Final Year B. Tech. (CSE) – I : 2021-22
4CS462 : PE2 - Data Mining Lab
Assignment No. 8

Group id: DM21G12
Group members:
Abhishek More(2018BTECS00037)
Sushil Wagh(2018BTECS00031)

Title : Apriori algorithm

Objective/Aim :

1. Implement the Apriori algorithm for generating Association Rules
2. Experiment with different values of support, confidence, and maximum rule length.
3. Tabulate the results containing frequent item sets, total number of rules generated for different support and confidence.
4. Find the interesting rules from above obtained rules using following metrics/measures
    a. Lift
    b. Chi-Square Test χ
    c. All_confidence measure
    d. Max_confidence measure
    e. Kulczynski measure
    f. Cosine measure

Introduction:

Apriori is an algorithm for frequent item set mining and association rule learning over relational databases. It proceeds by identifying the frequent individual items in the database and extending them to larger and larger item sets as long as those item sets appear sufficiently often in the database. The frequent item sets determined by Apriori can be used to determine association rules which highlight general trends in the database: this has applications in domains such as market basket analysis.

Theory/Algorithms:

Definitions of various terms

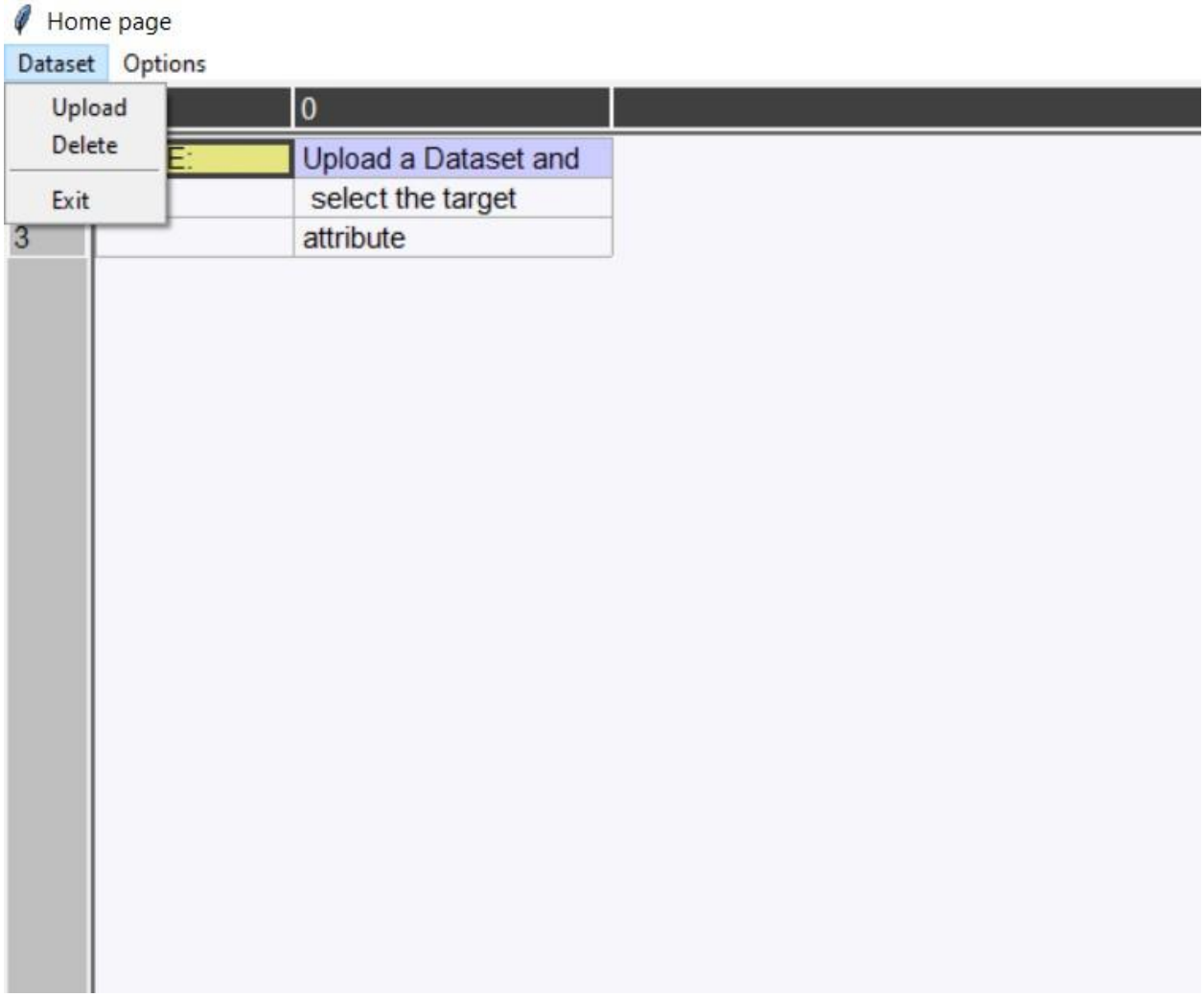
- ***Set of items:*** $I=\{I_1,I_2,\ldots,I_m\}$
- ***Transactions:*** $D=\{t_1,t_2,\ldots,t_n\}$, $t_j \subseteq I$
- ***Itemset:*** $\{I_{i1},I_{i2},\ldots,I_{ik}\} \subseteq I$
- ***Support of an itemset:*** Percentage of transactions which contain that itemset.
- ***Large (Frequent) itemset:*** **Itemset** whose number of occurrences is above a threshold.

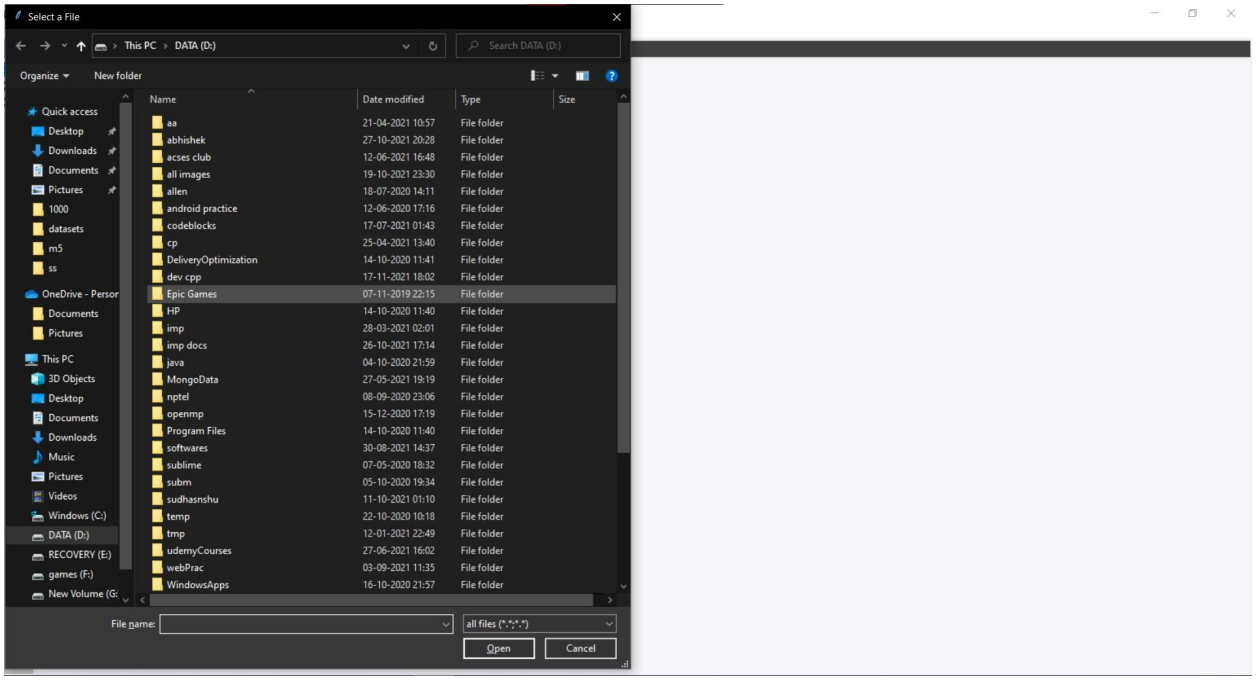
Apriori Algorithm

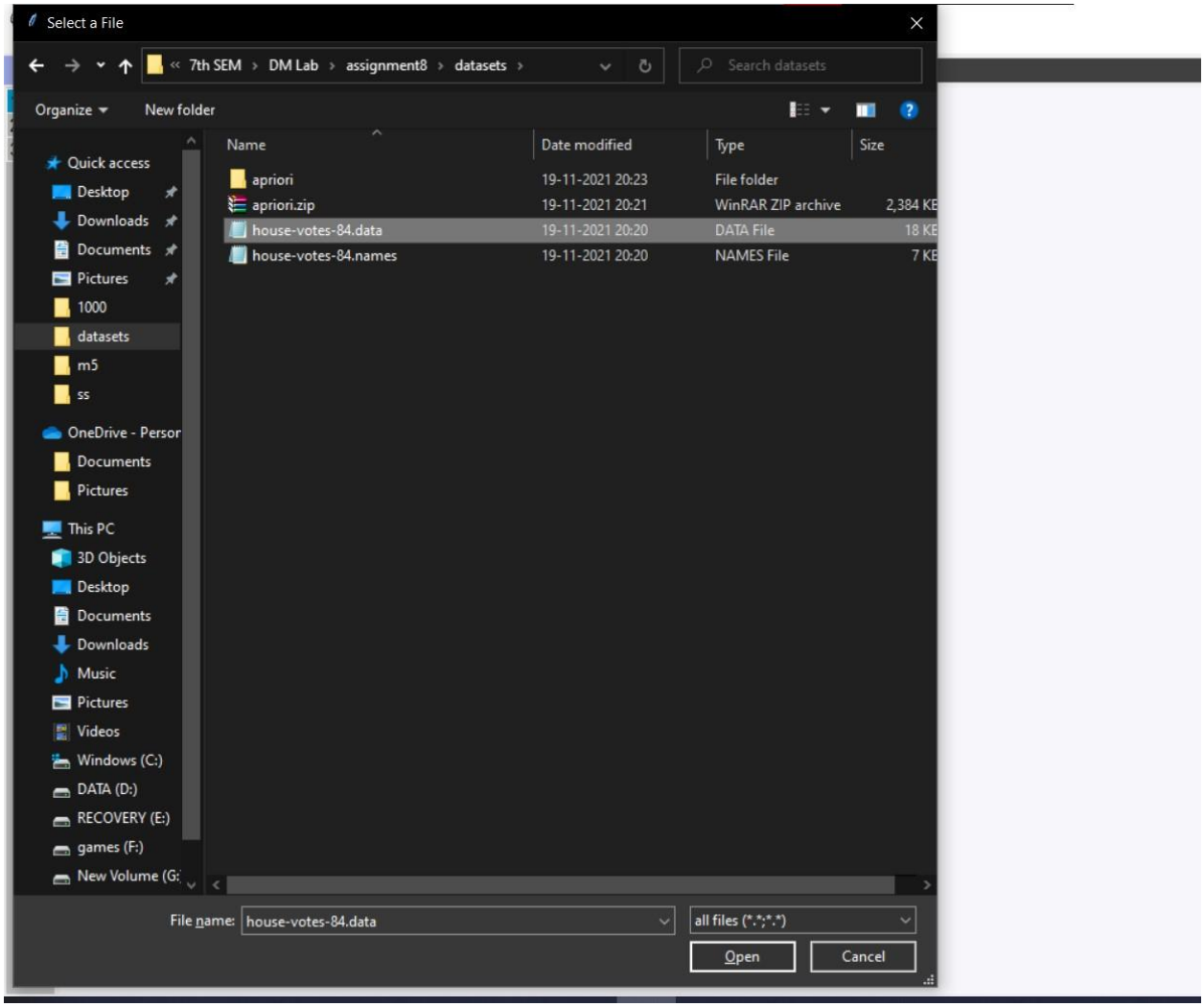
1. C_1 = Itemsets of size one in I;
2. Determine all large itemsets of size 1, L_1;
3. i = 1;
4. Repeat
5. i = i + 1;
6. C_i = Apriori-Gen(L_{i-1});
7. Count C_i to determine L_i;
8. until no more large itemsets found;

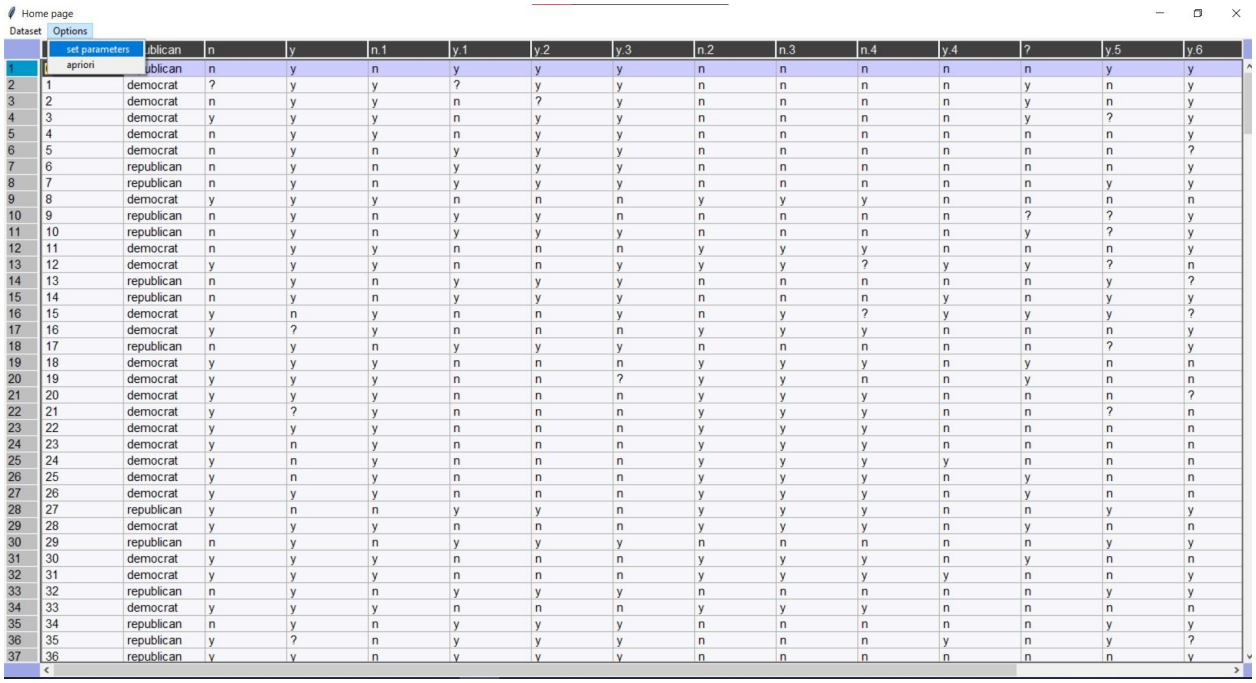
Result/Observations/Screenshots:


Home page
Dataset
Options
0
1
NOTE:
Upload a Dataset and
2
select the target
3
attribute


Home page
Dataset
Options
Upload
Delete
Exit
0
E:
Upload a Dataset and
select the target
attribute
3


Select a File
This PC
DATA (D:)
Search DATA (D:)
Organize
New folder
Quick access
Desktop
Downloads
Documents
Pictures
1000
datasets
m5
ss
OneDrive - Persor
Documents
Pictures
This PC
3D Objects
Desktop
Documents
Downloads
Music
Pictures
Videos
Windows (C:)
DATA (D:)
RECOVERY (E:)
games (F:)
New Volume (G:
Name
Date modified
Type
Size
aa 21-04-2021 10:57 File folder
abhishek 27-10-2021 20:28 File folder
acses club 12-06-2021 16:48 File folder
all images 19-10-2021 23:30 File folder
allen 18-07-2020 14:11 File folder
android practice 12-06-2020 17:16 File folder
codeblocks 17-07-2021 01:43 File folder
cp 25-04-2021 13:40 File folder
DeliveryOptimization 14-10-2020 11:41 File folder
dev cpp 17-11-2021 18:02 File folder
Epic Games 07-11-2019 22:15 File folder
HP 14-10-2020 11:40 File folder
imp 28-03-2021 02:01 File folder
imp docs 26-10-2021 17:14 File folder
java 04-10-2020 21:59 File folder
MongoData 27-05-2021 19:19 File folder
nptel 08-09-2020 23:06 File folder
openmp 15-12-2020 17:19 File folder
Program Files 14-10-2020 11:40 File folder
softwares 30-08-2021 14:37 File folder
sublime 07-05-2020 18:32 File folder
subm 05-10-2020 19:34 File folder
sudhasnshu 11-10-2021 01:10 File folder
temp 22-10-2020 10:18 File folder
tmp 12-01-2021 22:49 File folder
udemyCourses 27-06-2021 16:02 File folder
webPrac 03-09-2021 11:35 File folder
WindowsApps 16-10-2020 21:57 File folder
File name:
all files (*.*;*.*)
Open
Cancel


Select a File
7th SEM › DM Lab › assignment8 › datasets ›
Search datasets
Organize
New folder
Quick access
Desktop
Downloads
Documents
Pictures
1000
datasets
m5
ss
OneDrive - Persor
Documents
Pictures
This PC
3D Objects
Desktop
Documents
Downloads
Music
Pictures
Videos
Windows (C:)
DATA (D:)
RECOVERY (E:)
games (F:)
New Volume (G:
Name
Date modified
Type
Size
apriori
19-11-2021 20:23
File folder
apriori.zip
19-11-2021 20:21
WinRAR ZIP archive
2,384 KE
house-votes-84.data
19-11-2021 20:20
DATA File
18 KE
house-votes-84.names
19-11-2021 20:20
NAMES File
7 KE
File name:
house-votes-84.data
all files (*.*;*.*)
Open
Cancel



Home page

Dataset Options

		republican	n	y	n.1	y.1
1	0	republican	n	y	n	y
2	1	democrat	?	y	y	?
3	2	democrat	n	y	y	n
4	3					y	n
5	4					y	n
6	5					n	y
7	6					n	y
8	7					n	y
9	8					y	n
10	9					n	y
11	10	republican	n	y	n	y
12	11	democrat	n	y	y	n
13	12	democrat	y	y	y	n
14	13	republican	n	y	n	y
15	14	republican	n	y	n	y
16	15	democrat	y	n	y	n

For values of paramt...

Enter Value of min_support 0.2

Enter Value of min_confidence 0.3

Enter Value of max_length 4

Save K

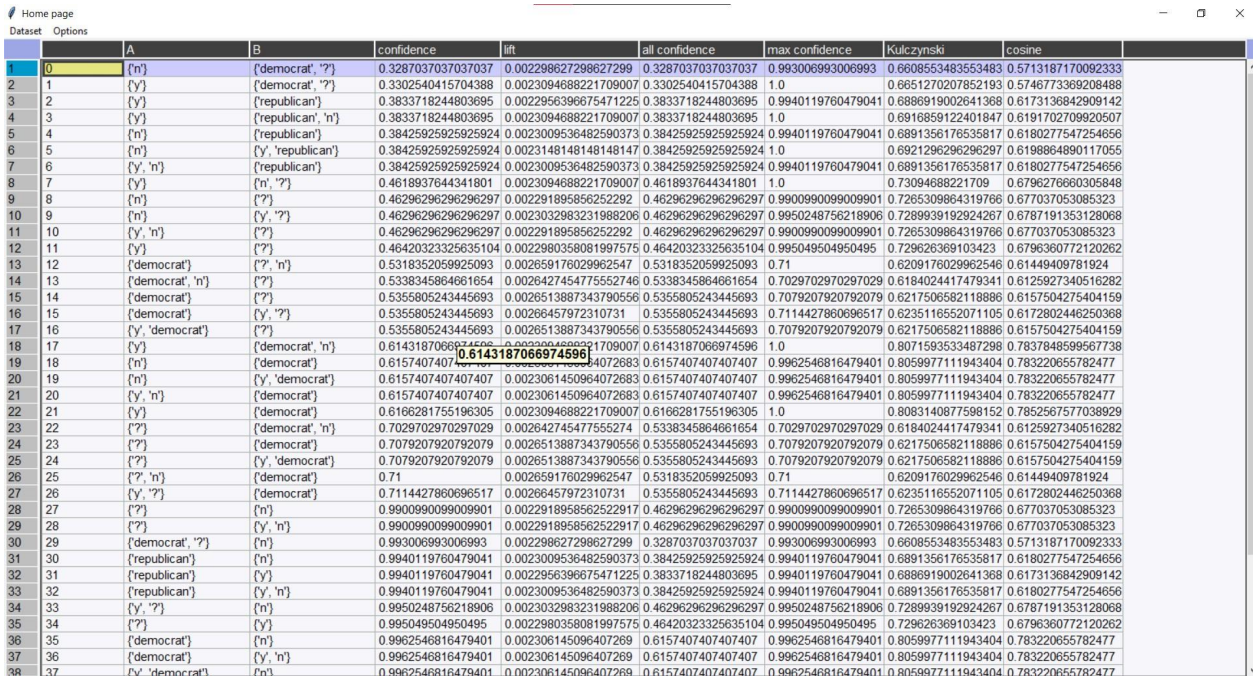
Home page

Dataset Options

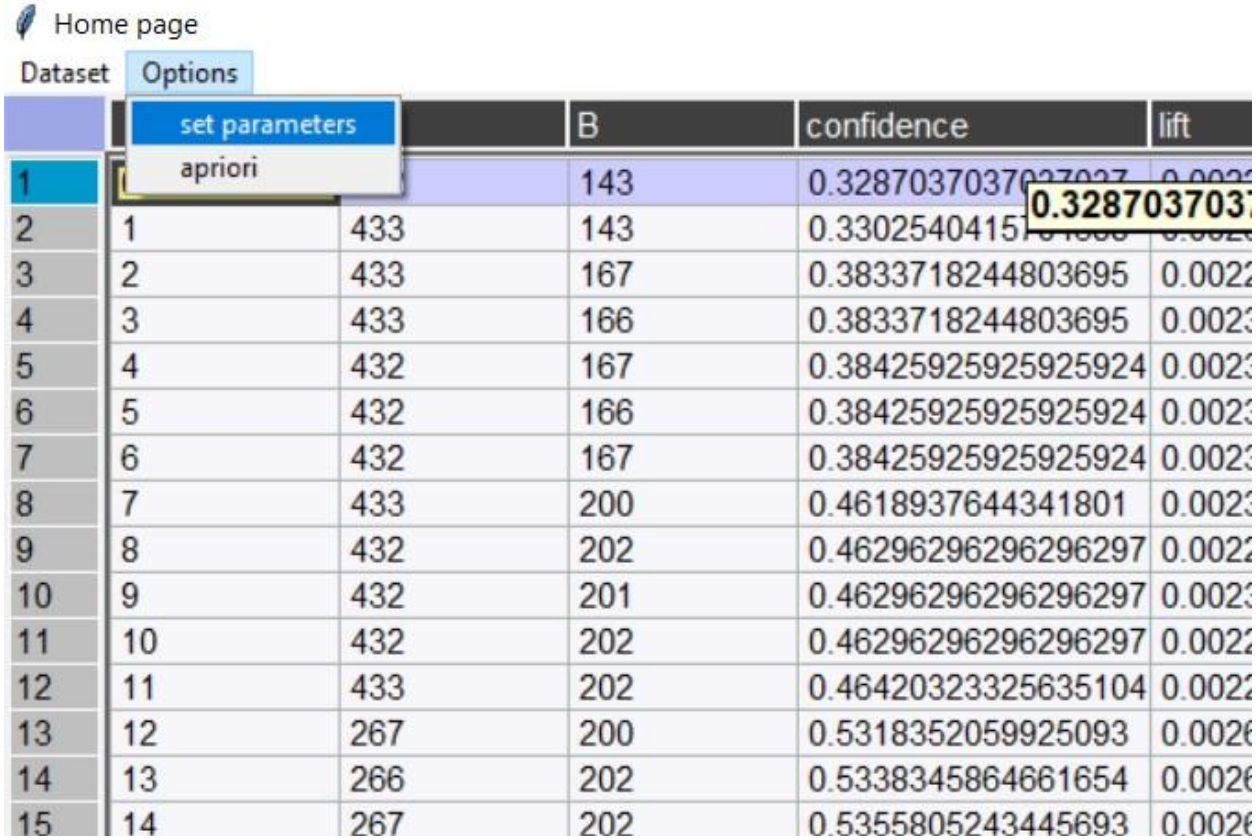
set parameters

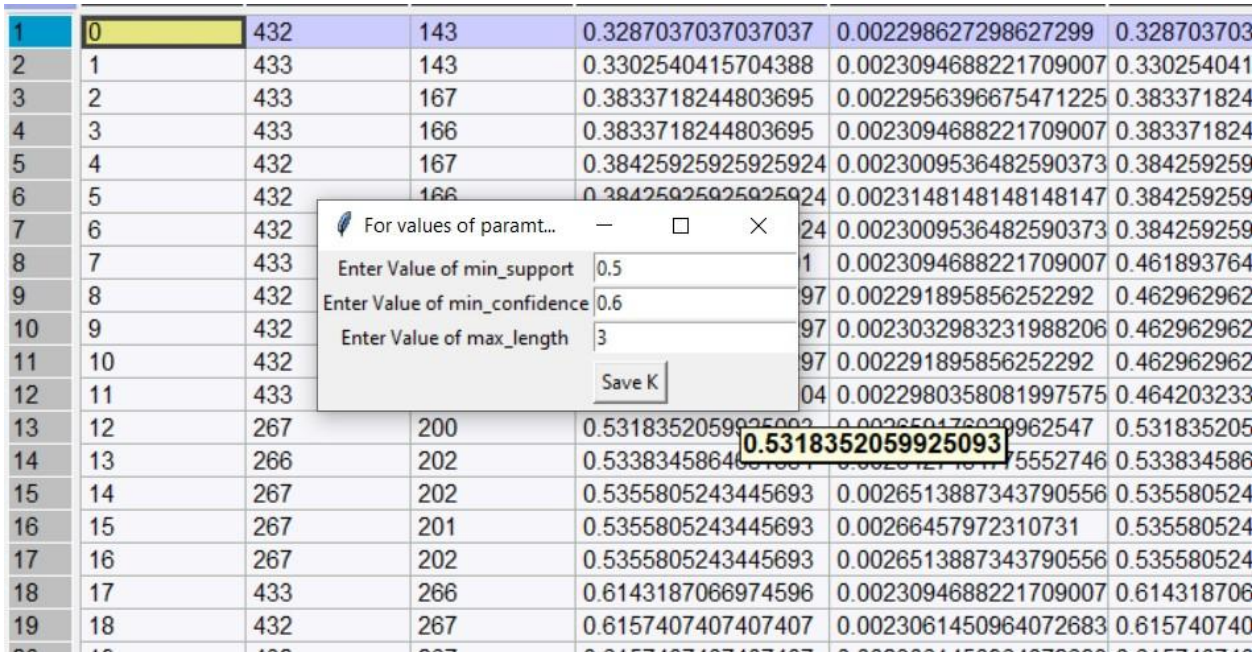
apriori

		ublican	n	y	n.1	y.1
1		ublican	n	y	n	y
2	1	democrat	?	y	y	?
3	2	democrat	n	y	y	n
4	3	democrat	y	y	y	n
5	4	democrat	n	y	y	n
6	5	democrat	n	y	n	y
7	6	republican	n	y	n	y
8	7	republican	n	y	n	y
9	8	democrat	y	y	y	n
10	9	republican	n	y	n	y
11	10	republican	n	y	n	y
12	11	democrat	n	y	y	n
13	12	democrat	y	y	y	n
14	13	republican	n	y	n	y
15	14	republican	n	y	n	y
16	15	democrat	y	n	y	n
17	16	democrat	y	?	y	n
18	17	republican	n	y	n	y
19	18	democrat	y	y	y	n
20	19	democrat	y	y	y	n
21	20	democrat	y	y	y	n
22	21	democrat	y	?	y	n
23	22	democrat	y	y	y	n

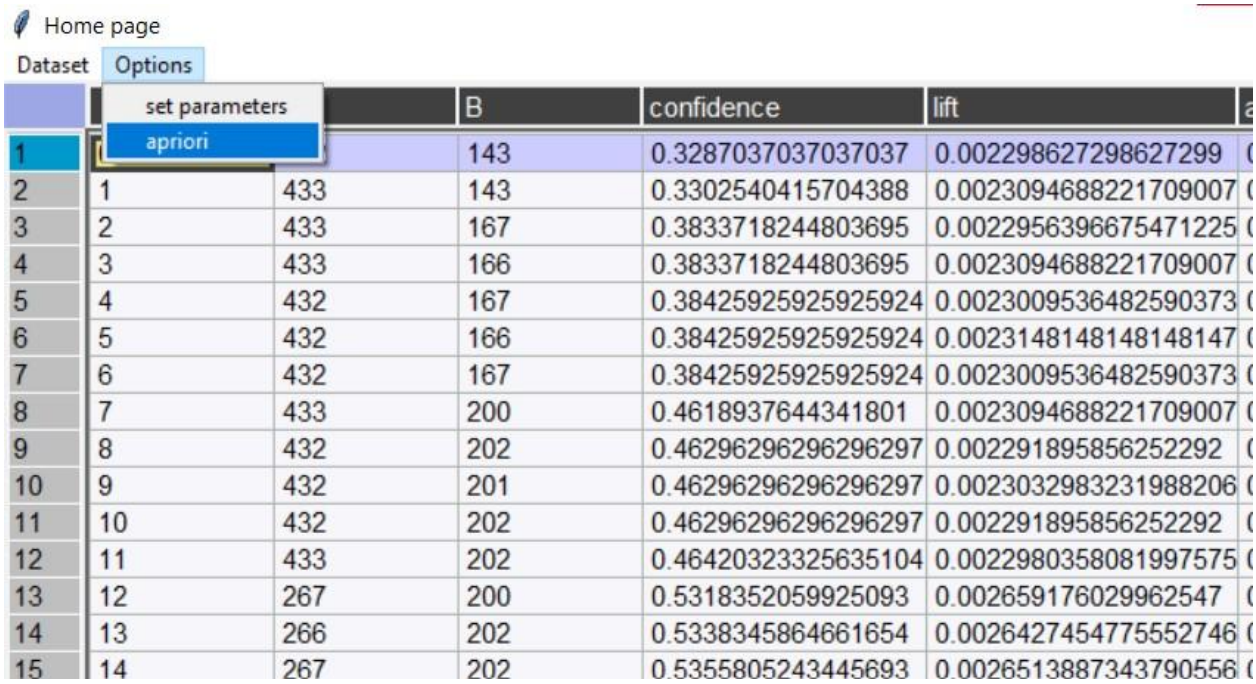


		A	B	confidence	lift	all confidence	max confidence	Kulczynski	cosine
1	0	{'n'}	{'democrat', '?'}	0.3287037037037037	0.002298627298627299	0.3287037037037037	0.993006993006993	0.6608553483553483	0.5713187170092333
2	1	{'y'}	{'democrat', '?'}	0.3302540415704388	0.0023094688221709007	0.3302540415704388	1.0	0.6651270207852193	0.5746773369208488
3	2	{'y'}	{'republican'}	0.3833718244803695	0.0022956396675471225	0.3833718244803695	0.9940119760479041	0.6886919002641368	0.6173136842909142
4	3	{'y'}	{'republican', 'n'}	0.3833718244803695	0.0023094688221709007	0.3833718244803695	1.0	0.6916859122401847	0.6191702709920507
5	4	{'n'}	{'republican'}	0.38425925925925924	0.0023009536482590373	0.38425925925925924	0.9940119760479041	0.6891356176535817	0.6180277547254656
6	5	{'n'}	{'y', 'republican'}	0.38425925925925924	0.0023148148148148147	0.38425925925925924	1.0	0.6921296296296297	0.6198864890117055
7	6	{'y', 'n'}	{'republican'}	0.38425925925925924	0.0023009536482590373	0.38425925925925924	0.9940119760479041	0.6891356176535817	0.6180277547254656
8	7	{'y'}	{'n', '?'}	0.4618937644341801	0.0023094688221709007	0.4618937644341801	1.0	0.73094688221709	0.6796276660305848
9	8	{'n'}	{'?'}	0.46296296296296297	0.002291895856252292	0.46296296296296297	0.9900990099009901	0.7265309864319766	0.677037053085323
10	9	{'n'}	{'y', '?'}	0.46296296296296297	0.0023032983231988206	0.46296296296296297	0.995024875621890	0.7289939192924267	0.6787191353128068
11	10	{'y', 'n'}	{'?'}	0.46296296296296297	0.002291895856252292	0.46296296296296297	0.9900990099009901	0.7265309864319766	0.677037053085323
12	11	{'y'}	{'?'}	0.46420323325635104	0.0022980358081997575	0.46420323325635104	0.995049504950495	0.729626369103423	0.6796360721202262
13	12	{'democrat'}	{'?', 'n'}	0.5318352059925093	0.002659176029962547	0.5318352059925093	0.71	0.6209176029962546	0.61449409781924
14	13	{'democrat', 'n'}	{'?'}	0.5338345864661654	0.0026427454775552746	0.5338345864661654	0.7029702970297029	0.6184024417479341	0.6125927340516282
15	14	{'democrat'}	{'?'}	0.5355805243445693	0.0026513887343790556	0.5355805243445693	0.7079207920792079	0.6217506582118886	0.6157504275404159
16	15	{'democrat'}	{'y', '?'}	0.5355805243445693	0.0026645797231073 1	0.5355805243445693	0.7114427860696517	0.6235116552071105	0.6172802446250368
17	16	{'y', 'democrat'}	{'?'}	0.5355805243445693	0.0026513887343790556	0.5355805243445693	0.7079207920792079	0.6217506582118886	0.6157504275404159
18	17	{'y'}	{'democrat', 'n'}	0.6143187066974596	[illegible]	0.6143187066974596	1.0	0.8071593533487298	0.7837848599567738
19	18	{'n'}	{'democrat'}	0.6157407407407407	[illegible]	0.615740740740 7407	0.9962546816479401	0.8059977111943404	0.783220655782477
20	19	{'n'}	{'y', 'democrat'}	0.6157407407407407	0.0023061450964072683	0.6157407407407407	0.9962546816479401	0.8059977111943404	0.783220655782477
21	20	{'y', 'n'}	{'democrat'}	0.6157407407407407	0.0023061450964072683	0.6157407407407407	0.9962546816479401	0.8059977111943404	0.783220655782477
22	21	{'y'}	{'democrat'}	0.6166281755196305	0.0023094688221709007	0.6166281755196305	1.0	0.8083140877598152	0.7852567577038929
23	22	{'?'}	{'democrat', 'n'}	0.7029702970297029	0.002642745775555274	0.5338345864661654	0.7029702970297029	0.6184024417479341	0.6125927340516282
24	23	{'?'}	{'democrat'}	0.7079207920792079	0.0026513887343790556	0.5355805243445693	0.7079207920792079	0.6217506582118886	0.6157504275404159
25	24	{'?'}	{'y', 'democrat'}	0.7079207920792079	0.0026513887343790556	0.5355805243445693	0.7079207920792079	0.6217506582118886	0.6157504275404159
26	25	{'?', 'n'}	{'democrat'}	0.71	0.002659176029962547	0.5318352059925093	0.71	0.6209176029962546	0.61449409781924
27	26	{'y', '?'}	{'democrat'}	0.7114427860696517	0.0026645797231073 1	0.5355805243445693	0.7114427860696517	0.6235116552071105	0.6172802446250368
28	27	{'?'}	{'n'}	0.9900990099009901	0.002291895856252291 7	0.4629629629629297	0.9900990099009901	0.7265309864319766	0.677037053085323
29	28	{'?'}	{'y', 'n'}	0.9900990099009901	0.002291895856252291 7	0.4629629629629297	0.9900990099009901	0.7265309864319766	0.677037053085323
30	29	{'democrat', '?'}	{'n'}	0.99306993006993	0.002298627298627299	0.3287037037037037	0.993006993006993	0.6608553483553483	0.5713187170092333
31	30	{'republican'}	{'n'}	0.9940119760479041	0.0023009536482590373	0.38425925925925924	0.9940119760479041	0.6891356176535817	0.6180277547254656
32	31	{'republican'}	{'y'}	0.9940119760479041	0.0022956396675471225	0.3833718244803695	0.9940119760479041	0.6886919002641368	0.6173136842909142
33	32	{'republican'}	{'y', 'n'}	0.9940119760479041	0.0023009536482590373	0.38425925925925924	0.9940119760479041	0.6891356176535817	0.6180277547254656
34	33	{'y', '?'}	{'n'}	0.995024875621890	0.0023032983231988206	0.4629629629629297	0.995024875621890	0.7289939192924267	0.6787191353128068
35	34	{'?'}	{'y'}	0.995049504950495	0.0022980358081997575	0.46420323325635104	0.995049504950495	0.729626369103423	0.6796360721202262
36	35	{'democrat'}	{'n'}	0.9962546816479401	0.002306145096407269	0.6157407407407407	0.9962546816479401	0.8059977111943404	0.783220655782477
37	36	{'democrat'}	{'y', 'n'}	0.9962546816479401	0.002306145096407269	0.6157407407407407	0.9962546816479401	0.8059977111943404	0.783220655782477
38	37	{'y', 'democrat'}	{'n'}	0.9962546816479401	0.002306145096407269	0.6157407407407407	0.9962546816479401	0.8059977111943404	0.783220655782477

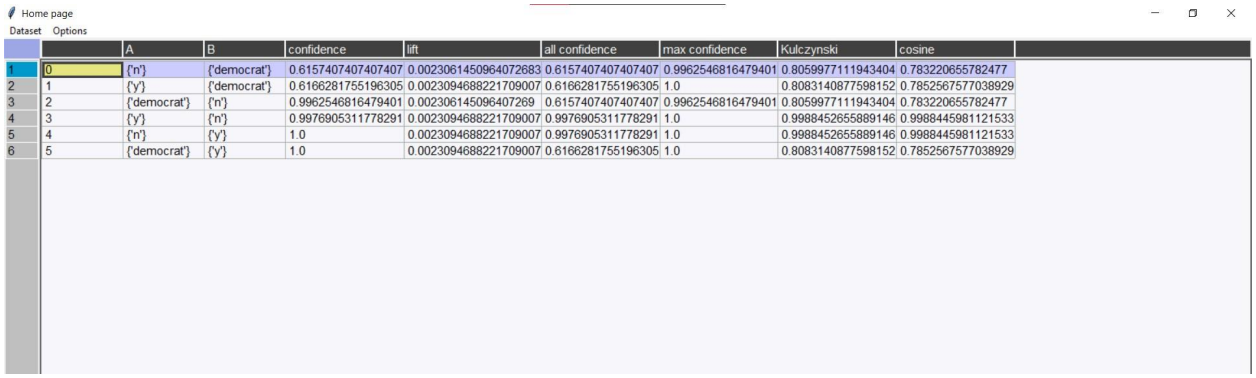




1	0	432	143	0.3287037037037037	0.002298627298627299	0.328703703
2	1	433	143	0.3302540415704388	0.0023094688221709007	0.330254041
3	2	433	167	0.3833718244803695	0.0022956396675471225	0.383371824
4	3	433	166	0.3833718244803695	0.0023094688221709007	0.383371824
5	4	432	167	0.38425925925925924	0.0023009536482590373	0.384259259
6	5	432	166	0.38425925925925924	0.0023148148148148147	0.384259259
7	6	432	167	0.38425925925925924	0.0023009536482590373	0.384259259
8	7	433	200	0.4618937644341801	0.0023094688221709007	0.461893764
9	8	432	202	0.4629629629629629	0.002291895856252292	0.462962962
10	9	432	201	0.4629629629629629	0.0023032983231988206	0.462962962
11	10	432	202	0.4629629629629629	0.002291895856252292	0.462962962
12	11	433	202	0.46420323325635104	0.0022980358081997575	0.464203233
13	12	267	200	0.5318352059925093	0.002659176029962547	0.531835205
14	13	266	202	0.5338345864661654	0.0026427454775552746	0.533834586
15	14	267	202	0.5355805243445693	0.0026513887343790556	0.535580524
16	15	267	201	0.5355805243445693	0.00266457972310731	0.535580524
17	16	267	202	0.5355805243445693	0.0026513887343790556	0.535580524
18	17	433	266	0.6143187066974596	0.0023094688221709007	0.614318706
19	18	432	267	0.6157407407407407	0.0023061450964072683	0.615740740



			B	confidence	lift	a
1	0	432	143	0.3287037037037037	0.002298627298627299	0
2	1	433	143	0.3302540415704388	0.0023094688221709007	0
3	2	433	167	0.3833718244803695	0.0022956396675471225	0
4	3	433	166	0.3833718244803695	0.0023094688221709007	0
5	4	432	167	0.38425925925925924	0.0023009536482590373	0
6	5	432	166	0.38425925925925924	0.0023148148148148147	0
7	6	432	167	0.38425925925925924	0.0023009536482590373	0
8	7	433	200	0.4618937644341801	0.0023094688221709007	0
9	8	432	202	0.4629629629629629	0.002291895856252292	0
10	9	432	201	0.4629629629629629	0.0023032983231988206	0
11	10	432	202	0.4629629629629629	0.002291895856252292	0
12	11	433	202	0.46420323325635104	0.0022980358081997575	0
13	12	267	200	0.5318352059925093	0.002659176029962547	0
14	13	266	202	0.5338345864661654	0.0026427454775552746	0
15	14	267	202	0.5355805243445693	0.0026513887343790556	0



		A	B	confidence	lift	all confidence	max confidence	Kulczynski	cosine
1	0	{'n'}	{'democrat'}	0.6157407407407407	0.00230614509640726 83	0.6157407407407407	0.9962546816479401	0.8059977111943404	0.783220655782477
2	1	{'y'}	{'democrat'}	0.6166281755196305	0.00230946882217090 07	0.6166281755196305	1.0	0.8083140877598152	0.7852567577038929
3	2	{'democrat'}	{'n'}	0.9962546816479401	0.00230614509640726 9	0.6157407407407407	0.9962546816479401	0.8059977111943404	0.783220655782477
4	3	{'y'}	{'n'}	0.9976905311778291	0.00230946882217090 07	0.9976905311778291	1.0	0.9988452655889146	0.9988445981121533
5	4	{'n'}	{'y'}	1.0	0.00230946882217090 07	0.9976905311778291	1.0	0.9988452655889146	0.9988445981121533
6	5	{'democrat'}	{'y'}	1.0	0.00230946882217090 07	0.6166281755196305	1.0	0.8083140877598152	0.7852567577038929

Conclusion:

1. Implemented the Apriori algorithm for generating Association Rules
2. Experimented with different values of support, confidence, and maximum rule length.
3. Tabulated the results containing frequent item sets, total number of rules generated for different support and confidence.
4. Found the interesting rules from above obtained rules using following metrics/measures
    a. Lift
    b. Chi-Square Test χ
    c. All_confidence measure
    d. Max_confidence measure
    e. Kulczynski measure
    f. Cosine measure