

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

Оновлено та затверджено
на засіданні
кафедри обчислювальної математики
факультету прикладної математики та
інформатики
Львівського національного університету
імені Івана Франка
(протокол № 6 від 21 грудня 2023 р.)



Завідувач кафедри

Роман ХАПКО

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

Львів 2023 р.

| | |
|--|---|
| Назва дисципліни | Технології опрацювання та аналізу інформації |
| Адреса викладання дисципліни | Головний корпус ЛНУ ім. І. Франка м. Львів, вул. Університетська 1 |
| Факультет та кафедра, за якою закріплена дисципліна | Факультет прикладної математики та інформатики Кафедра обчислювальної математики |
| Галузь знань, шифр та назва спеціальності | 11 – математика і статистика 113 – прикладна математика |
| Викладачі дисципліни | Вавричук Василь Григорович, кандидат фізико-математичних наук, доцент кафедри обчислювальної математики |
| Контактна інформація викладачів | vasyi.vavrychuk@lnu.edu.ua, https://ami.lnu.edu.ua/employee/vavrychuk Головний корпус ЛНУ ім. І. Франка, каб. 360. м. Львів, вул. Університетська, 1 |
| Консультації з питань навчання по дисципліні відбуваються | Консультації в день проведення лекцій/практичних занять, за розкладом консультацій кафедри, а також в середовищі Microsoft Teams. |
| Сторінка курсу | https://ami.lnu.edu.ua/course/tekhnolohii-opratsiuвання-ta-analizu-informatsii-prykladna-matematyka-1-9 |
| Інформація про дисципліну | Дисципліна «Технології опрацювання та аналізу інформації» є вибірковою дисципліною для спеціальності 113 - прикладна математика для освітньо-наукової програми «Прикладна математика», яка викладається в 2-му семестрі в обсязі 3-ох кредитів (за Європейською Кредитно-Трансферною Системою ECTS). |
| Коротка анотація дисципліни | 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. |
| Мета та цілі дисципліни | Метою вивчення навчальної дисципліни є освоєння методів опрацювання та аналізу інформації та отримання відповідних навиків на базі платформи Kaggle. |
| Література для вивчення дисципліни | Основна література: <ol style="list-style-type: none"> https://pandas.pydata.org Wes McKinney. Python for Data Analysis: Data Wrangling with pandas, NumPy, and Jupyter. - O'Reilly Media. - 2022 Aurelien Geron. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 3rd Edition. - O'Reilly Media, Inc. - 2022 Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, Jonathan Taylor. An Introduction to Statistical Learning: with Applications in Python. - Springer. - 2023. - https://www.statlearning.com Trevor Hastie, Robert Tibshirani, Jerome Friedman. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second |

| | | | |
|---|--|--------------|----------------|
| | Edition. - https://hastie.su.domains/ElemStatLearn/ | | |
| Обсяг курсу | Загальний обсяг: 90 годин. Аудиторних занять: 32 год., з них 16 год. лекцій та 16 год. лабораторних робіт. Самостійної роботи: 58 год. | | |
| Очікувані результати навчання | Після завершення цього курсу студент буде знати: <ul style="list-style-type: none"> • методи вирішення задач з опрацювання та аналізу даних вміти: <ul style="list-style-type: none"> • навчати модель машинного навчання • підбирати параметри моделі • проводити попередню обробку даних | | |
| Ключові слова | ІТ, програмування, прикладна математика | | |
| Формат курсу | Очний | | |
| Теми | Подано нижче у таблиці «Схема курсу» | | |
| Підсумковий контроль, форма | Залік | | |
| Пререквізити | Для вивчення курсу студенти потребують базових знань з програмування, обчислювальної математики, статистики. | | |
| Навчальні методи та техніки, які будуть використовуватися під час викладання курсу | Презентації, лекції, модульний контроль, індивідуальні завдання. | | |
| Необхідне обладнання | Комп'ютер, Internet. | | |
| Критерії оцінювання (окремо для кожного виду навчальної діяльності) | Оцінювання проводиться за 100-бальною шкалою. | | |
| | Оцінка за шкалою ECTS | | Оцінка в балах |
| | A | Відмінно | 100 - 90 |
| | B | Дуже добре | 81- 89 |
| | C | Добре | 71 -80 |
| | D | Задовільно | 61 - 70 |
| | E | Достатньо | 51- 60 |
| | FX (F) | Незадовільно | 0 - 50 |
| | Оцінка за національною шкалою | | |
| | зараховано | | |
| | не зараховано | | |
| | Під час семестру студент може отримати 100 балів з них: <ul style="list-style-type: none"> • 60 балів за здачу індивідуальних завдань: 4 завдання по 15 балів • 40 балів за модульний контроль | | |
| | Критерії оцінювання індивідуальних завдань: 15 балів - Програма повністю реалізовує необхідний функціонал, осмислено спроектована з використанням архітектурних патернів, патернів проектування, принципів парадигми | | |

| | |
|--|--|
| | <p>програмування, ідіом; імена в програмі є осмисленими, підтримується єдиний стиль кодування. Студент компетентно пояснює програму та може внести зміни.</p> <p>12-14 балів - Програма не повно реалізовує необхідний функціонал. У архітектуру та код можна внести покращення. Студент компетентно пояснює програму та може внести зміни.</p> <p>7-11 балів - Програма реалізовує більшість необхідного функціоналу. Архітектура та код можуть бути дещо хаотичними. Студент може пояснити програму, але з внесенням змін можуть бути труднощі.</p> <p>3-6 бали - Програма реалізовує найпростіший юзкейс. Студент може пояснити програму, але з внесенням змін можуть бути труднощі.</p> <p>1-2 бали - Програма наявна, студент може її запустити та продемонструвати, проте пояснення її роботи є незадовільним.</p> <p>0 балів - програми немає або вона не запускається.</p> <p>Критерії оцінювання модульного контролю (20 завдань по 2б.):</p> <p>2 бали: повністю вірна та повна відповідь;</p> <p>1 бал: не повністю повна відповідь або відповідь з несуттєвими неточностями, проте зрозуміла суть;</p> <p>0 балів: відповідь на питання неправильна.</p> <p>Академічна доброчесність: 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>Відвідання занять є важливою складовою навчання. Очікується, що всі студенти відвідають усі лекції та практичні 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>Література. 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>Політика виставлення балів. Враховуються бали набрані за здачу індивідуальних завдань та бали за модульний контроль. При цьому 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.</p> <p>Жодні форми порушення академічної доброчесності не толеруються.</p> |
| <p>Питання на модульний контроль.</p> | <ol style="list-style-type: none"> 1. Чому обрані параметри моделі є оптимальними. 2. Продемонструйте недонавчання вашої моделі. 3. Продемонструйте перенавчання вашої моделі. 4. Основні засоби бібліотеки pandas. |

| | |
|-------------------|---|
| | <p>5. Основні засоби бібліотеки scikit-learn.</p> <p>6. Робота з пропусками у вхідних даних.</p> <p>7. Робота з категоріальними вхідними даними.</p> <p>8. Регуляризація, бутстрепінг, бустінг.</p> |
| Опитування | 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 | Тема 1. Вступ Огляд завдань аналізу даних та платформи Kaggle. | Лекція (2 год.) | [4-5], https://www.kaggle.com | Опрацювання лекційного матеріалу (3 год.) | 1 тиждень |
| 2 | Тема 1. Вступ Налаштування робочого середовища, Python, jupyter. Індивідуальне завдання №1: застосування методу машинного навчання. | Практичне (2 год.) | https://code.visualstudio.com/docs/datascience/overview | Виконання індивідуального завдання №1 (3 год.) | 1 тиждень |
| 3 | Тема 2. Модель машинного навчання, недонавчання, перенавчання Перевірка точності моделі на тестових даних, поділ датасету, регресія, класифікація. | Лекція (2 год.) | [4-5] | Опрацювання лекційного матеріалу (4 год.) | 1 тиждень |
| 4 | Тема 2. Модель машинного навчання, недонавчання, перенавчання Приклади роботи регресії та класифікації. | Практичне (2 год.) | [4-5] | Виконання індивідуального завдання №1 (4 год.) | 6 тижнів |
| 5 | Тема 3. Python бібліотека для аналізу даних pandas Зчитування даних, DataFrame, NumPy, | Лекція (2 год.) | [1-2] | Опрацювання лекційного матеріалу (4 год.) | 1 тиждень |

| | | | | | |
|----|--|--------------------|---|--|-----------|
| | вибірка. | | | | |
| 6 | Тема 3. Python бібліотека для аналізу даних pandas Приклади роботи бібліотеки. | Практичне (2 год.) | [1-2] | Виконання індивідуального завдання №1 (4 год.) | 4 тижні |
| 7 | Тема 4. Python бібліотека для машинного навчання scikit-learn Регресія та класифікація у бібліотеці scikit-learn. | Лекція (2 год.) | [3] | Опрацювання лекційного матеріалу (4 год.) | 1 тиждень |
| 8 | Тема 4. Python бібліотека для машинного навчання scikit-learn Деревовидні моделі. | Практичне (2 год.) | [3] | Виконання індивідуального завдання №1 (4 год.) | 2 тижнів |
| 9 | Тема 5. Попередня обробка даних Обробка пропусків, перетворення, категоріальні дані. | Лекція (2 год.) | https://www.kaggle.com/learn/intermediate-machine-learning | Опрацювання лекційного матеріалу (4 год.) | 1 тиждень |
| 10 | Тема 5. Попередня обробка даних Здача індивідуального завдання №1. Індивідуальне завдання №2: попередня обробка даних. | Практичне (2 год.) | https://www.kaggle.com/learn/intermediate-machine-learning | Виконання індивідуального завдання №2 (4 год.) | 2 тижні |
| 11 | Тема 6. Регуляризація, крос-валідація Види регуляризації, реалізація та переваги крос-валідації, підбір параметрів моделі. | Лекція (2 год.) | [4-5] | Опрацювання лекційного матеріалу (4 год.) | 1 тиждень |
| 12 | Тема 2-5. Моделі машинного навчання. Модульний контроль. Здача індивідуального завдання №2. Індивідуальне завдання | Практичне (2 год.) | [4-5] | Виконання індивідуального завдання №3 (4 год.) | 2 тижні |

| | | | | | |
|----|--|--------------------|-------|--|-----------|
| | №3: оптимізація моделі машинного навчання регуляризацією та крос-валідацією. | | | | |
| 13 | Тема 7. Бутстрепінг, бустінг Ансамблеві методи, баггінг. | Лекція (2 год.) | [4-5] | Опрацювання лекційного матеріалу (4 год.) | 1 тиждень |
| 14 | Тема 2-5. Моделі машинного навчання. 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 | Тема 8. Підсумкова лекція Огляд альтернативних методів аналізу даних. | Лекція (2 год.) | [1-5] | Опрацювання лекційного матеріалу (4 год.) | 1 тиждень |
| 16 | Тема 8. Підсумкова лекція Здача індивідуального завдання №4. | Практичне (2 год.) | [1-5] | | |