



Силабус дисципліни «Байєсівський аналіз даних»



Викладач: *Калініна Ірина Олександрівна*

д-р техн. наук, доцент, в. о. професора кафедри інтелектуальних інформаційних систем

Очікувані результати навчання

В результаті вивчення дисципліни студент

має знати:

- математичну формалізацію та особливості застосування байєсівської методології аналізу даних;
- основні типи ймовірнісно-статистичних моделей, що використовуються у байєсівському аналізі даних;
- методи та процедури врахування невизначеностей у ймовірнісно-статистичному байєсівському моделюванні;
- математичну формалізацію та особливості застосування теореми та мережі Байєса для дискретних і неперервних даних;
- особливості ймовірнісних методів та підходів до поглибленого аналізу даних, виконання теоретичних і прикладних наукових досліджень, удосконалення знань і навичок стосовно практичного використання отриманих наукових результатів для підтримки прийняття відповідних рішень системного характеру згідно з постановками задач;
- особливості порівняльного аналізу результатів, отриманих іншими дослідниками, створення нових концепцій, методів, теоретичних основ і алгоритмів ймовірнісно-статистичного моделювання, прогнозування та підтримки прийняття рішень в економіці, фінансах, бізнесі і на виробництві.

має вміти:

Обсяг дисципліни: 5 кредити ECTS (18 годин лекцій, 36 годин практичних робіт).

Мета: розвиток загальних і фахових компетентностей стосовно здійснення збору і аналізу даних байєсівськими ймовірнісно-статистичними методами, освоєння аналітичних ймовірнісних методів та підходів до поглибленого аналізу даних, виконання теоретичних і прикладних наукових досліджень, удосконалення знань і навичок стосовно практичного використання отриманих наукових результатів для підтримки прийняття відповідних рішень системного характеру згідно з постановками задач; порівняльного аналізу результатів, отриманих іншими дослідниками, створення нових концепцій, методів, теоретичних основ і алгоритмів ймовірнісно-статистичного моделювання, прогнозування та підтримки прийняття рішень в економіці, фінансах, бізнесі і на виробництві.

Оригінальність навчальної дисципліни:

Авторський курс

Зміст дисципліни

Тема 1. Вступ до байєсівської методології аналізу даних. Методологія байєсівського програмування. Структурні компоненти байєсівської методології аналізу даних. Особливості байєсівської методології аналізу даних та експертних оцінок. Порівняння класичного ймовірнісного аналізу даних та байєсівського підходу.

Тема 2. Основні типи ймовірнісно-статистичних моделей, що використовуються у байєсівському аналізі даних. Особливості використання байєсівських моделей різних типів. Означення ймовірнісно-статистичної моделі. Структура байєсівської моделі та її оцінювання. Послідовність побудови ймовірнісно-статистичної моделі байєсівського типу.

Тема 3. Невизначеності у ймовірнісно-статистичному моделюванні. Зв'язок з принципами системного аналізу. Типи невизначеностей, що зустрічаються у

- самостійно проводити дослідницьку діяльність з використанням ймовірісно-статистичних байєсівських методів обробки даних, включаючи поглиблений аналіз задач, проблем, постановку цілей і завдань, вибір засобів та методів дослідження, а також аналізувати оцінки його якості;
- започаткувати, планувати, реалізовувати та корегувати послідовний процес наукового дослідження з використанням ймовірісно-статистичного байєсівського моделювання, прогнозування та підтримки прийняття рішень.

Пререквізити

Викладання навчальної дисципліни базується на знаннях, отриманих у результаті вивчення попередніх навчальних дисциплін та набуття компетенцій після завершення навчання на рівні бакалавра, потребує базових знань з математичних та інформаційних дисциплін, достатніх для сприйняття методів і моделей, що ґрунтуються на використанні ймовірісно-статистичного байєсівського підходу до моделювання, прогнозування та підтримки прийняття рішень. Зокрема, це «Вища математика», «Системний аналіз», «Інтелектуальний аналіз даних», «Теорія ймовірностей та математична статистика», «Теорія прийняття рішень», «Методи і системи машинного навчання».

Пореквізити

Дисципліна надає здобувачам третього ступеня вищої освіти необхідні знання та практичні навички для накопичення та аналізу наукової інформації стосовно ймовірісно-статистичних методів аналізу статистичних/ експериментальних даних за темою кваліфікаційної роботи, побудови та аналізу адекватності створених ймовірісно-статистичних моделей, призначених для розв'язання задач прогнозування коротко- та середньострокового прогнозування розвитку процесів та подій і підтримки прийняття відповідних управлінських рішень.

Семестровий контроль: залік

Оцінювання:

ймовірісно-статистичному моделюванні. Ідентифікація невизначеностей, пов'язаних із статистичними даними та експертними оцінками. Аналіз невизначеностей у методології системного аналізу.

Тема 4. Методи та процедури врахування невизначеностей у ймовірісно-статистичному байєсівському моделюванні. Врахування невизначеностей ймовірісного типу. Статистичні невизначеності та їх врахування у процедурах моделювання. Інтервальне оцінювання та його використання.

Тема 5. Функція і принцип правдоподібності, означення та застосування. Приклади обчислення. Особливості ймовірісно-статистичного моделювання з використанням функцій правдоподібності. Використання принципу правдоподібності у методах системного аналізу. Обчислення оптимальних оцінок параметрів математичних моделей.

Тема 6. Теорема Байєса для дискретних даних і дискретних параметрів. Теорема Байєса для неперервних даних і дискретних параметрів. Формулювання, інтерпретація та можливості застосування. Аналіз складових елементів теореми. Аналіз прикладів застосування.

Тема 7. Мережі Байєса для дискретних і неперервних даних та експертних оцінок. Формулювання, побудова, інтерпретація та можливості застосування. Послідовність побудови ймовірісно-статистичної моделі у формі байєсівської мережі. Дискретизація неперервних даних. Формування ймовірісного висновку за мережею та його інтерпретація. Приклади застосування.

Тема 8. Представлення моделей у просторі станів. Переваги та недоліки. Методи простору станів: фільтри Калмана для лінійної гаусової моделі; приховані марковські моделі; Байєсівський структурний часовий ряд. Приклади представлення моделей у просторі станів.

Тема 9. Байєсівські структурні часові ряди (BSTS). Байєсівські принципи аналізу часових рядів. Алгоритм навчання BSTS-моделі з предикторами та без предикторів.

Критерії оцінювання лабораторних/практичних/індивідуальних/робіт/ доповідей/проєктів

Максимальна кількість балів – студент з високою якістю самостійно виконав весь обсяг робіт, відповідає на всі питання, пов'язані з виконаними роботами, та робить додаткові розрахунки, які йому пропонує викладач. У викладача немає претензій щодо реалізації та вимог до виконання роботи.

За семестр: 70 балів

За залік: 30 балів

Види робіт:

Практичні роботи

Технічне забезпечення

Заняття з дисципліни проводяться у дистанційному форматі (*GoogleMeet, Zoom* тощо) або у комп'ютерних класах з використанням:

- PC зі швидкісним доступом до мережі Internet;
- ОС Windows 10;
- Браузери: Chrome / Opera / Mozilla Firefox / MS Edge.
- Ноутбук, проектор, екран.
- Комплект слайд-презентацій по курсу.
- Програмне забезпечення для демонстрацій слайд-презентацій.
- Мови та середовища програмування Python, R, RStudio

Політика щодо дедлайнів

Роботи, які здаються із порушенням термінів без поважних причин, оцінюються на нижчу оцінку.

Політика щодо академічної доброчесності

Передбачає самостійне виконання лабораторних робіт. Списування під час іспиту (в т. ч. із використанням мобільних пристроїв) заборонено. У разі виявлення плагіату або списування роботи не зараховуються.

70%-99% від максимальної кількості балів – студент з достатньою якістю виконав всі завдання, але в процесі роботи він робив деякі помилки, які, після вказування на них викладачем, самостійно виправляв. На деякі питання він відповідає з похибкою. Запропоновані викладачем додаткові розрахунки робить з деякою потугою. Не всі вимоги до виконання роботи дотримані.

40%-69% від максимальної кількості балів – студент самостійно виконав всі роботи, але якість реалізації недостатня (помилки при розрахунках, не всі вимоги до роботи дотримані). На питання щодо виконання робіт відповідає не зовсім чітко. Є помилки при відповідях.

1%-39% від максимальної кількості балів – студент самостійно виконав не всі роботи, при цьому якість реалізації недостатня (помилки при розрахунках, не дотримується вимог до оформлення роботи). На питання щодо виконання робіт відповідає не чітко. Є грубі помилки при відповідях.

0 балів – студент не виконав весь обсяг робіт, або виконав з грубими помилками. Він має проблеми з розрахунками, не знає теоретичного матеріалу, програмна реалізація не відповідає поставленим вимогам.

Для зарахування практичної роботи здобувач має змогу надати сертифікат (або інший документ) про проходження курсів на платформах *Prometheus, Coursera* тощо та за умови співпадіння направленості курсів блоку освітньої компоненти.