

ПОЛТАВСЬКИЙ УНІВЕРСИТЕТ ЕКОНОМІКИ І ТОРГІВЛІ
Навчально-науковий інститут денної освіти
Кафедра комп'ютерних наук та інформаційних технологій

ЗАТВЕРДЖУЮ

Завідувач кафедри



Олена ОЛЬХОВСЬКА

« 28 » 06 2024 р.

РОБОЧА ПРОГРАМА

навчальної дисципліни
освітня програма
спеціальність
галузь знань
ступінь вищої освіти

«Машинне навчання»
Комп'ютерні науки
122 Комп'ютерні науки
12 Інформаційні технології
магістр

Робоча програма навчальної дисципліни «Машинне навчання» рекомендована до використання в освітньому процесі на засіданні кафедри комп'ютерних наук та інформаційних технологій
Протокол № 13 від 28.06.2024 року.

Полтава 2024

Укладачі:

Кошова Оксана Петрівна, доцент кафедри комп'ютерних наук та інформаційних технологій, к.пед.н.

ПОГОДЖЕНО:

Гарант освітньої програми «Комп'ютерні науки»

спеціальності 122 Комп'ютерні науки
ступеня магістра, к.ф.-м.н, доцент



Олена ОЛЬХОВСЬКА

« 28 » __ 06 ____ 2024 року

Розділ 1. Опис навчальної дисципліни

Таблиця 1 - Опис навчальної дисципліни «Машинне навчання»

Місце у структурно-логічній схемі підготовки	<i>Пререквізити:</i> Спеціальні (фахові) компетентності зі спеціальності 122 Комп'ютерні науки галузі знань 12 Інформаційні технології для першого (бакалаврського) рівня вищої освіти. Отримані навички з ОК Розробка та аналіз алгоритмів, Моделювання об'єктів та процесів, Хмарні технології та великі дані, Технології обробки та аналізу даних. <i>Постреквізити:</i> Кваліфікаційна робота	
Мова викладання	Українська	
Статус дисципліни	Обов'язкова	
Курс/семестр вивчення	2/3	
Кількість кредитів ECTS/кількість модулів	3/2	
Денна форма навчання:		
Кількість годин: 90 год – загальна кількість: 1 семестр – 90 год.		
- Лекції: 16 год.		
- Практичні (семінарські, лабораторні) заняття: 20 год.		
- Самостійна робота: 54 год.		
- Вид підсумкового контролю (ПМК, екзамен): 1 семестр - екзамен		
Заочна форма навчання:		
Кількість годин: 90 год – загальна кількість: 1 семестр – 90 год.		
- Лекції: 6 год.		
- Практичні (семінарські, лабораторні) заняття: 2 год.		
- Самостійна робота: 82 год.		
- Вид підсумкового контролю (ПМК, екзамен): 1 семестр - екзамен		

Розділ 2. Перелік компетентностей, які забезпечує дана навчальна дисципліна, програмні результати навчання

Метою навчальної дисципліни «Машинне навчання» є формування у студентів вміння до ефективного застосування методів машинного навчання та оволодіння практичними навичками вирішення прикладних завдань інтелектуального аналізу даних.

Методи навчання:

Словесні – використання проблемних лекцій, лекцій-дискусій, мозкового штурму, обговорення проблемних питань лекції у форматі «The World Cafe», методу проєктів, роботи в малих групах та ін.

Наочні – візуалізація алгоритмів, роботи нейронних мереж, використання інструментів для обробки і візуалізації даних типу matplotlib, генеративних AI, Sentiment Analysis Charts та ін.

Практичні – використання методу проєктів під час виконання практичних завдань та самостійної роботи; відвідування майстер класів запрошених лекторів та фахівців із IT компаній; система дистанційного навчання ПУЕТ; сучасне програмне забезпечення типу Tesseract OCR; віртуальні лабораторії Jupyter Notebook, Google Colab; інструмент AWS CLI (Command Line Interface) та ін.

Таблиця 2 - Перелік компетентностей та програмні результати навчання, які забезпечує навчальна дисципліна «Машинне навчання»

<i>Програмні результати навчання</i>	<i>Компетентності, якими повинен оволодіти здобувач</i>
<p>РН2. Мати спеціалізовані уміння/навички розв'язання проблем комп'ютерних наук, необхідні для проведення досліджень та/або провадження інноваційної діяльності з метою розвитку нових знань та процедур.</p> <p>РН7. Розробляти та застосовувати математичні методи для аналізу інформаційних моделей.</p> <p>РН8. Розробляти математичні моделі та методи аналізу даних (включно з великим).</p> <p>РН9. Розробляти алгоритмічне та програмне забезпечення для аналізу даних (включно з великими).</p> <p>РН11. Створювати нові алгоритми розв'язування задач у сфері комп'ютерних наук, оцінювати їх ефективність та обмеження на їх застосування</p> <p>РН16. Виконувати дослідження у сфері комп'ютерних наук.</p>	<p>ЗК01. Здатність до абстрактного мислення, аналізу та синтезу.</p> <p>ЗК02. Здатність застосовувати знання у практичних ситуаціях.</p> <p>ЗК03. Здатність спілкуватися державною мовою як усно, так і письмово.</p> <p>ЗК05. Здатність вчитися й оволодівати сучасними знаннями.</p> <p>ЗК07. Здатність генерувати нові ідеї (креативність).</p> <p>СК1. Усвідомлення теоретичних засад комп'ютерних наук</p> <p>СК03. Здатність використовувати математичні методи для аналізу формалізованих моделей предметної області.</p> <p>СК6. Здатність застосовувати існуючі і розробляти нові алгоритми розв'язування задач у галузі комп'ютерних наук.</p> <p>СК7. Здатність розробляти програмне забезпечення відповідно до сформульованих вимог з урахуванням наявних ресурсів та обмежень.</p> <p>СК11. Здатність ініціювати, планувати та реалізовувати процеси розробки інформаційних та комп'ютерних систем та програмного забезпечення, включно з його розробкою, аналізом, тестуванням, системною інтеграцією, впровадженням і супроводом.</p>

Розділ 3. Програма навчальної дисципліни

Модуль 1: Обробка природної мови та тексту

Тема 1: Машинне навчання в NLP. Задачі NLP

Машинне навчання та його алгоритми. Використання математичних моделей та методів аналізу інформаційних моделей, даних та Big Data. Особливості створення алгоритмів для вирішення задач машинного навчання та сфери ІТ і, зокрема, для обробки природної мови. Детальний огляд задач обробки природної мови, таких як класифікація тексту, розпізнавання іменованих сутностей (NER), переклад тексту та аналіз настроїв. Розгляд концепцій та визначень, а також аналіз найпоширеніших проблем і викликів, з якими стикаються дослідники та розробники. Використання Amazon Comprehend для аналізу тексту і настроїв, а також Google Cloud Natural Language API для виконання аналогічних задач, включаючи автоматизований переклад з Google Translate.

Тема 2: Етапи вирішення NLP завдань. Еволюція архітектур

Розбір послідовних етапів вирішення завдань NLP, включаючи збір та підготовку

даних, вибір і налаштування моделей, їх навчання та оцінювання. Аналіз розвитку архітектур NLP від класичних статистичних методів до сучасних підходів, таких як глибоке навчання та трансформери. Застосування AWS SageMaker та Google AI Platform для розгортання моделей та їх навчання.

Тема 3: Методи обробки тексту. Збір текстових даних

Розгляд методів обробки тексту, включаючи токенізацію, лематизацію, видалення стоп-слів та інші методи очищення даних. Огляд технологій збору текстових даних з різних джерел, таких як веб-скрейпінг та оптичне розпізнавання символів (OCR), з використанням Amazon Textract для автоматизованого витягування тексту з документів, Tesseract як відкритого інструменту для OCR, та Google Cloud Storage для зберігання текстових даних.

Тема 4: Попередня обробка та векторизація тексту

Дослідження методів нормалізації тексту та видалення шуму для підготовки даних до подальшого аналізу. Обговорення процесу векторизації тексту, що включає перетворення тексту у числові вектори для обробки моделями машинного навчання. Розгляд технік, таких як TF-IDF, word embeddings та використання бібліотек, таких як spaCy і NLTK для попередньої обробки тексту.

Модуль 2: Практичне застосування машинного навчання та генеративний AI

Тема 5: Алгоритми аналізу настроїв та витяг інформації

Аналіз алгоритмів для визначення емоційної тональності тексту та методів витягу інформації з текстових даних, включаючи ідентифікацію імен, дат, місць та інших важливих елементів. Використання Amazon Comprehend, а також spaCy для обробки іменованих сутностей та аналізу настроїв. Використання Sentiment Analysis Charts.

Тема 6: Тематичне моделювання та багатомовні задачі

Особливості розробки та застосування математичних методів та моделей для аналізу інформації. Огляд методів тематичного моделювання для виявлення основних тем у великих текстових корпусах, зокрема латентний семантичний аналіз (LSA) та латентний Діріхле розподіл (LDA). Обговорення проблем та викликів при роботі з багатомовними текстами, зокрема виявлення мов та автоматизований переклад. Використання Amazon Comprehend, Google Cloud Translation для багатомовних задач та Gensim для тематичного моделювання. Особливості створення нових алгоритмів, оцінка їх ефективності та обмеження на їх застосування.

Тема 7: Генеративний AI та великі мовні моделі (LLMs)

Аналіз концепцій генеративного AI та великих мовних моделей (LLMs), як об'єктів досліджень у сфері комп'ютерних наук, таких як GPT та BERT, із детальним розглядом архітектури трансформерів та параметрів конфігурації. Обговорення практичних прикладів інтеграції LLMs у додатки для автоматизації процесів, генерації тексту та інших завдань, з використанням AWS SageMaker, Hugging Face Transformers та OpenAI API.

Тема 8. Використання алгоритмів машинного навчання у імітаційному моделюванні та прийнятті рішень.

Методи машинного навчання у задачах інтелектуального аналізу даних та прийнятті рішень. Використання машинного навчання у DDDM (прийняття рішень на основі даних). Приклади використання алгоритмів машинного навчання у різних сферах бізнесу, логістиці, імітаційному моделюванні роботи техніки та моделюванні біологічних систем. Особливості роботи із Big data для аналізу ефективності роботи бізнес компаній. Особливості розробки алгоритмічного та програмного забезпечення для аналізу даних (включно з великими) та його використання для підтримки бізнес сектору у сфері логістики, фінансів, робототехніки та ін.

Розділ 4. Тематичний план навчальної дисципліни

Таблиця 3. Тематичний план навчальної дисципліни для студентів денної форми навчання

1	2	3	4	5	6
Назва теми (лекції) та питання теми (лекції)	К-ть год	Назва теми та питання семінарського, практичного або лабораторного заняття	К - т ь г о д	Завдання самостійної роботи в розрізі тем	К-ть год
Модуль 1: Обробка природної мови та тексту					
Тема 1: Машинне навчання в NLP. Задачі NLP Проблемна лекція. Машинне навчання та його алгоритми. Використання математичних моделей та методів аналізу інформаційних моделей, даних та Big Data. Особливості створення алгоритмів для вирішення задач машинного навчання та сфери IT і, зокрема, для обробки природної мови. Детальний огляд задач обробки природної мови, таких як класифікація тексту, розпізнавання іменованих сутностей (NER), переклад тексту та аналіз настроїв. Розгляд концепцій та визначень, а також аналіз найпоширеніших проблем і викликів, з якими стикаються дослідники та розробники. Використання Amazon Comprehend для аналізу тексту і настроїв, а також Google Cloud Natural Language API для виконання аналогічних задач, включаючи автоматизований переклад з Google Translate.	2	Практичне заняття 1. Розгляд концепцій та визначень, а також аналіз найпоширеніших проблем і викликів, з якими стикаються дослідники та розробники. Використання Amazon Comprehend для аналізу тексту і настроїв, а також Google Cloud Natural Language API для виконання аналогічних задач, включаючи автоматизований переклад з Google Translate. Робота у малих групах	2	Розпізнавання іменованих сутностей (NER): методи розпізнавання (CRF, LSTM, BERT) Аналіз настроїв: методи та алгоритми (лексичні підходи, нейронні мережі)	6
Тема 2: Етапи вирішення NLP завдань. Еволюція архітектур Лекція дискусія. Розбір послідовних етапів вирішення завдань NLP, включаючи збір та підготовку даних, вибір і налаштування моделей, їх навчання та оцінювання. Аналіз розвитку архітектур NLP від класичних статистичних методів до сучасних підходів, таких як глибоке навчання та трансформери. Застосування AWS SageMaker та Google AI Platform для розгортання моделей та їх навчання.	2	Практичне заняття 2 Застосування ML до проблеми НЛП Застосування AWS SageMaker та Google AI Platform для розгортання моделей та їх навчання.	2	Збір та підготовка даних: методи збору даних для NLP; очищення та попередня обробка текстових даних (токенізація, стемінг, лемматизація, видалення стоп-слів) Вибір і налаштування моделей: критерії вибору моделей для різних задач NLP; налаштування гіперпараметрів моделей. Оцінювання моделей: метрики оцінювання якості моделей (точність, F1-міра, ROC-AUC)	6

1	2	3	4	5	6
Назва теми (лекції) та питання теми (лекції)	К-ть год	Назва теми та питання семінарського, практичного або лабораторного заняття	К - т ь г о д	Завдання самостійної роботи в розрізі тем	К-ть год
Тема 3: Методи обробки тексту. Збір текстових даних Розгляд методів обробки тексту, включаючи токєнізацію, лематизацію, видалення стоп-слів та інші методи очищення даних. Огляд технологій збору текстових даних з різних джерел, таких як веб-скрейпінг та оптичне розпізнавання символів (OCR), з використанням Amazon Textract для автоматизованого витягування тексту з документів, Tesseract як відкритого інструменту для OCR, та Google Cloud Storage для зберігання текстових даних.	2	Практичне заняття 3 Вилучення тексту з веб-сторінок і зображень із використанням Amazon Textract	2	Токєнізація: токєнізації тексту; інструменти та бібліотеки для токєнізації (NLTK, spaCy) Лематизація та стемінг: відмінності між лематизацією та стемінгом Веб-скрейпінг: інструменти для веб-скрейпінгу (BeautifulSoup, Scrapy) Оптичне розпізнавання символів (OCR) використання Tesseract для OCR	6
Тема 4: Попередня обробка та векторизація тексту Дослідження методів нормалізації тексту та видалення шуму для підготовки даних до подальшого аналізу. Обговорення процесу векторизації тексту, що включає перетворення тексту у числові вектори для обробки моделями машинного навчання. Розгляд технік, таких як TF-IDF, word embeddings та використання бібліотек, таких як spaCy і NLTK для попередньої обробки тексту	2	Практичне заняття 4 Процес векторизації тексту, що включає перетворення тексту у числові вектори для обробки моделями машинного навчання. Опрацювання тексту із Natural Language Toolkit (NLTK) library Кодування та векторизація тексту із scikit-learn library Робота у малих групах Практичне заняття 5 МКР1	2 2	Нормалізація тексту: методи нормалізації тексту ; інструменти для нормалізації тексту (NLTK, spaCy) Видалення шуму: Визначення та видалення шуму з текстових даних (HTML-теги, спеціальні символи); Використання регулярних виразів для видалення шуму	6
Модуль 2: Практичне застосування машинного навчання та генеративний AI.					
Тема 5: Алгоритми аналізу настроїв та витяг інформації Аналіз алгоритмів для визначення емоційної тональності тексту та методів витягу інформації з текстових даних, включаючи ідентифікацію імен, дат, місць та інших важливих елементів. Використання Amazon Comprehend, а також spaCy для обробки іменованих сутностей та аналізу настроїв. Використання Sentiment Analysis Charts	2	Практичне заняття 6-7. Впровадження аналізу настроїв. Використання Amazon Comprehend, а також spaCy для обробки іменованих сутностей та аналізу настроїв. Впровадження вилучення інформації із використання Amazon Comprehend, а також spaCy. Робота з сутностями	4	Лексичні методи аналізу настроїв: Методи машинного навчання для аналізу настроїв. Глибоке навчання для аналізу настроїв. Витяг іменованих сутностей (NER). Витяг ключових слів та фраз. Використання Amazon Comprehend.	6

1	2	3	4	5	6
Назва теми (лекції) та питання теми (лекції)	К-ть год	Назва теми та питання семінарського, практичного або лабораторного заняття	К - т ь г о д	Завдання самостійної роботи в розрізі тем	К-ть год
Тема 6: Тематичне моделювання та багатомовні задачі Проблемна лекція. Особливості розробки та застосування математичних методів та моделей для аналізу інформації. Огляд методів тематичного моделювання для виявлення основних тем у великих текстових корпусах, зокрема латентний семантичний аналіз (LSA) та латентний Діріхле розподіл (LDA). Обговорення проблем та викликів при роботі з багатомовними текстами, зокрема виявлення мов та автоматизований переклад. Використання Amazon Comprehend, Google Cloud Translation для багатомовних задач та Gensim для тематичного моделювання. Особливості створення нових алгоритмів, оцінка їх ефективності та обмеження на їх застосування	2	Практичне заняття 8. Впровадження тематичного моделювання за допомогою Amazon Comprehend, нейронної тематичної моделі (NTM) та LDA, впровадження тематичного моделювання. Використання Amazon Comprehend, Google Cloud Translation для багатомовних задач та Gensim для тематичного моделювання. Впровадження багатомовного рішення	2	опрацьовують матеріал лекцій; готуються до практичних завдань відповідно до теми заняття; виконують домашні роботи; працюють із літературою; самостійно вивчають: Алгоритми локалізації та картування: Kalman Filter, Particle Filter, Graph-based SLAM. Фільтри Калмана та комплементарні фільтри. Алгоритми обробки даних LIDAR та камер	6

1	2	3	4	5	6
Назва теми (лекції) та питання теми (лекції)	К-ть год	Назва теми та питання семінарського, практичного або лабораторного заняття	К-ть год	Завдання самостійної роботи в розрізі тем	К-ть год
<p>Тема 7: Генеративний AI та великі мовні моделі (LLMs) Аналіз концепцій генеративного AI та великих мовних моделей (LLMs), як об'єктів досліджень у сфері комп'ютерних наук, таких як GPT та BERT, із детальним розглядом архітектури трансформерів та параметрів конфігурації. Обговорення практичних прикладів інтеграції LLMs у додатки для автоматизації процесів, генерації тексту та інших завдань, з використанням AWS SageMaker, Hugging Face Transformers та OpenAI API.</p> <p>Тема 8. Використання алгоритмів машинного навчання у імітаційному моделюванні та прийнятті рішень. Обговорення у форматі світового кафе. Методи машинного навчання у задачах інтелектуального аналізу даних та прийнятті рішень. Використання машинного навчання у DDDM (прийняття рішень на основі даних). Приклади використання алгоритмів машинного навчання у різних сферах бізнесу, логістиці, імітаційному моделюванні роботи техніки та моделюванні біологічних систем. Особливості роботи із Big data для аналізу ефективності роботи бізнес компаній. Особливості розробки алгоритмічного та програмного забезпечення для аналізу даних (включно з великими) та його використання для підтримки бізнес сектору у сфері логістики, фінансів, робототехніки та ін.</p>	2 2	<p>Практичне заняття 9. Використання алгоритмів машинного навчання у прийнятті рішень. Робота над проектом у командах, обраних викладачем</p> <p>Практичне заняття 10 МКР2</p>	2 2	<p>Основи трансформерів: self-attention механізм, Multi-Head Attention. Детальний розгляд архітектури GPT та BERT. Огляд Multi-Head Attentionта її впливу на ефективність моделей.</p> <p>Використання машинного навчання у DDDM (прийняття рішень на основі даних). Використання алгоритмів машинного навчання у різних сферах бізнесу та моделюванні роботи техніки</p>	8 10
	16		20		54

Таблиця 4 - Тематичний план навчальної дисципліни для студентів заочної форми навчання

1	2	3	4	5	6
Назва теми (лекції) та питання теми (лекції)	К-ть год	Назва теми та питання семінарського, практичного або лабораторного заняття	К-ть год	Завдання самостійної роботи в розрізі тем	К-ть год
Модуль 1: Обробка природної мови та тексту					

1	2	3	4	5	6
Назва теми (лекції) та питання теми (лекції)	К-ть год	Назва теми та питання семінарського, практичного або лабораторного заняття	К-ть год	Завдання самостійної роботи в розрізі тем	К-ть год
<p>Тема 1: Машинне навчання в NLP. Задачі NLP</p> <p>Машинне навчання та його алгоритми. Використання математичних моделей та методів аналізу інформаційних моделей, даних та Big Data. Особливості створення алгоритмів для вирішення задач машинного навчання та сфери IT і, зокрема, для обробки природної мови. Детальний огляд задач обробки природної мови, таких як класифікація тексту, розпізнавання іменованих сутностей (NER), переклад тексту та аналіз настроїв. Розгляд концепцій та визначень, а також аналіз найпоширеніших проблем і викликів, з якими стикаються дослідники та розробники. Використання Amazon Comprehend для аналізу тексту і настроїв, а також Google Cloud Natural Language API для виконання аналогічних задач, включаючи автоматизований переклад з Google Translate.</p>	2	<p>Практичне заняття 1.</p> <p>Розгляд концепцій та визначень, а також аналіз найпоширеніших проблем і викликів, з якими стикаються дослідники та розробники.</p> <p>Використання Amazon Comprehend для аналізу тексту і настроїв, а також Google Cloud Natural Language API для виконання аналогічних задач, включаючи автоматизований переклад з Google Translate. Робота у малих групах</p>		<p>Розпізнавання іменованих сутностей (NER): методи розпізнавання (CRF, LSTM, BERT)</p> <p>Аналіз настроїв: методи та алгоритми (лексичні підходи, нейронні мережі)</p>	10
<p>Тема 2: Етапи вирішення NLP завдань. Еволюція архітектур</p> <p>Розбір послідовних етапів вирішення завдань NLP, включаючи збір та підготовку даних, вибір і налаштування моделей, їх навчання та оцінювання. Аналіз розвитку архітектур NLP від класичних статистичних методів до сучасних підходів, таких як глибоке навчання та трансформери. Застосування AWS SageMaker та Google AI Platform для розгортання моделей та їх навчання.</p>		<p>Практичне заняття 2</p> <p>Застосування ML до проблеми НЛП</p> <p>Застосування AWS SageMaker та Google AI Platform для розгортання моделей та їх навчання.</p>		<p>Збір та підготовка даних: методи збору даних для NLP; очищення та попередня обробка текстових даних (токенізація, стемінг, лемматизація, видалення стоп-слів)</p> <p>Вибір і налаштування моделей: критерії вибору моделей для різних задач NLP; налаштування гіперпараметрів моделей.</p> <p>Оцінювання моделей: метрики оцінювання якості моделей (точність, F1-міра, ROC-AUC)</p>	10

1	2	3	4	5	6
Назва теми (лекції) та питання теми (лекції)	К-ть год	Назва теми та питання семінарського, практичного або лабораторного заняття	К-ть год	Завдання самостійної роботи в розрізі тем	К-ть год
Тема 3: Методи обробки тексту. Збір текстових даних Розгляд методів обробки тексту, включаючи токенизацію, лематизацію, видалення стоп-слів та інші методи очищення даних. Огляд технологій збору текстових даних з різних джерел, таких як веб-скрейпінг та оптичне розпізнавання символів (OCR), з використанням Amazon Texttract для автоматизованого витягування тексту з документів, Tesseract як відкритого інструменту для OCR, та Google Cloud Storage для зберігання текстових даних.	2	Практичне заняття 3 Вилучення тексту з веб-сторінок і зображень із використанням Amazon Texttract	2	Токенизація: токенизації тексту; інструменти та бібліотеки для токенизації (NLTK, spaCy) Лематизація та стемінг: відмінності між лематизацією та стемінгом Веб-скрейпінг: інструменти для веб-скрейпінгу (BeautifulSoup, Scrapy) Оптичне розпізнавання символів (OCR) використання Tesseract для OCR	10
Тема 4: Попередня обробка та векторизація тексту Дослідження методів нормалізації тексту та видалення шуму для підготовки даних до подальшого аналізу. Обговорення процесу векторизації тексту, що включає перетворення тексту у числові вектори для обробки моделями машинного навчання. Розгляд технік, таких як TF-IDF, word embeddings та використання бібліотек, таких як spaCy і NLTK для попередньої обробки тексту		Практичне заняття 4 Процес векторизації тексту, що включає перетворення тексту у числові вектори для обробки моделями машинного навчання. Опрацювання тексту із Natural Language Toolkit (NLTK) library Кодування та векторизація тексту із scikit-learn library Робота у малих групах Практичне заняття 5 МКР1		Нормалізація тексту: методи нормалізації тексту ; інструменти для нормалізації тексту (NLTK, spaCy) Видалення шуму: Визначення та видалення шуму з текстових даних (HTML-теги, спеціальні символи); Використання регулярних виразів для видалення шуму	10
Модуль 2: Практичне застосування машинного навчання та генеративний AI.					
Тема 5: Алгоритми аналізу настроїв та витяг інформації Аналіз алгоритмів для визначення емоційної тональності тексту та методів витягу інформації з текстових даних, включаючи ідентифікацію імен, дат, місць та інших важливих елементів. Використання Amazon Comprehend, а також spaCy для обробки іменованих сутностей та аналізу настроїв.		Практичне заняття 6-7. Впровадження аналізу настроїв. Використання Amazon Comprehend, а також spaCy для обробки іменованих сутностей та аналізу настроїв. Впровадження вилучення інформації із використання Amazon Comprehend, а також spaCy. Робота з сутностями		Лексичні методи аналізу настроїв: Методи машинного навчання для аналізу настроїв. Глибоке навчання для аналізу настроїв. Витяг іменованих сутностей (NER). Витяг ключових слів та фраз. Використання Amazon Comprehend.	10

1	2	3	4	5	6
Назва теми (лекції) та питання теми (лекції)	К-ть год	Назва теми та питання семінарського, практичного або лабораторного заняття	К-ть год	Завдання самостійної роботи в розрізі тем	К-ть год
<p>Тема 6: Тематичне моделювання та багатомовні задачі</p> <p>Особливості розробки та застосування математичних методів та моделей для аналізу інформації. Огляд методів тематичного моделювання для виявлення основних тем у великих текстових корпусах, зокрема латентний семантичний аналіз (LSA) та латентний Діріхле розподіл (LDA). Обговорення проблем та викликів при роботі з багатомовними текстами, зокрема виявлення мов та автоматизований переклад. Використання Amazon Comprehend, Google Cloud Translation для багатомовних задач та Gensim для тематичного моделювання. Особливості створення нових алгоритмів, оцінка їх ефективності та обмеження на їх застосування</p>	2	<p>Практичне заняття 8.</p> <p>Впровадження тематичного моделювання за допомогою Amazon Comprehend, нейронної тематичної моделі (NTM) та LDA, впровадження тематичного моделювання.</p> <p>Використання Amazon Comprehend, Google Cloud Translation для багатомовних задач та Gensim для тематичного моделювання.</p> <p>Впровадження багатомовного рішення</p>		<p>опрацьовують матеріал лекцій; готуються до практичних завдань відповідно до теми заняття; виконують домашні роботи; працюють із літературою; самостійно вивчають: Алгоритми локалізації та картування: Kalman Filter, Particle Filter, Graph-based SLAM. Фільтри Калмана та комплементарні фільтри. Алгоритми обробки даних LIDAR та камер</p>	10

1	2	3	4	5	6
Назва теми (лекції) та питання теми (лекції)	К-ть год	Назва теми та питання семінарського, практичного або лабораторного заняття	К-ть год	Завдання самостійної роботи в розрізі тем	К-ть год
<p>Тема 7: Генеративний AI та великі мовні моделі (LLMs) Аналіз концепцій генеративного AI та великих мовних моделей (LLMs), як об'єктів досліджень у сфері комп'ютерних наук, таких як GPT та BERT, із детальним розглядом архітектури трансформерів та параметрів конфігурації. Обговорення практичних прикладів інтеграції LLMs у додатки для автоматизації процесів, генерації тексту та інших завдань, з використанням AWS SageMaker, Hugging Face Transformers та OpenAI API.</p> <p>Тема 8. Використання алгоритмів машинного навчання у імітаційному моделюванні та прийнятті рішень. Обговорення у форматі світового кафе. Методи машинного навчання у задачах інтелектуального аналізу даних та прийнятті рішень. Використання машинного навчання у DDDM (прийняття рішень на основі даних). Приклади використання алгоритмів машинного навчання у різних сферах бізнесу, логістиці, імітаційному моделюванні роботи техніки та моделюванні біологічних систем. Особливості роботи із Big data для аналізу ефективності роботи бізнес компаній. Особливості розробки алгоритмічного та програмного забезпечення для аналізу даних (включно з великими) та його використання для підтримки бізнес сектору у сфері логістики, фінансів, робототехніки та ін.</p>		<p>Практичне заняття 9. Використання алгоритмів машинного навчання у прийнятті рішень Робота над проектом у командах, обраних викладачем</p> <p>Практичне заняття 10 МКР2</p>		<p>Основи трансформерів: self-attention механізм, Multi-Head Attention. Детальний розгляд архітектури GPT та BERT. Огляд Multi-Head Attentionта її впливу на ефективність моделей.</p> <p>Використання машинного навчання у DDDM (прийняття рішень на основі даних). Використання алгоритмів машинного навчання у різних сферах бізнесу та моделюванні роботи техніки</p>	<p>10</p> <p>12</p>
	6		2		82

Розділ 5. «Система оцінювання знань студентів»

Таблиця 5.1 - Розподіл балів за результатами вивчення навчальної дисципліни
Підсумкова оцінка за вивчення навчальної дисципліни розраховується через поточне оцінювання

Форма роботи	Вид роботи	Бали
Навчальна	1. Виконання практичних завдань (5*8=40 балів)	40
	2. Виконання МКР (10 балів за 1 МКР, 10*2=20 балів)	20
	3. Іспит	40
Всього з курсу		100
Зарахування Тем 3-7 при опрацюванні та наявності сертифікату з курсу Machine Learning for Natural Language Processing на платформі AWS Academy		

Таблиця 5.2 – Система нарахування додаткових балів за видами робіт з вивчення навчальної дисципліни

Форма роботи	Вид роботи	Бали
Науково-дослідна	Опрацювання додаткових навчальних курсів, написання та публічний захист наукового реферату на теми, що передбачені у завданнях самостійної роботи у розрізі тем Написання наукової статті для фахового збірника наукових робіт	10*

*За додаткові види навчальних робіт студент може отримати не більше 10 балів. Додаткові бали додаються до загальної підсумкової оцінки за вивчення навчальної дисципліни, але загальна підсумкова оцінка не може перевищувати 100 балів.

Таблиця 6 – Шкала оцінювання знань здобувачів вищої освіти за результатами вивчення навчальної дисципліни

Сума балів за всі види навчальної діяльності	Оцінка за шкалою ЄКТС	Оцінка за національною шкалою
90-100	A	Відмінно
82-89	B	Дуже добре
74-81	C	Добре
64-73	D	Задовільно
60-63	E	Задовільно достатньо
35-59	FX	Незадовільно з можливістю повторного складання
0-34	F	Незадовільно з обов'язковим повторним вивченням навчальної дисципліни

Розділ 6. Інформаційні джерела

1. Brockman G., et al. Deep Reinforcement Learning Hands-On / Maxim Lapan. - Packt Publishing, 2020. - 354p.
2. Brownlee J. Machine Learning Mastery with Python / Jason Brownlee. - Machine Learning Mastery, 2020. - 500p.
3. Hands-On Genetic Algorithms with Python: Apply genetic algorithms to solve real-world AI and machine learning problems by Eyal Wirsansky, Edition 2, 2024
4. C. Aggarwal. Neural Networks and Deep Learning A Textbook. 2023. 529 c.
5. Chollet F. Deep Learning with Python / François Chollet. - Manning Publications,

2018. - 384p.

6. Feature Engineering / S. Sumathi та ін. Machine Learning for Decision Sciences with Case Studies in Python. Boca Raton, 2022. С. 351–371. URL: <https://doi.org/10.1201/9781003258803-7>.
7. Foster D. Generative Deep Learning: Teaching Machines to Paint, Write, Compose, and Play. O'Reilly Media, Incorporated, 2023.
8. Géron A. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow / Aurélien Géron. - O'Reilly Media, 2019. - 818p.
9. Goodfellow I., Bengio Y., Courville A. Deep Learning / Ian Goodfellow, Yoshua Bengio, Aaron Courville. - MIT Press, 2023. - 800p.
10. Jones A. The Unsupervised Learning Workshop / A. Jones, K. Christopher, B. Johnston., 2020. – 549 с. – (Packt Publishing).
11. Keras T. Keras Documentation and Tutorials / François Chollet, et al. - Keras Documentation, 2021.
12. Machine Learning Mastery. MachineLearningMastery.com. URL: <https://machinelearningmastery.com/>
13. Machine Learning. Google for Developers. URL: <https://developers.google.com/machine-learning>.
14. Müller A.C., Guido S. Introduction to Machine Learning with Python / Andreas C. Müller, Sarah Guido. - O'Reilly Media, 2019. - 400p.
15. Peters M.E., Schuster M., Rotman B., et al. Transformers for Natural Language Processing / Michael E. Peters, Matt Schuster, et al. - Packt Publishing, 2021. - 410p.
16. Pramod Singh. Deploy Machine Learning Models to Production: With Flask, Streamlit, Docker, and Kubernetes on Google Cloud Platform. Apress. - 2021. - 161 с.
17. Raschka S., Mirjalili V. Python Machine Learning / Sebastian Raschka, Vahid Mirjalili. - Packt Publishing, 2021. - 500p.
18. Vaswani A., Shazeer N., Parmar N., et al. Attention is All You Need / Ashish Vaswani, Noam Shazeer, et al. - 2017. (Репринт з доповненнями в 2020)
19. Machine Learning, Data Science and Generative AI with Python. URL: <https://ua.udemy.com/course/data-science-and-machine-learning-with-python-hands-on/learn/lecture/4020720#overview>
20. AWS Certified Machine Learning Specialty 2024 - Hands On! URL: <https://ua.udemy.com/course/aws-machine-learning/>
21. Стрілець В.Є. Методи машинного навчання у задачах системного аналізу і прийняття рішень : монографія / В. Є. Стрілець, С. І. Шматков, М. Л. Угрюмов та ін. – Харків : Харківський національний університет імені В. Н. Каразіна, 2020. – 160 с.
22. Кононова К. Ю. Машинне навчання: методи та моделі: підручник для бакалаврів, магістрів та докторів філософії спеціальності 051 «Економіка» / К. Ю. Кононова. – Харків: ХНУ імені В. Н. Каразіна, 2020. – 301 с.
23. Complete Machine Learning,NLP Bootcamp MLOPS & Deployment. URL: <https://ua.udemy.com/course/complete-machine-learning-nlp-bootcamp-mlops-deployment/>
24. Кошова, О. П., Ольховська, О. В., & Бражніченко, А. О. (2024). ПІДВИЩЕННЯ ЕФЕКТИВНОСТІ ПРИЙНЯТТЯ УПРАВЛІНСЬКИХ РІШЕНЬ В ІТ-ПРОЕКТАХ МЕТОДАМИ ІМІТАЦІЙНОГО МОДЕЛЮВАННЯ. Таврійський науковий вісник. Серія: Технічні науки, (3), 39-50. <https://doi.org/10.32782/tnv-tech.2024.3.5>
25. Кошова О.П. Застосування принципів об'єктно-орієнтованих мов

програмування для моделювання клітинного поділу / О.П. Кошова, Ю.Ф. Олексійчук, А.В. Оборожний, О.О. Черненко // Збірник наукових праць національного університету кораблебудування імені адмірала Макарова, 2024, № 3, С. 87-94. <http://znp.nuos.mk.ua/3-2024> <http://znp.nuos.mk.ua/archives/2024/3/15.pdf>

Розділ 7. Програмне забезпечення навчальної дисципліни

Персональний комп'ютер або ноутбук зі сталим доступом до мережі Інтернет.

Система дистанційного навчання ПУЕТ та його засоби відео-конференц зв'язку, дистанційний курс з ОК «Машинне навчання».

Продукти, програми та служби Microsoft.

Web-browser (Chrome, Firefox, Safari, тощо)

AWS CLI (Command Line Interface)

Jupyter Notebook або JupyterLab

Python та необхідні бібліотеки: NumPy, Pandas, scikit-learn, spaCy, NLTK, Gensim, TensorFlow, PyTorch, Hugging Face Transformers.

IDE або текстовий редактор (PyCharm, Visual Studio Code, Sublime Text, тощо)

Tesseract OCR