

TMO_RK_1

ugapanyuk edited this page 15 days ago · 5 revisions

Рубежный контроль №1

Необходимо подготовить отчет по рубежному контролю и разместить его в Вашем репозитории. Вы можете использовать титульный лист, или в начале ноутбука в текстовой ячейке указать Ваши Ф.И.О. и группу.

Тема: Технологии разведочного анализа и обработки данных.

Варианты заданий.

Номер варианта	Номер задачи	Номер набора данных, указанного в задаче
1	1	1
2	1	2
3	1	3
4	1	4
5	1	5
6	1	6
7	1	7
8	1	8
9	2	1
10	2	2
11	2	3
12	2	4
13	2	5
14	2	6
15	2	7
16	2	8
17	3	1
18	3	2
19	3	3
20	3	4
21	3	5
22	3	6
23	3	7
24	3	8
25	4	1
26	4	2
27	4	3
28	4	4
29	4	5
30	4	6
31	4	7
32	4	8

Для студентов групп ИУ5-61Б, ИУ5-62Б, ИУ5-63Б, ИУ5-64Б, ИУ5-65Б, РТ5-61Б номер варианта = номер в списке группы.

Для студентов групп ИУ5Ц-81Б, ИУ5Ц-82Б, ИУ5Ц-83Б, ИУ5Ц-84Б номер варианта = 25 + номер в списке группы.

Для студентов группы ИУ5И-61Б номер варианта = 30 + номер в списке группы.

Дополнительные требования по группам:

- Для студентов групп ИУ5-61Б, ИУ5Ц-81Б, ИУ5И-61Б - для пары произвольных колонок данных построить график "Диаграмма рассеяния".
- Для студентов групп ИУ5-62Б, ИУ5Ц-82Б - для произвольной колонки данных построить гистограмму.
- Для студентов групп ИУ5-63Б, ИУ5Ц-83Б - для произвольной колонки данных построить график "Ящик с усами (boxplot)".
- Для студентов группы ИУ5-64Б, ИУ5Ц-84Б - для произвольной колонки данных построить график "Скрипичная диаграмма (violin plot)".
- Для студентов группы ИУ5-65Б - для набора данных построить "парные диаграммы".
- Для студентов группы РТ5-61Б - для пары произвольных колонок данных построить график "Jointplot".

Задача №1.

Для заданного набора данных проведите корреляционный анализ. В случае наличия пропусков в данных удалите строки или колонки, содержащие пропуски. Сделайте выводы о возможности построения моделей машинного обучения и о возможном вкладе признаков в модель.

Наборы данных:

- https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_boston.html#sklearn.datasets.load_boston
- https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_iris.html#sklearn.datasets.load_iris
- https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_wine.html#sklearn.datasets.load_wine
- <https://www.kaggle.com/carlolepelaars/toy-dataset>
- <https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset>
- <https://www.kaggle.com/mohansacharya/graduate-admissions> (файл Admission_Predict.csv)
- <https://www.kaggle.com/mohansacharya/graduate-admissions> (файл Admission_Predict_Ver1.1.csv)
- <https://www.kaggle.com/lava18/google-play-store-apps>

Задача №2.

Для заданного набора данных проведите обработку пропусков в данных для одного категориального и одного количественного признака. Какие способы обработки пропусков в данных для категориальных и количественных признаков Вы использовали? Какие признаки Вы будете использовать для дальнейшего построения моделей машинного обучения и почему?

Наборы данных:

- <https://www.kaggle.com/karangadiya/fifa19>
- <https://www.kaggle.com/fivethirtyeight/fivethirtyeight-comic-characters-dataset> (файл dc-wikia-data.csv)
- <https://www.kaggle.com/fivethirtyeight/fivethirtyeight-comic-characters-dataset> (файл marvel-wikia-data.csv)
- <https://www.kaggle.com/noriuk/us-education-datasets-unification-project> (файл states_all.csv)
- <https://www.kaggle.com/noriuk/us-education-datasets-unification-project> (файл states_all_extended.csv)
- <https://www.kaggle.com/rhuebner/human-resources-data-set>
- <https://www.kaggle.com/san-francisco/sf-restaurant-scores-lives-standard>
- <https://www.kaggle.com/mathan/fifa-2018-match-statistics>

Задача №3.

Для заданного набора данных произведите масштабирование данных (для одного признака) и преобразование категориальных признаков в количественные двумя способами (label encoding, one hot encoding) для одного признака. Какие методы Вы использовали для решения задачи и почему?

Наборы данных:

- <https://www.kaggle.com/karangadiya/fifa19>
- <https://www.kaggle.com/fivethirtyeight/fivethirtyeight-comic-characters-dataset> (файл dc-wikia-data.csv)
- <https://www.kaggle.com/fivethirtyeight/fivethirtyeight-comic-characters-dataset> (файл marvel-wikia-data.csv)
- <https://www.kaggle.com/noriuk/us-education-datasets-unification-project> (файл states_all.csv)
- <https://www.kaggle.com/noriuk/us-education-datasets-unification-project> (файл states_all_extended.csv)
- <https://www.kaggle.com/rhuebner/human-resources-data-set>
- <https://www.kaggle.com/san-francisco/sf-restaurant-scores-lives-standard>
- <https://www.kaggle.com/mathan/fifa-2018-match-statistics>

Задача №4.

Для заданного набора данных постройте основные графики, входящие в этап разведочного анализа данных. В случае наличия пропусков в данных удалите строки или колонки, содержащие пропуски. Какие графики Вы построили и почему? Какие выводы о наборе данных Вы можете сделать на основании построенных графиков?

Наборы данных:

- https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_boston.html#sklearn.datasets.load_boston
- https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_iris.html#sklearn.datasets.load_iris
- https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_wine.html#sklearn.datasets.load_wine
- <https://www.kaggle.com/carlolepelaars/toy-dataset>
- <https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset>
- <https://www.kaggle.com/mohansacharya/graduate-admissions> (файл Admission_Predict.csv)
- <https://www.kaggle.com/mohansacharya/graduate-admissions> (файл Admission_Predict_Ver1.1.csv)
- <https://www.kaggle.com/lava18/google-play-store-apps>

▼ Pages **20**

Find a Page...

▶ Home

▶ COURSE_MMO

▶ COURSE_TMO

▶ DSLIST

▶ DZ_MMO

▶ IDE

▶ LAB_MMO_DATA_STORY

▶ LAB_MMO_FEATURES

▶ LAB_MMO_FEATURES_2

▶ LAB_MMO_RECOM

▶ LAB_TMO_EDA_VISUALIZATION

▶ LAB_TMO_KNN

▶ LAB_TMO_MISSING

▶ LAB_TMO_TREES

▶ MMO_RK_1

Show 5 more pages...

Clone this wiki locally

https://github.com/ugapanyuk/ml_

