



Вступ до інтелектуального аналізу даних

Робоча програма навчальної дисципліни (Силабус)

Реквізити навчальної дисципліни

Рівень вищої освіти	<i>Перший (бакалаврський)</i>
Галузь знань	<i>12 Інформаційні технології</i>
Спеціальність	<i>122 Комп'ютерні науки</i>
Освітня програма	<i>Цифрові технології в енергетиці</i>
Статус дисципліни	<i>Нормативна</i>
Форма навчання	<i>очна(денна)</i>
Рік підготовки, семестр	<i>3 курс, весняний семестр</i>
Обсяг дисципліни	<i>120 годин (4 кредита ЄКТС) 36 лек. 18 лаб. 66 СРС</i>
Семестровий контроль/ контрольні заходи	<i>Екзамен, мкр</i>
Розклад занять	<i>http://rozklad.kpi.ua/</i>
Мова викладання	<i>Українська</i>
Інформація про керівника курсу / викладачів	<i>Лектор: д.т.н., Путренко Віктор Валентинович, putrenko@ukr.net, telegram, viber, Zoom session Лабораторні: д.т.н., Путренко Віктор Валентинович, putrenko@ukr.net, telegram, viber, Zoom session</i>
Розміщення курсу	<i>https://classroom.google.com/c/NDU0OTk1MzQ1OTY2?cjc=sn7abx2</i>

Програма навчальної дисципліни

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

Програму навчальної дисципліни «Вступ до інтелектуального аналізу даних» складено відповідно до освітньо-професійної програми підготовки бакалавра з комп'ютерних наук першого рівня вищої освіти спеціальності 122 «Комп'ютерні науки». Навчальна дисципліна належить до циклу професійної підготовки. Статус навчальної дисципліни - нормативна. Обсяг навчальної дисципліни становить 4 кредити ЄКТС.

Мета вивчення дисципліни – формування у студентів загальних та фахових компетентностей.

Предмет вивчення.

Методи, підходи та технології щодо видобування даних, їхньої попередньої обробки та подальшого аналізу різноманітними методами інтелектуального аналізу даних.

Основні завдання кредитного модуля.

Згідно з вимогами програми навчальної дисципліни аспіранти після засвоєння кредитного модуля мають продемонструвати такі результати навчання:

знання:

- основних понять, задач, стадій інтелектуального аналізу даних;
- підходів до збереження, представлення даних у сучасних інформаційних системах;
- аналізу залежностей у даних;
- методи та інструменти інтелектуального аналізу даних;
- сучасні програмні засоби, застосовувані для інтелектуального аналізу даних;

- концепції сховищ даних, підходи до їх аналітичної обробки.

уміння:

- обґрунтовувати вибір конкретного типу моделі та методу інтелектуального аналізу даних при вирішенні поставленої практичної задачі;
- проводити необхідну попередню обробку даних, визначати тип задачі аналізу, вирішувати її адекватно обраним методом з оптимально визначеними параметрами, оцінювати результати, робити змістовні висновки та інтерпретацію;
- використовувати сучасні програмні засоби для проектування та дослідження систем інтелектуального аналізу даних;
- застосовувати технології роботи зі сховищами даних, здійснювати їх аналітичну обробку та інтелектуальний аналіз для забезпечення надійної роботи інформаційних систем;

Очікувані результати навчання

ЗК 7. Здатність до пошуку, оброблення та аналізу інформації з різних джерел

ЗК 8. Здатність генерувати нові ідеї (креативність)

ЗК 11. Здатність приймати обґрунтовані рішення

ФК 2. Здатність до виявлення статистичних закономірностей недетермінованих явищ, застосування методів обчислювального інтелекту, зокрема статистичної, нейромережевої та нечіткої обробки даних, методів машинного навчання та генетичного програмування тощо

ФК 4. Здатність використовувати сучасні методи математичного моделювання об'єктів, процесів і явищ, розробляти моделі й алгоритми чисельного розв'язування задач математичного моделювання, враховувати похибки наближеного чисельного розв'язування професійних задач

ФК 6. Здатність до системного мислення, застосування методології системного аналізу для дослідження складних проблем різної природи, методів формалізації та розв'язування системних задач, що мають суперечливі цілі, невизначеності та ризику

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

Міждисциплінарні зв'язки: вивчення даної дисципліни забезпечується дисциплінами обов'язкового компоненту циклів загальної та професійної підготовки: «Теорія ймовірностей, ймовірносні процеси та математична статистика», «Проектування та аналіз обчислювальних алгоритмів» та «Математичний аналіз». Вивчення дисципліни «Вступ до інтелектуального аналізу даних» забезпечує засвоєння дисциплін нормативного блоку таких, як «Методи та системи штучного інтелекту» та «Моделювання систем в енергетиці», «Чисельні методи в моделюванні енергетичних процесів», тощо.

3. Зміст навчальної дисципліни

<i>Перелік розділів і тем дисципліни</i>
<i>Тема 1. Основи інтелектуального аналізу даних. Визначення Data Mining і область застосування. Задачі, моделі та методи Data Mining. Методи, стадії, задачі Data Mining. Поняття Business Intelligence.</i>
<i>Тема 2. Інтелектуальний аналіз даних у енергетиці Приклади застосування у різних сферах. Застосування інтелектуального аналізу для вирішення складних задач.</i>
<i>Тема 3. Етапи інтелектуального аналізу даних. Загальні закономірності при аналізі даних. Групи задач аналізу даних. Класифікація методів аналізу.</i>
<i>Тема 4. Методи аналізу часових рядів Поняття нечітких часових рядів. Методи</i>

<i>моделювання часових рядів. Методи аналізу та прогнозування поведінки часових рядів.</i>
<i>Тема 5. Вирішення задачі кластеризації Постановка задачі кластеризації та представлення результатів. Види кластерів. Міри близькості, засновані на відстанях. Базові алгоритми кластеризації. Адаптивні методи кластеризації.</i>
<i>Тема 6. Класифікація. Постановка задачі класифікації. Контрольована непараметрична класифікація. Контрольована непараметрична нейромережева класифікація. Класифікація за методом опорних векторів. Дерева рішень. Неконтрольована класифікація.</i>
<i>Тема 7. Регресія. Поняття регресії. Основні етапи регресійного аналізу. Методи відновлення регресії.</i>
<i>Тема 8. Вирішення задачі пошуку асоціативних правил Постановка задачі пошуку асоціативних правил та представлення результатів. Секвенціальний аналіз. Різновиди задач пошуку асоціативних правил. Методи подання результатів. Алгоритми пошуку асоціативних правил. Методи пошуку асоціативних правил: метод Apriori, побудова FP-дерев пошуку шаблонів даних. Min-tax асоціації у базах даних. Побудова hash-дерев.</i>
<i>Тема 9. Застосування нейронних мереж. Самоорганізаційна карта Кохонена. Початкова ініціалізація карти. Навчання. Застосування алгоритму.</i>
<i>Тема 10. Застосування генетичних алгоритмів. Основні визначення. Схема генетичного алгоритму. Генетичні операції. Представлення даних. Відбір.</i>
<i>Тема 11. Способи візуального представлення даних. Методи візуалізації. Представлення даних в одному, двох та трьох вимірах. Паралельні координати. Якість візуалізації. Основні тенденції в області візуалізації.</i>
<i>Тема 12. Очистка та предобробка даних. Оцінка якості даних. Технології та методи оцінки якості даних. Очистка та предобробка. Фільтрація даних. Відновлення пропусків даних. Скорочення розмірності. Скорочення кількості ознак та записів. Семпінг.</i>
<i>Тема 13. Ансамблі моделей. Вступ. Беггінг. Бустінг. Альтернативні методи побудови ансамблів.</i>
<i>Тема 14. Порівняння моделей. Оцінка ефективності та порівняння моделей. Оцінка помилки моделі. ROC-аналіз. Навчання в умовах незбалансованості класів.</i>
<i>Тема 15. Оперативний аналіз даних Розгортання OLAP-кубів. Операції над OLAP-кубами (зріз, обертання, консолідація, деталізація). Архітектура OLAP-систем: MOLAP, ROLAP, HOLAP.</i>
<i>Тема 16. Сховища даних 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</i>
<i>Тема 17. Інтелектуальний аналіз просторових даних. Класифікація супутникових даних</i>
<i>Тема 18. Відкриті бібліотеки інструментів інтелектуального аналізу даних. Програмні засоби інтелектуального аналізу даних</i>

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. Барсегян А. А. , Куприянов М. С. Степаненко В. В., Холод И. И. Технологии анализа данных: Data Mining, Visual Mining, Text Mining, OLAP /— 2-е изд.,перераб. и доп. — СПб.: БХВ-Петербург, 2010. — 384 с:
3. Дюк В., Самойленко А. Data Mining: учебный курс. - СПб: Питер, 2001. -368 с.
4. Ситник В.Ф. Системи підтримки прийняття рішень: Навч.посібник. - К.:КНЕУ, 2003.

5. Филиппов В.А. *Интеллектуальный анализ данных: методы и средства.* – М.: Эдиториал, 2001.
6. Барсегян А. А. *Методы и модели анализа данных: OLAP и Data Mining* / А. А. Барсегян, М. С. Куприянов, В. В. Степаненко та ін. – 2-е изд., перераб. и доп. – СПб. : БХВ-Петербург, 2004. – 336 с.
7. Корнеев В. В. *Базы данных. Интеллектуальная обработка информации* / В. В. Корнеев, А. Ф. Гареев, С. В. Васютин и др. – М. : Издатель Могачева С. В. ; Издательство Нолидж, 2001. – 496 с.
8. Рассел С. *Искусственный интеллект: современный подход* / Рас-сел С., Норвинг П. – М. : Издательский дом "Вильямс", 2006. – 1408 с.
9. Вагин В. Н. *Дедукция и обобщение в системах принятия решений* Вагин, В. Н. – М. : Наука, 1988. – 383 с.
10. Дюк В. *"Обработка данных на ПК в примерах"* / Дюк В. – СПб. : Питер, 1997. – 240 с.
11. Загоруйко Н. Г. *Прикладные методы анализа данных и знаний* Н. Г. Загоруйко. – Новосибирск : Изд. Ин-та математики СО РАН, 1999. – 270 с.
12. Каллан Р. *Основные концепции нейронных сетей* / Р. Каллан. Пер. с англ. – М. : Издательский дом "Вильямс", 2001. – 292 с.
13. Круглов В. В. *Искусственные нейронные сети: Теория и практика* /В. В. Круглов, В. В. Борисов. – М. : Горячая линия – Телеком, 2001. – 383 с.
14. Лбов Г. С. *Анализ данных и знаний : учебн. пособ.* / Г. С. Лбов. – Новосибирск : Изд. НГТУ, 2001. – 90 с.
15. Романов А. Н. *Советующие информационные системы в экономике : учебн. пособ. для вузов* / А. Н. Романов, Б. Е. Одинцов. – М. : ЮНИТИ-ДАНА, 2000. – 488 с.
16. Тюрин Ю. Н. *Статистический анализ данных на компьютере* / Ю. Н. Тюрин, А. А. Макаров ; под ред. В. Э. Фигурнова. – М. : ИНФРА-М, 1998. – 528 с.
17. Дюк В. А. *Применение технологий интеллектуального анализа данных в естественнонаучных, технических и гуманитарных областях [Электронный ресурс]* / В. А. Дюк, А. В. Флегонтов, И. К. Фомина. – Режим доступа : ftp://lib.herzen.spb.ru/text/dyuk_138_77_84.pdf.
18. Шапот М. *Интеллектуальный анализ данных в системах поддержки принятия решений [Электронный ресурс]* / М. Шапот. – Режим доступа : <http://www.osp.ru/os/1998/01/179360/>.

Додаткова література:

1. Agrawa R. *Fast Discovery of Association Rules* / R. Agrawal, R. Srikant // In Proc. of the 20th International Conference on VLDB, Santiago, Chile, September 1994.
2. Agrawal R. *Mining Associations between Sets of Items in Massive Databases* / R. Agrawal, T. Imielinski, A. Swami // In Proc. of the 1993 ACM-SIGMOD Int'l Conf. on Management of Data, 207-216.
3. Agrawal R. *Mining Generalized Association Rules* / R. Srikant, R. Agrawal // In Proc. of the 21th International Conference on VLDB, Zurich, Switzerland, 1995.
4. Agrawal R. *Mining quantitative association rules in large relational tables* / R. Srikant, R. Agrawal // In Proceedings of the ACM SIGMOD Conference on Management of Data, Montreal, Canada, June 1996.
5. Babcock B. *Maintaining variance and k-medians over data stream windows* / B. Babcock, M. Datar, R.
6. Motwani, and L. O'Callaghan // Proc. ACM Symp. on Principles of Database Systems, pp. 234–243, 2003.
7. Baestaens D.E. *Neural Network Solution for Trading in Financial Markets* / Dirk Emma Baestaens, Willem Max Van Den Bergh, Douglas Wood // Pitman publishing

9. Berry Michael J. A. *Data Mining techniques: for marketing, sales, and customer relationship management* / Michael J.A. Berry, Gordon Linoff // 2nd ed.
10. BIlmes J. *A Gentle Tutorial on the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models* / Tech. Report ICSI- TR-97-021. — 1997.
11. Bradley P.S. *Scaling EM (Expectation-Maximization) Clustering to Large Databases* / Paul S. Bradley, Usama M. Fayyad, Cory A. // Microsoft Research. — 1999. Wiley Publishing, Inc" 2006.
12. Agrawal R., Srikant R. *Mining Sequential Patterns* // Journal Intelligent Systems, 1997. Vol. 9. N21, P. 33-56,
13. Agrawal R" Mannila H., Srikant R., Toivonen H., Verkamo A.], *Fast Discovery of Associations Rules* // Advances in Knowledge Discovery and Data Mining. American Association for Artificial Intelligence, 1996. P. 307-328,
14. Anderson R, *The Credit Scoring Toolkit* // Theory and Practice for Retail Credit Risk Management and Decision Automation. OXFORD University Press, 2007.
15. Berthold M., Hand D.]. (Eds.) *Intelligent Data Analysis*. Berlin; Heidelberg: Springer Verlag, 2007,
16. Chiu S., Tavella D. *Data Mining and Market Intelligence for Optimal Marketing Returns*. Elsevier Inc., 2003,
17. Hamerly G.Elkan C. *Learning in the K-Means* //Neural Information Processing Systems. MIT Press, 2003,

Навчальний контент

5. Методика опанування навчальної дисципліни (освітнього компонента)

Лекції, практичні заняття, виконання аналітично-розрахункових завдань, дискусії, аналіз конкретних ситуацій, командна робота, модульна контрольна робота.

Структура кредитного модуля

Назви розділів і тем	Кількість годин				
	Всього	у тому числі			
		Лекції	Практичні (семінарські)	Лабораторні	СРС
1	2	3	4	5	6
«Вступ до інтелектуального аналізу даних»					
Тема 1. Основи інтелектуального аналізу даних. Визначення Data Mining і область застосування. Задачі, моделі та методи Data Mining. Методи, стадії, задачі Data Mining. Поняття Business Intelligence.	7	2		1	4
Тема 2. Інтелектуальний аналіз даних у бізнесі Приклади застосування у різних сферах. Застосування інтелектуального аналізу для вирішення складних задач.	7	2		1	4
Тема 3. Етапи інтелектуального аналізу даних. Загальні закономірності при аналізі даних. Групи задач аналізу даних. Класифікація методів аналізу.	7	2		1	4
Тема 4. Методи аналізу часових рядів Поняття нечітких часових рядів. Методи моделювання часових рядів.	7	2		1	4

Методи аналізу та прогнозування поведінки часових рядів.					
Тема 5. Вирішення задачі кластеризації. Постановка задачі кластеризації та представлення результатів. Види кластерів. Міри близькості, засновані на відстанях. Базові алгоритми кластеризації. Адаптивні методи кластеризації.	7	2		1	4
Тема 6. Класифікація. Постановка задачі класифікації. Контрольована непараметрична класифікація. Контрольована непараметрична нейромережева класифікація. Класифікація за методом опорних векторів. Дерева рішень. Неконтрольована класифікація.	7	2		1	4
Тема 7. Регресія. Поняття регресії. Основні етапи регресійного аналізу. Методи відновлення регресії.	7	2		1	4
Тема 8. Вирішення задачі пошуку асоціативних правил. Постановка задачі пошуку асоціативних правил та представлення результатів. Секвенціальний аналіз. Різновиди задач пошуку асоціативних правил. Методи подання результатів. Алгоритми пошуку асоціативних правил. Методи пошуку асоціативних правил: метод Аргіогі, побудова FP-дерев пошуку шаблонів даних. Міп-тах асоціації у базах даних. Побудова hash-дерев.	7	2		1	4
Тема 9. Застосування нейронних мереж. Самоорганізаційна карта Кохонена. Початкова ініціалізація карти. Навчання. Застосування алгоритму.	7	2		1	4
Тема 10. Застосування генетичних алгоритмів. Основні визначення. Схема генетичного алгоритму. Генетичні операції. Представлення даних. Відбір.	7	2		1	4
Тема 11. Способи візуального представлення даних. Методи візуалізації. Представлення даних в одному, двох та трьох вимірах. Паралельні координати. Якість візуалізації. Основні тенденції в області візуалізації.	7	2		1	4
Тема 12. Очистка та предобробка даних. Оцінка якості даних. Технології та методи оцінки якості даних. Очистка та предобробка. Фільтрація даних. Відновлення пропусків даних. Скорочення розмірності. Скорочення кількості ознак та записів. Семпінг.	7	2		1	4
Тема 13. Ансамблі моделей. Вступ.	7	2		1	4

Бегінг. Бустінг. Альтернативні методи побудови ансамблів.					
Тема 14. Порівняння моделей. Оцінка ефективності та порівняння моделей. Оцінка помилки моделі. ROC-аналіз. Навчання в умовах незбалансованості класів.	7	2		1	4
Тема 15. Оперативний аналіз даних Розгортання OLAP-кубів. Операції над OLAP-кубами (зріз, обертання, консолідація, деталізація). Архітектура OLAP-систем: MOLAP, ROLAP, HOLAP.	5	2		1	2
Тема 16. Сховища даних 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	5	2		1	2
Тема 17. Інтелектуальний аналіз просторових даних. Класифікація супутникових даних	5	2		1	2
Тема 18. Відкриті бібліотеки інструментів інтелектуального аналізу даних. Програмні засоби інтелектуального аналізу даних	7	2		1	4
Разом за розділом	120	36		18	66
<i>Екзамен</i>					
Всього годин	120	36		18	66

Лекційні заняття

№ з/п	Назва теми лекції та перелік основних питань
1.	Основи інтелектуального аналізу даних. Визначення Data Mining і область застосування. Задачі, моделі та методи Data Mining. Методи, стадії, задачі Data Mining. Поняття Business Intelligence.
2.	Інтелектуальний аналіз даних у бізнесі Приклади застосування у різних сферах. Застосування інтелектуального аналізу для вирішення складних задач.
3.	Етапи інтелектуального аналізу даних. Загальні закономірності при аналізі даних. Групи задач аналізу даних. Класифікація методів аналізу.
4.	Методи аналізу часових рядів Поняття нечітких часових рядів. Методи моделювання часових рядів. Методи аналізу та прогнозування поведінки часових рядів.
5.	Вирішення задачі кластеризації Постановка задачі кластеризації та представлення результатів. Види кластерів. Міри близькості, засновані на відстанях. Базові алгоритми кластеризації. Адаптивні методи кластеризації.
6.	Класифікація. Постановка задачі класифікації. Контрольована непараметрична класифікація. Контрольована непараметрична нейромережева класифікація. Класифікація за методом опрних векторів. Дерева рішень. Неконтрольована класифікація.
7.	Регресія. Поняття регресії. Основні етапи регресійного аналізу. Методи відновлення

	регресії.
8.	Вирішення задачі пошуку асоціативних правил Постановка задачі пошуку асоціативних правил та представлення результатів. Секвенціальний аналіз. Різновиди задач пошуку асоціативних правил. Методи подання результатів. Алгоритми пошуку асоціативних правил. Методи пошуку асоціативних правил: метод Apriori, побудова FP-дерев пошуку шаблонів даних. Min-max асоціації у базах даних. Побудова hash-дерев.
9.	Застосування нейронних мереж. Самоорганізація карта Кохонена. Початкова ініціалізація карти. Навчання. Застосування алгоритму.
10.	Застосування генетичних алгоритмів. Основні визначення. Схема генетичного алгоритму. Генетичні операції. Представлення даних. Відбір.
11.	Способи візуального представлення даних. Методи візуалізації. Представлення даних в одному, двох та трьох вимірах. Паралельні координати. Якість візуалізації. Основні тенденції в області візуалізації.
12.	Очистка та предобробка даних. Оцінка якості даних. Технології та методи оцінки якості даних. Очищення та предобробка. Фільтрація даних. Відновлення пропусків даних. Скорочення розмірності. Скорочення кількості ознак та записів. Семпінг.
13.	Ансамблі моделей. Вступ. Беггінг. Бустінг. Альтернативні методи побудови ансамблів.
14.	Порівняння моделей. Оцінка ефективності та порівняння моделей. Оцінка помилки моделі. ROC- аналіз. Навчання в умовах незбалансованості класів.
15.	Оперативний аналіз даних Розгортання OLAP-кубів. Операції над OLAP-кубами (зріз, обертання, консолідація, деталізація). Архітектура OLAP-систем: MOLAP, ROLAP, HOLAP.
16.	Сховища даних 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
17.	Інтелектуальний аналіз просторових даних. Класифікація супутникових даних
18.	Відкриті бібліотеки інструментів інтелектуального аналізу даних. Програмні засоби інтелектуального аналізу даних

Лабораторні заняття

Основні завдання циклу лабораторних занять набути практичних навичок застосування методів інтелектуального аналізу даних

№ з/п	Назва теми заняття
1	Методи, стадії, задачі Data Mining. Поняття Business Intelligence. Застосування інтелектуального аналізу для вирішення задач у різних сферах
2	Етапи інтелектуального аналізу даних. Загальні закономірності при аналізі даних. Методи аналізу та моделювання часових рядів
3	Вирішення задачі кластеризації
4	Класифікація. Дерева рішень.
5	Вирішення задачі пошуку асоціативних правил
6	Застосування нейронних мереж
7	Способи візуального представлення даних. Методи візуалізації
8	Ансамблі моделей. Беггінг. Бустінг. Альтернативні методи побудови ансамблів.
9	Сховища даних, використання. Архітектура сховища даних

6. Самостійна робота студента/аспіранта

До самостійної роботи студентів відноситься підготовка до аудиторних занять, опанування тем, що винесені на самостійне вивчення, проведення обчислень та написання програм, за первинними даними, отриманими на практичних заняттях, підготовка до модульної контрольної роботи та екзамену. Відповідно до навчального плану на самостійну роботу відводиться 66 годин.

Самостійна робота

№ з/п	Назви тем і питань, що виносяться на самостійне опрацювання та посилання на навчальну літературу	Кількість годин СРС
1	Основи інтелектуального аналізу даних. Визначення Data Mining і область застосування. Задачі, моделі та методи Data Mining. Методи, стадії, задачі Data Mining. Поняття Business Intelligence.	4
2	Інтелектуальний аналіз даних у бізнесі Приклади застосування у різних сферах. Застосування інтелектуального аналізу для вирішення складних задач.	4
3	Етапи інтелектуального аналізу даних. Загальні закономірності при аналізі даних. Групи задач аналізу даних. Класифікація методів аналізу.	4
4	Методи аналізу часових рядів Поняття нечітких часових рядів. Методи моделювання часових рядів. Методи аналізу та прогнозування поведінки часових рядів.	4
5	Вирішення задачі кластеризації Постановка задачі кластеризації та представлення результатів. Види кластерів. Міри близькості, засновані на відстанях. Базові алгоритми кластеризації. Адаптивні методи кластеризації.	4
6	Класифікація. Постановка задачі класифікації. Контрольована непараметрична класифікація. Контрольована непараметрична нейромережева класифікація. Класифікація за методом опорних векторів. Древа рішень. Неконтрольована класифікація.	4
7	Регресія. Поняття регресії. Основні етапи регресійного аналізу. Методи відновлення регресії.	4
8	Вирішення задачі пошуку асоціативних правил Постановка задачі пошуку асоціативних правил та представлення результатів. Секвенціальний аналіз. Різновиди задач пошуку асоціативних правил. Методи подання результатів. Алгоритми пошуку асоціативних правил. Методи пошуку асоціативних правил: метод Apriori, побудова FP-дерев пошуку шаблонів даних. Min-max асоціації у базах даних. Побудова hash-дерев.	4
9	Застосування нейронних мереж. Самоорганізація карта Кохонена. Початкова ініціалізація карти. Навчання. Застосування алгоритму.	4
10	Застосування генетичних алгоритмів. Основні визначення. Схема генетичного алгоритму. Генетичні операції. Представлення даних. Відбір.	4
11	Способи візуального представлення даних. Методи візуалізації. Представлення даних в одному, двох та трьох вимірах. Паралельні координати. Якість візуалізації. Основні тенденції в області візуалізації.	4
12	Очистка та предобробка даних. Оцінка якості даних. Технології та методи оцінки якості даних. Очищення та предобробка. Фільтрація даних. Відновлення пропусків даних. Скорочення розмірності. Скорочення кількості ознак та записів. Семпінг.	4
13	Ансамблі моделей. Вступ. Беггінг. Бустінг. Альтернативні методи побудови ансамблів.	4
14	Порівняння моделей. Оцінка ефективності та порівняння моделей. Оцінка	4

	помилки моделі. ROC- аналіз. Навчання в умовах незбалансованості класів.	
15	Оперативний аналіз даних Розгортання OLAP-кубів. Операції над OLAP-кубами (зріз, обертання, консолідація, деталізація). Архітектура OLAP-систем: MOLAP, ROLAP, HOLAP.	2
16	Сховища даних 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	2
17	Інтелектуальний аналіз просторових даних. Класифікація супутникових даних	2
18	Відкриті бібліотеки інструментів інтелектуального аналізу даних. Програмні засоби інтелектуального аналізу даних	4

Політика та контроль

7. Політика навчальної дисципліни (освітнього компонента)

Під час занять (як лекцій, так і практичних) студентам надаються інтерактивні вправи у вигляді задач, оцінюється як активність, так і успішність освоєння матеріалу.

(в тому числі під час виконання домашніх завдань і підготовки до лекцій), призначаються заохочувальні бали.

Всі завдання з описом відповідних етапів їх виконання, а також відповідних контрольних строків завантажені у систему Moodle дисципліни. Усі роботи студенти мають завантажувати в особистому кабінеті Moodle. Дедлайни кожного завдання позначені в завданнях у Moodle. Захисти індивідуальних робіт відбуваються на практичних заняттях. Роботи мають бути виконані з дотриманням академічної доброчесності. У період роботи в дистанційному режимі лектор може запропонувати студентам пройти запропоновані ним онлайн-курси на платформі Coursera. Сертифікати/проміжні результати цих курсів зараховуються бути частково зараховані згідно до Положення розміщеного на сайті КПІ ім. Ігоря Сікорського (https://document.kpi.ua/files/2020_7-124.pdf).

Крім того, передбачається, в межах вивчення навчальної дисципліни, участь в конференціях, форумах, круглих столах тощо.

Політика та принципи академічної доброчесності визначені у розділі 3 Кодексу честі Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського». Детальніше: <https://kpi.ua/code>.

Норми етичної поведінки студентів і працівників визначені у розділі 2 Кодексу честі Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського». Детальніше: <https://kpi.ua/code>.

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

Семестрова атестація проводиться у виді екзамену. Для оцінювання результатів навчання застосовується 100-бальна рейтингова система і університетська шкала оцінювання.

Рейтинг студента з кредитного модуля розраховується зі 100 балів, з них 52 бали складає стартова шкала. Стартовий рейтинг (протягом семестру) складається з балів, що студент отримує за:

- роботу на лабораторних заняттях (9 занять);*

- виконання завдань індивідуального навчально-дослідного завдання.

Критерії нарахування балів:

1. Робота на лабораторних заняттях:

- активна творча робота – 4 бали;
- плідна робота – 3 бали.

2. Виконання індивідуальних навчально-дослідних завдань:

- творча робота – 16-15 балів;
- роботу виконано з незначними недоліками – 14-12 балів;
- роботу виконано з певними помилками – 11-10 балів;
- роботу не зараховано (завдання не виконане або є грубі помилки) – 0 балів.

3. Умовою першої атестації є отримання не менше 8 балів та виконання завдань всіх лабораторних робіт (на час атестації). Умовою другої атестації – отримання не менше 22 балів, виконання завдань всіх лабораторних робіт (на час атестації) та зарахування виконання індивідуальної роботи.

4. Умовою допуску до екзамену є зарахування всіх лабораторних робіт, індивідуальної роботи та стартовий рейтинг не менше 26 балів.

5. На екзамені студенти виконують письмову контрольну роботу. Кожне завдання містить два теоретичних запитання (завдання) і одне практичне. Перелік запитань наведений у Рекомендаціях до засвоєння кредитного модуля. Кожне запитання (завдання) оцінюється у 16 балів за такими критеріями:

- «відмінно», повна відповідь, не менше 90% потрібної інформації (повне, безпомилкове розв'язування завдання) – 16-15 балів;
- «добре», достатньо повна відповідь, не менше 75% потрібної інформації або незначні неточності (повне розв'язування завдання з незначними неточностями) – 14-12 балів;
- «задовільно», неповна відповідь, не менше 60% потрібної інформації та деякі помилки (завдання виконане з певними недоліками) – 11-10 балів;
- «незадовільно», відповідь не відповідає умовам до «задовільно» – 0 балів.

6. Сума стартових балів та балів за залікову контрольну роботу переводиться до залікової оцінки згідно з таблицею:

Бали: практичні заняття + ІДЗ + + залікова контрольна робота	Оцінка	Бали ECTS
100...95	Відмінно	A
94...85	Дуже добре	B
84...75	Добре	C
74...65	Задовільно	D
64...60	Достатньо	E
Менше 60	Незадовільно	FX
Є не зараховані лабораторні роботи або стартовий рейтинг менше 26 балів	Не допущено	F

9. Додаткова інформація з дисципліни (освітнього компонента)

У студентів є можливість отримати позитивну оцінку по дисципліні за умов отримання сертифікатів проходження дистанційних чи онлайн курсів за відповідною тематикою.

У додатку до Силабусу знаходиться перелік питань до екзамену

Робочу програму навчальної дисципліни (силабус):

Складено професором кафедри цифрових технологій в енергетиці, д.т.н., с.н.с. Путренко В.В.

Ухвалено кафедрою ЦТЕ (протокол № 20 від 10.05.23)

Погоджено Методичною комісією НН ІАТЕ КПІ ім. Ігоря Сікорського (протокол № 9 від 26.05.23)