

Чернівецький національний університет імені Юрія Федьковича

(повне найменування закладу вищої освіти)

економічного факультету

(назва інституту/факультету)

Кафедра економіко-математичного моделювання

(назва кафедри)

СИЛАБУС

навчальної дисципліни

Обробка та аналіз великих даних

(вказати назву навчальної дисципліни (іноземною, якщо дисципліна викладається іноземною мовою))

вибіркова

(вказати: обов'язкова)

Освітньо-професійна програма Економічна кібернетика

(назва програми)

Спеціальність 051 Економіка

(вказати: код, назва)

Галузь знань 05 «Соціальні та поведінкові науки»

(вказати: шифр, назва)

Рівень вищої освіти перший (бакалаврський)

(вказати: перший (бакалаврський)/другий (магістерський)/третій (освітньо-науковий))

економічний факультет

(назва факультету/інституту, на якому здійснюється підготовка фахівців за вказаною освітньо-професійною програмою)

Мова навчання українська

(вказати: на яких мовах читається дисципліна)

Розробники: Вінничук О.Ю., доцент кафедри економіко-математичного моделювання, к.е.н.

(вказати авторів (викладач (ів)), їхні посади, наукові ступені, вчені звання)

Профайл викладача (-ів) <https://emm.cv.ua/teachers/vinnichuk-olena-yuriyivna/>

Контактний тел. +380505667274

E-mail: o.vinnichuk@chnu.edu.ua

Сторінка курсу в Moodle <https://classroom.google.com/c/MTUzMjc5NzkzMTA2>

Консультації
Очні консультації відповідно до графіку консультацій
он-лайн консультації: вівторок, 11:00 – 11:30
(за попередньою домовленістю)

1. Анотація дисципліни (призначення навчальної дисципліни).

Протягом останніх років всі розуміють, що майбутнє за рішеннями, які використовують сучасні підходи та інструменти роботи з даними (аналітика, machine learning, Big Data, штучний інтелект та ін.). Дисципліна «Обробка та аналіз великих даних» включає різні методи обробки структурованих і неструктурованих даних великого обсягу для використання їх із метою вирішення різних завдань.

Дисципліна «Обробка та аналіз великих даних» є вибірковою компонентою у підготовці студентів освітньо-професійної програми «Економічна кібернетика» та спрямована на формування у них системного та наукового уявлення про методи обробки та аналізу великих даних, як їх трактувати й узагальнювати та якими принципами керуватися, щоб використати зібрану інформацію для розвитку бізнесу.

2. Мета навчальної дисципліни:

Формування теоретичних знань і практичних навиків за методологією обробки та аналізу великих даних.

3. Пререквізити.

Теоретико-методологічною базою для вивчення дисципліни є статистика, теорія ймовірності та математична статистика, вища математика, алгоритмізація та програмування.

4. Результати навчання

Засвоєння теоретичних положень та опанування практичних навичок щодо обробки та аналізу великих даних.

Відповідно до освітньо-професійної програми підготовки бакалаврів галузі знань 05 Соціальні та поведінкові науки за спеціальністю 051 Економіка (освітня програма: «Економічна кібернетика») вивчення дисципліни «Економетрика» сприяє формуванню компетентностей та програмних результатів навчання:

Загальні та фахові компетентності:

ЗК3. Здатність до абстрактного мислення, аналізу та синтезу.

ЗК4. Здатність застосовувати знання у практичних ситуаціях.

СК1. Здатність виявляти знання та розуміння проблем предметної області, основ функціонування сучасної економіки на мікро-, мезо-, макро- та міжнародному рівнях.

СК4. Здатність пояснювати економічні та соціальні процеси і явища на основі теоретичних моделей, аналізувати і змістовно інтерпретувати отримані результати.

СК6. Здатність застосовувати економіко-математичні методи та моделі для вирішення економічних задач.

СК9. Здатність прогнозувати на основі стандартних теоретичних та економетричних моделей соціально економічні процеси.

СК15. Знання та уміння використовувати сучасне програмне забезпечення для обробки економічних даних та аналізу економіко-математичних моделей.

Програмні результати навчання:

ПРН 5. Застосовувати аналітичний та методичний інструментарій для обґрунтування пропозицій та прийняття управлінських рішень різними економічними агентами (індивідуумами, домогосподарствами, підприємствами та органами державної влади).

ПРН 8. Застосовувати відповідні економіко-математичні методи та моделі для вирішення економічних задач.

ПРН 10. Проводити аналіз функціонування та розвитку суб'єктів господарювання, визначати функціональні сфери, розраховувати відповідні показники які характеризують результативність їх діяльності.

ПРН 21. Вміти абстрактно мислити, застосовувати аналіз та синтез для виявлення ключових характеристик економічних систем різного рівня, а також особливостей поведінки їх суб'єктів.

У підсумку вивчення навчальної дисципліни дає можливість студентам:

знати:

- застосовувати знання обробки та аналізу великих даних у практичних ситуаціях;

- виявляти знання та розуміння проблем предметної області, основ функціонування сучасної економіки на мікро-, мезо-, макро- та міжнародному рівнях на основі аналізу великих даних;

- пояснювати економічні та соціальні процеси і явища на основі теоретичних моделей, аналізувати і змістовно інтерпретувати отримані результати;

- застосовувати економіко-математичні методи та моделі для вирішення економічних задач;

- прогнозувати на основі стандартних теоретичних та економетричних моделей соціально економічні процеси;

- використовувати сучасне програмне забезпечення для обробки економічних даних та аналізу економіко-математичних моделей;

уміти:

- застосовувати аналітичний та методичний інструментарій для обґрунтування пропозицій та прийняття управлінських рішень різними економічними агентами (індивідуумами, домогосподарствами, підприємствами та органами державної влади) на основі використання інструментарію обробки та аналізу даних;

- застосовувати відповідні економіко-математичні методи та моделі для вирішення економічних задач;

- проводити аналіз функціонування та розвитку суб'єктів господарювання, визначати функціональні сфери, розраховувати відповідні показники які характеризують результативність їх діяльності;

- абстрактно мислити, застосовувати аналіз та синтез для виявлення ключових характеристик економічних систем різного рівня, а також особливостей поведінки їх суб'єктів, використовуючи методи обробки та аналізу даних.

5. Опис навчальної дисципліни

5.1. Загальна інформація

Назва навчальної дисципліни <u>Обробка та аналіз великих даних</u>												
Форма навчання	Рік підготовки	Семестр	Кількість			Кількість годин						Вид підсумкового контролю
			кредитів	годин	змістових модулів	лекції	практичні	семінарські	лабораторні	самостійна робота	індивідуальні завдання	
Денна	3-й	5	3	90	2	15	-	-	30	45	-	іспит
Заочна	-	-	-	-	-	-	-	-	-	-	-	-

5.2. Дидактична карта навчальної дисципліни

Назви змістових модулів і тем	Кількість годин												
	денна форма						заочна форма						
	усього	у тому числі					усього	у тому числі					
		л	п	лаб	інд	с.р.		л	п	лаб	інд	с.р.	
1	2	3	4	5	6	7	8	9	10	11	12	13	
Теми лекційних занять	Змістовий модуль 1. Основні поняття й завдання аналізу даних												
Тема 1. Вступ: дата-аналітичне мислення	16	4	-	6	-	6	-	-	-	-	-	-	-
Тема 2. Технології Big data: ключові характеристики, особливості та переваги	20	3	-	6	-	11	-	-	-	-	-	-	-

Разом за ЗМ1	36	7	-	12	-	17	-	-	-	-	-	-
Теми лекційних занять	Змістовий модуль 2. Обробка та аналіз великих даних											
Тема 3. Очищення та попередня обробка даних. Описова статистика	18	4	-	6	-	8	-	-	-	-	-	-
Тема 4. Візуалізація даних в Python	18	2	-	6	-	10						
Тема 5. Вступ у прогностичне моделювання: від кореляції до контрольованої сегментації	18	2	-	6	-	10						
Разом за ЗМ 2	54	8	-	18	-	28	-	-	-	-	-	-
Усього годин	90	15	-	30	-	45	-	-	-	-	-	-

5.3. Зміст завдань для самостійної роботи

№	Назва теми
1	Тема 1. Вступ: дата-аналітичне мислення
2	Тема 2. Технології Big data: ключові характеристики, особливості та переваги
3	Тема 3. Очищення та попередня обробка даних. Описова статистика
4	Тема 4. Візуалізація даних в Python
5	Тема 5. Вступ у прогностичне моделювання: від кореляції до контрольованої сегментації

6. Система контролю та оцінювання

Види та форми контролю

Формами поточного контролю є усна чи письмова (тестування, контрольні роботи, презентації) відповідь студента.

Формами підсумкового контролю є **екзамен**.

Засоби оцінювання

Засобами оцінювання та демонстрування результатів навчання є:

- тестові завдання;
- виконані завдання в лабораторних роботах;
- проекти;
- презентації результатів виконаних завдань та досліджень;
- студентські презентації та виступи на наукових заходах;
- інші види індивідуальних та групових завдань.

Критерії оцінювання результатів навчання з навчальної дисципліни

Політика щодо дедлайнів та перескладання: Роботи, які здаються із порушенням термінів без поважних причин, оцінюються на нижчу оцінку (-10 балів). Перескладання модулів відбувається із дозволу деканату за наявності поважних причин (наприклад, лікарняний).

Політика щодо академічної доброчесності: Усі письмові роботи перевіряються на наявність плагіату і допускаються до захисту із коректними текстовими запозиченнями не більше 20%. Списування під час контрольних робіт та екзаменів заборонені (в т.ч. із використанням мобільних девайсів).

Політика щодо відвідування: Відвідування занять є обов'язковим компонентом оцінювання, за яке нараховуються бали. За об'єктивних причин (наприклад, хвороба, міжнародне стажування) навчання може відбуватись в он-лайн формі за погодженням із керівником курсу.

Критеріями оцінювання є:

при усних відповідях: повнота розкриття питання; логіка викладання матеріалу; використання основної та додаткової літератури; аналітичні міркування, уміння робити порівняння, висновки;

при виконанні письмових завдань: повнота розкриття питання, аргументованість і логіка викладання матеріалу, використання літературних джерел, прикладів та фактичного матеріалу тощо; цілісність, системність, логічність, уміння формулювати висновки; акуратність оформлення письмової роботи.

Максимальна оцінка знань студента під час навчальних занять за кожну тему (опитування, тестування, розв'язання задач), виконанні завдань для самостійної роботи, підготовці есе – до 5 балів.

Проведення підсумкового контролю здійснюється у формі екзамену в обсязі навчального матеріалу, визначеного навчальною програмою дисципліни і в терміни, передбачені графіком навчального процесу. Загальна підсумкова оцінка з дисципліни (максимум 100 балів) визначається як сума балів поточного і модульного контролю та результатів екзамену (як можливість отримання додаткових балів, якщо набрані протягом семестру бали не влаштовують студентів). У випадку отримання менше 50 балів за результатами загального підсумкового контролю, студент обов'язково здійснює перескладання для ліквідації академічної заборгованості.

Шкала оцінювання: національна та ЄКТС

Оцінка за національною шкалою	Оцінка за шкалою ECTS	
	Оцінка (бали)	Пояснення за розширеною шкалою
Відмінно	A (90-100)	відмінно
Добре	B (80-89)	дуже добре
	C (70-79)	добре
Задовільно	D (60-69)	задовільно
	E (50-59)	достатньо
Незадовільно	FX (35-49)	(незадовільно) з можливістю повторного складання
	F (1-34)	(незадовільно) з обов'язковим повторним курсом

Розподіл балів, які отримують студенти

Поточне оцінювання (аудиторна та самостійна робота)						Кількість балів (екзамен)	Сумарна к-ть балів
Змістовий модуль 1			Змістовий модуль 2			40	100
ЛР 1	ЛР 2	ЛР 3	ЛР 4	ЛР 5	Тестові завдання		
10	20	10	5	5	10		

5. Рекомендована література -основна

1. Методи аналізу даних : навчальний посібник для студентів / В.Є. Бахрушин. – Запоріжжя : КПУ, 2011. – 268 с.

2. Кеті О'Ніл. BIG DATA. Зброя математичного знищення. Як великі дані збільшують нерівність і загрожують демократ, 2019. 336 с.

3. Ф. Провост, Т. Фоусет. Data Science для бізнесу. Як збирати, аналізувати і використовувати дані, 2019. 400 с.

4. Т. Ерл. Основы Big Data: концепции, алгоритмы и технологии, 2018. 320 с.

5. Ульман Дж., Раджараман А., Лесковец Ю. Анализ больших наборов данных, 2016. 498 с.

6. Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data, 2015. 432 p.

7. Теоретический минимум по Big Data. Всё, что нужно знать о больших данных. – СПб.: Питер, 2019. – 208 с.

6. Інформаційні ресурси

1. Інтернет – джерела

1. Data Cleaning and Preprocessing
<https://medium.com/analytics-vidhya/data-cleaning-and-preprocessing-a4b751f4066f>
2. Аналіз даних
https://stud.com.ua/93298/statistika/analiz_danih
3. Відкритий посібник з відкритих даних
<https://socialdata.org.ua/manual4/>
4. Python для статистиків
<https://python-school.ru/python-for-statisticians/>
5. Составление гистограммы Python
<https://www.codeflow.site/ru/article/python-histograms>
6. Основы статистики с Python: описательная статистика
<https://tproger.ru/translations/basic-statistics-in-python-descriptive-statistics/>
7. Описательная статистика на Python (часть 1)
<http://chel-center.ru/python-yfc/2020/02/11/opisatel'naya-statistika-na-python-chast-1/>
8. Описательная статистика на Python (часть 2)
<https://coderlessons.com/tutorials/python-technologies/vyuchit-python-panda/python-pandas-opisatel'naya-statistika>
9. Линейная регрессия на Python: объясняем на пальцах
<http://chel-center.ru/python-yfc/2020/02/13/opisatel'naya-statistika-na-python-chast-2/>
10. Простые регрессионные модели в Python
<https://nagorny.me/courses/data-science/regression/>
11. Линейные модели: простая регрессия
<https://habr.com/ru/post/279117/>
12. Учебник по Python. Классификация клиентов на основе кластеризации методом k-средних с помощью машинного обучения SQL
<https://docs.microsoft.com/ru-ru/sql/machine-learning/tutorials/python-clustering-model?view=sql-server-ver15>
13. Обзор методов классификации в машинном обучении с помощью Scikit-Learn
<https://tproger.ru/translations/scikit-learn-in-python/>

2. Dataset

- <https://www.kaggle.com/datasets>
- <http://mlr.cs.umass.edu/ml/>
- <https://github.com/awesomedata/awesome-public-datasets>
- <https://www.data.gov/>
- <https://data.gov.uk/>
- <http://data.worldbank.org/>