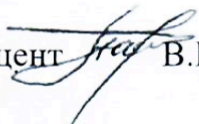


Міністерство освіти і науки України
Львівський національний університет природокористування
Факультет механіки, енергетики та інформаційних технологій
Кафедра інформаційних технологій



ЗАТВЕРДЖЕНО

Гарант освітньо-професійної
програми «Комп'ютерні науки»
першого (бакалаврського) рівня
вищої освіти

к.т.н., доцент  В.В. Пташник

**СИЛАБУС
НАВЧАЛЬНОЇ ДИСЦИПЛІНИ
«ПІДГОТОВКА ТА ОБРОБКА ВЕЛИКИХ ДАНИХ»**

освітньо-професійна програма «Комп'ютерні науки»
спеціальність 122 «Комп'ютерні науки»
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 23-річним досвідом, автор та співавтор понад 300 наукових статей, 4 – підручників та навчальних посібників, 8 монографій, 3 патентів України на винаходи і корисні моделі, 65 навчально-методичних розробок.

Читає курси: Обчислювальний інтелект, Інтелектуальний аналіз даних, Підготовка та обробка великих даних, Управління ІТ проектами, Інформаційні технології в наукових дослідженнях. Сфера наукових інтересів: проектування інтелектуальних інформаційних систем, розробка інструментарію управління проектами та програмами.

ЛЬВІВ 2024

Рівень вищої освіти – перший (бакалаврський)

Галузь знань: 12 «Інформаційні технології»

Спеціальність: 122 «Комп'ютерні науки»

Освітньо-професійна програма «Комп'ютерні науки»

Кількість кредитів – 4

Рік підготовки, семестр – 4 рік, 8 семестр

Компонент освітньої програми: вибіркова

Мова викладання: українська

Опис дисципліни

Дисципліна «Підготовка та обробка великих даних» спрямована на ознайомлення студентів з фундаментальними знаннями з теорії та практики в області розробки і використання систем обробки і підготовки великих масивів даних. Дисципліна знайомить з базовими поняттями великих даних і висвітлює питання підготовки великих даних та пов'язаних з ними технічних, концептуальних та етичних проблем. Здобувачі вищої освіти набувають практичні навички з розробки та представлення концепції великих даних для конкретних предметних областей реального світу. Дисципліна містить практичні завдання для ознайомлення студентів з форматом великих даних, практичний досвід підготовки та обробки великих, складних структур даних.

Програма дисципліни «Підготовка та обробка великих 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 «Підготовка та обробка великих 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», «Бази даних», «Інтелектуальний аналіз даних», «Нейронні мережі», «Системи штучного інтелекту».

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

Предметом вивчення освітньої компоненти «Підготовка та обробка великих даних» є методи та технології збору, зберігання, підготовки, обробки, аналізу і візуалізації великих обсягів даних для виявлення закономірностей, тенденцій і взаємозв'язків. Це включає використання сучасних інструментів і підходів для роботи з великими даними, зокрема машинного навчання, статистичних методів, алгоритмів підготовки та обробки даних, баз даних та засобів програмування.

Метою вивчення освітньої компоненти «Підготовка та обробка великих даних» є формування у студентів знань, умінь і навичок, необхідних для ефективної роботи з великими обсягами даних.

Основними завданнями освітньої компоненти «Підготовка та обробка великих даних» є: вивчення основних понять, принципів і характеристик великих даних, а також розуміння важливості і застосувань великих даних у різних галузях, ознайомлення з технологіями та інструментами для збору великих обсягів даних з різних джерел, а також з методами їх ефективного зберігання, включаючи бази даних, сховища даних та хмарні технології, вивчення алгоритмів і методів для обробки великих обсягів даних, включаючи статистичний аналіз, машинне навчання, кластеризацію, класифікацію та інші аналітичні підходи, використання інструментів і методів для візуалізації великих обсягів даних, що допомагає зрозуміти та інтерпретувати результати аналізу, а також представити їх у зрозумілому вигляді для прийняття рішень.

Структура курсу

Години аудиторних занять (лек./ практи.)	Тема	Результати навчання	Завдання
2/2	Тема 1. Аналіз та обробка великих даних.	Розуміти актуальність даних в сучасному світі. Розуміти поняття великих даних. Знати характеристики великих даних. Знати та вміти застосовувати методи аналізу та обробки великих даних.	Питання, практична робота
2/2	Тема 2. Методи та засоби інтеграції даних.	Розуміти актуальність задачі обробки різнотипних даних. Розуміти визначення поняття «консолідація» та знати його зв'язок з процесами обробки різнотипних даних. Знати та вміти застосовувати методи обробки даних з різних джерел. Знати стандарти обробки різнотипних даних. Знати та вміти застосовувати засоби інтеграції даних з різнотипних джерел.	Питання, практична робота
2/2	Тема 3. Методи та засоби забезпечення якості даних.	Знати причини забезпечення якості консолідованих даних. Знати стандарти забезпечення якості даних. Вміти аналізувати та застосовувати методи та засоби відновлення відсутніх даних в наборах великих даних.	Питання, практична робота
2/2	Тема 4. Відновлення відсутніх даних.	Знати та вміти застосовувати моделі великих даних для задачі відновлення відсутніх даних. Знати та вміти застосовувати методи відновлення відсутніх даних з використанням продукційних залежностей і правил асоціації. Вміти оцінювати складність алгоритмів відновлення відсутніх даних.	Питання, практична робота
2/2	Тема 5. Бази даних та сховища даних.	Розуміти особливості та вміти застосовувати бази даних NoSQL. Знати принципи організації та функціонування і вміти застосовувати MongoDB, PostgreSQL. Вміти проводити дослідження швидкодії роботи різних баз даних. Вміти проводити обробку запитів на локальній машині. Вміти проводити обробку запитів на машині на сервісах Amazon.	Питання, практична робота
2/2	Тема 6. Розподілені платформи.	Знати особливості та вміти застосовувати інфраструктури розподілених обчислень Hadoop, Spark, Dryad, Message Passing Interface.	Питання, практична робота

2/2	Тема 7. Інструментальні засоби.	Знати вимоги до програмних систем. Знати особливості та вміти застосовувати системи Rapid Miner, Orange, Weka, Knime, Alteryx. Знати особливості та вміти застосовувати мови програмування Java, Scala, R, Python. Вміти проводити порівняльний аналіз мов програмування. Знати особливості та вміти застосовувати бібліотеки мови програмування Python.	Питання, практична робота
2/2	Тема 8. Застосування глибоких нейронних мереж для аналізу та обробки великих даних.	Знати особливості архітектури глибоких нейронних мереж. Знати алгоритми навчання глибоких нейронних мереж. Вміти застосовувати глибокі нейронні мережі для аналізу та обробки великих даних.	Питання, практична робота
2/2	Тема 9. Методи аналізу великих даних на основі глибоких нейронних мереж.	Вміти класифікувати мережеві пакети на основі глибоких нейронних мереж. Вміти розпізнавати об'єкти на зображеннях супутникових знімків. Вміти розпізнавати об'єкти на зображеннях текстових документів.	Питання, практична робота
2/2	Тема 10. Методи та засоби підвищення ефективності підготовки та обробки великих даних.	Вміти створювати глибокі нейронні мережі на основі еволюційного підходу. Знати методи та засоби підвищення швидкодії нейронної мережі. Знати методи та засоби підвищення швидкодії прийняття рішень на основі нечіткої логіки.	Питання, практична робота
2/2	Тема 11. Побудова інформаційної технології підготовки та обробки великих даних.	Вміти розробляти структурну модель інформаційної технології інтелектуального аналізу та обробки великих даних. Вміти розробляти архітектури систем інтелектуального аналізу та обробки великих даних. Вміти оцінювати ефективність алгоритмів інтелектуального аналізу та обробки великих даних.	Питання, практична робота
2/2	Тема 12. Когнітивний аналіз даних.	Знати та вміти застосовувати технологій когнітивного аналізу даних. Розуміти проект DeepQA. Знати та вміти застосовувати когнітивну систему типу IBM Watson.	Питання, практична робота

Навчальний контент

Формування програмних компетентностей

Індекс в матриці ОПП	Програмні компоненти
ЗК6	Здатність вчитися й оволодівати сучасними знаннями.
ЗК7	Здатність до пошуку, оброблення та аналізу інформації з різних джерел.
СК2	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.
СК3	Здатність до логічного мислення, побудови логічних висновків, використання формальних мов і моделей алгоритмічних обчислень, проектування, розроблення й аналізу алгоритмів, оцінювання їх ефективності та складності, розв'язності та нерозв'язності алгоритмічних проблем для адекватного моделювання предметних областей і створення програмних та інформаційних систем.
СК11	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.
ПРН3	Використовувати знання закономірностей випадкових явищ, їх властивостей та операцій над ними, моделей випадкових процесів та сучасних програмних середовищ для розв'язування задач статистичної обробки даних і побудови прогнозних моделей.
ПРН5	Проектувати, розробляти та аналізувати алгоритми розв'язання обчислювальних та логічних задач, оцінювати ефективність та складність алгоритмів на основі застосування формальних моделей алгоритмів та обчислюваних функцій.
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

Базова

1. Томас Єрл, Ваджид Хаттак, Пол Булер. Основи Big Data: Концепції, алгоритми та технології / Пер.з англ. Анатолія Гладуна; За наук. ред. Олексія Найдю. Дніпро: «Баланс Бізнес Букс», 2018. 320 с.
2. Кучеров Д.П. Методи аналізу великих даних «Big Data». Київ. 2020. 237 с.
3. Ланде, Д. В. Оброблення надвеликих масивів даних (Big Data) [Електронний ресурс] : навчальний посібник для використання у навчальному процесі з підготовки фахівців другого (магістерського) рівня вищої освіти зі спеціальності 122 «Комп'ютерні науки» / Д. В. Ланде, І.Ю. Субач, А. Я. Гладун ; КПІ ім. Ігоря Сікорського. Електронні текстові дані. Київ : КПІ ім. Ігоря Сікорського, 2021. 168 с.

4. Навчальний посібник з дисципліни “Технології Big Data” для студентів спеціальності 123 “Комп’ютерна інженерія” / Таран В.І., Гордієнко Ю.Г., Стіренко С.Г. Київ: КПІ, 2022. 56 с.

Допоміжна

5. Технології оброблення великих даних: конспект лекцій з дисципліни «Технології оброблення великих даних» [Електронний ресурс] : навч. посіб. / Л.М. Олещенко; КПІ ім. Ігоря Сікорського. Електронні текстові дані. Київ: КПІ ім. Ігоря Сікорського, 2021. 227 с.

6. Zgurovsky M.Z., Zaychenko Y.P. Big Data: Conceptual Analysis and Applications. Springer, 2020. 298 p.

7. Wiktorski Tomasz. Data-intensive Systems: Principles and Fundamentals using Hadoop and Spark. Springer, 2019. 105 p.

8. Wang C., Shakhovska N., Sachenko A., Komar M. A New Approach for Missing Data Imputation in Big Data Interface. Information Technology and Control. 2020. Vol. 49. No 4. P. 541-555.

9. Комплект методичних посібників виданих кафедрою, конспект лекцій.

Інформаційні ресурси в Інтернеті

10. The latest in machine learning. Papers With Code [Електронний ресурс]. Електрон. дан. Режим доступу: World Wide Web. URL: <https://paperswithcode.com/>

11. Платформа для змагань з аналітики та передбачувального моделювання. [Електронний ресурс]. Режим доступу: <https://www.kaggle.com/>

12. Портал відкритих даних України. [Електронний ресурс]. Режим доступу: <https://data.gov.ua/>

13. Shaw J. Why “Big Data” Is a Big Deal [Електронний ресурс]. Режим доступу: <http://harvardmag.com/pdf/2014/03-pdfs/0314-HarvardMag.pdf>

14. Schutt P. What is Big Data? [Електронний ресурс]. Режим доступу: <https://blogs.oracle.com/bigdata/big-data-andanalytic-top-10-trends-for-2014/>

15. Відкритий посібник з відкритих даних [Електронний ресурс]. Режим доступу: <https://socialdata.org.ua/manual/>

16. Big Data Та Блокчейн – Прорив В Області Аналізу Даних [Електронний ресурс]. Режим доступу: <https://busines.in.ua/big-data-ta-blokchejn-proryv-v-oblasti-analizu-danyh/>

17. Weka Machine learning software to solve data mining problems [Електронний ресурс]. Режим доступу: https://sourceforge.net/projects/weka/?source=typ_redirect

18. Books Ngram Viewer [Електронний ресурс]. Режим доступу: <https://books.google.com/ngrams>

19. Мова програмування R [Електронний ресурс]. Режим доступу: <https://cran.r-project.org>

20. Середовище для розробки програм на R – R Studio [Електронний ресурс]. Режим доступу: <http://www.r-studio.com>

21. Manyika James and others. Big data: The next frontier for innovation, competition, and productivity [Електронний ресурс]. Режим доступу: <http://www.mckinsey.com/business-functions/business-technology/our-insights/big-data-the-next-frontier-for-innovation>

22. IBM Analytics [Електронний ресурс]. Режим доступу: <http://www.ibm.com/analytics/us/en/technology/hadoop/hadoop-trials.html>

23. IBM Cloud [Електронний ресурс]. Режим доступу: https://www.ibm.com/cloud-computing/bluemix/?lnk=hp_trials_uauk

24. IBM Bluemix Promo Code - 6 Month Trial [Електронний ресурс]. Режим доступу: <https://ibm.onthehub.com/WebStore/OfferingDetails.aspx?o=bb3528b7-2b63-e611-9420-b8ca3a5db7a1>

25. Hadoop: Built for big data, insights, and innovation [Електронний ресурс]. Режим доступу: <http://www.ibm.com/analytics/us/en/technology/hadoop/>

26. IBM BigInsights [Електронний ресурс]. Режим доступу: <http://www.ibm.com/analytics/us/en/technology/biginsights/>

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). Перескладання проміжних модулів відбувається за наявності поважних причин (наприклад, лікарняний).

Політика щодо академічної доброчесності: Списування під час тестування, виконання контрольних робіт або підсумкового контролю заборонені (в т.ч. із використанням мобільних девайсів та генеративного інтелекту). Мобільні пристрої дозволяється технічно використовувати лише під час он-лайн тестування та підготовки до виконання завдань.

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 керівником 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 8 лабораторних робіт по 5 бали за кожну роботу ($8 \times 5 = 40$) та 10 балів за індивідуальну роботу, яка оцінюється під час її захисту (співбесіда із викладачем).

Поточне тестування та самостійна робота (разом 50 балів)		Самостій- на робота	Підсумковий контроль	Сума
Модуль 1 (20 балів)	Модуль 2 (20 балів)		екзамен	
Л1- Л4	Л5- Л8			
4 x 5 = 20	4 x 5 = 20	10	50	100

Л1, Л2 ... Л8 – лабораторні 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) Завдання для підсумкової роботи, питання на іспит
- 4) Електронне навчання у системі MODLE.