Decision Tree

Features   Counter with 2-gram and tf-idf
Maximum   25000

criterion gini — max_depth 100

Feature number		Time to train	TP	TN	FP	FN	Accuracy	Precision	Recall	F1-Score
1	39527	429.4031811	1114	1594	381	111	0.84625	0.74515	0.90939	0.81912
2	39527	419.2259033	1080	1590	346	184	0.83438	0.75736	0.85443	0.80297
3	39527	260.6240873	1153	1613	324	110	0.86438	0.78064	0.91291	0.84161
4	39527	93.09207344	1103	1563	388	146	0.83313	0.73977	0.88311	0.80511
5	39527	102.1296356	1084	1555	353	208	0.82469	0.75435	0.83901	0.79443
Average		260.8949761	1106.8	1583	358.4	151.8	0.84056	0.75545	0.87977	0.81265

criterion entropy — max_depth 100

Feature number		Time to train	TP	TN	FP	FN	Accuracy	Precision	Recall	F1-Score
1	39527	91.57152629	1066	1651	298	185	0.84906	0.78152	0.85212	0.8153
2	39527	92.02545428	994	1659	346	201	0.82906	0.74179	0.8318	0.78422
3	39527	88.18855762	1079	1629	305	187	0.84625	0.77962	0.85229	0.81434
4	39527	94.66225863	1064	1597	311	228	0.83156	0.77382	0.82353	0.7979
5	39527	85.87591124	1120	1632	279	169	0.86	0.80057	0.86889	0.83333
Average		90.46474161	1064.6	1633.6	307.8	194	0.84319	0.77547	0.84573	0.80902

criterion gini — max_depth 200

Feature number		Time to train	TP	TN	FP	FN	Accuracy	Precision	Recall	F1-Score
1	39527	137.3900383	1162	1540	378	120	0.84438	0.75455	0.9064	0.82353
2	39527	132.9106324	1134	1547	400	119	0.83781	0.73924	0.90503	0.81378
3	39527	132.3292778	1172	1568	351	109	0.85625	0.76953	0.91491	0.83595
4	39527	114.8673561	1128	1611	356	105	0.85594	0.76011	0.91484	0.83033
5	39527	132.3715451	1119	1591	365	125	0.84688	0.75404	0.89952	0.82038
Average		129.9737699	1143	1571.4	370	115.6	0.84825	0.75549	0.90814	0.82479

criterion entropy — max_depth 200

Feature number		Time to train	TP	TN	FP	FN	Accuracy	Precision	Recall	F1-Score
1	39527	103.6855941	1165	1614	306	115	0.86844	0.79198	0.91016	0.84696
2	39527	106.8156304	1133	1639	324	104	0.86625	0.77763	0.91593	0.84113
3	39527	103.9283895	1110	1654	297	139	0.86375	0.78891	0.88871	0.83584
4	39527	108.5672193	1149	1621	320	110	0.86563	0.78216	0.91263	0.84238
5	39527	104.0673733	1150	1621	311	118	0.86594	0.78713	0.90694	0.8428
Average		105.4128413	1141.4	1629.8	311.6	117.2	0.866	0.78556	0.90687	0.84182

criterion gini — max_depth 400

Feature number		Time to train	TP	TN	FP	FN	Accuracy	Precision	Recall	F1-Score
1	39527	132.86023	1132	1573	364	131	0.84531	0.75668	0.89628	0.82059
2	39527	133.6318045	1125	1598	365	112	0.85094	0.75503	0.90946	0.82508
3	39527	138.7623982	1162	1547	370	121	0.84656	0.75849	0.90569	0.82558
4	39527	160.9456882	1153	1585	357	105	0.85563	0.76358	0.91653	0.83309
5	39527	150.1404922	1136	1578	370	116	0.84813	0.75432	0.90735	0.82379
Average		143.2681226	1141.6	1576.2	365.2	117	0.84931	0.75762	0.90706	0.82562

criterion entropy — max_depth 400

Feature number		Time to train	TP	TN	FP	FN	Accuracy	Precision	Recall	F1-Score
1	39527	107.7346308	1152	1583	342	123	0.85469	0.77108	0.90353	0.83207
2	39527	107.9468927	1112	1669	302	117	0.86906	0.78642	0.9048	0.84147
3	39527	102.2973237	1124	1626	306	144	0.85938	0.78601	0.88644	0.83321
4	39527	109.0716522	1119	1615	326	140	0.85438	0.77439	0.8888	0.82766
5	39527	101.5984204	1133	1614	324	129	0.85844	0.77763	0.89778	0.83339
Average		105.729784	1128	1621.4	320	130.6	0.85919	0.77911	0.89627	0.83356

criterion gini — max_depth 600

Feature number		Time to train	TP	TN	FP	FN	Accuracy	Precision	Recall	F1-Score
1	39527	158.0925345	1122	1600	381	97	0.85063	0.74651	0.92043	0.82439
2	39527	151.3102052	1150	1584	377	89	0.85438	0.75311	0.92817	0.83153
3	39527	157.8437214	1180	1529	369	122	0.84656	0.76178	0.9063	0.82778
4	39527	157.46083	1121	1583	372	124	0.845	0.75084	0.9004	0.81885
5	39527	154.000463	1200	1562	350	88	0.86313	0.77419	0.93168	0.84567
Average		155.7415508	1154.6	1571.6	369.8	104	0.85194	0.75729	0.91739	0.82964

criterion entropy — max_depth 600

Feature number		Time to train	TP	TN	FP	FN	Accuracy	Precision	Recall	F1-Score
1	39527	114.749965	1122	1605	325	148	0.85219	0.7754	0.88346	0.82591
2	39527	105.3174195	1110	1657	317	116	0.86469	0.77786	0.90538	0.83679
3	39527	107.0216668	1161	1595	317	127	0.86125	0.78552	0.9014	0.83948
4	39527	106.5190003	1142	1631	321	106	0.86656	0.78059	0.91506	0.84249
5	39527	110.5286832	1139	1581	358	122	0.85	0.76086	0.90325	0.82596
Average		108.8273469	1134.8	1613.8	327.6	123.8	0.85894	0.77604	0.90171	0.83413

criterion gini — max_depth MAX (~800)

Feature number		Time to train	TP	TN	FP	FN	Accuracy	Precision	Recall	F1-Score
1	39527	153.7395298	1180	1556	346	118	0.855	0.77326	0.90909	0.83569
2	39527	166.029485	1138	1540	393	129	0.83688	0.74331	0.89818	0.81344
3	39527	147.8999882	1119	1617	343	121	0.855	0.76539	0.90242	0.82828
4	39527	187.4034297	1122	1614	346	118	0.855	0.76431	0.90484	0.82866
5	39527	122.3378849	1140	1599	353	108	0.85594	0.76356	0.91346	0.83181
Average		155.4820635	1139.8	1585.2	356.2	118.8	0.85156	0.76197	0.9056	0.82758

criterion entropy — max_depth MAX(~800)

Feature number		Time to train	TP	TN	FP	FN	Accuracy	Precision	Recall	F1-Score
1	39527	108.3459818	1150	1608	307	135	0.86188	0.78929	0.89494	0.8388
2	39527	105.8573973	1132	1628	320	120	0.8625	0.77961	0.90415	0.83728
3	39527	108.5925591	1160	1617	312	111	0.86781	0.78804	0.91267	0.84579
4	39527	107.6706195	1127	1637	316	120	0.86375	0.78101	0.90377	0.83792
5	39527	107.1375256	1135	1605	357	103	0.85625	0.76072	0.9168	0.8315
Average		107.5208167	1140.8	1619	322.4	117.8	0.86244	0.77974	0.90647	0.83826