Technical Information:

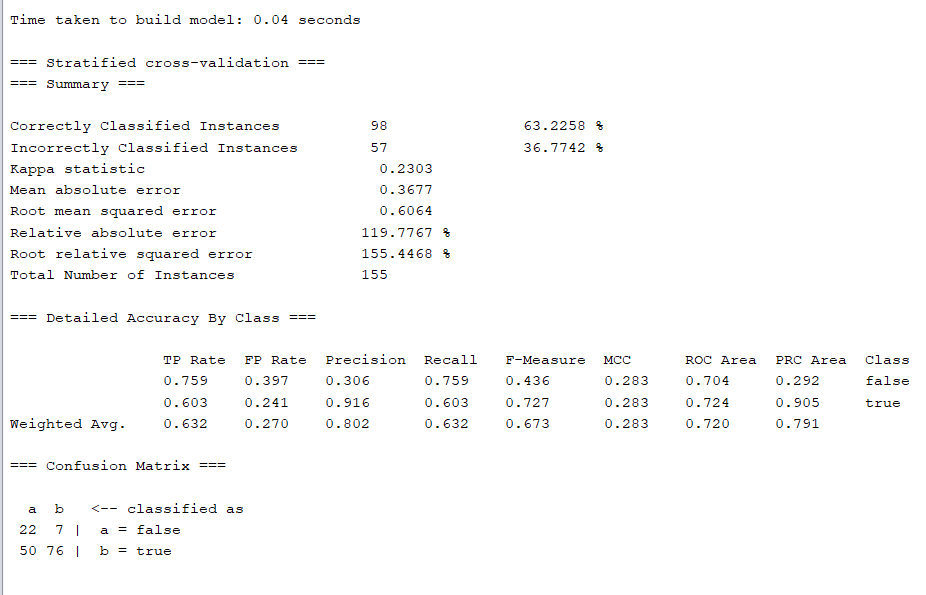
* Node used: M4.large()
* Number of Instances: 5
* Number of Jobs: 4
* Finish Time: 12 minutes
* DPMin: 5
* Ngarm corpus: biarcs00

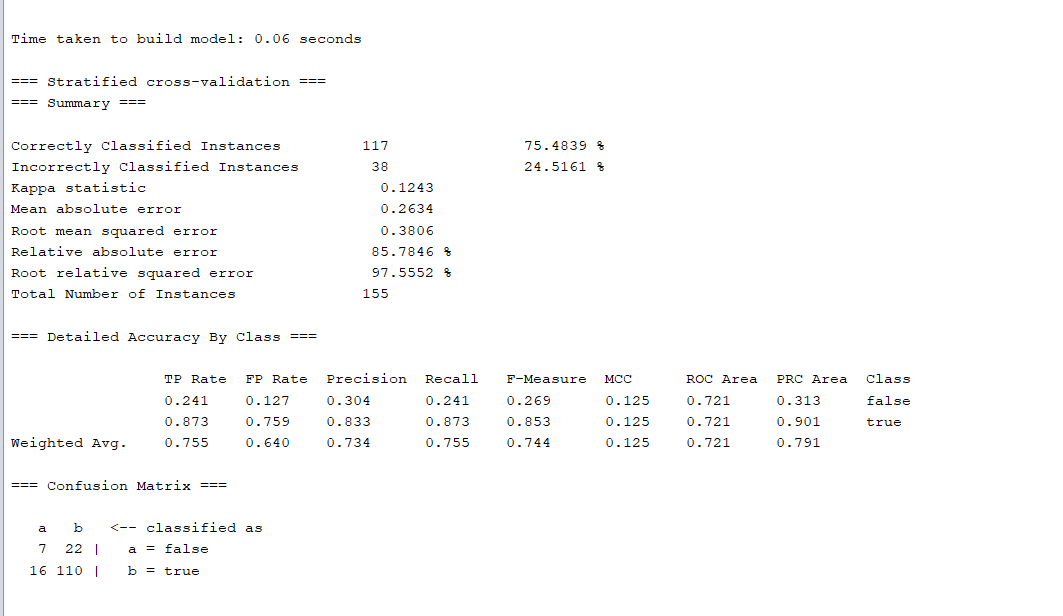
Communication:

|  |  |
| --- | --- |
| Step | : |
| Step1: | Map input records=16145663  Map output records=2010136  Map output bytes=54467524 |
| Step2: | Map input records=759801  Map output records=759801  Map output bytes=20671842 |
| Step3: | Map input records=12235  Map output records=12235  Map output bytes=475311 |
| Step4: | Map input records=14025  Map output records=14025  Map output bytes=360154 |

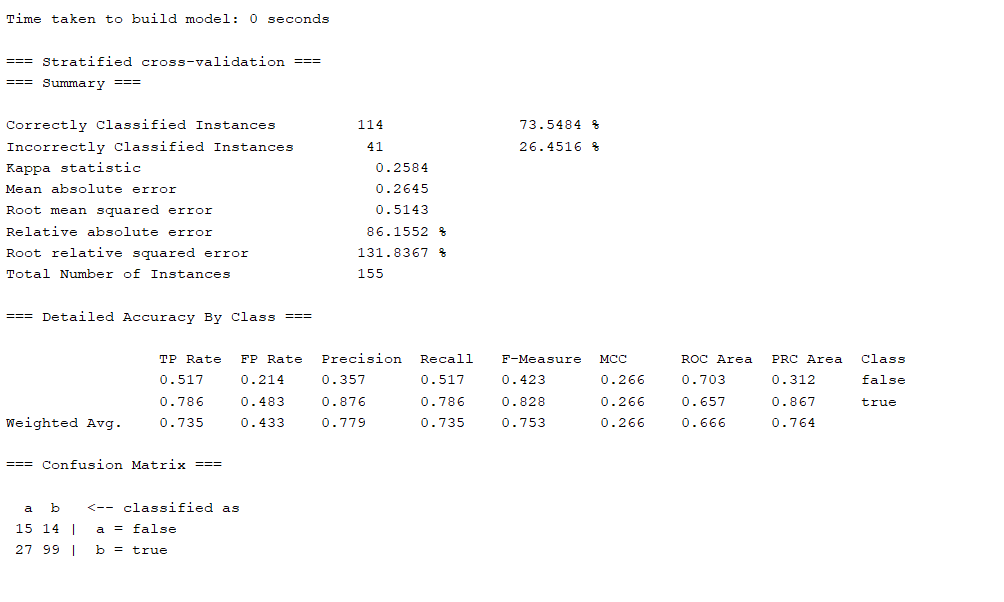
Results – 3 Algorithems:

**NaiveBayes:**



**Ba**yesNet:

**NaiveBayes Multinomial:**



Analysis(Based on NaïveBayes):

True-Positive(said true and its true):

|  |  |
| --- | --- |
| Noun1 | Noun2 |
| understand | abil |
| Think | abil |
| Power | abil |
| hypertrophi | abnorm |
| histori | studi |

False-Positive(said false and its false):

|  |  |
| --- | --- |
| Noun1 | Noun2 |
| garrison | abolitionist |
| equilibrium | abil |
| engin | 2 |
| countri | people |
| concept | abil |

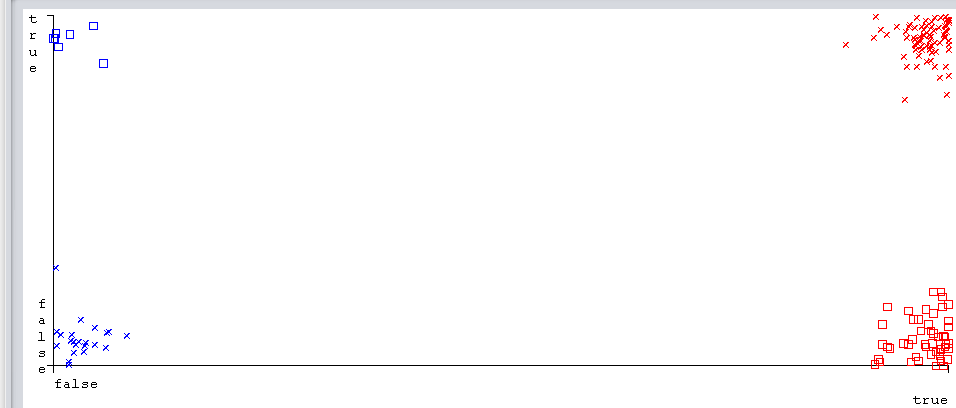
True-Negative(said true but its false):

|  |  |
| --- | --- |
| Noun1 | Noun2 |
| attent | abil |
| wisdom | abil |
| screen | a |
| potenti | abil |
| gift | abil |

False-Negative(said false but its true):

|  |  |
| --- | --- |
| Noun1 | Noun2 |
| anomali | abnorm |
| account | abil |
| aberr | abnorm |
| vision | abil |
| vallei | abbei |

**Yaxis – Predicted, Xaxis – Expected**

First explanation for the phenomena would be the algorithm itself – as seen in this report we’ve seen that NaiveBayes algorithm gave us the worst results between all 3 algorithm we tried.  
Another reason is the amount of data acquired from the EMA cluster -> we may be lack of data because we ran only on biarcs00(500mb).  
We’ve looked into the features vector of the false cases and we noticed what may cause the phenomena:

* One of the nouns dependency path with many occurrences has a big weight can cause True Negative:  
  For example: attent:abil -> N:partmod:GROUP:prep:ON:pobj:N – 310 occurrences
* Low noun occurrences counter on a any dp may cause False Negative:  
  For example:   
  anomali:abnorm: it’s occurences number is very low – 15-24 occurences on 5 dependency paths.