PART A.

2. a.

i)

	trendy	wood	chair	old	discount	blue	jeans	navy	sweater
A	1	1	1	0	0	0	0	0	0
B	0	0	0	2	1	1	1	1	0
C	1	0	0	1	1	1	0	0	1

ii)

	trendy	wood	chair	old	discount	blue	jeans	navy	sweater
A	1	1	1	0	0	0	0	0	0
B	0	0	0	1	1	1	1	1	0
C	1	0	0	1	1	1	0	0	1

iii)

	trendy	wood	chair	old	discount	blue	jeans	navy	sweater
IDF	2	2.5	2.5	2	2	2	2.5	2.5	2.5
TF-A	1	1	1	0	0	0	0	0	0
TF-B	0	0	0	2	1	1	1	1	0
TF-C	1	0	0	1	1	1	0	0	1

b. Q = "old jeans"

Let $ED(A,B)$ be Euclidean distance between A and B.

Then $ED(B,Q) = 2$ and $ED(C,Q) \approx 2.236$.

Since $ED(B,Q)$ is smaller, I recommend product B.

c. Q = "discount chair"

Let $CS(A,B)$ be cosine similarity between A and B.

Then $CS(B,Q) \approx 0.2683$ and $CS(C,Q) \approx 0.3202$

~~Sine CS(B,Q) is smaller, I recommend~~

d. ① Euclidean Distance is sensitive to magnitude of sentences → Accounts for different document length

② Euclidean Distance performs poorly when high-dimensional space. → Use dot product, which handles high-dimensional data better.

3.

a. i) $\text{accuracy} = \dfrac{2+9}{2+2+2+9} = \dfrac{11}{15}$

ii) $\text{precision} = \dfrac{2}{2+2} = \dfrac{1}{2}$

iii) $\text{Recall} = \dfrac{2}{2+2} = \dfrac{1}{2}$

iv) $F1 = 2 \cdot \dfrac{P \cdot R}{P+R} = 2 \cdot \dfrac{\frac{1}{2} \cdot \frac{1}{2}}{\frac{1}{2}+\frac{1}{2}} = \dfrac{\frac{1}{2}}{1} = \dfrac{1}{2}$

b.

		Predicted: Threat	Predicted: No threat
Actual	Threat	2	2 (FN)
	No Threat	2 (FP.)	9

$\dfrac{TP}{TP+FN}$.

→ Recall. The model should be more accepting of FP since cost of FN is extremely high.

c. ~~Accuracy. Majority of classes are either True threat or True No threat. Accuracy captures how many cases are classified correctly.~~

d. ~~0.5 or 7/15. If it is a complete random guess, then the accuracy is 0.5, and in our case half of it can be rounded down to 7.~~

11/15. can be achieved by predicting everything as No threat.

PART A.

1C. "This looks bad"

i) prior :

$$p(Y = IB) = \frac{1}{2}$$
$$p(Y = NIB) = \frac{1}{2}$$

ii) Evidence.

$$p(X) = p(X \mid Y=IB) \cdot p(Y=IB) + p(X \mid Y=NIB) \cdot p(Y=NIB)$$
$$= \frac{1}{9} \cdot \frac{1}{2} + \frac{1}{9} + \frac{1}{2} = 1/9$$

iii) Likelihood.

$$p(X \mid Y=IB) = p(X=\text{"looks"} \mid Y=IB) \cdot p(X=\text{"bad"} \mid Y=IB)$$
$$= \frac{1}{3} \cdot \frac{1}{3}$$
$$= \frac{1}{9}$$

$$p(X \mid Y=NIB) = p(X=\text{"looks"} \mid Y=NIB) \cdot p(X=\text{"bad"} \mid Y=NIB)$$
$$= \frac{1}{3} \cdot \frac{1}{3}$$
$$= \frac{1}{9}.$$

iv) posterior:

$$P(Y = ZB \mid X) = \frac{P(X \mid Y = ZB) \cdot P(Y = ZB)}{P(X)}$$

$$= \frac{1/9 \cdot 1/2}{1/9} = 1/2$$

$$P(Y = NZB \mid X) = \frac{P(X \mid Y = NZB) \cdot P(Y = NZB)}{P(X)}$$

$$= \frac{1/9 \cdot 1/2}{1/9} = 1/2$$