

## Силабус курсу Аналіз даних

Освітньо-професійна програма: 124 «Системний аналіз»  
Ступінь вищої освіти – магістр  
Спеціальність: 124 «Системний аналіз»

Рік навчання: I, Семестр: 2

Кількість кредитів: 5 Мова викладання: українська

### Керівник курсу

ППП

д.т.н., професор Пасічник Роман Мирославович  
к.е.н., доцент Данилюк Ірина Вадимівна

Контактна інформація [roman.pasi@gmail.com](mailto:roman.pasi@gmail.com)

### Опис дисципліни

Даний курс знайомить студентів з предметною областю великих даних (big data), а також показує взаємозв'язки з наукою про дані (data science) та аналізом даних (data analytics) та забезпечує: оволодіння навичками оперування великими даними шляхом використання спеціалізованих програмно-апаратних засобів, зокрема хмарних сервісів, спеціалізованих систем зберігання даних, розподілених файлових систем, тощо; розвиток уміння працювати з великим обсягом даних з урахуванням їх ключових характеристик: обсягу, різноманітності, мінливості й забезпечення відповідного рівня швидкості опрацювання; розвиток навичок ефективного використання математичного, алгоритмічного та програмного забезпечення для розв'язки основних задач предметної області великих даних.

Студенти зможуть ефективно використовувати парадигми паралельного опрацювання даних, зокрема MapReduce та системи Apache Spark, відповідні хмарні служби Amazon Web Services та IBM Bluemix; розгортати надійні та швидкі сховища для надвеликих обсягів даних; використовувати програмні бібліотеки та фреймворки з ефективними алгоритмами опрацювання надвеликих обсягів даних.

### Структура курсу

Години (лек. / сем.)	Тема	Результати навчання	Завдання
----------------------------	------	---------------------	----------

2/1	1. Лінійний і квадратичний дискримінантний аналіз	Вміти використати лінійний дискримінантний аналіз (LDA) із застосуванням генеративного підходу для класифікації. Знати моделі класів та їх коваріацій. Уміти застосувати квадратний дискримінантний аналіз (QDA) із індивідуальною коваріаційною матрицею на практиці	Поточне опитування
4/1	2. Машини опорних векторів	Знати мету алгоритму опорних векторів. Орієнтуватися у зозділяючій гіперплощині. Знати що таке м'ягкі розділювачі, регуляризація.	Поточне опитування Практичні завдання
4/1	3. Стохастичний градієнтний спуск	Знати що таке рух за антиградієнтом. Розуміти обмеження на крок спуску, стохастична точка спуску. Множина точок пошуку. Уміти використати стохастичний градієнтний спуск у лінійних моделях машинного навчання	Поточне опитування Практичні завдання
4/1	4. Метод найближчих сусідів	Знати базову класифікацію точок на основі класифікації найближчих сусідів. Уміти здійснювати підбір кількості сусідів, підбір кількості сусідів на невеликій навчальній вибірці. Здійснювати вибір метрики у відстанях між сусідами.	Поточне опитування Практичні завдання
4/2	5. Суміші нормальних розподілів	Знати Гаусівські процеси. Орієнтуватися у базисній функції (RBF). Здійснювати розподіл імовірностей для нових спостережень. Уміти прогнозувати Гаусівський процес.	Поточне опитування Практичні завдання
2/1	6. Наївний баєсівський класифікатор	Знати теорему Байєса, докази. Розуміти класифікаційну імовірність. Уміти застосувати види класифікаторів: нормальний, мультиноміальний, Бернуллі.	Поточне опитування Практичні завдання
2/1	7. Деревя рішень	Знати структуру дерева рішень: вузли, гілки, листки, функції, правила, результати. Вміти застосувати алгоритм дерева класифікації та регресії. Орієнтуватися у сильних та слабких сторонах підходу дерева рішень	Поточне опитування Практичні завдання

4/2	8. Ансамблеві методи	Знати призначення ансамблевих методів. Розуміти що таке бутстрапінг як метод випадкового створення вибірок, беггінг як метод пакетування із агрегацією бутстрапінгу. Використовувати моделі випадкового лісу, як удосконалений беггінг.	Поточне опитування Практичні завдання
4/2	9. Напівконтрольоване навчання	Знати основні відмінності між контрольованим та неконтрольованим навчанням. Використовувати невеликої кількості розмічених та великої кількості нерозмічених даних. Уміти здійснювати маркування аудіофайлів, веб-контенту, білкових структур.	Поточне опитування Практичні завдання
4/2	10. Ізотонічна регресія	Орієнтуватися у проблемі знаходження монотонної функції, що мінімізує похибку на експериментальних даних. Уміти покроково використовувати динамічне програмування. Застосувати на практиці ізотонічної регресії. Уміти здійснювати корекцію прогнозованих даних. Знати моделювання порядкових змінних.	MP

## Літературні джерела

1. <https://rb.ru/howto/chto-takoe-big-data/>
2. <https://postnauka.ru/faq/46974>
3. <https://www.datacenterknowledge.com/archives/2015/03/30/big-data-bubble-set-burst>
4. [http://www.tadviser.ru/index.php/Статья:Большие\\_данные \(Big\\_Data\)](http://www.tadviser.ru/index.php/Статья:Большие_данные_(Big_Data))
5. [https://ru.wikipedia.org/wiki/Большие\\_данные](https://ru.wikipedia.org/wiki/Большие_данные)
6. <https://intellect.ml/big-data-6821>
7. [http://www.mckinsey.com/insights/business\\_technology/big\\_data\\_the\\_next\\_frontier\\_f or\\_innovatin](http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_f or_innovatin)
9. <http://www.ogcs.com.ua/index.php/articles/121-big-data-v-promyshlennosti-innovatsii-k-kotorym-pridetsya-privykat>

## Політика оцінювання

- **Політика щодо дедлайнів та перескладання:** Роботи, які здаються із порушенням термінів без поважних причин, оцінюються на нижчу оцінку (-20 балів). Перескладання модулів відбувається із дозволу деканату за наявності поважних причин (наприклад, лікарняний).

- **Політика щодо академічної доброчесності:** Усі письмові роботи перевіряються на наявність плагіату і допускаються до захисту із коректними текстовими запозиченнями не більше 20%. Списування під час контрольних робіт та екзаменів заборонені (в т.ч. із використанням мобільних девайсів). Мобільні пристрої дозволяється використовувати лише під час он-лайн тестування (наприклад, програма Kahoot).

- **Політика щодо відвідування:** Відвідування занять є обов'язковим компонентом оцінювання, за яке нараховуються бали. За об'єктивних причин (наприклад, хвороба, міжнародне стажування) навчання може відбутись в он-лайн формі за погодженням із керівником курсу.

## Оцінювання

Модуль 1		Модуль 2	Модуль 3
40%	40%	5%	15%
Поточне оцінювання	Модульний контроль	Тренінги	Самостійна робота
Визначається як середнє арифметичне оцінок, отриманих на практичних заняттях (кожен здобувач має отримати не менше 4 оцінок)	Підсумкова контрольна робота, яка охоплює всі теми	Визначається як оцінка, отримана за виконанні завдання під час тренінгу	Визначається як оцінка, отримана за виконання завдання самостійної роботи

### Шкала оцінювання:

За шкалою ЗУНУ	За національною шкалою	За шкалою ECTS
90–100	відмінно	A (відмінно)
85–89	добре	B (дуже добре)
75-84		C (добре)
65-74		D (задовільно)
60-64	задовільно	E (достатньо)
35-59		FX (незадовільно з можливістю повторного складання)
1-34	незадовільно	F (незадовільно з обов'язковим повторним курсом)