

Міністерство освіти і науки України
Львівський національний університет природокористування
Факультет механіки, енергетики та інформаційних технологій
Кафедра інформаційних технологій



СИЛАБУС
НАВЧАЛЬНОЇ ДИСЦИПЛІНИ
«ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ»
(вибіркова загальноуніверситетська)

освітньо-професійні програми
усіх спеціальностей
перший (бакалаврський) рівень вищої освіти

ВИКЛАДАЧ

Тригуба Анатолій Миколайович



Електронна пошта: *trianamik@gmail.com*

Телефон *+380680506725*

Завідувач кафедри інформаційних технологій Львівського національного університету природокористування, доктор технічних наук, професор. Викладач з 22-річним досвідом, автор та співавтор понад 3000 наукових статей, 4 – підручників та навчальних посібників, 8 монографій, 3 патентів України на винаходи і корисні моделі, 55 навчально-методичних розробок.

Читає курси: Інтелектуальний аналіз даних, Основи проектування інформаційних систем, Обчислювальний інтелект, Інформаційні технології в наукових дослідженнях. Сфера наукових інтересів: проектування інтелектуальних інформаційних систем, розробка інструментарію управління проектами та програмами, обчислювальний інтелект.

Галузь знань:
Спеціальність:
Освітньо-професійна програма
Рівень вищої освіти – перший (бакалаврський)
Кількість кредитів – 3
Рік підготовки, семестр – 4 рік, 7 семестр
Компонент освітньої програми: обов'язкова
Мова викладання: українська

Опис дисципліни

В результаті розвитку інформаційних технологій кількість даних, накопичених в електронному вигляді, зростає швидкими темпами. Ці дані існують в різних форматах: тексти, зображення, аудіо, відео, гіпертекстові документи, реляційні бази даних і т.д. Однак переважна частина доступної інформації не несе для конкретної людини будь-якої користі, так як вона не в змозі переробити таку кількість інформації. Виникає проблема вилучення корисної для користувача інформації з великого обсягу даних. Дисципліна «Інтелектуальний аналіз даних» розглядає технології інтелектуального аналізу даних (Data Mining) – однієї із активно розвинутих областей інформаційних технологій, призначених для виявлення корисних знань з баз даних різної природи.

Програма дисципліни «Інтелектуальний аналіз даних» відноситься до дисциплін професійної підготовки та складена відповідно до освітньо-професійної програми «Комп'ютерні науки» першого (бакалаврського) рівня вищої освіти.

Міждисциплінарні зв'язки: освітня компонента «Інтелектуальний аналіз даних» є складовою частиною циклу професійної підготовки для здобувачів освітньо-професійної програми «Комп'ютерні науки» першого (бакалаврського) рівня вищої освіти. Вивчення дисципліни передбачає наявність систематичних та ґрунтовних знань із суміжних курсів – «Вища математика», «Теорія ймовірності та математична статистика», «Алгоритмізація та програмування», «Інженерія даних та знань», «Основи проектування інформаційних систем», «Системи штучного інтелекту».

Вимоги до знань та умінь визначаються галузевими стандартами вищої освіти України.

Предметом вивчення освітньої компоненти «Інтелектуальний аналіз даних» є процес навчання і підготовки фахівця за освітньо-професійною програмою «Комп'ютерні науки» першого (бакалаврського) рівня вищої освіти, який дозволить використовувати методи та засоби інтелектуального аналізу даних, а також використання практичних інструментів для виявлення знань із великих баз даних.

Метою вивчення освітньої компоненти «Інтелектуальний аналіз даних» є теоретична та практична підготовка здобувачів вищої освіти у напрямку розв'язання задач обробки великих масивів інформації, проектування інформаційного забезпечення інформаційних систем та розробки сценаріїв можливих дій в умовах невизначеності з використанням інтелектуальних методів обчислень.

Основними завданнями освітньої компоненти «Інтелектуальний аналіз даних» є: надання комплексу знань, умінь та навичок здобувачами вищої освіти технологій інтелектуального аналізу даних, підготовка до їх вибору, реалізації і використання при вирішенні прикладних задач, ознайомлення зі станом та перспективами розвитку інтелектуальних методів обчислень як одного з напрямів штучного інтелекту.

Структура курсу

Години аудиторних занять (лек./ практи.)	Тема	Результати навчання	Завдання
2/2	Тема 1. Предмет дисципліни «Інтелектуальний аналіз даних».	Розуміти поняття інтелектуального аналізу даних, методи, стадії та завдання інтелектуального аналізу даних. Знати особливості виконання Data Mining, а також його місце на ринку інформаційних технологій.	Питання, практична робота
2/2	Тема 2. Препроцесінг інформації.	Знати про ентропію і кількість інформації, способи нормалізації і стандартизації вихідних значень. Розуміти особливості виключення неінформативних чинників. Володіти знаннями із застосування аналітико-евристичних алгоритмів визначення вагомих інформативних ознак лінійних регресій у машинному навчанні.	Питання, практична робота
2/4	Тема 3 Кореляційний і регресійний аналіз.	Розуміти принципи кореляційного і регресійного аналізу. Знати форми, тісноту та спрямованість статистичного взаємозв'язку. Вміти застосовувати графічний метод та виконувати оцінення функціональних і статистичних взаємозв'язків.	Питання, практична робота
2/2	Тема 4. Поняття та методи класифікації.	Знати основні задачі та види класифікації. Володіти методами, що застосовуються для розв'язання задач класифікації. Вміти використовувати засоби класифікації.	Питання, практична робота
2/2	Тема 5. Лінійний дискримінантний аналіз. Побудова канонічних та класифікаційних функцій.	Знати особливості виконання лінійного дискримінантного аналізу схем роботи, зв'язок між дискримінантним, дисперсійним і регресійним аналізом. Володіти методами методами дискримінації даних та моделями дискримінантного аналізу.	Питання, практична робота
4/2	Тема 6. Кластеризація. Кластерний аналіз. Ієрархічна та секційна кластеризація.	Знати основні задачі кластеризації, складові кластерного аналізу та ієрархічний кластерний аналіз. Володіти знаннями щодо особливостей застосування методів кластерного аналізу. Вміти використовувати засоби кластерного аналізу.	Питання, практична робота
2/2	Тема 7. Пошук асоціативних правил. Древа рішень.	Знати основні поняття теорії асоціативних правил та напрямки застосування дерев рішень. Вміти використовувати програмні засоби пошуку асоціативних правил.	Питання, практична робота
2/4	Тема 8. Статистична обробка часових рядів і прогнозування.	Знати основні елементи часового ряду та випадкові процеси та часові ряди. Володіти знаннями щодо розкладання (декомпозиції) часового ряду. Вміти виконувати прогнозування часових рядів з використанням ARIMA-моделей, а також прогнозування тенденції на основі згладжування часових рядів.	Питання, практична робота

2/2	Тема 9. Візуалізація даних.	Знати базові поняття, принципи і цілі візуалізації. Володіти знаннями щодо метафори і критеріїв змістовної візуалізації, конвеєра візуалізації як композиції трансформацій прикладних даних. Вміти застосовувати сучасні концепції візуалізації даних.	Питання, практична робота
2/2	Тема 10. Аналіз текстової інформації Text Mining.	Знати задачі та етапи аналізу текстів Text Mining. Володіти знаннями щодо опису процесу отримання понять з тексту, методів класифікації та кластеризації текстових документів: ієрархічні, бінарні. Вміти застосовувати засоби аналізу текстової інформації.	Питання, практична робота
2/2	Тема 11. Добування знань із Web – Web Mining.	Володіти загальними поняттями добування знань з Web - Web Mining. Знати категорії Web Mining. Вміти використовувати методи витягу Web-контента, а також витягу Web-структур.	Питання, практична робота
4/2	Тема 12. Засоби аналізу процесів Process Mining.	Знати технологію Process Mining, основні задачі та методи Process Mining. Вміти використовувати стандарт запису протоколів MXML. Володіти інструментальними засобами для виконання Process Mining.	Питання, практична робота

Навчальний контент

Формування програмних компетентностей

Індекс в матриці ОПШ	Програмні компоненти
ЗК02	Здатність застосовувати знання у практичних ситуаціях.
ЗК07	Здатність до пошуку, оброблення та аналізу інформації з різних джерел.
ФК02	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.
ФК05	Здатність здійснювати формалізований опис задач дослідження операцій в організаційно-технічних і соціально-економічних системах різного призначення, визначати їх оптимальні розв'язки, будувати моделі оптимального управління з урахуванням змін економічної ситуації, оптимізувати процеси управління в системах різного призначення та рівня ієрархії.
ФК06	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 ризики.
ФК11	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.
ПРН1	Застосовувати знання основних форм і законів абстрактно-логічного мислення, основ методології наукового пізнання, форм і методів вилучення, аналізу, обробки та синтезу інформації в предметній області комп'ютерних наук.

ПРН4	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.
ПРН12	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.

Літературні джерела

Базові

1. Гладун А. Я., Рагушина Ю. В. Data Mining: пошук знань в даних. К.: ТОВ «ВД «АДЕФ Україна», 2016. 452 с.
2. Ситник В. Ф., Краснюк М.Т. Інтелектуальний аналіз даних (дейтамайнінг): Навч. посібник. К: КНЕУ, 2007. 376 с.
3. Sarkar D., Bali R., Sharma T. Practical Machine Learning with Python. Apress, 2018. 545 p. URL: www.apress.com/978-1-4842-3206-4.
4. Zaki M. J., Meira W. Jr. Data mining and analysis : Fundamental Concepts and Algorithms. New York : Cambridge University Press, 2014. 604 p.

Допоміжні

1. Литвин В. В., Пасічник В. В., Нікольський Ю. В. Аналіз даних та знань : навчальний посібник. Львів: «Магнолія 2006», 2015. 276 с.
2. Tryhuba, A., Koval, N., Tryhuba, I., Boiarchuk, O. Application of Sarima Models in Information Systems Forecasting Seasonal Volumes of Food Raw Materials of Procurement on the Territory of Communities. CEUR Workshop Proceedings. 2022. 3295, p. 64-75.
3. Tryhuba A., Boyarchuk V., Tryhuba I., Ftoma O., Padyuka R., Rudynets M. Forecasting the risk of the resource demand for dairy farms basing on machine learning (MoMLeT&DS-2020) In: CEUR Workshop Proceedings, 2020, 2631, pp. 327-340.
4. Бахрушин В. Є. Методи аналізу даних : навчальний посібник для . Запоріжжя: КПУ, 2011. 268 с.
5. Інтелектуальний аналіз даних : Комп'ютерний практикум : навч. посібник / О. О. Сергєєв-Горчинський, Г. В. Іщенко. К. : КПІ ім. Ігоря Сікорського, 2018. 73 с.
6. Інтелектуальний аналіз даних : практикум / М. Т. Фісун, І. О. Кравець, П. П. Казмірчук, С. Г. Ніколенко. Л. : «Новий світ-2000», 2016. 162 с.
7. Снитюк В. Є. Прогнозування. Моделі. Методи. Алгоритми : навчальний посібник. К.: Маклаут, 2008. 364 с.
8. Литвин В.В., Пасічник В.В., Яцишин Ю.В. Інтелектуальні системи : підручник. Львів: Новий світ – 2000, 2009. 406с.
9. Математичне забезпечення інформаційно-керуючих систем: підручник / А. М. Гуржій, З. В. Дудар, В. М. Левикін, Б. В. Шамша. Х. : Компанія Сміт, 2006. 448 с.
10. Vikram Dayal. An Introduction to R for Quantitative Economics: Graphing, Simulating and Computing. Springer, 2015. ISBN 978-81-322-2340-5. <http://www.springer.com/978-81-322-2340-5>
11. K. Soetaert, J. Cash, and F. Mazzia. Solving Differential Equations in R. Use R! Springer, 2012. ISBN 978-3-642-28070-2
12. Kantarzic M. Data Mining. Concepts, Models, Methods and Algorithms / M. Kantarzic, 3rd Ed. Publisher : Wiley, 2019. 672 p.
13. Комплект методичних посібників виданих кафедрою, конспект лекцій.

Інформаційні ресурси в Інтернеті

1. Бібліотечно-інформаційні ресурси – книжковий фонд, періодика та фонди на електронних носіях бібліотеки ЛНУП, державних органів науково-технічної інформації, наукових, науково-технічних бібліотек та інших наукових бібліотек України.
2. Віртуальне навчальне середовище ЛНУП. URL: <https://moodle.lnup.edu.ua/>
3. Електронні інформаційні ресурси мережі інтернет:
 - Прискорена гібридна хмарна платформа даних. URL: <https://www.weka.io/accelerated-cloud-data-platform/>
 - JupyterLab. URL: <https://jupyter.org/>
 - Evolutionary Complexity Research Group (Eplex): <http://eplex.cs.ucf.edu/>
 - Матеріали відкритого курсу OpenDataScience [Електронний ресурс]. Електрон. дан. Режим доступу: World Wide Web. URL: <https://habr.com/ru/company/ods/blog/344044>.
 - The latest in machine learning. Papers With Code [Електронний ресурс]. Електрон. дан. Режим доступу: World Wide Web. URL: <https://paperswithcode.com/>.
 - Платформа для змагань з аналітики та передбачувального моделювання. [Електронний ресурс]. Режим доступу: <https://www.kaggle.com/>.
 - Портал відкритих даних України. URL: <https://data.gov.ua/>
 - Weka Machine learning software to solve data mining problems [Електронний ресурс]. URL: https://sourceforge.net/projects/weka/?source=typ_redirect.

Політика оцінювання

Політика щодо дедлайнів та перескладання: Роботи, які здаються із порушенням термінів без поважних причин, оцінюються на нижчу оцінку (75% від можливої максимальної кількості балів за вид діяльності балів). Перескладання модулів відбувається за наявності поважних причин (наприклад, лікарняний).

Політика щодо академічної доброчесності: Списування під час контрольних робіт заборонені (в т.ч. із використанням мобільних девайсів). Мобільні пристрої дозволяється використовувати лише під час он-лайн тестування та підготовки практичних завдань під час заняття.

Політика щодо відвідування: Відвідування занять є обов'язковим компонентом оцінювання. За об'єктивних причин (наприклад, хвороба, працевлаштування, міжнародне стажування) навчання може відбутись в он-лайн формі за погодженням із ведучим викладачем курсу.

Оцінювання

Остаточна оцінка за курс розраховується наступним чином: поточний контроль оцінюється в 50 балів, та складається із двох модулів по 25 балів кожен. В суму балів кожного модуля входять бали за підготовку, виконання та захисту 10 практичних робіт по 4 бали за кожну роботу ($10 \times 4 = 40$) та 2 бали за самостійну роботу, яка оцінюється усна компонента під час здачі модуля (співбесіда із лектором) ($10 \times 2 = 20$) та 20 балів розрахункова робота.

Поточне тестування та самостійна робота (разом 80 балів)				Розрахункова робота	Сума
Модуль 1 (25 балів)		Модуль 2 (25 балів)			
П1- П5	СР	П6- П10	СР		
5 x 6 =30	10	5 x 6 =30	10	20	100

П1, П2 ... П10 – практичні роботи; СР – самостійна робота.

До Силабусу також готуються матеріали навчально-методичного комплексу:

- 1) Навчальний контент (розширений план лекцій);
- 2) Тематика та зміст практичних робіт;

- 3) Тематика та методичні рекомендації до виконання курсової роботи;
- 4) Завдання для підсумкової роботи, питання на іспит;
- 5) Електронне навчання у віртуальному навчальному середовищі ЛНУП (<https://moodle.lnup.edu.ua/>).