Отчёт по домашнему заданию №9 (анализ тональности)

Для выполнения задания я выбрал классификаторы:

* Gaussian Naive Bayes
* Multinomial Naive Bayes
* Bernoulli Naive Bayes
* Support Vector Classification
* Linear Support Vector Classification

Файл с цитатами разбивается на список предложений и список меток.

Из этих списков удалил цитаты с метками отличными от +,-,0

Предложения векторизуются встроенным механизном векторизации в виде:

* частот(tf)
* tf-idf
* булевых векторов

Для каждого из видов векторов проведена классификация по всем методам.

Обучение проводится на trane (fit(вектор обработанных цитат, метки из trane))

Классификация на test (predict(вектор обработанных цитат))

Затем по результатам вычисляется макро и микро F-меры.

Используется встроенная функция f1\_score(выдача классификатора, метки из test).

К отчёту прилагается программа, которая реализует все описанные действия и выдаёт такие результаты:

F-measures

Gaussian Naive Bayes with tf:

Micro: 0.346617035253

Macro: 0.247976531197

Gaussian Naive Bayes with tf-idf:

Micro: 0.347492883731

Macro: 0.251440659131

Gaussian Naive Bayes with bin:

Micro: 0.33807751259

Macro: 0.221969286029

Bernulli Naive Bayes with tf:

Micro: 0.442741405737

Macro: 0.376174269901

Bernulli Naive Bayes with tf-idf:

Micro: 0.442741405737

Macro: 0.376174269901

Bernulli Naive Bayes with bin:

Micro: 0.442741405737

Macro: 0.376174269901

Multinominal Naive Bayes with tf:

Micro: 0.459601488942

Macro: 0.406142807964

Multinominal Naive Bayes with tf-idf:

Micro: 0.428070943727

Macro: 0.24787169952

Multinominal Naive Bayes with bin:

Micro: 0.427851981607

Macro: 0.341141397576

Support Vector Classification with tf:

Micro: 0.411320341581

Macro: 0.40926133936

Support Vector Classification with tf-idf:

Micro: 0.413400481717

Macro: 0.194990963078

Support Vector Classification with bin:

Micro: 0.441646595139

Macro: 0.299850470936

Linear SVC with tf:

Micro: 0.36763739873

Macro: 0.292245032415

Linear SVC with tf-idf:

Micro: 0.498795708342

Macro: 0.455395612673

Linear SVC with bin:

Micro: 0.469892708561

Macro: 0.411128008529