



# Байєсівський аналіз даних у наукових дослідженнях

## Робоча програма навчальної дисципліни (силабус)

### Реквізити навчальної дисципліни

Рівень вищої освіти	<i>Третій (доктор філософії PhD)</i>
Галузь знань	<i>12 Інформаційні технології</i>
Спеціальність	<i>122 Комп'ютерні науки</i>
Освітня програма	<i>Комп'ютерні науки</i>
Статус дисципліни (код)	<i>Вибіркова</i>
Форма навчання	<i>очна(денна)/дистанційна/змішана</i>
Рік підготовки, семестр	<i>2 курс, осінній семестр</i>
Обсяг дисципліни	<i>4 кредитів ЄКТС</i>
Семестровий контроль/ контрольні заходи	<i>Залік</i>
Розклад занять	<i>Rozklad.kpi.ua</i>
Мова викладання	<i>Українська</i>
Інформація про керівника курсу / викладачів	Лектор: <i>д.т.н., професор, Бідюк Петро Іванович, pbidyuke_00@ukr.net</i> Практичні: <i>д.т.н., професор, Бідюк Петро Іванович</i>
Розміщення курсу	<i>Googleclassroom</i>

### Програма навчальної дисципліни

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

*Курс призначено для здобувачів третього рівня вищої освіти (PhD) навчання очної та заочної форми. Висвітлено основні поняття байєсівського підходу до аналізу статистичних та експериментальних даних. Викладено принципи, особливості застосування, методи, структуру і технології теоретичних та практичних занять з байєсівського аналізу даних (БДА) та його зв'язку з принципами системного аналізу. Викладення матеріалу акцентовано на системному підході до збору, аналізу даних, побудови ймовірно-статистичних моделей, аналізу їх адекватності та подальшого практичного застосування. Значну увагу приділено особливостям теоретичного і практичного аналізу даних за допомогою ймовірнісних байєсівських методів моделювання і прогнозування в умовах статистичної невизначеності, розвитку навичок аналізу даних за сучасними теоретичними положеннями ймовірно-статистичного підходу, що відносяться до курсу БДА. Докладно висвітлено вимоги до створення ймовірно-статистичних моделей багатовимірних систем та аналізу їх адекватності за системним підходом. Курс розроблено таким чином, щоб надати необхідні теоретичні знання для опанування основних методів байєсівського аналізу даних та їх практичного використання для побудови моделей, оцінювання ймовірнісного висновку, коротко- і середньострокового прогнозування розвитку нелінійних нестационарних процесів у різних галузях діяльності людини. Тому у курсі представлено як загальні методи, моделі і методики виконання досліджень з*

використанням байєсівських методів, так і конкретні кроки з метою формування практичних навичок стосовно розв'язання конкретних завдань ймовірно-статистичного моделювання та підтримки прийняття рішень на основі отриманих результатів моделювання і прогнозування.

Метою викладання дисципліни є розвиток загальних і фахових компетентностей стосовно здійснення збору і аналізу даних байєсівськими ймовірно-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 ймовірно-статистичного моделювання, прогнозування та підтримки прийняття рішень в економіці, фінансах, бізнесі і на виробництві.

У процесі навчання здобувачі третього рівня вищої освіти мають оволодіти такими компетентностями: ЗК4 «Здатність самостійно проводити дослідницьку діяльність з використанням ймовірно-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», ЗК5 «Здатність до започаткування, планування, реалізації та корегування послідовного процесу ґрунтовного наукового дослідження з використанням ймовірно-статистичного байєсівського моделювання, прогнозування та підтримки прийняття рішень, а також дотриманням належної академічної доброчесності», ФК8 «Здатність планувати і проводити наукові дослідження з використанням ймовірно-статистичного байєсівського моделювання, прогнозування та підтримки прийняття рішень, готувати, презентувати та публікувати результати науково-дослідницької діяльності, отримані у процесі застосування ймовірно-статистичних байєсівських методів моделювання, прогнозування та підтримки прийняття рішень».

По завершенню курсу здобувачі третього рівня вищої освіти мають набути такі програмні результати навчання: ЗН5 «Знати основи організації дослідницького наукового процесу для розв'язання значущих прикладних задач і проблем в галузі системного аналізу об'єктів і процесів, використовуючи сучасні методи ймовірно-статистичного байєсівського моделювання, прогнозування та підтримки прийняття рішень, визначати і застосовувати ключові методологічні принципи та методи дослідження системного аналізу відповідно до об'єкту і предмету конкретного дослідження, використовуючи відповідні міждисциплінарні підходи»; УМ5 «Уміти впроваджувати результати наукових досліджень, які ґрунтуються на основі методів і принципів системного аналізу та ймовірно-статистичного байєсівського моделювання, прогнозування та підтримки прийняття рішень»; УМ7 Уміти з нових дослідницьких позицій формулювати і використовувати для розв'язання практичних задач загальну методологічну і конкретну базу власного наукового дослідження, ймовірно-статистичні методи байєсівського моделювання, прогнозування і підтримки прийняття рішень та інші методи із суміжних областей, усвідомлювати його актуальність та необхідність, мету і значення для розвитку інших галузей науки».

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

**Пререквізити:** викладання навчальної дисципліни базується на знаннях, отриманих у результаті вивчення попередніх навчальних дисциплін та набуття компетенцій після завершення навчання на рівні бакалавра і магістра з системного аналізу, потребує базових знань з математичних та інформаційних дисциплін, достатніх для сприйняття методів і моделей, що ґрунтуються на використанні ймовірно-статистичного байєсівського підходу до моделювання, прогнозування та підтримки прийняття рішень. Зокрема, це теорія

ймовірностей і прикладна статистика, математичний аналіз, диференціальні рівняння, системи підтримки прийняття рішень.

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

### **3. Зміст навчальної дисципліни**

#### **РОЗДІЛ 1. Вступ до байєсівської методології аналізу даних.**

**Тема 1.1. Методологія байєсівського програмування. Елементи байєсівської методології аналізу даних.**

Тема розкриває такі питання: Структурні компоненти байєсівської методології аналізу даних. Особливості байєсівської методології аналізу даних та експертних оцінок. Порівняння класичного ймовірнісного аналізу даних та байєсівського підходу.

**Тема 1.2. 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 1.3. Невизначеності у ймовірнісно-статистичному моделюванні.** Зв'язок з принципами системного аналізу.

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

**Тема 1.4. Методи та процедури врахування невизначеностей у ймовірнісно-статистичному байєсівському моделюванні.**

Розглядаються питання: врахування невизначеностей ймовірнісного типу. Статистичні невизначеності та їх врахування у процедурах моделювання. Інтервальне оцінювання та його використання.

#### **РОЗДІЛ 2. Варіанти теореми Байєса та мережі Байєса.**

**Тема 2.1. Функція і принцип правдоподібності, означення та застосування.** Приклади обчислення.

В темі розглянуто особливості ймовірнісно-статистичного моделювання з використанням функцій правдоподібності. Використання принципу правдоподібності у методах системного аналізу. Обчислення оптимальних оцінок параметрів математичних моделей.

**Тема 2.2. Теорема Байєса для дискретних даних і дискретних параметрів.**

Формулювання теореми для дискретних даних і дискретних параметрів. Аналіз елементів теореми. Теорема для взаємодоповнюючих подій. Приклади застосування.

**Тема 2.3. Теорема Байєса для неперервних даних і дискретних параметрів. Формулювання, інтерпретація та можливості застосування.**

Формулювання теореми для неперервних даних і дискретних параметрів. Аналіз складових елементів теореми. Аналіз прикладів застосування.

**Тема 2.4. Мережі Байєса для дискретних і неперервних даних та експертних оцінок. Формулювання, побудова, інтерпретація та можливості застосування.**

Послідовність побудови ймовірнісно-статистичної моделі у формі байєсівської мережі. Дискретизація неперервних даних. Формування ймовірнісного висновку за мережею та його інтерпретація. Приклади застосування.

#### 4. Навчальні матеріали та ресурси

##### Базова

1. Згуровський М.З., Бідюк П.І., Терентьєв О.М., Просянкін-Жарова Т.І. Байєсівські мережі у системах підтримки прийняття рішень. – Київ: «Едельвейс», 2015. – 300 с.
2. Bolstad W.M. *Understanding computational Bayesian statistics*. – Hoboken, NJ: John Wiley & Sons Inc., 2010. – 334 p.
3. Jensen F.V., Nielsen T.D. *Bayesian networks and decision graphs*. – Berlin: Springer, 2007. – 457 p.
4. Koop G. *Bayesian econometrics*. – Hoboken, NJ: John Wiley & Sons Inc., 2003. – 373 p.
5. Pourret O., Naim P., Marcot B. *Bayesian networks – practical guide to application*. – Chichester: John Wiley & Sons Ltd, 2008. – 430 с.
6. Pole A., West M., Harrison J. *Applied Bayesian forecasting and time series analysis*. – London: Chapman & Hall/CRC, 1999. – 410 p.
7. Greene W.H. *Econometric analysis*. – London: Prentice Hall International, 2014. – 1004 p.

##### Додаткова

1. Магнус Я.Р., Нейдеккер Х. *Матричное дифференциальное исчисление с приложениями к статистике и эконометрике*. – Москва: Физматлит, 2002. – 496 с.
2. Кимбл Г. *Как правильно пользоваться статистикой*. – Москва: Финансы и статистика, 1982. – 295 с.
3. Ликеш И., Ляга Й. *Основные таблицы математической статистики*. – Москва: Финансы и статистика, 1985. – 356 с.
4. Press S.J. *Subjective and objective Bayesian statistics*. – New York: A John Wiley & Sons, Inc., Publications, 2003. – 558 p.
5. Tsay R.S. *Financial time series analysis*. – New York: John Wiley & Sons, Inc., 2010. – 715 p.
6. Документація. Звіти у сфері науки і техніки. Структура і правила оформлення: ДСТУ 3008-95 (ГОСТ 7.32-91). – Чинний від 2006-01-01. К.: Держспоживстандарт України, 2006. – 213 с. – (Національний стандарт України).

#### Навчальний контент

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

##### Лекційні заняття

№	Назва теми лекції та перелік основних питань
1	<b>Методологія байєсівського програмування. Елементи байєсівської методології аналізу даних.</b> <i>Структурні компоненти байєсівської методології аналізу даних. Особливості байєсівської методології аналізу даних та експертних оцінок. Порівняння класичного ймовірнісного аналізу даних та байєсівського підходу.</i> Рекомендована література: [1] – С. 20 – 45, [5] – С. 10 – 35, [6] – С.22 – 35, [1] – С. 15 –25; [4] – С. 15 – 34.

2	<p><b>Типи ймовірно-статистичних моделей, що використовуються у байєсівському аналізі даних.</b></p> <p>Означення ймовірно-статистичної моделі. Поняття структури байєсівської моделі та її оцінювання. Послідовність побудови ймовірно-статистичної моделі байєсівського типу.</p> <p>Рекомендована література: [1] – С. 30 –55; [5] – С. 8 – 24; [6] – С. 29 – 36. [1] – С. 66 –77; [6] – С.37 – 50.</p>
3	<p><b>Невизначеності у ймовірно-статистичному моделюванні.</b></p> <p>Зв'язок аналізу невизначеностей з принципами системного аналізу.</p> <p>Типи невизначеностей, що зустрічаються у ймовірно-статистичному моделюванні. Ідентифікація невизначеностей, пов'язаних із статистичними даними та експертними оцінками. Аналіз невизначеностей у методології системного аналізу.</p> <p>Рекомендована література: [1] – С. 46 –54; [5] – С.37 – 50.</p>
4	<p><b>Методи та процедури врахування невизначеностей у ймовірно-статистичному байєсівському моделюванні.</b></p> <p>Розглядаються питання: врахування невизначеностей ймовірного типу. Статистичні невизначеності та їх врахування у процедурах моделювання. Точкове та інтервальне оцінювання та їх використання в аналізі невизначеності.</p> <p>Рекомендована література: [1] – С. 60–68; [4] – С. 72 – 78; [6]- С. 23 – 28.</p>
5	<p><b>Функція і принцип правдоподібності, означення та застосування. Приклади обчислення.</b></p> <p>Розглядаються особливості ймовірно-статистичного моделювання з використанням функцій правдоподібності. Використання принципу правдоподібності у методах системного аналізу. Обчислення оптимальних оцінок параметрів математичних моделей.</p> <p>Рекомендована література: [1] – С. 55 –60; [5] – С. 51 – 68; [6]- С. 37 – 55.</p>
6	<p><b>Теорема Байєса для дискретних даних і дискретних параметрів.</b></p> <p>Формулювання теореми для дискретних даних і дискретних параметрів. Аналіз елементів теореми. Теорема для взаємодоповнюючих подій. Приклади застосування.</p> <p>Рекомендована література: [1] – С. 30 –35; [4] – С. 180 – 214; [6]- С. 257 – 288.</p>
7	<p><b>Теорема Байєса для неперервних даних і дискретних параметрів. Формулювання, інтерпретація та можливості застосування.</b></p> <p>Формулювання теореми для неперервних даних і дискретних параметрів. Аналіз та інтерпретація складових елементів теореми. Аналіз прикладів застосування.</p> <p>Рекомендована література: [1] – С. 30 –35; [4] – С. 185 – 230; [6]- С. 260 – 320.</p>
8	<p><b>Мережі Байєса для дискретних і неперервних даних та експертних оцінок. Формулювання, побудова, інтерпретація та можливості застосування.</b></p> <p>Послідовність побудови ймовірно-статистичної моделі у формі байєсівської мережі. Дискретизація неперервних даних. Формування ймовірного висновку за мережею та його інтерпретація. Приклади застосування.</p> <p>Рекомендована література: [1] – С. 30 –35; [4] – С. 185 – 230; [5]- С. 20 – 220.</p>
9	<p><b>Нормативно-правове забезпечення проведення та оформлення результатів виконання наукових дослідження.</b></p> <p>В темі розглянуто Закони України, Постанови Кабінету Міністрів України, Державні стандарти України .</p> <p>Рекомендовані ресурси: <a href="#">Про затвердження Вимог до оформле...   від 12.01.2017 № 40 (rada.gov.ua)</a>, <a href="#">Про затвердження Вимог до оформле...   від 12.01.2017 № 40 (rada.gov.ua)</a>, <a href="https://rada.kpi.ua/node/1136">https://rada.kpi.ua/node/1136</a></p>



## Практичні заняття

№	Назва теми занять
1	Методи обробки статистичних даних та експертних оцінок з використанням методології байєсівського аналізу.
2	Застосування методики побудови ймовірісно-статистичних моделей за фактичними даними. Побудова та аналіз багатовимірних розподілів.
3	Аналіз критеріальної бази байєсівського підходу до аналізу даних.
4	Побудова прогнозних розподілів з використанням байєсівського підходу.
5	Побудова байєсівської мережі за експертними оцінками.
6	Побудова байєсівської мережі за статистичними даними.
7	Модульна контрольна робота

### 6. Самостійна робота здобувача третього рівня вищої освіти (PhD) (94 год.)

Самостійна робота здобувача складається з виконання індивідуального завдання – виконання самостійного завдання на одну із запропонованих тем:

1. Побудова і використання функцій правдоподібності для оцінювання параметрів математичних моделей.
2. Оцінювання параметрів ймовірісно-статистичних моделей за допомогою методу Монте-Карло для марковських ланцюгів.
3. Побудова узагальнених лінійних моделей з використанням байєсівського підходу до аналізу даних.
4. Прогнозування розподілів за байєсівським методом.
5. Побудова функцій прогнозування на основі моделей байєсівського типу.
6. Оцінювання структури і параметрів математичних моделей нелінійних нестационарних процесів.
7. Побудова ймовірісно-статистичної моделі у формі мережі Байєса на основі експертних оцінок.
8. Побудова мережі Байєса для розв'язання задачі короткострокового прогнозування.
9. Побудова мережі Байєса для розв'язання задачі технічної діагностики.
10. Побудова мережі Байєса для розв'язання задачі медичної діагностики.
11. Побудова мережі Байєса для розв'язання задачі фінансово-економічної діагностики.
12. Побудова мережі Байєса для розв'язання задачі підтримки прийняття рішень у бізнесі.
13. Побудова комбінованої моделі на основі мережі Байєса і регресійної моделі.
14. Застосування методів фільтрації при побудові ймовірісно-статистичних моделей.
15. Застосування оптимальної та гранулярної фільтрації для попередньої обробки даних.

## Політика та контроль

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

Здобувачі вищої освіти (PhD) не мають право пропускати лекційні та практичні заняття без поважних причин. На кожному практичному занятті здобувачі повинні активно залучатися до аналізу, обговорення та розв'язування поставлених задач. Для цього викладач на кожній лекції повинен приділяти увагу до застосування викладених теоретичних основ прочитаних тем в різних галузях прикладної науки. Захист індивідуального завдання повинен виявити наскільки здобувач може не тільки абстрактно та логічно мислити, а й аналізувати результати виконаного практично спрямованого дослідження. Усі роботи здобувачі мають прикріплювати в особистому кабінеті гугл-класу. Роботи мають бути виконані з дотриманням академічної доброчесності.

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

Семестровий контроль: залік

Семестровий рейтинг з дисципліни «Байєсівський аналіз даних у наукових дослідженнях» складається з елементів, поданих нижче.

1. Система оцінювання з виконання наукових досліджень та представлення наукових результатів з використанням байєсівського аналізу даних складається з рейтингових балів (табл.1), і не перевищує  $R_{\max} = 100$ . В семестрі здобувач може набрати 60 балів, відповідно на заліку – 40 балів.

Таблиця 1. Система рейтингових балів.

№	Контрольний захід	Бали
1.	Виконання самостійного завдання	30
2.	Модульна контрольна робота «Методи байєсівського аналізу даних – практичне застосування»	15
3.	Активна робота на практичних заняттях	15

2. Самостійне завдання зараховується тільки за умови його захисту. Для захисту самостійного завдання надається не більше трьох спроб. Залежно від того, з якої спроби була захищена робота, нараховується така кількість балів:

- захист з першої спроби - 30 балів;
- захист з другої спроби -20 балів;
- захист з третьої спроби (останній) – 10 балів.

3. Здобувач допускається до заліку при виконанні таких умов:

- поточний рейтинг за семестр складає не нижче 30 балів;
- захищена самостійне завдання.

Відповідно сумарної кількості балів, що набрані в семестрі та на заліку, здобувач вищої освіти (PhD) отримує оцінку згідно таблиці 2.

Таблиця 2 відповідності рейтингових балів оцінкам за університетською шкалою:

Рейтинг	Оцінка ECTS	Традиційна оцінка
95 - 100	<b>A</b> — відмінно	Відмінно
85 - 94	<b>B</b> — дуже добре	Добре
75 - 84	<b>C</b> — добре	
65 - 74	<b>D</b> — задовільно	Задовільно
60 - 64	<b>E</b> — достатньо	
менше 60 балів	<b>FX</b> — незадовільно	Незадовільно
менше 30 балів	<b>F</b> — не допущено	Не допущено

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

Теоретичні питання

1. Система методів, методик, моделей і алгоритмів у предметній області байєсівського аналізу даних.

2. *Методи і способи виконання наукових досліджень на емпіричному та теоретичному рівнях з використанням байєсівського аналізу даних.*
3. *Методи байєсівського аналізу даних у системному аналізі і системах підтримки прийняття рішень.*
4. *Оцінювання ефективності результатів застосування байєсівського аналізу даних при виконання наукових досліджень.*
5. *План проведення досліджень з використанням байєсівського аналізу даних.*
6. *Методика застосування теоретичних основ байєсівського аналізу даних.*
7. *Вимоги до представлення результатів дослідження з використанням байєсівського аналізу даних.*
8. *Правила публікацій результатів наукових досліджень з використанням методології байєсівського аналізу даних.*
9. *Загальні правила оформлення тез доповідей, презентацій на конференціях, наукових семінарах на тематику, пов'язану з байєсівським аналізом даних.*
10. *Виступи на наукових семінарах, конференціях за тематикою байєсівського аналізу даних.*
11. *Порядок викладення результатів теоретичних досліджень та практичних результатів комп'ютерних експериментів з використанням байєсівської методології моделювання, прогнозування та підтримки прийняття рішень.*

**Робочу програму навчальної дисципліни (силабус):**

**Склав:** *д.т.н., професор, Бідюк Петро Іванович*

**Ухвалено** кафедрою ММСА (протокол № 9 від 24.06.2020)

**Погоджено** Методичною комісією ІПСА (протокол № 9 від 25.06.2020)