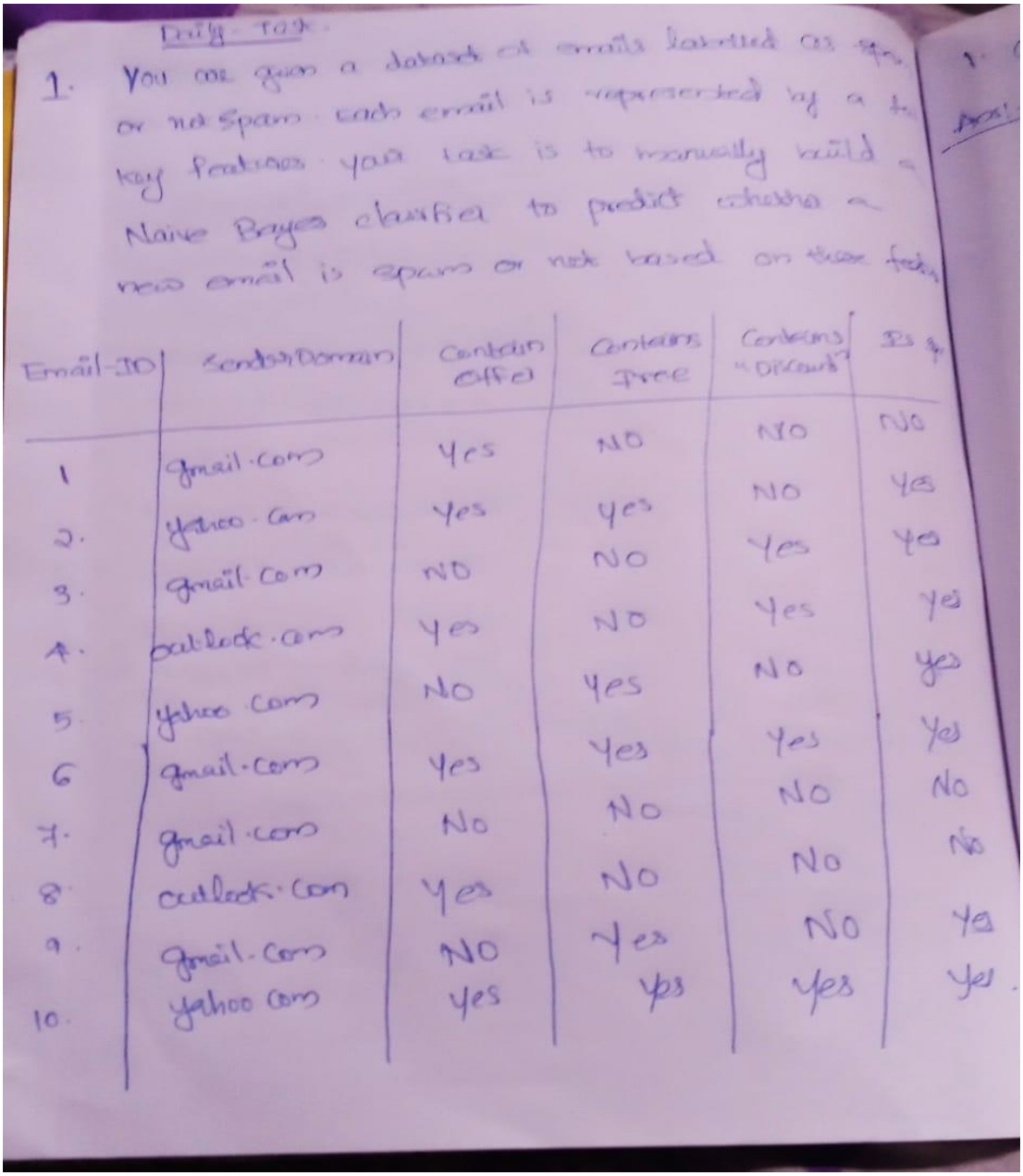
1. Calculate the prior probabilities, spam and not spam
2. Use the Naive Bayes classifier to predict whether a new email from gmail.com that contains "Free" but no "Discount" is spam or not



Daily Task.

1. You are given a dataset of emails labelled as spam or not spam. Each email is represented by a few key features. Your task is to manually build a Naive Bayes classifier to predict whether a new email is spam or not based on these features.

Email-ID	Sender Domain	Contain Offer	Contains Free	Contains "Discount"	Is Spam
1	gmail.com	Yes	NO	NO	NO
2.	yahoo.com	Yes	yes	NO	yes
3.	gmail.com	NO	NO	Yes	yes
4.	outlook.com	Yes	NO	Yes	yes
5.	yahoo.com	NO	Yes	NO	yes
6	gmail.com	Yes	Yes	Yes	Yes
7.	gmail.com	No	NO	NO	No
8.	outlook.com	Yes	NO	No	No
9.	gmail.com	NO	Yes	NO	Yes
10.	yahoo.com	Yes	Yes	Yes	Yes

1. Calculate the prior probabilities, spam & not spam.

Ans:- Prior probability of spam (P(Spam))

$$P(\text{Spam}) = \frac{\text{Num of spam emails}}{\text{Total num of emails}}$$

$$= \frac{7}{10} = 0.7$$

$$P(\text{not spam}) = \frac{\text{Num of not spam emails}}{\text{Total num of emails}}$$

$$= \frac{3}{10} = 0.3$$

2. Use the naive bayes classifier to predict whether a new email from gmail.com that contains free but no discount is spam or not.

gmail, Contain Free (Yes), Discount (No) = spam or not spam ?

I

Step 1:- Sender Domain gmail.com.

$$P(\text{gmail} \mid \text{Spam}) = \frac{\text{Num of spam emails from gmail.com}}{\text{Total number of spam emails.}}$$

$$= \frac{3}{7} = 0.42$$

$$P(\text{gmail} \mid \text{Not spam}) = \frac{\text{Num of not spam emails from gmail}}{\text{Total num of not spam email}}$$

$$= \frac{2}{3} = \underline{0.66}$$

Step 2 :- find Contain Free yes | spam or not spam

→ $$P(\text{Contain Free} = \text{Yes} \mid \text{Spam}) = \frac{\text{Num of spam emails containing Free}}{\text{Total num of spam email}}$$

$$= \frac{5}{7} = \underline{0.71}$$

→ $$P(\text{Contain Free} = \text{yes} \mid \text{not spam}) = \frac{\text{Num of Not spam emails containing free}}{\text{Total num of not spam email.}}$$

$$= \frac{0}{3} = \underline{0}$$

Step 3: Find Contain Discount = No | Spam & not

→ P(Contain Discount = No | Spam)

= Num of spam emails Contain no Discount / Total num of spam email

= 3/7 = 0.42

→ P(Contain Discount = No | not spam)

= Num of Not spam emails Contain Discount / Total num of not spam email.

= 3/3 = 1

II Calculate posterior probabilities using Naive Bayes

→ Probability of every features with all spam

P(x | spam) = 0.42 × 0.71 × 0.42

= 0.125

→ Probability of every features with not spam

$P(x|\text{not spam}) = 0.66 \times 0 \times 1$

$= \underline{0}$

III $

→ probability of this spam into corresponding spam

$P(x|\text{spam}) \times P(\text{spam})$

$= 0.125 \times 0.7$

$= \underline{0.08}$

→ probability of this not spam into corresponding not spam

$P(x|\text{not spam}) \times P(\text{not spam})$

$= 0 \times 0.3$

$= \underline{0}$

$\therefore P(x|\text{spam}) > P(x|\text{Not spam})$

The new email from gmail.com that contains "free" but no "Discount" is predicted to be spam.