

**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**

**Аналіз даних**

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

вибіркова

(вказати: обов'язкова)

**Освітньо-професійна програма «Інформаційні технології та моделювання в економіці»**

(назва програми)

**Спеціальність** 051 Економіка

(вказати: код, назва)

**Галузь знань** 05 Соціальні та поведінкові науки

(вказати: шифр, назва)

**Рівень вищої освіти** другий магістерський

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

економічний

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

**Мова навчання** українська

(вказати: на яких мовах читається дисципліна)

**Розробники:** д.е.н., доц. Білоскурський Р.Р.

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

**Профайл викладача (-ів)**

<https://emm.cv.ua/teachers/biloskurskij-ruslan-romanovich-2/>

**E-mail:**

[r.biloskursky@chnu.edu.ua](mailto:r.biloskursky@chnu.edu.ua)

**Консультації**

Щотижня у вівторок з 09.00 до 10.00 год, 64 а.

Індивідуальні консультації (за попередньою домовленістю)

у вівторок з 13.00 год, 64 а.

### 1. Анотація дисципліни (призначення навчальної дисципліни).

Розкриття змісту, методів розв'язування та можливостей практичного застосування різних типів моделей машинного навчання, а також візуалізації даних.

2. Метою навчальної дисципліни є формування у студентів системи знань з методології та інструментарію побудови, аналізу і використання алгоритмів машинного навчання та візуалізації даних.

3. Пререквізити. Для успішного вивчення дисципліни студенти повинні опанувати такі курси: «Вища математика», «Інформатика».

### 4. Результати навчання

У результаті вивчення навчальної дисципліни студент повинен

#### знати:

- теоретичні положення машинного навчання та аналізу даних;
- практичні можливості мови програмування Python та її бібліотек для аналізу даних.
- основні алгоритми машинного навчання та аналізу даних.
- метрики якості, які використовуються в алгоритмах машинного навчання.

#### вміти:

- самостійно будувати і досліджувати лінійні моделі машинного навчання на розмічених даних;
- використовувати бібліотеку NumPy для аналізу даних;
- використовувати бібліотеку Pandas для аналізу даних;
- проводити форматування та очищення наборів вхідних даних для алгоритмів машинного навчання;
- будувати графіки та візуалізувати дані з допомогою бібліотек Matplotlib та Seaborn;
- оцінювати точність алгоритмів машинного навчання за допомогою метрик якості.
- математично оцінювати, обґрунтовувати прийняття рішень в сфері економіки, моделювати економічні процеси та прогнозувати ефективність результатів.

### 5. Опис навчальної дисципліни

#### 5.1. Дидактична карта навчальної дисципліни

Форма навчання	Рік підготовки	Семестр	Кількість		Кількість годин					Вид підсумкового контролю	
			кредитів	годин	лекції	практичні	семінарські	лабораторні	самостійна робота		індивідуальні завдання
Денна		9	4.0	120	15			15	90		Іспит

## 5.2. Структура 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
<b>Змістовий модуль 1. Основи аналізу та візуалізації даних</b>												
Тема 1. Основи машинного моделювання на розмічених даних	16	2		2		12	-					
Тема 2. Бібліотеки Python для аналізу даних: Pandas та NumPy	18	2		2		14	-					
Тема 3. Візуалізація даних в Python	16	2		2		12	-					
Тема 4. Інженерія даних. Проблеми перенавчання. Метрики якості	16	2		2		12	-					
Разом за змістовим модулем 1	66	8		8		50	-					
<b>Змістовий модуль 2. Алгоритми машинного навчання</b>												
Тема 5. Аналіз даних інструментами бібліотеки Scikit-learn	18	3		2		13	-					
Тема 6. Дерева рішень та випадкові ліси	18	2		2		14	-					
Тема 7. Градієнтний бустинг в задачах регресії	18	2		3		13	-					
Разом за змістовим модулем 2	54	7		7		40	-					
Усього годин	120	15		15		90	-					

### 5.3. Зміст завдань для самостійної роботи

№	Назва теми
1	Встановлення та налаштування середовища для програмування в Python
2	Робота з масивами в NumPy
3	Графічні можливості бібліотеки Matplotlib
4	Нормування даних
5	Робота з часовими рядами
6	Моделі логістичної регресії
7	Нейромережеві технології

## 6. Система контролю та оцінювання

### 1. Порядок проведення заліку

1.1. З навчальної дисципліни проводиться залік. Студенти отримують завдання, що містить 2 теоретичні питання з дисципліни та 2 практичних завдання.

У 40 балів, що можливо отримати на заліку входить 2 теоретичних питання (по 10 балів) та два практичних завдання (по 10 балів). 10 балів ставиться за повне розкриття теоретичного питання. У випадку не повної відповіді зберігається відсоткове відношення розкриття відповіді та бального оцінювання. Практичне завдання оцінюється таким чином: 10 балів – за повністю розв'язану задачу; 8 – за правильно вказані формули, однак неправильне розв'язання; 5 – за правильно вказані формули, що необхідно для розв'язку, яке однак не має числового розв'язку; 3 – з правильною хід обчислення, однак не виконаний в повному обсязі. В іншому випадку задача оцінюється в 0 балів.

1.2. Після підготовки студентами конспекту відповіді та розв'язання задачі відбувається його усна відповідь викладачеві. У процесі цього студентів можуть бути задані додаткові питання.

### 2. Критерії оцінювання відповідей

Оцінкою "А" оцінюється повна та аргументована відповідь на теоретичне запитання, а також подано правильний розв'язок задачі, що розкриває суть матеріалу, що свідчить про вміння аналізувати матеріал та робити змістовні висновки. Відповідь повинна бути чіткою, логічною і послідовною. Відповідь оцінюється на "В" за умови розкриття теоретичного питання білету та практичного завдання, але містить неточності, що не суттєво впливають на зміст завдання. Відповідь оцінюється на "С" за умови повного та правильного розкриття одного з питань білету, але у відповіді не достатньо правильно сформульовано визначення. У той же час практичні завдання вирішені на належному рівні. Якщо підхід викладення матеріалу правильний, але виявляється недостатнє його розуміння, і в той же час практичне завдання розв'язано з деякими неточностями виставляється оцінка "D". Відповідь оцінюється на "E" у випадку правильного підходу до викладення теоретичного матеріалу та розв'язання практичного завдання. В усіх інших випадках відповідь оцінюється на "Fх".

### 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

	Е (50-59)	достатньо
<b>Незадовільно</b>	FX (35-49)	(незадовільно) з можливістю повторного складання
	F (1-34)	(незадовільно) з обов'язковим повторним вивченням дисципліни

## 5. Засоби оцінювання

Засобами оцінювання та демонстрування результатів навчання можуть бути:

- лабораторні роботи;
- усне опитування;
- реферати;
- студентські презентації та виступи на наукових заходах.

## 6. Форми поточного та підсумкового контролю

**Поточний контроль** здійснюється під час проведення лекційних та лабораторних робіт і має на меті перевірку рівня підготовленості студента до виконання конкретної роботи.

При вивченні курсу «Аналіз даних» застосовуються методи усного контролю та задачі лабораторних робіт, серед яких слід виокремити: експрес опитування лекційного матеріалу, аудиторне розв'язування задач до теми, задача лабораторних робіт, проведення поточних контрольних робіт, тестування, перевірка виконання самостійних та індивідуальних завдань.

**При здійсненні рубіжного модульного контролю** застосовується метод письмового контролю – проведення модульних контрольних робіт.

**Підсумковий контроль** проводиться у вигляді заліку.

## 7. Рекомендована література

### 7.1. Базова (основна)

1. Guido S. Introduction to Machine Learning with Python. Sebastopol, United States: O'Reilly Media, Inc, USA, 2016. – 392 с.
2. Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning. Springer, 2014. — 739 p.
3. Stephen Marsland. Machine Learning: An Algorithmic Perspective, 452 p., 2015.
4. Christopher M Bishop. Pattern recognition. Machine Learning, 128 p., 2006.
5. Ethem Alpaydin. Introduction To Machine Learning, 584 p., 2009.
6. Yaser S. Abu-Mostafa. Learning from data, 215 p., 2017
7. Alex Smola. Introduction to Machine Learning, 234 p., 2008.
8. Люгер Дж.Ф. Искусственный интеллект: стратегии и методы решения сложных проблем. – М.: Вильямс, 2005. – 864 с.
9. Trevor Hastie, Robert Tibshirani, Jerome Friedman. The Elements of Statistical Learning: Data Mining, Inference and Prediction, 764 p., 2008.

### 7.2. Допоміжна

1. Козлов А., Мхитарян В., Шишов В. Статистический анализ данных в MS Excel. – М.: Инфра-М, 2013. – 175 с.
2. Clarence Chio, David Freeman. Machine Learning and Security. – Sebastopol: O'Reilly Media, Inc., 2018. – 367 p.
3. Луис Педро Коэльо, Вилли Ричарт. Построение систем машинного обучения на языке Python. – М.: ДМК, 2016. – 301 с.

### 8. Інформаційні ресурси

1. Наукова бібліотека Чернівецького національного університету імені Юрія Федьковича / [www.library.chnu.edu.ua](http://www.library.chnu.edu.ua).
2. Національна бібліотека України ім. В.І. Вернадського / [www.nbuv.gov.ua](http://www.nbuv.gov.ua).
3. Державний комітет статистики України / [www.ukrstat.gov.ua/](http://www.ukrstat.gov.ua/)
4. Національний банк України / <http://www.nbuv.gov.ua>
5. Пошукова система / <http://uk.wikipedia.org>

### Додатково

#### Розподіл балів, які отримують студенти

Поточне оцінювання ( <i>аудиторна та самостійна робота</i> )							Кількість балів (залік)	Сумарна к-ть балів
Змістовий модуль 1				Змістовий модуль 2			40	100
T1	T2	T3	T4	T5	T6	T7		
8	8	8	8	8	10	10		