

Report

Goal: to study and verify the influence of various factors on the accuracy of the program (model)

Experiment №1

Table with results to division 90/10, 80/20, 70/30, 60/40, 50/50 and normalization by features or by samples:

Tab. 1

	90/10	80/20	70/30	60/40	50/50
Without normalization (only /255)	97.5%	96.25%	97.5%	94.375%	93.5%
Normalization by features	97.5%	96.25%	97.5%	94.375%	92.5%
Normalization by samples	95.0%	92.5%	95.0%	92.5%	91.0%

Inference:

- 1) Normalization by features gives the better results than by samples;
- 2) Best result we have when division 90/10 or 70/30 and without normalization or with normalization by features – 97,5%;
- 3) For next experiments we will use model without normalization (only /255) and division 80/20.

Experiment №2

Table with results PCA =0,95 , PCA =0,98 , PCA =0,99:

Tab. 2

	PCA=0,95	PCA=0,98	PCA=0,99
Without normalization (only /255), 80/20	46.25% K = 8	71.25% K = 45	92.5% K = 102

Inference:

- 1) PCA is effective algorithm to data reduction, but requires times to count PCA matrix (about 2 hours with Google Colab).
- 2) With PCA =0,99 accuracy is 92,5% relatively 96,25 without PCA.

Inference:

- 1) Best accuracy we have when use the model without normalization or with normalization by features , division 90/10 or 70/30, without PCA = 97,5%