

МІЖРЕГІОНАЛЬНА АКАДЕМІЯ УПРАВЛІННЯ ПЕРСОНАЛОМ



МАУП

Факультет комп'ютерно-інформаційних технологій

Кафедра комп'ютерних інформаційних систем та технологій

Затверджую:

Декан ФКІТ

Чолишкіна О.Г.

“20” серпня 2020 р.



Схвалено на засіданні кафедри

Протокол № 12 від 20 серпня 2020 р.

Завідувач кафедри  Кавун С.В.

СИЛАБУС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Інтелектуальний аналіз даних

спеціальності: 122 Комп'ютерні науки
(шифр і назва спеціальності)

освітнього рівня Бакалавр
(назва освітнього рівня, ОКР)

освітньої програми: Комп'ютерні науки
(шифр і назва освітньої програми)

спеціалізації: _____
(за наявності) (назва спеціалізації)

Київ МАУП 2020

Розробник (-и) силябусу навчальної дисципліни:

Кавун Сергій Віталійович, доктор технічних наук, професор кафедри комп'ютерних-інформаційних систем та технологій.

Викладач:

Кавун Сергій Віталійович, доктор технічних наук, професор кафедри комп'ютерних-інформаційних систем та технологій.

Силябус розглянуто і затверджено на засіданні кафедри КІСТ
Протокол від 20 серпня 2020 № 12

Завідувач кафедри _____ Кавун С.В.
(підпис)

Силябус погоджено з гарантом (керівником) освітньої програми 122
Комп'ютерні науки
(шифр, назва освітньої програми)

20.08. 2020 р.

Керівник (гарант) освітньої програми _____
(підпис)

Пролонговано:

на 20__/20__ н.р. _____ (_____), «__» 20__ р., протокол № ____
(підпис) (ПІБ)

на 20__/20__ н.р. _____ (_____), «__» 20__ р., протокол № ____
(підпис) (ПІБ)

на 20__/20__ н.р. _____ (_____), «__» 20__ р., протокол № ____
(підпис) (ПІБ)

на 20__/20__ н.р. _____ (_____), «__» 20__ р., протокол № ____

ПрАТ «ВНЗ «Міжрегіональна Академія управління персоналом» <http://maup.com.ua/>
 Факультет комп'ютерно-інформаційних технологій <http://itmaup.com.ua>
 Кафедра комп'ютерних інформаційних систем та технологій
http://itmaup.com.ua/?page_id=589

Назва дисципліни	<i>Інтелектуальний аналіз даних</i>
Викладач (-і)	<i>Кавун С.В.</i>
Портфоліо викладача (-ів)	http://itmaup.com.ua/?page_id=589
Контактний тел.	+380677095577
E-mail:	
Сторінка дисципліни на сайті	http://itmaup.com.ua/?page_id=1725
Консультації	<i>Пт: 14:20-15:50 корпус 24, аудиторія 10</i>

1. Коротка анотація до дисципліни. Курс призначений для вивчення основ інтелектуального аналізу даних та машинного навчання, передбачає застосування сучасних інструментальних засобів аналізу даних для вирішення завдань класифікації, регресії, пошуку асоціацій та послідовності (послідовної асоціації) процесів у реальному часі, які відбуваються на об'єктах дослідження.

Обговорюються відмінності Data Mining від класичних статистичних методів аналізу й OLAP-систем. Описується сфера застосування Data Mining. Докладно розглядаються методи, інструментальні засоби та застосування Data Mining у сучасних програмних продуктах для різноманітних сфер діяльності. Вивчається бібліотека мови програмування Python Scikit-Learn.

На практичних та лабораторних заняттях за допомогою засобів бібліотеки Scikit-Learn вирішуються задачі класифікації, регресії та прогнозування на реальних відкритих наборах даних з таких джерел, як Kaggle, UCI Machine Learning Repository, VisualData, Amazon, Google's Datasets Search Engine, Computer Vision Datasets, Public Government Datasets for Machine Learning

2. Мета та завдання дисципліни. Метою викладання навчальної дисципліни є вивчення методів інтелектуального аналізу даних (Knowledge Discovery in Data, KDD), аналітичного дослідження великих масивів інформації з метою виявлення нових раніше невідомих, практично корисних знань і закономірностей, необхідних для прийняття рішень; огляд методів, програмних продуктів і різних інструментальних засобів, які використовуються KDD; розгляд практичних прикладів застосування KDD на реальних відкритих наборах даних.

Цілі курсу:

- опанувати базові принципи побудови моделей даних;
- ознайомитися з концепцією Knowledge Discovery in Data (виявлення знань в даних);
- навчитися ефективно використовувати методи здобуття знань з великих масивів даних;
- ознайомитися з основними типами задач, що можуть бути вирішені за допомогою методів інтелектуального аналізу даних;

- отримати практичні навички з використання інструментальних засобів інтелектуального аналізу даних при вирішенні прикладних задач та навчитися інтерпретувати отримані результати.

3. Формат курсу

Курс передбачає теоретичні та практичні заняття.

У межах теоретичної частини забезпечуються знання:

- основні поняття, задачі та стадії інтелектуального аналізу даних;
- підходи до збереження, подання та обробки інформації в сучасних інформаційних системах;
- методи побудови моделей та аналізу залежностей у великих масивах даних;
- основні концепції маніпуляції з даними та формування датасетів;
- критерії порівняння моделей і методів інтелектуального аналізу даних;
- сучасні програмні засоби для проектування й розробки систем інтелектуального аналізу даних

Проходження практичної частини курсу формує вміння:

- проводити найбільш раціональний відбір методів аналізу даних для вирішення конкретних завдань;
- виконувати формалізацію задачі на базі змістовного опису;
- проводити необхідну попередню обробку даних;
- визначати тип задачі аналізу, вирішувати її адекватно обраним методом з оптимально визначеними параметрами;
- оцінювати результати, робити змістовні висновки та інтерпретацію;
- використовувати сучасні програмні засоби (зокрема дистрибутив Anaconda, бібліотеки NumPy, Pandas, Scikit-Learn) для проектування та дослідження систем / модулів інтелектуального аналізу даних, які використовуються під час досліджень в R&D центрах ІТ-компаній.

4. Програмні результати навчання (інтегральні, фахові компетентності).

Вивчення навчальної дисципліни передбачає формування та розвиток у студентів компетентностей:

загальних:

- здатність до абстрактного мислення, аналізу та синтезу
- здатність застосовувати знання у практичних ситуаціях;
- знання та розуміння предметної області та розуміння професійної діяльності;
- здатність вчитися й оволодівати сучасними знаннями;
- здатність до пошуку, оброблення та аналізу інформації з різних джерел
- здатність генерувати нові ідеї (креативність);
- здатність працювати в команді;
- здатність приймати обґрунтовані рішення;
- здатність оцінювати та забезпечувати якість виконуваних робіт.

фахових:

- здатність до виявлення статистичних закономірностей недетермінованих явищ, застосування методів обчислювального інтелекту, зокрема статистичної, нейромережевої та нечіткої обробки даних, методів машинного навчання та генетичного програмування тощо;
- здатність до логічного мислення, побудови логічних висновків, використання формальних мов і моделей алгоритмічних обчислень, проектування, розроблення й аналізу алгоритмів, оцінювання їх ефективності та складності, розв'язності та нерозв'язності алгоритмічних

проблем для адекватного моделювання предметних областей і створення програмних та інформаційних систем;

- здатність використовувати сучасні методи математичного моделювання об'єктів, процесів і явищ, розробляти моделі й алгоритми чисельного розв'язування задач математичного моделювання, враховувати похибки наближеного чисельного розв'язування професійних задач;
- здатність до системного мислення, застосування методології системного аналізу для дослідження складних проблем різної природи, методів формалізації та розв'язування системних задач, що мають суперечливі цілі, невизначеності та ризики;
- здатність застосовувати теоретичні та практичні основи методології та технології моделювання для дослідження характеристик і поведінки складних об'єктів і систем, проводити обчислювальні експерименти з обробкою й аналізом результатів;
- здатність проєктувати та розробляти програмне забезпечення із застосуванням різних парадигм програмування: узагальненого, об'єктно-орієнтованого, функціонального, логічного, з відповідними моделями, методами й алгоритмами обчислень, структурами даних і механізмами управління;
- 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;
- здатність використовувати методи Data Mining для аналізу та структуризації багатовимірних даних і подальшого їхнього використання при розв'язанні прикладних задач.

Результати навчання цієї дисципліни деталізують такі програмні результати навчання:

- застосовувати знання основних форм і законів абстрактно-логічного мислення, основ методології наукового пізнання, форм і методів вилучення, аналізу, обробки та синтезу інформації в предметній області комп'ютерних наук;
- використовувати сучасний математичний апарат неперервного та дискретного аналізу, лінійної алгебри, аналітичної геометрії, в професійній діяльності для розв'язання задач теоретичного та прикладного характеру в процесі проєктування та реалізації об'єктів інформатизації;
- використовувати знання закономірностей випадкових явищ, їх властивостей та операцій над ними, моделей випадкових процесів та сучасних програмних середовищ для розв'язування задач статистичної обробки даних і побудови прогнозних моделей;
- 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;
- проєктувати, розробляти та аналізувати алгоритми розв'язання обчислювальних та логічних задач, оцінювати ефективність та складність алгоритмів на основі застосування формальних моделей алгоритмів та обчислюваних функцій;
- розробляти програмні моделі предметних середовищ, вибирати парадигму програмування з позицій зручності та якості застосування для реалізації методів та алгоритмів розв'язання задач в галузі комп'ютерних наук;
- 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. Опис навчальної дисципліни :

Найменування показників	Характеристика дисципліни за формами навчання	
	денна	заочна
<i>Інтелектуальний аналіз даних</i>		
Курс	4	4
Семестр	7	7
Обсяг кредитів	3	3
Обсяг годин, в тому числі:	90	90
Аудиторні	34	6
Модульний контроль	2	-
Семестровий контроль	2	2
Самостійна робота	56	84
Форма семестрового контролю	<i>іспит</i>	<i>іспит</i>

6. Статус дисципліни: обов'язкова

7. Пререквізити – вивчення дисципліни ґрунтується на знаннях дисциплін організація баз даних та знань та математичні методи дослідження операцій.

8. Технічне й програмне забезпечення /обладнання - у якості програмних засобів рекомендується використовувати дистрибутив Anaconda (www.anaconda.com), бібліотеки NumPy (numpy.org), Pandas (pandas.pydata.org), Scikit-Learn (scikit-learn.org). Також можна використовувати засоби Google Colaboratory (colab.research.google.com) або IDE PyCharm (www.jetbrains.com/pycharm)- додатково вимагає інсталяції всіх необхідних Python бібліотек окремо. Ознайомитися із наявними в Інституті програмами і сервісами можна: ауд. №5: Тищенко Максим Вадимович.

9. Політика курсу - Для успішного проходження курсу та складання контрольних заходів необхідним є виконання таких обов'язків:

- не запізнюватися на заняття;
- не пропускати заняття (як лабораторні, практичні так і лекційні), у разі хвороби мати довідку або її ксерокопію;
- самостійно опрацьовувати весь лекційний матеріал та ресурси для самостійної роботи;
- конструктивно підтримувати зворотній зв'язок з викладачем на всіх етапах проходження курсу (особливо під час виконання курсового проєкту);
- своєчасно і самостійно виконувати всі передбачені програмою лабораторні та практичні завдання;
- не користуватися мобільним телефоном під час аудиторних занять;
- брати очну участь у контрольних заходах;
- будь-яке відтворення результатів чужої праці (виключаючи практичну роботу над командним проєктом), в тому числі використання завантажених з Інтернету матеріалів, як власних результатів кваліфікується як порушення норм і правил академічної доброчесності та передбачає притягнення до відповідальності у порядку, визначеному чинним законодавством.

10. Схема курсу (Навчально-методична карта дисципліни)

НАВЧАЛЬНО-МЕТОДИЧНА КАРТА ДИСЦИПЛІНИ «Інтелектуальний аналіз даних»

Разом: 90 год., з них 17 год. – лекції, 17 год. – практична робота; самостійна робота – 56 год., модульний контроль – 2 год

Кількість балів за семестр	100					
Модулі	Змістовий модуль I					
Назва модуля	Основи бібліотеки Scikit-Learn					
Лекції	1	2	3	4		
Теми лекцій	Тема 1. Знайомство з бібліотекою Scikit-Learn	Тема 2. Проектування ознак: категоріальні ознаки, текстові ознаки, ознаки для зображень, конвеєри ознак.	Тема 3. Гіперпараметри та перевірка моделі.	Тема 4. Постановка задачі класифікації та представлення результатів. Методи побудови правил класифікації. Методи побудови дерев рішень		
Практичні заняття	1	2	3	4		
Теми практичних занять	Основи роботи з Scikit-Learn.	ZeroR-алгоритм в Scikit-Learn. Confusion Matrix	Проектування ознак	Наївні байєсовські класифікатори в Scikit-Learn		
Поточний контроль	модульна контрольна робота №1 (20 балів)					
Модулі	Змістовий модуль II					
Назва модуля	Методи класифікації та регресії					
Лекції	5	6	7	8	9	10
Теми лекцій	Тема 5. Наївна байєсівська класифікація	Тема 6. Лінійна модель регресії та метод найменших	Тема 7 Дерева рішень. Структура дерев рішень. Алгоритм побудови	Тема 8. Метод опорних векторів та метод k -	Тема 9. Постановка задачі	Тема 10. Методи якості

		квадратів	дерева рішень. Навчання дерева рішень	найближчих сусідів	пошуку асоціативних правил та представле ння результатів . Методи пошуку асоціативн их правил: метод Аргіорі, побудова FR-дерева пошуку шаблонів даних. Секвенціал ьний аналіз.	моделі та їх обчислен ня
Практичні заняття	5 6	7 8	9 10	11 12	13	14
Теми практичних занять	К-блочна крос- перевірка засобами Scikit-Learn	Задача лінійної регресії в Scikit- Learn. Задача логістичної регресії в Scikit-Learn	Дерева рішень в Scikit- Learn	Метод опорних векторів. Метод k - найближчих сусідів	Методи пошуку асоціативн их правил	Секвенція льний аналіз
Поточний контроль	модульна контрольна робота №1 (20 балів)					
Підсумковий контроль	Іспит					

11. Система оцінювання навчальних досягнень

Вид діяльності студента	Максимальна кількість балів за одиницю	Модуль 1		Модуль 2	
		кількість одиниць	максимальна кількість балів	кількість одиниць	максимальна кількість балів
Відвідування лекцій	1	4	4	6	6
Відвідування практичних занять	1	4	4	6	6
Виконання практичних завдань	10	4	40	6	60
Виконання завдань для самостійної роботи (додавати завдання за потребою)	5	1	5	1	5
Виконання модульної роботи	20	1	20	1	20
Разом за кожним модулем	-		73	-	97
Максимальна кількість балів					170
170:100=1,70. Студент набрав X балів; Розрахунок: X:1,70 = загальна кількість балів.					
Усього за навчальною дисципліною					100

12. Завдання для самостійної роботи та критерії її оцінювання

№ з/п	Зміст самостійної роботи студента	Обсяг СРС (годин)
1.	Опрацювання лекційного матеріалу	16
2.	Підготовка до практичних занять	32

3.	Підготовка до модульних контрольних робіт №1, №2	8
Усього за навчальною дисципліною		56

Критерії оцінювання:

- змістовність – 1 бал
- відповідність темі та стилю оформлення – 1 бал

13. Форми проведення модульного контролю та критерії оцінювання

Модульний контроль здійснюється шляхом написання модульної контрольної роботи.

Студентами пропонуються варіанти МКР, кожен білет містить по 2 теоретичних запитання та 2 практичних завдання.

Критерії оцінювання:

- виконання МКР – максимум 20 балів.

Критеріями оцінки правильності виконання модульних контрольних завдань є:

№ п/п	Критерії оцінки знань студента виявлених під час проведення модульного контролю	кількість балів
1.	студент в процесі письмової відповіді дає правильні відповіді на всі поставлені запитання , виявляє високий рівень знань теоретичного та практичного матеріалу. Викладає свою відповідь системно та логічно, упевнено і правильно аргументує власну позицію, робить висновки, тощо.	20 балів
2.	студент має належний рівень знань теоретичного та практичного матеріалу, на поставлені запитання відповіді дає, переважно, правильні, однак допускає певні неточності у визначеннях, не завжди належно (коректно) аргументує відповідь або правильно відповідає лише на половину поставлених запитань , тощо.	15 балів
3	студент має задовільний рівень знань теоретичного та практичного матеріалу, на поставлені запитання відповідає, але не на всі, допускає певні неточності у визначеннях базових категорій, не завжди належно (коректно) аргументує або правильно дає відповідь на 1/3 (одну третину) поставлених запитань тощо.	10 балів
4.	студент дає неправильні відповіді на поставлені запитання, виявляє неналежний рівень знань теоретичного та практичного матеріалу, неспроможний послідовно і правильно аргументувати свою відповідь або взагалі не в змозі відповісти на поставлені запитання , тощо.	0 балів

14. Форми проведення семестрового контролю та критерії оцінювання: Іспит**15. Орієнтовний перелік питань для семестрового комплексного контролю.**

1. Бібліотека Scikit-Learn.
2. Проектування ознак: категоріальні ознаки, текстові ознаки, ознаки для зображень, конвеєри ознак.
3. Гіперпараметри.

4. Методи побудови правил класифікації.
5. Методи побудови дерев рішень.
6. Наївна байєсівська класифікація
7. Лінійна модель регресії.
8. метод найменших квадратів.
9. Дерева рішень.
10. Структура дерев рішень.
11. Алгоритм побудови дерева рішень.
12. Метод опорних векторів та метод k - найближчих сусідів.
13. Методи пошуку асоціативних правил.
14. Секвенціальний аналіз.
15. Методи якості моделі та їх обчислення
16. Задача лінійної регресії в Scikit-Learn.
17. Задача логістичної регресії в Scikit-Learn
18. Дерева рішень в Scikit-Learn.
19. Методи пошуку асоціативних правил
20. Навчання дерева рішень.
21. Проектування ознак.
22. Основи бібліотеки Scikit-Learn.
23. ZeroR-алгоритм в
24. Scikit-Learn. Confusion Matrix.
25. Гіперпараметри та перевірка моделі.
26. Методи побудови правил класифікації.
27. К-блочна крос- перевірка засобами Scikit-Learn.
28. Методи якості моделі та їх обчислення

16. Шкала відповідності оцінок

Сума балів за всі види навчальної діяльності	Оцінка ECTS	Оцінка за національною шкалою	
		для екзамену, курсового проекту (роботи), практики	для заліку
90 – 100	A	відмінно	зараховано
82-89	B	добре	
75-81	C		

68-74	D	задовільно	
60-67	E		
35-59	FX	незадовільно з можливістю повторного складання	не зараховано з можливістю повторного складання
1-34	F	незадовільно з обов'язковим повторним вивченням дисципліни	не зараховано з обов'язковим повторним вивченням дисципліни

17. Рекомендовані джерела (література):

1. 32.81 Л64 Литвин, В. В. Аналіз даних та знань : навч. посіб. / В. В. Литвин, В. В. Пасічник, Ю. В. Нікольський. – Л. : Магнолія 2006, 2015. – 276 с.
2. 32.973 М33 Матвієнко, М. П. Комп'ютерна логіка : навч. посіб. / М. П. Матвієнко. – К. : Ліра-К, 2015. – 288 с.
3. 32.973 Ш32 Шаховська, Н. Б. Алгоритми і структури даних : посіб. / Н. Б. Шаховська, Р. О. Голощук. – Л. : Магнолія 2006, 2016. – 214 с
4. 1. 32.973 Б83 Боровиков, В. Statistica: искусство анализа данных на компьютере / В. Боровиков. – СПб : Питер, 2001. – 656 с.
5. 2. 65.290-21 Г29 Гейк, П. Вчись аналізувати ринок / П. Гейк, П. Джексон ; пер.з англ. – Л. : Сейбр-Світло, 1995. – 270с.
6. 3. 22.172 Г69 Горкавий, В. К. Математична статистика : Навч. посіб. / В. К. Горкавий, В. В. Ярова. – К. : Професіонал, 2004. – 384 с.
7. 4. 32.81 Л64 Литвин, В. В. Аналіз даних та знань : навч. посіб. / В. В. Литвин, В. В. Пасічник, Ю. В. Нікольський. – Л. : Магнолія 2006, 2015. – 276 с.
8. 5. 32.81 Л64 Литвин, В. В. Інтелектуальні системи : підруч. / В. В. Литвин, В. В. Пасічник, Ю. В. Яцишин. – Львів : Новий Світ-2000, 2009. – 406 с. – (Комп'ютинг).
9. 6. 32.81 Л64 Литвин, В. В. Інтелектуальні системи : підруч. / В. В. Литвин, В. В. Пасічник, Ю. В. Яцишин. – Л. : Новий Світ-2000, 2011. – 406 с.

Електронні джерела

1. <http://docs.scipy.org/doc/>
2. <http://scikit-learn.org/stable/>
3. www.anaconda.com
4. numpy.org
5. pandas.pydata.org
6. colab.research.google.com
7. www.jetbrains.com/pycharm