Association Rules Mining

Rajesh K. Jat 2021kpad1001@iiitkota.ac.in

Statistics for Data Science (Winter, 2021-22)

Dataset:-

An individual's annual income results from various factors.

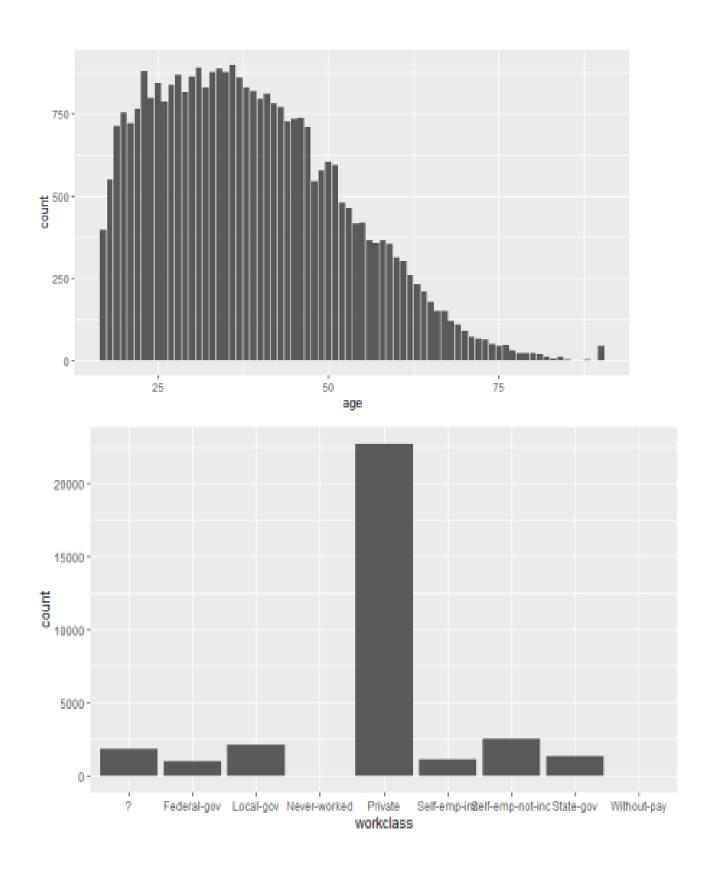
Income is influenced by the individual's education level, age, gender, occupation, etc.

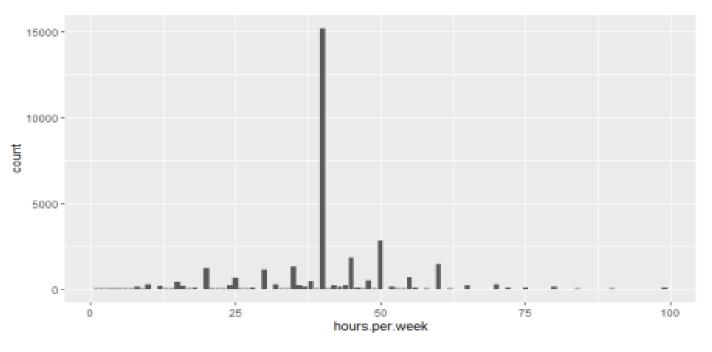
Fields: The dataset contains 16 columns.

Target Field: Income

The income is divided into two classes: <=50K and >50K.

Data Visualization:-





Data Discretization:

Here we discretized 4 columns: age, capital.gain, capital.loss, hours.per.week

Terms Used:

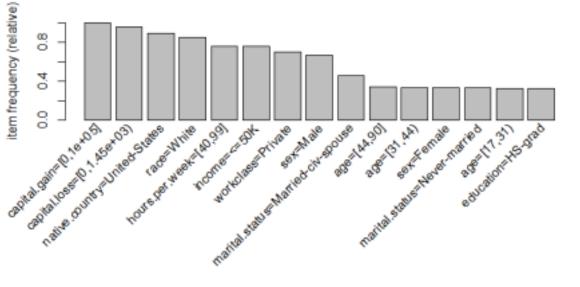
 $\textbf{Lift} \rightarrow \! \textbf{Probability of togetherness}$

Gini → Info Provide

 $\textbf{Phi} \rightarrow \text{Binary Correlation}$

 $\textbf{Support} \rightarrow \text{frequency of items occurred}.$

Confidence → How Reliable the rue is.



Maximal frequent itemset:

Maximal frequent itemset shows that none of its immediate supersets are frequent.

Here, the total no. of maximal frequent itemsets is 8. Some of them are:

Items	Support	Count
{race=White,	0.5709	18592
capital.gain=[0,1e+05], capital.loss=[0,1.45e+03), hours.per.week=[40,99], native.country=United States}		

{race=White,	0.5645	18382		
capital.gain=[0,1e+05], capital.loss=[0,1.45e+03), native.country=United States, income=<=50K}				
{workclass=Private,	0.5307	17281		
capital.gain=[0,1e+05], capital.loss=[0,1.45e+03), income=<=50K}				
{capital.gain=[0,1e+05],	0.5279	17192		

capital.loss=[0,1.45e+03), hours.per.week=[40,99], income=<=50K}		
--	--	--

Closest frequent itemset:

Closet frequent itemset refers that none of it's immediate superset has same support.

Here are total 48 itemsets of such type, some of them are:

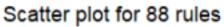
Item	Support	Count
{capital.gain=[0,1e+05]}	1.000	32561
{capital.gain=[0,1e+05], capital.loss=[0,1.45e+03)}	0.9559	31127

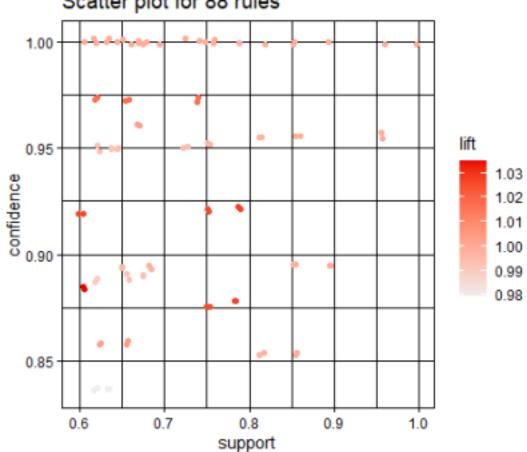
Association rule:

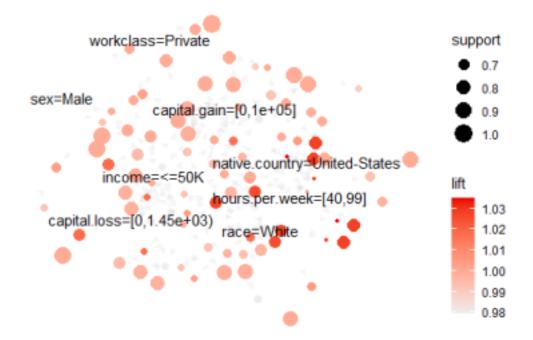
support = 0.6

confidence = 0.8

Total number of association rules is 88.

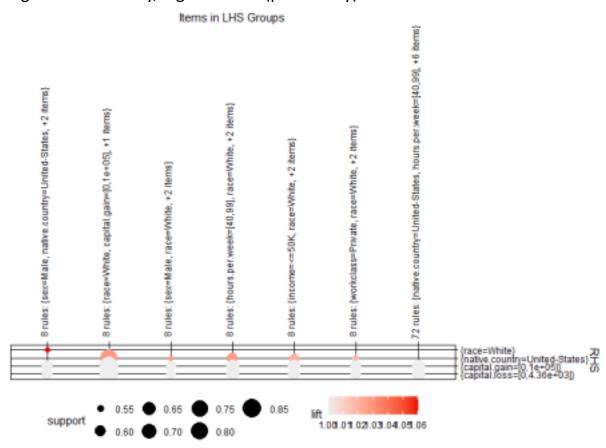






Bigger the circle, higher the support.

Higher the intensity, higher the lift(probability).



This										
native.country=United-States	ı		1hs		rhs	support	confidence	coverage	lift	count
[2] {capital.gain=[0,1e+05], hours.per.week=[40,99], native.country=United-States} => {race=white} = 0.6021314		[1]								
hours.per.week=[40,99], native.country=united-states} => {race=white} => {native.country=united-states} 0.7868616 0.9210886 0.8542735 1.028165 25621 [4] {race=white} => {native.country=united-states} 0.7868616 0.9210886 0.8542735 1.028165 25621 [4] {race=white} => {native.country=united-states} 0.7868616 0.9210886 0.8542735 1.028165 25621 [5] {native.country=united-states} => {native.country=united-states} 0.7868616 0.8783339 0.8958570 1.028165 25621 [6] {capital.gain=[0,1e+05], native.country=united-states} => {native.country=united-states} 0.7868616 0.8783339 0.8958570 1.028165 25621 [7] {race=white, native.country=united-states} => {native.country=united-states} 0.6021314 0.9203399 0.6542489 1.027329 19606 [8] {race=white, capital.gain=[0,1e+05], hours.per.week=[40,99]} => {native.country=united-states} 0.6021314 0.9203399 0.6542489 1.027329 19606 [9] {race=white, capital.gain=[0,1e+05], capital.loss=[0,1.45e+03]} => {native.country=united-states} 0.7499770 0.9198780 0.8153005 1.026813 24420 [10] {race=white, capital.loss=[0,1.45e+03]} => {native.country=united-states} 0.7499770 0.9198780 0.8153005 1.026813 24420 [11] {capital.loss=[0,1.45e+03], native.country=united-states} => {race=white} 0.7499770 0.8762739 0.8558705 1.025753 24420 [12] {capital.gain=[0,1e+05], capital.loss=[0,1.45e+03], native.country=united-states} => {capital.loss=[0,1.45e+03]} 0.7388287 0.9731796 0.7591904 1.018013 24057 [13] {native.country=united-states} => {capital.loss=[0,1.45e+03]} 0.7388287 0.9731796 0.7591904 1.018013 24057 [15] {native.country=united-states} => {capital.loss=[0,1.45e+03]} 0.7388287 0.9731796 0.7591904 1.018013 24057 [15] {native.country=united-states} => {capital.loss=[0,1.45e+03]} 0.7388287 0.9731796 0.7591904 1.018013 24057 [15] {native.country=united-states} => {capital.loss=[0,1.45e+03]} 0.7388287 0.9731796 0.7591904 1.018013 24057 [15] {native.country=united-states} => {capital.loss=[0,1.45e+03]} 0.7388287 0.9731796 0.7591904 1.018013 24057 [15] {native.country=		F-3.3		=>	{race=White}	0.6021314	0.8841887	0.6809987	1.035018	19606
native.country=United-States} => {race=White} => 0.6021314 0.8841887 0.6809987 1.035018 19606 [3] {race=White} => {native.country=United-States} 0.7868616 0.9210886 0.8542735 1.028165 25621 [4] {race=White, capital.gain=[0,1e+05]} => {native.country=United-States} 0.7868616 0.9210886 0.8542735 1.028165 25621 [5] {native.country=United-States} => {race=White} 0.7868616 0.8783339 0.8958570 1.028165 25621 [6] {capital.gain=[0,1e+05], native.country=United-States} => {native.country=United-States} 0.7868616 0.8783339 0.8958570 1.028165 25621 [7] {race=White, capital.gain=[0,1e+05], hours.per.week=[40,99]} => {native.country=United-States} 0.6021314 0.9203399 0.6542489 1.027329 19606 [8] {race=White, capital.loss=[0,1.45e+03)} => {native.country=United-States} 0.6021314 0.9203399 0.6542489 1.027329 19606 [9] {race=White, capital.loss=[0,1.45e+03)} => {native.country=United-States} 0.7499770 0.9198780 0.8153005 1.026813 24420 [10] {race=White, capital.loss=[0,1.45e+03]} => {native.country=United-States} 0.7499770 0.9198780 0.8153005 1.026813 24420 [11] {capital.loss=[0,1.45e+03], native.country=United-States} => {race=White} 0.7499770 0.8762739 0.8558705 1.025753 24420 [12] {capital.gain=[0,1e+05], capital.loss=[0,1.45e+03]} => {race=White} 0.7499770 0.8762739 0.8558705 1.025753 24420 [13] {income=<50K} => {capital.loss=[0,1.45e+03]} 0.7388287 0.9731796 0.7591904 1.018013 24057 [15] {native.country=United-States, experimental plane pl		[2]								
[3] {race=white}					[raca-White]	0.6021214	A 00/1007	0.6900097	1 035010	10606
[4] {race=white, capital.gain=[0.1e+05]}		[3]								
<pre>capital.gain=[0,1e+05]} => {native.country=United-States} 0.7868616 0.9210886 0.8542735 1.028165 25621 [5] {native.country=United-States} => {race=white}</pre>				-,-	(nacrie country-onreco-scaces)	0.7000020	0.3220000	0.0342/33	1.010103	27722
[5] {native.country=United-States} => {race=white} 0.7868616 0.8783339 0.8958570 1.028165 25621 [6] {capital.gain=[0,1e+05], native.country=United-States} => {race=white} 0.7868616 0.8783339 0.8958570 1.028165 25621 [7] {race=white, hours.per.week=[40,99]} => {native.country=United-States} 0.6021314 0.9203399 0.6542489 1.027329 19606 [8] {race=white, capital.gain=[0,1e+05], hours.per.week=[40,99]} => {native.country=United-States} 0.6021314 0.9203399 0.6542489 1.027329 19606 [9] {race=white, capital.loss=[0,1.45e+03)} => {native.country=United-States} 0.7499770 0.9198780 0.8153005 1.026813 24420 [10] {race=white, capital.gain=[0,1e+05], capital.loss=[0,1.45e+03)} => {native.country=United-States} 0.7499770 0.9198780 0.8153005 1.026813 24420 [11] {capital.loss=[0,1.45e+03]} => {native.country=United-States} 0.7499770 0.8762739 0.8558705 1.025753 24420 [12] {capital.gain=[0,1e+05], capital.loss=[0,1.45e+03], native.country=United-States} => {race=white} 0.7499770 0.8762739 0.8558705 1.025753 24420 [13] {income=<=50K} => {race=white} 0.7499770 0.8762739 0.8558705 1.025753 24420 [13] {income=<=50K} => {capital.loss=[0,1.45e+03)} 0.7388287 0.9731796 0.7591904 1.018013 24057 [15] {native.country=United-States}, => {capital.loss=[0,1.45e+03)} 0.7388287 0.9731796 0.7591904 1.018013 24057 [15] {native.country=United-States}, => {capital.loss=[0,1.45e+03)} 0.7388287 0.9731796 0.7591904 1.018013 24057 [15] {native.country=United-States}, => {capital.loss=[0,1.45e+03)} 0.7388287 0.9731796 0.7591904 1.018013 24057 [15] {native.country=United-States}, => {capital.loss=[0,1.45e+03)} 0.7388287 0.9731796 0.7591904 1.018013 24057 [15] {native.country=United-States}, => {capital.loss=[0,1.45e+03)} 0.7388287 0.9731796 0.7591904 1.018013 24057 [15] {native.country=United-States}, => {capital.loss=[0,1.45e+03)} 0.7388287 0.9731796 0.7591904 1.018013 24057 [15] {native.country=United-States}, => {capital.loss=[0,1.45e+03]} 0.7388287 0.9731796 0.7591904 1.018013 24057 [15] {native.country=United-States}, => {capital.loss=[0,1.45e+03]}		1.7		=>	{native.country=United-States}	0.7868616	0.9210886	0.8542735	1.028165	25621
[6] {capital.gain=[0,1e+05], native.country=United-States} => {race=white}		[5]								
[7] {race=white, hours.per.week=[40,99]}			{capital.gain=[0,1e+05],							
hours.per.week=[40,99]}				*>	{race=white}	0.7868616	0.8783339	0.8958570	1.028165	25621
[8] {race=white, capital.gain=[0,1e+05], hours.per.week=[40,99]} => {native.country=United-States} 0.6021314 0.9203399 0.6542489 1.027329 19606 [9] {race=white, capital.loss=[0,1.45e+03)} => {native.country=United-States} 0.7499770 0.9198780 0.8153005 1.026813 24420 [10] {race=white, capital.gain=[0,1e+05], capital.loss=[0,1.45e+03)} => {native.country=United-States} 0.7499770 0.9198780 0.8153005 1.026813 24420 [11] {capital.loss=[0,1.45e+03), native.country=United-States} => {race=white} 0.7499770 0.8762739 0.8558705 1.025753 24420 [12] {capital.gain=[0,1e+05], capital.loss=[0,1.45e+03), native.country=United-States} => {race=white} 0.7499770 0.8762739 0.8558705 1.025753 24420 [13] {income=<=50K} => {race=white} 0.7499770 0.8762739 0.8558705 1.025753 24420 [13] {income=<=50K} => {capital.loss=[0,1.45e+03)} 0.7388287 0.9731796 0.7591904 1.018013 24057 [14] {capital.gain=[0,1e+05], income=<=50K} => {capital.loss=[0,1.45e+03)} 0.7388287 0.9731796 0.7591904 1.018013 24057 [15] {native.country=United-States},		[7]								
<pre>capital.gain=[0,1e+05], hours.per.week=[40,99]}</pre>		Fall		=>	{native.country=United-States}	0.6021314	0.9203399	0.6542489	1.027329	19606
hours.per.week=[40,99]}		[8]								
[9] {race=white, capital.loss=[0,1.45e+03)}					(native country-United-States)	0 6021314	U 05U5500	0.6542489	1 027320	10606
<pre>capital.loss=[0,1.45e+03)}</pre>		[9]		-,-	(macrie: country-offices-states)	0.0022324	0.3603333	0.0342403	2.02.323	23000
<pre>[10] {race=white,</pre>		1-1		=>	{native.country=United-States}	0.7499770	0.9198780	0.8153005	1.026813	24420
<pre>capital.loss=[0,1.45e+03)}</pre>		[10]			,,					
<pre>[11] {capital.loss=[0,1.45e+03),</pre>										
native.country=United-States} => {race=White}				=>	{native.country=United-States}	0.7499770	0.9198780	0.8153005	1.026813	24420
[12] {capital.gain=[0,1e+05], capital.loss=[0,1.45e+03), native.country=United-States} => {race=white} 0.7499770 0.8762739 0.8558705 1.025753 24420 [13] {income=<=50K} => {capital.loss=[0,1.45e+03)} 0.7388287 0.9731796 0.7591904 1.018013 24057 [14] {capital.gain=[0,1e+05], income=<=50K} => {capital.loss=[0,1.45e+03)} 0.7388287 0.9731796 0.7591904 1.018013 24057 [15] {native.country=United-States,		[11]								
<pre>capital.loss=[0,1.45e+03), native.country=United-States} => {race=white}</pre>		F4 3 3		=>	{race=White}	0.7499770	0.8762739	0.8558705	1.025753	24420
native.country=United-States} => {race=white} 0.7499770 0.8762739 0.8558705 1.025753 24420 [13] {income=<=50K} => {capital.loss=[0,1.45e+03)} 0.7388287 0.9731796 0.7591904 1.018013 24057 [14] {capital.gain=[0,1e+05], income=<=50K} => {capital.loss=[0,1.45e+03)} 0.7388287 0.9731796 0.7591904 1.018013 24057 [15] {native.country=United-States,		[12]	[capital.gain=[0,1e+05],							
[13] {income=<=50K} => {capital.loss=[0,1.45e+03)} 0.7388287 0.9731796 0.7591904 1.018013 24057 [14] {capital.gain=[0,1e+05], income=<=50K} => {capital.loss=[0,1.45e+03)} 0.7388287 0.9731796 0.7591904 1.018013 24057 [15] {native.country=United-States,					(race-white)	0.7499770	0.8762730	0.8558705	1.025753	24420
<pre>[14] {capital.gain=[0,1e+05],</pre>		[13]								
income=<=50K} => {capital.loss=[0,1.45e+03)} 0.7388287 0.9731796 0.7591904 1.018013 24057 [15] {native.country=United-States,					(***********	********		
				=>	{capital.loss=[0,1.45e+03)}	0.7388287	0.9731796	0.7591904	1.018013	24057
income=<=50K} => {capital.loss=[0,1.45e+03)} 0.6572280 0.9727715 0.6756242 1.017586 21400		[15]	{native.country=United-States,							
			income=<=50K}	*>	{capital.loss=[0,1.45e+03)}	0.6572280	0.9727715	0.6756242	1.017586	21400