

1

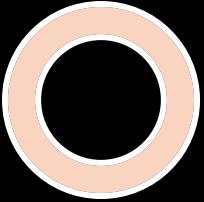
Прикладная аналитика для решения проблем бизнеса

Семинар - 1





Инфо обо мне



Аника Джха

- DSBA'19 Graduate
- Master of Business Analytics – University of Wollongong in Dubai
- Business Analyst at XPANCEO, Dubai

Контакты:

- a.dzhkha@hse.ru
- Tg: t.me/indiaaan_girl



Цели курса:

01

**Понимание
бизнес-
проблем**

02

**Сбор и
обработка
данных**

03

**Подбор
правильных
инструментов**

04

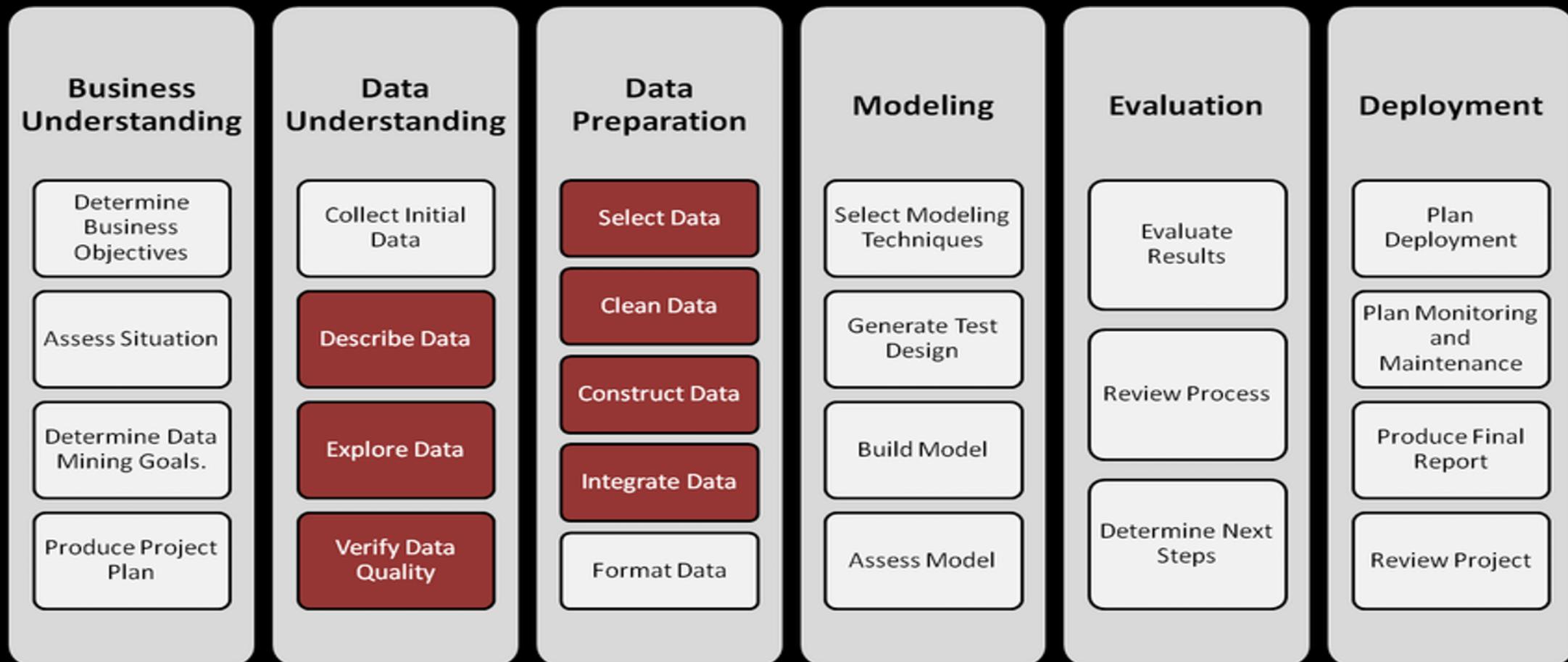
**Интерпретация
результатов**

05

**Выполнение
полного цикла
решения
бизнес-задач**

CRISP DM Process

“CROSS Industry Standard Process for Data Mining”



Бизнес-кейс: Прогнозирование кредитного риска



ЦЕЛЬ: ПОСТРОИТЬ МОДЕЛЬ, ПРЕДСКАЗЫВАЮЩУЮ
ДЕФОЛТ ПО КРЕДИТАМ КЛИЕНТОВ, ДАТЬ
РЕКОМЕНДАЦИИ БИЗНЕСУ НА ОСНОВЕ ПОЛУЧЕННЫХ
РЕЗУЛЬТАТОВ.



ДАТАСЕТ: KAGGLE (OPEN-SOURCE).

1. Понимание бизнес-проблемы



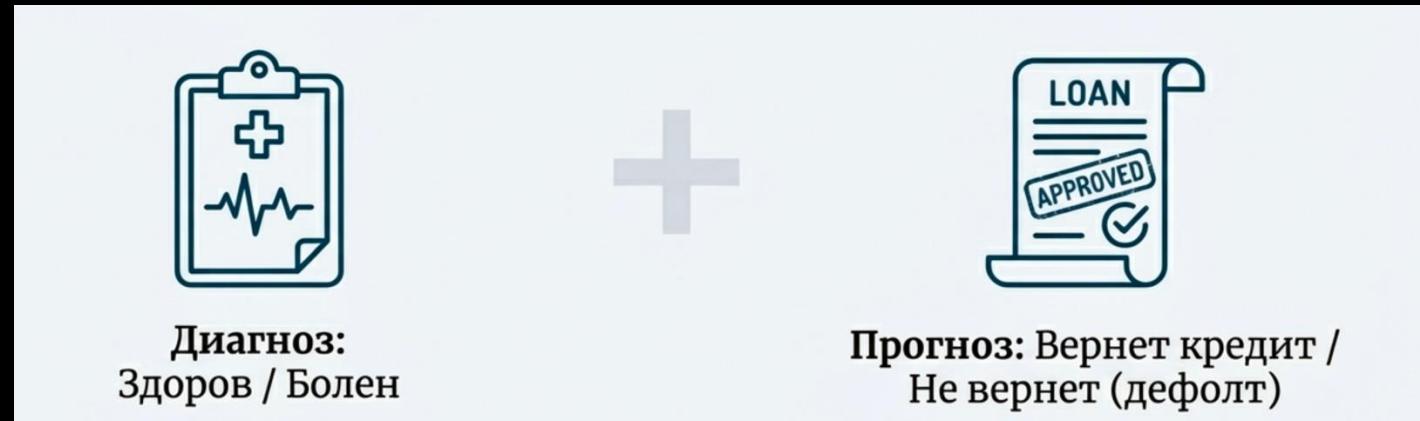
Как отличить надежного заемщика от того, кто с высокой вероятностью не вернет кредит?

Это одна из ключевых задач для любого банка. Цена ошибки - миллионы. Наша цель - использовать данные и инструменты для анализа, чтобы минимизировать эту ошибку.

Аналитический подход: Бинарная Классификация

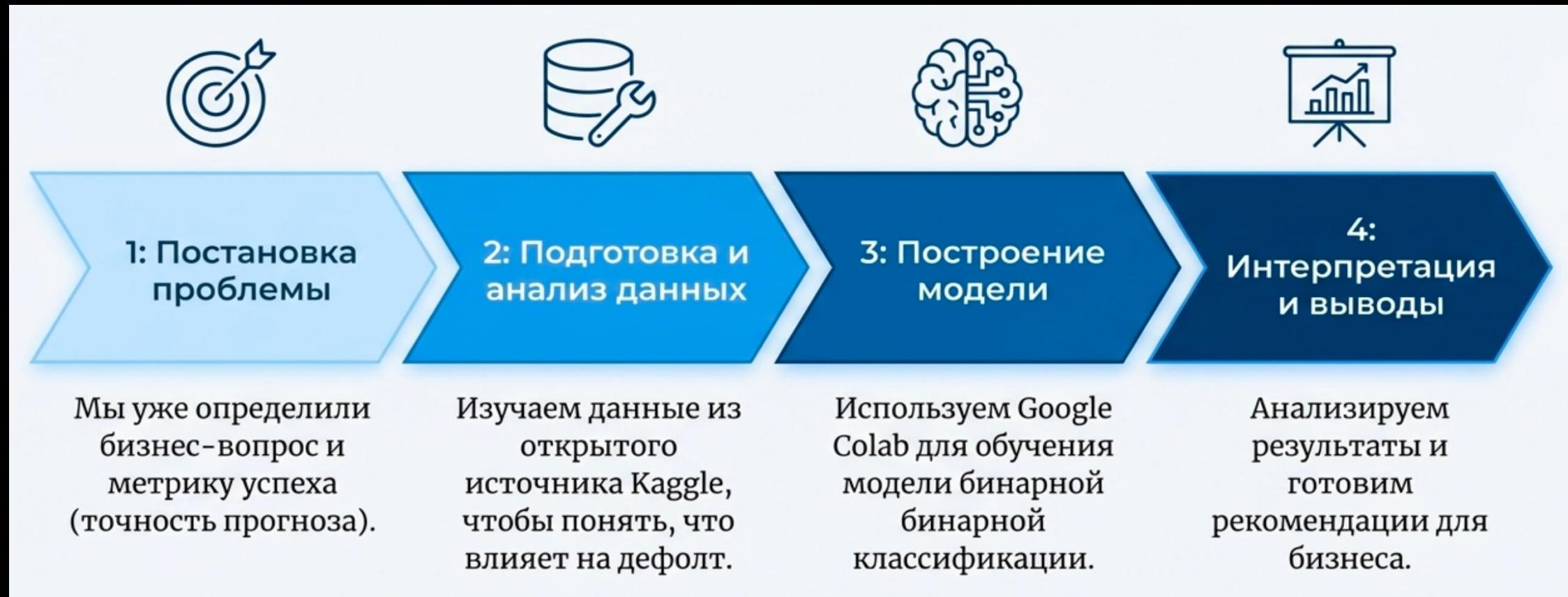
Метод, который разделяет объекты на два класса. Врач ставит диагноз: «болен» или «здоров». Email-фильтр решает : «спам » или « не спам ».

Наша модель будет делать то же самое только для заемщиков.



Задача — научить алгоритм принимать это решение на основе данных клиента.

Полный цикл решения задачи





“Модель дает вам ответ. Хороший аналитик объясняет, почему этот ответ именно такой.”

Важно не просто получить процент точности. Мы **должны понять, какие факторы** (например, баланс, история платежей, демография) **сильнее всего влияют на вероятность дефолта**. Именно эти знания превращают **анализ в ценное бизнес-решение**.

Мы вместе пройдем все шаги на примере одного набора данных.

Теперь ваша очередь применить полученные знания. Работая в группах, вы решите похожую бизнес-задачу.



Ваше задание: по шагам



1.

1. Изучите новую проблему: Вы получите новый датасет с Kaggle и краткое описание бизнес-контекста.



2.

2. Адаптируйте код: Используйте наш Google Colaboratory notebook как основу. Ваша задача — адаптировать его для работы с новыми данными.



3.

3. Сравните модели: (**Дополнительные баллы**) Попробуйте несколько алгоритмов классификации и сравните их эффективность. Какой из них лучше решает поставленную задачу?



4.

4. Подготовьте презентацию: Представьте свои результаты в формате короткой презентации.

Что такое отличный результат?

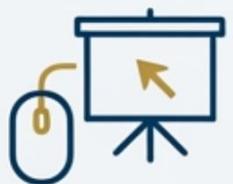
Ваша презентация — это отчет для бизнес-заказчика. Он должен быть ясным, убедительным и полезным.

-  **Четкая постановка проблемы:** Какую бизнес-задачу вы решали?
-  **Прозрачная методология:** Какие данные и инструменты вы использовали?
-  **Главные выводы (инсайты):** Какие факторы оказались наиболее важными? Что удивило в данных?
-  **Бизнес-рекомендации:** Что вы предлагаете делать на основе вашего анализа? (Например, на какие сегменты клиентов обратить внимание).

Формат и сроки



Работа в группах: Распределитесь на небольшие группы.



Инструменты: Google Colaboratory для кода, Google Slides/PowerPoint для презентации.



Презентация: 5-7 минут на выступление.



Срок: Презентации пройдут в конце этой пары либо на следующем семинаре.

Этот опыт - от понимания проблемы до презентования выводов - является ключевым навыком, который ценится в ведущих компаниях по всему миру.

