**Оценка важности признаков (текстовые данные) 2.0**

Цель работы – провести бинарную классификацию для двух лююых классов текстовых сообщений из датасета fetch\_20newsgroups с последующей интерпретацией результатов с помощью метода explain\_instance из класса LimeTextExplainer.

В качестве примера выбраны следующие классы, которые содержатся в датасете: 1) rec.autos (об автомобилях) и 2) talk.politics.guns (о политике и оружии).

Для достижения поставленной цели было необходимо выполнить определенные действия. Для начала, необходимо установить и импортировать все нужные библиотеки, файлы .py с записанными функциями.

Далее происходит выбор классов из датасета. В нашем случае это rec.autos и talk.politics.guns

Затем мы импортируем записанную функцию load\_data из файла tool\_1 для загрузки данных. После этого происходит разделение на обучающую и тестовую выборки.

Дальше необходимо провести преобразовать тексты из наших классов в числовые данные с помощью CountVectorizer, чтобы модель могла их обработать.

Для классификации использована модель логистической регрессии. Она моделирует взаимосвязь между признаками и вероятностью принадлежности к определённому классу с помощью логистической функции.

Затем идет этап с проверкой качества модели. Он необходим для того, чтобы оценить, насколько эффективно она работает после обучения.

Наконец, код завершается визуализацией результатов работы модели, в котором показаны различия в лексике между двумя выбранными классами, к которым наша модель должна отнести тот или иной текст. Здесь нам пригодится импорт функции explain\_prediction из файла tool\_2. В idx ставим любое число от 0 до 284.

В заключение, можно сделать вывод, что модель успешно прошла свое обучение и способна отличать тексты двух разных тематик с хорошей точностью. Более того, визуализация результатов работы показывает, что модель способна проанализировать состав лексики и на основе анализа формулирует вероятность принадлежности текста к определенному классу.