| | | Overall | Measurement | Data | Time Constraint | Temperature | Mass | Size | Non-label Data | Label Data |
|---|---|---|---|---|---|---|---|---|---|---|
| BERT base vs. BERT large | Precision | p-value: 0.16772384227816506, Effect size: 0.6275, Small | p-value: 0.046791615346000 49, Effect size: 0.31625, negligible | p-value: 0.46517480 36184682, Effect size: 0.5675, Small | p-value: 0.04986369 021054049, Effect size: 0.68125, Medium | p-value: 4.228627942 2220743e-07, Effect size: 0.9675, Large | p-value: 0.01383349 8919863277, Effect size: 0.7275, Large | p-value: 0.54277170 15170068, Effect size: 0.44375, negligible | p-value: 6.3018482213 92269e-08, Effect size: 1.0, Large | p-value: 0.09619629 624919565, Effect size: 0.65375, Medium |
| | Recall | p-value: 0.4818673682837039, Effect size: 0.565, Small | p-value: 0.93532245 00925261, Effect size: 0.4925, negligible | p-value: 0.01929237 9706561613, Effect size: 0.71625, Large | p-value: 0.04248307 982571583, Effect size: 0.6875, Medium | p-value: 8.512328416 895721e-08, Effect size: 0.995, Large | p-value: 0.00028928 3716172173 85, Effect size: 0.835, Large | p-value: 0.44074988 359834266, Effect size: 0.57125, Small | p-value: 6.3018482213 92269e-08, Effect size: 1.0, Large | p-value: 0.56085169 66002302, Effect size: 0.55375, negligible |
| | F-score | p-value: 0.01729348 6431260986, Effect size: 0.72, Large | p-value: 0.00305662 9228174791, Effect size: 0.22625, negligible | p-value: 0.00802755 3206315485, Effect size: 0.745, Large | p-value: 0.01606401 314826911, Effect size: 0.7225, Large | p-value: 7.326768199 47719e-08, Effect size: 0.9975, Large | p-value: 0.00035614 8463090723 2, Effect size: 0.83, Large | p-value: 0.63594455 11344555, Effect size: 0.45625, negligible | p-value: 6.3018482213 92269e-08, Effect size: 1.0, Large | p-value: 0.13681554 418853384, Effect size: 0.6375, Small |
| BERT base vs. ALBERT | Precision | p-value: 0.02149031 7747919475, Effect size: 0.7125, Large | p-value: 0.96763460 83712113, Effect size: 0.50374999 99999999, negligible | p-value: 0.00053535 8151340559 7, Effect size: 0.82, Large | p-value: 0.78677493 20074033, Effect size: 0.525, negligible | p-value: 6.261396055 8707055e-06, Effect size: 0.9175, Large | p-value: 0.29144660 464450534, Effect size: 0.5975, Small | p-value: 0.63594455 11344555, Effect size: 0.45625000 000000004, negligible | p-value: 0.0834150833 5601154, Effect size: 0.66, Medium | p-value: 0.65536143 90103917, Effect size: 0.54125, negligible |
| | Recall | p-value: 0.03851496 043112118, Effect size: 0.69125, Medium | p-value: 0.00015246 9453821355 6, Effect size: 0.85, Large | p-value: 0.30399527 78179398, Effect size: 0.405, Small | p-value: 0.25591315 01098182, Effect size: 0.395, Small | p-value: 6.795615128 173358e-08, Effect size: 0.99875, Large | p-value: 0.86043097 37352772, Effect size: 0.48375, negligible | p-value: 0.00016098 1339659183 95, Effect size: 0.84875, Large | p-value: 0.0764309455 7954021, Effect size: 0.66375, Medium | p-value: 0.16359572 56111319, Effect size: 0.62875, Small |
| | F-score | p-value: 1.03770175 92246101e-05, Effect size: 0.9075, Large | p-value: 0.00279860 2190907080 4, Effect size: 0.77625, Large | p-value: 0.62632806 30826087, Effect size: 0.545, negligible | p-value: 0.31040225 614615824, Effect size: 0.40625, Small | p-value: 6.301848221 392269e-08, Effect size: 1.0, Large | p-value: 0.83923221 30718609, Effect size: 0.48124999 999999996, negligible | p-value: 0.08341508 335601154, Effect size: 0.66, Medium | p-value: 0.0658554842 3829655, Effect size: 0.67, Medium | p-value: 0.17621399 555130546, Effect size: 0.625, Small |
| BERT base vs. RoBERTa | Precision | p-value: 0.79719741 92691748, Effect size: 0.52375, negligible | p-value: 0.25591315 01098182, Effect size: 0.395, Small | p-value: 0.62632806 30826087, Effect size: 0.545, negligible | p-value: 0.77639050 55505131, Effect size: 0.47375, negligible | p-value: 0.310402256 14615824, Effect size: 0.59375, Small | p-value: 0.11049878 401993231, Effect size: 0.35250000 000000004, Medium | p-value: 0.36484165 4608659, Effect size: 0.41625, Small | p-value: 0.1104987840 1993231, Effect size: 0.6475, Medium | p-value: 0.63594455 11344555, Effect size: 0.54375, negligible |
| | Recall | p-value: 0.52498691 01903615, Effect size: 0.44125000 000000003, negligible | p-value: 4.41597598 995539e-05, Effect size: 0.12250000 000000005, Large | p-value: 0.33691523 838231996, Effect size: 0.41125, Small | p-value: 0.72509989 43601429, Effect size: 0.5325, negligible | p-value: 0.797197419 2691748, Effect size: 0.523749999 9999999, negligible | p-value: 0.22869435 423745565, Effect size: 0.38875000 000000004, Small | p-value: 0.01547860 2442838574, Effect size: 0.27625, Large | p-value: 0.0834150833 5601154, Effect size: 0.66, Medium | p-value: 0.00050907 2918078457, Effect size: 0.17874999 999999996, Large |
| | F-score | p-value: 0.49033426 450951045, Effect size: 0.56375, Small | p-value: 1.17533049 53705375e-05, Effect size: 0.09499999 999999997, Large | p-value: 0.72509989 43601429, Effect size: 0.4675, negligible | p-value: 0.91383676 24408553, Effect size: 0.51, negligible | p-value: 0.913836762 4408553, Effect size: 0.51, negligible | p-value: 0.14784658 271436188, Effect size: 0.36624999 999999996, Small | p-value: 0.00396623 8640223918, Effect size: 0.23375, Large | p-value: 0.0294408844 1208284, Effect size: 0.70125, Medium | p-value: 0.00148097 7185796863 6, Effect size: 0.20625000 000000004, Large |