

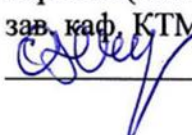
**ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ АВТОМОБІЛЬНО-ДОРОЖНИЙ
УНІВЕРСИТЕТ**

Система забезпечення якості освітньої діяльності та вищої освіти

Кафедра комп'ютерних технологій і мехатроніки

ЗАТВЕРДЖЕНО

Гарант освітньо-професійної програми
«Інженерія програмного забезпечення»
першого(бакалаврського) рівня освіти:

зав. каф. КТМ, д.т.н., проф.
 Ніконов О.Я.

**СИЛАБУС
ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ /
DATA MINING**

освітній ступінь	бакалавр / bachelor
галузь знань	12 Інформаційні технології / Information Technology
спеціальність	121 Інженерія програмного забезпечення / Software Engineering
освітня програма	Програмне забезпечення систем / Systems Software

Харків 2020

Автор: Шуляков Владислав Миколайович, асистент кафедри комп'ютерних технологій і мехатроніки.

Силабус розглянуто та затверджено на засіданні кафедри комп'ютерних технологій і мехатроніки, протокол № 20 від «28» серпня 2020 р.

СИЛАБУС

ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ /

DATA MINING

SYLLABUS

освітній ступінь	бакалавр / bachelor
галузь знань	12 Інформаційні технології / Information Technology
спеціальність	121 Інженерія програмного забезпечення / Software Engineering
освітня програма	Програмне забезпечення систем / Systems Software

Анотація курсу

1. Викладачі

1.1. Лектор: Шуляков Владислав Миколайович

- асистент кафедри комп'ютерних технологій і мехатроніки;
- педагогічний стаж – 13 років;
- контактний телефон +38-093-18-57-595;
- e-mail: jasonvlad2013@gmail.com;
- наукові інтереси: інтелектуальний аналіз даних, програмування Java, штучні нейронні мережі та нечітка логіка, створення Web сайтів, SEO, HTML/CSS/JS.

1.2. Асистент лектора:

2. Дисципліна «Інтелектуальний аналіз даних»

- рік навчання: 3;
- семестр навчання: 6;
- кількість годин за семестр: 150, в т. ч.
 - лекційних: 16;
 - практичних занять: 32;
 - на самостійне опрацювання: 102;
- кількість аудиторних годин на тиждень
 - лекційних: 2 (раз на два тижні);
 - практичних занять: 4 (раз на два тижні).

3. Час та місце проведення

- аудиторні заняття – відповідно до розкладу ХНАДУ, ауд. 214, 216;
- позааудиторна робота – самостійна робота студента із використанням засобу аналізу даних Weka.

4. Мета, завдання та результати вивчення дисципліни (Компетентності), її місце в освітньому процесі

4.1. Метою вивчення дисципліни «Інтелектуальний аналіз даних» є теоретична та практична підготовка студентів до вивчення систем обробки даних та принципів інтелектуального аналізу даних на основі методів та алгоритмів Data Mining.

4.2. Завданнями дисципліни «Інтелектуальний аналіз даних» є теоретична та практична підготовка майбутніх фахівців з наступних питань:

- формування у студентів уяви про застосування технологій зберігання та організації даних;
- методи та алгоритми Data Mining;
- процеси виявлення знань;
- принципи побудови сховищ даних;
- способи візуального представлення даних.

4.3. Предметом вивчення дисципліни є створення та дослідження математичних та програмних моделей інтелектуального аналізу даних, пов'язаних із функціонуванням об'єктів професійної діяльності.

4.4. Результати вивчення дисципліни «Інтелектуальний аналіз даних».

В результаті вивчення дисципліни студенти повинні:

- Знати кодекс професійної етики, розуміти соціальну значимість та культурні аспекти інженерії програмного забезпечення і дотримуватись їх в професійній діяльності;
- Знати основні процеси, фази та ітерації життєвого циклу програмного забезпечення;
- Знати і застосовувати відповідні математичні поняття, методи доменного, системного і об'єктно-орієнтованого аналізів та математичного моделювання для розробки програмного забезпечення;
- Знати і застосовувати на практиці фундаментальні концепції, парадигми і основні принципи функціонування мовних, інструментальних і обчислювальних засобів інженерії програмного забезпечення;
- Знати та вміти розробляти та реалізовувати сучасні інноваційні інформаційні технології проектування в області інтелектуальних транспортних систем та мехатронних систем і комплексів.

Дисципліна «Інтелектуальний аналіз даних» базується на знаннях, отриманих студентами під час вивчення таких навчальних дисциплін:

- «Вища математика»;
- «Дискретна математика»;

- «Теорія ймовірностей та випадкові процеси»;
- «Теорія алгоритмів»;
- «Чисельні методи»;
- «Математичні методи дослідження операцій»;
- «Організація баз даних та знань».

Викладання дисципліни розраховано на виконання студентами лабораторних робіт та самостійних завдань у програмному середовищі Text Miner, Deductor Studio Academic та Weka.

5. План вивчення дисципліни

Результати навчання	Навчальна діяльність	Робочий час студента	Оцінювання в балах
Тема 1. Основи інтелектуального аналізу даних.			
Загальні та спеціальні компетентності: Здатність брати участь у проектуванні програмного забезпечення, включаючи проведення моделювання (формальний опис) його структури, поведінки та процесів функціонування. Результати навчання: Знати кодекс професійної етики, розуміти соціальну значимість та культурні аспекти інженерії програмного забезпечення і дотримуватись їх в професійній діяльності.	Лекція №1 <i>План лекції:</i> 1. Вступ. Мета та завдання дисципліни, її місце в освітньому процесі. 2. Поняття інтелектуального аналізу даних. 3. Методи, стадії та завдання інтелектуального аналізу даних. 4. Data Mining як частина ринку інформаційних технологій. Список рекомендованих джерел: Основний: 1-4. Додатковий: 6,8 Інтернет-ресурси: 9	2	
	Задання для самостійної роботи: Вивчення та доповнення матеріалів лекції на основі самостійного опрацювання основних літературних джерел, зазначених у списку та електронних матеріалів, які надаються викладачем. Питання винесені на самостійне опрацювання: 1. Розвиток технології баз даних. 2. Порівняння статистики, машинного навчання та Data Mining.	12	4
	Практичне заняття 1 <i>План заняття:</i> 1. Первинний розвідувальний аналіз даних. 2. Освоєння методів статистичної обробки даних.	4	8

Тема 2. Регресійний аналіз та нелінійні моделі			
Загальні та спеціальні компетентності: Володіння знаннями про інформаційні моделі даних, здатність створювати програмне забезпечення для зберігання, видобування та опрацювання даних. Результати навчання: Знати основні процеси, фази та ітерації життєвого циклу програмного забезпечення.	Лекція №2 <i>План лекції:</i> 1. Методи первинної обробки даних. 2. Методи дослідження структури даних: візуалізація та автоматичне групування даних. 3. Описова статистика. 4. Кореляційний аналіз. 5. Регресійний аналіз. Список рекомендованих джерел: Основний: 1,2,3 Додатковий: 6,7,8 Інтернет-ресурси: 9	2	
	Задання для самостійної роботи: Вивчення та доповнення матеріалів лекції на основі самостійного опрацювання основних літературних джерел, зазначених у списку та електронних матеріалів, які надаються викладачем. Питання винесені на самостійне опрацювання: 1. Множинний регресійний аналіз. 2. Нелінійна множинна регресійна модель. 3. Нелінійне оцінювання параметрів.	12	4
	Практичне заняття 2 План заняття: 1. Кластерний аналіз даних. 2. Освоєння методу побудови моделей кластеризації. 3. Виявлення в даних неочевидних закономірностей.	4	8
Тема 3. Методи багатомірного розвідувального аналізу			
Загальні та спеціальні компетентності: Здатність розв'язувати складні спеціалізовані завдання або практичні проблеми інженерії програмного забезпечення, що характеризуються комплексністю та невизначеністю умов, із застосуванням теорій та методів інформаційних технологій.	Лекція №3 <i>План лекції:</i> 1. Методи кластерного аналізу. 2. Заходи подібності. 3. Методи об'єднання або зв'язку. 4. Ієрархічний кластерний аналіз в SPSS. 5. Ітеративні методи. 6. Алгоритм k-середніх (k-means). 7. Лінійний дискримінантний аналіз. Список рекомендованих джерел: Основний: 1,2,4,5 Додатковий: 6,7 Інтернет-ресурси: 9	2	
	Задання для самостійної роботи: Вивчення та доповнення матеріалів лекції на основі самостійного опрацювання основних літературних джерел, зазначених у списку та електронних матеріалів, які надаються викладачем. Питання винесені на самостійне опрацювання:	12	4

Результати навчання: Знати і застосовувати відповідні математичні поняття, методи доменного, системного і об'єктно-орієнтованого аналізів та математичного моделювання для розробки програмного забезпечення.	1. Алгоритм РАМ. 2. Попереднє скорочення розмірності. 3. Факторний аналіз. 4. Ітеративна кластеризація в SPSS. 5. Растрова кластеризація об'єктів.		
	Практичне заняття 3 План заняття: 1. Застосування дерев класифікації у вирішенні завдань інтелектуального аналізу даних. 2. Підготовка даних та відбір ознак.	4	8
Тема 4. Методи класифікації			
Загальні та спеціальні компетентності: Здатність брати участь у проектуванні програмного забезпечення, включаючи проведення моделювання (формальний опис) його структури, поведінки та процесів функціонування. Результати навчання: Знати і застосовувати на практиці фундаментальні концепції, парадигми і основні принципи функціонування мовних, інструментальних і обчислювальних засобів інженерії програмного забезпечення.	Лекція №4 План лекції: 1. Завдання класифікації. 2. Процес класифікації. 3. Методи, що застосовуються для вирішення задач класифікації. 4. Точність класифікації: оцінка рівня помилок. 5. Оцінювання класифікаційних методів. 6. Дерева рішень. Список рекомендованих джерел: Основний: 1-5 Додатковий: 6-8 Інтернет-ресурси: 9	2	
	Задання для самостійної роботи: Вивчення та доповнення матеріалів лекції на основі самостійного опрацювання основних літературних джерел, зазначених у списку та електронних матеріалів, які надаються викладачем. Питання винесені на самостійне опрацювання: 1. Методи опорних векторів, «найближчого сусіда», або системи міркувань на основі аналогічних випадків. 2. Метод Байєса. Лінійний SVM. 3. Класифікація об'єктів у випадку невідомих розподілень даних.	12	4
	Практичне заняття 4 План заняття: 1. Прогнозування часових рядів. 2. Виділення періодичних складових. 3. Аналіз часового ряду.	4	8

Тема 5. Методи прогнозування			
Загальні та спеціальні компетентності: Володіння знаннями про інформаційні моделі даних, здатність створювати програмне забезпечення для зберігання, видобування та опрацювання даних. Результати навчання: Знати та вміти розробляти та реалізовувати сучасні інноваційні інформаційні технології проектування в області інтелектуальних транспортних систем та мехатронних систем і комплексів.	Лекція №5 <i>План лекції:</i> 1. Завдання прогнозування. 2. Порівняння задач прогнозування та класифікації. 3. Прогнозування і тимчасові ряди. 4. Завдання та методи візуалізації. Список рекомендованих джерел: Основний: 1,2,4 Додатковий: 7,8 Інтернет-ресурси: 9	2	
	Задання для самостійної роботи: Вивчення та доповнення матеріалів лекції на основі самостійного опрацювання основних літературних джерел, зазначених у списку та електронних матеріалів, які надаються викладачем. Питання винесені на самостійне опрацювання: 1. Рішення завдання прогнозування. 2. Аналіз багатомірних угруповань. 3. Статистична обробка тимчасових рядів і прогнозування.	12	5
	Практичне заняття 5 План заняття: 1. Кореляційно-регресивний аналіз. 2. Знаходження залежності результативної ознаки від однієї або кількох ознак-факторів.	4	8
Тема 6. Методи пошуку шаблонів даних			
Загальні та спеціальні компетентності: Здатність брати участь у проектуванні програмного забезпечення, включаючи проведення моделювання (формальний опис) його	Лекція №6 <i>План лекції:</i> 1. Асоціативні правила. 2. Послідовне відображення шаблонів даних. 3. Методи пошуку асоціативних правил. 4. Додатки із застосуванням асоціативних правил. 5. Характеристики асоціативних правил. Список рекомендованих джерел: Основний: 1-3, 5 Додатковий: 7, 8 Інтернет-ресурси: 9	2	

структури, поведінки та процесів функціонува. Результати навчання: Знати кодекс професійної етики, розуміти соціальну значимість та культурні аспекти інженерії програмного забезпечення і дотримуватись їх в професійній діяльності.	Задання для самостійної роботи: Вивчення та доповнення матеріалів лекції на основі самостійного опрацювання основних літературних джерел, зазначених у списку та електронних матеріалів, які надаються викладачем. Питання винесені на самостійне опрацювання: 1. Межі підтримки і достовірності асоціативного правила. 2. Методи пошуку асоціативних правил. 3. Різновиди алгоритму Apriori.	14	5
	Практичне заняття 6 План заняття: 1. Обробка текстових даних за допомогою Text Miner. 2. Аналіз текстових даних за допомогою Text Miner.	4	8
Тема 7. OLAP та Data Mining			
Загальні та спеціальні компетентності: Володіння знаннями про інформаційні моделі даних, здатність створювати програмне забезпечення для зберігання, видобування та опрацювання даних. Результати навчання: Знати основні процеси, фази та ітерації життєвого циклу програмного забезпечення.	Лекція №7 <i>План лекції:</i> 1. Методи, стадії задачі Data Mining. 2. Упровадження Data Mining. 3. OLAP і сховищ даних у СППР. Список рекомендованих джерел: Основний: 3-5 Додатковий: 7 Інтернет-ресурси: 9	2	
	Задання для самостійної роботи: Вивчення та доповнення матеріалів лекції на основі самостійного опрацювання основних літературних джерел, зазначених у списку та електронних матеріалів, які надаються викладачем. Питання винесені на самостійне опрацювання: 1. Процес Data Mining. 2. Стандарти Data Mining. 3. Інструменти Data Mining.	14	5
	Практичне заняття 7 План заняття: 1. Описова і передбачувальна аналітика за допомогою платформи Deductor Studio Academic.	4	8

Тема 8. Способи візуального представлення даних			
Загальні та спеціальні компетентності: Здатність брати участь у проектуванні програмного забезпечення, включаючи проведення моделювання (формальний опис) його структури, поведінки та процесів функціонування. Результати навчання: Знати і застосовувати відповідні математичні поняття, методи доменного, системного і об'єктно-орієнтованого аналізів та математичного моделювання для розробки програмного забезпечення.	Лекція №8 <i>План лекції:</i> 1. Візуалізація інструментів Data Mining. 2. Візуалізація Data Mining моделей. 3. Методи візуалізації. 4. Представлення даних в одному, двох і трьох вимірах. Список рекомендованих джерел: Основний: 1-4 Додатковий: 7 Інтернет-ресурси: 9	2	
	Задання для самостійної роботи: Вивчення та доповнення матеріалів лекції на основі самостійного опрацювання основних літературних джерел, зазначених у списку та електронних матеріалів, які надаються викладачем. Питання винесені на самостійне опрацювання: 1. Паралельні координати. Якість візуалізації. 2. Подання просторових характеристик. 3. Основні тенденції в області візуалізації.	14	5
	Практичне заняття 8 План заняття: 1. Звітність та візуалізація оброблених даних за допомогою платформи Deductor Studio Academic.	4	8
	Разом	150годин/ 5 кредити	100 балів
Підсумковий контроль		Іспит	

6. Список рекомендованих джерел

Основний

1. Барсегян А.А. Анализ данных и процессов / Барсегян А.А., Куприянов М.С., Холод И.И., Тесс М.Д., Елизаров С.И. // 3-е изд. – СПб.: БХВ-Петербург, 2009. – 512 с. – ISBN: 978-5-9775-0368-6
2. Чубукова И. А. Data Mining: учебное пособие / И.А. Чубукова. - 2-е изд., испр. – М.: INTUIT; БИНОМ. Лаборатория знаний, 2008. – 382 с. – ISBN: 978-5-

9556-0064-2.

3. Дюк В. Data Mining: учебный курс / Дюк В., Самойленко А. – СПб: Изд. Питер 2001. — 368 с. – ISBN: 5-318-00227-7.

4. Барсегян А.А. Технологии анализа данных. Data Mining, Visual Mining, Text Mining, OLAP / Барсегян А.А., Куприянов М.С., Степаненко В.В., Холод И. И. // 2-е изд. – СПб.: БВХ–Петербург, 2007, – 384 с. – 3-е изд. – ISBN 5-94157-991-8.

5. Барсегян А. А. Методы и модели анализа данных OLAP и Data Mining / Барсегян А. А., Куприянов М. С., Степаненко В. В., Холод И. И. // СПб.: БВХ–Петербург, 2004, – 336 с. – ISBN 5-94157-522-Х.

Додатковий

6. Спирли Эрик – Корпоративные хранилища данных. Планирование, разработка и реализация. Том 1 – Пер с англ. – М.: Издательский дом «Вильямс», 2001. – 400с.: ил.

7. Корнеев В. В. Базы данных. Интеллектуальная обработка информации / В. В. Корнеев, А. Ф. Гареев, С. В. Васютин и др. – М. : Изда-тель Могачева С. В. ; Издательство Нолидж, 2001. – 496 с.

8. Рассел С. Искусственный интеллект: современный подход / Рассел С., Норвинг П. – М. : Издательский дом "Вильямс", 2006. – 1408 с.

Интернет-ресурси

9. Дистанційний курс-ресурс «Інтелектуальний аналіз даних» [Електронний ресурс]. – Режим доступу: <http://dl.khadi.kharkov.ua/course/view.php?id=946>.