

Алгоритми та програмні засоби реалізації інтелектуального аналізу потоків даних

Автор: Калій Д. О. Тв-61м

Керівник: Гагарін О. О.

Визначення

Data Mining - виявлення прихованих закономірностей або взаємозв'язків між змінними у великих масивах необроблених даних. Зазвичай поділяють на задачі класифікації, моделювання та прогнозування.

Актуальність

На сьогоднішній день актуальність використання методів інтелектуального аналізу даних лише збільшується, оскільки обміни накопиченої інформації та її диференціація стають все більш значними. Тому методи її аналізу також мають постійно вдосконалюватися. Через значну кількість методів одні краще підходять для реалізації одних задач, а другі інших. Саме коло задач, що можуть бути розв'язані подібними методами також постійно розширюється.

Мета роботи

Мета дослідження:

Дослідити існуючі методи інтелектуального аналізу даних та існуючі підходи до їх використання у сфері вищої освіти, визначити аспекти навчального процесу, аналіз яких дасть можливість підвищити ефективність роботи вищого навчального закладу та запропонувати власний алгоритм, що міг би, аналізуючи попередньо визначені аспекти функціонування ВНЗ, вносити рекомендації щодо підвищення ефективності ВНЗ в цілому.

- ▶ **Наукова частина:** удосконалення алгоритму, що має здійснювати обробку даних вищого навчального закладу, використовуючи методи інтелектуального аналізу даних, візуалізувати результати аналізу та виносити рекомендації, щодо подальшого підвищення ефективності роботи ВНЗ.
- ▶ **Практична частина:** розробка інструментарію для проведення досліджень запропонованого алгоритму з метою його подальшого покращення, розширення та можливої інтеграції з більш складними системами.

Задачі

- ▶ проаналізувати існуючі методи інтелектуального аналізу даних та визначити ті методи, що найбільш відповідають специфіці даних сфери вищої освіти;
- ▶ проаналізувати існуючі підходи до оцінки ефективності роботи вищих навчальних закладів та визначити підхід, що може бути реалізований обраними методами аналізу даних і зможе принести відчутний результат;
- ▶ удосконалити обраний підхід, реалізувавши аналіз максимальної кількості параметрів, зміна значень яких може впливати на підвищення якості вищої освіти;
- ▶ розробити програмний застосунок, що реалізує обраний підхід до аналізу ефективності вищої освіти методами Data Mining, може показувати зрозумілі рядовому користувачу результати та візуалізувати їх.

Об'єкт та предмет дослідження

Об'єкт дослідження : Методи інтелектуального аналізу даних (Data Mining)

Предмет дослідження : Задача оцінки ефективності навчальної програми ВНЗ методами Data Mining

Методи технології Data Mining

- Лінійна регресія
- Нейронні мережі
- Дерева прийняття рішень
- К - ближчого сусіда

Порівняльна характеристика алгоритмів Data Mining

Алгоритм	Точність	Масштабованість	Інтерпритованість	Придатність до використання
Лінійна регресія	Нейтральна	Висока	Нейтральна	Нейтральна
Нейронні мережі	Висока	Нейтральна	Низька	Висока
Дерева рішень	Низька	Висока	Висока	Висока
К-ближчого сусіда	Висока	Дуже низька	Висока	Нейтральна

Переваги використання методу побудови дерев

- Швидкий процес навчання
- Генерація правил в областях, де знання важко формалізувати
- Зрозуміла класифікаційна модель
- Висока точність прогнозу

Підходи до оцінки ефективності освіти у ВНЗ

- ▶ Аналіз успішності студентів та їх участі у різних активностях в університеті;
- ▶ Аналіз викладацьких методик та особливостей внутрішнього стану ВНЗ;
- ▶ Аналіз відповідності навчальної програми реаліям ринку праці.

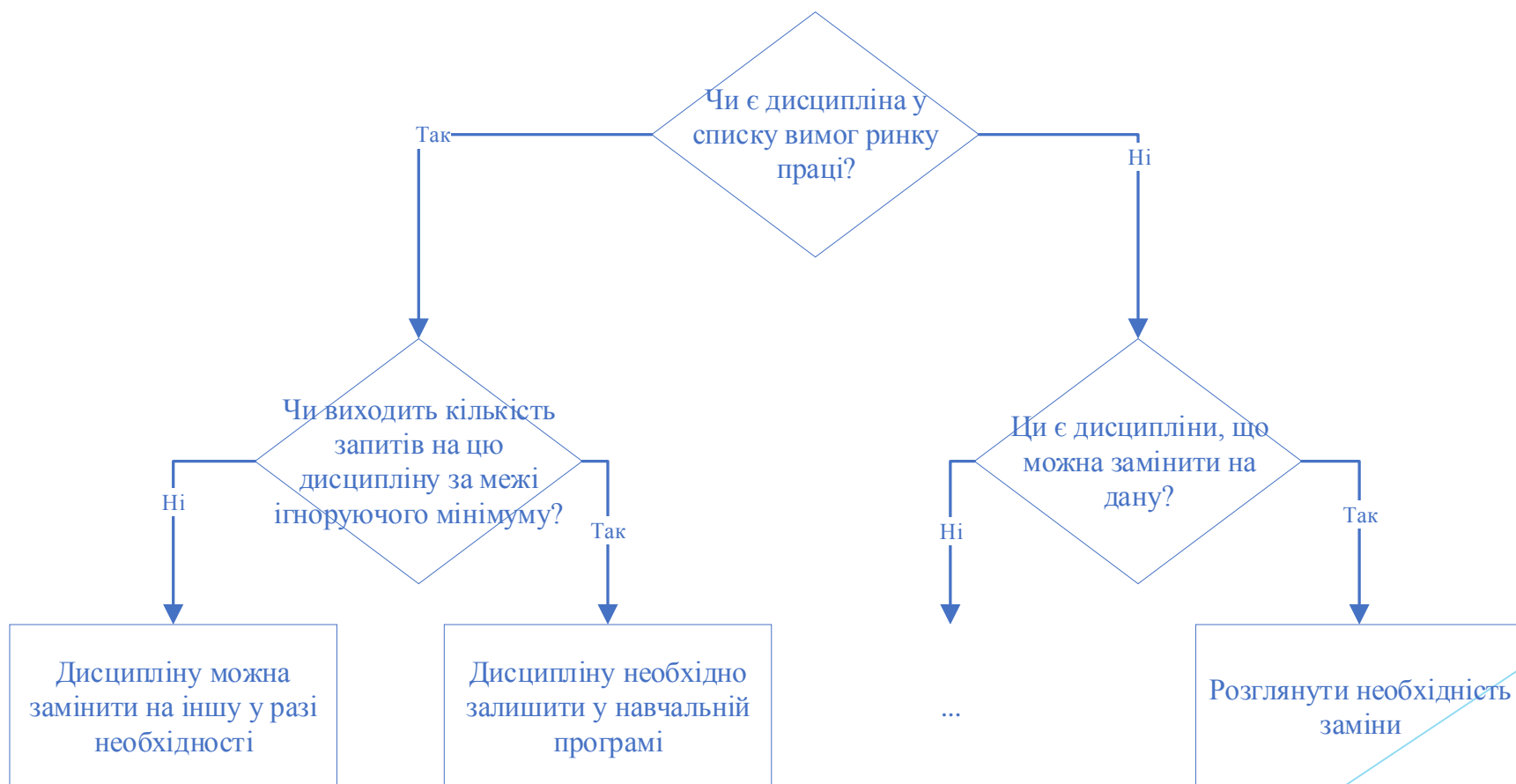
Переваги підходу оцінки начальної програми

- ▶ є найбільш зручним для реалізації методом дерев прийняття рішень
- ▶ Відсутній фактор суб'єктивної оцінки
- ▶ має порівняно низьку кількість параметрів.

Основні програмні модулі

- Модуль завантаження, збереження та редагування даних
- Модуль оцінки відповідності навчальної програми реаліям ринку праці
- Модуль візуалізації результатів

Приклад дерева рішень



Приклад роботи програми

Availability : General

Специфіка ВНЗ: Характер професії:

Університет:

Дані навчальної програми університету:

- ASP.NET / .NET
- Assembler
- C
- C#
- C++
- Delphi
- Erlang
- Java, J2EE, JSP
- Javascript/JScript
- MySQL
- Oracle
- PHP
- Unix / Linux
- Visual Basic

Дані щодо вимог ринку праці:

- Javascript/JScript (183) YES
- PHP (171) YES
- Java, J2EE, JSP (115) YES
- MySQL (115) YES
- C++ (86) YES
- C# (73) YES
- Python (69) NO
- ASP.NET / .NET (59) YES
- Unix / Linux (48) YES
- 1С-Битрикс (46) NO
- 1C (43) NO
- PL/SQL (42) NO
- Microsoft SQL (40) NO
- Ruby (38) NO
- Oracle (35) YES
- Ajax (30) NO
- C (29) YES
- Delphi (15) YES
- Objective - C (15) NO
- T - SQL (15) NO
- Perl (10) NO

Порівняти

Аналітика Розширений аналіз

Результати аналізу:

Число необхідних дисциплін : 10

Число надлишкових дисциплін : 3

Можливість заміщення : часткова

Візуалізація:

Кількість позицій :

Додати для візуалізації

Діаграма Таблиця

Завантажити навчальну програму

Завантажити дані ринку праці

Граничний відсоток:

Записів оброблено : 1339

Результати порівняння:

Загальна відповідність : 38%

Відповідність з урахуванням граничного відсотку : 55%

Рівень відповідності : середній/середній

Вихід

Приклад результатів оцінки

Результати аналізу

Специфіка ВНЗ : технічний Записів розглянуто : 1339

Характер професії : розробник

Університет : НТУУ

Необхідні дисципліни

- Python(69)
- Microsoft SQL(40)
- Ruby(38)
- 1С-Битрикс(36)
- 1С(33)
- PL/SQL(32)
- Ajax(30)
- Objective - C(15)
- T - SQL(15)

Надлишкові дисципліни

- Erlang(2)
- Visual Basic(2)
- Assembler(1)

Рекомендації :

Замінити дисципліни **на дисципліни**

- Erlang
- Visual Basic
- Assembler

- Python
- Microsoft SQL
- Ruby

Загальна відповідність у випадку заміни : 55% -> 70%

Рівень відповідності у випадку заміни : середній -> високий

Зберегти результат **Переглянути збережене**

Назад

Приклад числового вираження результатів

Назва ВНЗ (скорочено)	Відповід-ність (%)	Значення граничного коефіцієнту (%)	Відповід-ність з коефіцієнт-ом (%)	Відповідність після коригування алгоритмом (%)
НТУУ ім. Ігоря Сікорського	38	1	55	70
НАУ	30	1	50	67
НУ Києво-могилянська академія	45	1	70	70
КНУ імені Тараса Шевченка	36	1	53	78

Приклад візуалізації результатів порівняння



Висновки

- Проаналізовано придатність використання методів технології Data Mining у сфері освіти та відносна ефективність їх використання.
- Проаналізовано підходи до оцінки ефективності освіти у ВНЗ.
- Розроблено та протестовано програмний застосунок, що відображає ефективність освітнього процесу на основі аналізу ринку праці

Недоліки

- Відсутність інтегрованого алгоритму збору та фільтрації даних
- Залежність програми від джерел та обсягів даних
- Відсутність урахування окремих критеріїв, що можуть вплинути на кінцевий результат

Подальші задачі

- ▶ Розробка механізму збору та фільтрації інформації
- ▶ Ускладнення процедури оцінки, шляхом включення нових параметрів
- ▶ Інтеграція у більш комплексну систему

Дякую за увагу!