UNIVERSITY OF MOHAMED BOUDIAF – M'SILA
FACULTY OF MATHEMATICS AND INFORMATICS
DEPARTMENT OF COMPUTER SCIENCE
2nd year Master (IDO)

Time duration: 1h: 30m - Biannual Exam of Information Retrieval & Data Mining - University year: 2020/2021
By Dr. B. LOUNNAS

Exercise 01: Course question (06pt)

1. Describe in short definition exact match retrieval and best match retrieval? (1pt)
2. Describe shortly what is the important of Data Mining? (1pt)
3. CRISP-DM defined and validated a data mining process that could be applicable in any domain application sectors.
   a. Cite some benefits of using such process? (1pt)
4. What is the result of using Clustering algorithm? (1pt)
5. What's the benefit of building decision tree? (1pt)
6. What is the role of association rules algorithm? (1pt)

Exercise 02: Information Retrieval Models (04pt)

You have the following values of *tf* of three (3) documents from a collection of 806791 documents, and the values of *tf* and *Idf* of four (4) terms:

	Doc 1	**Doc 2**	**Doc 3**
car	12	25	39
auto	2	33	1
insurance	0	31	20
best	10	0	12

term	**df_t**	**idf_t**
car	18165	1.65
auto	6723	2.08
insurance	19241	1.62
best	25235	X

- First, calculate the value of **X1**?

- Compute the two top scoring documents on the query "car car insurance" for the following SMART notation scheme:

  - nnc.ntc

Exercise 03: Decision tree (04pt)

Roger Federer is one of the greatest tennis player since tennis have been invented. We want to learn a little about what makes R. Federer win or lose a match. We would like to be able to predict the outcome of a match. Knowing the conditions in which a tennis match takes place, we would like to predict whether R. Federer will win or lose the match. To do so, we gathered data from games played by R. Federer, shown in table below

Time	Match type	Court surface	Best effort	Outcome
Morning	Master	Grass	Yes	*Win*
Afternoon	Grand Slam	Clay	Yes	*Win*
Night	Friendly	Hard	No	*Win*
Afternoon	Friendly	Mixed	No	*Loose*
Afternoon	Master	Clay	Yes	*Loose*
Afternoon	Grand Slam	Grass	Yes	*Win*
Afternoon	Grand Slam	Hard	Yes	*Win*
Night	Master	Mixed	Yes	*Loose*
Afternoon	Master	Grass	Yes	*Win*
Afternoon	Grand Slam	Hard	Yes	*Win*
Afternoon	Master	Clay	Yes	*Loose*
Afternoon	Grand Slam	Hard	Yes	*Win*
Morning	Master	Grass	Yes	*Win*
Afternoon	Grand Slam	Grass	Yes	*Loose*
Night	Friendly	Hard	No	*Win*
Afternoon	Grand Slam	Clay	Yes	*Win*

1. Build the decision tree?

Exercise 04: Association rules (06pt)

The 'database' below has four transactions. What association rules can be found in this set, if the minimum support is 3 transaction and the minimum confidence is 80%?

T1	K, A, D, B
T2	D, A, C, E, B
T3	C, A, B, E
T4	B, A, D