



ML-ге кіріспе. CRISP-DM әдіснамасы

Апта: № 1
Дәріс

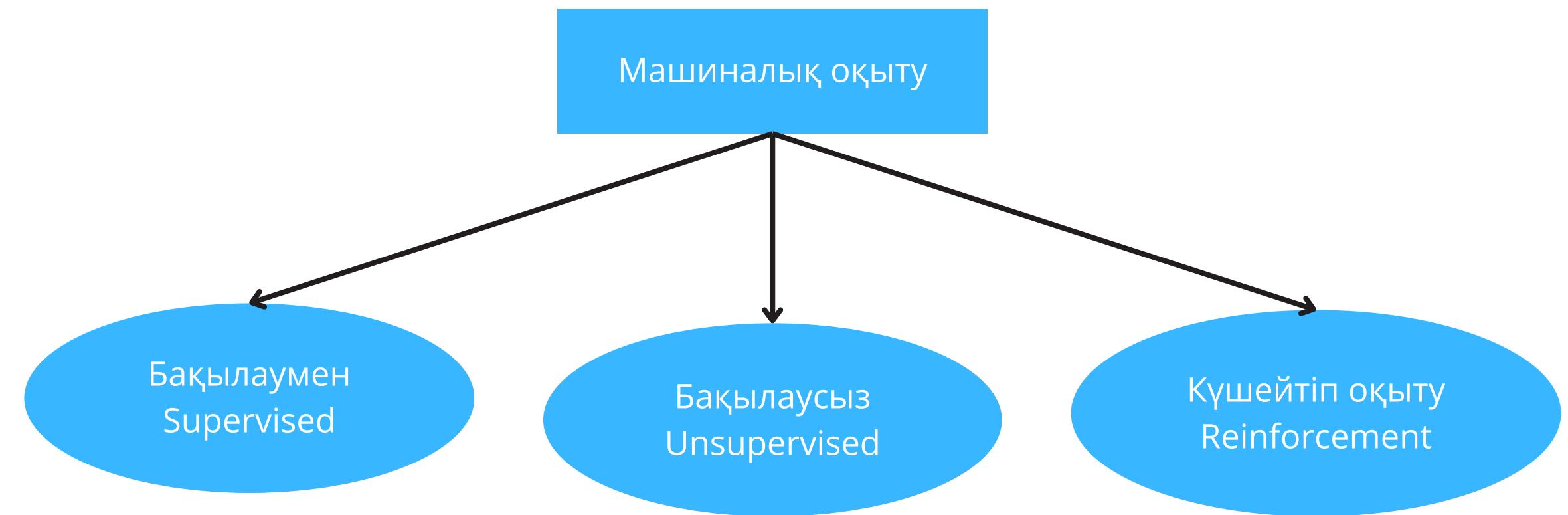
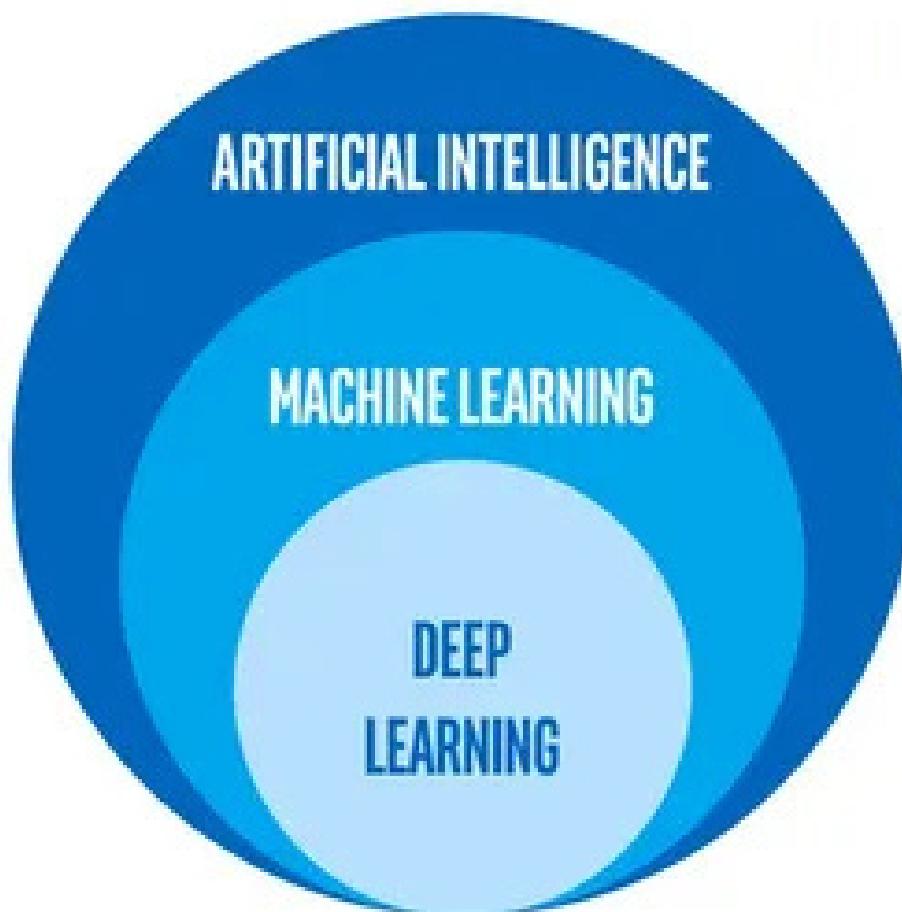
Мақсат

- Машиналық оқытудың не екенін және оның түрлері
- Міндет түрлері
- Деректер типтері мен шкалалары
- CRISP-DM әдіснамасының кезеңдері
- Практикалық құралдармен танысамыз

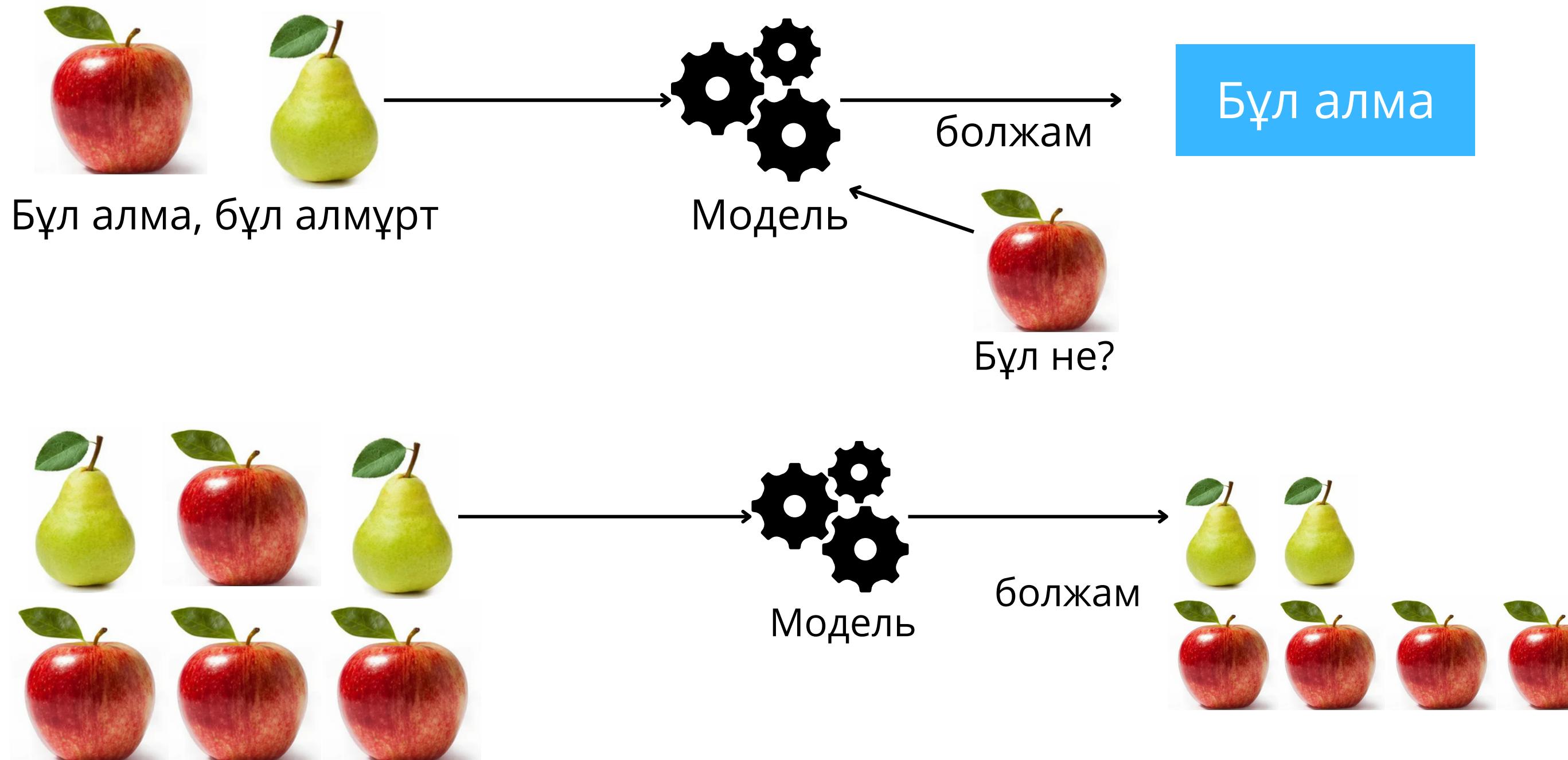


ML деген не?

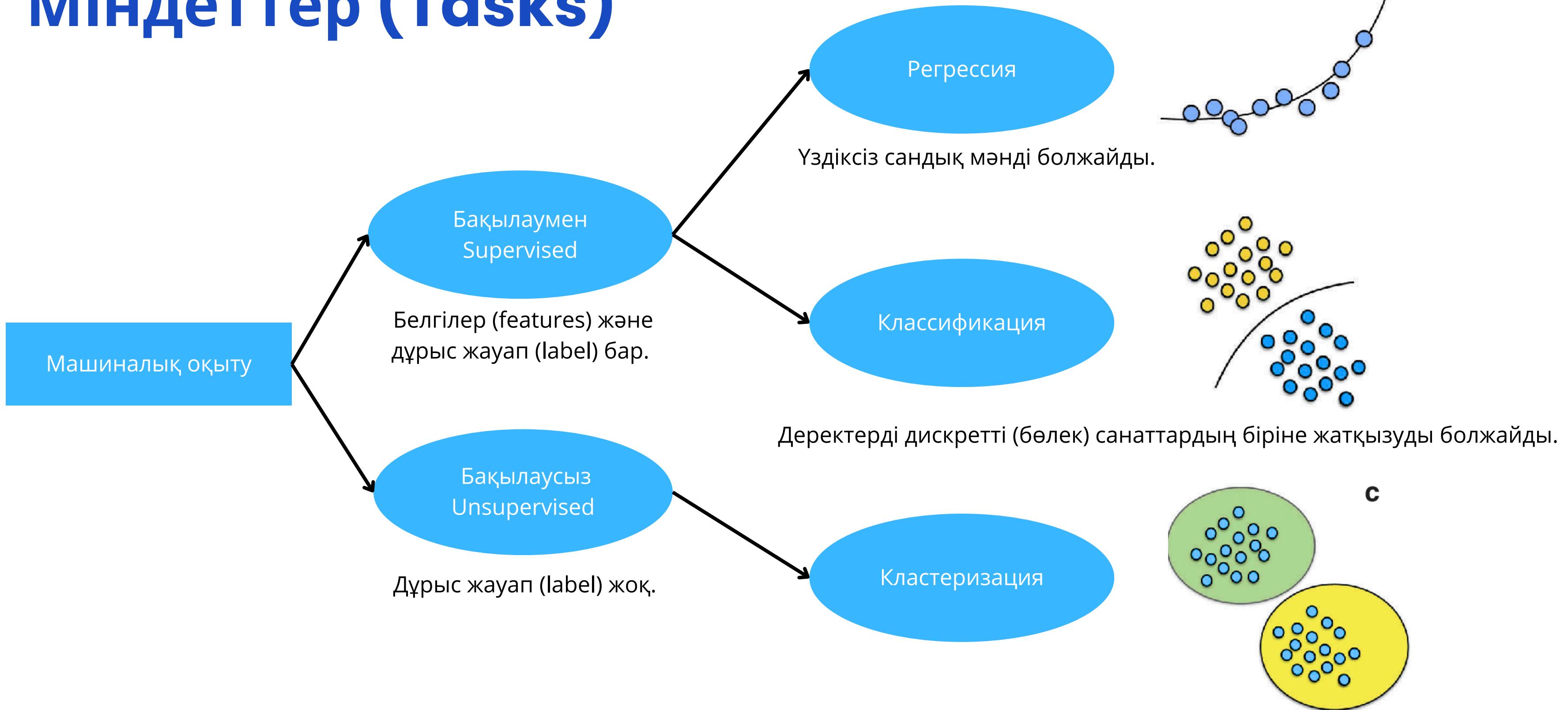
Машиналық оқыту деректерден заңдылық табу тәсілі
Бағдарламашы ережелер жазбайды, модель өзі
үйренеді



Бақылаумен және бақылаусыз оқыту



Міндеттер (Tasks)

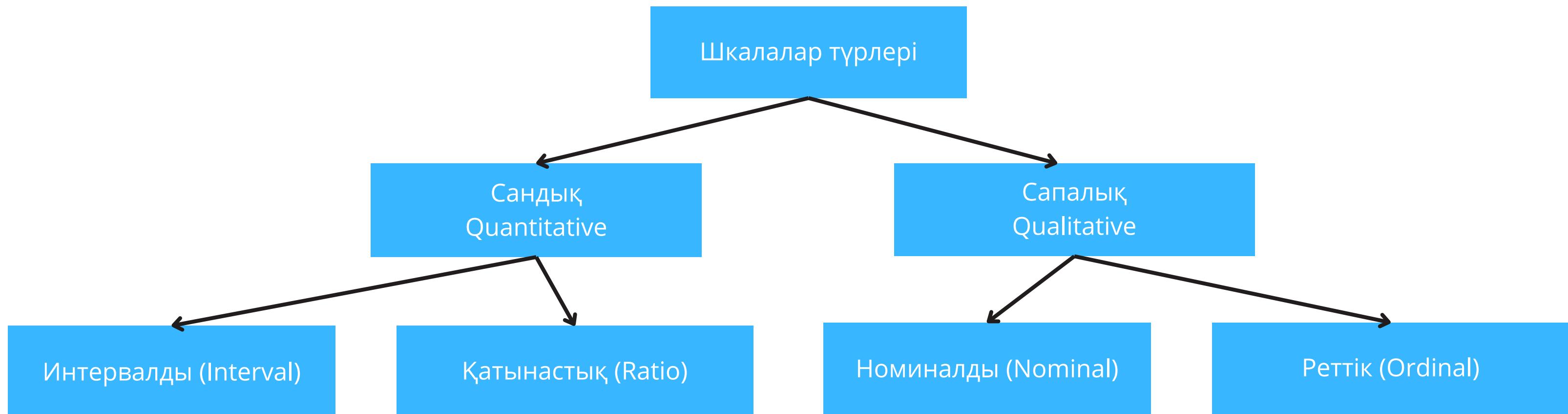


Міндет Түрлері

Міндет Түрі	Сипаттамасы	Мысалдар
Классификация	Деректерді дискретті (бөлек) санаттардың біріне жатқызуды болжайды.	Электрондық пошта спам ба, жоқ па? Суретте мысық па, ит пе? Клиент қарызыды қайтара ма?
Регрессия	Үздіксіз сандық мәнді болжайды.	Үйдің бағасын болжау. Келесі айдағы сатылым көлемін болжау. Температуралы болжау.
Болжау	Уақыт қатарындағы келесі мәндерді алдын ала айту. Жиі регрессияның бір түрі ретінде қарастырылады.	Акция бағасының келесі аптадағы қозғалысын болжау. Келесі тоқсандағы электр энергиясын тұтыну.
Ранжирлеу	Нысандарды олардың маңыздылығына немесе қалауына қарай реттеу (тізімдеу).	Іздеу жүйесіндегі сұранысқа сәйкес беттерді реттілікпен көрсету. Пайдаланушыға ұсынылатын тауарлардың реті.



Шкалалар түрлері



айырмашылық бар, бірақ
абсолютті нөл жоқ (мысалы:
температура °C)

айырмашылық пен қатынас бар,
абсолютті нөл бар (мысалы:
салмақ, баға).

тек атаулар мен
категориялар (мысалы:
көз түсі, қала атауы)

реттілік бар, бірақ
айырмашылық өлшенбейді
(мысалы: білім деңгейі)



Мысал

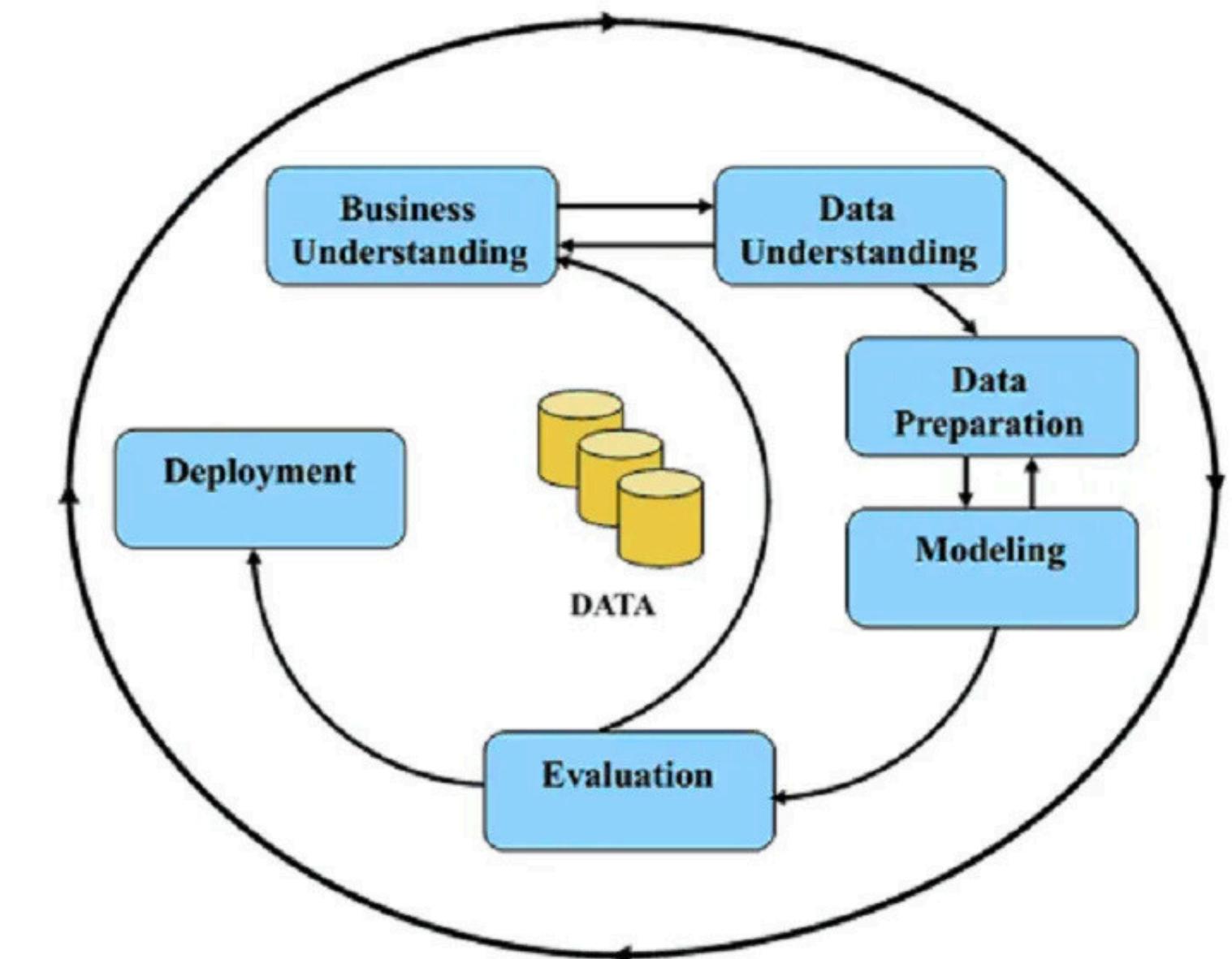
№	Студент аты	Білім деңгейі	Туған жылы	GPA
1	Айдана	Орта	2001	3.2
2	Нұржан	Жоғары	1999	2.9
3	Дана	Бастауыш	2003	3.4



CRISP-DM әдіснамасы

CRISP-DM (Cross-Industry Standard Process for Data Mining) – бұл деректерді талдау және ML-жобаларын жүргізуге арналған салааралық стандартты процесс.

- Бизнесті Түсіну: Жобаның мақсаттарын, бизнес талаптарын анықтау.
- Деректерді Түсіну: Деректерді жинау, олардың сапасын тексеру, бастапқы талдау жүргізу.
- Деректерді Дайындау: Таңдау, тазалау, өндіру, түрлендіру және қажетті деректерді форматтау. Бұл ең көп уақытты алатын кезең.
- Модельдеу: Қолданылатын ML-алгоритмдерін таңдау және оларды оқыту.
- Бағалау: Модельдің бизнес-мақсаттарға сәйкестігін бағалау және оның тиімділігін тексеру.
- Енгізу: Модельді жұмыс ортасына (өндіріске) енгізу және оны тұрақты бақылауды қамтамасыз ету.



Дерек қорлары



kaggle

- Google dataset
- <https://datasetsearch.research.google.com/>

- Kaggle
- <https://www.kaggle.com/>

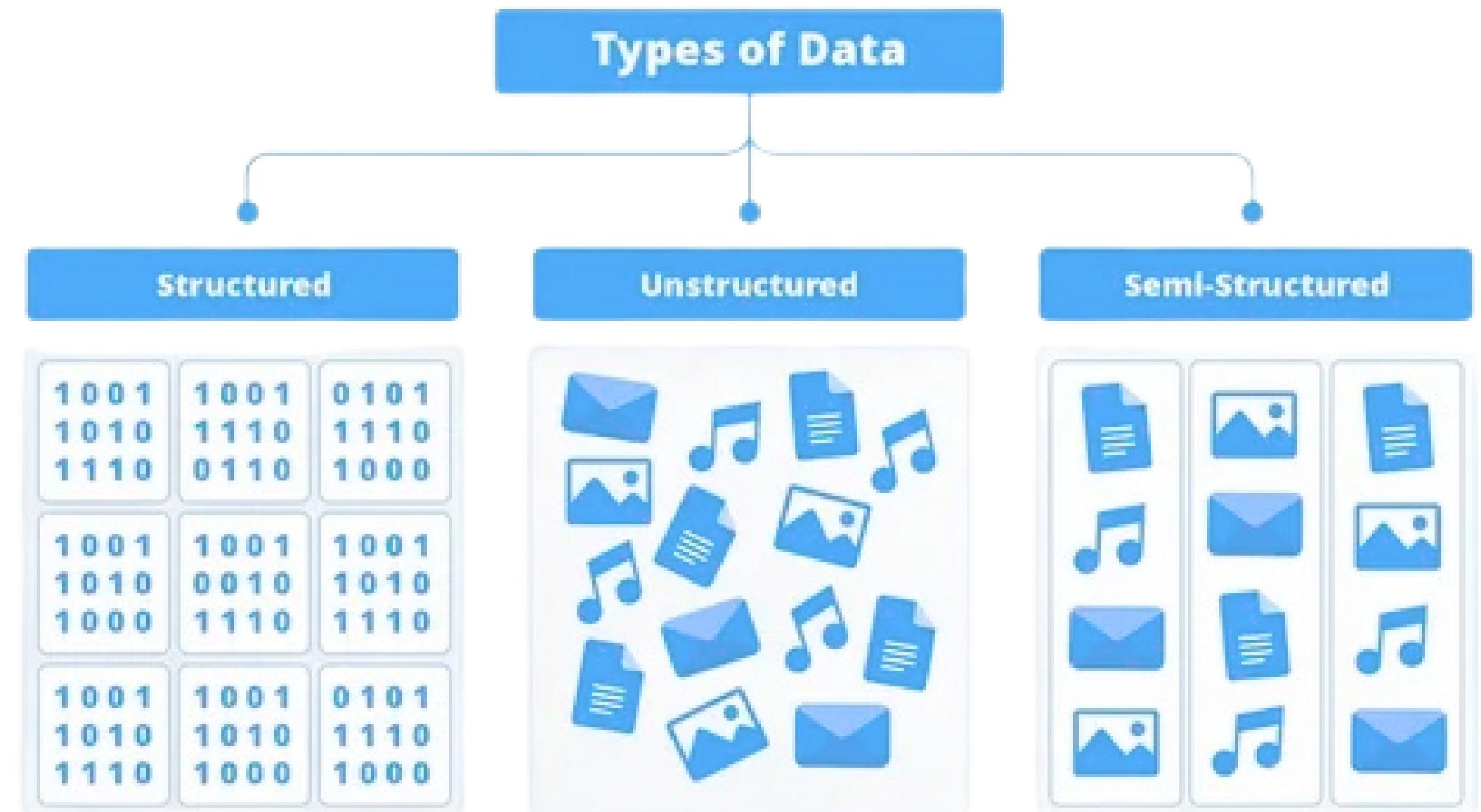


- Data gov.kz
- <https://data.egov.kz/>
- <https://stat.gov.kz/>



Дерек типтері

- Құрылымданған
- Құрылымданбаған
- Жартылай құрылымданған



Кестелік деректер

Баған 1 (Өріс)	Баған 2 (Өріс)	Баған 3 (Өріс)
Жол 1 (Жазба)	Ұяшық 1.2	Ұяшық 1.3
Жол 2 (Жазба)	Ұяшық 2.2	Ұяшық 2.3
Жол 3 (Жазба)	Ұяшық 3.2	Ұяшық 3.3

Жолдар (Қатарлар) (Rows)

Бағандар (Колонкалар) - (Columns)

Ұяшықтар - (Cells)

Өрістер - (Fields/Attributes)

Жазбалар - (Records)

Бастапқы кілт - Primary Key)

Сыртқы кілт - (Foreign Key)

Схема - (Schema)

Деректердің типі - (Data Type)



Қолданатын құралдар

Google Colab / Jupyter Notebook

Онлайн орта, Python кодын жазуға және орындауға мүмкіндік береді.
Ештеңе орнатудың қажеті жоқ — тек Google аккаунт керек.
colab.research.google.com

<https://colab.research.google.com>

Kaggle

Машиналық оқыту және деректер ғылымы бойынша платформа.
Деректер жиынтығын (dataset) жүктеп, талдау мен модель құруға болады.
www.kaggle.com

<https://www.kaggle.com>

GitHub / Classroom

GitHub — кодты сақтау, бөлісу және топпен жұмыс істеу платформа.

[classroom.github.com](https://classroom.github.com/classrooms)

<https://classroom.github.com/classrooms>



Репозиторий

 MO2026_1 Public

Edit Pins Watch 0

main 1 Branch 0 Tags Go to file Add file Code

Yerassyl04 Update README.md 3997e86 · 20 hours ago 7 Commits

README.md Update README.md 20 hours ago

README

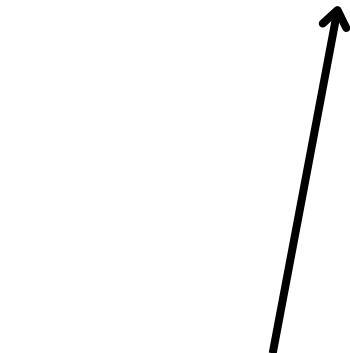
Машиналық оқыту курсы | ITECHRESEARCH

Курс туралы

Бұл ITECHRESEARCH мектебінің Машиналық оқыту (МО2026) курсының толық 26 апталық бағдарламасы. Курс теориялық білім мен практикалық дағдыларды біріктіріп, студенттерді қазіргі заманғы ML технологияларымен таныстырады.

MACHINE LEARNING PYTHON STATUS ACTIVE

Дәріс	Практика	Дәріс сөж	Практика сөж
 Дәріс жазбасы	 Практика	 сөж	 сөж



Сілтемелерге өту

https://github.com/ItechResearch26/MO2026_1?tab=readme-ov-file





Назарларыңызға рақмет!

(Сұрақ-жаяп)