

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 Разбиение набора данных

Разделите исходный набор данных на обучающую и тестирующую выборки оптимальным образом. Приведите обоснование разбиения

2.2 Визуализация зависимостей данных

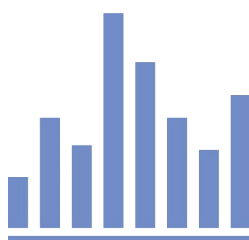
Используя программные средства, визуализируйте зависимости атрибутов в наборе данных. Визуализация должна отражать влияние атрибутов на определение классов – уровня опасности посещения страны для туриста. Произведите расчеты зависимостей по выбранным алгоритмам. Приведите интерпретацию полученным результатам

2.3 Классификация исходных компетенций

Выберите модель классификации данных по уровням опасности посещения для туристов. Приведите обоснование выбора модели.

2.4 Обучение

Проведите обучение выбранной модели на обучающей выборке. Протестируйте работу обученной модели на тестовой выборке. Определите показатели точности работы выбранной модели, сравните с остальными рассматриваемыми моделями.



2.5 Feature Engineering

Путём преобразования набора данных, добейтесь более точной работы выбранной модели. Опишите приемы генерации новых данных и результаты, к которым они привели, рассматривая все ранее определенные показатели точности

2.6 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

