



АНАЛІЗ ВЕЛИКИХ ДАНИХ ТА ШТУЧНИЙ ІНТЕЛЕКТ

Робоча програма навчальної дисципліни (Силабус)

Реквізити навчальної дисципліни

Рівень вищої освіти	<i>Другий (магістерський)</i>
Галузь знань	<i>05 Соціальні та поведінкові науки</i>
Спеціальність	<i>054 Соціологія</i>
Освітня програма	<i>Аналітика соціальних даних</i>
Статус дисципліни (код)	<i>Нормативна, цикл професійної підготовки</i>
Форма навчання	<i>очна(денна)</i>
Рік підготовки, семестр	<i>2 курс, осінній семестр</i>
Обсяг дисципліни	<i>4 кредити ЄКТС, 120 годин лекції 18 год., практичні заняття 36 год., самостійна робота 66 год.</i>
Семестровий контроль/ контрольні заходи	<i>Залік, МКР</i>
Розклад занять	<i>rozklad.kpi.ua 1 год. лекційних занять та 2 год. комп'ютерних практикумів на тиждень</i>
Мова викладання	<i>Українська</i>
Інформація про керівника курсу / викладачів	<i>Лектор: к.ф.-м.н., доцент, Пишнограєв Іван Олександрович, pyshnograiev@wdc.org.ua Комп. практикуми: к.ф.-м.н., доцент, Пишнограєв Іван Олександрович</i>
Розміщення курсу	<i>Google classroom https://classroom.google.com/c/NzA3MzMzMzY5NjE5?cjc=rzzvmuh</i>

Програма навчальної дисципліни

1. Опис навчальної дисципліни, її мета, предмет вивчення та результати навчання

Дисципліна є нормативною в освітній програмі. Вивчення навчальної дисципліни націлено на формування, розвиток та закріплення у здобувачів таких загальних та фахових компетентностей:

ЗК 01 Здатність до абстрактного мислення, аналізу та синтезу,

ФК 02 Здатність виявляти, діагностувати та інтерпретувати соціальні проблеми українського суспільства та світової спільноти,

ФК 04 Здатність збирати та аналізувати емпіричні дані з використанням сучасних методів соціологічних досліджень,

ФК 11 Здатність здійснювати аналіз відкритих джерел даних (OSINT), аналізувати якісну інформацію, текстові дані, використовувати інтелектуальний аналіз для соціальних даних,

ФК 12 Здатність застосовувати сучасні статистичні методи, моделі, цифрові технології, спеціалізоване програмне забезпечення для моделювання соціальних процесів.

Внаслідок вивчення курсу студент повинен бути здатний продемонструвати такий програмний результат навчання ОНП:

ПРН 04 Застосовувати наукові знання, соціологічні та статистичні методи, цифрові технології, спеціалізоване програмне забезпечення для розв'язування складних задач соціології та суміжних галузей знань,

ПРН 05 Здійснювати пошук, аналізувати та оцінювати необхідну інформацію в науковій літературі, банках даних та інших джерелах,

ПРН 12 Здійснювати аналіз відкритих джерел даних (OSINT), аналізувати якісну інформацію, текстові дані, використовувати інтелектуальний аналіз для соціальних даних,

ПРН 14 Застосовувати мови програмування R та Python для аналізу соціальних даних.

У кінці вивчення курсу студент повинен **знати**:

- особливості роботи з великими даними;
- методи обробки та аналізу великих даних;
- особливості застосування штучного інтелекту при аналізі соціальних даних;
- методи створення і способи застосування моделей машинного навчання для аналізу даних.

вміти:

- аналізувати великі дані за допомогою мов програмування R та Python;
- створювати моделі для аналізу соціальних даних.

2. Пререквізити та постреквізити дисципліни (місце в структурно-логічній схемі навчання за відповідною освітньою програмою)

Дисципліна базується на знаннях та навичках суміжних дисциплін, що вивчаються на попередньому році та освітньому рівні. Дана дисципліна передує ОК ПО 12 «Виконання магістерської дисертації» та може являтися однією з головних її складових.

3. Зміст навчальної дисципліни:

Розділ 1. Вступ до аналізу великих даних та штучного інтелекту

Тема 1.1. Основні поняття великих даних.

1. Характеристика та відмінності великих даних;
2. Використання великих даних в Data Science;
3. Базові інструменти для роботи з великими даними;
4. Обмеження використання великих даних.

Тема 1.2. Основні поняття штучного інтелекту.

1. Основні визначення та сфери застосування;
2. Етичні проблеми при використанні штучного інтелекту;
3. Базові інструменти для роботи з моделями штучного інтелекту.

Розділ 2. Використання великих даних та штучного інтелекту в соціальних дослідженнях

Тема 2.1. Обчислювальні методи в соціальних дослідженнях.

1. Основні задачі і виклики;
2. Приклади застосування машинного навчання та штучного інтелекту в соціальних дослідженнях.

Тема 2.2. Аналіз соціальних мереж.

1. Постановка задачі та формалізація мережі;
2. Методи та інструменти аналізу соціальних мереж;
3. Аналіз та прогнозування розвитку соціальних мереж.

Тема 2.3. Агентне моделювання в соціальних дослідженнях.

1. Постановка задачі та основні визначення;
2. Моделювання поведінки замкнутого суспільства;
3. Побудова та дослідження штучного суспільства.

Тема 2.4. Етапи створення проєкту під час виконання кількісного соціального дослідження.

Перелік комп'ютерних практикумів:

1. Розв'язання базових задач з використанням великих даних та штучного інтелекту.
2. Аналіз соціальної мережі для вирішення поставленої проблеми.
3. Моделювання поведінки штучного суспільства.
4. Проведення соціального дослідження з використанням штучного інтелекту.

4. Навчальні матеріали та ресурси

Базова:

1. Zgurovsky, M., . Zaychenko, Y. (2020). *Big Data: Conceptual Analysis and Applications*. Cham, Switzerland: Springer. <https://link.springer.com/book/10.1007/978-3-030-14298-8>
2. Олещенко, Л. М. Технології оброблення великих даних. Конспект лекцій [Електронний ресурс] : навчальний посібник / Л. М. Олещенко ; КПІ ім. Ігоря Сікорського. – Електронні текстові дані (1 файл: 5,55 Мбайт). – Київ : КПІ ім. Ігоря Сікорського, 2021. – 227 с. <https://ela.kpi.ua/handle/123456789/42206>
3. SpringerLink (Online service), Zomaya, A. Y., & Sakr, S. (2017). *Handbook of Big Data Technologies* (1st ed. 2017.). Cham: Springer International Publishing. <https://link.springer.com/book/10.1007/978-3-319-49340-4>
4. Методи аналізу «великих даних» : методичні рекомендації з курсу «Прикладні задачі аналізу великих даних» / уклад. Н. М. Кізілова. – Харків : ХНУ імені В. Н. Каразіна, 2023. – 92 с. <https://ekhnuir.karazin.ua/items/6b5998f2-7e48-4dfb-8fa7-99aa3a03a150>
5. Методи та системи штучного інтелекту: Навчальний посібник для студентів/Уклад. : А.С. Савченко, О. О. Синельников. – К. : НАУ, 2017. – 190 с. https://er.nau.edu.ua/bitstream/NAU/40676/1/Методи%20та%20системи%20штучного%20інтелекту%20_Навч_посібн.pdf

Допоміжна:

6. R програмування // Електронний ресурс. Режим доступу: <https://coderlessons.com/tutorials/mashinnoe-obuchenie/r-programmirovaniye/r-programmirovaniye>
7. Virtualization Technology // Електронний ресурс. Режим доступу: <https://www.sciencedirect.com/topics/computer-science/virtualization-technology>
8. Apache Hadoop // Електронний ресурс. Режим доступу: <http://hadoop.apache.org/>
9. Stuart Russel, Peter Norvig. *Artificial Intelligence: A Modern Approach 4rd Edition*, 2020, 1408 p.

Навчальний контент

5. Методика опанування навчальної дисципліни (освітнього компонента)

Лекційні заняття

Розділ 1. Вступ до аналізу великих даних та штучного інтелекту

Лекція 1. Феномен великих даних: характеристики та роль у сучасному світі

Визначення поняття Big Data. Модель «5V» (Volume, Velocity, Variety, Veracity, Value) та її розширення. Відмінність великих даних від традиційних статистичних даних. Джерела генерації даних: соціальні мережі, IoT, транзакційні системи. Обмеження використання великих даних: проблема репрезентативності, шум, доступність. Роль Data Science як науки про методи обробки та аналізу даних.

Завдання на СРС:

1. Знайти та проаналізувати кейс успішного використання Big Data в комерційному або державному секторі.
2. Скласти порівняльну таблицю «Традиційна статистика vs Big Data» за критеріями: обсяг, метод збору, структурованість, генеральна сукупність.

Лекція 2. Екосистема та інструментарій технологій Big Data

Огляд архітектури для роботи з великими даними. Поняття розподілених обчислень. Базові інструменти та фреймворки (Hadoop, Spark – оглядово). Типи сховищ даних: Data Warehouse та Data Lake. Хмарні рішення для Big Data. Мови програмування для аналізу даних (R, Python): переваги та недоліки в контексті великих масивів. Візуалізація як інструмент інтерпретації великих даних.

Завдання на СРС:

1. Ознайомитися з документацією Apache Spark або екосистеми Hadoop, визначити основні компоненти.
2. Встановити середовище розробки (RStudio або Jupyter Notebook) та завантажити тестовий датасет для перевірки налаштувань.

Лекція 3. Штучний інтелект: концепції, історія та етика

Визначення штучного інтелекту (AI). Класифікація ШІ: слабкий (Narrow AI), сильний (General AI) та штучний суперінтелект. Основні сфери застосування. Етичні виклики: упередженість алгоритмів (bias), прозорість прийняття рішень («black box»), приватність даних, авторське право. Регулювання ШІ (EU AI Act та інші ініціативи).

Завдання на СРС:

1. Проаналізувати етичну дилему на прикладі реального скандалу (наприклад, Cambridge Analytica або дискримінація в алгоритмах найму).
2. Ознайомитися з основними положеннями концепції «Trustworthy AI».

Лекція 4. Моделі та інструменти штучного інтелекту

Огляд підходів до створення ШІ: логічний підхід, еволюційні методи, нейронні мережі. Поняття глибокого навчання (Deep Learning). Великі мовні моделі (LLM) та генеративний ШІ: принципи роботи та промпт-інжиніринг. Бібліотеки для роботи з моделями (TensorFlow, PyTorch, scikit-learn – оглядово). Використання API для інтеграції ШІ-рішень.

Завдання на СРС:

1. Зареєструватися на платформі Hugging Face або OpenAI, протестувати роботу однієї з відкритих моделей.
2. Скласти перелік задач у власній науковій діяльності, які можна автоматизувати за допомогою існуючих інструментів ШІ.

Розділ 2. Використання великих даних та штучного інтелекту в соціальних дослідженнях

Лекція 5. Обчислювальні методи в соціальних дослідженнях (Computational Social Science)

Парадигма обчислювальних соціальних наук. Основні задачі: класифікація, кластеризація, регресія в контексті соціальних даних. Машинне навчання (Machine Learning) в соціології: навчання з учителем та без учителя. Аналіз текстів (NLP): тональність, тематичне моделювання. Виклики валідності даних отриманих цифровим шляхом.

Завдання на СРС:

1. Знайти наукову статтю, де застосовується метод NLP (обробка природної мови) для аналізу соціальних настроїв.
2. Описати різницю між навчанням з учителем та без учителя на прикладі соціологічної задачі (наприклад, сегментація електорату).

Лекція 6. Аналіз соціальних мереж (SNA): формалізація та метрики

Теорія графів як основа SNA. Основні елементи мережі: вузли, ребра, спрямованість, вага зв'язків. Матриці суміжності. Рівні аналізу: рівень вузла, рівень групи, рівень мережі. Базові метрики центральності (Centrality measures): ступінь (Degree), близькість (Closeness), посередництво (Betweenness). Візуалізація графів.

Завдання на СРС:

1. Побудувати его-мережу власного оточення (схематично або за допомогою ПЗ) та виділити ключових акторів.

Лекція 7. Поглиблений аналіз та прогнозування соціальних мереж

Виявлення спільнот (Community detection): алгоритми та інтерпретація кластерів. Феномен «тісного світу» та безосштабні мережі. Поширення інформації та дифузія інновацій у мережах. Прогнозування зв'язків (Link Prediction). Інструменти для SNA (Gephi, igraph, tidygraph).

Завдання на СРС:

1. Завантажити та встановити Gephi (або відповідний пакет в R), ознайомитись з інтерфейсом.
2. Знайти приклад дослідження поширення дезінформації або вірусного контенту через методи мережевого аналізу.

Лекція 8. Агентне моделювання (ABM) соціальних процесів

Поняття агента та середовища. Принципи агентного моделювання: автономність, взаємодія, обмежена раціональність. Від мікро-поведінки до макро-феноменів (емерджентність). Класичні моделі: модель сегрегації Шелінга, модель «Хижак-Жертва» в соціальному контексті. Інструментарій ABM (NetLogo, спеціалізовані бібліотеки).

Завдання на СРС:

1. Запустити симуляцію моделі сегрегації Шелінга, змінюючи параметри толерантності, та описати зміни в результатах.
2. Запропонувати ідею для агентної моделі, що описує поведінку натовпу або поширення чуток.

Лекція 9. Життєвий цикл проекту кількісного соціального дослідження з використанням Big Data

Етапи Data Science проекту: від постановки гіпотези до впровадження. ETL-процеси (Extract, Transform, Load) в соціальних дослідженнях. Очищення даних (Data Wrangling) та робота з пропусками. Побудова аналітичних дашбордів (Shiny, Tableau, PowerBI) для презентації результатів стейкхолдерам. Інтерпретація результатів машинного навчання для соціальних наук.

Завдання на СРС:

1. Розробити концептуальну схему (pipeline) власного дослідницького проекту: визначити джерело даних, методи обробки та очікуваний результат.
2. Ознайомитися з прикладами інтерактивних дашбордів, що відображають соціальні або демографічні показники.

Практичні заняття (комп'ютерні практикуми)

Комп'ютерний практикум 1. Розв'язання базових задач з використанням великих даних та штучного інтелекту. (6 год.)

Мета: Навчитись налаштовувати середовище для роботи з даними, виконувати первинну та застосовувати базові алгоритми машинного навчання.

План

1. Підготовка середовища та даних.
2. Попередня обробка даних (очищення, розрахунки описової, очищення пропусків).
3. Застосування методів ШІ для розв'язання задачі.
4. Розрахунок базових метрик точності. Інтерпретація отриманих результатів, коефіцієнтів або важливості ознак.

Комп'ютерний практикум 2. Аналіз соціальної мережі для вирішення поставленої проблеми. (8 год.)

Мета: Опанувати методи мережевого аналізу (SNA) для виявлення впливових агентів та спільнот у соціальній структурі.

План

1. Формалізація задачі, пошук і завантаження даних.
2. Побудова мережі та розрахунок спеціалізованих метрик.
3. Виявлення «спільнот» для розв'язання задачі.
4. Побудова візуалізації мережі з розфарбуванням вузлів та розміром залежно від поставленої задачі.

Комп'ютерний практикум 3. Моделювання поведінки штучного суспільства. (8 год.)

Мета: Створити та дослідити агентну модель для розуміння емерджентних соціальних явищ.

План

1. Постановка задачі. Визначення агентів та їх властивостей.
2. Ініціалізація моделі з випадковим розподілом агентів, спостереження за динамікою.
3. Проведення експерименту, аналіз наслідків змін властивостей агентів.
4. Аналіз отриманих кінцевих станів.

Комп'ютерний практикум 4. Проведення соціального дослідження з використанням штучного інтелекту. (10 год.)

Мета: Реалізувати повний цикл міні-дослідження з використанням методів ШІ для аналізу соціальних процесів.

План

1. Постановка задачі. Збір неструктурованих даних (текстів) та структурованих даних.
2. Data Mining. Побудова концептуальної моделі, вибір інструментів для проведення дослідження.
3. Проведення дослідження з використанням моделей машинного навчання, LLM, агентного моделювання тощо.
4. Формулювання соціологічних висновків на основі отриманих закономірностей.

Модульна контрольна робота (2 год)

Залік (2 год)

6. Самостійна робота студента

Самостійна робота здобувача включає:

- підготовка до аудиторних занять – 56 год;
- підготовка до модульної контрольної роботи – 4 год.
- підготовка до заліку - 6 год.

Загалом – 66 год.

Політика та контроль

7. Політика навчальної дисципліни (освітнього компонента)

Усі роботи студенти мають прикріплювати в особистому кабінеті гугл-класу. Дедлайни кожного завдання позначені в завданнях у гугл-класі. Роботи мають бути виконані з дотриманням академічної доброчесності. Політика та принципи академічної доброчесності, етична поведінка студентів визначені у Кодексі честі <https://kpi.ua/code>. Лектор може запропонувати студентам пройти онлайн-курси на платформі Coursera. Також сертифікати цих курсів можуть бути частково зараховані згідно до Положення.

Тематика робіт спрямована на поглиблення засвоєного матеріалу лекцій. На заняттях комп'ютерного практикуму розв'язуються задачі та вправи по темах лекцій.

8. Види контролю та рейтингова система оцінювання результатів навчання (PCO)

Семестровий контроль: **залік**.

Семестровий рейтинг студента з дисципліни виставляється лектором та складається з балів, що він отримує за:

- ~ виконання модульної контрольної;
- ~ виконання 4 комп'ютерних практикумів.

Критерії нарахування балів за семестр:

- 1) Модульна контрольна робота оцінюється у 20 балів.
- 2) Кожен з практикумів оцінюється в 20 балів.

Критерії нарахування балів за контрольні заходи:

- "відмінно": 95 -100% - здобувач виявив всебічні, систематичні та глибокі знання навчального матеріалу з дисципліни; продемонстрував уміння вільно виконувати всі завдання, передбачені програмою; засвоїв основну та додаткову літературу; проявив творчі здібності в розумінні, в логічному, чіткому, стислому та ясному трактуванні навчального матеріалу; засвоїв взаємозв'язок основних понять дисципліни, їх значення для подальшої професійної діяльності
- "дуже добре": 85-94% - здобувач виявив систематичні знання навчального матеріалу з дисципліни вище середнього рівня; продемонстрував уміння добре виконувати всі завдання, передбачені програмою, допустивши незначні помилки; засвоїв основну та додаткову літературу; засвоїв взаємозв'язок основних понять дисципліни, їх значення для подальшої професійної діяльності
- "добре": 75-84% - здобувач виявив загалом добрі знання навчального матеріалу при виконанні передбачених програмою завдань, але припустив ряд помітних помилок; засвоїв основну літературу; показав систематичний характер знань з дисципліни; здатний до їх самостійного використання та поповнення в процесі подальшої навчальної роботи і професійної діяльності
- "задовільно": 65-74% - здобувач виявив знання основного навчального матеріалу з дисципліни в обсязі, необхідному для подальшого навчання та майбутньої професійної діяльності; ознайомився з основною літературою; впорався з виконанням завдань, передбачених програмою, але припустив значну кількість помилок або недоліків на запитання при співбесіді, тестуванні та при виконанні завдань тощо, принципів з яких може усунути самостійно
- "достатньо": 60-64% - здобувач виявив знання основного навчального матеріалу з дисципліни в мінімальному обсязі, необхідному для подальшого навчання та майбутньої професійної діяльності; ; ознайомився з основною літературою; в основному виконав завдання, передбачені програмою, але припустив помилки у відповіді на запитання при співбесідах,

тестуванні та при виконанні завдань тощо, які він може усунути лише під керівництвом та за допомогою викладача

- “незадовільно”: 30-54% - здобувач мав значні прогалини в знаннях основного навчального матеріалу; допускав принципові помилки при виконанні передбачених програмою завдань, але спроможний самостійно доопрацювати програмний матеріал і підготуватися для перездачі дисципліни
- “незадовільно”: 0-29% - здобувач не мав знань зі значної частини навчального матеріалу з дисципліни; допускав принципові помилки при виконанні більшості передбачених програмою завдань або не виконував ці завдання

Умовою першої атестації є поточний рейтинг не менше 30% запланованих балів за семестр. Умова другої атестації – поточний рейтинг не менше 50% запланованих балів.

Необхідною умовою допуску до заліку є зарахування всіх комп’ютерних практикумів та 60 балів семестрового рейтингу. В цьому випадку залік виставляється за результатом семестрового рейтингу. Якщо здобувач не набрав 60 балів або хоче покращити свою оцінку, можливо написати залікову роботу. На залікову роботу виноситься 100 балів. Залік проводиться у вигляді письмової роботи, в якій три теоретичних питання та два практичних. Кожне завдання оцінюється в 20 балів за такими критеріями:

- «відмінно», повна відповідь, не менше 90% потрібної інформації, що виконана згідно з вимогами до рівня «умінь», (повне, безпомилкове розв’язування завдання) – 18-20 балів;
- «добре», достатньо повна відповідь, не менше 75% потрібної інформації, що виконана згідно з вимогами до рівня «умінь» або є незначні неточності (повне розв’язування завдання з незначними неточностями) – 14-16 балів;
- «задовільно», неповна відповідь, не менше 60% потрібної інформації, що виконана згідно з вимогами до «стереотипного» рівня та деякі помилки (завдання виконане з певними недоліками) – 12 балів;
- «незадовільно», відповідь не відповідає умовам до «задовільно» – 0 балів.

Сума рейтингових балів, отриманих здобувачем протягом семестру, переводиться до підсумкової оцінки згідно з таблицею.

Таблиця відповідності рейтингових балів оцінкам за університетською шкалою:

Бали:	Оцінка
100...95	Відмінно
94...85	Дуже добре
84...75	Добре
74...65	Задовільно
64...60	Достатньо
Менше 60	Незадовільно
Невиконання умов допуску до семестрового контролю	Не допущено

Робочу програму навчальної дисципліни (силабус):

Складено доцент, к.ф.-м.н., доцент Пишнограєв Іван Олександрович



Ухвалено кафедрою ШІ (протокол № 14 від 11.05.2024)

Погоджено Методичною комісією НН ІПСА (протокол № 10 від 24.06.2024)