

# Домашнее задание:

Условия по выполнению домашнего задания по итогам модуля 6 (Введение в NLP):

Вам предоставлен датасет **wine\_reviews.csv** в котором содержится описания различных марок вин, их происхождение, название, описание, цена и оценка.

Так же вам предоставлен юпитер ноутбук, с базовыми набросками для помощи, чтобы решить задачу. Это не теймплейт с теми методами, которые вы должны обязательно использовать, это лишь набросок как вы можете поступить к решению проблемы. Ищите свой путь, если не устраивает подход к решению, экспериментируйте, может вы найдете что-то намного лучшее.

Для реализации задачи вам необходимы 4 основных признака: **decsription, points, title, variety (+/- cost** по желанию).

Рекомендации по выполнению: используйте уже готовые натренированные модели, желательно фреймворк [spacy](#) либо [gensim](#).

Итого ваша задача - построить решение и выполнить следующие задачи (в jupyter notebook в качестве исследования):

1. Задача **summarization** описания вина:  
В качестве инпута у вас описание вина, на выходе должны получить сокращенную версию данного описания для целей упрощения задачи. Методы вы знаете, обученные модели есть в свободном доступе. Если хотите показать себя, обучите рекуррентную нейронную сеть с подходом seq2seq для задачи сокращения текста (либо трансформер, на ваш выбор). За это +1 балл. Но это займет у вас много времени, поэтому на ваше усмотрение.
2. Как только вы получили саммари вина, постройте **doc2vec эмбединг (либо другой подход который вы считаете правильным) по саммари** для следующих целей: Реализуйте поиск похожих вин, по описанию, на входе у вас описание, на выходе у вас список вин (top-5), которые плюс минус подходят для пользователя, с выводом должна присутствовать оценка для данной категории вина и **по желанию средняя стоимость за бутылку**.
3. После того, как вы реализовали прототипирование и исследование в юпитер ноутбуке, вам необходимо реализовать модуль (загрузка и обработка текста, тренировка модели? и прочее).  
  
Реализуйте взаимодействие с командной строкой в следующем виде:  
  
а) Получение саммари, на входе пользователь вводит небольшое описание вина а на выходе получает саммари  
б) Получение рекомендаций, на входе пользователь вводит небольшое описание вина, на выходе получает топ-5 вин подходящих под описание с тайтлом, variety и points.

Помимо текущего задания, реализуйте еще те задания, которые связаны с информационным поиском (были на занятии).

Критерии оценки, помимо правильности выполнения задания, будет оцениваться также логика построения модуля (архитектура проекта, гранулярные скрипты отвечающие за логику кода, стиль и оформления кода, правильность и логика написания классов, логика и правильность консольной программы).

На выполнение отводится **10 дней, то есть с 31.08.2023 по 9.09.2023 (включительно)**.