**Table 3.** Relaxed match f1 scores, true positive, false positive, false negative, true negative values, and Cohen's Kappa of entities, calculated by Spacy benchmark / evaluate command

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Model name** | **Entity name** | **Prec** | **Recall** | **F1-Score** | **Supp (TP+FN)** | **TP** | **FP** | **FN** | **TN** | **Kappa** |
| Spacy-transformer | ANAT | 0.99 | 0.99 | 0.99 | 3144 | 3116 | 25 | 28 | 5474 | 0.99 |
| Gemini p 1.5 nshot | ANAT | 0.99 | 0.99 | 0.99 | 3078 | 3038 | 45 | 40 | 5408 | 0.98 |
| Gemini p 2.5 nshot | ANAT | 0.99 | 0.99 | 0.99 | 3259 | 3232 | 24 | 27 | 5229 | 0.99 |
| Gemini p 2.5 5shot | ANAT | 0.99 | 0.97 | 0.98 | 2939 | 2850 | 24 | 89 | 5302 | 0.97 |
| Spacy-transformer | OBS-P | 0.96 | 0.98 | 0.97 | 2727 | 2676 | 103 | 51 | 5813 | 0.96 |
| Gemini p 1.5 nshot | OBS-P | 0.96 | 0.96 | 0.96 | 2681 | 2587 | 113 | 94 | 5737 | 0.94 |
| Gemini p 2.5 nshot | OBS-P | 0.97 | 0.98 | 0.98 | 2535 | 2496 | 68 | 39 | 5909 | 0.97 |
| Gemini p 2.5 5shot | OBS-P | 0.95 | 0.98 | 0.97 | 2525 | 2480 | 133 | 45 | 5607 | 0.95 |
| Spacy-transformer | OBS-A | 0.98 | 0.96 | 0.97 | 562 | 540 | 10 | 22 | 8071 | 0.97 |
| Gemini p 1.5 nshot | OBS-A | 0.96 | 0.96 | 0.96 | 560 | 539 | 21 | 21 | 7950 | 0.96 |
| Gemini p 2.5 nshot | OBS-A | 0.95 | 0.98 | 0.96 | 540 | 527 | 30 | 13 | 7942 | 0.96 |
| Gemini p 2.5 5shot | OBS-A | 0.91 | 0.98 | 0.94 | 549 | 537 | 53 | 12 | 7663 | 0.94 |
| Spacy-transformer | IMP | 0.87 | 0.87 | 0.87 | 485 | 420 | 62 | 65 | 8096 | 0.86 |
| Gemini p 1.5 nshot | IMP | 0.89 | 0.90 | 0.89 | 457 | 410 | 51 | 47 | 8023 | 0.89 |
| Gemini p 2.5 nshot | IMP | 0.85 | 0.93 | 0.89 | 457 | 423 | 74 | 34 | 7981 | 0.88 |
| Gemini p 2.5 5shot | IMP | 0.77 | 0.92 | 0.84 | 474 | 437 | 132 | 37 | 7659 | 0.83 |
| Spacy-transformer | OBS-U | 0.78 | 0.55 | 0.65 | 145 | 80 | 23 | 65 | 8475 | 0.64 |
| Gemini p 1.5 nshot | OBS-U | 0.84 | 0.66 | 0.74 | 124 | 82 | 16 | 42 | 8391 | 0.74 |
| Gemini p 2.5 nshot | OBS-U | 0.81 | 0.66 | 0.73 | 118 | 78 | 18 | 40 | 8736 | 0.73 |
| Gemini p 2.5 5shot | OBS-U | 0.78 | 0.60 | 0.68 | 133 | 80 | 23 | 53 | 8109 | 0.67 |
| Spacy-transformer | O | 0.97 | 0.97 | 0.97 | 1580 | 1537 | 51 | 43 | 7012 | 0.96 |
| Gemini p 1.5 nshot | O | 0.96 | 0.96 | 0.96 | 1631 | 1566 | 63 | 65 | 6837 | 0.95 |
| Gemini p 2.5 nshot | O | 0.98 | 0.94 | 0.96 | 1603 | 1513 | 29 | 90 | 6880 | 0.95 |
| Gemini p 2.5 5shot | O | 0.98 | 0.90 | 0.94 | 1645 | 1482 | 34 | 163 | 6586 | 0.92 |

**Table 4.** Relaxed match precision, recall, f1 scores, accuracy, macro average and weighted average, and Cohen's Kappa of models, calculated by Spacy benchmark / evaluate command. Note that sum of False positives equals to sum of false negatives

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Model name** | **Prec** | **Recall** | **F1-Score** | **Support** | **Support**  **(Seqeval)** | **TP** | **FP** | **FN** | **Kappa** |
| **Spacy-transformer** |  |  |  |  |  |  |  |  |  |
| Accuracy |  |  | 0.97 | 8643 | 7063 | 8369 | 274 | 274 | 0.96 |
| Macro Avg | 0.93 | 0.89 | 0.90 | 8643 | 7063 |  |  |  |  |
| Weighted Avg | 0.97 | 0.97 | 0.97 | 8643 | 7063 |  |  |  |  |
| **Gemini p 1.5 nshot** |  |  |  |  |  |  |  |  |  |
| Accuracy |  |  | 0.96 | 8531 | 6900 | 8222 | 309 | 309 | 0.95 |
| Macro Avg | 0.91 | 0.91 | 0.92 | 8531 | 6900 |  |  |  |  |
| Weighted Avg | 0.96 | 0.96 | 0.96 | 8531 | 6900 |  |  |  |  |
| **Gemini p 2.5 nshot** |  |  |  |  |  |  |  |  |  |
| Accuracy |  |  | 0.97 | 8512 | 6909 | 8269 | 243 | 243 | 0.96 |
| Macro Avg | 0.93 | 0.91 | 0.92 | 8512 | 6909 |  |  |  |  |
| Weighted Avg | 0.97 | 0.97 | 0.97 | 8512 | 6909 |  |  |  |  |
| **Gemini p 2.5 5shot** |  |  |  |  |  |  |  |  |  |
| Accuracy |  |  | 0.95 | 8265 | 6620 | 7866 | 399 | 399 | 0.93 |
| Macro Avg | 0.90 | 0.89 | 0.89 | 8265 | 6620 |  |  |  |  |
| Weighted Avg | 0.95 | 0.95 | 0.95 | 8265 | 6620 |  |  |  |  |