

Ex. Frequent Pattern

Dataset

1. {A,B,C}
2. {B,C,D}
3. {C,D}
4. {A,B,D,E}
5. {A, E}
6. {A,B,C,E}
7. {C,D,E}
8. {E,F}
9. {B,E,F}
10. {B,C,D,E}

- a) Extract frequent pattern with minimum support equal to 30%
- b) Extract Association Rules with minimum confidence equal to 70%
- c) Compute the interest of the rules extracted in b)

Ex. Classification

Consider the following datasets: training (left), test (right)

	Minutes	Internet	Giga	ChangeContract		Minutes	Internet	Giga	ChangeContract
0	210	4G	<10	NO	0	160	3G	<10	NO
1	120	3G	>10	YES	1	150	3G	>10	YES
2	200	4G	>10	NO	2	220	3G	<10	NO
3	160	3G	<10	YES	3	120	4G	>10	NO
4	125	4G	>10	NO	4	125	4G	>10	YES
5	150	3G	>10	YES					
6	220	4G	<10	NO					
7	120	4G	>10	NO					
8	150	3G	>10	YES					
9	220	3G	<10	NO					

Use the training dataset for building a decision tree based on misclassification rate for the variable "ChangeContract", expanding the nodes of the tree until the precision is not improved locally, i.e., no split provides a gain.

Use the test set to provide the confusion matrix and to evaluate the accuracy, precision and recall of the tree. Provide the formula of each evaluation measure.