

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ**  
**Львівський національний університет імені Івана Франка**  
**Факультет прикладної математики та інформатики**  
**Кафедра кібербезпеки**

**Затверджено**

На засіданні кафедри кібербезпеки  
факультету прикладної математики та  
інформатики  
Львівського національного університету  
імені Івана Франка  
(Протокол № 15/23 від 29 серпня 2023 р.)



Завідувач кафедри \_\_\_\_ П.С.Венгерський

**Силабус з навчальної дисципліни**  
**“Методи інтелектуального аналізу даних”,**  
**що викладається в межах ОПП Кібербезпека**  
**першого (бакалаврського) рівня вищої освіти для здобувачів з**  
**спеціальності 125 – кібербезпека та захист інформації**

Львів 2023 р.

<b>Назва дисципліни</b>	Методи інтелектуального аналізу даних
<b>Адреса викладання дисципліни</b>	Головний корпус ЛНУ ім. І. Франка м. Львів, вул. Університетська 1
<b>Факультет та кафедра, за якою закріплена дисципліна</b>	Факультет прикладної математики та інформатики Кафедра кібербезпеки
<b>Галузь знань, шифр та назва спеціальності</b>	12 – інформаційні технології 125 – кібербезпека та захист інформації
<b>Викладачі дисципліни</b>	Гірна Олександра Йосипівна, канд. фіз.-мат. наук, доцент, доцент кафедри вищої математики
<b>Контактна інформація викладачів</b>	<a href="mailto:oleksandra.hirna@lnu.edu.ua">oleksandra.hirna@lnu.edu.ua</a> , Головний корпус ЛНУ ім. І. Франка, каб. 380. м. Львів, вул. Університетська, 1
<b>Консультації з питань навчання по дисципліні відбуваються</b>	Консультація проводиться за розкладом консультацій викладача. Можливі дистанційні консультації за попередньою домовленістю.
<b>Сторінка курсу</b>	<a href="https://e-learning.lnu.edu.ua/course/view.php?id=6132">https://e-learning.lnu.edu.ua/course/view.php?id=6132</a>
<b>Інформація про дисципліну</b>	Дисципліна “ Методи інтелектуального аналізу даних ” є вибірковою дисципліною для спеціальності 125 – кібербезпека та захист інформації для освітньої програми Кібербезпека, яка викладається в 7-му семестрі в обсязі 5-ох кредитів (за Європейською Кредитно-Трансферною Системою ECTS).
<b>Коротка анотація дисципліни</b>	Навчальну дисципліну розроблено так, щоб надати учасникам розуміння базових концепцій інтелектуального аналізу даних та алгоритмів розділу штучного інтелекту – машинного навчання, що займається розбудовою розумних моделей, а також їх реалізацій засобами мови програмування Python. У дисципліні представлені алгоритми та інструментальні засоби первинної статистичної обробки та очищення даних, різноманітні засоби групування та візуалізації; алгоритми класичного навчання з вчителем та без вчителя, а також ансамблеві методи покращення якості моделі.
<b>Мета та цілі дисципліни</b>	<b>Метою</b> навчальної дисципліни є формування у студентів системи знань базових концепцій та алгоритмів інтелектуального аналізу даних і практичних навичок їх застосування для побудови та аналізу моделей на основі вибіркового даних.  <b>Завданням</b> вивчення навчальної дисципліни є сформувати у студентів теоретичні знання та практичні навички в області обробки та представлення даних; методів класифікації, регресії, кластеризації, пошуку асоціативних правил, побудови ансамблів.
<b>Література для вивчення дисципліни</b>	<b>Основна література:</b> 1. Харченко В. О. Основи машинного навчання : навч. посіб. / В. О. Харченко. – Суми : СДУ, 2023. – 264 с. 2. Шарадкін Д.М., Субач І.Ю., Микитюк А.В. Інструментальні засоби Python для моделювання та системного аналізу часових рядів при вирішенні задач кіберзахисту інформаційно-комунікаційних систем: навч. пос. / Шарадкін Д.М., Субач І.Ю., Микитюк А.В.; ІСЗЗІ КПІ ім. Ігоря Сікорського. – Київ : КПІ ім. Ігоря Сікорського, 2023. - 139 с.

	<p>3. Інтелектуальний аналіз даних та машинне навчання. Частина Базові методи та засоби аналізу даних / Я. В. Іванчук, В. І. Месюра, А. А. Яровий, О. Д. Манжілевський – Вінниця : ВНТУ, 2021. – 69 с.</p> <p>4. Susan E. McGregor Practical Python Data Wrangling and Data Quality / O'Reilly Media, 2022. – 578 p.</p> <p>5. Bernd Klein Machine Learning with Python Tutorial / Bodensee, 2021. – 453 p.</p> <p style="text-align: center;"><b>Додаткова література:</b></p> <p>6. Ланде Д.В., Субач І.Ю., Бояринова Ю.Є. Основи теорії і практики інтелектуального аналізу даних у сфері кібербезпеки: навчальний посібник. - К.: КПІ ім. Ігоря Сікорського», 2018. - 300 с.</p> <p>7. Черняк О.І. Інтелектуальний аналіз даних: Підручник / О.І. Черняк, П.В. Захарченко ; Київ. нац. ун-т ім. Т. Шевченка. — К. : Знання, 2014. - 599 с.</p> <p>8. Гороховатський В.О., Творошенко І.С. Методи інтелектуального аналізу та оброблення даних: навч. посібник. – Харків: ХНУРЕ, 2021. - 92 с.</p> <p>9. Ланде Д.В., Субач І.Ю., Бояринова Ю.Є. Основи теорії і практики інтелектуального аналізу даних у сфері кібербезпеки: навчальний посібник. — К.: ІСЗІ КПІ ім. Ігоря Сікорського», 2018. - 300 с.</p> <p>10. Ситник В. Ф. Інтелектуальний аналіз даних (дейтамайнінг): Навч. Посібник/ В. Ф. Ситник, М.Т. Краснюк - К: КНЕУ, 2007. - 376 с.</p> <p style="text-align: center;"><b>Інформаційні ресурси в Інтернет</b></p> <p>11. Сайт ЛНУ ім.Івана Франка <a href="http://www.mmf.lnu.edu.ua/ar/1739">http://www.mmf.lnu.edu.ua/ar/1739</a></p> <p>12. Національний інститут стандартів і технологій (NIST), Стандарти науки про дані та аналізу великих даних: <a href="https://bigdatawg.nist.gov/standards/">https://bigdatawg.nist.gov/standards/</a></p> <p>13. Група спеціальних інтересів з виявлення знань та інтелектуального аналізу даних (SIGKDD) Асоціації обчислювальної техніки (Association for Computing Machinery, ACM): <a href="https://www.kdd.org/">https://www.kdd.org/</a></p> <p>14. Процес міжгалузевого стандарту для інтелектуального аналізу даних (CRISP-DM): <a href="https://www.datascience-pm.com/crisp-dm-2/">https://www.datascience-pm.com/crisp-dm-2/</a></p> <p>15. Data Mining Group (DMG): <a href="https://www.dmg.org/">https://www.dmg.org/</a></p> <p>16. Стандарти мови розмітки предиктивних моделей (PMML): <a href="http://dmg.org/pmml/v4-3/GeneralStructure.html">http://dmg.org/pmml/v4-3/GeneralStructure.html</a></p>
<b>Обсяг курсу</b>	Загальний обсяг: 150 годин. Аудиторних занять: 64 год., з них 32 годин лекцій та 32 години лабораторних занять. Самостійної роботи: 86 годин.
<b>Очікувані результати навчання</b>	<p><b>У результаті вивчення навчальної дисципліни студент повинен знати:</b></p> <ul style="list-style-type: none"> <li>• основні концепції маніпуляції з даними та формування масивів даних;</li> <li>• основні поняття та методи машинного навчання;</li> <li>• методи моделювання, прогнозування та керування на базі гібридних систем;</li> <li>• обчислювального інтелекту;</li> <li>• сутність базових алгоритмів класичного навчання з учителем: та без учителя машинного навчання;</li> <li>• сутність базових алгоритмів побудови ансамблів;</li> <li>• методи та прийоми перевірки, навчання та покращення якості моделі.</li> </ul> <p><b>вміти :</b></p> <ul style="list-style-type: none"> <li>• застосовувати методи: <ul style="list-style-type: none"> <li>✓ класифікації, групування, очищення та візуалізації даних, у тому числі враховувати особливості роботи з часовими рядами;</li> <li>✓ дерев рішень;</li> </ul> </li> </ul>

	<ul style="list-style-type: none"> <li>✓ регресійних моделей;</li> <li>✓ вибору та інженерії ознак моделі;</li> <li>✓ зменшення розмірності моделі;</li> <li>✓ кластеризації;</li> <li>✓ асоціативних правил;</li> <li>✓ побудови ансамблів;</li> </ul> <ul style="list-style-type: none"> <li>• обирати алгоритм навчання системи штучного інтелекту та реалізувати його засобами мови програмування Python для вирішення конкретної практичної задачі.</li> </ul> <p><b>Курс забезпечує набуття таких компетентностей: ІК, КЗ 1, КЗ 5, КФ 2, КФ 5, КФ 12 та програмних результатів навчання: ПРН 2-3, ПРН 4, ПРН 6, ПРН 10-13, ПРН 16, ПРН 28-30.</b></p>
<b>Ключові слова</b>	Інтелектуальний аналіз даних, машинне навчання, навчання з учителем, навчання без учителя, криві навчання.
<b>Формат курсу</b>	Очний, дистанційний
<b>Теми</b>	Завдання та стандарти інтелектуального аналізу даних Дослідницький аналіз даних та задачі описової статистики Основні поняття методів машинного навчання Навчання з учителем: Дерева рішень. Метод k-найближчих сусідів. Регресія. Логістична регресія. Метод максимуму правдоподібності Перевірка моделі та криві навчання Навчання без учителя: Метод головних компонент. Кластеризація. Метод k-середніх. Пошук асоціативних правил Ансамблеві методи Аналіз часових рядів
<b>Підсумковий контроль, форма</b>	Залік у кінці семестру
<b>Пререквізити</b>	Для вивчення дисципліни студенти потребують базові знання з курсів “Моделі та методи дискретної математики”, “Основи математичного аналізу та їх застосування”, “Прикладні основи обчислювальної геометрії та алгебри”, “Застосування теорії ймовірностей в кібербезпеці”, “Програмування”.
<b>Навчальні методи та техніки, які будуть використовуватися під час викладання курсу</b>	Презентації, лекції, лабораторні роботи, індивідуальні завдання, індивідуальні доповіді, самостійна робота. Лекції та лабораторні: інформаційно-рецептивний метод, репродуктивний метод, евристичний метод, метод проблемного викладу. Самостійна робота: репродуктивний метод, дослідницький метод.
<b>Необхідне обладнання</b>	Комп’ютер, Internet.
<b>Критерії оцінювання (окремо для кожного виду навчальної діяльності)</b>	Оцінювання проводиться за 100-бальною шкалою. Бали нараховуються за наступним співвідношенням: <ul style="list-style-type: none"> <li>• лабораторні роботи: 50% семестрової оцінки; максимальна кількість балів – 50;</li> <li>• реферат і доповідь: 20% семестрової оцінки; максимальна кількість балів – 20;</li> <li>• контрольний тест: по 30% семестрової оцінки; максимальна кількість балів – 30.</li> </ul> <p><b>Академічна доброчесність:</b> Очікується, що роботи студентів будуть їх самостійними дослідженнями чи міркуваннями. Виявлення ознак академічної недоброчесності в письмовій роботі студента є підставою для її незарахування викладачем, незалежно від масштабів плагіату чи обману.</p>

	<p><b>Відвідування занять</b> є важливою складовою навчання. Усі студенти зобов'язані відвідувати усі лекції та лабораторні заняття курсу, дотримуватися термінів виконання індивідуальних завдань.</p> <p><b>Література.</b> Уся література, яку студенти не зможуть знайти самостійно, буде надана викладачем виключно в освітніх цілях без права її передачі третім особам. Студенти заохочуються до використання також й іншої літератури та джерел, яких немає серед рекомендованих.</p> <p><b>Політика виставлення балів.</b> Враховуються бали набрані при поточному опитуванні, виконанні самостійних робіт, бали тестувань. Враховуються активність студентів під час занять, своєчасність виконання завдань, не допускається списування та плагіат.</p> <p>Жодні форми порушення академічної доброчесності не толеруються.</p>
<b>Питання до контролю</b>	<ol style="list-style-type: none"> <li>1. Засоби та стандартизовані процеси аналізу даних</li> <li>2. Основні етапи CRISP-метолології</li> <li>3. Описовий аналіз даних з Pandas</li> <li>4. Виявлення та обробка відсутніх значень, дублікатів, викидів</li> <li>5. Групування, таблиці спряженості</li> <li>6. Шкалювання даних</li> <li>7. Описові статистики для числових та категорійних ознак</li> <li>8. Візуалізація даних з Seaborn, Matplotlib, Plotly</li> <li>9. Розподіли ознак, гістограми, коробкові діаграми, точкові графіки</li> <li>10. Особливості обробки часових рядів. Модулі time, datetime</li> <li>11. Основні складові, види та задачі машинного навчання</li> <li>12. Дерева рішень. Критерії якості: ентропія, невизначеність Джині.</li> <li>13. Критерій зупинення алгоритму. Відсікання гілок. Вилучення правил</li> <li>14. Лінійна регресія. Метод найменших квадратів. Теорема Гауса-Маркова</li> <li>15. Декомпозиція дисперсії із зміщенням. Регуляризація моделі</li> <li>16. Метод максимуму правдоподібності</li> <li>17. Лінійна класифікація. Логістична регресія</li> <li>18. L<sub>2</sub>-регуляризація логістичної моделі</li> <li>19. Зменшення помилок моделі та поняття ансамблю. Теорема Кондорсе</li> <li>20. Ідеї методів бутстрепа та бегінгу</li> <li>21. Алгоритм випадкового лісу. Параметри підвищення точності моделі</li> <li>22. Метод k-найближчих сусідів. Визначення класу нового об'єкта. Вибір параметра k.</li> <li>23. Значущість та вибір ознак моделі</li> <li>24. Лассо-регресія і рідж-регресія</li> <li>25. L<sub>1</sub>-регуляризація</li> <li>26. Метод головних компонент. Зменшення розмірності даних</li> <li>27. Методи кластеризації</li> <li>28. Метод k-середніх. Покрокова реалізація. Вибір кількості кластерів</li> </ol>
<b>Опитування</b>	Анкету-оцінку з метою оцінювання якості курсу буде надано по завершенні курсу.

### Схема курсу

Тиж.	Тема заняття	Форма діяльності (заняття)	Література	Заняття, год.	Самост. робота, год.
1.	Засоби та стандартизовані процеси аналізу даних. Можливості Python	лекція	[1-10]	2	5
	Описовий аналіз даних з Pandas Групування, таблиці спряженості	лабораторна		2	

2.	Шкалювання даних. Описові статистики для числових та категорійних ознак	лекція	[1-10]	2	5
		лабораторна		2	
3.	Візуалізація даних з Seaborn, Matplotlib, Plotly Розподіли ознак, гістограми, коробкові діаграми, точкові графіки	лекція	[1-10]	2	6
		лабораторна		2	
4.	Основні складові, види та задачі машинного навчання Виявлення та обробка відсутніх значень, дублікатів, викидів. Особливості обробки часових рядів	лекція	[1-10]	2	5
		лабораторна		2	
5.	Навчання з учителем. Класифікація: Дерева рішень. Метод k-найближчих сусідів Дерева рішень. Критерії якості та зупинення алгоритму. Відсікання гілок. Вилучення правил	лекція	[1-10]	2	5
		лабораторна		2	
6.	Ентропія, невизначеність Джині. Ключові параметри дерева Захист ЛР №1	лекція	[1-10]	2	5
		лабораторна		2	
7.	Лінійна регресія. Метод найменших квадратів. Теорема Гауса-Маркова	лекція	[1-10]	2	6
		лабораторна		2	
8.	Декомпозиція дисперсії із зміщенням. Перевірка та криві навчання. Регуляризація моделі. Метод максимуму правдоподібності	лекція	[1-10]	2	6
		лабораторна		2	
9.	Лінійна класифікація. Логістична регресія. L <sub>2</sub> -регуляризація логістичної моделі	лекція	[1-10]	2	5
		лабораторна		2	
10.	Зменшення помилок моделі та поняття ансамблю. Теорема Кондорсе. Ідеї методів бутстрепа та бегінгу Захист ЛР №2	лекція	[1-10]	2	6
		лабораторна		2	
11.	Алгоритм випадкового лісу. Параметри підвищення точності моделі	лекція	[1-10]	2	5
		лабораторна		2	
12.	Значущість ознак. Вибір ознак моделі. Лассо-регресія і рідж-регресія. L <sub>1</sub> -регуляризація	лекція	[1-10]	2	6
		лабораторна		2	
13.	Навчання без учителя. Метод головних компонент. Зменшення розмірності даних	лекція	[1-10]	2	5
		лабораторна		2	
14.	Методи кластеризації. Метод k-середніх. Покрокова реалізація. Вибір кількості кластерів	лекція	[1-10]	2	5
		лабораторна		2	
15.	Пошук асоціативних правил. Алгоритм Apriori Доповідь за рефератом	лекція	[1-10]	2	5
		лабораторна		2	
16.	Аналіз часових рядів. Методи прогнозування. Згладжування. Лаги часових рядів. Стаціонарність, одиничний корінь Тест	лекція	[1-10]	2	6
		лабораторна		2	
<b>Всього 150 год.</b>				<b>64</b>	<b>86</b>