



Силабус освітнього компонента

Програма навчальної дисципліни

ВИЯВЛЕННЯ АНОМАЛІЙ У ДАНИХ І ЧАСОВИХ РЯДАХ



Шифр та назва спеціальності
113 Прикладна математика

Інститут
ННІ Комп'ютерних наук та інформаційних
технологій

Освітня програма
Інтелектуальний аналіз даних

Кафедра
Комп'ютерної математики і аналізу даних

Рівень освіти
Бакалавр

Тип дисципліни
Спеціальна (фахова), Вибіркова

Семестр
8

Мова викладання
Українська

Викладачі, розробники



Прізвище Ім'я По батькові

Гардер Сергій Євгенійович

Sergei.Garder@khi.edu.ua

Кандидат технічних наук, доцент кафедри комп'ютерної математики і аналізу даних НТУ «ХПІ».

Досвід роботи – 32 років. Автор 89 наукових та навчально-методичних праць. Провідний лектор з дисциплін: «Аналіз даних», «Теорія часових рядів», «Штучні нейронні мережі»

[Детальніше про викладача на сайті кафедри](#)

Загальна інформація

Анотація

Мета викладання дисципліни – Забезпечити підготовку фахівців в галузі прикладної математики, кібербезпеки здатних формулювати, розв'язувати та узагальнювати практичні задачі у своїй професійній діяльності з використанням фундаментальних та спеціальних прикладних методів аналізу даних, теорій часових рядів, нейромережових технологій. Розробляти математичні моделі, алгоритми, створювати та експлуатувати відповідне програмне забезпечення..

Мета та цілі дисципліни

У зрозумілій для студента формі презентовані цілі курсу, знання та навички, які можна здобути в результаті навчання.

Набуття необхідних компетентностей для використання методів виявлення аномалій у даних, ідентифікації і застосування адекватних методів їх усунення:

1. Здатність формалізувати постановку завдань, перевіряти математичну коректність постановки.
2. Здатність обирати та застосовувати спеціальні математичні та чисельні методи для розв'язання практичних задач дослідження та виправлення даних для навчання ШНМ, моделювання, та прогнозування.
3. Здатність розробляти алгоритми та структури даних, програмні засоби та програмну документацію.

4. Здатність до проведення математичного і комп'ютерного моделювання, аналізу та обробки даних, обчислювального експерименту, розв'язання формалізованих задач за допомогою спеціалізованих програмних засобів

Формат занять

Лекції, лабораторні роботи, самостійна робота, консультації. Підсумковий контроль – іспит.

Компетентності

Загальні компетентності:

ЗК 1. Здатність учитися й оволодівати сучасними знаннями.

ЗК 3. Здатність генерувати нові ідеї (креативність).

ЗК 5. Здатність проведення досліджень на відповідному рівні.

ЗК 6. Здатність до абстрактного мислення, аналізу і синтезу.

ЗК 8. Знання і розуміння предметної області та розуміння професійної діяльності.

ЗК 10. Навички у використанні інформаційних і комунікаційних технологій.

Спеціальні (фахові компетентності):

СК 3. Здатність обирати та застосовувати математичні методи для розв'язання прикладних задач, моделювання, аналізу, проектування, керування, прогнозування, прийняття рішень.

СК 18. Здатність обирати та застосовувати математичні моделі та методи для статистичного та інтелектуального аналізу даних в умовах невизначеності.

СК 19. Здатність застосовувати математичні методи та алгоритми машинного навчання, м'яких обчислень і обчислювального інтелекту для аналізу невизначених даних, прогнозування та прийняття рішень.

СК 22. Здатність до використання інформаційних технологій статистичного та інтелектуального аналізу даних, прогнозування, прийняття рішень, інформаційного пошуку та видобування знань.

Результати навчання

РН 1. Демонструвати знання й розуміння основних концепцій, принципів, теорій прикладної математики і використовувати їх на практиці.

РН 2. Володіти основними положеннями та методами математичного, комплексного та функціонального аналізу, лінійної алгебри та теорії чисел, аналітичної геометрії, теорії диференціальних рівнянь, зокрема рівнянь у частинних похідних, теорії ймовірностей, математичної статистики та випадкових процесів, чисельними методами.

РН 3. Формалізувати задачі, сформульовані мовою певної предметної галузі; формулювати їх математичну постановку та обирати раціональний метод вирішення; розв'язувати отримані задачі аналітичними та чисельними методами, оцінювати точність та достовірність отриманих результатів.

РН 14. Виявляти здатність до самонавчання та продовження професійного розвитку.

РН 21. Знати та розуміти сучасні методи розв'язання математичних задач статистичного і інтелектуального аналізу даних, прогнозування тощо.

РН 23. Вміти застосовувати існуючі та розробляти нові алгоритми і програмні засоби для статистичного і інтелектуального аналізу невизначених даних.

РН 24. Вміти застосовувати існуючі та розробляти нові алгоритми і програмні засоби обробки даних вимірювань та спостережень, текстів, сигналів та зображень.

Обсяг дисципліни

Загальний обсяг дисципліни 120 год. (4 кредити ECTS): лекції – 20 год., лабораторні роботи – 20 год., самостійна робота – 80 год.

Передумови вивчення дисципліни (пререквізити)

«Чисельні методи», «Аналіз даних та часових рядів», «Нейромережеві технології».

Особливості дисципліни, методи та технології навчання

При викладанні дисципліни «Виявлення аномалій у даних і часових рядах» передбачено застосування активних та інтерактивних методів навчання лекцій, лабораторних занять на реальних даних у активній формі, колективного обговорення задач. Ефективність навчального процесу виявляється у підвищенні самосвідомості студентів; формуванню здатності ухвалювати самостійні рішення, придбанні навичок колективного обговорення проблем; розвитку здібностей до аналізу і логік. Навчальні матеріали доступні студентам через OneNote, Teams.

Програма навчальної дисципліни

Теми лекційних занять

Тема 1. Вступ Поняття часового ряду. Часовий ряд, як дискретний випадковий процес Генезис даних часового ряду, завдання і цілі дослідження часових рядів. Методи згладжування часового ряду. Адитивна модель часового ряду. Алгоритм потрібного згладжування, S (ARIMA), бустинг. Аналітичні методи виключення детермінованої складової. Ідентифікація тренду часового ряду аналітичними методами і аналіз ряду залишків. Трендовий сплайн-аналіз

Тема 2. Завдання виявлення аномалій у даних.

Області застосування. Прогнозування поломок обладнання, Прогнозування шахрайських дій, виявлення аномальних споживчих патернів, Виявлення аномального попиту та навантаження Первинний аналіз часового ряду. Критерій Ірвіна для пошуку аномальних значень.

Тема 3. Види аномалій

Завдання виявлення аномалій у даних

точкові аномалії, коли спостерігається відхилення у поведінці в окремих точках; групові аномалії, в яких аномально поведуться група точок, аномалії контексту, коли аномалія пов'язана із зовнішніми даними, що не стосуються значень ряду (наприклад, негативна температура на вулиці влітку).

Тема 4 Методи виявлення аномалій.

Метод опорних векторів з одним класом One-Class SVM Метод ізолюючого лісу – isolate forest Метод "Elliptic envelope" та статистичні методи.

Тема 5 Методи виявлення аномалій.

Метричні методи (алгоритми "к найближчих сусідів", "к-го найближчого сусіда", ABOD (angle-based outlier detection) чи LOF (local outlier factor)). Кластерні методи.

Тема 5 (продовження). Методи виявлення аномалій. Метод головних компонентів.

Тема 6 Застосування рекуррентних нейромереж

Тема 5. (продовження). Регресійні методи виявлення аномалій.

Теми практичних занять

Теми лабораторних робіт

1. Методи згладжування часового ряду. Первинна перевірка наявності аномальних значень та їх виключення.
2. Аналітичні методи виключення детермінованої складової. Ідентифікація тренду
3. Метод опорних векторів з одним класом One-Class SVM
4. алгоритми "к найближчих сусідів", "к-го найближчого сусіда".
5. Метод головних компонентів.
6. Регресійні методи виявлення аномалій.
7. Кластеризація для пошуку аномалій .
8. Кластеризація для пошуку аномальних значень (з використанням ШНМ Кохонена, карта Кохонена).
9. Моделювання часових рядів за допомогою рекуррентної мережі Елмана.

Самостійна робота

Курс передбачає виконання індивідуального розрахункового завдання з моделювання (з використанням PИTON) реального часового ряду або набору даних по вибіру студента. Результат розрахунків та моделювання оформлюється у письмовий звіт. Студентам також рекомендуються додаткові матеріали (статті [9], [10], [11]) для самостійного вивчення та аналізу.

Література та навчальні матеріали

1. Chandola V., Banerjee A., Kumar V. Anomaly detection: A survey. – 2009. – Access mode: https://web.archive.org/web/20140211053131/http://www.cs.umn.edu/tech_reports_upload/tr2007/07-017.pdf
2. Prajakta S.K. Time series Forecasting using HoltWinters Exponential Smoothing. – Access mode: https://caohock24.files.wordpress.com/2012/11/04329008_exponentialsMOOTHING.pdf.
3. Kingsland S. E. Modeling nature. – 1995. – Access mode: <https://press.uchicago.edu/ucp/books/book/chicago/M/bo3630803.html>
4. “TadGAN: Time Series Anomaly Detection Using Generative Adversarial Networks”, <https://arxiv.org/abs/2009.07769>
5. “An Introductory Study on Time Series Modeling and Forecasting”, описание SARIMA <https://arxiv.org/ftp/arxiv/papers/1302/1302.6613.pdf>
6. <https://arxiv.org/abs/1808.03314>
7. Cycle Consistency Loss, <https://paperswithcode.com/method/cycle-consistency-loss>
8. [https://github.com/CyberLympha/Examples/tree/main/Разбор статей/TadGAN](https://github.com/CyberLympha/Examples/tree/main/Разбор%20статей/TadGAN)
9. Бібліотека для розпізнавання аномалій в сигналах, <https://github.com/signals-dev/Orion>
10. Алгоритм Dynamic Time Warping, <https://towardsdatascience.com/dynamic-time-warping-3933f25fcdd>
11. <https://medium.com/mit-data-to-ai-lab/time-series-anomaly-detection-in-the-era-of-deep-learning-dccb2fb58fd>
12. <https://medium.com/mit-data-to-ai-lab/time-series-anomaly-detection-in-the-era-of-deep-learning-f0237902224a>

Система оцінювання

Критерії оцінювання успішності студента та розподіл балів

Опис структури підсумкової оцінки, обов'язкових завдань та процедури нарахування балів, особливо звертаючи увагу на самостійну роботу та індивідуальні завдання.

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
31.08.2023 р.



Завідувач кафедри
Олена АХІСЗЕР

Дата погодження, підпис
31.08.2023 р.



Гарант ОП
Олена АХІСЗЕР