

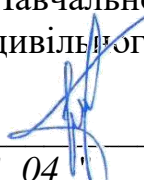
ЛЬВІВСЬКИЙ ДЕРЖАВНИЙ УНІВЕРСИТЕТ БЕЗПЕКИ ЖИТТЄДІЯЛЬНОСТІ

**КАФЕДРА УПРАВЛІННЯ ПРОЕКТАМИ, ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ ТА
ТЕЛЕКОМУНІКАЦІЙ**

ЗАТВЕРДЖУЮ

Голова Вченої ради

Навчально-наукового інституту
цивільного захисту

 Василь ПОПОВИЧ
" 04 " вересня 2020р.

ОК 2.17 ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ

ПРОГРАМА

навчальної нормативної дисципліни

підготовки бакалавра

спеціальності: 122 Комп'ютерні науки

за освітньою програмою: Комп'ютерні науки

Львів
2020 рік

Розробник програми:


Анатолій Тригуба, професор кафедри управління проектами, інформаційних технологій та телекомунікацій, докт. тех. наук, професор

Рецензент: Роман Дунець, завідувач кафедри спеціальних комп'ютерних систем, Національного університету «Львівська політехніка», доктор технічних наук, професор

Програму рекомендовано кафедрою управління проектами, інформаційних технологій та телекомунікацій

Протокол від “27” серпня 2020 року № 1

Начальник (завідувач) кафедри управління проектами, інформаційних технологій та телекомунікацій


(підпис)

Олександр ПРИДАТКО
(ім'я та прізвище)

Схвалено Вченою радою навчально-наукового інституту цивільного захисту

Протокол від “04” вересня 2020 року № 1

ВСТУП

Програма вивчення нормативної навчальної дисципліни «Інтелектуальний аналіз даних» складена відповідно до освітньо-професійної програми підготовки бакалаврів зі спеціальності 122 «Комп'ютерні науки».

Предметом вивчення навчального курсу є процес навчання і підготовки фахівця зі спеціальності 122 «Комп'ютерні науки» за вищою освітою, який дозволить використовувати методи та засоби інтелектуального аналізу даних, а також використання практичних інструментів для виявлення знань із великих баз даних.

Міждисциплінарні зв'язки. Дисципліна «Інтелектуальний аналіз даних» є складовою частиною циклу загальної підготовки для студентів за спеціальністю «Комп'ютерні науки». Вивчення дисципліни передбачає наявність систематичних та ґрунтовних знань із суміжних курсів – «Математичний аналіз», «Алгоритми та структури даних», «Основи програмування», «Теорія ймовірності та математична статистика», «Бази даних і знань», «Системний аналіз та теорія прийняття рішень», «Основи штучного інтелекту», «Системи підтримки прийняття рішень».

Програма навчальної дисципліни складається з таких **змістових модулів та тем:**

Змістовий модуль 1. Концептуальні основи інтелектуального аналізу даних

Тема 1. Предмет дисципліни «Інтелектуальний аналіз даних».

Тема 2. Препроцесінг інформації.

Тема 3. Кореляційний і регресійний аналіз.

Тема 4. Методи класифікації.

Тема 5. Лінійний дискримінантний аналіз. Побудова канонічних та класифікаційних функцій.

Тема 6. Кластерний аналіз. Ієрархічна та секційна кластеризації.

Змістовий модуль 2. Особливості виконання інтелектуального аналізу даних

Тема 7. Пошук асоціативних правил

Тема 8. Статистична обробка часових рядів і прогнозування.

Тема 9. Візуалізація даних.

Тема 10. Аналіз текстової інформації Text Mining.

Тема 11. Добування знань із Web – Web Mining.

Тема 12. Засоби аналізу процесів Process Mining.

1. МЕТА ТА ЗАВДАННЯ НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

1.1. Метою навчальної дисципліни «Інтелектуальний аналіз даних» є теоретична та практична підготовка здобувачів вищої освіти у напрямку розв'язання задач обробки великих масивів інформації, проектування інформаційного забезпечення інформаційних систем та розробки сценаріїв

можливих дій в умовах невизначеності з використанням інтелектуальних методів обчислень.

1.2. Основними завданнями вивчення дисципліни «Інтелектуальний аналіз даних» є забезпечити розуміння і засвоєння здобувачами вищої освіти технологій інтелектуального аналізу даних, підготовка до їх вибору, реалізації і використання при вирішенні прикладних задач, ознайомлення зі станом та перспективами розвитку інтелектуальних методів обчислень як одного з напрямів штучного інтелекту.

1.3. Програмні результати навчання:

- застосовувати знання основних форм і законів абстрактно-логічного мислення, основ методології наукового пізнання, форм і методів вилучення, аналізу, обробки та синтезу інформації в предметній області комп'ютерних наук;

- використовувати знання закономірностей випадкових явищ, їх властивостей та операцій над ними, моделей випадкових процесів та сучасних програмних середовищ для розв'язування задач статистичної обробки даних і побудови прогнозних моделей;

- використовувати методи обчислювального інтелекту, машинного навчання, нейромережевої та нечіткої обробки даних, генетичного та еволюційного програмування для розв'язання задач розпізнавання, прогнозування, класифікації, ідентифікації об'єктів керування тощо;

- застосовувати методи та алгоритми обчислювального інтелекту та інтелектуального аналізу даних в задачах класифікації, прогнозування, кластерного аналізу, пошуку асоціативних правил з використанням програмних інструментів підтримки багатовимірного аналізу даних на основі технологій DataMining, TextMining, WebMining.

На вивчення навчальної дисципліни відводиться 90 годин(и)/ 3,0 кредити ECTS.

2. ІНФОРМАЦІЙНИЙ ОБСЯГ НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

ЗМІСТОВИЙ МОДУЛЬ 1.

КОНЦЕПТУАЛЬНІ ОСНОВИ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ.

Тема 1. Предмет дисципліни «Інтелектуальний аналіз даних»

Загальні поняття інтелектуального аналізу даних. Визначення інтелектуального аналізу даних (Data Mining), історія та передумови виникнення. Data Mining і витяг знань з даних (Knowledge Discovery in Databases). Місце і роль Data Mining в процесі прийняття рішень. Основні завдання Data Mining, види моделей. Data Mining і статистичний аналіз. Data Mining і сховища даних. Data Mining і OLAP. Data Mining як самостійна галузь індустрії програмного забезпечення. Необхідність попередньої обробки даних. Причини, що вимагають очистки даних. Сутність методів попередньої обробки даних.

Тема 2. Препроцесінг інформації

Нормалізація і стандартизація вихідних значень. Аналітико-евристичні алгоритми визначення інформативних ознак. Алгоритм «вибілювання» входів. Нейромережеве визначення значущих факторів. Методика «box-counting».

Тема 3. Кореляційний і регресійний аналіз

Поняття зв'язку між величинами. Кореляційний аналіз. Загальні положення регресійного аналізу. Метод найменших квадратів. Парна лінійна регресія. Множинна лінійна регресія. Тестування і усунення мультиколінеарності. Тестування і усунення гетероскедастичності. Автокореляція: причини і наслідки. Множинна нелінійна регресія. Алгоритм Брандона.

Тема 4. Методи класифікації

Постановка задачі класифікації даних. Правила класифікації. Методи побудови правил класифікації. Точність класифікації: оцінка рівня помилок. Алгоритм 1- rule. Наївний баєсівський класифікатор. Поняття дерева рішень. Методика «розділяй та володарюй». Алгоритм ID3. Алгоритм C4.5. Алгоритм CART. Метод опорних векторів. Метод «найближчого сусіда». Застосування штучних нейронних мереж для класифікації.

Тема 5. Лінійний дискримінантний аналіз. Побудова канонічних та класифікаційних функцій

Основні поняття дискримінантного аналізу. Канонічні дискримінантні функції. Класифікаційні функції. Врахування апіорних ймовірностей. Класифікація за допомогою канонічних дискримінантних функцій. Класифікаційна матриці. Обґрунтування за допомогою розбиття вибірки

Тема 6. Кластерний аналіз. Ієрархічна та секційна кластеризації

Постановка задачі кластеризації. Метрики визначення відстані між об'єктами. Ієрархічні агломеративні методи кластеризації. Ієрархічні дивізимні методи кластеризації. Неієрархічні методи кластеризації: процедура Мак-Кіна, метод k-середніх, сітчасті методи кластеризації, алгоритм fuzzy c-means, алгоритм Гюстафсона-Кеселля.

ЗМІСТОВИЙ МОДУЛЬ 2.

ОСОБЛИВОСТІ ВИКОНАННЯ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ.

Тема 7. Пошук асоціативних правил

Постановка задачі пошуку асоціативних правил. Секвенційний аналіз. Різновиди задач пошуку асоціативних правил. Метод Apriori. Побудова FPдерев пошуку шаблонів даних. Min-max асоціації у базах даних. Побудова hash-дерев.

Тема 8. Статистична обробка часових рядів і прогнозування

Аналіз часових рядів з допомогою Python. Експоненціальне згладжування. Модель Хольта-Вінтерса. Крос-валідація на часових рядах. Лінійні та нелінійні моделі на часових рядах. Вилучення ознак. Алгоритм XGBoost.

Тема 9. Візуалізація даних

Постановка задачі візуалізації даних. Етапи візуального аналізу даних. Методи геометричних перетворень. Методи, орієнтовані на пікселі. Ієрархічні образи. OLAP-аналіз. Засоби візуалізації даних, призначені для оцінки якості моделей: матриця класифікації, діаграма розсіювання, ретропрогноз, графіки контролю процесу навчання.

Тема 10. Аналіз текстової інформації Text Mining

Постановка задачі аналізу текстової інформації. Етапи аналізу текстів. Добування ключових понять із тексту. Класифікація текстових документів. Методи кластеризації текстових документів. Анотування текстів. Засоби аналізу текстової інформації.

Тема 11. Добування знань із Web – Web Mining

Основні поняття Web Mining. Методи добування web-контенту. Добування web-структур. Дослідження використання web-ресурсів.

Тема 12. Засоби аналізу процесів Process Mining

Автоматизація виконання бізнесів-процесів. Технологія аналізу процесів. Методи Process Mining. Бібліотека алгоритмів Process Mining – ProM.

3. РЕКОМЕНДОВАНА ЛІТЕРАТУРА

Базова

1. **Інтелектуальний аналіз даних: практикум** / Фісун М.Т., Кравець І.О., Казмірчук П.П., Ніколенко С.Г. – Л.: «Новий Світ – 2000», 2020. – 162 с.
2. **Методичні вказівки до виконання лабораторних робіт з дисципліни «Інтелектуальний аналіз даних»** для студентів спеціальності 122 «Комп'ютерні науки» / А.М. Тригуба. – Львів: ЛНАУ, 2020. – 54 с.
3. **Інтелектуальні системи: підручник** / Литвин В.В., Пасічник В.В., Яцишин Ю.В.. Львів: Новий світ – 2000, 2009. – 406с.
4. **Data Mining: пошук знань в даних** / Гладун А. Я., Рагушина Ю. В. – К.: ТОВ «ВД «АДЕФ Україна», 2016. – 452 с.
5. **Introduction to Data Mining** / P. Tan, M. Steinbach, V. Kumar. – Boston: Addison Wesley, 2006. – 769 p.
6. **Аналіз даних у R (II)**. / А. Б. Шипунов, А. И. Коробейников, Е. М. Балдин. – 2016. – 60. Режим доступу: <http://www.inp.nsk.ru/~baldin/DataAnalysis/R/R-06-gui.pdf>

Допоміжна

1. **Hulida, E., Pasnak, I., Koval, O., Tryhuba, A.** (2019), “Determination of the Critical Time of Fire in the Building and Ensure Successful Evacuation of People”, *Periodica Polytechnica Civil Engineering*, No. 63(1), pp. 308–316.
2. **Sokulskyi O.** The Internet of Things Solutions in the Investigation of Urban Passenger Traffic and Passenger Service Quality / Sokulskyi O., Hilevska K., Chumakevych V., Ptashnyk V., Tryhuba A., Sachenko A. // 2020 IEEE European Technology and Engineering Management Summit (E-TEMS), 2020. pp. 1-5.
3. **Аналіз даних та знань : навчальний посібник** / Литвин В. В., Пасічник В. В., Нікольський Ю. В. . Львів: «Магнолія 2006», 2015. – 276 с.
4. **Practical Machine Learning with Python** / Sarkar D., Bali R., Sharma T.. Apress, 2018. 545 p. URL: www.apress.com/978-1-4842-3206-4.
5. **Data mining and analysis : Fundamental Concepts and Algorithms** / Zaki M. J., Meira W. Jr.. New York : Cambridge University Press, 2014. – 604 p.
6. **Data Mining for the Masses**. Matthew North / North M. – Global Text Project, 2012. – 264 p. – Режим доступу: <http://dl.dropboxusercontent.com/u/31779972/DataMiningForTheMasses.pdf>
7. **Data Mining: Concepts and Techniques**: 3rd ed. / Jiawei Han, Micheline Kamber, Jian Pei. – NY: Elsevier, 2012. – 740 p.
8. **Методи аналізу даних : навчальний посібник** / Бахрушин В. Є. Запоріжжя: КПУ, 2011. – 268 с.
9. **Прогнозування. Моделі. Методи. Алгоритми : навчальний посібник** / Снитюк В. Є. – К.: Маклаут, 2008. – 364 с.
10. **Наочна статистика. використовуємо R!**/ А. Б Шипунов. – 2016. – 296. Режим доступу: <https://cran.r-project.org/doc/contrib/Shipunov-rbook.pdf>

11. **Бизнес-аналитика : от данных к знаниям** / Паклин Н. Б., Орешков В. И.. СПб.: Питер, 2009. – 624 с.

12. **Нечеткие модели и методы в интеллектуальных системах : учебное пособие** / Зайченко Ю. П. – К.: Слово, 2008. 344с.

13. **Математичне забезпечення інформаційно-керуючих систем: підручник** / А. М. Гуржій, З. В. Дудар, В. М. Левикін, Б. В. Шамша. Х. : Компанія Сміт, 2006. – 448 с.

Інформаційні ресурси

1. Prometheus. Курс «**Аналіз даних та статистичне виведення на мові R**». [Електронний ресурс]. – Доступний з https://courses.prometheus.org.ua/courses/IRF/Stat101/2016_T3/about

2. Prometheus. Курс «**Машинне навчання**». [Електронний ресурс]. – Доступний з https://courses.prometheus.org.ua/courses/IRF/ML101/2016_T3/about

3. Prometheus. Курс «**Візуалізація даних**». [Електронний ресурс]. – Доступний з https://courses.prometheus.org.ua/courses/IRF/DV101/2016_T3/about

4. **Wiki-портал** [Електронний ресурс]. – Доступний з <http://www.machinelearning.ru>

5. Журнал «**Нечіткі системи і м'які обчислення**». [Електронний ресурс]. – Доступний з <http://fuzzy.tversu.ru/>.

6. **Evolutionary Complexity Research Group (EPlEx)**: [Електронний ресурс]. – Доступний з <http://eplex.cs.ucf.edu/>

4. КРИТЕРІЇ УСПІШНОСТІ НАВЧАННЯ ТА ФОРМА ПІДСУМКОВОГО КОНТРОЛЮ

При оцінюванні результатів навчання здобувачів освіти потрібно керуватися такими критеріями успішності навчання:

Бали	Оцінка	Критерії оцінювання
91–100	Відмінно	Здобувач демонструє повні й вичерпні знання навчального матеріалу в обсязі, що відповідає робочій програмі дисципліни, правильно й обґрунтовано приймає необхідні рішення в різних нестандартних ситуаціях. Вміє реалізувати теоретичні положення дисципліни при розв'язуванні практичних завдань, може аналізувати і співставляти навчальний матеріал з даної та суміжних дисциплін. Знає сучасні технології та методи рішення прикладних завдань з дисципліни. За час навчання при проведенні практичних занять, виконанні індивідуальних завдань проявив вміння самостійно вирішувати поставлені завдання, активно включатись в дискусії, може відстоювати власну позицію в питаннях та рішеннях, що розглядаються.

Бали	Оцінка	Критерії оцінювання
		Зменшення 100-бальної оцінки може бути пов'язане з недостатнім розкриттям питань, що стосується дисципліни яка вивчається, але виходить за рамки об'єму матеріалу передбаченого робочою програмою, або здобувач проявляє невпевненість в тлумаченні теоретичних положень чи рішенні складних практичних завдань.
81–90	Добре	Здобувач демонструє добрі та вичерпні знання, володіє матеріалом, що відповідає робочій програмі дисципліни, робить на основі здобутих знань аналіз можливих ситуацій та вміє застосовувати теоретичні положення при рішенні практичних завдань, проте допускає окремі неточності. Вміє самостійно виправляти допущені помилки, кількість яких є незначною. Знає сучасні технології та методи рішення практичних завдань з дисципліни. За час навчання при проведенні практичних занять, виконанні індивідуальних завдань та поясненні прийнятих рішень, дає вичерпні пояснення.
71–80	Добре	Здобувач в загальному добре володіє матеріалом, знає основні положення матеріалу, що відповідають робочій програмі дисципліни, робить на їх основі аналіз можливих ситуацій та використовує для рішення характерних/типових прикладних завдань з дисципліни. Вміє пояснити основні положення виконаних завдань та давати правильні відповіді про зміну результату при заданій зміні вихідних параметрів. Помилки у відповідях / рішеннях / розрахунках не є системними. Розуміє основні положення, що мають визначальне значення для практичних занять, виконанні індивідуальних завдань в межах дисципліни.
61–70	Задовільно	Здобувач засвоїв основний теоретичний матеріал, передбачений робочою програмою дисципліни, та розуміє постановку стандартних практичних завдань, має пропозиції щодо напрямку їх вирішень. Розуміє основні положення, що є визначальними в курсі, може вирішувати завдання подібні тим, що розглядалися на заняттях, проте допускає значну кількість неточностей і помилок, усунути які здатен лише за допомогою викладача.
51–60	Задовільно	Здобувач володіє певними знаннями та основними положеннями, передбаченими робочою програмою дисципліни, на мінімально допустимому рівні для подальшого засвоєння результатів навчання в рамках освітньої програми. З використанням основних теоретичних положень здобувач з труднощами пояснює правила вирішення практичних завдань дисципліни. Виконання практичних / контрольних / індивідуальних завдань, значно формалізовано: є відповідність алгоритму, проте відсутнє глибоке розуміння самої роботи.
35–50	Незадовільно	Здобувач може відтворити окремі фрагменти знань з курсу. Незважаючи на те, що програму навчальної дисципліни

Бали	Оцінка	Критерії оцінювання
		здобувач виконав, працював він пасивно, його відповіді під час практичних робіт та результати поточного контролю в більшості є невірними та/або необґрунтованими. Цілісність розуміння матеріалу з дисципліни у здобувача відсутні, що створює перепони для подальшого засвоєння результатів навчання в рамках освітньої програми.
0–34	Незадовільно	Здобувач повністю не виконав вимог робочої програми навчальної дисципліни. Його відповіді під час практичних робіт та результати поточного контролю є невірними та/або необґрунтованими. Його знання на підсумкових етапах навчання є фрагментарними.

Формою підсумкового контролю екзамен.

5. ЗАСОБИ ДІАГНОСТИКИ УСПІШНОСТІ НАВЧАННЯ

Під час вивчення дисципліни передбачено індивідуальний поточний контроль, фронтальний контроль, контроль за виконанням самостійної роботи, самоконтроль та індивідуальний підсумковий контроль у формі екзамену.

Поточний контроль здійснюється у формі виконання тестових завдань на базі платформи віртуального навчального середовища та письмового опитування.

Самоконтроль організовано шляхом надання здобувачам освіти другої спроби для складання тестових завдань (можливість надолуження пройденого матеріалу та перевірки рівня його засвоєння).

Фронтальний контроль передбачає проведення наскрізного тестування або усного опитування під час лекційних занять з метою визначення якості засвоєння нового матеріалу.

Під час практичних занять або/та консультацій викладач здійснює контроль за самостійною роботою здобувачів освіти шляхом прийняття індивідуальних практичних завдань.

Індивідуальний підсумковий контроль здійснюється у формі екзамену.

Усі форми контролю включено до 100-бальної шкали оцінювання. Оцінка із 100-бальної шкали в національну переводиться відповідно до діючого положення про освітній процес (91–100 – «відмінно», 71–90 – «добре», 51–70 – «задовільно», менше 51 – «незадовільно»).