрация



The background features a teal-to-dark-blue gradient. In each of the four corners, there is a decorative pattern of overlapping hexagons and lines, some of which are highlighted with small blue dots.

Благодаря за вниманието!

Изготвил: Атанас Василев, 62577