Пәні: Деректерді интеллектуалды талдау және визуализациялау

Деректерді интеллектуалды талдауды әртүрлі зерттеу салаларында қолдану

Дәріс №15

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

2026-2027 оқу жылы

Дәрістің мақсаты

Деректерді интеллектуалды талдау әдістерінің түрлі ғылыми және қолданбалы зерттеу салаларындағы рөлін түсіндіру; медицина, білім беру, экономика, экология, инженерия және әлеуметтік ғылымдардағы типтік міндеттерді қарастыру; зерттеу мақсатына сәйкес Data Mining әдістерін таңдаудың логикасын көрсету және нәтижелерді дұрыс интерпретациялау қағидаларын қалыптастыру.

Дәріс жоспары

1. Зерттеулердегі Data Mining орны және негізгі міндеттері
2. Медицина, биология және денсаулық сақтау салаларындағы қолдану
3. Білім беру, экономика, бизнес және әлеуметтік зерттеулердегі қолдану
4. Экология, инженерия және ғылыми зерттеулердегі қолдану
5. Нәтижелерді бағалау, этика және практикалық шектеулер

Data Mining зерттеуде қандай міндет атқарады?

Заңдылықтарды табу

Үлкен деректер жиынынан жасырын байланыстар мен тұрақты үлгілерді анықтау.

Болжау

Белгілі деректер негізінде болашақ мәндерді, тәуекелді немесе нәтижені бағалау.

Топтау

Объектілерді ұқсастық белгілері бойынша кластерлерге бөлу.

Аномалия

Қалыпты үлгіден ауытқитын ерекше жағдайларды анықтау.

Нәтижесінде зерттеуші гипотеза қалыптастырады, модель құрады және дәлелді шешім қабылдауға негіз алады.

Зерттеу сұрағы мен Data Mining әдісінің сәйкестігі

Зерттеу сұрағы	Әдіс	Нәтиже
Қай сыныпқа жатады?	Классификация	Санат / класс
Қанша болады?	Регрессия	Сандық болжам
Қандай табиғи топтар бар?	Кластерлеу	Кластерлер
Не нәрселер бірге кездеседі?	Ассоциация	Ережелер
Қалыптан тыс жағдай бар ма?	Аномалияны анықтау	Күдікті объектілер

Зерттеудегі интеллектуалды талдау процесі



Зерттеу сұрағы — Деректер жинау — Алдын ала өңдеу — Модель құру — Бағалау — Интерпретация

Процесс итеративті: бағалау нәтижесіне қарай зерттеуші деректерге, белгілерге немесе модель параметрлеріне қайта оралуы мүмкін.

Медицинада Data Mining қолдану

Диагностика

Зертханалық көрсеткіштер мен симптомдар негізінде ауру ықтималдығын бағалау.

Тәуекел стратификациясы

Пациенттерді асқыну қаупі бойынша төмен, орта және жоғары топтарға бөлу.

Ем нәтижесін болжау

Терапияға жауап, қайта госпитализация немесе қалпына келу ықтималдығын бағалау.

Медициналық кескіндер

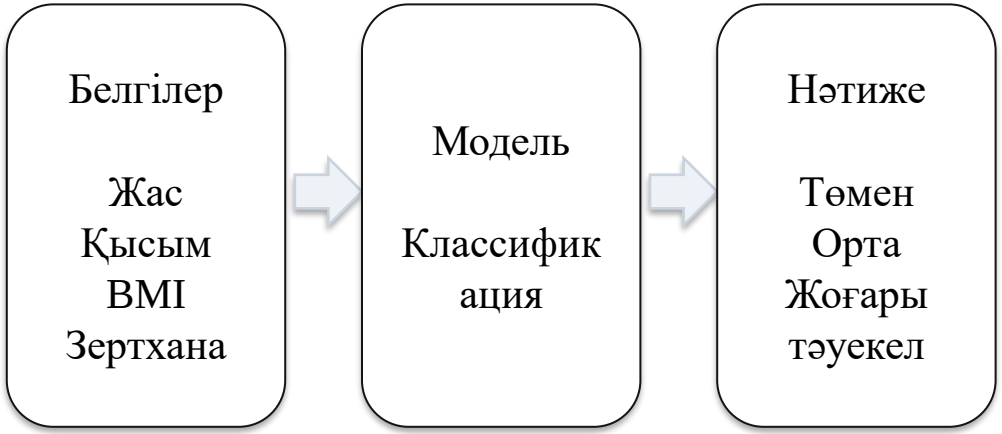
Рентген, КТ, МРТ және гистология кескіндерінен белгілерді анықтау.

Дәрі-дәрмек қауіпсіздігі

Жағымсыз реакциялар мен қауіпті комбинацияларды анықтау.

Мысал: ауру қаупін болжау моделі

- Кіріс белгілері: жас, қан қысымы, BMI, зертханалық көрсеткіштер, өмір салты факторлары.
- Модельдер: логистикалық регрессия, шешім ағашы, Random Forest, градиенттік бустинг.
- Нәтиже: жеке пациент үшін тәуекел ықтималдығы немесе тәуекел класы.
- Бағалау: accuracy жеткіліксіз болуы мүмкін; sensitivity, specificity, ROC-AUC және calibration маңызды.



Биоинформатика және геномика

- Ген экспрессиясы деректерінен ауруға байланысты маңызды гендерді анықтау.
- Пациенттерді молекулалық профильдері бойынша кластерлеу және аурудың кіші типтерін табу.
- Биомаркерлерді іріктеу үшін feature selection және dimensionality reduction әдістерін қолдану.
- Дәрілік жауапты генетикалық ерекшеліктермен байланыстыру.
- Көпөлшемді деректерде интерпретация мен репликацияның маңызы өте жоғары.

Білім беру зерттеулерінде қолдану

Оқу нәтижесін болжау

Бағалар, қатысу, тапсырма белсенділігі арқылы үлгерім тәуекелін ерте анықтау.

Студенттерді топтау

Оқу стилі, белсенділік немесе білім деңгейі бойынша сегменттерді табу.

Оқу траекториясы

Келесі тақырыпты немесе ресурсты ұсыну үшін мінез-құлық деректерін талдау.

Learning Analytics деректері LMS журналдары, тест нәтижелері, форумдар және цифрлық оқу әрекеттерінен алынуы мүмкін.

Экономика және бизнестегі қолдану

Несие скорингі

Клиенттің дефолт тәуекелін бағалау.

Алаяқтықты анықтау

Төлемдердегі күдікті транзакцияларды табу.

Сұранысты болжау

Сатылым мен қор қажеттілігін болжау.

Клиент сегментациясы

Мінез-құлық пен құндылық бойынша топтау.

Churn талдауы

Клиенттің кету ықтималдығын бағалау.

Market basket

Бірге сатып алынатын тауарларды анықтау.

Әлеуметтік ғылымдардағы зерттеулер

- Сауалнама деректерін кластерлеу арқылы респонденттердің типтік профильдерін анықтау.
- Мәтіндік деректерден тақырыптарды, эмоциялық тонды және қоғамдық пікір тенденцияларын табу.
- Әлеуметтік желілердегі байланыстарды графтық талдау арқылы зерттеу.
- Қоғамдық мінез-құлықтың уақыт бойынша өзгерісін анықтау.
- Маңызды шектеу: алгоритмдік нәтиже себеп-салдар байланысын автоматты түрде дәлелдемейді.

Мәтіндерді интеллектуалды талдау

Міндет	Мысал	Әдіс
Тональдылық	Пікір позитивті ме?	Классификация
Тақырыптарды табу	Қандай мәселелер жиі талқыланады?	Topic modeling / кластерлеу
Құжаттарды топтау	Ұқсас мақалаларды біріктіру	Кластерлеу
Кілттік ұғымдар	Маңызды терминдерді табу	NLP / статистикалық әдістер
Спамды анықтау	Қажетсіз хабарды бөлу	Классификация

Экология және қоршаған орта

Ауа сапасы

Ластаушы заттардың концентрациясын болжау және қауіпті кезеңдерді анықтау.

Климаттық деректер

Температура, жауын-шашын және экстремалды құбылыстар трендтерін талдау.

Биоалуантүрлілік

Түрлердің таралуын, популяция өзгерісін және мекен ету ортасын модельдеу.

Қашықтан зондтау

Спутниктік кескіндерден жер жамылғысын классификациялау.

Экологиялық тәуекел

Орман өрті, су тасқыны немесе ластану қаупін бағалау.

Инженерия және өндірістегі қолдану

- Predictive maintenance: сенсор деректерінен жабдықтың істен шығуын алдын ала болжау.
- Сапаны бақылау: өндірістік параметрлер мен ақаулар арасындағы заңдылықтарды анықтау.
- Аномалияларды анықтау: қалыпты жұмыс режимінен ауытқуларды ерте табу.
- Процестерді оңтайландыру: энергия, уақыт және материал шығындарын азайту.
- Computer Vision: өнім кескіндерінен визуалды ақауларды автоматты анықтау.

Мысал: Predictive Maintenance



Негізгі артықшылығы: жоспардан тыс тоқтауды азайту және техникалық қызмет көрсетуді нақты қажеттілікке қарай жоспарлау.

Басқа ғылыми зерттеу салалары

Астрономия

Нысандарды классификациялау, сирек құбылыстарды іздеу.

Химия

Молекула қасиеттерін болжау, қосылыстарды топтау.

Ауыл шаруашылығы

Өнімділік, ауру және суару қажеттілігін болжау.

Киберқауіпсіздік

Шабуылдар мен қалыптан тыс желілік белсенділікті анықтау.

Көлік

Ағынды болжау, маршруттарды оңтайландыру.

Геология

Минералдық ресурстар мен сейсмикалық деректерді талдау.

Әдісті таңдау критерийлері

- Зерттеу мақсаты: сипаттау ма, түсіндіру ме, әлде болжау ма?
- Нысаналы айнымалының болуы: supervised немесе unsupervised тәсіл.
- Деректер түрі: сандық, категориялық, мәтін, кескін, уақыттық қатар.
- Деректер көлемі мен белгілер саны.
- Интерпретация қажеттілігі: қарапайым модель ме, әлде күрделі жоғары дәлдікті модель ме?
- Есептеу ресурстары және модельді практикада енгізу мүмкіндігі.

Модель сапасын бағалау көрсеткіштері

Міндет	Негізгі метрикалар	Ескерту
Классификация	Accuracy, Precision, Recall, F1, ROC-AUC	Класс дисбалансын ескеру керек
Регрессия	MAE, RMSE, R^2	Қателер шкаласына назар аудару
Кластерлеу	Silhouette, Davies–Bouldin	Салалық интерпретация қажет
Уақыттық қатар	MAE, RMSE, MAPE	Уақыттық бөлуді сақтау керек

Зерттеу нәтижесінің сенімділігі

Train / Test бөлу

Модель бұрын көрмеген деректерде тексерілуі тиіс.

Cross-validation

Бірнеше бөлік бойынша қайталанған бағалау нәтижені тұрақтандырады.

External validation

Модельді басқа деректер жиынында тексеру жалпылану қабілетін көрсетеді.

Data leakage және overfitting зерттеу нәтижесін жасанды түрде жақсартып көрсетуі мүмкін.

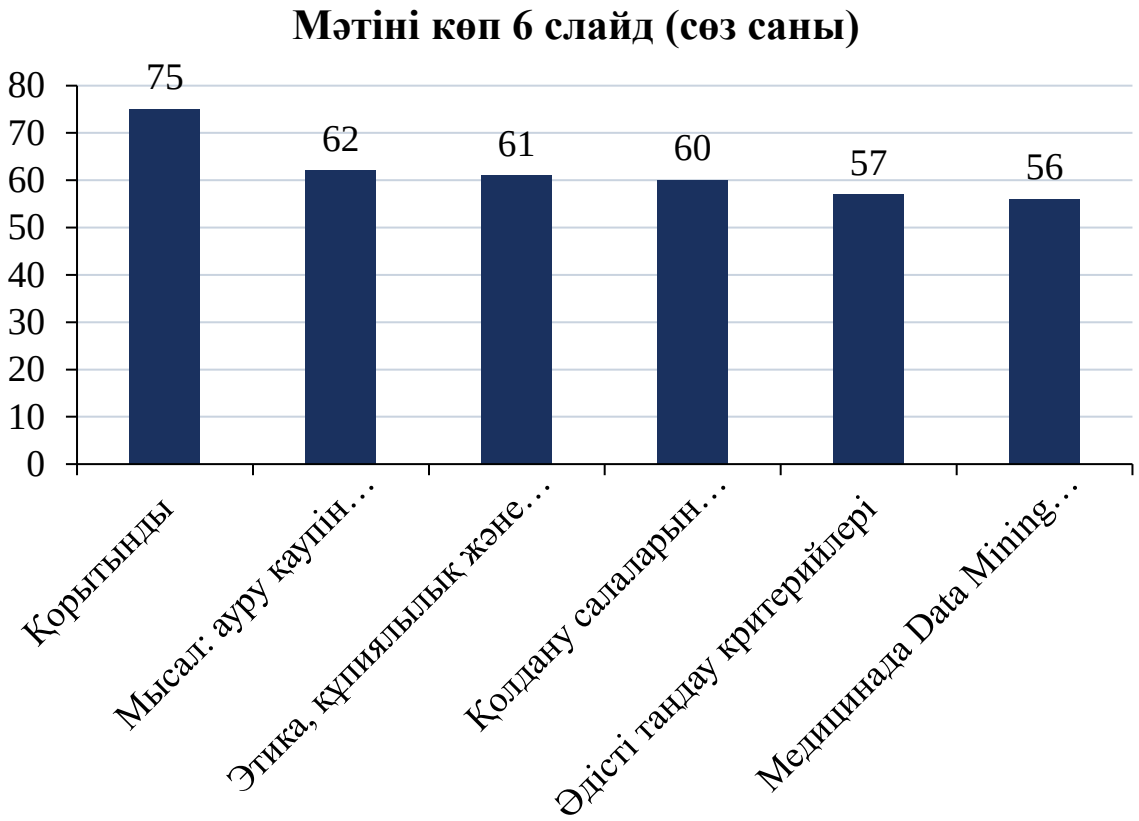
Этика, құпиялылық және алгоритмдік әділеттілік

- Жеке және медициналық деректерді қорғау талаптарын сақтау.
- Деректер жинауда ақпараттандырылған келісім мен мақсаттың айқындығын қамтамасыз ету.
- Bias: деректердегі теңсіздік модель нәтижесінде қайталануы мүмкін.
- Түсіндірілетін модельдер маңызды шешімдерде сенімділікті арттырады.
- Автоматты шешім адамның кәсіби бағасын толық алмастырмауы тиіс.
- Зерттеу нәтижелерін репродукциялау үшін әдіс, параметр және дерек өңдеу қадамдарын құжаттау қажет.

Қолдану салаларын салыстыру

Сала	Негізгі деректер	Типтік міндет
Медицина	Клиникалық, геномдық, кескін	Тәуекел / диагноз / болжам
Білім беру	Баға, LMS, тест	Үлгерім және траектория
Экономика	Транзакция, клиент, нарық	Скоринг / сұраныс / сегмент
Экология	Сенсор, спутник, климат	Мониторинг / болжам
Инженерия	Сенсор, өндіріс	Ақау / сапа / оңтайландыру
Әлеуметтік ғылым	Сауалнама, мәтін, желі	Профиль / пікір / байланыс

Диаграммалар





Қорытынды

1. Data Mining әртүрлі зерттеу салаларында үлкен және күрделі деректерден пайдалы білім алуға мүмкіндік береді.

2. Әдісті таңдау зерттеу сұрағына, деректердің түріне және нәтижені интерпретациялау талабына тәуелді.

3. Медицина, білім беру, экономика, экология және инженерияда классификация, регрессия, кластерлеу, ассоциация және аномалияларды анықтау кең қолданылады.

4. Модельдің жоғары дәлдігі ғана жеткіліксіз: сенімді бағалау, этика, құпиялылық және салалық интерпретация міндетті.

5. Data Mining зерттеушінің шешімін алмастырмайды, бірақ дәлелді шешім қабылдауды күшейтетін аналитикалық құрал болып табылады.

Деректерді интеллектуалды талдауды әртүрлі зерттеу салаларында қолдану

Дәріс №15