	Amazon31	Banking77	Claude9
6.25% VS CFT	0.685, 0.685, 0.683, 0.682	0.551, 0.551, 0.497, 0.497	0.904, 0.904, 0.149, 0.874
Supervised Only	0.660, 0.660, 0.659, 0.658	0.234, 0.234, 0.167, 0.167	0.907, 0.907, 0.110, 0.864
12.5% VS CFT	0.711, 0.711, 0.710, 0.709	0.676, 0.676, 0.650, 0.650	0.906, 0.906, 0.177, 0.877
Supervised Only	0.701, 0.701, 0.702, 0.701	0.378, 0.378, 0.318, 0.318	0.908, 0.908, 0.120, 0.866
25% VS CFT	0.742, 0.742, 0.741, 0.740	0.751, 0.751, 0.742, 0.742	0.906, 0.906, 0.351, 0.895
Supervised Only	0.736, 0.736, 0.736, 0.735	0.567, 0.567, 0.530, 0.530	0.906, 0.906, 0.173, 0.870
50% VS CFT	0.769, 0.769, 0.768, 0.767	0.82, 0.82, 0.818, 0.818	0.914, 0, 914, 0.493, 0.910
Supervised Only	0.766, 0.766, 0.766, 0.765	0.753, 0.753, 0.745, 0.745	0.911, 0.911, 0.256, 0.884
100% VS CFT	0.791, 0.791, 0.791, 0.790	0.866, 0.866, 0.865, 0.865	0.914, 0.914, 0.558, 0.916
Supervised Only	0.790, 0.790, 0.790, 0.789	0.846, 0.846, 0.844, 0.844	0.913, 0.913, 0.349, 0.900

Table 1: Accuracy, F1_micro, F1_macro, F1_weighted of the dataset with different validation size (VS).

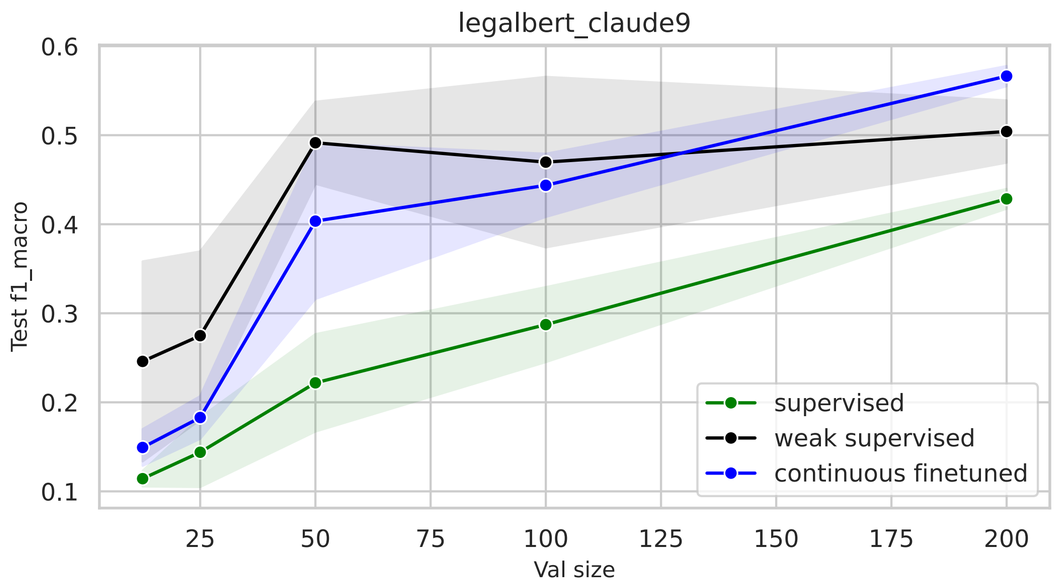

Figure 1: F1 Macro vs. Val size running three pipelines on Claude9 with Legal BERT

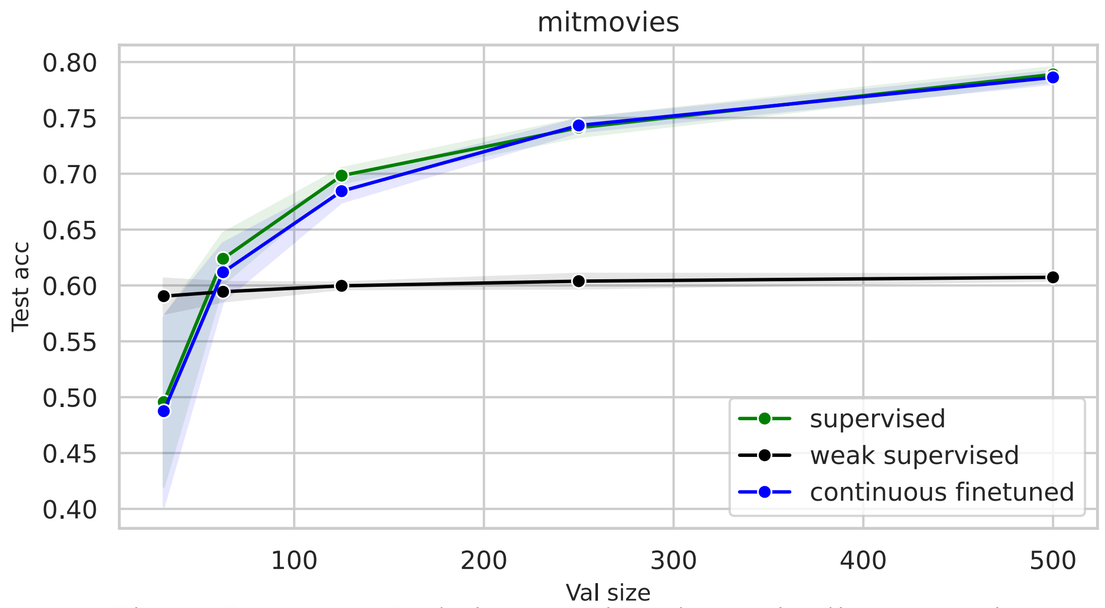

Figure 2: Acc vs. Val size running three pipelines on mit-movies with Sequential BERT

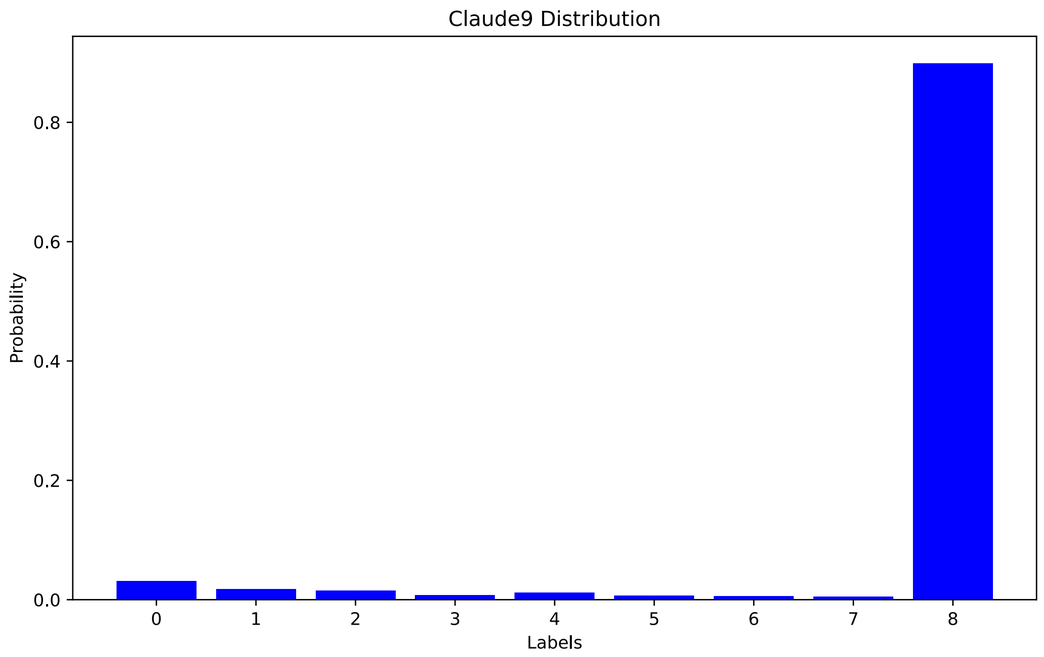

Figure 3: Label Distribution of Claude9

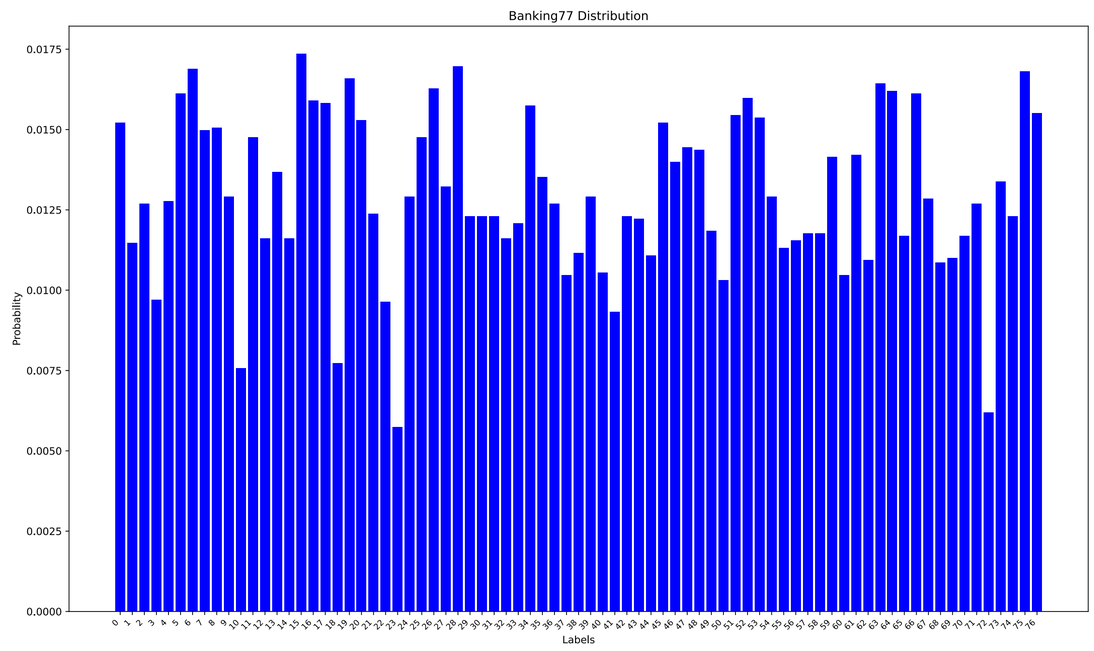

Figure 4: Label Distribution of Banking77

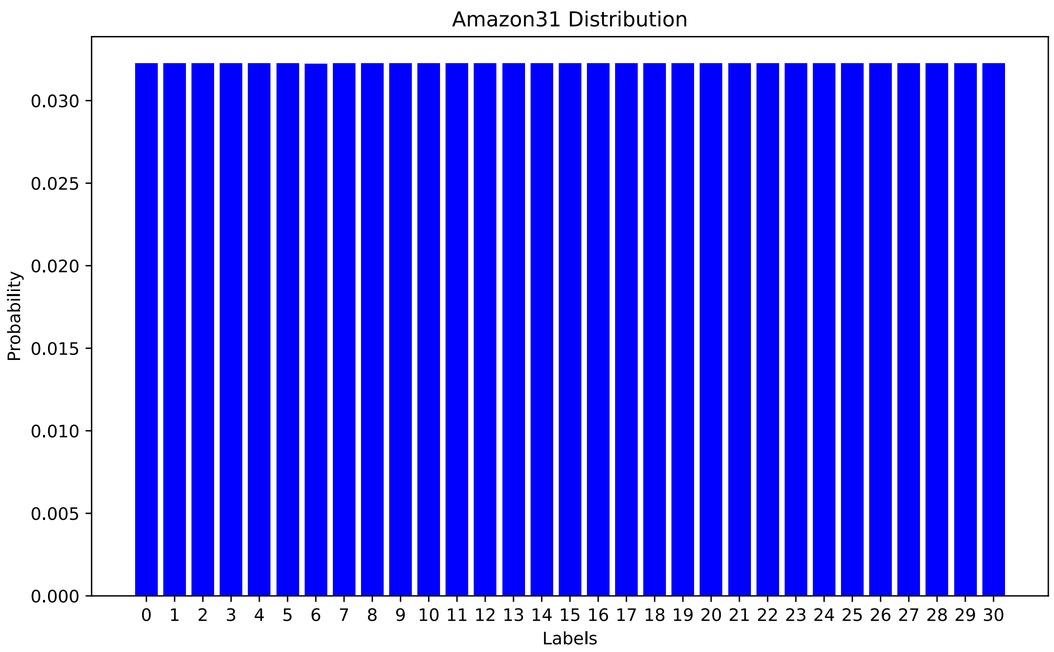

Figure 5: Label Distribution of Amazon31

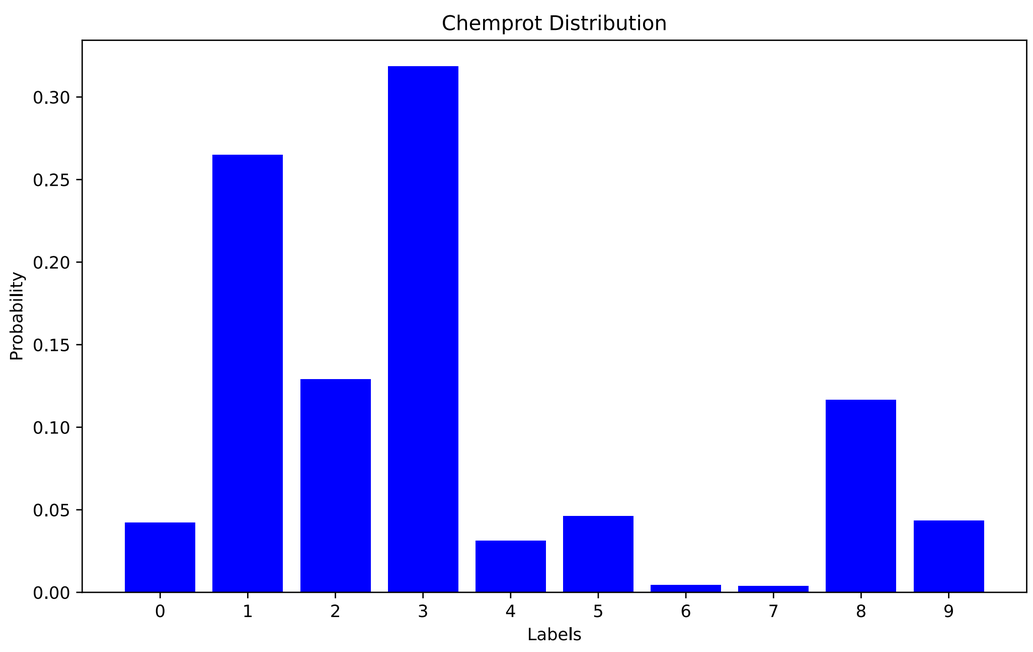

Figure 6: Label Distribution of ChemProt