



Інтелектуальний аналіз даних великих обсягів

Робоча програма навчальної дисципліни (Силабус)

Реквізити навчальної дисципліни

Рівень вищої освіти	<i>Другий (магістерський)</i>
Галузь знань	<i>11 Математика та статистика</i>
Спеціальність	<i>113 Прикладна математика</i>
Освітня програма	<i>Наука про дані та математичне моделювання</i>
Статус дисципліни	<i>Нормативна</i>
Форма навчання	<i>очна(денна)</i>
Рік підготовки, семестр	<i>2 курс, осінній</i>
Обсяг дисципліни	<i>4 кредитів</i>
Семестровий контроль/ контрольні заходи	<i>екзамен/колоквіум, лабораторні роботи</i>
Розклад занять	<i>лекція – 2 год. на тиждень, лабораторні роботи – 1 год. на тиждень</i>
Мова викладання	<i>Українська</i>
Інформація про керівника курсу / викладачів	Лектори: <i>Мальчиков Володимир Вікторович, mavr2k@gmail.com</i> Лабораторні: <i>Мальчиков Володимир Вікторович, mavr2k@gmail.com</i>
Розміщення курсу	Посилання на дистанційний ресурс https://app.slack.com/client/T019Y2DJMU4/C019B4ZRN87

Програма навчальної дисципліни

1. Опис навчальної дисципліни, її мета, предмет вивчення та результати навчання

Дисципліна «Інтелектуальний аналіз даних великих обсягів» є базовим курсом при підготовці фахівців з прикладної математики та науки про дані.

Метою курсу є викладення основних понять і методів, потрібних для обробки великих обсягів даних в різних проблемних областях.

Предметом вивчення дисципліни є методи кластеризації даних великих обсягів та методи візуалізації даних великих обсягів.

Згідно з вимогами освітньо-професійної програми, студенти після засвоєння дисципліни «Інтелектуальний аналіз даних великих обсягів» мають продемонструвати такі результати навчання:

компетентності

- демонструвати знання й розуміння основних концепцій, принципів, теорій прикладної математики і використовувати їх на практиці;

знання:

- підходів до роботи з даними великих обсягів;
- методів кластеризації даних великих обсягів;
- методів візуалізації даних великих обсягів;

уміння:

- виділення груп у даних великих обсягів;
- графічного представлення багатовимірних даних;

навички:

- візуалізації багатовимірних даних;

- визначення центроїдів даних великих обсягів;
досвід:
- виділення кластерів в даних великих обсягів;
- візуального представлення багатовимірних даних.

2. Пререквізити та постреквізити дисципліни (місце в структурно-логічній схемі навчання за відповідною освітньою програмою)

Дисципліна «Інтелектуальний аналіз даних великих обсягів» вивчається на 2 курсі в осінньому семестрі.

Вивченню дисципліни «Інтелектуальний аналіз даних великих обсягів» має передувати вивчення дисциплін «Інтелектуальний аналіз даних» та «Машинне навчання».

3. Зміст навчальної дисципліни

РОЗДІЛ 1. КЛАСТЕРИЗАЦІЯ ДАНИХ ВЕЛИКИХ ОБСЯГІВ

- Тема 1.1. Алгоритми BIRCH, Batch K-means
- Тема 1.2. Алгоритми DBScan, Cure
- Тема 1.3. Алгоритми WaveCluster, CLARA, Clarans
- Тема 1.4. Алгоритми Боруєвки та Форел
- Тема 1.5. Карти Кохонена

РОЗДІЛ 2. ВІЗУАЛІЗАЦІЯ ДАНИХ ВЕЛИКИХ ОБСЯГІВ

- Тема 2.1. Основні принципи візуалізації за Тафтом
- Тема 2.2. Обличчя Чернова
- Тема 2.3. Паралельні координати
- Тема 2.4. Пелюсткові діаграми

4. Навчальні матеріали та ресурси

Базова література

1. Ситник В. Ф., Краснюк М. Т. Інтелектуальний аналіз даних (дейтамайнінг): Навч. посібник. — К.: КНЕУ, 2007. — 376 с.
2. Гороховатський В.О. Методи інтелектуального аналізу та оброблення даних. — Харків, ХНУРЕ, 2021. — 90 с.

Допоміжна література

3. Han, Jiawei, Micheline Kamber, and Jian Pei. *Data mining: concepts and techniques*. Morgan kaufmann, 2006.
4. Guo, Yike; and Grossman, Robert (editors) (1999); *High Performance Data Mining: Scaling Algorithms, Applications and Systems*, Kluwer Academic Publishers
5. Ian H. Witten, Eibe Frank and Mark A. Hall. *Data Mining: Practical Machine Learning Tools and Techniques*. — 3rd Edition. — Morgan Kaufmann, 2011. — P. 664. — ISBN 9780123748560.
6. Soukup T. and Davidson I. *Visual Data Mining: Techniques and Tools for Data Visualization and Mining*. John Wiley & Sons, 2002, 382 p

Матеріали курсу представлені в <https://app.slack.com/client/T019Y2DJMU4/C019B4ZRN87>, <http://login.kpi.ua>.

Навчальний контент

5. Методика опанування навчальної дисципліни (освітнього компонента)

Лекційні заняття

- 1 **Алгоритм BIRCH**
- 2 **Алгоритм Batch K-means**
- 3 **Алгоритм DBScan**
- 4 **Алгоритм Cure**
- 5 **Алгоритм WaveCluster**
- 6 **Алгоритми CLARA, Clarans**
- 7 **Алгоритми Борувки та Форел**
- 8 **Карти Кохонена**
- 9 **Основні принципи візуалізації за Тафтом**
- 10 **Обличчя Чернова**
- 11 **Паралельні координати та пелюсткові діаграми**

Лабораторні роботи

Основні завдання циклу лабораторних робіт: закріплення студентами на практиці знань, отриманих ними під час лекцій та самостійної роботи, та отримання практичних навичок застосування вивченого матеріалу.

Треба відпрацювати такі теми:

1. **Алгоритми BIRCH, Batch K-means**
2. **Алгоритми DBSCAN, CURE**
3. **Алгоритми Wave;Cluster, CLARA**
4. **Алгоритми Борувки та Форел**
5. **Візуалізація даних за допомогою облич Чернова**
6. **Візуалізація даних за допомогою паралельних координат та пелюсткових діаграм**

6. Самостійна робота студента

До самостійної роботи студента виносяться:

- підготовка до аудиторних занять – систематично до 2 годин на заняття з урахуванням повторення лекційного матеріалу;
- підготовка до колоквіуму – до 4 годин самостійної роботи;
- підготовка до екзамену – до 10 годин самостійної роботи;
- самостійно ознайомитися з темами: інші методи кластеризації даних великих обсягів, асоціативні правила інтелектуального аналізу даних.

7. Політика навчальної дисципліни (освітнього компонента)

Вимоги, які ставляться перед студентом під час опанування навчальної дисципліни:

- систематичне відвідування занять;
- вчасне виконання лабораторних робіт;
- заохочувальні бали нараховуються за відповіді на запитання лектора до загальної аудиторії, за знаходження помилок та описок у лекціях, за задавання питань, які свідчать про вдумливу роботу студента із навчальним матеріалом;
- у випадку недостатньої кількості балів, що набрані за семестр, для допуску до екзамену, дається декілька завдань, для досягнення допуску;
- за списування або розмови під час контрольних заходів знімаються штрафні бали, за списування на екзамені студент усувається із аудиторії.

8. Види контролю та рейтингова система оцінювання результатів навчання (PCO)

Поточний контроль: експрес-опитування, колоквиум,.

Календарний контроль: провадиться двічі на семестр як моніторинг поточного стану виконання вимог силабусу; студент отримує «зараховано», якщо його поточний рейтинг не менше 50 % від максимальної кількості балів, яку може отримати студент до даного календарного контролю

Семестровий контроль: екзамен

Умови допуску до семестрового контролю: семестровий рейтинг не менше ніж 25 балів.

Таблиця відповідності рейтингових балів оцінкам за університетською шкалою:

Кількість балів	Оцінка
100-95	Відмінно
94-85	Дуже добре
84-75	Добре
74-65	Задовільно
64-60	Достатньо
Менше 60	Незадовільно
Не виконані умови допуску	Не допущено

9. Додаткова інформація з дисципліни (освітнього компонента)

- додаток до силабусу – перелік питань та типи практичних задач, які виносяться на семестровий контроль;
- екзамен проходить у письмовій формі по індивідуальних білетах; завдання в білетах різні, не повторюються, розраховані на час проведення екзамену й загалом однакові по складності;
- кожний білет містить три теоретичних питання і одне питання прикладного характеру;
- кожне теоретичне питання та практичне заняття оцінюється у 10 балів;
- для встановлення ступеня знання студентом матеріалу після перевірки письмової відповіді додатково задаються додаткові запитання, які оцінюються у сукупності у 10 балів;
- за умови дистанційного семестрового контролю екзаменаційний білет буде складатися з двох теоретичних запитання та одного завдання прикладного характеру, що оцінюються по 10 балів кожне (у режимі письмової контрольної); додаткові запитання оцінюються у 20 балів та задаються кожному студенту особисто в режимі відеоконференції.

Робочу програму навчальної дисципліни (силабус):

Складено старший викладач Мальчиков Володимир Вікторович

Ухвалено кафедрою ПМА (протокол № 13 від 16 червня 2022 р.)

Погоджено Методичною комісією факультету прикладної математики (протокол № 9 від 24 червня 2022 р.)