

**ДЕРЖАВНИЙ ТОРГОВЕЛЬНО-ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ**

**СИСТЕМА УПРАВЛІННЯ ЯКІСТЮ**

Система забезпечення якості освітньої діяльності та якості вищої освіти  
*сертифікована на відповідність ДСТУ ISO 9001:2015 / ISO 9001:2015*

**Кафедра комп'ютерних наук та інформаційних систем**

**ЗАТВЕРДЖЕНО**

вченою радою ДТЕУ

(пост. П. 9 від « 29 » 09 2022 р.)

Ректор



**Анатолій МАЗАРАКІ**

**МАШИННЕ НАВЧАННЯ/  
MACHINE LEARNING**

**ПРОГРАМА /  
COURSE SUMMARY**

**Київ 2022**

**Розповсюдження і тиражування без офіційного дозволу ДТЕУ заборонено**

Автори: О.І. ПУРСЬКИЙ, доктор фізико-математичних наук, професор,  
Т.О. Філімонова, кандидат фізико-математичних наук, доцент  
А.В. Селіванова, старший викладач.

Програму розглянуто і затверджено на засіданні кафедри комп'ютерних наук та інформаційних систем 28 червня 2022р., протокол №35

Рецензенти: Г.Т. САМОЙЛЕНКО, кандидат фізико-математичних наук,  
доцент.  
В.М. ВОЛОХОВ, директор ТОВ «МККУ-МЕРЕЖІ», кандидат  
фізико-математичних наук, доцент, Лауреат Державної премії  
України в галузі науки і техніки

## **МАШИННЕ НАВЧАННЯ/ MACHINE LEARNING**

### **ПРОГРАМА / COURSE SUMMARY**

## ВСТУП

Програма дисципліни «Машинне навчання» призначена для здобувачів першого рівня вищої освіти ОС «Бакалавр», галузі знань 12 «Інформаційні технології», спеціальностей 122 «Комп'ютерні науки» та 126 «Інформаційні системи та технології», спеціалізацій «Комп'ютерні науки» та «Інформаційні системи та технології».

Програму підготовлено відповідно до вимог Стандартів вищої освіти України та відповідних освітньо-професійних програм підготовки бакалаврів.

Програма складається з таких розділів:

1. Мета, завдання та предмет дисципліни.
2. Передумови вивчення дисципліни як вибіркової компоненти освітньої програми.
3. Результати вивчення дисципліни.
4. Зміст дисципліни.
5. Список рекомендованих джерел.

### 1. МЕТА, ЗАВДАННЯ ТА ПРЕДМЕТДИСЦИПЛІНИ

*Метою* вивчення дисципліни «Машинне навчання» є засвоєння основ методології машинного навчання і надання студентами теоретичних знань та формування практичних навичок побудови систем машинного навчання з метою аналізу складних соціально-економічних систем.

*Завданням* вивчення дисципліни «Машинне навчання» є теоретична та практична підготовка з питань: сутності понять і категорій методології машинного навчання та сфер його застосування; організації процесу машинного навчання; застосування алгоритмів машинного навчання; застосування сучасних бібліотек машинного навчання; побудови етапів розробки системи машинного навчання; застосування різних типів процесу машинного навчання; методів вирішення основних типів задач машинного навчання - регресія, кластеризація, класифікація, зниження розмірності; застосування різних механізмів навчання нейронних мереж; визначення адекватності побудованих моделей машинного навчання.

*Предметом* вивчення дисципліни є моделі, методи та інформаційні технології машинного навчання.

### 2. ПЕРЕДУМОВИ ВИВЧЕННЯ ДИСЦИПЛІНИ ЯК ВИБІРКОВОЇ КОМПОНЕНТИ ОСВІТНЬОЇ ПРОГРАМИ

*Знання:*

- предмета і сутності машинного навчання та сфери його застосування;
- типових задач машинного навчання;
- сучасних бібліотек машинного навчання;
- методів машинного навчання;
- основних алгоритмів машинного навчання;
- механізмів навчання нейронних мереж;

- типів архітектур нейронних мереж;
- методів оцінки адекватності моделей машинного навчання;

*Вміння:*

- здійснювати вибір методів машинного навчання для розв'язку задач аналізу та обробки даних;
- застосовувати сучасні бібліотеки машинного навчання в різних алгоритмах машинного навчання;
- роз'язувати типові задачі машинного навчання - регресія, кластеризація, класифікація, зниження розмірності;
- застосовувати різні архітектури нейронних мереж для побудови моделей машинного навчання;
- проводити верифікацію та оцінку якості методів машинного навчання на основі існуючих критеріїв;
- самостійно здійснювати: підготовку (представлення) даних, конструювання алгоритму навчання; тренування алгоритму на наявних даних та валідацію алгоритму на тестових даних;
- вирішувати на основі машинного навчання задачі розпізнавання образів на зображенні та задачі сентимент-аналізу;
- творчо застосовувати набуті знання в процесі вирішення практичних задач аналізу даних.

### 3. РЕЗУЛЬТАТИ ВИВЧЕННЯ ДИСЦИПЛІНИ

Дисципліна «Машинне навчання», як вибіркова компонента освітньо-професійної програми, забезпечує оволодіння студентами загальними та фаховими компетентностями і досягнення ними програмних результатів навчання за відповідними освітньо-професійними програмами:

*Комп'ютерні науки (ОС бакалавр)*

Номер в освітній програмі	Зміст компетентності	Номер теми, що розкриває зміст компетентності
<i>Загальні компетентності за освітньо-професійною програмою</i>		
ЗК 3	Знання та розуміння предметної області та розуміння професійної діяльності.	1,2
ЗК 7	Здатність до пошуку, оброблення та аналізу інформації з різних джерел	1,2,3,5
ЗК 9	Здатність працювати в команді.	2,3,9,10,11,12
<i>Фахові компетентності за освітньо-професійною програмою</i>		
СК 2	Здатність до виявлення статистичних закономірностей недетермінованих явищ, застосування методів обчислювального інтелекту, зокрема	3,4,5,6,7,8,9,10,11,12

	статистичної, нейромережевої та нечіткої обробки даних, методів машинного навчання та генетичного програмування тощо.	
СК 11	Здатність до інтелектуального аналізу даних на основі методів обчислювального інтелекту включно з великими та погано структурованими даними, їхньої оперативної обробки та візуалізації результатів аналізу в процесі розв'язування прикладних задач.	5,6,7,8,9,10,11,12
<i>Програмні результати навчання за освітньо-професійною програмою</i>		
ПР 4	Використовувати методи обчислювального інтелекту, машинного навчання, нейромережевої та нечіткої обробки даних, генетичного та еволюційного програмування для розв'язання задач розпізнавання, прогнозування, класифікації, ідентифікації об'єктів керування тощо.	3,4,5,6,7,9,10,11,12
ПР 12	Застосовувати методи та алгоритми обчислювального інтелекту та інтелектуального аналізу даних в задачах класифікації, прогнозування, кластерного аналізу, пошуку асоціативних правил з використанням програмних інструментів підтримки багатовимірного аналізу даних на основі технологій DataMining, TextMining, WebMining.	5,6,7,8,9,10,11,12
<i>Інформаційні системи та технології (ОС бакалавр)</i>		
<i>Загальні компетентності за освітньо-професійною програмою</i>		
КЗ 5	Здатність вчитися і оволодівати сучасними знаннями.	1,2
КЗ 6	Здатність до пошуку, оброблення та узагальнення інформації з різних джерел.	1,2,3
<i>Фахові компетентності за освітньо-професійною програмою</i>		
КС 4	Здатність проектувати, розробляти та використовувати засоби реалізації інформаційних систем, технологій та інфокомунікацій (методичні, інформаційні, алгоритмічні, технічні,	1,2, 3,8,9,11

	програмні та інші).	
КС 6	Здатність використовувати сучасні інформаційні системи та технології (виробничі, підтримки прийняття рішень, інтелектуального аналізу даних та інші), методики й техніки кібербезпеки під час виконання функціональних завдань та обов'язків.	3,4,5,6,7,8
<i>Програмні результати навчання за освітньо-професійною програмою</i>		
ПР 6	Демонструвати знання сучасного рівня технологій інформаційних систем, практичні навички програмування та використання прикладних і спеціалізованих комп'ютерних систем та середовищ з метою їх запровадження у професійній діяльності.	5,6,7,8,9,10,11,12
ПР 7	Обґрунтовувати вибір технічної структури та розробляти відповідне програмне забезпечення, що входить до складу інформаційних систем та технологій.	4,5,7,8,9,10,11,12

#### 4. ЗМІСТ ДИСЦИПЛІНИ

##### **Тема 1. Загальні поняття дисципліни «Машинне навчання» (Machine learning).**

Теорія обчислювального навчання. Типи навчання. Індуктивне навчання на основі емпіричних даних (навчальній вибірці). Дедуктивне навчання на основі формалізації знань експертів. Типи машинного навчання: навчання з вчителем, навчання без вчителя, навчання з підкріпленням. Задачі машинного навчання: кластеризація, регресія, зниження розмірності.

##### **Список рекомендованих джерел**

*Основний:* 1[с.3-25],2[с.4-10], 3[с.3-7], 4[с. 23-135]

*Додатковий:*2[с.19-36]

*Інтернет-ресурси:* 1,4

##### **Тема 2. Введення в машинне навчання на мові програмування Python.**

Машинне навчання та Python. Переваги використання Python в процедурах машинного навчання. Інсталяція Python. Пакет прикладних програм Anaconda X.X. Загальні поняття мови програмування Python. Вбудовані бібліотеки Python - Jupyter Notebook, NumPy, SciPy, Matplotlib, Pandas, Mglearn. Бібліотека Scikit-learn – призначення та інсталяція. Бібліотека TensorFlow – призначення та інсталяція. Бібліотека Theano – призначення та інсталяція. Бібліотека CUDA – призначення та

інсталяція. Бібліотека Keras – призначення та інсталяція. Бібліотека PyTorch – призначення та інсталяція.

### **Список рекомендованих джерел**

*Основний:* 2[с.12-18], 3[с. 7-23]

*Додатковий:* 2[с.43-96]

*Інтернет-ресурси:* 2,3,6

### **Тема 3. Етапи розробки моделі машинного навчання.**

Етап підготовки (представлення) даних. Ознакове описання об'єкту. Вибір ключових ознак. Етап конструювання алгоритму (вибір типу задачі). Етап тренування та тестування моделі. Перенавчання. Тренувальна та тестувальна вибірки на прикладі нейронної мережі. Типи ознак (бінарні, номінальні, порядкові, кількісні). Етап валідації алгоритму на тестових даних. Функціонал якості, квадратична та абсолютна похибки. Мінімізація помилки. Зміна сили зв'язків.

### **Список рекомендованих джерел**

*Основний:* 1[с.133-145],3[с.17-23]

*Додатковий:* 2[с.515-517],2[с.57-77]

*Інтернет-ресурси:* 1,4,6

### **Тема 4. Методи вирішення типових задач в машинному навчанні.**

Візуалізація даних (використання бібліотеки Matplotlib). Лінійна алгебра (вектори, матриці, тензори). Статистика (Опис одиночного набору даних. Показники центру розподілу. Показники варіації. Кореляція. Парадокс Сімпсона). Теорія ймовірностей (Залежність і незалежність. Умовна ймовірність. Теорема Байєса. Випадкові величини. Неперервні розподіли. Нормальний розподіл. Центральна гранична теорема). Гіпотеза і висновок. Перевірка статистичних гіпотез. Р-значення. Довірчі інтервали. Підгонка Р-значення. Проведення А/В-тестування. Байєсівський статистичний висновок. Лінійна регресія. Градієнтний спуск. Логістична регресія. Навчання логістичної регресії. Софт-макс регресія.

### **Список рекомендованих джерел**

*Основний:* 1[с.251-305],2[с.12-18], 3[с. 7-23]

*Додатковий:* 2[с.43-96]

*Інтернет-ресурси:* 2,3,4

### **Тема 5. Машинне навчання з учителем.**

Класифікація і регресія. Узагальнююча здатність, перенавчання та недонавчання. Взаємозв'язок між складністю моделі і розміром набору даних. Алгоритми машинного навчання з учителем: набори даних, метод k найближчих сусідів, лінійні моделі, наївні байєсовські класифікатори, дерева рішень. Ансамблі дерев рішень, метод опорних векторів, нейронні мережі.

Оцінки невизначеності для класифікаторів. Функція рішень. Прогнозування ймовірностей. Невизначеність в мультикласовій класифікації.

#### **Список рекомендованих джерел**

*Основний: 3[с.70-94]*

*Додатковий: 3[с.36-56]*

*Інтернет-ресурси: 4*

#### **Тема 6. Машинне навчання без учителя.**

Типи машинного навчання без учителя. Проблеми машинного навчання без учителя. Попередня обробка даних і масштабування. Різні види попередньої обробки даних. Застосування перетворень даних. Масштабування навчального і тестового наборів. Вплив попередньої обробки даних на машинне навчання без учителя. Зниження розмірності, виділення ознак і множинне навчання. Аналіз головних компонент (PCA). Факторизація невід'ємних матриць (NMF). Множинне навчання за допомогою алгоритму t-SNE. Кластеризація. Кластеризація k-середніх. Агломеративна кластеризація. Кластеризація DBSCAN. Порівняння і оцінка якості алгоритмів кластеризації.

#### **Список рекомендованих джерел**

*Основний: 3[с.168-208]*

*Додатковий: 3[с.5-35]*

*Інтернет-ресурси: 4*

#### **Тема 7. Машинне навчання з підкріпленням.**

Навчання з підкріпленням. Етапи та основні проблеми. Функція підкріплення. Середовища та ресурси. Кластеризація та зменшення розмірності в машинному навчанні з підкріпленням. Автокодувальники.

#### **Список рекомендованих джерел**

*Основний: 3[с.131-208]*

*Додатковий: 1[с. 140],3[с.5-33]*

*Інтернет-ресурси: 4*

#### **Тема 8. Нейронні мережі, як найпоширеніший метод машинного навчання.**

Нейронні мережі. Перцептрони. Нейронні мережі прямого розповсюдження. Нейронні мережі зі згорткою (конволютивні). Мах-пулінг. Процес навчання нейронної мережі методом зворотного поширення помилки. Оцінка помилки. Зворотне поширення помилки. Посилення зв'язків. Зміна величини кроку. Адаптивна зміна величини кроку. Локальні мінімуми. VDG-архітектура нейронної мережі. Нейронна мережа Елмана. Архітектура рекурентної нейронної мережі Елмана, приклад роботи. LSTM-архітектура нейронних мереж, приклад роботи.

#### **Список рекомендованих джерел**

*Основний: 1[с.172-200], 3[с.276-296]*

*Додатковий: 3[с. 64-70]*



*Інтернет-ресурси: 6,7*

### **Тема 9. Типи даних і конструювання ознак.**

Категоріальні змінні. Пряме кодування (dummy-змінні). Біннінг, дискретизація, лінійні моделі і дерева. Взаємодії і поліноми. Одномірні нелінійні перетворення. Автоматичний відбір ознак. Одномірні статистики. Відбір ознак на основі моделі. Ітеративний відбір ознак. Застосування експертних знань.

#### **Список рекомендованих джерел**

*Основний: 3[с.70-94]*

*Додатковий: 3[с.36-44]*

*Інтернет-ресурси: 2,3*

### **Тема 10. Оцінка і поліпшення якості моделі.**

Перехресна перевірка. Перехресна перевірка в scikit-learn. Переваги перехресної перевірки. Стратифікована k-блокова перехресна перевірка та інші стратегії. Решітчастий пошук. Небезпека перенавчання параметрів і перевірочний набір даних. Решітчастий пошук з перехресної перевіркою. Метрики якості моделі і їх обчислення. Метрики для бінарної класифікації. Метрики для мультикласової класифікації. Метрики регресії. Використання метрик оцінки для відбору моделі.

#### **Список рекомендованих джерел**

*Основний: 1 [с.71-121], 2[с.35-43], 3[с.251-302]*

*Додатковий: 2[с.558-571]*

*Інтернет-ресурси: 4*

### **Тема 11. Об'єднання алгоритмів в ланцюги та конвеєри.**

Відбір параметрів з використанням попередньої обробки. Побудова конвеєрів. Використання конвеєра, поміщеного в об'єкт GridSearchCV. Загальний інтерфейс конвеєра. Метод побудови конвеєрів за допомогою функції `make_pipeline`. Робота з атрибутами етапів. Оптимальні параметри етапів конвеєра. Вибір оптимальної моделі конвейера.

#### **Список рекомендованих джерел**

*Основний: 3[с.305-320]*

*Додатковий: 3[с.57-77]*

*Інтернет-ресурси: 1,2*

### **Тема 12. Робота з текстовими і графічними даними.**

Строкові типи даних. Приклад застосування: аналіз тональності кіно відгуків. Подання текстових даних у вигляді «кошика слів». Застосування моделі «кошика слів» до синтетичного набору даних. Модель «мішка слів» для кіноотзивов. Стоп-слова. Масштабування даних за допомогою TensorFlow-idf. Дослідження коефіцієнтів моделі. Модель «кошика слів» для послідовностей з декількох слів (n-

грам). Просунута токенізація, стемінг і лематизація. Моделювання тем і кластеризація документів. Латентне розміщення Діріхле. Введення в обробку зображень. Завантаження та показ зображень. Бінаризація. Гаусове розмивання. Проста класифікація зображень. Обчислення ознак по зображенню. Створення власних ознак. Використання ознак для пошуку подібних зображень. Локальні представлення ознак. Прикладна задача: конвеєр розпізнавання облич. Ознаки в методі HOG. Метод HOG - простий детектор обличь.

### Список рекомендованих джерел

*Основний: 1[с.234-237]*

*Додатковий: 3[с.71-77]*

*Інтернет-ресурси: 7*

## 5. СПИСОК РЕКОМЕНДОВАНИХ ДЖЕРЕЛ

### Основний:

1. Chollet F. Deep Learning with Python / F. Chollet – Manning Publications Co., 2021 – 504 p.
2. Galea A. Beginning Data Science with Python and Jupyter / A. Galea. – Packt Publishing Ltd, 2018. – 194 p.
3. Muller A.C. Introduction to Machine Learning with Python / A.C. Muller, S. Guido. – Published by O'Reilly Media, Inc., 2017. – 376 p.
4. Falk K. Practical Recommender Systems / K. Falk. – Manning, 2019. – 432 p.

### Додатковий:

1. Pursky O.I. Identifying customer segments in e-trade with using system analysis and clustering methods: Monograph / O.I. Pursky. – Agenda Publishing House, Coventry, United Kingdom, 2018. - 140 p.
2. Johansson R. Numerical Python: Scientific Computing and Data Science Applications with Numpy, SciPy and Matplotlib / R.Johansson. – Urayasu-shi, Chiba, Japan, – 2019. – 709p.
3. Олещенко Л.М. Машинне навчання: комп'ютерний практикум: навч. посіб. / Л.М. Олещенко. Київ: КПІ ім. Ігоря Сікорського, 2022. – 92 с.

### Інтернет-ресурси:

1. Jupyter [Електронний ресурс]. – Режим доступу: <https://jupyter.org/>
2. NumPy [Електронний ресурс]. – Режим доступу: <https://numpy.org/>
3. Pandas [Електронний ресурс]. – Режим доступу: <https://pandas.pydata.org/>
4. Scikit-learn [Електронний ресурс]. – Режим доступу: <https://scikit-learn.org/stable/>
5. Surprise [Електронний ресурс]. – Режим доступу: <https://surpriselib.com/>
6. Tensorflow [Електронний ресурс]. – Режим доступу: <https://www.tensorflow.org/>
7. Keras [Електронний ресурс]. – Режим доступу: <https://keras.io/>

\*- Курсивом виділені джерела, що є в бібліотеці ДТЕУ