МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Харківський національний автомобільно-дорожній університет

«ЗАТВЕРДЖУЮ»
Заступник ректора ХНАДУ
професор _____ Гладкий І.П.
«___» _____ 2017 року

П Р О Г Р А М А

навчальної дисципліни Інтелектуальний аналіз даних
(назва навчальної дисципліни згідно навчального плану)

підготовки бакалавра
(назва освітньо-кваліфікаційного рівня)

галузі знань 12 Інформаційні технології
(шифр і назва галузі знань)

спеціальності 121 Інженерія програмного забезпечення
(шифр і назва напряму підготовки)

кваліфікація 3121 Фахівець з інформаційних технологій
(шифр і назва кваліфікації для бакалавра, спеціальності – для магістра)

(шифр № 1.3.05)

2017 рік

Розроблено та внесено: кафедрою комп'ютерних технологій та мехатроніки

Розробники програми: професор кафедри комп'ютерних технологій та мехатроніки, д.т.н., доцент Клец Дмитро Михайлович

Обговорено та рекомендовано до затвердження на засіданні кафедри.
Протокол № 1 від 28.08.2017 р.

Завідувач кафедри докт. техн. наук, доцент _____ Клец Д.М.

"Узгоджено"
Завідувач кафедри комп'ютерних технологій та мехатроніки
доцент _____ Клец Д.М.

28 серпня 2017 року

"Узгоджено"
Декан механічного факультету

професор _____ Кириченко І.Г.

28 серпня 2017 року

© Клец Д.М., 2017 рік
© Клец Д.М., 2022 рік

ВСТУП

Програма вивчення навчальної дисципліни **"Інтелектуальний аналіз даних"** складена відповідно до освітньо-кваліфікаційної характеристики та навчального плану підготовки **бакалавра** спеціальності **121 Інженерія програмного забезпечення**.

1. Мета, предмет та завдання навчальної дисципліни

1.1. **Метою** викладання дисципліни є теоретична та практична підготовка студентів до вивчення систем обробки даних та принципів інтелектуального аналізу даних на основі методів та алгоритмів Data Mining.

1.2. **Предмет дисципліни** включає створення та дослідження математичних та програмних моделей інтелектуального аналізу даних, пов'язаних із функціонуванням об'єктів професійної діяльності.

1.3. **Основними завданнями** є формування у студентів уяви про застосування технологій зберігання та організації даних, методів та алгоритмів Data Mining, процесів виявлення знань, принципів побудови сховищ даних, а також способів візуального представлення даних.

1.4. В результаті вивчення навчальної дисципліни студент повинен:

знати:

- базові визначення й поняття, основні етапи аналізу та інтерпретації даних, загальну структуру і функціонування інтегрованих інформаційно-аналітичних систем (систем підтримки прийняття рішень);
- основи організації й технології обробки експериментальних і статистичних даних на ЕОМ;
- змістовну й математичну постановку основних завдань аналізу даних, способи їхнього розв'язання;
- сутність і властивості основних алгоритмів інтелектуального аналізу;
- приклади рішення практичних завдань аналізу даних.

вміти:

- зводити словесні постановки завдань до типових математичних й відносити їх до відповідних розділів математики та зв'язувати з відомими засобами обробки експериментальних даних;
- створювати і користуватися типовими математичними моделями для дослідження випадкових явищ та процесів за результатами спостережень при розв'язанні завдань: порівняння, виявлення й відновлення закономірностей, класифікації, прогнозування;
- ставити завдання, вибирати засоби й методи обробки експериментальних даних для ефективного застосування;
- представляти результати аналізу в зручному для сприйняття вигляді, інтерпретувати їх відповідно до поставленого завдання.

Міждисциплінарні зв'язки:

Вивченню дисципліни

передують: «Вища математика», «Дискретна математика», «Теорія ймовірностей, ймовірні процеси», «Теорія алгоритмів», «Чисельні методи», «Математичні

методи дослідження операцій», «Організація баз даних та знань».
потребують її вивчення: «Проектування інформаційних систем», «Методи та системи штучного інтелекту», «Системна інженерія», «Моделювання транспортних потоків», «Переддипломна практика», «Дипломне проектування».

1.ОПИС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Найменування показників	Характеристика навчальної дисципліни	
	денна форма навчання	**заочна (дистанційна) форма навчання**[1]
Кількість кредитів – 4 **Кількість годин** - 120	нормативна (нормативна, за вибором ВНЗ, за вибором студента)	
Семестр викладання дисципліни	7	
Вид контролю:	екзамен (залік, екзамен)	
Розподіл часу:		
- лекції (годин)	16	
- практичні, семінарські (годин)		
- лабораторні роботи (годин)	32	
- самостійна робота студентів (годин)	72	
- курсовий проект (годин)		
- курсова робота (годин)		
- розрахунково-графічна робота (контрольна робота)		
- підготовка та складання екзамену		

2. ІНФОРМАЦІЙНИЙ ОБСЯГ НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Розділ 1. Первинна обробка даних

Тема 1. Основи інтелектуального аналізу даних.

Поняття інтелектуального аналізу даних. Методи, стадії та завдання інтелектуального аналізу даних. Data Mining як частина ринку інформаційних технологій. Розвиток технології баз даних. Порівняння статистики, машинного навчання та Data Mining.

Тема 2. Регресійний аналіз та нелінійні моделі

Методи первинної обробки даних. Методи дослідження структури даних: візуалізація та автоматичне групування даних. Описова статистика. Кореляційний аналіз. Регресійний аналіз. Множинний регресійний аналіз. Нелінійна множинна регресійна модель. Нелінійне оцінювання параметрів.

Тема 3. Методи багатомірного розвідувального аналізу

Методи кластерного аналізу. Заходи подібності. Методи об'єднання або зв'я-

[1] Якщо дисципліна на заочній (дистанційній) формі навчання не викладається, то графа "заочна форма навчання" відсутня.

зку. Ієрархічний кластерний аналіз в SPSS. Ітеративні методи. Алгоритм k-середніх (k-means). Алгоритм PAM. Попереднє скорочення розмірності. Факторний аналіз. Ітеративна кластеризація в SPSS. Растрова кластеризація об'єктів. Лінійний дискримінантний аналіз.

Розділ 2. Класифікація та прогнозування

Тема 4. Методи класифікації

Завдання класифікації. Процес класифікації. Методи, що застосовуються для вирішення задач класифікації. Точність класифікації: оцінка рівня помилок. Оцінювання класифікаційних методів. Дерева рішень. Методи опорних векторів, «найближчого сусіда», або системи міркувань на основі аналогічних випадків. Метод Байєса. Лінійний SVM. Класифікація об'єктів у випадку невідомих розподілень даних.

Тема 5. Методи прогнозування

Завдання прогнозування. Порівняння задач прогнозування та класифікації. Прогнозування і тимчасові ряди. Завдання та методи візуалізації. Рішення завдання прогнозування. Аналіз багатомірних угрупувань. Статистична обробка тимчасових рядів і прогнозування.

Тема 6. Методи пошуку шаблонів даних

Асоціаційні правила. Послідовне відображення шаблонів даних. Методи пошуку асоціативних правил. Додатки із застосуванням асоціативних правил. Характеристики асоціативних правил. Межі підтримки і достовірності асоціативного правила. Методи пошуку асоціативних правил. Різновиди алгоритму Apriori.

Розділ 3. Упровадження моделі та представлення даних

Тема 7. OLAP та Data Mining

Методи, стадії задачі Data Mining. Упровадження Data Mining, OLAP і сховищ даних у СППР. Процес Data Mining. Стандарти Data Mining. Інструменти Data Mining.

Тема 8. Способи візуального представлення даних

Візуалізація інструментів Data Mining. Візуалізація Data Mining моделей. Методи візуалізації. Представлення даних в одному, двох і трьох вимірах. Паралельні координати. Якість візуалізації. Подання просторових характеристик. Основні тенденції в області візуалізації.

Теми лабораторних занять:

1. Первинний розвідувальний аналіз даних (4 години).
2. Кластерний аналіз даних (4 години).
3. Застосування дерев класифікації у вирішенні завдань інтелектуального аналізу даних (4 години).

4. Прогнозування часових рядів (4 години).
5. Кореляційно-регресивний аналіз (4 години).
6. Обробка та аналіз текстових даних за допомогою Text Miner (4 години).
7. Описова і передбачувальна аналітика за допомогою платформи Deductor Studio Academic (4 години).
8. Звітність та візуалізація оброблених даних за допомогою платформи Deductor Studio Academic (4 години).

3. РЕКОМЕНДОВАНА ЛІТЕРАТУРА ТА ІНФОРМАЦІЙНІ РЕСУРСИ

Базова

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-X.

Допоміжна

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. : Издатель 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.

4. ЗАСОБИ ДІАГНОСТИКИ УСПІШНОСТІ НАВЧАННЯ

(вказати перелік засобів контролю успішності навчання студентів, які застосовуються: тести, екзаменаційні білети, тощо)

Підсумковий контроль знань та компетентностей студентів з навчальної дисципліни здійснюється на підставі захисту курсової роботи та проведення семестрового екзамену.

Екзаменаційний білет включає теоретичні питання та практичні завдання, які оцінюються відповідно до положення "Про порядок оцінювання результатів навчання студентів за накопичувальною бально-рейтинговою системою"