

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ
ІМЕНІ СЕМЕНА КУЗНЕЦЯ

"ЗАТВЕРДЖУЮ"
Заступник керівника
(проректор з науково-педагогічної роботи)



М. В. Афанасьєв

"ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ"
робоча програма навчальної дисципліни

Галузьзнань 12 "ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ"
Спеціальність 125 "КІБЕРБЕЗПЕКА"
Освітнійрівень другий (магістерський)
Освітняпрограма "КІБЕРБЕЗПЕКА"

Виддисципліни вибіркова
Мова викладання, навчання та оцінювання українська

Завідувач кафедри кібербезпеки
та інформаційних технологій

Євсєєв С.П.

Харків
ХНЕУ ім. С. Кузнеця
2019

ЗАТВЕРДЖЕНО
на засіданні кафедри кібербезпеки
та інформаційних технологій
Протокол № 1 від 26.08.2019 р.

Розробник(-и):
Мілов О.В., к.т.н., доцент кафедри КІТ

**Лист оновлення та перезатвердження
робочої програми навчальної дисципліни**

Навчальний рік	Дата засідання кафедри – розробника РПНД	Номер протоколу	Підпис завідувача кафедри

1. Вступ

Анотація навчальної дисципліни:

Подано тематичний план навчальної дисципліни й її змістовність за модулями та темами, вміщено плани лекцій і лабораторних занять, матеріал щодо закріплення знань (завдання для самостійної роботи, контрольні запитання), методичні рекомендації та оцінювання знань студентів.

Дисципліна “Інтелектуальний аналіз даних” знайомить з технологією DataMining, докладно розглядаються методи, інструментальні засоби й застосування DataMining. Опис кожного методу супроводжується конкретним прикладом його використання. Обговорюються відмінності DataMining від класичних статистичних методів аналізу й OLAP-систем, розглядаються типи закономірностей, що виявляють DataMining (асоціація, класифікація, послідовність, кластеризація, прогнозування). Описується сфера застосування DataMining. Вводиться поняття WebMining. Докладно розглядаються методи DataMining: нейронні мережі, дерева рішень, методи обмеженого перебору, генетичні алгоритми, еволюційне програмування, кластерні моделі, комбіновані методи. Знайомство з кожним методом проілюстровано рішенням практичного завдання за допомогою інструментального засобу, що використовує технологію DataMining. Викладаються основні концепції сховищ даних і місця DataMining у їхній архітектурі. Вводяться поняття OLTP, OLAP, ROLAP, MOLAP. Обговорюється процес аналізу даних за допомогою технології DataMining. Докладно розглядаються етапи цього процесу. Аналізується ринок аналітичного програмного забезпечення, описуються продукти від провідних виробників DataMining, обговорюються їхні можливості.

Мета навчальної дисципліни:

Метою викладання дисципліни “Інтелектуальний аналіз даних” є навчання студентів принципам інтелектуального аналізу систем та процесів захисту інформації для формування контуру безпеки бізнес-процесів в комп'ютерних системах на основі технологій data-mining та запобіганню зовнішніх кібер-загроз.

Курс	1	
Семестр	1	
Кількість кредитів ECTS	5	
Аудиторні навчальні заняття	лекції	20
	семінарські, практичні	–
	лабораторні	20
Самостійна робота		110
Форма підсумкового контролю	екзамен	

Структурно-логічна схема вивчення навчальної дисципліни:

Попередні дисципліни	Наступні дисципліни
Вища математика	Основи криптографічного захисту
Основитеоріїбаз даних	Основи технічного захисту інформації
Технологіїобробкиінформації	Забезпечення інформаційної безпеки

2. Компетентності та результати навчання задисципліною:

Компетентності	Результати навчання
Здатність до використання інформаційних і комунікаційних технологій з метою пошуку нової інформації, створення баз даних, аналізу розподілених АС, каналів зв'язку, систем управління процесами, баз даних, оперативного планування роботи систем на основі аналізу інформаційних потоків та їх оптимізації.	виконувати комплексний аналіз захищеності систем безпеки об'єктів критичної інфраструктури
Здатність до використання технологій datamining з метою пошуку нової інформації, аналізу розподілених АС, виявлення можливих загроз, аналіз стану захищеності систем управління процесами, баз даних, оперативного планування роботи систем на основі аналізу інформаційних потоків та їх оптимізації.	вирішувати завдання захисту програм та даних ІТС програмно-апаратними засобами та давати оцінку якості прийнятих рішень; вибирати основні методи та способи захисту інформації відповідно до вимог сучасних стандартів інформаційної безпеки щодо критеріїв безпеки інформаційних технологій, застосовуючи системний підхід та знання основ теорії інформаційної безпеки
Здатність прогнозувати, виявляти та оцінювати стан інформаційної безпеки об'єктів і систем	аналізувати комплексні систему захисту інформації АС організації (підприємства) відповідно до вимог нормативних документів системи технічного захисту інформації
Здатність прогнозувати можливі кібернапади на систему та класифікувати можливі загрози с метою їх запобігання та вибору для цього відповідних засобів протидії	здійснювати оцінку можливості проникнення в ІТ системи та мережі шляхом експлуатації наявних вразливостей

3. Програма навчальної дисципліни

Тема 1. Інтелектуальний аналіз даних – DataMining

Докладно розглядаються поняття DataMining. Описано виникнення, перспективи, проблеми Datamining. Наведен погляд на технологію DataMining як на частину ринку інформаційних технологій. Докладно розглядається поняття даних. Пояснюється значення понять об'єкт та атрибут, вибірка, залежна й незалежна змінна. Докладно обговорюються типи шкал. Приводяться різні типи наборів даних. Коротко розглянуті поняття бази даних і СУБД. Описуються стадії DataMining та дії, які виконуються в рамках цих стадій. Розглянуто відомі класифікації методів

DataMining. Наведено порівняльну характеристику деяких методів, заснована на їхніх властивостях. Характеризується основна суть *завдань DataMining* й їхня класифікація. Докладно розглянуті поняття "інформація", "знання", а також зіставлення й порівняння цих понять.

Тема 2. Візуальний аналіз даних — VisualMining

Розглядаються питання візуального аналізу даних. Наведені характеристики засобів візуалізації даних, методів візуалізації та методів геометричних перетворень. Порівнюються методи, орієнтовані на пікселі, а також методи аналізу ієрархічних образів та відображення іконок. Розглядаються методи й засоби візуального подання інформації, зокрема, способи подання інформації в одне-, дво-, тривимірному вимірах, а також способи відображення інформації в більш ніж трьох вимірах. Описано принципи якісної візуалізації. Викладено основні тенденції в області візуалізації.

Тема 3. Аналіз текстової інформації — TextMining

Формулюються задачі аналізу текстів (етапи аналізу текстів, попередня обробка тексту, завачі TextMining). Розглядаються етапи аналізу текстів, такі як витяг ключових понять із тексту (загальний опис процесу витягу понять із тексту, стадія локального аналізу, стадія інтеграції й висновку понять), класифікація текстових документів (опис задач класифікації текстів, методи класифікації текстових документів), методи кластеризації текстових документів (наведення текстових документів, ієрархічні методи кластеризації текстів, бінарні методи кластеризації текстів), анотування текстів (виконання анотування текстів, методи витягу фрагментів для анотації). Порівнюються різноманітні засоби аналізу текстової інформації (засоби Oracle - OracleText, засоби від IBM - IntelligentMinerforText, засоби SAS Institute - TextMiner, засоби Мега-комп'ютерИнтеллидженс - TextAnalyst).

Тема 4. Витяг знань з Web - WebMining

Розглянуті проблеми аналізу інформації з Web, етапи WebMining, WebMining та інші інтернет-технології, а також категорії WebMining. Описані методи витягу Web-контента (витяг Web-контента в процесі інформаційного пошуку, витяг Web-контента для формування баз даних), а також методи витягу Web-структур (представлення Web-структур, оцінка важливості Web-структур, пошук Web-документів з урахуванням гіперпосилань, кластеризація Web-структур). Наведені результати досліджень використання Web-ресурсів (дослідницька інформація, етап препроцесінгу, етап витягу шаблонів, етап аналізу шаблонів та їхнє застосування).

Тема 5. Засоби аналізу процесів - ProcessMining

Розглянуті засоби автоматизації виконання бізнес-процесів (бізнес-процеси, формалізація бізнес-процесів, Workflow-системи, сервісно-орієнтована архітектура, проектування бізнес-процесів). Виконан аналіз процесів (технологія ProcessMining, аналіз протоколів, стандарт запису протоколів MXML, задачі ProcessMining, проблеми аналізу протоколів). Порівнюються методи ProcessMining (перші імовірнісні методи ProcessMining, метод побудови диз'юнктивної Workflow-схеми, (α -алгоритм, методи на основі генетичних алгоритмів). Описана бібліотека алгоритмів Process Mining-Pro (архітектура Pro, ProImport Framework).

Тема 6. Пошук асоціативних правил - RulesMining

Виконана постановка задачі. Розглянуті форми подання результатів (правила класифікації, дерева класифікації, математичні функції), методи побудови правил класифікації (алгоритм побудови 1-правил, метод NaiveBayes), а також методи побудови дерев класифікації (методика "розділяй і пануй", алгоритм покриття),

методи побудови математичних функцій (загальний вид, лінійні методи, метод найменших квадратів, нелінійні методи, SupportVectorMachines (SVM), регуляризаційні мережі (RegularizationNetworks), дискретизації й рідкі сітки). Розглянута постановка задачі пошуку асоціативних правил (формальна постановка задачі, секвенціальний аналіз, різновиди задач пошуку асоціативних правил), алгоритми (алгоритм Apriori, різновид алгоритму Apriori).

Тема 7. Розподілений аналіз даних

Розглядаються системи мобільних агентів (основні поняття, стандарти багатоагентних систем, системи мобільних агентів, система мобільних агентів JADE). Продемонстровано використання мобільних агентів для аналізу даних (проблеми розподіленого аналізу даних, агенти-аналітики, варіанти аналізу розподілених даних). Побудована система аналізу розподілених даних (загальний підхід до реалізації системи, агент для збору інформації про базу даних, агент для збору статистичної інформації, агент для вирішення одного завдання інтелектуального аналізу даних, агент для вирішення інтегрованого завдання інтелектуального аналізу даних).

Теми лабораторних робіт

Лабораторна робота 1. Навички роботи з програмою Deductor.

Лабораторна робота 2. Дерева класифікації.

Лабораторна робота 3. Візуальний аналіз даних — VisualMining

Лабораторна робота 4. Витяг знань з Web – WebMining

Лабораторна робота 5. Засоби аналізу процесів – ProcessMining. Використання ProM.

Лабораторна робота 6. Пошук асоціативних правил – RulesMining. Програма See 5.

Лабораторна робота 7. Розподілений аналіз даних. Агентне моделювання.

4. Порядок оцінювання результатів навчання

Система оцінювання сформованих компетентностей у студентів враховує види занять, які згідно з програмою навчальної дисципліни передбачають лекційні та лабораторні заняття, а також виконання самостійної роботи. Оцінювання сформованих компетентностей у студентів здійснюється за накопичувальною 100-бальною системою. Відповідно до Тимчасового положення “Про порядок оцінювання результатів навчання студентів за накопичувальною бально-рейтинговою системою” ХНЕУ ім. С. Кузнеця, контрольні заходи включають:

поточний контроль, що здійснюється протягом семестру під час проведення лекційних та лабораторних занять і оцінюється сумою набраних балів (максимальна сума – 60 балів; мінімальна сума, що дозволяє студенту скласти іспит, – 35 балів);

модульний контроль, що проводиться у формі колоквиуму як проміжний міні-екзамен з ініціативи викладача з урахуванням поточного контролю за відповідний змістовий модуль і має на меті *інтегровану* оцінку результатів навчання студента після вивчення матеріалу з логічно завершеної частини дисципліни – змістового модуля;

підсумковий/семестровий контроль, що проводиться у формі семестрового екзамену, відповідно до графіку навчального процесу.

Порядок проведення поточного оцінювання знань студентів. Оцінювання знань студента під час семінарських, практичних і лабораторних занять та виконання індивідуальних завдань проводиться за такими критеріями:

розуміння, ступінь засвоєння теорії та методології проблем, що розглядаються; ступінь засвоєння фактичного матеріалу навчальної дисципліни; ознайомлення з рекомендованою літературою, а також із сучасною літературою з питань, що розглядаються; вміння поєднувати теорію з практикою при розгляді виробничих ситуацій, розв'язанні задач, проведенні розрахунків у процесі виконання індивідуальних завдань та завдань, винесених на розгляд в аудиторії; логіка, структура, стиль викладу матеріалу в письмових роботах і при виступах в аудиторії, вміння обґрунтовувати свою позицію, здійснювати узагальнення інформації та робити висновки; арифметична правильність виконання індивідуального та комплексного розрахункового завдання; здатність проводити критичну та незалежну оцінку певних проблемних питань; вміння пояснювати альтернативні погляди та наявність власної точки зору, позиції на певне проблемне питання; застосування аналітичних підходів; якість і чіткість викладення міркувань; логіка, структуризація та обґрунтованість висновків щодо конкретної проблеми; самостійність виконання роботи; грамотність подачі матеріалу; використання методів порівняння, узагальнення понять та явищ; оформлення роботи.

Загальними критеріями, за якими здійснюється оцінювання позааудиторної самостійної роботи студентів, є: глибина і міцність знань, рівень мислення, вміння систематизувати знання за окремими темами, вміння робити обґрунтовані висновки, володіння категорійним апаратом, навички і прийоми виконання практичних завдань, вміння знаходити необхідну інформацію, здійснювати її систематизацію та обробку, самореалізація на практичних та семінарських заняттях.

Підсумковий контроль знань та компетентностей студентів з навчальної дисципліни здійснюється на підставі проведення семестрового екзамену, завданням якого є перевірка розуміння студентом програмного матеріалу в цілому, логіки та взаємозв'язків між окремими розділами, здатності творчого використання накопичених знань, вміння формулювати своє ставлення до певної проблеми навчальної дисципліни тощо.

Екзаменаційний білет охоплює програму дисципліни і передбачає визначення рівня знань та ступеня опанування студентами компетентностей.

Кожен екзаменаційний білет складається із 2 теоретичних та одного практичного завдання, які передбачають вирішення типових професійних завдань фахівця на робочому місці та дозволяють діагностувати рівень теоретичної підготовки студента і рівень його компетентності з навчальної дисципліни.

Результат семестрового екзамену оцінюється в балах (максимальна кількість – 40 балів, мінімальна кількість, що зараховується, – 25 балів) і проставляється у відповідній графі екзаменаційної "Відомості обліку успішності".

Студента слід **вважати атестованим**, якщо сума балів, одержаних за результатами підсумкової/семестрової перевірки успішності, дорівнює або перевищує 60. Мінімумально можлива кількість балів за поточний і модульний контроль упродовж семестру – 35 та мінімумально можлива кількість балів, набраних на екзамені, – 25.

Підсумкова оцінка з навчальної дисципліни розраховується з урахуванням балів, отриманих під час екзамену, та балів, отриманих під час поточного контролю за накопичувальною системою. Сумарний результат у балах за семестр складає: "60 і більше балів – зараховано", "59 і менше балів – не зараховано" та заноситься у залікову "Відомість обліку успішності" навчальної дисципліни.

Розподіл балів за тижнями
(вказати засоби оцінювання згідно з технологічною картою)

Теми змістового модуля		Лекційні заняття	Захист лабораторних робіт	Завдання за темами	Поточні КР	Усього	
	Тема 1	1 тиждень	1	-	-	1	
		2 тиждень			-		
	Тема 1	3 тиждень	1	4	1	6	
		4 тиждень			-		
	Тема 2	5 тиждень	1	4	1	6	
		6 тиждень			-		
	Тема 3	7 тиждень	1	4	1	6	
		8 тиждень			5	5	
	Тема 4	9 тиждень	1	4	1	6	
		10 тиждень			-		
	Тема 5	11 тиждень	1	4	1	6	
		12 тиждень			-		
	Тема 5	13 тиждень	1	4	1	6	
		14 тиждень			-		
	Тема 6	15 тиждень	1	4	1	6	
		16 тиждень	1		-	1	
	Тема 7	17 тиждень	1	4	1	5	11
	Екзамен	18 тиждень					40
Усього			10	32	8	10	100

Шкала оцінювання: національна та ЄКТС

Сума балів за всі види навчальної діяльності	Оцінка ЄКТС	Оцінка за національною шкалою	
		для екзамену, курсового проекту (роботи), практики	для заліку
90 – 100	A	відмінно	зараховано
82 – 89	B	добре	
74 – 81	C		
64 – 73	D	задовільно	
60 – 63	E		
35 – 59	FX	незадовільно	не зараховано
1 – 34	F		

5. Рекомендована література

5.1 Основна

1. В. Дюк, А. Самойленко. «DataMining»: Учебный курс, - СПб: - Питер, 2001,- 368 с., ил.
2. Барсемян А. А. и др. Технологии анализа данных: DataMining, VisualMining, TextMining, OLAP. 2-е изд., перераб. и доп. — СПб.: БХВ-Петербург, 2007.

3. Рассел С., Норвинг П. Искусственный интеллект: современный подход. – М.: Издательский дом «Вильямс», 2006. – 1408 с.
4. В.В. Корнеев, А.Ф. Гареев, С.В. Васютин, В.В. Райх. Базы данных. Интеллектуальная обработка информации. – М. Издатель Могачева С.В. , Издательство Нолидж, 2001. – 496 с., ил.
5. Задания и методические рекомендации к выполнению лабораторных работ по теме «Нейронные сети» курса «Прогнозирование социально-экономических процессов» для студентов специальности 8.050102 «Экономическая кибернетика» дневной форм обучения / Сост. Н.А. Дубровина, А.В. Милов, С.В. Прокопович. – Харьков. Изд. ХНЭУ, 2005. – 60 с.

5.2. Додаткова

6. В. А. Дюк. Обработка данных на ПК в примерах. -СПб: Питер, 1997,- 240 с.
7. Н.Г. Загоруйко. Прикладные методы анализа данных и знаний. - Новосибирск: Изд-во Ин-та математики СО РАН, 1999. -270 с.
8. Вагин В. Н., Головина Е. Ю., Загорянская А. А. Достоверный и правдоподобный вывод в интеллектуальных системах. — М.: Физматлит, 2004.
9. *Зиновьев А. Ю.* Визуализация многомерных данных. — Красноярск: Изд-во КГТУ, 2000.
10. Хайкин С. Нейронные сети: полный курс. 2-е изд. / Пер. с англ. — М.: Вильямс, 2006.
11. Тюрин Ю.Н., Макаров А.А. Статистический анализ данных на компьютере / Под ред. В.Э. Фигурнова. - М.: ИНФРА-М. 1998.-528 с, ил.
12. Лбов Г.С , Анализ данных и знаний: Учеб. пособие. – Новосибирск: Изд-во НГТУ, 2001. – 90 с.
13. Романов А.Н., Одинцов Б.Е. Советующие информационные системы в экономике: Учеб. пособие для вузов. – М.: ЮНИТИ-ДАНА, 2000. – 488 с.
14. Круглое В. В., Борисов В. В. Нейронные сети: теория и практика. — М.: Горячая линия Телеком, 2001.
15. Ханк Д. Э., Уичерн Д. У., Райте А. Дж. Бизнес-прогнозирование. 7-е изд. / Пер. с англ. — М.: Вильямс, 2003.
16. Каллан Р. Основные концепции нейронных сетей.: Пер. с англ. – М.: Издательский дом «Вильямс», 2001. – 292 с.

5.3. Інформаційні ресурси в мережі Інтернет

17. <http://all-ebooks.org>
18. <http://financepro.ru>
19. <http://www.classs.ru>