

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ПРИКАРПАТСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ВАСИЛЯ СТЕФАНИКА

Факультет *математики та інформатики*
Кафедра *диференціальних рівнянь і прикладної математики*

СИЛАБУС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ
АНАЛІЗ ДАНИХ

Рівень освіти: _____ Перший (бакалаврський) _____
Галузь знань: _____ 11 Математика та статистика _____
Спеціальність: _____ 113 Прикладна математика _____
Освітня програма: _____ Комп'ютерне моделювання та _____
_____ технології програмування _____
Факультет: _____ Математики та інформатики _____

Затверджено на засіданні кафедри
Протокол № 1 від “31” серпня 2021 р.

ЗМІСТ

1. Загальна інформація
2. Анотація до курсу
3. Мета та цілі курсу
4. Компетентності
5. Результати навчання
6. Організація навчання курсу
7. Система оцінювання курсу
8. Політика курсу
9. Рекомендована література

| 1. Загальна інформація | |
|---|---|
| Назва дисципліни | Аналіз даних |
| Рівень вищої освіти | Бакалавр |
| Викладач (-і) | Гой Тарас Петрович |
| Контактний телефон викладача | 8(0342) 596027 |
| E-mail викладача | taras.goy@pnu.edu.ua |
| Формат дисципліни | Лекції та практичні заняття |
| Обсяг дисципліни | 3 кредити |
| Посилання на сайт дистанційного навчання | https://d-learn.pnu.edu.ua |
| Консультації | середа, 15.00-16.00 |
| 2. Анотація до курсу | |
| Під час навчання студенти ознайомляться з основними поняттями, методами та засобами попередньої підготовки, статистичної обробки, візуалізації та аналізу даних, основами машинного навчання. На лабораторних заняттях навчатимуться використовувати основні бібліотеки мови Python для роботи з даними (NumPy, SciPy, Matplotlib, Pandas). | |
| 3. Мета та цілі курсу | |
| Метою навчальної дисципліни є формування у студентів теоретичних знань і практичних здатностей застосовувати загальні методи і засоби підготовки, статистичної обробки, візуалізації та аналізу даних, використовуючи основні бібліотеки мови Python для роботи з даними. Дисципліна «Аналіз даних» належить до спеціальних дисциплін, що забезпечують формування системи теоретичних та практичних знань та навичок щодо застосування механізмів інтелектуального аналізу для вирішення прикладних задач обробки великих даних. Завданнями дисципліни є вивчення принципів, методів та інструментарію аналізу даних та способів вирішення типових задач аналізу даних. | |
| 4. Компетентності | |
| ЗК 02. Здатність застосовувати знання у практичних ситуаціях. ЗК 06. Знання і розуміння предметної області та розуміння професійної діяльності. ФК 01. Здатність використовувати і адаптувати математичні теорії, методи та прийоми для доведення математичних тверджень і теорем. ФК 03. Здатність обирати та застосовувати математичні методи до розв'язування прикладних задач, моделювання, аналізу, проектування, керування, прогнозування, прийняття рішень. ФК 07. Здатність експлуатувати та обслуговувати програмне забезпечення автоматизованих та інформаційних систем різного призначення. ФК 09. Здатність до проведення математичного і комп'ютерного моделювання, аналізу та обробки даних, обчислювального експерименту, розв'язання формалізованих задач за допомогою спеціалізованих програмних засобів. ФК 14. Здатність сформулювати математичну постановку задачі, спираючись на постановку мовою предметної галузі, та обирати метод її розв'язання, що забезпечує потрібну точність і надійність результату. | |
| 5. Результати навчання | |
| РН 01. Демонструвати знання й розуміння основних концепцій, принципів, теорій прикладної математики і використовувати їх на практиці. РН 08. Поєднувати методи математичного та комп'ютерного моделювання з неформальним процедурами експертного аналізу для пошуку оптимальних рішень. РН 10. Володіти методиками вибору раціональних методів та алгоритмів розв'язання математичних задач оптимізації, дослідження операцій, оптимального керування і прийняття рішень, аналізу даних. РН 13. Використовувати у практичній роботі спеціалізовані програмні продукти та програмні системи комп'ютерної математики. | |
| 6. Організація навчання курсу | |
| Обсяг курсу | |
| Вид заняття | Загальна кількість годин |
| лекції | 20 |
| лабораторні | 40 |
| самостійна робота | 30 |

| Ознаки курсу | | | | | |
|---|----------------------|------------------------|--|----------------|---------------------|
| Семестр | Спеціальність | Курс (рік навчання) | Нормативний / вибірковий | | |
| 7 | Прикладна математика | 4 | нормативний | | |
| Тематика курсу | | | | | |
| Тема, план | Форма заняття | Література | Завдання, год | Вага оцінки | Термін виконання |
| Основні поняття обробки даних. Бібліотеки мови Python для аналізу даних. | Лекція 1 | [1, 2, 6] | Опрацювати матеріал лекції з рек. літературою, 2 год | 0 | 1-й тиждень |
| Методи статистичної обробки даних. Статистична обробка даних засобами Python | Лекція 2 | [1, 2, 5] | Опрацювати матеріал лекції з рек. літературою, 2 год | 0 | 1-й тиждень |
| Бібліотеки для роботи з даними у Python | Лаб. зан. 1 | [2, 7, 8] | Виконати завдання лаб. роботи, 2 год | 0 | 2-й тиждень |
| Статистична обробка даних засобами Python | Лаб. зан. 2 | [1–4, 7, 8] | Виконати завдання лаб. роботи, 2 год | 0 | 2-й тиждень |
| Попередня обробка даних. Виявлення аномальних даних. Попередня обробка даних в Pандас | Лекція 3 | [1, 3, 5] | Опрацювати матеріал лекції з рек. літературою, 2 год | 0 | 3-й тиждень |
| Структури та представлення даних | Лаб. зан. 3 | [1–4, 7, 8] | Виконати завдання лаб. роботи, 2 год | 0 | 3-й тиждень |
| Попередня обробка даних засобами Python. Виявлення аномалій та залежностей. | Лаб. зан. 4 | [1–4, 7, 8] | Виконати завдання лаб. роботи, 2 год | 0 | 4-й тиждень |
| Кореляційних аналіз даних. Згладжування даних | Лекція 4 | [1–5] | Опрацювати матеріал лекції з рек. літературою, 2 год | 0 | 4-й тиждень |
| Кореляційних аналіз даних | Лаб. зан. 5 | [1–4, 7, 8] | Виконати завдання лаб. роботи, 2 год | 0 | 5-й тиждень |
| Згладжування даних | Лаб. зан. 6 | [1–4, 7, 8] | Виконати завдання лаб. роботи, 2 год | 0 | 5-й тиждень |
| Часові ряди | Лекція 5 | [1–5] | Опрацювати матеріал лекції з рек. літературою, 2 год | 0 | 6-й тиждень |
| Статистична обробка часових рядів | Лаб. зан. 7 | [1–5] | Виконати завдання лаб. роботи, 2 год | 0 | 6-й тиждень |
| Аналіз часових рядів засобами Pандас. | Лаб. зан. 8 | [1–3, 7] | Виконати завдання лаб. роботи, 2 год | 0 | 7-й тиждень |
| Візуальне представлення даних. Візуалізація даних з Matplotlib. | Лекція 6 | [1–3, 7] | Опрацювати матеріал лекції з рек. літературою, 2 год | 0 | 7-й тиждень |
| Візуалізація даних засобами | Лаб. зан. 9 | [1–3, 7] | Виконати | 0 | 8-й |

| | | | | | |
|---|--------------|-------------|--|-----|--------------|
| Python | | | завдання лаб. роботи, 2 год | | тиждень |
| Використання бібліотеки Seaborn для візуалізації | Лаб. зан. 10 | [1–5] | Виконати завдання лаб. роботи, 2 год | 0 | 8-й тиждень |
| Основи інтелектуального аналізу даних. | Лекція 7 | [1–4, 7, 8] | Опрацювати матеріал лекції з рек. літературою, 2 год | 0 | 9-й тиждень |
| Інтелектуальний аналіз даних. | Лаб. зан. 11 | [1–5] | Виконати завдання лаб. роботи, 2 год | 0 | 9-й тиждень |
| Контрольна робота 1 | Лаб. зан. 12 | [1–8] | | 0.2 | 10-й тиждень |
| Основи машинного навчання. Підготовка даних для навчання. | Лекція 8 | [1–4, 7, 8] | Опрацювати матеріал лекції з рек. літературою, 2 год | 0 | 10-й тиждень |
| Машинне навчання засобами бібліотеки Scikit-learn | Лаб. зан. 13 | [2–4, 7, 8] | Виконати завдання лаб. роботи, 2 год | 0 | 11-й тиждень |
| Машинне навчання засобами бібліотеки Scikit-learn | Лаб. зан. 14 | [3, 4, 7] | Виконати завдання лаб. роботи, 2 год | 0 | 11-й тиждень |
| Методи класифікації даних | Лекція 9 | [1–5, 7] | Опрацювати матеріал лекції з рек. літературою, 2 год | 0 | 12-й тиждень |
| Основні методи класифікації даних, їхні особливості та критерії вибору | Лаб. зан. 15 | [2–4, 6, 8] | Виконати завдання лаб. роботи, 2 год | 0 | 12-й тиждень |
| Наївний байєсівський класифікатор | Лаб. зан. 16 | [2–4, 7, 8] | Виконати завдання лаб. роботи, 2 год | 0 | 13-й тиждень |
| Метод опорних векторів. Основи кластерного аналізу | Лекція 10 | [1–3, 6] | Опрацювати матеріал лекції з рек. літературою, 2 год | 0 | 13-й тиждень |
| Реалізація байєсівського класифікатора у Scikit-learn. Використання та реалізація лінійної регресії. Метод опорних векторів | Лаб. зан. 17 | [3, 4, 7] | Виконати завдання лаб. роботи, 2 год | 0 | 14-й тиждень |
| Основні методи кластерного аналізу. Метод k-середніх. | Лаб. зан. 18 | [3, 4, 7] | Виконати завдання лаб. роботи, 2 год | 0 | 14-й тиждень |
| Контрольна робота 2 | Лаб. зан. 19 | [1-8] | Індивід. завдання, 2 год | 0.2 | 15-й тиждень |
| Підсумкове заняття | Лаб. зан. 20 | [1-8] | | | 15-й тиждень |

7. Система оцінювання курсу

| | |
|-----------------------------------|--|
| Загальна система оцінювання курсу | Підсумкова оцінка з дисципліни є сумою оцінок за кожен з таких видів робіт: своєчасне виконання і захист лабораторних робіт, виконання контрольних робіт, підсумковий контроль (екзамен). Підсумкова оцінка визначається за різними шкалами (100-бальна, ECTS, національна). |
|-----------------------------------|--|

| | |
|--|---|
| Вимоги до контрольної роботи | Пакети індивідуальних завдань для проведення контрольних робіт містять 5 завдань у кожному. Максимальна сумарна оцінка за виконання контрольних робіт становить 15 балів. |
| Лабораторні заняття | Максимальна оцінка за своєчасне виконання і захист лабораторних робіт становить 15 балів. |
| Умови допуску до підсумкового контролю | Загальна кількість балів за навчальну (аудиторну) і практичну роботу становить не менше 25 балів. |

8. Політика курсу

- відвідування лекційних та лабораторних занять є обов'язковою складовою вивчення матеріалу;
 - на лекції викладач користується власним презентаційним матеріалом; використовує гугл-диск для викладання матеріалу поточної лекції, додаткової інформації, завдань до лабораторних робіт та інше;
 - звіти з лабораторних робіт завантажуються напередодні захисту; для захисту лабораторної роботи необхідно продемонструвати роботу відповідної завданню програми та відповісти на питання щодо програми та контрольні питання;
 - контрольні роботи пишуться без застосування допоміжних засобів (мобільні телефони, планшети та ін.); результат завантажується у файлі через гугл-форму до відповідної директорії гугл-диску;
 - заохочувальні бали виставляються за: участь у факультетських олімпіадах з навчальних дисциплін, участь у конкурсах робіт, підготовка оглядів наукових праць тощо. Кількість заохочувальних балів – не більше 10.
- Кожен студент зобов'язаний дотримуватися принципів академічної доброчесності.

9. Рекомендована література

1. Марченко О.О., Россада Т.В. Актуальні проблеми Data Mining: Навчальний посібник для студентів факультету комп'ютерних наук та кібернетики. – Київ. – 2017. – 150 с.
2. Ланде Д.В., Субач І.Ю., Бояринова Ю.Є. Основи теорії і практики інтелектуального аналізу даних у сфері кібербезпеки: навчальний посібник. – К.: ІСЗЗІ КПІ ім. Ігоря Сікорського», 2018. – 297 с.
3. Бахрушин В.Є. Методи аналізу даних : навчальний посібник для студентів. – Запоріжжя : КПУ, 2011. – 268 с.
4. Василенко О. А., Сенча І. А. Математично-статистичні методи аналізу у прикладних дослідженнях: навч. посіб. – Одеса: ОНАЗ ім. О. С. Попова, 2011. – 166 с.
5. Ситник В. Ф., Краснюк М. Т. Інтелектуальний аналіз даних (дейтамайнінг). – К.: КНЕУ, 2007. – 376 с.
6. Черняк О. І., Захарченко П. В. Інтелектуальний аналіз даних: підручник. – К.: Знання, 2014. – 599 с.
7. Інтелектуальний аналіз даних: методичні вказівки до виконання лабораторних робіт для студентів напряму підготовки 6.050101 «Комп'ютерні науки» денної форми навчання. Част. 1 / уклад. Н.М. Ліщина, А.А. Ящук. – Луцьк : Луцький НТУ, 2015. – 102 с.
8. Joel Grus. Data Science from Scratch. – O'Reilly Media, 2019, 513 p.

Викладач _____ Т.П. ГОЙ