*Федеральное государственное бюджетное образовательное учреждение* *высшего профессионального образования*

|  |  |
| --- | --- |
| **Gerb-BMSTU_01** | ***«Московский государственный технический университет  имени Н.Э. Баумана»***  ***(МГТУ им. Н.Э. Баумана)*** |

КАФЕДРА

Системы обработки информации и управления

**Отчет**

**по рубежному контролю № 2**

# Дисциплина: Технологии машинного обучения

Студент гр. **РТ5-61Б \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ А.Г. Обрезкова**

(Подпись, дата) (И.О. Фамилия)

Преподаватель  **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Ю.Е. Гапанюк**

(Подпись, дата) (И.О. Фамилия)

Москва, 2020

**Вариант №1. Классификация текстов на основе методов наивного Байеса.**

**Задание**

Необходимо решить задачу классификации текстов на основе любого выбранного Вами датасета. Классификация может быть бинарной или многоклассовой. Целевой признак из выбранного Вами датасета может иметь любой физический смысл, примером является задача анализа тональности текста.

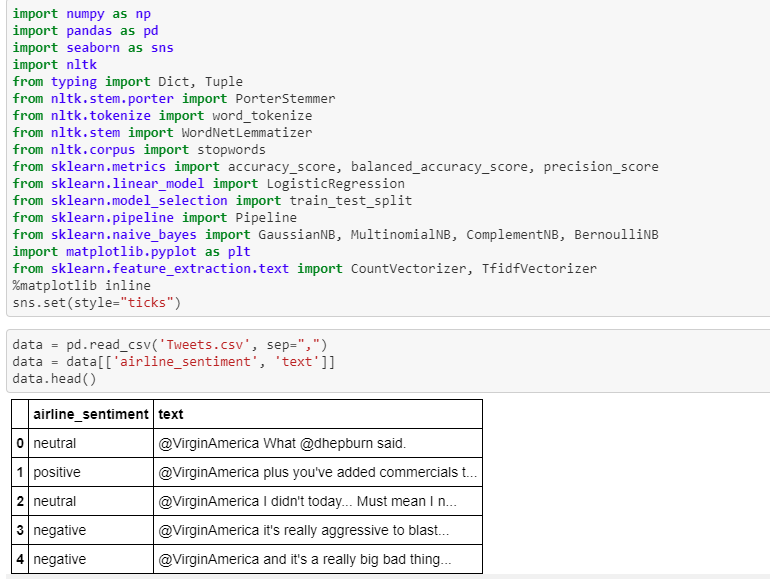
Необходимо сформировать признаки на основе CountVectorizer или TfidfVectorizer.

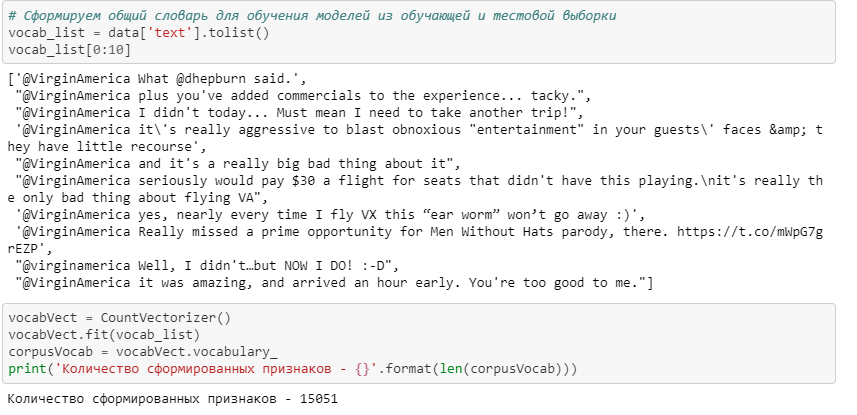
В качестве классификаторов необходимо использовать один из классификаторов, не относящихся к наивным Байесовским методам (например, LogisticRegression), а также Multinomial Naive Bayes (MNB), Complement Naive Bayes (CNB), Bernoulli Naive Bayes.

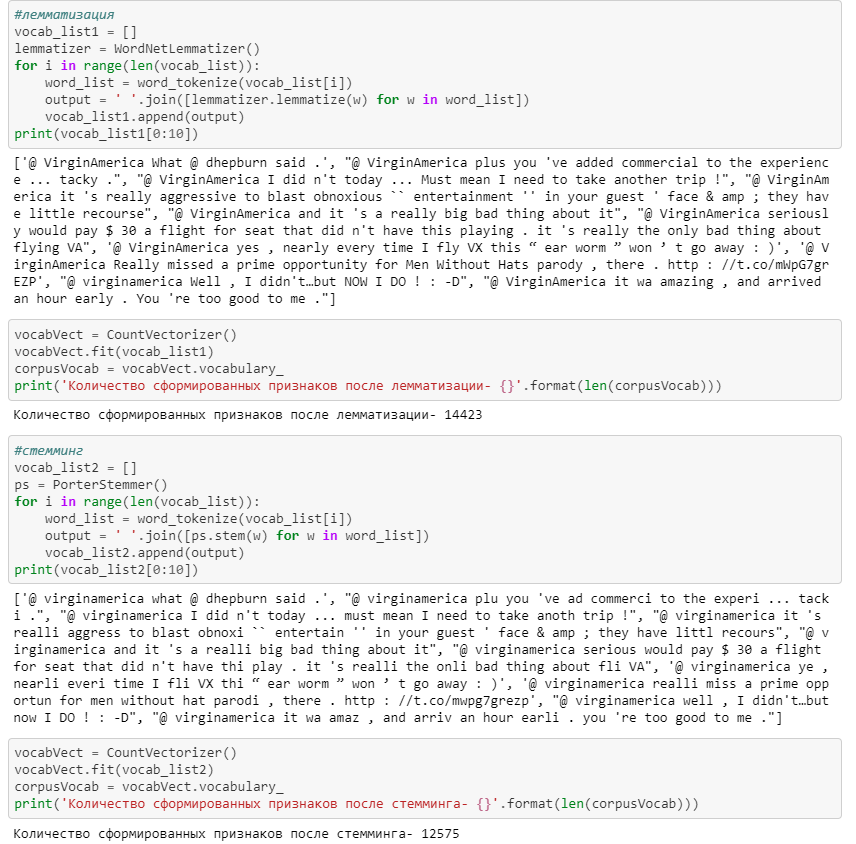
Для каждого метода необходимо оценить качество классификации с помощью хотя бы одной метрики качества классификации (например, Accuracy).

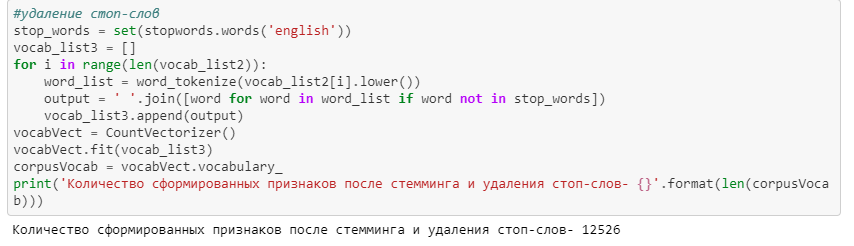
Сделате выводы о том, какой классификатор осуществляет более качественную классификацию на Вашем наборе данных.

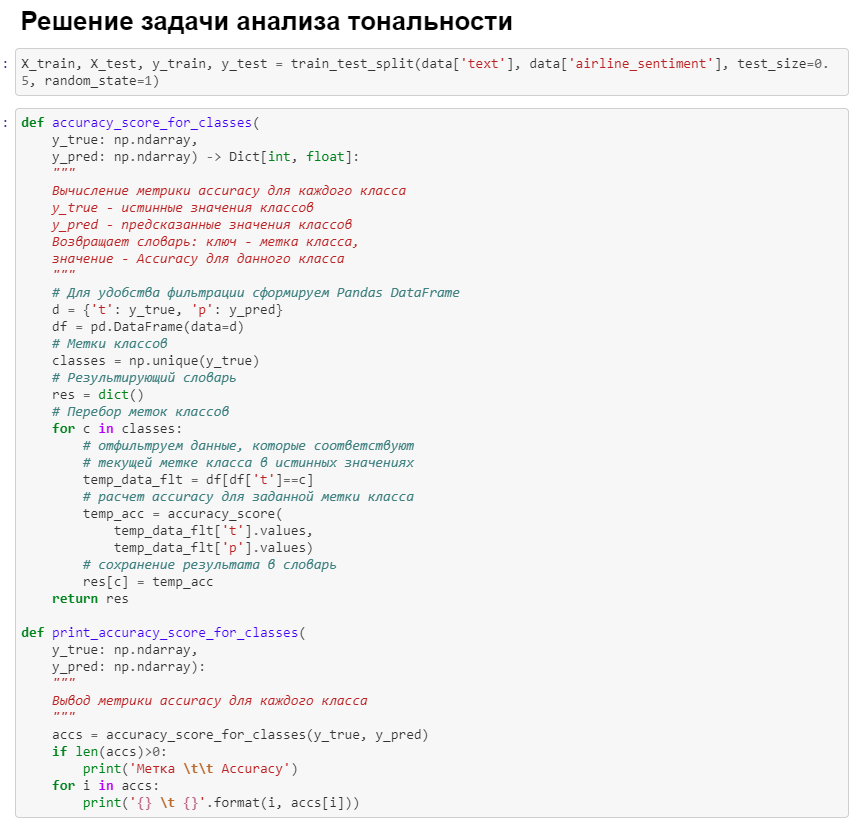
**Выполнение**

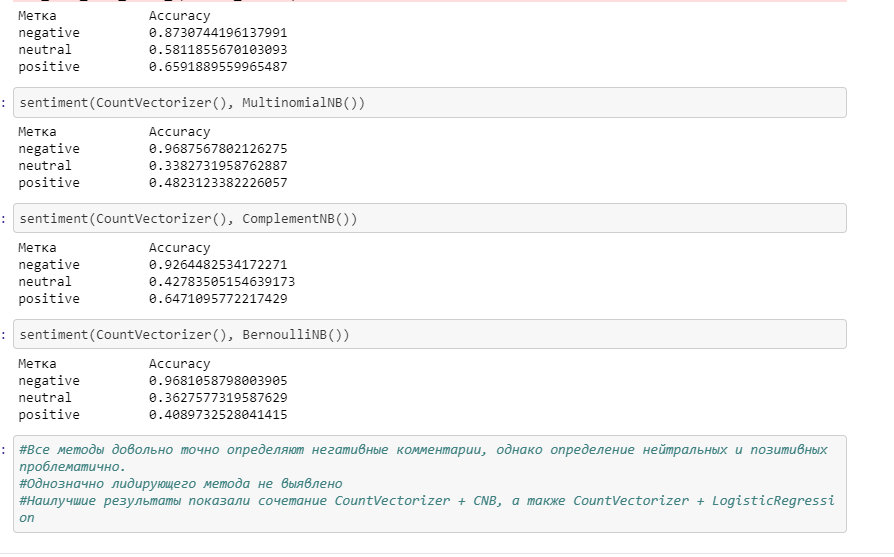
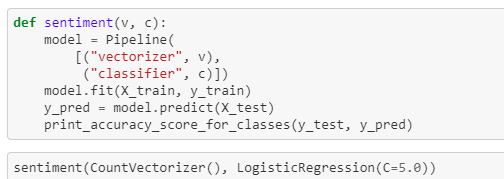
****

****

****

****

****

****