

Міністерство освіти і науки України

Харківський національний університет імені В.Н. Каразіна

Кафедра прикладної математики

“ЗАТВЕРДЖУЮ”

Проректор

в науково-педагогічній роботі

Олександр ГОЛОВКО



„серпень 2022 р.

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Алгоритми в науці про дані

(назва навчальної дисципліни)

рівень вищої освіти другий (магістерський)

галузь знань 11 – математика та статистика

спеціальність 113 – прикладна математика
(шифр і назва)

освітня програма прикладна математика
(шифр і назва)

спеціалізація _____
(шифр і назва)

вид дисципліни обов'язкова
(обов'язкова / за вибором)

факультет математики і інформатики

2022 / 2023 навчальний рік

Програму рекомендовано до затвердження Вченою радою факультету математики і інформатики

“29” серпня 2022 року, протокол № 7

РОЗРОБНИКИ ПРОГРАМИ: **Степанова Катерина Вадимівна**, кандидат фізико-математичних наук, доцент кафедри прикладної математики

Програму схвалено на засіданні кафедри Прикладної математики

Протокол від “29” серпня 2022 року №11

Завідувач кафедри Прикладної математики



(підпис)

Валерій КОРОБОВ

(прізвище та ініціали)

Програму погоджено з гарантом освітньо-професійної програми (керівником проектної групи)

Прикладна математика

назва освітньої програми

Гарант освітньо-професійної програми

(керівник проектної групи) Прикладна математика



(підпис)

Світлана ІГНАТОВИЧ

(прізвище та ініціали)

Програму погоджено з гарантом освітньо- наукової програми (керівником проектної групи)

Прикладна математика

назва освітньої програми

Гарант освітньо-наукової програми

(керівник проектної групи) Прикладна математика



(підпис)

Валерій КОРОБОВ

(прізвище та ініціали)

Програму погоджено науково-методичною комісією факультету математики і інформатики Харківського національного університету імені В.Н. Каразіна

назва факультету, для здобувачів вищої освіти якого викладається навчальна дисципліна

Протокол від “29” серпня 2022 року, протокол № 1

Голова науково-методичної комісії факультету математики і інформатики



(підпис)

Ольга АНОЩЕНКО

(прізвище та ініціали)

ВСТУП

Програма навчальної дисципліни «**Алгоритми в науці про дані**» складена відповідно до освітньо-професійної та освітньо-наукової програм підготовки

магістр

Спеціальності (напрям) 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	
56 год.	–
в 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 даних
- синтаксис мови програмування

вміти:

– застосовувати доцільно вивчені алгоритми у різних прикладних задачах

2. Тематичний план навчальної дисципліни

Тема 1. *Введення до алгоритмів науки даних. Логістична та гребнева регресії.*

Основні поняття науки даних. Схема базових алгоритмів для науки даних 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 confusion matrix.

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 даних. 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 (семінарських, лабораторних) занять

№ з/п	Назва теми	Кількість годин
1	Передбачення достовірності позитивної та негативної події на базі алгоритму логістичної регресії за допомогою пакета <code>sklearn.LogisticRegression</code> . Збереження прогнозованих значень як <code>DataFrame</code> .	4
2	Прогнозування довільних значень за допомогою лінійної регресії. Оцінка якості моделі за допомогою <code>R2 Score</code> .	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

4	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
5	Застосування алгоритму KMeans, що кластеризує дані. Відокремлення зразків 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. Індивідуальні завдання

Не передбачені

7. Методи навчання

Частково-пошуковий, дослідницький, пояснювально-ілюстративний, репродуктивний. Проблемні лекції, пояснення, практичні роботи.

8. Методи контролю

- перевірка виконання домашніх завдань

- поточне опитування за лекційним матеріалом
- перевірка індивідуальних завдань
- перевірка звітів-результатів самостійної роботи студентів
- перевірка контрольної роботи
- проведення та перевірка залікової роботи

9. Схема нарахування балів

Поточний контроль, самостійна робота, індивідуальні завдання					Контрольна робота, передбачена навчальним планом	Залікова робота	Сума
T1	T2	T3	T4	T5			
10	10	10	10	10	10	40	100

T1, T2 ... – теми.

Критерії оцінювання залікових робіт з курсу «Алгоритми в науці даних»

Білет складається з 2 теоретичних питань.

Максимальна оцінка за кожне питання 20 балів, разом за обидва - 40 балів.

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

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 Holt and 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”. Harper Collins. 2009. - 304 p.
5. Прийма С. М. Теорія алгоритмів: Навчальний посібник. – Мелітополь: ФОП Однорог Т. В., 2018. – 116 с.
6. Костів О.В. Вступ до інформатики: Тексти лекцій / О.В.Костів, С.А.Ярошко. – Львів : ЛНУ імені Івана Франка, 2009

11. Посилання на інформаційні ресурси в Інтернеті, відео-лекції, інше методичне забезпечення

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. Ігоря Сікорського, 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