

Data science и инструменты анализа данных

2024/2025 учебный год

Доцент кафедры ИВЭ, Махно В.В.

©Создано при помощи <https://sberuniversity.ru/>



Лекция 2. Задача классификации

Пример с вакансиями.

Нужно построить алгоритм, который позволит системе определить, есть ли в резюме кандидата необходимые параметры. Если есть – отправить в папку «Собеседование», если нет – в папку «Отказать».

TRUE – алгоритм ответил правильно FALSE – неправильно

POSITIVE – подходящие NEGATIVE - неподходящие

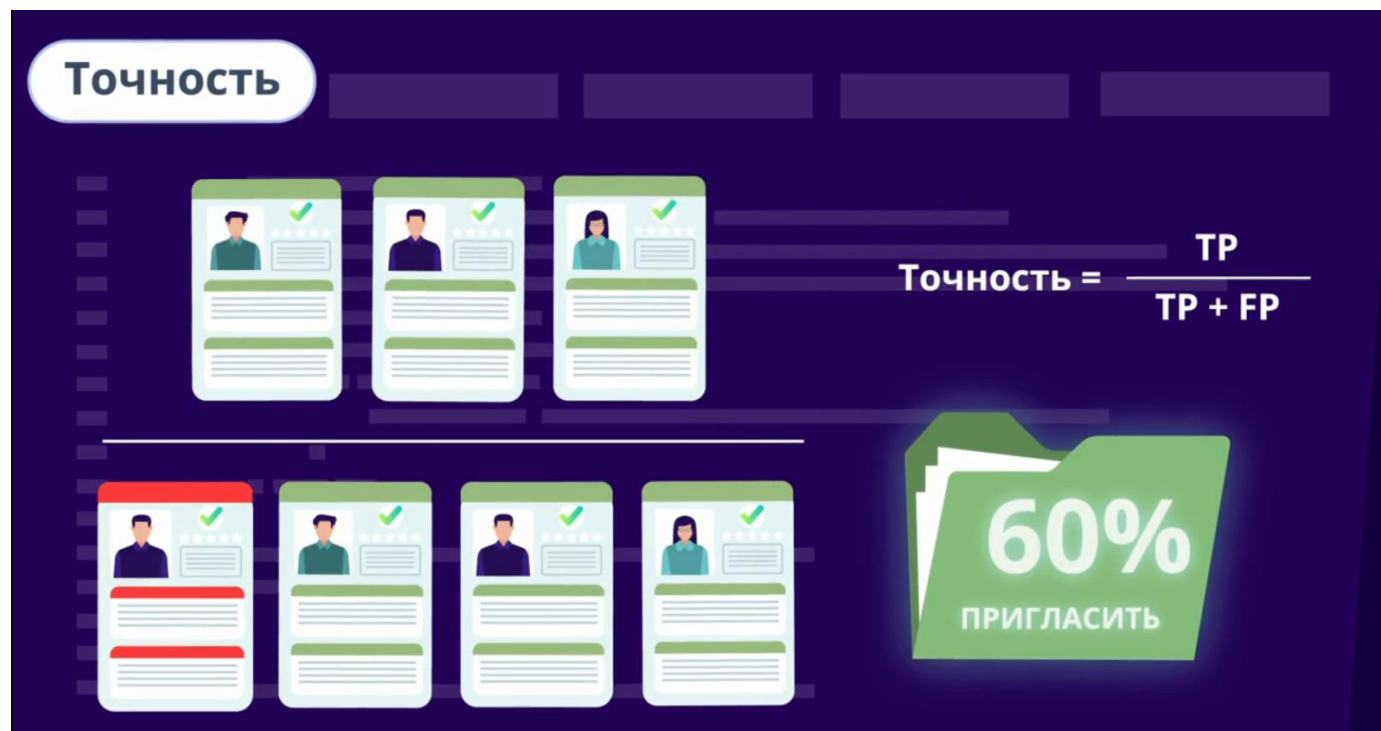
- TP — true positive, алгоритм верно пометил резюме как подходящее
- TN — true negative, алгоритм верно отнес резюме к неподходящим
- FP — false positive, алгоритм ошибочно считает подходящим резюме, в котором нет нужных качеств
- FN — false negative, алгоритм ошибочно отбраковал подходящее резюме



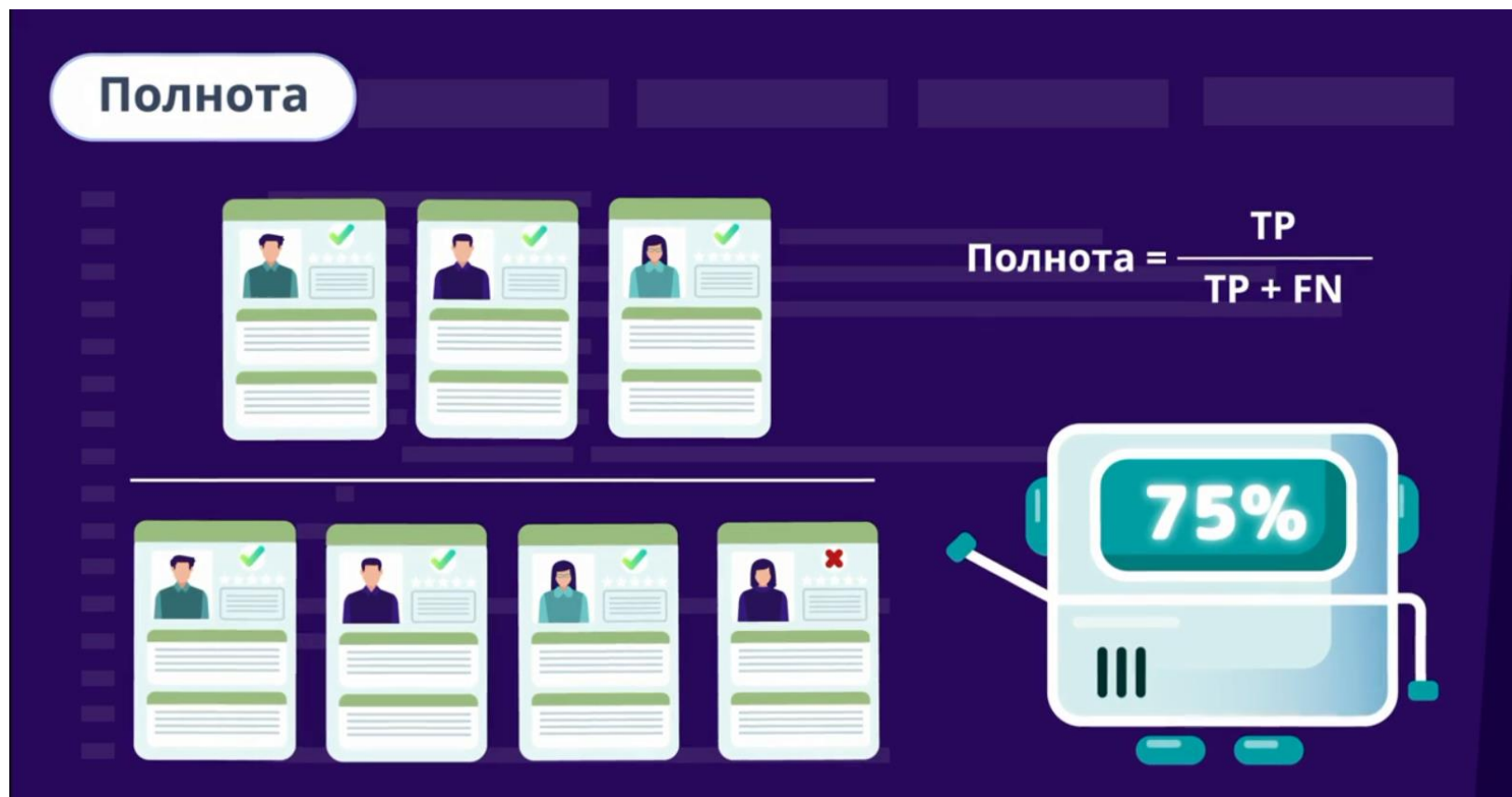
Метрики

- Самая простая метрика – **доля правильных предсказаний**: сколько раз прогноз машины и разметка программиста совпали между собой.
- Другая метрика – **точность**. Она показывает отношение количества верно угаданных подходящих резюме к количеству тех, кого машина вообще отнесла к группе «собеседование».
- Кроме точности есть еще метрика **полноты**. Она показывает отношение количества верно угаданных подходящих резюме, к другому значению: количеству кандидатов, которых следовало пригласить по мнению программиста.

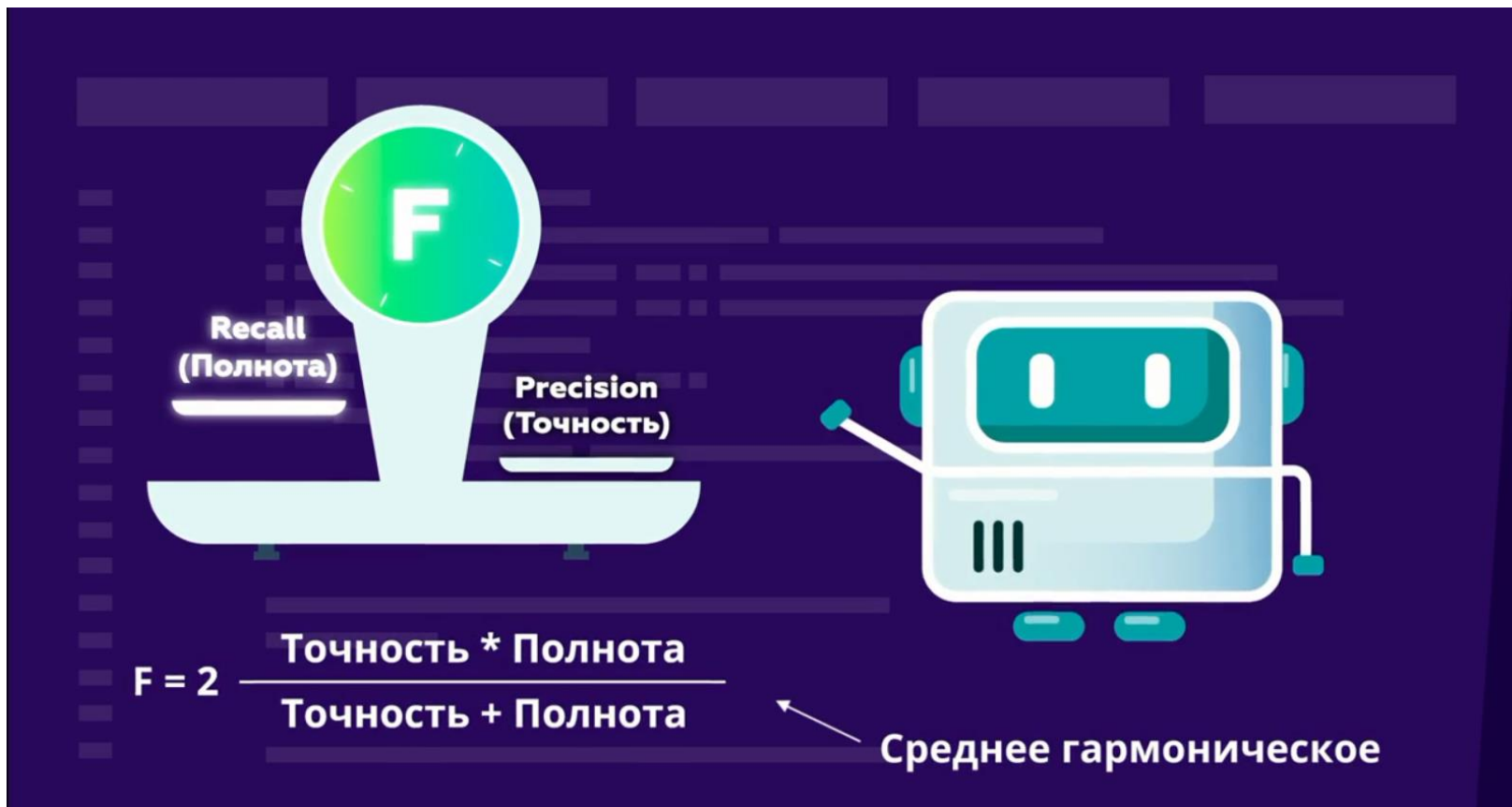
Метрика точность



Метрика полнота



F-метрика



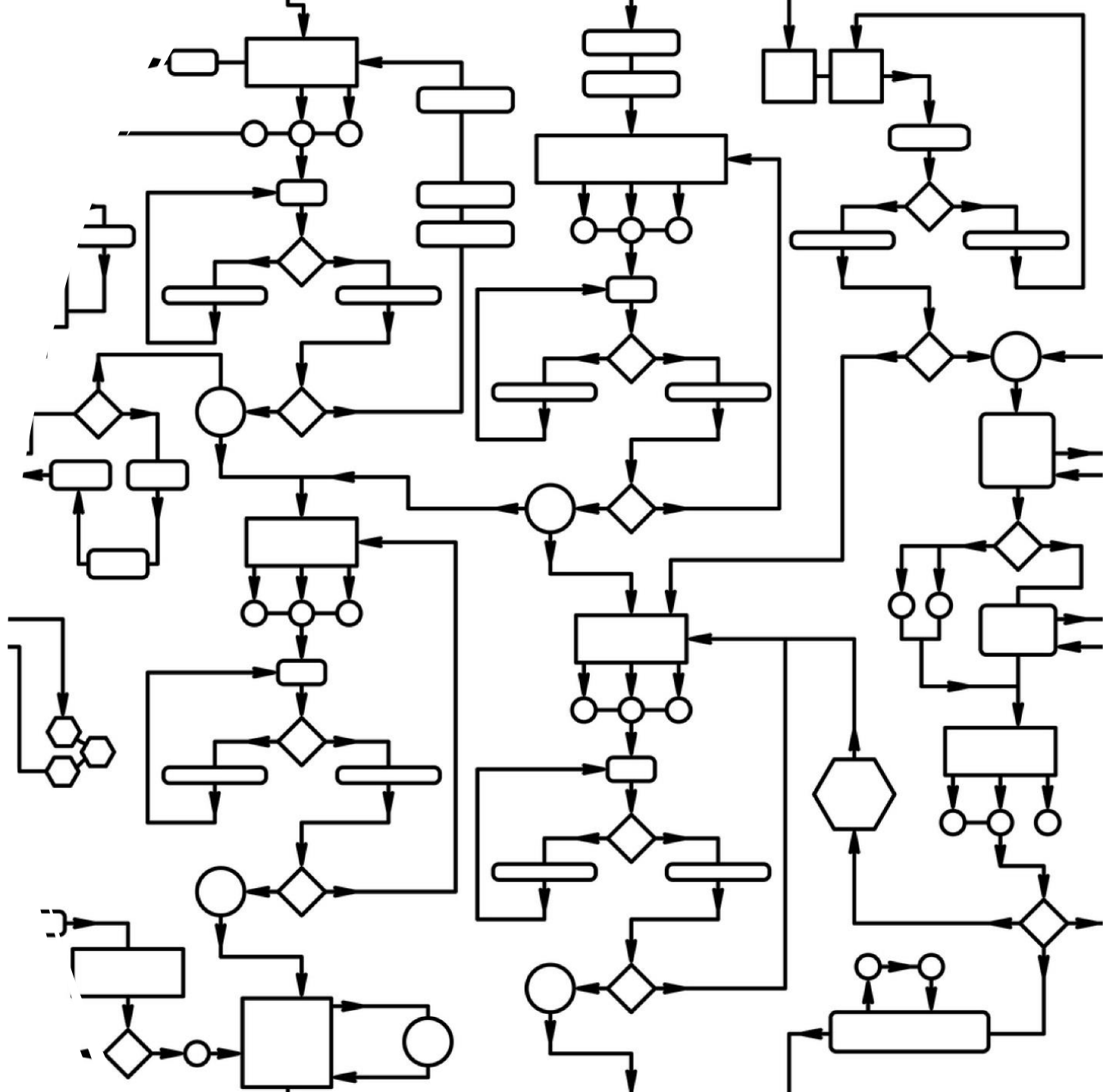
Табличные данные

Зарботная плата	Возраст	Должность	Уровень образования	Город проживания	Стаж работы (годы)	Вернет ли клиент кредит
100000	26	Риэлтор	Высшее	Санкт-Петербург	5	Да
50000	20	Продавец-консультант	Высшее	Москва	1	Нет
35000	39	Автомеханик	Среднее специальное	Воронеж	8	Нет
25000	23	Программист	Высшее	Самара	2	Да
75000	41	Юрист	Среднее	Москва	14	Да

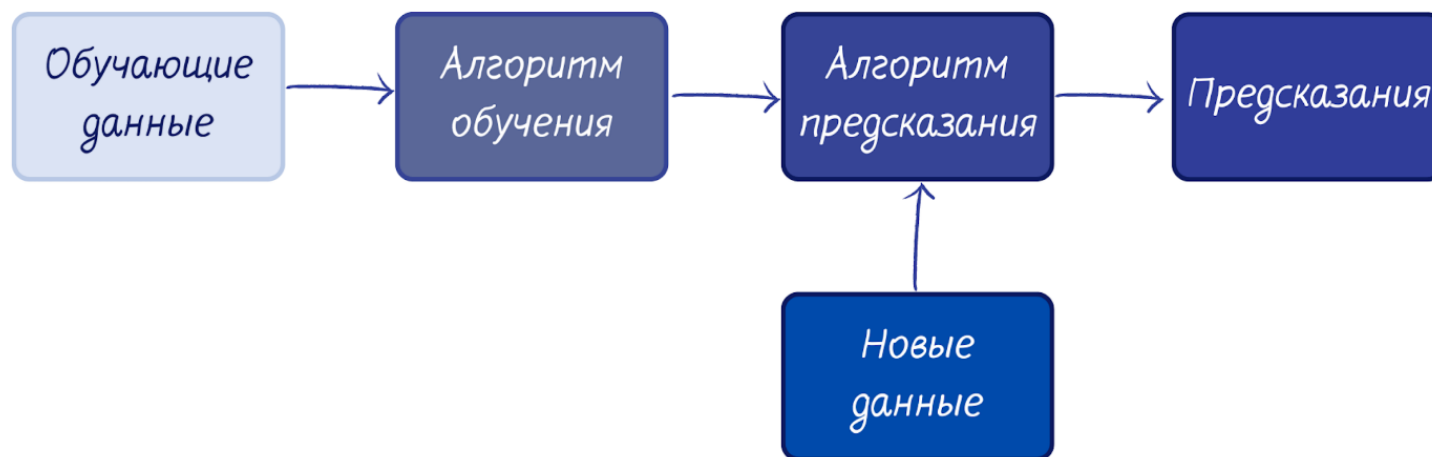
Создание алгоритма классификации

В этой задаче каждому объекту (строке в таблице данных) соответствует класс — значение из заданного набора классов.

Задача классификации состоит в том, чтобы разработать алгоритм, который по признакам объекта будет предсказывать класс



Машинное обучение



Онлайн-курс СберУниверситета

Генеративное искусство

Подробнее о курсе

Бесплатный курс от Сбера
по генеративному
искусству

https://courses.sberuniversity.ru/generative-art?utm_source=telegram&utm_medium=organic&utm_campaign=courses&utm_content=geni&utm_term=01_09_2023