

Tutorial 4 Classification II

School of Computer Science and Electrical Engineering

INFS4203/7203 – Data Mining

Question 1.

- (a) Perform ID3 Algorithm to derive a decision tree and rules for motor diagnosis. The attribute MOTOR is the outcome (class) of the training set (show all detailed calculations).

LINE	SPEED	AGE	TEMPERATURE	MOTOR
	low	old	high	bad
	low	old	normal	bad
	normal	new	normal	good
	normal	old	high	bad
	high	old	high	bad
	high	old	normal	good
	normal	new	normal	good
	low	new	high	bad
	low	new	high	bad

- (b) Perform ID3 Algorithm to derive a decision tree and rules for playing Golf. The attribute Outcome is the class of the training set (“No.” is not an attribute in the table but a row number).

No.	OUTLOOK	HUMIDITY	TEMPERATURE	WINDY	outcome
1	sunny	70	75	yes	do
2	sunny	90	80	yes	dont
3	sunny	85	85	no	dont
4	sunny	95	72	no	dont
5	sunny	70	69	no	do
6	overcast	90	72	yes	do
7	overcast	78	83	no	do
8	overcast	65	64	yes	do
9	overcast	75	81	no	do
10	rain	80	71	yes	dont
11	rain	70	65	yes	dont
12	rain	80	75	no	do
13	rain	80	68	no	do
14	rain	96	70	no	do

- (c) Discuss the problems and necessary improvements to ID3 algorithm in dealing with *noisy data*, *unknown attribute values*, and *excessive branching*. Modify or insert new values in the table of (b) to illustrate the problems (hint: read the reference papers about C4.5 or C5).