

**КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ТАРАСА ШЕВЧЕНКА**

ФАКУЛЬТЕТ КОМП'ЮТЕРНИХ НАУК ТА КІБЕРНЕТИКИ

Кафедра теорії та технології програмування

«ЗАТВЕРДЖУЮ»

Заступник декана
з навчальної роботи

_____ Кашпур О.Ф.

«___» _____ 2018 року

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ
"ОСНОВИ DATA MINING"
для студентів

галузь знань **12 «Інформаційні технології»**
(цифр і назва)

спеціальність **122 «Комп'ютерні науки»**
(цифр і назва спеціальності)

освітній рівень **бакалавр**
(молодший бакалавр, бакалавр, магістр)

освітня програма **«Інформатика»**
(назва освітньої програми)

вид дисципліни **обов'язкова**

| | |
|--|-------------------|
| Форма навчання | денна |
| Навчальний рік | 2018/2019 |
| Семестр | 7 |
| Кількість кредитів ECTS | 3 |
| Мова викладання, навчання та оцінювання | українська |
| Форма заключного контролю | залік |

Викладачі: **к.ф.-м.н., ас. Федорова М.В.**

Пролонговано: на 20__/20__ н.р. _____ (_____) «__» ____ 20__р.
(підпис, ПІБ, дата)

на 20__/20__ н.р. _____ (_____) «__» ____ 20__р.
(підпис, ПІБ, дата)

КИЇВ – 2018

Розробник: Федорова Марія Вікторівна, к.ф.-м.н., асистент кафедри «Теорії та технології програмування»

ЗАТВЕРДЖЕНО

В. о. зав. кафедри «Теорії та технології програмування»

_____ (Панченко Т.В.)
(підпис) (прізвище та ініціали)

Протокол № ____ від «____» _____ 20__ р.

Схвалено науково-методичною комісією факультету комп'ютерних наук та кібернетики

Протокол від «____» _____ 20__ року № ____

Голова науково-методичної комісії _____ (Хусаїнов Д.Я.)
(підпис) (прізвище та ініціали)

«____» _____ 20__ року

Затверджено вченою радою факультету комп'ютерних наук та кібернетики

Протокол від «____» _____ 2018 року № ____

Голова вченої ради факультету _____ А.В. Анісімов
(підпис) (прізвище та ініціали)

ВСТУП

1. Мета дисципліни – знайомство з теоретичними аспектами технології Data Mining, методами, можливостями їх застосування.

2. Попередні вимоги до опанування або вибору навчальної дисципліни (за наявності):

1. Знати:

- математичний апарат, принципи програмування;
- методи розробки алгоритмів, конструювання програмного забезпечення та структур даних і знань.

2. Вміти:

- застосовувати на практиці інструментальні засоби проектування та розробки програмного забезпечення;
- формулювати та досліджувати математичні моделі, зокрема дискретні математичні моделі, обґрунтовувати вибір методів і підходів для розв'язування теоретичних і прикладних задач у галузі комп'ютерних наук.

3. Володіти навичками:

- програмування;
- проектування та розробки програмного забезпечення.

3. Анотація навчальної дисципліни:

Навчальна дисципліна “Основи Data Mining” є складовою програми підготовки фахівців за першим (бакалаврським) рівнем вищої освіти *галузі знань* 12 „Інформаційні технології” зі *спеціальності* 122 „Комп'ютерні науки”, *освітньо-професійної програми* – „Інформатика”.

Дана дисципліна є обов'язковою навчальною дисципліною за *програмою “Інформатика”*.

Викладається у 7 семестрі 4 курсу в **обсязі – 90 год.**

(**3 кредити ECTS**) зокрема: *лекції – 40 год., консультації – 2 год., самостійна робота – 48 год.* У курсі передбачено **2 змістових модулі** та **2 модульні контрольні роботи**. Завершується дисципліна – **заліком у 7 семестрі**.

В результаті вивчення навчальної дисципліни студент повинен:

знати основні поняття та теоретичні аспекти методів Data Mining.

вміти застосовувати методи та алгоритми інтелектуального аналізу даних для задач класифікації, прогнозування, кластерного аналізу, пошуку асоціативних правил з використанням програмних інструментів підтримки багатовимірного аналізу даних на основі використання технологій DataMining, TextMining, WebMining.

Для допуску до дисципліни «Основи Data Mining» освітньо-професійної програми «Інформатика» студент повинен опанувати компетентності та результати навчання, які надають дисципліни «Програмування», «Дискретна математика», «Математичний аналіз», «Математична логіка», «Теорія ймовірностей», «Математична статистика», «Диференційні рівняння», «Чисельні методи» програми «Інформатика». Дисципліна «Основи Data Mining» є базовою для засвоєння дисциплін спеціалізації, дисциплін вільного вибору студента програмістського спрямування програми «Інформатика», зокрема «Композиційної семантики SQL-подібних мов» та виробничої практики "Інформаційні системи та технології".

4. Завдання (навчальні цілі):

набуття знань, умінь та навичок (компетентностей) на рівні новітніх досягнень у програмуванні та аналізі даних, відповідно до кваліфікації фахівців з інформаційних технологій. Зокрема, розвивати:

- здатність до алгоритмічного та логічного мислення;
- здатність до побудови логічних висновків, використання формальних мов і моделей алгоритмічних обчислень, проектування, розроблення й аналізу алгоритмів, оцінювання їх ефективності та складності, розв'язності та нерозв'язності алгоритмічних проблем для адекватного моделювання предметних областей і створення програмних та інформаційних систем;
- здатність до інтелектуального багатовимірного аналізу даних та їхньої оперативної аналітичної обробки з візуалізацією результатів аналізу в процесі розв'язування прикладних задач.

5. Результати навчання за дисципліною:

| Результат навчання (РН) (1. знати; 2. вміти; 3. комунікація; 4. автономність та відповідальність) | | Форми (та/або методи і технології) викладання і навчання | Методи оцінювання та пороговий критерій оцінювання (за необхідності) | Відсоток у підсумковій оцінці з дисципліни |
|--|--|--|--|--|
| Код | Результат навчання | | | |
| РН 1.1 | Знати основні поняття Data Mining, методи та алгоритми інтелектуального аналізу даних для задач класифікації та регресії; пошуку асоціативних правил; кластеризації. | Лекції, самостійна робота | Тест | 15% |
| РН 1.2 | Знати основні методи Data Mining: нечітка логіка, генетичні алгоритми, нейронні мережі. | Лекції, самостійна робота | Тест | 15% |
| РН 1.3 | Знати основні методи класифікації та регресії, методи побудови правил класифікації, дерев рішень, функцій класифікації та регресії. | Лекції, самостійна робота | Тест | 15% |
| РН 2.1 | Вміти застосовувати методи та алгоритми інтелектуального аналізу даних для задач класифікації, регресії, прогнозування, кластерного аналізу. | Самостійна робота | Тест (5%), захист проєктів (20%) | 25% |
| РН 2.2 | Вміти застосовувати технології DataMining, TextMining, WebMining. | Самостійна робота | Тест (3%), захист проєктів (7%) | 10% |
| РН 4.1 | Розвинути навички командної роботи та проєктної діяльності при розробці комплексного програмного забезпечення для інтелектуального аналізу даних. | Самостійна робота | Захист проєктів | 20% |

6. Співвідношення результатів навчання дисципліни із програмними результатами навчання

| Результати навчання дисципліни | 1.1 | 1.2 | 1.3 | 2.1 | 2.2 | 4.1 |
|---|-----|-----|-----|-----|-----|-----|
| Програмні результати навчання | | | | | | |
| <i>(з опису освітньої програми)</i> | | | | | | |
| ПРН12. Вміти застосовувати методи та алгоритми інтелектуального аналізу даних для задач класифікації, прогнозування, кластерного аналізу, пошуку асоціативних | + | + | + | + | + | + |

| | | | | | |
|--|--|--|---|---|---|
| правил з використанням програмних інструментів підтримки багатовимірного аналізу даних на основі використання технологій DataMining, TextMining, WebMining. | | | | | |
| ПРН18.2. Аналізувати, оцінювати і вибирати інструментальні та обчислювальні засоби, парадигми, технології, алгоритмічні і програмні рішення при проектуванні та розробці програмних систем. | | | + | + | + |

7. Схема формування оцінки.

7.1 Форми оцінювання студентів:

- семестрове оцінювання:

1. Тести: РН 1.1., РН 1.2, РН 1.3, РН 2.1, РН 2.2 — 53/31,8 балів.
2. Самостійна робота: РН 2.1, РН 2.2, РН 4.1 — 47/28,2 балів.

- підсумкове оцінювання у формі заліку. Виставляється за результатами роботи студентами впродовж усього семестру та не передбачає додаткових заходів оцінювання для успішних студентів.

Запитання для підготовки до заліку

1. Поняття Data Mining. Задачі Data Mining: класифікації та регресії; пошуку асоціативних правил; кластеризації. Практичне використання Data Mining. Методи Data Mining: нечітка логіка, генетичні алгоритми, нейронні мережі. Засоби Data Mining
2. Постановка задачі класифікації та регресії.
3. Методи побудови правил класифікації: алгоритм побудови елементарних правил (1-rule), алгоритм Naive Bayes.
4. Методи побудови дерев рішень: «розділяй та пануй», алгоритм покриття.
5. Методи побудови функцій класифікації та регресії: лінійні, нелінійні, SVM, Regularization Networks.
6. Постановка задачі пошуку асоціативних правил, її різновиди. Представлення результатів.
7. Алгоритм Аpriori та його різновиди.
8. Постановка задачі кластеризації. Базові алгоритми кластеризації. Адаптивні методи кластеризації.
9. Методи візуалізації. Стандарти Data Mining: CWM, CRISP, PMML

7.2 Організація оцінювання:

Терміни проведення форм оцінювання:

1. Тест РН 1.1. : до 8 тижня семестру.
2. Тест РН 1.2. : до 14 тижня семестру.

У випадку відсутності студента з поважних причин відпрацювання та перездачі МКР здійснюються у відповідності до „Положення про порядок оцінювання знань студентів при кредитно-модульній системі організації навчального процесу” від 1 жовтня 2010 року.

У разі неякісного виконання лабораторної роботи, викладач має право не зарахувати лабораторну роботу, або знизити за неї бали.

Студент має право здавати лабораторні роботи після закінчення визначеного для них терміну, але з втратою одного балу за кожен тиждень, який пройшов з моменту закінчення терміну її здачі.

7.3 Шкала відповідності оцінок

| | |
|----------------------------------|--------|
| Відмінно / Excellent | 90-100 |
| Добре / Good | 75-89 |
| Задовільно / Satisfactory | 60-74 |
| Незадовільно / Fail | 0-59 |
| Зараховано / Passed | 60-100 |
| Не зараховано / Fail | 0-59 |

8. Структура навчальної дисципліни. Тематичний план лекцій і лабораторних занять

| № лекції | Назва лекції | Кількість годин | |
|----------|--|-----------------|----------------|
| | | Лекції | Самост. робота |
| | Змістовий модуль 1. Задачі класифікації та регресії | | |
| 1 | Тема 1. Традиційні мови програмування для Data Mining. Огляд основних особливостей мови програмування Python. | 2 | 6 |
| 2 | Тема 2. Поняття Data Mining. Задачі Data Mining: класифікації та регресії; пошуку асоціативних правил; кластеризації. | 2 | 2 |
| 3 | Тема 3. Практичне використання Data Mining. Методи Data Mining: нечітка логіка, генетичні алгоритми, нейронні мережі. Засоби Data Mining. | 2 | 2 |
| 4 | Тема 4. Методи класифікації та регресії. Методи побудови правил класифікації: алгоритм побудови елементарних правил (1-rule). | 2 | 2 |
| 5 | Тема 5. Методи класифікації та регресії. Методи побудови правил класифікації: алгоритм Naïve Bayes. | 2 | 2 |
| 6 | Тема 6. Методи побудови дерев рішень: «розділяй та пануй», алгоритм покриття. | 2 | 2 |
| 7 | Тема 7. Методи побудови функцій класифікації та регресії: лінійні та нелінійні. | 2 | 2 |
| 8 | Тема 8. Методи побудови функцій класифікації та регресії: SVM та Regularization Networks. | 2 | 2 |
| 9 | Тема 9. Багатовимірний аналіз. Теплові карти, відношення, згладжування. | 2 | 2 |
| 10 | Тема 10. Обробка даних: нормалізація, викиди, заповнення пропусків даних. | 2 | 2 |
| | Змістовий модуль 2. Задачі пошуку асоціативних правил та кластеризації | | |
| 11 | Тема 11. Постановка задачі пошуку асоціативних правил, її різновиди. Представлення результатів. | 2 | 2 |
| 12 | Тема 12. Алгоритм Apriori та його різновиди. | 2 | 2 |
| 13 | Тема 13. Постановка задачі кластеризації. | 2 | 2 |
| 14 | Тема 14. Базові алгоритми кластеризації. | 2 | 4 |
| 15 | Тема 15. Адаптивні методи кластеризації. | 2 | 2 |
| 16 | Тема 16. Методи візуалізації. Аналіз просторових даних. Побудова карт. | 2 | 2 |
| 17 | Тема 17. Особливості аналізу неструктурованої текстової інформації. Text Mining. | 2 | 4 |
| 18 | Тема 18. Стандарти Data Mining: CWM. | 2 | 2 |
| 19 | Тема 19. Стандарти Data Mining: CRISP. | 2 | 2 |
| 20 | Тема 20. Стандарти Data Mining: PMML. | 2 | 2 |
| | ВСЬОГО | 40 | 48 |

Загальний обсяг 90 год., в тому числі:

Лекцій – **40 год.**

Консультації – **2 год.**

Самостійна робота - **48 год.**

9. Рекомендовані джерела:

Основні:

1. Leskovec J., Rajaraman A., Ullman J. D. Mining of massive datasets. – Cambridge university press, 2014.
2. Барсегян А. Методы и модели анализа данных: OLAP и Data Mining. – БХВ-Петербург, 2004.
3. Паклин Н.Б. Бизнес-аналитика: от данных к знаниям. – Питер, 2013..

Додаткові:

1. Tan P. N., Steinbach M., Kumar V. Introduction to data mining. 1st. – 2005.
2. Han J., Pei J., Kamber M. Data mining: concepts and techniques. – Elsevier, 2011.
3. Zaki M. J., Meira Jr W., Meira W. Data mining and analysis: fundamental concepts and algorithms. – Cambridge University Press, 2014.
4. Aggarwal C. C. Data mining: the textbook. – Springer, 2015.
5. Friedman J., Hastie T., Tibshirani R. The elements of statistical learning. – New York, NY, USA: Springer series in statistics, 2001. – Т. 1. – №. 10.
6. <https://www.deeplearning.ai/courses/>
7. <https://www.coursera.org/browse/data-science>