

СЕССИЯ 2

Исходные файлы:

- | | |
|------------------|---------------------------|
| 1) Data.zip | (Предобработанные данные) |
| 2) Сессия 2.docx | (Инструкция ко 2 сессии) |

Результаты работы:

- | | |
|-------------------------------|------------------------------|
| 1) Data.zip | (Результаты сессии) |
| 2) Report.html + Report.ipynb | (Отчет о проделанной работе) |
| 3) Readme.txt | (Дополнительные комментарии) |

ВВЕДЕНИЕ

В этой сессии вы продолжаете работать с данными, подготовленными в предыдущей сессии. Предобработанные данные предстоит классифицировать и визуализировать зависимости данных. Какая-либо работа, обусловленная задачами предыдущей сессии, выполненная в ходе текущей, оцениваться не будет, поэтому проделывайте её только в случае необходимости.

ЗАДАНИЕ

2.1 Формирование дополнительных атрибутов

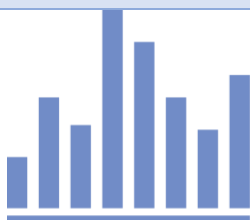
В представленных данных по некоторым адресам происходило несколько дорожно-транспортных происшествий. Для каждого такого адреса необходимо вычислить индекс, отражающий зависимость от количества происшествий, частоты их происхождения, тяжести последствий, доли по отношению ко всем происшествиям в регионе. Проанализируйте возможность определения наиболее опасных дорожных участков на основе вычисленного индекса.

2.2 Кластеризация набора данных

Выберите модель кластеризации данных. Задача кластеризации – определить группы дорожных участков с различным уровнем опасности, для которых могут потребоваться различные меры (например, успокоение трафика, сужение проезжей части, реорганизация движения, установка светофорных переходов, и т.д.). В результате кластеризации может получиться несколько уровней опасности, но в класс с максимальным уровнем, требующим принятия срочных решений, должны попасть не более 10% адресов от всего количества адресов в наборе по региону. Приведите обоснование выбора модели.

2.3 Разбиение набора данных

Разделите исходный набор данных на обучающую и тестирующую выборки оптимальным образом так, чтобы в обучающей и тестирующей выборках были данные по каждому региону.



Приведите обоснование разбиения

2.4 Подготовка отчета

Подготовьте отчет о проделанной работе по итогам сессии в котором будут представлены результаты, выводы и обоснования выбора по каждому разделу задания. В отчете также опишите содержимое результирующих файлов архива Data.zip

