

**КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ 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



**МАШИННЕ НАВЧАННЯ /
MACHINE LEARNING**

**ПРОГРАМА /
COURSE SUMMARY**

Київ 2020

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

Автори: О.І. ПУРСЬКИЙ, доктор фізико-математичних наук, професор,
С.М. Шклярський, канд. техн. наук, доцент
А.В. Селіванова, старший викладач.

Програму розглянуто і затверджено на засіданні кафедри комп'ютерних наук та інформаційних систем 11.05.2020р., протокол № 9

Рецензенти: Г.Т. САМОЙЛЕНКО, кандидат фізико-математичних наук, доцент.
А. А. Роскладка, доктор економічних наук, професор
Волохов В.М., 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», 124 «Системний аналіз» та 126 «Інформаційні системи та технології», спеціалізацій «Комп'ютерні науки», «Інформаційні технології та бізнес-аналітика (Data Science)» та «Інформаційні системи та технології».

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

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	1,2,5
ЗК 9	Здатність працювати в команді.	1
<i>Фахові компетентності за освітньо-професійною програмою</i>		
СК 2	Здатність до виявлення статистичних закономірностей недетермінованих явищ, застосування методів	3,4,5,6,7,8,9,10,11

	обчислювального інтелекту, зокрема 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
<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
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

Інформаційні технології та бізнес-аналітика (Data Science) (ОС бакалавр)

<i>Загальні компетентності за освітньо-професійною програмою</i>		
К09	Здатність до адаптації та дії в новій ситуації	1,2,3
К11	Здатність генерувати нові ідеї (креативність)	2,4,6
К20	Здатність визначати основні чинники, які впливають на розвиток фізичних, економічних, соціальних процесів, виокремлювати в них стохастичні та	5,6,8,9

	невизначені показники, формулювати їх у вигляді випадкових або нечітких величин, векторів, процесів та досліджувати залежності між ними.	
K22	Здатність до комп'ютерної реалізації математичних моделей реальних систем і процесів; проектувати, застосовувати і супроводжувати програмні засоби моделювання, прийняття рішень, оптимізації, обробки інформації, інтелектуального аналізу даних.	1,6,10
K23	Здатність використовувати сучасні інформаційні технології для комп'ютерної реалізації математичних моделей та прогнозування поведінки конкретних систем а саме: об'єктно-орієнтований підхід при проектуванні складних систем різної природи, прикладні математичні пакети, застосування баз даних і знань.	3,4,5,6,8
K28	Здатність розуміти та уміло використовувати теорію і методи Data Science.	5,7,9,10
K30	Здатність використовувати програмне забезпечення для аналізу даних (мови програмування, аналітичні платформи) з метою проведення математичних та методологічних досліджень	4,11
<i>Програмні результати навчання за освітньо-професійною програмою</i>		
ПР01	Знати і вміти застосовувати на практиці диференціальне та інтегральне числення, ряди та інтеграл Фур'є, аналітичну геометрію, лінійну алгебру та векторний аналіз, функціональний аналіз та дискретну математику в обсязі, необхідному для вирішення типових завдань системного аналізу	1,2,3
ПР04	Знати та вміти застосовувати базові методи якісного аналізу та інтегрування звичайних диференціальних рівнянь і систем, диференціальних рівнянь в частинних похідних, в тому числі	2,3,7

	рівнянь математичної фізики.	
ПР12	Застосовувати методи і засоби роботи з даними і знаннями, методи математичного, логіко-семантичного, об'єктного та імітаційного моделювання, технології системного і статистичного аналізу.	4,5,9
ПР14	Розуміти і застосовувати на практиці методи статистичного моделювання і прогнозування, оцінювати вихідні дані.	2,6,7,9
ПР18	Володіти достатніми знаннями математичних моделей і методів аналітики даних, мов моделювання та програмних засобів для виконання практичних завдань бізнес-аналізу.	5,6,8
ПР19	Володіти математичними методами розробки та дослідження алгоритмів розв'язування задач бізнес-аналітики, моделювання об'єктів і процесів, розробки алгоритмів функціонування систем.	5,8,10,11

Інформаційні системи та технології (ОС бакалавр)

<i>Загальні компетентності за освітньо-професійною програмою</i>		
КЗ 1	Здатність до абстрактного мислення, аналізу та синтезу.	1,2,3
КЗ 2	Здатність застосовувати знання у практичних ситуаціях.	1,3,5,6,7
КЗ 3	Здатність до розуміння предметної області та професійної діяльності.	1,2
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
КЗ 8	Здатність оцінювати та забезпечувати якість виконуваних робіт.	1,5,6,7
<i>Фахові компетентності за освітньо-професійною програмою</i>		
КС 3	Здатність до проектування, розробки, налагодження та вдосконалення системного, комунікаційного та програмно-апаратного забезпечення інформаційних систем та технологій,	2,5,6,7

	Інтернету речей (IoT), комп'ютерно-інтегрованих систем та системної мережної структури, управління ними.	
КС 4	Здатність проектувати, розробляти та використовувати засоби реалізації інформаційних систем, технологій та інфокомунікацій (методичні, інформаційні, алгоритмічні, технічні, програмні та інші).	4,5,6,7,8,9,10,11
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.	4,5,6,7,8
КС 13	Здатність проводити обчислювальні експерименти, порівнювати результати експериментальних даних і отриманих рішень.	3,4,5,6,7,8,9,10,11,12
<i>Програмні результати навчання за освітньо-професійною програмою</i>		
ПР 1	Знати лінійну та векторну алгебру, диференціальне та інтегральне числення, теорію функцій багатьох змінних, теорію рядів, диференціальні рівняння для функції однієї та багатьох змінних, операційне числення, теорію ймовірностей та математичну статистику в обсязі, необхідному для розробки та використання інформаційних систем, технологій та інфокомунікацій, сервісів та інфраструктури організації.	1,3,4,11
ПР 2	Застосовувати знання фундаментальних і природничих наук, системного аналізу та технологій моделювання, стандартних алгоритмів та дискретного аналізу при розв'язанні задач проектування і використання інформаційних систем та технологій.	2,3
ПР 3	Використовувати базові знання інформатики й сучасних інформаційних систем та технологій, навички	2,3,4,5,6,7,8,9,10

	програмування, технології безпечної роботи в комп'ютерних мережах, методи створення баз даних та інтернет-ресурсів, технології розроблення алгоритмів і комп'ютерних програм мовами високого рівня із застосуванням об'єктно-орієнтованого програмування для розв'язання задач проектування і використання інформаційних систем та технологій.	
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

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: 1,2

Додатковий:4

Інтернет-ресурси: 2

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, Mlglearn. 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: 1,4

Додатковий: 3

Інтернет-ресурси: 1

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: 1,2

Додатковий: 1,2

Інтернет-ресурси: 4

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

Додатковий: 2

Інтернет-ресурси: 2

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: 1,2

Додатковий: 2,4

Інтернет-ресурси: 2

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: 1,2,3

Додатковий: 1,4

Інтернет-ресурси: 2

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: 1,5

Додатковий: 4

Інтернет-ресурси: 2

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

Нейронні мережі. Перцептрони. Нейронні мережі прямого розповсюдження. Нейронні мережі зі згортою (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: 1,2,3

Додатковий: 2,3

Інтернет-ресурси: 2,5

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: 1,2

Додатковий: 2,4

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: 1,2

Додатковий: 3,4

Інтернет-ресурси: 2,4

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: 1,2

Додатковий: 3,4

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 токенизація, 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: 1,2,3

Додатковий: 3,4

Інтернет-ресурси: 1,2

5. СПИСОК РЕКОМЕНДОВАНИХ ДЖЕРЕЛ

Основний:

1. Muller A.C. Introduction to Machine Learning with Python / A.C. Muller, S. Guido. — Published by O'Reilly Media, Inc., 2017. — 376 p.
2. Коельо Л.П. Побудова систем машинного навчання на мові Python / Л.П. Коельо, В. Річарт. – К.: Видавнича група BHV, 2016. – 302 с.
3. Жерон О. Прикладне машинне навчання за допомогою Scikit-Learn і TensorFlow: концепції, інструменти і техніки для створення інтелектуальних систем / О. Жерон. – К: Діалектика, 2018. – 688 с.
4. Lutz M. Learning Python / M. Lutz. – 4th. Ed. – O'Reilly Media, 2009. – 1280 p.
5. Raschka S. Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow / S. Raschka, V. Mirjalili. – 2nd Ed. – Packt Publishing, 2017. – 622 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. Hastie T. The Elements of Statistical Learning / T. Hastie, R. Tibshirani, J. Friedman. – Springer, 2014. – 739 p.
3. McKinney W. Python for Data Analysis: Data Wrangling with Pandas, NumPy, and Ipython / W. McKinney. – 2nd. Ed. – O'Reilly Media, 2017. – 550 p.
4. Bishop C.M. Pattern Recognition and Machine Learning / C.M. Bishop. – New York: Springer, 2006. – 738 p.

Інтернет-ресурси:

1. Interactive course. Deep Learning in Python [Електронний ресурс]. – Режим доступу: <https://www.datacamp.com/courses/deep-learning-in-python>.
2. Machine Learning [Електронний ресурс]. – Режим доступу: <https://www.kaggle.com/learn/machine-learning>.
3. Datasets [Електронний ресурс]. – Режим доступу: <https://www.kaggle.com/datasets>.

4. BrainBasket Foundation. Відео-курс “Intro to Data Science” [Електронний ресурс].
– Режим доступу: <https://brainbasket.org/data-science-3/>.
5. Машинне навчання і нейронні мережі: бібліотека PHP-ML [Електронний ресурс].
– Режим доступу: <http://echo.lviv.ua/dev/5469>.

*- Курсивом виділені джерела, що є в бібліотеці КНТЕУ