islington college
(इस्लिङ्टन कलेज)

LONDON METROPOLITAN UNIVERSITY

Module Code & Module Title

CU6051NI ARTIFICIAL INTELLIGENCE

Assessment Type

Lab Assignment 2

Year and Semester

2021-22 Autumn

Student Name: Mimona Karki

London Met ID: 19031154

College ID: NP01CP4A190214

Assignment Due Date: December 17, 2021

Assignment Submission Date: December 17, 2021

I confirm that I understand my coursework needs to be submitted online via Google Classroom under the relevant module page before the deadline for my assignment to be accepted and marked. I am fully aware that late submissions will be treated as non-submission and a mark of zero will be awarded.

Table of Contents

List of Tables

1. Training Data

Congrats, You have won!! reply to our sms for a free nokia mobile + free camcorder.	Spam
Congrats! 1 year special cinema pass for 2 is yours. reply to this sms to claim your prize.	Spam
I am pleased to tell you that you are awarded with a 1500 Bonus Prize, reply to this sms to claim your prize.	Spam
Dont worry. I guess he is busy.	Not Spam
Going for dinner. msg you later.	Not Spam
Ok, I will call you up when I get some cash.	Not Spam

2. Test Data

Text	Label
I am busy. I will msg you later.	?
Congrats! You are awarded a free mobile.	?

MIMONA KARKI

3. Vocabulary

Congrats, you, have, won, reply, to, our, sms, for, a, free, nokia, mobile, free, camcorder, year, special, cinema, pass, is, yours, this, claim, your, prize, I, am, pleased, tell, that, are, awarded, with, bonus, don't, worry, guess, he, busy, going, dinner, msg, later, ok, will, call, up, when, get, some, cash.

According to the given data set,

Total no. of unique words (Vocabulary) = 51

No. of words in spam text, n(spam) = 53

No. of words in non-spam text, n(notspam) = 24

Probability of spam text,

P(spam) = 3/6 = 0.5

Probability of not spam text,

P(notSpam) = 3/6 = 0.5

4. Word Tables

Table 1: Word table part 1

Text no.	Congrats	you	have	won	reply	to	our	sms	for	a	free	nokia	mobile	camcorder	year	special	cinema	Label
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	Spam
2	1	0	0	0	1	1	0	1	0	0	0	0	0	0	0	1	1	Spam
3	0	2	0	0	1	1	0	1	0	1	0	0	0	0	0	0	0	Spam
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Not spam
5	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Not spam
6	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Not spam

Table 2: Word table part 2

Text no.	pass	is	yours	this	claim	your	prize	I	am	pleased	tell	that	are	awarded	with	bonus	Label
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Spam
2	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	Spam
3	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	Spam
4	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Not spam
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Not spam
6	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	Not spam

Table 3: Word table part 3

Text no.	Don't	worry	guess	he	busy	going	dinner	msg	later	ok	will	call	up	when	get	some	cash	Label
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Spam
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Spam
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Spam
4	1	1	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	Not spam
5	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	1	1	Not spam
6	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	Not spam

The formula to calculate the probability that a given word comes from a spam text is:

$$p(w_k|spam) = \frac{n_k + 1}{n + |Vocabulary|}$$

The formula to calculate the probability that a given word comes from a non-spam text is:

$$p(w_k|notspam) = \frac{n_k + 1}{n + |Vocabulary|}$$

The Classification formula is,

$$y_{new} = argmax_{y\epsilon(+,-)}\ p(y) \prod_{w \in words} p(w|y)$$

5. Finding the label for the first text in the Test Data table.

"I am busy. I will msg you later."

Using classification formula, comparing the words with the word table where the label is "spam", We get,

$$\boldsymbol{y_{spam}} = p(spam)p(I|spam)p(am|spam)\ p(busy|spam)p(I|spam)\ p(will|spam)$$
$$p(msg|spam)p(you|spam)p(later|spam)$$

$$\boldsymbol{y_{spam}} = 0.5\times0.0192\times0.0192\times0.0096\times0.0192\times0.0096\times0.0096\times\ 0.0384\times0.0096$$
$$\boldsymbol{y_{spam}} = 1.16910912E\text{-}15$$

Comparing the words with the word table where the label is "not spam" using classification formula,

$$\boldsymbol{y_{notSpam}} = p(notSpam)p(I|notSpam)p(am|notSpam)$$
$$p(busy|notSpam)p(I|notSpam)\ p(will|notSpam)$$
$$p(msg|notSpam)p(you|notSpam)p(later|notSpam)$$

$$\boldsymbol{y_{notSpam}} = 0.5\times0.0533\times0.0133\times0.0266\times0.0533\times0.0266\times0.0266\times\ 0.04\times0.0266$$
$$\boldsymbol{y_{notSpam}} = 3.78322875E\text{-}13$$

Comparing $\boldsymbol{y_{spam}}$ and $\boldsymbol{y_{notSpam}}$, $\boldsymbol{y_{notSpam}}$ is closer to 1 than $\boldsymbol{y_{spam}}$.

Hence, the text "*"I am busy. I will msg you later."* is classified as not spam.

6. Finding the label for the second text in the Test Data table.

"*Congrats! You are awarded a free mobile.*"

Using classification formula, comparing the words with the word table where the label is "spam", we get,

$$\boldsymbol{y_{spam}} = p(spam)p(Congrats|spam)p(You|spam)\ p(are|spam)p(awarded|spam)$$
$$p(a|spam) \quad p(free|spam)p(mobile|spam)$$

$$\boldsymbol{y_{spam}} = 0.5\times0.0288\times0.0384\times0.0192\times0.0192\times0.0288\times0.0192\times0.0192$$
$$\boldsymbol{y_{spam}} = 2.18856200E\text{-}12$$

Comparing the words with the word table where the label is "not spam" using classification formula,

$$\boldsymbol{y_{notSpam}} = p(notSpam)p(Congrats|notSpam)p(You|notSpam)$$
$$p(are|notSpam)p(awarded|notSpam)\ p(a|notSpam)$$
$$p(free|notSpam)p(mobile|notSpam)$$

$$\boldsymbol{y_{notSpam}} = 0.5\times0.0133\times0.04\times0.0133\times0.0133\times0.0133\times0.0133\times0.0133$$
$$\boldsymbol{y_{notSpam}} = 1.10698017E\text{-}13$$

Comparing $\boldsymbol{y_{spam}}$ and $\boldsymbol{y_{notSpam}}$, $\boldsymbol{y_{spam}}$ is closer to 1 than $\boldsymbol{y_{notSpam}}$.

Hence, the text "*Congrats! You are awarded a free mobile."* is classified as spam.