|  |  |
| --- | --- |
|  | **Министерство науки и высшего образования Российской Федерации Федеральное государственное бюджетное образовательное учреждение высшего образования**  **«Московский государственный технический университет имени Н.Э. Баумана**  **(национальный исследовательский университет)» (МГТУ им. Н.Э. Баумана)** |

ФАКУЛЬТЕТ \_Информатика и системы управления КАФЕДРА Системы обработки информации и управления

## Рубежный контроль №2

«Классификация текстов на основе методов наивного Байеса.» по курсу «Технологии машинного обучения»

Выполнил: Студент группы ИУ5Ц-81Б

Гаранин Антон

Проверил: Преподаватель кафедры ИУ5

Гапанюк Ю.Е.

**Москва 2020**

# Рубежный контроль №2

Тураев Глеб, ИУ5Ц-81Б, Вариант №5, Задача №1

# Задание

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

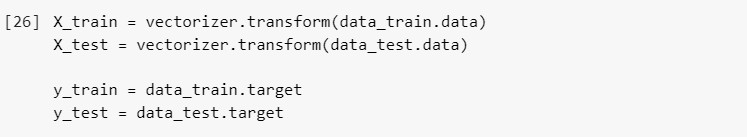
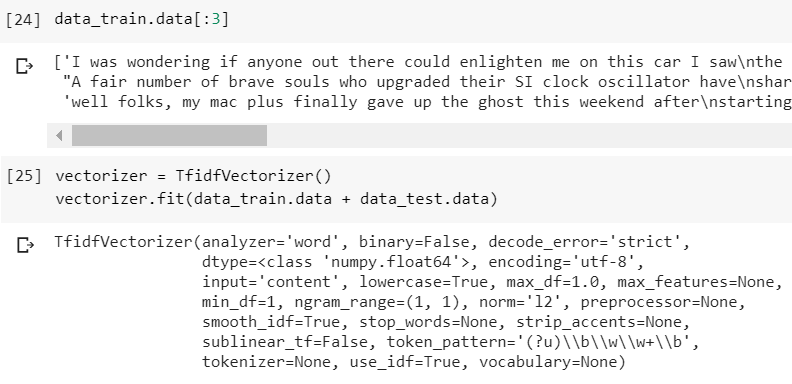
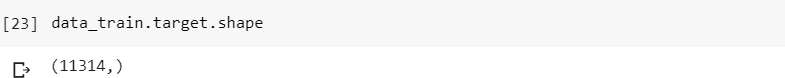
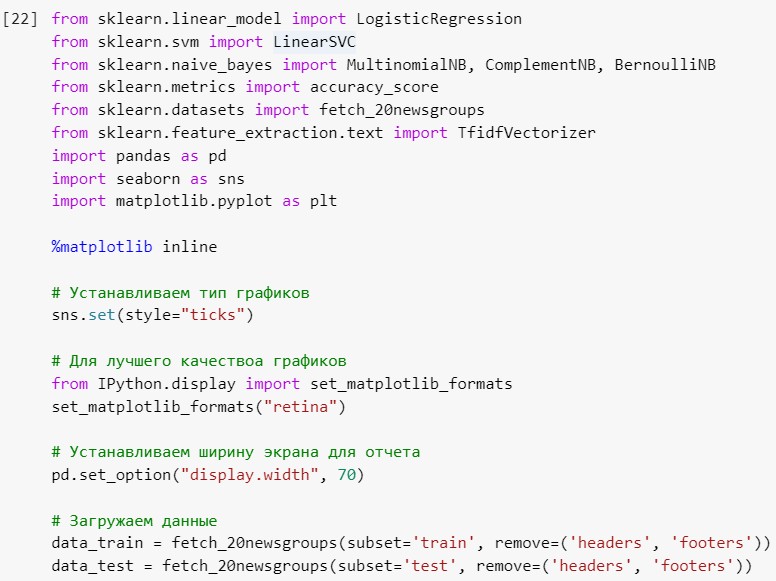
Необходимо сформировать признаки на основе CountVectorizer или TfidfVectorizer. В качестве классификаторов необходимо использовать два классификатора, не относящихся к наивным Байесовским методам (например, LogisticRegression, LinearSVC), а также Multinomial Naive Bayes (MNB), Complement Naive Bayes (CNB), Bernoulli Naive Bayes.

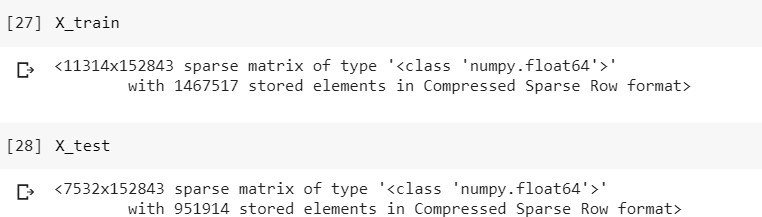
Для каждого метода необходимо оценить качество классификации с помощью хотя бы одной метрики качества классификации (например, Accuracy).

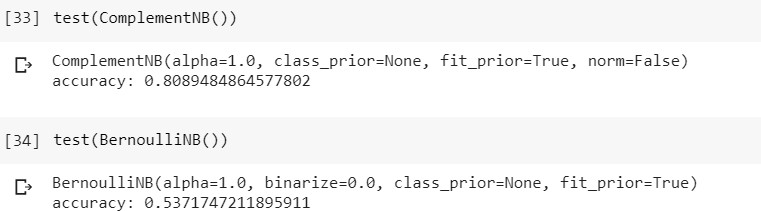
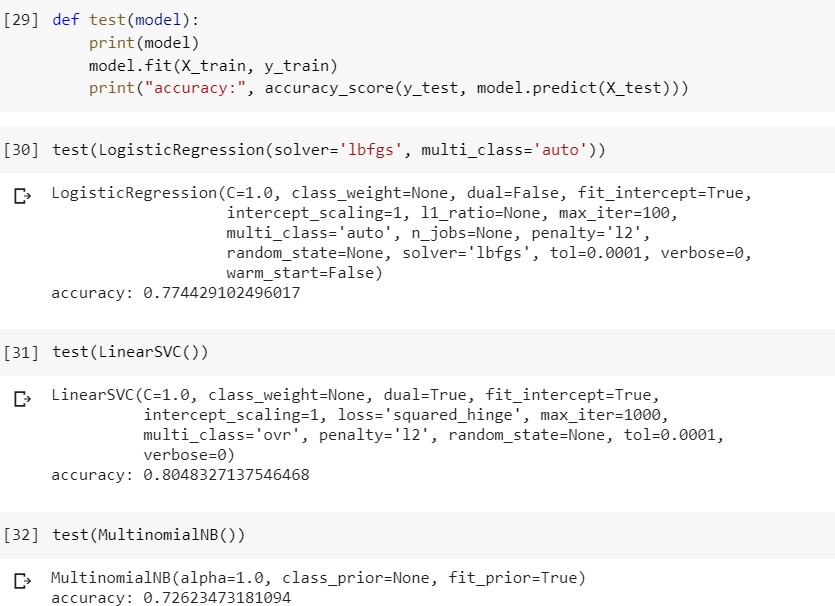
Сделайте выводы о том, какой классификатор осуществляет более качественную классификацию на Вашем наборе данных.

# Выполнение рубежного контроля

Подключим необходимые библиотеки и загрузим набор данных







**Вывод:** Видим, что метод Complement Naive Bayes лучше всего решает поставленную задачу многоклассовой классификации в условиях дисбаланса классов, но стоит отметить, что LinearSVC также показал отличный результат и показал такой же результат, как метод Complement Naive Bayes.