



# Інтелектуальний аналіз великих сховищ даних (Big data analytics)

## Робоча програма навчальної дисципліни (Силабус)

### Реквізити навчальної дисципліни

Рівень вищої освіти	<i>Другий (магістерський)</i>
Галузь знань	12 Інформаційні технології
Спеціальність	124 «Системний аналіз»
Освітня програма	«Системний аналіз і управління»
Статус дисципліни	Вибіркова
Форма навчання	очна(денна)
Рік підготовки, семестр	1 курс, весняний семестр
Обсяг дисципліни	5 кредитів (150 годин), 36 год.лекції, 18 год.практичні заняття, 96 год.СРС
Семестровий контроль/ контрольні заходи	Екзамен / МКР, письмово
Розклад занять	Щотижня лекція, один раз на два тижня практикум, щотижня заняття з прийому завдань СРС, <a href="https://schedule.kpi.ua/">https://schedule.kpi.ua/</a>
Мова викладання	Українська
Інформація про керівника курсу / викладачів	Лектор: д.т.н., доцент, доц. Недашківська Надія Іванівна, n.nedashkivska@gmail.com Практичні заняття: д.т.н., доцент, доц. Недашківська Надія Іванівна, n.nedashkivska@gmail.com
Розміщення курсу	Платформа дистанційного навчання «Сікорський», Googleclassroom, код курсу nu3ese2

### Програма навчальної дисципліни

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

Метою кредитного модуля є формування у студентів здатностей:

- застосовувати сучасні моделі і алгоритми інтелектуального аналізу даних і машинного навчання, глибокі нейронні мережі прямого розповсюдження, згорткові мережі, структурні ймовірнісні моделі, автокодувальники, рекурентні нейронні мережі, методи регуляризації;
- виконувати попередню обробку даних і побудову навчальних наборів;
- будувати моделі кодувальник-декодувальник, породжуючі змагальні мережі, моделі рекомендаційних систем, навчати вказані моделі, оцінювати якість їх роботи, використовуючи програмне забезпечення python;
- розв'язувати задачі класифікації, структурного виводу, синтезу і вибірки, породження нових даних, пошуку асоціативних правил, навчання ранжуванню, прогнозування покупок, надання персоналізованих рекомендацій.

Після засвоєння кредитного модуля мають продемонструвати такі результати навчання:

**компетентності:**

здатність застосовувати знання в практичних ситуаціях, здатність абстрактно мислити, застосовувати методи аналізу і синтезу, здатність знати та розуміти предметну область і професійну діяльність, здатність до пошуку, оброблення та аналізу інформації з різних джерел, здатність до адаптації та дії в новій ситуації, здатність забезпечувати та оцінювати якість виконуваних робіт,

здатність використовувати системний аналіз в якості сучасної міждисциплінарної методології, заснованої на прикладах математичних методів та сучасних інформаційних технологіях, і орієнтована на вирішення задач аналізу і синтезу технічних, економічних, соціальних, екологічних та інших складних систем,

здатність будувати математично коректні моделі статичних та динамічних процесів і систем із зосередженими та розподіленими параметрами із врахуванням невизначеності зовнішніх та внутрішніх факторів,

здатність до комп'ютерної реалізації математичних моделей реальних систем і процесів; проектувати, застосовувати і супроводжувати програмні засоби моделювання, прийняття рішень, обробки інформації, інтелектуального аналізу даних,

здатність розробляти експериментальні та спостережувальні дослідження і аналізувати дані, отримані в них,

застосовувати методи і засоби роботи з даними і знаннями, методи математичного моделювання, технології системного і статичного аналізу,

проектувати, реалізовувати, тестувати, впроваджувати, супроводжувати, експлуатувати програмні засоби роботи з даними і знаннями в комп'ютерних системах і мережах,

розуміти і застосовувати на практиці методи статичного моделювання і прогнозування, оцінювати вихідні дані;

#### **ЗНАННЯ:**

методів попередньої обробки даних і побудови навчальних наборів, оцінювання важливості ознак і вибору значущих ознак, теорії глибоких нейронних мереж прямого розповсюдження, багатозарового перцептрона, згорткових нейронних мереж; поняття перенавчання (overfitting) моделі та регуляризації; кривих навчання і перевірки; U-кривої; методу перехресної перевірки моделі (cross-validation, CV); алгоритмів решітчастого пошуку Grid Search CV та рандомізованого пошуку Random Search CV; метрик якості алгоритмів навчання з вчителем; теореми Байєса; максимальної апостеріорної гіпотези (MAP); функцій втрат в задачах класифікації на основі перехресної ентропії; функцій активації: сигмоїдна, ReLU, LeakyReLU, ELU, Swish та ін.; технологій проектування архітектури глибоких нейронних мереж, скритих і вихідних шарів нейронної мережі; проблем оптимізації нейронних мереж: погана обумовленість, локальні мінімуми, плато, довгострокові залежності та інші; алгоритмів навчання градієнтними методами: стохастичного градієнтного спуску, з адаптивною швидкістю навчання: AdaGrad, Adadelta, RMSProp, Adam, пакетних і міні-пакетних алгоритмів, алгоритму зворотного розповсюдження помилки, методів оптимізації другого порядку: Ньютона, Гауса-Ньютона, спряжених градієнтів, квазіньютонівських; шляхів розв'язання проблеми вибору швидкості навчання; стратегій оптимізації і метаалгоритмів: нормування на основі міні-батчів, покоординатний спуск, усереднення Поляка; як проектувати моделі з урахуванням простоти оптимізації; засобів регуляризації глибоких моделей: штрафи за нормою параметрів, робастність відносно шуму, поповнення набору даних, рання зупинка, зв'язування і розділення параметрів, багатозадачне навчання, розріджені представлення, ансамблеві методи, змагальне навчання; сучасних методів ініціалізації ваг: Ксав'є і Хе; методу нормалізація за міні-батчами; дропауту; операції згортки і субдискретизації, ефективні алгоритми згортки; проблеми безструктурного моделювання; застосування графів для описання структури моделі: орієнтовані моделі, неорієнтовані моделі, факторні графи, енергетичні моделі; переваги структурного моделювання; методи Монте-Карло; вибірка за значимістю; вибірка за

Гіббсом; понижуючі та регуляризовані автокодувальники; стохастичні кодувальники і декодувальники; варіаційний автокодувальник; вибірка з автокодувальників; марковська мережа, асоційована з довільним шумоподавляючим автокодувальником; архітектури породжуючих змагальних мереж (GAN), згорткової мережі GAN – мережі DCGAN, мережі GAN Вассерштейна, функцій втрат генератора і дискримінатора; видів рекомендаційних систем; методів колаборативної фільтрації, SVD, контентної фільтрації; проблем розробки рекомендаційних систем; метрик оцінювання якості рекомендацій; алгоритмів надання рекомендацій на основі наборів даних транзакцій: Apriori, Eclat, FP-growth; метрик оцінювання якості асоціативних правил; гібридних моделей побудови прогнозу в рекомендаційних системах, моделі нейронної колаборативної фільтрації; основи рекурентних нейронних мереж RNN; модель в просторі станів; задачі обробки послідовностей; проблеми навчання рекурентних нейронних мереж; моделі довгої короткотермінової пам'яті (Long Short-Term Memory, LSTM) та GRU; моделі рекомендаційних систем на основі рекурентних нейронних мереж типу кодувальник-декодувальник;

технологій реалізації глибоких нейронних мереж в TensorFlow Python для розв'язання задач класифікації, регресії, структурного виводу, синтезу і вибірки, породження нових даних, пошуку асоціативних правил, побудови рекомендаційних систем;

#### **уміння:**

застосовувати описані вище сучасні моделі і алгоритми інтелектуального аналізу великих сховищ даних і машинного навчання, виконувати попередню обробку даних і побудову навчальних наборів, оцінювати важливість ознак і обирати значущі ознаки, виконувати прогнозування на основі глибоких нейромережевих моделей, обирати функції активації, алгоритм оптимізації, обирати гіперпараметри моделі, оцінювати навчену модель на тестовому наборі даних, зберігати і повторно завантажувати навчені моделі, будувати і навчати понижуючий, регуляризований і варіаційний автокодувальники, породжувати нові зображення змагальними мережами GAN, DCGAN, GAN Вассерштейна; виконувати прогнозування на основі наборів даних транзакцій, використовуючи алгоритми Apriori, Eclat, FP-growth, оцінювати якість побудованих асоціативних правил; будувати моделі рекомендаційних систем на основі рекурентних нейронних мереж RNN типу кодувальник-декодувальник, оцінювати якість рекомендацій, використовуючи програмне забезпечення Python;

#### **досвід:**

теоретичний та практичний досвід аналізу і обробки даних у різних форматах з метою підтримки прийняття рішень, побудови прогнозів, використання програмного забезпечення Python для інтелектуального аналізу даних та машинного навчання в практичній роботі.

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

При вивченні дисципліни використовуються знання дисциплін «Теорія ймовірностей», «Математична статистика», «Математичний аналіз», «Лінійна алгебра», «Методи оптимізації», «Чисельні методи», «Об'єктно-орієнтовне програмування», «Дискретна математика», «Математична логіка», «Інтелектуальний аналіз даних», «Інтелектуальні системи підтримки прийняття рішень», знають синтаксис мови програмування Python.

Знання, набуті при вивченні цієї дисципліни, використовуються в дипломному проектуванні, у практичній самостійній роботі випускника в галузі інтелектуального аналізу даних під час аналізу великих і надвеликих баз даних та масивів тексту, при побудові прогнозів на основі статистичних даних та оцінок експертів, при розробці корпоративних інформаційно-аналітичних систем в державних і приватних управлінських структурах.

### **3. Зміст навчальної дисципліни**

#### **Розділ 1. Основи інтелектуального аналізу даних (ІАД) та машинного навчання (МН)**

Тема 1. Загальні відомості про ІАД. Досвід в задачах ІАД: навчання з вчителем – *supervised learning*, без вчителя – *unsupervised learning*, з частковим залученням вчителя – *semi-supervised learning*. Класифікація, регресія, структурний вивід, синтез і вибірка, оцінка функції ймовірності та функції щільності ймовірності, навчання ранжуванню, кластеризація, сегментація, зниження розмірності, пошук асоціативних правил. Огляд методів ІАД та машинного навчання.

Тема 2. Оцінювання якості алгоритмів навчання з вчителем. Поняття перенавчання (*overfitting*) моделі та регуляризації. Компроміс між систематичною помилкою і дисперсією моделі. Криві навчання і перевірки. U-крива. Перехресна перевірка моделі (*cross-validation, CV*). Вибір гіперпараметрів моделі методами решітчастого пошуку *Grid Search CV* та рандомізованого пошуку *Random Search CV*. Метрики якості алгоритмів навчання з вчителем. Реалізація в *scikit-learn python*.

Тема 3. Теорема Байеса. Максимальна апостеріорна гіпотеза. Метод максимальної правдоподібності.

Тема 4. Практичне застосування ІАД.

#### **Розділ 2. Попередня обробка даних і побудова навчальних наборів**

Тема 5. Обробка категоріальних даних: їх кодування за допомогою *pandas*, кодування міток класів, відображення порядкових ознак, виконання унітарного кодування на іменних ознаках. Приведення ознак до одного масштабу.

Тема 6. Розв'язання проблеми з відсутніми даними: їх ідентифікація в таблицях, підходи до розрахунку даних, що відсутні.

Тема 7. Оцінювання важливості ознак і вибір значущих ознак.

#### **Розділ 3. Розпаралелювання процесу навчання нейронних мереж за допомогою TensorFlow Python. Механіка TensorFlow**

Тема 8. Проблеми, пов'язані з продуктивністю навчання. Створення тензорів в TensorFlow і робота з ними. API Dataset TensorFlow.

Тема 9. Побудова нейромережових моделей багатошарового перцептрона для класифікації і регресії в TensorFlow. API Keras. Вибір функцій активації для глибоких нейронних мереж: *SoftMax, tanh, RELU* та її модифікації, *Swish*. Оцінювання навченої моделі на тестовому наборі даних. Збереження і повторне завантаження навченої моделі.

Тема 10. Графи обчислень в TensorFlow: створення графу, перенос графу, завантаження вхідних даних в модель. Декоратори функцій. Об'єкти *Variable* для збереження і оновлення параметрів моделі.

Тема 11. Створення власних класів, використовуючи *tf.Module*, *tf.keras.Model*, *tf.keras.layers.Layer*.

Тема 12. Розрахунок градієнтів за допомогою автоматичного диференціювання і *GradientTape*. Використання оцінщиків TensorFlow: *tf.estimator*.

#### **Розділ 4. Оптимізація в навчанні глибоких моделей. Регуляризація**

Тема 13. Проблеми оптимізації нейронних мереж. Градієнтні методи з адаптивною швидкістю навчання: *AdaGrad, Adadelta, RMSProp, Adam*. Вибір алгоритму оптимізації. Реалізація в TensorFlow і Keras.

Тема 14. Методи оптимізації другого порядку: Ньютона, Гауса-Ньютона, спряжених градієнтів, квазіньютонівські.

Тема 15. Стратегії оптимізації і метаалгоритми: нормування на основі міні-батчів, покоординатний спуск, усереднення Поляка. Проектування моделей з урахуванням простоти оптимізації.

Тема 16. Регуляризація глибоких моделей: штрафи за нормою параметрів, робастність відносно шуму, поповнення набору даних, рання зупинка, зв'язування і розділення параметрів, багатозадачне навчання, розріджені представлення, ансамблеві методи, змагальне навчання.

### **Розділ 5. Структурні ймовірнісні моделі у глибокому навчанні**

Тема 17. Проблема безструктурного моделювання. Застосування графів для описання структури моделі: орієнтовані моделі, неорієнтовані моделі, факторні графи, енергетичні моделі. Переваги структурного моделювання.

Тема 18. Вибірка і методи Монте-Карло. Вибірка за значимістю. Методи Монте-Карло за схемою марковської мережі. Вибірка за Гіббсом.

Тема 19. Наближений вивід. Вивід як оптимізація. MAP-вивід і розріджене кодування. Варіаційний вивід і навчання: дискретні і неперервні латентні змінні. Взаємодія між навчанням і виводом. Навчений наближений вивід.

### **Розділ 6. Автокодувальники**

Тема 20. Понижуючі та регуляризовані автокодувальники. Стохастичні кодувальники і декодувальники. Репрезентативна здатність, розмір шару і глибина. Варіаційний автокодувальник.

Тема 21. Вибірка з автокодувальників. Марковська мережа, асоційована з довільним шумоподавляючим автокодувальником.

Тема 22. Застосування автокодувальників. Побудова і навчання автокодувальників в TensorFlow.

### **Розділ 7. Орієнтовані породжуючі моделі**

Тема 23. Породжуючі змагальні мережі (GAN). Функції втрат генератора і дискримінатора. Реалізація породжуючої змагальної мережі в TensorFlow.

Тема 24. Згорткові мережі GAN – мережі DCGAN. Мережі GAN Вассерштейна. Реалізація в TensorFlow.

Тема 25. Інші модифікації GAN.

Тема 26. Застосування GAN.

Тема 27. Сигмоїдні мережі довіри. Авторегресивні мережі.

### **Розділ 8. Рекомендаційні системи**

Тема 28. Види рекомендаційних систем. Навчання ранжуванню. Колаборативна фільтрація. Алгоритм SVD. Проблеми розробки рекомендаційних систем.

Тема 29. Метод контентної фільтрації.

Тема 30. Метрики оцінювання якості рекомендацій.

Тема 31. Задача аналізу ринкових кошиків. Алгоритми надання рекомендацій на основі наборів даних транзакцій: Apriori, Eclat, FP-growth. Реалізація цих алгоритмів на python. Метрики оцінювання якості асоціативних правил. Шаблони послідовностей.

Тема 32. Гібридні алгоритми побудови прогнозу в рекомендаційних системах. Нейронна колаборативна фільтрація.

Тема 33. Основи рекурентних нейронних мереж. Модель в просторі станів. Задачі обробки послідовностей. Алгоритми зворотного розповсюдження в часі (BackPropagation Throught Time) для навчання рекурентних мереж. Проблеми навчання рекурентних нейронних мереж.

Тема 34. Модель довгої короткотермінової пам'яті (Long Short-Term Memory, LSTM). Модель GRU. Модифікації LSTM.

Тема 35. Моделі рекомендаційних систем на основі рекурентних нейронних мереж типу кодувальник-декодувальник. Реалізація в TensorFlow.

Напрямки розвитку та перспективи подальших досліджень в області ІА великих сховищ даних та машинного навчання. Невирішені проблеми.

#### 4. Навчальні матеріали та ресурси

##### Базова

1. Н.І. Недашківська. Слайди лекцій з кредитного модуля «Інтелектуальний аналіз великих сховищ даних», 2021, <https://classroom.google.com/c/MjQ3NzY0ODQzMjc?cjc=3ogbesx>
2. Н.І. Недашківська. Методичні вказівки до виконання практичних робіт з кредитного модуля «Інтелектуальний аналіз великих сховищ даних», 2021, <https://classroom.google.com/c/MjQ3NzY0ODQzMjc?cjc=3ogbesx>
3. Н.І. Недашківська. Слайди лекцій з кредитного модуля «Інтелектуальний аналіз великих сховищ даних», 2022, <https://classroom.google.com/c/MjI3Nzc5NzMyMDU5?cjc=nu3ese2>
4. Н.І. Недашківська. Методичні вказівки до виконання практичних робіт з кредитного модуля «Інтелектуальний аналіз великих сховищ даних», 2022, <https://classroom.google.com/c/MjI3Nzc5NzMyMDU5?cjc=nu3ese2>

Зазначені в п. 1 і 2 матеріали є обов'язковими для прочитання, їх можна знайти на Платформі дистанційного навчання «Сікорський», Googleclassroom, коди курсів **3ogbesx** і **nu3ese2**.

##### Додаткова література

5. Ian Goodfellow, Yoshua Bengio, Aaron Courville. Deep Learning. The MIT Press Cambridge, Massachusetts London, England, 2017.
6. Sebastian Raschka, Vahid Mirjalili. Python Machine Learning. Third Edition. Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow 2. Packt Publishing, 2019.
7. Scikit-Learn Documentation. Режим доступу: <https://scikit-learn.org/>, 2022.
8. TensorFlow Documentation. Режим доступу: <https://www.tensorflow.org> . 2022.
9. Keras Documentation. Режим доступу: <https://keras.io> . 2022.
10. Jake VanderPlas. Python Data Science Handbook. Essential Tools for Working with Data. O'Reilly Media Inc., 2017. 576 p.
11. Aurelien Geron. Hands-On Machine Learning with Scikit-Learn and TensorFlow. O'Reilly Media Inc., Sebastopol, CA, 2017.
12. Rodolfo Bonnin. Building Machine Learning Projects with TensorFlow. Packt Publishing Ltd., Birmingham, Uk, 2016.
13. Ramsundar B., Zadeh R.B.. TensorFlow for Deep Learning. O'Reilly Media Inc., Sebastopol, CA, 2018.
14. Wes McKinney. Python for Data Analysis. O'Reilly Media Inc., 2013. 482 p.
15. Henrik Brink Joseph W. Richards Mark Fetherolf. Real-World Machine Learning. Manning Publications. 2016. 336 p.

16. Andreas C. Mueller and Sarah Guido. *An Introduction to Machine Learning with Python*. O'Reilly Media Inc., 2017. 392 p.
17. Davy Cielen, Arno Meysman, Mohamed Ali. *Introducing Data Science: Big Data, Machine Learning, and more, using Python tools*. Manning Publications, 2016. 320 p.

Використовується сучасне комп'ютерне та мультимедійне обладнання, платформа дистанційного навчання «Сікорський».

Для виконання практичних робіт використовується open-source програмне забезпечення Python (<https://www.python.org/>), Scikit-Learn 1.2.1 – open source, commercially usable – BSD license (<https://scikit-learn.org/>), TensorFlow v.2.11.0 – Apache-2.0 license (<https://www.tensorflow.org/>), Keras – Apache-2.0 license (<https://keras.io>)

## Навчальний контент

### 5. Методика опанування навчальної дисципліни (освітнього компонента)

#### Лекційні заняття

Лекція 1. Загальні відомості про ІАД. Досвід в задачах ІАД: навчання з вчителем – *supervised learning*, без вчителя – *unsupervised learning*, з частковим залученням вчителя – *semi-supervised learning*. Класифікація, регресія, структурний вивід, синтез і вибірка, оцінка функції ймовірності та функції щільності ймовірності, кластеризація, сегментація, зниження розмірності, породження нових даних, пошук асоціативних правил, навчання ранжуванню і побудова рекомендаційних систем. Огляд методів ІАД та машинного навчання. Теорема Байеса. Максимальна апостеріорна гіпотеза. Метод максимальної правдоподібності. [1 – 6]

Лекція 2. Оцінювання якості алгоритмів навчання з вчителем. Поняття перенавчання (*overfitting*) моделі та регуляризації. Компроміс між систематичною помилкою і дисперсією моделі. Криві навчання і перевірки. U-крива. Перехресна перевірка моделі (*cross-validation*, CV). Вибір гіперпараметрів моделі. Методи решітчастого пошуку *Grid Search CV* та рандомізованого пошуку *Random Search CV*. Метрики якості алгоритмів навчання з вчителем. Реалізація в *scikit-learn python*. [1 – 7]

Лекція 3. Обробка категоріальних даних: їх кодування за допомогою *pandas*, кодування міток класів, відображення порядкових ознак, виконання унітарного кодування на іменних ознаках. Приведення ознак до одного масштабу. [10 – 17]

Лекція 4. Оцінювання важливості ознак і вибір значущих ознак. [10 – 17]

Лекція 5. Проблеми, пов'язані з продуктивністю навчання. Створення тензорів в *TensorFlow* і робота з ними. *API Dataset TensorFlow*. Побудова нейромережових моделей багатoshарового перцептрона для класифікації і регресії в *TensorFlow*. Оцінювання навченої моделі на тестовому наборі даних. Збереження і повторне завантаження навченої моделі. [3 – 9]

Лекція 6. Графи обчислень в *TensorFlow*: створення графу, перенос графу, завантаження вхідних даних в модель. Декоратори функцій. Об'єкти *Variable* для збереження і оновлення параметрів моделі. Створення власних класів, використовуючи *tf.Module*, *tf.keras.Model*, *tf.keras.layers.Layer*. Розрахунок градієнтів за допомогою автоматичного диференціювання і *GradientTape*. Використання оцінщиків *TensorFlow*: *tf.estimator*. [3 – 9]

Лекція 7. Проблеми оптимізації нейронних мереж. Градієнтні методи з адаптивною швидкістю навчання: *AdaGrad*, *Adadelta*, *RMSProp*, *Adam*. Вибір алгоритму оптимізації. Реалізація в *TensorFlow* і *Keras*. Методи оптимізації другого порядку: Ньютон, Гауса-Ньютон, спряжених градієнтів, квазіньютонівські. [1 – 5, 12 – 17]

*Лекція 8. Стратегії оптимізації і метаалгоритми: нормування на основі міні-батчів, покоординатний спуск, усереднення Поляка. Проектування моделей з урахуванням простоти оптимізації. [1 – 5]*

*Лекція 9. Регуляризація глибоких моделей: штрафи за нормою параметрів, робастність відносно шуму, поповнення набору даних, рання зупинка, зв'язування і розділення параметрів, багатозадачне навчання, розріджені представлення, ансамблеві методи, змагальне навчання. [1 – 5, 12 – 17]*

*Лекція 10. Проблема безструктурного моделювання. Застосування графів для описання структури моделі: орієнтовані моделі, неорієнтовані моделі, факторні графи, енергетичні моделі. Переваги структурного моделювання. [1 – 5]*

*Лекція 11. Понижуючі та регуляризовані автокодувальники. Стохастичні кодувальники і декодувальники. Репрезентативна здатність, розмір шару і глибина. [1 – 5, 12 – 17]*

*Лекція 12. Варіаційний автокодувальник. Застосування автокодувальників. Побудова і навчання автокодувальників в TensorFlow. [1 – 6, 8, 9]*

*Лекція 13. Породжуючі змагальні мережі (GAN). Функції втрат генератора і дискримінатора. Реалізація породжуючої змагальної мережі в TensorFlow. Згорткові мережі GAN – мережі DCGAN. Мережі GAN Вассерштейна. Реалізація в TensorFlow. [1 – 6, 8, 9]*

*Лекція 14. Види рекомендаційних систем. Навчання ранжуванню. Колаборативна фільтрація. Алгоритм SVD. Метод контентної фільтрації. Проблеми розробки рекомендаційних систем. Метрики оцінювання якості рекомендацій.*

*Лекція 15. Задача аналізу ринкових кошиків. Алгоритми надання рекомендацій на основі наборів даних транзакцій: Apriori, Eclat, FP-growth. Реалізація цих алгоритмів на python. Метрики оцінювання якості асоціативних правил. Шаблони послідовностей. [1 – 4]*

*Лекція 16. Основи рекурентних нейронних мереж. Модель в просторі станів. Задачі обробки послідовностей. Алгоритми зворотного розповсюдження в часі (BackPropagation Through Time) для навчання рекурентних мереж. Проблеми навчання рекурентних нейронних мереж. [1 – 6]*

*Модель довгої короткотермінової пам'яті (Long Short-Term Memory, LSTM). Модель GRU. Модифікації LSTM [1 – 6]*

*Лекція 17. Гібридні алгоритми побудови прогнозу в рекомендаційних системах. Нейронна колаборативна фільтрація. [5]*

*Лекція 18. Моделі рекомендаційних систем на основі рекурентних нейронних мереж типу кодувальник-декодувальник. Реалізація в TensorFlow. [5, 6, 8, 9]*

*Напрямки розвитку та перспективи подальших досліджень в області ІА великих сховищ даних та машинного навчання. Невирішені проблеми.*

### **Практичні / лабораторні роботи**

*Метою практичних/ лабораторних робіт є закріплення теоретичних положень навчальної дисципліни, отримання практичних навичок створення і навчання моделей інтелектуального аналізу великих сховищ даних на python. В результаті виконання робіт студенти отримують практичні навички застосовувати сучасні моделі і алгоритми інтелектуального аналізу даних і машинного навчання, глибокі нейронні мережі прямого розповсюдження, згорткові мережі, структурні ймовірнісні моделі, автокодувальники, рекурентні нейронні мережі, методи регуляризації. Будуть вміти виконувати попередню обробку даних і побудову навчальних наборів; розв'язувати задачі класифікації, породження нових даних, пошуку асоціативних правил, прогнозування і надання персоналізованих рекомендацій; будувати моделі кодувальник-декодувальник, породжуючі моделі, моделі*



рекомендаційних систем, оцінювати якість їх роботи, використовуючи програмне забезпечення python.

№ з/п	Назва роботи	Кількість ауд. годин
1	Попередня обробка даних і побудова навчальних наборів. Оцінювання важливості ознак і вибір значущих ознак.	2
2	Класифікація і регресія на основі глибоких нейромережових моделей прямого розповсюдження сигналу. Вибір функцій активації. Вибір алгоритму оптимізації. Вибір гіперпараметрів моделі. Оцінювання навченої моделі на тестовому наборі даних. Збереження і повторне завантаження навченої моделі.	2
3	Побудова і навчання понижуючого, регуляризованого і варіаційного автокодувальників.	4
4	Породження нових зображень змагальними мережами GAN, DCGAN, GAN Вассерштейна.	4
5	Прогнозування на основі наборів даних транзакцій, використовуючи алгоритми Apriori, Eclat, FP-growth. Оцінювання якості побудованих асоціативних правил.	2
6	Побудова моделі рекомендаційних систем на основі рекурентних нейронних мереж типу кодувальник-декодувальник. Оцінювання якості рекомендацій.	4

Для виконання практичних робіт використовується open-source програмне забезпечення Python (<https://www.python.org/>), Scikit-Learn 1.2.1 – open source, commercially usable – BSD license (<https://scikit-learn.org/>), TensorFlow v.2.11.0 – Apache-2.0 license (<https://www.tensorflow.org>), Keras – Apache-2.0 license (<https://keras.io>)

## 6. Самостійна робота студента

Самостійна робота студента включає підготовку до практичних/ лабораторних робіт, підготовку до модульної контрольної роботи, в тому числі опрацювання окремих частин наступних тем:

Тема 1. Загальні відомості про ІАД. Досвід в задачах ІАД: навчання з вчителем – supervised learning, без вчителя – unsupervised learning, з частковим залученням вчителя – semi-supervised learning. Класифікація, регресія, структурний вивід, синтез і вибірка, оцінка функції ймовірності та функції щільності ймовірності, кластеризація, сегментація, зниження розмірності, породження нових даних, пошук асоціативних правил, навчання ранжуванню і побудова рекомендаційних систем. Огляд методів ІАД та машинного навчання.

Тема 2. Оцінювання якості алгоритмів навчання з вчителем. Поняття перенавчання (overfitting) моделі та регуляризації. Компроміс між систематичною помилкою і дисперсією моделі. Криві навчання і перевірки. U-крива. Перехресна перевірка моделі (cross-validation, CV). Вибір гіперпараметрів моделі методами решітчастого пошуку Grid Search CV та рандомізованого пошуку Random Search CV. Метрики якості алгоритмів навчання з вчителем.

Тема 3. Теорема Байеса. Максимальна апостеріорна гіпотеза. Метод максимальної правдоподібності.

Тема 4. Практичне застосування ІАД.

Тема 5. Обробка категоріальних даних: їх кодування за допомогою pandas, кодування міток класів, відображення порядкових ознак, виконання унітарного кодування на іменних ознаках. Приведення ознак до одного масштабу.

Тема 6. Розв'язання проблеми з відсутніми даними: їх ідентифікація в таблицях, підходи до розрахунку даних, що відсутні.

Тема 7. Оцінювання важливості ознак і вибір значущих ознак.

Тема 8. Проблеми, пов'язані з продуктивністю навчання. Створення тензорів в TensorFlow і робота з ними. API Dataset TensorFlow.

Тема 9. Побудова нейромережових моделей багатошарового перцептрона для класифікації і регресії в TensorFlow. API Keras. Вибір функцій активації для глибоких нейронних мереж: SoftMax, tanh, RELU та її модифікації, Swish. Оцінювання навченої моделі на тестовому наборі даних. Збереження і повторне завантаження навченої моделі.

Тема 10. Графи обчислень в TensorFlow: створення графу, перенос графу, завантаження вхідних даних в модель. Декоратори функцій. Об'єкти Variable для збереження і оновлення параметрів моделі.

Тема 11. Створення власних класів, використовуючи tf.Module, tf.keras.Model, tf.keras.layers.Layer.

Тема 12. Розрахунок градієнтів за допомогою автоматичного диференціювання і GradientTape. Використання оцінщиків TensorFlow: tf.estimator.

Тема 13. Проблеми оптимізації нейронних мереж. Градієнтні методи з адаптивною швидкістю навчання: AdaGrad, Adadelta, RMSProp, Adam. Вибір алгоритму оптимізації. Реалізація в TensorFlow і Keras.

Тема 14. Методи оптимізації другого порядку: Ньютона, Гауса-Ньютона, спряжених градієнтів, квазіньютонівські.

Тема 15. Стратегії оптимізації і метаалгоритми: нормування на основі міні-батчів, покоординатний спуск, усереднення Поляка. Проектування моделей з урахуванням простоти оптимізації.

Тема 16. Регуляризація глибоких моделей: штрафи за нормою параметрів, робастність відносно шуму, поповнення набору даних, рання зупинка, зв'язування і розділення параметрів, багатозадачне навчання, розріджені представлення, ансамблеві методи, змагальне навчання.

Тема 17. Проблема безструктурного моделювання. Застосування графів для описання структури моделі: орієнтовані моделі, неорієнтовані моделі, факторні графи, енергетичні моделі. Переваги структурного моделювання.

Тема 18. Вибірка і методи Монте-Карло. Вибірка за значимістю. Методи Монте-Карло за схемою марковської мережі. Вибірка за Гіббсом.

Тема 19. Наближений вивід. Вивід як оптимізація. MAP-вивід і розріджене кодування. Варіаційний вивід і навчання: дискретні і неперервні латентні змінні. Взаємодія між навчанням і виводом. Навчений наближений вивід.

Тема 20. Понижуючі та регуляризовані автокодувальники. Стохастичні кодувальники і декодувальники. Репрезентативна здатність, розмір шару і глибина. Варіаційний автокодувальник.

Тема 21. Вибірка з автокодувальників. Марковська мережа, асоційована з довільним шумоподавляючим автокодувальником.

Тема 22. Застосування автокодувальників. Побудова і навчання автокодувальників в TensorFlow.

Тема 23. Породжуючі змагальні мережі (GAN). Функції втрат генератора і дискримінатора. Реалізація породжуючої змагальної мережі в TensorFlow.

Тема 24. Згорткові мережі GAN – мережі DCGAN. Мережі GAN Вассерштейна. Реалізація в TensorFlow.

Тема 25. Інші модифікації GAN.

Тема 26. Застосування GAN.

Тема 27. Сигмоїдні мережі довіри. Авторегресивні мережі.

Тема 28. Види рекомендаційних систем. Навчання ранжуванню. Колаборативна фільтрація. Алгоритм SVD. Проблеми розробки рекомендаційних систем.

Тема 29. Метод контентної фільтрації.

Тема 30. Метрики оцінювання якості рекомендацій.

Тема 31. Задача аналізу ринкових кошиків. Алгоритми надання рекомендацій на основі наборів даних транзакцій: Apriori, Eclat, FP-growth. Реалізація цих алгоритмів на python. Метрики оцінювання якості асоціативних правил. Шаблони послідовностей.

Тема 32. Гібридні алгоритми побудови прогнозу в рекомендаційних системах. Нейронна колаборативна фільтрація.

Тема 33. Основи рекурентних нейронних мереж. Модель в просторі станів. Задачі обробки послідовностей. Алгоритми зворотного розповсюдження в часі (BackPropagation Through Time) для навчання рекурентних мереж. Проблеми навчання рекурентних нейронних мереж.

Тема 34. Модель довгої короткотермінової пам'яті (Long Short-Term Memory, LSTM). Модель GRU. Модифікації LSTM [1, 3]

Тема 35. Моделі рекомендаційних систем на основі рекурентних нейронних мереж типу кодувальник-декодувальник. Реалізація в TensorFlow.

## Політика та контроль

### 7. Політика навчальної дисципліни (освітнього компонента)

Система вимог, які ставляться перед студентом:

- при захисті практичних / лабораторних робіт забороняється використання засобів зв'язку для пошуку інформації на гугл-диску викладача чи в інтернеті;
- при написанні модульної контрольної роботи забороняється використання засобів зв'язку для пошуку інформації на гугл-диску викладача чи в інтернеті;
- на заняттях бажано відключати телефони;
- за несвоєчасну подачу звіту з практичної/ лабораторної роботи нараховуються штрафні бали:  $-0.3$  балів за кожний тиждень запізнення.

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

**Поточний контроль: модульна контрольна робота.**

Модульна контрольна робота складається з двох частин – КР№1 і КР№2.

Кожна КР містить два завдання. Оцінки за теоретичні питання визначаються за шкалою:

- «відмінно», повна відповідь (не менше 95% потрібної інформації) – 4.8-5 балів;
- «добре», достатньо повна відповідь (не менше 75% потрібної інформації), або повна відповідь з незначними неточностями – 3.7 – 4.8 балів;
- «задовільно», неповна відповідь (не менше 60% потрібної інформації) та значні помилки – 3 – 3.7 балів;
- «незадовільно», незадовільна відповідь (не відповідає вимогам на «задовільно») – 0 – 3 бали.

Максимальна оцінка за кожну частину МКР складає 10 балів. **Максимальна кількість балів за дві частини МКР складає  $2 \cdot 10 = 20$  балів.**

**Календарний контроль:** проводиться двічі на семестр як моніторинг поточного стану виконання вимог силябусу.

**Семестровий контроль:** екзамен.

Умови допуску до семестрового контролю: семестровий рейтинг не менше 40 балів.

### **Рейтингова система оцінювання результатів навчання**

Рейтинг студента з кредитного модуля складається з балів, які він отримує за:

- 1) виконання та захист 6 (шести) практичних робіт – максимум 60 балів;
- 2) опрацювання літератури та виконання творчих робіт за однією з тем дисципліни – максимум 20 балів;
- 3) виконання модульної контрольної роботи – максимум 20 балів.

**1. Практичні/ лабораторні роботи.** Упродовж семестру студент має виконати 6 (шість) практичних/ лабораторних робіт (ПР).

Рейтингова оцінка кожної ПР складається з 2 частин, які оцінюються окремо.

а. Якість підготовки до роботи, її виконання та оформлення звіту.

- за умови правильно оформленого звіту з точним виконанням завдання ПР – 5 балів;
- за наявності несуттєвих неточностей в оформленні або процедурі виконання ПР – 4 – 4.5 балів;
- за наявності порушень в оформленні, неповного або неточного виконання – 3-4 бали.

б. Якість захисту матеріалу. В цій частині оцінюється ступінь володіння теоретичним і практичним матеріалом за темою роботи.

- відмінне володіння матеріалом – 5 балів;
- добре володіння матеріалом – 4 – 4.5 балів;
- задовільне володіння матеріалом – 3 – 4 бали.

Максимальна кількість балів за всі ПР дорівнює:  $6 \cdot 10 = 60$  балів.

**2. Опрацювання літератури та виконання творчих робіт за однією з тем дисципліни** оцінюється в **20 балів**.

**3. Модульна контрольна робота.** Модульна контрольна робота складається з двох частин – КР№1 і КР№2. Кожна КР містить два завдання. Оцінки за кожне завдання визначаються за шкалою:

- «відмінно», повна відповідь (не менше 95% потрібної інформації) – 4.8-5 балів;
- «добре», достатньо повна відповідь (не менше 75% потрібної інформації), або повна відповідь з незначними неточностями – 3.7 – 4.8 балів;
- «задовільно», неповна відповідь (не менше 60% потрібної інформації) та значні помилки – 3 – 3.7 балів;
- «незадовільно», незадовільна відповідь (не відповідає вимогам на «задовільно») – 0 – 3 бали.

Максимальна кількість балів за дві частини модульної КР складає  $2 \cdot 10 = 20$  балів.

За результатами навчальної роботи за перші 8 тижнів станом на 24.03 «ідеальний студент» має набрати 49 балів. На першому календарному контролі (8-й тиждень, 24.03) студент отримує «зараховано», якщо його поточний рейтинг не менше  $49/2 = 25$  балів.

За результатами 15 тижнів навчання станом на 12.05 «ідеальний студент» має набрати 100 балів. На другій атестації (15-й тиждень, 12.05) студент отримує «зараховано», якщо його поточний рейтинг не менше 60 балів.

**Максимальна сума балів за роботу в семестрі складає 100.** Необхідною умовою допуску до екзамену є отримання рейтингу 40 балів і вище. Для отримання екзамену з кредитного модуля «автоматом» потрібно мати рейтинг не менше 60 балів.

Студенти, які наприкінці семестру мають рейтинг менше 60 балів, а також ті, хто хоче підвищити оцінку, виконують екзаменаційну роботу. При цьому до балів за лабораторні роботи додаються бали за екзаменаційну роботу, і ця рейтингова оцінка є остаточною.

Завдання екзаменаційної контрольної роботи складається з чотирьох завдань різних розділів силабусу. Кожне завдання контрольної роботи оцінюється у 5 балів відповідно до системи оцінювання:

- «відмінно», повна відповідь (не менше 95% потрібної інформації) – 4.8-5 балів;
- «добре», достатньо повна відповідь (не менше 75% потрібної інформації), або повна відповідь з незначними неточностями – 3.7 – 4.8 балів;
- «задовільно», неповна відповідь (не менше 60% потрібної інформації) та значні помилки – 3 – 3.7 балів;
- «незадовільно», незадовільна відповідь (не відповідає вимогам на «задовільно») – 0 – 3 бали.

Таблиця відповідності рейтингових балів оцінкам за університетською шкалою:

<i>Кількість балів</i>	<i>Оцінка</i>
100-95	Відмінно
94-85	Дуже добре
84-75	Добре
74-65	Задовільно
64-60	Достатньо
Менше 60	Незадовільно
Менше 40	Не допущено

## **9. Додаткова інформація з дисципліни (освітнього компонента)**

*Сертифікати проходження дистанційних чи онлайн курсів за тематикою дисципліни можуть бути зараховані з додатковими 5 – 10 балами до загального рейтингу студента.*

### **Робочу програму навчальної дисципліни (силабус):**

**Складено** доцент, д.т.н., доц. Недашківська Надія Іванівна

**Ухвалено** кафедрою ММСА НН ІПСА (протокол № 11 від 08.07.2022)

**Погоджено** Методичною комісією факультету<sup>1</sup> (протокол № \_\_ від \_\_\_\_\_)

---

<sup>1</sup>Методичною радою університету– для загальноуніверситетських дисциплін.