

Data Mining and Warehousing Lab

Lab Assignment: 3

Tasks to perform (in a programming language of your choice without using the inbuilt functions):

1. Using Bayesian classification, predict the class (Target wait) for the following sample:

X = (alt = T, Bat = T, Fri = F, Hun = T, Pat = Some, Price = \$\$\$, Rain = T, Res = T, Type = Italian, Est > 60).

Note: Consider the sample in DATASET.rar file to solve the problem.

2. Given the Table A, predict a class label using naïve Bayesian classification for the tuple:

X = {age = "<= 30", income = "medium", student = "yes", credit_rating = "fair"}

Table A

RID	Age	Income	Student	Credit_rating	Class : buys_computer
1	<=30	High	No	Fair	No
2	<=30	High	No	Excellent	No
3	31 ...40	High	No	Fair	Yes
4	>40	Medium	No	Fair	Yes
5	>40	Low	Yes	Fair	Yes
6	>40	Low	Yes	Excellent	No
7	31 ...40	Low	Yes	Excellent	Yes
8	<=30	Medium	No	Fair	No
9	<=30	Low	Yes	Fair	Yes
10	>40	Medium	Yes	Fair	Yes
11	<=30	Medium	Yes	Excellent	Yes
12	31 ...40	Medium	No	Excellent	Yes
13	31 ...40	High	Yes	Fair	Yes
14	>40	Medium	No	Excellent	No

3. Using the “C4.5 Algorithm”, construct a decision tree for the following table:

Day	Outlook	Temperature	Humidity	Winter	Play_ball
1	sunny	85	High	weak	No
2	sunny	80	High	strong	No
3	Overcast	83	High	weak	Yes
4	Rain	70	High	weak	Yes
5	Rain	68	Normal	weak	Yes
6	Rain	65	Normal	strong	No
7	Overcast	64	Normal	strong	Yes
8	sunny	72	High	weak	No
9	sunny	69	Normal	weak	Yes
10	Rain	75	Normal	weak	Yes
11	sunny	75	Normal	strong	Yes
12	Overcast	72	High	strong	Yes
13	Overcast	81	Normal	weak	Yes
14	Rain	71	High	strong	No

Hint: Use Bayes Theorem, Entropy and Information Gain