Target	Apache Ant	ArgoUML	Columba	EMF	Hibernate	JEdit	JFreeChart	JMeter	JRuby	SQuirrel	Ave-? → 1
Apache Ant	-	0.108	0.207	0.224	0.189	0.178	0.201	0.181	0.206	0.156	0.183
ArgoUML	0.681	-	0.657	0.741	0.647	0.659	0.699	0.706	0.642	0.692	0.680
Columba	0.511	0.249	-	0.485	0.438	0.360	0.349	0.556	0.360	0.411	0.413
EMF	0.443	0.125	0.160	-	0.123	0.140	0.214	0.193	0.221	0.173	0.199
Hibernate	0.756	0.410	0.588	0.747	-	0.581	0.629	0.686	0.575	0.630	0.622
JEdit	0.374	0.252	0.313	0.279	0.327	-	0.379	0.348	0.422	0.416	0.346
JFreeChart	0.262	0.209	0.338	0.546	0.368	0.291	-	0.405	0.397	0.246	0.340
JMeter	0.674	0.297	0.436	0.477	0.461	0.522	0.512	-	0.477	0.482	0.482
JRuby	0.648	0.428	0.610	0.542	0.635	0.534	0.513	0.624	-	0.507	0.560
SQuirrel	0.505	0.205	0.286	0.480	0.353	0.318	0.435	0.436	0.382	-	0.378
Ave-1 →?	0.539	0.254	0.399	0.502	0.393	0.398	0.437	0.459	0.409	0.413	-‘

Table 1: Our approach: Precision in 1 → 1 Setting (row ? → 1 represents the nine experiments of the same testing project and column 1 →? represents the nine experiments of the same training project.) The best F1-scores of each method are in bold and the worst ones are underlined.

Target	Apache Ant	ArgoUML	Columba	EMF	Hibernate	JEdit	JFreeChart	JMeter	JRuby	SQuirrel	Ave-? → 1
Apache Ant	-	0.471	0.559	0.353	0.451	0.539	0.412	0.373	0.569	0.373	0.456
ArgoUML	0.640	-	0.739	0.882	0.752	0.834	0.824	0.747	0.724	0.804	0.768
Columba	0.695	0.820	-	0.758	0.805	0.820	0.695	0.781	0.797	0.797	0.774
EMF	0.419	0.716	0.554	-	0.541	0.581	0.500	0.459	0.514	0.459	0.527
Hibernate	0.623	0.658	0.700	0.743	-	0.769	0.714	0.700	0.706	0.724	0.704
JEdit	0.344	0.554	0.415	0.297	0.446	-	0.410	0.328	0.503	0.456	0.417
JFreeChart	0.436	0.752	0.752	0.525	0.762	0.564	-	0.743	0.762	0.554	0.650
JMeter	0.681	0.677	0.713	0.784	0.770	0.755	0.702	-	0.702	0.745	0.725
JRuby	0.480	0.742	0.809	0.554	0.841	0.611	0.679	0.836	-	0.548	0.678
SQuirrel	0.488	0.632	0.577	0.662	0.652	0.672	0.632	0.562	0.582	-	0.607
Ave-1 →?	0.530	0.669	0.646	0.618	0.669	0.683	0.619	0.614	0.651	0.607	-

Table 2: Our approach: Precision in 1 → 1 Setting (row ? → 1 represents the nine experiments of the same testing project and column 1 →? represents the nine experiments of the same training project.) The best F1-scores of each method are in bold and the worst ones are underlined.

Target	Apache Ant	ArgoUML	Columba	EMF	Hibernate	JEdit	JFreeChart	JMeter	JRuby	SQuirrel	Ave-? → 1	Ave-? → 1 (w/o ft)	Imp.-2
Apache Ant	-	0.152	0.277	0.338	0.353	0.255	0.299	0.329	0.312	0.279	0.288	0.183	57.50%
ArgoUML	0.591	-	0.568	0.548	0.625	0.594	0.585	0.595	0.610	0.585	0.589	0.68	-13.38%
Columba	0.372	0.175	-	0.491	0.277	0.259	0.311	0.375	0.268	0.309	0.315	0.413	-23.68%
EMF	0.466	0.315	0.305	-	0.494	0.443	0.437	0.466	0.500	0.437	0.429	0.199	115.69%
Hibernate	0.726	0.486	0.631	0.666	-	0.640	0.639	0.675	0.671	0.695	0.648	0.622	4.13%
JEdit	0.402	0.294	0.387	0.409	0.424	-	0.432	0.375	0.493	0.434	0.406	0.346	17.21%
JFreeChart	0.472	0.260	0.355	0.489	0.450	0.368	-	0.462	0.461	0.461	0.420	0.34	23.46%
JMeter	0.690	0.390	0.561	0.591	0.563	0.517	0.595	-	0.623	0.644	0.575	0.482	19.27%
JRuby	0.634	0.476	0.504	0.569	0.661	0.639	0.567	0.627	-	0.630	0.590	0.56	5.30%
SQuirrel	0.480	0.262	0.429	0.436	0.406	0.461	0.421	0.475	0.410	-	0.420	0.378	11.11%
Ave-1 →?	0.537	0.312	0.446	0.504	0.473	0.464	0.476	0.487	0.483	0.497	-	-	-
Ave-1 →? (w/o ft)	0.539	0.254	0.399	0.502	0.393	0.398	0.437	0.459	0.409	0.413	-	-	-
Imp.-1	-0.37%	22.92%	11.86%	0.42%	20.24%	16.58%	8.98%	6.00%	18.12%	20.37%	-	-	-

Table 3: Traditional Text Mining: Precision in $1 \rightarrow 1$ Setting with 20% Fine-Tuning Data (row $? \rightarrow 1$ represents the nine experiments of the same testing project and column $1 \rightarrow ?$ represents the nine experiments of the same training project.

Target	Apache Ant	ArgoUML	Columba	EMF	Hibernate	JEdit	JFreeChart	JMeter	JRuby	SQuirrel	Ave-? → 1	Ave-? → 1 (w/o ft)	Imp.-2
Apache Ant	-	0.530	0.614	0.602	0.578	0.578	0.602	0.590	0.494	0.639	0.581	0.456	27.36%
ArgoUML	0.651	-	0.811	0.743	0.778	0.754	0.796	0.769	0.732	0.813	0.761	0.768	-0.94%
Columba	0.508	0.571	-	0.884	0.492	0.667	0.730	0.667	0.540	0.730	0.643	0.774	-16.90%
EMF	0.853	0.842	0.683	-	0.884	0.947	0.842	0.863	0.926	0.905	0.861	0.527	63.29%
Hibernate	0.610	0.643	0.660	0.797	-	0.740	0.803	0.733	0.687	0.753	0.714	0.704	1.42%
JEdit	0.342	0.542	0.452	0.465	0.394	-	0.471	0.348	0.477	0.406	0.433	0.417	3.84%
JFreeChart	0.453	0.600	0.733	0.573	0.480	0.573	-	0.560	0.707	0.547	0.581	0.650	-10.67%
JMeter	0.721	0.616	0.707	0.769	0.664	0.716	0.738	-	0.672	0.734	0.704	0.725	-2.88%
JRuby	0.786	0.743	0.757	0.868	0.770	0.855	0.750	0.786	-	0.836	0.795	0.678	17.19%
SQuirrel	0.551	0.526	0.558	0.628	0.500	0.673	0.583	0.603	0.551	-	0.575	0.607	-5.31%
Ave-1 →?	0.608	0.624	0.664	0.703	0.616	0.723	0.702	0.658	0.643	0.707	-	-	-
Ave-1 →? (w/o ft)	0.53	0.669	0.646	0.618	0.669	0.683	0.619	0.614	0.651	0.607	-	-	-
Imp.-1	14.78%	-6.78%	2.77%	13.79%	-7.99%	5.79%	13.35%	7.11%	-1.25%	16.47%	-	-	-

Table 4: Traditional Text Mining: Recall in 1 → 1 Setting with 20% Fine-Tuning Data (row ? → 1 represents the nine experiments of the same testing project and column 1 →? represents the nine experiments of the same training project.