

DynAA - Statistical Tests

It was generate 100 random samples of the predicted class on the test sample. For each sample, we compute the F1-macro, than with 100 points it was calculated the mean, the T-test and p-value against the DynAA.

By dataset

	DynAA_mean	char_mean	char_ttest	char_pvalue	CNN_mean	CNN_ttest	CNN_pvalue	BERT_mean	BERT_ttest	BERT_pvalue
dataset										
lyrics	0.48	0.47	11.09	4.71e-27	0.21	100.42	0.00e+00	0.20	106.37	0.00e+00
pan18_eval	0.71	0.63	34.77	1.31e-207	0.35	89.16	0.00e+00	0.22	145.91	0.00e+00
socialaa	0.68	0.57	35.28	1.30e-230	0.17	108.51	0.00e+00	0.35	65.90	0.00e+00

By dataset and language

		DynAA_mean	char_mean	char_ttest	char_pvalue	CNN_mean	CNN_ttest	CNN_pvalue	BERT_mean	BERT_ttest	BERT_pvalue
dataset	language										
lyrics	en	0.50	0.49	3.57	3.85e-04	0.19	105.47	0.00e+00	0.23	78.54	4.72e-283
	pt	0.46	0.44	12.90	4.82e-33	0.22	58.16	1.94e-224	0.18	72.58	2.44e-267
pan18_eval	en	0.79	0.75	12.14	4.46e-29	0.36	58.97	4.30e-199	0.30	95.46	6.90e-277
	fr	0.69	0.63	15.71	1.27e-43	0.33	46.86	2.61e-164	0.19	85.84	2.48e-259
	it	0.71	0.67	13.14	4.85e-33	0.36	61.17	8.06e-205	0.32	85.69	4.83e-259
	pl	0.47	0.44	10.82	4.39e-24	0.32	29.05	1.80e-100	0.10	77.67	5.61e-243
	sp	0.87	0.66	50.25	1.19e-174	0.37	68.19	5.41e-222	0.20	125.53	7.91e-323
socialaa	en	0.68	0.57	35.28	1.30e-230	0.17	108.51	0.00e+00	0.35	65.90	0.00e+00

By dataset and nauthors

		DynAA_mean	char_mean	char_ttest	char_pvalue	CNN_mean	CNN_ttest	CNN_pvalue	BERT_mean	BERT_ttest	BERT_pvalue
dataset	nauthors										
lyrics	5	0.59	0.59	0.72	4.71e-01	0.41	23.40	4.74e-59	0.42	39.84	8.72e-97
	10	0.57	0.55	8.37	9.99e-15	0.22	134.10	4.55e-197	0.20	95.46	3.76e-168
	15	0.46	0.41	26.92	3.01e-68	0.17	52.67	8.14e-119	0.16	72.05	1.57e-144
	20	0.42	0.42	4.51	1.09e-05	0.13	145.61	4.11e-204	0.15	134.65	2.04e-197
	25	0.35	0.37	-12.04	1.99e-25	0.10	92.94	6.92e-166	0.09	112.58	3.77e-182
pan18_eval	5	0.71	0.62	16.78	1.95e-50	0.44	31.32	7.02e-120	0.28	63.38	7.98e-241
	10	0.75	0.64	26.25	4.83e-96	0.36	62.08	7.36e-237	0.27	68.65	2.27e-256
	15	0.69	0.65	11.49	2.75e-27	0.36	42.79	4.10e-169	0.17	85.25	1.37e-299
	20	0.68	0.61	18.52	1.11e-58	0.25	60.14	8.80e-231	0.17	86.32	4.08e-302
socialaa	5	0.84	0.63	20.49	2.45e-75	0.27	59.11	5.95e-294	0.47	25.27	5.23e-104
	10	0.71	0.59	22.01	2.83e-84	0.28	48.81	5.94e-242	0.40	29.91	1.82e-132
	20	0.66	0.59	19.10	2.83e-67	0.11	68.87	0.00e+00	0.32	53.74	1.56e-267
	50	0.51	0.46	29.81	7.38e-132	0.02	47.86	7.25e-237	0.20	45.15	4.16e-222

By dataset and ndocs

		DynAA_mean	char_mean	char_ttest	char_pvalue	CNN_mean	CNN_ttest	CNN_pvalue	BERT_mean	BERT_ttest	BERT_pvalue
dataset	ndocs										
lyrics	23	0.36	0.36	0.31	7.56e-01	0.12	104.43	2.48e-237	0.11	132.60	1.10e-267
	24	0.38	0.36	11.60	3.79e-20	0.09	108.50	1.01e-104	0.10	98.30	1.61e-100
	25	0.54	0.51	8.06	1.89e-14	0.29	34.53	2.46e-106	0.27	56.14	7.52e-161
	26	0.57	0.55	8.37	9.99e-15	0.22	134.10	4.55e-197	0.20	95.46	3.76e-168
	29	0.58	0.58	NaN	nan	0.30	49.76	7.33e-72	0.41	27.72	1.89e-48

		DynAA_mean	char_mean	char_ttest	char_pvalue	CNN_mean	CNN_ttest	CNN_pvalue	BERT_mean	BERT_ttest	BERT_pvalue
dataset	ndocs										
pan18_eval	7	0.71	0.63	34.77	1.31e-207	0.35	89.16	0.00e+00	0.22	145.91	0.00e+00
socialaa	10	0.61	0.50	24.28	3.76e-111	0.18	66.18	0.00e+00	0.33	39.50	1.05e-238
	50	0.75	0.63	25.61	1.64e-121	0.16	95.86	0.00e+00	0.37	55.44	0.00e+00

By dataset and nchar

		DynAA_mean	char_mean	char_ttest	char_pvalue	CNN_mean	CNN_ttest	CNN_pvalue	BERT_mean	BERT_ttest	BERT_pvalue
dataset	nchar										
lyrics	1000	0.48	0.47	11.09	4.71e-27	0.21	100.42	0.00e+00	0.20	106.37	0.00e+00
pan18_eval	3162	0.71	0.63	34.77	1.31e-207	0.35	89.16	0.00e+00	0.22	145.91	0.00e+00
socialaa	31	0.43	0.16	16.83	3.65e-45	0.05	29.60	7.02e-91	0.31	7.55	5.37e-13
	100	0.45	0.38	17.04	6.08e-59	0.08	57.95	0.00e+00	0.25	29.87	1.04e-149
	316	0.91	0.79	28.48	1.15e-144	0.26	113.39	0.00e+00	0.44	79.53	0.00e+00

By dataset and nword

		DynAA_mean	char_mean	char_ttest	char_pvalue	CNN_mean	CNN_ttest	CNN_pvalue	BERT_mean	BERT_ttest	BERT_pvalue
dataset	nword										
lyrics	210	0.49	0.48	5.47	1.37e-07	0.34	27.30	3.35e-69	0.28	48.85	9.14e-113
	215	0.57	0.52	26.44	1.16e-46	0.23	100.02	2.95e-101	0.15	103.85	7.37e-103
	220	0.38	0.36	11.60	3.79e-20	0.09	108.50	1.01e-104	0.10	98.30	1.61e-100
	230	0.36	0.36	-0.99	3.24e-01	0.07	129.55	2.70e-112	0.08	129.69	2.43e-112
	275	0.58	0.58	NaN	nan	0.30	49.76	7.33e-72	0.41	27.72	1.89e-48
	285	0.34	0.37	-22.26	2.58e-40	0.12	105.20	2.08e-103	0.11	111.31	8.15e-106
	305	0.57	0.58	-4.48	2.03e-05	0.21	108.50	1.01e-104	0.24	100.28	2.28e-101

		DynAA_mean	char_mean	char_ttest	char_pvalue	CNN_mean	CNN_ttest	CNN_pvalue	BERT_mean	BERT_ttest	BERT_pvalue
dataset	nword										
	310	0.50	0.47	11.03	2.22e-22	0.17	107.28	4.76e-178	0.19	91.66	1.02e-164
pan18_eval	765	0.45	0.36	23.22	7.65e-42	0.25	37.69	1.59e-60	0.09	54.56	1.09e-75
	780	0.51	0.51	NaN	nan	0.30	35.43	4.87e-58	0.07	84.31	5.35e-94
	785	0.62	0.57	11.25	1.02e-24	0.38	20.90	2.00e-60	0.23	42.39	1.83e-128
	790	0.66	0.54	26.26	2.11e-46	0.12	88.64	4.02e-96	0.11	88.69	3.80e-96
	795	0.62	0.61	2.57	1.07e-02	0.41	45.96	1.42e-137	0.16	89.05	2.80e-217
	800	0.59	0.57	5.68	1.36e-07	0.29	32.14	3.48e-54	0.12	47.27	9.60e-70
	810	0.68	0.65	10.26	3.01e-17	0.35	88.34	5.63e-96	0.34	77.38	2.32e-90
	815	0.77	0.67	15.23	3.07e-35	0.25	60.72	2.21e-130	0.17	71.88	2.49e-144
	820	0.77	0.74	6.00	9.35e-09	0.27	75.63	1.42e-148	0.31	68.62	1.79e-140
	825	0.86	0.75	19.54	3.70e-48	0.32	87.64	6.17e-161	0.31	38.00	3.55e-93
	830	0.89	0.69	24.93	1.84e-44	0.53	40.75	1.13e-63	0.18	184.02	2.52e-127
	840	0.92	0.61	51.19	4.86e-73	0.51	31.12	6.31e-53	0.40	55.00	5.05e-76
	850	0.89	0.79	13.38	6.25e-24	0.61	16.35	6.91e-30	0.35	44.10	6.79e-67
socialaa	10	0.29	0.17	11.69	2.41e-20	0.03	27.65	2.35e-48	0.32	-2.46	1.58e-02
	15	0.30	0.18	11.74	9.01e-29	0.05	28.77	5.46e-115	0.19	14.10	3.41e-39
	20	0.37	0.37	NaN	nan	0.01	53.44	8.05e-75	0.14	35.60	3.11e-58
	30	0.50	0.39	16.68	1.38e-51	0.07	55.60	1.17e-238	0.25	31.94	1.77e-131
	35	0.85	0.70	9.32	2.21e-17	0.21	43.17	5.39e-103	0.51	10.20	6.53e-20
	55	1.00	1.00	NaN	nan	0.27	75.01	4.75e-89	0.59	39.74	1.18e-62
	60	0.84	0.75	17.78	4.36e-57	0.19	68.33	5.71e-285	0.33	94.78	0.00e+00
	65	1.00	0.48	24.41	1.11e-43	0.28	39.58	1.69e-62	0.00	inf	0.00e+00
	105	1.00	0.89	12.76	1.69e-31	0.38	56.15	1.59e-191	0.61	25.39	2.59e-85
	110	0.90	0.78	27.64	9.73e-95	0.25	50.14	2.46e-174	0.51	58.24	3.57e-197

By dataset and problem

		DynAA_mean	char_mean	char_ttest	char_pvalue	CNN_mean	CNN_ttest	CNN_pvalue	BERT_mean	BERT_ttest	BERT_pvalue
dataset	problem										
lyrics	problem00001	0.61	0.60	0.72	4.72e-01	0.52	18.09	3.80e-33	0.43	28.87	5.12e-50
	problem00002	0.57	0.52	26.44	1.16e-46	0.23	100.02	2.95e-101	0.15	103.85	7.37e-103
	problem00003	0.38	0.35	21.75	1.72e-39	0.16	68.80	2.13e-85	0.13	71.32	6.45e-87
	problem00004	0.38	0.36	11.60	3.79e-20	0.09	108.50	1.01e-104	0.10	98.30	1.61e-100
	problem00005	0.36	0.36	-0.99	3.24e-01	0.07	129.55	2.70e-112	0.08	129.69	2.43e-112
	problem00006	0.58	0.58	NaN	nan	0.30	49.76	7.33e-72	0.41	27.72	1.89e-48
	problem00007	0.57	0.58	-4.48	2.03e-05	0.21	108.50	1.01e-104	0.24	100.28	2.28e-101
	problem00008	0.54	0.47	37.74	1.42e-60	0.17	142.24	2.71e-116	0.19	119.70	6.46e-109
	problem00009	0.47	0.48	-18.26	1.85e-33	0.17	104.36	4.58e-103	0.19	92.33	7.43e-98
	problem00010	0.34	0.37	-22.26	2.58e-40	0.12	105.20	2.08e-103	0.11	111.31	8.15e-106
pan18_eval	problem00001	0.70	0.70	NaN	nan	0.28	66.17	9.24e-84	0.20	79.50	1.66e-91
	problem00002	0.72	0.70	3.88	1.92e-04	0.31	62.69	1.71e-81	0.23	75.10	4.26e-89
	problem00003	0.83	0.78	6.61	1.98e-09	0.27	90.25	6.90e-97	0.42	42.36	2.97e-65
	problem00004	0.89	0.79	13.38	6.25e-24	0.61	16.35	6.91e-30	0.35	44.10	6.79e-67
	problem00005	0.66	0.54	26.26	2.11e-46	0.12	88.64	4.02e-96	0.11	88.69	3.80e-96
	problem00006	0.66	0.66	2.00	4.86e-02	0.45	29.69	4.35e-51	0.11	86.64	3.74e-95
	problem00007	0.72	0.60	13.95	4.09e-25	0.36	71.42	5.60e-87	0.16	76.10	1.17e-89
	problem00008	0.73	0.73	NaN	nan	0.37	20.67	1.08e-37	0.36	27.39	5.40e-48
	problem00009	0.68	0.65	10.26	3.01e-17	0.35	88.34	5.63e-96	0.34	77.38	2.32e-90
	problem00010	0.71	0.68	4.74	7.15e-06	0.44	36.50	3.14e-59	0.31	61.90	5.88e-81
	problem00011	0.86	0.79	9.15	7.75e-15	0.37	59.53	2.51e-79	0.50	39.65	1.46e-62
	problem00012	0.59	0.57	5.68	1.36e-07	0.29	32.14	3.48e-54	0.12	47.27	9.60e-70
	problem00013	0.51	0.51	NaN	nan	0.30	35.43	4.87e-58	0.07	84.31	5.35e-94

		DynAA_mean	char_mean	char_ttest	char_pvalue	CNN_mean	CNN_ttest	CNN_pvalue	BERT_mean	BERT_ttest	BERT_pvalue
dataset	problem										
	problem00014	0.49	0.50	-7.84	5.27e-12	0.34	31.33	3.49e-53	0.08	99.21	6.54e-101
	problem00015	0.45	0.36	23.22	7.65e-42	0.25	37.69	1.59e-60	0.09	54.56	1.09e-75
	problem00016	0.42	0.38	5.55	2.40e-07	0.41	0.95	3.46e-01	0.17	40.16	4.37e-63
	problem00017	0.83	0.63	48.20	1.51e-70	0.19	199.20	1.01e-130	0.12	255.29	2.28e-141
	problem00018	0.86	0.72	24.43	1.02e-43	0.27	113.70	1.01e-106	0.12	165.55	8.62e-123
	problem00019	0.89	0.69	24.93	1.84e-44	0.53	40.75	1.13e-63	0.18	184.02	2.52e-127
	problem00020	0.92	0.61	51.19	4.86e-73	0.51	31.12	6.31e-53	0.40	55.00	5.05e-76
socialaa	problem00001	0.29	0.29	-0.08	9.37e-01	0.12	8.20	8.79e-13	0.12	15.36	5.97e-28
	problem00002	0.71	0.12	26.27	2.01e-46	0.11	25.54	2.31e-45	0.48	5.63	1.69e-07
	problem00003	0.31	0.31	NaN	nan	0.03	31.63	1.49e-53	0.22	7.05	2.44e-10
	problem00004	0.29	0.17	11.69	2.41e-20	0.03	27.65	2.35e-48	0.32	-2.46	1.58e-02
	problem00005	0.28	0.20	19.92	2.11e-36	0.02	38.20	4.66e-61	0.12	23.10	1.17e-41
	problem00006	0.37	0.37	NaN	nan	0.01	53.44	8.05e-75	0.14	35.60	3.11e-58
	problem00007	0.09	0.11	-16.32	8.07e-30	0.00	37.04	8.06e-60	0.11	-7.63	1.45e-11
	problem00008	0.13	0.07	19.12	5.17e-35	0.00	46.92	1.93e-69	0.06	22.49	1.08e-40
	problem00009	0.70	0.70	NaN	nan	0.11	28.05	6.54e-49	0.73	-1.06	2.90e-01
	problem00010	1.00	0.48	24.41	1.11e-43	0.28	39.58	1.69e-62	0.00	inf	0.00e+00
	problem00011	1.00	0.71	12.39	7.54e-22	0.45	21.98	7.15e-40	0.69	13.79	8.79e-25
	problem00012	1.00	0.71	12.39	7.51e-22	0.30	35.44	4.75e-58	0.29	35.05	1.31e-57
	problem00013	1.00	1.00	NaN	nan	0.27	42.28	3.60e-65	0.46	25.46	3.04e-45
	problem00014	1.00	1.00	NaN	nan	0.50	20.56	1.66e-37	1.00	NaN	nan
	problem00015	0.39	0.21	20.84	5.61e-38	0.09	34.44	6.53e-57	0.38	0.89	3.77e-01
	problem00016	0.85	0.85	NaN	nan	0.64	12.30	1.18e-21	0.38	41.70	1.31e-64
	problem00017	0.84	0.75	6.74	1.07e-09	0.58	21.98	7.27e-40	0.61	15.05	2.42e-27
	problem00018	1.00	0.60	33.34	1.25e-55	0.20	99.91	3.28e-101	0.48	51.54	2.53e-73

		DynAA_mean	char_mean	char_ttest	char_pvalue	CNN_mean	CNN_ttest	CNN_pvalue	BERT_mean	BERT_ttest	BERT_pvalue
dataset	problem										
	problem00019	1.00	1.00	NaN	nan	0.27	75.01	4.75e-89	0.59	39.74	1.18e-62
	problem00020	1.00	0.84	16.69	1.57e-30	0.39	49.25	1.94e-71	0.23	88.23	6.33e-96
	problem00021	0.41	0.41	NaN	nan	0.05	40.99	6.50e-64	0.24	20.99	3.16e-38
	problem00022	0.85	0.55	36.02	1.06e-58	0.05	126.68	2.45e-111	0.36	53.15	1.35e-74
	problem00023	1.00	0.85	22.69	5.16e-41	0.34	85.91	8.55e-95	0.62	61.60	9.31e-81
	problem00024	0.47	0.48	-2.18	3.17e-02	0.06	48.70	5.65e-71	0.25	28.60	1.20e-49
	problem00025	0.93	0.85	16.09	2.18e-29	0.13	134.21	8.32e-114	0.31	57.72	4.94e-78
	problem00026	1.00	1.00	NaN	nan	0.19	142.92	1.70e-116	0.52	55.94	9.91e-77
	problem00027	0.28	0.27	3.37	1.09e-03	0.01	63.40	5.76e-82	0.07	45.53	3.36e-68
	problem00028	0.69	0.66	7.31	7.03e-11	0.04	139.73	1.57e-115	0.22	79.74	1.24e-91
	problem00029	0.77	0.64	38.52	2.12e-61	0.05	148.27	4.54e-118	0.33	95.83	1.95e-99
	problem00030	0.47	0.40	29.93	2.10e-51	0.02	93.02	3.60e-98	0.10	69.97	4.10e-86
	problem00031	0.71	0.62	28.71	8.52e-50	0.03	156.80	1.83e-120	0.26	87.51	1.41e-95
	problem00032	0.97	0.88	27.18	1.06e-47	0.02	433.32	4.28e-164	0.46	100.15	2.59e-101