



Силабус освітнього компонента

Програма навчальної дисципліни



Аналіз даних

Шифр та назва спеціальності

122 – Комп'ютерні науки

Інститут

ННІ Комп'ютерних наук та інформаційних технологій

Освітня програма

Комп'ютерні науки

Кафедра

Системного аналізу та інформаційно-аналітичних технологій

Рівень освіти

Бакалавр

Тип дисципліни

Спеціальна (фахова), Вибіркова

Семестр

5

Мова викладання

Українська

Викладачі, розробники



Безменов Микола Іванович

mykola.bezmenov@khpі.edu.ua

Кандидат технічних наук, доцент, професор

Автор понад 100 наукових та навчально-методичних праць. Провідний лектор з дисциплін: «Алгоритмізація та програмування», «Інформатика і програмування», «Основи візуального програмування», «Аналіз даних», «Статистичний аналіз даних», «Статистична обробка соціально-економічної інформації».

[Детальніше про викладача на сайті кафедри](#)

Загальна інформація

Анотація

У рамках курсу розглядаються питання використання апарату теорії ймовірностей та математичної статистики для статистичного аналізу експериментальних даних для подальшої побудови математичних моделей складних систем.

Мета та цілі дисципліни

Освоєння підходів та засобів аналізу емпіричних даних за допомогою теоретичних та евристичних методів для подальшого використання результатів аналізу при побудові математичних моделей складних технічних і соціально економічних систем.

Формат занять

Лекції, лабораторні роботи, розрахункове завдання, консультації. Підсумковий контроль - іспит.

Компетентності

ВКП2.1. Здатність проектувати, розробляти та вико-ристовувати інформаційно-аналітичні системи та те-хнології в різноманітних галузях людської діяльно-сті...

Результати навчання

РНП2.1. Вміти проектувати, розробляти та використовувати інформаційно-аналітичні системи, зокрема із застосуванням хмарних технологій та розподілених обчислень..

Обсяг дисципліни

Загальний обсяг дисципліни 150 год. (5.0 кредитів ECTS): лекції – 32 год., лабораторні роботи – 32 год., самостійна робота – 86 год.

Передумови вивчення дисципліни (пререквізити)

Для успішного проходження курсу необхідно мати знання та практичні навички з наступних дисциплін: «Алгебра та геометрія», «Фізика», «Теорія ймовірностей та математична статистика».

Особливості дисципліни, методи та технології навчання

Вивчення дисципліни орієнтоване на розгляд лекційного матеріалу та відпрацювання його під час лабораторних занять. Лабораторні роботи можуть виконуватися в будь-якому середовищі програмування за розсудом студента або з використанням можливостей Excel або MathCAD.

Програма навчальної дисципліни

Теми лекційних занять

Тема 1. Вступ. Предмет і задачі дисципліни її місце у навчальній програмі. Дані й виміри. Типи даних і шкал.

Тема 2. Аналіз дихотомічних спостережень.

Аналіз дихотомічних спостережень з використанням біноміального розподілу. Аналіз дихотомічних спостережень з використанням z-статистики. Перевірка гіпотез про належність одному з декількох класів. Використання статистики χ^2 у випадку двох класів.

Тема 3. Перевірка гіпотез у випадку неперервних змінних.

Особливості перевірки гіпотез у випадку неперервних змінних. Критерій узгодження χ^2 . Критерій узгодження Колмогорова – Смірнова. Перевірка гіпотез і довірчій інтервал для μ . Перевірка гіпотез і довірчій інтервал для σ^2 .

Тема 4. Аналіз двох випадкових величин.

Дві випадкові величини з однієї сукупності. Аналіз однієї випадкової величини на двох сукупностях. Аналіз K випадкових величин на одній сукупності. Аналіз однієї випадкової величини на K сукупностях.

Тема 5. Аналіз таблиць спряженості ознак.

Тема 6. Міри зв'язності Гудмена – Краскала.

Тема 7. Обробка аномальних спостережень для одновимірних даних.

Вінзоризовані та усечені оцінки. Кусочно-лінійна M-оцінка Гампеля.

Тема 8 Обробка аномальних спостережень для багатовимірних даних.

Аналіз викидів у багатовимірному випадку. Алгоритми обробки даних з наявністю викидів.

Тема 9. Перевірка гіпотез та класифікація об'єктів засобами багатовимірного аналізу.

Перевірка гіпотез для багатовимірних об'єктів. Класифікація об'єктів засобами багатовимірного аналізу. Баєсівський підхід до класифікації.

Тема 10. Перевірка гіпотез при багатовимірному біноміальному і нормальному розподілах.

Перевірка гіпотез про вектори середніх при багатовимірному біноміальному і нормальному розподілах.

Тема 11. Основи дискримінантного аналізу.

Тема 12. Покроковий дискримінантний аналіз.

Тема 13. Класифікація за допомогою методів теорії інформації.

Поняття ентропії та інформації. Використання інформаційних показників для розв'язання задач класифікації.

Тема 14. Основи кластерного аналізу.

Загальні поняття кластерного аналізу. Гіпотеза компактності. Послідовні та паралельні алгоритми кластерного аналізу.

Тема 15. Методи головних компонент.

Поняття головних компонент. Класифікація за допомогою метода головних компонент.

Тема 16. Основи факторного аналізу.

Основні поняття факторного аналізу. Обчислення головних факторів та обертання факторів.

Теми практичних занять

Практичні заняття в рамках дисципліни не передбачені.

Теми лабораторних робіт

Тема 1. Перевірка гіпотез про ймовірність для дискретних параметрів.

Використання біноміального розподілу.

Тема 2. Перевірка гіпотез про ймовірність для дискретних параметрів.

Використання z-статистики і статистики χ^2 у випадку двох класів.

Тема 3. Перевірка гіпотези про відповідність розподілу неперервного параметра деякому закону.

Критерій узгодження χ^2 . Критерій узгодження Колмогорова – Смірнова.

Тема 4. Перевірка гіпотез про математичне очікування і дисперсії.

Тема 5. Аналіз двох випадкових величин із однієї сукупності.

Тема 6. Аналіз випадкової величини на двох сукупностях.

Тема 7. Перевірка гіпотези про незалежність двох дискретних параметрів.

Тема 8. Міри зв'язності Гудмена – Краскала.

Тема 9. Оцінювання математичного очікування при наявності викидів у вимірах.

Тема 10. Аналіз викидів у багатовимірному випадку.

Тема 11. Перевірка гіпотез про вектори середніх у багатовимірному випадку.

Тема 12. Перевірка гіпотез про вектори середніх у багатовимірному випадку.

Тема 13. Класифікація багатовимірних об'єктів при наявності інформації про закон розподілу.

Тема 14. Класифікація багатовимірних об'єктів при наявності інформації про закон розподілу.

Тема 15. Класифікація багатовимірних об'єктів за допомогою дискримінантних функцій при невідомому розподілі векторів спостережень.

Тема 16. Класифікація багатовимірних об'єктів з використанням інформаційних показників.

Самостійна робота

Самостійна робота присвячена освоєнню середовища MathCAD, бібліотечних функцій MathCAD і програмуванню в цьому середовищі. Крім того, для самостійного вивчення виноситься деякий матеріал, що не входить в лекційні заняття, а саме, методи нормування змінних, Методи екстремального групування параметрів, методи зменшення розмірності факторного простору. У навчальному плані дисципліни передбачено також виконання індивідуального розрахункового завдання. Розрахункові завдання здаються студентами окремо і враховуються у разі виставлення оцінки за рейтингом.

Методичні вказівки до лабораторних робіт студенти можуть отримати звернувшись до сайту науково-технічної бібліотеки НТУ "ХПІ".

Література та навчальні матеріали

Основна література:

1. Бахрушин В. Є. Методи аналізу даних : навч. посіб. для студ. Запоріжжя: КПУ, 2011. 268 с. URL: https://moodle.znu.edu.ua/pluginfile.php/464907/mod_folder/intro/%D0%9F%D0%BE%D1%81%D1%96%D0%B1%D0%BD%D0%B8%D0%BA_1.1.pdf

2. Гороховатський В. О., Творошенко І. С. Методи інтелектуального аналізу та оброблення даних : навч. посіб. Харків : ХНУРЕ, 2021. 92 с. URL: <https://openarchive.nure.ua/server/api/core/bitstreams/2e55d639-52fd-48d9-b7b7-14989f49f291/content>
3. Перегуда О. В., Капустян О. А., Курилко О. Б. Статистична обробка даних [Електронний ресурс]: навч. посіб. Київ, 2022. 103 с. URL: https://www.mechmat.univ.kiev.ua/wp-content/uploads/2022/02/navch_pos_perehuda.pdf
4. Олещенко Л.М. Технології оброблення великих даних: конспект лекцій з дисципліни «Технології оброблення великих даних» [Електронний ресурс] : навч. посіб. Київ: КПІ ім. Ігоря Сікорського, 2021. 227 с. URL: <https://ela.kpi.ua/bitstream/123456789/42206/1/%D0%9AonspLekts Tekhnolohii-obroblennia-velykykh-danykh %D0%9Eleshchenko.pdf>
5. Бондаренко Я. С., Кравченко С. В. Посібник до вивчення дисципліни «Статистичний аналіз даних». Дніпро: Ліра, 2018. 40 с. URL: https://www.researchgate.net/publication/335340316_Posibnik_do_vivcenna_disciplini_Statisticni_j_analiz_danih
6. Мхітарян В. С. Аналіз даних. URL: https://stud.com.ua/93298/statistika/analiz_danih
7. Єлісеєва В. І. Статистика. URL: <https://stud.com.ua/74988/statistika/statistika>
8. Кононова К. Ю. Машинне навчання: методи та моделі: підручник. Харків: ХНУ імені В. Н. Каразіна, 2020. 301 с. URL: https://moodle.znu.edu.ua/pluginfile.php/593075/mod_folder/intro/%D0%91%D0%B0%D0%B7%D0%BE%D0%B2%D0%B8%D0%B9%20%D0%BF%D1%96%D0%B4%D1%80%D1%83%D1%87%D0%BD%D0%B8%D0%BA 2%20%28%D0%9A%D0%BE%D0%BD%D0%BE%D0%BD%D0%BE%D0%B2%D0%B0%20%D0%9A.%20%D0%AE.%20%D0%9C%D0%B0%D1%88%D0%B8%D0%BD%D0%BD%D0%B5%20%D0%BD%D0%B0%D0%B2%D1%87%D0%B0%D0%BD%D0%BD%D1%8F%20-%20%D0%BC%D0%B5%D1%82%D0%BE%D0%B4%D0%B8%20%D1%82%D0%B0%20%D0%BC%D0%BE%D0%B4%D0%B5%D0%BB%D1%96%29.pdf2
9. Єременко В.С., Куц Ю.В., Мокійчук В.М., Самойліченко О.В. Статистичний аналіз даних вимірювань: навч. посіб. Київ: НАУ, 2013. 320 с. URL: https://ela.kpi.ua/bitstream/123456789/44902/1/statistical_analysis.pdf
10. Черняк О. І. Інтелектуальний аналіз даних: підруч. Київ: Знання, 2014. 599 с.
11. Данильченко О. М., Данильченко А. О. Інтелектуальний аналіз даних: навч. посіб. Житомир: ЖДТУ, 2009. 405 с.
12. Черняк О. І., Захарченко П. О. Інтелектуальний аналіз даних: підручник. Київ, 2010. URL: https://moodle.znu.edu.ua/pluginfile.php/984539/mod_resource/content/3/0040761.pdf
13. Ситник В. Ф., Краснюк М. Т. Інтелектуальний аналіз даних (дейтамайнінг): навч. посіб. Київ: КНЕУ, 2007. 376 с. URL: <https://www.twirpx.com/file/1504742/>

Додаткова література:

1. Безменов М. І., Безменова О. М. Вступ до Mathcad : навч.-метод. посіб. для самост. роб. з курсу «Аналіз даних» для студ. спеціальностей 122 – Комп’ютерні науки, 124 – Системний аналіз, 186 – Видавництво та поліграфія. Харків: НТУ «ХПІ»; ТОВ «Планета-Принт», 2019. 68 с. <https://repository.kpi.kharkov.ua/server/api/core/bitstreams/dab858ab-4594-46df-a990-35c9fdb6a947/content>
2. Безменов М. І., Безменова О. М. Методичні вказівки до лабораторної роботи «Перевірка гіпотез про ймовірність для дискретних параметрів» з курсу «Аналіз даних» для студ. напряму 6.040302 – Інформатика і курсу «Аналіз даних і знань» для студ. напряму 6.040303 – Системний аналіз. НТУ «ХПІ», 2015. 24 с.
3. Безменов М. І., Безменова О. М., Калінін Д. В. Методичні вказівки до лабораторних робіт за темою «Перевірка гіпотез для неперервних параметрів» з курсу «Аналіз даних» для студ. спеціальностей 122 – Комп’ютерні науки та інформаційні технології, 124 – Системний аналіз, 186 – Видавництво та поліграфія. Харків: НТУ «ХПІ», 2017. 44 с.
4. Безменов М. І., Безменова О. М., Калінін Д. В. Методичні вказівки до лабораторних робіт за темою «Оцінювання наявності зв’язку між дискретними параметрами» з курсу «Аналіз даних» для студ. спеціальностей 122 – Комп’ютерні науки та інформаційні технології, 124 – Системний аналіз, 186 – Видавництво та поліграфія. Харків: НТУ «ХПІ», 2017. 57 с.

5. Безменов М. І., Безменова О. М., Калінін Д. В. Методичні вказівки до лабораторних робіт за темою «Обробка аномальних спостережень» з курсу «Аналіз даних» для студ. спеціальностей 122 – Комп'ютерні науки та інформаційні технології, 124 – Системний аналіз, 186 – Видавництво та поліграфія. Харків: НТУ «ХПІ», 2017. 33 с.
6. Безменов М. І., Безменова О. М., Калінін Д. В. Методичні вказівки до лабораторних робіт за темою «Перевірка гіпотез про вектори середніх і класифікація об'єктів для багатовимірного випадку» з курсу «Аналіз даних» для студ. спеціальностей 122 – Комп'ютерні науки та інформаційні технології, 124 – Системний аналіз, 186 – Видавництво та поліграфія. Харків: НТУ «ХПІ», 2017. 64 с.
7. Куссуль Н. М., Шелестов А. Ю., Тарасенко С. А., Яйлимова Г. О. Аналіз даних. Лабораторний практикум [Електронний ресурс] : навч. посіб. Київ : КПІ ім. Ігоря Сікорського, 2022. 28 с. URL: https://www.researchgate.net/publication/364353207_Analiz_danih
8. Сергеев-Горчинський О. О., Іщенко Г. В. Інтелектуальний аналіз даних: Комп'ютерний практикум [Електронний ресурс] : навч. посіб. Київ : КПІ ім. Ігоря Сікорського, 2018. 73 с. URL: https://ela.kpi.ua/bitstream/123456789/24971/1/Komp_prakt.pdf
9. Шелестов А. Ю., Куссуль Н. М. Засоби підготовки та аналізу даних. Лабораторний практикум [Електронний ресурс]: навч. посіб. Київ : КПІ ім. Ігоря Сікорського, 2022. 31 с. URL: https://ela.kpi.ua/bitstream/123456789/43491/1/Shelestov_Zasoby-pidhotovky-ta-analizu-danykh_LabPrakt.pdf

Система оцінювання

Критерії оцінювання успішності студента та розподіл балів

100% підсумкової оцінки складаються з результатів оцінювання у вигляді екзамену (40%) та поточного оцінювання (60%). Екзамен: два запитання з теорії і одна практична задача з усною доповіддю, причому розв'язання практичної задачі є обов'язковим. Поточне оцінювання: оцінювання виконання лабораторних робіт (40%), 2 контрольні роботи (по 5%) та розрахункове завдання (по 10%).

Шкала оцінювання

Сума балів	Національна оцінка	ECTS
90–100	Відмінно	A
82–89	Добре	B
75–81	Добре	C
64–74	Задовільно	D
60–63	Задовільно	E
35–59	Незадовільно (потрібне додаткове вивчення)	FX
1–34	Незадовільно (потрібне повторне вивчення)	F

Норми академічної етики і політика курсу

Студент повинен дотримуватися «Кодексу етики академічних взаємовідносин та доброчесності НТУ «ХПІ»: виявляти дисциплінованість, вихованість, доброзичливість, чесність, відповідальність. Конфліктні ситуації повинні відкрито обговорюватися в навчальних групах з викладачем, а при неможливості вирішення конфлікту – доводитися до відома співробітників дирекції інституту. Нормативно-правове забезпечення впровадження принципів академічної доброчесності НТУ «ХПІ» розміщено на сайті: <http://blogs.kpi.kharkov.ua/v2/nv/akademichna-dobrochesnist/>

Погодження

Силабус погоджено

Дата погодження, підпис

Завідувач кафедри
Юрій ДОРОФЄЄВ

Дата погодження, підпис

Гарант ОП
Олена ЛОБАЧ

