

MSc in Software Engineering and Database Technologies

CT621 Artificial Intelligence

**Student Name:** Cristina Borges

**NUIG ID Number:** 20234955

**Course:** CT621 Artificial Intelligence

**Workshop No:** Week 7

**Assignments:** Week 7

**Date of Submission:** 04/07/2021

Contents

1.0 Assignment 2

1.1 Question 1 3

Summarising Information Provided 3

Assumptions 3

Calculation 3

2.1 Question 2 4

Summarise Data 4

New Instance 4

Arc Probabilities to be calculated 4

Classify new instance 5

Other Comments 5

Appendix 1 7

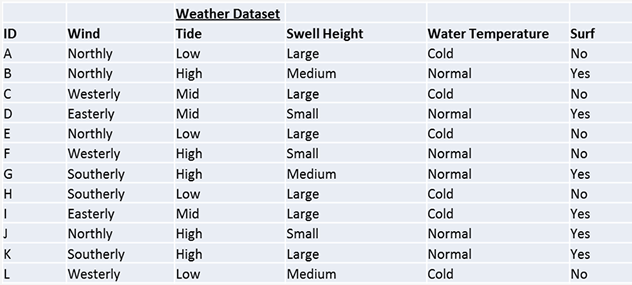
# Assignment

**Question 1.**

A company employs a total of 800 staff at two locations A and B. Location A has 200 female staff members and 100 male staff members. Location B has 300 female staff and 200 male staff. A male employee wins the employee of the month award. What is the probability that the employee is from location A? Please show all workings.

**Question 2.**

The image depicts the Surf dataset which is maintained by an avid surfer. He uses the database to help him/her decided whether to go surfing or not given specific conditions.



There are twelve examples in the dataset, each with an ID from A to L.

The dataset has five attributes.

1. Wind attribute can have one of four values: Northly, Westerly, Southerly, Easterly
2. Tide attribute can have one of three values Low, High, Mid
3. Swell Height can have one of three values: Small, Medium, Large
4. Water Temperature can have one of two values: Normal, Cold
5. Surf attribute can have one of two values: Yes, No

For each entry in the dataset each of the attributes will have a specific value and the output is either will surf "Yes" or will not surf "No". For example, in the first line of the dataset the Wind attribute is Northly, the Tide attribute is Low, the Swell Height is Large, and the Water Temperature is Cold. Consequently, the decision to surf is No.

Consider the following new data instance:

Wind = Easterly, Tide = High, Swell Height = Medium, Water Temperature = Cold: Surf = ?

Evaluate the probability that surf = yes and surf = no using a Naive Bayes Classifier for the new data instance. To solve this problem, you will need to calculate the arc probabilities for Surf = "Yes" and Surf = "No" for the Surf dataset and classify the new data instance. Workshop 7 Section 2 provides an example of this process applied to the Weather dataset.

## 1.1 Question 1

## Summarising Information Provided

Locations – A & B

Genders – Male & Female

Number of Male at Location A – 100

Number of Male at Location B – 200

Number of Female at Location A – 200

Number of Female at Location B – 300

P(employee gender = Male) = 300/800 =0.375

P(employee gender = female) = 500/800 n= 0.625

P(employee location = A) = 300/800 = 0.375

P(employee location = B) = 500/800 = 0.625

P(location A & Gender = Male) = 100/800 = 0.125

P(location A & Gender = Female) = 200/800 = 0.250

P(location B & Gender = Male) = 200/800 = 0.250

P(location B & Gender = Female) = 300/800 = 0.375

## Assumptions

Assume all employees equally likely to be employee of the month

Assume single winner

Assume award is memoryless i.e. any previous results do not impact current month

## Calculation

P(employee month Location = A I gender of winner = Male) = [P(location A & Gender = Male) ] / P(gender = Male)

= (100 / 800) / (300 / 800)

= (100 / 800) x (800 / 300)

= (100 / 300)

= 0.333

## 2.1 Question 2

## Summarise Data

I have summarised in the following table the weather dataset that has been provided

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ID | Wind | Tide | Swell Height | Water Temperature | Surf |
| A | Northly | Low | Large | Cold | No |
| B | Northly | High | Medium | Normal | Yes |
| C | Westerly | Mid | Large | Cold | No |
| D | Easterly | Mid | Small | Normal | Yes |
| E | Northly | Low | Large | Cold | No |
| F | Westerly | High | Small | Normal | No |
| G | Southerly | High | Medium | Normal | Yes |
| H | Southerly | Low | Large | Cold | No |
| I | Easterly | Mid | Large | Cold | Yes |
| J | Northly | High | Small | Normal | Yes |
| K | Southerly | High | Large | Normal | Yes |
| L | Westerly | Low | Medium | Cold | No |

## New Instance

Wind = Easterly

Tide = High

Swell Height = Medium

Water Temperature = Cold

Surf = ?

## Arc Probabilities to be calculated

**Wind**

P(Wind=Northly I Surf = Yes) = 2/6

P(Wind=Northly I Surf = No ) = 2/6

P(Wind=Southerly I Surf = Yes) = 2/6

P(Wind=Southerly I Surf = No ) = 1/6

P(Wind=Easterly I Surf = Yes) = 2/6

P(Wind=Easterly I Surf = No ) = 0/6

P(Wind=Westerly I Surf = Yes) = 0/6

P(Wind=Westerly I Surf = No ) = 3/6

**Tide**

P(Tide= High I Surf = Yes) = 4/6

P(Tide= High I Surf = No) = 1/6

P(Tide= Medium I Surf = Yes) = 2/6

P(Tide= Medium I Surf = No) = 1/6

P(Tide= Low I Surf = Yes) = 0/6

P(Tide= Low I Surf = No) = 4/6

P(Swell height = Large I Surf = Yes) = 2/6

P(Swell height = Large I Surf = No) = 4/6

P(Swell height = Medium I Surf = Yes) = 2/6

P(Swell height = Medium I Surf = No) = 1/6

P(Swell height = Small I Surf = Yes) = 2/6

P(Swell height = Small I Surf = No) = 1/6

P(Water temperature = Normal I Surf = Yes) = 5/6

P(Water temperature = Normal I Surf = No) = 1/6

P(Water temperature = Cold I Surf = Yes) = 1/6

P(Water temperature = Cold I Surf = No) = 5/6

## Classify new instance

Using the general formula, we then classify the probability of each possible class value given the values of the other attributed

P(Surf = Y) x P(Wind = Easterly) x P(Tide = High) x P(Swell Height = Medium) x P(Water temperature = Cold)

= 6/12 x 2/6 x 4/6 x 2/6 x 1/6 = 1 / 162

P(Surf = N) x P(Wind = Easterly) x P(Tide = High) x P(Swell Height = Medium) x P(Water temperature = Cold)

= 6/12 x 0/6 x 1/6 x 1/6 x 5/6 = 0

Next step is to normalise

P(Surf = Yes I Data) = (1/162) / (1/162 + 0) = 1

P(Surf = No I Data) = (0) / (1/162 + 0) = 0

## Other Comments

Naïve Bayes classifier result gives a measure of the certainty of each outcome. In this case, it suggests it is 100% certain that for this instance Surf = Yes

The size of the data set means that we have seen an example of one of the main disadvantages of the Naïve Bayes classifier in this question. As the training dataset did not have an instance of Wind = Easterly and Surf = No, it assigned a zero probability to this event meaning that the algorithm is unable to make a prediction.

Some possible solutions to this would be to either assign a low probability to the events with zero observations or remove that classifier entirely.

If we assigned P(Wind=Easterly I Surf = No ) = 1/6 – the lowest probability, our results would become:

P(Surf = Y) x P(Wind = Easterly) x P(Tide = High) x P(Swell Height = Medium) x P(Water temperature = Cold)

= 6/12 x 2/6 x 4/6 x 2/6 x 1/6 = 1 / 162

P(Surf = N) x P(Wind = Easterly) x P(Tide = High) x P(Swell Height = Medium) x P(Water temperature = Cold)

= 6/12 x 1/6 x 1/6 x 1/6 x 5/6 = 5/2592

Next step is to normalise again

P(Surf = Yes I Data) = (1/162) / (1/162 + 5/2592) = 0.7619

P(Surf = No I Data) = (5/2592) / (1/162 + 5/2592) = 0.2381

Another option is to remove Wind direction, in which case our calculation becomes:

P(Surf = Y) x ~~P(Wind = Easterly)~~ x P(Tide = High) x P(Swell Height = Medium) x P(Water temperature = Cold)

= 6/12 x ~~2/6~~ x 4/6 x 2/6 x 1/6 = 1 / 54

P(Surf = N) x ~~P(Wind = Easterly)~~ x P(Tide = High) x P(Swell Height = Medium) x P(Water temperature = Cold)

= 6/12 x ~~0/6~~ x 1/6 x 1/6 x 5/6 = 5/432

Next step is to normalise again

P(Surf = Yes I Data) = (1/54) / (1/54 + 5/432) = 0.6154

P(Surf = No I Data) = (5/432) / (1/54 + 5/432) = 0.3846

Using both of these alternative methods, we still arrive at the same outcome for our new instance i.e. surf = yes, however, the level of certainty is no longer equal to 1.

## Appendix 1

**Assignment 1**

**References**

CT621 Artificial Intelligence (2021) Uncertainty in AI Workshop 7 Section 1

Available at: NUIG Blackboard

Accessed 3rd July 2021

CT621 Artificial Intelligence (2021) Reasoning with Probabilities Workshop 7 Section 2

Available at: NUIG Blackboard

Accessed 3rd July 2021

Sharma N (2020) Understanding the Mathematics Behind Naïve Bayes

Available at: <https://heartbeat.fritz.ai/understanding-the-mathematics-behind-naive-bayes-ab6ee85f50d0?gi=c8a8ffc1ad40>

Accessed 3rd July 2021