



Силабус дисципліни

Інтелектуальний аналіз даних

Викладач: Болюбаш Надія Миколаївна
доцент, к.пед.н., доцент кафедри ІС

Очікувані результати навчання

В результаті вивчення дисципліни студент

має знати:

на *понятійному рівні* – основні поняття інтелектуальної технології Data Mining, чітко оперувати і володіти її понятійно-категоріальним апаратом (*фактичне знання*);

на *фундаментальному рівні* – основні принципи, категорії Data Mining, основні задачі Data Mining та етапи й інструментальні засоби їх розв’язання, моделі й структурні компоненти технології Data Mining (*концептуальне знання*);

на *практичному рівні* – етапи та алгоритми попереднього аналізу й обробки вхідних даних; основні алгоритми та методи інтелектуального аналізу при розв’язанні задач класифікації, кластеризації, візуалізації даних, пошуку асоціативних правил та послідовностей, прогнозування; застосування програмних та інструментальних засобів для розв’язання задач інтелектуального аналізу даних.

на *творчому рівні* – контекстні та умовні когнітивні завдання, способи запам’ятовування фактів, стратегії розуміння сприйнятого, планування, методи самопізнання (*метакогнітивне знання*).

має вміти:

на *репродуктивному рівні* – орієнтуватися, упізнавати та пригадувати основні поняття, алгоритми,

Обсяг: буде визначено після здійснення процедури вибору студентами вибіркового дисциплін.

Мета: надання теоретичної та практичної підготовки майбутнім ІТ-фахівцям в області інтелектуального аналізу даних (Data Mining), формування у них умінь та навичок використання програмних продуктів, інструментальних засобів, алгоритмів та методів Data Mining при розв’язанні задач аналітичного дослідження великих масивів інформації з метою виявлення знань, необхідних для прийняття рішень у певній предметній області.

Оригінальність навчальної дисципліни:

Авторський курс

Зміст дисципліни

Тема 1. Основи інтелектуального аналізу даних.

Основні поняття, методи та стадії, задачі Data Mining. Области застосування, проблеми та перспективи Data Mining.

Тема 2. Попередня обробка даних.

Набір даних та їх атрибутів. Дані та шкали. Перетворення даних (нормалізація, стандартизація, дискредитація, агрегація), очищення даних.

Тема 3. Первинний статистичний аналіз даних.

Вибірка даних, варіаційний ряд та його характеристики. Візуалізація даних (полігон, гістограма, кумулята, теплові карти). Характеристики та закони розподілу неперервної і дискретної випадкової величини. Оцінка параметрів розподілу. Статистичні гіпотези та критерії їх прийняття.

Тема 4. Кластерний аналіз даних.

Матриця відстаней, міри близькості, метрики. Таблиці спряженості. Етапи кластерного аналізу, представлення результатів. Оцінка якості та забезпечення надійності кластеризації. Алгоритми кластеризації (ієрархічні, Hard C-Means, Fuzzy C-Means).

Тема 5. Класифікація.

Постановка задачі класифікації та регресії. Навчаюча та тестова множини. Класифікатор.

методи технології Data Mining, розуміти можливості їх використання при аналітичному дослідженні великих масивів інформації, здійснювати розв'язання задач інтелектуального аналізу даних з використанням інструментальних засобів Data Mining;

на реконструктивному рівні – розуміти знання з інтелектуального аналізу даних на фундаментальному та практичному рівнях і здійснювати з їх застосуванням при розв'язанні задач Data Mining організацію та співвіднесення: інтерпретувати, наводити приклади, визначати тип задачі аналізу та здійснювати її постановку, робити класифікації, узагальнення, порівняння, пояснення, умовиводи, представлення розв'язків у вигляді моделей, алгоритмів і проектів для прийняття рішень, розробки та реалізації інтелектуальних інформаційних систем;

на творчому рівні – оцінювати та перевіряти особливості використання алгоритмів, методів, засобів технології Data Mining при розв'язанні конкретних задач інтелектуального аналізу у різних сферах діяльності; обґрунтовувати вибір алгоритму, методу, моделі, інструментального та програмного засобу для вирішення поставленої задачі та оцінювати якість її розв'язання; здійснювати оцінку, генерацію (набір альтернативних рішень на основі запропонованих критеріїв), розробку та реалізацію проектів Data Mining; здійснювати оцінку проблем при інтелектуальному аналізі та визначати шляхи їх вирішення; формувати ефективні методичні стратегії з метою донесення ідей, проблем, рішень у сфері інтелектуального аналізу до суспільства.

Пререквізити

Вивчення дисципліни базується на знанні основ вищої та дискретної математики, безпосередньо спирається на теорію ймовірностей і математичну статистику, теорію алгоритмів, програмування, організацію баз даних та знань,

Етапи класифікації. Дискримінантний аналіз. Метод k-ближніх сусідів (kNN). Методи побудови математичних функцій, правил класифікації (1R, Naive Bayes), дерев рішень (алгоритми покриття, ID3).

Тема 6. Пошук асоціативних правил та послідовностей.

Постановка задачі. Асоціативні правила, їх оцінки (підтримка, достовірність, ліфт, левередж, поліпшення. Алгоритми пошуку асоціативних правил, алгоритм Apriori.

Тема 7. Кібернетичні методи Data Mining.

Основні поняття штучних нейронних мереж та їх архітектура. Методи навчання нейронних мереж. Застосування нейронних мереж для розв'язання задач класифікації, кластеризації, прогнозування. Генетичні алгоритми як інструменти Data Mining.

Тема 8. Виявлення зв'язків та закономірностей.

Кореляційний аналіз даних. Коефіцієнт кореляції Пірсона. Дисперсійний аналіз даних, його етапи. Перевірка гіпотез про рівність генеральних середніх і дисперсій, F-критерій Фішера. Регресійний аналіз даних, його методи, етапи. Перевірка моделі на адекватність. Оцінка моделі регресії, індекс кореляції, детермінації. Основні положення та методи, етапи факторного аналізу даних.

Тема 9. Прогнозування.

Задача прогнозування в Data Mining. Поняття часового ряду, його структура. Етапи побудови прогнозу по часовим рядам. Трендові моделі, їх вибір та оцінка адекватності.

Тема 10. Перспективні напрями технології Data Mining.

Інтелектуальний аналіз текстової інформації (Text Mining), освітніх даних (Educational Data Mining, Learning Analytics). Web Mining: добування знань з WWW. Data Mining в реальному часі. Мобільні агенти у розв'язанні задач Data Mining.

Критерії оцінювання лабораторних/практичних/індивідуальних/робіт/ доповідей/проектів

1. При оцінювання завдань з лабораторних робіт максимальна кількість балів (3-4 бали за окрему лабораторну роботу) буде нарахована у випадку, якщо студент продемонстрував знання теоретичного матеріалу, необхідного для виконання завдань, засвоїв основні алгоритми аналізу даних, необхідні для їх розв'язання та виконав у повному обсязі усі поставлені

математичні методи дослідження операцій, що знаходить відображення у міжпредметних зв'язках цих дисциплін зі змістом курсу «Інтелектуальний аналіз даних».

Пореквізити

Знання, отримані під час проходження дисципліни, можуть бути використані при вивченні дисциплін професійного спрямування та у подальшій професійній роботі майбутніх ІТ-фахівців.

Семестровий контроль: іспит

Оцінювання:

За семестр: 60 балів

За іспит: 40 балів

Види робіт:

При вивченні дисципліни передбачено виконання лабораторних робіт, підсумкової контрольної роботи та виконання курсової роботи (яка оцінюється окремо максимально в 100 балів). По кожній з тем, яка вивчається на лабораторних заняттях передбачено виконання індивідуальних самостійних завдань.

Технічне забезпечення

Лекції проводяться у аудиторії, обладнаній проектором та комп'ютером з можливістю демонстрації слайд-лекцій.

При проведенні лабораторних занять у комп'ютерних класах використовується наступне програмне та апаратне забезпечення:

1) віддалений Windows-сервер, віддалений Unix-сервер, термінальні клієнтські станції, мережеве обладнання (маршрутизатори, комутатори, проху-сервери);

2) операційна система Windows 10, система управління навчанням LMS Moodle, Microsoft Office 2016, інструментальні засоби для написання програм на засвоєних

завдання лабораторної роботи.

№	Тема лабораторної роботи	Максимальна кількість балів
1	Обробка даних у середовищі MatLab. Нормалізація, стандартизація даних	3
2	Набори даних. Шкали. Попередня обробка даних	4
3	Первинний статистичний аналіз даних	3
4	Закони розподілу. Оцінка параметрів розподілу	3
5	Ієрархічна кластеризація даних в SPSS	3
6	Здійснення ієрархічного, чіткого та нечіткого кластерного аналізу у середовищі MatLab	3
7	Класифікація. Дискримінантний аналіз	3
8	Методи розв'язання задачі класифікації	3
9	Пошук асоціативних правил	3
10	Застосування нейронних мереж для розв'язання задач кластеризації та класифікації	4
11	Дисперсійний аналіз	4
12	Регресійний аналіз. Лінійна регресія	4
13	Аналіз часових рядів	4
14	Факторний аналіз	4

2. При оцінюванні підсумкової контрольної роботи максимальна кількість балів – 12, буде нарахована у випадку, якщо студент продемонстрував сформованість практичних умінь та навичок застосування алгоритмів інтелектуального аналізу для розв'язання усіх поставлених завдань.

3. Під час підсумкового оцінювання при складанні іспиту максимальна кількість балів – 40 балів буде нарахована у випадку, якщо студент продемонстрував знання теоретичного матеріалу при розкритті перших трьох питань білету та сформованість практичних умінь і навичок застосування алгоритмів інтелектуального аналізу для розв'язання двох практичних завдань білету.

мовах програмування (C++, Java, PHP, Python);

3) програмні засоби з вбудованими алгоритмами інтелектуального аналізу даних SPSS, MatLab, MathCAD, MS Excel.

Політика щодо дедлайнів

Роботи, які здаються із порушенням термінів без поважних причин, оцінюються на нижчу оцінку.

Політика щодо академічної доброчесності

Передбачає самостійне виконання лабораторних робіт. У разі виявлення плагіату або списування роботи не зараховуються.