

4. Анализ тональности отзывов

Не всегда в вашем распоряжении уже есть готовая выборка для построения модели и оценки ее качества, иногда ее нужно собрать самому. На этой неделе мы будем учиться парсить веб-страницы.

Задание, оцениваемое сокурсниками: Парсинг веб-страниц

Парсинг веб-страниц

Важная информация

Необходимо отправить это задание до завершения срока сдачи, июнь 25, 11:59 вечера PDT, поскольку другие учащиеся должны его оценить. Если поздно отправить задание, для его проверки может не оказаться сокурсников. Это затрудняет оценивание, а в некоторых случаях делает его невозможным. Чтобы избежать этих рисков, сдавайте задания вовремя.

1. [Инструкции](#)
2. [Моя работа](#)
3. [Обсуждения](#)

Инструкции

В этом задании вы потренируетесь парсить веб-страницы. Это умение пригодится вам, когда возникнет потребность самостоятельно собрать выборку для построения модели, будь то тексты или какая-то другая информация с общедоступных веб-страниц. В рамках данного проекта это случится уже в задании следующей недели.

Инструкция

1. Изучите блокнот с примерами парсинга страниц (в аттаче)
2. Поэкспериментируйте с парсингом любой страницы на ваш вкус, бегло просмотрите документацию библиотек `beautiful soup` и `requests`. При парсинге вам в любом случае потребуется смотреть на `html`-код страницы, чтобы понять, какие элементы вас интересуют.
3. Чтобы продемонстрировать, что вы разобрались, как использовать `requests` и `beautiful soup`, распарсите:
а) из статьи https://en.wikipedia.org/wiki/Bias-variance_tradeoff все заголовки верхнего уровня; б) со страницы https://en.wikipedia.org/wiki/Category:Machine_learning_algorithms названия всех статей в категории `Machine Learning Algorithms`

PageParsing.ipynb

Review criteriaменьше

В качестве ответа в этом задании вам нужно загрузить ноутбук. Позаботьтесь об оформлении, чтобы вашим сокурсникам при проверке было понятно, что является вашим результатом в пункте 3.