

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ**  
**Львівський національний університет імені Івана Франка**  
**Факультет прикладної математики та інформатики**  
**Кафедра обчислювальної математики**

**Затверджено**

На засіданні  
кафедри обчислювальної математики  
факультету прикладної математики та  
інформатики  
Львівського національного університету  
імені Івана Франка  
(протокол № 1 від 29 серпня 2023 р.)



Завідувач кафедри

Роман ХАПКО

**Силабус з навчальної дисципліни**  
**“Технології опрацювання та аналізу інформації”,**  
**що викладається в межах ОПП Прикладна математика**  
**другого (магістерського) рівня вищої освіти для здобувачів з**  
**спеціальності 113 – прикладна математика**

Львів 2023 р.

<b>Назва дисципліни</b>	Технології опрацювання та аналізу інформації
<b>Адреса викладання дисципліни</b>	Головний корпус ЛНУ ім. І. Франка м. Львів, вул. Університетська 1
<b>Факультет та кафедра, за якою закріплена дисципліна</b>	Факультет прикладної математики та інформатики Кафедра обчислювальної математики
<b>Галузь знань, шифр та назва спеціальності</b>	11 – математика і статистика 113 – прикладна математика
<b>Викладачі дисципліни</b>	Вавричук Василь Григорович, кандидат фізико-математичних наук, доцент кафедри обчислювальної математики
<b>Контактна інформація викладачів</b>	vasyi.vavrychuk@lnu.edu.ua, <a href="https://ami.lnu.edu.ua/employee/vavrychuk">https://ami.lnu.edu.ua/employee/vavrychuk</a> Головний корпус ЛНУ ім. І. Франка, каб. 360. м. Львів, вул. Університетська, 1
<b>Консультації з питань навчання по дисципліні відбуваються</b>	Консультації в день проведення лекцій/практичних занять, за розкладом консультацій кафедри, а також в середовищі Microsoft Teams.
<b>Сторінка курсу</b>	<a href="https://ami.lnu.edu.ua/course/tekhnolohii-opratsiuвання-ta-analizu-informatsii-prykladna-matematyka">https://ami.lnu.edu.ua/course/tekhnolohii-opratsiuвання-ta-analizu-informatsii-prykladna-matematyka</a>
<b>Інформація про дисципліну</b>	Дисципліна "Технології опрацювання та аналізу інформації" є вибірковою дисципліною для спеціальності 113 - прикладна математика для освітньої програми "Прикладна математика", яка викладається в 2-му семестрі в обсязі 3-ох кредитів (за Європейською Кредитно-Трансферною Системою ECTS).
<b>Коротка анотація дисципліни</b>	Штучний інтелект та наука про дані є одними із найбільш актуальних спеціальностей на даний час, що перебувають на перетині прикладної математики та ІТ. Платформа Kaggle для змагань з науки про дані була заснована у 2010 році як стартап, а тепер є дочірньою компанією Google. Ця платформа є ефективним майданчиком для здобуття практики з аналізу даних. У курсі розглядається вирішення проблем з аналізу даних, поставлених на платформі Kaggle. Для цього викладаються основи застосування методів машинного навчання, попередньої обробки даних, дослідження труднощів пов'язаних з недонавчанням та перенавчанням.
<b>Мета та цілі дисципліни</b>	Метою вивчення навчальної дисципліни є освоєння методів опрацювання та аналізу інформації та отримання відповідних навиків на базі платформи Kaggle.
<b>Література для вивчення дисципліни</b>	Основна література: <ol style="list-style-type: none"> <li><a href="https://pandas.pydata.org">https://pandas.pydata.org</a></li> <li>Wes McKinney. Python for Data Analysis: Data Wrangling with pandas, NumPy, and Jupyter. - O'Reilly Media. - 2022</li> <li>Aurelien Geron. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 3rd Edition. - O'Reilly Media, Inc. - 2022</li> <li>Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, Jonathan Taylor. An Introduction to Statistical Learning: with Applications in Python. - Springer. - 2023. - <a href="https://www.statlearning.com">https://www.statlearning.com</a></li> <li>Trevor Hastie, Robert Tibshirani, Jerome Friedman. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second</li> </ol>

	Edition. - <a href="https://hastie.su.domains/ElemStatLearn/">https://hastie.su.domains/ElemStatLearn/</a>
<b>Обсяг курсу</b>	Загальний обсяг: 90 годин. Аудиторних занять: 32 год., з них 16 год. лекцій та 16 год. лабораторних робіт. Самостійної роботи: 58 год.
<b>Очікувані результати навчання</b>	Після завершення цього курсу студент буде знати: <ul style="list-style-type: none"> <li>• методи вирішення задач з опрацювання та аналізу даних</li> </ul> вміти: <ul style="list-style-type: none"> <li>• навчати модель машинного навчання</li> <li>• підбирати параметри моделі</li> <li>• проводити попередню обробку даних</li> </ul>
<b>Ключові слова</b>	ІТ, програмування, прикладна математика
<b>Формат курсу</b>	Очний
<b>Теми</b>	Подано нижче у таблиці «Схема курсу»
<b>Підсумковий контроль, форма</b>	Залік
<b>Пререквізити</b>	Для вивчення курсу студенти потребують базових знань з програмування, обчислювальної математики, статистики.
<b>Навчальні методи та техніки, які будуть використовуватися під час викладання курсу</b>	Презентації, лекції, модульний контроль, індивідуальні завдання.
<b>Необхідне обладнання</b>	Комп'ютер, Internet.
<b>Критерії оцінювання (окремо для кожного виду навчальної діяльності)</b>	<p>Під час семестру студент може отримати 100 балів з них:</p> <ul style="list-style-type: none"> <li>• 60 балів за здачу індивідуальних завдань: 4 завдання по 15 балів</li> <li>• 40 балів за модульний контроль</li> </ul> <p><b>Академічна доброчесність:</b> Очікується, що роботи студентів будуть їх оригінальними дослідженнями чи міркуваннями. Відсутність посилань на використані джерела, фабрикування джерел, списування, втручання в роботу інших студентів становлять, але не обмежують, приклади можливої академічної недоброчесності. Виявлення ознак академічної недоброчесності в письмовій роботі студента є підставою для її незарахування викладачем, незалежно від масштабів плагіату чи обману.</p> <p><b>Відвідання занять</b> є важливою складовою навчання. Очікується, що всі студенти відвідають усі лекції та лабораторні заняття курсу. Студенти повинні інформувати викладача про неможливість відвідати заняття. У будь-якому випадку студенти зобов'язані дотримуватися термінів визначених для виконання всіх видів письмових робіт та індивідуальних завдань, передбачених курсом.</p> <p><b>Література.</b> Уся література, яку студенти не зможуть знайти самостійно, буде надана викладачем виключно в освітніх цілях без права її передачі третім особам. Студенти заохочуються до використання також й іншої літератури та джерел, яких немає серед рекомендованих.</p> <p><b>Політика виставлення балів.</b> Враховуються бали набрані за здачу індивідуальних завдань та бали за модульний контроль. При цьому</p>

	обов'язково враховуються присутність на заняттях та активність студента під час практичного заняття; недопустимість пропусків та запізнь на заняття; користування мобільним телефоном, планшетом чи іншими мобільними пристроями під час заняття в цілях не пов'язаних з навчанням; списування та плагіат; несвоєчасне виконання поставленого завдання і т. ін. Жодні форми порушення академічної доброчесності не толеруються.
<b>Питання на модульний контроль.</b>	<ol style="list-style-type: none"> <li>1. Чому обрані параметри моделі є оптимальними.</li> <li>2. Продемонструйте недонавчання вашої моделі.</li> <li>3. Продемонструйте перенавчання вашої моделі.</li> <li>4. Основні засоби бібліотеки pandas.</li> <li>5. Основні засоби бібліотеки scikit-learn.</li> <li>6. Робота з пропусками у вхідних даних.</li> <li>7. Робота з категоріальними вхідними даними.</li> <li>8. Регуляризація, бутстреппінг, бустінг.</li> </ol>
<b>Опитування</b>	Анкету-оцінку з метою оцінювання якості курсу буде надано по завершенню курсу.

### Схема курсу “Технології опрацювання та аналізу інформації”

Ти ж.	Тема, план, короткі тези	Форма діяльності (заняття)	Література . Ресурси в інтернеті	Завдання, год.	Термін виконання
1	<b>Тема 1. Вступ</b> Огляд завдань аналізу даних та платформи Kaggle.	Лекція (2 год.)	[4-5], <a href="https://www.kaggle.com">https://www.kaggle.com</a>	Опрацювання лекційного матеріалу (3 год.)	1 тиждень
2	<b>Тема 1. Вступ</b> Налаштування робочого середовища, Python, jupyter.  Індивідуальне завдання №1: застосування методу машинного навчання.	Практичне (2 год.)	<a href="https://code.visualstudio.com/docs/datascience/overview">https://code.visualstudio.com/docs/datascience/overview</a>	Виконання індивідуального завдання №1 (3 год.)	1 тиждень
3	<b>Тема 2. Модель машинного навчання, недонавчання, перенавчання</b>  Перевірка точності моделі на тестових даних, поділ датасету, регресія, класифікація.	Лекція (2 год.)	[4-5]	Опрацювання лекційного матеріалу (4 год.)	1 тиждень
4	<b>Тема 2. Модель машинного навчання, недонавчання,</b>	Практичне (2 год.)	[4-5]	Виконання індивідуального завдання №1 (4 год.)	6 тижнів

	<b>перенавчання</b> Приклади роботи регресії та класифікації.				
5	<b>Тема 3. Python бібліотека для аналізу даних pandas</b> Зчитування даних, DataFrame, NumPy, вибірка.	Лекція (2 год.)	[1-2]	Опрацювання лекційного матеріалу (4 год.)	1 тиждень
6	<b>Тема 3. Python бібліотека для аналізу даних pandas</b> Приклади роботи бібліотеки.	Практичне (2 год.)	[1-2]	Виконання індивідуального завдання №1 (4 год.)	4 тижні
7	<b>Тема 4. Python бібліотека для машинного навчання scikit-learn</b> Регресія та класифікація у бібліотеці scikit-learn.	Лекція (2 год.)	[3]	Опрацювання лекційного матеріалу (4 год.)	1 тиждень
8	<b>Тема 4. Python бібліотека для машинного навчання scikit-learn</b> Деревовидні моделі.	Практичне (2 год.)	[3]	Виконання індивідуального завдання №1 (4 год.)	2 тижнів
9	<b>Тема 5. Попередня обробка даних</b> Обробка пропусків, перетворення, категоріальні дані.	Лекція (2 год.)	<a href="https://www.kaggle.com/learn/intermediate-machine-learning">https://www.kaggle.com/learn/intermediate-machine-learning</a>	Опрацювання лекційного матеріалу (4 год.)	1 тиждень
10	<b>Тема 5. Попередня обробка даних</b> Здача індивідуального завдання №1. Індивідуальне завдання №2: попередня обробка даних.	Практичне (2 год.)	<a href="https://www.kaggle.com/learn/intermediate-machine-learning">https://www.kaggle.com/learn/intermediate-machine-learning</a>	Виконання індивідуального завдання №2 (4 год.)	2 тижні
11	<b>Тема 6. Регуляризація, крос-валідація</b> Види регуляризації, реалізація та переваги крос-валідації, підбір	Лекція (2 год.)	[4-5]	Опрацювання лекційного матеріалу (4 год.)	1 тиждень

	параметрів моделі.				
12	<b>Тема 2-5. Моделі машинного навчання.</b>  Модульний контроль.  Здача індивідуального завдання №2.  Індивідуальне завдання №3: оптимізація моделі машинного навчання регуляризацією та крос-валідацією.	Практичне (2 год.)	[4-5]	Виконання індивідуального завдання №3 (4 год.)	2 тижні
13	<b>Тема 7. Бутстрепінг, бустінг</b>  Ансамблеві методи, баггінг.	Лекція (2 год.)	[4-5]	Опрацювання лекційного матеріалу (4 год.)	1 тиждень
14	<b>Тема 2-5. Моделі машинного навчання.</b>  Здача індивідуального завдання №3.  Індивідуальне завдання №4: оптимізація моделі машинного навчання ансамблевими методами.	Практичне (2 год.)	[4-5]	Виконання індивідуального завдання №4 (4 год.)	2 тижні
15	<b>Тема 8. Підсумкова лекція</b>  Огляд альтернативних методів аналізу даних.	Лекція (2 год.)	[1-5]	Опрацювання лекційного матеріалу (4 год.)	1 тиждень
16	<b>Тема 8. Підсумкова лекція</b>  Здача індивідуального завдання №4.	Практичне (2 год.)	[1-5]		