УДК 005.311:004.6

ЗАСТОСУВАННЯ МЕТОДІВ АНАЛІЗУ ДАНИХ ДЛЯ ОЦІНЮВАННЯ РЕЗУЛЬТАТІВ НАВЧАННЯ ЗА ПІДСУМКАМИ СКЛАДАННЯ НМТ З МАТЕМАТИКИ

Гибкіна Н.В., Шарай К.В.

Харківський національний університет радіоелектроніки, Харків

Сучасний світ висуває низку серйозних викликів системі освіти загалом, і системі шкільної освіти зокрема. В умовах стрімкого технічного прогресу та цифровізації процесів у повсякденному житті та професійній діяльності навчальні програми та класичні методики викладання швидко застарівають. Загальнодоступність сервісів, заснованих на штучному інтелекті, підміняє процес освоєння інформації та отримання знань споживанням матеріалів у вигляді готових відповідей на конкретні (часто некоректно сформульовані) запитання, як правило, без критичного осмислення та аналізу. Військові дії та спричинені ними кризи у всіх сферах життя руйнують сформовану систему шкільної освіти, що тягне за собою підсилення проблем у сфері вищої освіти, пов'язаних з нестачею абітурієнтів та низьким рівнем їх знань, і тим самим у найближчі роки може призвести до суттєвого недоліку кваліфікованої робочої сили, яка відіграватиме ключову роль у повоєнному відновленні та реконструкції країни.

Одним з багатьох факторів, який слід враховувати при організації якісного освітнього процесу у закладах вищої освіти (ЗВО), є рівень знань, які майбутні здобувачі вищої освіти набувають протягом навчання у закладах середньої освіти (ЗСО). Виходячи з цього, актуальним є всебічний контроль за якістю освітнього процесу, застосування сучасних методик аналізу до оцінки результатів навчання та отримання висновків щодо слабких та сильних сторін окремих ЗСО з метою підтримки належного рівня процесу надання знань. Необхідну для такого аналізу інформацію можуть дати результати незалежного мультипредметного тесту, які складають випускники

для можливості вступу у ЗВО.

Розглянемо методи оцінювання якості рівня набутих знань та їх динаміки на прикладі результатів НМТ з математики 2024 року м. Харкова. Основою аналізу є деперсоналізовані дані про кожного випускника, включаючи особисту інформацію, місце розташування та тип ЗСО, в якому він навчався, дату, місце та результат складання ним тесту НМТ. Підготовчий етап аналізу полягає у виділенні з цих даних необхідних для аналізу ознак: назва та тип ЗСО, бал складання НМТ з математики; групування даних за окремими ЗСО та обчислення загальної кількості їх випускників, що проходили тестування. Для інтелектуального аналізу використовуються набір даних $\{\vec{x}^{(i)}\}_{i=1}^{n}$, де n – кількість ЗСО, $\vec{x} = (x_1, \dots, x_m)$ – вектор стандартизованих ознак кожного ЗСО (кількість ознак $m = 7$), що містить інформацію по цьому ЗСО щодо відсотків випускників: які зареєструвалися на НМТ, але не писали його; набрали менше 100 балів; склали НМТ на 100 – 120 балів (низькі), 120 – 140 балів (нижче середнього), 140 – 160 балів (середні), 160 – 180 балів (вище середнього), 180 –200 балів (високі) відповідно.

Метою аналізу є виділення серед розглядуваних ЗСО груп закладів зі схожими тенденціями у результатах складання НМТ. Для цього пропонується використовувати методи неієрархічної кластеризації k-means та k-medians. Ідея неієрархічної кластеризації полягає у виділенні з множини об’єктів підготовленого набору даних k об’єктів, що призначаються центрами кластерів, і розбитті решти об’єктів з $\{\vec{x}^{(i)}\}_{i=1}^{n}$ по цим кластерам, виходячи з того, до якого з центрів якнайближче знаходиться кожен об’єкт. Після такого розбиття центри кластерів оновлюються з врахуванням їх вмісту і оцінюється відстань об’єктів набору даних до нових центрів. Якщо виявляється, що при цьому деякий об’єкт стає ближчим до центру чужого кластера, він переноситься до нього. Ця ітераційна процедура повторюється до того моменту, коли склад кластерів перестає змінюватись або виконана максимально припустима кількість ітерацій [1].

Центром кожного кластера у методі k-means є центроїд, тобто точка $\vec{\mu} = (\mu_1, \dots, \mu_m)$ з координатами $\mu_i = \frac{1}{n_c} \sum_{r=1}^{n_c} x_i^{(r)}$, $i = 1, \dots, m$, де сума обчислюється тільки за тими об'єктами $\vec{x}$, що належать до даного кластера, а n_c – кількість об'єктів у ньому, $c = 1, \dots, k$. У методі k-medians центром є медоїд, тобто той з об'єктів $\vec{x}$ цього кластера, для якого загальна сума відстаней до всіх інших об'єктів у кластері мінімальна.

Для обчислення близькості між об'єктами та центрами кластерів використовується евклідова метрика $d(\vec{x}^{(i)}, \vec{x}^{(j)}) = \sqrt{\sum_{r=1}^{m} \left(x_r^{(i)} - x_r^{(j)}\right)^2}$, де $\vec{x}^{(i)}, \vec{x}^{(j)}$ – об'єкти, між якими обчислюється відстань.

Результатом неієрархічної кластеризації є розбиття набору ЗСО з результатами НМТ з математики у 2024 році по 5 кластерам. Кластеризація набору даних ЗСО дозволила виділити деякі тенденції у розподілі балів НМТ з математики. До кластера 1 увійшли ЗСО із загальним стабільно низьким рівнем балів НМТ і з незначною часткою тих, хто отримав оцінки на середньому рівні чи вище середнього, високих результатів не було взагалі. Цей кластер містить 22% аналізованих ЗСО. Виходячи з цього можливим є висновок про низьку якість навчання та/або рівень підготовки учнів, відсутність особистої зацікавленості у переважної кількості учнів цих ЗСО щодо вивчення дисципліни. Є незначна частка тих, результати яких середні. До кластера 2 увійшли ЗСО, у яких близько половини випускників не з'явились на тестування або отримали менше 100 балів. Решта продемонстрували середні, нижче середнього та низькі результати, причому переважають низькі та середні. Цей кластер містить 22% аналізованих ЗСО. Можливим є висновок про те, що у ЗСО цього кластера не приділяється належної уваги вивченню математики, переважна більшість учнів не мають мотивації та, можливо, наявні організаційні проблеми. При цьому ті, хто приділяє увагу навчанню, демонструють відносно непогані результати на середньому рівні. До кластера 3 увійшли ЗСО, які можна вважати закладами з стабільним середнім рівнем викладання математики та рівнем учнівської

підготовки, з присутністю відносно незначної частки «сильних» та «слабких» учнів. Цей кластер містить 38% аналізованих ЗСО. Кластер 4 містить ЗСО, у яких математика викладається на низькому рівні та/або більшість учнів мають невисокий рівень загальної підготовки, але незначна частина учнів отримали результати вище середнього або високі. До цього кластера увійшли 10% аналізованих ЗСО. Порівняно з кластером 1 кластер 4 є кращим майже за всіма рівнями оцінок, зокрема відсоток низьких результатів значно менший, ніж результатів нижче середнього; більший відсоток тих, хто отримав результати вище середнього, та майже ніхто не продемонстрував результат менше 100 балів. ЗСО цього кластера можуть бути перспективними для підвищення рівня викладання математики та зацікавленості учнів у її вивченні. Кластер 5 включає ЗСО, у яких приділяється велика увага вивченню математики та учні мають високий рівень знань, основна маса учнів демонструє прийнятний рівень знань. Результати нижче середнього та гірші за них продемонструвала лише незначна частка випускників цих ЗСО, а результати середні, вищі за середні та високі – найвищі порівняно з усіма кластерами. До цього кластера увійшли 17% аналізованих ЗСО.

Додатковим елементом аналізу може бути застосування методів зниження розмірності для візуального відображення результатів кластеризації. що дозволить наочно оцінити схожість (близькість у сенсі обраної метрики) ЗСО у межах кожного кластера.

Отримані результати можуть бути корисними при розробці стратегій підвищення рівня викладання аналізованих дисциплін з урахуванням особливостей, притаманних конкретним ЗСО, при аналізі загального рівня підготовки випускників та виявленні недоліків у організації та реалізації систем оцінювання.

Література:

1. Харченко В. О. Основи машинного навчання. Суми, Україна: Сумський державний університет, 2023.