**Data Science проект (золотодобывающая отрасль).**

**Компания-заказчик: "Цифра" (zyfra.com)**

**Задача:** подготовить прототип модели машинного обучения.

Компания разрабатывает решения для эффективной работы промышленных предприятий.

Модель должна предсказать коэффициент восстановления золота из золотосодержащей руды.

**Данные:** параметры добычи и очистки.

**Цель исследования:** с помощью модели оптимизировать производство, чтобы не запускать предприятие с убыточными характеристиками.

**Ход исследования:**

Подготовка данных;

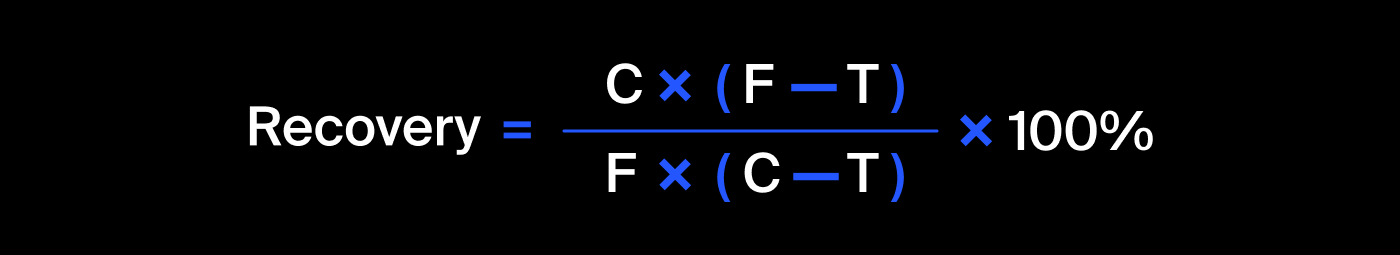
Исследовательский анализ данных;

Построение и обучение модели.

Расчёт эффективности

**Необходимо**: смоделировать процесс восстановления золота из золотосодержащей руды.

Эффективность обогащения рассчитывается по формуле



где:

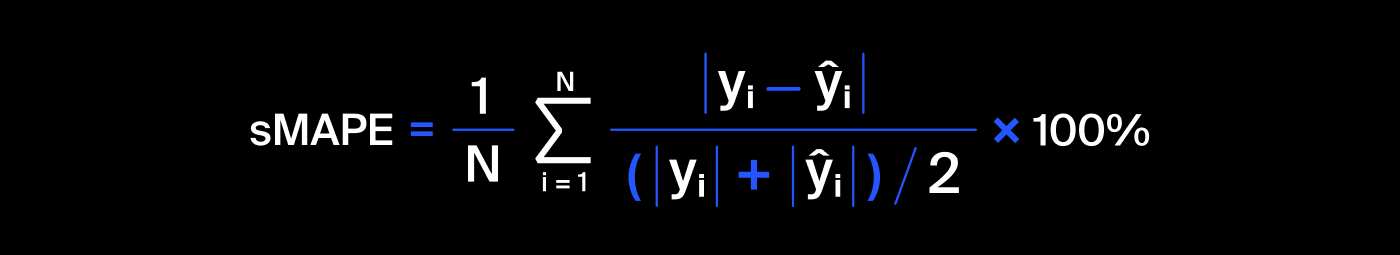
* *C* — доля золота в концентрате после флотации/очистки;
* *F* — доля золота в сырье/концентрате до флотации/очистки;
* *T* — доля золота в отвальных хвостах после флотации/очистки.

Для прогноза коэффициента нужно найти долю золота в концентратах и хвостах. Причём важен не только финальный продукт, но и черновой концентрат.

**Метрика качества**

Для решения задачи использована метрика качества — **sMAPE** (англ. Symmetric Mean Absolute Percentage Error, «симметричное среднее абсолютное процентное отклонение»).

Аналог MAE, но выражается не в абсолютных величинах, а в относительных. Почему симметричная? Она одинаково учитывает масштаб и целевого признака, и предсказания. (см. sMAPE.jpg)



**Обозначения**:



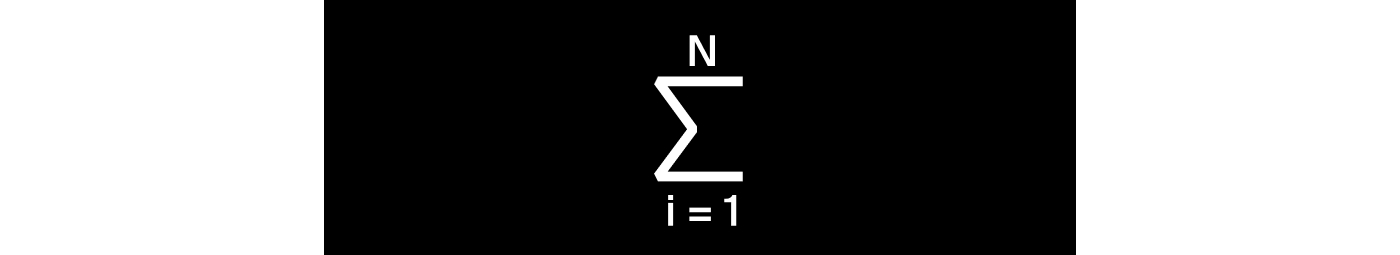
* Значение целевого признака для объекта с порядковым номером ***i*** в выборке, на которой измеряется качество.



* Значение предсказания для объекта с порядковым номером ***i***, например, в тестовой выборке.



* Количество объектов в выборке.



* Суммирование по всем объектам выборки (***i*** меняется от 1 до *N*).

Нужно спрогнозировать сразу две величины:

1. эффективность обогащения чернового концентрата rougher.output.recovery;
2. эффективность обогащения финального концентрата final.output.recovery.

Итоговая метрика складывается из двух величин:

