**CU6051NI - Artificial Intelligence**

**Lab work – 5**

For this lab work, you will be working with the Naïve Bayes Classifier for spam detection.

Given below are 2 tables containing training data and test data respectively.

Training examples consist of text (sms) labeled as spam or not spam. Use the examples to build

the vocabulary for the classifier. Then using the bag of words approach, transform the texts

into feature vectors.

Then following the algorithm for the Naïve Bayes Classifier, classify the 2 texts in the test data

**Training Data:**

|  |  |
| --- | --- |
| **Text** | **Label** |
| 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 |

**Text Classification of Training Data of Spam and Non-Spam via code:**

spamDict=dict()

nonSpamDict=dict()

#test\_string\_spam = "Congrats, You have won!! reply to our sms for a free nokia mobile free camcorder. Congrats! 1 year special cinema pass for 2 is yours. reply to this sms to claim your prize. I am pleased to tell you that you are awarded with a 1500 Bonus Prize, reply to this sms to claim your prize."

def word\_count\_spam(str):

    counts = dict()

    spam=dict()

    words = str.split()

    for word in words:

        if word in counts:

            counts[word] += 1

        else:

            counts[word] = 1

        spam[word]=(counts[word]+1)/109

    spamDict=spam

    return spam

spamForOthers= 1/109

print( word\_count\_spam('congrats you have won reply to our sms for a free nokia mobile free camcorder congrats 1 year special cinema pass for 2 is yours reply to this sms to claim your prize I am pleased to tell you that you are awarded with a 1500 bonus prize reply to this sms to claim your prize'))

print("Spam classification value for other words not in Spam: " + str(spamForOthers))

print("...........................................................NOT SPAM BELOW.........................................................................")

def word\_count\_notspam(str):

    counts = dict()

    spam=dict()

    words = str.split()

    for word in words:

        if word in counts:

            counts[word] += 1

        else:

            counts[word] = 1

        spam[word]=(counts[word]+1)/76

    return spam

nonSpamForOthers=1/76

print("Not-Spam classification value for other words not in Not Spam like Congrats etc.: " + str(nonSpamForOthers))

print( word\_count\_notspam('dont worry I guess he is busy going for dinner msg you later ok I will call you up when I get some cash'))

Output:

{'congrats': 0.027522935779816515, 'you': 0.03669724770642202, 'have': 0.01834862385321101, 'won': 0.01834862385321101, 'reply': 0.03669724770642202, 'to': 0.06422018348623854, 'our': 0.01834862385321101, 'sms': 0.03669724770642202, 'for': 0.027522935779816515, 'a': 0.027522935779816515, 'free': 0.027522935779816515, 'nokia': 0.01834862385321101, 'mobile': 0.01834862385321101, 'camcorder': 0.01834862385321101, '1': 0.01834862385321101, 'year': 0.01834862385321101, 'special': 0.01834862385321101, 'cinema': 0.01834862385321101, 'pass': 0.01834862385321101, '2': 0.01834862385321101, 'is': 0.01834862385321101, 'yours': 0.01834862385321101, 'this': 0.027522935779816515, 'claim': 0.027522935779816515, 'your': 0.027522935779816515, 'prize': 0.03669724770642202, 'I': 0.01834862385321101, 'am': 0.01834862385321101, 'pleased': 0.01834862385321101, 'tell': 0.01834862385321101, 'that': 0.01834862385321101, 'are': 0.01834862385321101, 'awarded': 0.01834862385321101, 'with': 0.01834862385321101, '1500': 0.01834862385321101, 'bonus': 0.01834862385321101}

Spam classification value for other words not in Spam: 0.009174311926605505

...........................................................NOT SPAM BELOW.........................................................................

Not-Spam classification value for other words not in Not Spam like Congrats etc.: 0.013157894736842105

{'dont': 0.02631578947368421, 'worry': 0.02631578947368421, 'I': 0.05263157894736842, 'guess': 0.02631578947368421, 'he': 0.02631578947368421, 'is': 0.02631578947368421, 'busy': 0.02631578947368421, 'going': 0.02631578947368421, 'for': 0.02631578947368421, 'dinner': 0.02631578947368421, 'msg': 0.02631578947368421, 'you': 0.039473684210526314, 'later': 0.02631578947368421, 'ok': 0.02631578947368421, 'will': 0.02631578947368421, 'call': 0.02631578947368421, 'up': 0.02631578947368421, 'when': 0.02631578947368421, 'get': 0.02631578947368421, 'some': 0.02631578947368421, 'cash': 0.02631578947368421}

**Test Data**

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

**Classifying the sentence of Test Data as spam or not-spam:**

“I am busy. I will msg you later.”

Y(spam)=p(spam)p(I/spam)p(am/spam)p(busy/spam)p(will/spam)p(msg/spam)p(you/spam)(later/spam)

= 0.5 \* 0.018 \* 0.018 \* 0.009 \* 0.009 \* 0.036 \* 0.009

= 0.000000000000425

Y(non-spam) = p(non-spam)p(I/ non-spam)p(am/ non-spam)p(busy/ non-spam)p(will/ non-spam)p(msg/ non-spam)p(you/ non-spam)(later non-/spam)

= 0.5 \* 0.0526 \* 0.0131 \* 0.0262 \* 0.0263 \* 0.0263 \* 0.0394 \* 0.0263

= 0.00000000000649

The sentence will be classified as a not spam. Since the value of non-spam is greater than spam.

“Congrats! You are awarded a free mobile”

Y(spam)=p(spam)p(congrats/spam)p(you/spam)p(are/spam)p(awarded/spam)p(a/spam)p(free/spam)p(mobile/spam)

= 0.5\*0.0275\*0.0366\*0.0183\*0.018\*0.0275\*0.0275\*0.0183

= 0.000000000336

Y(non-spam)=p(non-spam)p(congrats/non-spam)p(you/non-spam)p(are/non-spam)p(awarded/non spam)p(a/non-spam)p(free/non-spam)p(mobile/non-spam)

= 0.5\*0.0131\*0.0394\*0.0131\*0.0131\*0.0131\*0.0131\*0.0131

= 0.00000000000000995

The sentence will be classified as a spam. Since the value of spam is greater than non-spam.

**Final Output**

The final output of the test data:

|  |  |
| --- | --- |
| **Text** | **Label** |
| I am busy. I will msg you later. | Non-Spam |
| Congrats! You are awarded a free mobile. | Spam |