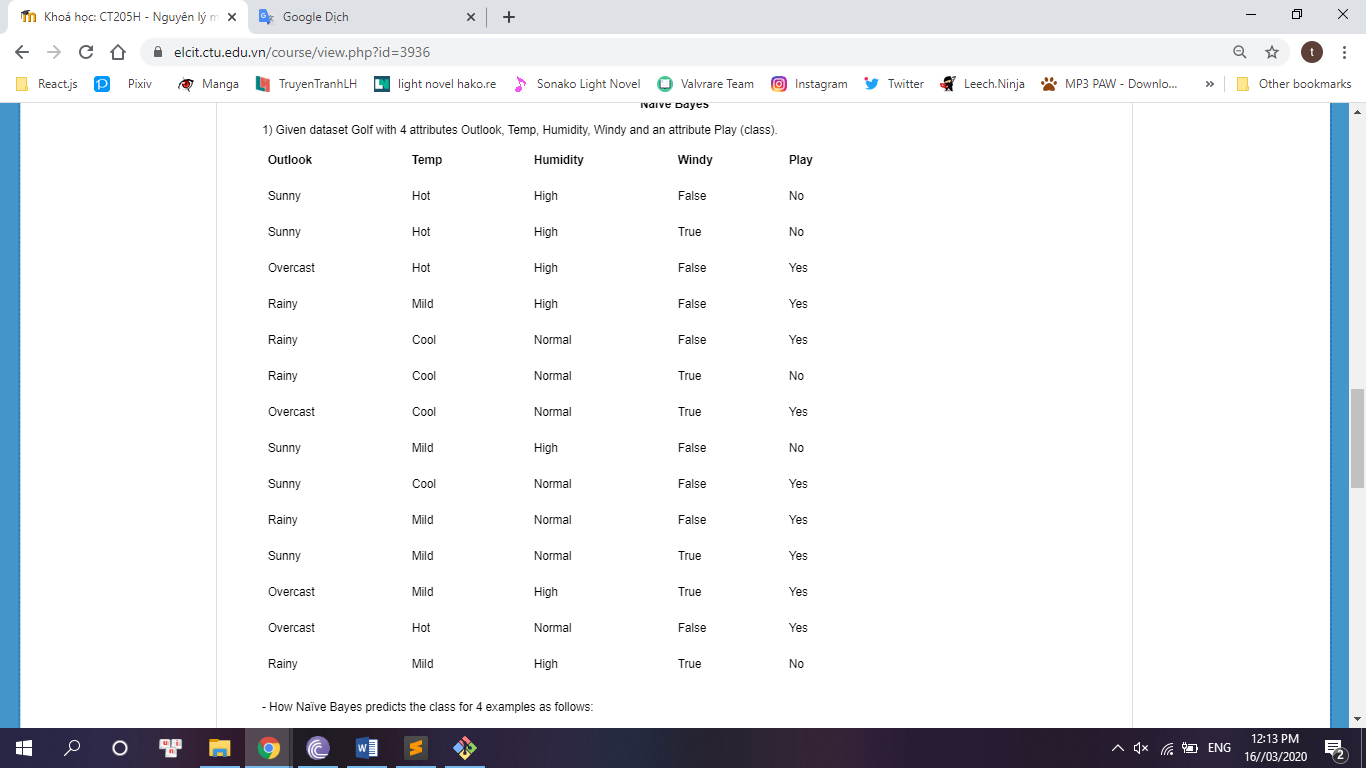
Naïve Bayes

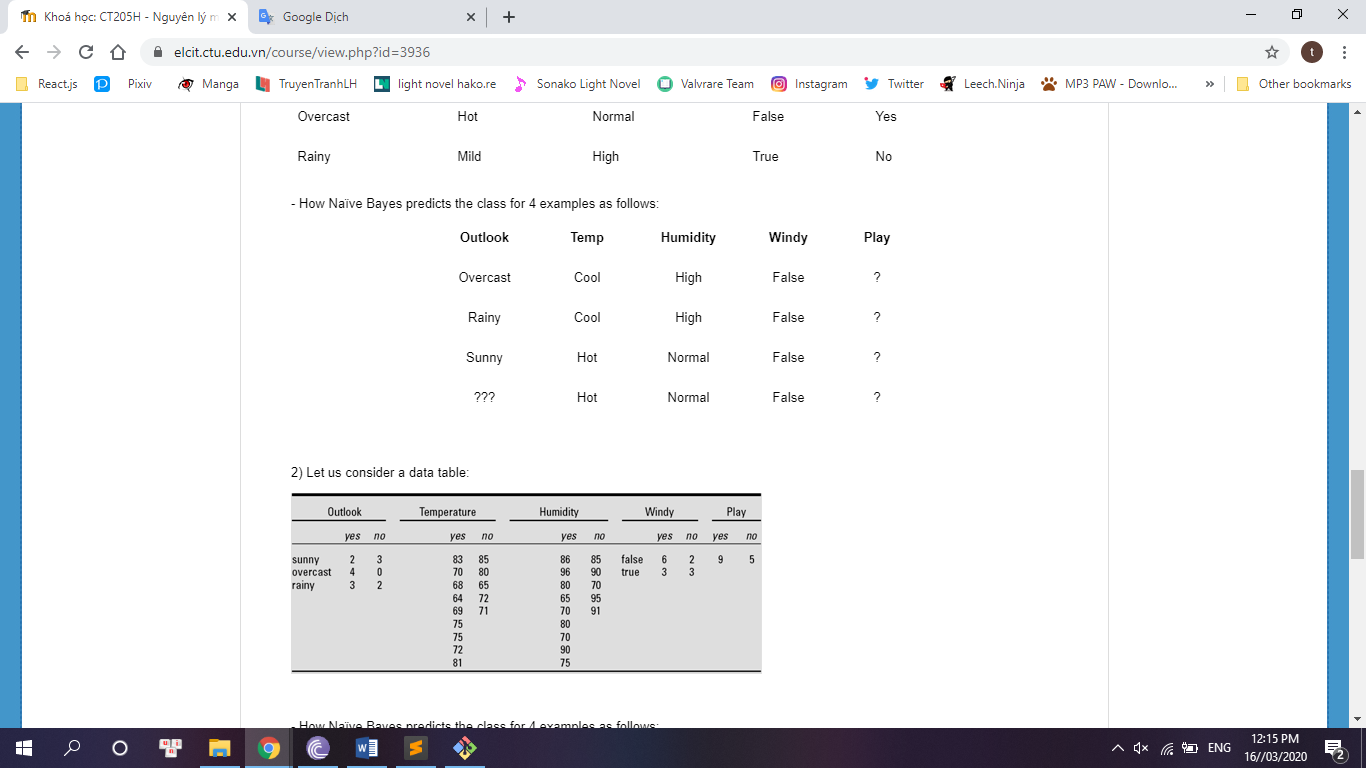
Student name: Tran Chien Thanh

Student code: B1605365

1. **Given dataset Golf with 4 attributes Outlook, Temp, Humidity, Windy and an attribute Play (class).**



**- How Naïve Bayes predicts the class for 4 examples as follows:**



|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Outlook** | | | **Temperature** | | | **Humidity** | | | **Windy** | | | **Play** | | |
|  | Yes | No |  | Yes | No |  | Yes | No |  | Yes | No |  | Yes | No |
| Sunny | 2 | 3 | Hot | 2 | 2 | High | 3 | 4 | False | 6 | 2 |  | 9 | 5 |
| Overcast | 4 | 0 | Mild | 4 | 2 | Normal | 6 | 1 | True | 3 | 3 |  |  |  |
| Rainy | 3 | 2 | Cool | 3 | 1 |  |  |  |  |  |  |  |  |  |
| Sunny | 2/9 | 3/5 | Hot | 2/9 | 2/5 | High | 3/9 | 4/5 | False | 6/9 | 2/5 |  | 9/14 | 5/14 |
| Overcast | 4/9 | 0/5 | Mild | 4/9 | 2/5 | Normal | 6/9 | 1/5 | True | 3/9 | 3/5 |  |  |  |
| Rainy | 3/9 | 2/5 | Cool | 3/9 | 1/5 |  |  |  |  |  |  |  |  |  |

* **Row 1:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Overcast | Cool | High | False | ? |

Likelihood(yes) = 4/9 \* 3/9 \* 3/9 \* 6/9 \* 9/14 = 0.021

Likelihood(no) = 0/5 \* 1/5 \* 4/5 \* 2/5 \* 5/14 = 0

Because probability be 0 will never occur, so we need to use Laplace estimator

|  |  |  |
| --- | --- | --- |
| Sunny | 3/12 | 4/8 |
| Overcast | 5/12 | 1/8 |
| Rainy | 4/12 | 3/8 |

=> Likelihood(yes) = 5/12 x 3/9 x 3/9 x 6/9 x 9/14 = 0.0198

Likelihood(no) = 1/8 x 1/5 x 4/5 x 2/5 x 5/14 = 0.0029

Probability:

P(yes) = 0.0198 / (0.0198 + 0.0029) = 0.8722

P(no) = 0.0029 / (0.0198 + 0.0029) = 0.1278

P(yes) > P(no) => Weather conditions are suitable for playing a game of golf

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Overcast | Cool | High | False | Yes |

* **Row 2:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Rainy | Cool | High | False | ? |

Likelihood(yes) = 4/12 \* 3/9 \* 3/9 \* 6/9 \* 9/14 = 0.0159

Likelihood(no) = 3/8 \* 1/5 x 4/5 x 2/5 x 5/14 = 0.0086

Probability:

P(yes) = 0.0159 / (0.0159 + 0.0086) = 0.649

P(no) = 0.0086/ (0.0159 + 0.0086) =0.351

P(yes) > P(no) => Weather conditions are suitable for playing a game of golf

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Rainy | Cool | High | False | Yes |

* **Row 3:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Sunny | Hot | Normal | False | ? |

Likelihood(yes) = 3/12 \* 2/9 \* 6/9 \* 6/9 \* 9/14 = 0.0159

Likelihood(no) = 4/8 \* 2/5 \* 1/5 \* 2/5 \* 5/14 = 0.0057

Probability:

P(yes) = 0.0159 / (0.0159 + 0.0057) = 0.7361

P(no) = 0.0057/ (0.0159 + 0.0057) = 0.2639

P(yes) > P(no) => Weather conditions are suitable for playing a game of golf

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Sunny | Hot | Normal | False | Yes |

* **Row 4:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ??? | Hot | Normal | False | ? |

Likelihood(yes) = 2/9 \* 6/9 \* 6/9 \* 9/14 = 0.0634

Likelihood(no) = 2/5 \* 1/5 \* 2/5 \* 5/14 = 0.0114

Probability:

P(yes) = 0.0634 / (0.0634 + 0.0114) = 0.8476

P(no) = 0.0114 / (0.0634 + 0.0114) = 0.1524

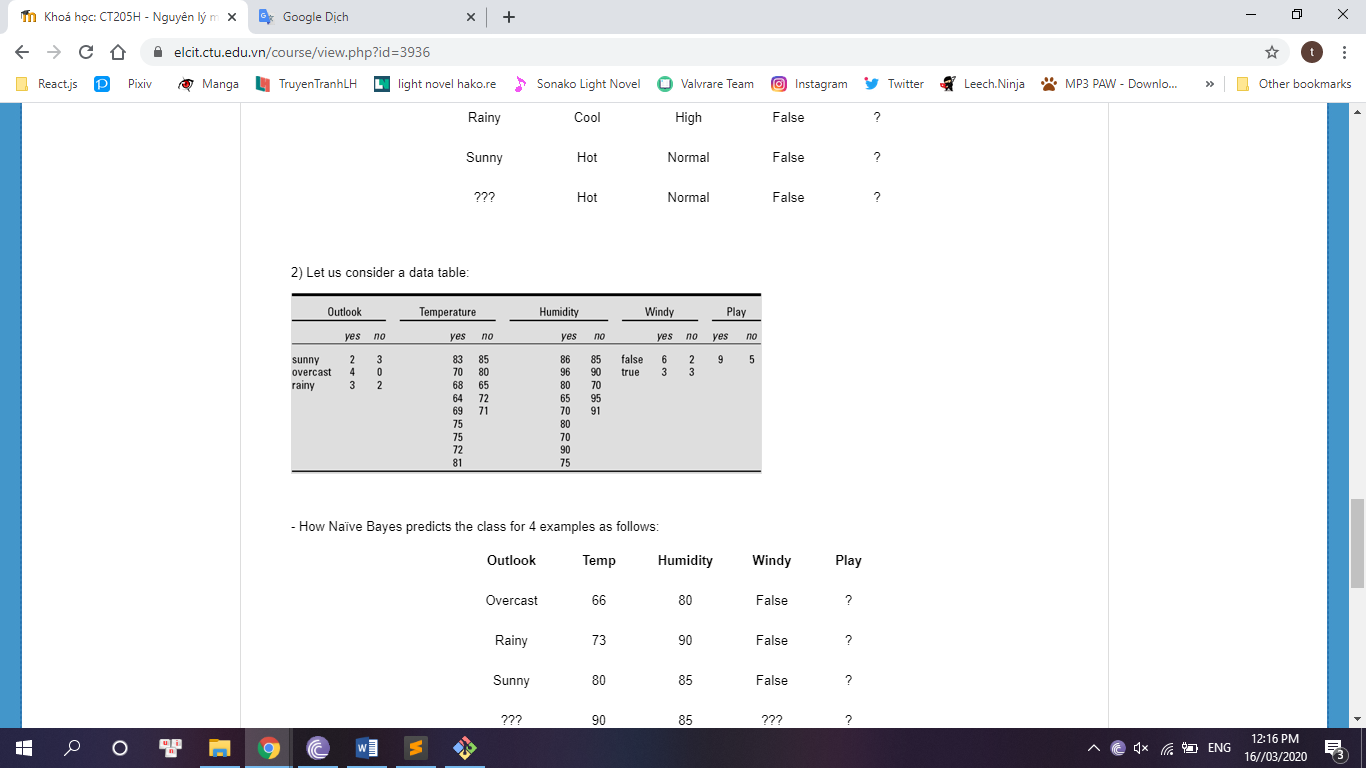
P(yes) > P(no) => Weather conditions are suitable for playing a game of golf

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ??? | Hot | Normal | False | Yes |

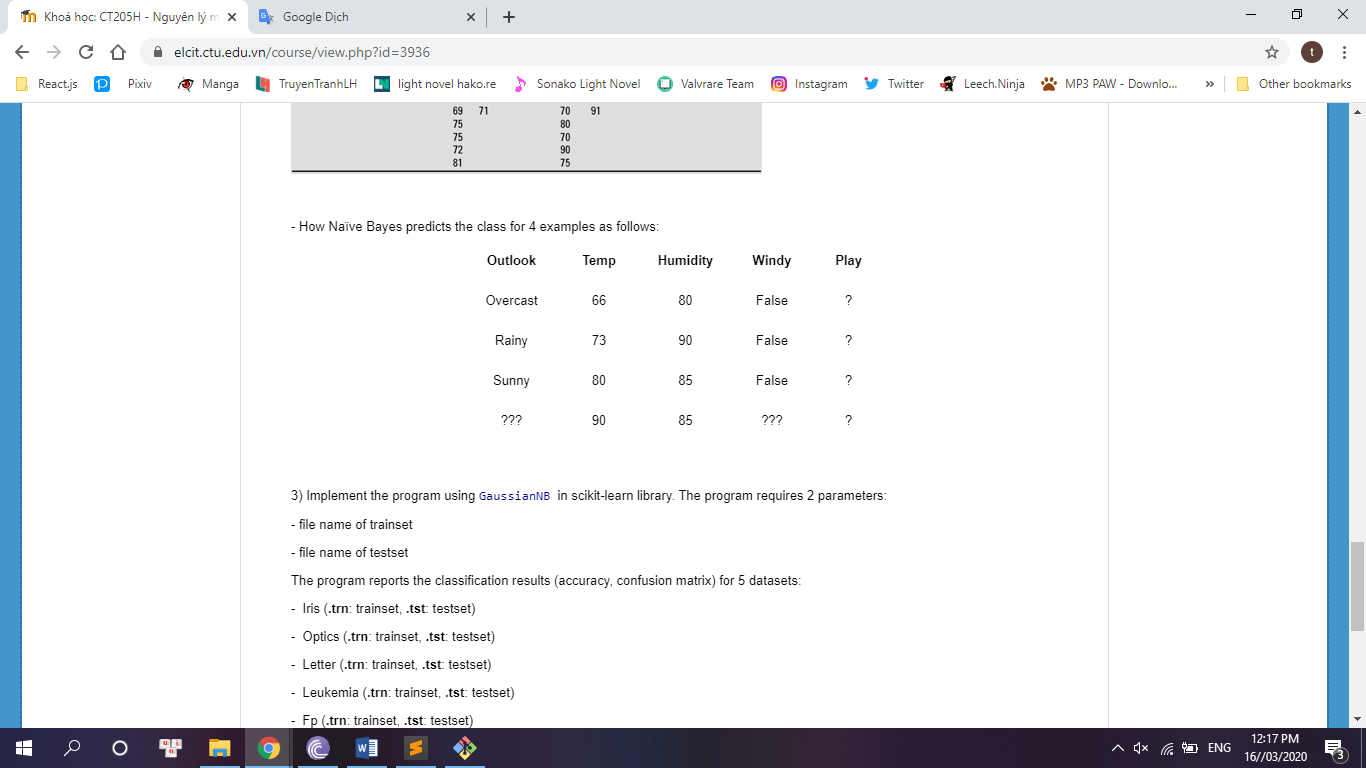
Summary:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Outlook** | **Temperature** | **Humidity** | **Windy** | **Play** |
| Overcast | Cool | High | False | Yes |
| Rainy | Cool | High | False | Yes |
| Sunny | Hot | Normal | False | Yes |
| ??? | Hot | Normal | False | Yes |

1. **Let us consider a data table**



**- How Naïve Bayes predicts the class for 4 examples as follows:**



|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Outlook** | | | **Temperature** | | | **Humidity** | | | **Windy** | | | **Play** | | |
|  | Yes | No |  | Yes | No |  | Yes | No |  | Yes | No |  | Yes | No |
| Sunny | 2 | 3 |  | 83 | 85 |  | 86 | 85 | False | 6 | 2 |  | 9 | 5 |
| Overcast | 4 | 0 |  | 70 | 80 |  | 96 | 90 | True | 3 | 3 |  |  |  |
| Rainy | 3 | 2 |  | 68 | 65 |  | 80 | 70 |  |  |  |  |  |  |
|  |  |  |  | 64 | 72 |  | 65 | 95 |  |  |  |  |  |  |
|  |  |  |  | 69 | 71 |  | 70 | 91 |  |  |  |  |  |  |
|  |  |  |  | 75 |  |  | 80 |  |  |  |  |  |  |  |
|  |  |  |  | 75 |  |  | 70 |  |  |  |  |  |  |  |
|  |  |  |  | 72 |  |  | 90 |  |  |  |  |  |  |  |
|  |  |  |  | 81 |  |  | 75 |  |  |  |  |  |  |  |
| Sunny | 2/9 | 3/5 |  |  |  |  |  |  | False | 6/9 | 2/5 |  | 9/14 | 5/14 |
| Overcast | 4/9 | 0/5 |  |  |  |  |  |  | True | 3/9 | 3/5 |  |  |  |
| Rainy | 3/9 | 2/5 |  |  |  |  |  |  |  |  |  |  |  |  |

* **Temperature**

Mean P(yes) = (83 + 70 + 68 + 64 + 69 + 75 + 75 + 72 + 81) / 9 = 73

Mean P(no) = (85 + 80 + 65 + 72 + 71) / 5 = 74.6

Standard deviation P(yes) =

= 6.2

Standard deviation P(no) =

= 7.9

* **Humidity**

Mean P(yes) = (86 + 96 + 80 + 65 + 70 + 80 + 70 + 90 + 75) / 9 = 79.1

Mean P(no) = (85 + 90 + 70 + 95 + 91) / 5 = 86.2

Standard deviation P(yes) =

= 10.2

Standard deviation P(no) =

= 9.7

Result:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Sunny | 2/9 | 3/5 | mean | 73 | 74.6 | mean | 79.1 | 86.2 | False | 6/9 | 2/5 |  | 9/14 | 5/14 |
| Overcast | 4/9 | 0/5 | Std.dev | 6.2 | 7.9 | Std.dev | 10.2 | 9.7 | True | 3/9 | 3/5 |  |  |  |
| Rainy | 3/9 | 2/5 |  |  |  |  |  |  |  |  |  |  |  |  |

As explained in exercise 1, we need to use Laplace estimator

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Sunny | 3/12 | 4/8 | mean | 73 | 74.6 | mean | 79.1 | 86.2 | False | 6/9 | 2/5 |  | 9/14 | 5/14 |
| Overcast | 5/12 | 1/8 | Std.dev | 6.2 | 7.9 | Std.dev | 10.2 | 9.7 | True | 3/9 | 3/5 |  |  |  |
| Rainy | 4/12 | 3/8 |  |  |  |  |  |  |  |  |  |  |  |  |

* **Row 1:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Overcast | 66 | 80 | False | ? |

Density probability of temperature and humidity:

0.034

0.0279

0.039

0.0335

Likelihood(yes) = 5/12 \* 0.034 \* 0.039 \* 6/9 \* 9/14 = 0.2368 \*

Likelihood(no) = 1/8 \* 0.0279 \* 0.0335 \* 2/5 \* 5/14 = 0.017 \*

Probability:

P(yes) = 0.2368 \* / (0.2368 \* + 0.017 \* ) = 0,933

P(no) = 0.017 \* / (0.2368 \* + 0.017 \* ) = 0. 067

P(yes) > P(no) => Weather conditions are suitable for playing a game of golf

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Overcast | 66 | 80 | False | yes |

* **Row 2:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Rainy | 73 | 90 | False | ? |

Density probability of temperature and humidity:

0.0643

0.0495

0.022

0.038

Likelihood(yes) = 4/12 \* 0.0643 \* 0.022 \* 6/9 \* 9/14= 0.202 \*

Likelihood(no) = 3/8 \* 0.0495 \* 0.038 \* 2/5 \* 5/14 = 0.1008 \*

Probability:

P(yes) = 0.202 \* / (0.202 \* + 0.1008 \* ) = 0.6671

P(no) = 0.1008 \* / (0.202 \* + 0.1008 \* ) = 0.3329

P(yes) > P(no) => Weather conditions are suitable for playing a game of golf

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Rainy | 73 | 90 | False | yes |

* **Row 3:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Sunny | 80 | 85 | False | ? |

Density probability of temperature and humidity:

0.034

0.04

0.033

0.0408

Likelihood(yes) = 3/12 \* 0.034 \* 0.033 \* 6/9 \* 9/14 = 0.1202 \*

Likelihood(no) = 4/8 \* 0.04 \* 0.0408 \* 2/5 \* 5/14 = 0.1166 \*

Probability:

P(yes) = 0.1202 \* / (0.1202 \* + 0.1166 \* ) = 0.508

P(no) = 0.1166 \* / (0.1202 \* + 0.1166 \* ) = 0.492

P(yes) > P(no) => Weather conditions are suitable for playing a game of golf

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Sunny | 80 | 85 | False | yes |

* **Row 4:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ??? | 90 | 85 | ??? | ? |

Density probability of temperature and humidity:

1.5 \* 10-3

7.6 \* 10-3

0.033

0.0408

Likelihood(yes) = 1.5 \* 10-3 \* 0.033 \* 9/14 = 0.3182 \*

Likelihood(no) = 7.6 \* 10-3 \* 0.0408 \* 5/14= 1.1074 \*

Probability:

P(yes) = 0.3182 \* / (0.3182 \* + 1.1074 \* ) = 0.2232

P(no) = 1.1074 \* / (0.3182 \* + 1.1074 \* ) = 0.7768

P(yes) < P(no) => Weather conditions are unsuitable for playing a game of golf

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ??? | 90 | 85 | ??? | no |

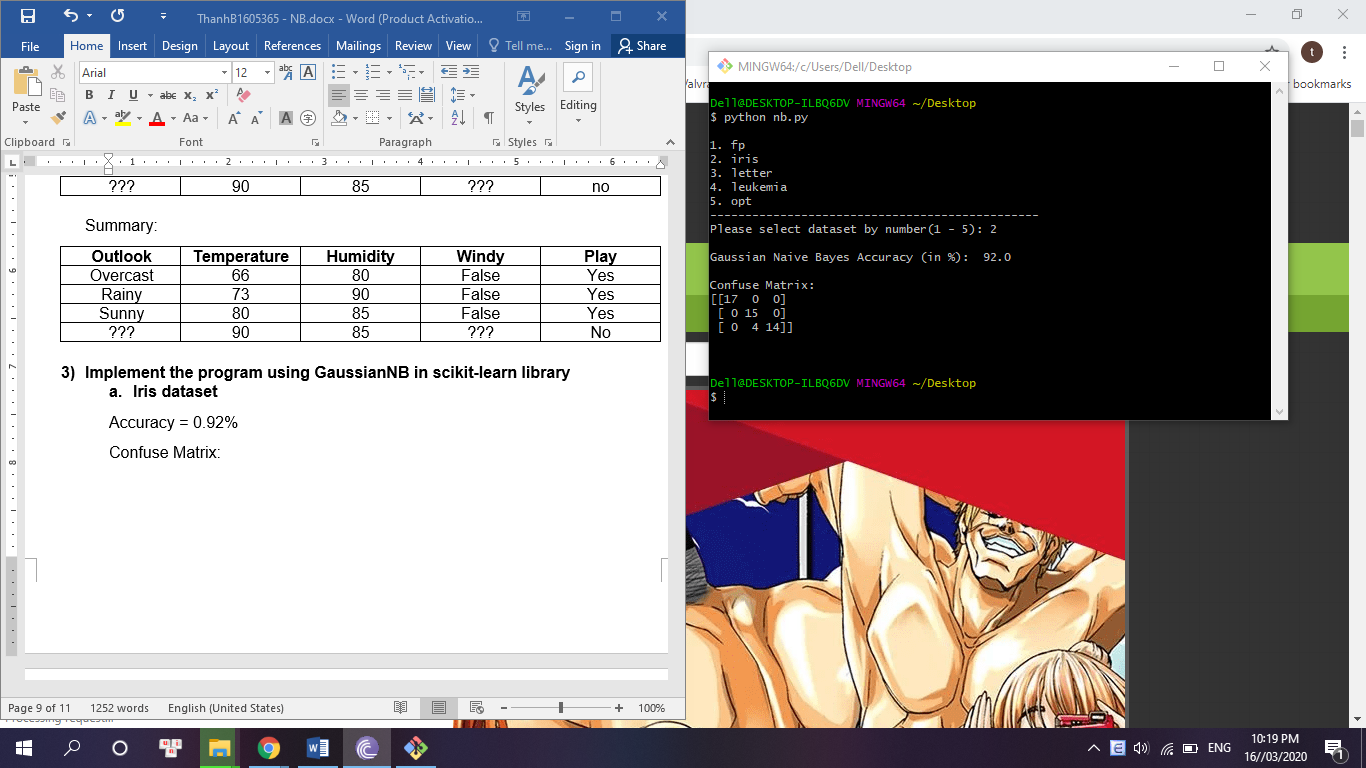
Summary:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Outlook** | **Temperature** | **Humidity** | **Windy** | **Play** |
| Overcast | 66 | 80 | False | Yes |
| Rainy | 73 | 90 | False | Yes |
| Sunny | 80 | 85 | False | Yes |
| ??? | 90 | 85 | ??? | No |

1. **Implement the program using GaussianNB in scikit-learn library**
   1. **Iris dataset**

Accuracy = 92%

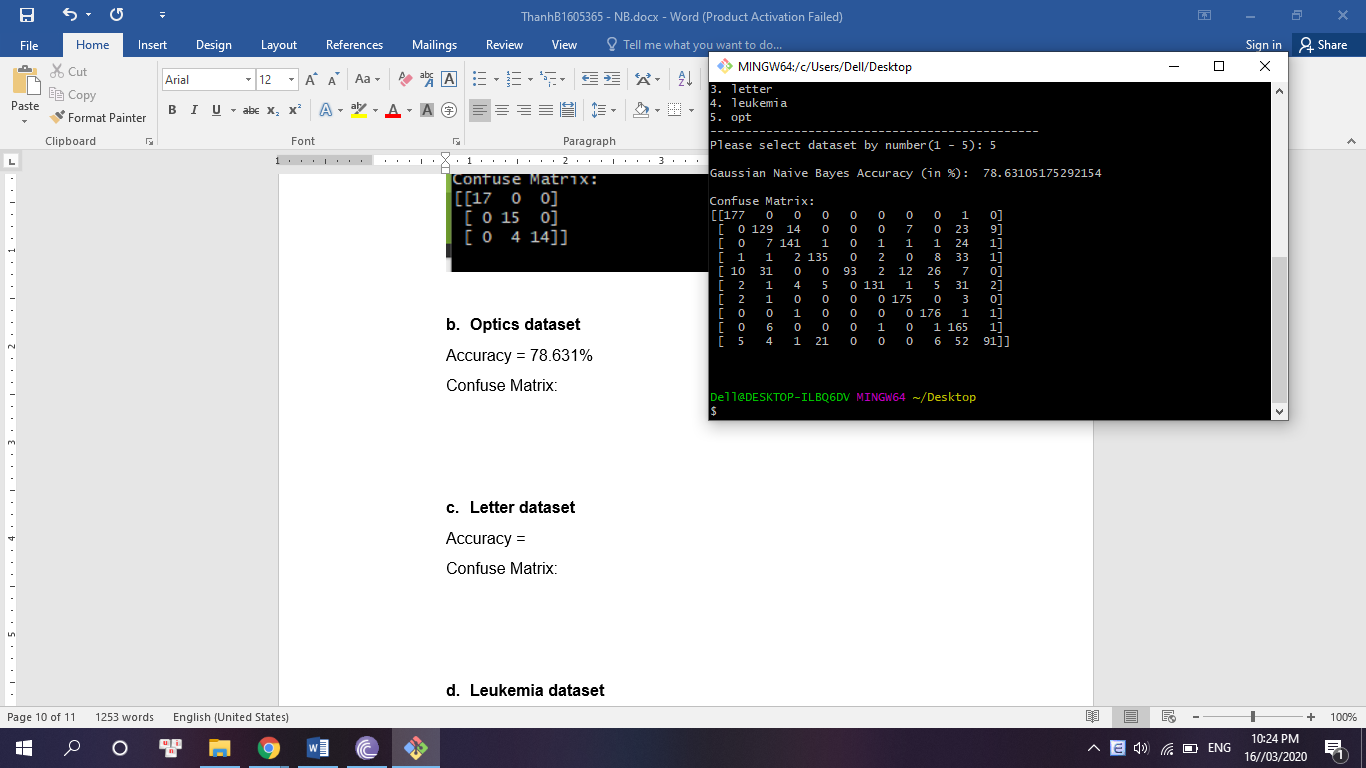
Confuse Matrix:



* 1. **Optics dataset**

Accuracy = 78.631%

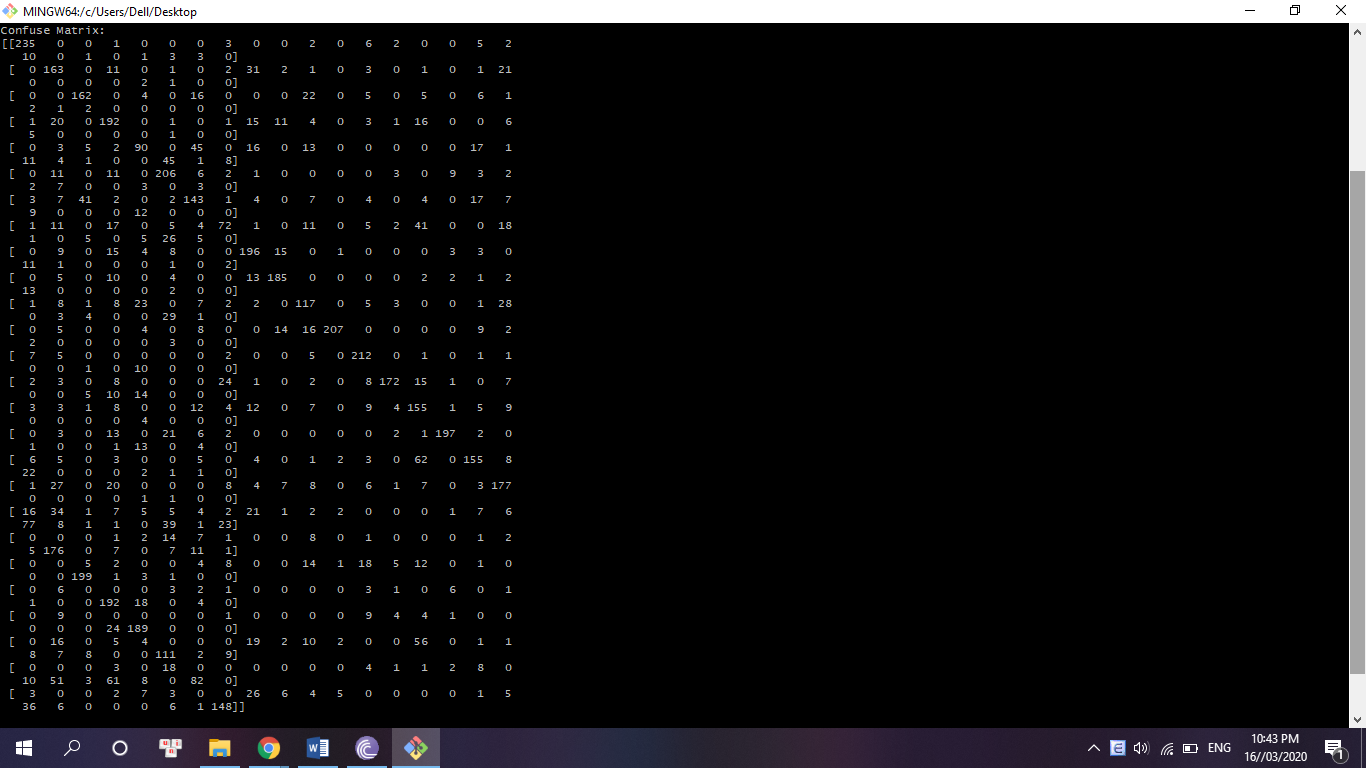
Confuse Matrix:



* 1. **Letter dataset**

Accuracy = 63.156

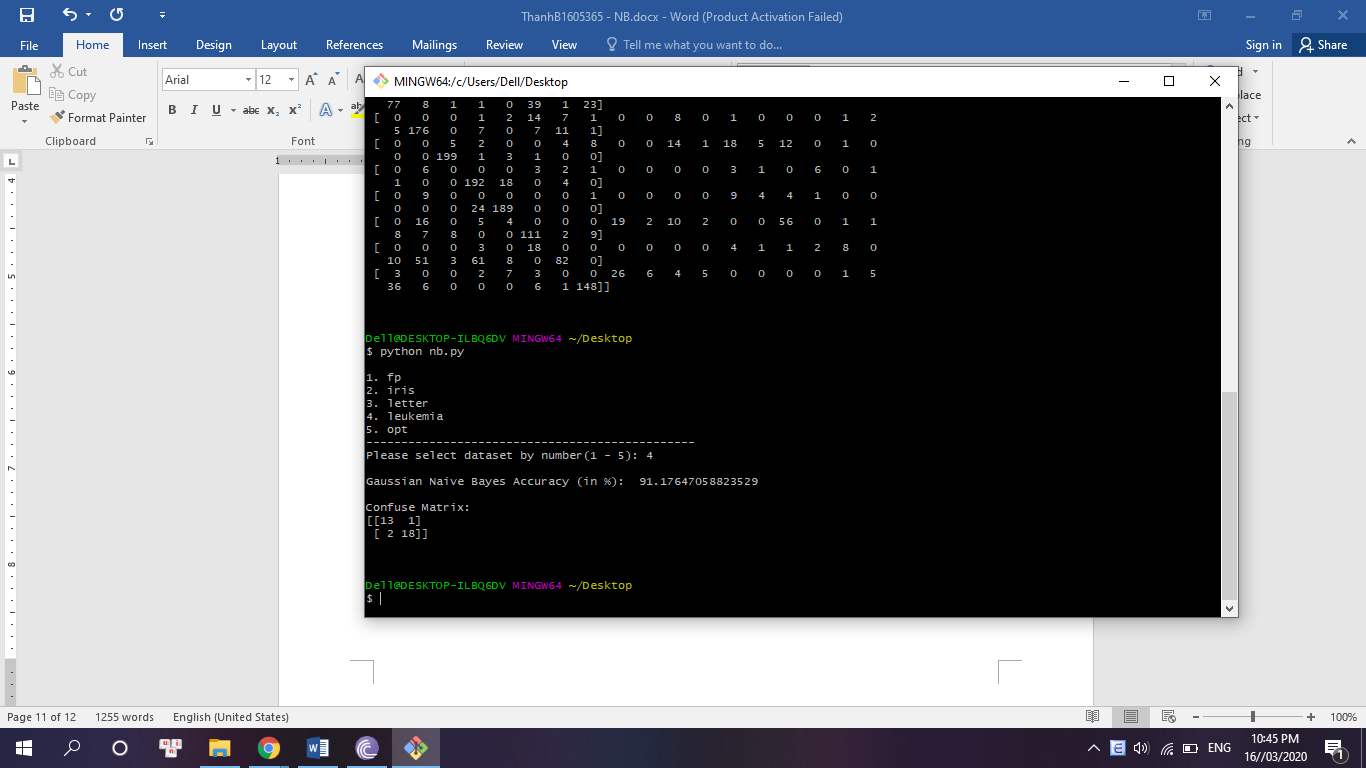
Confuse Matrix:



* 1. **Leukemia dataset**

Accuracy = 94.176

Confuse Matrix:



* 1. **Fp dataset**

Accuracy = 75

Confuse Matrix:

