Пәні: Деректерді интеллектуалды талдау және визуализациялау

Деректерді өңдеудің негізгі кезеңдері

Дәріс №13

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

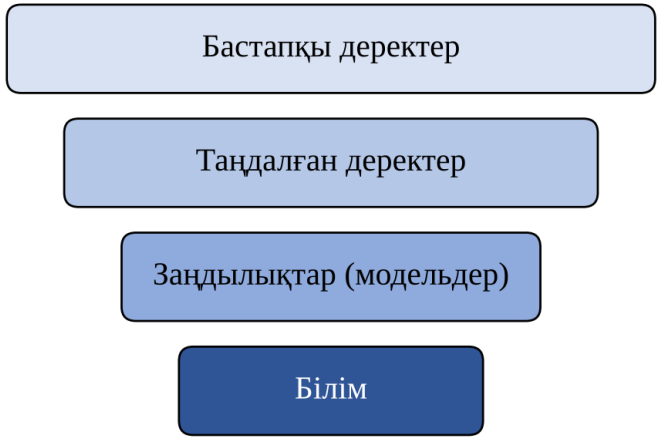
2026-2027 оқу жылы

Дәрістің мақсаты

Data Mining процесінің негізгі кезеңдерін, олардың мазмұнын, орындалатын әрекеттерін және өзара байланысын меңгеру

Міндеттері:

- еркін іздеу кезеңінде заңдылықтардың қалай анықталатынын түсіну;
- болжамдық модельдеудегі жіктеу және болжау есептерін ажырату;
- индуктивті және дедуктивті тұжырымдардың айырмашылығын білу;
- ауытқуларды талдау тәсілдерін үйрену.



Лекция жоспары

Data Mining процесінің кезеңдері

Еркін іздеу (Discovery)

Болжамдық модельдеу (Predictive Modeling)

Ерекше жағдайларды талдау (Forensic Analysis)

1. Data Mining процесінің кезеңдері

1. Data Mining процесінің кезеңдері

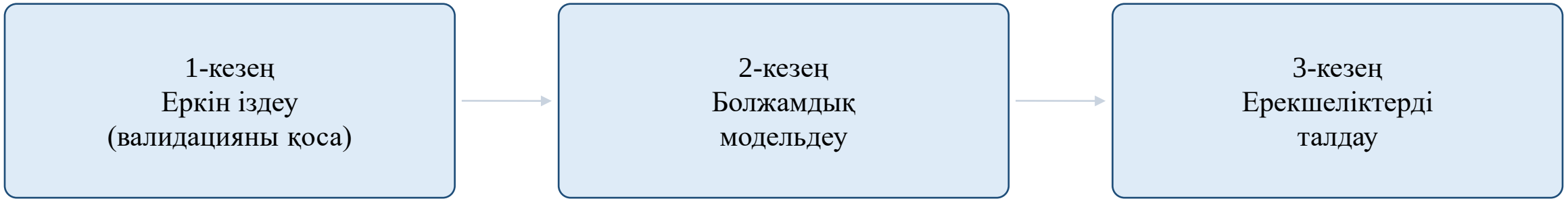
2. Еркін іздеу (Discovery)

3. Болжамдық модельдеу (Predictive Modeling)

4. Ерекше жағдайларды талдау (Forensic Analysis)

Data Mining процесінің кезеңдері

Data Mining процесі екі немесе үш кезеңнен тұруы мүмкін. Толық түрде ол келесі кезеңдер тізбегімен ұсынылады:



- 1-кезең - заңдылықтарды анықтау (еркін іздеу);
- 2-кезең - белгісіз мәндерді болжау үшін анықталған заңдылықтарды қолдану;
- 3-кезең - заңдылықтардағы ауытқуларды анықтау және түсіндіру.

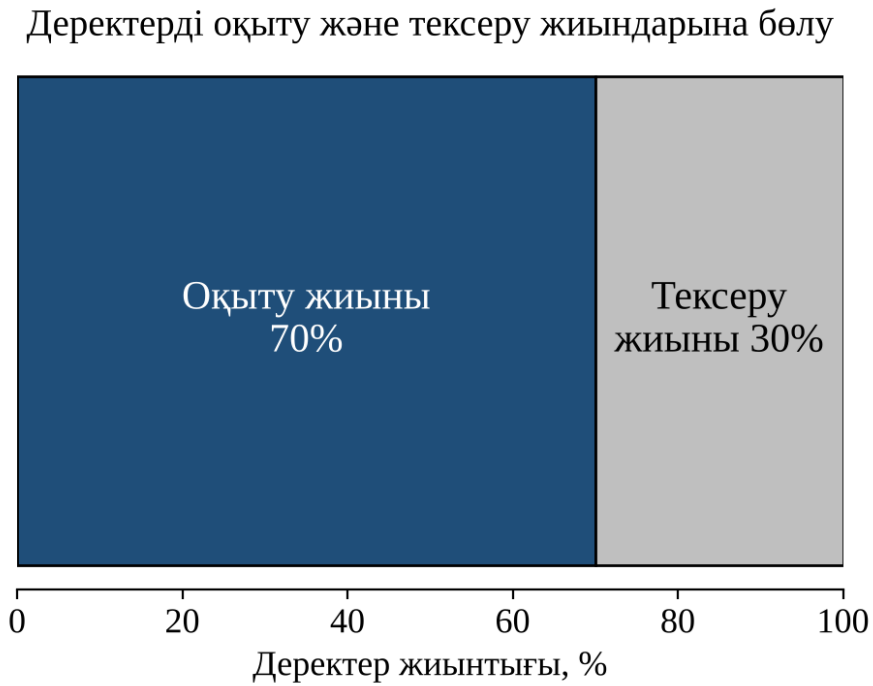
Кезеңдердің салыстырмалы сипаттамасы

Кезең	Ағылшынша атауы	Мақсаты	Негізгі әрекет
Еркін іздеу	Discovery	Жасырын заңдылықтарды табу	Ережелерді индукциялау
Болжамдық модельдеу	Predictive Modeling	Белгісіз мәндерді болжау	Жіктеу және болжау
Ерекшеліктерді талдау	Forensic Analysis	Ауытқуларды анықтау және түсіндіру	Deviation detection

Әр кезең алдыңғы кезеңнің нәтижесіне сүйенеді.

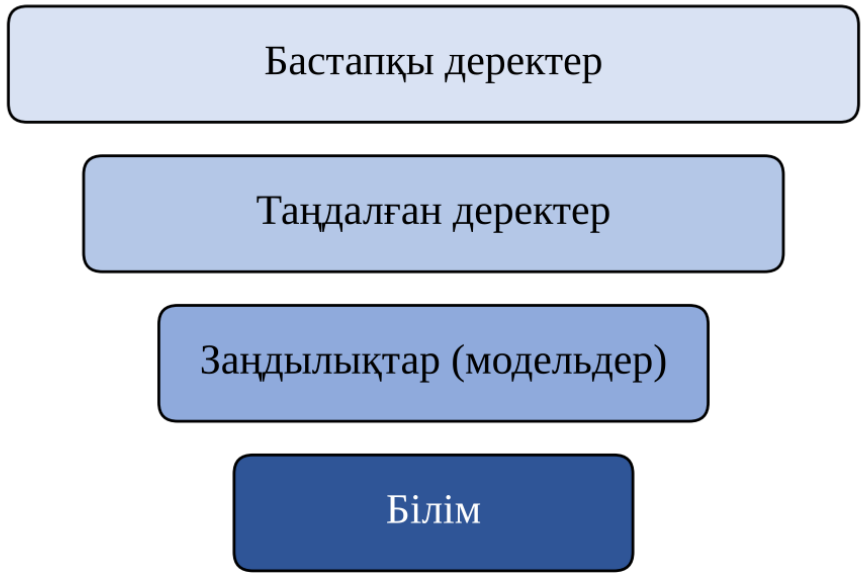
Валидация - заңдылықтардың дұрыстығын тексеру

- Кейде еркін іздеуден кейін жеке тексеру кезеңі енгізіледі.
- Тексерудің мақсаты - табылған заңдылықтардың дұрыстығын тексеру.
- Валидацияны бірінші кезеңнің бөлігі деп санаймыз.
- Нейрондық желілер мен шешім ағаштарында деректер оқыту және тексеру жиындарына бөлінеді.



Деректерді бөлу мысалы

Деректерден білімге дейінгі жол



Data Mining нәтижесінің қалыптасуы

- Бастапқы деректер- дерекқордағы үлкен көлемді жазбалар.
- Таңдалған деректер - талдауға қажетті бөлігі.
- Заңдылықтар - еркін іздеу кезеңінде табылған ережелер мен модельдер.
- Білім - болжау мен шешім қабылдауға қолданылатын нәтиже.

2. Еркін іздеу (Discovery)

1. Data Mining процесінің кезеңдері

2. Еркін іздеу (Discovery)

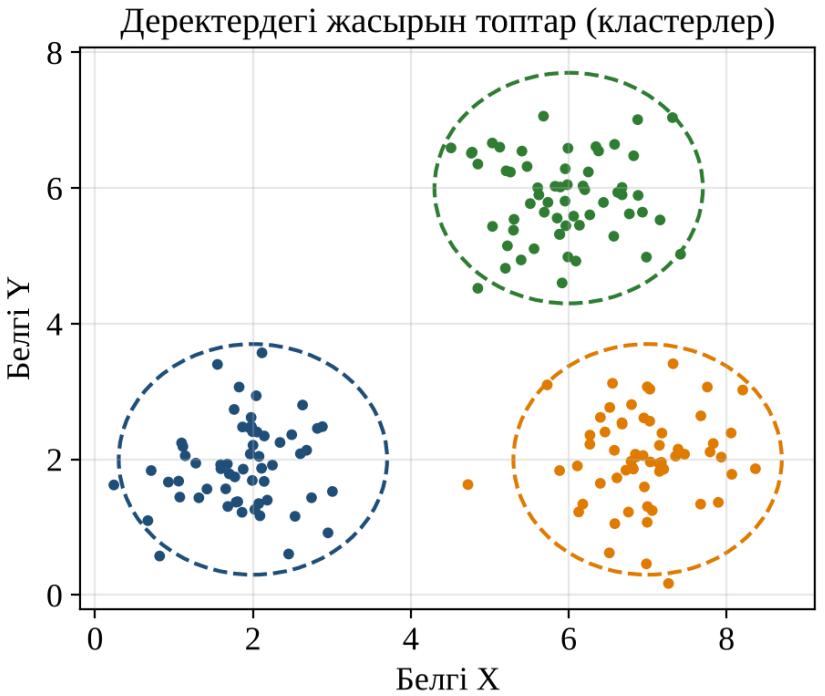
3. Болжамдық модельдеу (Predictive Modeling)

4. Ерекше жағдайларды талдау (Forensic Analysis)

Еркін іздеу кезеңінің мәні

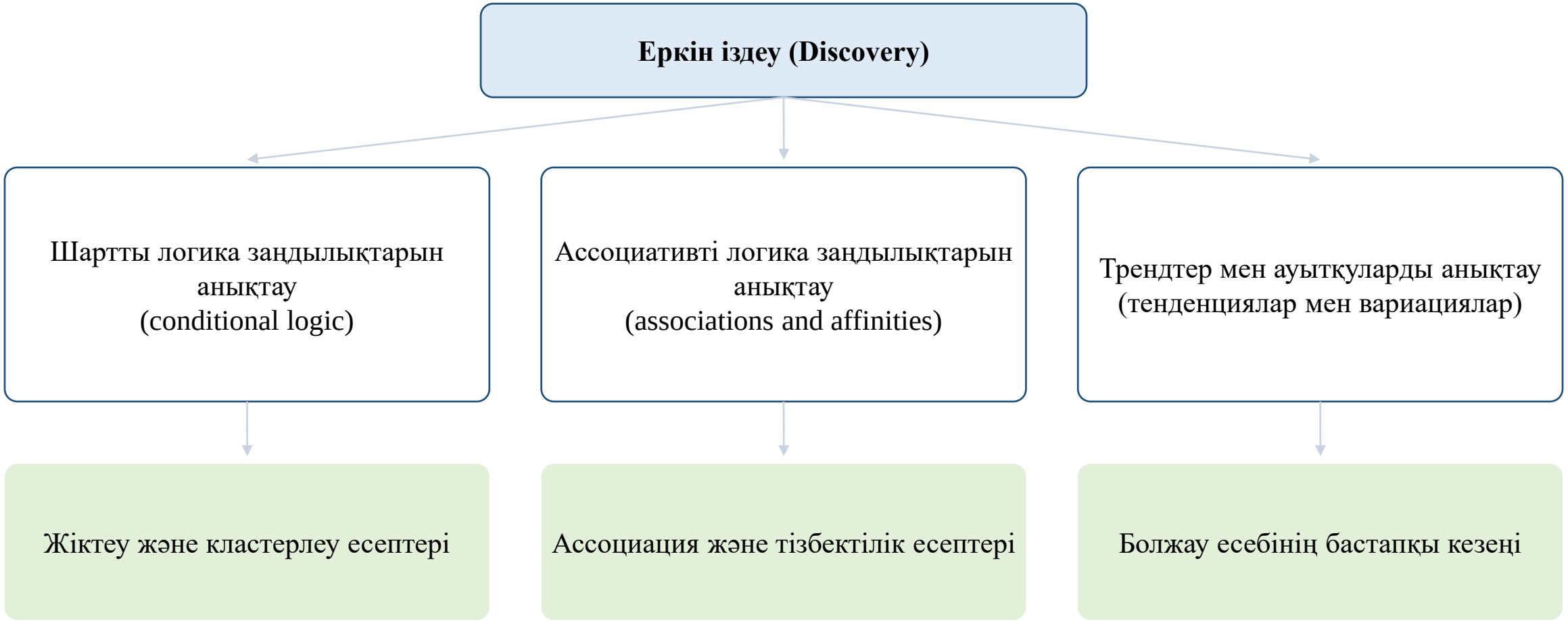
- Деректер жиынтығы жасырын заңдылықтарды іздеу үшін зерттеледі.
- Заңдылықтардың түрі туралы алдын ала гипотеза қойылмайды.

Заңдылық (law) - құбылыстар мен процестердің қалыптасуы мен дамуының кезеңдері мен нысандарын анықтайтын маңызды, үнемі қайталанатын өзара байланыс



Жасырын заңдылықтар

Еркін іздеу кезеңінің әрекеттері



Мысал: кадрлық агенттіктің дерекқоры

Дерекқорда мамандық, еңбек өтілі, жасы және қажетті сыйақы деңгейі туралы мәліметтер бар.

Өз бетінше сұрау (OLAP)

- Аналитик көптеген сұрауды өзі құрастырады.
- Нәтиже: «25–35 жас аралығындағы мамандардың орташа сыйақысы 1200 ш.б.»

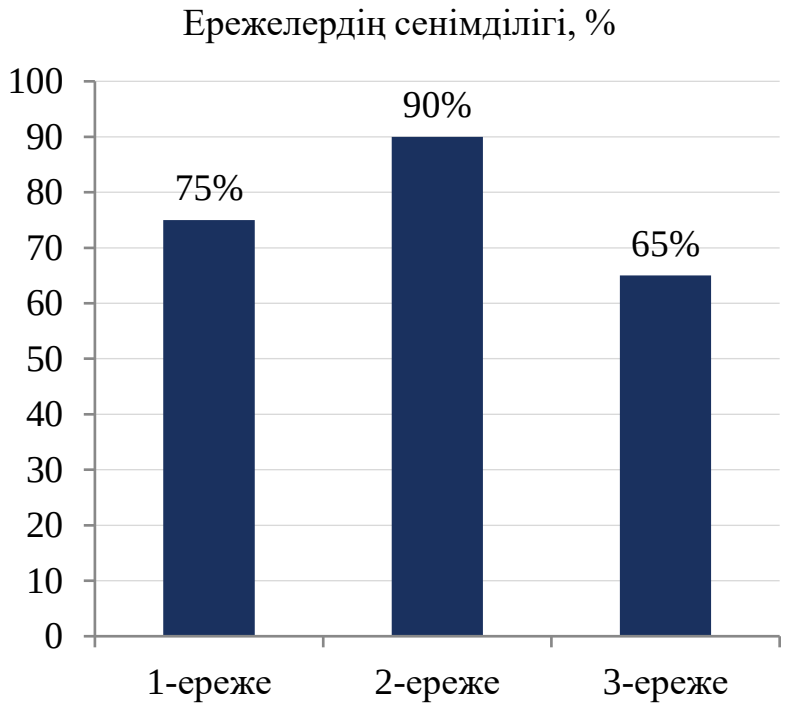
Еркін іздеу (Data Mining)

- Жүйе үлгілерді өзі іздейді, тек мақсатты айнымалы беріледі.
- Нәтиже: «егер ..., онда ...» түріндегі логикалық ережелер жиынтығы.

Еркін іздеу нәтижесінде табылған ережелер

Шарт («егер...»)	Қорытынды («онда...»)	Сенімділік
жасы < 20 және сыйақы > 700 ш.б.	бағдарламашы жұмысын іздейді	75%
жасы > 35 және сыйақы > 1200 ш.б.	жетекші жұмысын іздейді	90%
жетекші жұмысын іздейді және өтілі > 15 жыл	жасы > 35 жас	65%

Мақсатты айнымалы: 1-2 ережеде - мамандық, 3-ережеде - жасы.



3. Болжамдық модельдеу (Predictive Modeling)

1. Data Mining процесінің кезеңдері

2. Еркін іздеу (Discovery)

3. Болжамдық модельдеу (Predictive Modeling)

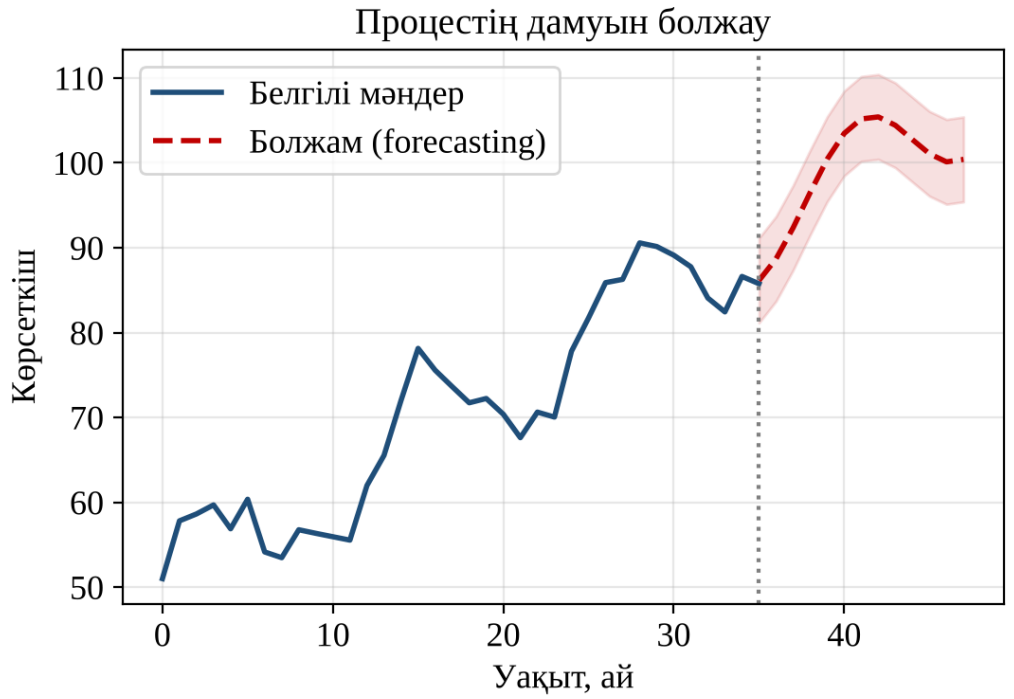
4. Ерекше жағдайларды талдау (Forensic Analysis)

Болжамдық модельдеудің мазмұны

Бірінші кезеңнің нәтижелері қолданылады: анықталған заңдылықтар тікелей болжау үшін пайдаланылады.

Белгісіз мәндерді болжау
(outcome prediction)

Процестердің дамуын болжау
(forecasting)



Процесс дамуын болжау

Жіктеу және болжау есептері

Критерий	Жіктеу	Болжау
1-кезеңнің қолданылатын нәтижесі	Ережелер индукциясы	Тренд немесе тербеліс
Не анықталады	Жаңа объектінің алдын ала белгілі сыныбы	Мақсатты айнымалының белгісіз (болашақ) мәндері
Мысал	Ізденуші – «жетекші» сыныбы (90%)	Келесі кезеңдегі сұраныс деңгейі

Логика тұрғысынан салыстыру: индукция және дедукция

Еркін іздеу - индукция

Жалпы ереже: «өтініш берушілер – бағдарламашылар»

жасы < 20

сыйақы > 700

Жеке → жалпы

Болжамдық модельдеу - дедукция

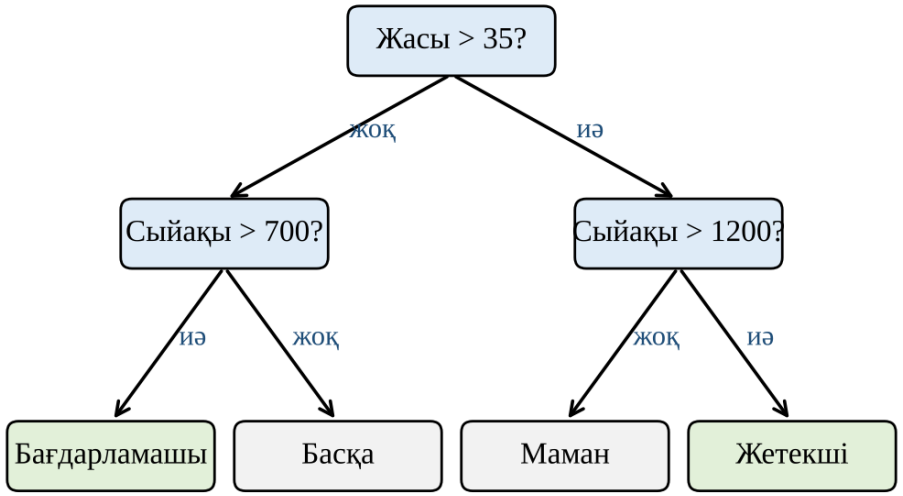
Жалпы ереже: «жетекші жұмыс + өтілі > 15 жыл»

Жеке тұлға: жасы > 35 (65%)

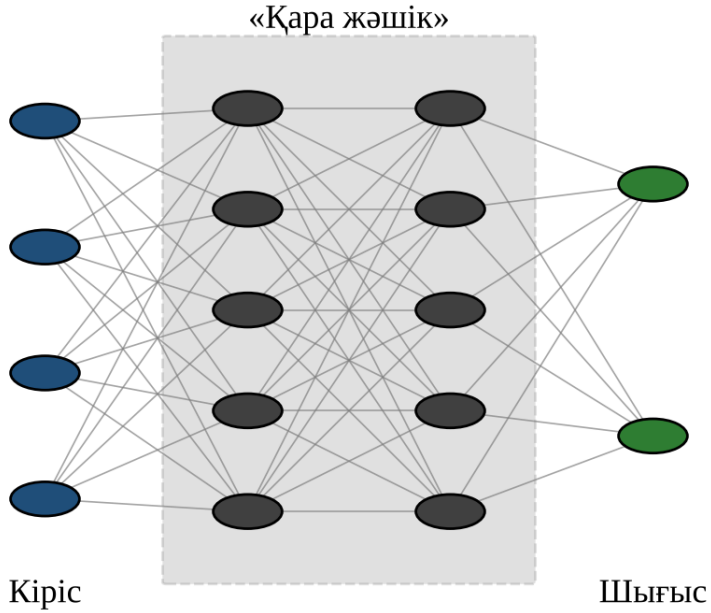
Жалпы → жеке

Индукция сынып туралы жалпы білім береді, дедукция – нақты объект туралы жаңа білім.

Заңдылықтардың көрінісі: мөлдір модель және «қара жәшік»



Шешім ағашы



Нейрондық желі

Мөлдір модельді аналитик түсіндіре алады, ал нейрондық желінің ішкі логикасы жасырын болады.

4. Ерекше жағдайларды талдау (Forensic Analysis)

1. Data Mining процесінің кезеңдері

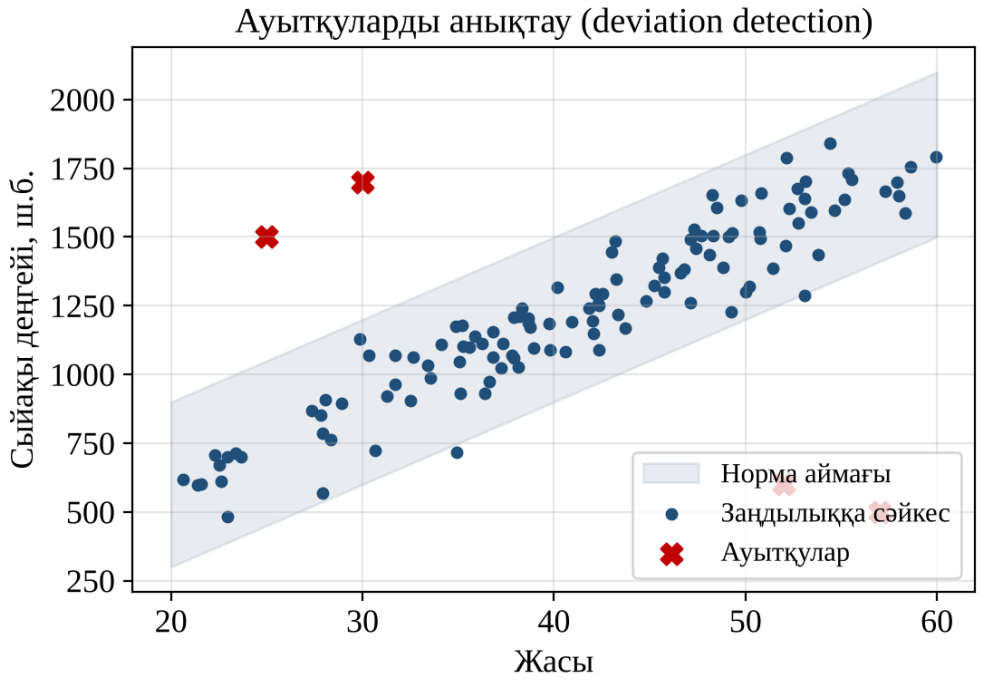
2. Еркін іздеу (Discovery)

3. Болжамдық модельдеу (Predictive Modeling)

4. Ерекше жағдайларды талдау (Forensic Analysis)

Ерекше жағдайларды талдау кезеңі

- Табылған заңдылықтардағы ерекшеліктер мен ауытқулар талданады.
- Негізгі әрекет - ауытқуларды анықтау (deviation detection).
- Ауытқуды табу үшін алдымен норманы анықтау қажет.
- Норма еркін іздеу кезеңінде есептеледі.

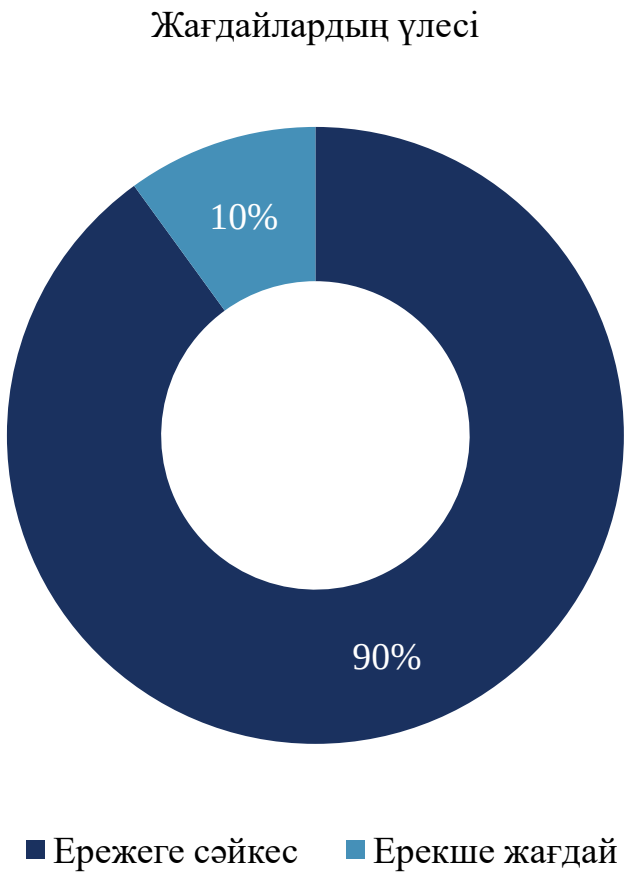


Нормадан ауытқулар

Мысал: ережеге сәйкес келмейтін жағдайлар

«Егер жасы > 35 және сыйақы > 1200 ш.б. болса, онда 90% жағдайда ізденуші жетекші жұмысын іздейді»

Сұрақ: қалған 10% жағдайды неге жатқызуға болады?



Ерекше жағдайларды түсіндірудің екі нұсқасы

10% ерекше жағдай

1-нұсқа: логикалық түсініктеме бар

2-нұсқа: бастапқы деректердегі қателер

Жаңа ереже түрінде тұжырымдалады

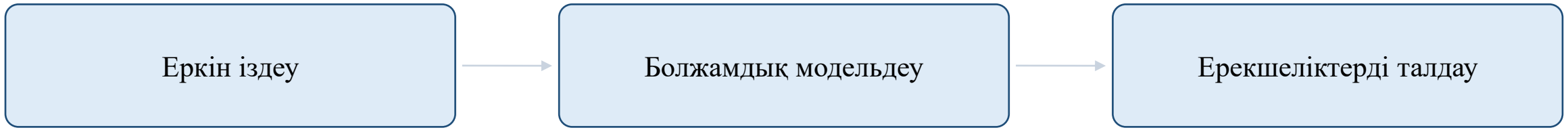
Кезең деректерді тазарту ретінде қолданылады

Data Mining кезеңдерінің жиынтық сипаттамасы

Кезең	Ойлау логикасы	Нәтижесі
Еркін іздеу	Индуктивті (жеке → жалпы)	Ережелер мен заңдылықтар
Болжамдық модельдеу	Дедуктивті (жалпы → жеке)	Белгісіз мәндердің болжамы
Ерекшеліктерді талдау	Нормадан ауытқуды талдау	Ауытқулардың түсіндірмесі, тазартылған деректер

Қорытынды

1. Data Mining процесі үш кезеңнен тұрады: еркін іздеу, болжамдық модельдеу және ерекшеліктерді талдау.
2. Еркін іздеу индуктивті: жеке деректерден жалпы ережелер алынады, валидация осы кезеңге кіреді.
3. Болжамдық модельдеу дедуктивті: жалпы ережелер жаңа объектілерге қолданылады.
4. Ауытқуларды талдау жаңа ережелерді табуға немесе деректерді тазартуға мүмкіндік береді.



Деректерді өңдеудің негізгі кезеңдері

Дәріс №13