**Initial results (Using default configuration/without optimization)**

**Methodology**

1. Preprocessing

* Remove row data if na value exists on column translation\_assessment
* Remove column username,word\_origin,word\_translation
* Partition dataset using 70/30 proportion, random sampling with consideration of distribution on each target class

1. Model Creation by assign algorithms to learn training dataset
2. Performance assessment by assign generated model to predict test dataset
3. Compare prediction with real target value in test dataset

**In summary**

Naïve Bayes - 63.97%

SVM

* One-class classification using assumptions “best translations” criterion (target value in training\_data is colum translation\_assessment=5) :
* Multiclass classification – in progress, results not satisfying yet

Decision Tree

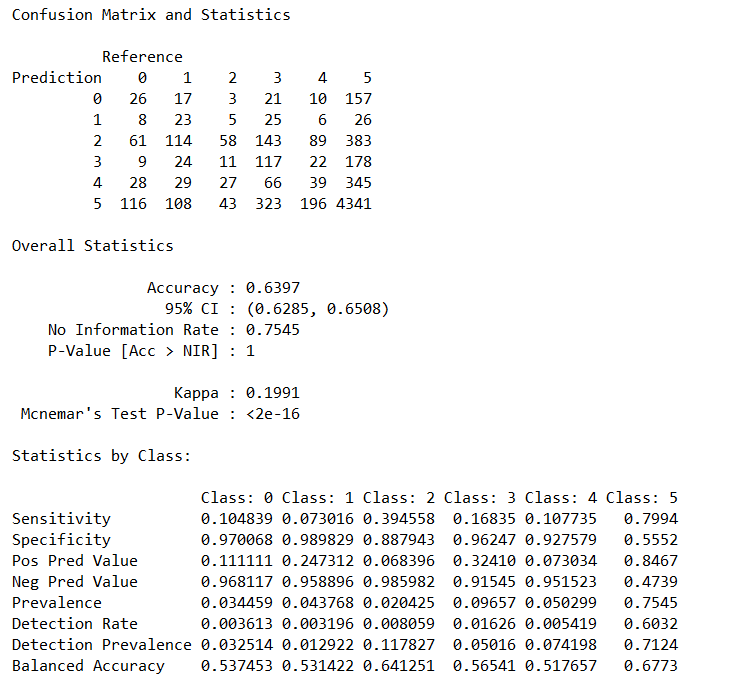
* CART
* C4.5
* C50

Random Forest - 88.86%

**To do list**

* Hyper-parameter optimization
* Multiclass SVM to achieve acceptable prediction accuracy

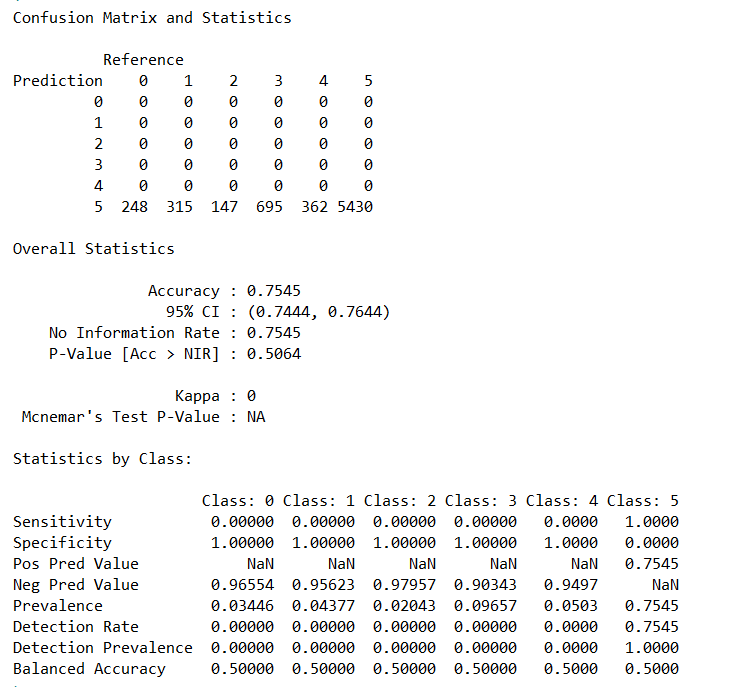
**Naïve bayes**

****

**SVM**

**Decision tree**

* **Cart**

****

* **C45**
* **C50**

**Random forest**

