



# DATA SCIENCE

## Робоча програма навчальної дисципліни (Силабус)

### Реквізити навчальної дисципліни

Рівень вищої освіти	<i>Другий (магістерський)</i>
Галузь знань	<i>12 Інформаційні технології</i>
Спеціальність	<i>122 Комп'ютерні науки та інформаційні технології</i>
Освітня програма	<i>Системний аналіз і управління</i>
Статус дисципліни	<i>Вибіркова</i>
Форма навчання	<i>очна (денна)</i>
Рік підготовки, семестр	<i>1 курс, весняний семестр</i>
Обсяг дисципліни	<i>150 годин (5 кредитів ЄКТС)</i>
Семестровий контроль/ контрольні заходи	<i>Екзамен, МКР</i>
Розклад занять	<i><a href="https://schedule.kpi.ua/">https://schedule.kpi.ua/</a> Лекції та практики кожного тижня</i>
Мова викладання	<i>Українська</i>
Інформація про керівника курсу / викладачів	<i>д.т.н., доцент, Терент'єв Олександр Миколайович, пошта: <a href="mailto:o.terentiev@gmail.com">o.terentiev@gmail.com</a></i>
Розміщення курсу	<i>Система Moodle</i>

### Програма навчальної дисципліни

#### 1. Опис навчальної дисципліни, її мета, предмет вивчення та результати навчання

*Програму навчальної дисципліни «Data Science» складено відповідно до освітньо-професійної програми підготовки магістрів з комп'ютерних наук другого рівня вищої освіти спеціальності 122 «Комп'ютерні науки». Навчальна дисципліна належить до циклу професійної підготовки. Статус навчальної дисципліни - вибіркова. Обсяг навчальної дисципліни становить 5 кредитів ЄКТС.*

*Предмет вивчення дисципліни становлять, головним чином, методи, підходи та технології обробки даних в умовах великих обсягів і високого рівня паралелізму, статистичні методи, методи інтелектуального аналізу даних і застосування штучного інтелекту для роботи з даними, а також методи проектування та розробки баз даних, методи підготовки даних для аналізу, перевірки їх на якість та аномалії, побудови математичних моделей методами прогнозного моделювання.*

*Вивчення навчальної дисципліни націлено на формування, розвиток та закріплення у здобувачів таких загальних та фахових **компетентностей**: ЗК 1 Здатність до абстрактного мислення, аналізу та синтезу, ЗК 7 Здатність до пошуку, оброблення та аналізу інформації з різних джерел, ЗК 11 Здатність приймати обґрунтовані рішення, ФК 2 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, ФК 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, ФК 22 Здатність*

використовувати мови штучного інтелекту при розробці програмного забезпечення інтелектуальних інформаційних систем, здатність орієнтуватися в різних типах інтелектуальних систем і технологій; ставити завдання побудови інтелектуальних систем для вирішення завдання вибору варіантів в проблемній області, що погано формалізується.

По завершенню курсу студент має набути наступні **програмні результати навчання**: 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, 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 багатовимірної аналізу даних на основі технологій DataMining, TextMining, WebMining, ПР 19 Розробляти системи розпізнавання образів та класифікації в різних предметних областях, обґрунтовано вибирати та використовувати алгоритми розпізнавання образів та проводити навчання систем розпізнавання образів, ПР 24 Розуміти різні типи інтелектуальних систем і технологій; ставити завдання побудови інтелектуальних систем для вирішення завдання вибору варіантів в проблемній області, що погано формалізується.

Після засвоєння дисципліни студенти мають продемонструвати такі результати навчання:

знання:

- основних понять, задач, стадій інтелектуального аналізу даних;
- підходів до збереження, представлення даних у сучасних інформаційних системах;
- аналізу залежностей у даних;
- методи та інструменти інтелектуального аналізу даних;
- сучасні програмні засоби, застосовувані для інтелектуального аналізу даних;
- концепції сховищ даних, підходи до їх аналітичної обробки.

уміння:

- обґрунтовувати вибір конкретного типу моделі та методу інтелектуального аналізу даних при вирішенні поставленої практичної задачі;
- 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 для проектування та дослідження систем інтелектуального аналізу даних;
- застосовувати технології роботи зі сховищами даних, здійснювати їх аналітичну обробку та інтелектуальний аналіз для забезпечення надійної роботи інформаційних систем;

## **2. Пререквізити та постреквізити дисципліни (місце в структурно-логічній схемі навчання за відповідною освітньою програмою)**

Міждисциплінарні зв'язки: вивчення даної дисципліни забезпечується дисциплінами обов'язкового компоненту циклів загальної та професійної підготовки: «теорія ймовірностей», «математична логіка і теорія алгоритмів» та «математичний аналіз». Знання, отримані при вивченні дисципліни, можуть бути використані студентами при написанні магістерської дисертаційної роботи

### 3. Зміст навчальної дисципліни

<i>Перелік розділів і тем дисципліни</i>
<i>Тема 1. Наука о даних, як міждисциплінарна галузь про наукові методи, процеси і системи, які стосуються добування знань із даних у різних формах, як структурованих так і неструктурованих. Сфери застосування та типи задач, що вирішуються.</i>
<i>Тема 2. Топ-10 підходів та методів Data Science для аналізу даних. Квадранти Gartner та цикли розвитку інтелектуальних технологій. Найкращі платформи машинного навчання за рейтингом консалтингової компанії Gartner.</i>
<i>Тема 3. Основні етапи та кроки Data Science проекту щодо аналізу даних. Підходи щодо формалізації аналітичної задачі. Методи збору даних – класичні та сучасні. Статистичні методи аналізу якості даних.</i>
<i>Тема 4. Прогнозне моделювання із використанням регресії. Вибір вхідних змінних регресії за допомогою статистичних тестів та в автоматичному режимі, методами forward, backward та stepwise.</i>
<i>Тема 5. Оптимізація складності моделі регресії. Принцип розбиття даних на навчальну, валідаційну та тестові вибірки.</i>
<i>Тема 6. Побудова нових та перетворення вже існуючих змінних моделі, з метою усунення проблем мультиколінеарності та зміщення.</i>
<i>Тема 7. Застосування категоріальних змінних при побудові регресійних моделей. Побудова фіктивних змінних аналізу, групування ознак за методами WOE та SWOE.</i>
<i>Тема 8. Використання поліноміальної регресії. Проблема вибору важливих змінних аналізу в умовах обмеженості обчислювального ресурсу.</i>
<i>Тема 9. Побудова скорингових карт із використанням логіт та пробіт регресійних моделей.</i>
<i>Тема 10. Основи використання дерев рішень. Використання статистики <math>\chi^2</math>-квадрат р-значення, як міри визначення точки розділення змінної аналізу.</i>
<i>Тема 11. Методи автономної побудови дерев рішень. Поправки Бонферотті та за глибиною.</i>
<i>Тема 12. Використання нейронних мереж, для вирішення задач прогнозного моделювання. Типи нейронних мереж, види зав'язків між змінними. Проблема складності побудови мережі та вибору топології.</i>
<i>Тема 13. Проблема вибору вхідних змінних аналізу для побудови нейронної мережі. Оптимізація нейронної мережі – за збіжністю функціонала або зупинене навчання.</i>
<i>Тема 14. Використання методу головних компонентів як окремого підходу для моделювання, так і у сукупності з іншими методами. Власні вектори та власні числа.</i>
<i>Тема 15. Використання ансамблю моделей та сурогатне моделювання.</i>
<i>Тема 16. Застосування методів аналізу текстової інформації, з метою видобування ознак для подальшого використання в задачах прогнозного моделювання.</i>
<i>Тема 17. Порівняння моделей. Оцінка ефективності та порівняння моделей. Оцінка помилки моделі. Статистики підгонки моделей та статистичні графіки. ROC-аналіз.</i>
<i>Тема 18. Навчання в умовах незбалансованості класів. Застосування поправок для роздільних вибірок в задачах прогнозного моделювання. Застосування матриць дохідностей.</i>

### 4. Навчальні матеріали та ресурси

Базова література:

1. Бідюк П.І. Прикладна статистика (навчальний посібник) / П.І. Бідюк, О.М. Терентьєв, Т.І. Просянкіна-Жарова. – Вінниця : ПП “ТД Едельвейс і К”, 2013. – 288 с. –

ISBN 978-966-2462-21-0. – Режим доступу: <https://ela.kpi.ua/handle/123456789/19563>

2. Згуровський М.З. Байєсівські мережі в системах підтримки прийняття рішень (навчальний посібник) / Згуровський М.З., Бідюк П.І., Терент'єв О.М., Просянкіна-Жарова Т.І. – К: ТОВ “Видавниче підприємство Едельвейс”, 2015. – 300 с. – ISBN 978-966-2748-73-4. – Режим доступу: <https://ela.kpi.ua/handle/123456789/19582>
3. Mohammed J. Zaki, Wagner Meira, Jr., *Data Mining and Machine Learning: Fundamental Concepts and Algorithms, 2nd Edition, Cambridge University Press, March 2020. ISBN: 978-1108473989.* – Режим доступу: [https://dataminingbook.info/book\\_html](https://dataminingbook.info/book_html)

#### Додаткова література:

1. Сергеев-Горчинський О. О., Іщенко Г. В. Інтелектуальний аналіз даних (комп'ютерний практикум) [Електронний ресурс] : навч. посіб. для студ. спеціальності 122 «Комп'ютерні науки та інформаційні технології», спеціалізацій «Інформаційні системи та технології проектування», «Системне проектування сервісів» / О. О. Сергеев-Горчинський, Г. В. Іщенко ; КПІ ім. Ігоря Сікорського. – Електронні текстові дані (1 файл: 1,72 Мбайт). – Київ : КПІ ім. Ігоря Сікорського, 2018. – 73 с. – Режим доступу: [https://ela.kpi.ua/bitstream/123456789/24971/1/Komp\\_prakt.pdf](https://ela.kpi.ua/bitstream/123456789/24971/1/Komp_prakt.pdf)
2. O. M. Trofymchuk, P. I. Bidyuk, T. I. Prosyankina-Zharova, O. M. Terentiev. *Decision Support Systems for Modeling, Forecasting and Risk Estimation (Monograph).* – LAP LAMBERT Academic Publishing, 2019. – 176 p. – ISBN-13: 978-3-659-61214-5. – ISBN-10: 3659612146. – EAN: 9783659612145. – Режим доступу: <https://www.morebooks.shop/store/gb/book/decision-support-systems-for-modeling,-forecasting-and-risk-estimation/isbn/978-3-659-61214-5>
3. Pang-Ning Tan Michael Steinbach Vipin Kumar. *Introduction to Data Mining.* Pearson Education Limited 2014. – 719 p. – ISBN 10: 1-292-02615-4. – ISBN 13: 978-1-292-02615-2. – Режим доступу: [https://www.ceom.ou.edu/media/docs/upload/Pang-Ning\\_Tan\\_Michael\\_Steinbach\\_Vipin\\_Kumar\\_-\\_Introduction\\_to\\_Data\\_Mining-Pe\\_NRDK4fi.pdf](https://www.ceom.ou.edu/media/docs/upload/Pang-Ning_Tan_Michael_Steinbach_Vipin_Kumar_-_Introduction_to_Data_Mining-Pe_NRDK4fi.pdf)

### Навчальний контент

#### 5. Методика опанування навчальної дисципліни (освітнього компонента)

Лекції, практичні заняття, виконання аналітично-розрахункових завдань, дискусії, аналіз конкретних ситуацій, командна робота, модульна контрольна робота.

#### Структура дисципліни

Назви розділів і тем	Кількість годин				
	Всього	у тому числі			
		Лекції	Практичні (семінарські)	Лабораторні	СРС
1	2	3	4	5	6
<b>«Вступ до інтелектуального аналізу даних»</b>					
Тема 1. Наука о даних, як міждисциплінарна галузь про наукові методи, процеси і системи, які стосуються добування знань із даних у різних формах, як структурованих так і неструктурованих. Сфери	8	2	1		5

застосування та типи задач, що вирішуються.					
Тема 2. Топ-10 підходів та методів Data Science для аналізу даних. Квадранти Gartner та цикли розвитку інтелектуальних технологій. Найкращі платформи машинного навчання за рейтингом консалтингової компанії Gartner.	8	2	1		5
Тема 3. Основні етапи та кроки Data Science проекту щодо аналізу даних. Підходи щодо формалізації аналітичної задачі. Методи збору даних – класичні та сучасні. Статистичні методи аналізу якості даних.	8	2	1		5
Тема 4. Прогнозне моделювання із використанням регресії. Вибір вхідних змінних регресії за допомогою статистичних тестів та в автоматичному режимі, методами forward, backward та stepwise.	8	2	1		5
Тема 5. Оптимізація складності моделі регресії. Принцип розбиття даних на навчальну, валідаційну та тестові вибірки.	8	2	1		5
Тема 6. Побудова нових та перетворення вже існуючих змінних моделі, з метою усунення проблем мультиколінеарності та зміщення.	8	2	1		5
Тема 7. Застосування категоріальних змінних при побудові регресійних моделей. Побудова фіктивних змінних аналізу, групування ознак за методами WOE та SWOE.	8	2	1		5
Тема 8. Використання поліноміальної регресії. Проблема вибору важливих змінних аналізу в умовах обмеженості обчислювального ресурсу.	8	2	1		5
Тема 9. Побудова скорингових карт із використанням логіт та пробіт регресійних моделей.	8	2	1		5
Тема 10. Основи використання дерев рішень. Використання статистики $\chi^2$ -квадрат р-значення, як міри визначення точки	8	2	1		5



розділення змінної аналізу.					
Тема 11. Методи автономної побудови дерев рішень. Поправки Бонферотті та за глибиною.	8	2	1		5
Тема 12. Використання нейронних мереж, для вирішення задач прогнозного моделювання. Типи нейронних мереж, види зав'язків між змінними. Проблема складності побудови мережі та вибору топології.	8	2	1		5
Тема 13. Проблема вибору вхідних змінних аналізу для побудови нейронної мережі. Оптимізація нейронної мережі – за збіжністю функціонала або зупинене навчання.	9	2	1		6
Тема 14. Використання методу головних компонентів як окремого підходу для моделювання, так і у сукупності з іншими методами. Власні вектори та власні числа.	9	2	1		6
Тема 15. Використання ансамблю моделей та сурогатне моделювання.	9	2	1		6
Тема 16. Застосування методів аналізу текстової інформації, з метою видобування ознак для подальшого використання в задачах прогнозного моделювання.	9	2	1		6
Тема 17. Порівняння моделей. Оцінка ефективності та порівняння моделей. Оцінка помилки моделі. Статистики підгонки моделей та статистичні графіки. ROC-аналіз.	9	2	1		6
Тема 18. Навчання в умовах незбалансованості класів. Застосування поправок для роздільних вибірок в задачах прогнозного моделювання. Застосування матриць дохідностей.	9	2	1		6
<b>Разом за розділом</b>	<b>150</b>	<b>36</b>	<b>18</b>		<b>96</b>
<i>Залік</i>					
<b>Всього годин</b>	<b>150</b>	<b>36</b>	<b>18</b>		<b>96</b>

### Лекційні заняття

№ з/п	Назва теми лекції
1.	Тема 1. Наука о даних, як міждисциплінарна галузь про наукові методи, процеси і системи, які стосуються добування знань із даних у різних формах, як структурованих так і неструктурованих. Сфери застосування та типи задач, що вирішуються.
2.	Тема 2. Топ-10 підходів та методів Data Science для аналізу даних. Квадранти Gartner та цикли розвитку інтелектуальних технологій. Найкращі платформи машинного навчання за рейтингом консалтингової компанії Gartner.
3.	Тема 3. Основні етапи та кроки Data Science проекту щодо аналізу даних. Підходи щодо формалізації аналітичної задачі. Методи збору даних – класичні та сучасні. Статистичні методи аналізу якості даних.
4.	Прогнозне моделювання із використанням регресії. Вибір вхідних змінних регресії за допомогою статистичних тестів та в автоматичному режимі, методами forward, backward та stepwise.
5.	Оптимізація складності моделі регресії. Принцип розбиття даних на навчальну, валідаційну та тестові вибірки.
6.	Побудова нових та перетворення вже існуючих змінних моделі, з метою усунення проблем мультиколінеарності та зміщення.
7.	Застосування категоріальних змінних при побудові регресійних моделей. Побудова фіктивних змінних аналізу, групування ознак за методами WOE та SWOE.
8.	Використання поліноміальної регресії. Проблема вибору важливих змінних аналізу в умовах обмеженості обчислювального ресурсу.
9.	Побудова скорингових карт із використанням логіт та пробіт регресійних моделей.
10.	Основи використання дерев рішень. Використання статистики $\chi^2$ -квадрат р-значення, як міри визначення точки розділення змінної аналізу.
11.	Методи автономної побудови дерев рішень. Поправки Бонферотті та за глибиною.
12.	Використання нейронних мереж, для вирішення задач прогнозного моделювання. Типи нейронних мереж, види зав'язків між змінними. Проблема складності побудови мережі та вибору топології.
13.	Проблема вибору вхідних змінних аналізу для побудови нейронної мережі. Оптимізація нейронної мережі – за збіжністю функціонала або зупинене навчання.
14.	Використання методу головних компонентів як окремого підходу для моделювання, так і у сукупності з іншими методами. Власні вектори та власні числа.
15.	Використання ансамблю моделей та сурогатне моделювання.
16.	Застосування методів аналізу текстової інформації, з метою видобування ознак для подальшого використання в задачах прогнозного моделювання.
17.	Порівняння моделей. Оцінка ефективності та порівняння моделей. Оцінка помилки моделі. Статистики підгонки моделей та статистичні графіки. ROC-аналіз.
18.	Навчання в умовах незбалансованості класів. Застосування поправок для роздільних вибірок в задачах прогнозного моделювання. Застосування матриць дохідностей.

### Практичні заняття

*Основні завдання циклу практичних занять набути практичних навичок застосування методів інтелектуального аналізу даних*

№ з/п	Назва теми заняття
1	Написання програми для вивантаження даних за API з біржи Binance.
2	Формалізація аналітичної задачі щодо прогнозування напрямку зміни курсу крипто-валюти. Побудова цільової та вхідної змінних.
3	Написання програми для прогнозування напрямку зміни крипто-валюти на основі

	методу експоненційного згладжування.
4	Вирішення аналітичної задачі, щодо кластеризації крипто-активів за їх текстовими описами. Завантаження даних, парсинг, фільтрація та кластеризація.
5	Побудова моделі нейронної мережі для прогнозування курсу крипто-валюти. Типи нейронних мереж та функцій зав'язків, що використовуються.
6	Застосування моделей множинної регресії для прогнозування.
7	Побудова мереж Байеса на прикладах із використанням програмного забезпечення Netica та Hugin.
8	Практичний приклад використання дерева рішень для прогнозування бінарного відгуку цільової змінної, за методом Кааса.
9	Використання моделі логістичної регресії для прогнозування ризику неповернення кредиту. Поняття скорингової карти. Групування рівнів змінних за методами WOE та SWOE.

## 6. Самостійна робота студента/аспіранта

*До самостійної роботи студентів відноситься підготовка до аудиторних занять, опанування тем, що винесені на самостійне вивчення, проведення обчислень та написання програм, за первинними даними, отриманими на практичних заняттях, підготовка до модульної контрольної роботи та екзамену. Відповідно до навчального плану на самостійну роботу відводиться 96 годин.*

### Самостійна робота

№ з/п	Назви тем і питань, що виносяться на самостійне опрацювання та посилання на навчальну літературу	Кількість годин СРС
1	Наука о даних, як міждисциплінарна галузь про наукові методи, процеси і системи, які стосуються добування знань із даних у різних формах, як структурованих так і неструктурованих. Сфери застосування та типи задач, що вирішуються.	5
2	Топ-10 підходів та методів Data Science для аналізу даних. Квадранти Gartner та цикли розвитку інтелектуальних технологій. Найкращі платформи машинного навчання за рейтингом консалтингової компанії Gartner.	5
3	Основні етапи та кроки Data Science проекту щодо аналізу даних. Підходи щодо формалізації аналітичної задачі. Методи збору даних – класичні та сучасні. Статистичні методи аналізу якості даних.	5
4	Прогнозне моделювання із використанням регресії. Вибір вхідних змінних регресії за допомогою статистичних тестів та в автоматичному режимі, методами forward, backward та stepwise.	5
5	Оптимізація складності моделі регресії. Принцип розбиття даних на навчальну, валідаційну та тестові вибірки.	5
6	Побудова нових та перетворення вже існуючих змінних моделі, з метою усунення проблем мультиколінеарності та зміщення.	5
7	Застосування категоріальних змінних при побудові регресійних моделей. Побудова фіктивних змінних аналізу, групування ознак за методами WOE та SWOE.	5
8	Використання поліноміальної регресії. Проблема вибору важливих змінних аналізу в умовах обмеженості обчислювального ресурсу.	5
9	Побудова скорингових карт із використанням логіт та пробіт регресійних моделей.	5
10	Основи використання дерев рішень. Використання статистики $\chi^2$ -квадрат $p$ -	5



	значення, як міри визначення точки розділення змінної аналізу.	
11	Методи автономної побудови дерев рішень. Поправки Бонферотті та за глибиною.	5
12	Використання нейронних мереж, для вирішення задач прогнозного моделювання. Типи нейронних мереж, види зав'язків між змінними. Проблема складності побудови мережі та вибору топології.	5
13	Проблема вибору вхідних змінних аналізу для побудови нейронної мережі. Оптимізація нейронної мережі – за збіжністю функціонала або зупинене навчання.	6
14	Використання методу головних компонентів як окремого підходу для моделювання, так і у сукупності з іншими методами. Власні вектори та власні числа.	6
15	Використання ансамблю моделей та сурогатне моделювання.	6
16	Застосування методів аналізу текстової інформації, з метою видобування ознак для подальшого використання в задачах прогнозного моделювання.	6
17	Порівняння моделей. Оцінка ефективності та порівняння моделей. Оцінка помилки моделі. Статистики підгонки моделей та статистичні графіки. ROC-аналіз.	6
18	Навчання в умовах незбалансованості класів. Застосування поправок для роздільних вибірок в задачах прогнозного моделювання. Застосування матриць дохідностей.	6

## Політика та контроль

### 7. Політика навчальної дисципліни (освітнього компонента)

*Під час занять (як лекцій, так і практичних) студентам надаються інтерактивні вправи у вигляді задач, оцінюється як активність, так і успішність освоєння матеріалу.*

*(в тому числі під час виконання домашніх завдань і підготовки до лекцій), призначаються заохочувальні бали.*

*Всі завдання з описом відповідних етапів їх виконання, а також відповідних контрольних строків завантажені у систему Moodle дисципліни. Усі роботи студенти мають завантажувати в особистому кабінеті Moodle. Дедлайни кожного завдання позначені в завданнях у Moodle. Захисти індивідуальних робіт відбуваються на практичних заняттях. Роботи мають бути виконані з дотриманням академічної доброчесності. У період роботи в дистанційному режимі лектор може запропонувати студентам пройти запропоновані ним онлайн-курси на платформі Coursera. Сертифікати/проміжні результати цих курсів зараховуються бути частково зараховані згідно до Положення розміщеного на сайті КПІ ім. Ігоря Сікорського ([https://document.kpi.ua/files/2020\\_7-124.pdf](https://document.kpi.ua/files/2020_7-124.pdf)).*

*Крім того, передбачається, в межах вивчення навчальної дисципліни, участь в конференціях, форумах, круглих столах тощо.*

*Політика та принципи академічної доброчесності визначені у розділі 3 Кодексу честі Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського». Детальніше: <https://kpi.ua/code>.*

*Норми етичної поведінки студентів і працівників визначені у розділі 2 Кодексу честі Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського». Детальніше: <https://kpi.ua/code>.*

## 8. Види контролю та рейтингова система оцінювання результатів навчання (PCO)

Семестрова атестація проводиться у виді екзамену. Для оцінювання результатів навчання застосовується 100-бальна рейтингова система і університетська шкала оцінювання.

Рейтинг студента з дисципліни розраховується зі 100 балів, з них 52 бали складає стартова шкала. Стартовий рейтинг (протягом семестру) складається з балів, що студент отримує за:

- роботу на практичних заняттях (9 занять);
- виконання завдань індивідуального навчально-дослідного завдання;
- виконання модульної контрольної роботи – максимум 10 балів

Критерії нарахування балів:

### 1. Робота на практичних заняттях:

- активна творча робота – 4 бали;
- плідна робота – 3 бали.

### 2. Виконання індивідуальних навчально-дослідних завдань:

- творча робота – 16-15 балів;
- роботу виконано з незначними недоліками – 14-12 балів;
- роботу виконано з певними помилками – 11-10 балів;
- роботу не зараховано (завдання не виконане або є грубі помилки) – 0 балів.

3. Модульна контрольна робота оцінюється із 10 балів. Контрольне завдання оцінюється за такими критеріями:

- «відмінно» – повна відповідь (не менше 90% потрібної інформації), надані відповідні обґрунтування та особистий погляд – 10...9 балів;
- «добре» – достатньо повна відповідь (не менше 75% потрібної інформації), що виконана згідно з вимогами до рівня «умінь», або незначні неточності – 8...7 балів;
- «задовільно» – неповна відповідь (не менше 60% потрібної інформації, що виконана згідно з вимогами до «стереотипного» рівня та деякі помилки) – 6...5 балів;
- «незадовільно» – незадовільна відповідь – 0 балів.

4. Умовою першої атестації є отримання не менше 8 балів та виконання завдань всіх практичних робіт (на час атестації). Умовою другої атестації – отримання не менше 22 балів, виконання завдань всіх практичних робіт (на час атестації) та зарахування виконання індивідуальної роботи.

5. Умовою допуску до екзамену є зарахування всіх практичних робіт, індивідуальної роботи та стартовий рейтинг не менше 26 балів.

6. На екзамені студенти виконують письмову контрольну роботу. Кожне завдання містить два теоретичних запитання (завдання) і одне практичне. Перелік запитань наведений у Рекомендаціях до засвоєння дисципліни. Кожне запитання (завдання) оцінюється у 16 балів за такими критеріями:

- «відмінно», повна відповідь, не менше 90% потрібної інформації (повне, безпомилкове розв'язування завдання) – 16-15 балів;
- «добре», достатньо повна відповідь, не менше 75% потрібної інформації або незначні неточності (повне розв'язування завдання з незначними неточностями) – 14-12 балів;

- «задовільно», неповна відповідь, не менше 60% потрібної інформації та деякі помилки (завдання виконане з певними недоліками) – 11-10 балів;
- «незадовільно», відповідь не відповідає умовам до «задовільно» – 0 балів.

7. Сума стартових балів та балів за екзаменаційну контрольну роботу переводиться до рейтингової оцінки згідно з таблицею:

Бали: практичні заняття + ІДЗ + + екзаменаційна контрольна робота	Оцінка	Бали ECTS
100...95	Відмінно	A
94...85	Дуже добре	B
84...75	Добре	C
74...65	Задовільно	D
64...60	Достатньо	E
Менше 60	Незадовільно	FX
Є не зараховані практичні роботи або не зарахована розрахункова робота або стартовий рейтинг менше 26 балів	Не допущено	F

### 9. Додаткова інформація з дисципліни (освітнього компонента)

*У студентів є можливість отримати позитивну оцінку по дисципліні за умов отримання сертифікатів проходження дистанційних чи онлайн курсів за відповідною тематикою.*

*У додатку до силабусу знаходиться перелік питань до екзамену.*

### Робочу програму навчальної дисципліни (силабус):

**Складено** доцентом кафедри математичних методів системного аналізу, д.т.н., доцентом Терентьєвим О.М.

**Ухвалено** кафедрою математичних методів системного аналізу (протокол № 13 від 05.06.2024)

**Погоджено** Методичною комісією НН ІПСА (протокол № 10 від 24.06.2024)

## Перелік питань до екзамену.

1. Суть та мета Data Science?
2. Дослідження систем на основі аналітичного підходу?
3. Способи візуалізації даних?
4. Основні групи аналітичних задач?
5. Процес побудови аналітичної моделі?
6. Формалізація даних та вимоги до них, при вирішенні задач прогнозного моделювання?
7. Особливості технологій KDD, Data Mining та Data Science?
8. Сучасні аналітичні платформи для вирішення задач Data Science?
9. Методи та алгоритми Data Science. Загальні характеристики кожного з методів.
10. Місце та роль Text Mining при вирішенні задач Data Science?
11. Автоматичне реферування.
12. Правило Парето і деякі його наслідки.
13. Закони Зіпфа.
14. Закономірність Бредфорда.
15. Модель як засіб дослідження об'єкту, системи, явища.
16. Основні підходи до побудови моделей даних.
17. Властивості моделі.
18. Дані, їх види та форми їх представлення.
19. Структуровані дані як основна форма представлення відомостей в базах даних.
20. Типи структурованих даних.
21. Особливості підготовки даних до аналізу.
22. Навчальна вибірка. Правила формування.
23. Навчання з учителем та без.
24. Помилки, що виникають у процесі навчання.
25. Перенавчання. Ефект перенавчання
26. Які типи невизначеностей враховуються у мережах Байєса?
27. Які два типи навчання необхідно виконати для побудови мереж Байєса?
28. Чим відрізняються дискретні змінні від неперервних?
29. Чи може мережа Байєса містити одночасно дискретні і неперервні змінні?
30. Наведіть недоліки і переваги використання мереж Байєса?
31. Основні типи з'єднань в SQL? Наведіть приклади.
32. Що таке Self Join та для чого використовується оператор UNION?
33. Яким чином працюють підстановочні знаки в SAS Base? Наведіть приклади.
34. Для чого використовуються аліаси (aliases) в бібліотеках та базах даних? Наведіть приклади.
35. Що таке нормалізація та денормалізація? Наведіть приклади.
36. Які типи даних використовуються в Python? Скільки байт займають різні типи та як зберігаються?
37. Інформаційно-пошукові мови. Ранжування.
38. Векторно-просторова та ймовірнісна модель пошуку.
39. Ймовірнісна модель пошуку
40. Моделі динаміки інформаційних потоків.
41. Фрактальні властивості інформаційних потоків. Параметр Херста.
42. Вікі (корпоративна технологія і Вікіпедія). Розмітка документів в Вікі з метою збагачення даних при вирішенні задач Data Science.
43. Основні компоненти інформаційно-аналітичної системи.

44. З яких компонентів складається типова аналітична платформа для вирішення задач Data Science? Надайте опис кожного з компонентів.
45. Побудова ER-моделі бази даних для автоматизації роботи бібліотеки?
46. Особливості та концепції сховищ даних.
47. Деталізовані та агреговані дані.
48. Архітектура сховищ даних.
49. Багатовимірні та гібридні сховища даних.
50. Що таке OLTP, OLAP, OPT системи? Наведіть приклади.