

Manifestoberta Performance Report

Sentence Version 2024a (56Topics)

2024-07-31

Summary

- Performance was measured on 203 manifestos, which represent 200920 annotated quasi-sentences
- Overall the model manages to assign the correct category to quasi-sentences in the test data set with an accuracy of 57.12%. In 73.48% of the cases, the true category is among the two most confident predictions of the model, and in 81.07% among the top three. (Table 1)
- Lower macro averaged F1, Precision, and Recall reflect problems with some individual categories, especially rare/exotic categories like 102, 409, or 702 (Table 1 and Table 2)
- The overall distribution and frequency of individual category predictions isn't as closely aligned with the true distribution of categories as they are in our context model. The model is over and under predicting some codes (Table 2)
- The model performs in a acceptable range across all countries/languages present in the test data set, with the lowest accuracy value of 46.07% in Mexico (Table 3).
- Probability estimates of the model are well calibrated and properly reflect the likelihood of a right prediction. If the model reports a confidence of 95% or higher (which happened for 6.76% of all quasi-sentences in our test set) it was, in fact, right in 95.54% of those cases. (Table 4)
- True rile values and rile values calculated based on model predictions are rather strongly correlated (Plot 1).

Usage Recommendations

As the model performs worse than the context model version yet is easier to apply, it is primarily intended for small ad-hoc analyses and prototyping. Generally, we recommend using the more capable context model for more robust and reliable results.

Results

accuracy	0.57
top2_acc	0.73
top3_acc	0.81
precision	0.48
recall	0.43
f1macro	0.45
mcc	0.55
cross-entropy	1.47

Table 1: Classification Results - Overall

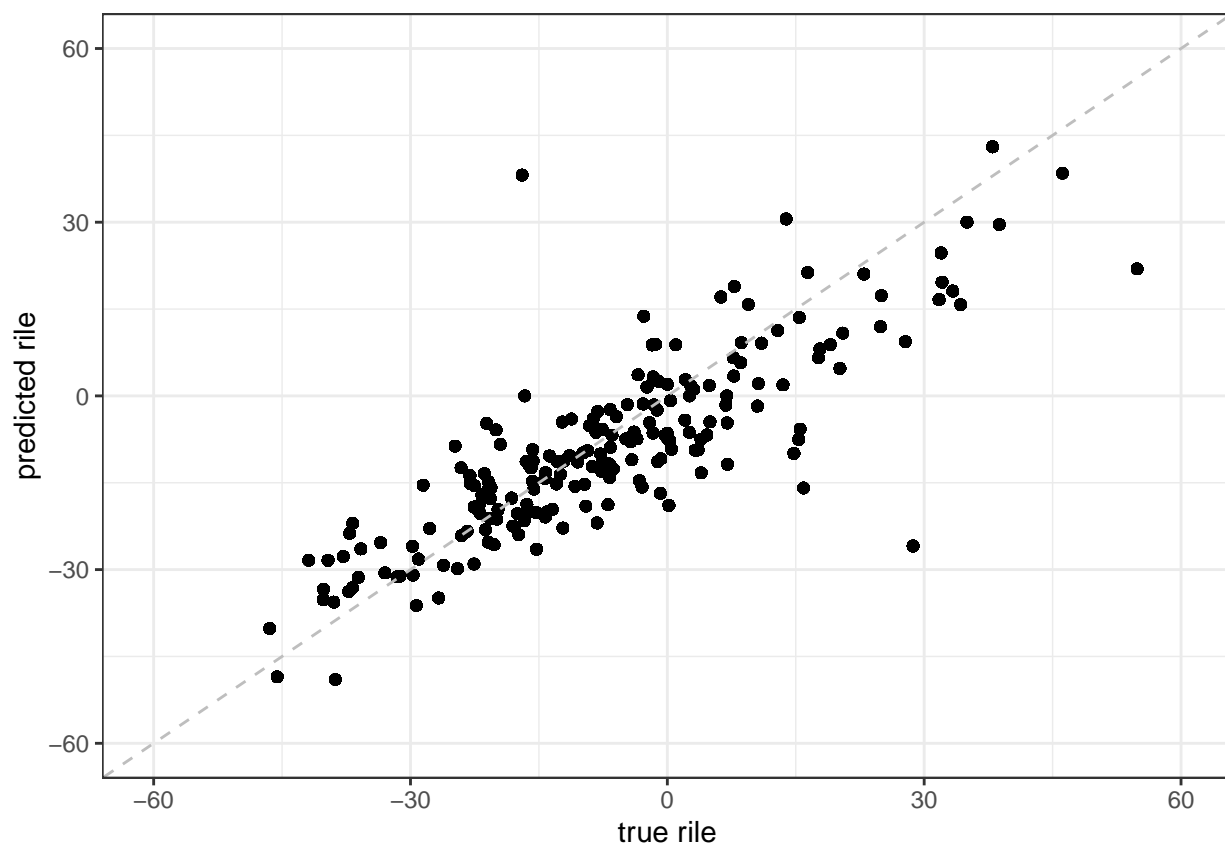


Figure 1: True rle values vs. predicted rle values

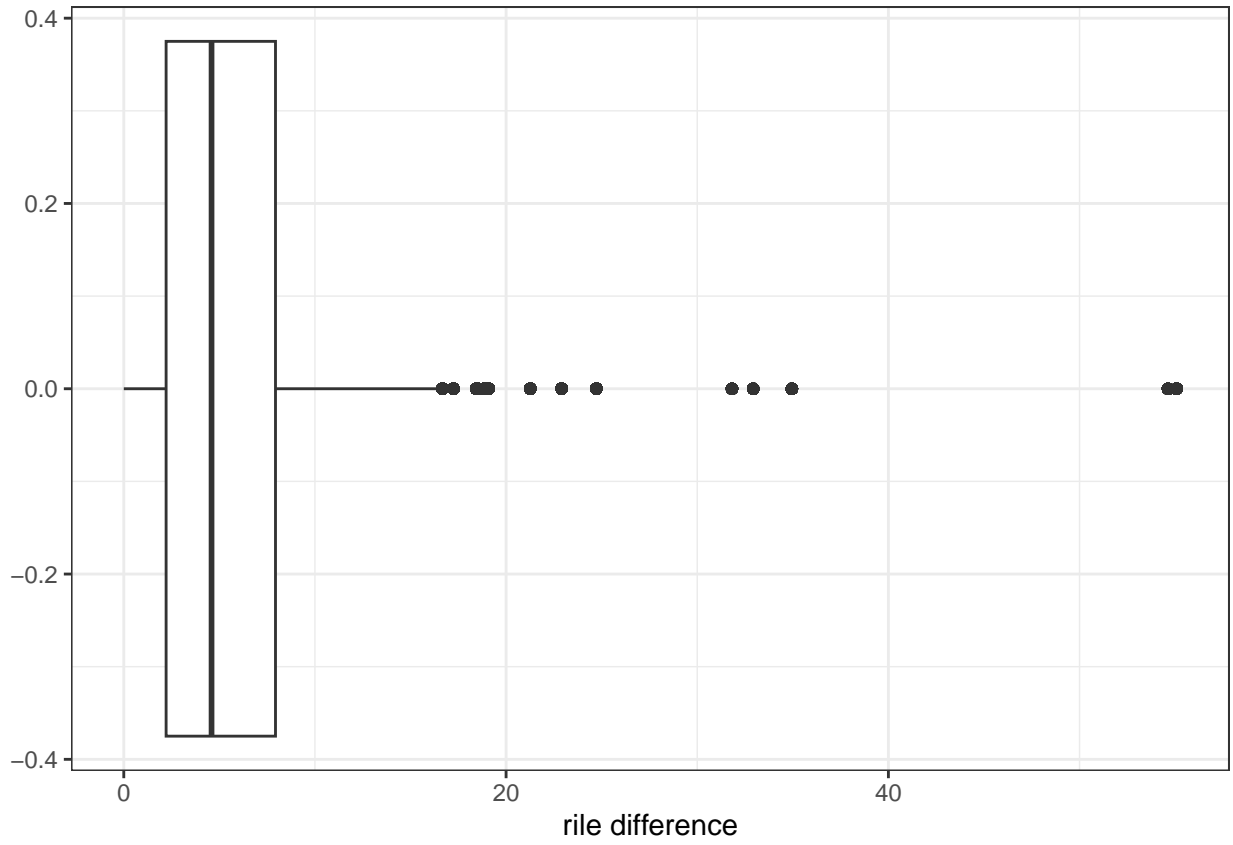


Figure 2: Absolute rle differences between true and predicted cmp codes

Category	Precision	Recall	F1	n(%)	n_predicted(%)
101	0.38	0.37	0.38	0.29%	0.28%
102	0.47	0.38	0.42	0.07%	0.06%
103	0.37	0.34	0.35	0.26%	0.24%
104	0.66	0.68	0.67	1.55%	1.60%
105	0.55	0.67	0.60	0.34%	0.40%
106	0.51	0.60	0.55	0.34%	0.40%
107	0.55	0.59	0.57	2.35%	2.50%
108	0.56	0.59	0.57	1.24%	1.31%
109	0.34	0.20	0.25	0.16%	0.09%
110	0.53	0.46	0.49	0.43%	0.38%
201	0.53	0.54	0.53	2.20%	2.23%
202	0.59	0.54	0.56	3.57%	3.31%
203	0.43	0.37	0.40	0.18%	0.15%
204	0.53	0.43	0.48	0.21%	0.17%
301	0.54	0.59	0.57	2.01%	2.20%
302	0.31	0.11	0.17	0.17%	0.06%
303	0.49	0.50	0.49	4.35%	4.52%
304	0.66	0.52	0.58	1.55%	1.22%
305	0.44	0.51	0.47	2.06%	2.37%
401	0.36	0.26	0.30	1.11%	0.79%
402	0.49	0.51	0.50	2.50%	2.61%
403	0.45	0.44	0.44	3.19%	3.09%
404	0.36	0.24	0.29	0.59%	0.39%
405	0.37	0.23	0.28	0.25%	0.16%
406	0.37	0.28	0.32	0.40%	0.31%
407	0.45	0.39	0.42	0.47%	0.41%
408	0.30	0.16	0.21	1.52%	0.83%
409	0.16	0.08	0.11	0.28%	0.15%
410	0.50	0.47	0.49	2.00%	1.88%
411	0.63	0.69	0.66	8.50%	9.38%
412	0.43	0.15	0.23	0.63%	0.22%
413	0.46	0.54	0.50	0.37%	0.42%
414	0.51	0.50	0.50	1.22%	1.20%
415	0.33	0.21	0.26	0.13%	0.09%
416	0.54	0.37	0.43	3.07%	2.09%
501	0.64	0.74	0.69	5.72%	6.56%
502	0.71	0.78	0.74	3.24%	3.57%
503	0.56	0.55	0.55	6.00%	5.91%
504	0.60	0.72	0.66	9.47%	11.31%
505	0.48	0.23	0.31	0.68%	0.33%
506	0.69	0.75	0.72	5.54%	6.03%
507	0.35	0.06	0.10	0.12%	0.02%
601	0.49	0.44	0.46	1.55%	1.39%
602	0.26	0.21	0.23	0.30%	0.23%
603	0.53	0.56	0.54	1.10%	1.16%
604	0.51	0.45	0.47	0.49%	0.43%
605	0.61	0.67	0.64	3.94%	4.33%
606	0.47	0.42	0.45	1.48%	1.32%
607	0.58	0.55	0.57	1.44%	1.38%
608	0.38	0.44	0.41	0.26%	0.30%
701	0.59	0.64	0.61	3.53%	3.85%
702	0.56	0.20	0.30	0.09%	0.03%
703	0.72	0.71	0.72	3.06%	3.02%
704	0.47	0.18	0.26	0.37%	0.15%
705	0.40	0.17	0.24	0.86%	0.37%
706	0.41	0.27	0.33	1.19%	0.80%

Table 2: Classification Results - Categories

country	accuracy	precision	recall	f1	n
Argentina	60.78%	0.51	0.56	0.54	1,160
Armenia	62.04%	0.51	0.59	0.62	108
Australia	62.56%	0.45	0.47	0.49	6,215
Austria	54.68%	0.40	0.44	0.41	7,798
Belgium	48.80%	0.37	0.36	0.34	16,221
Bolivia	48.70%	0.36	0.35	0.39	924
Bosnia-Herzegovina	46.62%	0.39	0.38	0.36	1,746
Bulgaria	53.03%	0.40	0.39	0.41	1,005
Canada	52.91%	0.47	0.40	0.42	3,551
Chile	47.04%	0.34	0.39	0.35	6,405
Colombia	59.22%	0.33	0.46	0.38	7,534
Costa Rica	65.51%	0.46	0.51	0.49	6,796
Croatia	67.72%	0.51	0.61	0.62	3,194
Cyprus	57.37%	0.46	0.42	0.44	2,977
Czech Republic	55.94%	0.45	0.42	0.43	5,009
Denmark	53.87%	0.39	0.40	0.36	2,376
Dominican Republic	64.99%	0.53	0.55	0.52	4,107
Ecuador	53.06%	0.46	0.45	0.52	360
Estonia	67.61%	0.54	0.45	0.52	2,865
Finland	58.99%	0.47	0.45	0.45	2,975
France	56.22%	0.42	0.43	0.43	2,236
Germany	55.26%	0.45	0.42	0.41	11,226
Greece	59.55%	0.45	0.48	0.45	1,424
Hungary	59.51%	0.47	0.44	0.45	3,650
Iceland	59.20%	0.49	0.48	0.51	375
Israel	64.31%	0.56	0.54	0.56	1,695
Italy	46.56%	0.34	0.33	0.32	7,606
Latvia	73.45%	0.66	0.68	0.78	113
Lithuania	66.23%	0.47	0.47	0.52	2,055
Mexico	46.07%	0.35	0.42	0.38	777
Montenegro	57.65%	0.53	0.50	0.53	562
Netherlands	52.18%	0.39	0.35	0.36	10,835
New Zealand	62.71%	0.50	0.47	0.46	8,819
North Macedonia	64.29%	0.56	0.46	0.53	9,116
Norway	59.56%	0.37	0.46	0.40	5,505
Panama	60.61%	0.48	0.52	0.49	952
Peru	63.67%	0.48	0.49	0.48	4,790
Poland	54.01%	0.46	0.49	0.47	822
Portugal	53.42%	0.43	0.40	0.39	7,290
Romania	55.53%	0.41	0.46	0.46	398
Russia	57.47%	0.55	0.50	0.50	1,164
Serbia	60.21%	0.53	0.54	0.51	382
Slovakia	54.06%	0.46	0.42	0.42	3,302
Slovenia	56.53%	0.40	0.42	0.44	1,668
South Africa	65.84%	0.44	0.50	0.50	802
South Korea	58.31%	0.35	0.43	0.46	626
Spain	61.88%	0.49	0.45	0.45	9,514
Sweden	60.47%	0.48	0.45	0.48	2,398
Switzerland	53.02%	0.51	0.51	0.55	215
Turkey	63.67%	0.50	0.49	0.48	6,601
Ukraine	56.95%	0.51	0.47	0.49	439
United Kingdom	58.64%	0.50	0.48	0.48	2,998
United States	52.61%	0.43	0.44	0.43	3,657
Uruguay	48.63%	0.32	0.36	0.34	3,582

Table 3: Classification Results - Countries

parfam	accuracy	precision	recall	f1	n
10	53.35%	0.41	0.38	0.37	17,701
20	55.87%	0.42	0.41	0.40	25,211
30	59.24%	0.47	0.45	0.44	39,639
40	55.85%	0.45	0.42	0.42	28,942
50	58.19%	0.48	0.43	0.44	22,771
60	58.77%	0.48	0.43	0.44	31,692
70	55.70%	0.47	0.44	0.43	6,810
80	63.28%	0.42	0.50	0.44	5,390
90	56.45%	0.47	0.43	0.43	8,255
95	51.29%	0.41	0.39	0.38	11,726
98	67.44%	0.46	0.49	0.48	2,377
999	57.64%	0.54	0.54	0.55	406

Table 4: Classification Results - Parfam

prob_estimates	accuracy	n(%)	cum_n(%)
> 95%	95.54%	6.76%	6.76%
90%-95%	88.82%	7.64%	14.41%
85%-90%	82.55%	6.82%	21.22%
80%-85%	76.45%	6.34%	27.56%
75%-80%	72.35%	6.02%	33.58%
70%-75%	65.99%	5.91%	39.49%
65%-70%	61.93%	5.95%	45.44%
60%-65%	56.83%	5.98%	51.42%
55%-60%	52.36%	6.17%	57.59%
50%-55%	47.80%	6.52%	64.11%
45%-50%	44.04%	6.79%	70.90%
40%-45%	39.76%	6.51%	77.41%
35%-40%	34.67%	6.04%	83.46%
30%-35%	31.36%	5.28%	88.74%
25%-30%	25.47%	4.47%	93.21%
20%-25%	21.69%	3.35%	96.56%
15%-20%	16.38%	2.16%	98.72%
10%-15%	13.03%	1.13%	99.85%
5%-10%	9.15%	0.15%	100.00%

Table 5: Model Calibration - Probability Groups