Submission sh Assignment II	eet	t				
			-		bout the operations applied to I submit it on Studium.	
TASK 1: warm u	p					
Number of record	s	9816 (after removing dupl	icates)			
Number of attribu	tes	(excluding User and	Query)	500		
TASK 2: frequen	t it	emsets				
What is the most t	frec	quent item (i.e., single	keywor	d)? ["c	of"	
In how many quer	ies	does it occur: 955 ->	without rem	oving dı	uplicates (927 with removing duplicates)	
Which support val	lue	corresponds to a supp	ort coun	t of 10	$00 \text{ records}? \boxed{100/98} = 0.0101874491$	
How many freque	nt i	temsets have you four	nd with th	nis mi	n-support? 28	
What is the maxin	nun	n size of your frequer	nt itemse	ts? 2	2	
TASK 3: impact	of 1	the support paramet	ter			
Number of freque	nt i	temsets for some sele	ected valu	ues of	support:	
Mininum support]	Number of Frequent I	temsets			
0.001		855				
0.004		140				
0.007		55				
0.01		27				
			iscovery	of ab	out 150 itemsets? 0.004 -> leads to 140 ite	mset
TASK 4: rule ger	ier	ation				
How many rules h	ave	e you identified with	min-conf	fidenc	ee .8? 4	
Indicate a high-co	nfi	dence rule X -> Y wh	ere Y ->	X has	s lower confidence.	
		York		\rightarrow	new	

Why is the confidence of $Y \rightarrow X$ lower than the confidence of $X \rightarrow Y$?

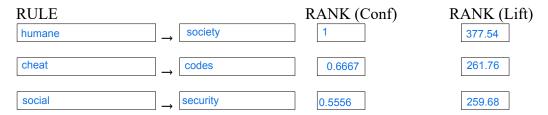
TASK 5: impact of confidence

Number of rules for some selected values of minimum confidence:

Mininum confidence	Number of Rules			
0.1	18			
0.3	14			
0.5	7			
0.7	5			
0.9	1			

TASK 6: rule interpretation

What are the top-3 interesting rules, and what is their rank w.r.t. confidence and lift?



What are the top-3 unexpected rules, and what is their rank w.r.t. confidence and lift?

