

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Волинський національний університет імені Лесі Українки
Факультет інформаційних технологій і математики
Кафедра загальної математики та методики навчання інформатики

СИЛАБУС

вибіркового освітнього компонента

МЕТОДИ АНАЛІЗУ ДАНИХ

(назва освітнього компонента)

Підготовки бакалавра

Предметної спеціальності 014.09 Середня освіта (Інформатика)

освітньо-професійної програми Середня освіта. Інформатика

Силабус освітнього компонента «Методи аналізу даних» підготовки бакалавра, галузі знань 01 Освіта, предметної спеціальності Середня освіта (Інформатика), за освітньою програмою «Середня освіта. Інформатика»

Розробник: Хомяк Марія Ярославівна, кандидат фізико-математичних наук, доцент

Погоджено

Гарант освітньо-професійної програми:



Світлана ЯЦЮК

Силабус освітнього компонента затверджено на засіданні кафедри загальної математики та методики навчання інформатики
протокол №7 від 03 лютого 2026 р.

Завідувач кафедри:



Марія ХОМЯК

I. Опис освітнього компонента

Найменування показників	Галузь знань, спеціальність, освітньо-професійна /освітньо-наукова/освітньо-творча програма, освітній рівень	Характеристика освітнього компонента
Денна (очна) форма здобуття освіти	01 Освіта 014.09 Середня освіта (Інформатика) Середня освіта. Інформатика бакалавр	Нормативний
Кількість годин/кредитів 150/5		Рік навчання 3
		Семестр 6-ий
		Лекції 10 год.
		Практичні (семінарські) Лабораторні 20 год.
		Індивідуальні _____ год.
		Самостійна робота 110 год.
ІНДЗ: <u>немає</u>	Консультації 10 год.	
	Форма контролю: залік	
Мова навчання		українська

II. Інформація про викладача (- ів)

Хомяк Марія Ярославівна, кандидат фізико-математичних наук, доцент

Телефон: 097-9655693

Аудиторні заняття проводяться за розкладом:

<http://194.44.187.20/cgi-bin/timetable.cgi>

III. Опис освітнього компонента

1. Анотація ОК.

Силабус вибіркового освітнього компонента «МЕТОДИ АНАЛІЗУ ДАНИХ» складено з урахуванням можливості формування індивідуальної освітньої траєкторії здобувачів освіти першого(бакалаврського) рівня.

2. Мета і завдання освітнього компонента.

Мета - формування у здобувачів вищої освіти системи теоретичних знань та практичних навичок щодо застосування математичних методів та алгоритмів для дослідження великих масивів даних, виявлення в них прихованих закономірностей, прогнозування процесів та прийняття обґрунтованих рішень у професійній сфері.

Для досягнення мети передбачається вирішення таких основних **завдань**:

- **Теоретичні:**

Ознайомлення з основними концепціями Data Science та життєвим циклом аналізу даних.

Вивчення математичних основ статистичного аналізу та машинного навчання.

Розуміння принципів роботи алгоритмів класифікації, кластеризації та регресійного аналізу.

- **Практичні:**

Оволодіння інструментами попередньої обробки («очищення») даних та роботи з пропущеними значеннями.

Набуття навичок візуалізації даних для виявлення аномалій та трендів.

Побудова та валідація аналітичних моделей за допомогою сучасних мов програмування (наприклад, Python або R) та спеціалізованих бібліотек.

Навчання інтерпретації отриманих результатів для розв'язання прикладних задач.

3. Soft skills.

Знання статистичних методів обробки інформації та визначення залежностей між випадковими величинами Уміння користуватись сучасними інформаційними технологіями, включаючи методи отримання, обробки та зберігання наукової інформації Уміння адаптувати, інтерпретувати та узагальнювати результати сучасних математичних та статистичних досліджень для розв'язання теоретичних та прикладних проблем.

4. Структура освітнього компонента.

Назви змістових модулів і тем	Усього	Лек.	Лабор.	Сам. роб.	Конс.	Форма контролю/ Бали
Змістовий модуль 1. Основи роботи з даними та описова статистика						
Тема 1. Вступ до Data Science: Роль аналітика в сучасному світі. Життєвий цикл аналітичного проєкту.	15	1	2	11	1	Лб 9
Тема 2. Типи та структури даних: Кількісні, якісні, часові ряди. Робота з різними форматами (CSV, JSON, SQL).	15	1	2	11	1	Лб 9
Тема 3. Exploratory Data Analysis (EDA): Розрахунок мір центральної тенденції (середнє, медіана, мода) та варіативності.	15	1	2	11	1	Лб 9
Тема 4. Попередня обробка: Очищення даних від «шуму», робота з аномаліями (outliers) та пропущеними значеннями.	15	1	2	11	1	Лб 9
Разом за модулем 1	60	4	8	44	4	36
Змістовий модуль 2. Статистичний висновок та перевірка гіпотез						
Тема 5. Ймовірнісні розподіли: Нормальний розподіл та його значення для аналізу.	15	1	2	11	1	Лб 9
Тема 6. Статистичні гіпотези: Формулювання нульової та альтернативної гіпотез. Поняття p-value.	15	1	2	11	1	Лб 9
Тема 7. Порівняння груп: Використання t-критеріїв Стьюдента та ANOVA (дисперсійний аналіз).	15	1	2	11	1	Лб 9
Тема 8. Кореляційний та регресійний аналіз: Побудова лінійних моделей та прогнозування трендів.	15	1	2	11	1	Лб 9
Разом за модулем 2	60	4	8	11	4	36
Змістовий модуль 3. Візуалізація та Storytelling						
Тема 9. Принципи ефективної візуалізації: Вибір правильного типу графіка (Bar, Line, Scatter, Heatmap).	15	1	2	11	1	Лб 9

Тема 10. Побудова інтерактивних дашбордів: Робота з інструментами візуалізації (наприклад, Seaborn/Plotly або BI-системи).	15	1	2	11	1	Лб 9
Разом за модулем 3	30	2	4	22	2	18
Види підсумкових робіт						Бал
Контрольна робота						10
Всього годин/Балів	150	10	20	110	10	100

5. Завдання для самостійного опрацювання.

- Пошук та вибір датасету:** Самостійний пошук наборів даних на ресурсах (Kaggle, Google Dataset Search, Open Data Portal) для підготовки курсового проекту.
- Скрапінг та імпорт:** Опрацювання методів завантаження даних із різних джерел (API, SQL-бази, HTML-таблиці).
- Feature Engineering:** Вивчення методів створення нових ознак на основі наявних (наприклад, виділення дня тижня з дати).
- Аналіз розподілів:** Самостійне дослідження реальних даних на відповідність закону нормального розподілу (тест Шапіро-Вілка).
- Кейс-стаді «А/В тестування»:** Розрахунок необхідного розміру вибірки та аналіз результатів маркетингового тесту за допомогою p -value.
- Непараметричні методи:** Вивчення критеріїв Манна-Уїтні та Вілкоксона для випадків, коли дані не є нормальними.
- Регуляризація моделей:** Опрацювання методів Lasso та Ridge для запобігання перенаванчання (overfitting) регресійних моделей.
- Оцінка якості класифікації:** Побудова та аналіз матриці помилок (Confusion Matrix), розрахунок метрик Precision, Recall та F1-score.
- Порівняльний аналіз:** Тестування різних алгоритмів (наприклад, дерево рішень проти k-NN) на одному наборі даних.
- Створення інтерактивного дашборду:** Побудова звіту в MS Power BI, Tableau або за допомогою Python-бібліотеки Streamlit.
- Критичний огляд візуалізацій:** Аналіз «поганих» графіків (Misleading visualization) та їх виправлення відповідно до принципів Едварда Тафті.
- Підготовка Executive Summary:** Написання короткого резюме за результатами аналізу, орієнтованого на керівника бізнесу.

IV. Політика оцінювання

Оцінювання знань здобувачів освіти здійснюється згідно «ПОЛОЖЕННЯ про поточне та підсумкове оцінювання знань здобувачів вищої освіти Волинського національного університету імені Лесі Українки».

Освітній компонент складається з трьох змістових модулів та його вивчення передбачає виконання лабораторних робіт, а також контрольного тесту. У цьому випадку підсумкова оцінка за 100-бальною шкалою складається із сумарної кількості балів за усі виконані завдання. Якщо за результатами виконання завдань ОК накопичено не менше 60 балів і здобувач погоджується із цим результатом, то виставляється оцінка за семестр. В іншому випадку здобувач освіти має можливість скласти залік під час ліквідації академічної заборгованості. Ліквідація академічної заборгованості із практики передбачає виконання того ж набору індивідуальних задач із подальшим захистом результатів практики.

Здобувачу освіти також можуть бути зараховані результати навчання, здобуті у процесі

формальної, неформальної та/або інформальної освіти відповідно до «ПОЛОЖЕННЯ про визнання результатів навчання, отриманих у формальній, неформальній та/або інформальній освіті у Волинському національному університеті імені Лесі Українки». Визнанню можуть підлягати результати навчання, що відповідають тематиці освітнього компоненту, його окремому розділу, темі (темам) або індивідуальному завданню, які здобувач освіти самостійно набув, вивчаючи освітні ресурси (семінари, інтернет-курси, професійні стажування та ін.) на онлайн-платформах Prometheus (<https://prometheus.org.ua>), EdEra (<https://www.ed-era.com>) та інших, і підтвердив відповідними сертифікатами.

V. Підсумковий контроль

Із вибіркових освітніх компонентів форма контролю – залік. Оцінювання знань здобувачів освіти здійснюється під час поточного контролю за результатами виконання тих видів робіт, які передбачені силабусом.

Питання, завдання заліку.

1. Етапи життєвого циклу аналізу даних згідно з методологією CRISP-DM.
2. Різниця між структурованими, напівструктурованими та неструктурованими даними.
3. Методи ідентифікації та обробки пропущених значень (imputation).
4. Що таке аномалії (outliers), як вони впливають на результат та методи їх виявлення.
5. Нормалізація та стандартизація даних: коли і навіщо вони застосовуються?
6. Міри центральної тенденції: математичне сподівання, медіана, мода. У яких випадках медіана краща за середнє?
7. Міри мінливості: дисперсія, стандартне відхилення, розмах, інтерквартильний розмах.
8. Статистичні гіпотези: нульова H_0 та альтернативна H_1 . Помилки I та II роду.
9. Рівень значущості та показник p-value: сутність та правила прийняття рішень.
10. Кореляційний аналіз: коефіцієнт Пірсона та Спірмена. Чому кореляція не означає причинно-наслідковий зв'язок?
11. Лінійна регресія: сутність методу та інтерпретація коефіцієнтів моделі.
12. Сутність задачі класифікації. Огляд алгоритму k-найближчих сусідів (k-NN).
13. Логістична регресія: чому вона використовується для класифікації, а не для прогнозування чисел?
14. Кластерний аналіз: метод k-середніх (k-means). Як обрати оптимальну кількість кластерів?
15. Розбиття даних на навчальну (train) та тестову (test) вибірки: мета, пропорції та роль валідації в оцінці якості моделі.
16. Матриця помилок (Confusion Matrix): аналіз Precision, Recall та Accuracy.
17. F1-міра як гармонійне середнє між точністю та повнотою.
18. Основні типи графіків (Scatter plot, Box plot, Histogram) та їх призначення в аналізі.
19. Принципи Data Storytelling: як підготувати звіт для нетехнічного замовника.
20. Етичні аспекти аналізу даних та конфіденційність інформації.

Шкала оцінювання знань здобувачів освіти з освітніх компонентів, де формою контролю є залік

Оцінка в балах	Лінгвістична оцінка
90–100	Зараховано
82–89	
75–81	
67–74	
60–66	
0–59	Незараховано (необхідне перескладання)

VI. Рекомендована література та інтернет-ресурси.

Методичне забезпечення ОК

1. Maria Khomyak & Svitlana Yatsyuk. SOCIOMETRIC TEST AS A METHOD OF STUDYING FOR THE INDIVIDUAL AND THE TEAM. PROCESSING OF RESEARCH RESULTS AND USING THEM IN THE EDUCATIONAL PROCESS. *Proceeding Book of III International Congress of modern studies in social sciences and humanities*, June 7-10, 2022, Karabagh, Azerbaijan. Vol.3, P. 58-62.
2. Khomyak M. Analysis of data on the organization of distance learning. *Middle East International conference on contemporary scientific studies-V*, March 27-28, 2021, Ankara, Turkey. Vol.II. P. 384-386.
3. Хомяк М.Я. Дистанційний курс Moodle: Математичні методи обробки даних психолого-педагогічних досліджень (в тестовому режимі). Режим доступу: <https://moodle.vnu.edu.ua/course/view.php?id=>
4. Хомяк М.Я. Мова програмування R як засіб навчання математичної статистики майбутніх ІТ-фахівців та вчителів інформатики. *Математика. Інформаційні технології. Освіта: тези доп. XI міжнар. наук.-практ. конф. (м. Луцьк, 3-5 червн. 2022 р.)*. Луцьк, 2022.. С. 171-173.
5. Хомяк М. Я. Математичні методи аналізу емпіричних даних: методичні вказівки. Луцьк: ВНУ ім. Лесі Українки, 2021. 16 с.
6. Хомяк М. Я. Математичні методи обробки даних психолого-педагогічних досліджень: методичні вказівки. Луцьк : ВНУ ім. Лесі Українки, 2022. 16 с.
7. Яцюк С.М., Хомяк М.Я., Юнчик В.Л., Чепрасова Т.І. Методика використання цифрових освітніх ресурсів у процесі підготовки майбутніх учителів інформатики. *Професіоналізм педагога: теоретичні й методичні аспекти*. 2021. №16. С.15-25.
8. Хомяк М. Я. Математичні методи обробки даних психолого-педагогічних досліджень: лабораторний практикум для здобувачів освіти спеціальності 014 Середня освіта (Інформатика) другого (магістерського) рівня освіти : Електронний ресурс. Луцьк : ВНУ ім. Лесі Українки, 2025. 32 с.
9. Maria Khomyak Statistics: Course Description. Lutsk : Lesia Ukrainka VNU, 2022. 26 p.

Рекомендована література та інтернет-джерела

1. Маханець Л.Л., Вінничук О.Ю., Григорків М.В. Статистика: лабораторний практикум у STATISTICA 12: навч. посіб. Чернівці : Чернівець. нац. ун-т ім. Ю. Федьковича, 2023. 161 с.
2. Петровська І.Р., Островська К.О. Математично-статистичні методи обробки емпіричних даних психолого-педагогічних досліджень: навчальний посібник. Львів: Друкарня «Справи Кольпінга в Україні», 2021. 140 с
3. Технології оброблення великих даних: конспект лекцій з дисципліни «Технології оброблення великих даних» [Електронний ресурс] : навч. посіб. для студ. спеціальності 121 «Інженерія програмного забезпечення» (освітня програма «Інженерія програмного забезпечення мультимедійних та інформаційно-пошукових систем»)/ Л.М. Олещенко; КПІ ім. Ігоря Сікорського. Електронні текстові дані (1 файл: 5,55 Мбайт). Київ: КПІ ім. Ігоря Сікорського, 2021. 227 с.

Інтернет-ресурси та бази даних

- **Kaggle (kaggle.com):** Найбільша платформа для Data Science проєктів, що містить тисячі датасетів та прикладів коду (Kernels).
- **Towards Data Science (towardsdatascience.com):** Провідний блог на Medium з актуальними статтями про нові методи аналізу.
- **Google Dataset Search (<https://www.google.com/search?q=datasetsearch.google.com>):** Спеціалізований пошуковик для знаходження наборів даних для лабораторних робіт.
- **Our World in Data (ourworldindata.org):** Джерело якісних соціально-економічних даних для аналітичних проєктів.
- **Coursera / edX:** Спеціалізовані курси від Google (Data Analytics Professional Certificate) та IBM (Data Science Professional Certificate), оновлені у 2024-2025 роках.