



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 ЕКТС) 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 ЕКТС.

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 1 Здатність до абстрактного мислення, аналізу та синтезу;
- ЗК 7 Здатність до пошуку, оброблення та аналізу інформації з різних джерел;
- 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 11 Здатність до інтелектуального аналізу даних на основі методів обчислювального інтелекту включно з великими та погано структурованими даними, їхньої оперативної обробки та візуалізації результатів аналізу в процесі розв'язування прикладних задач;
- ПР 4 Використовувати методи обчислювального інтелекту, машинного навчання, нейромережевої та нечіткої обробки даних, генетичного та еволюційного програмування для розв'язання задач розпізнавання, прогнозування, класифікації, ідентифікації об'єктів керування тощо;
- ПР 12 Застосовувати методи та алгоритми обчислювального інтелекту та інтелектуального аналізу даних в задачах класифікації, прогнозування, кластерного аналізу, пошуку асоціативних правил з використанням програмних інструментів підтримки багатовимірного аналізу даних на основі технологій DataMining, TextMining, WebMining.

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».

3. Зміст навчальної дисципліни

<i>Перелік розділів і тем дисципліни</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.
Тема 2. Інтелектуальний аналіз даних у енергетиці Приклади застосування у різних сферах. Застосування інтелектуального аналізу для вирішення складних задач.
Тема 3. Етапи інтелектуального аналізу даних. Загальні закономірності при аналізі даних. Групи задач аналізу даних. Класифікація методів аналізу.

Тема 4. Методи аналізу часових рядів Поняття нечітких часових рядів. Методи моделювання часових рядів. 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.
Тема 6. Класифікація. Постановка задачі класифікації. Контрольована непараметрична класифікація. Контрольована непараметрична нейромережева класифікація. Класифікація за методом опорних векторів. Дерева рішень. Неконтрольована класифікація.
Тема 7. Регресія. Поняття регресії. Основні етапи регресійного аналізу. Методи відновлення регресії.
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-max асоціації у базах даних. Побудова hash-дерев.
Тема 9. Застосування нейронних мереж. Самоорганізаційна карта Кохонена. Початкова ініціалізація карти. Навчання. Застосування алгоритму.
Тема 10. Застосування генетичних алгоритмів. Основні визначення. Схема генетичного алгоритму. Генетичні операції. Представлення даних. Відбір.
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.
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.
Тема 13. Ансамблі моделей. Вступ. Беггінг. Бустінг. Альтернативні методи побудови ансамблів.
Тема 14. Порівняння моделей. Оцінка ефективності та порівняння моделей. Оцінка помилки моделі. ROC-аналіз. Навчання в умовах незбалансованості класів.
Тема 15. Оперативний аналіз даних Розгортання OLAP-кубів. Операції над OLAP-кубами (зріз, обертання, консолідація, деталізація). Архітектура OLAP-систем: MOLAP, ROLAP, HOLAP.
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
Тема 17. Інтелектуальний аналіз просторових даних. Класифікація супутникових даних
Тема 18. Відкриті бібліотеки інструментів інтелектуального аналізу даних. Програмні засоби інтелектуального аналізу даних

4. Навчальні матеріали та ресурси

Базова література:

1. Литвин В.В. Аналіз даних та знань: підручник/ В.В. Литвин, В.В. Пасічник, Ю.В. Нікольський.- Л.: Магнолія, 2020.- 276с.
2. Інтелектуальний аналіз даних: практикум/ М.Т. Фісун, І.О. Кравець, П.П. Казмірчук.- Л.: Новий Світ-2000, 2020.- 162с.
3. Гладун А.Я., Рогушина Ю. В. Data Mining: пошук знань в даних. Київ. ТОВ «ВД «АДЕФ-Україна», 2016. — 452 с..

4. Акіменко В.В. Прикладні задачі інтелектуального аналізу даних (DATA MINING). К.: КНУ ім. Тараса Шевченка, 2018. 152 с.
5. Марченко О. О., Россада Т.В. Актуальні проблеми Data Mining: навчальний посібник для студентів факультету комп'ютерних наук та кібернетики. – Київ. – 2017. – 150 с.
6. Ситник В. Ф.,Краснюк М.Т. Інтелектуальний аналіз даних (дейтамайнінг): Навч. посібник. - К: КНЕУ, 2007. - 376 с.
7. Ситник В.Ф. Системи підтримки прийняття рішень: Навч.посібник. - К.:КНЕУ, 2003.
8. Hadley Wickham, Garrett Golemund R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. Sebastopol: O'Reilly Media, 2017. 520 p.
9. Julia Silge, David Robinson. Text Mining with R. O'Reilly Media, Inc., 2017
10. Jan Žižka, František Dařena, Arnošt Svoboda. Text Mining with Machine Learning. Principles and Techniques CRC Press, 2020. 368 p.
11. Kantarjic M. Data Mining. Concepts, Models, Methods and Algorithms / M. Kantarjic, 3rd Ed. – Publisher : Wiley, 2019. – 672 p.
12. Бахрушин В. Є. Методи аналізу даних: Навч. посібник / В. Є. Бахрушин. – Запоріжжя: КПУ, 2011. – 268 с.

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
8. Berry Michael J. A. Data Mining techniques: for marketing, sales, and customer relationship management / Michael J.A. Berry, Gordon Linoff // 2nd ed.
9. 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.
10. 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.
11. Agrawal R., Srikant R. Mining Sequential Patterns // J ournal Intelligent Systems, 1997. Vol. 9. N21, P. 33-56,
12. Agrawal R" Маппіла н., Srikant R., Тоівопеп н., Verkamo A.], Fast Discovery of As sociations Rules // / Advances in Knowledge Discovery and Data Mining. American Association for Artificial Intelligence, 1996. P. 307-328,
13. Anderson R, The Credit Scoring Toolkit // / Theory and Practice for Retail Credit Risk Management and Decision Automation. OXFORD University Press, 2007.
14. Berthold м., Hand D.]. (Eds.) Intelligent Data Analysis. Berlin; Heidelberg: Springer Verlag, 2007,
15. Chiu S., Tavella D. Data Mining and Market Intelligence for Optimal Marketing Returns. Elsevier Inc., 2003,

16. 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, лабораторні заняття, 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. Методи аналізу та прогнозування поведінки часових рядів.	7	2		1	4
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 Apriori, побудова FP-дерев пошуку шаблонів даних. Min-max 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. Бегінг. Бустінг. Альтернативні методи побудови ансамблів.	7	2	1	4
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).	5	2	1	2

Вітрини даних, куби даних, багатомірна модель даних					
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
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 опорних 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. Самоорганізаційна 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 лабораторних 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 опрних векторів. Дерева 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 гентичного 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 помилки моделі. ROC- аналіз. Навчання в умовах незбалансованості класів.	4
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 лабораторних) студентам надаються інтерактивні вправи у вигляді задач, оцінюються 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 лабораторних 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)

1. Максимальна кількість балів з дисципліни дорівнює 100.

2. Нарахування балів по окремих видах робіт:

- роботу на лабораторних заняттях (9 занять);
- написання контрольної роботи (МКР).

Система рейтингових (вагових) балів та критерії оцінювання

1. Виконання лабораторних робіт

Оцінюються 9 робіт, передбачених робочою програмою. Максимальний ваговий бал гЛР =63

Сума вагових балів лабораторних робіт:

№ п.п.	Назва лабораторної роботи	Максимальний ваговий бал
1	Методи, стадії, задачі Data Mining. Поняття Business Intelligence. Застосування інтелектуального аналізу для вирішення задач у різних сферах	7
2	Етапи інтелектуального аналізу даних. Загальні закономірності при аналізі даних. Методи аналізу та моделювання часових рядів	7
3	Вирішення задачі кластеризації	7
4	Класифікація. Древа рішень.	7
5	Вирішення задачі пошуку асоціативних правил	7
6	Застосування нейронних мереж	7
7	Способи візуального представлення даних. Методи візуалізації	7
8	Ансамблі моделей. Бегінг. Бустінг. Альтернативні методи побудови ансамблів.	7
9	Сховища даних, використання. Архітектура сховища даних	7
Разом		63

Оцінювання лабораторних робіт:

– якщо робота виконана невчасно знімається 10-30% від максимальної кількості балів (кількість процентів залежить від терміну запізнення);

– якщо робота виконана не самостійно та простежується не індивідуальне виконання то знімається 50% від максимальної кількості балів;

–якщо в програмі не витримані основні правила створення програмних продуктів (модульність, дружній інтерфейс, наявність коментарів та т.п.) знімається 5%.

2. Модульний контроль

На одному з лекційних занять проводиться модульна контрольна робота: Максимальний ваговий бал гМКР = 10.

Оцінювання модульної контрольної роботи виконується наступним чином:

–якщо на всі питання дані повні та чітко аргументовані відповіді, контрольна виконана охайно, з дотримання основних правил, то виставляється 9 - 10 балів;

–якщо методика виконання запропонованого завдання розроблена вірно, але допущені неprincipпові помилки у теоретичному описі або розрахунках, то виставляється 6 - 8 балів;

–від 3 до 5 балів нараховується, якщо методика виконання завдання розроблена в основному вірно, але допущені деякі з наступних помилок: помилки у представленні вихідних даних, не обгрунтовані теоретичні рішення, помилки у методиці розрахунків;

–нижче 3 балів нараховується, якщо завдання не виконане або допущені грубі помилки.

3. Екзамен

Екзамен відбувається у письмовій формі. Максимальна оцінка за екзамен складає гЕК = 27 балів.

Умови позитивної проміжної атестації

Для отримання „зараховано” з першої проміжної атестації студент повинен мати не менше, ніж 12 балів (за умови, що за 8 тижнів згідно з календарним планом контрольних заходів студент повинен отримати 24 бали).

Для отримання „зараховано” з другої проміжної атестації студент повинен мати не менше, ніж 40 балів (за умови, що за 14 тижнів згідно з календарним планом контрольних заходів студент повинен отримати 76 балів).

Розрахунок шкали (R) рейтингу:

Сума вагових балів контрольних заходів протягом семестру складає:

$$R=63+10+27 = 100 \text{ балів}$$

Таким чином, рейтингова шкала з кредитного модуля складає 100 балів.

Умови допуску до іспиту: зарахування всіх лабораторних робіт, а також стартовий рейтинг $r \geq 40$ балів.

Для отримання студентом відповідних оцінок (ECTS та традиційних) його рейтингова оцінка RD переводиться згідно таблиці:

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)

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)