

|   |  |   |        |     |  |
|---|--|---|--------|-----|--|
|  | Центральнoукраїнський державний університет імені Володимира Винниченка  | <b>Силабус навчальної дисципліни</b>                                      |        |     |  |
|   |  | <b>Аналіз даних</b>   |        |     |  |
|   |  | Статус дисципліни <i>обов'язковий компонент (цикл фахової підготовки)</i> |        |     |  |
| Галузь знань  | Е Природничі науки, математика та статистика   |   |        |     |  |
| Спеціальність   | Е8 Статистика  |   |        |     |  |
| Освітня програма  | Інтелектуальний аналіз даних та цифрова екрнрміка  |   |        |     |  |
| Рівень вищої освіти   | перший (бакалаврський)   |   |        |     |  |
| Форма навчання  | денна  |   |        |     |  |
| Курс  | 3  |   |        |     |  |
| Семестр   | 5, 6   |   |        |     |  |
| Обсяг дисципліни  | Кредити  | 8   | Години | 240 |  |
|   | Лекційні   |   |        | 40  |  |
|   | Практичні  |   |        | 40  |  |
|   | Лабораторні  |   |        |     |  |
|   | Самостійна робота  |   |        | 160 |  |
| Семестровий контроль  | Екзамен, екзамен   |   |        |     |  |
| Викладач  | Гуртовий Юрій Валерійович, кандидат фізико-математичних наук, доцент кафедри математики та цифрових технологій, доцент   |   |        |     |  |
| Контактна інформація  | y.v.hurtovyi@cuspu.edu.ua  |   |        |     |  |
| Кафедра   | Математики, фізики та методик викладання   |   |        |     |  |
| Факультет   | Інформаційних технологій математики та природничих наук  |   |        |     |  |
| Предмет навчання<br>(Що буде вивчатися)   | <ol style="list-style-type: none"> <li>Збір та підготовка: навчитися імпортувати дані з різних джерел та очищувати їх від помилок і пропусків.</li> <li>Візуалізація: Освоїти побудову графіків для вивчення структури та аномалій у даних.</li> <li>Моделювання: опанувати побудову регресійних, класифікаційних та кластерних моделей.</li> <li>Прогнозування: навчитися аналізувати часові ряди та оцінювати точність отриманих прогнозів.</li> <li>Валідація: вміти перевіряти адекватність моделей та їхню придатність для реальних задач.</li> </ol> |   |        |     |  |
| Мета (Чому це цікаво/потрібно вивчати)  | Мета вивчення дисципліни формування теоретичних знань та практичних навичок використання статистичних методів для обробки великих масивів даних, виявлення закономірностей та прийняття рішень.  |   |        |     |  |
| Компетентності  | ЗК-3. Знання й розуміння предметної області та професійної діяльності.<br>ЗК-5. Здатність спілкуватися іноземною мовою та використовувати її знання в предметній області.  |   |        |     |  |

ЗК-7. Здатність вчитися і оволодівати сучасними знаннями.  
ЗК-8. Здатність до пошуку, обробки та аналізу інформації з різних джерел.  
ЗК-9. Здатність приймати обґрунтовані рішення.  
ЗК-12. Здатність працювати автономно.  
ЗК-14. Здатність застосовувати знання у практичних ситуаціях  
ФК-5. Здатність до кількісно-статистичного мислення.  
ФК-7. Здатність робити якісні висновки з кількісних даних.  
ФК-8. Уміння працювати з інформаційними базами даних.  
ФК-11. Здатність використання обчислювальної техніки, спеціалізованих мов програмування та програмних засобів для розв'язання задач і здобуття додаткової інформації.  
ФК-14. Здатність подавати статистичні процедури та результати їхнього застосування у формі, придатній для цільової аудиторії, до якої звертаються, як усно, так і письмово.  
ФК-15. Здатність до аналізу основ і властивостей статистичних алгоритмів та розуміння переваг тих чи інших підходів, у тому числі до оцінки їх обґрунтованості й ефективності.

**Програмні результати**  
(Чому можна навчитися)

У результаті вивчення навчальної дисципліни студенти зможуть:

ПРН-1. Здійснювати професійну письмову й усну комунікації українською мовою та, принаймні, однією з іноземних мов.

ПРН-2. Вміти працювати зі спеціальною літературою іноземною мовою.

ПРН-10. Вміти здійснювати статистичне точкове, інтервальне оцінювання параметрів розподілів випадкових величин і процесів, непараметричне оцінювання, тестувати статистичні гіпотези.

ПРН-11. Вміти аналізувати та прогнозувати лінійні статистичні моделі та моделі регресії, оцінювати їхні параметри.

ПРН-14. Володіти сучасними інформаційними технологіями для створення презентацій, роботи з базами даних, пошуку інформації та обміну нею.

ПРН-17. Знати методи моделювання природничих, соціальних, фінансово-економічних, страхових процесів.

ПРН-18. Вміти застосовувати ймовірно-статистичні моделі та методи для розв'язання прикладних проблем і задач.

## Зміст дисципліни

### Модуль 1. Інструментарій та підготовка даних (Data Wrangling)

Тема 1. Вступ до аналізу даних та налаштування середовища

Роль аналізу даних у сучасній статистиці.

Огляд мов програмування (R проти Python) та IDE (RStudio, Jupyter Notebook).

Типи даних: структуровані, неструктуровані, напівструктуровані.

Встановлення пакетів та робота з документацією.

Тема 2. Маніпуляція даними та очищення (Data Cleaning)

Імпорт даних з різних форматів (CSV, Excel, SQL, API).

Робота з пропущеними значеннями (Missing Data): видалення vs імпутація.

Трансформація даних: фільтрація, сортування, створення нових змінних.

Об'єднання таблиць (Joins) та перегрупування (Reshaping: Wide vs Long format).

### Модуль 2. Розвідувальний аналіз (EDA)

Тема 3. Описова статистика та статистичні розподіли

Розрахунок мір центральної тенденції (середнє, медіана, мода) та мір розсіювання.

Оцінка форми розподілу: асиметрія (skewness) та ексцес (kurtosis).

Виявлення та обробка викидів (Outliers) методами Тюкі та Z-оцінок.

Тема 4. Графічний аналіз та візуалізація

Побудова гістограм та графіків щільності розподілу.

Аналіз взаємозв'язків: діаграми розсіювання (Scatter plots) та матричні графіки.

Візуалізація категоріальних даних (Bar charts, Box-plots).

Принципи побудови ефективних графіків (Grammar of Graphics).

Тема 5. Статистичне висновування в аналізі даних

Довірчі інтервали для параметрів розподілу.

Статистичні гіпотези: нульова та альтернативна, рівні значущості (-value).

Параметричні тести: тест Стьюдента, тест Фішера.

### Модуль 3. Статистичне моделювання взаємозв'язків

Тема 6. Лінійна регресія та діагностика моделей

Парна та множинна лінійна регресія: МНК-оцінки.

Перевірка передумов регресійного аналізу: нормальність залишків, гомоскедастичність.

Мультиколінеарність та фактор інфляції дисперсії (VIF).

Інтерпретація коефіцієнтів та оцінка якості (Adjusted ).

Тема 7. Узагальнені лінійні моделі (GLM)

Логістична регресія: моделювання ймовірності бінарної події.

Відношення шансів (Odds Ratio) та їх інтерпретація.  
Оцінка якості класифікації: матриця помилок (Confusion Matrix), точність, повнота, F1-міра.  
Побудова та аналіз ROC-кривої (AUC).

#### Модуль 4. Аналіз структур та часових залежностей

Тема 8. Кластерний аналіз та навчання без учителя

Міри близькості (Евклідова відстань, Манхеттенська відстань).  
Ієрархічна кластеризація та метод К-середніх (-means).  
Вибір оптимальної кількості кластерів (метод «ліктя», силуетний аналіз).

Тема 9. Зниження розмірності даних

Метод головних компонент (PCA): математична основа та візуалізація.

Факторний аналіз: концепція прихованих змінних.

Інтерпретація головних компонент та їх використання у регресії.

Тема 10. Аналіз часових рядів (Time Series)

Компоненти часового ряду: тренд, сезонність, циклічність, шум.  
Перевірка стаціонарності (тест Діккі-Фуллера).

Автокореляція (ACF) та часткова автокореляція (PACF).

Вступ до моделей ARIMA та експоненційного згладжування.

#### Модуль 5. Сучасні алгоритми та валідація

Тема 11. Древа рішень та ансамблеві методи

Алгоритми побудови дерев (CART, ID3).

Проблема перенавчання (Overfitting) та прунінг (обрізка дерев).

Випадковий ліс (Random Forest) як метод покращення стабільності моделі.

Тема 12. Валідація моделей та відбір ознак

Розподіл вибірки на навчальну, валідаційну та тестову.

Крос-валідація (K-fold Cross-validation).

Методи регуляризації (Ridge, Lasso) для боротьби з мультиколінеарністю

**Критерії оцінювання роботи студентів**

**Підсумкова кількість балів з дисципліни** (максимум 100 балів) визначається як сума балів за виконання практичних завдань, самостійної роботи та бали підсумкового контролю (екзамену). **Поточний контроль** – це оцінювання навчальних досягнень здобувача (рівень теоретичних знань та практичні навички з тем) під час проведення аудиторних занять, практичних занять, організації самостійної роботи, на консультаціях. Поточний контроль реалізується у формі опитування, виконання практичних завдань, а також контролю засвоєння матеріалу, запланованого на самостійне опрацювання тощо. Завданням **підсумкового контролю** є перевірка розуміння студентом програмного матеріалу в цілому, логіки та взаємозв'язків між окремими розділами, здатності творчого використання накопичених знань, уміння сформулювати своє

|                                   |   |
|-----------------------------------|---|
|                                   | ставлення до певної проблеми навчальної дисципліни тощо.  |
| Політика курсу                    | <p><b>Форми участі здобувача у навчальному процесі, які підлягають поточному контролю:</b></p> <ul style="list-style-type: none"> <li>- виконання практичних завдань;</li> <li>- участь у дискусіях, інтерактивних формах організації заняття;</li> <li>- самостійне опрацювання тем;</li> <li>- підготовка конспектів;</li> <li>- систематичність роботи на практичних заняттях, активність під час обговорення питань на лекційних заняттях.</li> </ul> <p><b>Вимоги викладача:</b></p> <ul style="list-style-type: none"> <li>- обов'язкове відвідування навчальних занять (лекційних, практичних);</li> <li>- активність здобувача під час лекційних та практичних занять;</li> <li>- своєчасне та якісне виконання усіх завдань, у тому числі завдань самостійної роботи.</li> </ul> <p><b>Не допускається:</b></p> <ul style="list-style-type: none"> <li>- пропуск занять без поважних причин;</li> <li>- запізнення на заняття;</li> <li>- користування мобільним телефоном, планшетом чи іншими мобільними пристроями під час занять, за винятком використання їх з навчальною метою з дозволу викладача;</li> <li>- списування, плагіат.</li> </ul> |
| Інформаційне забезпечення         | <p>На веб-сторінці бібліотеки на основі системи Ірбіс функціонує електронний каталог, а також інституційний репозитарій. Здобувачі мають доступ до зарубіжних баз періодики.</p> <p>Комп'ютерні навчальні лабораторії, які підключені до загальної університетської мережі та мережі Інтернет і мають актуальне програмне забезпечення.</p>   |
| Матеріально-технічне забезпечення | Лекційні аудиторії з мультимедійним обладнанням, комп'ютерні лабораторії з сучасними робочими станціями.  |