



Урок №47

Р

На прошлом уроке

Генеративное обучение

Л

Дипломная работа

Требования, чек-лист.

А

Практика

Пайплайны подготовки данных, решение задач

Н

Вопросы



Диплом

Варианты

1. Любая задача взятая из интернета по стандартному ML
2. Собственная задача

Оформление:

репозиторий на GitHub
содержащий файл README.md
содержащий файл с **теоретической частью**
содержащий ссылки на источники данных
или непосредственно источники данных
разработанные скрипты в отдельных папках
списки зависимостей

репозиторий должен клонироваться и запускаться локально
либо в Collab.



Диплом

Пример репозитория Француа Шолле

<https://github.com/keras-team/keras>



Диплом

Теоретическая часть

1. Постановка задачи
2. Ваш подход к решению
3. Описание применяемых алгоритмов и методик оценки качества моделей.
4. Интерпретация результатов
5. Выводы



Диплом

Задачи на выбор

1. ГРЗ - привести в порядок репозиторий
2. Генеративная текстовая сеть (сделать из той что была на прошлом уроке)
3. Разобраться со сложной моделью CV и реализовать подсчет людей/предметов в кадре)
4. Любая задача классического ML



Диплом

Где брать данные

<https://tproger.ru/translations/the-best-datasets-for-machine-learning-and-data-science/>

kaggle

Трансформер

Пишем собственный простейший трансформер

```
class some_transformer (BaseEstimator, TransformerMixin):  
  
    def __init__(self, remove_origin):  
        self.remove_origin = remove_origin  
  
    def fit(self, X, y = None):  
        return self  
  
    def transform(self, X, y = None):  
        # Transformation  
  
        if self.remove_origin:  
            # Remove original fields  
  
        return X
```



Трансформер

Пишем трансформаторы для всех переменных.

Объединяем трансформеры в пайплайн.

Выносим код в модуль подготовки данных.



Практика

Пишем проект диплома
на примере простой задачи классификации.

Датасет удовлетворенность клиентов авиакомпании

Применяемые модели

Логистическая регрессия

Дерево

SVM (разные ядра)

Случайный лес

Градиентный бустинг

Адаптивный бустинг

Выбираем лучшие модели - тюним параметры.

К стати как будем понимать какая модель лучше?

The image features a clean white background with abstract, flowing blue lines at the top and bottom, resembling waves or smoke. The word "THANKS" is centered in a blue, sans-serif font with a subtle gradient.

THANKS