

**Інформаційний пакет освітніх компонент навчального плану
освітньо-професійної програми «Системний аналіз»**

(назва)

Освітнього рівня бакалавр

Спеціальності 124 Системний аналіз

Галузь знань 12 Інформаційні технології

1. Назва освітньої компоненти Основи Data Science
(назва дисципліни)

2. Тип основна

3. Обсяг:	Кредитів ECTS	Годин	За видами занять:				
			Лекцій	Семінар	Практичних занять	Лабораторних занять	Самостійна підготовка
	5	150	18	-	18	18	96

4. Взаємозв'язок у структурно-логічній схемі

Освітні компоненти, які передують вивченню	<ol style="list-style-type: none"> 1. Логіка та алгоритми обробки даних. 2. Системний аналіз та проектування інформаційних систем 3. Організація баз даних та знань. 4. Системи підтримки прийняття рішень. 5. Технології Business Intelligence
--	--

Освітні компоненти для яких є базовою	Кваліфікаційна робота
---------------------------------------	-----------------------

5. Компетенції відповідно до ОПП та вимог роботодавців:

Компетенції відповідно до ООП

ПК4. Здатність виділяти основні чинники, які впливають на розвиток фізичних, економічних, соціальних процесів, відокремлювати в них стохастичні та невизначені фактори, формулювати ці фактори у вигляді випадкових або нечітких величин, векторів, процесів та досліджувати залежності між ними.

ПК6. Здатність до комп'ютерної реалізації математичних моделей реальних систем і процесів; проектувати, застосовувати і супроводжувати програмні засоби моделювання, прийняття рішень, оптимізації обробки інформації, інтелектуального аналізу даних.

ПК10. Здатність розробляти експериментальні та спостережувальні дослідження і аналізувати дані, отримані в них

Можливості Power BI

Компетенції відповідно до вимог роботодавців

1. Досліджувати предметну область в бізнес завданнях і пропонувати рішення з використанням концепцій Data Science і Machine Learning.
2. Застосовувати ключові концепції збору, підготовки, дослідження та візуалізації даних.
3. Використовувати сучасні методи машинного навчання для вирішення завдань прогнозування і класифікації.
4. Сегментувати великі дані, створювати рекомендаційні моделі для полегшення прийняття стратегічних рішень.

6. Результати навчання відповідно до ОПП

ПРН 16. Застосовувати методи і засоби роботи з даними і знаннями, методи математичного, логіко-семантичного, об'єктного та імітаційного моделювання, технології системного і статистичного аналізу.

ПРН 21. Знати базові визначення й поняття, основні етапи аналізу та інтерпретації даних, загальну структуру і функціонування інтегрованих інформаційно-аналітичних систем; основи організації вимірювань й технології обробки експериментальних і статистичних даних на ЕОМ; змістовну й математичну постановку основних завдань аналізу даних, способи їхнього розв'язання; сутність і властивості основних алгоритмів інтелектуального аналізу; приклади рішення практичних завдань аналізу даних.

ПРН37. Вміти застосовувати на практиці ключові концепції збору, підготовки, дослідження та візуалізації даних; сегментувати великі дані, створювати рекомендаційні моделі для полегшення прийняття стратегічних рішень; використовувати сучасні методи машинного навчання для вирішення завдань прогнозування і класифікації.

7. План вивчення освітньої компоненти

Змістовний розділ	Вид заняття	Тема	Знати	Вміти	План заняття	Лекція, методична розробка
Розділ 1						
	Лекція 1	Методологія Data Science рішень.	<ol style="list-style-type: none"> 1. Означення Data Science 2. Моделі в Data Science 	Аналізувати предметну область, обирати засоби та інструменти для data science	http://dl.dut.edu.ua/mod/resource/view.php?id=89634	http://dl.dut.edu.ua/mod/resource/view.php?id=89634
	Лекція 2	Попередня обробка даних. Конструювання ознак. Підходи і методики для візуалізації даних.	<ol style="list-style-type: none"> 1. Розуміти поняття даних 2. Особливості обробки даних 	Використовувати методи візуалізації даних та конструювання даних. Вміти працювати з атрибутами.	http://dl.dut.edu.ua/mod/resource/view.php?id=89635	http://dl.dut.edu.ua/mod/resource/view.php?id=89635
	Лекція 3	Логіка найближчого сусіда. Лінійна регресія.	<ol style="list-style-type: none"> 1. Розуміти просту лінійну регресію. 2. Задачу багатомірної 	Використовувати методи регресії для роботи із багатовимірними даними	http://dl.dut.edu.ua/mod/resource/view.php?id=89636	http://dl.dut.edu.ua/mod/resource/view.php?id=89636

Можливості Power BI

		Функція втрат.	регресії.			
Лекція 4	Поділ даних.		<ol style="list-style-type: none"> 1. Розуміти певну приховану закономірність 2. Тестувальна вибірка відрізняється від тренувальної вибірки 	Використовувати тренувальну вибірку для вирішення задач поділу даних.	http://dl.dut.edu.ua/mod/resource/view.php?id=89637	http://dl.dut.edu.ua/mod/resource/view.php?id=89637
Лекція 5	Мультикласовий класифікатор. Ранжування. Класифікація і дерево регресії.		<ol style="list-style-type: none"> 1. Задача класифікації. 2. Розуміти побудову дерева регресії. 	Методи, застосовувані для вирішення задач класифікації.	http://dl.dut.edu.ua/mod/resource/view.php?id=89638	http://dl.dut.edu.ua/mod/resource/view.php?id=89638
Лекція 6	Бустінг: Adaboost, градієнтний бустінг (GBM) ансамблі моделей.		<ol style="list-style-type: none"> 1. Постановка ML завдання 2. Розуміти реалізації GBM 	Вирішувати завдання відновлення функції в загальному контексті навчання з учителем.	http://dl.dut.edu.ua/mod/resource/view.php?id=89639	http://dl.dut.edu.ua/mod/resource/view.php?id=89639
Лекція 7	Байєсова навчання.		<ol style="list-style-type: none"> 1. Розуміти метод статистичного висновування 2. Правило Баєса 	Будувати Баєсові алгоритми класифікації.	http://dl.dut.edu.ua/mod/resource/view.php?id=89640	http://dl.dut.edu.ua/mod/resource/view.php?id=89640
Лекція 8	Data Science на основі нейронних мереж. Навчання нейронних мереж.		<ol style="list-style-type: none"> 1. Алгоритм побудови нейронних мереж. 2. Використання нейронних мереж в Data Science 	Штучні нейронні мережі в машинному навчанні. Метод зворотного розподілу помилки.	http://dl.dut.edu.ua/mod/resource/view.php?id=89641	http://dl.dut.edu.ua/mod/resource/view.php?id=89641
Лекція 9	Кластеризація: метод K-середніх виявлення аномалій.		<ol style="list-style-type: none"> 1. Відбір еталонних об'єктів. 2. Методи самоорганізації 	Використовувати метод найближчого сусіда та його узагальнення.	http://dl.dut.edu.ua/mod/resource/view.php?id=89642	http://dl.dut.edu.ua/mod/resource/view.php?id=89642
Лабораторне заняття 1-2	Підготовка даних для проекту		<ol style="list-style-type: none"> 1. Аналітичні технології 2. Перетворення стандартними техніками 	Вміти визначати вплив вибору набору ознак на результати	http://dl.dut.edu.ua/mod/resource/view.php?id=89644	http://dl.dut.edu.ua/mod/resource/view.php?id=89644

Можливості Power BI

	Лабораторне заняття 3-4	Моделювання логіки "найближчого сусіда" на прикладі аналітики по WHISKY	<ol style="list-style-type: none"> 1. Використовувати метод найближчого сусіда та його узагальнення. 2. Узагальнений метричний класифікатор 	Будувати метод найближчих сусідів та його варіації.	http://dl.dut.edu.ua/mod/resource/view.php?id=89649	http://dl.dut.edu.ua/mod/resource/view.php?id=89649
	Лабораторне заняття 5-7	Використання класифікатора Байєса для задач класифікації.	<ol style="list-style-type: none"> 1. Структура існуючих систем класифікації 2. Систематизація методів відбору ознак 	Будувати алгоритми класифікації і розробляти методи класифікації	http://dl.dut.edu.ua/mod/resource/view.php?id=89650	http://dl.dut.edu.ua/mod/resource/view.php?id=89650
	Лабораторне заняття 8-9	Створення моделей кластеризації і рекомендаційних моделей	<ol style="list-style-type: none"> 1. Постановка задачі кластеризації. 2. Використовувати рекомендаційні моделі 	Розробляти представлення кластерної стратегії	http://dl.dut.edu.ua/mod/resource/view.php?id=89646	http://dl.dut.edu.ua/mod/resource/view.php?id=89646
	Практичне заняття 1-2	Візуалізація даних за допомогою Power BI і R	<ol style="list-style-type: none"> 1. Моделювати звіти за допомогою R 2. Можливості Power BI 	Створювати наочні інтерактивні звіти за допомогою засобів візуальної аналітики	http://dl.dut.edu.ua/mod/resource/view.php?id=89643	http://dl.dut.edu.ua/mod/resource/view.php?id=89643
	Практичне заняття 3-4	Процес наукових досліджень даних на прикладі метеорологічних даних.	<ol style="list-style-type: none"> 1. Структура системи, закономірності взаємодії елементів у середині системи і поза нею 2. Процес аналізу даних 	Дослідження, динаміки та аналіз метеорологічних даних.	http://dl.dut.edu.ua/mod/resource/view.php?id=89648	http://dl.dut.edu.ua/mod/resource/view.php?id=89648
	Практичне заняття 5-7	Пошук нетривіальних практично корисних закономірностей у даних за допомогою Data Mining-процесу.	<ol style="list-style-type: none"> 1. Методи дослідження даних. 2. Сутність і мету технології Data Mining 	Виявлення кореляції, тенденцій, шаблонів, зв'язків і категорій, пошуку прихованих закономірностей.	http://dl.dut.edu.ua/mod/resource/view.php?id=89651	http://dl.dut.edu.ua/mod/resource/view.php?id=89651
	Практичне заняття 8-9	Процес створення	<ol style="list-style-type: none"> 1. Основні класифікації алгоритмів кластеризації 	Моделювати квібірку об'єктів, представлену у	http://dl.dut.edu.ua/mod/resource/view.php?id=89645	http://dl.dut.edu.ua/mod/resource/view.php?id=89645

Можливості Power BI

	програмних моделей для кластеризації, рекомендаційних алгоритмів.	2. Структура рекомендаційних алгоритмів.	вигляді графа, будувати графові алгоритмів кластеризації		
--	---	--	--	--	--

8. Мова вивчення освітньої компоненти

українська

9. Інформаційне забезпечення освітньої компоненти

1. Cristianini N., Shawe-Taylor J. An introduction to support vector machines (and other kernelbased learning method). - Cambridge UP, 2000. - 187 p
2. Cesa-Bianchi N., Lugosi G. Prediction, Learning and Games. - New York: Cambridge University Press, 2006. - 394 p.
3. Clarence Chio, David Freeman. Machine Learning and Security. – Sebastopol: O’Reilly Media, Inc., 2018. – 367 p
4. Lean Analytics: Use Data to Build a Better Startup Faster (Lean Series) Hardcover – March 26, 2013
5. Business value in the ocean of data —Ласло Чер, Тамаш Фехер Alinea Kiadó, 2013
6. Naked Statistics: Stripping the Dread from the Data, Paperback, 282 pages Published January 13th 2014 by W. W. Norton Company, 2012
7. Data Science Кеті Онил, Рейчл Шатт, Орелли 2019
8. Data Science at the Command Line: Facing the Future with Time-Tested Tools, Жерон Янсенс, Орели 2014

10. Методи оцінювання, підсумкові звітності за освітньою компонентою

залік

11. Матеріально-технічне забезпечення освітньої компоненти

Програмне середовище для мови програмування R, Rstudio. IDE для мови програмування Python. Пробна користувальницька версія Microsoft Power BI