Пәні: Деректерді интеллектуалды талдау және визуализациялау

Data Mining міндеттері

Data Mining - ақпараттық технологиялар нарығының құрамдас бөлігі

Дәріс №2

Шаяхметова А.С., PhD, қауымдастырылған профессор
Жасанды интеллект және Big Data кафедрасы

2026-2027 оқу жылы

Жұмыстың мақсаты

Data Mining ұғымы мен міндеттерінің жіктелуін қарастыру. Негізгі міндеттерді (классификация, кластерлеу, регрессия, ассоциация) сипаттау. Data Mining-тің ақпараттық технологиялар нарығындағы орнын көрсету. Қолдану салалары мен даму бағыттарын талдау.

Жұмыс жоспары

Data Mining ұғымы және міндеттерінің жіктелуі

Data Mining-тің негізгі міндеттері

Data Mining - ақпараттық технологиялар нарығының құрамдас бөлігі

Қолдану салалары және даму бағыттары

Data Mining анықтамасы

Data Mining (деректерді интеллектуалды талдау) - үлкен көлемді шикі деректерден бұрын белгісіз, практикада пайдалы және түсіндірілетін білімді табу процесі.

Деректер

кестелер, мәтін, журналдар, суреттер

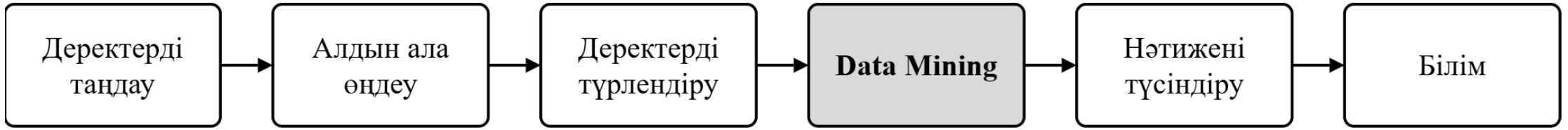
Әдістер

статистика, машиналық оқыту, дерекқорлар

Нәтиже

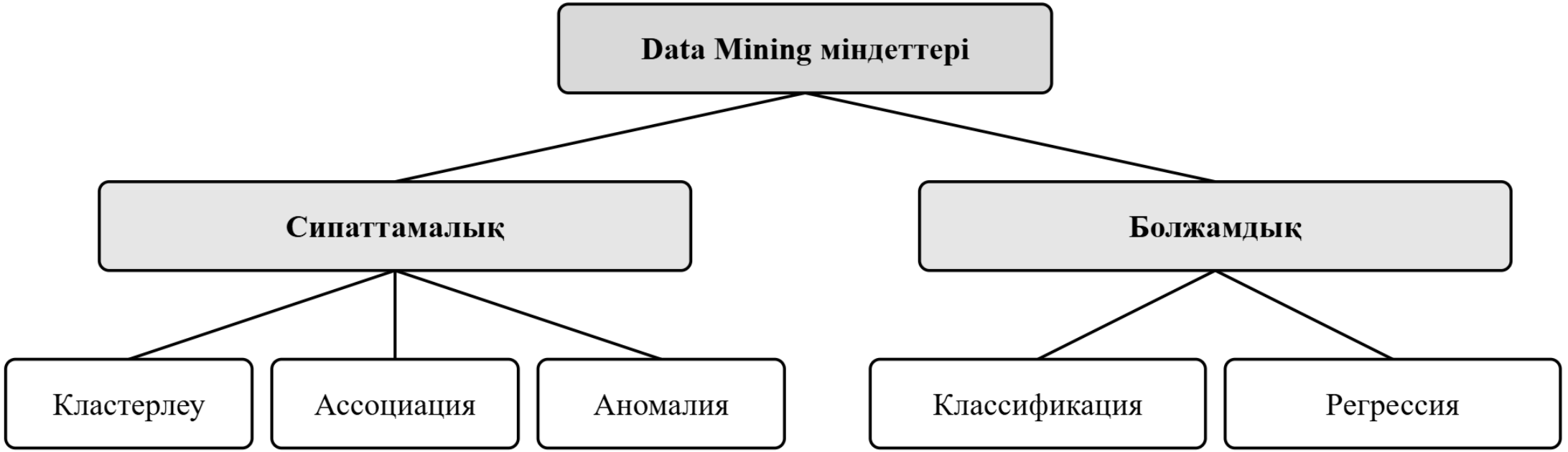
модельдер, заңдылықтар, болжамдар

Деректерден білім табу процесі (KDD)



Data Mining - білімді табудың орталық кезеңі. Оған дейін деректер дайындалады, одан кейін нәтиже бағаланып, түсіндіріледі.

Data Mining міндеттерінің жіктелуі



Сипаттамалық міндеттер деректердегі құрылымды ашады, болжамдық міндеттер белгісіз мәнді алдын ала айтады.

Бақыланатын және бақыланбайтын оқыту

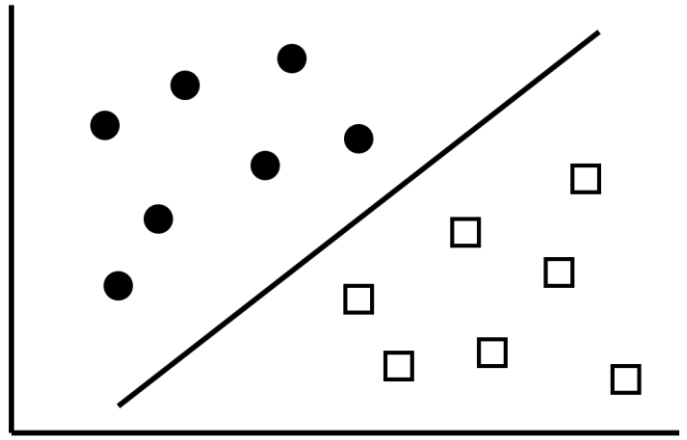
Оқыту түрі	Нәтиже белгілі ме?	Міндеттер
Бақыланатын	Иә, үлгілерде жауап бар	Классификация, регрессия
Бақыланбайтын	Жоқ, құрылым өзі табылады	Кластерлеу, ассоциация, аномалия

Тәсілді таңдау деректерде алдын ала белгіленген жауаптың (белгі, класс) бар-жоғына байланысты.

Классификация

Объектіні алдын ала белгілі сыныптардың біріне жатқызу.
Модель белгісі бар үлгілерден үйренеді.

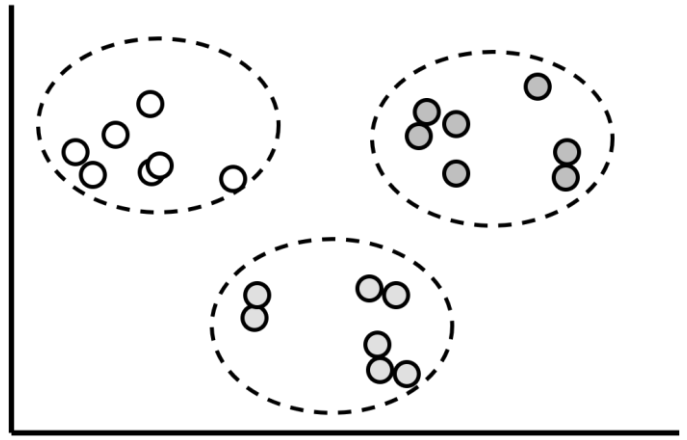
- Мысал: хаттың спам екенін анықтау;
- әдістер: шешім ағашы, k-көрші, нейрондық желі.



Кластерлеу

Алдын ала сыныптарсыз ұқсас объектілерді топтарға біріктіру. Топтар деректердің өзінен табылады.

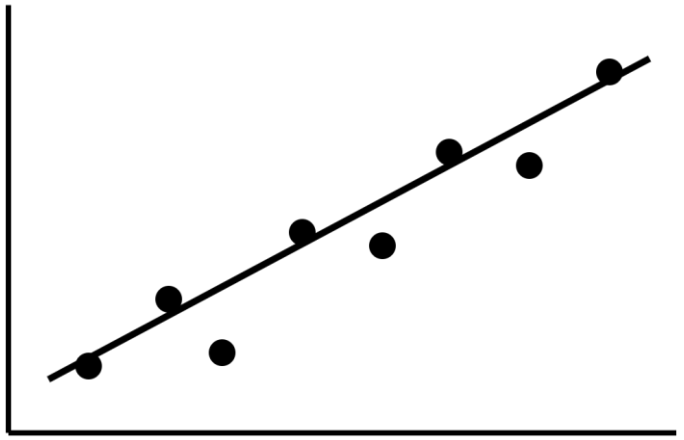
- Мысал: тұтынушыларды сегменттеу;
- әдістер: k-means, иерархиялық кластерлеу.



Регрессия

Сандық мәнді басқа атрибуттар бойынша болжау. Нәтиже - үздіксіз сан.

- Мысал: пәтер бағасын болжау;
- әдістер: сызықтық регрессия, ағаш негізіндегі модельдер.



Ассоциациялық ережелер

Бірге жиі кездесетін элементтерді табу. Ереже түрі: «егер A болса, онда B болады».

Көрсеткіш	Мағынасы
Support (қолдау)	A мен B бірге кездесетін жазбалар үлесі
Confidence (сенімділік)	A бар жазбаларда B-ның кездесу ықтималдығы
Lift (көтерілу)	Ереженің кездейсоқтан қаншалықты күшті екені

Мысал: «нан алған сатып алушы сүт те алады». Әдіс: Apriori алгоритмі.

Басқа міндеттер

Аномалияларды табу

қалыптыдан ерекшеленетін жазбаларды анықтау (алаяқтық)

Жинақтау (summarization)

деректердің қысқа сипаттамасын құру

Тізбектерді талдау

оқиғалар реті бойынша заңдылықтарды табу

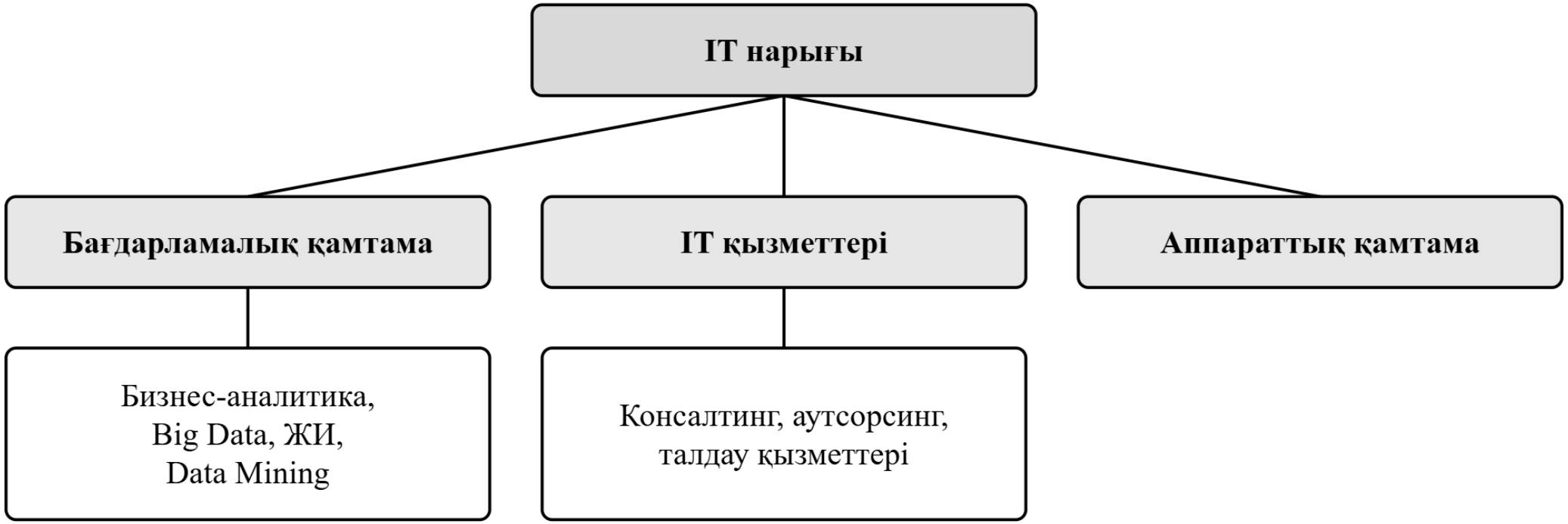
Уақыттық қатарларды болжау

өткен мәндер бойынша келесілерін айту

Міндеттер мен әдістер

Міндет	Әдіс мысалы	Нәтиже
Классификация	Шешім ағаштары, k-көрші	Сынып белгісі
Кластерлеу	k-means	Объектілер топтары
Регрессия	Сызықтық регрессия	Сандық болжам
Ассоциация	Apriori	«Егер-онда» ережелері
Аномалия	Статистикалық әдістер, изоляция ормандары	Күдікті жазбалар

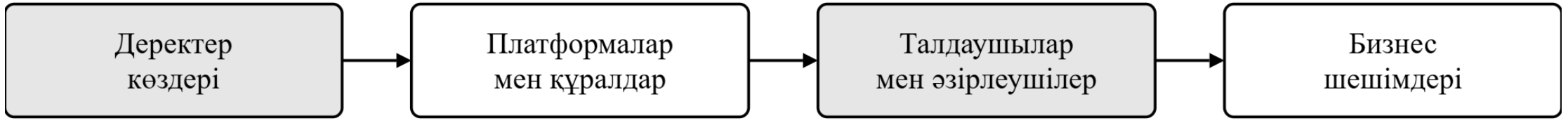
Data Mining-тің IT нарығындағы орны



Data Mining өнімдері мен қызметтері

Санат	Сипаттамасы	Мысал
Аналитикалық платформалар	Визуалды жұмыс ортасы	KNIME, RapidMiner
Бағдарламалау құралдары	Кітапханалар мен тілдер	Python (scikit-learn), R, Weka
Бұлттық сервистер	Модельді бұлтта оқыту	Azure ML, Amazon SageMaker
Қызметтер	Консалтинг, жобалық талдау	Талдау командалары, аутсорсинг

Data Mining құндылық тізбегі



- Деректер көздері - компаниялардың ішкі жүйелері, сенсорлар, интернет;
- құралдар мен платформалар деректерді өңдеуге мүмкіндік береді;
- мамандар модельдер құрып, нәтижені бизнес тіліне аударады.

Саладағы мамандықтар

Data Analyst

деректерді талдайды, есеп пен визуализация жасайды

Data Scientist

модельдер құрып, болжамдар жасайды

Data Engineer

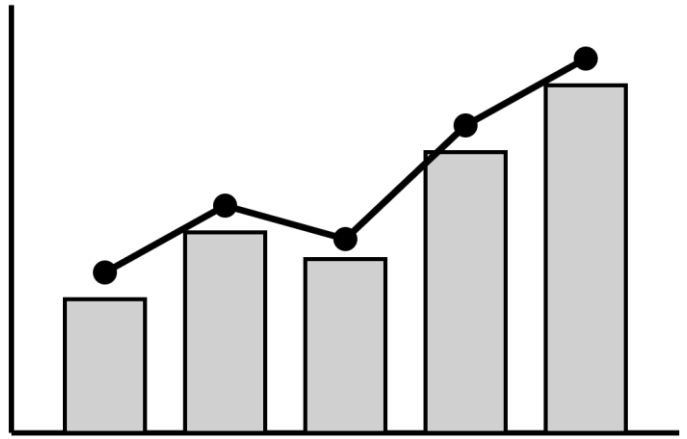
деректер жинау мен өңдеу жүйесін құрады

ML Engineer

модельдерді өндірістік жүйеге енгізеді

Қолдану салалары

Сала	Міндет	Мысал
Банк ісі	Классификация	Несиелік тәуекел
Сауда	Ассоциация	Тауар жиынтығы
Медицина	Классификация	Диагностика
Телеком	Кластерлеу	Абонент сегменті
Қауіпсіздік	Аномалия	Алаяқтықты анықтау



Даму бағыттары және қиындықтар

Мүмкіндіктер

AutoML, бұлттық платформалар, жасанды интеллектпен интеграция

Қиындықтар

деректер сапасы және сақтау құны

Құпиялылық

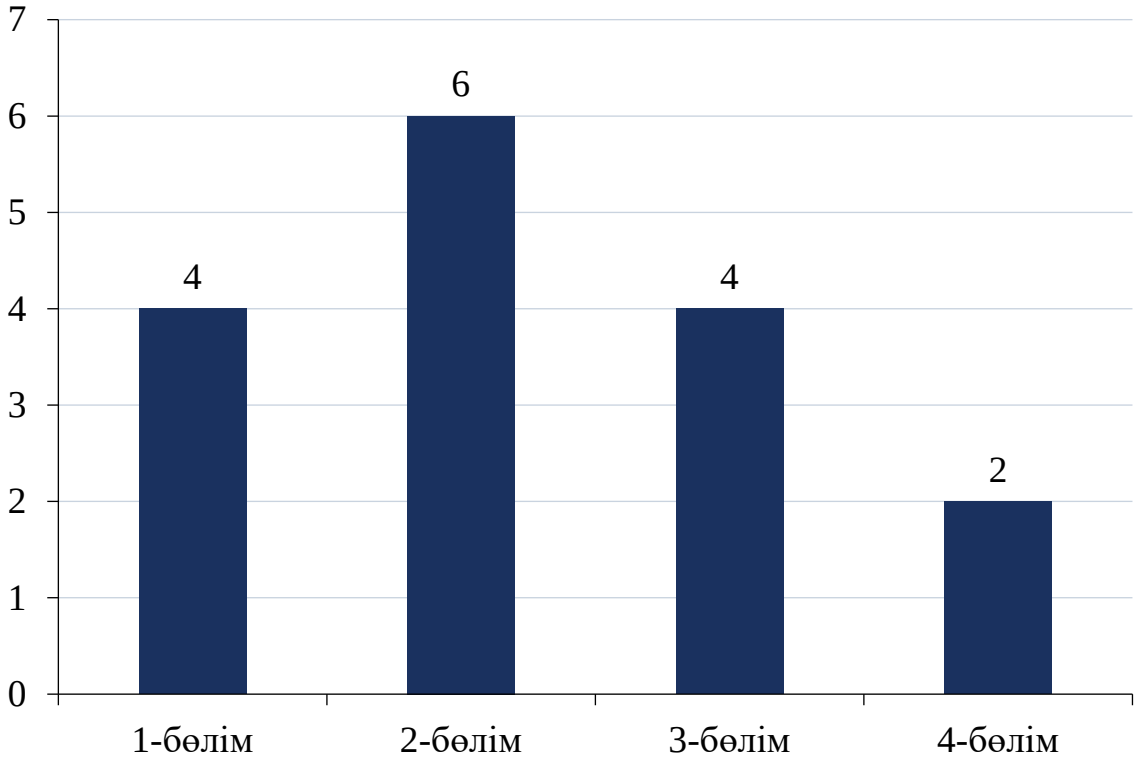
жеке деректерді қорғау талаптары

Түсіндірілетін модельдер

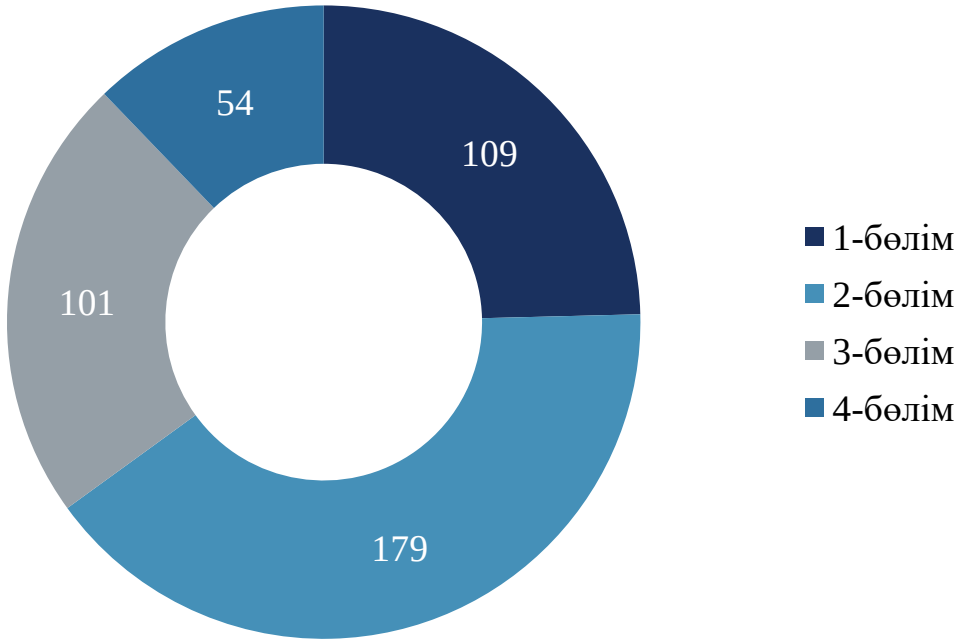
нәтижеге сенімді болу үшін қажет

Диаграммалар

Бөлімдер бойынша слайд саны



Бөлімдер бойынша мәтін көлемі (сөз саны)



Қорытынды

1. Data Mining - деректерден бұрын белгісіз пайдалы білімді табу процесі;
2. Міндеттер сипаттамалық (кластерлеу, ассоциация, аномалия) және болжамдық (классификация, регрессия) болып бөлінеді;
3. Data Mining - IT нарығындағы бағдарламалық қамтама мен қызметтер сегментінің бір бөлігі;
4. Қолдану аясы банк, сауда, медицина, телеком салаларына дейін кеңейіп келеді.



Data Mining міндеттері

Data Mining – ақпараттық технологиялар нарығының құрамдас бөлігі
Дәріс №2