



Силабус освітнього компонента Програма навчальної дисципліни



Інтелектуальний аналіз даних та видобування знань

Шифр та назва спеціальності

126 – Інформаційні системи та технології

Інститут

ННІ Комп'ютерних наук та інформаційних технологій

Освітня програма

Програмне забезпечення інформаційних систем

Кафедра

Інформаційні системи та технології(329)

Рівень освіти

Магістр

Тип дисципліни

Профільна, вибіркова

Семестр

3

Мова викладання

Українська

Викладачі, розробники



Москаленко Валентина Володимирівна

Valentyna.Moskalenko@khpi.edu.ua

Д.т.н., професор, професор кафедри ICT

Кількість наукових та навчальних публікацій – більше 130.
(<https://publons.com/researcher/1588564/valentyna-moskalenko/>;
Web of Science ResearcherID R-9960-2018;
[https://scholar.google.com.ua/citations?user=eUidJHIAAAA&hl](https://scholar.google.com.ua/citations?user=eUidJHIAAAA&hl;);
<https://www.scopus.com/authid/detail.uri?authorId=36021571200>;
<https://orcid.org/0000-0002-9994-5404>).

Провідний лектор з дисциплін: «Теорія ймовірності та математична статистика», «Основи інформаційних систем та технологій», «Методи бізнес-аналізу для управління вимогами», «Наукові напрямки дослідження інформаційних систем та технологій», «Аналіз та управління вимогами до програмного забезпечення інформаційних систем», «Вступ до Big Data», «Машинне навчання Big Data».

Наукові напрямки: розробка інформаційних систем для стратегічного управління компанією; застосування методів та моделей обчислюваного інтелекту для розв'язання задач управління складними організаційними системами; бізнес -аналітика..

[Детальніше про викладача на сайті кафедри](#)

Загальна інформація

Анотація

Завданнями дисципліни є формування у студентів комплексу знань щодо моделей та методів Data Mining; процесів виявлення знань, способів візуального представлення даних, проектування та використання сховищ даних для аналізу даних, а також знань методів інтелектуального аналізу тексту. Предметом вивчення дисципліни є створення та дослідження математичних та програмних моделей інтелектуального аналізу даних, пов'язаних із функціонуванням об'єктів професійної діяльності.

Мета та цілі дисципліни

Метою вивчення дисципліни є формування теоретичних і практичних знань студентів щодо інтелектуального аналізу даних, а також отримання навичок обґрунтовування вибору конкретного типу моделі інтелектуального аналізу даних та методу видобування знань при розв'язанні поставленої практичної задачі у процесі проектування та впровадження програмного забезпечення інформаційних систем.

Формат занять

Лекції, лабораторні роботи, самостійна робота, консультації. Підсумковий контроль – залік.

Компетентності

ЗК01. Здатність до абстрактного мислення, аналізу та синтезу.

СК04. Здатність розробляти математичні, інформаційні та комп'ютерні моделі об'єктів і процесів інформатизації.

СК05. Здатність використовувати сучасні технології аналізу даних для оптимізації процесів в інформаційних системах.

Результати навчання

РН01. Відшуковувати необхідну інформацію в науковій і технічній літературі, базах даних, інших джерелах, аналізувати та оцінювати цю інформацію

РН08. Розробляти моделі інформаційних процесів та систем різного класу, використовувати методи моделювання, формалізації, алгоритмізації та реалізації моделей з використанням сучасних комп'ютерних засобів.

РН09. Розробляти і використовувати сховища даних, здійснювати аналіз даних для підтримки прийняття рішень

Додатково для освітньо-наукових програм:

РН13. Планувати та виконувати наукові дослідження у сфері ICT, формулювати і перевіряти гіпотези, обирати методи, обґрунтовувати висновки, презентувати результати.

Обсяг дисципліни

Загальний обсяг дисципліни 120 год. (4 кредити ECTS): лекції – 30 год., лабораторні роботи – 30 год., самостійна робота – 60 год.

Передумови вивчення дисципліни (пререквізити)

Для ефективного засвоєння дисципліни необхідні знання дискретної математики, теорії ймовірності та математичної статистики, баз даних, програмування, планування, аудит та супровід інформаційних систем.

Особливості дисципліни, методи та технології навчання

Методи викладання та навчання:

інтерактивні лекції з презентаціями, дискусії, лабораторні заняття, командна робота, кейс-метод, метод зворотного зв'язку з боку студентів, проблемне навчання.

Форми оцінювання:

оцінювання знань на лабораторних заняттях (CAS), експрес-опитування (CAS)

Програма навчальної дисципліни

Теми лекційних занять

Тема 1. Основи інтелектуального аналізу даних.

Мета та завдання дисципліни. Поняття інтелектуального аналізу даних. Інструментальні засоби Data Mining. Методи дослідження структури даних: візуалізація даних.

Варіанти опису процесу Data mining: CRISPDM (Cross Industry Standard of Processes of Data Mining), SEMMA (Sample, Explore, Modify, Model and Assess), DMAIC (Define, Measure, Analyze, Improve and Control).

Тема 2. Методи первинної обробки даних. Регресійний аналіз.

Методи первинної обробки даних. Методи дослідження структури даних: візуалізація та автоматичне групування даних. Способи візуального представлення даних. Візуалізація інструментів Data Mining. Візуалізація Data Mining моделей. Кореляційний аналіз. Регресійний аналіз.

Тема 3. Методи багатомірного розвідувального аналізу

Методи кластерного аналізу. Ієрархічний кластерний аналіз. Ітеративні методи. Алгоритм k-середніх (k-means). Дискримінантний аналіз

Тема 4. Методи пошуку шаблонів даних

Асоціативні правила. Послідовне відображення шаблонів даних. Методи пошуку асоціативних правил. Додатки із застосуванням асоціативних правил. Характеристики асоціативних правил.

Тема 5. Інтелектуальний аналіз тексту (Text Mining)

Видобуток тексту. Представлення текстових документів. Побудова моделі векторного простору. Аналіз тексту та категоризація. Технологія обробки природної мови (NLP)

Тема 6. Сховища даних та оперативний аналіз даних (OLAP)

Визначення сховища даних, порівняння з базами даних, використання. Архітектура сховища даних. ETL-процеси (добування, перетворення й завантаження даних). Вітрини даних, куби даних, багатомірна модель даних. Розгортання OLAP-кубів. Операції над OLAP-кубами (зріз, обертання, консолідація, деталізація). Архітектура OLAP-систем: MOLAP, ROLAP, HOLAP

Теми практичних занять

Практичні заняття не передбачені навчальним планом.

Теми лабораторних робіт

Тема 1. Попередня підготовка та обробка даних. Регресійний аналіз

Тема 2. Методи розв'язання задачі кластеризації

Тема 3. Методи розв'язання задачі класифікації

Тема 4. Підготовка та інтелектуальний аналіз даних текстового формату

Самостійна робота

Тема 2. Методи первинної обробки даних. Регресійний аналіз.

Описова статистика. Кореляційний аналіз. Регресійний аналіз

Тема 3. Методи багатомірного розвідувального аналізу

Заходи подібності. Методи об'єднання або зв'язку. Лінійний дискримінантний аналіз

Тема 5. Інтелектуальний аналіз часових рядів

Завдання прогнозування. Прогнозування і часові ряди.

Тема 8. Інтелектуальний аналіз тексту (Text Mining)

Алгоритми веб -майнінгу. Отримання та аналіз веб -документів

Індивідуальних завдань не передбачено навчальним планом.

Студентам рекомендовані додаткові матеріали (відео, статті) для самостійного вивчення та опрацювання

Література та навчальні матеріали

Основна література

1. Jiawei Han, Jian Pei, Hanghang Tong. Data Mining: Concepts and Techniques. 4th Ed. Publisher: The Morgan Kaufmann Series in Data Management Systems. 2022, 752 p.
2. Pang-Ning Tan, Michael Steinbach. Introduction to Data Mining. 2nd Ed. Publisher: Pearson, 2018, 864p.
3. Mohammed J. Zaki, Wagner Meira, Jr. Data Mining and Machine Learning: Fundamental Concepts and Algorithms. Sec. Ed. Cambridge University Press, 2020.
- 4 Daniel T. Larose Data Mining and Predictive Analytics. Wiley Series on Methods and Applications in Data Mining, 2nd Ed., 2015, 824 p.

5. Charu C. Aggarwal Data Mining: The Textbook. Springer, 2015, 763 p

Додаткова література

1. Christopher Pal, Mark Hall, Eibe Frank, Ian Witten. Data Mining: Practical Machine Learning Tools and Techniques, 4rd ed. Morgan Kaufmann, 2016.
2. Jennifer Reese, Richard Reese. Java for Data Science. Packt Publishing, 2017.
3. Bostjan Kaluza. Machine Learning in Java. Packt Publishing, 2016.
4. Іванов С.М., Максишко Н.К., Бречко Д.О. Інтелектуальний аналіз даних: конспект лекцій. Запоріжжя: ЗНУ, 2020, 156 с.
5. Jason Bell. Machine Learning: Hands-On for Developers and Technical Professionals. John Wiley & Sons, 2014.

Web-ресурси

1. Data Mining Concepts // <https://learn.microsoft.com/en-us/analysis-services/data-mining/data-mining-concepts?view=asallproducts-allversions>
2. Data Mining Algorithms (Analysis Services - Data Mining) // <https://learn.microsoft.com/en-us/analysis-services/data-mining/data-mining-algorithms-analysis-services-data-mining?view=asallproducts-allversions>
3. Data Mining (SSAS) // <https://learn.microsoft.com/en-us/analysis-services/data-mining/data-mining-ssas?view=asallproducts-allversions>
4. Data Mining Extensions (DMX) Reference // <https://learn.microsoft.com/en-us/sql/dmx/data-mining-extensions-dmx-reference?view=sql-server-ver16>
5. DATA MINING MATLAB PROJECTS // <https://matlabsimulation.com/data-mining-matlab-projects/>
6. Data Mining Tutorial // <https://www.javatpoint.com/data-mining>
7. Data mining // https://www.sas.com/nl_nl/insights/analytics/data-mining.html
8. The 13 Best Deep Learning Courses and Online Training for 2023 // <https://solutionsreview.com/business-intelligence/the-best-deep-learning-courses-and-online-training/>
9. What Is Text Mining? A Beginner's Guide // <https://monkeylearn.com/text-mining/>

Система оцінювання

Критерії оцінювання успішності студента та розподіл балів

100% підсумкове оцінювання у вигляді заліку
100% поточне оцінювання:
4 лабораторні роботи (по 25%)

Шкала оцінювання

Сума балів	Національна оцінка	ECTS
90–100	Відмінно	A
82–89	Добре	B
75–81	Добре	C
64–74	Задовільно	D
60–63	Задовільно	E
35–59	Незадовільно (потрібне додаткове вивчення)	FX
1–34	Незадовільно (потрібне повторне вивчення)	F

Норми академічної етики і політика курсу

Студент повинен дотримуватися «Кодексу етики академічних взаємовідносин та доброчесності НТУ «ХПІ»: виявляти дисциплінованість, вихованість, доброзичливість, чесність, відповідальність. Конфліктні ситуації повинні відкрито обговорюватися в навчальних групах з викладачем, а при неможливості вирішення конфлікту – доводитися до відома співробітників дирекції інституту.

Нормативно-правове забезпечення впровадження принципів академічної доброчесності НТУ «ХПІ» розміщено на сайті: <http://blogs.kpi.kharkov.ua/v2/nv/akademichna-dobrochesnist/>

Погодження

Силабус погоджено

30.08.2023



Завідувач кафедри
Олена НІКУЛІНА

30.08.2023



Гарант ОП
Олена НІКУЛІНА