



МЕТОДОЛОГІЯ DATA SCIENCE

Робоча програма навчальної дисципліни (Силабус)

Реквізити навчальної дисципліни

Рівень вищої освіти	<i>Перший (бакалаврський)</i>
Галузь знань	<i>12 – Інформаційні технології</i>
Спеціальність	<i>122 Комп'ютерні науки</i>
Освітня програма	<i>Комп'ютерні технології в біології та медицині</i>
Статус дисципліни (код)	<i>Вибіркова</i>
Форма навчання	<i>очна(денна)</i>
Рік підготовки, семестр	<i>3 курс, осінній семестр</i>
Обсяг дисципліни	<i>4 кредитів ECTS (120 год). Лекції - 28 год. Практичні заняття -26 год. СР - 66 год.</i>
Семестровий контроль/ контрольні заходи	<i>Залік, МКР</i>
Розклад занять	<i>https://schedule.kpi.ua 2 год. лекційних занять, 1 год. комп'ютерного практикуму на тиждень</i>
Мова викладання	<i>Українська</i>
Інформація про керівника курсу / викладачів	<i>Лектор: к.ф.-м.н., доцент, Пишнограєв Іван Олександрович, pyshnograiev@gmail.com Комп. практикум: к.ф.-м.н., доцент, Пишнограєв Іван Олександрович</i>
Розміщення курсу	<i>Google classroom https://classroom.google.com/c/Nzc1MjQyOTk2ODEx</i>

Програма навчальної дисципліни

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

Незважаючи на постійне збільшення обчислювальної потужності та полегшення доступу до даних за останні пару десятиліть, наша здатність використовувати дані в процесі прийняття рішень далеко не завжди реалізується ефективно. Матеріали дисципліни допоможуть правильно організувати процес розв'язання предметної задачі з точки зору Data Science.

Дисципліна є вибірковою в освітній програмі. Вивчення навчальної дисципліни націлено на формування, розвиток та закріплення у здобувачів таких загальних та фахових компетентностей:

ЗК 2 Здатність застосовувати знання у практичних ситуаціях,

ЗК 7 Здатність до пошуку, оброблення та аналізу інформації з різних джерел,

ФК 4 Здатність використовувати сучасні методи математичного моделювання об'єктів, процесів і явищ, розробляти моделі й алгоритми чисельного розв'язування задач математичного моделювання, враховувати похибки наближеного чисельного розв'язування професійних задач,

ФК 7 Здатність застосовувати теоретичні та практичні основи методології та технології моделювання для дослідження характеристик і поведінки складних об'єктів і систем, проводити обчислювальні експерименти з обробкою й аналізом результатів,

ФК 11 Здатність до інтелектуального аналізу даних на основі методів обчислювального інтелекту включно з великими та погано структурованими даними, їхньої оперативної обробки та візуалізації результатів аналізу в процесі розв'язування прикладних задач,

ФК 21 Здатність використовувати математичні методи для прийняття ефективних рішень під час розв'язання професійних задач в процесі проектування інтелектуальних систем підтримки прийняття рішень (ІСППР).

Внаслідок вивчення курсу студент повинен бути здатний продемонструвати такий програмний результат навчання ОПП:

ПР12 Застосовувати методи та алгоритми обчислювального інтелекту та інтелектуального аналізу даних в задачах класифікації, прогнозування, кластерного аналізу, пошуку асоціативних правил з використанням програмних інструментів підтримки багатовимірного аналізу даних на основі технологій DataMining, TextMining, WebMining,

ПР 24 Розуміти різні типи інтелектуальних систем і технологій; ставити завдання побудови інтелектуальних систем для вирішення завдання вибору варіантів в проблемній області, що погано формалізується.

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

Дисципліна базується на знаннях та навичках з: теорії ймовірностей та математичної статистики, лінійної алгебри, чисельних методів, основ програмування. Дана дисципліна може слугувати базисом для вивчення дисципліни «Методи і системи штучного інтелекту» та дипломного проєктування.

3. Зміст навчальної дисципліни:

Розділ 1. Математичний та програмний інструментарій

Тема 1.1. Вступ та основи синтаксису мови R.

Вступ в дисципліну. Основні характеристики мови програмування R. Розгляд базових операцій та функцій. Пакети. Описовий аналіз даних. Основні алгоритмічні елементи мови. Типи даних.

Тема 1.2. Базовий математичний апарат Data Science.

Елементи теорії ймовірності та математичної статистики. Закони розподілу. Перевірка гіпотез. Регресія.

Розділ 2. Основи Data Science

Тема 2.1. Аналіз вибірки однієї та двох змінних.

Описовий аналіз. Перетворення даних. Візуалізація даних. Аналіз факторних змінних. Викиди даних.

Тема 2.2. Багатовимірний статистичний аналіз.

Описовий аналіз. Факторний аналіз. Теплові карти. Кластерний аналіз.

Тема 2.3. Регресія та прогноз.

Лінійна та нелінійна регресія. Проблеми регресійних моделей. Побудова прогнозів. Аналіз часових рядів. Порівняння моделей.

Тема 2.4. Обробка наборів даних.

Постановка задачі. Проблеми пошуку та адекватності джерела. Обробка помилок та пропусків даних. Перевірки гіпотез. Нормалізація даних.

Тема 2.5. Управління наборами даних

Принципи FAIR та їх реалізація. Принципи відкритої науки та відкритих даних. Створення метаданих. Інструменти для організації роботи з даними.

Розділ 3. Просунутий інструментарій Data Science

Тема 3.1. Моделі машинного навчання з вчителем.

Дерева та ліси. Елементарні нейронні мережі. Адекватність, вибір моделей. Розв'язання задачі класифікації.

Тема 3.2. Основи просторового аналізу даних.

Постановка задачі. Особливості аналізу та візуалізації просторових даних. Побудова карт. Здійснення базового просторового аналізу.

Тема 3.3. Основи текстової аналітики.

Постановка задачі. Особливості аналізу та візуалізації текстових даних. Обробка тексту. Побудова хмар термінів. Виділення основної тематики тексту.

Тема 3.4. Побудова інтерактивних дашбордів.

Використання засобів Shiny для візуалізації та аналізу даних. Виконання індивідуального завдання.

4. Навчальні матеріали та ресурси

Базова:

1. R for Data Science by Hadley Wickham, Garrett Grolemund [en]. URL: <https://r4ds.had.co.nz/>.

2. *Advanced R by Hadley Wickham [en]. URL: <https://adv-r.hadley.nz/index.html>*
3. *Майборода Р.Є Комп'ютерна статистика. Професійний старт. Навчальний посібник. Київський університет», 2020. – 482 с. <http://probability.univ.kiev.ua/userfiles/mre/compsta1.pdf>*
4. *Методи і моделі інтелектуального аналізу даних. Практикум [Електронний ресурс] : навчальний посібник для студентів, які навчаються за спеціальністю 122 «Комп'ютерні науки», освітньої програми «Системи і методи штучного інтелекту» / Н. І. Недашківська ; КПІ ім. Ігоря Сікорського. – Київ : КПІ ім. Ігоря Сікорського, 2019. – 71 с. <https://ela.kpi.ua/handle/123456789/53764>*
5. *Турчин В.М. Теорія ймовірностей і математична статистика,- Дніпропетровськ, ІМА-пресс, 2014. - 566 с. (НТБ ім. Г.І. Денисенка)*
6. *Ланде Д.В., Субач І.Ю., Бояринова Ю.Є. Основи теорії і практики інтелектуального аналізу даних у сфері кібербезпеки: навчальний посібник. – К.: ІСЗІ КПІ ім. Ігоря Сікорського», 2018. — 297 с. (НТБ ім. Г.І. Денисенка)*

Допоміжна:

7. *An introduction to machine learning with applications in R. Michael Clark. https://web.ipac.caltech.edu/staff/fmasci/home/astro_refs/ML_inR.pdf*
8. *Machine Learning Using R. Karthik Ramasubramanian, Abhishek Singh. <https://link.springer.com/content/pdf/10.1007/978-1-4842-2334-5.pdf>*
9. *The R Manuals. <https://cran.r-project.org>*

Навчальний контент

5. Методика опанування навчальної дисципліни (освітнього компонента)

Навчальна дисципліна охоплює 36 годин лекцій та 18 годин комп'ютерних практикумів, а також виконання модульної контрольної роботи, яка складається з тестових питань за всіма розділами.

Лекційний матеріал викладається за переліком тем п. 3, допоміжні матеріали для самостійної та/або асинхронної роботи розміщуються щотижнево в Google classroom.

Комп'ютерні практикуми з дисципліни проводяться з метою закріплення теоретичних положень лекційного матеріалу. Здобувачі виконують індивідуальні завдання, розміщені в Google classroom та захищають результати виконання перед викладачем.

Перелік лабораторних робіт:

1. *Основи мови R. Попередній аналіз соціально-економічних показників. Виділення груп об'єктів дослідження (6 годин);*
2. *Дослідження взаємозв'язків параметрів. Побудова моделей (4 години);*
3. *Побудова композитних індикаторів (Частина 1) (4 години);*
4. *Побудова композитних індикаторів (Частина 2) (4 години).*

5. Самостійна робота студента/аспіранта

Самостійна робота здобувача охоплює такі складники як підготовка до поточних опитувань, підготовки і виконання лабораторних робіт, підготовки до модульної контрольної роботи та екзамену.

№	Назва виду СРС	Кількість годин СРС
1	<i>Опрацювання лекційного матеріалу</i>	18
2	<i>Підготовка та виконання комп'ютерних практикумів</i>	36
3	<i>Підготовка до написання модульної контрольної роботи</i>	12
	<i>Разом</i>	66

Політика та контроль

6. Політика навчальної дисципліни (освітнього компонента)

Відвідування занять. Відсутність на аудиторному занятті не передбачає нарахування штрафних балів, оскільки фінальний рейтинговий бал здобувача формується виключно на основі оцінювання результатів навчання.

Пропущені контрольні заходи оцінювання. Кожен здобувач має право відпрацювати пропущені з поважної причини (лікарняний, мобільність тощо) заняття за рахунок самостійної роботи. Детальніше за посиланням: <https://kpi.ua/files/n3277.pdf>.

Процедура оскарження результатів контрольних заходів оцінювання. Здобувач може підняти будь-яке питання, яке стосується процедури контрольних заходів та очікувати, що воно буде розглянуто згідно із наперед визначеними процедурами. Здобувачі мають право аргументовано оскаржити результати контрольних заходів, пояснивши з яким критерієм не погоджуються відповідно до оціночного. Календарний контроль проводиться з метою підвищення якості навчання здобувачів та моніторингу виконання здобувачем вимог силабусу двічі протягом семестру (7 та 13-й тижні навчання).

Академічна доброчесність. Політика та принципи академічної доброчесності визначені у розділі 3 Кодексу честі Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського». Детальніше: <https://kpi.ua/code>.

Норми етичної поведінки. Норми етичної поведінки здобувачів і працівників визначені у розділі 2 Кодексу честі Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського». Детальніше: <https://kpi.ua/code>.

Інклюзивне навчання. Засвоєння знань та умінь в ході вивчення дисципліни може бути доступним для більшості осіб з особливими освітніми потребами, окрім здобувачів з серйозними вадами зору, які не дозволяють виконувати завдання за допомогою персональних комп'ютерів, ноутбуків та/або інших технічних засобів.

Навчання іноземною мовою. У ході виконання завдань здобувачем може бути рекомендовано звернутися до англomовних джерел.

Підготовка до комп'ютерних практикумів та контрольних заходів здійснюється під час самостійної роботи здобувачів з можливістю консультування з викладачем у визначений час консультацій або за допомогою електронного листування (електронна пошта, месенджери).

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

Семестровий контроль: залік.

Семестровий рейтинг студента з дисципліни виставляється лектором та складається з 100 балів, що він отримує за:

- ~ виконання модульної контрольної роботи;
- ~ виконання комп'ютерних практикумів (п'ять комп'ютерних практикумів).

Критерії нарахування балів за семестр:

- 1) Модульна контрольна робота оцінюється у 20 балів.
- 2) Кожний з комп'ютерних практикумів робіт оцінюється в 20 балів.

За кожний тиждень запізнення з поданням лабораторної роботи на перевірку нараховується штрафний -1 бал (до 4 балів включно).

Критерії нарахування балів за контрольні заходи:

- "відмінно": 95 -100% - здобувач виявив всебічні, систематичні та глибокі знання навчального матеріалу з дисципліни; продемонстрував уміння вільно виконувати всі завдання, передбачені програмою; засвоїв основну та додаткову літературу; проявив творчі здібності в розумінні, в логічному, чіткому, стислому та ясному трактуванні навчального матеріалу; засвоїв взаємозв'язок основних понять дисципліни, їх значення для подальшої професійної діяльності
- "дуже добре": 85-94% - здобувач виявив систематичні знання навчального матеріалу з дисципліни вище середнього рівня; продемонстрував уміння добре виконувати всі завдання, передбачені програмою, допустивши незначні помилки; засвоїв основну та додаткову літературу; засвоїв взаємозв'язок основних понять дисципліни, їх значення для подальшої професійної діяльності
- "добре": 75-84% - здобувач виявив загалом добрі знання навчального матеріалу при виконанні передбачених програмою завдань, але припустив ряд помітних помилок; засвоїв основну літературу; показав систематичний характер знань з дисципліни; здатний до їх самостійного використання та поповнення в процесі подальшої навчальної роботи і професійної діяльності
- "задовільно": 65-74% - здобувач виявив знання основного навчального матеріалу з дисципліни в обсязі, необхідному для подальшого навчання та майбутньої професійної діяльності; ознайомився з основною літературою; впорався з виконанням завдань, передбачених програмою, але припустив значну кількість помилок або недоліків на запитання при співбесіді, тестуванні та при виконанні завдань тощо, принципів з яких може усунути самостійно

- “достатньо”: 60-64% - здобувач виявив знання основного навчального матеріалу з дисципліни в мінімальному обсязі, необхідному для подальшого навчання та майбутньої професійної діяльності; ознайомився з основною літературою; в основному виконав завдання, передбачені програмою, але припустив помилки у відповіді на запитання при співбесідах, тестуванні та при виконанні завдань тощо, які він може усунути лише під керівництвом та за допомогою викладача
- “незадовільно”: 30-54% - здобувач мав значні прогалини в знаннях основного навчального матеріалу; допускав принципові помилки при виконанні передбачених програмою завдань, але спроможний самостійно доопрацювати програмний матеріал і підготуватися для перездачі дисципліни
- “незадовільно”: 0-29% - здобувач не мав знань зі значної частини навчального матеріалу з дисципліни; допускав принципові помилки при виконанні більшості передбачених програмою завдань або не виконував ці завдання

Умовою першої атестації в календарному контролі є поточний рейтинг не менше 30% запланованих балів за семестр (зданій 1 комп'ютерний практикум). Умова другої атестації ~ поточний рейтинг не менше 50% запланованих балів (здані 2 комп'ютерні практикуми).

Необхідною умовою допуску до заліку є зарахування всіх комп'ютерних практикумів та 50 балів семестрового рейтингу. Студенти, які мають менше 50 балів або не здали комп'ютерні практикуми, не допускаються до здачі заліку. Залік виставляється за результатом семестрового рейтингу.

Якщо студент не набрав 60 балів або хоче покращити свою оцінку, можливо написати залікову роботу. На залікову роботу виноситься 50 балів, які додаються до семестрового рейтингу, помноженого на коефіцієнт 0,5. Залік проводиться у вигляді письмової роботи, в якій три теоретичних питання та два практичних. Кожне завдання оцінюється в 10 балів за такими критеріями:

- «відмінно», повна відповідь, не менше 90% потрібної інформації, що виконана згідно з вимогами до рівня «умінь», (повне, безпомилкове розв'язування завдання) – 9-10 балів;
- «добре», достатньо повна відповідь, не менше 75% потрібної інформації, що виконана згідно з вимогами до рівня «умінь або є незначні неточності (повне розв'язування завдання з незначними неточностями) – 8 балів;
- «задовільно», неповна відповідь, не менше 60% потрібної інформації, що виконана згідно з вимогами до «стереотипного» рівня та деякі помилки (завдання виконане з певними недоліками) – 7 балів;
- «незадовільно», відповідь не відповідає умовам до «задовільно» – 0 балів.

Сума рейтингових балів, отриманих студентом протягом семестру, переводиться до підсумкової оцінки згідно з таблицею.

Таблиця відповідності рейтингових балів оцінкам за університетською шкалою:

Бали:	Оцінка
100...95	Відмінно
94...85	Дуже добре
84...75	Добре
74...65	Задовільно
64...60	Достатньо
Менше 60	Незадовільно
Невиконання умов допуску до семестрового контролю	Не допущено

Робочу програму навчальної дисципліни (силабус):

Складено доцент, к.ф.-м.н., доцент *Пишнограєв Іван Олександрович*



Ухвалено кафедрою ШІ (протокол № 14 від 24.06.2025)

Погоджено Методичною комісією НН ІПСА (протокол № 7 від 25.06.2025)