

1-2-gem-vault_idf

Features: Counter with 1-2-gem and idf
Maximum number of features considered: 100

Logistic Regression

Feature number	Time to 1 TP	TN	FP	FN	Accuracy	Precision	Recall	F1-Score		
1	200	0.046	691	1513	432	574	0.0802	0.01186	0.5023	0.3717
2	200	0.0462	672	1548	379	604	0.08281	0.03639	0.5095	0.3737
3	200	0.0397	699	1517	441	573	0.0813	0.0227	0.5085	0.3688
4	200	0.0396	699	1516	411	584	0.0806	0.02836	0.5124	0.3767
5	200	0.0467	675	1534	419	577	0.0803	0.017	0.5113	0.3768
Average	0.0409	677.2	1525	416.4	514	581.6	0.08019	0.01946	0.50909	0.3738

S2D Classifier

Feature number	Time to 1 TP	TN	FP	FN	Accuracy	Precision	Recall	F1-Score
1	200	0.0436	701	1523	401	575	0.0802	0.5007
2	200	0.04298	701	1523	401	575	0.0805	0.5037
3	200	0.04198	777	1420	538	485	0.08856	0.59087
4	200	0.04290	730	1465	462	543	0.08584	0.61242
5	200	0.04198	492	1657	286	755	0.07485	0.63239
Average	0.04250	714.6	1472	489.4	544	0.68331	0.60625	0.58772

Linear SVC

Feature number	Time to 1 TP	TN	FP	FN	Accuracy	Precision	Recall	F1-Score		
1	200	0.06048	701	1516	429	554	0.08281	0.02023	0.50857	0.38794
2	200	0.10692	692	1533	391	544	0.08219	0.0306	0.50448	0.38057
3	200	0.11593	693	1510	448	559	0.08531	0.02089	0.50992	0.37564
4	200	0.11693	702	1493	434	571	0.08594	0.01796	0.51545	0.38281
5	200	0.12593	694	1531	422	563	0.08219	0.01844	0.50852	0.38139
Average	0.0366	699.4	1515.6	426.8	503.2	0.0809	0.01925	0.5089	0.3717	

NaïveBayes (gemmar+idf)

Feature number	Time to 1 TP	TN	FP	FN	Accuracy	Precision	Recall	F1-Score		
1	200	32.6862	740	1393	552	515	0.08056	0.57278	0.58964	0.58108
2	200	35.8643	726	1469	515	550	0.08719	0.58501	0.56997	0.57688
3	200	33.2531	705	1435	523	536	0.08906	0.57445	0.56944	0.57143
4	200	32.6803	768	1370	557	505	0.08813	0.57962	0.60333	0.59122
5	200	31.8468	726	1450	503	521	0.088	0.59072	0.5822	0.58643
Average		33.2594	733.2	1411.4	503	524.4	0.08709	0.58501	0.58251	0.58141

SVC (gemmar+idf)

Feature number	Time to 1 TP	TN	FP	FN	Accuracy	Precision	Recall	F1-Score		
1	200	38.1262	240	1826	119	1015	0.04563	0.08852	0.19124	0.2914
2	200	37.7095	220	1849	75	1056	0.04656	0.74376	0.17241	0.28038
3	200	37.7034	209	1873	85	1033	0.05083	0.71088	0.16828	0.27214
4	200	38.2643	222	1848	79	1051	0.04608	0.73754	0.17349	0.28308
5	200	38.7795	203	1957	86	1044	0.04688	0.75042	0.15879	0.26432
Average		36.873	218.8	1852.6	88.8	1039.8	0.04731	0.71363	0.17382	0.2792

Gaussian Naive Bayes

Feature number	Time to 1 TP	TN	FP	FN	Accuracy	Precision	Recall	F1-Score		
1	200	0.07297	895	1192	753	350	0.05531	0.54584	0.72112	0.62135
2	200	0.08199	904	1223	701	362	0.06156	0.5605	0.70063	0.62278
3	200	0.06896	901	1105	793	341	0.04563	0.53188	0.72544	0.61376
4	200	0.06794	923	1143	784	350	0.04563	0.54071	0.72506	0.61946
5	200	0.06794	899	1181	772	348	0.05	0.5306	0.72093	0.61618
Average		0.07196	904.4	1180.8	760.6	354.2	0.05163	0.54539	0.71863	0.61871

Decision Tree

1	200	0.05097	997	1506	439	258	0.79219	0.09429	0.79442	0.74099
2	200	0.06156	996	1527	397	310	0.77906	0.07673	0.75705	0.7321
3	200	0.04052	951	1525	433	291	0.77375	0.08714	0.7657	0.7243
4	200	0.02451	989	1484	443	284	0.77281	0.09094	0.7769	0.73124
5	200	0.04451	956	1523	430	291	0.77449	0.08975	0.76994	0.72617
Average		0.05341	971.8	1513	428.4	289.8	0.77765	0.09411	0.77214	0.73996

RandomForest (10 estimators)

	1	200	0.47512	997	1536	409	258	0.79156	0.7091	0.79442	0.74934
	2	200	0.46871	989	1549	375	287	0.79213	0.72507	0.77308	0.74924
	3	200	0.44674	946	1550	365	296	0.79025	0.71994	0.75167	0.74222
	4	200	0.44673	985	1554	373	285	0.79435	0.72594	0.77612	0.74019
	5	200	0.44673	981	1530	422	285	0.79469	0.69672	0.79899	0.7401
Average			0.45701	980.2	1551.8	386.6	274.4	0.79125	0.71575	0.7788	0.74582

AdaBoost

1	200	1.8176	695	1476	469	560	0.6784	0.50708	0.55378	0.57462
2	200	1.8106	694	1563	361	612	0.05904	0.0478	0.52038	0.57714
3	200	1.7409	615	1508	372	637	0.08781	0.0231	0.49017	0.50162
4	200	1.6702	658	1548	391	612	0.08875	0.01323	0.51569	0.5682
5	200	1.725	683	1534	419	564	0.08281	0.01979	0.54579	0.56812
Average		1.7797	663	1541	460.4	595.6	0.08875	0.02421	0.52679	0.57096

XGBost

1	200	0.45045	595	1627	318	595	0.0802	0.03085	0.4352	0.32033
2	200	0.45012	593	1628	246	743	0.0805	0.04772	0.41771	0.31873
3	200	0.45031	593	1628	246	743	0.0805	0.04763	0.43078	0.32045
4	200	0.08152	547	1681	246	726	0.0825	0.08079	0.42069	0.32593
5	200	0.08045	544	1687	266	650	0.70031	0.07561	0.44427	0.33604
Average		0.31738	546.8	1673.8	267.8	711.8	0.09385	0.07198	0.43463	0.32747

Multi-Layer Perceptrons

		200	23.5805	1020	1950	440	236	0.79103	0.7046	0.81662	0.7551
		200	23.631	969	991	1937	317	0.76813	0.7265	0.79157	0.7683
		200	23.3834	1009	1950	430	236	0.78583	0.65603	0.80566	0.7494
		200	23.4585	971	1548	379	320	0.78719	0.71926	0.76277	0.74037
	4	200	23.8078	976	1496	457	271	0.77725	0.68109	0.73658	0.72836
Average			24.028	988.2	1924	417.4	270.4	0.78506	0.70565	0.78538	0.74716

Features: Counter with 1-2-gem and idf
Maximum: 1000

Logistic Regression

1	2000	0.10792	748	1612	393	447	0.7375	0.03537	0.02034	0.04041	1	10000	0.27032	936
2	2000	0.10764	820	1596	375	447	0.7425	0.03619	0.0418	0.05058	2	10000	0.24974	938
3	2000	0.09685	846	1540	374	482	0.7325	0.02821	0.02519	0.03236	3	10000	0.24729	938
4	2000	0.09684	790	1584	379	487	0.72938	0.02799	0.01653	0.04303	4	10000	0.2826	940
5	2000	0.10384	805	1549	375	477	0.73563	0.0322	0.03338	0.03554	5	10000	0.24576	938
Average		0.10334	791.4	1562.2	379.2	467.2	0.7355	0.02889	0.03076	0.05143	Average		0.25917	944.6

S2D Classifier

Feature number	Time to 1 TP	TN	FP	FN	Accuracy	Precision	Recall	F1-Score	Feature number	Time to 1 TP	TN	FP	FN	Accuracy	Precision	Recall	F1-Score	
1	2000	0.20802	800	1710	359	411	0.74106	0.03333	0.03027	1	10000	5.04104	1888	868	0.9898	0.9898	0.9898	0.9898
2	2000	0.20807	807	1451	480	362	0.73085	0.03333	0.71474	2	10000	5.34309	1918	868	0.9898	0.9898	0.9898	0.9898
3	2000	0.20806	741	1544	470	400	0.72531	0.03108	0.69196	3	10000	1.29762	1033	848	0.9898	0.9898	0.9898	0.9898
4	2000	0.20806	741	1595	338	526	0.73	0.08675	0.58853	4	10000	1.39008	1048	848	0.9898	0.9898	0.9898	0.9898
5	2000	0.20807	807	1474	450	397	0.73331	0.06615	0.68887	5	10000	1.29917	994	848	0.9898	0.9898	0.9898	0.9898
Average	0.20806	804.8	1534.8	406.6	453.8	0.73115	0.08615	0.67395	0.64488	Average	1.32628	996	848	848	0.9898	0.9898	0.9898	0.9898

Linear SVC

1	2000	0.22002	754	1589	410	417	0.174156	0.05333	0.05007	0.0547	1	10000	0.44158	1060
2	2000	0.24704	854	1502	429	402	0.73033	0.0821	0.08005	0.07447	2	10000	0.47950	1060
3	2000	0.20035	854	1520	394	432	0.74185	0.08429	0.05601	0.07403	3	10000	0.33101	1079
4	2000	0.22907	826	1515	418	439	0.73219	0.08453	0.05351	0.05807	4	10000	0.44719	1071
5	2000	0.22187	852	1510	414	394	0.7475	0.08206	0.09122	0.08585	5	10000	0.43235	1080
Average		0.22427	842.4	1527.2	416.2	416.2	0.7405	0.07018	0.06915	0.0686	Average		0.41235	1080

NaïveBayes (gemmar+idf)

1	2000	291.089	833	1543	462	362	0.7425	0.04334	0.09377	0.08026	1	1000	1.05295	949
2	2000	277.724	907	1448	483	362	0.7354	0.03232	0.7414	0.0321	2	1000	1.64778	914
3	2000	279.055	891	1483	431	395	0.74188	0.07388	0.08025	0.08326	3	1000	1.39797	923
4	2000	280.058	871	1483	450	366	0.7365	0.05938	0.08745	0.07311	4	1000	1.63746	977
5	2000	278.122	924	1483	481	352	0.74084	0.08175	0.72414	0.08448	5	1000	1.42785	1045
Average		281.636	885.2	1484	487.4	373.4	0.74038	0.06925	0.73035	0.08040	Average		1.55277	971.6

SVC (gemmar+idf)

1	2000	298.263	0	2005	0	1195	0.62656	0	0	0	1	10000	59.537	1121
2	2000	298.263	0	1921	0	1269	0.62444	0	0	0	2	10000	63.3021	1145
3	2000	298.056	0	1914	0	1286	0.59813	0	0	0	3	10000	60.8553	1132
4	2000	298.003	0	1923	0	1267	0.60406	0	0	0	4	10000	60.3947	1130
5	2000	298.311	0	1924	0	1276	0.60125	0	0	0	5	10000	62.0795	1173
Average	298.266	0	1941.4	0	1256.6	0.60669	0	0	0	0	Average	61.9797	1140.2	