Пәні: Деректерді интеллектуалды талдау және визуализациялау

Интеллектуалды талдаудың негізгі кезеңдері

Міндет қою, деректерді өңдеу, модельдерді құру және енгізу

Дәріс № 12

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

2026-2027 оқу жылы

Дәрістің мақсаты

Мақсаты

Интеллектуалды талдаудың негізгі алты кезеңін түсіндіру және оларды SQL Server құралдарымен орындау ерекшеліктерін көрсету.

Дәріс соңында білім алушы:

- Талдау міндетін нақты тұжырымдап, деректерді дайындау жолдарын сипаттайды.
- Деректерді зерттеу, модель құру және оқыту кезеңдерін түсіндіреді.
- Модель сапасын тексеріп, оны орналастыру мен жаңартудың мәнін біледі.

Дәрістің жоспары

1. Талдау міндетін қою және деректерді дайындау

2. Деректерді зерттеу және модельдерді құру

3. Модельдерді тексеру және бағалау

4. Модельдерді орналастыру және жаңарту

1. Талдау міндетін қою және деректерді дайындау

Талдаудың мақсатын, болжамды көрсеткішті және күтілетін нәтижені анықтау.

Деректер көздерін таңдау, олардың өзара байланысын талдау.

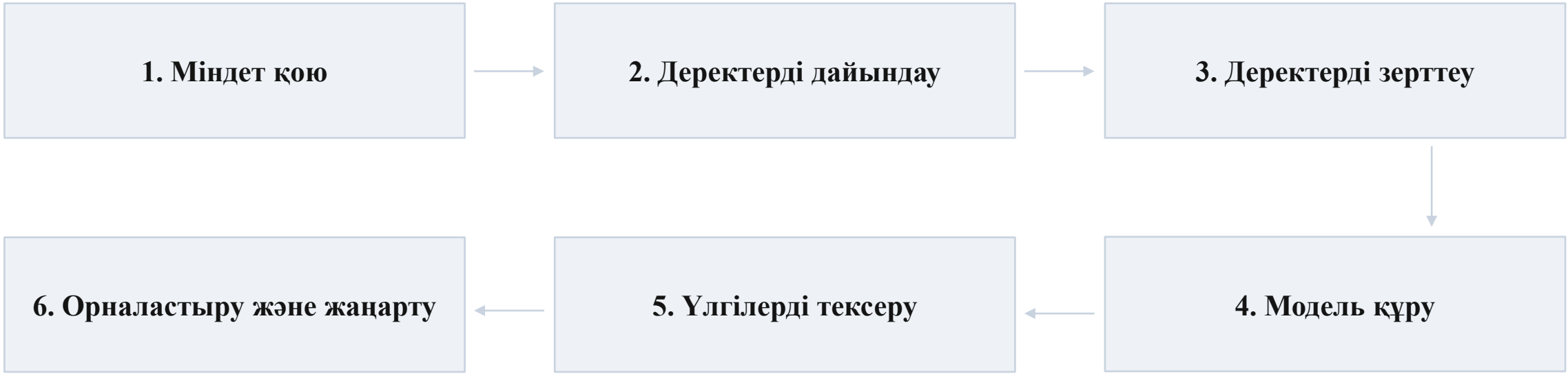
Бірнеше форматтағы ақпаратты біріктіру және түрлендіру.

Қате, жетіспейтін және аномальды мәндерді тазарту.

Міндет дұрыс қойылса, модель құру кезеңі де нәтижелі болады.

Интеллектуалды талдаудың алты негізгі кезеңі

Лекцияда Microsoft ұсынған алты кезеңдік тәсіл қарастырылады. Кезеңдер арасындағы байланыс итерациялық сипатқа ие.



Қажет болса, жаңа ақпарат алу үшін алдыңғы кезеңдерге қайта оралады.

Талдау кезеңдерінің өзара байланысы

Лекциядағы 1-сурет алты кезеңнің байланысын және MS SQL Server қызметтерімен орындалуын көрсетеді.



Сурет 1. Деректерді интеллектуалды талдау кезеңдері

Назар аударатын тұстар

- Кезеңдер бірінен соң бірі ғана жүрмейді.
- Деректер жеткіліксіз болса, міндет қайта анықталады.
- Шешім талдаушы мен пәндік сала мамандарының жұмысын қажет етеді.

Талдау міндетін дұрыс тұжырымдау

Бірінші кезеңде деректерден қандай нәтиже алынатыны және оның қай мәселені шешетіні анықталады.

Негізгі сұрақ	**Нақтылайтын мазмұн**
Не анықтау керек?	Болжау ма, әлде деректердегі заңдылықтарды табу ма?
Нені болжаймыз?	Болжамға арналған мақсатты атрибут қайсы?
Бағандар қалай байланысты?	Бір немесе бірнеше кестенің байланысын анықтау.
Дерек өкілді ме?	Маусымдылықты, үлестірімді және пәндік салаға сәйкестігін тексеру.

Нәтиже: талдаудың мақсаты және деректерге қойылатын талаптар нақтыланады.

Деректерді дайындау және біріктіру

Екінші кезеңде бірнеше ортадағы ақпарат бір жиынтыққа біріктіріліп, талдауға жарамды түрде түрлендіріледі.



SSIS деректерді біріктіру және түрлендіру әрекеттерін автоматтандыруға көмектеседі.

Деректерді тазарту және түрлендіру

Біріктірілген мәліметтерде қателер, жоқ мәндер және шектен тыс бақылаулар болуы мүмкін. Олар модель сапасына әсер етеді.

Өңдеу түрі	**Мәселе**	**Қолданылатын әрекет**
Қате мәндер	Дұрыс емес жазбалар	Түзету немесе алып тастау
Жоқ мәндер	Кей атрибуттар бос	Жоқ мәндерді өңдеу
Шығарындылар	Аномальды, шектен тыс мәндер	Тексеру және қажет болса жою
Түрлендіру	Өлшемдер әртүрлі	Қалыпқа келтіру немесе басқа сандық түрлендіру

Деректерді тазарту нәтижесі - біркелкі және талдауға қолайлы жиынтық.

Міндет қою мен дерек дайындаудың нәтижелері

Өлшем	Міндет қою	Деректерді дайындау
Негізгі мақсат	Не анықталатынын белгілеу	Қажетті мәліметтерді қалыптастыру
Негізгі сұрақ	Қай нәтижеге жету керек?	Ақпаратты қайдан және қалай аламыз?
Орындалатын жұмыс	Болжау нысаны, байланыстар мен маусымдылықты талдау	Жинау, біріктіру, тазарту, түрлендіру
Кезең нәтижесі	Нақты тұжырымдалған есеп	Сапасы тексерілген деректер жиынтығы

Бастапқы кезеңдер сапалы орындалмаса, келесі модель сенімсіз болады.

2. Деректерді зерттеу және модельдерді құру

Бөлімнің негізгі мәселелері

Жиынтықтың пәндік саланы қаншалықты дұрыс сипаттайтынын бағалау.

Ең аз, ең көп мәндерді және статистикалық сипаттамаларды зерттеу.

Бір құрылым үшін бірнеше модель мен сүзгіні қарастыру.

Алгоритмді таңдау, модельді оқыту және нәтижелерді сақтау.

Бұл кезеңнің өзегі - дерек құрылымынан пайдалы заңдылықты алу.

Деректерді зерттеу: негізгі әрекеттер

Үшінші кезең дайындалған деректердің зерттелетін пәндік салаға сәйкестігін тексеруге бағытталған.

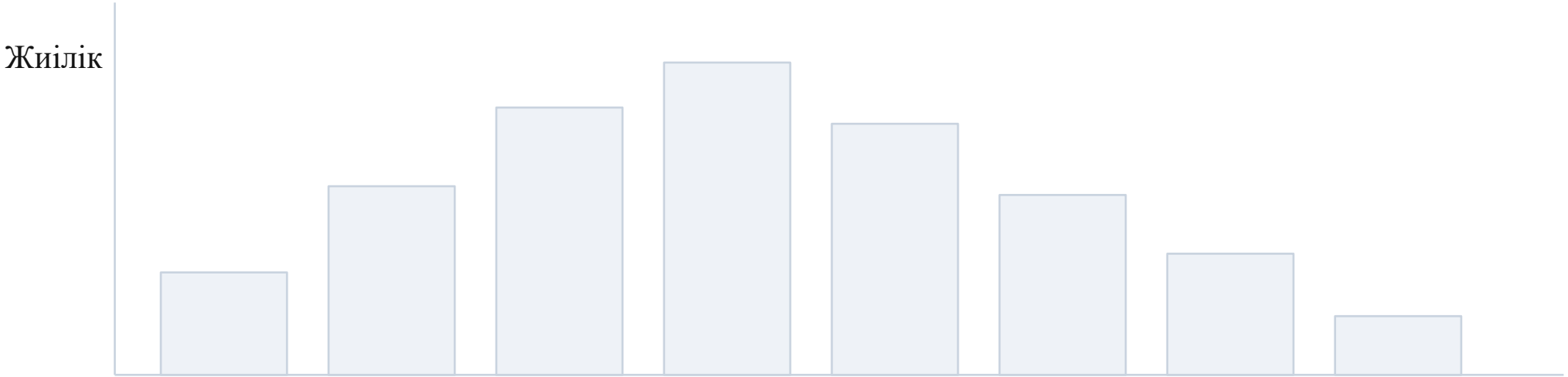
Талдау әрекеті	Нені көрсетеді?	Нәтижеге әсері
Минимум және максимум	Атрибуттың мәндер диапазоны	Күтілмеген шектен тыс мәнді табу
Таралуды зерттеу	Көрсеткіштердің жиілігі мен үлестірімі	Құрылымды және ерекшеліктерді байқау
Статистикалық сипаттама	Жалпы сандық қасиеттер	Деректерді салыстырып сипаттау
Пәндік саламен салыстыру	Нақты құбылысқа сәйкестік	Модельге дейін қателікті аңғару

Талдаудың бұл кезеңі модель құруға негіз болатын сапалы сипаттама береді.

Деректер таралуын визуалды зерттеу

Таралу кестелері мен диаграммалар атрибуттар мәндерінің қалай өзгеретінін түсінуге көмектеседі.

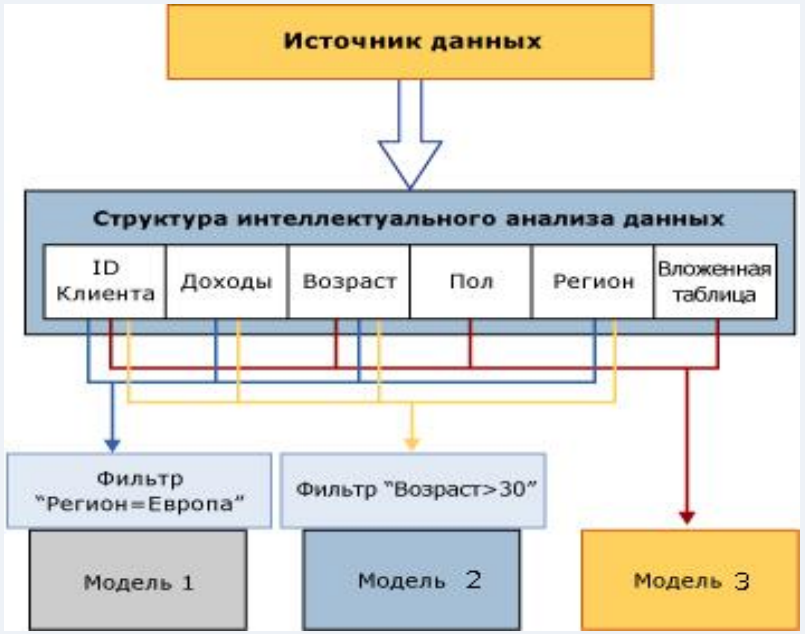
Гистограмма - көрсеткіш мәндерінің таралуына мысал



Ескерту: гистограмма түсіндіруге арналған шартты мысал.

Бір деректер құрылымынан бірнеше модель жасау

Төртінші кезеңде алдымен деректер құрылымы анықталып, содан кейін әртүрлі атрибуттар мен сүзгілер негізінде модельдер жасалады.



Сурет 2. Бір құрылымда жасалған бірнеше модель

Лекциядағы негізгі тұжырым

- Бір құрылым бірнеше модельге ортақ болады.
- Әр модельдің бағандар жиынтығы әртүрлі болуы мүмкін.
- Қажетті жазбалар сүзгі арқылы таңдалады.

Модель құрылымы, бағандар және сүзгілер

Талдауға арналған деректер құрылымы

Модель 1	Модель 2	Модель 3
Бағандар: A, B, C	Бағандар: A, C, D	Бағандар: B, C, E
Сүзгі: аймақ бойынша	Сүзгі: жас бойынша	Басқа атрибуттар

Сызба шартты: нақты бағандар мен сүзгілер есеп мазмұнына сай таңдалады.

Модельді құру және оқыту логикасы

Модель деректерді талдау алгоритмін, оның параметрлерін және қолданылатын мәліметтерді қамтиды.

Деректер құрылымы

Бағандарды және сүзгілерді белгілеу

→

Алгоритм және параметр

Модель тәсілін нақтылау

→

Оқыту жиынтығы

Алгоритмге деректер беру

→

Оқытылған модель

Анықталған заңдылықтарды сақтау

Модельді оқыту аяқталған соң, деректерден табылған заңдылықтар модельде сақталады.

Microsoft құралдары және модельдерді салыстыру

Құрал немесе тіл	Лекциядағы міндеті
BI Development Studio	Жаңа интеллектуалды талдау моделін құру ортасы.
Data Mining Wizard	Модельді шебер арқылы анықтау.
DMX	Деректерді интеллектуалды талдау модельдерімен жұмыс тілі.
Analysis Services	Модель құру, оқыту және тексеруді ұйымдастыру.

Бірнеше алгоритмді салыстыру ең тиімді модельді таңдауға мүмкіндік береді.

3. Модельдерді тексеру және бағалау

Бөлімнің негізгі мәселелері

Модельді өндірістік ортада қолданар алдында оның сапасын анықтау.

Деректерді оқыту және тексеру жиынтықтарына бөлу.

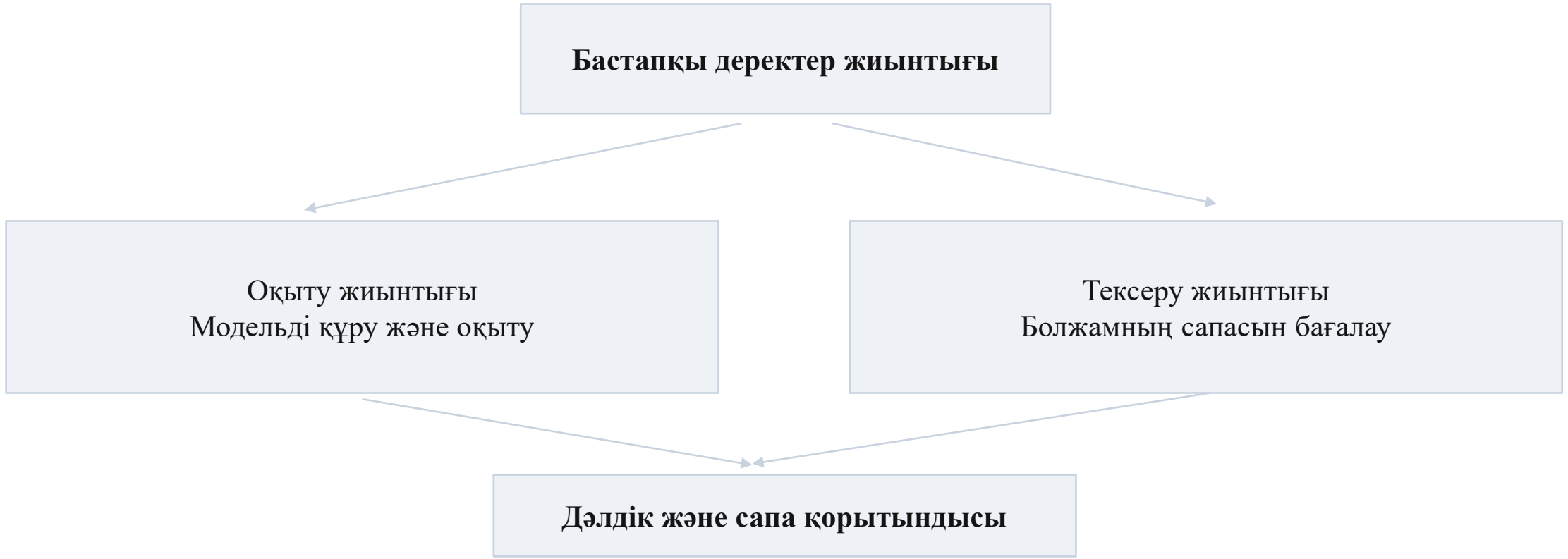
Болжау дәлдігі диаграммасы мен жіктеу матрицасын пайдалану.

Кросс-тексеру арқылы нәтижелердің тұрақтылығын бағалау.

Модельді тексерудің мақсаты - сенімді нәтиже беретін нұсқаны таңдау.

Оқыту және тексеру жиынтықтары

Модельді бағалау үшін болжанатын атрибуттың нақты мәндері белгілі бақылаулар қолданылады.



Модельді тексеру құралдары

Лекцияда болжам сапасын бағалаудың екі көрнекі құралы атап өтіледі: болжам дәлдігі диаграммасы және жіктеу матрицасы.

Болжам дәлдігі диаграммасы

Нақты / болжанған мән

Жіктеу матрицасының үлгісі

Нақты / Болжам	Оң	Теріс
Оң	TP	FN
Теріс	FP	TN

TP, TN - дұрыс; FP, FN - қате жіктеулер.

Көрсетілген сызбалар - бағалау құралдарының шартты иллюстрациялары.

Кросс-тексеру (Cross-validation)

Кросс-тексеру кезінде деректер бірнеше ішкі жиынға бөлінеді. Модель әр бөлікті тексеру үшін қолданып, нәтижелер салыстырылады.

Төрт бөлікке бөлудің шартты сызбасы

Тексеру 1	Тексеру	Оқыту	Оқыту	Оқыту
Тексеру 2	Оқыту	Тексеру	Оқыту	Оқыту
Тексеру 3	Оқыту	Оқыту	Тексеру	Оқыту
Тексеру 4	Оқыту	Оқыту	Оқыту	Тексеру

Кросс-тексеру болжау және сипаттамалық есептерде де қолданылады.

4. Модельдерді орналастыру және жаңарту

Тексеруден өткен тиімді модельді өндірістік ортаға енгізу.

Модельді пайдаланушы қолданбаларымен біріктіру.

Нәтижені сұранысқа жауап беретін есеп түрінде ұсыну.

Пәндік сала өзгерсе, модельді қайта оқыту немесе жаңасын жасау.

Модель пайдалы болуы үшін оны қолдану және уақытында жаңарту керек.

Модельді енгізу, есеп беру және жаңарту

Ең тиімді модель таңдалған соң, ол өндірістік ортада қолданылады және талдау нәтижелері пайдаланушыға есеп түрінде беріледі.

Тексерілген модель
Бағалаудан өткен нұсқа

→

Өндірістік орта
Қолданбамен интеграция

→

Пайдаланушы есебі
Reporting Services

Өзгерген деректерге жауап

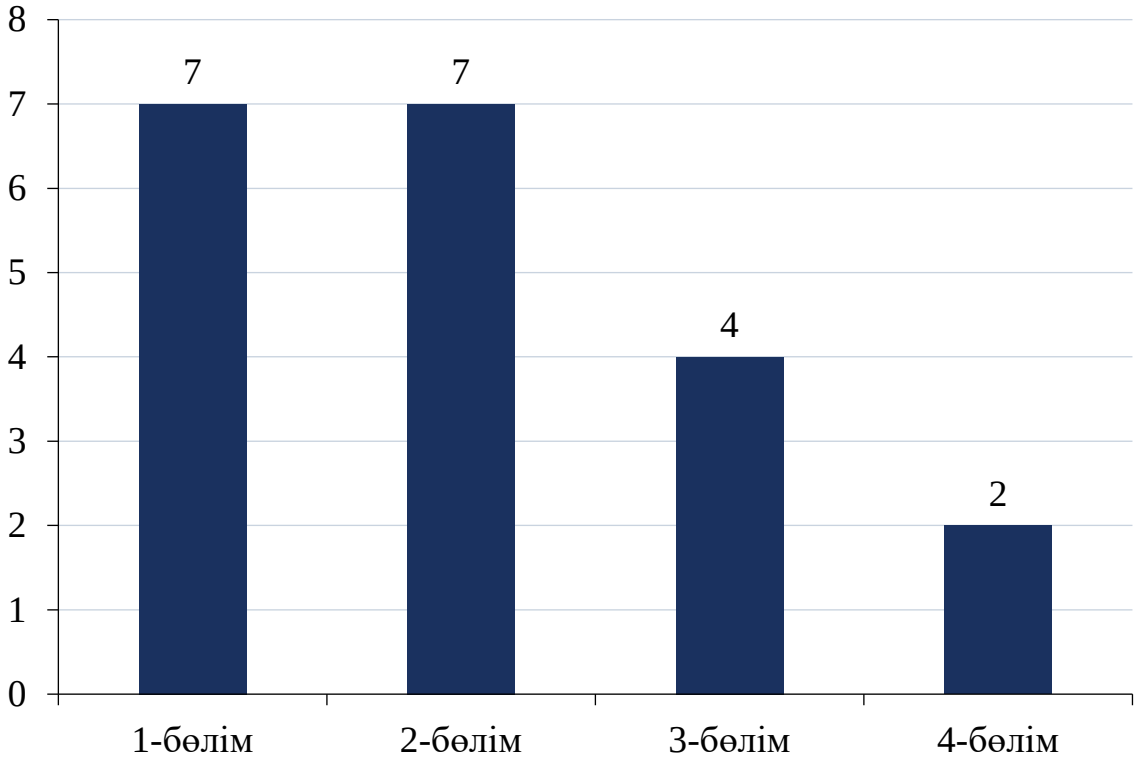
Модельді қайта оқыту

Жаңа модель әзірлеу

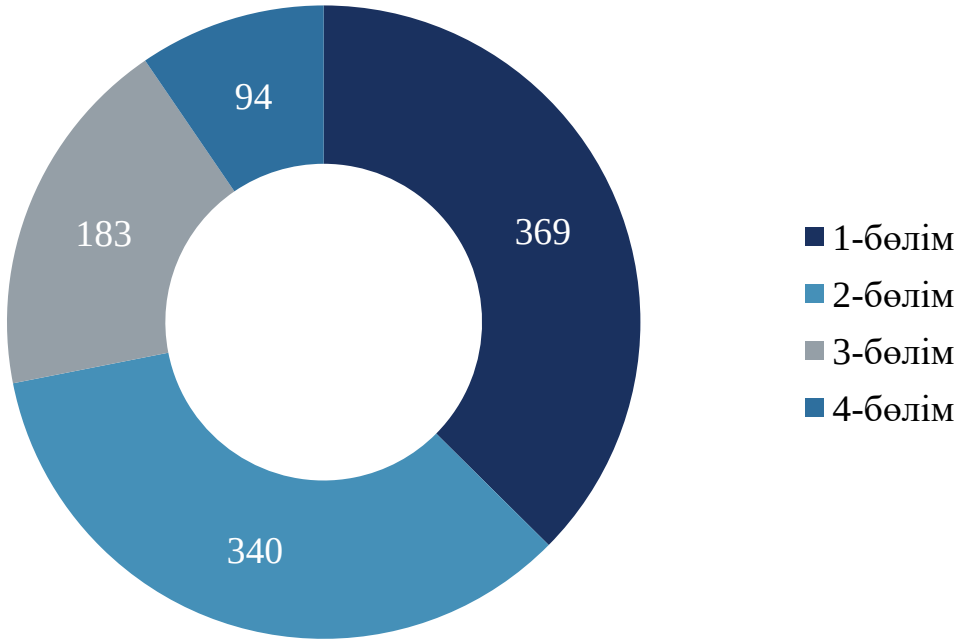
SQL Server Integration Services жаңарту үдерісін автоматтандыруы мүмкін.

Диаграммалар

Бөлімдер бойынша слайд саны



Бөлімдер бойынша мәтін көлемі (сөз саны)



Қорытынды

Интеллектуалды талдау алты кезеңнен тұрады және қажет болғанда алдыңғы кезеңге қайта оралады.
Талдау міндеті мен сапалы деректер модельдің пайдалы болуына негіз болады.

Деректерді зерттеу, модель құру және салыстыру арқылы заңдылықтар анықталады.
Тексеру, орналастыру және тұрақты жаңарту модельдің практикалық қолданылуын қамтамасыз етеді.

Негізгі түйін: интеллектуалды талдау - деректерді дайындаудан бастап енгізуге дейінгі тұтас цикл.

Интеллектуалды талдаудың негізгі кезеңдері

Міндет қою, деректерді өңдеу, модельдерді құру және енгізу
Дәріс №12