

Marks: 20

Question No. 1

Dataset	DocID	Features- Words in documents	Class Fruit=Yes/No
Training set	1	Orange, Orange, Lemon	No
	2	Orange, Red, Blue	No
	3	Apricot, Apple, Mango	Yes
	4	Apple, Banana , Orange	Yes
	5	Apple, Orange, Melon	Yes
Test set	6	Orange, Mango, Melon	?
	7	Orange, Red, Lemon	?

$$P(\text{Fruit}=\text{Yes}) = 3/5 = 0.6$$

$$P(\text{Fruit}=\text{No}) = 2/5 = 0.4$$

P(Orange/Fruit)	1/6	P(Orange/~Fruit)	4/15
P(Mango/Fruit)	1/9	P(Mango/~Fruit)	1/15
P(Melon/Fruit)	1/9	P(Melon/~Fruit)	1/15
P(Red/Fruit)	1/18	P(Red/~Fruit)	2/15
P(Lemon/Fruit)	1/18	P(Lemon/~Fruit)	2/15

P(d6/Fruit)	$0.6 * (1/6) * (1/9) * (1/9) = 0.001$
P(d6/~Fruit)	$0.4 * (4/15) * (1/15) * (1/15) = 0.0004$

P(d7/Fruit)	$0.6 * (1/6) * (1/18) * (1/18) = 0.0003$
P(d7/~Fruit)	$0.4 * (4/15) * (2/15) * (2/15) = 0.001$

Document d7 belongs to class Fruit=No.

$$C2 = \{P4, P5, P6\} = (7, 3)$$