Машинное обучение – это отрасль ИИ, которая фокусируется на использовании математических данных и алгоритмов для выявления закономерностей, и адаптации собственного алгоритма путем изменений.

Классификация – получение категориального ответа на основе набора признаков.

Категориальное значение – это перечисление ограниченного числа данных.

Кластеризация – это распределение данных на группы по схожим свойствам.

Регрессия – прогноз на основе выборки различных данных.

Понижение размерности данных – это сведения большего числа свойств/записей к меньшему.

Поиск аномальных значений – это отделение аномальных данных от стандартных случаев.

Обучение с учителем – имеются исходные данные и правильные решения этих данных, и требуется найти закономерность между ними.

Обучение с частичным привлечением учителя – имеются исходные данные и правильные решения для некоторых данных.

Обучение без учителя – на основе исходных данных необходимо сгруппировать объекты.

Обучение с подкреплением – на основе исходных данных (в среде) алгоритм машинного обучения делает решение и в зависимости от выбранного решения получает обратную связь.

В задачах классификации применяется обучение с учителем.

В задачах кластеризации применяется обучение без учителя.

В задачах регрессии применяется обучение с учителем.

В задачах понижения размерности применяется обучение без учителя.

DataSet – это набор исследуемых данных.

Первый вид датасета – табличный вид.

Подвиды простой записи:

* Транзакционные
* Матрица данных
* Матрица разрежонных данных

Второй тип – графы

Третий тип – упорядоченные данные

* Последовательные данные
* Временной ряд
* Пространственные

Признак (Feature-X) – свойства или же характеристики набора входных данных.

Метка (Y) – то, что выдает машина.

Целевая переменная (Yф) – фактические данные.

* Качественная переменная – задача кластеризации
* Полиномиальная - от трех и более
* Бинарная – два вида ответа
* Количественная – задача регресии
* Дискретная – есть диапазон
* Непрерывная – нет диапазона

Целевая переменная это признак который предстоит прогнозировать модели машинного обучения.

Процесс обучения модели – подача данных для алгоритма машинного обучения для процесса в ходе которого алгоритм должен найти закономерность исходных данных

Характеристики DataSet

* Размерность
* Разреженность