

Правила выполнения курсовой работы

В качестве курсовой работы может быть выполнено одно из следующих заданий.

- 1) Решение задачи машинного обучения на реальном датасете. Источником таких датасетов могут служить реальные данные, данные из соревнований Kaggle и других подобных платформ.
- 2) Реализация алгоритма машинного обучения из статьи, для которого нет референсной или общеиспользуемой реализации на Python. Наличие какой-то реализации не является запретом на выполнение проекта, но в этом случае должно быть проведено сравнение вашей реализации и приведенной.
- 3) Какой-то свой проект, если его обсудить с преподавателем.

Работу разрешено выполнять в группах до трех человек. В случае более одного участника в группе для каждого участка работы должен быть человек, который способен внятно объяснить, что происходит в этом участке и наоборот, для каждого участника должен быть существенный фрагмент, за который он ответственный.

Сроки

1. Защита работ будет проходить в конце декабря.
2. Срок по выбору проекта (датасета, статьи,...) и объединения в группы — конец ноября.

Подача заявки

Напишите письмо преподавателю, в котором кратко опишите, что планируете сделать (датасет, статья, какие-то идеи по методам) и состав участников.

Требования к работе в виде решения задачи машинного обучения

1. Датасет должен быть достаточно большим (не менее 10 тыс. примеров)
2. Количество опробованных методов решения задачи должно быть не меньше $1.5N$, где N — число человек, выполняющих работу.
3. Для задачи должен быть выбран метод измерения качества модели и объяснена его применимость в данном случае.
4. Должен быть проведен визуальный анализ данных.
5. Для каждого метода должна быть выполнена предобработка данных или показана, что она не требуется.
6. Для каждого метода должны быть подобраны гиперпараметры.
7. Для задачи регрессии должен быть проведен анализ целевой переменной и применено преобразование или объяснено, почему оно не требуется.
8. Решение должно быть воспроизводимым.

Требования к работе в виде реализации алгоритма

1. Должны быть разобраны теоретические предположения, стоящие за алгоритмом
2. Должна быть показана корректность реализации (тесты на простых случаях)
3. Должен быть проведен теоретический анализ вычислительной сложности алгоритма и требуемой памяти.
4. Должно быть сравнение с другими методами на задаче и каком-то датасете.
5. Должно быть измерено время работы и потребляемая память в зависимости от размера датасета (можно брать подвыборки одного достаточно большого датасета).