

TABLE I: Performance of stand-alone models on dataset 1 (Pr = Precision, Rc = Recall, F1 = F1-score, Sup = Support)

Dataset 1 Label	Baseline				LightGBM				BERT				RoBERTa			
	Pr	Rc	F1	Sup	Pr	Rc	F1	Sup	Pr	Rc	F1	Sup	Pr	Rc	F1	Sup
Action on Issue	0.45	0.39	0.42	61	0.52	0.43	0.45	61	0.76	0.67	0.69	61	0.7	0.7	0.69	61
Bug Reproduction	0.53	0.47	0.5	245	0.55	0.51	0.52	245	0.51	0.5	0.5	245	0.49	0.59	0.53	245
Contribution	0.25	0.19	0.21	83	0.25	0.14	0.18	83	0.58	0.48	0.51	83	0.53	0.52	0.51	83
Expected Behaviour	0.42	0.28	0.33	124	0.47	0.4	0.42	124	0.34	0.32	0.33	124	0.42	0.42	0.42	124
Investigation	0.6	0.65	0.62	377	0.65	0.67	0.66	377	0.43	0.47	0.45	377	0.46	0.5	0.48	377
Motivation	0.56	0.53	0.54	288	0.6	0.52	0.55	288	0.31	0.36	0.34	288	0.38	0.41	0.39	288
Observed Bug Behaviour	0.56	0.7	0.62	131	0.63	0.74	0.67	131	0.4	0.27	0.32	131	0.33	0.28	0.3	131
Potential New Issues	0.67	0.66	0.66	230	0.75	0.75	0.75	230	0.23	0.23	0.23	230	0.26	0.21	0.23	230
Social Conversation	0.63	0.62	0.63	798	0.64	0.64	0.64	798	0.82	0.78	0.8	798	0.84	0.81	0.83	798
Solution Discussion	0.68	0.71	0.69	1411	0.72	0.75	0.73	1411	0.61	0.58	0.59	1411	0.63	0.55	0.59	1411
Task Progress	0.27	0.14	0.18	125	0.28	0.14	0.18	125	0.47	0.43	0.44	125	0.5	0.5	0.48	125
Usage	0.65	0.82	0.73	368	0.71	0.8	0.75	368	0.49	0.55	0.51	368	0.46	0.6	0.52	368
Workarounds	0.58	0.45	0.49	89	0.61	0.48	0.5	89	0.31	0.31	0.31	89	0.33	0.38	0.35	89
Weighted avg/total	0.61	0.62	0.61	4330	0.64	0.65	0.64	4330	0.55	0.54	0.54	4330	0.57	0.56	0.56	4330

TABLE II: Performance of stand-alone models on dataset 2 (Pr = Precision, Rc = Recall, F1 = F1-score, Sup = Support)

Dataset 2 Label	Baseline				LightGBM				BERT				RoBERTa			
	Pr	Rc	F1	Sup	Pr	Rc	F1	Sup	Pr	Rc	F1	Sup	Pr	Rc	F1	Sup
Action on Issue	0.37	0.23	0.27	96	0.26	0.36	0.3	96	0.65	0.76	0.7	96	0.69	0.79	0.73	96
Bug Reproduction	0.36	0.17	0.23	176	0.28	0.31	0.29	176	0.6	0.56	0.58	176	0.55	0.61	0.58	176
Contribution	0.26	0.13	0.17	109	0.19	0.2	0.19	109	0.62	0.61	0.6	109	0.61	0.61	0.59	109
Expected Behaviour	0.19	0.17	0.17	24	0.27	0.33	0.27	24	0.21	0.17	0.18	24	0.33	0.38	0.33	24
Investigation	0.63	0.81	0.71	1891	0.71	0.6	0.65	1891	0.67	0.7	0.69	1891	0.7	0.69	0.69	1891
Motivation	0.57	0.5	0.5	18	0.48	0.67	0.53	18	0.28	0.17	0.2	18	0.28	0.17	0.2	18
Observed Bug Behaviour	0.44	0.31	0.35	81	0.22	0.31	0.26	81	0.27	0.25	0.26	81	0.33	0.35	0.33	81
Social Conversation	0.43	0.38	0.4	469	0.35	0.44	0.39	469	0.64	0.56	0.59	469	0.67	0.58	0.62	469
Solution Discussion	0.61	0.58	0.59	856	0.6	0.65	0.62	856	0.56	0.58	0.57	856	0.56	0.59	0.58	856
Task Progress	0.36	0.18	0.23	233	0.26	0.29	0.27	233	0.43	0.38	0.4	233	0.5	0.52	0.51	233
Testing	0.56	0.31	0.39	261	0.45	0.45	0.45	261	0.65	0.58	0.61	261	0.65	0.59	0.62	261
Workarounds	0.22	0.07	0.11	27	0.21	0.22	0.21	27	0.56	0.59	0.53	27	0.72	0.74	0.68	27
Weighted avg/total	0.55	0.58	0.55	4241	0.55	0.53	0.54	4241	0.61	0.61	0.61	4241	0.63	0.63	0.63	4241

TABLE III: Performance of stand-alone models on dataset 3 (Pr = Precision, Rc = Recall, F1 = F1-score, Sup = Support)

Dataset 3 Label	Baseline				LightGBM				BERT				RoBERTa			
	Pr	Rc	F1	Sup	Pr	Rc	F1	Sup	Pr	Rc	F1	Sup	Pr	Rc	F1	Sup
Action on Issue	0.4	0.31	0.35	157	0.34	0.35	0.34	157	0.72	0.68	0.69	157	0.67	0.78	0.72	157
Bug Reproduction	0.53	0.36	0.42	421	0.39	0.43	0.4	421	0.56	0.51	0.53	421	0.51	0.59	0.54	421
Contribution	0.34	0.16	0.22	192	0.22	0.19	0.2	192	0.65	0.54	0.58	192	0.59	0.59	0.59	192
Expected Behaviour	0.47	0.3	0.35	148	0.45	0.37	0.39	148	0.44	0.29	0.35	148	0.44	0.38	0.39	148
Investigation	0.58	0.77	0.67	2268	0.66	0.68	0.67	2268	0.61	0.67	0.64	2268	0.64	0.65	0.64	2268
Motivation	0.58	0.57	0.57	306	0.53	0.59	0.56	306	0.31	0.3	0.3	306	0.39	0.35	0.37	306
Observed Bug Behaviour	0.52	0.57	0.54	212	0.49	0.61	0.54	212	0.37	0.23	0.28	212	0.37	0.27	0.31	212
Potential New Issues	0.73	0.61	0.66	242	0.7	0.76	0.73	242	0.24	0.19	0.21	242	0.23	0.16	0.18	242
Social Conversation	0.57	0.51	0.54	1267	0.54	0.53	0.54	1267	0.75	0.7	0.72	1267	0.76	0.72	0.74	1267
Solution Discussion	0.65	0.63	0.64	2267	0.72	0.6	0.66	2267	0.56	0.6	0.58	2267	0.58	0.59	0.59	2267
Task Progress	0.31	0.18	0.23	358	0.22	0.24	0.23	358	0.4	0.35	0.37	358	0.41	0.4	0.4	358
Testing	0.57	0.36	0.44	278	0.41	0.5	0.45	278	0.57	0.54	0.55	278	0.55	0.53	0.54	278
Usage	0.66	0.77	0.71	372	0.66	0.84	0.74	372	0.54	0.53	0.53	372	0.5	0.59	0.54	372
Workarounds	0.56	0.4	0.46	116	0.51	0.47	0.48	116	0.46	0.36	0.4	116	0.43	0.39	0.4	116
Weighted avg/total	0.58	0.59	0.57	8604	0.59	0.58	0.58	8604	0.57	0.57	0.57	8604	0.58	0.59	0.58	8604

TABLE IV: Performance of hierarchical classification (Pr = Precision, Rc = Recall, F1 = F1-score, Sup = Support)

	Dataset 1				Dataset 2				Dataset 3			
	Pr	Rc	F1	Sup	Pr	Rc	F1	Sup	Pr	Rc	F1	Sup
Cluster Detection	0.74	0.74	0.74	4330	0.73	0.74	0.73	4241	0.72	0.72	0.72	8604
Cluster 1	0.73	0.74	0.73	1297	0.56	0.54	0.55	907	0.67	0.66	0.66	2216
Cluster 2	0.75	0.75	0.75	1165	0.79	0.82	0.79	2451	0.76	0.75	0.75	3633
Cluster 3	0.92	0.92	0.92	1868	0.96	0.97	0.96	883	0.93	0.93	0.93	2755

TABLE V: Performance of ensemble classifier (Pr = Precision, Rc = Recall, F1 = F1-score, Sup = Support)

Label	Dataset 1				Dataset 2				Dataset 3			
	Pr	Rc	F1	Sup	Pr	Rc	F1	Sup	Pr	Rc	F1	Sup
Action on Issue	0.71	0.7	0.68	61	0.66	0.81	0.72	96	0.65	0.78	0.7	157
Bug Reproduction	0.62	0.64	0.63	245	0.64	0.57	0.6	176	0.63	0.62	0.62	421
Contribution	0.5	0.37	0.41	83	0.66	0.54	0.59	109	0.69	0.55	0.61	192
Expected Behaviour	0.59	0.44	0.5	124	0.54	0.25	0.33	24	0.6	0.43	0.49	148
Investigation	0.63	0.64	0.63	377	0.73	0.82	0.77	1891	0.69	0.79	0.73	2268
Motivation	0.57	0.51	0.54	288	0.44	0.39	0.39	18	0.64	0.64	0.63	306
Observed Bug Behaviour	0.63	0.64	0.63	131	0.45	0.28	0.34	81	0.57	0.57	0.56	212
Potential New Issues	0.72	0.53	0.6	230					0.72	0.71	0.71	242
Social Conversation	0.79	0.7	0.74	798	0.66	0.52	0.58	469	0.78	0.65	0.71	1267
Solution Discussion	0.67	0.77	0.72	1411	0.69	0.66	0.67	856	0.71	0.72	0.72	2267
Task Progress	0.5	0.26	0.33	125	0.53	0.41	0.46	233	0.51	0.31	0.38	358
Testing					0.65	0.62	0.63	261	0.65	0.59	0.62	278
Usage	0.68	0.78	0.72	368					0.7	0.82	0.76	372
Workarounds	0.62	0.48	0.53	89	0.69	0.7	0.67	27	0.6	0.64	0.61	116
Weighted avg/total	0.67	0.67	0.66	4330	0.68	0.69	0.68	4241	0.69	0.69	0.68	8604