

Міністерство освіти і науки України

Харківський національний університет імені В.Н. Каразіна

Кафедра прикладної математики

“ЗАТВЕРДЖУЮ”

Декан факультету
математики і інформатики

Григорій ЖОЛТКЕВИЧ

“29” серпня 2024 р.



РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Алгоритми в науці про дані

рівень вищої освіти _____ другий (магістерський) рівень _____

галузь знань _____ 11 – Математика та статистика _____

спеціальність _____ 113 Прикладна математика _____

освітня програма _____ Прикладна математика _____

спеціалізація _____

вид дисципліни _____ обов'язкова _____

факультет _____ математики і інформатики _____

2024 / 2025 навчальний рік

Програму рекомендовано до затвердження Вченою радою факультету математики і інформатики

“27” серпня 2024 року, протокол № 8

РОЗРОБНИКИ ПРОГРАМИ: *Стпанова Катерина Вадимівна*, канд. фіз.-мат. наук, доцент, доцент закладу вищої освіти кафедри прикладної математики.

Програму схвалено на засіданні кафедри прикладної математики
Протокол від “26” серпня 2024 року № 8

Завідувач кафедри прикладної математики



Валерій КОРОБОВ

Програму погоджено з гарантом
освітньо-професійної програми «Прикладна математика»

Гарант освітньо-професійної програми «Прикладна математика»



Світлана ІГНАТОВИЧ

Програму погоджено з гарантом
освітньо-наукової програми «Прикладна математика»

Гарант освітньо-наукової програми «Прикладна математика»



Валерій КОРОБОВ

Програму погоджено науково-методичною комісією
факультету математики і інформатики

Протокол від “27” серпня 2024 року № 1

Голова науково-методичної комісії факультету математики і інформатики



Євген МЕНЯЙЛОВ

ВСТУП

Програма навчальної дисципліни “**Алгоритми в науці про дані**” складена відповідно до освітньо-професійної програми підготовки магістр

Спеціальності (напряму) 113 Прикладна математика
спеціалізації _____

1. Опис навчальної дисципліни

1.1. Метою викладання навчальної дисципліни є надання майбутнім спеціалістам теоретичних знань та практичних навичок з основ розробки і аналізу алгоритмів для розв’язання різних прикладних задач.

1.2. Основні завдання вивчення дисципліни:

- ознайомити студентів із основними поняттями науки даних;
- ознайомлення з основними методами аналізу алгоритмів;
- ознайомлення з деякими найбільш важливими алгоритмами;
- створення ефективних алгоритмів при розв’язанні прикладних задач.

1.3. Кількість кредитів - 6

1.4. Загальна кількість годин - 180

1.5. Характеристика навчальної дисципліни	
Обов’язкова	
Денна форма навчання	Заочна (дистанційна) форма навчання
Рік підготовки	
1-й	
Семестр	
1-й	
Лекції	
32 год.	
Практичні, семінарські заняття	
32 год.	
Лабораторні заняття	
Самостійна робота	
116 год.	
у тому числі індивідуальні завдання	

1.6. Заплановані результати навчання.

У результаті вивчення даного курсу студент повинен

знати:

- методи створення і аналізу алгоритмів, типи структур даних
- основні принципи роботи базових алгоритмів науки про дані
- синтаксис мови програмування

вміти:

- застосовувати доцільно вивчені алгоритми у різних прикладних задачах

2. Тематичний план навчальної дисципліни

Тема 1. Введення до алгоритмів науки про дані. Логістична та гребнева регресія.

Основні поняття науки даних. Схема базових алгоритмів для науки про дані та сфера їх застосування. Регресія. Гребнева регресія. Алгоритм логістичної регресії. Передбачення достовірностей подій.

Тема 2. Лінійна регресія.

Прогнозування довільних значень за допомогою лінійної регресії. Оцінка якості моделі за допомогою R2 Score.

Тема 3. Мульти-лінійна регресія.

Розподіл даних щодо вибірки на тренувальну множину та тестову множину. Означення крос-валідації. Визначення доцільних факторів для мульти-лінійної регресії та застосування цього алгоритму для прикладних задач.

Тема 4. Класифікація.

Класифікація даних за допомогою алгоритму дерева рішень. Візуалізація побудови рішень на базі впливових факторів за допомогою graphviz. Обґрунтування отриманих результатів за допомогою confusionmatrix.

Тема 5. Кластеризація.

Застосування алгоритму KMeans, що кластеризує дані. Відокремлення зразків в n групах рівної дисперсії. Збереження та загрузка побудованої моделі за допомогою модуля pickle.

3. Структура навчальної дисципліни

Назви розділів і тем	Кількість годин					
	денна форма					
	усього	у тому числі				
л		п	лаб.	інд.	с. р.	
1	2	3	4	5	6	7
Тема 1. Введення до алгоритмів для науки про дані. Логістична регресія	33	4	4			25
Тема 2. Лінійна регресія	32	6	6			20
Тема 3. Мульти-лінійна регресія	47	12	10			25
Тема 4. Класифікація	32	6	6			20
Тема 5. Кластеризація	34	4	4			26
<i>Контрольна робота</i>	2		2			
Усього годин	180	32	32			116

4. Теми практичних занять

№ з/п	Назва теми	Кількість годин
1	Передбачення достовірності позитивної та негативної події на базі алгоритму логістичної регресії за допомогою пакета sklearn. Logistic Regression. Збереження прогнозованих значень як DataFrame.	4
2	Прогнозування довільних значень за допомогою лінійної регресії. Оцінка якості моделі за допомогою R2 Score.	6
3	Розподіл даних щодо вибірки на тренувальну множину та тестову множину. Застосування крос-валідації. Визначення доцільних факторів для мульти-лінійної регресії та застосування цього алгоритму для конкретної прикладної задачі.	10
4	<i>Контрольна робота</i>	2
5	Класифікація даних за допомогою алгоритму дерева рішень. Візуалізація побудови рішень на базі впливових факторів за допомогою graphviz. Обґрунтування отриманих результатів за допомогою confusion matrix.	6
6	Застосування алгоритму KMeans, що кластеризує дані. Відокремлення з разків в n групах рівної дисперсії. Збереження та загрузка побудованої моделі за допомогою модуля pickle.	4
	Разом	32

5. Завдання для самостійної роботи

№ з/п	Види, зміст самостійної роботи	Кількість годин
1	Виконання домашніх завдань з визначенням цільових напрямків моделі даних	25
2	Виконання домашніх завдань з розробки моделі даних, яка задовольняє вимогам масштабування	20
3	Виконання домашніх завдань з побудови структури алгоритму на базі моделі даних	25
4	Виконання домашніх завдань з реалізації та аналізу одного із алгоритмів на програмній мові Python	20
5	Виконання домашніх завдань із застосуванням алгоритму KMeans, збереженням та загрузкою побудованої моделі за допомогою модуля pickle.	26
	Разом	116

6. Методи навчання

Частково-пошуковий, дослідницький, пояснювально-ілюстративний, репродуктивний. Лекційні та практичні заняття.

7. Методи контролю

- перевірка виконання домашніх завдань
- поточне опитування за лекційним матеріалом
- перевірка контрольної роботи
- перевірка звітів-результатів самостійної роботи студентів
- залік

8. Схема нарахування балів

Поточний контроль, самостійна робота, індивідуальні завдання							Залік	Сума
T1	T2	T3	T4	T5	Контрольна робота	Разом		
6	6	6	6	6	10	60	40	100

T1, ..., T5 – теми.

Критерії оцінювання

Поточний контроль:

бали нараховуються протягом курсу за виконання домашніх завдань, контрольної роботи і активність студентів під час лекцій (через їх поточне опитування) та практичних занять (де в

тому числі схвалюється, враховується та винагороджується балами їх самостійна робота й ініціативи).

За контрольну роботу бали нараховуються таким чином:

- максимальний бал у разі правильної обґрунтованої відповіді без помилок;
- за незначні помилки оцінка зменшується від 10 до 30 відсотків;
- за значні логічні помилки оцінка зменшується до 50 відсотків, якщо хід розв'язання/міркування в цілому правильний,
- у разі частково правильних міркувань за відсутності обґрунтованої відповіді виставляється до 30 відсотків від максимальної кількості балів
- якщо відповідь не відповідає жодному з критеріїв, які сформульовані вище, – виставляється 0 балів.

Для допуску до складання підсумкового контролю (заліку) здобувач вищої освіти повинен набрати не менше 10 балів з навчальної дисципліни протягом семестру.

Залік:

залікова робота складається із двох частин/блоків, що відповідає двом питанням/завданням. Максимальна оцінка за кожне – 20 балів, разом за обидва – 40 балів.

Для отримання максимальної оцінки (20 балів) за одне питання потрібно вміти сформулювати відповідні означення та/або навести алгоритм(и), надати аналіз щодо структури даних, на яких буде реалізовуватися цей алгоритм та обов'язково продемонструвати володіння синтаксисом відповідного алгоритму на прикладі(ах). Замість прикладів можуть бути наведені міркування щодо місту відповідних понять і результатів. Викладення основних концепцій та принципів роботи базового(их) алгоритму(ів) науки про дані та логіки (побудова діаграм та альтернативна візуалізація результатів) є обов'язковою.

Формулювання неповне або зі значними помилками: 5-15 балів.

Формулювання без прикладів або міркувань: 10-15 балів.

Шкала оцінювання (дворівнева)

Сума балів за всі види навчальної діяльності протягом семестру	Оцінка
50-100	зараховано
1-49	не зараховано

10. Рекомендована література

Основна література

1. [Brian Steele, John Chandler, Swarna Reddy](#). “Algorithms for Data Science”. Springer. 2016. - 430 p.

2. [Steven S Skiena](#). “The Algorithm Design Manual”. Springer Science & Business Media. 2009. - 730 p.
3. [Brian Christian, Tom Griffiths](#). “Algorithms to Live By: The Computer Science of Human Decisions”. Henry Holtand Company. 2016. - 368 p.
4. [Christopher Steiner](#). “Automate This: How Algorithms Took Over Our Markets, Our Jobs, and the World”. Penguin. 2012. - 256 p.
5. Michele Chambers, Christine Doig, Ian Stokes - Rees. “Breaking Data Science Open”. O'Reilly Media, Inc. 2017. - 78 p.

Допоміжна література

1. [Foster Provost, Tom Fawcett](#). “Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking”. O'Reilly Media, Inc. 2013 - 414 p.
2. [Cathy O'Neil, Rachel Schutt](#). “Doing Data Science: Straight Talk from the Frontline”. O'Reilly Media, Inc. 2013 - 408 p.
3. [Barbara Oakley, PhD](#). “A Mind For Numbers: How to Excel at Math and Science”. Penguin. 2014. - 336 p.
4. [Barry Schwartz](#). “The Paradox of Choice: Why More Is Less”. HarperCollins. 2009. - 304 p.
5. Прийма С. М. Теорія алгоритмів: Навчальний посібник. – Мелітополь: ФОП Однорог Т. В., 2018. – 116 с.
6. 4. Костів О.В. Вступ до інформатики: Тексти лекцій / О.В.Костів, С.А.Ярошко. – Львів:ЛНУ імені Івана Франка, 2009

10. Посилання на інформаційніресурси в Інтернеті, відео-лекції, інше мет.заб.

1. Sklearn Documentation

<http://scikit-learn.org/stable/documentation.html>

2. Learn Python

<https://www.codecademy.com>

3. Сайт Вікіпедії, сторінка про алгоритми. — <https://uk.wikipedia.org/wiki/Алгоритм>
4. Середовище Blockly. — <https://blockly-demo.appspot.com/static/demos/code/index.html?lang=uk>
5. Інформаційні моделі. Алгоритми. — <https://sites.google.com/site/kursosnoviinformatiki/lekcii/lekcia-3>
6. Алгоритмічні моделі. — [http://ksu.vntu.edu.ua/files/matmod/Part3/03_3.5\(gr\).pdf](http://ksu.vntu.edu.ua/files/matmod/Part3/03_3.5(gr).pdf)
7. Алгоритми та їх властивості. Вступ до інформатики. —

sites.google.com/site/6klasintervpravu/home/6-klas/tema-1-algoritmi-ta-ieh-vlastivosti

8. Формальні виконавці алгоритму. — <https://disted.edu.vn.ua/courses/learn/601>

9. Новотарський М.А. Алгоритми та методи обчислень [Електронний ресурс] :навч.посіб. для студ. спеціальностей 121 «Інженерія програмного забезпечення», спеціалізації «Програмне забезпечення високопродуктивних комп'ютерних систем та мереж» та 123 «Комп'ютерна інженерія», спеціалізації «Комп'ютерні системи та мережі» / М. А. Новотарський; КПІ ім. Ігоря Сікорського. – Електронні текстові дані(1 файл: 4648 Кбайт). – Київ : КПІ ім. ІгоряСікорського, 2019. – 407 с. —

https://ela.kpi.ua/bitstream/123456789/27864/1/Alhorytmy_ta_metody_obchislenn.pdf

10. Детермінований скінченний автомат. —

https://uk.wikipedia.org/wiki/Детермінований_автомат

11. Недетермінований скінченний автомат. —

https://uk.wikipedia.org/wiki/Недетермінований_скінченний_автомат