

### Numerical values of different dataset

Boosting iteration	Initial dataset	Test dataset	Training dataset	Validation dataset	Timestamp value	Augmented dataset
1	F567EF620D 351A56310C	0568A463A3 E34AD4110C	5AF324D4E4 67A24A3583	5876E39CF7 0534567A97	620A 4ED6	DF6A8E8565 6F5D1265A5
2	6A8EA9C78A 8F5A8E0D5E	C35853269F 357F2A167A	575F7E7589 0F7A64E8A8	858F54A58D 698709AF77	620A 4ED7	6F825E65E9 0E8D116A18
3	7F532A3C31 0463D2EA01	1AD357FD36 86352357A6	65E7A85E9A 8853A90560	3578908754 323568D256	620A 4ED8	A652EA6314 568575951E
4	A6389323C2 1A7EE85643	094567F4A1 360734A6256	57460F743E D48F8E985A	4528ED53A6 4E3643A890	620A 4ED9	10AF687A1A 1A7A58681E
5	3ED568524A CA6E975354	E64F345A34 65FA08431A	7E68E96D46 A84D63EC75	5616904345 2A7F4E65E8	620A 4EDA	DA5485645A 615D7A5A9F
6	843785DF43 A4598E1309	1A46FE3A57 386013C54A	5F75843E78 05D5467815	257F37E8E8 F6D36A0976	620A 4EDB	568A56DE83 231AD8568A
7	56AA578945 2F753CE12A	06E5793FF1 545670CD14	568F53EA68 F907456832	975468D6E5 75A953E74A	620A 4EDC	96867AEF56 30147CF568
8	685246323F 131232EA0E	13685E32D6 4327A35505	F7AF7E7A85 E98F8E8F8E	154289F684 E5A7478A86	620A 4EDD	324568A68A 568AD56547
9	E579568033 2AF56EA789	6A794F353D 43CA1576F2	C460F7743E D40F7A6940	2A6E678E85 E75E656880	620A 4EDE	356AE98756 123A560E89
10	3C6585A508 6423567F1A	1E46F542D4 6A6F743A05	32475F790F 7A6964E3F7	35C26839E7 538075D32D	620A 4EDF	356A58C586 F59874568E
11	147E16FE97 65C457D345	D3467F4D35 6D33DCE3E3	F324E7AC88 53A5390F85	8754323568 754A8F5687	620A 4EE0	851D25657E 4D6F02345F
12	623587E46E 54750E2560	467954F325 7A45F64E32	7F7E430ED4 8F8E8E9745	53A64709AF 3A64E36453	620A 4EE1	56897AD459 3256457E8F
13	34EA543678 6509CC7645	3456A3A578 5C42357C45	37A86806A8 4D6533E384	90390568D2 8392A7F443	620A 4EE2	967FF6587A 12348A957E
14	257D7E2586 4A853AC640	A4543D33A5 73A5FC723E	830F75890F 7A6946785D	D254E643A8 739F6D36E8	620A 4EE3	965A78E458 25A658E7D1
15	568589DA30 9754F2AC06	4A46784253 2A5F53D4A0	58095F75F7 E758456745	9042AF4E65 D7975A95D6	620A 4EE4	3658A6589D 2568DF568E
16	864296988A FDEA086436	1A463235A2 34F63F3DE7	F58A58990F 71E9A4853E	37EF636A09 555E5A74F6	620A 4EE5	9568E6578A 2565A456E1
17	16A7F6D36A 583A5D315A	135F44A532 74A6436831	7F7EE9A748 5343ED48F4	75575953E7 5E6E75E68E	620A 4EE6	6897AD575E 23569AED58
18	93A06789DF 463A674A63	36F42D7C7A 474279E4A3	37A8435ED4 8FD46A84D6	542897478A 5A85380739	620A 4EE7	3698D4E58D 5698AD5874
19	A07A657A96 46AE8E0874	C35853269F 357F2A167A	4D0FD466A8 4DE7805D57	57890709AF 7563235687	620A 4EE8	36E84F458E 1798A758EF
20	7F5Φ3E35A7 179AD9A178	1AD357FD36 86352357A6	5809E78505 D55A689078	528ED568D2 8A64E36453	620A 4EE9	598A658E4D 54D68E6897

### Results of generating augmented datasets with one training epoch

Boosting iteration	The ratio of the number of malware code segments to the total code in the test dataset (%)	The value of the coefficient of the boosting algorithm	Augmented datasets, not tested on <i>virustotal</i> (total number of generated datasets)	Augmented datasets after checking on the <i>virustotal</i> resource (number of undetected software datasets with an embedded malicious component)
1.	0,5	$0,5\alpha$	3	-
2.	1	$\alpha$	3	-
3.	1,5	$1,5\alpha$	3	-
4.	2	$2\alpha$	5	-
5.	2,5	$2,5\alpha$	7	1
6.	3	$3\alpha$	8	-
7.	4	$4\alpha$	10	1
8.	4,5	$4,5\alpha$	12	2
9.	5	$5\alpha$	16	2
10.	6	$6\alpha$	19	6
11.	6,5	$6,5\alpha$	24	4
12.	8	$8\alpha$	30	12
13.	8,5	$8,5\alpha$	37	10
14.	9	$9\alpha$	43	11
15.	10	$10\alpha$	56	23
16.	10,5	$10,5\alpha$	60	20
17.	11,5	$11,5\alpha$	75	27
18.	12,5	$12,5\alpha$	83	21
19.	15	$15\alpha$	88	29
20.	17	$17\alpha$	94	34

### Results of generating augmented datasets with second training epoch

Boosting iteration	The ratio of the number of malware code segments to the total code in the test dataset (%)	The value of the coefficient of the boosting algorithm	Augmented datasets, not tested on <i>virustotal</i> (total number of generated datasets)	Augmented datasets after checking on the <i>virustotal</i> resource (number of undetected software datasets with an embedded malicious component)
1.	0,5	$0,5\alpha$	5	3
2.	1	$\alpha$	3	1
3.	1,5	$1,5\alpha$	6	2
4.	2	$2\alpha$	8	5
5.	2,5	$2,5\alpha$	2	-
6.	3	$3\alpha$	4	1
7.	4	$4\alpha$	10	7
8.	4,5	$4,5\alpha$	9	6
9.	5	$5\alpha$	15	9
10.	6	$6\alpha$	16	11
11.	6,5	$6,5\alpha$	6	5
12.	8	$8\alpha$	14	12
13.	8,5	$8,5\alpha$	26	22
14.	9	$9\alpha$	32	28
15.	10	$10\alpha$	16	14
16.	10,5	$10,5\alpha$	35	29
17.	11,5	$11,5\alpha$	15	13
18.	12,5	$12,5\alpha$	23	20
19.	15	$15\alpha$	29	22
20.	17	$17\alpha$	38	29

### Results of generating augmented datasets with third training epoch

Boosting iteration	The ratio of the number of malware code segments to the total code in the test dataset (%)	The value of the coefficient of the boosting algorithm	Augmented datasets, not tested on <i>virustotal</i> (total number of generated datasets)	Augmented datasets after checking on the <i>virustotal</i> resource (number of undetected software datasets with an embedded malicious component)
1.	0,5	$0,5\alpha$	7	5
2.	1	$\alpha$	4	2
3.	1,5	$1,5\alpha$	6	4
4.	2	$2\alpha$	8	6
5.	2,5	$2,5\alpha$	3	1
6.	3	$3\alpha$	6	4
7.	4	$4\alpha$	9	8
8.	4,5	$4,5\alpha$	9	7
9.	5	$5\alpha$	16	11
10.	6	$6\alpha$	15	14
11.	6,5	$6,5\alpha$	8	6
12.	8	$8\alpha$	11	9
13.	8,5	$8,5\alpha$	29	22
14.	9	$9\alpha$	28	25
15.	10	$10\alpha$	35	29
16.	10,5	$10,5\alpha$	37	32
17.	11,5	$11,5\alpha$	27	24
18.	12,5	$12,5\alpha$	22	18
19.	15	$15\alpha$	17	14
20.	17	$17\alpha$	46	31