

УДК 378.1:005.6

[https://doi.org/10.33296/2707-0255-10\(19\)-08](https://doi.org/10.33296/2707-0255-10(19)-08)

<https://orcid.org/0000-0001-9879-0319>

БОНДАРЕНКО ТЕТЯНА

кандидат педагогічних наук, доцент
доцент кафедри інформаційних
комп'ютерних технологій і математики
Українська інженерно-педагогічна академія
(м. Харків)

ВИДІЛЕННЯ ЛАТЕНТНИХ ЗМІННИХ ІЗ BIG DATA В ІНСТИТУЦІЙНИХ ДОСЛІДЖЕННЯХ ІНЖЕНЕРІВ-ПЕДАГОГІВ ЯК ЕЛЕМЕНТ МОНІТОРИНГУ ПРОФЕСІЙНОЇ КОМПЕТЕНТНОСТІ ФАХІВЦІВ

Анотація. У статті описана методика «стиснення» інформації на основі виділення латентних змінних з Big Data з результатами опитувань. Особливість даних, які використовуються для стиснення інформації - це наявність різних типів відповідей в довільній текстовій формі, що не дозволяє використовувати існуючі методики, в яких оцінювання відповідей респондентів здійснюється тільки на основі шкали Лайкерта. На основі описаної методики можна перетворити будь-яку базу даних з результатами опитувань до виду придатного для статистичної обробки і забезпечити стиснення вихідного масиву даних до розмірів, які дозволяють проводити порівняльний аналіз результатів опитувань на тривалому проміжку часу. Побудована трирівнева модель індексу задоволеності студентів (ІЗС). Латентні змінні першого рівня цієї моделі деталізують склад латентних змінних другого рівня. Латентні змінні другого рівня визначають індекс ІЗС. Латентні змінні другого рівня це - ставлення, навчальний процес, взаємовідносини, умови. Латентні змінні першого рівня: ставлення до академії, ставлення до спеціальності, зміст освітніх програм, організація навчального процесу, практична підготовка, інформаційне обслуговування, взаємовідносини з адміністрацією, взаємовідносини з викладачами та співробітниками, взаємини зі студентами, умови проживання, харчування, дозвілля. На етапі фільтрування масиву даних з результатами опитувань з бази даних виключені питання, які не мають відношення до теми дослідження і питання, які були включені в анкети не постійно, а час від часу. На етапі перетворення даних і подання їх у вигляді придатному для оброблення виконувалося перетворення текстів відповідей в цифровий вигляд. Зі всіх питань анкети для перетворень були виділені: дихотомічні запитання, питання на основі шкали Лайкерта, питання з текстами відповідей у

довільній формі. У статті наведено результати перевірки методики на прикладі побудови індексу задоволеності студентів. Результати аналізу даних підтвердили достовірність і надійність отриманих результатів.

Ключові слова: інституційні дослідження, опитування студентів, індекс задоволеності студентів, прихована змінна, шкала питань, оцифрування анкети, перевірка сумісності відповідей, валідність і надійність опитування.

Вступ. Сьогодні заклади вищої освіти є масштабні, складно організовані, багатофункціональні системи, управління якими стає проблематичним за відсутності спеціальних досліджень, які дозволяють вибудовувати і модифікувати стратегію ЗВО в умовах конкуренції і гнучко реагувати на зовнішні зміни і запити зацікавлених сторін.

Такими спеціальними дослідженнями є інституційні дослідження (ІД), що досліджують вплив різних змін (таких, як глобалізація, демографічні зрушення, збільшення кількості студентів, швидка технологічна трансформація та ринкове середовище) на підтримку прийняття рішень та природу інституційних досліджень у вищій освіті. ІД - приклад областей знання, де поєднуються практичні та власні наукові інтереси з проблемою вищої освіти та управління університетами.

Напрями досліджень в рамках ІД численні і відповідають всім основним сферам університетської діяльності: дослідження процесу вступу абітурієнтів; вивчення процесу навчання студентів; вивчення освітніх програм і систем атестації студентів; дослідження професійної адаптації випускників; оцінювання професорсько-викладацького складу; оцінювання ефективності фінансового планування і розподілу фінансових коштів тощо. В процесі інституційних досліджень співробітники відділів ІД використовують результати власних внутрішньо університетських досліджень (опитування студентів, викладачів, допоміжного персоналу та інших зацікавлених осіб), університетську статистику, національну освітню статистику і результати загальнонаціональних досліджень університетів.

Одна з ключових проблем в рамках ІД як університет впливає на життя, розвиток, освіту студента? Це питання пов'язане з іншими: в чому полягають позитивні ефекти від вищої освіти і як вони можуть бути виміряні? Завдання дослідника полягає в пошуку комбінації «сприятливі умови - сприятливий результат», тобто визначенні тих чинників, які сприяють ефективності освітнього процесу.

Дослідження незмінно показують, що студенти є надійними джерелами інформації про якість викладання і найбільш корисним джерелом ідей про те, як його поліпшити. Опитування студентів також більш точні, ніж спостереження в класі, при прогнозуванні досягнень. Ніхто не зацікавлений в дієвості навчання більше, ніж студенти. Немає і кращих експертів по тому, як навчання сприймається передбачуваними бенефіціарами. Але тільки нещодавно практики прийшли до розуміння того, що студенти можуть бути важливим джерелом інформації про якість викладання і середовище навчання. Тому опитування студентів в межах ІД проводяться регулярно як на університетському, так і на національному та міжнародному рівнях. Прикладом таких досліджень є масштабний проект Національне опитування залученості студентів (НОЗС), в якому з початку 2000 року взяло участь понад 1300 університетів США і Канади.

Природно, що за весь період спостереження в світі накопичено значна кількість даних за різними опитуваннями студентів, проте дослідники при цьому використовують традиційний підхід до оцінювання результатів опитування. В основному це представлення відповідей респондентів у вигляді діаграм, на яких з кожного питання відображається процентне співвідношення різних варіантів відповідей або порівняння за елементами (як приклад можна привести річний звіт «Результати опитування щодо якості переддипломної освіти НОЗС за 2018 рік» [1]).

В Українській інженерно-педагогічній академії (УІПА) опитування майбутніх інженерів-педагогів проводяться з 2007 року. За цей час в

опитуваннях взяли участь понад 7000 студентів, в базі даних накопичено понад 500 тисяч відповідей з різних аспектів навчання і студентського життя. Основним методом аналізу даних результатів цих опитувань є порівняння за елементами, а також побудова графіків і діаграм, які ілюструють відповіді на окремі питання. Однак у зв'язку з великими обсягом аналізованих даних ці методи малоефективні і не забезпечують можливість порівняння результатів опитувань за роками і різними аспектам студентського життя. У зв'язку з чим постає актуальність даного дослідження.

Метою даної статті запропонувати та описати метод стиснення інформації, який дозволить проводити порівняльний аналіз наявних даних в різних розрізах.

Аналіз досліджень і публікацій. Для вилучення неочевидних, об'єктивних і корисних на практиці закономірностей з великих обсягів даних останнім часом широко використовуються технології Data Mining. Однак, за твердженням спеціалістів (наприклад, [2]) процес реалізації Data Mining на практиці виявляється більш складним, ніж очікується і практичне застосування даної технології для обробки результатів опитувань має певні обмеження [3; 4].

По-перше, результати Data Mining в великій мірі залежать від рівня підготовки даних і форми їх подання. Наприклад, обробка текстів інформації (а відповіді на запитання анкет часто представлені саме в такому вигляді) в технології Data Mining поки ще недостатньо розроблені.

По-друге, Data mining не може замінити аналітика. Вона всього лише дає йому потужний інструмент для полегшення і покращення його роботи і так як технологія Data Mining є мультидисциплінарною прикладною галуззю для проведення досліджень необхідно ще задіяти фахівців з різних областей, а також забезпечити їх якісну взаємодію.

По-третє, якість результатів роботи методів Data Mining достатньо складно перевірити. Математико-статистичні оцінки якості не гарантують

успішної інтерпретації результатів. Багато фахівців стверджують, що Data Mining-додатки можуть видавати велику кількість статистично недостовірних результатів. Щоб цього уникнути, необхідна перевірка адекватності отриманих моделей на тестових даних, що не завжди можливо. Додатковим обмеженням для використання цієї технології є висока вартість додатків Data Mining.

Таким чином, хоча ідея перекласти обробку великого масиву зібраних даних на Data Mining була дуже привабливою, нам довелося звернутися до традиційного методу редукції ("стиснення") соціальної інформації - методу індексів. Цей метод широко використовується в соціології для зведення соціальної інформації до єдиного індексу - зведеному числовому показнику, який отримано в результаті аналізу і узагальнення комбінації індикаторів (в нашому випадку питань в анкеті). Індекс слугує як єдиний показник властивостей латентної змінної (або деяких змінних), яка не підлягає безпосередньому виміру. При виборі латентної змінної дослідники виходять з того, що латентна змінна, не вимірювана змінна, впливає на відповіді ряду елементів опитування.

Як приклад використання індексів в соціологічних дослідженнях в освіті можна привести індекс освіти.

Організація Об'єднаних Націй щорічно публікує Індекс людського розвитку, який складається з індексу освіти, індексу ВВП і індексу очікуваної тривалості життя. Індекс освіти вимірює рівень освіти, і з 2010 року він вимірюється шляхом об'єднання середніх років навчання дорослих в школі з очікуваними роками навчання дітей, кожен з яких отримує 50% ваги. В інституційних дослідженнях широко використовується індекс задоволеності студентів (ІЗС) [5-9].

Виклад основного матеріалу. Для результатів опитувань, які були отримані в УПА, проблема побудови індексів визначається такими факторами. Дані, які розглядаються були зібрані на замовлення адміністрації академії, їх

збір планувався і здійснювався протягом тривалого часу для оцінювання ставлення студентів до різних аспектів навчання і діяльності академії відповідно до завдань конкретного дослідження. При такому підході досягається максимум відповідності зібраних первинних даних вимогам і цілям проведеного аналізу.

Для нашого дослідження ці дані є вторинними, оскільки зібрані для інших цілей. З одного боку, це дозволяє скоротити витрати на дорогий збір первинних даних, ці дані можуть використовуватися для повного або часткового вирішення поставленого завдання, оскільки дають дослідникам якусь відправну точку, будучи при цьому відносно дешевими і легкодоступними. Але, з іншого боку, недоліком вторинних даних є невідповідність наявної інформації поставленій меті дослідження. В даному випадку ця невідповідність визначається формою подання зібраної інформації. База даних з результатами опитувань інженерів-педагогів не відповідає вимогам, які необхідні для виділення латентних змінних при побудові індексів (наприклад, не всі індикатори вимірюються на інтервальній шкалі, шкали питання не однакові для всіх питань, в опитуванні використовуються дихотомічні питання тощо).

У зв'язку з цим ми запропонували методику виділення латентних змінних з баз даних з результатами опитувань, які містять різні типи питань з різною формою подання відповідей. Методика включає такі етапи: визначення складу латентних змінних, які можуть бути розраховані на основі наявного масиву даних, фільтрування масиву даних з результатами опитувань, перетворення даних і подання їх у вигляді придатному для обробки, розрахунок значень латентних змінних, аналіз отриманих результатів і оцінювання їх надійності.

На етапі визначення складу латентних змінних ми проаналізували вміст питань в анкетах і прийшли до висновку, що загальна спрямованість опитування - визначення ставлення студентів до навчання і середовища

навчання. В результаті аналізу та узагальнення комбінації питань в анкеті ми можемо вибрати склад латентних змінних і на їх підставі визначити ІЗС.

Огляд робіт, в яких розглядається проблема побудови ІЗС, показав, що єдиного підходу до оцінювання задоволеності студентів різними сторонами освітнього процесу поки не вироблено. Модель ІЗС в статті [5] - це структурна модель, заснована на припущенні, що задоволеність викликана деякими чинниками, такими як сприйняття якості, яка сприйняття цінності, очікування студентів і імідж фірми. Кожен фактор у моделі - це прихована конструкція, яка реалізується безліччю індикаторів. Дослідження [5] спрямовано на з'ясування впливу вартості навчання на залучення студентів до університету шляхом залучення сервісу якості, репутації університету та розумної вартості навчання для задоволення студентів, і студентської лояльності, яка в кінцевому підсумку вплине на конкурентну перевагу і фінансове виконання. У дослідженні [8] набір латентних змінних містить репутацію освітнього закладу, очікування студентів, активність студентів, якість сприйняття, цінність сприйняття, задоволеність студентів і підсумкову лояльність студентів.

На основі аналізу переліку питань в анкетах побудована трирівнева модель ІЗС. Латентні змінні першого рівня цієї моделі деталізують склад латентних змінних другого рівня. Латентні змінні другого рівня визначають індекс ІЗС. Латентні змінні другого рівня це Ставлення (1), Навчальний процес (2), Взаємовідносини (3), Умови (4). Латентні змінні першого рівня: Ставлення до академії (1.1), Ставлення до спеціальності (1.2), Зміст освітніх програм (2.1), Організація навчального процесу (2.2), Практична підготовка (2.3), Інформаційне обслуговування (2.4), Взаємовідносини з адміністрацією (3.1), Взаємовідносини з викладачами та співробітниками (3.2), Взаємини зі студентами (3.3), Умови проживання (4.1), Харчування (4.2), Дозвілля (4.3).

На етапі фільтрування масиву даних з результатами опитувань з бази даних були виключені питання, які не мають відношення до теми дослідження і

питання, які були включені в анкети не постійно, а час від часу. В результаті фільтрування в анкеті з понад 100 питань було відібрано 58.

Етап перетворення даних і подання їх у вигляді придатному для оброблення є найбільш трудомістким і відповідальним. Від його результатів багато в чому залежить надійність отриманої інформації і висновків, зроблених на її основі. На цьому етапі виконується перетворення текстів відповідей в цифровий вигляд. З 58 питань анкети для перетворень були виділені: дихотомічні запитання - 5, питання на основі 3-бальної шкали Лайкерта - 1, питання на основі 4-бальної шкали Лайкерта - 20, питання на основі 5-бальної шкали Лайкерта - 12, питання з текстами відповідей у довільній формі - 24.

Основне правило перетворення - для співставлення всі відповіді оцінюються на шкалі єдиного розміру від 0 до 1. Тому оцифрування відповідей питань за шкалою типу Лайкера не складало проблем.

Дихотомічні питання також дозволяють визначити ступінь згоди або незгоди студента з кожним судженням. Тільки на відміну від шкали Лайкерта ми маємо тільки дві рівні оцінки: «повністю згоден» (оцінка 1) і «повністю не згоден» (оцінка 0). Наприклад, якщо студент на питання «Ви б порадили своїм знайомим вступати до УПА?» відповідає «Так, порадив би», це свідчить про його позитивне ставлення до академії і ми оцінюємо таку відповідь в 1 бал.

Найбільшу складність при оцифрування відповідей склали питання з відповідями, які не можна розташувати за рейтинговою шкалою. Розглянемо як приклад оцифрування питання «Якби у Вас була можливість знову вибрати професію та навчальний заклад, Ви б повторили свій вибір?». В анкеті пропонуються наступні варіанти відповідей на це питання: 1. Так, повторив би, 2. Вибрав би іншу спеціальність в УПА, 3. Вибрав би цю ж спеціальність в іншому навчальному закладі, 4. Вибрав би іншу спеціальність і в іншому навчальному закладі. Перший варіант відповіді відповідає позитивній оцінці навчального закладу та обраної спеціальності, другий позитивну оцінку навчального закладу і негативною для обраної спеціальності, третій позитивну

оцінку спеціальності і негативною оцінкою навчального закладу, четвертий негативною оцінкою спеціальності і негативною оцінкою навчального закладу. На протилежних кінцях шкали розташовуються четвертий варіант відповіді (0 балів) і перший варіант відповіді (1 бал). Оцифрування варіантів відповідей, які викликають труднощі, виконувала група експертів з 5 чоловік. Другий варіант відповіді експерти запропонували оцінити в 0,6 бали, третій - в 0,3 (пояснення експертів - розмір навчального закладу більше розміру спеціальності). Таким чином, в разі даного дослідження виконується правило: чим краще ставиться респондент до елемента дослідження, тим вище буде сума його перетворених відповідей. При цьому середня сума балів за всіма відповідями знаходиться в діапазоні від 0 до 1 і легко переводиться у відсотки.

Для перевірки сумісності відповідей для кожного питання була розрахована кореляція між результатами відповідей на нього і рядком сум відповідей. Для більшості відповідей ця кореляція знаходиться в межах від 0,75 до 0,95 що означає, що ті, хто в цілому добре ставиться до об'єкта, позитивно оцінили дане твердження, і навпаки. Два питання, для яких значення кореляції було менше 0,5, з подальшого аналізу видалили в зв'язку з відсутністю зв'язку між загальним ставленням і оцінкою конкретного твердження. Для двох питань, пов'язаних з оцінкою спеціальності були отримані негативні значення кореляції, що свідчить про помилку експертів в оцінці позитивних факторів при оцифрування відповідей на дані питання. В оцифрування цих відповідей були внесені необхідні корективи.

Несподівані результати, які підтверджують надійність перетворених даних, були отримані при порівняльному аналізі підсумків опитувань з різних років. Так значення кореляцій між підсумковими даними за три останні роки виявилися близькими до 1. Анкетні опитування проводяться щорічно для всіх студентів академії і цей результат, з одного боку, підтверджує стійкість результатів, а, з іншого боку, свідчить про те, що протягом періоду навчання думка студентів про об'єкт дослідження істотно не змінюється. У зв'язку з цим

для порівняльного аналізу ми взяли дані з інтервалом в чотири роки (2011, 2014, 2018), який відповідає зміні поколінь студентів і, відповідно, дає можливість оцінити зміни в оцінюванні студентів про об'єкт дослідження. Для сусідніх років кореляція між даними опитувань дорівнює 0,72, а кореляція між даними опитувань в 2011 і в 2018 роках дорівнює 0,445. Таким чином, з'являється привід з'ясувати, що вплинуло на зміну ставлення студентів до об'єкта дослідження і в який бік змінилося це ставлення.

На першому етапі оброблення результатів була виконана оцінювання ступеню розкиду точок даних щодо їх середнього для кожного індикатора. Отримані значення стандартного відхилення знаходяться в межах від 0,006 (дані на досліджуваному періоді практично не змінювалися) до 0,21. На рис. 1 наведена зміна значень трьох індикаторів, які мають найбільший розкид значень за роками.

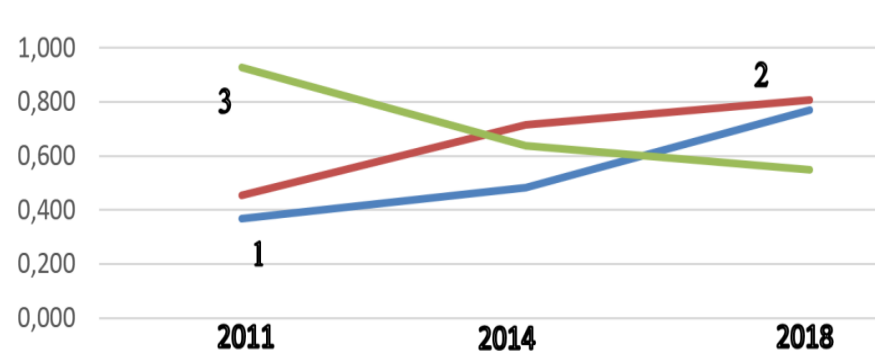


Рис. 1. Оцінювання ступеню розкиду точок даних щодо їх середнього значення за кожним показником

Найбільші зміни в оцінках студентів відносяться до питань роботи деканату (оцінка 0,37 - у 2011 році та 0,77 - у 2018 році), роботи кафедри (оцінка 0,46 - у 2011 році та 0,81 - у 2018 році) та продовження навчання в академії після отримання диплома бакалавра (оцінка 0,93 - в 2011 році та 0,55 - у 2018). У перших двох випадках спостерігається позитивна тенденція при зміні оцінки (збільшення в два рази), що говорить про прийняття адміністрацією академії відповідних заходів за результатами опитування. У третьому випадку спостерігається третя тенденція зниження кількості бажаючих продовжити

навчання в академії для отримання диплома магістра. Тут є ряд об'єктивних причин, не пов'язаних з рішеннями адміністрації: зміна правил прийому в магістратуру, зниження привабливості вищої освіти для молодих людей тощо. Незважаючи на це адміністрації академії слід звернути увагу на цю тенденцію та зі свого боку вжити упереджувальних заходів.

Ще одне питання, оцінювання відповідей, яке відрізняється майже в два рази, - це «Чи часто Ви відвідуєте сайт академії?». Це питання можна розглядати як перевірку, оскільки позитивна динаміка числа відвідувань сайту академії пояснюється об'єктивними причинами. Серед цих причин розвиток ІТ інфраструктури академії, підвищення ступеня оснащеності студентів засобами для роботи з ІТ технологіями тощо. Таким чином, відповіді на це питання є ще одним підтвердженням того, що отримані в результаті попереднього оброблення дані є достовірними. Результати розрахунку значень латентних змінних (ЛЗ) наведені в таблиці 1.

Таблиця 1

Результати розрахунку значень латентних змінних

Рівень 1		Рівень 2			Рівень 3	
Номер ЛЗ	Кількість індикаторів	Значення ЛЗ			Назва ЛЗ і її значення	Значення ІЗС
		2011	2014	2018		
1.1	3	0,66	0,67	0,67	2011 - 0,61 2014 - 0,63 2018 - 0,65	
1.2	5	0,62	0,61	0,58		
2.1	3	0,74	0,71	0,64		
2.2	4	0,59	0,64	0,71		
2.3	3	0,73	0,62	0,65		
2.4	4	0,63	0,72	0,79		
3.1	6	0,63	0,67	0,70		
3.2	7	0,66	0,69	0,73		
3.3	6	0,70	0,75	0,68		
4.1	5	0,39	0,40	0,44		
4.2	4	0,42	0,42	0,51		
4.3	6	0,61	0,65	0,69		

Перше, на що слід звернути увагу, це те, що розглянуті нами вище дворазові зміни оцінок студентів на кордонах періоду спостереження (2011 - 2018 роки) для окремих індикаторів при розрахунку латентних змінних компенсуються змінами інших індикаторів, які входять в групу оцінок. В результаті ми отримуємо незначні коливання зміни латентних змінних. Підсумкова оцінка ІЗС зростає за роками, але це зростання незначне в межах похибки вимірювання. Аналіз результатів показує, що стримуючим фактором для зростання оцінки задоволеності студентів є умови навчання (проживання і харчування).

Висновок. На основі описаної методики виділення латентних змінних можна перетворити будь-яку базу даних з результатами опитувань до виду придатного для статистичної обробки і забезпечити стиснення вихідного масиву даних до розмірів, які дозволяють проводити порівняльний аналіз результатів опитувань на тривалому проміжку часу. Аналіз отриманих латентних змінних дозволить керівництву ЗВО оцінювати запити майбутніх інженерів-педагогів, постійно відстежувати тенденції в зміні пропонованих вимог, виявляти слабкі місця в організації діяльності ЗВО і своєчасно проектувати і здійснювати відповідні коригувальні та запобіжні дії.

Новизна методики полягає в тому, що аналізовані опитування можуть містити різні типи питань з різною формою подання відповідей на відміну від наявних методик, в яких оцінювання відповідей респондентів здійснюється тільки на основі шкали Лайкерта. Результати аналізу даних підтвердили достовірність і надійність отриманих результатів, однак для остаточної перевірки працездатності методики слід провести оцінювання ІЗС на основі традиційної методики і порівняти отримані результати з результатами, які наведені вище. Ця робота заплановано на наступний навчальний рік і, сподіваємося, що результати цієї перевірки будуть представлені вже в наступному році.

Список використаних джерел

1. Survey Findings on the Quality of Undergraduate Education. NSSE Annual Results. 2018. URL: <http://bit.ly/3iZCbqa>.
2. Shmueli G., Peter C. Bruce P. C, Yahav I., Patel N. R., Kenneth C. Lichtendahl K. C Jr. Data Mining for Business Analytics: Concepts, Techniques, and Applications in R. John Wiley & Sons, 2017. 577 p.
3. Attewell P., Monaghan D.: Data Mining for the Social Sciences: An Introduction. University of California Press, 2015. 264 p.
4. Mosyagin A.: Using the data mining methodology for solving the problems of social data processing. *Scientific journal "Monitoring of Public Opinion"*. 2015. #2 (126). pp. 143-147.
5. Temizer L., Turkyilmazb A. Implementation of student satisfaction index model in higher education institutions. *Procedia - Social and Behavioral Sciences*. 2012. #46. pp. 3802–3806.
6. Adityo O.H., Sofianti T.D., Saraswati T. Students' Satisfaction Index: A Case Study in SGU. *Proceedings of the International Conference on Innovation, Entrepreneurship and Technology*. BSD City, Indonesia. 2015. pp. 293-300.
7. Adnan Al-R., Mohamed Al-F., Tarek A., Mun S., Hosny H. Measuring Student Satisfaction with Performance Enhancement Activities: Evidence from Business Education. *International Journal of Information and Education Technology*, 2016. Vol. 6. #10. pp. 741-754.
8. Duarte P.O., Raposo M.B., Alves H.B. Using a Satisfaction Index to Compare Students' Satisfaction During and After Higher Education Service Consumption. *Tertiary Education and Management*. 2015. Vol. 18. # 1. pp. 17-40.
9. Dharmayanti D., Samuel H., Devie. The Students Satisfaction, Student Loyalty, Competitive Advantage And Financial Sustainability On Private Universities In Surabaya. *Advances in Social Sciences Research Journal*. 2018. Vol. 5. #10. pp. 266-275.

References

1. Survey Findings on the Quality of Undergraduate Education. NSSE Annual Results. 2018. URL: <http://bit.ly/3iZCbqa>.
2. Shmueli G., Peter C. Bruce P. C, Yahav I., Patel N. R., Kenneth C. Lichtendahl K. C Jr. Data Mining for Business Analytics: Concepts, Techniques, and Applications in R. John Wiley & Sons, 2017. 577 p.
3. Attewell P., Monaghan D.: Data Mining for the Social Sciences: An Introduction. University of California Press, 2015. 264 p.
4. Mosyagin A.: Using the data mining methodology for solving the problems of social data processing. *Scientific journal "Monitoring of Public Opinion"*. 2015. #2 (126). pp. 143-147.
5. Temizer L., Turkyilmazb A. Implementation of student satisfaction index model in higher education institutions. *Procedia - Social and Behavioral Sciences*. 2012. #46. pp. 3802–3806.

6. Adityo O.H., Sofianti T.D., Saraswati T. Students' Satisfaction Index: A Case Study in SGU. *Proceedings of the International Conference on Innovation, Entrepreneurship and Technology*. BSD City, Indonesia. 2015. pp. 293-300.
7. Adnan Al-R., Mohamed Al-F., Tarek A., Mun S., Hosny H. Measuring Student Satisfaction with Performance Enhancement Activities: Evidence from Business Education. *International Journal of Information and Education Technology*, 2016. Vol. 6. #10. pp. 741-754.
8. Duarte P.O., Raposo M.B., Alves H.B. Using a Satisfaction Index to Compare Students' Satisfaction During and After Higher Education Service Consumption. *Tertiary Education and Management*. 2015. Vol. 18. # 1. pp. 17-40.
9. Dharmayanti D., Samuel H., Devie. The Students Satisfaction, Student Loyalty, Competitive Advantage And Financial Sustainability On Private Universities In Surabaya. *Advances in Social Sciences Research Journal*. 2018. Vol. 5. #10. pp. 266-275.

БОНДАРЕНКО ТАТЬЯНА

кандидат педагогических наук, доцент
доцент кафедры информационных
компьютерных технологий
и математики
Украинская инженерно-
педагогическая академия,
г. Харьков, Украина

ВЫДЕЛЕНИЕ ЛАТЕНТНЫХ ПЕРЕМЕННЫХ С BIG DATA В ИНСТИТУЦИОННЫХ ИССЛЕДОВАНИЯХ ИНЖЕНЕРОВ-ПЕДАГОГОВ КАК ЭЛЕМЕНТ МОНИТОРИНГА ПРОФЕССИОНАЛЬНОЙ КОМПЕТЕНТНОСТИ СПЕЦИАЛИСТОВ

Аннотация. В статье описана методика «сжатия» информации на основе выделения латентных переменных с Big Data результатам опросов. Особенность данных, используемых для сжатия информации это наличие различных типов ответов в произвольной текстовой форме не позволяет использовать существующие методики, в которых оценка ответов респондентов осуществляется только на основе шкалы Лайкерта. На основе описанной методики можно превратить любую базу данных с результатами опросов к виду пригодного для статистической обработки и обеспечить сжатие исходного массива данных до размеров, которые позволяют проводить сравнительный анализ результатов опросов на длительном промежутке времени. Построена трехуровневая модель индекса удовлетворенности студентов (ИУС). Латентные переменные первого уровня этой модели детализируют состав латентных переменных второго уровня. Латентные переменные второго уровня определяют индекс ИУС. Латентные переменные второго уровня - это

отношение, учебный процесс, взаимоотношения, условия. Латентные переменные первого уровня: отношение к академии, отношение к специальности, содержание образовательных программ, организация учебного процесса, практическая подготовка, информационное обслуживание, взаимоотношения с администрацией, взаимоотношения с преподавателями и сотрудниками, взаимоотношения со студентами, условия проживания, питания, досуга. На этапе фильтрации массива данных с результатами опросов из базы данных исключены вопросы, не имеющие отношения к теме исследования и вопросы, которые были включены в анкеты не постоянно, а время от времени. На этапе преобразования данных и представления их в виде пригодном для обработки выполнялось преобразование текстов ответов в цифровой вид. По всем вопросам анкеты для преобразований были выделены: дихотомические вопрос, вопрос на основе шкалы Лайкерта, вопрос с текстами ответов в произвольной форме. В статье приведены результаты проверки методики на примере построения индекса удовлетворенности студентов. Результаты анализа данных подтвердили достоверность и надежность полученных результатов.

Ключевые слова: институциональные исследования, опросы студентов, индекс удовлетворенности студентов, латентная переменная, шкала вопросов, оцифровка анкеты, проверка совместимости ответов, валидность и надежность опроса.

BONDARENKO TETIANA
PhD in Pedagogical Sciences,
Associate Professor, Associate
Professor Department of Information
Computer Technology and Mathematics
Ukrainian Engineering
Pedagogics Academy
Kharkiv, Ukraine

DETECTION OF LATENT VARIABLES FROM BIG DATA IN INSTITUTIONAL RESEARCH OF ENGINEERS TEACHERS AS AN ELEMENT OF MONITORING OF PROFESSIONAL COMPETENCE OF SPECIALISTS

Abstract. The article describes a method of "compression" of information based on the allocation of latent variables with Big Data poll results. The peculiarity of the data used to compress information is the presence of various types of answers in free text form does not allow the use of existing methods, in which the assessment of respondents' answers is carried out only on the basis of the Likert scale. On the basis of the described methodology, it is possible to transform any database with survey results into a form suitable for statistical processing and provide compression of the original data array to sizes that allow comparative analysis of survey results

over a long period of time. A three-level model of the student satisfaction index (IUS) has been built. The first-level latent variables of this model detail the composition of the second-level latent variables. Latent variables of the second level determine the ICS index. Latent variables of the second level are attitudes, educational process, relationships, conditions. Latent variables of the first level: attitude to the academy, attitude to the specialty, content of educational programs, organization of the educational process, practical training, information services, relationships with the administration, relationships with teachers and staff, relationships with students, living conditions, food, leisure. At the stage of filtering the data array with the results of the polls from the database, questions that are not related to the research topic and questions that were included in the questionnaires not constantly, but from time to time were excluded from the database. At the stage of data transformation and their presentation in a form suitable for processing, the response texts were converted into digital form. For all the questions, the questionnaires for transformations were highlighted: a dichotomous question, a question based on a Likert scale, a question with free-form answers. The article presents the results of testing the methodology on the example of constructing an index of student satisfaction. The results of data analysis confirmed the validity and reliability of the results.

Key words: institutional research, student surveys, student satisfaction index, latent variable, scale of questions, digitization of the questionnaire, checking the compatibility of answers, validity and reliability of the survey.