



Қорытынды дәріс: ИДТ жобалары

Бұл дәріс студенттердің курста меңгерген білімдерін біріктіруге, ИДТ жобасының толық циклін түсінуге және алынған білімді практикалық есептерде қолдануға дайындауға арналған.

Оқытушы: Сарсембаева Т.С.

Негізгі сұрақтар

1

ИДТ тұжырымдамалары

Курстың негізгі идеялары мен әдістері.

2

Модельдерді таңдау

Модельдерді таңдау және бағалау принциптері.

3

KDD/CRISP-DM циклі

ИДТ жобасының толық циклі.

4

Жоба талаптары

Қорытынды жобаның құрылымы мен талаптары.

5

Ensembling стратегиялары

Әдістерді біріктіру.

6

Болашақ бағыттар

ИДТ саласының даму перспективалары.

Курстың негізгі тұжырымдамалары



Әдістер жиынтығы

Классификация, кластерлеу, ансамбльдік және терең оқыту әдістері — түрлі мәселелерге арналған құралдар.



Деректер сапасы

Модельдің табыстылығының негізгі факторы. Тазалау мен ерекшелік инженериясы дәлдікті арттырады.



Модель таңдау

Мәселенің сипатына байланысты: бинарлы/көпсыныпты классификация, кескіндер, уақыттық қатарлар.



Бағалау метрикалары

Accuracy-ден басқа Precision, Recall, F1 және AUC қолданылады.



Модельді таңдау және бағалау

Модельді таңдау

- Бинарлы/көпсыныпты классификация:
Логистикалық регрессия,
Random Forest.
- Кескіндер: CNN.
- Уақыттық қатарлар: ARIMA,
LSTM.

Модельді таңдау мәселенің сипатына байланысты.

Бағалаудың рөлі

- Дәлдік (Accuracy) жеткіліксіз.
- Precision, Recall, F1-Score, AUC метрикаларын қолдану.
- Кросс-валидация арқылы тексеру.
- Қателіктің құнына байланысты метриканы дұрыс таңдау маңызды.

ИДТ жобасының жалпы циклі (KDD/CRISP-DM)

01

Мәселені түсіну

Нені болжау керек? Бизнеске/медицинаға пайдасы? Қандай метриканы қолдану керек?

02

Деректерді түсіну

Деректерді жинау, бастапқы талдау, сапаны тексеру.

03

Деректерді дайындау

Жетіспейтін мәндерді өңдеу, аномалияларды жою, ерекшеліктерді инженериялау.

04

Модельдеу

Бірнеше модельді (базалық, ансамбльдік, терең оқыту) сынақтан өткізу.

05

Бағалау

Модельдерді метрикалар бойынша салыстыру. Үздік модельді таңдау.

06

Даму/Жасақтау

Модельді өндірістік ортаға енгізу (концепциясын түсіну қажет).

Қорытынды практикалық жобалар

Қорытынды жоба студенттерге курста алған білімдерін нақты, ашық деректер жиынтығына қолдануға мүмкіндік береді. Негізгі міндет — әдістерді біріктіру және терең талдауды көрсету.

Мәселенің анықтамасы

Міндеттің, деректердің және мақсатты метриканың нақты сипаттамасы.

Деректерді талдау (EDA)

Бастапқы талдау, визуализациялар және тазалау стратегиясы.

Модельдік портфолио

Базалық, ансамбльдік және терең оқыту модельдерін енгізу.

Нәтижелерді салыстыру

Кросс-валидация және метрикалар негізінде модельдерді салыстыру.

Түсіндіру

Үздік модельдің болжамдарын түсіндіру (XAI құралдары).

Қорытынды және ұсыныстар

Жоба нәтижелерін жинақтау және нақты өмірдегі әрекеттерге қолдану.

Әдістерді біріктіру стратегиялары (Model Ensembling)

Әртүрлі әдістерді бір шешімге біріктіру.

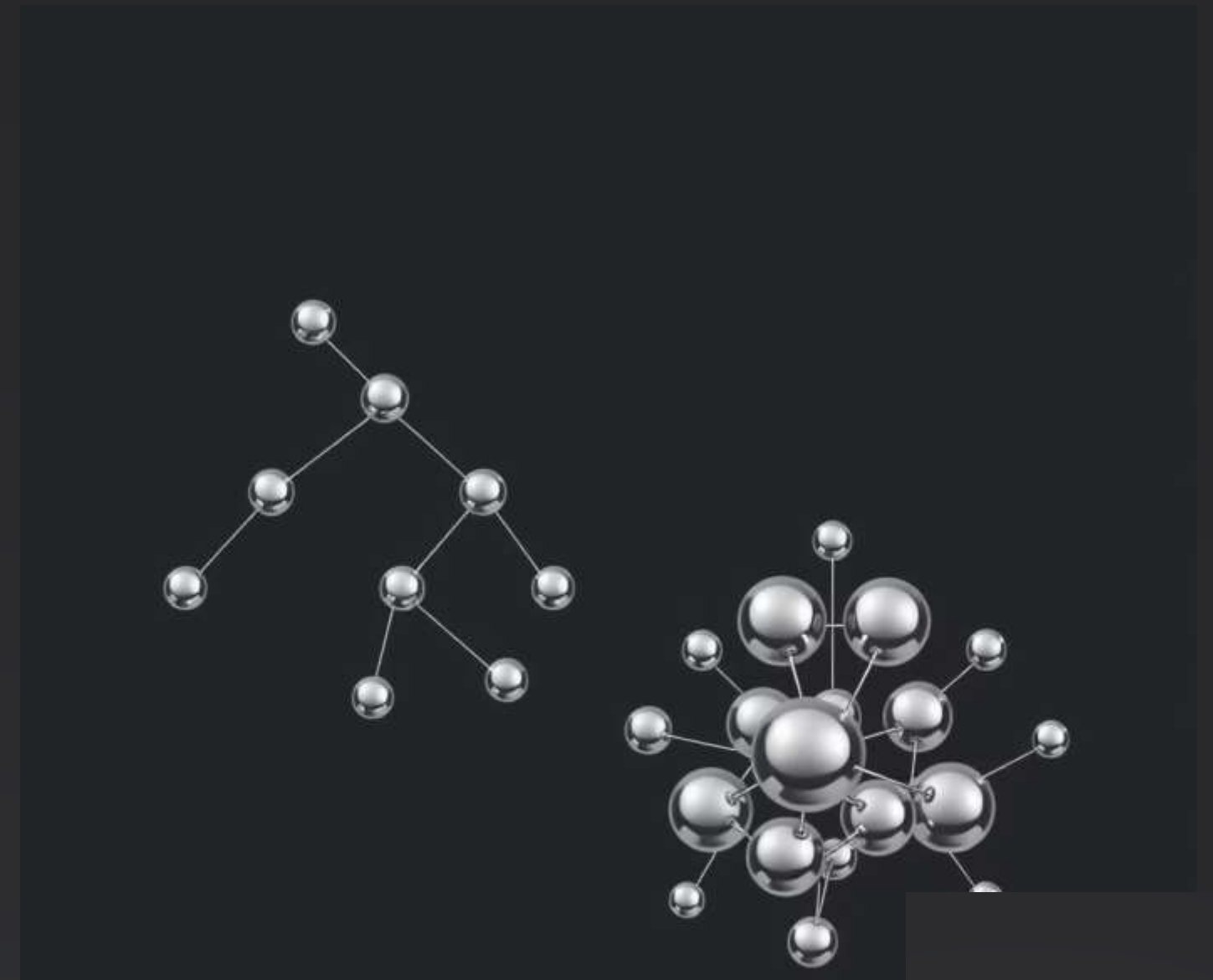
Стекинг (Stacking)

Бірнеше базалық модельдердің болжамдарын пайдаланып, осы болжамдарды кіріс ретінде қабылдайтын екінші деңгейдегі (**Meta**) модельді оқыту. Бұл әртүрлі модельдердің күшті жақтарын біріктіруге мүмкіндік береді.



Сыныпқа арналған арнайы біріктіру

Мысалы, аномалияларды анықтау үшін кластерлеу (DBSCAN) әдісін қолданып, содан кейін қалған деректерді классификациялау үшін **Random Forest** қолдану. Бұл әдістерді бір-біріне жалғау арқылы күрделі мәселелерді шешу.



ИДТ саласының болашақ бағыттары

ИДТ саласы үнемі дамып келеді. Курсты сәтті аяқтаған студенттердің болашақта терең білім алуы үшін келесі ілгері деңгейдегі бағыттар ұсынылады:



Трансформерлер

NLP саласындағы ең соңғы жетістіктер (BERT, GPT), олар RNN/LSTM-ді ығыстырып шығарды.



Күшейтіп оқыту (RL)

Агенттің ортамен өзара әрекеттесуіне негізделген оқыту (автоматтандыру, робототехника).



Үлкен деректер технологиялары

Spark, Hadoop сияқты бөлінген есептеулер жүйелерімен жұмыс істеу.



Қорытынды

ИДТ курсы сізге тек математикалық әдістерді ғана емес, сонымен бірге деректер негізінде сыншыл тұрғыдан ойлау және нақты әлемдегі күрделі мәселелерді жүйелі түрде шешу дағдыларын берді. Бұл дағдылар сізді **Data Scientist** немесе **Data Analyst** ретіндегі кәсіби жолыңыздың негізгі іргетасы болады.

Бақылау сұрақтары және әдебиеттер

Бақылау сұрақтары

1. ИДТ жобасының негізгі кезеңдерін атаңыз.
2. Модель сапасын қандай метрикалармен бағалаймыз?
3. Ensemble әдістерінің мәні неде?
4. Қорытынды жобада қандай элементтер қамтылуы тиіс?
5. ИДТ саласының келешек бағыттарын атаңыз.

Ұсынылатын әдебиеттер

- Han, J., Kamber, M., & Pei, J. Data Mining: Concepts and Techniques. Elsevier, 2011.
- Kuhn, M., & Johnson, K. Feature Engineering and Selection. CRC Press, 2019.
- Hastie, T., Tibshirani, R., & Friedman, J. The Elements of Statistical Learning. Springer, 2009.
- Provost, F., & Fawcett, T. Data Science for Business. O'Reilly Media, 2013.

