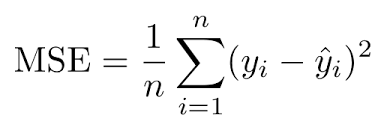
Общая структура систем интеллектуального анализа данных

Процесс разработки систем ML

**Машинное обучение с учителем:**

Принимает на вход признаками и ответами для него и принимая некоторую форма (модель) решения для него пытается его решать.

**Процесс разработки:**

1. Начать без машинного обучения (но это не точно);
2. Формализация задачи (Чёткая описание входящих и выходящих данных, от этого зависит модели ML (определить X Y и требования \*на сколько мы можем ошибаться\* к ним));
3. Метрики качества и приемки (берут X и Y и возвращает цифрки)
   1. Accuracy (X, Y, f) =
   2. 
   3. Метрики приёмки, метрики с которыми заказчики принимают модели.
4. Сбор бенчмарка (набор данных, на котором в процессе разработки мы будем считать метрики и смотрим качество, наша задача получить хороший результат) \*они маленькие 100-1000 примеров\*
5. Сбор обучающей выборки (100 - ∞)
   1. Разработка окружения для модели (го в 11 этап)
6. Очистка (поиск выбросов и шумов) и трансформация данных;
7. Разметка данных (если много, то краудсоурсинг)
8. Обучение и выбор модели (оценка модели)
9. Оценка качества (по метрики качества) \*если плохо, смотри 8й этап и го по новой (доверительный интервал)\*
10. Валидация (по бизнесовой метрики у заказчика), если не ок, идём 8
11. Деплой (мониторинг метрик качества, также мониторинг на данные, на сколько пользователи довольны нашей системой)

Краудсорсинг:

- человеческий фактор

- требования к квалификации

- байес по деньгам

- боты

Методы фикса ошибок:

- экзамен

- golden set (honeypot) \*даём кейсы на которые знаем ответы и смотрим, как он отвечает\*

- overtap

- алгоритм Девида Спина (оценивает акссесорам вес)