

Машиналық оқыту алгоритмдері

Болатбек М.А.

Мәліметтер

Кесте



Мәтін



Сигналдар



Дыбыс



Сурет



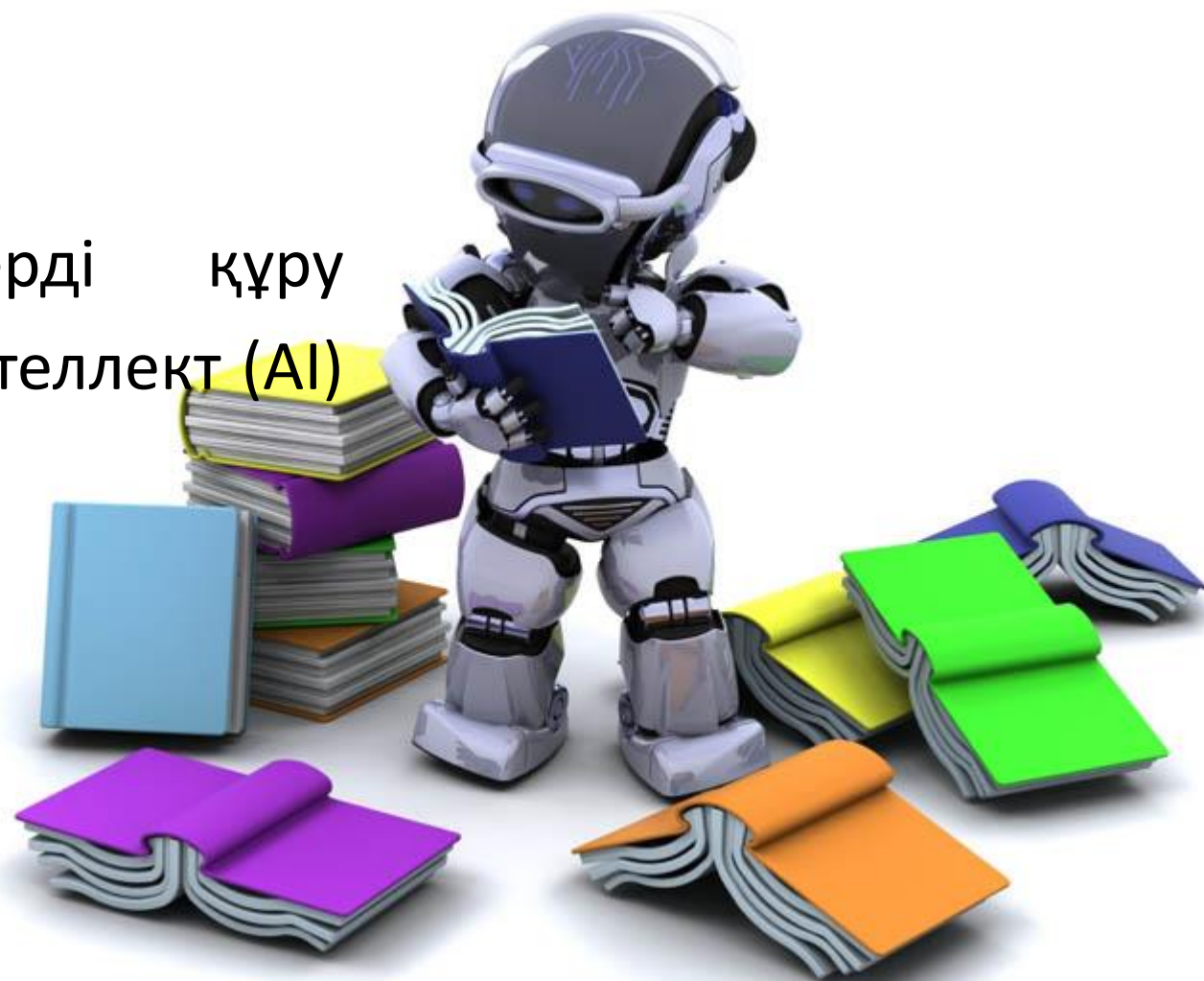
Видео



Машиналық оқыту дегеніміз не?

(ағыл. Machine learning)

оқуға қабілетті алгоритмдерді құру
әдістерін зерттейтін жасанды интеллект (AI)
бөлімі.



МО дегеніміз не?

- Машиналық оқыту - бұл адамның араласуынсыз өздігінен автоматты түрде талдайтын және шешім қабылдайтын алгоритмдер мен деректерді қамтитын жасанды интеллект қосымшасы.
- Ол алдыңғы тәжірибе негізінде компьютердің тапсырмаларды қалай дербес орындайтынын сипаттайды.
- Сондықтан, машиналық тілде жасанды интеллект тәжірибе негізінде жасалады деп айта аламыз.

Machine Learning (ML)

- ML-жасанды интеллект саласы:
- Деректерді түсіну үшін есептеу жүйелерін қолданады
- Үлгілерді алу, деректерді функцияларға сәйкестендіру, деректерді жіктеу және т. Б
- ML жүйелері үйреніп, жетілдіре алады
- Теориялық информатика мен Шу туралы нақты деректерді біріктіреді.

Компьютер vs ML

- Кәдімгі Компьютерлік бағдарламалық жасақтама мен машиналық оқытудың айырмашылығы-әзірлеуші адам жүйеге жағдайға қалай жауап беру керектігін үйрететін кодтарды бермейді, оның орнына көптеген мәліметтер оқытылады.

MO алгоритмдері:

- Neural Networks
- Random Forests
- Decision trees
- Genetic algorithm
- Radial basis function
- Sigmoid



МО қолданылуы

- Трафикті болжау
- Виртуалды Жеке Көмекші
- Сөйлеуді тану
- Спам мен зиянды бағдарламаларды электрондық пошта арқылы сүзу
- Биоинформатика
- Табиғи тілді өңдеу

МО нақты уақыт режимінде қолданылуы

- **ТРАФИКТІ БОЛЖАУ**
- **ВИРТУАЛДЫ ЖЕКЕ КӨМЕКШІ**
- **ОНЛАЙН ТАСЫМАЛДАУ**
- **ӘЛЕУМЕТТІК ЖЕЛІ ҚЫЗМЕТТЕРІ**
- **ЭЛЕКТРОНДЫҚ ПОШТА АРҚЫЛЫ ҚАЛАУСЫЗ ПОШТАНЫ СҮЗУ**
- **ӨНІМ ҰСЫНЫСЫ**
- **ИНТЕРНЕТТЕГІ АЛАЯҚТЫҚТЫ АНЫҚТАУ**

МО арналған бағдарламалау тілдері

Ең жақсы және жиі қолданылатын Машиналық оқыту бағдарламалары:

- Python,
- java,
- C,
- C++,
- Shell,
- R,
- JavaScript,
- Scala,
- Shell,
- Julia



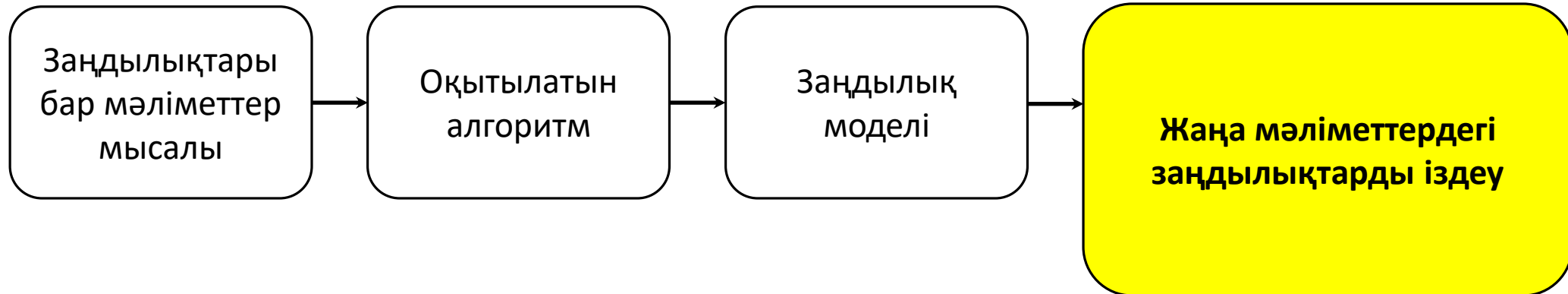
Машиналық Оқыту Мен Жасанды Интеллект Арасындағы Айырмашылық

Жасанды интеллект - бұл адамның мінез-құлқын ынталандыратын ақылды машиналарды құру тұжырымдамасы, ал Машиналық оқыту-бұл компьютерге бағдарламалаусыз мәліметтер негізінде үйренуге мүмкіндік беретін жасанды интеллекттің жиынтығы.

Машиналық оқыту

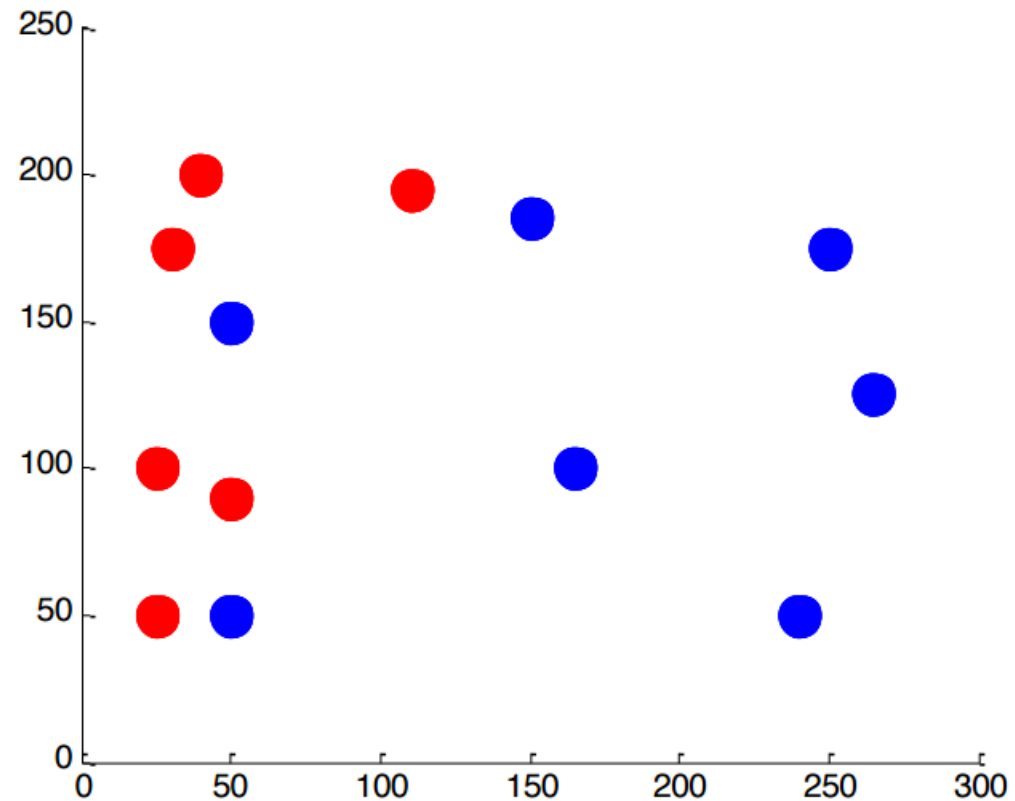


Машиналық оқыту



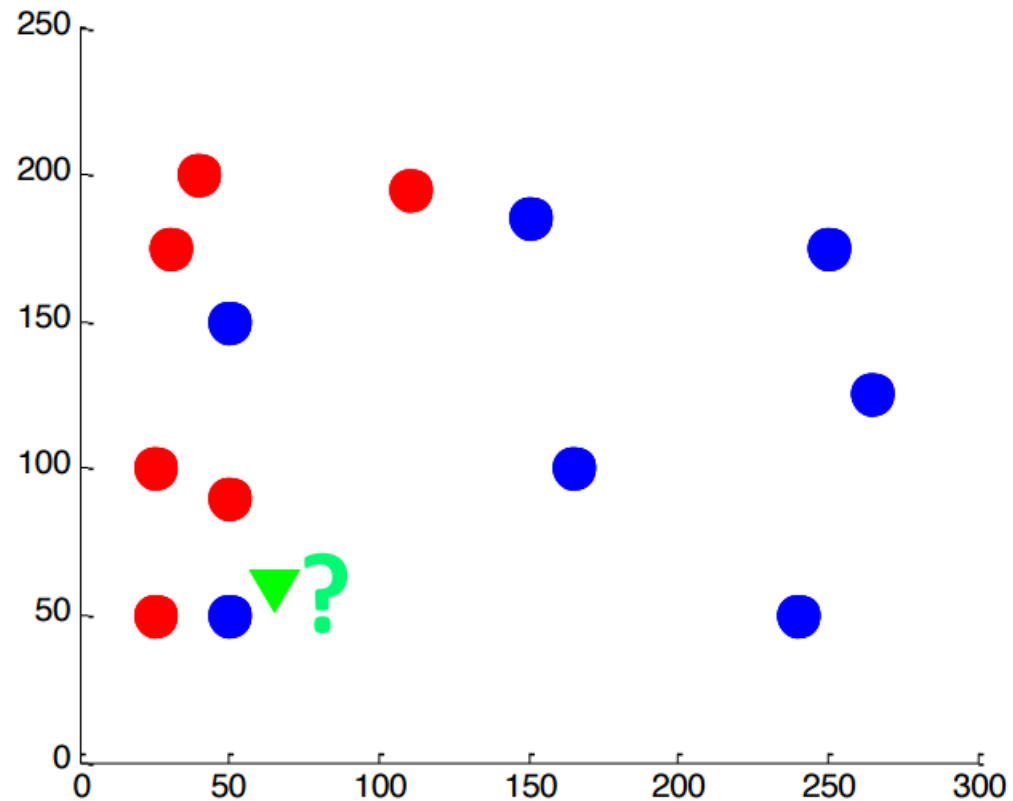
Қарапайым мысал

Бізге белгілі мәліметтермен көрсетілген «әлем сурет»

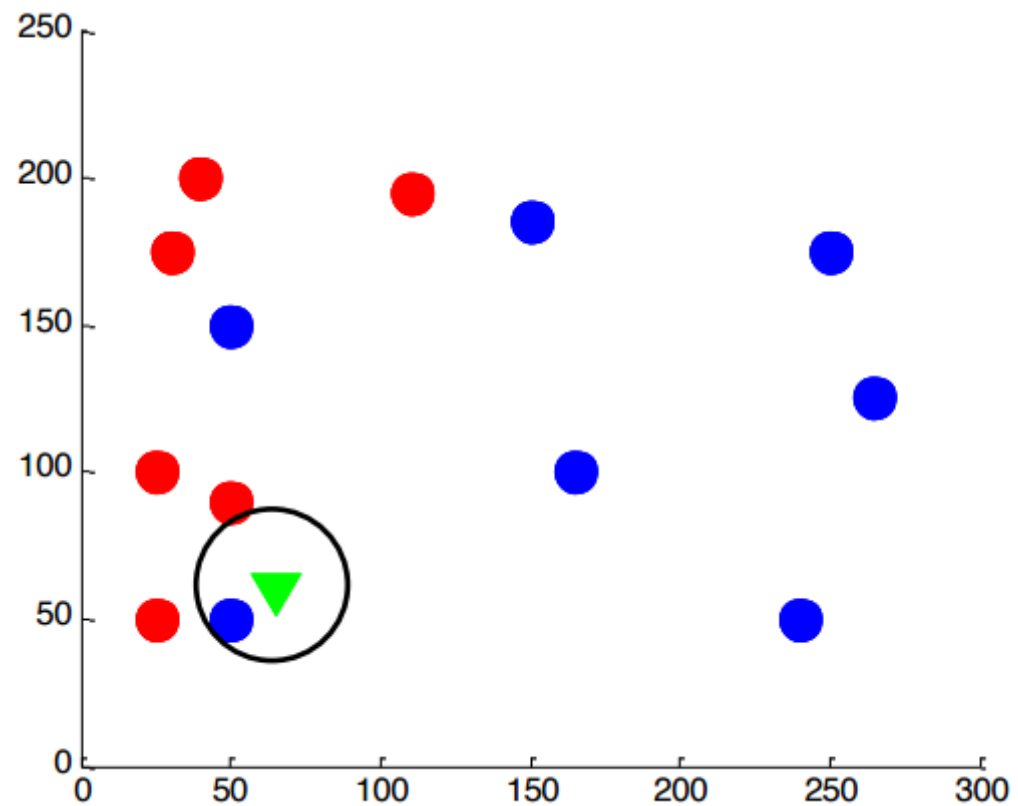


Қарапайым мысал

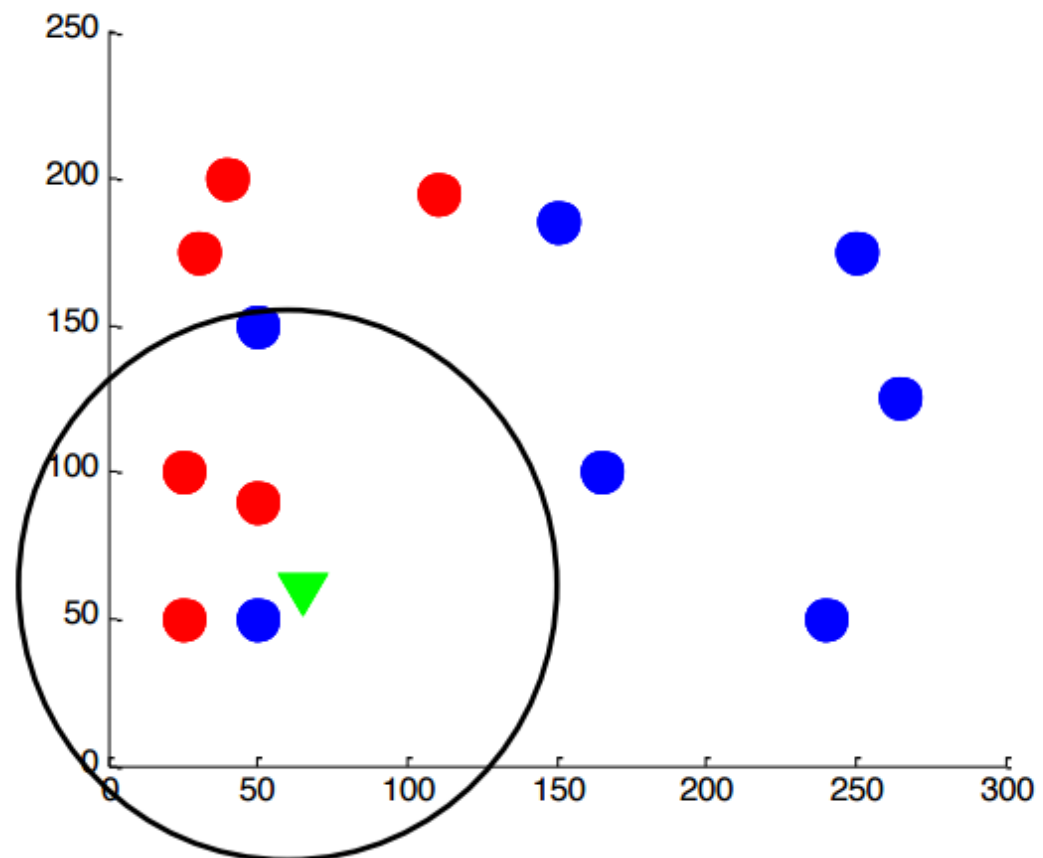
Жаңа нысан қызыл ма, көк пе?



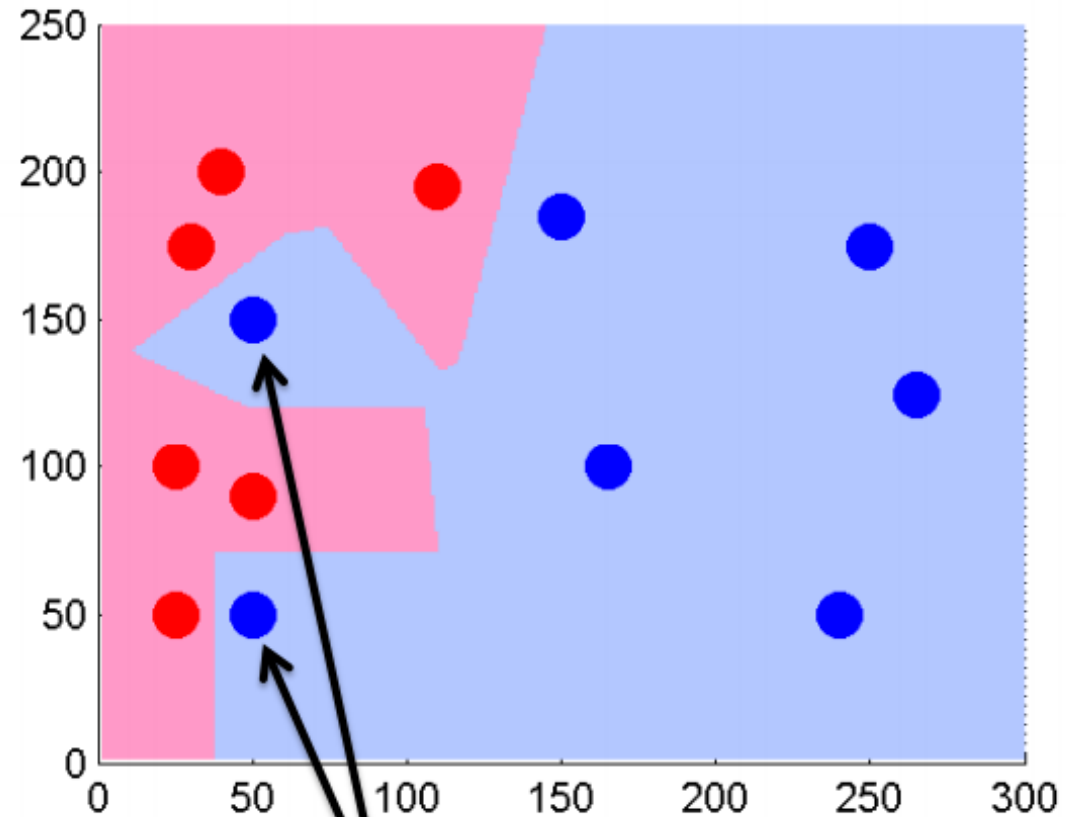
Қарапайым мысал



Қарапайым мысал

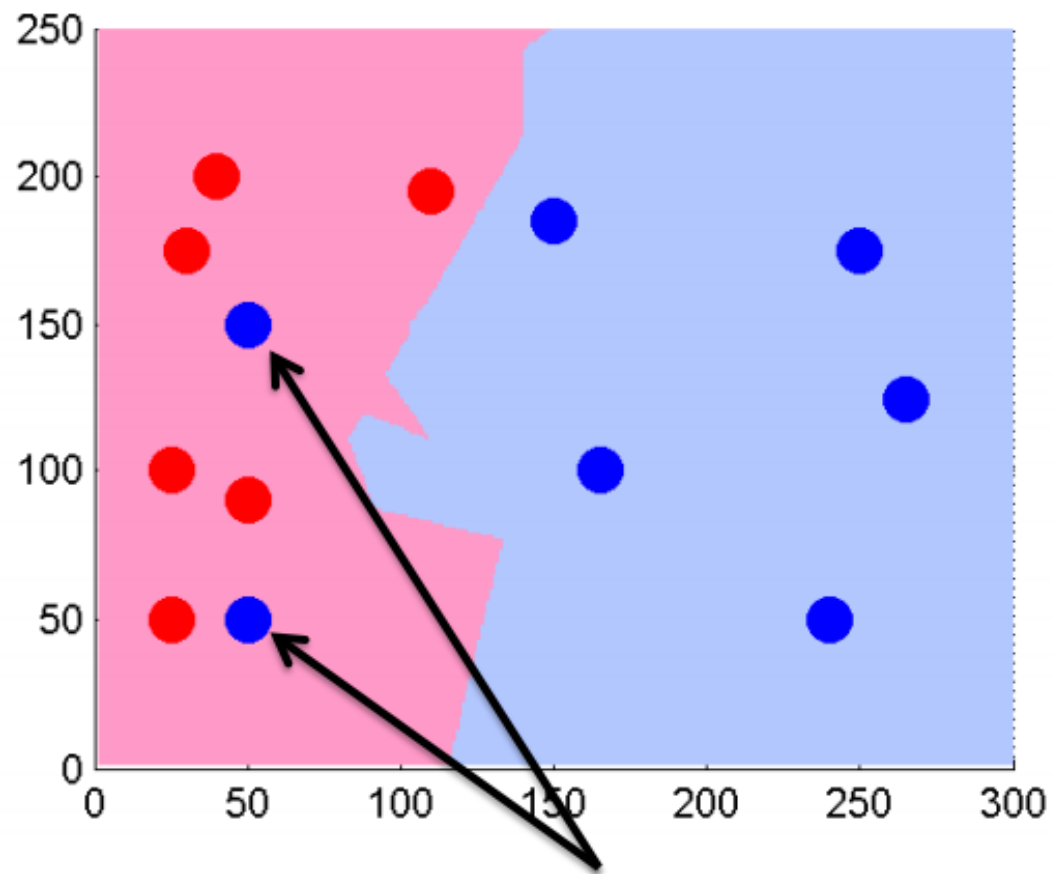


Қарапайым мысал



Шу ма, таңдама ма?

Қарапайым мысал



Қателік?

Машиналық оқыту

$$F: X \rightarrow Y$$

Белгілер (X)					Мақсатты айнымалылар (Y)	
Бойы	IQ	Дәрі мөлшері	Жынысы	Мектептегі орта баллы	Универдегі орта баллы	Шылым шеге ме?
170	128	0	М	4.7	4.4	Жоқ
195	90	40	М	3.3	3.1	Иә
160	111	2	Ж	4.0	3.9	Жоқ
183	143	0	М	4.8	4.7	Иә
158	118	25	Ж	3.7	3.5	Иә
178	132	10	М	4.4	?	?
170	120	4	Ж	3.9	?	?
165	100	20	М	3.5	?	?

Training. (rows 1-5)

Testing. (rows 6-8)

Машиналық оқыту

$$F: X \rightarrow Y$$

Белгілер (X)					Мақсатты айнымалы (Y)	
Бойы	IQ	Дәрі мөлшері	Жынысы	Мектептегі орта баллы	Универдегі орта баллы	Шылым шеге ме?
170	128	0	М	4.7	4.4	Жоқ
195	90	40	М	3.3	3.1	Иә
160	111	2	Ж	4.0	3.9	Жоқ
183	143	0	М	4.8	4.7	Иә
158	118	25	Ж	3.7	3.5	Иә
178	132	10	М	4.4	?	?
170	120	4	Ж	3.9	?	?
165	100	20	М	3.5	?	?

Train.

Test.

Нысандар

Машиналық оқыту

$$F: X \rightarrow Y$$

Белгілер (X)					Мақсатты айнымалы (Y)	
Бойы	IQ	Дәрі мөлшері	Жынысы	Мектептегі орта баллы	Универдегі орта баллы	Шылым шеге ме?
170	128	0	М	4.7	4.4	Жоқ
195	90	40	М	3.3	3.1	Иә
160	111	2	Ж	4.0	3.9	Жоқ
183	143	0	М	4.8	4.7	Иә
158	118	25	Ж	3.7	3.5	Иә
178	132	10	М	4.4	?	?
170	120	4	Ж	3.9	?	?
165	100	20	М	3.5	?	?

Нысандар **Белгілер**

Train.
 Test.

Машиналық оқыту

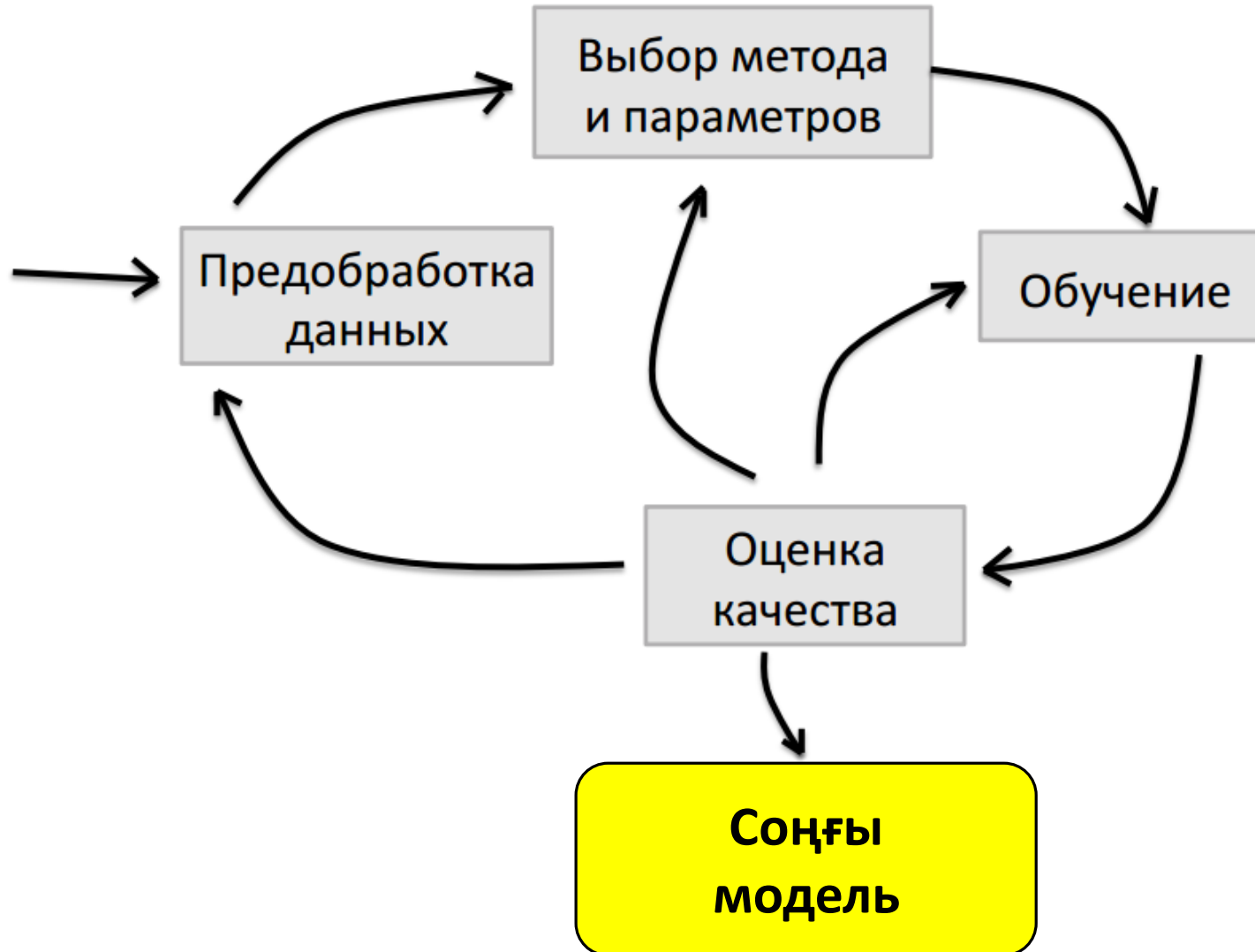
$$F: X \rightarrow Y$$

Белгілер (X)					Мақсатты айнымалы (Y)	
Бойы	IQ	Дәрі мөлшері	Жынысы	Мектептегі орта баллы	Универдегі орта баллы	Шылым шеге ме?
170	128	0	М	4.7	4.4	Нет
195	90	40	М	3.3	3.1	Да
160	111	2	Ж	4.0	3.9	Нет
183	143	0	М	4.8	4.7	Да
158	118	25	Ж	3.7	3.5	Да
178	132	10	М	4.4	?	?
170	120	4	Ж	3.9	?	?
165	100	20	М	3.5	?	?

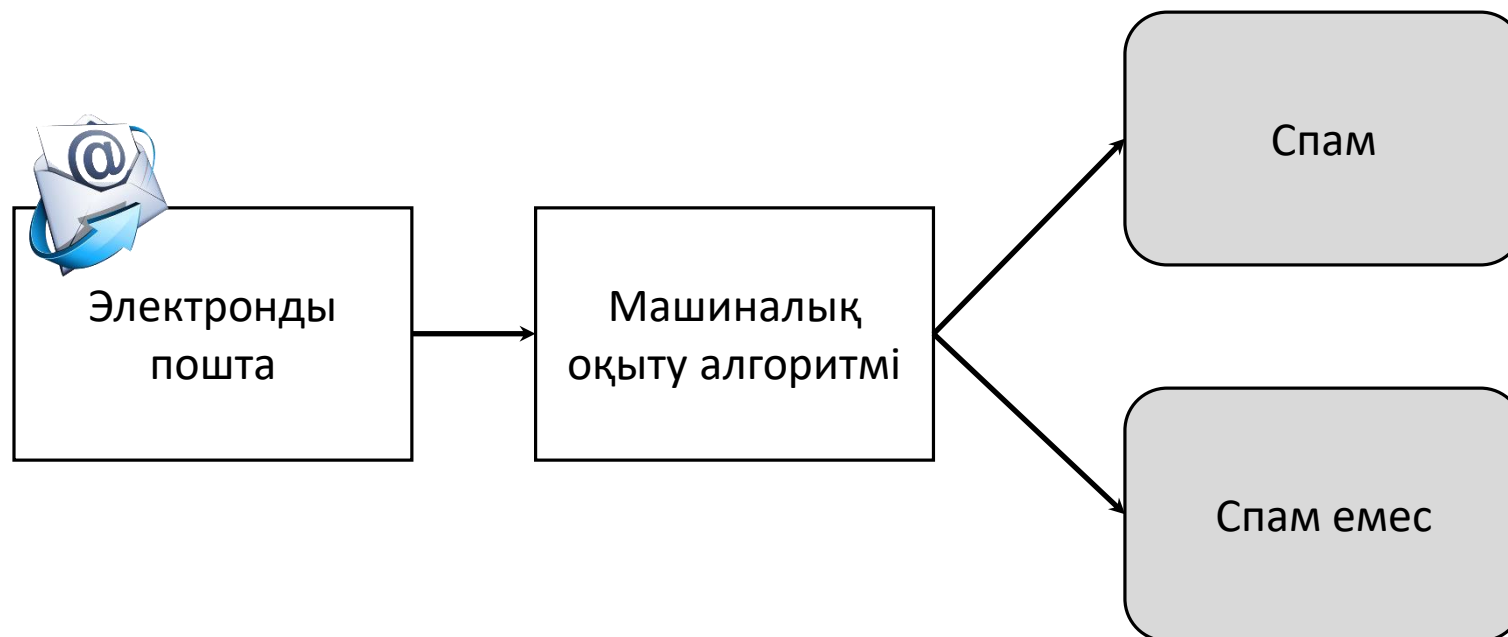
Нысандар **Регрессия** **Классификация**

Трен. Тест.

Бұл есеп қалай шешіледі?



Мысал

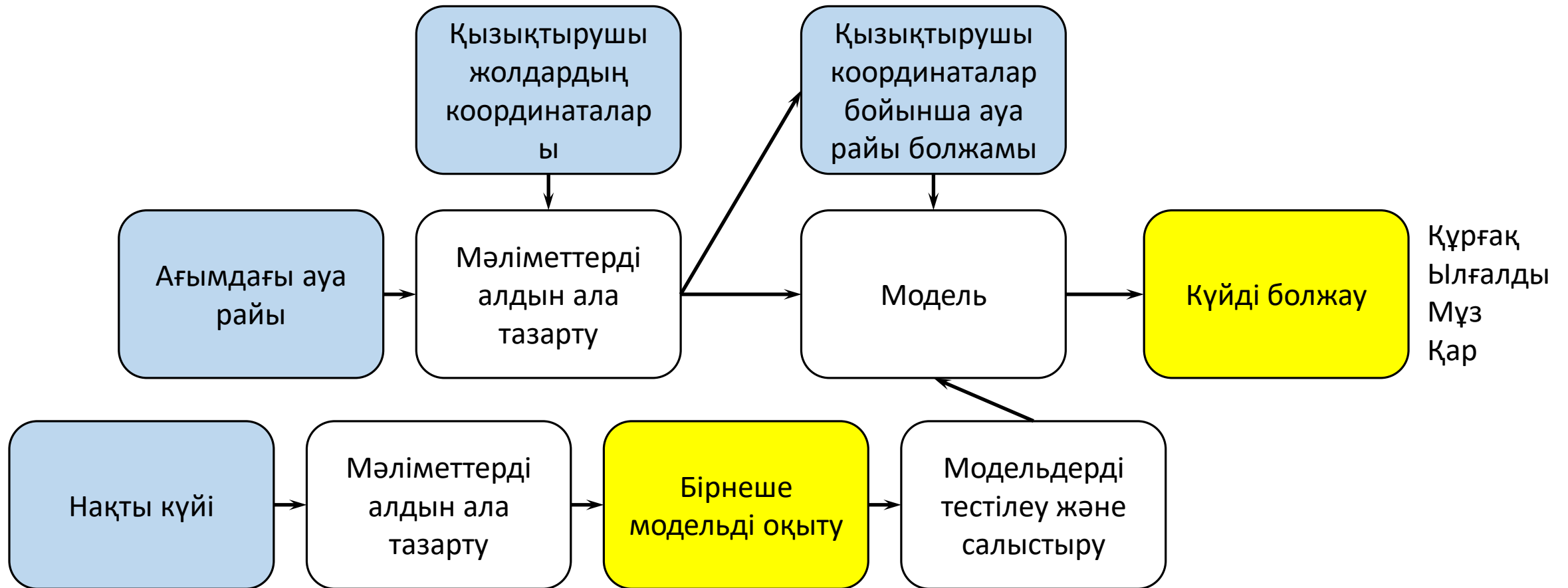


ML тағы қандай есептерде қолданылады?

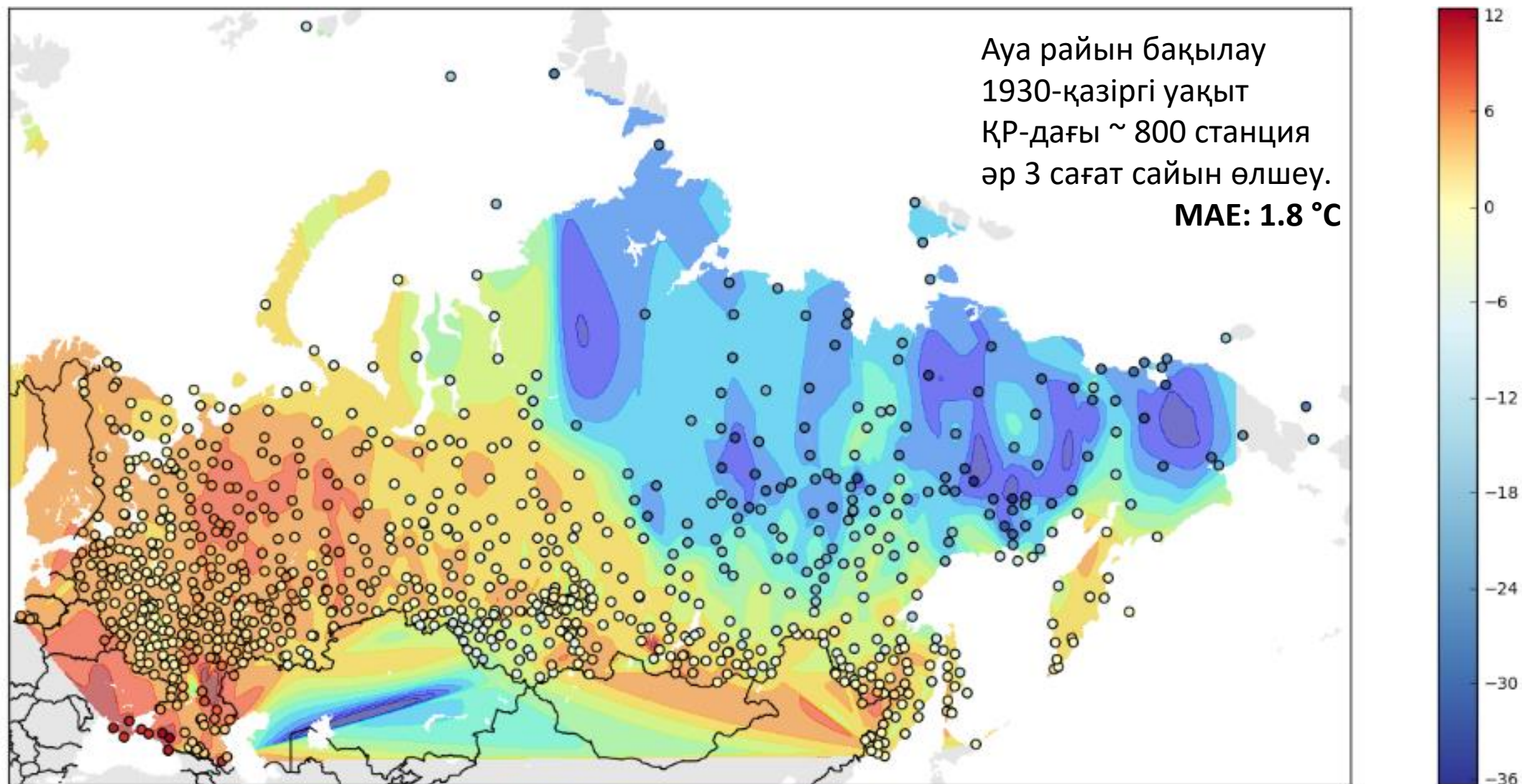
- Жіктеу
- Нақты жауап беру (регрессия)
- Болжау
- Ранжирлеу
- Таңдамаларды сүзу
- Ең маңызды көрсеткіштерді табу
- Ұсыныстар құрастыру және т.б.

Мысал

Жол төсемінің жағдайын 6 және 24 сағатқа болжау

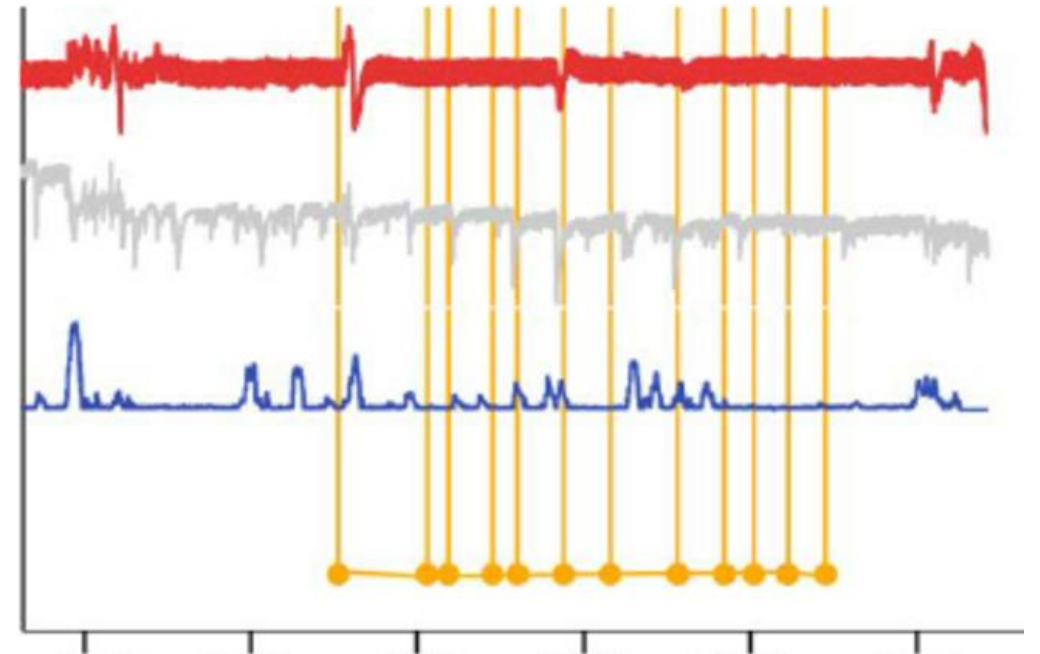
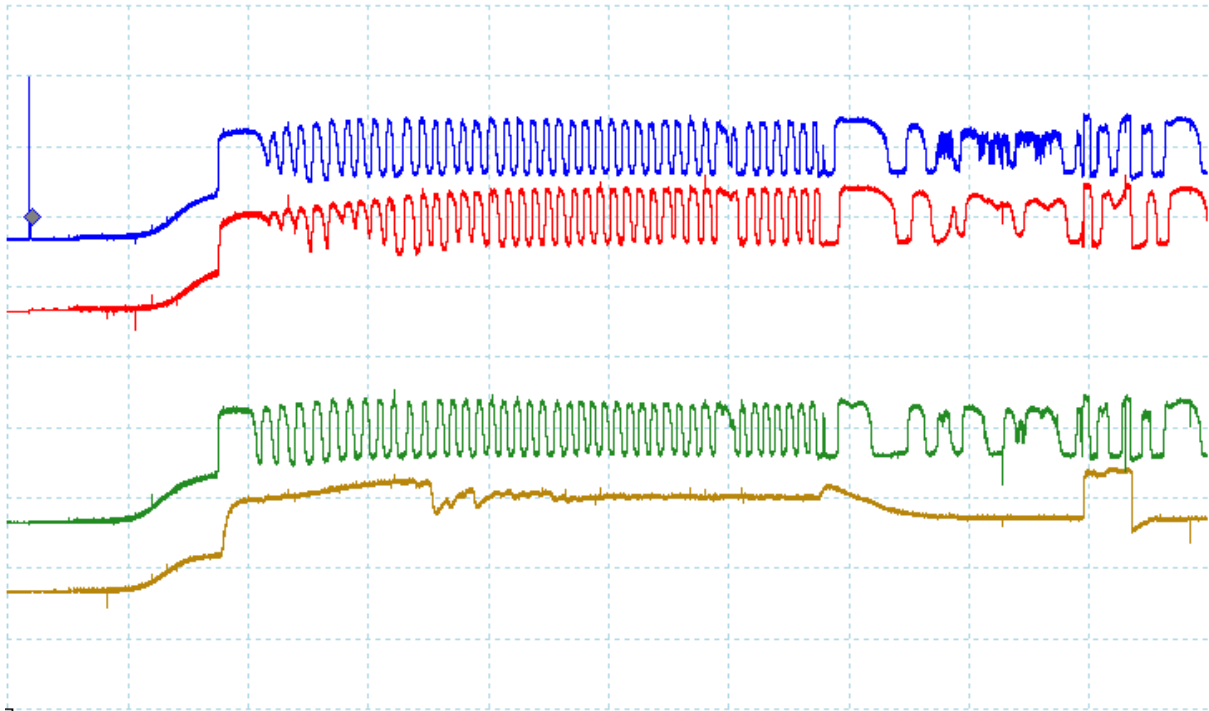


Мысал



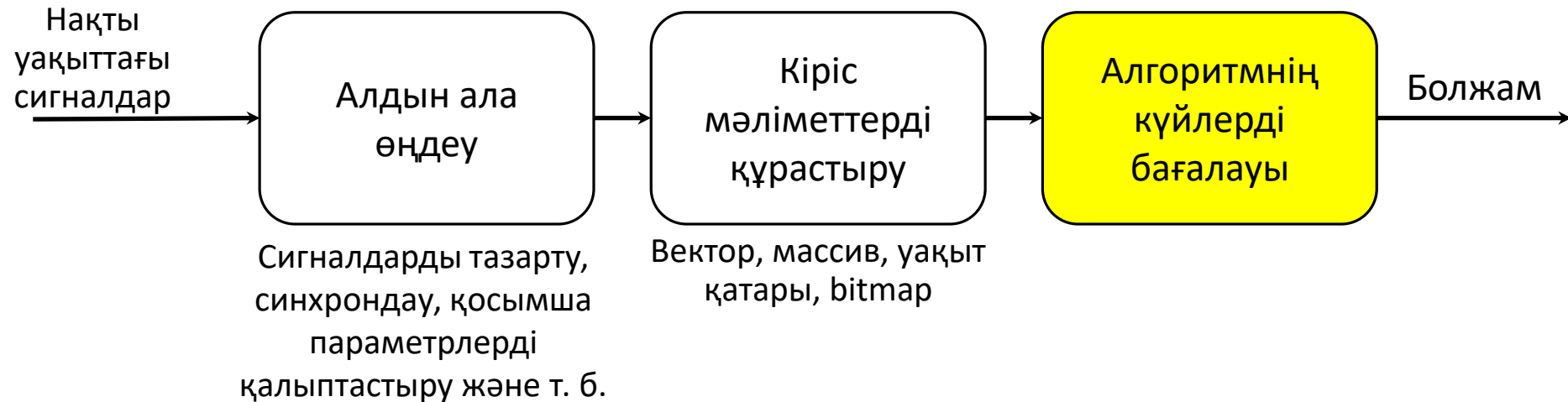
Мысал

Жабдықтың бұзылуын және штаттан тыс жағдайлардың туындауын болжау

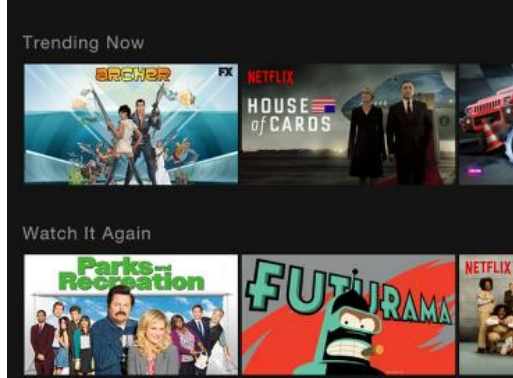
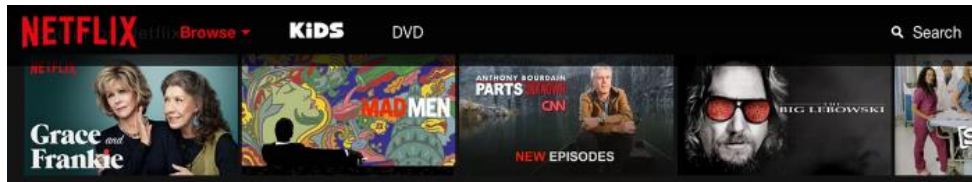


Мысал

Жабдықтың бұзылуын және штаттан тыс жағдайлардың туындауын болжау








ML қолданылуы



10 active competitions Sort By **Prize**

Active All Entered Main Site All Eval Metrics

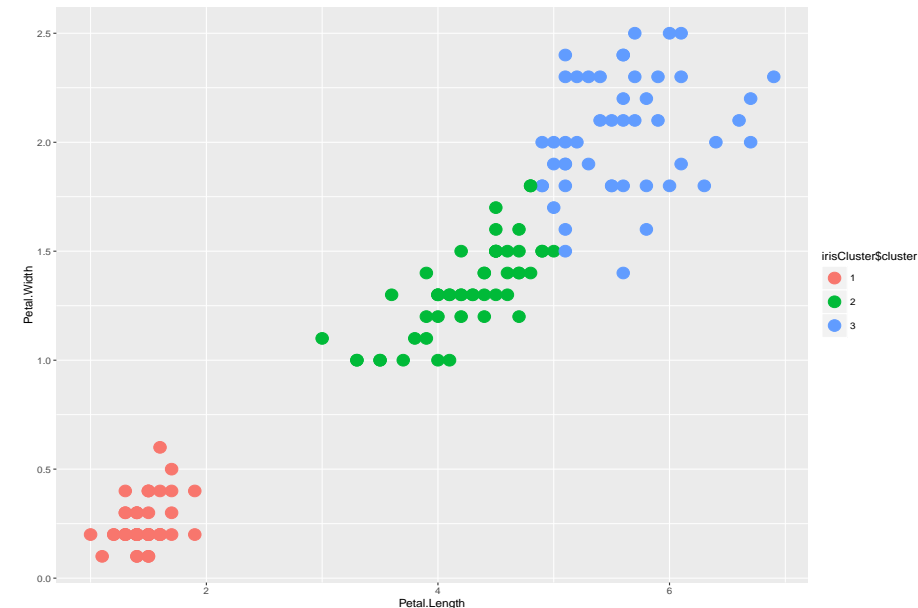
	Predicting Red Hat Business Value Classify customer potential A month to go · Featured	1,202 teams 1,062 kernels \$50,000
	Bosch Production Line Performance Reduce manufacturing failures 3 months to go · Featured	84 teams \$30,000
	TalkingData Mobile User Demographics Get to know millions of mobile device users 13 days to go · Featured	1,479 teams 2,446 kernels \$25,000
	Grupo Bimbo Inventory Demand Maximize sales and minimize returns of bakery goods 7 days to go · Featured	1,955 teams 2,714 kernels \$25,000
	Digit Recognizer Classify handwritten digits using the famous MNIST data 4 months to go · Getting Started	1,028 teams 5,710 kernels Knowledge

Оқытушымен және оқытушысыз МО

- Unsupervised Learning

- Алдын ала анықталған және белгілі нәтижелер жиынтығы жоқ
- Деректердегі жасырын заңдылықтар мен байланыстарды іздеу
- Типтік мысал: кластерлеу

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
1	5.1	3.5	1.4	0.2
2	4.9	3.0	1.4	0.2
3	4.7	3.2	1.3	0.2
4	4.6	3.1	1.5	0.2
5	5.0	3.6	1.4	0.2
6	5.4	3.9	1.7	0.4
7	4.6	3.4	1.4	0.3
8	5.0	3.4	1.5	0.2
9	4.4	2.9	1.4	0.2
10	4.9	3.1	1.5	0.1



Оқытушымен және оқытушысыз МО

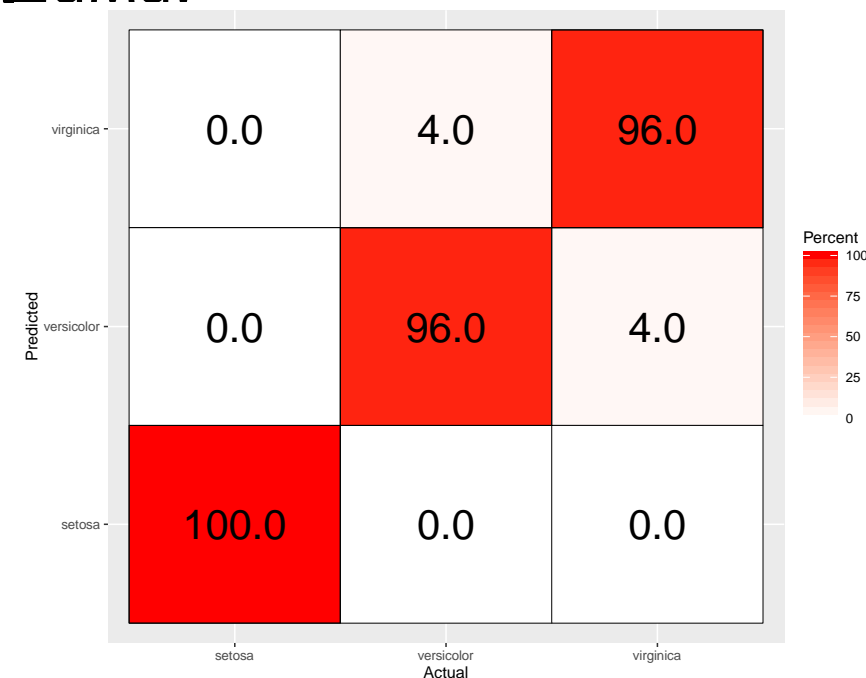
- Supervised Learning

- Әрбір мысал үшін деректерде әрқашан алдын-ала анықталған нәтиже болады
- Сипаттамалық функциялар жиынтығы мен мақсат арасындағы қатынасты модельдейді (деректерді функциямен салыстырады)
- 2 есептер тобы:
 - Классификация
 - Регрессия

Supervised Learning

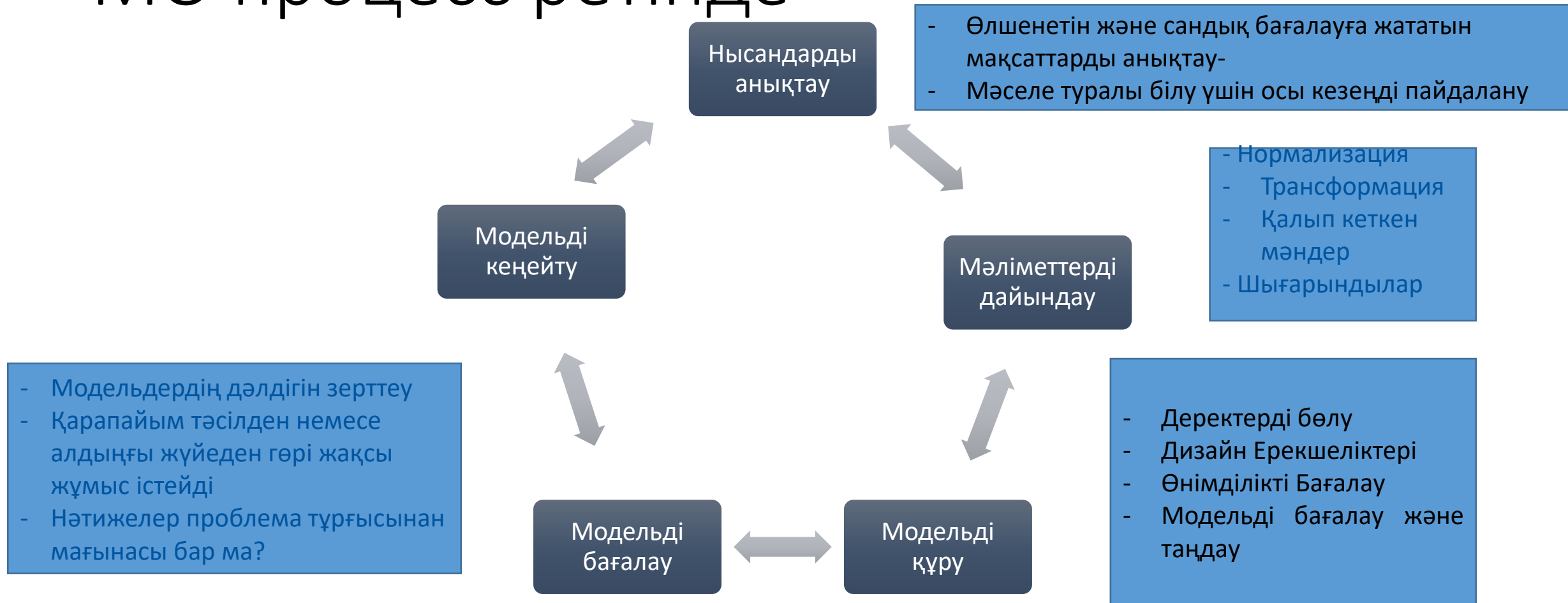
- Классификация
- Бұл деректерді іріктеу қай сыныпқа жататынын болжайды (сипаттамалық белгілерді іріктеу), дискретті шама.

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
1	5.1	3.5	1.4	0.2	setosa
2	4.9	3.0	1.4	0.2	setosa
3	4.7	3.2	1.3	0.2	setosa
4	4.6	3.1	1.5	0.2	setosa
5	5.0	3.6	1.4	0.2	setosa
6	5.4	3.9	1.7	0.4	setosa
7	4.6	3.4	1.4	0.3	setosa
8	5.0	3.4	1.5	0.2	setosa
9	4.4	2.9	1.4	0.2	setosa
10	4.9	3.1	1.5	0.1	setosa



- Регрессия
 - Үзіліссіз шамаларды болжайды

МО процесс ретінде



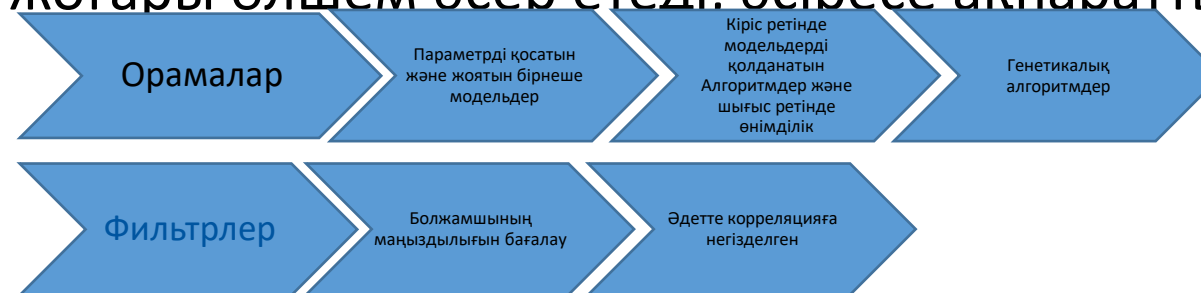
ML процесс ретінде: Мәліметтерді дайындау

- Аталған қажеттіліктің бірнеше себебі бар:
 - Кейбір модельдер деректерге қатаң талаптар қояды
 - Деректер масштабы, деректер нүктелері арасындағы интервалдар және т.б.
 - Кейбір деректер сипаттамалары модельдің жұмысына айтарлықтай әсер етуі мүмкін
 - Деректерді дайындау уақытын бағаламауға болмайды



ML процесс ретінде: Функцияларды әзірлеу

- Пайдаланылатын болжаушыларды (функцияларды) анықтау маңызды мәселелердің бірі болып табылады
- Кейде болжаушыларды қосу керек
- Санын Азайту: Болжаушылардың аз саны неғұрлым түсіндірілген модель және арзанырақ
- Көптеген модельдерге жоғары өлшем әсер етеді, әсіресе ақпараттық емес болжаушылар үшін



ML процесс ретінде: модель құру

- Деректерді бөлу
- Әртүрлі тапсырмалар бойынша деректерді тарату
- Модельдеу
- Оқыту тиімділігін бағалау
- Оқуға, тексеруге және тестілеуге арналған жиынтықтарды анықтау
- Функцияны таңдау (бұрын қабылданған шешімді қарап шығыңыз)
- Өнімділікті Бағалау
- Нәтижелерді визуализациялау-проблемалық кеңістіктің қызықты аймақтарын анықтау
- Статистика және тиімділік көрсеткіштері
- Модельді бағалау және таңдау

MO артықшылықтары

- Жылдам, Дәл, Тиімді
- Көптеген қосымшаларды автоматтандыру
- Нақты қолданудың кең спектрі
- Ақпараттық қауіпсіздік және спам табу есептеріндегі жоғары дәлдік
- Адамның араласуы қажет емес
- Көп өлшемді деректерді өңдеу

МО кемшіліктері

- Қателерді анықтау және түзету өте қиын
- Деректерді Жинау
- Нәтижелерді түсіндіру көп уақыт пен кеңістікті қажет етеді

Назарларыңызға рақмет!