



Статистичні методи машинного навчання

Робоча програма кредитного модуля (Силабус)

Реквізити навчальної дисципліни

Рівень вищої освіти	Перший (бакалаврський)
Галузь знань	12 Інформаційні технології
Спеціальність	121 Інженерія програмного забезпечення 123 Комп'ютерна інженерія 126 Інформаційні системи та технології
Освітня програма	
Статус дисципліни	Вибіркова
Форма навчання	Очна (денна)
Рік підготовки, семестр	4 курс, весняний семестр
Обсяг дисципліни	4 кредити, 120 годин – 54 години аудиторної роботи, 66 годин самостійної роботи
Семестровий контроль, контрольні заходи	Залік
Розклад занять	Згідно розкладу на весняний семестр поточного навчального року за посиланням schedule.kpi.ua
Мова викладання	Українська
Інформація про керівника курсу, викладачів	Асистент кафедри обчислювальної техніки, доктор філософії Шульга Максим Володимирович, shulha.maksym@iit.kpi.ua
Розміщення курсу	Онлайн лекції: https://bbb.comsys.kpi.ua/rooms/ik8-px6-c5a-qpl/join Методичні матеріали: https://classroom.google.com/c/ODE0MjM0OTAyMzA5?cjc=pchafitd

Програма навчальної дисципліни

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

Метою навчальної дисципліни є отримання студентами ґрунтовної підготовки з теоретичних, методологічних та практичних основ статистичних методів машинного навчання. Це включає розуміння, як здобувати та очищувати дані, проведення дослідницького аналізу даних, розуміння процесу статистичного висновування, використання регресійних моделей та застосування всього цього на практиці у завданнях машинного навчання.

Предметом навчальної дисципліни є статистичні методи машинного навчання, що пропонують студентам повне розуміння фундаментальних статистичних концепцій та їх ефективного застосування в області машинного навчання.

Завданням вивчення навчальної дисципліни є:

- ефективне управління та попередня обробка даних за допомогою R, отримання інформації з різноманітних джерел;

- створення інформативної аналітичної графіки та застосування статистичних методів для детального дослідження даних;
- отримання навичок перевірки гіпотез, включаючи побудову довірчих інтервалів та інтерпретацію р-значень;
- реалізація різноманітних регресійних моделей, що охоплюють лінійну, багатофакторну, логістичну та пуассонівську регресії;
- розробка прогнозних моделей, вмiла оцінка помилок і застосування методів перехресної перевірки;
- використання моделей на основі дерев, таких як випадкові ліси, та використання передбачень на основі моделей.

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

При вивченні цієї дисципліни використовуються знання студентів з дисциплін: “Вища математика”, “Аналітична геометрія”, “Теорія ймовірності та математична статистика”.

3. Зміст навчальної дисципліни

Розділ 1. Отримання і очищення даних

Тема 1.1. Пошук і зчитування даних, системи зберігання даних

Тема 1.2. Управління даними та маніпуляції з даними в R

Розділ 2. Дослідницький аналіз даних

Тема 2.1. Аналітична графіка та система побудови графіків в R

Тема 2.2. Статистичні методи для дослідницького аналізу даних

Розділ 3. Статистичне висновування

Тема 3.1. Ймовірність і очікувані значення

Тема 3.2. Дисперсія, розподіл і асимптотика

Тема 3.3. Інтервали, тестування гіпотез та р-значення

Тема 3.4. Потужність, бутстрепінг і пермутаційні тести

Розділ 4. Регресійні моделі

Тема 4.1. Метод найменших квадратів і лінійна регресія

Тема 4.2. Лінійна регресія, залишки та висновування

Тема 4.3. Множинна регресія

Тема 4.4. Логістична регресія та регресія Пуассона

Розділ 5. Практичне машинне навчання

Тема 5.1. Прогнозування, помилки та перехресна перевірка

Тема 5.2. Пакет caret

Тема 5.3. Прогнозування за допомогою дерев, випадкових лісів і прогнозування на основі моделі

Тема 5.4. Регуляризована регресія та комбінування предикторів

4. Навчальні матеріали та ресурси

Базові ресурси:

1. Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani: Introduction to Statistical Learning. – Ed. 1. – Springer, 2013. – 426 p.
2. Pratap Dangeti: Statistics for Machine Learning. – Packt Publishing, 2017. – 442 p.
3. Max Kuhn and Kjell Johnson: Applied Predictive Modeling. – Springer, 2013. – 600 p.
4. Roger D. Peng and Elizabeth Matsui: The Art of Data Science. – Leanpub, 2020. – 428 p.

5. Trevor Hastie, Robert Tibshirani, and Jerome Friedman: The Elements of Statistical Learning. – Springer, 2009. – 745 p.
6. Tom Mitchell: Machine Learning. – McGraw-Hill, 1997. – 414 p.
7. Michael Lavine: Introduction to Statistical Thought. // [Electronic resource]. – URL: <http://www.stat.cmu.edu/~cshalizi/350/>

Додаткові ресурси:

1. Aurélien Géron: Hands-On Machine Learning with Scikit-Learn and TensorFlow. – O'Reilly Media, 2019. – 856 p.
2. Sebastian Raschka and Vahid Mirjalili: Python Machine Learning. – Packt Publishing, 2017. – 454 p.
3. Hadley Wickham: Advanced R. – CRC Press, 2019. – 476 p.

Обладнання, що необхідне для проведення занять

Лекційні заняття проводяться в аудиторії, яка обладнана проектором, лабораторні заняття – в комп'ютерній лабораторії. Для дистанційних занять потрібен доступ до Інтернету та програмне й апаратне забезпечення для проведення відеоконференцій.

Навчальний контент

5. Методика опанування навчальної дисципліни (освітнього компонента)

Назви розділів, тем	Розподіл навчального часу				
	Всього	в тому числі			
		Лекції	Практичні (семінарські) заняття	Лабораторні роботи (комп'ютерний практикум)	CPC
1	2	3	4	5	6
Розділ 1. Отримання і очищення даних					
Тема 1.1. Пошук і зчитування даних, системи зберігання даних	4	2		1	1
Тема 1.2. Управління даними та маніпуляції з даними в R	4	2		1	1
Разом за розділом 1	8	4	0	2	2
Розділ 2. Дослідницький аналіз даних					
Тема 2.1. Аналітична графіка та система побудови графіків в R	5	2		1	2
Тема 2.2. Статистичні методи для дослідницького аналізу даних	5	2		1	2
Разом за розділом 2	10	4	0	2	4
Розділ 3. Статистичне висновлення					
Тема 3.1. Ймовірність і очікувані значення	7	2		1	4

1	2	3	4	5	6
Тема 3.2. Дисперсія, розподіл і асимптотика	7	2		1	4
Тема 3.3. Інтервали, тестування гіпотез та р-значення	7	2		1	4
Тема 3.4. Потужність, бутстрепінг і пермутаційні тести	7	2		1	4
Разом за розділом 3	28	8	0	4	16
Розділ 4. Регресійні моделі					
Тема 4.1. Метод найменших квадратів і лінійна регресія	7	2		1	4
Тема 4.2. Лінійна регресія, залишки та висновування	7	2		1	4
Тема 4.3. Множинна регресія	7	2		1	4
Тема 4.4. Логістична регресія та регресія Пуассона	7	2		1	4
Разом за розділом 4	28	8	0	4	16
Розділ 5. Практичне машинне навчання					
Тема 5.1. Прогнозування, помилки та перехресна перевірка	7	2		1	4
Тема 5.2. Пакет caret	7	2		1	4
Тема 5.3. Прогнозування за допомогою дерев, випадкових лісів і прогнозування на основі моделі	8	2		2	4
Тема 5.4. Регуляризована регресія та комбінування предикторів	8	2		2	4
Разом за розділом 5	30	8	0	6	16
Модульна контрольна робота	6				6
Залік	10	4			6
Разом за семестр	120	36	0	18	66

Лекційні заняття

№ з/п	Назва теми лекції та перелік основних питань (перелік дидактичних засобів, посилання на літературу та завдання на СРС)
1	2
1	Тема 1.1. Пошук і зчитування даних, системи зберігання даних Як структурований курс, що таке статистика та її цілі, як знаходити та читати дані з різних типів файлів, а також найпоширеніші системи зберігання даних і відповідні інструментами для отримання даних з Інтернету або з баз даних, таких як MySQL.
2	Тема 1.2. Управління даними та маніпуляції з даними в R Дослідження організації, об'єднання та управління зібраними даними, а також маніпуляції з текстом і датами в R.
3	Тема 2.1. Аналітична графіка та система побудови графіків в R Дослідження основ аналітичної графіки та базової системи побудови графіків у R, і деяких з більш просунутих систем побудови графіків, доступних у R: система Lattice та система ggplot2.
4	Тема 2.2. Статистичні методи для дослідницького аналізу даних Дослідження основних статистичних методів дослідницького аналізу, включаючи методи кластеризації та зменшення розмірності, інноваційні методи визначення кольорів у R.
5	Тема 3.1. Ймовірність і очікувані значення Дослідження основ статистики, включаючи ймовірність, випадкові величини, очікування тощо.
6	Тема 3.2. Дисперсія, розподіл і асимптотика Дослідження дисперсії, розподілів, меж та довірчих інтервалів.
7	Тема 3.3. Інтервали, тестування гіпотез та р-значення Дослідження інтервалів, тестування гіпотез та р-значень.
8	Тема 3.4. Потужність, бутстрепінг і пермутаційні тести Дослідження потужності, бутстрепінгу та пермутаційних тестів.
9	Тема 4.1. Метод найменших квадратів і лінійна регресія Дослідження найменших квадратів та лінійної регресії.
10	Тема 4.2. Лінійна регресія, залишки та висновування Продовження дослідження лінійної регресії, дослідження залишків та висновування у регресії.
11	Тема 4.3. Множинна регресія Дослідження множинної регресії.
12	Тема 4.4. Логістична регресія та регресія Пуассона Дослідження узагальнених лінійних моделей, включаючи бінарні результати та регресію Пуассона.
13	Тема 5.1. Прогнозування, помилки та перехресна перевірка Дослідження прогнозування, відносної важливості кроків, помилки та перехресної перевірки.

1	2
---	---

14	Тема 5.2. Пакет caret Ознайомлення з пакетом caret, інструментом для створення властивостей і попередньої обробки.
15	Тема 5.3. Прогнозування за допомогою дерев, випадкових лісів і прогнозування на основі моделі Дослідження різних алгоритмів машинного навчання.
16	Тема 5.4. Регуляризована регресія та комбінування предикторів Дослідження регуляризованої регресії та комбінованих предикторів.

Лабораторні заняття

№ з/п	Назва лабораторної роботи (комп'ютерного практикуму)	Кількість ауд. годин
1	2	3
1	Отримання і очищення даних (Розділ 1)	2
2	Дослідницький аналіз даних (Розділ 2)	2
3	Статистичне висновування (Розділ 3)	4
4	Регресійні моделі (Розділ 4)	4
5	Прогнозування (Розділ 5)	6
	Разом:	18

6. Самостійна робота студента

У процесі виконання індивідуальних завдань студенти повинні закріпити знання, отримані під час лекційних занять та лабораторних робіт, самостійно вивчати визначені теми, поглиблювати свої знання для подальшого навчання. Самостійна робота студентів полягає в наступному:

- підготовці до лекційних занять шляхом вивчення матеріалу, що розглядався на попередньому лекційному занятті;
- підготовці до лабораторних занять шляхом виконання лабораторних робіт з вивченням теоретичних питань, які розглядаються на лекційних заняттях, та оформленні звіту про виконання кожної лабораторної роботи;
- підготовці до модульної контрольної роботи, тестів та заліку.

Політика та контроль

7. Політика навчальної дисципліни (освітнього компонента)

Всі студенти повинні відвідувати лекційні та лабораторні заняття – як при очному навчанні (фізичне відвідування занять в аудиторіях), так і при дистанційному навчанні (віртуальне відвідування онлайн-занять).

Під час занять студенти повинні дотримуватись певних дисциплінарних правил:

- забороняється запізнюватись на заняття;
- не допускаються сторонні розмови або інший шум, що заважає проведенню занять;

- не допускається користування мобільними телефонами та іншими технічними засобами без дозволу викладача;
- не допускається займатися діяльністю, яка прямо не стосується навчальної дисципліни.

Результати виконання лабораторних робіт повинні бути оформлені у електронному форматі у вигляді файлів-звітів. Такі файли повинні містити результати виконання відповідних завдань та відповідати вимогам та методичним вказівкам для кожної роботи.

Усі письмові роботи перевіряються на наявність плагіату. Наявність плагіату призводить до незадовільної оцінки роботи (0 балів) без можливості повторної здачі роботи.

Лабораторні роботи виконуються командами. Кожна команда складається з не більше 3 (трьох) студентів. Якщо загальна кількість студентів, що вивчають дисципліну, не ділиться націло на 3 (три), то буде сформовано не більше 2 (двох) команд з 4 (чотирьох) студентів. Студенти самостійно формують команди протягом перших 2 (двох) тижнів навчання, після цього терміну студенти, які не сформували команди, будуть розподілені по командах випадковим чином.

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

Підсумкова рейтингова оцінка (R) студента складається з балів, які він отримує:

- за лабораторні роботи ($R_{\text{ЛАБ}}$);
- за тести (R_{T});
- за модульну контрольну роботу ($R_{\text{МКР}}$);
- за залікову роботу (R_3).

Відповідно до ["Положення про систему оцінювання результатів навчання в КПІ ім. Ігоря Сікорського"](#), затвердженого Наказом №1/273 від 14.09.2020 р., застосована система оцінювання типу PCO-1, для якої рейтингова оцінка складається з:

- стартової оцінки (R_C) – оцінювання заходів впродовж семестру: $R_C = R_{\text{ЛАБ}} + R_{\text{T}} + R_{\text{МКР}}$;
- оцінки за залікову роботу (R_3).

Таким чином, $R = R_C + R_3$.

Студенти, які мають стартову оцінку (R_C) не меншу за деяке значення мінімально можливої загальної позитивної оцінки ($R_{\text{МИН}}$) та згодні з тим, що їх стартова оцінка (R_C) дорівнюватиме підсумковій рейтинговій оцінці (R), звільняються від написання залікової роботи.

Студенти, які мають стартову оцінку (R_C) меншу за деяке значення мінімально можливої загальної позитивної оцінки ($R_{\text{МИН}}$) або бажають виконати залікову роботу, отримують стартову оцінку (R_C), що дорівнює оцінці за лабораторні роботи та модульну контрольну роботу ($R_{\text{ЛАБ}} + R_{\text{МКР}}$).

Протягом курсу студенти виконують 5 (п'ять) лабораторних робіт, 4 (чотири) тести та модульну контрольну роботу. Тест вважається контрольним заходом.

Розрахунок шкал оцінювання

Шкала оцінювання (максимально можлива оцінка) лабораторної роботи: 10 балів.

Шкала оцінювання всіх лабораторних робіт: $R_{\text{ЛАБ}} = 10 \text{ балів} \times 5 = 50 \text{ балів}$.

Шкала оцінювання (максимально можлива оцінка) тесту: 10 балів.

Шкала оцінювання всіх тестів: $R_{\text{T}} = 10 \text{ балів} \times 4 = 40 \text{ балів}$.

Шкала оцінювання модульної контрольної роботи: $R_{\text{МКР}} = 10 \text{ балів}$.

Шкала стартової оцінки: $R_C = R_{\text{ЛАБ}} + R_{\text{T}} + R_{\text{МКР}} = 50 \text{ балів} + 40 \text{ бали} + 10 \text{ балів} = 100 \text{ балів}$.

Шкала оцінювання залікової роботи: $R_3 = 40 \text{ балів}$.

Підсумкова рейтингова шкала: $R = R_c = 100$ балів або $R = R_{\text{ЛАБ}} + R_{\text{МКР}} + R_3 = 60$ балів + 40 балів = 100 балів.

Мінімально можлива загальна позитивна оцінка: $R_{\text{МИН}} = 60$ балів.

Нижня межа позитивного оцінювання кожного контрольного заходу (тест) є не менше 60% від балів, визначених для цього контрольного заходу ($10 \text{ балів} \times 0,6 = 6 \text{ балів}$), а негативний результат оцінюється у 0 балів. Якщо студент не проходив або не з'явився на контрольний захід (тест), його результат оцінюється у 0 балів.

Терміни здачі робіт

Термін здачі лабораторних робіт та модульної контрольної роботи визначається наступним чином:

- лабораторна робота 1 – з 3 (третього) до 6 (шостого) тижня навчання включно;
- лабораторна робота 2 – з 3 (третього) до 6 (шостого) тижня навчання включно;
- лабораторна робота 3 – з 5 (п'ятого) до 8 (восьмого) тижня навчання включно;
- лабораторна робота 4 – з 7 (сьомого) до 9 (дев'ятого) тижня навчання включно;
- лабораторна робота 5 – з 7 (сьомого) до 9 (дев'ятого) тижня навчання включно.

Термін здачі тестів визначається:

- часом кінця заняття, на якому проводиться тест, у разі очного навчання;
- 3 (трьома) днями після дня заняття, на якому проводиться тест, у разі дистанційного навчання.

Тести проводяться на наступних тижнях навчання: 2 (другий), 4 (четвертий), 6 (шостий), 8 (восьмий).

В разі виникнення ситуацій, що унеможливають вчасне виконання робіт, терміни здачі робіт можуть бути продовжені.

Заохочувальні та штрафні бали

Лабораторні роботи, які здаються до початку терміну здачі, оцінюються у заохочувальні +2 (плюс два) бали додатково до максимально можливої оцінки.

Лабораторні роботи, які здаються після терміну здачі без поважних причин, оцінюються у штрафні -2 (мінус два) бали додатково до максимально можливої оцінки.

Визначення умов допуску до заліку

Умовою допуску до заліку є виконання з певною успішністю та необхідної кількості завдань впродовж семестру, а також дотримання вимог академічної доброчесності, що описані в [Кодексі честі Національного технічного університету України "Київський політехнічний інститут імені Ігоря Сікорського"](#), затвердженого Вченою радою КПІ ім. Ігоря Сікорського (протокол №4 від 5 квітня 2021 р.). Тобто, студент повинен впродовж семестру набрати деяку початкову ненульову оцінку ($R_c > 0$), виконати модульну контрольну роботу, принаймні 3 (три) лабораторні роботи, принаймні 2 (два) тести, і письмові роботи студента не повинні містити плагіату. Якщо студент не допущений до заліку, то в екзаменаційній відомості виставляється статус "Не допущено" з оцінкою менше 60 балів.

Мінімальна початкова кількість балів для допуску до заліку ($R_{\text{МИНДОП}}$) визначається як 60% від мінімально можливої загальної позитивної оцінки ($R_{\text{МИН}}$). Звідси значення $R_{\text{МИНДОП}} = R_{\text{МИН}} \times 0,6 = 60 \text{ балів} \times 0,6 = 36 \text{ балів}$. Таким чином, умовою допуску до заліку буде $R_c \geq R_{\text{МИНДОП}}$, тобто, початкова кількість балів R_c має бути не менше 36 балів.

Підсумкова рейтингова оцінка переводиться в оцінку за університетською шкалою.

Таблиця переведення підсумкової рейтингової оцінки в університетську шкалу оцінок

Підсумкова рейтингова оцінка <i>R</i>	Університетська шкала оцінок
95...100	Відмінно
85...94	Дуже добре
75...84	Добре
65...74	Задовільно
60...64	Достатньо
36...59	Незадовільно
0...35	Не допущено

Робочу програму навчальної дисципліни (силабус):

Склав асистент кафедри обчислювальної техніки, доктор філософії Шульга М.В.

Ухвалено кафедрою обчислювальної техніки (протокол № 12 від 23 червня 2025 р.)

Погоджено Методичною комісією факультету інформатики та обчислювальної техніки (протокол № 11 від 27 червня 2025 р.)