**INF8245E: Machine Learning | Assignment #3**

Louis Plessis (1933334)

14 November 2021

Acknowledgement: I have briefly discussed about this assignment with Marie-Christine Paré and Joseph Battesti

1. **Data Pre-Processing**

The vocabulary can be found in “medical\_text-vocab.txt”. The data can be found in “medical\_text-train.txt”, “medical\_text-test.txt” and “medical\_text-valid.txt”.

1. **Binary bag-of-words (BBoW)**
2. **Random classifier performance (F1-score):** XXX

**Majority-class classifier performance (F1-score):** XXX

1. Please see Jupyter Notebook.
2. **Hyper-parameters**

* ***Naïve Bayes***

XXX

* ***Decision Trees***

XXX

* ***Logistic regression***

XXX

* ***Linear SVM***

XXX

1. **F1-score**

|  |  |  |  |
| --- | --- | --- | --- |
| ***Classifier*** | **Training** | **Validation** | **Testing** |
| *Naïve Bayes* |  |  |  |
| *Decision Trees* |  |  |  |
| *Logistic Regression* |  |  |  |
| *Linear SVM* |  |  |  |

1. **Performance of classifiers**

XXX

1. **Frequency bag-of-words (FBoW)**
2. Please see Jupyter Notebook.
3. **Hyper-parameters**

* ***Naïve Bayes***

XXX

* ***Decision Trees***

XXX

* ***Logistic regression***

XXX

* ***Linear SVM***

XXX

1. **F1-score**

|  |  |  |  |
| --- | --- | --- | --- |
| ***Classifier*** | **Training** | **Validation** | **Testing** |
| *Naïve Bayes* |  |  |  |
| *Decision Trees* |  |  |  |
| *Logistic Regression* |  |  |  |
| *Linear SVM* |  |  |  |

1. **Performance of classifiers**

XXX

1. **FBoW vs BBoW performance**

XXX

1. **Best representation**

XXX