

Міністерство освіти і науки України
Український державний університет науки та технологій
Кафедра інформаційних технологій і систем

ЗАТВЕРДЖУЮ
Перший проректор УДУНТ

Проф. _____Анатолій РАДКЕВИЧ
" ____ " _____ 2022 р.

Програма навчальної дисципліни
«Інтелектуальний аналіз даних»

| | |
|------------------------------------|---|
| Шифр та назва спеціальності | 121 – Інженерія програмного забезпечення |
| Назва освітньої програми (програм) | Інженерія програмного забезпечення у промисловості і бізнесі |
| Рівень вищої освіти | Перший (бакалаврський) рівень, 4 курс |
| Статус дисципліни | Обов'язкова навчальна дисципліна циклу фундаментальної підготовки |

Види навчальної роботи та її обсяг в акад. годинах (денна форма навчання)

| | Усього |
|---|--|
| Усього годин за навчальним планом | 90 |
| у тому числі: | |
| Аудиторні заняття | 40 |
| з них: | 16 |
| - лекції | |
| - лабораторні роботи | 24 |
| - практичні заняття | 0 |
| - семінарські заняття | 0 |
| Самостійна робота | 50 |
| у тому числі при : | 20 |
| - підготовці до аудиторних занять | |
| - підготовці до заходів модульного контролю | 9 |
| - виконанні курсових проектів (робіт) | - |
| - опрацюванні розділів програми, які не викладаються на лекціях | 21 |
| Семестровий контроль | середнє арифметичне 3-х модульних оцінок або іспит |

Характеристика дисципліни

Мета вивчення дисципліни «Інтелектуальний аналіз даних» – знайомить студентів з технологією Data Mining, докладно розглядаються методи, інструментальні засоби й застосування Data Mining. Опис кожного методу супроводжується конкретним прикладом його використання.

Обговорюються відмінності Data Mining від класичних статистичних методів аналізу й OLAP-систем, розглядаються типи закономірностей, що виявляють Data Mining (асоціація, класифікація, послідовність, кластеризація, прогнозування). Описується сфера застосування Data Mining.

Вводиться поняття Web Mining. Докладно розглядаються методи Data Mining: 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 Data Mining. Викладаються основні концепції сховищ даних і місця Data Mining у їхній архітектурі. Вводяться поняття OLTP, OLAP, ROLAP, MOLAP. Обговорюється процес аналізу даних за допомогою технології Data Mining. Докладно розглядаються етапи цього процесу. Аналізується ринок аналітичного програмного забезпечення, описуються продукти від провідних виробників Data Mining, обговорюються їхні можливості.

Компетентності, формування яких забезпечує навчальна дисципліна

| | |
|----------------------------|--|
| Інтегральна компетентність | Здатність розв'язувати складні спеціалізовані задачі та практичні проблеми у галузі комп'ютерних наук або у процесі навчання, що передбачає застосування теорій та методів комп'ютерних наук і характеризується комплексністю та невизначеністю умов. |
| Загальні компетентності | - ЗК1 Здатність до абстрактного мислення, аналізу та синтезу. - ЗК2 Здатність застосовувати знання у практичних ситуаціях. - ЗК5 Здатність вчитися і оволодівати сучасними знаннями. - ЗК6 Здатність до пошуку, оброблення та аналізу інформації з різних джерел. |
| Фахові компетентності | - СК7 Володіння знаннями про інформаційні моделі даних, здатність створювати програмне забезпечення для зберігання, видобування та опрацювання даних. - СК8 Здатність застосовувати фундаментальні і міждисциплінарні знання для успішного розв'язання завдань інженерії програмного забезпечення. - СК10 Здатність накопичувати, обробляти та систематизувати професійні знання щодо створення і супроводження програмного забезпечення та визнання важливості навчання протягом всього життя. - СК14 Здатність до алгоритмічного та логічного мислення. |

Завдання курсу - набуття студентами базових компетенцій у галузі професійного використання інформаційних технологій аналізу даних як інструменту для різних напрямків бізнесу.

У результаті вивчення дисципліни студент повинен:

мати уявлення:

- про роль та цілі застосування комп'ютеризованого аналізу й інтерпретації даних у пізнанні навколишнього світу та вирішенні практичних завдань;
- про прикладні методи аналізу даних і видобутку знань, основних напрямках та тенденцій їхнього розвитку;
- про можливості й перспективи застосування методів інтелектуального аналізу даних у дослідженні об'єктів і процесів професійної діяльності в області управління та організації;
- про організацію й застосування сучасних інформаційних технологій і програмних систем для обробки експериментальних і статистичних даних;

знати:

- предмет та об'єкти вивчення, структуру даного курсу, сутність використаних у ньому основних методів досліджень, його місце в навчальному плані для підготовки дипломованих фахівців з даного напрямку;
- базові визначення й поняття, основні етапи аналізу та інтерпретації даних, загальну структуру і функціонування інтегрованих інформаційноаналітичних 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;
- здійснювати самооцінку й самоконтроль при розв'язанні завдань обробки статистичних даних.

Дисципліна забезпечує досягнення таких **програмних результатів навчання:**

| | |
|-------------------------------|--|
| Програмні результати навчання | <ul style="list-style-type: none"> - ПР05 Знати і застосовувати відповідні математичні поняття, методи доменного, системного і об'єктно-орієнтованого аналізів та математичного моделювання для розробки програмного забезпечення. - ПР10 Проводити перед проектне обстеження предметної області, системний аналіз об'єкта проектування. |
|-------------------------------|--|

Заходи та методи оцінювання

Отримання позитивної оцінки при виконанні 3-х модульних контрольних робіт за 12-бальною шкалою.

Підсумкова оцінка навчальної дисципліни визначається як середнє арифметичне 3-х модульних оцінок за 12-бальною шкалою або іспитом.

Передумови вивчення дисципліни

Вивченню дисципліни має передувати вивчення дисциплін:

- Вища математика.
- Фізика.
- Основ теорії баз даних.
- Технології обробки інформації.
- Методи та системи штучного інтелекту.
- Моделі та методи нейро-нечітких технологій.

Структура дисципліни

| Модуль та назва | Тема заняття | Обсяг, годин |
|---|---|--------------|
| Модуль 1. | Лекції | 6 |
| Вступ в інтелектуальний аналіз даних. Поняття Data Mining. Методи Data Mining. Етапи аналізу даних. | 1. Вступ в інтелектуальний аналіз даних. Поняття Data Mining. Видобуток даних – Data Mining. Задачі Data Mining. Класифікація задач Data Mining. Задача класифікації й регресії (задача пошуку асоціативних правил, задача кластеризації). Практичне застосування Data Mining (інтернет-технології, торгівля, телекомунікації, промислове виробництво, медицина, банківська справа, страховий бізнес, інші області застосування). | 2 |
| | 2. Методи Data Mining (базові методи, нечітка логіка, генетичні алгоритми, нейронні мережі) | 6 |
| | 3. Процес виявлення знань (основні етапи аналізу, підготовка вихідних даних). Управління знаннями (Knowledge Management). Засоби Data Mining | 2 |
| | Лабораторні роботи | 8 |
| | 1. Надбудови інтелектуального аналізу даних | 4 |

| | | |
|---|--|-----------|
| | для Microsoft Office | |
| | 2. Використання інструментів "FillFromExample" та "Forecast"» | 4 |
| | Самостійна робота | 16 |
| | Підготовка до аудиторних занять | 7 |
| | Підготовка до заходів модульного контролю | 3 |
| | Опрацюванні розділів програми, які не викладаються на лекціях. Надбудови інтелектуального аналізу даних Microsoft SQL Server | 6 |
| | Усього: | 30 |
| Модулі 2. | Лекції | 4 |
| Візуальний аналіз даних – Visual Mining та Аналіз текстової інформації – Text Mining. | 1. Візуальний аналіз даних – Visual Mining. Виконання візуального аналізу даних. Характеристики засобів візуалізації даних. Методи візуалізації. Методи геометричних перетворень. Відображення іконок. Методи, орієнтовані на пікселі. Ієрархічні образи | 2 |
| | 2. Аналіз текстової інформації – Text Mining. 1. Задача аналізу текстів (етапи аналізу текстів, попередня обробка тексту, задачі Text Mining). Добування ключових понять із тексту (загальний опис процесу добування понять із тексту, стадія локального аналізу, стадія інтеграції й висновку понять). Класифікація текстових документів (опис задачі класифікації текстів, методи класифікації текстових документів). Методи кластеризації текстових документів (подання текстових документів, ієрархічні методи кластеризації текстів, бінарні методи кластеризації текстів). Задача анотування текстів (виконання анотування текстів, методи добування фрагментів для анотації). Засоби аналізу текстової інформації (засоб Oracle – Oracle Text, засоб від IBM – Intelligent Miner for Text, засоб SAS Institute – Text Miner, засоб Мега-комп'ютер Інтелідженс – Text Analyst). | 2 |
| | Лабораторні роботи | 8 |
| | 1. Використання інструментів "HighlightExceptions" та "ScenarioAnalysis" та Використання інструментів "Prediction Calculator" та "ShoppingbasketAnalysis" | 4 |
| | 2. Використання інструментів Data Mining Client для Excel 2007 для підготовки та | 4 |

| | | |
|---|--|-----------|
| | створення моделі інтелектуального аналізу даних» | |
| | Самостійна робота | 18 |
| | Підготовка до аудиторних занять | 6 |
| | Підготовка до заходів модульного контролю | 3 |
| | Опрацюванні розділів програми, які не викладаються на лекціях. Дерева рішень. Завдання класифікації, прогнозування та скорингу. | 9 |
| | Усього: | 30 |
| Модуль 3. | Лекції | 6 |
| Добування знань з Web – Web Mining та Засоби аналізу процесів – Process Mining. Пошук асоціативних правил – Rules Mining. | 1. Добування знань з Web – Web Mining. Web Mining (проблеми аналізу інформації з Web, етапи Web Mining, Web Mining й інші інтернет-технології, категорії Web Mining). Методи добування Web-контенту (добування Web-контенту в процесі інформаційного пошуку, добування Web-контенту для формування баз даних). Добування Web-структур (подання Web-структур, оцінка важливості Web-структур, пошук Web-документів з урахуванням гіперпосилань, кластеризація Web-структур). Дослідження використання Web-ресурсів (дослідницька інформація, етап препроцесінгу, етап добування шаблонів, етап аналізу шаблонів й їхнє застосування). | 2 |
| | 2. Засоби аналізу процесів – Process Mining. Автоматизація виконання бізнесів-процесів (бізнеси-процеси, формалізація бізнесів-процесів, Workflow-системи, сервісно-орієнтована архітектура, проектування бізнесів-процесів). Аналіз процесів (технологія Process Mining, аналіз протоколів, стандарт запису протоколів MXML, задачі Process Mining, проблеми аналізу протоколів). Методи Process Mining (перші імовірнісні методи Process Mining, метод побудови диз'юнктивної Workflow-схеми, -алгоритм, методи на основі генетичних алгоритмів). Бібліотека алгоритмів Process Mining— Pro (архітектура Pro, Pro Import Framework). | 2 |
| | 3. Пошук асоціативних правил – Rules Mining. Постановка задачі. Подання результатів (правила класифікації, дерева класифікації, математичні функції). Методи побудови | 2 |

| | |
|--|-----------|
| правил класифікації (алгоритм побудови правил, метод Naive Bayes). Методи побудови дерев класифікації (методика "розділяй і пануй", алгоритм покриття). Методи побудови математичних функцій (загальний вид, лінійні методи, метод найменших квадратів, нелінійні методи, Support Vector Machines (SVM), регуляризаційні мережі (Regularization Networks), дискретизації й рідкі сітки). Постановка задачі пошуку асоціативних правил (формальна постановка задачі, секвенціальний аналіз, різновиди задачі пошуку асоціативних правил). Подання результатів. Алгоритми (алгоритм Apriori, різновиди алгоритму Apriori). | |
| Лабораторні роботи | 8 |
| 1. Аналіз точності прогнозу та використання моделі інтелектуального аналізу | 4 |
| 2. Побудова моделі кластеризації, трасування та перехресна перевірка | 4 |
| Самостійна робота | 16 |
| Підготовка до аудиторних занять | 7 |
| Підготовка до заходів модульного контролю | 3 |
| Опрацюванні розділів програми, які не викладаються на лекціях. Розподілений аналіз даних. | 6 |
| Усього: | 30 |

Рекомендована література

Основна література:

1. Замятин, А.В. Интеллектуальный анализ данных: учебное пособие / А.В. Замятин / Национальный исследовательский Томский государственный университет. – Томск: Томский государственный университет, 2016. – 119 с.
2. Флах, П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных / П. Флах. – М.: ДМК Пресс, 2015. – 775 с.
3. Рашка С. Python и машинное обучение / С. Рашка. – М.: ДМК-Пресс, 2017. – 418 с.
4. Олійник А.О. Інтелектуальний аналіз даних: навчальний посібник / А.О. Олійник, С.О. Субботін, О.О. Олійник. – Запоріжжя : ЗНТУ, 2012. – 278 с.
5. Черняк О.І., Захарченко П.В. Інтелектуальний аналіз даних: Підручник. К., 2014. 599 с.
6. Інтелектуальний аналіз даних: Комп'ютерний практикум [Електронний ресурс]: навч. посіб. для студ. спеціальності 122 «Комп'ютерні науки та інформаційні технології», спеціалізацій «Інформаційні системи та технології проектування», «Системне проектування сервісів» / О.О. Сергеев-Горчинський,

Г. В. Іщенко ; КПІ ім. Ігоря Сікорського. – Електронні текстові данні (1 файл: 1,72 Мбайт). – Київ : КПІ ім. Ігоря Сікорського, 2018. – 73 с.: Іл.

Додаткова література:

1. Макленнен Д. Microsoft SQL Server 2008: Data Mining – интеллектуальный анализ данных / Д. Макленнен, Ч. Танг, Б. Криват. – СПб.: БХВ-Петербург, 2009. – 720 с.
2. Ian H. Witten, Eibe Frank and Mark A. Hall. Data Mining: Practical Machine Learning Tools and Techniques. - 3rd Edition. - Morgan Kaufmann, 2011. - P. 664
3. Data Mining and Image Processing Toolkits. – [Електронний ресурс]. – Режим доступу <http://datamining.itsc.uah.edu/adam/>.
4. Сайт з OLAP – технологій та сховищ даних - [Електронний ресурс]. – Режим доступу – <http://www.olap.ru>
5. Сайт одного із засновників Data Mining Г. Піатецького-Шапіро [Електронний ресурс]. – Режим доступу – www.kdnuggets.com
6. Електронна бібліотека. – Режим доступу : <http://all-ebooks.com>.
7. Свободна енциклопедія [Електронний ресурс]. – Режим доступу : http://ru.wikipedia.org/wiki/Data_mining.
8. Дюк В.А. Применение технологий интеллектуального анализа данных в естественнонаучных, технических и гуманитарных областях [Электронный ресурс] / В.А. Дюк, А. В. Флегонтов, И.К. Фомина. – Режим доступа : ftp://lib.herzen.spb.ru/text/dyuk_138_77_84.pdf.
9. Шапот М. Интеллектуальный анализ данных в системах поддержки принятия решений [Электронный ресурс] / М. Шапот. – Режим доступу : <http://www.osp.ru/os/1998/01/179360/>

Укладач:

к.т.н, доц., доцент кафедри ІТС _____ Катерина ОСТРОВСЬКА

Завідувач кафедри інформаційних технологій і систем (ІТС):

д.т.н., доц. _____ Вікторія ГНАТУШЕНКО

Робоча програма погоджена групою забезпечення якості освітньої програми «Інженерія програмного забезпечення у промисловості і бізнесі», спеціальність 121 «Інженерія програмного забезпечення» (Протокол №4/21-22 від 15 червня 2022 р.).

Гарант освітньої програми,

к.т.н, доц. _____ Тетяна СЕЛІВЬОРСТОВА

Погоджено:

Керівник навчального відділу _____ Володимир ПУЛЬПІНСЬКИЙ