



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>Системи і методи штучного інтелекту</i>
Статус дисципліни (код)	<i>Вибіркова</i>
Форма навчання	<i>очна(денна)/дистанційна/змішана</i>
Рік підготовки, семестр	<i>3 курс, осінній семестр</i>
Обсяг дисципліни	<i>4 кредити ЕКТС</i>
Семестровий контроль/ контрольні заходи	<i>Залік</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/NjgyMjcONTA5ODY3</i>

Програма навчальної дисципліни

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

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

В процесі вивчення дисципліни здобувачі вищої освіти матимуть змогу отримати: знання:

- етапів розв'язання задач з аналізу даних;
- методів Data Science, що використовується на кожному етапі розв'язання задачі;
- основних термінів та понять науки про дані.

уміння:

- аналізувати поставлені задачі та розбивати їх на етапи;
- застосовувати методи аналізу даних для вирішення задач на кожному етапі проєкту.

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.

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.

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.

3. 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-прес, 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

4. Методика опанування навчальної дисципліни (освітнього компонента)

У гугл-класі будуть щотижневі завдання з детальними інструкціями та необхідним матеріалом, які необхідно вчасно виконувати.

5. Самостійна робота студента/аспіранта

Індивідуальні завдання складаються з підготовки до комп'ютерних практикумів та опрацюванні лекційного матеріалу.

Політика та контроль

6. Політика навчальної дисципліни (освітнього компонента)

Усі роботи студенти мають прикріплювати в особистому кабінеті гугл-класу. Дедлайни кожного завдання позначені в щотижневих завданнях у гугл-класі. Роботи мають бути виконані з дотриманням академічної доброчесності. Політика та принципи академічної доброчесності, етична поведінка студентів визначені у Кодексі честі <https://kpi.ua/code>. Лектор може запропонувати студентам пройти запропоновані ним онлайн-курси на платформі Coursera. Також сертифікати цих курсів можуть бути частково зараховані згідно до [Положення](#).

Тематика лабораторних робіт комп'ютерного практикуму спрямована на поглиблення засвоєного матеріалу лекцій. На заняттях комп'ютерного практикуму розв'язуються задачі та вправи по темам лекції.

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 16 балів.

За кожний тиждень запізнення з поданням лабораторної роботи на перевірку нараховується штрафний -1 бал.

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 допрацювати 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 (здані 2 комп'ютерні практикуми). Умова другої атестації ~ поточний рейтинг не менше 50% запланованих балів (здані 4 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 виносяться 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:

<i>Бали:</i>	<i>Оцінка</i>
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 штучного інтелекту (протокол № 14 від 11.06.2024)

Погоджено Методичною комісією ННІПСА (протокол № 10 від 24.06.2024)