

Dept. of Computer Science and Software Engineering

**COMP 6721 : Applied Artificial Intelligence**

**Winter 2019-2020**

**Project 2**

Submitted To: **Dr. René Witte**

Submitted By:

Team **FL-G06**

|  |  |  |
| --- | --- | --- |
| **Name** | **Student ID** | **Email** |
| Mehrnaz Keshmirpour | 40063320 | keshmirpour@gmail.com |
| Naghmeh Shafiee Roudbari | 40129324 | naghmehshafiee@gmail.com |
| Jasmine Kaur | 40103309 | jasmine14concordia@gmail.com |

**Analysis**

* **Accuracy** = (number of instances correctly classified / Total number of instances) \* 100

= (733/800)\*100

= **91.63%**

|  |  |  |
| --- | --- | --- |
|  | **Instance in class Spam** | **Instance in class Ham** |
| **Model identified as spam** | 338 | 5 |
| **Model identified as ham** | 62 | 395 |

* **Precision(spam) =** number of instances that are in class Spam & model labeled as spam / total number of instances model labeled as Spam **=** 338/ (338+5) **= 0.985**

**Precision (ham) =** number of instances that are in class Ham & model labeled as ham / total number of instances model labeled as Ham = 395/ (395+62) **= 0.864**

* **Recall(spam) =** number of instances that are in class Spam & model labeled as spam / all instances in class Spam **=** 338/ (338+62) **= 0.845**

**Recall (ham) =** number of instances that are in class Ham & model labeled as ham / all instances in class Ham **=** 395/ (395+5) **= 0.987**

* **F-measure** (a weighted combination of precision & recall)

**F = (β2+1)\*PR / (β2P+R)**; β = 1 since, precision and recall have same importance

**=** (2\*0.845\*0.985) /(0.985+0.845) **=** 1.66465 / 1.83 **= 0.909**

Similarly, for class Ham, F = 1.705536 / 1.851 = **0.921**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Precision** | **Recall** | **F1-measure** |
| **SPAM class** | 0.985 | 0.845 | 0.909 |
| **HAM class** | 0.864 | 0.987 | 0.921 |

* **Confusion Matrix / Contingency Table**

|  |  |  |  |
| --- | --- | --- | --- |
| correct class  (that should have  been assigned) | classes assigned by the learner | | |
| **Ham** | **Spam** | **Total** |
| **Ham** | 395 | 5 | 400 |
| **Spam** | 62 | 338 | 400 |

This shows test dataset consist of equal distribution of spam and ham files. However, the model identifies 5 Ham files and 62 Spam files incorrectly i.e. it marked them to their opposite classes.

**References**