Московский авиационный институт
(Национальный исследовательский университет)

Факультет: «Информационные технологии и прикладная математика»
Кафедра: 806 «Вычислительная математика и программирование»
Дисциплина: «Искусственный интеллект»

Лабораторная работа № 1

Тема: Машинное обучение

Студент: Купцов Илья Владимирович
Группа: М80-307Б-18
Преподаватель: Ахмед Самир Халид
Дата:_____
Оценка:_____

Москва, 2021

1. Постановка задачи

Найти себе набор данных (датасет), для следующей лабораторной работы, и проанализировать его. Выявить проблемы набора данных, устранить их. Визуализировать зависимости, показать распределения некоторых признаков. Реализовать алгоритмы К ближайших соседа с использованием весов и Наивный Байесовский классификатор и сравнить с реализацией библиотеки sklearn.

2. Описание программы

Для данной лабораторной работы я нашел базу данных плохих и хороших отзывов о фильмах на сайте Kaggle. Датасет представляет из себя таблицу с двумя колоннами: отзыв и его качество: позитивный или негативный.

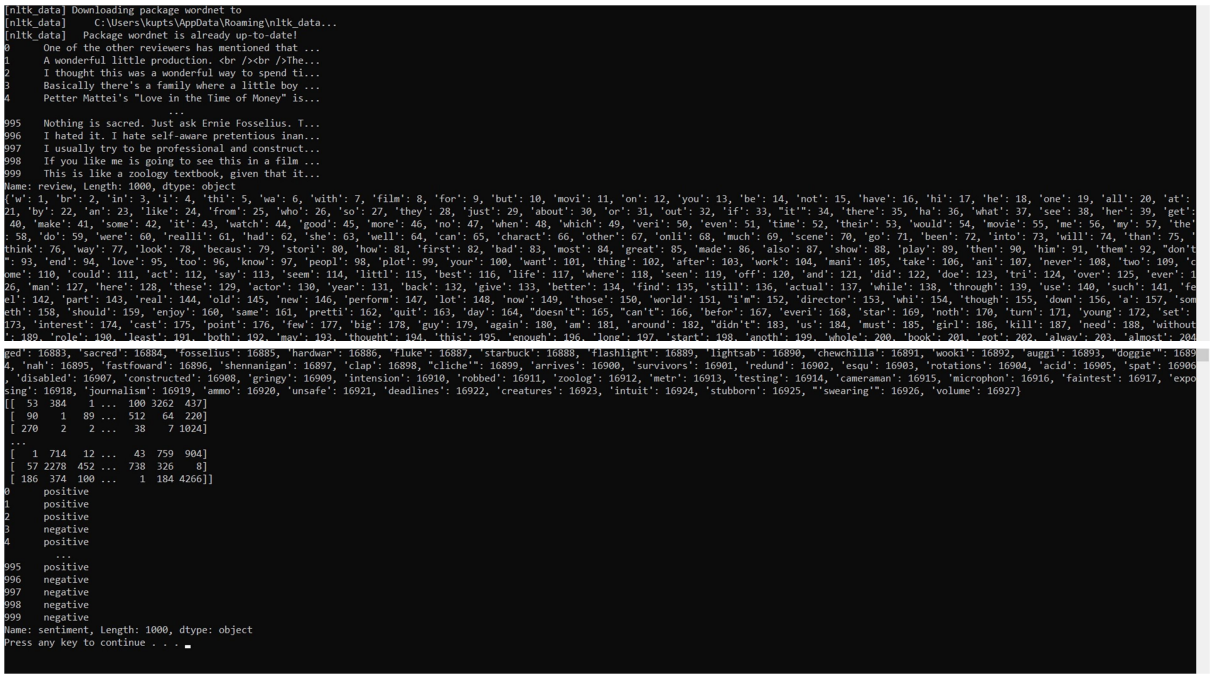
Я обработал эту базу данных для ее использования в машинном обучении. Для начала я проверил, полна ли база данных, проанализировав ее на пустые значения. К счастью, она сразу оказалась полной. После этого я отфильтровал весь текст, избавившись от лишних ненужных символов и преобразовав некоторые слова в более удобную форму. Затем я токенизировал все слова, представив их натуральными числами в некотором указанном мной диапазоне. Данный числовой формат уже пригоден для использования в машинном обучении.

Далее я реализовал алгоритм К ближайшего соседа, использовав стандартную документацию и язык Python. Чтобы проверить его корректность, я сравнил свою реализацию программы с реализацией модуля sklearn. Полученные результаты оказались схожи.

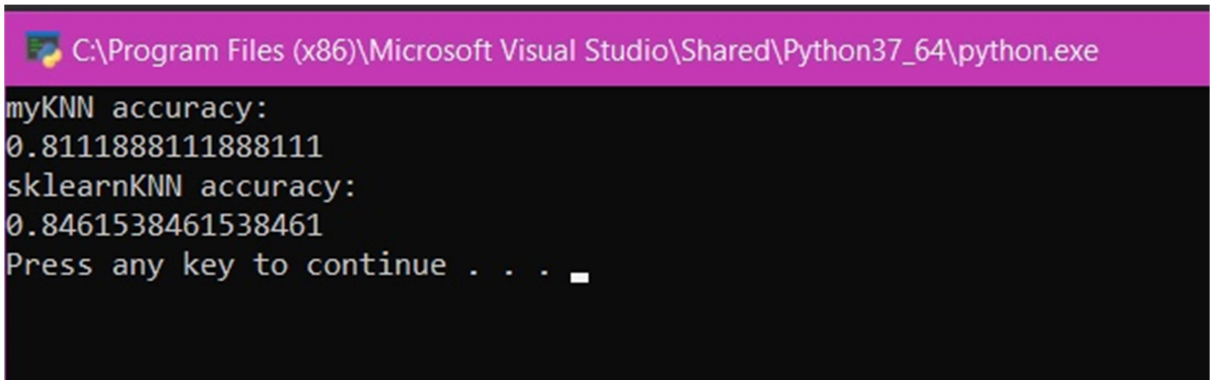
И, в заключение, я создал свой Байесовский классификатор на базе языка C++, аналогично, сравнив его работу с уже готовым Байесовским классификатором в библиотеке sklearn. Полученные результаты оказались схожи.

3. Результаты работы программы

1)Программа анализирующая данные:

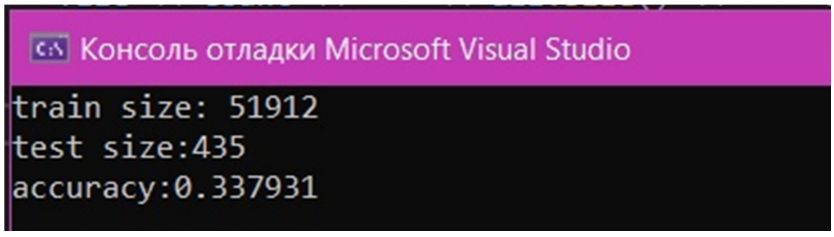


2) KNN

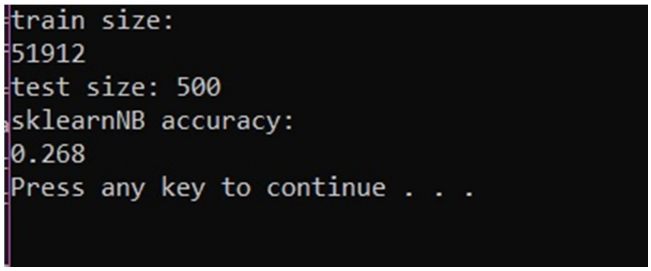


3) NB:

Моя программа:



Программа sklearn:



4. Вывод

В данной лабораторной работе я отыскал некоторую базу данных и, проанализировав ее, подготовил к алгоритмам машинного обучения. Также я реализовал и сами алгоритмы машинного обучения, такие как: К ближайших соседа и Байесовский классификатор. Сравнив мою реализацию с реализацией sklearn, я убедился в достоверной реализации своих программ.

В заключение хочу сказать, что полученные знания и опыт в данной лабораторной работе мне, несомненно, пригодятся, как в будущих студенческих проектах, так и далеко за пределами института.