

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Національний університет «Львівська політехніка»

Кваліфікаційна наукова
праця на правах рукопису

ЗУБ ХРИСТИНА ВІТАЛІЇВНА

УДК 004:378.091.212.2(043.5)

ДИСЕРТАЦІЯ

**ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ
ВСТУПНИКІВ ЗВО УКРАЇНИ**

126 Інформаційні системи та технології

12 – Інформаційні технології

Подається на здобуття наукового ступеня доктора філософії

Дисертація містить результати власних досліджень. Використання ідей,
результатів і текстів інших авторів мають посилання на відповідне джерело.

Х.В. Зуб

Науковий керівник:
Жежнич Павло Іванович
доктор технічних наук, професор

Львів – 2024

АНОТАЦІЯ

Зуб Х. В. Інформаційна технологія підтримки прийняття рішень вступників ЗВО України. – Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття ступеня доктора філософії за спеціальністю 126 «Інформаційні системи та технології» (12 – «Інформаційні технології»). – Національний університет «Львівська політехніка». Львів, 2024.

У дисертаційній роботі розв’язано актуальну науково-прикладну задачу в галузі інформаційних систем та технологій – підвищення рівня обізнаності вступників закладів вищої освіти (ЗВО) України щодо вибору освітніх програм відповідно до їхніх інтересів на основі розроблення нових та вдосконалення існуючих моделей, методів та інформаційних технологій підтримки прийняття рішень вибору освітніх програм з урахуванням особливостей процесу вступу до ЗВО України.

Дисертаційна робота складається зі вступу, чотирьох розділів, висновків, списку літературних джерел та додатків.

У першому розділі розглянуто основні поняття та засади проведення вступної кампанії ЗВО України. Проаналізовано поняття результативності вступної кампанії з погляду ЗВО та вступників. Проведено аналіз особливостей та принципів, подано атрибути та фази процесу прийняття рішень вступників у різні періоди вступної кампанії, інформаційні потреби вступника як особи, що приймає рішення, підтверджено складність та невизначеність цього процесу для вступника та визначено основні задачі, ефективного розв’язання яких потребує розроблення інформаційних систем підтримки прийняття рішень.

Аналіз сучасних відкритих веб-сервісів різних типів свідчить про актуальність розроблення додаткових інформаційних технологій для підтримки прийняття рішень вступників. Враховуючи специфіку процесу вступу до ЗВО України, практичний досвід закордонних університетів та використання ними комерційного програмного забезпечення не може бути повністю перенесене й застосоване у відповідних контекстах.

Враховуючи труднощі, які постають перед вступником, було проаналізовано низку наукових досліджень вітчизняних та закордонних авторів, що дало змогу виокремити особливості розроблення ефективних систем підтримки прийняття рішень вступників ЗВО. Результати аналізу вказують на важливість використання інтелектуалізованих компонентів для підтримки прийняття таких рішень.

Декомпозиція цілі та визначення основних інформаційних потреб вступника дали змогу виокремити дві основні задачі: вибір освітньої програми з можливих альтернатив, що відповідають інтересам вступника та передбачення успішності його вступу. Ефективне розв'язання таких задач можливе за умови застосування інформаційної технології на основі розроблення нових та вдосконалення існуючих моделей, методів і засобів підтримки прийняття рішень вступників ЗВО.

У другому розділі розроблено модель інформаційної системи підтримки прийняття рішень вступників ЗВО яка, на відміну від існуючих, використовує поєднання методу надання рекомендацій та удосконаленого методу передбачення успішності вступу, що підвищує рівень обізнаності вступників ЗВО України під час вибору освітніх програм відповідно до їхніх інтересів.

Обґрунтовано вибір методів надання рекомендацій вступникам і розроблено модель сумісного використання колаборативної фільтрації та фільтрації на основі вмісту для надання рекомендацій вступникам ЗВО альтернативних освітніх програм. Таке поєднання дасть змогу подолати обмеження індивідуальних методів рекомендацій та розширити сформовані результати відповідно до потреб вступника.

Розроблено технології збору та попереднього опрацювання даних для реалізації колаборативної фільтрації на основі архівних даних вступної кампанії та освітніх даних студентів, фільтрації на основі вмісту з використанням даних описів освітніх програм та методу передбачення успішності вступу до ЗВО України з використанням архівних даних вступної кампанії.

Обґрунтовано вибір методів машинного навчання для розв'язання задачі бінарної класифікації з метою передбачення успішності вступу до ЗВО України. Враховуючи особливості зібраного набору даних, а саме його незбалансованість, та їхній вплив на ефективність роботи засобів штучного інтелекту для розв'язання задачі класифікації, виконано вибір та обґрунтування критеріїв оцінювання ефективності роботи класифікаторів на основі машинного навчання. Це дало змогу виконати коректний та надійний аналіз а також інтерпретацію отриманих результатів.

У третьому розділі розроблено компоненти інформаційної системи підтримки прийняття вступників ЗВО, серед яких – метод надання рекомендацій альтернативних освітніх програм та метод передбачення успішності вступу до ЗВО.

В основу методу надання рекомендацій покладено сумісне використання колаборативної фільтрації (з використанням архівних даних вступної кампанії та освітніх даних), та фільтрації на основі вмісту (з використанням кластерів ключових слів освітніх програм). Для реалізації колаборативної фільтрації виконано опрацювання архівних даних вступної кампанії для визначення подібності інтересів вступників. Формування таких архівних даних виконувалось із використанням розробленої процедури розрахунку показника задоволеності студентів (зі значенням вище середнього). В його основі – процедура ранжування суми зважених значень таких критеріїв: успішність у навчанні, наукова діяльність та статус продовження навчання в магістратурі. Такий підхід дає змогу надати більш релевантні рекомендації вступнику щодо освітньої програми. Відбір ключових слів на основі описів освітніх програм та формування кластерів таких слів дали змогу розширити перелік рекомендацій освітніх програм методом фільтрації на основі вмісту згідно з інтересами вступників.

Виконано попереднє опрацювання архівних даних вступної кампанії задля їхнього подальшого використання методами машинного навчання під час

розв'язання задачі передбачення успішності вступу до ЗВО. Враховуючи переваги та недоліки різних методів визначення важливих ознак досліджуваного набору даних, було обґрунтовано використання деяких із них для розв'язання цієї задачі. Обґрунтовано необхідність вдосконалення існуючого інструментарію машинного навчання для розв'язання задачі передбачення успішності вступу до ЗВО за рахунок проведеного аналізу ефективності роботи різних класів методів машинного навчання та штучних нейронних мереж.

Удосконалено двоетапний метод класифікації даних для підвищення ефективності розв'язання задачі бінарної класифікації з метою зменшення похибки передбачення успішності вступу до ЗВО. Він ґрунтується на заміні класифікатора на основі машини опорних векторів на другому етапі методу на стекінгове об'єднання чотирьох класифікаторів на основі машини опорних векторів з різними ядрами та агрегуванням результатів їхньої роботи із використанням логістичної регресії. Подано структурно-функціональну схему роботи удосконаленого методу, описано алгоритми його навчання та застосування.

У четвертому розділі розроблено архітектуру інформаційної системи підтримки прийняття рішень вступників ЗВО України. Виконано оцінювання результатів моделювання компонент надання рекомендацій та передбачення успішності вступу до ЗВО.

Розроблена архітектура інформаційної системи підтримки прийняття рішень передбачає введення агентів попередньої обробки даних, що забезпечить можливість подальшого аналізу даних та їхнього використання в роботі інформаційної системи. Функціонування таких агентів передбачає проведення процесів попередньої обробки даних, що стосуються розрахунку показника задоволеності студента, а також відбір архівних даних вступників, формування ключових слів освітніх програм та їхніх кластерів, та створення набору атрибутів для передбачення успішності вступу.

Результати сумісного використання фільтрації на основі вмісту та колаборативної фільтрації розширюють перелік рекомендованих освітніх програм, які відповідають інтересам вступника ЗВО, що підвищує рівень його обізнаності щодо пропонованих ЗВО освітніх програм у межах вступної кампанії.

Проведене моделювання наявних інструментів машинного навчання з різних класів продемонструвало недостатню точність результатів його роботи, що зумовило необхідність удосконалення існуючого двоетапного методу машинного навчання для розв'язання поставленої задачі.

Виконано порівняння результатів роботи вдосконаленого двоетапного методу класифікації даних та існуючих методів із різних класів, результати якого показали підвищення точності вдосконаленого методу на 2,9 % (на основі F1-міри) під час розв'язання задачі передбачення успішності вступу до ЗВО. Враховуючи невелике значення похибки роботи останнього (1,8 % на основі F1-міри), обґрунтовано можливість його використання як компоненти інформаційної системи підтримки прийняття рішень з метою передбачення успішності вступу до ЗВО України.

Ключові слова: інформаційна технологія, інформаційна система, інтелектуальна інформаційна система, заклад вищої освіти, вступна кампанія, вступник, прийняття рішень, інформаційна система підтримки прийняття рішень, рекомендаційна система, колаборативна фільтрація, фільтрація на основі вмісту, машинне навчання, ієрархічні класифікатори, ансамблеві методи машинного навчання, штучна нейронна мережа.

ABSTRACT

Zub K. Information technology for supporting the decision-making of applicants to higher education institutions of Ukraine.

Dissertation for the Doctor of Philosophy degree in the speciality 126 “Information Systems and Technologies” (12 – “Information Technologies”) – Lviv Polytechnic National University, Lviv, 2024.

The dissertation solves an urgent scientific and applied problem in the field of information systems and technologies – expanding the awareness of applicants to Ukrainian higher education institutions to choose educational programs by their interests by building new and improving existing models, methods and information technologies to support decision-making on the choice of academic programs, taking into account the peculiarities of the process of admission to Ukrainian higher education institutions.

The dissertation consists of an introduction, four chapters, conclusions, a list of references and appendices.

The first chapter delves into the fundamental concepts and principles of conducting the admission campaign in higher education institutions in Ukraine. It thoroughly examines the notion of the effectiveness of the admission campaign from the perspectives of both higher education institutions and applicants. The analysis of the intricacies and principles of the decision-making process of applicants in various stages of the admission campaign, the attributes, phases, and information constraints for the applicant as a decision-maker, underscores the complexity and uncertainty of this process for the applicant. This complexity necessitates the development of information systems for decision support.

A scrutiny of diverse contemporary open web services underscores the uniqueness and relevance of developing additional information technologies to aid applicants' decision-making. Given the distinctiveness of the admission process to higher education institutions in Ukraine, the practical experience of foreign universities

and the potential use of commercial software cannot be directly transferred and applied in the relevant contexts. This highlights the originality and novelty of the research. Therefore, given the difficulties faced by applicants, several scientific studies by domestic and foreign authors were analysed, which made it possible to identify the features of building effective decision support systems for university applicants. The results of the analysis indicate the importance of using intellectualised components to support such decision-making.

The decomposition of goals and identification of the applicant's leading information constraints allowed us to identify two main tasks: choosing an educational program from possible alternatives that meet the applicant's interests and predicting the success of their admission. A practical solution to such tasks is possible if information technology is developed based on the development of new and improved existing models, methods, and tools to support the decision-making of university applicants.

The second chapter develops a model of an information system for supporting the decision-making of applicants to higher education institutions, which, unlike the existing ones, uses a combination of the method of providing recommendations and an improved method of predicting the success of admission, which expands the awareness of applicants to Ukrainian higher education institutions regarding the choice of educational programs by their interests.

The choice of methods for providing recommendations to applicants is substantiated, and a model for the joint use of collaborative filtering and content-based filtering to provide recommendations to applicants for alternative educational programs is developed. This will overcome individual recommendation methods' limitations and expand the results generated by the applicant's needs.

Data collection and pre-processing technologies have been developed to implement collaborative filtering based on archival data of the admission campaign and students' educational data, content-based filtering using data from educational program descriptions, and prediction of admission success to Ukrainian higher education institutions using archival data of the admission campaign.

The choice of machine learning methods to solve the binary classification problem and predict the success of admission to Ukrainian higher education institutions is substantiated. Considering the peculiarities of the collected data set, namely its imbalance and its impact on the efficiency of artificial intelligence tools for solving the classification problem, the criteria for evaluating the efficiency of machine learning-based classifiers were selected and justified. This allowed for a correct and reliable analysis and interpretation of the results.

In the third chapter, the components of the information system for supporting the admission of university applicants are developed, including a method for providing recommendations of alternative educational programs and a method for predicting the success of admission to a university.

The method of providing recommendations is based on the combined use of collaborative filtering, archival data from the admission campaign and educational data, and content-based filtering, using clusters of keywords of academic programs. To implement collaborative filtering, we processed the archival data of the admission campaign to determine the similarity of applicants' interests. Such archival data was formed using the developed algorithm for calculating the student satisfaction index (with an above-average value). It is based on ranking the weighted values of the following criteria: academic performance, research activity, and the status of continuing education in a master's program. This approach allows us to provide the applicant with more relevant recommendations for the educational program. The process of selecting keywords based on the descriptions of academic programs and forming clusters of such words was carried out, which allowed the expansion of the list of recommendations for educational programs by content-based filtering based on the applicants' interests.

Preliminary processing of the archival data of the admission campaign was carried out for their further use by machine learning methods in solving the problem of predicting the success of admission to a higher education institution. Considering the advantages and disadvantages of different methods for determining the essential

features of the studied data set, using some of them to solve this problem is substantiated. The necessity of improving the existing machine learning tools for solving the problem of predicting the success of admission to higher education institutions is substantiated by analysing the accuracy of different classes of machine learning methods and artificial neural networks.

A two-stage data classification method has been improved to improve the efficiency of solving the binary classification problem and increase the accuracy of predicting the success of university admission. It is based on replacing the support vector machine-based classifier at the second stage of the method with a stacked combination of four support vector machine-based classifiers with different kernels and aggregating the results of their work using logistic regression. The structural and functional diagram of the improved method is presented, and the algorithms for its training and application are described.

In the fourth chapter, the framework of the information system for supporting the decision-making of applicants to Ukrainian higher education institutions is developed. The results of modelling the components of providing recommendations and predicting the success of admission to higher education institutions are evaluated. The developed framework of the decision support information system provides for the introduction of data pre-processing agents, which will ensure the necessary data quality and the possibility of their further analysis. The functioning of such agents involves data pre-processing related to the calculation of student satisfaction, as well as the selection of archival data of applicants, the formation of keywords of educational programs and their clusters, and the creation of a set of attributes to predict the success of admission.

The results of the joint use of content-based and collaborative filtering expand the list of recommended educational programs for applicants that meet their interests, which increases the level of awareness of the academic programs offered by the university during the admission campaign.

The modelling of the existing machine learning tools from different classes demonstrated the insufficient accuracy of the results of its work, which confirmed the need to improve the existing two-stage machine learning method to solve the problem as the best method among the studied ones.

The improved two-stage data classification method results were compared to those of existing methods from different classes. The results showed an improved method's accuracy increase by 2.9% (based on the F1-measure) when solving the problem of predicting the success of an applicant's admission to a university. Considering the latter's small error, the possibility of its use as a component of a decision support information system to predict the success of admission to a Ukrainian university is substantiated.

Keywords: information technology, information system, intelligent information system, higher education institution, admission campaign, applicant, decision making, decision support information system, recommender system, collaborative filtering, content-based filtering, machine learning, hierarchical classifiers, ensemble machine learning methods, artificial neural network.

Список опублікованих праць за темою дисертації

1. Zub K., Zhezhnych P., Strauss C. Two-Stage PNN–SVM Ensemble for Higher Education Admission Prediction. *Big Data and Cognitive Computing*. 2023. Vol. 7, no. 2. P. 83. URL: <https://doi.org/10.3390/bdcc7020083>
2. Зуб Х., Жежнич П. Бустингові методи машинного навчання для прогнозування успішності вступу абітурієнтів ЗВО України. *Computer systems and information technologies*. 2023. № 1. С. 84–90. URL: <https://doi.org/10.31891/csit-2023-1-11>
3. Зуб Х. В., Жежнич П. І. Аналіз ефективності вступної кампанії закладів вищої освіти України та способів її підвищення шляхом впровадження інформаційних технологій. *Вісник Вінницького Політехнічного Інституту*. 2022. № 3. С. 52–59. URL: <https://DOI.org/10.31891/csit-2023-1-11>

4. Зуб Х. В., Жежнич П.І. Огляд сучасного стану систем підтримки прийняття рішень абітурієнтів закладів вищої освіти. Міжнародний науково-технічний журнал «Інформаційні технології та комп'ютерна інженерія». 2022. № 53(1), С. 28–36. URL: <https://doi.org/10.31649/1999-9941-2022-53-1-28-36>.
5. Зуб Х. Оцінювання шансів вступу вступником ЗВО на основі моделі стекінгового об'єднання машини опорних векторів. *Наукові записки. Українська академія друкарства*. 2021. Вип. 2, №. 63. С. 168–176. URL: <https://doi.org/10.32403/1998-6912-2021-2-63-168-176>.
6. Zub K., Zhezhnych P. Performance Evaluation of ML-based Classifiers for HEI Graduate Entrants. *CEUR Workshop Proceedings : Proceedings of the International Workshop of IT-professionals on Artificial Intelligence (ProfIT AI 2021)*, Kharkiv, Ukraine, September 20-21, 2021. Vol. 3003. P. 92–99. URL: <https://ceur-ws.org/Vol-3003/short3.pdf>
7. Zhezhnych P., Zub K., Berezko O., Shilinh A. A Comparison of Machine Learning Algorithms for Prediction Higher Education Institution's Entrants Admissions. *Advances in Artificial Systems for Logistics Engineering*. Cham : Springer International Publishing, 2021. Vol. 82. P. 171–179. DOI: 10.1007/978-3-030-80475-6_17
8. Zhezhnych, P., Berezko, O., Zub, K., Demydov, I. Analysis of Features and Abilities of Online Systems and Tools Meeting Information Needs of HEIs' Entrants. *International Workshop on Control, Optimisation and Analytical Processing of Social Networks*, 2020. Vol. 2616. P. 76–85. URL: <https://ceur-ws.org/Vol-2616/paper7.pdf>
9. Зуб Х. Рекомендаційна система для абітурієнтів закладів вищої освіти з використанням освітніх даних. *Інформація, комунікація, суспільство 2023* : матеріали 12-ї Міжнародної наукової конференції ICS-2023, 18-23 травня 2023, Львів : Видавництво Львівської політехніки, 2023. С. 172–173. URL: https://skid.lpnu.ua/wp-content/uploads/2023/05/ICS2023_Proceedings.pdf

10. Зуб Х. Рекомендаційні інформаційні системи як засіб підтримки прийняття рішень вступників ЗВО України. *Інформація, комунікація, суспільство 2022* : матеріали 11-ї Міжнародної наукової конференції ICS-2022, 19-21 травня 2022, Україна, Львів : Видавництво Львівської політехніки, 2022. С. 49–50. URL: http://skid.lpnu.ua/wp-content/uploads/2022/05/ICS2022_Proceedings.pdf

11. Zub K. Application of support vector machine method for prediction HEI entrants` admission success. *10th International Youth Science Forum 'Litteris et Artibus'*. 2021. URL: <https://nauka.international/lea-2021/application-support-vector-machine-method-prediction-hei-entrants-admission-success>

12. Зуб Х. Аналіз освітніх веб-ресурсів задоволення інформаційних потреб абітурієнта ЗВО України. *Інформація, комунікація, суспільство 2021*: матеріали 10-ої Міжнародної наукової конференції ІКС-2021, 20–22 травня 2021 року, Україна, Львів : Видавництво Львівської політехніки, 2021. С. 15–16. URL: http://ics.skid-lp.info/ics_2021.pdf

Зміст

АНОТАЦІЯ.....	2
Зміст	14
Перелік умовних скорочень	18
Список рисунків	19
Список таблиць	21
Вступ	22
Розділ 1. Аналіз інформаційних технологій підтримки прийняття рішень вступників закладів вищої освіти	30
1.1. Аналіз особливостей та принципів процесу прийняття рішень вступників закладів вищої освіти в різні періоди вступної кампанії	30
1.1.1. Аналіз основних понять, періодів вступної кампанії та принципів її проведення в закладах вищої освіти	30
1.1.2. Аналіз особливостей та принципів процесу прийняття рішень вступників закладів вищої освіти	35
1.2. Аналіз сучасних відкритих веб-ресурсів для вступників закладів вищої освіти	39
1.2.1. Аналіз сучасних вітчизняних веб-ресурсів для задоволення інформаційних потреб вступників	39
1.2.2. Аналіз сучасних відкритих веб-ресурсів для вступників закордонних університетів.....	43
1.3. Аналіз сучасних методів та засобів підтримки прийняття рішень вступників закладів вищої освіти	47
1.3.1. Аналіз інформаційних систем підтримки прийняття рішень в контексті їхнього застосування під час вступної кампанії.....	47
1.3.2. Аналіз наукових досліджень методів та засобів підтримки прийняття рішень вступників закладів вищої освіти	51
1.4. Формування вимог до інформаційної технології підтримки прийняття рішень вступників ЗВО.....	58

Висновки до розділу	61
Розділ 2. Модель інформаційної системи підтримки прийняття рішень вступників ЗВО та засоби її реалізації	63
2.1. Модель інформаційної системи підтримки прийняття рішень вступників ЗВО.....	63
2.2. Метод сумісного використання колаборативної фільтрації та фільтрації на основі вмісту для надання рекомендацій вступникам ЗВО альтернативних освітніх програм.....	67
2.2.1. Вибір та обґрунтування методів надання рекомендацій освітніх програм вступникам ЗВО.....	67
2.2.2. Технологія збору та попереднього опрацювання даних для колаборативної фільтрації.....	71
2.2.3. Технологія збору та попереднього опрацювання даних для фільтрації на основі вмісту	73
2.3. Методи машинного навчання для передбачення успішності вступу	74
2.3.1. Технології збору та опрацювання набору архівних даних вступної кампанії.....	74
2.3.2. Обґрунтування критеріїв оцінювання ефективності роботи класифікаторів на основі машинного навчання	76
2.3.3. Вибір інструментарію машинного навчання для розв'язання задачі класифікації	78
Висновки до розділу	82
Розділ 3. Розроблення компонент інформаційної системи підтримки прийняття рішень вступників ЗВО	83
3.1. Колаборативна фільтрація з використанням архівних даних вступу та освітніх даних.....	83
3.1.1. Обробка архівних даних заявок вступників для визначення подібності інтересів користувачів	83

3.1.2. Процедура розрахунку показника задоволеності на основі сформованого набору освітніх даних.....	84
3.2. Фільтрація на основі вмісту з використанням ключових слів та кластерів описів освітніх програм.....	87
3.2.1. Виділення ключових слів описів освітніх програм.....	87
3.2.2. Формування кластерів ключових слів освітніх програм	96
3.3. Удосконалення двоетапного методу класифікації даних для підвищення точності передбачення успішності вступу вступника ЗВО.....	99
3.3.1. Збір та попереднє опрацювання даних вступної кампанії ЗВО	99
3.3.2. Опис та аналіз базових компонентів удосконаленого методу	108
3.3.3. Алгоритмічна реалізація удосконаленого двоетапного методу класифікації даних	112
Висновки до розділу	114
Розділ 4. Розробка інформаційної системи підтримки прийняття рішень вступників ЗВО та апробація результатів її роботи.....	116
4.1. Архітектура інформаційної системи підтримки прийняття рішень	116
4.2. Апробація результатів роботи інформаційної системи надання рекомендацій вступникам ЗВО.....	125
4.2.1. Оцінювання рівня охоплення освітніх програм у наборі даних для надання рекомендацій	125
4.2.2. Результати роботи колаборативної фільтрації	127
4.2.3. Результати роботи фільтрації на основі вмісту	132
4.3. Оцінювання результатів роботи двоетапного методу класифікації даних.....	136
4.3.1. Результати попереднього опрацювання набору даних для удосконаленого класифікатора.....	136

4.3.2. Результати роботи класифікаторів на основі машинного навчання	138
4.3.3. Результати роботи удосконаленого двоетапного методу класифікації даних	141
4.3.4. Порівняння ефективності роботи удосконаленого методу та інтерпретація результатів його роботи	143
Висновки до розділу	145
Висновки	147
Література.....	149
Додаток А.	166
Додаток Б.	168
Додаток В.	171

Перелік умовних скорочень

Скорочення, термін, позначення	Пояснення
БД	База даних
ЗВО	Заклад вищої освіти
ЗНО	Зовнішнє незалежне оцінювання
ІС	Інформаційна система
ІСППР	Інформаційна система підтримки прийняття рішень
ЙНМ	Ймовірнісна нейронна мережа
КБ	Конкурсний бал
ЛР	Логістична регресія
МН	Машинне навчання
МОВ	Метод опорних векторів
НМТ	Національний єдиний мультипредметний тест
НУЛП	Національний університет «Львівська політехніка»
ОП	Освітня програма
ОПР	Особа, що приймає рішення
РС	Рекомендаційна система

Список рисунків

Рис.1.1 Оцінка комфортності проходження вступниками етапів вступної кампанії за результатами дослідження Державної служби якості освіти України.....	42
Рис.1.2 Джерела інформації, до яких звертались вступники у разі виникнення проблемних ситуацій.....	42
Рис.2.1 Контекстна діаграма інформаційної системи підтримки прийняття рішень.....	64
Рис.2.2 Діаграма потоків даних першого рівня	65
Рис.2.3 Діаграма діяльності інформаційної системи підтримки прийняття рішень вступників ЗВО	66
Рис.2.4 Метод сумісного використання колаборативної фільтрації та фільтрації на основі вмісту.....	69
Рис.2.5 Діаграма потоків даних процесів збору та опрацювання даних для колаборативної фільтрації	72
Рис.2.6 Діаграма потоків даних процесів збору та опрацювання даних фільтрації на основі вмісту.....	73
Рис.2.7 Діаграма потоків даних процесів збору та опрацювання даних для навчання методів машинного навчання	75
Рис.3.1 Результати вилучення ключових слів за допомогою ПЗ YAKE.....	94
Рис.3.2 Довжина текстового поля професійних профілів випускників в описах освітніх програм.....	95
Рис.3.3 Довжина текстового поля професійних профілів випускників в описах освітніх програм після доповнення інформації.....	95
Рис.3.4 Приклад даних видаленого кластеру.....	99
Рис.3.5 Діаграма розмаху із врахування конкурсного балу пільгових вступників	102
Рис.3.6 Діаграма розмаху без врахування конкурсного балу пільгових вступників	103

Рис.3.7 Топологія ймовірнісної нейронної мережі	110
Рис.3.8 Структурно-функціональна схема роботи стекінгового об'єднання МОВ з чотирьома різними ядрами.....	111
Рис.3.9 Структурно-функціональна схема роботи удосконаленого двоетапного методу класифікації даних.....	113
Рис.4.1 Трирівнева архітектура інформаційної системи підтримки прийняття рішень вступників.....	117
Рис.4.2 Архітектура інформаційної системи підтримки прийняття рішень вступників ЗВО	122
Рис.4.3 ERD діаграма бази даних ІСППР вступників ЗВО	124
Рис.4.4 Порівняння показника кількості інтервалів та загальної кількості вступників в межах освітніх програм.....	126
Рис.4.5 Порівняння показника задоволеності студента та загальної кількості вступників в межах освітніх програм.....	127
Рис.4.6 Часткова візуалізація результатів роботи колаборативної фільтрації	130
Рис.4.7 Частота вибору альтернативних освітніх програм студентами на основі архівних даних вступу.....	131
Рис.4.8 Візуалізація результатів перетину освітніх програм в межах виділених кластерів.....	133
Рис.4.9 Перелік освітніх програм за кластером «програмних»	135
Рис.4.10 Перелік освітніх програм за кластером «даних»	135
Рис.4.11 Порівняння ефективності роботи досліджуваних методів на основі F1-міри.....	144

Список таблиць

Таблиця 1.1. Розподіл проаналізованих наукових досліджень за завданнями	54
Таблиця 1.2. Ієрархічний процес ухвалення рішення вступниками	58
Таблиця 3.1. Характеристики початкової вибірки архівних даних вступу	106
Таблиця 3.2. Принципи роботи, переваги та недоліки методів визначення важливих ознак у наборі даних	107
Таблиця 4.1. Приклад результату роботи фільтрації на основі вмісту	133
Таблиця 4.2. Значення F1-міри під час класифікації даних при використанні МОВ з рбф ядром та різних методів вибору важливих ознак із заданого набору даних	137
Таблиця 4.3. Показники ефективності досліджуваних методів у режимі застосування під час розв'язання поставленої у роботі задачі.	139
Таблиця 4.4. Результати роботи удосконаленого методу та його основних компонентів в режимі застосування	142

Вступ

Актуальність. Швидкий прогрес інформаційних та комунікаційних технологій є невід’ємним та важливим явищем у різних областях, зокрема й у сфері вищої освіти. Одним із критично важливих напрямків діяльності ЗВО, що потребує постійного впровадження та удосконалення інформаційних технологій, є вступна кампанія. Така значимість зумовлена посиленням автономії, соціальної, інноваційної та економічної відповідальності ЗВО та конкурентної боротьби на ринку освітніх послуг.

Забезпечення належного рівня результативності вступної кампанії, який можна визначати як кількісний та якісний набір вмотивованих студентів, що здійснюється відповідно до специфічних цілей та можливостей ЗВО, є вкрай важливим та актуальним завданням кожного ЗВО. Тобто, ЗВО потрібен свідомий та виважений вибір вступника, оскільки це сприятиме з одного боку високій вмотивованості та успішності студентів, що є ціллю ЗВО, а з іншого – формуванню якісного фахівця на ринку, що є ціллю економіки та розвитку країни.

Водночас, вступна кампанія є складним та важливим процесом і для самих вступників. Значний обсяг правил прийому, створення нових освітніх програм та відсутність достатнього рівня знань для здійснення свідомого вибору щодо вступу створюють ризик невизначеності та хибних рішень під час вступної кампанії. Наслідками можуть бути занадто висока чи навпаки низька популярність освітніх програм, низька вмотивованість та низька академічна успішність студентів та, як результат, безробіття чи працевлаштування не за фахом для студента і низький рейтинг ЗВО.

Отже, виникає потреба у підтримці щодо прийняття рішень з метою зміни умов для прийняття рішень – з таких, що здійснюються в умовах невизначеності в такі, коли ймовірність вступу можна оцінити за умови її об’єктивності. Водночас, важливо позбавити вступників наявних обмежень відповідно до їхніх інформаційних потреб у прийнятті рішення щодо вибору освітніх програм,

тобто, забезпечити можливість аналізу та обробки потрібних даних, розширити знання про освітні програми та підвищити швидкість рішень.

Враховуючи складність процесу, умови прийняття рішень та об'єм даних, які необхідно опрацювати, саме використання інформаційних технологій дасть змогу вступникам моделювати та аналізувати необхідну інформацію. Так ЗВО зможе привернути увагу потенційних вступників та популяризувати ОП, чим підсилить імідж закладу та свою конкурентоспроможність.

Проте існуючі відкриті веб ресурси для вступників ЗВО забезпечують лише поширення інформації щодо вступної кампанії. До таких ресурсів належать: офіційні сайти Міністерства освіти та науки України (МОН), його підрозділів, самих ЗВО та їхні офіційні сторінки у соцмережах, освітні медіа ресурси. Автоматизацію процесів збору та обробки даних забезпечено лише для організації проведення процесу прийому на навчання (Єдиної державної електронної бази з питань освіти (ЄДЕБО), електронний кабінет студента).

Натомість, практики закордонних університетів свідчать про високий інтерес до надання ширших можливостей для вступників, а також про попит на комерційні програмні рішення. Зокрема, поширеною практикою є надання онлайн-консультацій вступникам, автоматизований пошук і фільтрація інформації щодо вступу, онлайн засоби для орієнтації вибору вступника тощо. Однак, для багатьох ЗВО України проблемою є висока вартість купівлі комерційних програмних продуктів та труднощі в їхньому обслуговуванні. Тому актуальною є потреба розробки ефективних інформаційних технологій підтримки прийняття рішень, що враховуватимуть особливості вступної кампанії саме в Україні.

Сучасні дослідження доволі широко розглядають організацію та проведення вступної кампанії. Проте лише незначна їхня частина зосереджена саме на інформаційних технологіях, їхній розробці та впровадженні з метою автоматизації підтримки прийняття рішень. До таких належать дослідження програмних засобів для вступників ЗВО, які здійснювали В. М. Можаровський ,

В. В. Осадчий, В. С. Круглик, і Д. О. Букреєв, В. Д. Каричковський, С. С. Федушко. Однак більшість релевантних вітчизняних досліджень не мали подальшого розвитку чи впровадження результатів.

Значно ширше задачі підтримки прийняття рішень вступників розглянуто в іноземних наукових дослідженнях. Так, серед найпоширеніших задач є передбачення, що визначають шанси вступу, зокрема у працях N. Chakrabarty, S. Chowdhury, S. Singhal і A. Sharma, D. J. Devarapalli, S. Sridhar, S. Mootha, and S. Kolagati, B. Wu; рекомендації альтернативних освітніх програм – P.-C. Chang, C.-H. Lin, and M.-H. Chen, Abiyoga, A. Wicaksana and N. M. S. Iswari; фактори, що впливають на тенденцію вступу T. Park and C. Kim, R. Ahlawat, S. Sahay, S. Sabitha, and A. Bansal. Автори наголошують на актуальності розроблення та використання інтелектуалізованих інформаційних технологій під час вступної кампанії, які б забезпечили наявність додаткових засобів отримання релевантної інформації що відповідає потребам вступника, під час прийняття рішення щодо вибору конкретної освітньої програми.

Сьогодні вступникам ЗВО України надаються недостатні можливості для прийняття обґрунтованого рішення щодо вибору освітньої програми, яка найбільш відповідає їхнім інтересам, із переліку запропонованих закладом. Отже, виникає необхідність розширити ці можливості шляхом розроблення нових та вдосконалення існуючих інформаційних технологій з урахуванням особливостей процесу вступу до ЗВО України.

Зв'язок роботи з науковими програмами, планами, темами. Тематика дисертаційної кваліфікаційної роботи визначена з урахуванням наукового напрямку кафедри соціальних комунікацій та інформаційної діяльності, в рамках зареєстрованої наукової тематики «Управління процесами соціальних комунікацій в глобальному інформаційному просторі» (номер державної реєстрації 0119U101870).

Мета і завдання дослідження. Метою дослідження є підвищення рівня обізнаності вступників ЗВО України відповідно до їхніх інтересів щодо

можливих альтернатив вибору освітніх програм на основі побудови нових і вдосконалення існуючих моделей, методів та інформаційних технологій підтримки прийняття рішень для вибору освітніх програм.

Для досягнення поставленої мети необхідно розв'язати такі задачі:

- проаналізувати особливості та принципи процесу прийняття рішень вступників ЗВО в різні періоди вступної кампанії та сучасний рівень забезпечення інформаційних технологій для вступників ЗВО, та основні тенденції у вітчизняних й іноземних наукових дослідженнях;
- розробити технологію збору й опрацювання даних для можливості проведення подальшого аналізу на їхній основі;
- розробити технологію надання розширених рекомендацій вступнику щодо вибору наявних у ЗВО освітніх програм, які відповідають його інтересам;
- удосконалити метод передбачення успішності вступу вступника на обрану освітню програму;
- розробити архітектуру інформаційної системи, провести експериментальні дослідження та здійснити апробацію отриманих результатів.

Об'єктом дослідження є процес підтримки прийняття рішень вступниками ЗВО України.

Предметом дослідження є моделі, методи та інформаційні технології підтримки прийняття рішень вступниками ЗВО.

Методи дослідження. Під час виконання досліджень дисертаційної роботи було використано теорію прийняття рішень, теорію та методи побудови інформаційних систем підтримки прийняття рішень, методи проектування інформаційних систем, статистичні методи вилучення ключових слів, семантичні методи кластеризації, методи математичної статистики, методи машинного навчання, теорію ансамблевого навчання.

Наукова новизна одержаних результатів полягає в науковому обґрунтуванні та розробленні інформаційної технології підтримки прийняття рішень. Серед інших, отримано такі основні наукові результати:

1. Вперше розроблено модель інформаційної системи підтримки прийняття рішень вступників ЗВО яка, на відміну від вже існуючих, використовує поєднання методу надання рекомендацій та удосконаленого методу передбачення успішності вступу, що підвищує рівень обізнаності вступників ЗВО України відповідно до їхніх інтересів.

2. Набуло подальшого розвитку використання технології надання розширених рекомендацій вступнику на основі сумісного застосування колаборативної фільтрації з використанням архівних даних вступної кампанії та освітніх даних, та фільтрації на основі вмісту з використанням кластерів ключових слів освітніх програм, що дало змогу розширити перелік альтернативних освітніх програм, які відповідають інтересам вступника.

3. Удосконалено двоетапний метод класифікації даних за рахунок заміни класифікатора на основі машини опорних векторів на другому етапі методу на стекінгове об'єднання чотирьох класифікаторів на основі машини опорних векторів із різними ядрами та агрегуванням результатів їхньої роботи логістичною регресією, що дозволило підвищити точність тестових даних під час розв'язання задачі передбачення успішності вступу до ЗВО.

Практичне значення одержаних результатів дисертаційної роботи полягає в тому, що розроблена інформаційна технологія дає змогу розширити спектр послуг з підтримки прийняття рішень для вступників ЗВО щодо вибору освітніх програм, які відповідають їхнім інтересам. Зокрема:

- розроблена технологія сумісного використання колаборативної фільтрації та фільтрації на основі вмісту забезпечує надання розширеного спектру рекомендацій освітніх програм вступнику до конкретного ЗВО. Зокрема, використання колаборативної фільтрації дає змогу сформувати перелік рекомендованих вступнику освітніх програм на основі врахування архівних даних уподобань випускників певної освітньої програми та їхнього рівня задоволеності під час навчання. Архівні дані в цьому випадку обмежуються тільки тими, для яких показник рівня задоволеності студентів після завершення

певної програми є вище середнього, та які відбираються шляхом ранжування зважених результатів їхньої семестрової успішності, наукової діяльності та статусу продовження навчання в магістратурі. Використання фільтрації на основі вмісту дає змогу сформувати додатковий перелік альтернативних освітніх програм, які відповідають інтересам вступника в межах переліку, що пропонується конкретним ЗВО під час проведення вступної кампанії. Такий перелік містить різні освітні програми і формується в межах виділеного кластеру ключових слів описів освітніх програм, що дає змогу враховувати в тому числі й освітні програми на перетині різних галузей знань, а також різні за рівнем популярності серед вступників;

- розроблені технології збору та опрацювання архівних даних вступної кампанії, описів освітніх програм, освітніх даних студентів покладено в основу реалізації методів колаборативної фільтрації, фільтрації на основі вмісту та передбачення успішності вступу до ЗВО;

- використання удосконаленого двоетапного класифікатора забезпечило підвищення точності його роботи на тестових даних на 2,9 % (згідно з F1-мірою) порівняно з точністю за існуючим методом під час розв'язання задачі передбачення успішності вступу до ЗВО. Враховуючи розмір тестового набору даних, таке підвищення точності (2,9 %) дало змогу видати правильні результати передбачення ще для 92-х вступників під час розв'язання поставленої задачі. Окрім цього, використання ансамблю з чотирьох методів машинного навчання в удосконаленому методі замість одного (як в базовому), забезпечило підвищення генералізаційних властивостей класифікатора та рівня довіри до результатів його роботи. Також, враховуючи мале значення похибки роботи вдосконаленого методу (1,8 % на основі F1-міри), його можна використовувати на практиці;

- розроблена архітектура інформаційної системи підтримки прийняття рішень забезпечила можливість використання колаборативної фільтрації та фільтрації на основі вмісту для надання розширеного спектру рекомендацій освітніх програм вступнику та вдосконаленого класифікатора для передбачення

успішності вступу. Така архітектура передбачає застосування агентів попереднього опрацювання даних, що покладені в основу роботи цих методів.

Результати дисертаційного дослідження впроваджено у Дрогобицькому державному педагогічному університеті імені Івана Франка та використано під час виконання наукової складової міжнародного проєкту Erasmus+ «Відкриті практики, прозорість та доброчесність для сучасної вищої школи» (ОРТІМА, 618940-EPP-1-2020-1-UA-EPPKA2-SVNE- JP), що підтверджено відповідними актами.

Особистий внесок здобувача. Усі наукові результати дисертаційної роботи отримані автором самостійно. У друкованих працях, опублікованих у співавторстві, автору належать: аналіз особливостей вступної кампанії ЗВО [14]; аналіз сучасного стану систем підтримки прийняття рішень вступників ЗВО [34]; дослідження ефективності використання інструментарію штучного інтелекту для інтелектуального аналізу даних вступної кампанії [83, 71, 72]; модель інформаційної системи підтримки прийняття рішень вступників ЗВО [4]; технології збору та опрацювання даних для подальшого інтелектуального аналізу на їхній основі [82, 65]; технологія надання розширених рекомендацій вступнику на основі сумісного застосування колаборативної фільтрації та фільтрації на основі вмісту [11, 69]; удосконалений двоетапний метод класифікації даних на основі сумісного використання ймовірнісної нейронної мережі та стекінгового об'єднання машини опорних векторів [84, 99].

Апробація результатів дисертації. Результати дисертаційної роботи апробовано на міжнародних наукових конференціях та форумах, зокрема:

- International Workshop of IT-professionals on Artificial Intelligence (ProfIT AI 2021), Kharkiv, Ukraine, September 20-21, 2021.
- International Conference on Artificial Intelligence and Logistics Engineering (ICAILE 2021), Kyiv, Ukraine, January 22 - 24, – 2021. E
- The 2nd International Workshop on Control, Optimisation and Analytical Processing of Social Networks (COAPSN 2020), Lviv, Ukraine, May 21, 2020.

- XII Міжнародна наукова конференція «Інформація, комунікація, суспільство - 2023» (ICS-2023), Львів, Україна, 18-20 травня, 2023.
- XI Міжнародна наукова конференція «Інформація, комунікація, суспільство – 2022» (ІКС-2022), Львів, Україна, 19–21 травня, 2022.
- 10th International Youth Science Forum «Litteris et Artibus» Національний університет «Львівська політехніка», Львів, Україна, 25-27 листопада, 2021.
- X Міжнародна наукова конференція «Інформація, комунікація, суспільство – 2021» (ІКС-2021), Львів, Україна, 20–22 травня, 2021

Публікації. Основні положення дисертації опубліковано у 12 наукових працях, з яких: 4 статті у наукових фахових виданнях України з технічних наук, 1 стаття у науковому періодичному виданні іншої держави, яка індексується у наукометричних базах Scopus Web of Science та віднесена до Q2, та 7 праць – у матеріалах і тезах конференцій та форуму.

Структура та обсяг роботи. Дисертаційна робота складається зі вступу, чотирьох розділів, висновків, списку використаних джерел (117 найменувань) та 3-х додатків. Основний зміст викладено на 149-ти сторінках друкованого тексту, містить 29 рисунків та 8 таблиць. Загальний обсяг роботи – 172 сторінки.

Розділ 1. Аналіз інформаційних технологій підтримки прийняття рішень вступників закладів вищої освіти

У першому розділі проаналізовано принципи проведення вступної кампанії ЗВО України та особливості підтримки прийняття рішень вступників у різні періоди вступної кампанії. Проаналізовано існуючі практичні рішення та інформаційні технології для задоволення інформаційних потреб вступників ЗВО. Здійснено аналіз існуючих методів та засобів підтримки прийняття рішень вступників ЗВО. Зокрема, інтелектуальних інформаційних систем підтримки прийняття рішень, методів та засобів вирішення задач підтримки прийняття рішень у вітчизняних та закордонних наукових дослідженнях. Обґрунтовано актуальність створення інформаційної системи підтримки прийняття рішень вступників ЗВО та визначено основні задачі такої системи.

1.1. Аналіз особливостей та принципів процесу прийняття рішень вступників закладів вищої освіти в різні періоди вступної кампанії

1.1.1. Аналіз основних понять, періодів вступної кампанії та принципів її проведення в закладах вищої освіти

Вступна кампанія як процес, відповідно до поставлених цілей та визначених можливостей ЗВО, включає такі елементи, як профорієнтаційну підготовку, нормативно-правове забезпечення, організаційні та рекламно-інформаційні процеси. Такі складові елементи вступної кампанії є взаємопов'язані та взаємозумовлені між собою.

Прямим суб'єктом відносин на ринку споживання освітніх послуг є сам ЗВО, як постачальник освітніх послуг, та вступники як споживачі. Іншими елементами інфраструктури ринку професійних освітніх послуг можна вважати державу, центри зайнятості, органи ліцензування освітніх закладів, бізнес. Початковим етапом взаємодії зі споживачами освітніх послуг є прийом на

навчання. Прийом на навчання до ЗВО здійснюється на конкурсній основі відповідно до Умов прийому на навчання для здобуття вищої освіти, затверджених центральним органом виконавчої влади у сфері освіти і науки [1].

Згідно з порядком прийому на навчання, кожен, хто має намір вступати до ЗВО, повинен зареєструватися та пройти вступні випробування, що мають на меті визначити рівень навчальних досягнень випускників загальноосвітніх навчальних закладів при вступі до ЗВО України.

У 2022 році через повномасштабне вторгнення росії в Україну було скасовано проведення зовнішнього незалежного оцінювання (ЗНО). Замість нього було вперше запроваджено проведення національного єдиного мультипредметного тесту (українська мова, математика, історія України) (НМТ).

Таким чином, відповідно до системи вищої освіти та строків прийому заяв та документів, конкурсного відбору та зарахування на навчання (що визначаються Порядком прийому можна виділити періоди прийому на навчання, які щороку визначаються конкретними строками [1]. Так, для вступу на перший курс для здобуття ступеня бакалавра на основі повної загальної середньої освіти (ПЗСО) за денною формою здобуття освіти можна виділити такі періоди:

- реєстрація електронних кабінетів вступників, завантаження необхідних документів;
- медичні огляди та інші доконкурсні процедури;
- прийом заяв та документів;
- реєстрація учасників творчих конкурсів, організація та проведення творчих конкурсів та співбесід;
- оприлюднення рейтингових списків;
- виконання вимог до зарахування вступниками, які отримали рекомендації;
- зарахування на навчання [1].

Згідно з порядком прийому на навчання для здобуття вищої освіти пояснення основних термінів наступне:

«вступник – вступник – особа, яка зареєструвала особистий електронний кабінет вступника та/або подала заяву(и) про допуск до участі в конкурсному відборі на одну (декілька) конкурсних пропозицій»;

«конкурсний відбір – процедура відбору вступників на конкурсні пропозиції на основі конкурсних балів, пріоритетів, мотиваційних листів відповідно до цього Порядку (незалежно від форми власності закладу освіти та джерел фінансування навчання)»;

«пріоритетність – визначена вступником під час подання заяв черговість (де 1 є найвищою пріоритетністю) їхнього розгляду у разі адресного розміщення бюджетних місць»;

«рейтинговий список вступників – список вступників, впорядкований за черговістю зарахування на навчання на конкурсну пропозицію з урахуванням джерела фінансування, що формується відповідно до цього Порядку та Правил прийому» [1].

Вступна кампанія як термін чітко не визначається законодавством України, проте вживається в межах законодавства та інформаційного наповнення офіційного сайту Міністерства освіти та науки України (МОН) [2]. Вважається, що початком вступної кампанії є реєстрація електронних кабінетів для завантаження необхідних документів. Однак з перспективи ЗВО, межі вступної кампанії не відповідають офіційним датам її проведення, оскільки підготування та організація наступної вступної кампанії розпочинається одразу після завершення попередньої [3]. Тому, поняття «вступник» у цьому дослідженні вживається у більш ширшому розумінні та позначає особу, зацікавлену у вступі, тобто ту, яка має на меті подати заявку на участь у конкурсі, або вже її подала.

Оскільки результативність вступної кампанії є важливою для обох сторін, як вступника так і самого ЗВО, доцільно розглянути це поняття детальніше.

Одним з інструментів оцінки та аналізу різних аспектів діяльності ЗВО, у тому числі і вступної кампанії, сьогодні є використання рейтингів. У контексті

постійної конкуренції в системі вищої освіти, висока позиція ЗВО у рейтингах стає необхідною складовою стратегій маркетингу. Інформація про результати вступної кампанії, яка доступна широкому загалу у мережі Інтернет, є помітним фактором, який впливає на формування привабливості ЗВО для потенційних вступників [4].

Більшість рейтингів, що аналізують результати вступної кампанії, базуються на кількісних даних і надають числову оцінку всім ЗВО за різними показниками. Незважаючи на те, що такі дані впливають на суспільну свідомість, жоден із них не може повністю передати об'єктивну оцінку результативності вступної кампанії конкретного ЗВО як окремого об'єкта. Крім того, варто відзначити, що показники або набір показників, що використовуються для складання рейтингів, є результатом суб'єктивного відбору і часто не відповідають розмірам і структурі конкретних ЗВО. Таким чином, існуючі загальнодоступні рейтинги не є інструментом, що забезпечує об'єктивну оцінку результативності вступної кампанії ЗВО. Тож, можна вважати, що оцінка результативності вступної кампанії повинна ґрунтуватися на поставлених цілях та пріоритетах ЗВО [4].

У рамках оцінки результатів вступної кампанії та загальної ефективності діяльності ЗВО, важливим аспектом є поняття якісного набору студентів. Показники формування якісного студентського контингенту мають прямий вплив на результативність ЗВО та є важливою передумовою високої якості та ефективності освітньої діяльності, а також на успішний процес підготовки кваліфікованих, професійно компетентних та мобільних фахівців [5]. Отже, одним із стратегічних завдань ЗВО під час вступної кампанії є відбір вступників з високим рівнем знань, що відповідає вимогам держави щодо підготовки необхідної кількості фахівців, яка визначена даними ліцензійного набору. Сучасні наукові дослідження визначають якісний контингент студентів за допомогою оцінки рівня їхніх знань, що виражається у співвідношенні певних кількісних показників, таких як бали, які враховуються під час вступу.

Проте, не менш важливими показниками якості набору є обдуманий вибір вступників та рівень його мотивації до обраної освітньої програми (ОП), який визначається пріоритетністю заяв. Різноманітність правил прийому та освітніх напрямів різних ЗВО (включаючи вагові коефіцієнти, відмінності між гуманітарними та технічними освітніми програмами) призводить до того, що визначення якісного контингенту студентів може проводитися з урахуванням індивідуальних цілей та особливостей набору студентів кожного конкретного ЗВО.

Таким чином, результативність вступної кампанії можна описати як кількісний та якісний набір вмотивованих студентів, що здійснюється відповідно до специфічних цілей та можливостей ЗВО. Для того, аби чіткіше дати оцінку ефективності потрібно аналізувати результати діяльності кожного окремого ЗВО відповідно для його цілей, освітніх пропозицій, структури та задач які виникають під час вступної кампанії.

Відповідно до особливостей вступної кампанії як процесу, формування кількісних та якісних показників залежить від очевидних та неочевидних впливів. Зокрема, це:

- фактори зовнішнього середовища – соціально-демографічні, економічні, потреби ринку праці, політично-правові, технологічні, культурні тощо (оскільки суспільство та ринок праці є непрямим суб'єктом процесу надання освітніх послуг);
- регулювання з боку державних органів виконавчої влади (оскільки держава є непрямим суб'єктом процесу надання освітніх послуг);
- встановлення вимог, організація та контролювання перебігу вступної кампанії самим ЗВО;
- рішення вступників ЗВО.

Очевидно, що вступники створюють попит на освітні програми та приймають рішення, які в кінцевому результаті формують кількісні показники вступної кампанії. Оскільки на інші показники, такі, як демографічні показники,

економіка країни та вимоги державних органів вплинути не можливо, доцільно розглянути процес прийняття рішень вступниками як один із основних чинників формування результатів вступної кампанії [4].

1.1.2. Аналіз особливостей та принципів процесу прийняття рішень вступників закладів вищої освіти

З точки зору теорії прийняття рішень вступник приймає рішення на підставі його знань про вступну кампанію, як вона змінюється з перебігом часу та за наявності множини показників, що характеризують ефективність та якість прийнятого рішення відповідно до поставленої мети. Тому, аби розуміти можливі труднощі цього процесу потрібно проаналізувати сам процес прийняття рішень вступниками та його особливості в межах вступної кампанії.

Вступна кампанія є складним та довгим процесом, де дії вступника відрізняються залежно від конкретного періоду. Для кращого розуміння задачі та процесу прийняття рішень та майбутньої побудови моделі прийняття рішень варто врахувати особливості у кожен з періодів. Відповідно до порядку прийому до ЗВО можна виділити такі основні періоди вступної кампанії:

- профорієнтація та прийняття рішення майбутніми вступниками щодо вибору ЗВО та напряму підготовки – протягом навчання в закладі загальної середньої освіти та до реєстрації;
- реєстрація на НМТ;
- складання НМТ;
- реєстрація в кабінетах та подання заявок;
- подання оригіналів та зарахування;
- узгодження запитань щодо навчальної діяльності університету та соціально-побутових умов.

Сам процес прийняття рішень вступників характеризується вибором серед альтернативних варіантів, а саме ОП, що формується добровільно відповідно до думок та мотивів з метою вступити до ЗВО (відповідно здобути вищу освіту) [6].

Відповідно, з точки зору процесу прийняття рішень, можна виділити певні атрибути такого процесу, що проаналізовані нижче.

1. Використання внутрішнього (розумового) представлення. Вступники не можуть використовувати попередній досвід, оскільки він відсутній (за типових умов вступу на базі ПЗСО). Тож, опираються на представлення, розуміння доступної їм інформації щодо вступу.

Умови вступу зумовлюють те, що будь-які рішення вступника, аж до подання оригіналів документів щодо його навчання в ЗВО, мають значні наслідки на подальшу траєкторію як студента. Рішення, що стоять перед вступниками відповідно до бажаного результату (мета вступу у кожного своя) полягають в виборі: предметів НМТ, ЗВО, освітньої програми, пріоритетів заяв, творчих конкурсів. Такі умови вступу надають альтернативи вибору вступниками, проте, водночас ускладнюють можливість належним чином оцінити усі фактори так, щоб кінцеве рішення про подання оригіналів документів справді було ефективним.

Ймовірність вступу, інфраструктура ЗВО, вартість навчання чи додаткові витрати на проживання та інші фактори, які може опустити вступник під час внутрішнього (розумового) представлення інформації якою володіє, призводять до того, що вступник згодом стає обмежений в виборі, а не володіє альтернативами та не в змозі оцінити свої шанси на вступ.

Великий обсяг правил прийому, впровадження нових освітніх програм та, водночас, відсутність достатнього рівня знань для прийняття вибору щодо вступу створюють ризик виникнення невизначеності та прийняття неоптимальних рішень під час проведення вступної кампанії. Потенційні наслідки цього можуть включати як перевищення, так і недоліки у популярності освітніх програм, знижену мотивацію та недостатній академічний успіх студентів, а також, відповідно, зміну рейтингу ЗВО, безробіття випускників, або їхнього працевлаштування не за фахом [4].

Отож прийняття рішень вступниками здійснюється в умовах невизначеності, оскільки вони не можуть самостійно оцінити імовірність потенційних результатів.

2. Досягнення мети. Ухвалення рішення, як у більшості випадків, починається з розгляду бажаного результату – мети, що можна доволі чітко описати, тобто здобуття рівня вищої освіти. Натомість сам процес, тобто будь-які рішення протягом вступної кампанії зв'язуються з кінцевим бажаним результатом та здійснюються на основі розумового представлення ситуацій чи умов під час прийняття рішень.

Власне кожен вступник має свою усвідомлену мету вступу, досягнення якої і можна вважати ефективністю для нього результатів вступної кампанії, що передбачає в зальному:

- цінності перегукуються із цінностями університету;
- особисту зацікавленість у вступі на певну освітню програму й відповідні очікування, досягнення в навчанні та інших видах діяльності;
- можливість побудувати професійні та навчальні цілі;
- є ідеї та плани на час навчання в університеті: участь у самоврядуванні, здобуття навичок, програми обміну та навчальні проєкти;
- є ідеї та плани після випуску і глобальна мета реалізації себе у професії.

Отже, ЗВО потрібен свідомий і виважений вибір вступника, оскільки це означатиме, як результат, висока вмотивованість та успішність студентів що є ціллю ЗВО та якісний фахівець на ринку що є ціллю економіки та розвитку країни. Водночас, основним питанням для вступника на базі ПЗСО, який вже визначився із галуззю як загальною сферою його майбутньої професійної діяльності, є вибір належної освітньої програми ЗВО. Цей вибір повинен відповідати особистим інтересам, здібностям і можливостям студента та, водночас, попиту на ринку праці у конкретній професійній сфері, з метою свідомого вибору майбутньої кар'єри [7].

3. Сприйняття інформації щодо вступу, яку вступник інтерпретує на основі наявних знань, досвіду, підготовки, природних здібностей. Принцип сприйняття взаємодіє з принципом внутрішнього представлення інформації, тобто з набутими знаннями вступника щодо вступу. Тут знову ж таки варто наголосити на тому, що навіть достатньо структурована інформація щодо вступу яка надається ЗВО для вступників може бути складною для сприйняття та аналізу.

Також важливо розглянути фази (етапи) прийняття рішення вступника в контексті вступної кампанії:

1. Фаза аналізу та постановки задачі. Це період визначення основних та допоміжних цілей прийняття рішень. Сюди можна віднести вибір напряму, відвідування додаткових (приватних) уроків. Вступник збирає усю релевантну інформацію, діагностує та виявляє проблеми. Це дає змогу визначити напрями пошуку рішень і відкинути ті, які не відповідають цілям.

2. Проектування альтернатив з огляду на інформаційні потреби та критерії (формування рішень). Сюди можна віднести вибір міста для навчання, форми навчання та фінансування, ЗВО, освітні програми для подачі заявок, пріоритети. На даному етапі відбувається визначення обмежень та відокремлення прийнятних варіантів від неприйнятних та визначення критеріїв, які сприяють вибору варіантів рішення, що в найбільшій мірі відповідають інтересам вступника [8].

3. Оцінка та вибір кращої альтернативи (вибір рішення) . Вступник обирає кращу альтернативу з використанням певних критеріїв та здійснює подальший остаточний вибір – освітня програма серед поданих заявок, що пройшли конкурсний відбір відповідно до визначених раніше критеріїв – форма навчання, оплати [6].

При цьому варто зазначити, що процес прийняття рішення вступником може мати загальні обмеження відповідно до їхніх інформаційних потреб. Для одержання ефективного результату, вступник не в змозі самостійно здійснити аналіз та обробку потрібних даних відповідно до усіх тактичних та стратегічних

цілей. Якість та швидкість рішень може бути недостатньо високою. Враховуючи кількість вступників, що беруть участь у конкурсі та кількість показників, що формують конкурсний відбір вступник також не в змозі самотійно здійснювати обробку числових даних, виконувати операції алгоритму прийняття рішень. Оскільки рішення може прийматись у реальному часі (подача заявки в конкретний момент з конкретною конкурсною ситуацією) спрогнозувати та домогтись бажаних результатів для вступника самотійно не можливо. Також вступники можуть проявляти мінливість, невпевненість, нелогічність, намагання суттєво спростити задачу. Тож, без допоміжних засобів перед вступником виникають загальні для прийняття рішення обмеження - робочої пам'яті людини (яка потрібна для опрацювання інформації), джерел інформації (для використання знань якими володіє), оброблення числових даних та складних процесів мислення [6].

1.2. Аналіз сучасних відкритих веб-ресурсів для вступників закладів вищої освіти

1.2.1. Аналіз сучасних вітчизняних веб-ресурсів для задоволення інформаційних потреб вступників

Вибір конкретних ОП фактично є продовженням профорієнтаційної діяльності, що здійснюється в закладах ПЗСО. Програма наскрізної профорієнтації, розроблена МОН, спрямована на введення учнів у поняття професійного спрямування, та ознайомлення з різноманітністю професій та виявлення їхніх талантів, а також має на меті орієнтацію на ринок праці в Україні в цілому. Проте, вступна кампанія в ЗВО є складнішою, оскільки відбір претендентів на навчання майже відсутній у системі загальної середньої освіти. У зв'язку з тим, що загальна середня освіта є обов'язковою і гарантованою державою, ситуація у сфері професійної освіти є іншою [9].

Серед існуючих відкритих веб-ресурсів для вступників, що забезпечені на національному рівні можна виділити наступні:

1. Офіційний веб-сайт МОН, на якому забезпечено актуальну інформацію щодо умов вступу в окремому порталі сайту, розділ питання/відповіді, новинний контентом щодо актуальних змін. Також є архівна версія попередніх вступних кампаній починаючи з 2017 року (<https://mon.gov.ua/ua/tag/vstupna-kampaniya-2023>).

2. Український центр оцінювання якості освіти - інформаційний супровід (<https://testportal.gov.ua>).

3. Особистий електронний кабінет вступника - що забезпечує подання документів, проходження НМТ, аж до підтвердити вибір місця навчання після надання рекомендацій (<https://cabinet.edbo.gov.ua>).

4. Сайт-моніторинг вступних кампаній до вищих навчальних закладів України (<https://vstup.edbo.gov.ua>). Дає змогу відслідкувати інформацію (з можливістю пошуку та фільтрування) щодо порядкового номера вступника у рейтингах протягом вступної кампанії та інформацію щодо загальної кількості заяв, що беруть участь у конкурсі за даними ЗВО та Єдиної державної електронної бази з питань освіти (ЄДБО).

5. ЄДЕБО – надає інформацію щодо ЗВО, документів про освіту та наукові ступені, перебігу вступної кампанії (<https://info.edbo.gov.ua/>).

Варто зазначити, розроблений план заходів Концепції державної системи професійної орієнтації населення включає події як для школи так і для ЗВО, та з 30-ти запропонованих є лише один, орієнтований на ІТ – створення протягом 2018 - 2023 років електронні бази даних профорієнтаційного спрямування що мають створення інформаційної бази для сприяння профорієнтації різних категорій населення, зокрема сільської молоді та представників національних меншин [10].

Достатньо популярними сьогодні ресурсами також є освітні портали. Прикладами таких є «Вступ.ОСВІТА.УА» – інформаційна система, що забезпечує пошук, відображення, постійне оновлення даних в період вступної кампанії. Також надає про інформацію про ЗВО, умови вступу до ЗВО України,

архівні дані – статистична інформація про результати вступу до ЗВО у минулих роках (<https://vstup.osvita.ua>).

Інформаційна інфраструктура ЗВО підсилюється та вдосконалюється за допомогою веб-сайтів, які виконують різноманітні освітні, інформаційні та комунікаційні функції у суспільстві [11]. Для аналізу ЗВО, було здійснено на основі рейтингу Webometrics [12] та консолідованого рейтингу вузів України 2023 року за даними ресурсу ОСВІТА.ua [13]. У всіх випадках онлайн системи це виключно інформаційно-пошукові портали. Варто наголосити, що є недоліки у поданні інформаційного наповнення, серед яких: відсутність консолідованого каталогу освітніх програм для вступу або ж засобів навігації та розширеного пошуку в межах сторінок переліку освітніх програм; відсутність детального опису пропозицій за освітніми рівнями; завантаження файлів замість перегляду веб-сторінок при перегляді статистики вступної кампанії, описів освітніх програм; немає архіву даних, або складно знайти.

Тож, існуючі відкриті веб-сервіси забезпечують:

- інформацією в галузі освіти щодо навчальних закладів, документів про освіту та наукові ступені;
- процес подання та розгляд заяв та прийом на навчання (в певних випадках) в електронній формі на участь у конкурсному відборі до ЗВО;
- збереження інформації та оперативне інформування щодо статистики перебігу вступної кампанії;
- обчислення конкурсного балу.

Варто зазначити, що згідно останніх досліджень щодо організації та проведення закладами освіти вступної кампанії в 2023 році, проведених Державною службою якості освіти України (<https://sqe.gov.ua/>) перші етапи вступної кампанії є більш складними. Ці етапи включають пошук загальної інформації щодо ЗВО та освітніх програм.

Онлайн-опитування, проведене наприкінці вересня 2023-го року передбачало додатковий блок запитань для респондентів з-поміж вступників

2023 року. Оцінка складності проходження вступниками кожного з етапів вступної кампанії дала змогу визначити які саме труднощі виникають перед вступниками. Зокрема, серед проблемних питань, які потребували додаткових зусиль та часу є великі обсяги оприлюдненої інформації і, відповідно, складний пошук конкретних (простих для розуміння) відомостей (Рис. 1.1).

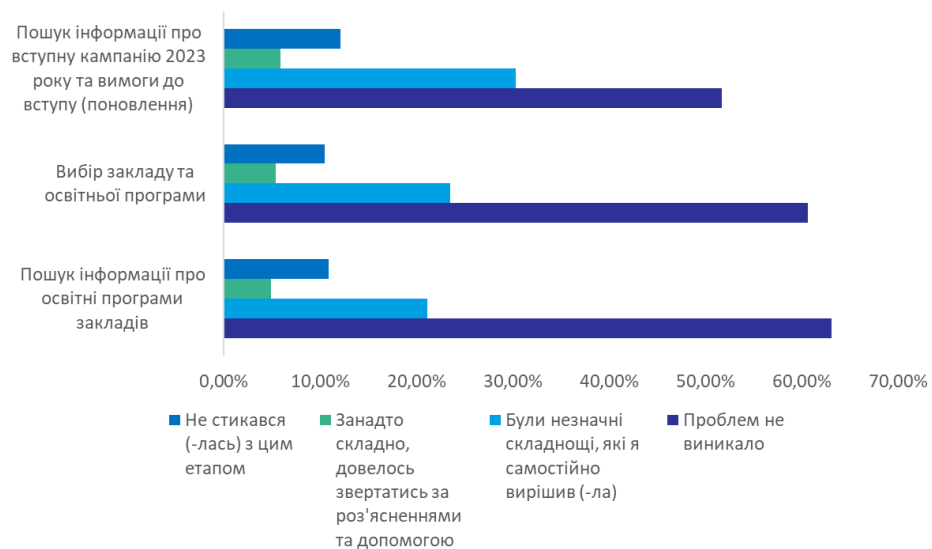


Рис.1.1 Оцінка комфортності проходження вступниками етапів вступної кампанії за результатами дослідження Державної служби якості освіти України



Рис.1.2 Джерела інформації, до яких звертались вступники у разі виникнення проблемних ситуацій

В більшості випадків респонденти задоволені рівнем інформаційного забезпечення. Проте в разі виникнення проблемних ситуацій вступники шукали відповіді звертаючись не лише до офіційних джерел (Рис. 1.2).

Отже, актуальним є розроблення нових та удосконалення існуючих інформаційних технологій з урахуванням особливостей процесу вступу до ЗВО України для задоволення інформаційних потреб вступників. Це дозволить закладам вищої освіти (ЗВО) адаптувати свою діяльність у частині організації вступної кампанії, що позитивно вплине на їхню репутацію та вдосконалисть стратегію набору вступників у майбутньому.

1.2.2. Аналіз сучасних відкритих веб-ресурсів для вступників закордонних університетів

Для кращого аналізу освітньої діяльності ЗВО, зокрема в частині проведення вступної кампанії доцільно проаналізувати також практики закордонних університетів. Тому за даними рейтингу «Webometrics» було проаналізовано досвід 30-ти найуспішніших університетів Європи [12].

Аналіз офіційних веб-сайтів обраних університетів, дав змогу зробити певні висновки щодо їхнього інформаційного наповнення та функціональних можливостей. Так, варто виділити наступні ознаки інформаційного наповнення що стосується вступу в навчальний заклад:

- усі заклади надають інформацію у відкритому доступі;
- дані про освітні пропозиції чітко структуровані відповідно до пропозицій закладу освіти;
- описи освітніх програм для вступу подано у повному та достатньому для розуміння обсязі. Основний акцент в інформаційному наповненні зроблено на те, щоб вступник зміг повністю ознайомитись зі структурою програми;
- формальні вимоги для кандидатів чітко сформульовані, аби не виникало непорозумінь щодо необхідних документів [14].

Процес прийому у всіх випадках передбачає електронний вступ, що включає всі етапи, в тому числі здійснення оплати. Вони є зручними у використанні, проте подачі заявок необхідно оплачувати.

Власний портал для подачі заявок та прийому на навчання мають Katholieke Universiteit Leuven (Бельгія), École Polytechnique Fédérale de Lausanne (Швейцарія), University of Groningen (Нідерланди), University of Oslo (Норвегія).

Слід зазначити, що в деяких випадках онлайн вступ здійснюється зовнішніми системами. Це зумовлено особливістю організації освітнього процесу певної країни. Так, у Великій Британії функціонує сертифікована служба подачі документів - Universities and Colleges Admissions Service, UCAS. Варто зазначити, що King's College London поєднує цей ресурс з власним програмним рішенням - King's Apply, де облікові записи користувачів інтегрується з UCAS. Вступникам надається можливість відстежувати стан заявок, дізнатись чи запросили на співбесіду, отримувати системні сповіщення, подати заявку на проживання та ін.. Подібно до UCAS для всіх неprivатних вищих навчальних закладів в Нідерландах діє схожий сервіс – Studielink, в Данії – Optagelse, Studyinfo.fi – в Фінляндії.

Поширеною практикою є онлайн консультації, що проводяться з метою координування кандидатів для раціонального вибору курсу навчання чи коледжу. Інформація про потреби вступників збирається попередньо за допомогою заповнення онлайн-анкети. Таку послугу для вступників запровадили університети Единбурга, Оксфорд, Кембрідж, університети Лондона. Наприклад, тут можуть порадижити конкретний коледж, якщо вступника цікавить спортивна діяльність, базуючись на тому, який коледж більш розвинений в цьому напрямі. Проте, недоліком є те, що залишається вплив людського фактору.

Utrecht University (Нідерланди) надає онлайн сервіс порівняння магістерських програм. На основі введених користувачем назв освітніх програм формується зведена таблиця формальних вимог. Такий формат має значну

перевагу у порівнянні з переглядом окремих сторінок. Проте така можливість надається лише для одного освітнього рівня.

Онлайн засіб для орієнтації вибору вступника пропонує *École Polytechnique Fédérale de Lausan* (Швейцарія). На основі набору з шести зображень, які обрав користувач система генерує галузі досліджень, які б його могли зацікавити. Очевидно, що такі рекомендації є швидше засобом маркетингу а рівень точності доволі низький.

Також протягом останніх 10-ти років на ринку інформаційних технологій почали активно створювати комерційні програмні рішення націлені на управління вступною кампанією. Сьогодні такі продукти є достатньо популярними серед потенційних клієнтів, серед яких, здебільшого заклади освіти. Аналіз інформації про покупців такого програмного забезпечення (ПЗ) (зі сторінок їхніх офіційних веб-сайтів) свідчить про існуючий попит на такі програмні рішення. Тому доцільно розглянути їх як засоби, що відповідають інформаційним потребам цільової аудиторії.

Ключовими постачальниками ПЗ на світовому ринку програмного забезпечення вважають такі компанії, як Blackbaud, Ellucian Company та Hyland Software. Окрім їхніх продуктів, було розглянуто також найбільш популярні платні рішення за даними сайта Capterra [11]. Функціонал таких систем доволі типовий та включає, зокрема:

- автоматизований процес прийому заявок;
- контроль дедлайнів в ході вступної кампанії, модуль сповіщень для користувачів;
- здійснення оплати за навчання онлайн;
- особистий кабінет користувача;
- зручний та брендований інтерфейс;
- версії мобільних додатків;
- збір, збереження, пошук даних, засоби для звітності та аналітики для адміністраторів.

Натомість, різниця полягає у наступному: групи кінцевих споживачів; кількість користувачів які система здатна обслуговувати; операційні системи на яких можливе розгортання та режим доступу (web-based або installed); вартість.

Серед переваг, з точки зору використання, можна виділити зручність і простоту у використанні, наявність мобільної версії та брендований інтерфейс для кожного окремого закладу освіти. З точки зору роботи адміністраторів беззаперечно такі рішення теж мають переваги, серед яких: швидкий та прозорий варіант ведення записів; можливість використання даних для подальших процесів; інтеграція з іншими інформаційними системами (школи чи коледжу, ERP, управління навчальним процесом ЗВО); зменшення потреби матеріальних та людських ресурсів в процесі проведення вступної кампанії.

Однак більшість рішень орієнтовані на сам процес прийому та опрацювання заяв від вступників. Окрім того залишається проблема конфіденційності та безпеки даних. Також для багатьох навчальних закладів проблемою є висока вартість купівлі та труднощі в обслуговуванні. Платні рішення здебільшого популярні в Сполучених Штатах Америки та лише частково в Європі. Це зумовлює необхідність внесення значних змін в ПЗ, що, наприклад, з точки зору українських ЗВО є фінансово не вигідним.

Тож, з огляду на специфіку процесу вступу до ЗВО України, практичний досвід закордонних університетів або ж використання комерційного програмного забезпечення не можуть бути повністю перенесені та застосовані в українському контексті.

Тому, актуальним є аналіз та пошук інформаційних технологій, які б забезпечили належним чином підтримку прийняття рішень відповідно до інформаційних потреб вступників ЗВО України.

1.3. Аналіз сучасних методів та засобів підтримки прийняття рішень вступників закладів вищої освіти

1.3.1. Аналіз інформаційних систем підтримки прийняття рішень в контексті їхнього застосування під час вступної кампанії

З огляду на визначені задачі дослідження надалі проаналізовано застосування такого типу ІС як системи підтримки прийняття рішень (далі ІСППР). Власне автоматизація процесів підтримки прийняття рішень зумовило появу СППР та різних підходів до їхньої побудови [15, 16, 17].

Проектування, порівняння та вибір альтернативних рішень різними способами для вступників, під час прийняття рішення можна забезпечити саме за допомогою обчислювальних засобів, що застосовують нові альтернативні підходи до формування і прийняття високоякісних рішень в ІСППР. У сучасному етапі розвитку та впровадження інформаційних технологій в різні сфери людської діяльності відбуваються значні зміни, оскільки з'являються потужні технічні ресурси для накопичення та обробки великих обсягів інформації [18]. Тож, інформаційні технології впроваджуються у всіх галузях діяльності суспільства та вдосконалюються відповідно до поставлених задач та цілей, зокрема, обробки та аналізу даних, обчислень та передбачення в режимі реального часу, для допомоги у прийнятті рішень, особливо з метою автоматизації інтелектуальної праці [19, 20, 21, 22, 23, 24, 25].

Проаналізувавши найбільш поширені означення, (які в основному відрізняються методами, моделями, засобами самої СППР) можна узагальнено описати ІСППР як – комп'ютерна інформаційна система, що використовує дані і моделі для розв'язання неструктурованих і слабоструктурованих проблем, дає змогу користувачам моделювати й аналізувати інформацію з метою підвищення ефективності прийняття певних специфічних рішень у різних галузях людської діяльності [19].

Сам Термін СППР з'явився у 70-х роках і належить А. Горрі і М. Скотт-Мортону [15]. Еволюція СППР передбачала кілька поколінь та зміну визначених типів таких систем. Так розподіл на два типи за Стевенем Альтером (Steven Alter) у період 1976—1980 років передбачав системи, орієнтовані на дані, які просто здійснюють вибирання інформації та системи, орієнтовані на моделі, що дійсно дають змогу підтримувати прийняття рішень. Пізніше такий поділ розширив Пауер (D. J. Power), доповнивши її новими типами СППР [26].

ІСППР на кожному етапі свого розвитку були тісно пов'язані з досягненнями обчислювальної техніки та інформатики. Сьогодні ж інтелектуалізація взаємодії користувача з такими системи та застосування інтелектуальних методів та алгоритмів опрацювання даних призвело до появи нового широкого класу – інтелектуальних інформаційних систем підтримки прийняття рішень.

Інтелектуальна система – це штучна система, яка працює як агент, тобто система сприймає своє середовище, діє в середовищі та взаємодіє з іншими агентами, і демонструє раціональну поведінку, тобто система діє раціонально (щоб максимізувати успіх своїх завдань) і виявляє раціональне мислення (обґрунтовує переконання через міркування) [26].

Таке визначення передбачає, що в загальному, інтелектуальна система працює як агент, що означає, що вона здатна діяти, використовуючи дані про стан навколишнього середовища. Використання датчиків і приводів, які можуть бути реальними або віртуальними, відокремлює тіло інтелектуальної системи від решти середовища. Поняття агента використовувалося різними авторами для ідентифікації характеристик інтелекту системи [26].

Так, використовуючи аналітику даних та алгоритми машинного навчання (МН), ІСППР може аналізувати великі обсяги даних, щоб приймати більш обґрунтовані та об'єктивні рішення в межах вступної кампанії. Така система може виявляти закономірності, тенденції та кореляції в даних вступників, які

можуть бути неочевидними для людей, що призводить до покращення критеріїв відбору та результатів прийняття рішень.

Такі системи можна налаштувати для адаптації до конкретних потреб і вимог різних ЗВО. Заклади можуть визначати свої особливості та потреби вступу, на основі своїх унікальних цілей, пріоритетів та профілів студентів. Також, перевагою є те що ІСППР можуть обробляти великі обсяги заявок і масштабуватися, щоб врахувати зміни під час пікових періодів вступу, зокрема в рейтингових списках.

Розробляючи агент попередньої обробки як частину інтелектуальної системи, систему можна зробити більш адаптивною до змін у джерелах даних або вимогах. Процеси попередньої обробки можна автоматизувати, щоб автоматично обробляти нові дані, зменшуючи потребу в ручному втручанні та оптимізуючи загальний робочий процес.

Інтелектуальна система також може відігравати і роль консультанта. У цьому випадку користувач приймає рішення про те, які дії слід виконувати в середовищі, і він також виконує ці дії. Вважається, що агент діє раціонально, якщо рішення, які він приймає щодо своїх дій, спрямовані на оптимізацію показника ефективності. Поняття раціональності в прийнятті рішень було описано в ШІ за допомогою принципу раціональності, а також використовується в інших дисциплінах, таких як філософія чи економіка [26].

Раціональні дії можна проаналізувати з зовнішньої точки зору штучної системи, не знаючи внутрішніх механізмів, які породжують таку поведінку. Показник ефективності визначає мету системи та може використовуватися для кількісної оцінки ступеня успіху, тобто для визначення того, наскільки добре система виконує завдання, для якого вона була розроблена. Зовнішній спостерігач може зробити висновок, що машина діє раціонально, якщо вона послідовно виконує дії, які успішно оптимізують міру продуктивності.

Щоб полегшити аналіз і проєктування інтелектуальних систем варто визначити тип процесів, які можуть породжувати таку поведінку. У цьому

контексті інтелектуальна система розуміється як система, тобто як інтеграція набору компонентів, які організовані відповідно до певної мети.

Тож необхідною є функціональна декомпозиція, що розбиває загальну поведінку інтелектуальної системи на менші частини, щоб визначити основні процеси, які зазвичай включають інтелектуальні системи.

Так, однією з важливих функцій інтелектуального агента є його здатність розуміти оточення. Глобальну функцію розуміння середовища включає, з одного боку, сприйняття середовища, метою якого є інтерпретація даних із середовища, щоб сформувані уявлення про поточну ситуацію. З іншого боку, розуміння навколишнього середовища може також включати оцінку поточної ситуації шляхом розгляду ширшого контексту (наприклад, міркування про минуле чи майбутнє) і шляхом присвоєння значення, яке виражає, наскільки хороша чи погана ситуація.

Іншою функцією інтелектуального агента є дія в навколишньому середовищі. Це включає, по-перше, рішення про те, які дії слід виконати в поточній ситуації, і, по-друге, виконання цих дій у середовищі. Функція, що полягає у спілкуванні передбачає можливість спілкування з іншими агентами, надсилаючи та отримуючи повідомлення за допомогою мови, письмових повідомлень, графічних зображень або іншими засобами.

Таким чином, використання ПСППР як інформаційна технології може дозволити вирішити основні задачі, що мають на меті підтримку прийняття рішень вступників. Роль системи полягає в тому, щоб давати поради, надаючи корисну інформацію для полегшення таких рішень користувачеві. Система, яка приймає роль консультанта, може виконувати завдання (такі як планування, проектування або розподіл ресурсів), які підказують користувачеві, які дії потрібно виконати.

1.3.2. Аналіз наукових досліджень методів та засобів підтримки прийняття рішень вступників закладів вищої освіти

Аналіз існуючих загальнодоступних інформаційних веб-ресурсів, свідчить про недостатній рівень забезпечення підтримки прийняття рішень вступників ЗВО України. Тож, доречно проаналізувати також існуючі наукові дослідження у цьому напрямі.

Здійснивши аналіз сучасних вітчизняних наукових досліджень можемо виділити певні їхні напрями, зокрема:

- дослідження соціальних мереж. У сучасних роботах досліджено соціальні мережі як засіб для підвищення ефективності вступної кампанії. Зокрема, було розроблено концепцію формування системи маркетингових комунікацій ЗВО в мережі [27-30];

- дослідження, що стосувались освітніх веб-форумів та розглядали їх як ефективне джерело задоволення інформаційно-комунікаційних потреб учасників освітньої діяльності. Зокрема, авторами було опрацьовано інформаційне наповнення веб-форумів, в залежності від періоду вступної кампанії, та проведено його комп'ютерно-лінгвістичний аналіз. Запропонована модель може бути застосована для ефективного планування освітніх послуг та підтримки зворотного зв'язку з вступниками через комунікативну діяльність [31-33];

- роль офіційних веб-сайтів ЗВО як засіб інформаційного забезпечення вступників [34-35];

- автоматизація процесу вступу та моніторинг вступної кампанії. Зокрема, було запропоновано автоматизовану систему моніторингу для відслідковування вступної кампанії ЗВО, в основу якої покладено аналіз ІС «Конкурс» (<https://www.vstup.info>) та запропоновано користувацький інтерфейс для перегляду рейтингових списків та заявок вступників та їх зручного впорядкування за допомогою відповідних ознак [36];

- також іншими науковцями було досліджено оптимізацію існуючих систем та інформаційних зав'язків у ЗВО та супровід вступної кампанії та запропоновано нову модель інформаційних зав'язків, що автоматизує в ЗВО процес роботи із заявами вступників [37].

Існує чимало досліджень, що аналізують особливості профорієнтаційної діяльності та процес прийняття рішень вступниками, проте їхні результати з огляду на галузь досліджень не пропонують застосування інформаційних технологій, а обмежені лише теоретичними висновками [38].

Актуальність застосування інформаційно-комунікаційних технологій в межах вступної кампанії ЗВО, в сучасних наукових дослідженнях розглядається та пояснюється в контексті маркетингових стратегій навчального закладу, організації інформаційних потоків (інформаційного супроводу вступників, забезпечення інформаційних потреб) та профорієнтаційної діяльності та підтверджують необхідність впровадження нових механізмів аналізу та розробки спеціалізованих моделей для кожного навчального закладу [39], [40]. Проте, більшість релевантних вітчизняних досліджень не мали подальшого розвитку.

Аналіз закордонних сучасних наукових досліджень, що мають на меті підтримати прийняття рішень вступників дозволив виділити три основні напрями – передбачення успішності вступу, надання рекомендацій освітніх програм або закладу освіти, та ті що досліджують фактори, що впливають на вибір вступника [41].

Нижче виокремлено та проаналізовано існуючі методи передбачення вступу вступників до певного університету. Так, S. Sridhar та інші автори запропонували модель стекінгового ансамблю, побудовану для визначення ймовірностей вступу до ЗВО [42]. Дослідження включає різні фактори, пов'язані з вступником, зокрема, його дослідницький та виробничий досвід тощо. Перевагою в даному випадку є те, що автори оцінювали кілька алгоритмів МН, включаючи інші методи глибокого навчання. Оскільки запропонована модель легко перевершує всі інші моделі і забезпечує високу точність.

В іншій роботі автори рекомендують використовувати максимально великий набір даних і методи обробки природної мови для оцінки есе і листів як нових деталей про вступників, щоб підвищити точність передбачення [42]. В іншому дослідженні вчені розгорнули регресійну модель з градієнтним підсиленням, яка враховує академічні успіхи студента та його рейтинг в університеті. Вони запропонували ймовірність вступу до цього університету як вихідний результат [43].

Використовуючи різні типи алгоритмів ШІ, інші автори запропонували фреймворк Graduate Admissions Prediction [44]. Крім того, було запропоновано користувацький інтерфейс, щоб побачити результат. Але із запропонованої роботи користувачі можуть визначити шанси на отримання місця без можливості отримати список університетів, в які вони можуть вступити.

В іншому дослідженні вчені звернули увагу на те, що вступники не завжди чітко розуміють, до якого навчального закладу вони можуть вступити, тому можуть витратити фінансові та часові ресурси [45]. Тому автори запропонували модель, яка надає майбутньому студенту інформацію про його шанси вступити до конкретного університету. Щоб зробити програму доступною для користувачів, автори розробили графічний інтерфейс.

М. А. Khan, М. Dixit та А. Dixit наголошують на необхідності заощадити час вступників та підвищити їхні шанси на вступ на певні магістерські програми [46]. Вони порівняли різні алгоритми МН, щоб визначити найбільш точні. Сформовані кластери схожих вступників були позначені, якщо їх було рекомендовано для вступу.

З метою підтримки прийняття рішень вступників, у дослідженнях достатньо часто розглядають рекомендаційні системи, що представляють собою підклас системи фільтрації інформації, яка будує рейтинговий список об'єктів, яким потенційний студент може віддати перевагу. Дані профілю користувача (набір уподобань та інтересів) та всі можливі набори даних використовуються для надання результатів процесу рекомендацій. Відповідно до дослідницьких

завдань (цілей) останніх досліджень, їх зазвичай поділяють на певні групи, такі як: вибір освітніх програм, вибір курсів, довгострокове академічне планування [47-50].

Іншим напрямом є аналіз та визначення тих факторів, що мають вплив на вступ до ЗВО та його результати. Основною метою цих досліджень є аналіз моделей вступу до ЗВО та факторів, що впливають на ці моделі, для підвищення загальної ефективності освітньої галузі [51-53]. Автори виявили знання з академічного профілю успішних вступників попереднього року з метою використання цих даних для майбутніх вступників. Вартість навчання, місцезнаходження, рейтинг закладу, матеріально-технічна база, можливості працевлаштування можуть мати значний вплив на результати прийняття рішень.

Проведений аналіз методів досліджень свідчить про те, що більшість досліджень, які мають на меті передбачення успішності вступу або рекомендації освітніх програм чи навчального закладу, базуються на інтелектуальному аналізі даних. У Таблиці 1.1 представлено розподіл таких досліджень за завданнями.

Таблиця 1.1. Розподіл проаналізованих наукових досліджень за завданнями

Основні задачі наукових праць	Основні результати наукових праць
Класифікація	<ul style="list-style-type: none"> • моделі передбачення, що визначають шанси студента на вступ до коледжу [54], університету [55]; • класифікація з метою визначення, чи може студент бути зарахованим до цільового університету чи ні [56];

	<ul style="list-style-type: none"> • визначення чинників, які мають високий вплив на прийняття рішень вступниками [54], [42]; • моделі змінних, які можуть передбачити майбутні події при (повторному) вступі до університету [57]; • рекомендації у виборі факультетів.
Передбачення	<ul style="list-style-type: none"> • передбачення шансів зарахування студента на певну програму коледжу чи університету [42], [43], [44], [45], [58], [59]; • передбачення прохідного балу до коледжу [60]; • регресійні моделі для передбачення вступу до вишів [61]; • розпізнавання та таргетування університетів, які найкраще підходять для профілю вступника, та передбачення шансів майбутнього студента на вступ до університету [49]; • передбачення ймовірності отримання гранту [62].
Кластеризація	кластеризація подібних університетів [55], фактори, що впливають на тенденцію вступу [63], [51].

Завдання, які ставилися, дослідники пояснювали кінцевою метою, як-от: економія часу та грошей вступників (кожна заява є платною), уникнення їхньої неуспішності в майбутньому навчанні, уникнення дострокового відрахування та виключення загалом, підвищення задоволеності обраною освітньою програмою, підвищення рівня працевлаштування. Для екстраполяції закономірностей та нових знань на основі зібраних дослідниками даних використовувалися різні методи інтелектуального аналізу даних.

Розглянуті дослідження проводилися відповідно до потреб вступників на перший курс [45, 54, 58, 64, 65], повторного вступу [57], магістратури [63], [43, 52, 56, 61, 66], аспірантів [42, 44, 47, 55], [41, 49], для студентів, які подають документи на вступ до закордонних університетів [30].

Дослідники використовували набори даних, які містили

- дані про минулі зарахування: наприклад, минулі дані про розподіл місць у коледжах, дані, наявні на Edulix (<https://www.edulix.com>) [42, 55, 65];
- набір даних про вступників попереднього року на подібних програмах [65], [51, 45, 47, 59, 61];
- результати опитувань [53, 57, 58, 63];
- дані про студентів: база даних студентів-випускників [43], академічні дані успішних студентів, які вже отримали можливість навчатися [65], академічна інформація про попередніх студентів бакалаврату та результати подачі заяв до цільових університетів [56], деяка особиста та академічна інформація про конкретного студента;
 - консультанти університетів [56];
 - поточні дані користувачів [48, 64, 65, 66].

Проте, однією із актуальних задач залишається підвищення точності результатів розв'язання задачі класифікації у кожному випадку поставлених задач. Базуючись на перевагах методів МН, як природного продовження традиційних статистичних підходів, які зосереджені в першу чергу на

передбаченнях, автоматичному визначенні закономірностей у даних і виконанні завдань, надалі досліджено саме ці методи [67].

Питаннями інтелектуального аналізу даних займалися такі провідні українські вчені як С.О. Субботін (методи опису, побудови та застосування моделей знань в системах штучного інтелекту та підтримки прийняття рішень), Р.О. Ткаченко (ансамблеві методи передбачення і класифікації даних), Н.Б. Шаховська (ієрархічні методи класифікації даних різних обсягів), Є.В. Бодянський (методи нечіткої кластеризації даних різних обсягів), Бідюк П.І. та Говорущенко Т. О. (інтелектуальні інформаційні системи підтримки прийняття рішень).

Перевага традиційних статистичних підходів полягає в тому, що вони прості для розуміння. З іншого боку, результати МН часто важко інтерпретувати. Крім того, алгоритми МН передбачають попередню обробку даних, навчання на наборах даних, вимагають великих наборів даних та ітеративне вдосконалення щодо реальної проблеми. Методи МН також можуть призвести до створення моделі, надто тісно пов'язаної з основним набором даних. Це явище може обмежити можливість узагальнення моделі на різні набори даних, а отже, і передбачення. Щоб уникнути цієї проблеми, необхідний відповідний баланс між навчальним набором і набором перевірки [67].

Результати аналізу наявних досліджень показують, що впровадження методів і прийомів машинного навчання дає високу точність результатів при їхньому використанні для підтримки вступників. Отримані результати є досить точними та дозволяють застосовувати методи дослідження при побудові інформаційних систем. Аналіз наукових досліджень щодо реалізації ІСППР свідчить про зростаючу актуальність розробки таких систем, що передбачає створення ІСППР на основі сучасних методів ШІ, у поєднанні з використанням експертних систем.

1.4. Формування вимог до інформаційної технології підтримки прийняття рішень вступників ЗВО

Для того аби забезпечити вирішення задачі підтримки прийняття рішень вступників, необхідно здійснити декомпозицію задачі на детальніші процеси та описати їх, враховуючи подані раніше відповідними фазами (етапами) прийняття рішення вступника в контексті вступної кампанії. Прив'язати задачі до конкретних цільових подій у відповідності до часових періодів вступної кампанії не доречно, оскільки вступник може цікавитись та обирати освітню програму ще задовго до початку. Натомість необхідно враховувати ієрархію подій і проміжних цілей. Таким чином відповідно до процесів руху до встановленої мети, ієрархічний процес ухвалення рішення вступником можна описати цільовими подіями, поданими в Таблиці 1.2. (кожну дію, що відбувається в процесі ухвалення рішення, називають цільовою подією).

Таблиця 1.2. Ієрархічний процес ухвалення рішення вступниками

Фаза прийняття рішення	Цільова подія
Фаза аналізу та постановки задачі.	профорієнтація та прийняття рішення вступниками щодо вибору напряму та ЗВО – протягом навчання в школі та до реєстрації на тестування
	реєстрація електронних кабінетів, на зовнішнє незалежне оцінювання НМТ
Проектування альтернатив з огляду на обмеження та критерії	складання вступних випробувань (ЗНО/НМТ, творчі конкурси)
	подання заявок до ЗВО та на ОП
Оцінка та вибір кращої альтернативи	вибір ОП після надання рекомендацій (за широким конкурсом)
	подання оригіналів документів та зарахування
	погодження питань щодо навчальної діяльності ЗВО, соціально-побутових

	УМОВ
--	------

Таким чином, підзадачами процесу прийняття рішення вступника є:

- ознайомитись та проаналізувати з напрямами, освітніми програми та здійснити їхній вибір;
- врахувати вимоги вступу та власні можливості (щодо навчання);
- обрати освітні програми до здачі вступних випробувань для подачі заявок (щоб написати тести);
- обрати освітню програму після здачі вступних випробувань для подачі оригіналів.

Нижче подано аналіз і визначення обмежень відповідно до інформаційних потреб вступника, які можуть виникнути при прийнятті рішення. Визначення слабких місць в процесі прийняття рішень дає змогу підібрати методи та засоби для виконання обчислювальних та логічних процедур з метою підсилення таких місць[6]. Потреба застосування методів підтримки прийняття рішень може бути зумовлена такими загальними моментами:

1. Неможливість прогнозувати процес.
2. Труднощі в об'єднанні атрибутів та цілей.
3. Труднощі з обробкою даних, необхідних для прийняття рішень.
4. Проблеми з аналізом і формуванням логічного висновку (прийняття рішення).
5. Труднощі з візуалізацією результатів.
6. Неточність процесу евристичних суджень [6].

Проведений аналіз особливостей та принципів процесу прийняття рішень вступників ЗВО в періоді вступної кампанії дає змогу визначити прийняття рішень вступниками як таке, що приймається в умовах погано структурованої проблеми, в умовах невизначеності та має прямий вплив на життєву його траєкторію. Тож, виникає потреба забезпечення підтримки прийняття рішень, з метою змінити умови прийняття рішень – з таких, що здійснюються в умовах

невизначеності в такі що здійснюються в умовах ризику – коли імовірність можна оцінити за умови її об'єктивності, та – з умов обмежених можливостей – в такі, де ці можливості розширені. Очевидно, що реалізація таких змін можлива за умови використання інформаційних технологій, що обґрунтовує необхідність пошуку та впровадження найбільш ефективних методів та засобів підтримки прийняття рішень вступників.

Тож, основною метою такої системи є підвищення рівня обізнаності вступників ЗВО України відповідно до їхніх інтересів щодо можливих альтернатив вибору освітніх програм та успішності вступу. Враховуючи проведений аналіз предметної області, завдань та обмежень відповідно до інформаційних потреб, що виникають перед вступником, можна виділити наступні вимоги ІСППР:

1. Інтеграція даних і доступ. Передбачає можливість інтегрувати дані з багатьох джерел, включаючи бази даних, електронні таблиці та зовнішні джерела та підтримку оновлень даних у режимі реального часу або за розкладом, щоб користувачі, які приймають рішення, мали доступ до актуальної інформації.

2. Аналіз даних. Необхідно застосувати інструменти для виконання визначених типів аналізу (залежно від обраних методів).

3. Можливості підтримки прийняття рішень для надання рекомендацій, розуміння та оперативної інформації для сприяння процесу прийняття рішень.

4. Адаптація системи до конкретних інформаційних потреб встановлених окремими користувачами та відповідно до вимог ЗВО.

5. Інтерактивність і залучення користувачів. Інтерактивні функції, такі як деталізація, фільтри та додаткова інформація, які дозволяють користувачам досліджувати дані та отримувати глибші відомості.

6. Продуктивність і масштабованість. Швидкий час відповіді на запити даних для забезпечення безперебійної взаємодії з користувачем. Масштабованість для обробки збільшення обсягу даних, одночасної роботи користувачів і навантаження на систему без зниження продуктивності.

На основі проведеного аналізу особливостей процесу прийняття рішень вступників ЗВО та принципів роботи ІСППР, можна зробити висновки, що основним задачами такої системи повинні бути:

- сприяння підтримки прийняття рішень вступників щодо вибору освітньої програми та здатності в коротший період приймати якісні рішення;
- доповнення існуючих технологій (інструментів), що існують сьогодні для вступників новими, з продуктивнішими можливостями;
- спрощення для вступників проходження етапів прийняття рішень, таких як збір інформації, проєктування та відбір альтернатив;
- упорядкування й полегшення аналізу можливих шляхів вибору освітньої програми;
- підвищення компетентності знань вступників щодо цього рішення засобами інформаційних технологій.

Тож, основою такої системи можна визначити дві компоненти: рекомендація альтернативних освітніх програм та передбачення успішності вступу. Для побудови моделі ІС наступним питанням аналіз та вибір методів та алгоритмів МН та способів виконання функцій збирання, зберігання, оброблення, передавання та використання знань. Визначені у попередньому розділі декомпоновані задачі та глобальні функції ІСППР дозволить далі обрати найбільш ефективні методи та алгоритму для реалізації системи. Тож надалі необхідним є розробка моделі ІСППР вступників ЗВО та технології збору та опрацювання даних для можливості проведення подальшого аналізу на їхній основі.

Висновки до розділу

У першому розділі здійснено аналіз інформаційних технологій підтримки прийняття рішень вступників ЗВО. Серед вагомих отриманих результатів виконаного аналізу слід виділити наступні:

- аналіз принципів та особливостей проведення вступної кампанії ЗВО України підтверджує складність та значущість майбутніх її результатів, що можуть вплинути на рівень задоволеності вибором студента та на оцінку ефективності діяльності ЗВО загалом;
- кількісний та якісний набір вмотивованих студентів, що здійснюється відповідно до специфічних цілей та можливостей ЗВО, що відображає результативність вступної кампанії, формується залежно від багатьох чинників, один з яких є вибір вступника щодо конкретної освітньої програми;
- свідомий і виважений вибір освітньої програми є важливим етапом вступної кампанії, що сприятиме високій вмотивованості та успішності студентів. Такий вибір є ціллю вступної кампанії ЗВО та відповідає інтересам та поставленій меті вступника, тож потребує необхідної підтримки, зокрема з використанням інформаційних технологій;
- аналіз сучасних відкритих веб сервісів різних типів, зокрема сайтів ЗВО, освітніх порталів, офіційних веб ресурсів на національному рівні свідчить про необхідність розробки додаткових інформаційних технологій підтримки прийняття рішень вступників. Водночас, практичний досвід закордонних університетів та комерційного програмного забезпечення не є актуальним враховуючи особливості прийому на навчання в ЗВО України;
- аналіз поняття, особливостей побудови та застосування ІСППР, а також результатів наукових досліджень вітчизняних та закордонних авторів, підтверджують актуальність розроблення та використання інтелектуальних засобів підтримки прийняття рішень під час вступної кампанії;
- декомпозиція цілі та визначення основних обмежень відповідно до інформаційних потреб вступника дали змогу виділити вибір освітньої програми та передбачення успішності вступу як основні задачі, що потребують підтримки прийняття рішень.

Розділ 2. Модель інформаційної системи підтримки прийняття рішень вступників ЗВО та засоби її реалізації

У другому розділі розроблено модель інформаційної системи підтримки прийняття рішень вступників ЗВО, яка використовує поєднання методу надання рекомендацій та удосконаленого методу передбачення успішності вступу, що підвищує рівень обізнаності вступників ЗВО України відповідно до їхніх інтересів.

Розроблено технологію збору та опрацювання даних для можливості проведення подальшого аналізу на їхній основі для надання рекомендацій освітніх програм. Також розроблено технологію збору та опрацювання даних та їхнього використання у роботі для апробації розробленої моделі машинного навчання. Здійснено вибір та обґрунтування критеріїв оцінювання ефективності роботи класифікаторів на основі машинного навчання.

2.1. Модель інформаційної системи підтримки прийняття рішень вступників ЗВО

Проведений аналіз вступної кампанії та процесу прийняття рішень вступників показує його складність та вплив на результативність вступної кампанії, що визначає необхідність впровадження технології для підтримки прийняття рішень вступників. Вирішення таких задач, як надання рекомендацій вступникам щодо вибору освітніх програм та передбачення успішності їхнього вступу, потребує обробки та аналізу даних, обчислень та передбачення для допомоги у прийнятті таких рішень, особливо з метою автоматизації інтелектуальної праці.

Як інструмент структурного аналізу та проектування інформаційної системи використано методологію діаграми потоків даних (DFD, Data Flow Diagram).

Для короткого опису проектованої інформаційної системи на Рис.2.1 Подано верхневрівневу контекстну діаграму.

На контекстній діаграмі (Рис. 2.1) відображено основний процес, зовнішні сутності, які з ним взаємодіють, зокрема: «Користувач», «Аналітик даних» та «БД інформаційних систем ЗВО» та основні потоки даних, якими обмінюються зовнішні сутності з інформаційною системою.

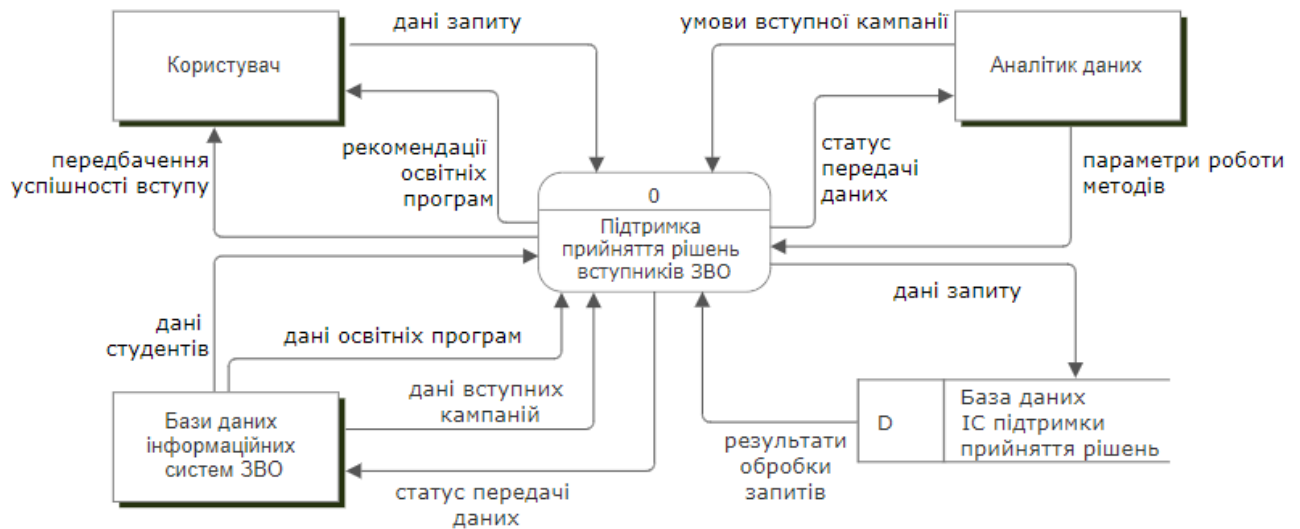


Рис.2.1 Контекстна діаграма інформаційної системи підтримки прийняття рішень

Розроблена модель передбачає взаємодію користувача з ІСППР для збору його даних та видачі результатів. Для реалізації поставлених задач така система потребує попередньої обробки різних наборів даних, тому іншою зовнішньою сутністю, яка взаємодіє з системою є аналітик даних. Джерелом даних, покладених в основу ІСППР є бази даних ЗВО.

Для детальнішої демонстрації того, як кожен процес проектованої інформаційної системи перетворює відповідні вхідні дані у вихідні та які відносини між цими процесами нижче подано декомпозицію діаграми.

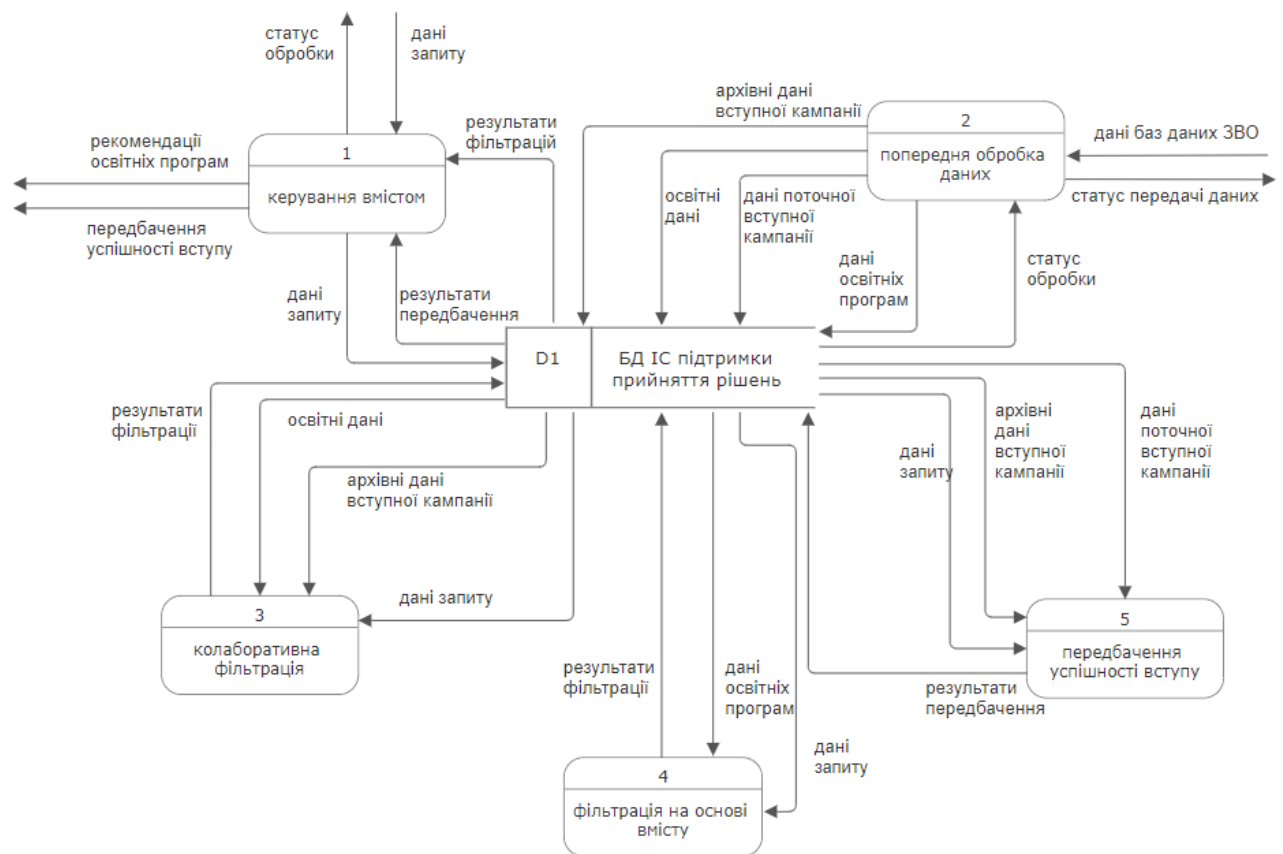


Рис.2.2 Діаграма потоків даних першого рівня

Розроблена модель ІСППР на основі отриманих від користувача даних щодо освітньої програми, яка найбільше цікавить вступника та значення його конкурсного балу надає можливість отримати рекомендації щодо альтернативних освітніх програм. Колаборативна фільтрація надає перелік освітніх програм, які обрали з іншими пріоритетами вступники, які були визначені як найбільш подібні до користувача та мали високий показник задоволеності свого вибору щодо освітньої програми. Результати таких рекомендації розширено також рекомендаціями на основі фільтрації вмісту, що відбирає альтернативні освітні програми на основі суміжності їхніх кластерів. Користувач має можливість оцінити свої шанси на вступ за допомогою методів машинного навчання.

Подальші етапи розробки компонентів інформаційної системи здійснено на основі реальних наборів даних Національного університету «Львівська політехніка» (НУЛП). Усі використані дані було анонімізовано.

Розроблена модель інформаційної системи зображена також за допомогою діаграми діяльності на Рис. 2.3.

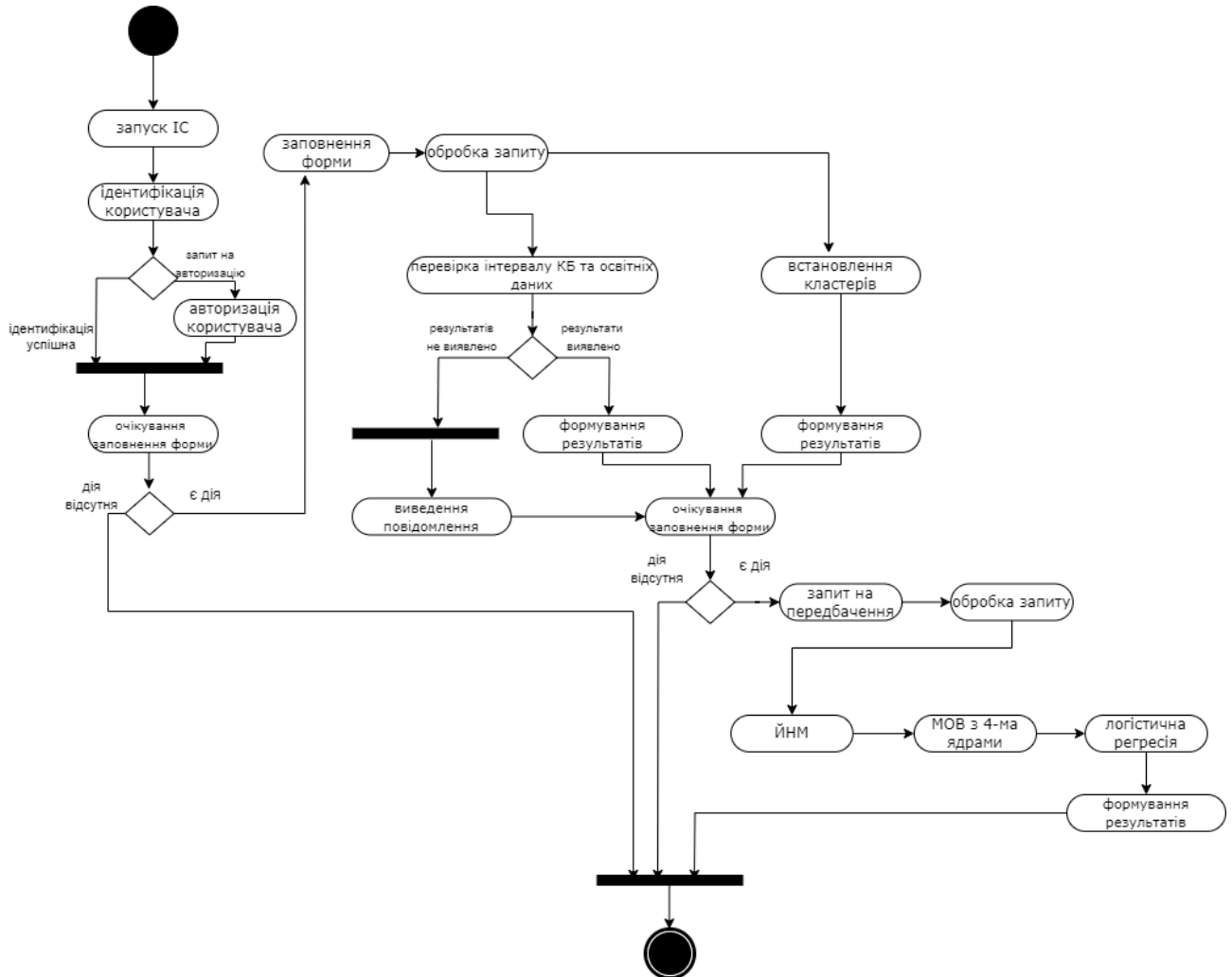


Рис.2.3 Діаграма діяльності інформаційної системи підтримки прийняття рішень вступників ЗВО

Таким чином, розроблена модель інформаційної системи підтримки прийняття рішень вступників ЗВО, на відміну від існуючих використовує поєднання методу надання рекомендацій та удосконаленого методу передбачення успішності вступу, що може підвищити рівень обізнаності вступників ЗВО України відповідно до їхніх інтересів.

2.2. Метод сумісного використання колаборативної фільтрації та фільтрації на основі вмісту для надання рекомендацій вступникам ЗВО альтернативних освітніх програм

2.2.1. Вибір та обґрунтування методів надання рекомендацій освітніх програм вступникам ЗВО

Сьогодні достатньо популярним та ефективним засобом для виявлення знань і прийняття рішень в різноманітних сферах людської діяльності є рекомендаційні системи (РС). Використання РС з метою фільтрування інформації онлайн є особливо актуальним з огляду на часті зміни вподобань користувачів та зростаючій популярності персоналізації контенту [68-71].

Ціллю алгоритмів РС є надання пропозицій релевантних елементів користувачам, тож кінцевим результатом роботи такої системи у даному випадку є рекомендація освітньої програми, яка найбільшою мірою відповідає інтересам вступника, тобто користувача. Тож, надалі досліджено можливі методи реалізації компоненти надання рекомендації для подальшого її використання з метою підтримки прийняття рішень.

Формування рекомендацій за рахунок високої популярності елементу серед усіх користувачів є найбільш простим підходом. Враховуючи одну або декілька ознак, визначається популярність елементів, перелік яких і надається користувачу. Наприклад, цей підхід може бути достатньо ефективним у випадку рекомендації статей онлайн-ресурсу, що мають найбільшу кількість переглядів.

Побудова рекомендації в такому підході здійснюється базуючись на вподобаннях усієї аудиторії, яка є достатньо різноманітна. Це, в свою чергу, позбавляє рекомендації відповідності персональним вподобанням користувача, що є недоліком. Використовувати для надання рекомендацій вступнику такий узагальнений підхід є неефективно, оскільки врахувати інтереси очевидно є необхідним. Водночас, це може сприяти підвищенню рівня незбалансованості конкурсу на окремі освітньої програми. Тому, надалі розглянуто більш складні

та, відповідно, ефективніші підходи. Тому, надалі розглянуто більш складні та відповідно більш ефективніші підходи.

Найпоширеніші основні парадигми рекомендаційних систем - це рекомендації на основі фільтрації вмісту, колаборативна фільтрація та гібридна фільтрація, яка комбінує обидві попередні. Те, яким чином інформація була відібрана для користувач відрізняє такі типи РС між собою [72].

Колаборативна фільтрація передбачає використання методів, які, для створення нових рекомендацій, ґрунтуються виключно на минулих взаємодіях між елементами та користувачами. Ці взаємодії зберігаються в так званій «матриці взаємодій користувач-елемент». Так, головна ідея, яка регулює такі методи, полягає у виявленні таких попередніх взаємодій між користувачем та елементом, та виявленні на їх основі подібних користувачів та/або елементів наданні рекомендацій.

На відміну від колаборативної фільтрації, підходи на основі фільтрації вмісту полягають у використанні додаткової інформації про користувачів та/або елементи. Так, ідея методів, що базуються на фільтрації вмісту, полягає в побудові моделі на основі доступних «функцій», що відображають наявну взаємодію між користувачем і елементом. При цьому необхідно встановити модель, яка базуватиметься на функціях користувача та/або елемента, які є в наборі даних («вміст» методу «на основі вмісту») [73-75].

Гібридні методи – це сукупність двох або більше методів, які використовуються в поєднанні для вирішення обмежень окремих методів рекомендацій. Включення різних технік може бути виконано різними способами. Гібридний алгоритм може включати результати, отримані за допомогою окремих методів, або він може використовувати фільтрацію на основі вмісту в колаборативній фільтрації, або ж використовувати колаборативну фільтрацію у фільтрації на основі вмісту. Це гібридне включення різних методів, як правило, призводить до підвищення продуктивності та підвищення точності в багатьох рекомендаційних системах [76].

Розроблена модель передбачає використання історичних даних вступу вступників, а саме пріоритетів заявок, що визначатимуть користувачів зі схожими вподобаннями. Для підвищення точності та якості наданих рекомендацій запропоновано використати також конкурсний бал та освітні дані їхньої життєвої траєкторії як студентів і таким чином оцінити міру рекомендації. Водночас, для розширення кінцевого переліку та, відповідно підвищення рівня обізнаності вступників, запропоновано додати рекомендації фільтрацією на основі вмісту – використовуючи кластери ключових слів описів ОП для визначення їхньої подібності (Рис. 2.4).

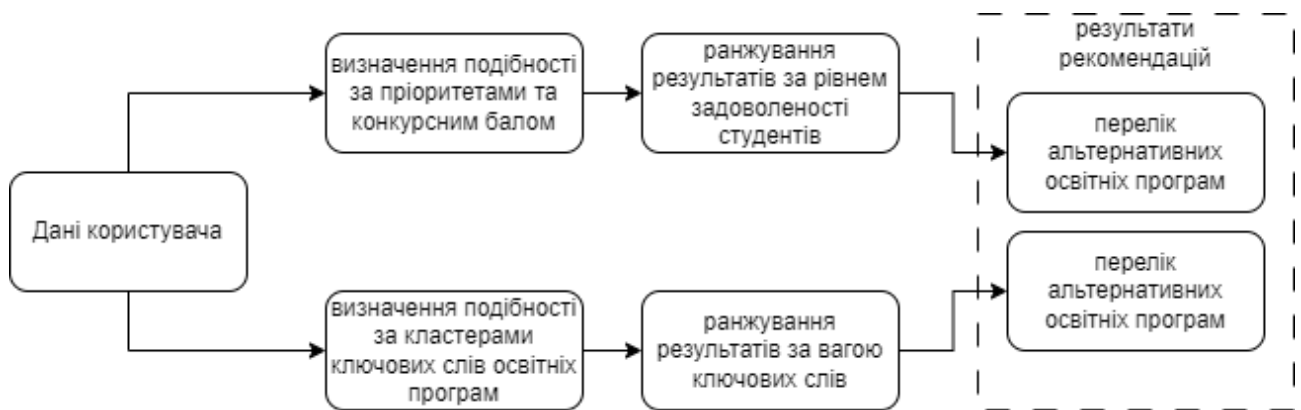


Рис.2.4 Метод сумісного використання колаборативної фільтрації та фільтрації на основі вмісту

Суть такої моделі полягає в тому, щоб подолати обмеження індивідуальних методів рекомендацій і покращити загальну якість рекомендацій. Такий підхід дасть змогу уникнути проблеми холодного старту, розрідженості та незбалансованості та підвищить рівень обізнаності вступника, рекомендуючи нові елементи та, відповідно, в цілому підвищить ефективність роботи технологій надання рекомендацій.

В основі методу колаборативної фільтрації запропоновано використати як матрицю взаємодії користувача та елемента дані заявок вступників, де елементами вважаємо ОП, користувачами - вступників, а поставлені пріоритети – вподобанням елементів. Для більшої релевантної отриманих результатів запропоновано ранжувати результати за показником задоволеності що аналізує

його життєву траєкторію як студента. Таким чином кінцевий користувач фактично отримує на цьому етапі рівень задоволення здійсненого вибору схожих вступників. Основними критеріями успішності вибору обрано успішність у навчанні, наукову діяльність та продовження навчання для здобуття ступеня магістра – як такі, що в найбільшій мірі описують наслідки здійсненого вибору. Тож колаборативна фільтрація передбачає наступні етапи:

1. Збір даних про вподобання користувача. Користувач повинен обрати з переліку освітню програму, для якої він визначає найвищий пріоритет та внести свій конкурсний бал.

2. Визначення подібності. Перевірка архівних даних вступу та користувача. Вибірка, що описує архівні дані вступу передбачає визначення для кожної заявки властивості, сформовані за наступними ознаками: освітня програма, пріоритет (перший), інтервал конкурсного балу. Результат включає освітні програми інших заявок вступників (архівних).

3. Відбір елементів. Для визначених на попередньому етапі користувачів класу відбір інших їхніх заяв (за наявності).

4. Ранжування результатів за показником задоволеності інших користувачів (студентів). Визначається за результатами показника, що оцінює успішність в навчанні, наукову діяльність та продовження навчання в магістратурі. Ранжування передбачає відбір користувачів, де міра задоволеності більша середньо.

5. Формування рекомендацій.

Метод фільтрації на основі вмісту базується на даних про елементи, тож необхідно забезпечити відповідні характеристики ОП. Оскільки ОП характеризуються їхніми текстовими описами, потрібен перелік властивостей у зручній для використання формі. Для цього було обрано ключові слова освітніх програм та їхні кластери, які покладені в основу рекомендацій освітніх програм.

Етапами фільтрації на основі вмісту є наступні:

1. Збір даних про вподобання користувача. Користувач повинен обрати з переліку освітню програму.
2. Визначення подібності. Пошук кластерів ключових слів ОП.
3. Ранжування результатів. Визначається за результатами відбору кластерів ключових слів з високою вагою.
4. Формування рекомендацій.

Процеси збору та попередньої обробки даних описано в наступних підрозділах.

2.2.2. Технологія збору та попереднього опрацювання даних для колаборативної фільтрації

Як було зазначено в попередньому підрозділі, метод колаборативної фільтрації передбачає використання архівних даних вступу та освітніх даних студентів. На Рис. 2.5. відображено процес збору та попередньої обробки даних, який забезпечить їхній подальший аналіз та використання та надання рекомендацій вступнику. Технологія збору та попереднього опрацювання даних забезпечить можливість використання показника задоволеності студентів та ранжування отриманих результатів для надання релевантних рекомендацій вступнику із використанням колаборативної фільтрації.

Джерелами збору даних є бази даних ІС ЗВО, що містять інформацію щодо семестрової успішності (БД ІС «Деканат»), наукової діяльності (БД ІС «Science LP») та архівних даних вступної кампанії (БД ІС «Абітурієнт»). Проміжні таблиці «student_temp_rating», «student_temp_master», «student_temp_science» та «admission_temp» на етапі збору даних містили початкову вибірку відповідно до кожного з набору даних. Процеси обробки даних вказують на відповідні обчислення показників, виконані під час розроблення відповідних компонент (описано в наступному розділі). Для зручного використання даних на подальших етапах, освітні дані було об'єднано в одну таблицю «students_rating_filtering» за допомогою запитів SQL. Архівні дані вступної кампанії містяться в таблиці «admission_filtering».

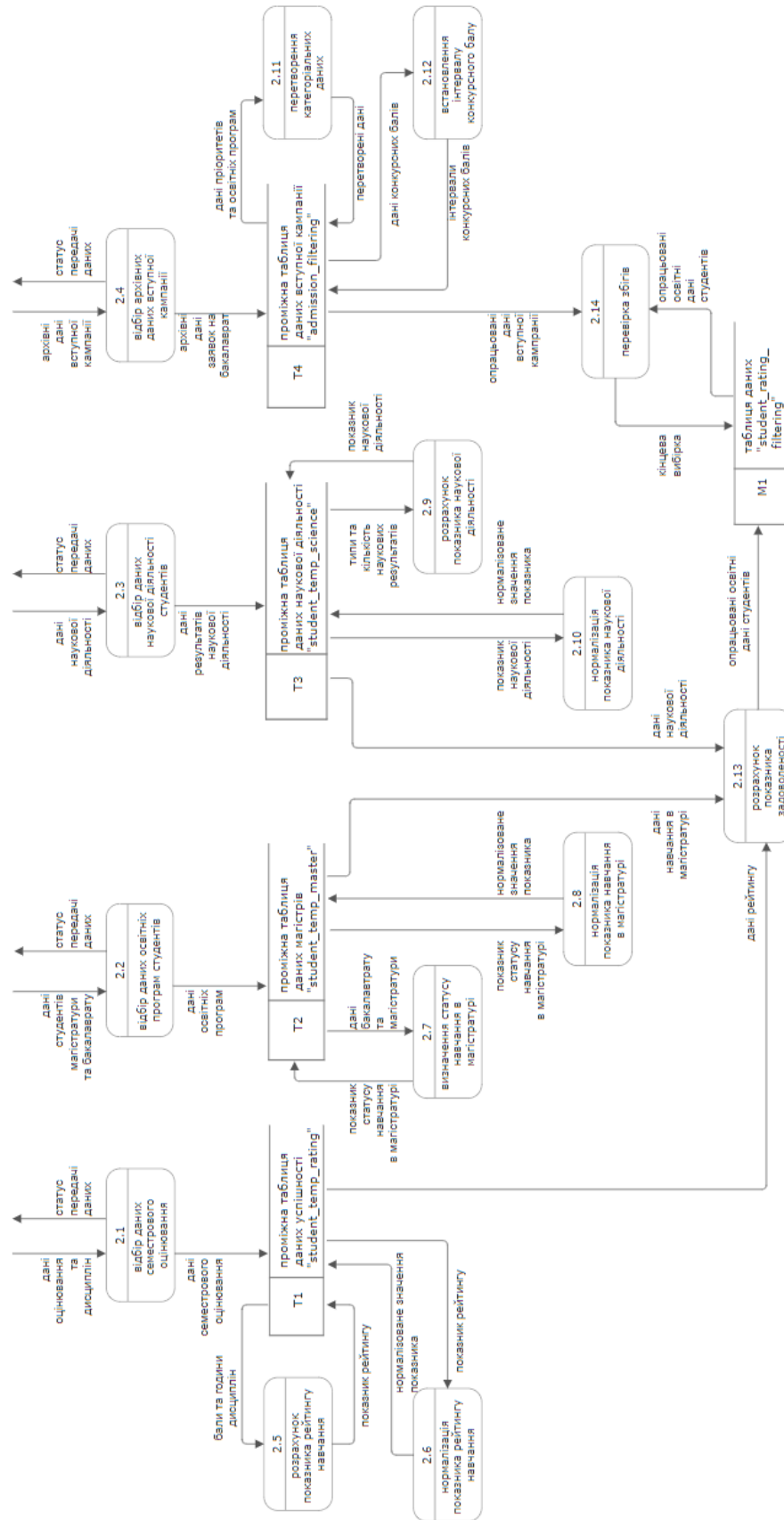


Рис.2.5 Діаграма потоків даних процесів збору та опрацювання даних для колаборативної фільтрації

2.2.3. Технологія збору та попереднього опрацювання даних для фільтрації на основі вмісту

Розроблена технологія збору та попереднього опрацювання даних для дозволяє на основі відбору описів професійних профілів випускників (з освітніх програм) сформувати набір ключових слів освітніх програм та їхніх кластерів (Рис. 2.6). Результати опрацювання цих даних покладено в основу надання рекомендацій альтернативних освітніх програм, які відповідають інтересам вступника із використанням фільтрації на основі вмісту .

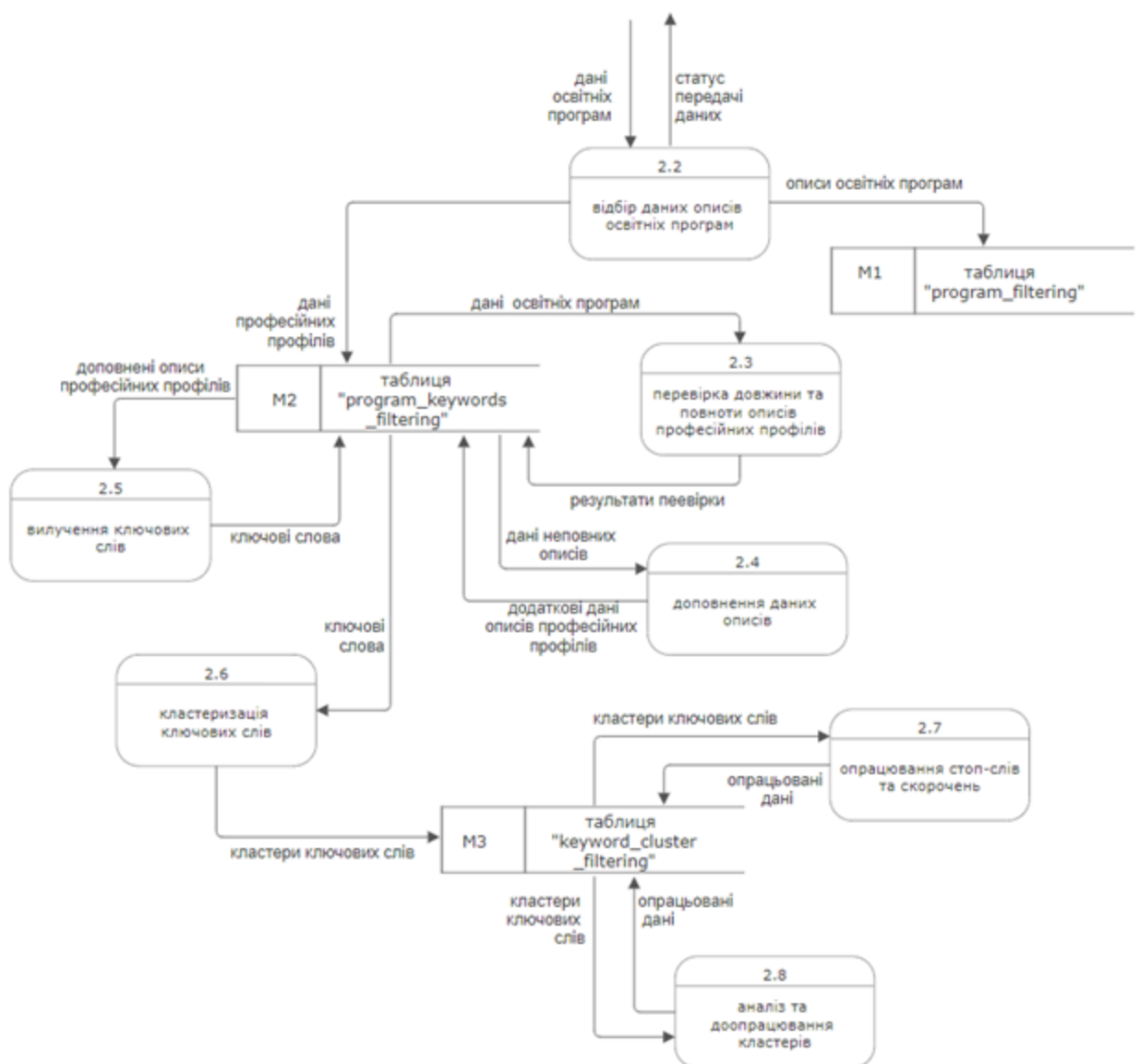


Рис.2.6 Діаграма потоків даних процесів збору та опрацювання даних фільтрації на основі вмісту

Джерелом для формування таких даних є БД ІС «Описи освітніх програм та дисциплін». На Рис. 2.4 відображено процес збору та попереднього опрацювання даних для фільтрації на основі вмісту. Проміжні таблиці «program_keywords_filtering» та «keywords_cluster_filtering» на етапі збору даних містили різні версії залежно від етапу опрацювання початкової вибірки. Процеси обробки даних вказують на відповідні процеси виділення ключових слів освітніх програм та їхніх кластерів, виконані під час розроблення відповідних компонент (описано в наступному розділі).

2.3. Методи машинного навчання для передбачення успішності вступу

2.3.1. Технології збору та опрацювання набору архівних даних вступної кампанії

Через повномасштабне російське вторгнення в Україну, яке відбулось 24-го лютого 2022 року, цього ж року МОН було змінено форму проведення вступного випробування на бакалаврат для здобуття вищої освіти в Україні – введено НМТ. НМТ є формою вступного випробування в році, яке передбачає оцінювання результатів навчання, здобутих на основі повної загальної середньої освіти. Тож, для можливості використання інструментарію машинного навчання на наступних вступних кампаніях, було опрацьовані архівні дані вступної кампанії НУЛП 2022-го року.

На Рис. 2.7 зображено розроблену технологію збору та опрацювання набору архівних даних вступної кампанії. Ці дані призначені для застосування в основі методів МН для передбачення успішності вступу.

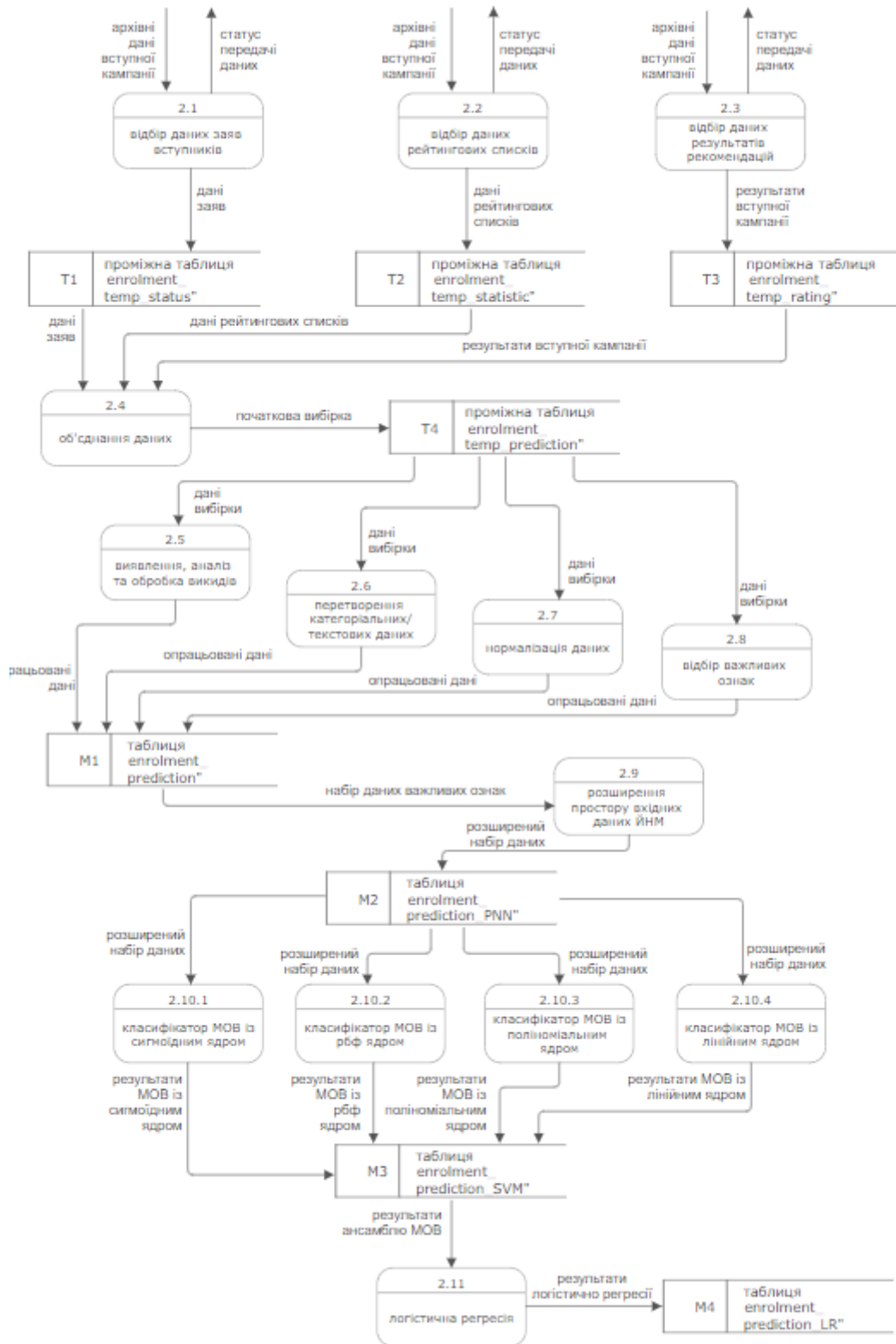


Рис.2.7 Діаграма потоків даних процесів збору та опрацювання даних для навчання методів машинного навчання

Джерелами збору даних є бази даних, що містять інформацію щодо архівних даних (БД ІС «Абітурієнт»). На першому етапі збору даних необхідно

відібрати інформацію про деталі поданих вступниками заяв, статистику вступної кампанії, характеристики освітніх пропозицій (кількість виділених місць), та інформацію за результатами вступної кампанії – статусами допущених до конкурсу заяв. Для цього, на етапі збору даних передбачено проміжні таблиці («enrolment_temp_rating», «enrolment_temp_statistic», «enrolment_temp_status», Рис. 2.7).

2.3.2. Обґрунтування критеріїв оцінювання ефективності роботи класифікаторів на основі машинного навчання

Правильний вибір інструменту машинного навчання для розв’язання задачі передбачення успішності вступу вступника ЗВО повинен базуватись на певних критеріях ефективності його роботи. У нашому випадку, серед таких критеріїв слід виділити точність роботи, яку можна обчислювати із використанням різних показників; генералізаційні властивості методу класифікації на основі машинного навчання; та власне тривалість процедури його навчання [77].

Враховуючи особливості зібраного набору даних, а саме його незбалансованість, ефективність засобів штучного інтелекту для розв’язання задачі класифікації доцільно обчислювати із використанням таких показників:

- влучність (precision);
- повнота (recall);
- F1-міра, як середнє гармонійне між влучністю та повнотою;
- коефіцієнт кореляції Метьюса;

Формули для їхнього обчислення подано нижче [76].

$$\text{Влучність} = \frac{TP}{(TP+FP)} \quad (2.1)$$

$$\text{Повнота} = \frac{TP}{(TP+FN)} \quad (2.2)$$

$$F1_{\text{міра}} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \quad (2.3)$$

$$\text{Коефіцієнт кореляції Метьюса} = \frac{(TP \cdot TN - FP \cdot FN)}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (2.4)$$

де: TP (True Positives) – це випадки, коли модель правильно передбачила екземпляри позитивного класу (істинно позитивні випадки); TN (True Negatives) – випадки, коли модель правильно передбачила екземпляри негативного класу (істинно негативні випадки); FP (False Positive) – випадки, коли модель класифікувала екземпляри позитивного класу, як негативні (хибнонегативні приклади); FN (False Negative) – випадки, коли модель класифікувала екземпляри негативного класу, як позитивні (хибнопозитивні випадки).

Слід зазначити, що такий показник як загальна точність роботи класифікатора (total accuracy), в цьому дослідженні не слід брати до уваги, оскільки вона не дає повної картини під час оцінювання результатів класифікації у випадку аналізу незбалансованого набору даних.

Сумісне використання таких показників як F1-міра та коефіцієнт кореляції Метьюса (2.3)-(2.4), забезпечать підвищення надійності під час оцінки роботи обраного класифікатора. Це пояснюється тим, що до сьогодні триває наукова дискусія який із показників краще обирати під час оцінювання результатів розв'язання задачі незбалансованої класифікації. Саме тому, щоб уникнути двозначності у інтерпретації отриманих результатів, у роботі використано обидва ці показники.

Оцінювання рівня генералізаційних властивостей обраного класифікатора можна виконувати шляхом порівняння різниці між точністю у режимі навчання методу (навчальні дані) та у режимі застосування методу (тестові дані). Мінімальне значення цієї різниці дозволить стверджувати про високі генералізаційні властивості методу, в той час як від'ємне значення цього показника демонструватиме низьку генералізацію, а велике додатне значення цієї різниці свідчатиме про ймовірність перенавчання обраної моделі.

Тривалість процедури навчання обраного класифікатора великою мірою залежить від обсягу та розмірності досліджуваного набору даних, характеристик ПК на якому виконується моделювання та різних особливостей, обраного методу машинного навчання (зокрема ітераційність алгоритму навчання). Оскільки, у

цій роботі використовується набір даних середнього обсягу, а обчислення для інтелектуалізованої компоненти інформаційної системи можуть відбуватися у хмарному середовищі, тривалість навчання обраного методу чи моделі повинна бути задовільною.

2.3.3. Вибір інструментарію машинного навчання для розв'язання задачі класифікації

Сучасний розвиток штучного інтелекту забезпечує потенційні можливості для адміністрації ЗВО та дослідників у цій області застосовувати численний арсенал алгоритмів, методів та моделей машинного навчання для розв'язання задач передбачення/класифікації різного рівня складності. У разі ефективності такого підходу, на базі них можна формувати інтелектуалізовані функціональні компоненти інформаційних систем різноманітного призначення (наприклад, для сприйняття, керування виконавчими механізмами, міркування, комунікації тощо). Однак вибір чи удосконалення існуючих або розроблення нових, ефективніших методів машинного навчання для розв'язання конкретної задачі на основі вищевказаних вхідних даних є важливим та складним етапом реалізації подібних компонентів.

Якщо говорити про задачу передбачення успішності вступу вступника ЗВО, то її можна розглядати як задачу класифікації. При такій постановці, є лише дві відповіді – вступить претендент чи ні. З точки зору машинного навчання – це задача бінарної класифікації. Базуючись на висновках, поданих у першому розділі роботи, для реалізації компоненти передбачення успішності вступу вступника ЗВО варто розглянути інструментарій машинного навчання, який використовується для розв'язання задачі бінарної класифікації.

У цій роботі проведено дослідження ефективності різних алгоритмів машинного навчання, штучних нейронних мереж, та ансамблів на їхній основі для розв'язання поставленої задачі. Дослідження проводилося в декілька етапів.

На першому етапі досліджувалася ефективність лінійних та нелінійних методів машинного навчання [79], [80]. Використання методів першого класу

пояснюється високою швидкістю їхньої роботи та простотою реалізації. Застосування методів другого класу характеризується підвищеною точністю роботи. Серед розглянутих методів слід виділити класифікатори на основі алгоритму стохастичного градієнтного спуску (СГС) та машини опорних векторів (МОВ) з різними ядрами, дерев рішень. Перший метод базується на використанні градієнта функції втрат для оновлення параметрів моделі з метою знаходження мінімуму цієї функції [81]. Незважаючи на те, що така процедура виконується ітераційно, цей метод дуже швидкий та забезпечує високу ефективність під час аналізу великих наборів даних. Класифікатор на основі МОВ з лінійним ядром передбачає побудову оптимальної гіперплощини яка повинна якомога точніше розділити класи задачі [82]. Він демонструє високу ефективність під час аналізу високорозмірних даних малого та середнього обсягу. Проте він має схильність до перенавчання та стає неефективним у випадку, коли дані не можна розділити гіперплощиною. Використання нелінійних ядер у МОВ забезпечують підвищення точності аналізу даних, проте правильний підбір ядра являється важливим етапом аналізу даних подібними методами [83]. Загалом, найбільшим недоліком усіх методів машинного навчання з класу лінійні є невисока точність розв'язання поставленої задачі у випадку аналізу складних даних, а нелінійних – правильний підбір параметрів роботи, зокрема ядра для методу МОВ.

Другий етап присвячено дослідженню ефективності штучних нейронних мереж, як універсальних апроксиматорів [84, 85, 86]. Тут розглядалася ймовірнісна нейронна мережа та багатошаровий перцептрон (БШП), як найбільш використовувані методи у літературі для розв'язання поставленої у роботі задачі [87]. Основною перевагою першого методу є високі узагальнюючі властивості за відсутності алгоритму навчання методу, як такого. Окрім цього, ЙМН дає можливість синтезу результатів процедури класифікації як у формі міток класів визначених задачею так і у формі ймовірностей приналежності до кожного із класів [88]. Останнє є великою перевагою під час розв'язання великого кола

прикладних задач. БШП у свою чергу забезпечує високу точність під час аналізу нелінійних даних, проте він є доволі часозатратний з огляду на ітераційний характер процедури його навчання [89], [90]. Це накладає ряд обмежень на можливість його практичного використання, у випадку високої чутливості інформаційної системи до часу навчання моделі [91].

На третьому етапі аналізувалися ансамблеві методи, які підвищують точність класифікації в порівнянні із кожним окремим методом які в сукупності формують ансамбль [91], [92]. Серед розглянутих методів слід виділити AdaBoost, як представник класу бустингових ансамблів [93], алгоритм Випадкового Лісу, як представник класу бегінгових ансамблів [94], об'єднання МОВ з чотирьома різними ядрами (рбф, сигмоїдне, поліноміальне та лінійне), як представник класу стекінгових ансамблів [95], та двоетапний класифікатор на основі сумісного використання ЙМН та МОВ як представник класу каскадних ансамблів [96].

Алгоритм машинного навчання AdaBoost (Adaptive Boosting) виконує ітеративний пошук і комбінування слабких класифікаторів (зазвичай дерев рішень), щоб створити сильний класифікатор. На кожній наступній ітерації він намагається врахувати помилки із попередньої для підвищення точності класифікації. Фінальний результат роботи методу це комбінування роботи усіх слабких моделей із врахуванням визначених методом вагових коефіцієнтів для кожного методу [93]. Серед основних недоліків методу в контексті поставленої у роботі задачі слід відзначити необхідність значних обчислювальних ресурсів для побудови багатьох слабких класифікаторів та їхніх комбінацій. Це справедливо і для інших ансамблевих методів з класу бустингу [97].

Алгоритм Випадкового лісу (Random Forest) як один із представників класу бегінг передбачає побудову кількох дерев рішень під час навчання та комбінуванні їхніх результатів для отримання більш точного прогнозу. Кожне дерево в цьому випадку будується на своїй підвибірці даних із випадковим вибором ознак для аналізу. Фінальний результат методу формується шляхом

голосування більшістю побудованих дерев [94]. Як і у попередньому випадку, методи цього класу вимагають чимало ресурсів для побудови великої кількості дерев. Окрім цього, у випадках, коли вибірка даних містить рідкісні екземпляри, точність цього методу може падати [95]. У нашому випадку це критично з огляду на можливість наявності у вибірці невеликої кількості даних про якусь із менш популярних освітніх програм.

Методи із класу стекінгових ансамблів передбачають запуск декількох різних класифікаторів на основі машинного навчання (у випадку гетерогенного стекінгу) та використання результатів їхньої роботи як нової вибірки для навчання метакласифікатора [96].

Серед переваг роботи методів цього класу слід відзначити підвищення точності роботи стекінгового об'єднання класифікаторів на основі машинного навчання у порівнянні із кожним класифікатором який включено до складу стекінгу [95]. Це за умови використання різноманітних класифікаторів які не корелюють між собою [98]. Серед недоліків методів цього класу, як для усіх, поданих вище – таких підхід вимагає чимало обчислювальних та часових ресурсів для реалізації моделі. Проте, враховуючи те, що базові класифікатори згідно стекінгового підходу можна запустити паралельно [99], цей тип ансамблю нівелюватиме останній недолік щодо часу навчання моделі. Це є важливою перевагою в контексті цієї роботи.

Останній клас ансамблевих методів – каскадні класифікатори передбачають побудову послідовних слабких класифікаторів, де кожен наступний враховує результати роботи попереднього чим покращує точність розв'язання поставленої задачі [100]. Серед інших їхніх переваг слід відзначити високу точність та швидкодію, а також зменшення впливу дисбалансу класів, що є важливим для поставленої у роботі задачі [101]. Серед недоліків, слід відзначити те, що кожен класифікатор у каскаді вимагає додаткових обчислювальних ресурсів та часу на навчання. Проте, у випадку, коли рівнів каскаду лише два, як у [96], цим недоліком можна нівелювати.

Висновки до розділу

У другому розділі подано модель інформаційної системи підтримки прийняття рішень вступників ЗВО та обґрунтовано її розроблення за рахунок сумісного використання колаборативної фільтрації та фільтрації на основі вмісту для розширення переліку альтернативних освітніх програм, що відповідають потребам вступників ЗВО у поєднанні з передбаченням успішності вступу на основі методів машинного навчання. Серед вагомих отриманих результатів слід виділити наступні:

- розроблено модель інформаційної системи підтримки прийняття рішень вступників ЗВО, для підвищення рівня обізнаності вступників ЗВО України щодо вибору освітніх програм відповідно до їхніх інтересів;
- обґрунтовано вибір методів надання рекомендації освітніх програм вступникам ЗВО для реалізації компоненти інформаційної системи підтримки прийняття рішень;
- розроблено метод сумісного застосування колаборативної фільтрації, із використанням архівних даних вступної кампанії і освітніх даних, та фільтрації на основі вмісту, із використанням кластерів ключових слів освітніх програм для розширення спектру альтернативних освітніх програм, які відповідають інтересам вступників;
- розроблено технології збору та опрацювання даних для можливості проведення подальшого аналізу на їхній основі для надання рекомендацій освітніх програм та передбачення успішності вступу;
- визначено та обґрунтовано основні метрики для оцінювання результатів роботи класифікаторів на основі машинного навчання у випадку аналізу незбалансованого набору даних, що дало змогу проводити коректний та надійний аналіз та інтерпретацію отриманих результатів;
- визначено та обґрунтовано вибір методів машинного навчання для розв'язання задачі бінарної класифікації з метою отримання найвищої точності результатів передбачення успішності вступу до ЗВО.

Розділ 3. Розроблення компонент інформаційної системи підтримки прийняття рішень вступників ЗВО

У третьому розділі розроблено процедуру розрахунку показника задоволеності студентів на основі їхніх освітніх даних та даних вступної кампанії для надання рекомендацій вступникам на основі колаборативної фільтрації. Для реалізації фільтрації на основі вмісту здійснено формування набору ключових слів описів освітніх програм та виконано формування кластерів цих слів для рекомендацій альтернативних освітніх програм, які відповідають інтересам вступника.

Здійснено формування множини атрибутів набору даних, які впливають на успішність вступу до ЗВО. Досліджено чотири методи визначення важливих ознак зібраного набору даних для проведення їхнього подальшого інтелектуального аналізу. Досліджено існуючий інструментарій машинного навчання з різних класів для розв'язання задачі класифікації, та обґрунтовано вибір найкращого методу. Удосконалено двоетапний метод класифікації даних на основі сумісного використання ймовірнісної нейронної мережі та стекінгового об'єднання машини опорних векторів з чотирьома різними ядрами. Подано структурно-функціональну схему удосконаленого методу, деталізовано алгоритми його навчання і застосування.

3.1. Колаборативна фільтрація з використанням архівних даних вступу та освітніх даних

3.1.1. Обробка архівних даних заявок вступників для визначення подібності інтересів користувачів

Збір та обробку даних для реалізації цього методу було здійснено на основі даних архівних даних вступної кампанії до НУЛП 2019-го року. Джерелом даних для реалізації цього методу є ІС «Абітурієнт». Вибірка містила ідентифікатори освітніх програм, студентів (як вступників 2019-го року) та визначені ними

пріоритети. Оскільки основним призначенням цих даних є порівняння вподобань пріоритетів, то опрацювання даних передбачало вибірку необхідних атрибутів та створення структури таблиці БД. Для отримання інтервалів класів КБ їх було згруповано за кількісною ознакою – інтервалом розміром 5 балів. Межі КБ від 130 до 200 (однаково для 2019 та 2023 року).

Таким чином кожному запису було призначено клас за такими ознаками: перший пріоритет, інтервал конкурсного балу та освітня програма.

3.1.2. Процедура розрахунку показника задоволеності на основі сформованого набору освітніх даних

Очищення даних передбачало видалення результатів заліку з фізичного виховання на першому курсі навчання. В наступних роках навчання було враховано фізичне виховання, оскільки це свідомий вибір студент з циклу професійної та практичної підготовки та враховується при рейтингуванні студентів.

Оскільки дані було відібрано з індивідуальних навчальних планів, котрі заповнюють працівники через відповідну ІС університету, могли бути пропуски. Тож за параметром кількості студентів в групі було перевірено повноту складу студентів а також, вибірково предметів по яких було виставлено середню оцінку по відношенні до кількості предметів в навчальному плані студента. Також для перевірки були пораховано кількість оцінок для кожного зі студентів так, щоб виявити чи немає значень менше восьми (необхідними є всі вісім семестрів за чотири роки навчання). Пропуски даних було доповнено.

Вибірка включала дані студента (ідентифікатор, без персональних даних), навчально-наукового інституту, семестру та року навчання, ОП студента, освітніх компонент (далі предмет) з їхніми балами за підсумковий семестровий контроль.

У цьому дослідженні успішність студента розраховано з врахуванням обсягу навчального навантаження як кількість академічних годин та результату підсумкового семестрового контролю як балів у національній шкалі оцінювання.

Обчислення рейтингу виконується шляхом пошуку середнього арифметичного значення успішності студента по кожній дисципліні (беручи до уваги кількість годин відповідної дисципліни) протягом усіх семестрів навчання:

$$\text{Рейтинг} = \frac{\sum_{j=1}^M \left(\frac{\sum_{i=1}^N \text{Оцінка}_i * \text{Години}_i}{\text{К-сть годин за семестр}} \right)}{M} \quad (3.1)$$

де: j - поточний семестр, $j=1, M$, i – поточна дисципліна у семестрі, $i=1, N$.

Іншим показником також було враховано чи залишився студент за тим же напрямом або ж змінив його, або ж не продовжив навчання в ЗВО відповідно до переліку освітніх ступенів та освітніх програм, за якими оголошується прийом на навчання у 2023 визначені правилами прийому до НУЛП [102].

Тож дані було доповнено новою властивістю у відповідності до прийняття рішення студента щодо навчання в магістратурі. Так, кожен з елементів було віднесено до одного з класів:

- відповідність галузі, освітньої програми та спеціалізації,
- *або* відповідність галузі, освітньої програми, АЛЕ зміна спеціалізації у зв'язку з їхньою відсутністю у пропозиціях програм магістратурі – навчання за спеціалізацією;
- *або* відповідність галузі, освітньої програми, АЛЕ зміна спеціалізації у зв'язку з їх відсутністю у пропозиціях програм магістратурі у тому ж інституті – значення «навчання в магістратурі ЛП». Якщо напрям той самий, а освітня програма відрізняється одним терміном, що вказує на деталізацію, то не вважаємо це зміною напрямку, а вибором однієї з альтернатив;
- відповідність галузі, освітньої програми але зміна спеціалізації за наявності такої ж програми і магістрату – «навчання в магістратурі ЛП за спеціальністю»;
- відповідність галузі але зміна освітньої програми та спеціалізації за наявності такої ж програми і магістрату – «навчання в магістратурі ЛП за галуззю»;

- зміна галузі та освітньої програми, а відповідно й спеціалізації – «зміна галузі ЛП»;
- якщо студент не продовжив навчання в магістратурі вказано статус «інше».

Причинами не продовження навчання в магістратурі можуть бути не лише прийняття рішення змінити спеціалізацію. Існує широкий спектр можливих чинників – зміна країни чи регіону проживання, вартість навчання, суспільно-економічні чинники. Окрім цього доступ до даних є обмежений, та не зберігається в межах того чи іншого ЗВО. Тому, можна вважати ці дані такими, які не доречно враховувати. Кожному з класів було призначено також числовий відповідник значенням від «1» до «5», де 1 вказує на те, що студент не продовжив навчання.

Як джерело даних також було використано БД ІС «Science LP» – інформаційної системи університету, що зберігає дані працівників та студентів, зокрема про результати їхньої наукової діяльності.

Використовуючи перелік студентів сформованого раніше набору даних було отримано вибірку за результатами наукової діяльності студентів. Отримані дані містять записи за наступними сімома типами наукової діяльності:

- участь у виконанні науково-дослідницької роботи (НДР) на кафедрі;
- публікація у наукових виданнях України;
- опубліковані тези доповідей наукової конференції;
- наявність нагород (обласні премії, стипендії Верховної Ради України, стипендії Президента України);
- участь у конкурсі студентських наукових робіт;
- участь у виставці студентських робіт;
- міжнародні стипендії.

Для кожного студента було підраховано кількість записів за типами наукової діяльності та присвоєно відповідні ідентифікатори: НДР – «1», публікації – «2», тези – «3», нагороди – «4», конкурс – «5», виставка – «6»,

стипендія. Для ефективнішого використання цих даних в подальшому, їхні значення було модифіковано. Кожному з показників було призначено коефіцієнт від 0 до 1. Сума добутків кількості кожного з показників є кінцевим значенням.

Коефіцієнти наступні: НДР – 1, (найвище оскільки НДР передбачає сукупність робіт, спрямованих на отримання нових знань та їхнє практичне застосування); публікації – 0,9; тези – 0,8 (проведено наукові дослідження); нагороди – 0,7; стипендія – 0,7; конкурс – 0,6; виставка – 0,6.

Відповідно до розподілу отриманих даних, за показниками результатів наукової діяльності кожному елементу було призначено один з класів: «низький» – значення від 0 до 2, «нижче середнього» – від 2 до 4, «середній» – від 4 до 6, «вище середнього» – від 6 до 8, та «високий» – більше 8.

Числові значення для кожного з показників було нормалізовано із використанням MinMax аби забезпечити однаковий масштаб та важливість ознак. Формула нормалізації подано нижче.

$$x_{norm} = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (3.2)$$

Значенням отриманих показників було присвоєно коефіцієнти та просумовано для отримання єдиного значення. Зокрема для показника успішності в навчанні було введено коефіцієнт 0.5, для показника продовження навчання в магістратурі – 0.3, для показника наукової діяльності – 0.2.

3.2. Фільтрація на основі вмісту з використанням ключових слів та кластерів описів освітніх програм

3.2.1. Виділення ключових слів описів освітніх програм

Іншою властивістю, необхідною для подальшого опрацювання даних є ключові терміни, що найбільшою мірою можуть передати суть окремих елементів набору даних – освітніх програм.

Як згадувалось раніше, джерелом відбору ключових слів є БД ІС «Опис освітніх програм та дисциплін». Дану ІС по кожній зі освітніх програм наповнює інформацією відповідний гарант – працівник ЗВО, який має кваліфікацію

відповідно до освітньої програми, науковий ступінь та/або вчене звання за відповідною або спорідненою до освітньої програми [103]. Гарант повинен відповідати за розробку освітньої програми та за забезпечення її якості і реалізовувати концепцію студоцентричного навчання. Так, згідно з положенням про гарантів освітніх програм НУЛП до їхніх функцій, зокрема, входять визначення змісту ОП, координування роботи з розробки ОП, забезпечення відповідності змісту ОП стандартам вищої освіти – професійним стандартам та вимогам національної рамки кваліфікацій, координування діяльності щодо актуалізації змісту ОП та її компонентів з урахуванням інтересів та пропозицій стейкхолдерів, на основі вивчення тенденцій розвитку освітніх програм та аналогічних вітчизняних та іноземних ОП, ринків праці, секторальних, галузевих та регіональних контекстів. Враховуючи специфіку роботи та обов'язків гаранта ОП, сформований нам та внесений до відповідної ІС опис освітньої програми можна вважати таким, який не потребує попередньої обробки чи аналізу - ключові поняття та найбільш релевантні ключові слова та фрази вже присутні в текстових даних [103].

Серед існуючих даних, що описують освітню програму, для відбору ключових слів/фраз обрано поле «Професійні профілі випускників» (серед інших – «Характеристика освітньої програми», «Набуті компетентності», «Програмні результати навчання»). Як і при створенні опису освітньої програми згідно методичних вказівок вміст цього поля описує придатність випускників до працевлаштування, коротко описує види економічної діяльності, професійні назви робіт (відповідно до Класифікатора професій ДК 003:2010) які може виконувати випускник відповідного рівня вищої освіти, первинні посади, які може займати, а також можливість та/або вимога професійної Сертифікації.

Для відображення основного змісту текстового елемента існує два основних способи – ручний аналіз, який полягає в обробці даних вручну і в тому, щоб почати реєструвати інформацію, яка здається важливою, та другий – автоматизований аналіз тексту – це інший підхід, який використовує

комп'ютерні інструменти для виконання аналізу тексту. Останній метод є більш доречним через великі обсяги даних, оскільки вручну обробити значний обсяг тексту буде неможливо.

Виділення ключових фраз і ключових речень є двома важливими підзадачами в задачі генерації та аналізу тексту [104]. В даний час вилучення ключових фраз широко використовується в багатьох галузях, таких як обробка природної мови, інформаційний пошук, аналіз тональності тексту, індексування документів чи класифікація документів. Метою виділення таких властивостей є визначення найбільш релевантних і важливих слів або комбінацій слів, які виражають усі ключові аспекти його змісту. Ключові фрази становлять стислий концептуальний виклад документа, який є корисним у системах управління цифровою інформацією для семантичного індексування, фасетного пошуку, кластеризації та класифікації документів [105].

Сучасні дослідження некерованих методів вилучення ключових фраз поділяються на дві групи – лінгвістичні та статистичні, до яких вже належать: методи, засновані на статистиці, засновані на графах, засновані на темах та засновані на мовних моделях.

Такий розподіл пояснений різницею загальноживаних показників, функцій, які впливають на вилучення ключових фраз, і основних методів неконтрольованого вилучення ключових фраз. Враховуючи характеристики даних та цілі аналізу було розглянуто саме статистичні методи. Мається на увазі те, що кінцевою метою використання ключових слів (як і решти попередньо описаних властивостей) є визначення нетопових подібностей між освітніми програмами. Тут, варто зазначити, що ключові фрази та ключові слова є спорідненими поняттями в контексті аналізу тексту, але вони мають різні значення.

Ключові слова – це окремі слова, які мають значення та мають значення в документі чи наборі даних та зазвичай використовуються в контексті пошукових систем, де користувачі вводять конкретні слова, щоб знайти відповідну

інформацію. Ключові фрази – це групи слів, які разом передають певне значення або представляють важливу концепцію в документі чи наборі даних та ефективні для одержання детальнішої та контекстно-насиченої інформації порівняно з окремими ключовими словами. Тобто, хоча ключові слова – це окремі слова, важливі в певному контексті, ключові фрази включають групи слів, які разом передають більш вичерпне значення.

Тож, з огляду на ціль аналізу та рівень деталізації вмісту текстових елементів у цьому дослідженні зосереджено увагу на вилученні саме ключових фраз, що в подальшому забезпечить користувачу більш детальне розуміння змісту.

Належним чином обраний та реалізований конкретний метод може підвищити продуктивність, точність і процесу роботи з текстовими даними, а також точність наступних процесів побудови модулів ІС, які залежать від вилучення ключових слів як етапу попередньої обробки. Тож, для порівняння ефективності результатів було проведено тестування існуючих онлайн застосунків вилучення ключових слів різними методами для декількох освітніх програм. При вилученні ключової фраз чи термінів характеристики наборів даних безпосередньо впливають на продуктивність методів, тому під час аналізу було використано текстові елементи різних освітніх програм.

Незважаючи на значну кількість існуючих методів, лише деякі з них були розроблені у вигляді програмних додатків з відкритим вихідним кодом або інструментів, які можна легко інтегрувати в інші програми для автоматизації процесу вилучення важливих ключових слів.

Так, вилучення ключових слів на основі частоти (Frequency-based methods) – є одним з найбільш поширених методів, що визначає ключові фрази на основі частоти слів у документі. Терміни, які зустрічаються частіше, вважаються більш важливими.

Цей метод було досліджено за допомогою наступних веб-застосунків: WordCounter (<https://wordcounter.net/>), MonkeyLearn's Online Text Analysis Tool

(<https://monkeylearn.com/keyword-extractor-online>). Освітні програми, які тестувались: Науки про Землю, Журналістика, Харчові технології.

Основним недоліком отриманих результаті є те, що ключові слова, які найбільшою мірою описують текст отримали найменшу міру важливості.

Іншим поширеним методом є статистичні вимірювання, зокрема, TF-IDF (Частота термінів – зворотна частота документа), що оцінює важливість термінів у документі відносно їхньої появи в усьому наборі даних.

Освітні програми, які тестувались: перша група: Науки про землю, Геодезія та землеустрій, Екологія, Технології захисту навколишнього середовища, друга група: Музейна, пам'яткоохоронна діяльність і культурний туризм, Право, Інженерія програмного забезпечення, Маркетинг. Цей метод було досліджено за допомогою веб-застосунку: TF-IDF Demo (<https://remykarem.github.io/tfidf-demo>).

Переглянувши результатами аналізу можна виділити декілька недоліків такого методу при використанні описів освітніх програм. Зокрема, такий метод не забезпечить виділення значення фраз або багатослівних виразів, чутливість до довжини документа (вміст використовуваних полів має різну довжину і привести до чітко однакових не можливо), рідкісним термінам (термінам, які зустрічаються лише в кількох документах) у TF-IDF може надаватися надто велика вага, що призводить до надмірного акцентування. Це може бути проблемою, якщо рідкісний термін присутній у кількох документах із високою частотою, правильна попередня обробка, включаючи токенізацію та стемінг, є важливою для ефективної роботи TF-IDF.

Важливість використання як статистичних даних, так і інформації про контекст підтверджується такими новітніми методами, як YAKE [106]. YAKE, окрім позиції/частоти терміну, також використовує нові статистичні метрики, які фіксують інформацію про контекст і поширення термінів у документі. По-перше, YAKE попередньо обробляє текст, розбиваючи його на окремі терміни. По-друге, для кожного окремого терміну обчислюється набір з 5 характеристик:

Оболонка (W_{case} , яка відображає аспект оболонки слова), Позиція слова ($W_{Position}$, яка має більше значення для тих слів, які більше значення мають слова, що зустрічаються на початку документа), частота вживання слова (W_{Freq}), зв'язок слова з контекстом (W_{Rel} , що обчислює кількість різних термінів, які зустрічаються зліва/справа від слова-кандидата) і різниця вживання слова ($W_{DifSentence}$, що визначає, як часто слово-кандидат з'являється в різних реченнях). Потім усі ці характеристики використовуються для обчислення оцінки $S(w)$ для кожного терміну (чим менше значення, тим важливішим буде слово w). Нарешті, генерується неперервна послідовність ключових слів-кандидатів з 1, 2 і 3 грамів. Кожному ключовому слову-кандидату kw присвоюється наступна оцінка:

$$S(kw) = \frac{\prod_{w \in kw} S(w)}{Tf(kw) * (1 + \sum_{w \in kw} S(w))} \quad (3.3)$$

Щоб уникнути труднощів у визначенні стислого списку описових термінів виключно на основі самого документа у цьому методі використано подвійний підхід: по-перше використано локальні статистичні характеристики, які виділяють інформативний зміст у тексті, щоб обчислити важливість окремих термінів; (2) по-друге, застосовано спеціальну модель побудови n грам для формування багатослівних термінів і використано евристичну міру для визначення їхньої релевантності.

Основними особливостями цього методу, що пояснюють доречність його використання є:

- підхід без вчителя алгоритму автоматичного вилучення ключових слів, який ґрунтується на локальних статистичних характеристиках тексту, отриманих з окремих документів, тобто не потребує жодного навчального корпусу;
- незалежність від корпусу, що дає змогу отримувати ключові слова лише з одного документа, без необхідності покладатися на зовнішню статистику збору документів, як це робить IDF; тобто його можна застосувати до будь-якого тексту;

- незалежність від домену та мови: YAKE! працює з доменами та мовами, для яких немає готових систем вилучення ключових слів, оскільки він не вимагає ані навчального корпусу, ані залежить від складних зовнішніх джерел або лінгвістичних інструментів, крім статичних список стоп-слів. Хоча мовою за замовчуванням у YAKE є англійська, ключові фрази можна видобувати й іншими мовами, встановивши параметр мови на потрібну мову;
- висока точність роботи порівняно з іншими сучасними методами. Експериментальні результати сучасних досліджень, виконані на основі понад двадцяти наборів даних, показують, що YAKE! значно перевершує інші неконтрольовані методи на текстах різних розмірів, мов і тематик [107];
- відкритий вихідний код: доступна демо версія (yake.inesctec.pt), програма в Google Play, а також API і пакет python (github.com/LIAAD/yake).

Дослідження роботи методу за допомогою онлайн-демо-версії показує ефективній відбір саме ключових фраз. Також перевагою є те, що результати застосування даного методу дають нормалізоване значення ваги ключової фрази.

Для використання методу було здійснено автономну інсталяцію на вільновикористовуваній операційній системі RockyLinux сімейства RedHat. Такий спосіб інсталяції рекомендується використовувати для розробки або інтеграції. при цьому необхідними є Python 3 та pip — система керування пакунками, яка використовується для встановлення та управління програмними пакетами, які написані на Python.

YAKE було встановлено через pip з повноваженнями звичайного користувача (не привеліюваного). Наступним кроком є перевірка довідки YAKE, яка виводить весь перелік всіх параметрів. Для вибору мови тексту, що буде аналізуватися використано параметр `-l`, а для визначення української — `-ua`. Наступний параметр - це файл, у якому знаходиться текст. Тестування роботи методу показано на Рис. 3.1

```

[redacted]@vmpPortalSites ~]$ yake -l ua -i ./text.txt -n 10 -v -df leve
keyword                                     score
-----
космічний моніторинг Землі                0.00672789
космічний моніторинг                      0.0334385
моніторинг Землі                          0.0382746
Професійний профіль                       0.0439051
методи космічного моніторингу             0.0573594
космічного моніторингу                   0.0632505
Проводить                                 0.0898839
Землі                                     0.0947657
Застосовує методи космічного моніторингу 0.0964751
методи космічного                         0.100471
[redacted]@vmpPortalSites ~]$ █

```

Рис.3.1 Результати вилучення ключових слів за допомогою ПЗ YAKE

Для роботи алгоритму було підготовано текст і такі параметри кількість n-грам, мову тексту (параметр мови тексту потрібен лише для ідентифікації конкретного списку стоп-слів), кількість ключових слів. Параметри було встановлено експериментально за допомогою онлайн демо-версії, зокрема: кількість n-грам - 3, к-ть ключових слів - 20.

Для визначення оптимальної кількості ключових слів, було досліджено текстові дані кількох освітніх програм, які містять середню кількість символів.

Перелік пропонованих для здобуття ступеня бакалавра освітніх програм у НУЛП налічує 80 позицій відповідно до правил прийому у 2023-му році у 15-тьох різних навчально-наукових інститутах.

Варто зазначити, що дані більшості освітніх програм знаходяться в межах середнього розміру даних. Проте, у тих випадках, де дані були відсутні, або значно коротші, доповнено даними відповідного поля з інформації на офіційному сайті університету, розділ «Про освітні програми» (<https://lpnu.ua/osvita/pro-osvitni-programy>). У цьому розділі містяться затвержені освітні програми у файлах .pdf. Це дозволило зменшити різницю довжини текстових даних. Зміни відносно середньої кількості символів до та після внесення змін зображено на рисунках нижче (Рис. 3.2, Рис. 3.3).

Аналіз та модифікацію даних було здійснено запитамі SQL – підраховано значення кількості символів для кожної окремо освітньої програми, визначено середнє значення (функції LENGTH та AVG).

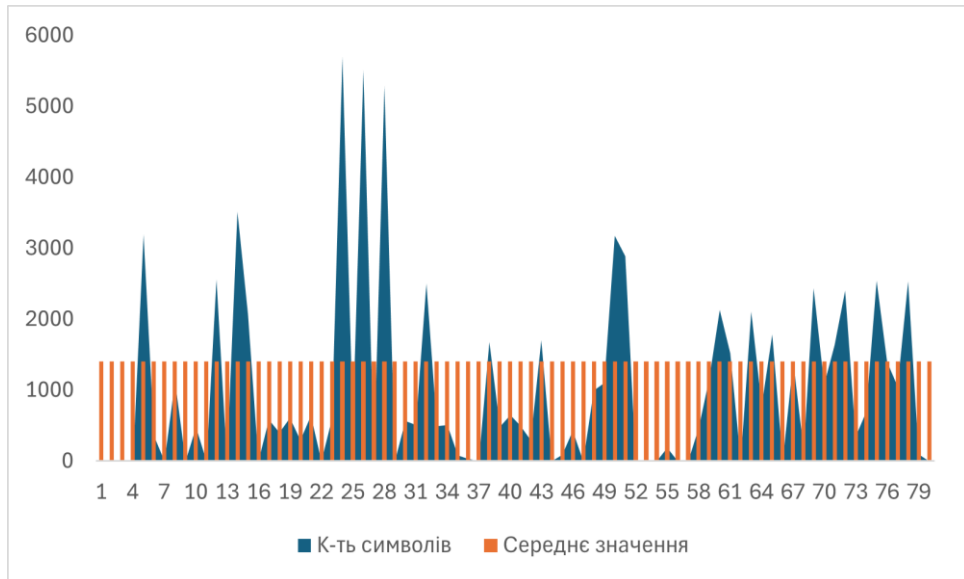


Рис.3.2 Довжина текстового поля професійних профілів випускників в описах освітніх програм

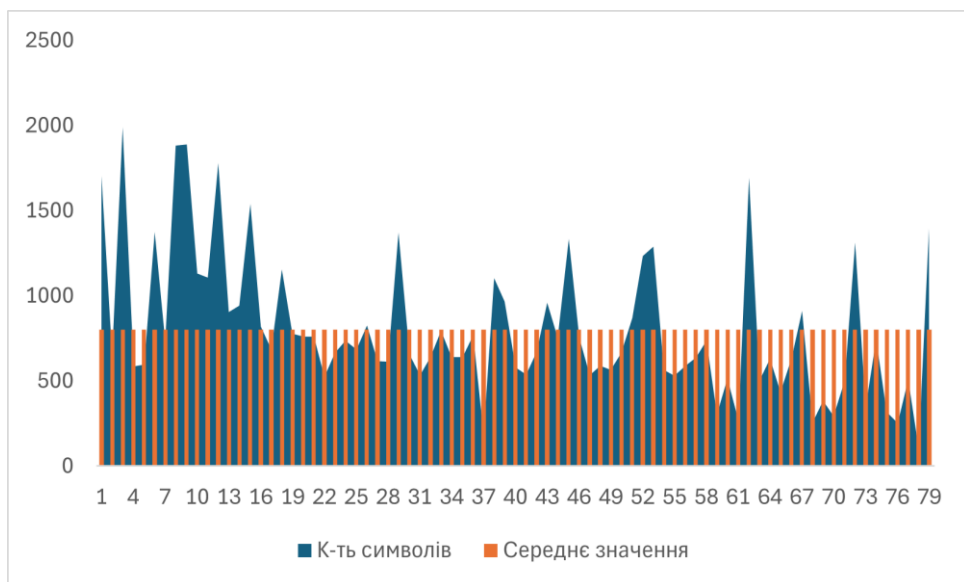


Рис.3.3 Довжина текстового поля професійних профілів випускників в описах освітніх програм після доповнення інформації

Після цього було здійснено додаткову обробку – серед отриманих ключових фраз було видалено такі як «національний класифікатор», та приведено до однакового вигляду скорочення та деякі терміни (наприклад «веб» та «web», «дані» та «data»).

3.2.2. Формування кластерів ключових слів освітніх програм

Оскільки організація і категоризація ключових слів має вирішальне значення для ефективного їхнього використання, отримані результати необхідно структурувати та зберегти таким чином, щоб вони мали сенс і забезпечували легкий доступ. Тож, задля належної організації ключових слів, а також пришвидшення роботи ІС отримані результати було додатково опрацьовано.

З метою здійснення аналізу та пошуку ознак, за якими можна об'єднати освітні програми, стиснення даних щодо ключових слів та пошуку нових спільних ознак, що не входять у жодну із наявних було застосовано також кластеризацію отриманого переліку ключових слів.

Мета полягає в тому, щоб згрупувати немарковані дані так, аби об'єкти даних, характеристики та атрибути яких схожі, були об'єднані в кластери таким чином, щоб схожість об'єктів даних в межах одного кластера була вищою порівняно з об'єктами даних в інших кластерах. Іншими словами, аналіз кластеризації даних класифікує немарковані дані, щоб забезпечити вищу внутрішньокластерну схожість і нижчу міжкластерну схожість [108].

Серед основних недоліків такого методу виділяють обчислювальну складність при великому наборі даних та обмежене розуміння контекстної інформації чи семантики слів. Проте, вибірка ключових слів цього дослідження не налічує лише 1256 записів, а коли акцент робиться на виявленні орфографічної та структурної схожості, а не на семантичних зв'язках.

З огляду на особливості даних та мету їхньої кластеризації було розглянуто методи, які не потребують попередньо визначеної кількості кластерів та обрано алгоритм Левенштейна (Levenshtein distance), який має такі переваги:

- принцип вимірювання схожості рядків враховує як семантичну, так і синтаксичну схожість;
- нечутливість до друкарських помилок, оскільки кількісно оцінює схожість між словами на основі відстані між ними;

- обробляє варіації форм слів (наприклад, однина/множина, часи дієслів тощо), фіксує схожість між словами незалежно від їхніх граматичних варіацій;
- не вимагає попередніх знань про словниковий запас або кількість кластерів;
- придатність для широкого спектру застосувань і наборів даних не залежно від мов і доменів;
- інтерпретованість: Кластери, створені за допомогою кластеризації відстаней Левенштейна, часто можна інтерпретувати, оскільки слова в межах одного кластера схожі за написанням і структурою. Це може полегшити подальший аналіз і розуміння основних закономірностей у даних [109].

Для застосування такого методу було використано онлайн інструмент Umbrellum Keyword Clustering є частиною набору інструментів для SEO, розробленого Воутером ван дер Меєм, SEO-консультантом (<https://umbrellum.com>). На основі отриманих результатів було розширено набір даних додатковими властивостями (проміжна таблиця «prof_profilesV03» на Рис.2.6)

Основні кроки алгоритму виділення кластерів ключових слів із використанням відстані Левенштейна наведено нижче:

1. Створення двовимірного масиву з усіх фраз, наявних у наборі даних по рядках ($phrase_i$, $i=1,N$) та всіх окремих слів з цих фраз по стовпцям ($word_j$, $j=1,M$). Перетину кожного рядка-стовпця присвоїти 1 (одиницю) у випадку, коли слово належить фразі, чи 0 (нуль) у інших випадках. Результатом виконання такої операції є вектор булевих значень який можна використати як набір ознак для кожного окремого слова із заданого набору даних (кожен окремий стовпець). На основі такого набору ознак можна формувати міри, які характеризуватимуть текст, на основі виділених ключових слів.

2. Пошук однакових слів у заданому наборі даних по отриманих векторах ознак для кожного окремого слова. Для цього необхідно виконати

пошук відповідних еквівалентних векторів в утвореному масиві даних. У випадку їхньої наявності, утворюємо групу (кластер) з назвою шуканого слова. До такого кластера входять усі фрази із набору даних, які містять виділені вище слова (які на перетині містять цифру 1). Таку процедуру необхідно виконувати циклічно, перебираючи усі наявні слова у масиві даних.

Виконання цього кроку забезпечує формування кластерів, в яких будуть міститися словосполучення, які завжди використовуються разом.

3. Виконати пошук подібних векторів в утвореному масиві даних із використанням відстані Левенштейна для доповнення кластерів знайдених на попередньому кроці:

$$lev_{a,b}(i,j) = \begin{cases} \max(i,j) & \text{if } \min(i,j) = 0, \\ \min \begin{cases} lev_{a,b}(i-1,j) + 1 \\ lev_{a,b}(i,j-1) + 1 \\ lev_{a,b}(i-1,j-1) + 1_{(a_i \neq b_j)} \end{cases} & \text{otherwise.} \end{cases} \quad (3.4)$$

Для обчислення цієї відстані можна використати певне порогове значення, яке можна обчислювати як десятковий логарифм із максимальної довжини двох векторів-слів які порівнюються [109].

$$\lim lev_{a,b}(i,j) = \lg(\max(|a|, |b|)) \quad (3.5)$$

Усі знайдені подібні слова виділилися в окрему групу – кластер. Відповідно, до такого кластера входять усі фрази із набору даних, які містять виділені вище подібні слова.

Описана вище процедура виконується циклічно, перебираючи усі наявні слова у масиві даних.

4. Після виконання кроків 2 та 3, отримано сформовані кластери слів та їхні елементи у вигляді фраз із початкового набору даних. Оскільки слова у наборі даних можуть бути неповнозначні, видалено кластери, назви яких містять менше трьох символів.

5. Перебір усіх слів із заданого набору даних згідно процедур 2 та 3 супроводжується появою дублікатів. Саме тому на цьому кроці виконується

дедублікація отриманих кластерів. Також, на цьому кроці видаляються кластери які містять всього один елемент.

б. Отримані кластери слів із відповідними їм фразами з описів освітніх програм співставлено власне з назвами освітніх програм і отримано набір даних для виконання фільтрації на основі вмісту.

Три кластери серед отриманих результатів було видалено (Рис. 3.4), оскільки вони не несуть семантичного значення, це «та» (71 запис), «на» (5 записів), «різних» (11 записів).

Keyword discovery Umbrellum	
Group	Keyword
на	технолог на підприємствах
на	технік-проектант на підприємствах
на	управління на рівні районів
на	управління рухомими об'єктами на транспорті
на	Інтернет речей на транспорті

Рис.3.4 Приклад даних видаленого кластеру

3.3. Удосконалення двоетапного методу класифікації даних для підвищення точності передбачення успішності вступу вступника ЗВО

3.3.1. Збір та попереднє опрацювання даних вступної кампанії ЗВО

Початкова вибірка містила дані вступників вступної кампанії НУЛП 2022-го року з обмеженнями за денною формою навчання, відкритою пропозицією для прийому на навчання на перший курс для здобуття ступеня бакалавра.

На першому етапі збору даних, відповідно до встановлених обмежень дані було відібрано із кількох сутностей, що містили інформацію про деталі поданих вступниками заяв, статистику вступної кампанії за кількістю поданих заяв та кількісних характеристик освітніх пропозицій (кількість місць), та інформацію за результатами вступної кампанії – статусами допущених до конкурсу заяв. Для цього, на етапі збору даних було створено проміжні таблиці

(«enrolment_temp_rating», «enrolment_temp_statistic», «enrolment_temp_status» Рис. 2.7). Кінцевий набір даних з усіма необхідними атрибутами можна використовувати для подальшого інтелектуального аналізу за умови успішного застосування попередньої обробки даних. Це один із найважливіших етапів, який передбачає підготовку даних – трансформацію, інтеграцію, очищення, зменшення розмірності, нормалізацію даних, тощо.

Ці кроки можуть відрізнятися залежно від конкретного набору даних і поставленого завдання машинного навчання. Основною задачею цього етапу є підготовка даних таким чином, щоб максимізувати ефективність моделі МН [104]. Під ефективністю в цьому випадку слід розуміти поєднання декількох критеріїв, які повинні забезпечити максимізацію точності та генералізації при мінімізації часових ресурсів та ресурсів пам'яті під час роботи того чи іншого методу машинного навчання.

Враховуючи особливості зібраного набору даних, основними процесами попередньої обробки даних на цьому етапі були наступні (Рис 2.7):

1. Анонімізація даних.
2. Видалення нерелевантних даних та даних, що дублюються.
3. Інтеграція даних – об'єднання даних в один набір.
4. Виявлення та обробка неповних даних та викидів.
5. Вибір найбільш вагомих ознак для аналізу

Усі операції з даними було здійснено за допомогою запитів SQL у середовищі phpMyAdmin. Запити SQL було використано у вирішенні завдань очищення та підготовки даних до подальшого аналізу завдяки своїй здатності ефективно відбирати, фільтрувати, оновлювати та видаляти дані. Запити щодо аналізу даних (виявлення дублів, неповних даних та викидів), їхнє очищення та доповнення здійснювались над проміжними таблицями, які було створено на етапі збору даних. Приклад процесу відбору даних показано на Рис. 3.5.

Початкова вибірка даних містила усі існуючі властивості (поля таблиць БД ІС які були джерелами даних) аби уникнути можливих втрат релевантних даних

вплив результатів пільгових вступників на рейтинговий список, що знову ж таки підтверджує необхідність видалення цих даних (Рис. 3.5, рис 3.6).

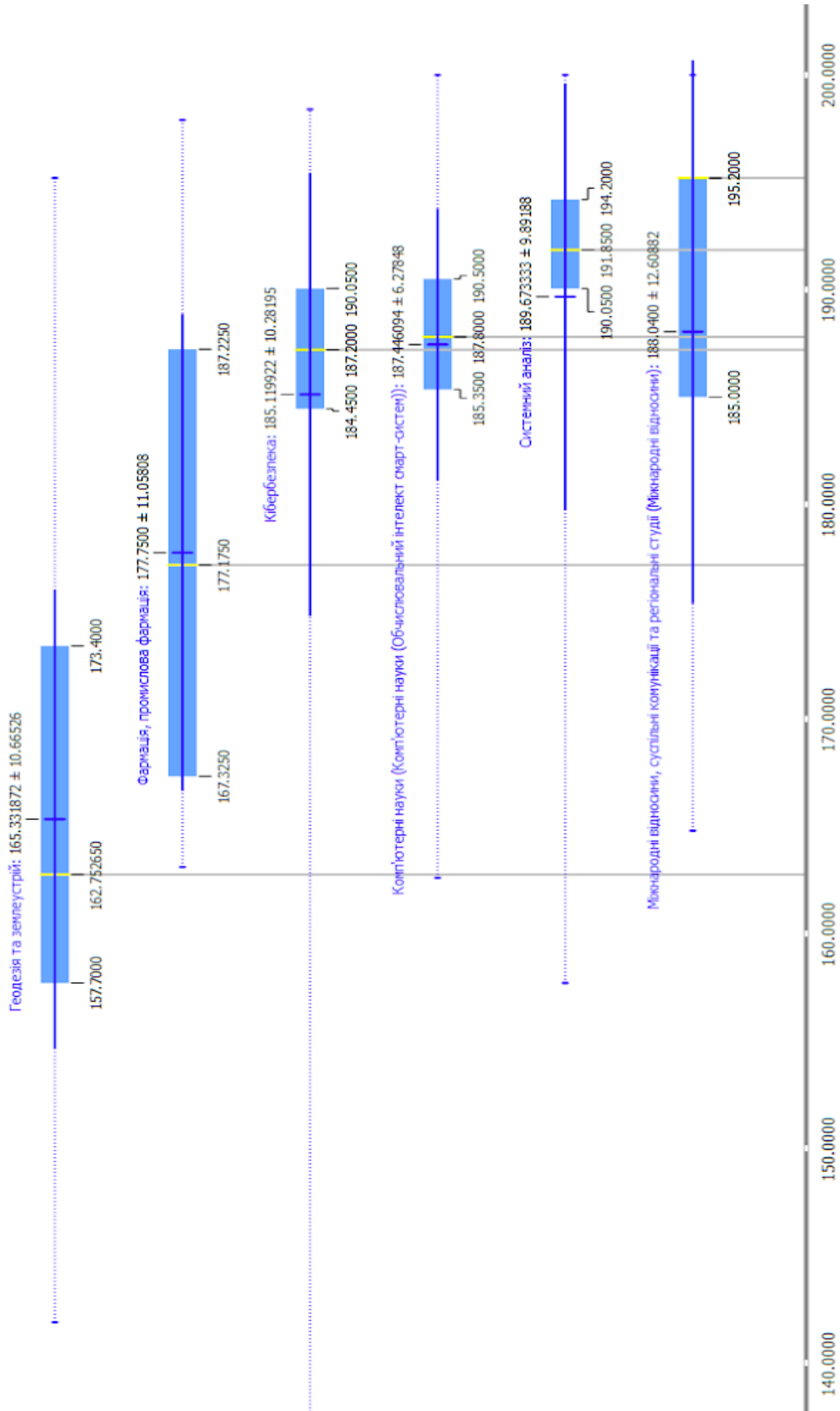


Рис.3.5 Діаграма розмаху із врахування конкурсного балу пільгових вступників

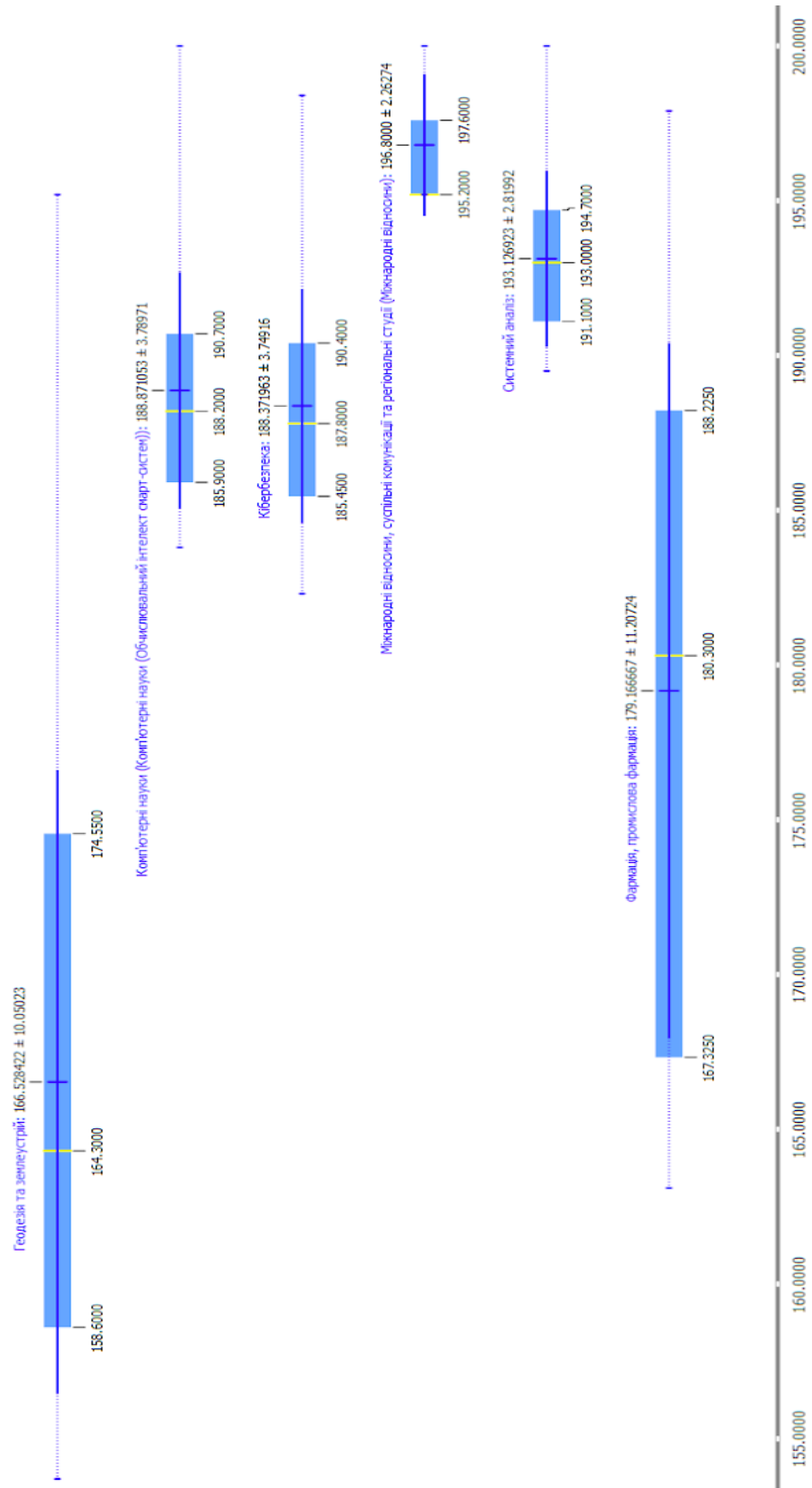


Рис.3.6 Діаграма розмаху без врахування конкурсного балу пільгових вступників

Після виявлення викидів та аномалій необхідно проаналізувати причини їхнього виникнення та здійснити їхнє видалення, нормалізуючи перетворення або ж залишити без змін. Так, з огляду на мету та контекст аналізу даних, та природу визначених викидів, їх кількість – найкращим варіантом буде вилучення рядків, які містять викиди. Це вступники, що були зараховані за пільговими умовами, тож бали можуть бути значно нижчими у деяких випадках. Такі значення не є релевантними та важливими для дослідницького питання. Оскільки дані були відібрано з кількох сутностей (бази даних БД ІС «Абітурієнт»), попередньо узгодивши типи та розмірності полів, дані було зведено в один набір даних.

Основним принципом відбору властивостей (атрибутів) набору даних є наявність їх як показників формування рейтингового списку вступників. Такий вибір зумовлений чітко встановленими правилами прийому до ЗВО, які прямо впливають на результат успішності вступу вступника. Так, серед визначених правилами прийому чинників, враховуючи вибірку цього дослідження (бакалавр на основі ПЗСО), було обрано наступні властивості:

- конкурсний бал (формує рейтинг вступників більшого до меншого);
- пріоритетність заяви (від меншої до більшої);
- оцінки з предметів (від більшого до меншого) [102].

Для конкурсного відбору для здобуття ступеня бакалавра, зараховуються бали НМТ з української мови (перший предмет), математики (другий предмет) та історії України (третій предмет), або бали ЗНО 2019-2021 років з трьох конкурсних предметів (у будь-яких комбінаціях), передбачених Правилами прийому [102].

Конкурсний бал – як комплексне оцінювання досягнень вступника, яке включає результати вступних випробувань та інші показники, розраховані відповідно до правил про вступ у кожному окремому ЗВО. Важливо, що окрім результатів НМТ (або ЗНО), формула також передбачає інтегральні вагові коефіцієнти, визначені для кожної освітньої програми, та коригувальні

коефіцієнти – регіональні (для тих, хто вступає до ЗВО області) та галузеві (для освітніх програм, які отримати спеціальну підтримку). рейтинговий список вступників впорядковується

Іншою важливою властивістю набору даних, що впливає на формування рейтингового списку відповідно до Правил прийому є пріоритет заяви. Визначена вступником кожній заявці пріоритетність визначає порядок заяв від 1 до 5, де найбільш пріоритетною є заява з першим пріоритетом та встановлюється вступником для заяв на бюджетну форму навчання. Відповідно до правил прийому, кількість вступників, які можуть бути зараховані обмежена кількістю місць державного замовлення, для відкритих пропозицій – це максимально можлива кількість місць державного замовлення, на які можна отримати рекомендацію. Характеристики початкового набору даних подано у Таблиці 3.1.

Відповідно до принципів відкритої конкурсної пропозиції бюджетні місця надаються вступникам в яких конкурсний бал найвищий. Конкурс відбувається між усіма вступниками усіх ЗВО, що оголосили набір на певну освітня програма. Таким чином за відкритими конкурсними пропозиціями здійснюється розподіл вступників, який визначається як адресне розміщення державного замовлення. Таке розміщення полягає у рекомендації до зарахування вступників з найкращим конкурсним балом до тих ЗВО, які визначили для них найбільший пріоритет.

Такий принцип розподілу свідчить про важливість таких кількісних показників обсяг державного замовлення ЗВО та відповідно, кількість заяв поданих на конкурс для кожної окремої ОП. Семе значення не було включено у вибірку, оскільки цю кількість стає відоме лише на останньому етапі вступної кампанії. Натомість, саме значення успішності вступу було враховано саме за надими рекомендаціями. Таким чином, початкова вибірка містила записи, що відповідають заявам вступників та складалась з 12 властивостей, характеристики яких подано в Таблиці 3.2.

Таблиця 3.1. Характеристики початкової вибірки архівних даних вступу

п/п	Властивість	Машинна назва	Тип даних	Базові статистики
1	конкурсний бал	mark	числовий	min = 125, max = 200 ;
2	бал за предмет 1	mark_1	числовий	min = 125 , max = 200;
3	ваговий коефіцієнт предмету 1	weightsubject1	числовий	min = 0,25 , max = 0,45 ;
4	бал за предмет 2	mark_2	числовий	min = 100, max = 200 ;
5	ваговий коефіцієнт предмету 2	weightsubject2	числовий	min = 0,25 , max = 0,5 ;
6	бал за предмет 3	mark_3	числовий	min = 113, max = 200 ;
7	ваговий коефіцієнт предмету 3	weightsubject3	числовий	min = 0,25 , max = 0,5 ;
8	пріоритет заяви	priority	категорійний	[1;5]
9	максимально можлива кількість місць державного замовлення визначена ЗВО	statecount	числовий	min = 5, max = 238;
10	кількість місць державного замовлення за квотою	quotastatecount	числовий	min = 2, max = 42;
11	освітня програма	speciality_name	текстовий	назва ОП
12	результати конкурсу	enrolment_status	числовий	0 або 1 Рекомендовано бюджет = 1, не рекомендовано = 0

Для таких властивостей як «Освітня програма» та «Пріоритет» було виконано перетворення текстових та категоріальних змінних відповідно до обох властивостей на числові (бінарні) змінні. Таким чином, розмірність кінцевої вибірки даних зростає і зібраний набір даних містить 71 властивість (атрибути).

У попередньому підрозділі описано процеси формування набору даних із множиною певних властивостей (атрибутів), які повинні впливати на передбачення успішності вступу вступника ЗВО. Якщо передбачення успішності вступу вступника ЗВО реалізовувати методами машинного навчання, то ключовим етапом ефективного інтелектуального аналізу є вибір важливих атрибутів (feature selection). Цей процес полягає у виборі найбільш суттєвих факторів, які впливають на цільову змінну, та їхнє використання під час побудови методу чи моделі машинного навчання. Такий підхід може підвищити ефективність роботи останнього за рахунок як зменшення вимірності простору вхідних даних задачі так і підвищенням точності класифікації.

Існує чимало методів вибору важливих атрибутів. У цій роботі використано 4 різних методів. По-перше, різні методи працюють з різними типами даних та мають ряд переваг і недоліків, а тому отримані результати краще аналізувати із використанням їхньої сукупності. По-друге, такий підхід забезпечить підвищення рівня надійності отриманих результатів. У Таблиці 3.2, на основі [110] узагальнено принципи роботи, переваги та недоліки кожного із досліджуваних методів.

Таблиця 3.2. Принципи роботи, переваги та недоліки методів визначення важливих ознак у наборі даних

Метод визначення важливих ознак	Принцип роботи	Переваги	Недоліки
Поріг дисперсії (VarianceThreshold)	Виконує вилучення ознак з низькою дисперсією	Метод простий у використанні	Метод не враховує кореляцію з

			ЦІЛЬОВОЮ ЗМІННОЮ
Фільтр кореляції Пірсона (Pearson correlation filter)	Виконує вибір ознак із значущим рівнем кореляції з цільовою змінною	Метод забезпечує ефективний вибір ознак	Метод чутливий до лінійних зв'язків у даних
Фільтр взаємної інформації (Mutual Information filter)	Виконує вибір тих ознак які мають високу взаємну інформацію з цільовою змінною	Метод враховує нелінійні зв'язки	Метод чутливий до незначних кореляцій у даних
Фільтр Anova F- score (Analysis of Variance F-score filter)	Виконує вибір числових ознак із високим рівнем значущості на основі аналізу варіантності	Метод забезпечує ефективний вибір числових ознак	Метод чутливий до нерівномірно розподілених ознак

3.3.2. Опис та аналіз базових компонентів удосконаленого методу

В основі базового методу класифікації даних з [96], покладено послідовне використання ЙНМ та МОВ з рбф ядром. З метою отримання більш точних та надійних результатів класифікації, у цьому підрозділі описано удосконалений метод двоетапної класифікації даних. Він передбачає сумісне використання ЙМН та стекінгового об'єднання з чотирьох МОВ з різними ядрами (рбф, сигмоїдне, поліноміальне та лінійне), результати яких агрегуються

метакласифікатором [111]. В якості метакласифікатора обрано найбільш поширений для цих цілей алгоритм – логістична регресія. Розглянемо кожен із компонентів удосконаленого двоетапного методу більше детально

Як вже зазначалося вище, в основі удосконаленого двоетапного методу класифікації даних покладено послідовне використання ЙМН та стекінгового об'єднання чотирьох МОВ з різними ядрами. Така комбінація повинна забезпечити підвищення точності розв'язання поставленої задачі.

На першій стадії методу відбувається застосування ЙМН. Топологія цієї штучної нейронної мережі (ШНМ) наведено на Рис. 3.7.

ЙНМ складається із чотирьох шарів: вхідного шару, шару перетворення, шару сумування та вихідного шару. Основними кроками алгоритмічної реалізації ЙНМ є наступні:

- пошук Евклідових відстаней від поточного спостереження з невідомим виходом та усіма спостереженнями опорної вибірки даних (опорна – це набір усіх спостережень для яких відоме значення виходу);
- пошук гаусівських функцій від Евклідових відстаней;
- пошук набору ймовірностей належності поточного спостереження до кожного із класів, визначених задачею. Слід зазначити, що цей процес відбувається у шарі сумування ЙНМ і у цій роботі використовувалася модифікована у [83] версія формування ймовірностей, яка забезпечує врахування незбалансованості вибірки даних;
- пошук мітки класу для поточного спостереження по найбільшій із ймовірностей, отриманих на попередньому кроці.

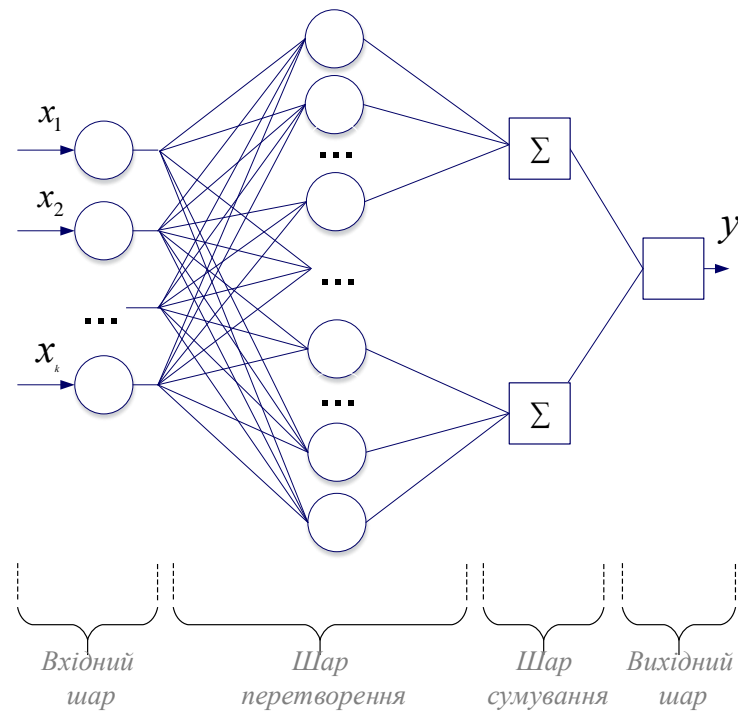


Рис.3.7 Топологія ймовірнісної нейронної мережі

Детальні математичні викладки цього методу подано у [112]. Як досліджено у [96], ЙНМ має ряд важливих переваг, зокрема відсутність процедури навчання, єдиний параметр який слід підбирати експериментальним шляхом, високі генералізаційні властивості та можливості представлення результату і як мітки класу і як набору ймовірностей належності до кожного із класів задачі. Все це служить базисом для використання ЙНМ як засобу попереднього опрацювання даних.

У цій роботі ЙНМ використовувалася як засіб розширення входів для методу другого етапу. З цією метою бралися дві ймовірності належності до обох класів задачі бінарної класифікації, які формуються у шарі сумування ЙНМ, утворюють повну систему подій та забезпечують врахування нерівномірності представлення обох класів у наборі даних [96].

Наступним компонентом удосконаленого методу є стекінгове об'єднання МОВ з чотирьох різних ядрами, зокрема рбф, сигмоїде, поліноміальне та лінійне [14]. Структурно-функціональну схему його роботи подано на Рис. 3.8.

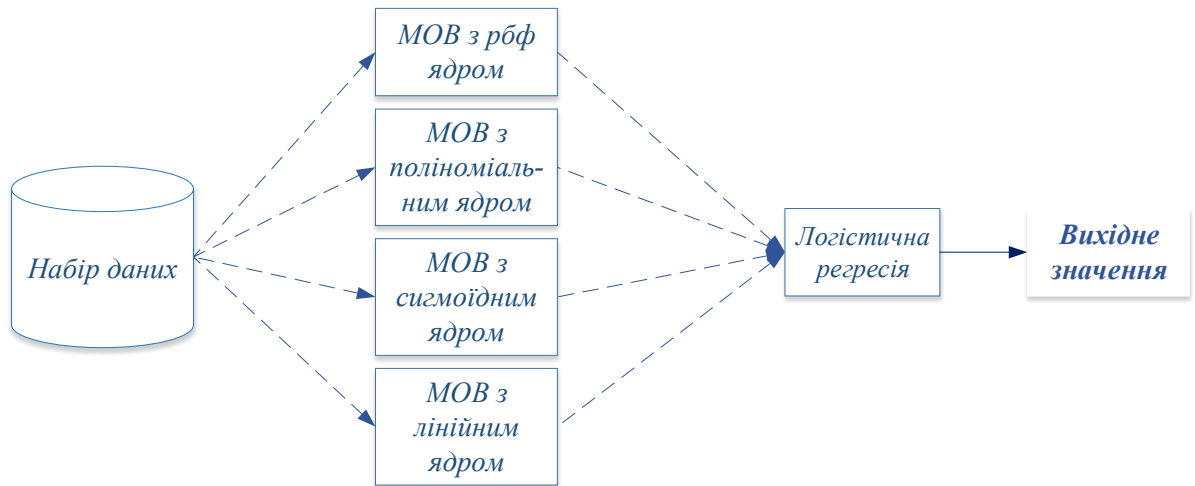


Рис.3.8 Структурно-функціональна схема роботи стекінгового об'єднання МОВ з чотирьома різними ядрами

Як видно з Рис. Рис. 3.8, наявний набір даних надходить до кожного із методів машинного навчання які передбачено стекінгом. У цьому випадку використовувалася МОВ з чотирьома ядрами. Саме використання різних ядер забезпечує декореляцію компонентів стекінгу, що дає змогу будувати ансамблевий метод з таких компонентів. Результати роботи кожного із чотирьох компонентів стекінового ансамблю формують новий набір даних, який подається на метакласифікатор. В цьому випадку використано логістичну регресію, яка забезпечує високу точність під час розв'язання поставленої задачі [95]. Власне вихід логістичної регресії і являється шуканою міткою класу для поточного спостереження.

Використання стекінгової моделі як компоненти удосконаленого двоетапного методу класифікації даних забезпечує зменшення впливу викидів, зменшення ризику перенавчання та підвищення точності результаті класифікації в порівнянні із кожною базовою моделлю яка покладено в основі стекінгового ансамблю [98]. Окрім цього, використання чотирьох, а не однієї моделі забезпечує більший рівень надійності отриманих результатів та вищий рівень довіри до отриманого результату [113], [114]

3.3.3. Алгоритмічна реалізація удосконаленого двоетапного методу класифікації даних

На Рис. 3.9 наведено структурно-функціональну схему роботи удосконаленого двоетапного методу класифікації даних [111].

Процедура навчання удосконаленого методу включає послідовне виконання таких кроків:

- застосування ЙНМ для синтезу набору ймовірностей належності до кожного класу задачі для кожного спостереження із навчального набору даних. Слід зазначити, що на цьому етапі, поточне спостереження із навчального набору даних для якого виконувалися обчислення – видалялося із набору даних;
- розширення початкового набору даних двома новими незалежними атрибутами – отриманими ймовірностями на попередньому кроці
- навчання чотирьох МОВ з різними ядрами на основі використання розширеного набору даних;
- застосування навчального набору даних на попередньо навчених МОВ з відповідними ядрами для синтезу вихідного значення мітки класу для кожного спостереження із навчального набору даних;
- формування нового набору даних із чотирьох стовпців, кожен з яких містить результати роботи відповідної МОВ стекінгового ансамблю.
- навчання логістичної регресії на новому наборі даних.

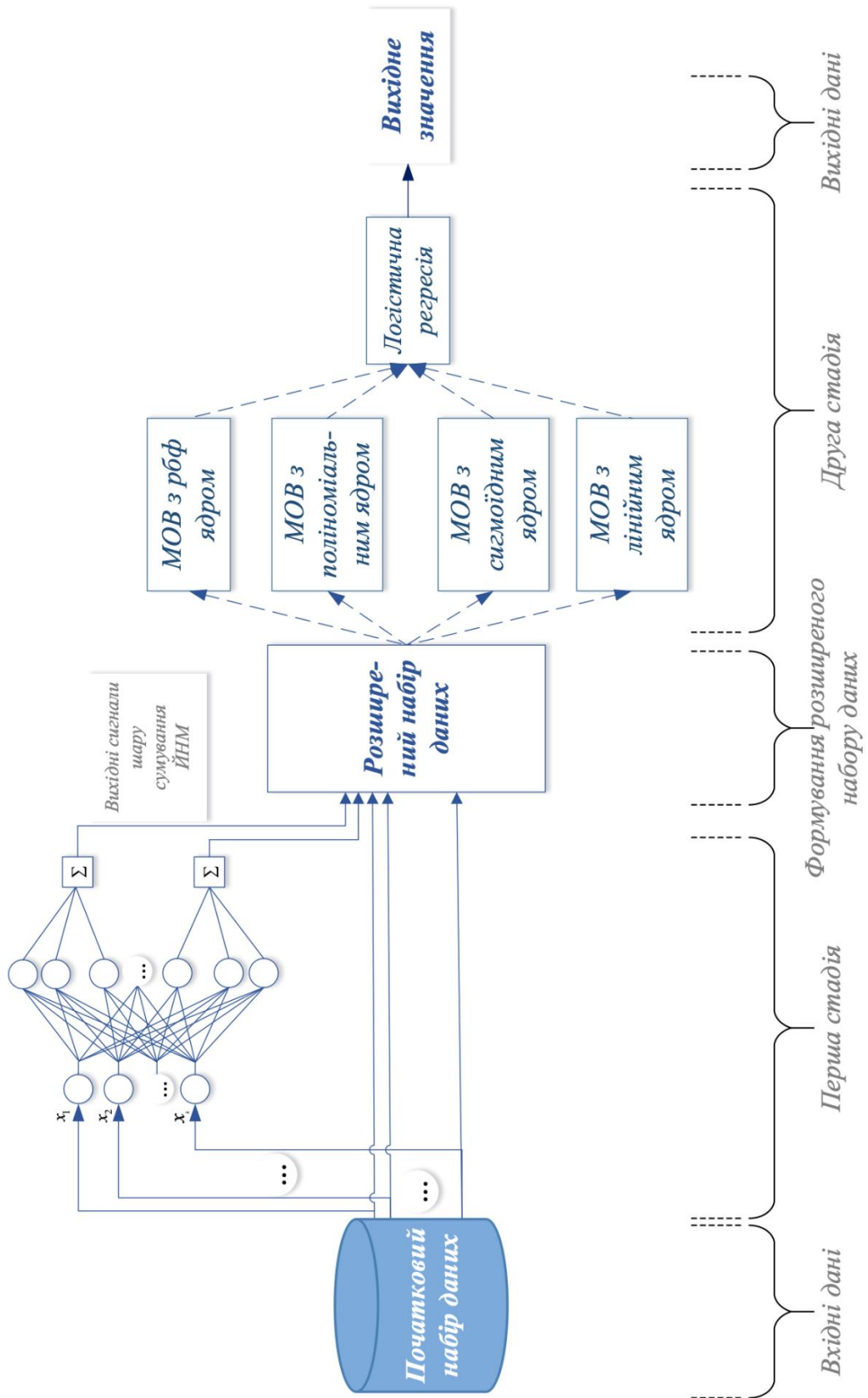


Рис.3.9 Структурно-функціональна схема роботи удосконаленого двоетапного методу класифікації даних

Процедура застосування удосконаленого методу для поточного спостереження з невідомим виходом включає послідовне виконання таких кроків:

- застосування ЙНМ для синтезу набору ймовірностей належності поточного спостереження до кожного із класів задачі;
- розширення поточного вектора двома новими ознаками – ймовірностями, отриманими на попередньому кроці;
- застосування розширеного поточного вектора даних на попередньонавчених МОВ з відповідними ядрами;
- формування нового вектора даних із чотирьох елементів, кожен з яких містить результати роботи відповідної МОВ;
- застосування попередньонавченої логістичної регресії на новому векторі даних і отримання шуканого результату

Висновки до розділу

Третій розділ дисертаційної роботи присвячено розробленню компонентів інформаційної системи підтримки прийняття вступників ЗВО, а саме компоненти для надання рекомендацій альтернативних освітніх програм та компоненти для передбачення успішності вступу до ЗВО. Серед вагомих отриманих результатів слід виділити наступні:

- розроблено процедуру розрахунку показника задоволеності студентів після завершення навчання на першому рівні вищої освіти за рахунок обчислення суми зважених результатів їхньої семестрової успішності, наукової діяльності та статусу продовження навчання в магістратурі, що дало змогу виконати ранжування отриманих результатів для надання релевантних рекомендацій вступнику із використанням колаборативної фільтрації;
- сформовано набір ключових слів освітніх програм на основі відбору описів професійних профілів їхніх випускників та виконано формування кластерів цих слів на основі відстані Левенштайна для рекомендацій

альтернативних освітніх програм, які відповідають інтересам вступника із використанням фільтрації на основі вмісту;

- проведено збір та виконано ряд процедур попереднього опрацювання архівних даних вступної кампанії, що стало основою для застосування методології машинного навчання під час розв'язання задачі передбачення успішності вступу до ЗВО;

- вибрано та обґрунтовано використання чотирьох методів визначення важливих для аналізу атрибутів зібраного набору архівних даних вступу, врахування результатів яких покладено в основі прийняття рішення щодо можливості зменшення кількості незалежних атрибутів задачі для підвищення точності та генералізаційних властивостей роботи класифікатора на основі машинного навчання під час аналізу такої вибірки даних.

- проведено аналіз ефективності роботи різних класів методів машинного навчання та штучних нейронних мереж, що показало необхідність у вдосконаленні існуючого інструментарію машинного навчання для розв'язання поставленої задачі;

- удосконалено двоетапний метод класифікації даних на основі послідовного використання ймовірнісної нейронної мережі та машини опорних векторів з рбф ядром, шляхом заміни останньої на стекінгове об'єднання машини опорних векторів з чотирьома різними ядрами і агрегуванням результатів їхньої роботи логістичною регресією, що забезпечить підвищення ефективності розв'язання задачі бінарної класифікації.

Розділ 4. Розробка інформаційної системи підтримки прийняття рішень вступників ЗВО та апробація результатів її роботи

У четвертому розділі розроблено архітектуру інформаційної системи підтримки прийняття рішень вступників ЗВО яка, на відміну від існуючих використовує поєднання методу надання рекомендацій на основі колаборативної фільтрації та фільтрації на основі вмісту, та удосконаленого методу передбачення успішності вступу.

Здійснено оцінку рівня охоплення освітніх програм у наборі даних колаборативної фільтрації та фільтрації на основі вмісту. Подано результати використання колаборативної фільтрації, що відображають можливі розширення рекомендацій на основі архівних даних вступу та освітніх даних студентів. Подано результати перетину освітніх програм в межах виділених кластерів, а відповідно і можливих рекомендацій отриманих за допомогою методу фільтрації на основі вмісту.

Проведено експериментальні дослідження, під час яких встановлено, що використання 71 незалежної ознаки для передбачення успішності вступу до ЗВО є обґрунтованим. Проведено порівняння ефективності роботи існуючих методів машинного навчання та штучних нейронних мереж. Проведено моделювання роботи удосконаленого двоетапного методу класифікації даних і шляхом порівняння з існуючими методами з різних класів встановлено підвищення точності його роботи під час розв'язання задачі передбачення успішності вступу до ЗВО.

4.1. Архітектура інформаційної системи підтримки прийняття рішень

У цьому дослідженні моделювання архітектури ІСППР здійснено як веб-сервіс, оскільки у порівнянні з традиційними системами на основі настільних

персональних комп'ютерів такий підхід забезпечить кращу доступність, масштабованість, безпеку, одночасно спрощуючи процеси розгортання та обслуговування.

Для логічного визначення зв'язків між сервером і стороною клієнта подано модель архітектури ІСППР вступників ЗВО. Запропонована архітектура відповідає парадигмі проектування трьох рівнів: рівня відображення, рівня логіки і рівня керування даними (Рис. 4.1).

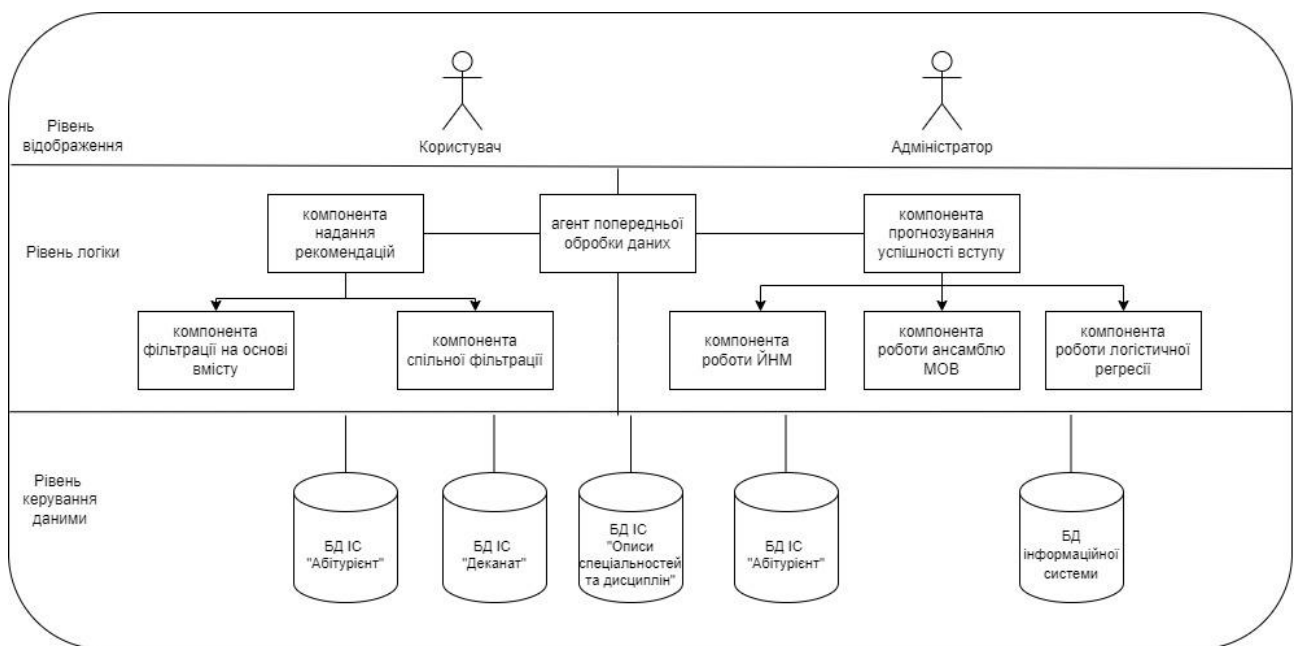


Рис.4.1 Трирівнева архітектура інформаційної системи підтримки прийняття рішень вступників

Рівень відображення передбачає взаємодію ІСППР зі вступником, що відповідає за представлення йому інформації та збір вхідних даних цього користувача, обробляє взаємодію користувача та відображає дані, опрацьовані на рівні логіки.

Рівень логіки – середній рівень, який містить основну логіку веб-сервісу та його функціональність. Обробляє запити користувачів, виконує бізнес-логіку та координує зв'язок між рівнями відображення та доступу до даних. Цей рівень включає обчислювальні процеси, надання рекомендацій та передбачення успішності вступу, які власне формують функціонал системи.

Рівень керування даними охоплює компоненти та процеси управління, зберігання, отримання та маніпулювання даними для підтримки прийняття рішень вступників. Цей рівень забезпечує доступ до відповідних джерел даних і забезпечуючи якість, цілісність і безпеку даних.

Враховуючи описані в попередніх розділах процеси попередньої обробки даних, архітектура системи повинна включати агенти часткової попередньої обробки даних. Дані, що забезпечують роботу компонент надання рекомендацій та передбачення успішності вступу можуть виконуватись автоматизовано для кожної вступної кампанії. Зокрема, це збір та попередня обробка:

- архівних даних вступу для визначення подібності інтересів вступників;
- архівних даних вступу для визначення інтервалів конкурсного балу;
- даних описів освітніх програм;
- освітніх даних для визначення рівня показника задоволеності студента;
- архівних даних вступу для роботи моделі МОВ-ЙНМ;

Тож виконання процедур щодо опрацювання вищезгаданих наборів даних доречно виконувати з використанням віртуальних таблиць (подання – Views), що базуватимуться на наборі результатів SELECT-запитів. Вона багато в чому поводитьься як таблиця, дозволяючи робити запити до неї, об'єднувати її з іншими таблицями та виконувати різні операції над нею, але сама не зберігає жодних даних. Натомість воно динамічно отримує дані з базових таблиць, визначених у запиті подання, щоразу, коли до нього звертаються.

Також, одиниці роботи або завдання, які можна запланувати на автоматичне виконання відповідно до заздалегідь визначеного розкладу варто виконувати за допомогою визначених завдань (Job). Це дозволить підвищити операційну ефективність, зменшити ручне втручання і гарантувати, що критичні завдання виконуються надійно і вчасно.

Операції з даними, які описані нижче, здійснювались за допомогою SQL – це стандартна мова програмування, яка використовується в управлінні реляційними базами даних і виконує важливі операції з їхніми даними. Створений у 70-х роках, SQL є найпоширенішим інструментом адміністраторів баз даних. Він також широко використовується для аналітичних запитів та написання сценаріїв для інтеграції даних аналітиками даних і розробниками [115]. Збір даних у цьому дослідженні здійснено методом вилучення з існуючих баз даних інформаційних систем НУЛП.

Процес перетворення даних, було здійснено з метою перетворення необроблених даних в придатний для використання формат шляхом видалення дублів, перетворення типів даних і зменшення набору даних. Перетворення формату або структури даних було здійснено з метою забезпечення сумісності з системою, де вони зберігаються. Це один з етапів процесу вилучення, перетворення, завантаження, який необхідний для доступу до даних і їхнього використання для прийняття рішень.

На етапі збору дані завантажуються в центральне сховище (в нашому випадку сервер бази даних та окрема БД) з різних джерел, оскільки початково вони знаходяться в необробленому вихідному вигляді, який непридатний для використання у даній системі (ці дані нормалізовані/денормалізовані для тих систем у яких вони використовуються). Так дані стають доступними, безпечними та зручнішими, що дає змогу використовувати їх різними способами. Щоб забезпечити зручність використання видобутих даних, їх було перетворено у потрібний формат.

Цей процес перетворення даних передбачає визначення структури, відображення даних, вилучення даних із вихідної системи, виконання перетворень, а потім збереження перетворених даних у відповідному наборі даних.

Так, по кожному з наборів даних було здійснено наступні трансформації:

- перевірка та очищення даних;

- перетворення форматів даних – частково змінено типи даних для забезпечення їхньої сумісності;
- структурування ключів – перетворено значення з вбудованими значеннями на загальні ідентифікатори для використання як унікальних ключів;
- стандартизація імен змінних;
- об'єднання одного набору даних з іншим;
- встановлення зав'язків між наборами даних.

Після цього з отриманими даним було здійснено ряд маніпуляцій, що передбачає створення нових значень із існуючих або зміну поточних даних за допомогою обчислень. Основні маніпуляції по кожному з набору даних описані нижче та передбачали:

- обчислення між стовпчиками для отримання нових значень
- узагальнення, яке агрегує значення;
- перетворення текстових/категоріальних даних;
- сортування, упорядкування та індексація даних для покращення ефективності пошуку;
- нормалізація значень шляхом їхнього розміщення на узгодженій шкалі.

Процес відділення (*separating*) здебільшого стосувався блоку даних освітніх програм та передбачав поділ значень даних на частини для детального аналізу. У цьому випадку розбиття передбачає поділ одного стовпця з кількома значеннями на окремі стовпці з кожним із цих значень. Це дає змогу фільтрувати на основі певних значень (ключові слова освітніх програм).

Таким чином запропонована архітектура передбачає введення агентів попередньої обробки даних, що забезпечить необхідну якість даних і гарантує, що подальший аналіз або обробка ґрунтується на точній інформації. Автоматизація завдань попередньої обробки даних може сприяти зниженню рівня витрат часу та зусиль, необхідних для очищення, трансформації та підготовки даних до аналізу, оскільки такі процеси є повторюваними (Рис. 4.2).

Робота таких агентів у запропонованій архітектурі передбачає, зокрема, процеси попередньої обробки даних щодо розрахунку показника успішності навчання, показника продовження навчання в магістратурі, наукової діяльності, формування вибірки архівних даних вступників, формування ключових слів та кластерів освітніх програм, формування множини атрибутів набору даних для передбачення вступу. Збережені у базі даних ІСППР такі набори даних забезпечать роботу розроблених компонент – надання рекомендацій альтернативних освітніх програм та передбачення успішності вступу.

Для кращого розуміння організації та структурування даних нижче подано діаграму, що включає сутності які було створено для роботи компонент ІСППР (Рис. 4.3.).

- «user», «webform_custom_prediction» «webform_custom_filtering» – містять внесені дані для надання рекомендацій та передбачення успішності вступу;
- «enrolment_prediction» – містить початковий набір даних для роботи ЙНМ (історичні дані вступної кампанії 2022-го року);
- «enrolment_prediction_pnn» – містить розширений набір даних після роботи ЙНМ;
- «enrolment_prediction_svm» – містить дані результатів роботи МОВ;
«enrolment_prediction_lr» - містить дані результатів роботи логістичної регресії;
- «admission_statistic_2023» – містить параметри ОП 2023 року вступу;
- «program_keywords_filtering» - містить дані ОП, ключових слів та кластерів;
- «keywords_cluster_filtering» – містить дані ОП, ключові слова та їхні кластери;
- «program_filtering» – містить перелік ОП 2023 року вступу;
- «students_rating_filtering» – містить освітні дані студентів;

- «admission_filtering» – містить архівні дані вступної кампанії 2019-го року;

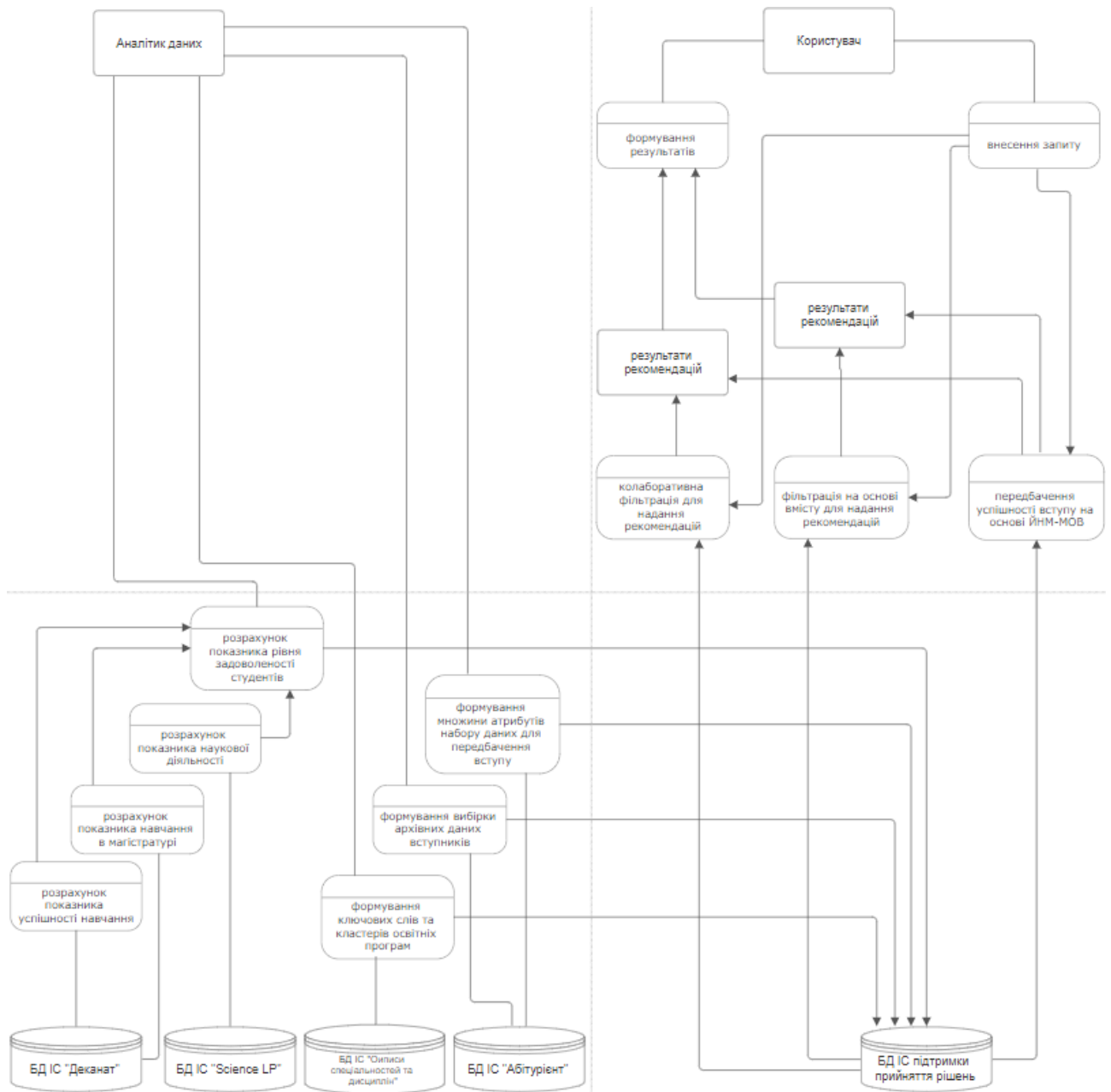


Рис.4.2 Архітектура інформаційної системи підтримки прийняття рішень вступників ЗВО

Сутності «users» та «webform_custom_insert» зберігають дані користувачів та їхніх запитів. Для збереження даних відповідно до потреб компонент було створено сутності, структура яких відповідає параметрам набору даних для роботи ЙНМ-МОВ та надання рекомендацій («webform_custom_prediction», «webform_custom_filtering»).

Робота такої структури передбачає, що для реалізації двоступеневої моделі ЙНМ-МОВ тестовими даними є вектор, що формується на основі запиту користувача та даних статистики вступної кампанії, а як навчальні дані використано сутність «enrolment_prediction». Результати роботи алгоритмів ЙНМ та чотирьох МОВ зберігаються у сутностях «enrolment_prediction_pnn» та «enrolment_prediction_svm», «enrolment_prediction_lr».

Сутність «enrolment_prediction» містить наступні властивості (для кращої візуалізації поля mark, priority, weightsubject, specialization_name вказано лише один раз):

- «priority_1», «priority_2», «priority_3», «priority_4», «priority_5» – значення пріоритетів;
- «mark1», «weightsubject1», «mark2», «weightsubject2», «mark3», «weightsubject3» – бали конкурсних предметів та їхні коефіцієнти та «mark» – конкурсний бал;
- усі властивості «specialization_name**» – відповідають освітнім програмам;
- «statecount» – максимально можлива кількість місць державного замовлення визначена ЗВО
- «quotastatecount» – кількість місць державного замовлення за квотою
- «enrolmentstatus_recomendation_name» – результати конкурсу

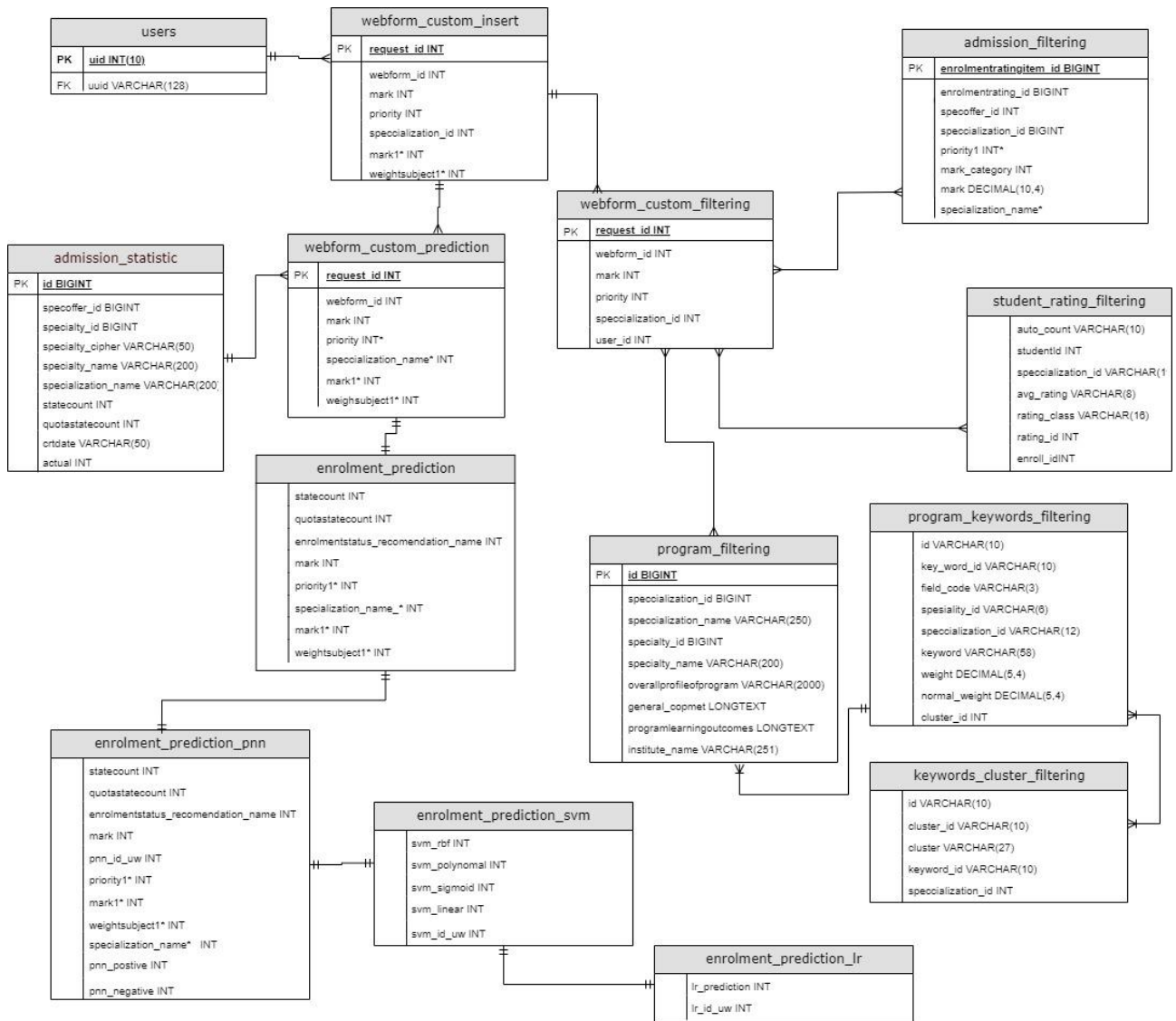


Рис.4.3 ERD діаграма бази даних ІСППР вступників ЗВО

Властивості сутності «admission_statistic_2023» наступні: «specoffer_id», «specialty_id», «specialty_cipher», «specialty_name», «specialization_name» описують дані освітньої програми. Значення даних за рештою властивостей відповідають таким як і в попередніх сутностях.

Сутність «program_keywords_filtering» містить дані освітньої програми – «field_code», «speciality_id», «specialization_id», дані ключових слів та ідентифікатор кластера – «key_word_id», «keyword, weight», «normal_weight», «cluster_id».

Сутність «keywords_cluster_filtering» містить ідентифікатори освітньої програми «specialization_id», та ключових слів «key_word_id» та дані кластера – «cluster» та «cluster_id».

Сутність «program_filtering» містить ідентифікатор програм «specialization_id» та кілька властивостей, що описують саму програму, зокрема «specialization_name», «specialty_name», «overallprofileofprogram», «general_compet», «programlearningoutcomes» - освітня програма, загальний опис програми, загальні компетентності, результати навчання, навчально-науковий інститут.

Сутність «student_rating_filtering» містить властивості, що описують результати оцінки наукової діяльності - «science_class», «sciencepoints», продовження навчання в магістртурі «master_status», «master», рейтинг успішності – «Rating_id», «rating_class», «AVG_Rating».

Сутність «admission_filtering» містить ідентифікатор програм «specialization_id», перетворені з категоріальних даних значення пріоритетів та освітніх програм – «priority1*», «specialization_name*» (для кращої візуалізації вказано лише один раз) та конкурсного балу – «mark_category», «mark».

4.2. Апробація результатів роботи інформаційної системи надання рекомендацій вступникам ЗВО

4.2.1. Оцінювання рівня охоплення освітніх програм у наборі даних для надання рекомендацій

Аналіз набору даних здійснено шляхом визначення міри охоплення пропонованих ЗВО ОП. Для цього було проаналізовано усю кількість запропонованих програм у порівнянні, до тих, які покриває певний показник, за яким здійснюється фільтрація. Зокрема, в межах кожної з ОП це наявність вступників з першим пріоритетом, що використовується для визначення подібностей, наявність різних інтервалів конкурсного балу та наявність більше однієї заяви з більшими пріоритетами.

Кількість пропонованих ЗВО освітніх програм – 6. Такі різниця полягаю у появі нових ОП у 2023. Цю різницю покрито результатами фільтрації на основі вмісту. Другим показником є інтервал конкурсного балу. Результати аналізу сформованої вибірки даних на Рис. 4.4.

На Рис.4.4 показано, що у тих ОП, де набір менший, покриття різними інтервалами КБ буде меншим. Проте, кожна з ОП покриває різну кількість інтервалів, де мінімальна кількість – чотири.

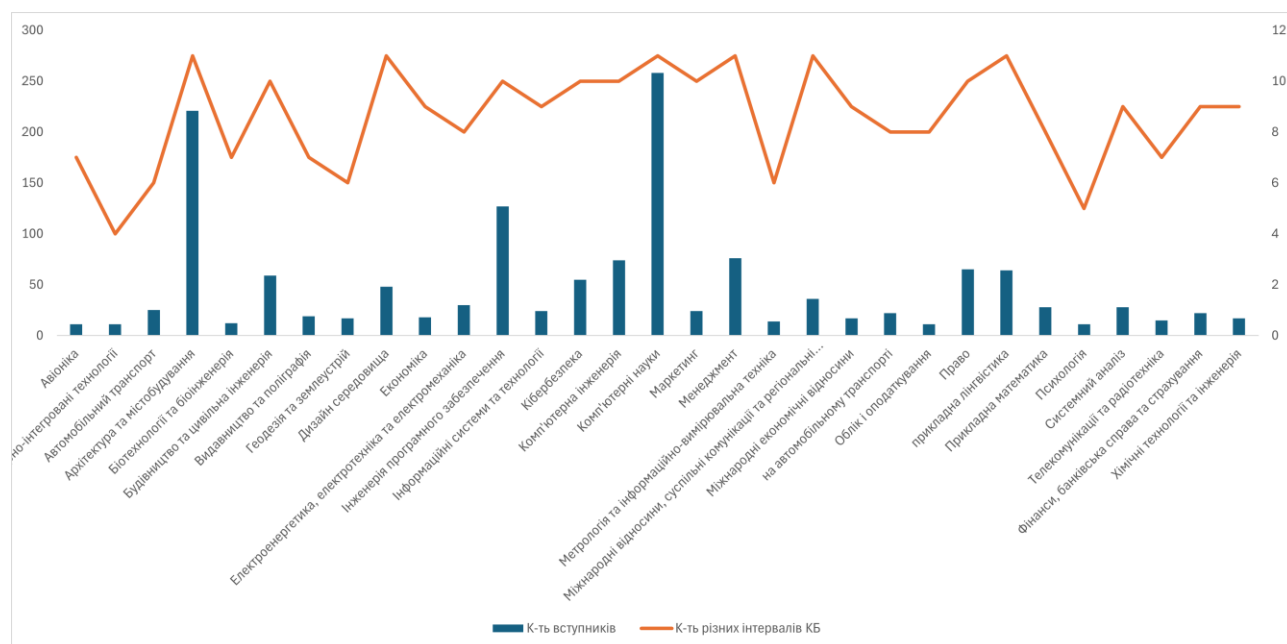


Рис.4.4 Порівняння показника кількості інтервалів та загальної кількості вступників в межах освітніх програм

Третій показник, що впливає на фільтрацію – показник рівня задоволеності. Вибірка даних користувачів з першим пріоритетом показує, що у всіх ОП є показник більше середнього. Діаграма на Рис. 4.5 відображає покриття високим рівнем показника задоволеності всіх ОП.

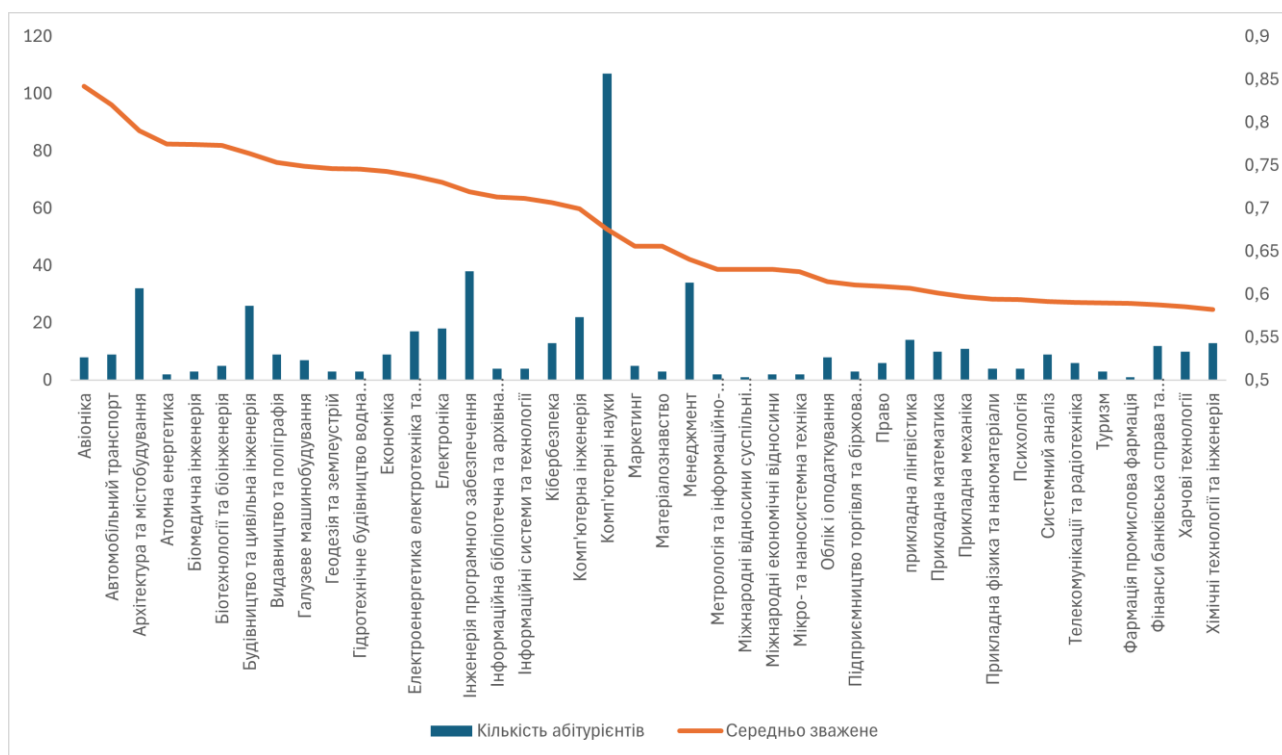


Рис.4.5 Порівняння показника задоволеності студента та загальної кількості вступників в межах освітніх програм

Оскільки ключові слова та кластери було сформовано на основі даних актуального переліку ОП ЗВО, результати повністю покривають пропонувані програми.

4.2.2. Результати роботи колаборативної фільтрації

Метою виконання колаборативної фільтрації є надання рекомендацій вступнику на основі історичних даних студентів, зібраних за 4 роки навчання бакалаврату та факту вступу в магістратуру. Врахування таких показників як: рівень успішності навчання на бакалавраті (за усі 4 роки навчання); рівень проведення наукової роботи (якщо така проводилася); та вступ у магістратуру за спеціальністю, на якій відбувалося навчання на бакалавраті, можуть характеризувати успішність та задоволеність навчанням на обраній освітній програмі. На основі цього можна виконати відбір набору даних мотивованих студентів, які задоволені вибором освітньої програми та процесом навчання на ній. Проте, для отримання більш точних результатів відбору даних, було введено

ряд коефіцієнтів які дозволяють зважити вплив кожного із трьох вищевказаних показника. Зокрема, для першого показника введено коефіцієнт 0,5, для другого – 0,2 і для третього – 0,3. Зваживши значення досліджуваних показників із використанням цих коефіцієнтів, знайшовши середнє значення суми їхніх значень на основі аналізу усіх доступних спостережень, було відкинуто тих випускників бакалаврату, для яких значення цього показника є нижче середнього рівня.

Початковий набір даних містив 1015 студентів з різних освітніх програм. Після фільтрування, із використанням методики описаної вище, при використанні порогового значення показника відбору на рівні 0,580565008, вдалося відкинути менш вмотивованих студентів. Загальна кількість спостережень, які залишилися у досліджуваному набору містить 502 записи.

Подальші дослідження базувалися на наданні рекомендацій альтернативних освітніх програм вступнику, які являються найближчими до тих, що залишилися у відфільтрованому наборі даних на основі таких показників як:

- освітня програма за якою навчався студент
- альтернативні освітні програми які він обирав під час вступу
- пріоритет заявки
- конкурсний бал у визначених інтервалах (130-135, 135-140, 140-145, 145-150, 150-155, 155-160, 160-165, 165-170, 170-175, 175-180, 180-185, 185-190, 190-195, 195-200).

На Рис. 4.6 подано візуалізацію результатів такої фільтрації. Для кращого розуміння результатів візуалізації зробимо деякі роз'яснення. На осі *ox* подано досліджувані ОП на яких навчалися студенти. На осі *oy* подано альтернативні ОП обрані кожним відповідним студентом під час вступу. Відповідно, на перетині обох осей міститимуться студент чи студенти із вказанням освітньої програми на якій вони вчилися, та усіх альтернативної освітньої програми яку вони обрали під час вступу. При цьому, різні фігури позначають різний інтервал середнього балу відповідного студента, а різних колір чітко визначає назви

альтернативних освітніх програм, вибраних ним під час вступу. Слід зазначити, що сірим кольором позначено усі освітні програми які не можна було візуалізувати через їхню велику кількість та обмеженість інструменту візуалізації.

Запропонований вид фільтрації забезпечує рекомендації освітніх програм від 1 до 5.

Проте, такий тип фільтрації має свої обмеження. Серед очевидних, слід зазначити обмеженість або взагалі неможливість надання рекомендації вступнику щодо вибору альтернативних програм у випадку їхньої невеликої кількості або взагалі їхньої відсутності у досліджуваному наборі історичних даних.

Зокрема, на Рис. 4.7 наведено інформацію про доступний набір даних на основі якого відбувається формування рекомендації згідно з колаборативною фільтрацією. Зокрема, тут наведено інформації про частоту вибору альтернативних освітніх програм студентами, відповідно до тої на якій вони навчалися.

На осі ox подано перелік освітніх програм на яких навчалися студенти. На осі oy подано частоту вибору альтернативних освітніх програм, які обрані кожним відповідним студентом під час вступу. З гістограми чітко видно, що привабливі для вступу освітньої програми мають чимало альтернатив для коректної роботи колаборативної фільтрації. Проте, для менш популярних освітніх програм, практично відсутні альтернативи, або взагалі немає. Це є важливим обмеженням цього типу фільтрації.

Саме тому, у цій роботі пропонується сумісне використання фільтрації цього типу та фільтрації на основі вмісту. Такий підхід дозволить нівелювати вищевказаний недолік та забезпечить можливість розширення переліку альтернативних освітніх програм, які відповідають інтересам вступника.

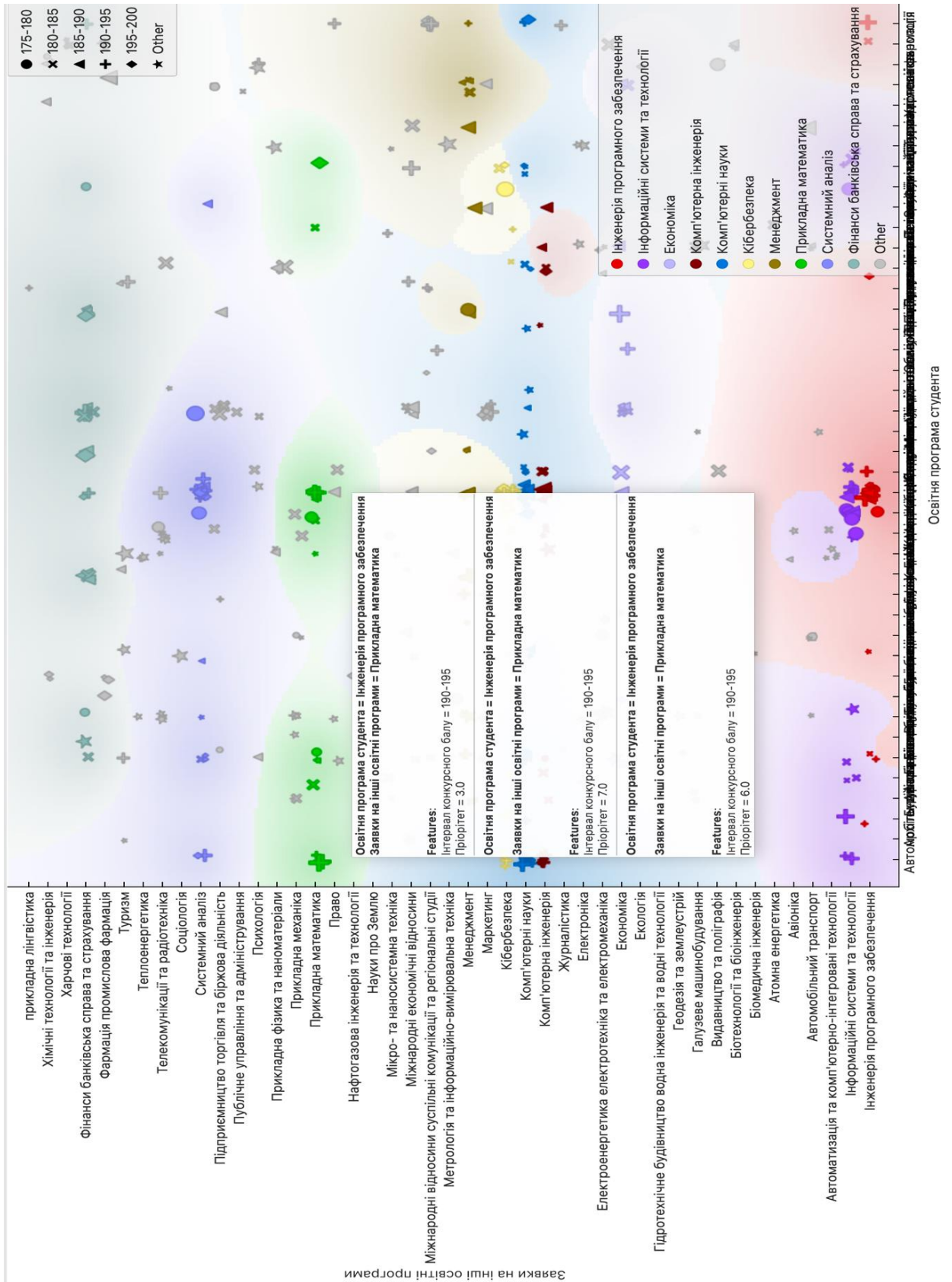


Рис.4.6 Часткова візуалізація результатів роботи колаборативної фільтрації

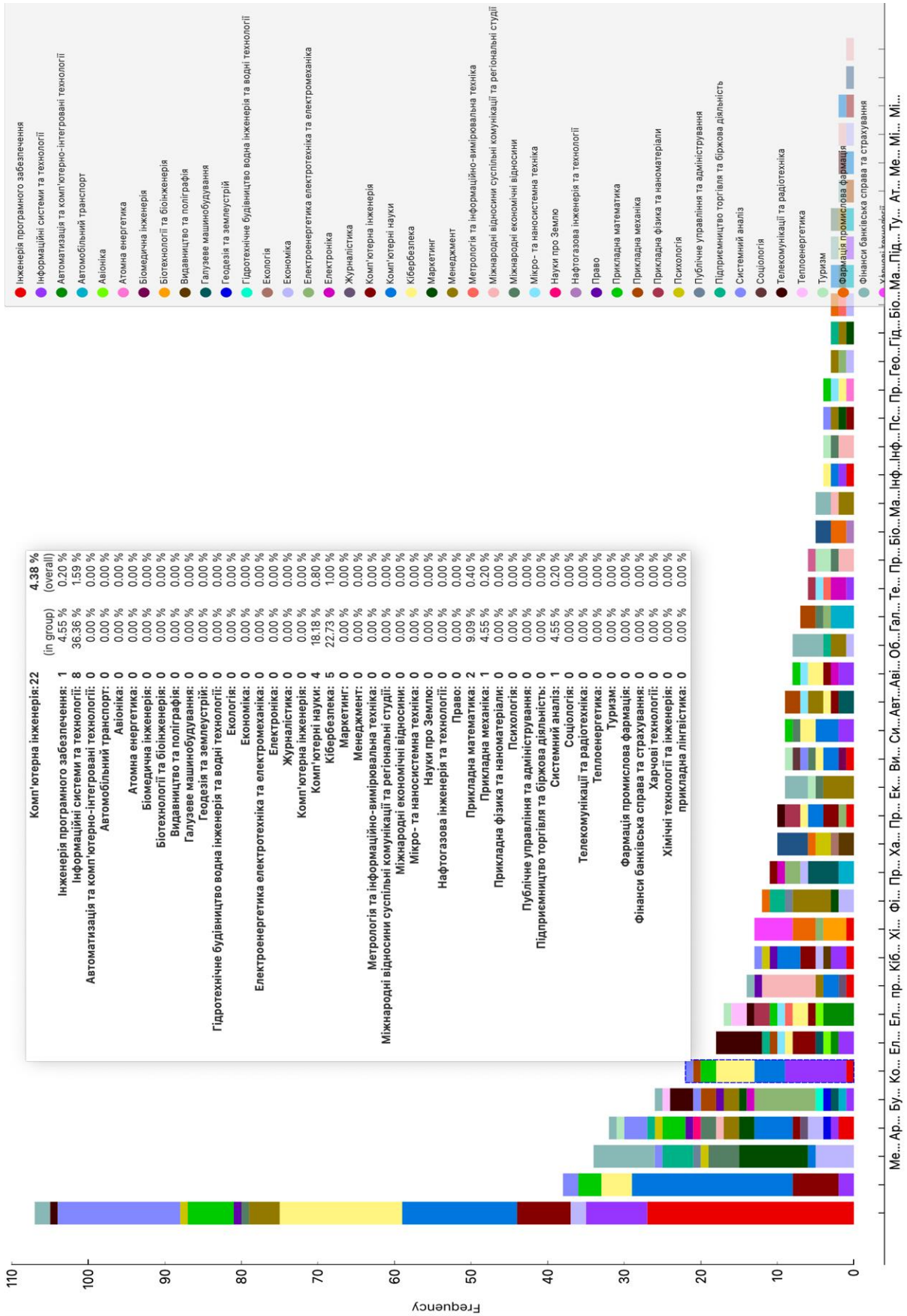


Рис.4.7 Частота вибору альтернативних освітніх програм студентами на основі архівних даних вступу

Освітня програма студента

4.2.3. Результати роботи фільтрації на основі вмісту

На основі попередньо оброблених та збережених наборів даних та розроблених моделей інформаційної технології було виконано основні вимоги щодо запропонованої відповідно до поставлених задач досліджень системи.

Зокрема, доповнено технології, що існують сьогодні для вступників НУЛП. Основним ресурсом інформації щодо переліку та опису освітніх програм є Каталог освітніх програм, що забезпечує перелік, опис та можливість пошуку за потрібними параметрами (<https://directory.lpnu.ua/>). Запропонована технологія надає продуктивніші можливості, та забезпечує спрощення для вступників проходження етапів прийняття рішень – збору інформації, проектування, відбору альтернатив. Іншим джерелом є інформація у вигляді файлів на офіційному сайті ЗВО – додатків до правил прийому, що публікуються та оновлюються відповідно до кожної вступної кампанії.

Візуалізація результатів перетину освітніх програм в межах виділених кластерів, а відповідно і можливих рекомендацій подано на Рис. 4.8.

Запропонована інформаційна технологія дає змогу забезпечити поширення інформації щодо менш популярних серед вступників освітніх програм. Такі програми мають менший конкурс, а відповідно більші шанси на вступ. Приклади таких результатів роботи системи на базі даних НУЛП описано нижче. Так, одним з прикладів роботи системи є результати за кластером «Менеджер». Кількість заяв у співвідношенні до кількості виділених місць за державною формою навчання таких освітніх програм у 2023-му році був нижчим ніж його мали інші, більш популярні серед вступників програми. Порівняння цих показників наведено у Таблиці 4.1.

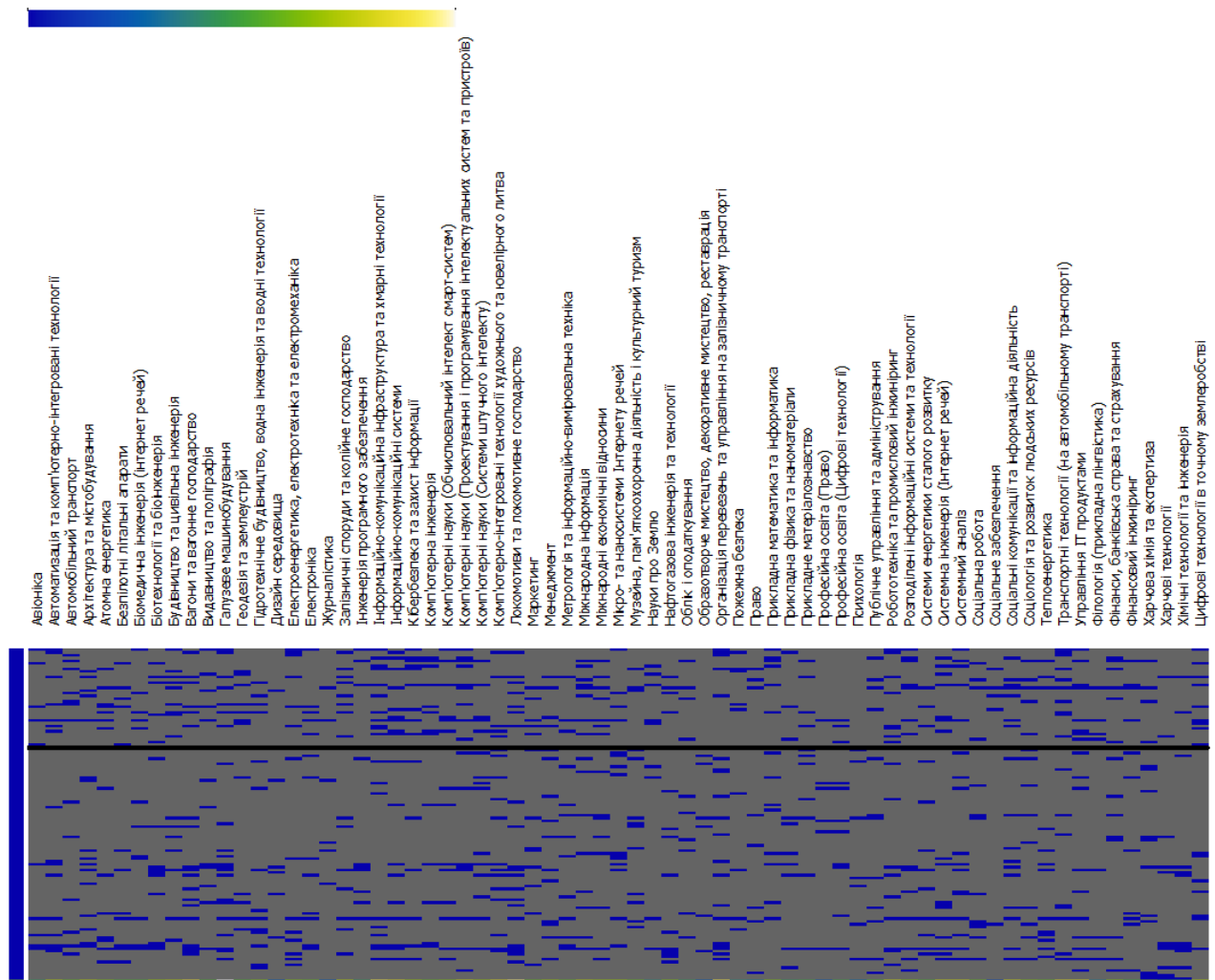


Рис.4.8 Візуалізація результатів перетину освітніх програм в межах виділених кластерів

Таблиця 4.1. Приклад результату роботи фільтрації на основі вмісту

Освітня програма	Частковий вміст описів професійних профілів випускників	Кількість заяв у співвідношенні до кількості виділених місць
«Музейна та архівна справа»	«Це молоді спеціалісти, які володіють широким спектром інформаційних технологій, основами маркетингу та	2.14

	менеджменту в сфері культурної політики, а також двома іноземними мовами. »	
«Соціальні комунікації та інформаційна діяльність»	«...менеджмент інформаційних продуктів та послуг, управління інформацією, PR-технології, керування спільнотами, адміністративний менеджмент, інформаційний консалтинг.»;	2.27
«Організація перевезень та управління на залізничному транспорті»	«...та займати первинні посади: менеджер з транспортно-експедиторської діяльності»	3.12
Менеджмент	«Робочі місця у сфері менеджменту...»	9.54

Також як рекомендації можуть потрапляти освітні програми з однієї галузі та приблизно рівним конкурсом. Проте, такі рекомендації також можна вважати об'єктивними, оскільки вони відповідають потребам користувача та розширюють альтернативу вибору. Прикладами таких є результати за кластером «веб»: Комп'ютерні науки (Системна інженерія Інтернет речей) та Видавництво та поліграфія.

Іншим варіантом використання є перелік програм, які отримує користувач якщо його цікавить робота з даними. У такому випадку перевагою системи є швидкість відбору достатньо великої кількості освітніх програм. Результати подано на Рис. 4.9.

Освітня програма	Ключове слово	Кластер
Інженерія програмного забезпечення	програмного забезпечення	програмного
Інженерія програмного забезпечення	тестування програмного забезпечення	програмного
Комп'ютерна інженерія	програмного забезпечення	програмного
Прикладна математика та інформатика	програмного забезпечення комп'ютерів	програмного
Прикладна математика та інформатика	інженер з програмного забезпечення	програмного
Прикладна математика та інформатика	програмного забезпечення	програмного
Системний аналіз	програмними засобами	програмного
Автоматизація та комп'ютерно-інтегровані технології	програмного забезпечення	програмного
Галузеве машинобудування	основних принципів програмування	програмного
Галузеве машинобудування	комп'ютерних програмних засобів	програмного
Комп'ютерно-інтегровані технології художнього та ювелірного литва	програмне забезпечення	програмного
Комп'ютерні науки (Обчислювальний інтелект смарт-систем)	тестування програмного забезпечення	програмного
Комп'ютерні науки (Проектування і програмування інтелектуальних систем та пристроїв)	програмного забезпечення	програмного
Комп'ютерні науки (Системи штучного інтелекту)	математичного програмного забезпечення	програмного
Комп'ютерні науки (Системи штучного інтелекту)	інформаційного та програмного забезпечення	програмного
Комп'ютерні науки (Системи штучного інтелекту)	прикладне програмне забезпечення	програмного
Видавництво та поліграфія	програмноапаратні інструментальні засоби	програмного
Системна інженерія (Інтернет речей)	програмного забезпечення пристроїв	програмного
Прикладна фізика та наноматеріали	основних мов програмування	програмного
Інформаційно-комунікаційні системи	програмування	програмного
Інформаційно-комунікаційні системи	тестування і впровадження програмного забезпечення	програмного
Інформаційно-комунікаційні системи	Програмні платформи	програмного

Рис.4.9 Перелік освітніх програм за кластером «програмних»

Також перевагою те, що серед освітніх програм, що потрапляють в перелік рекомендацій є ті, які належать до різних галузей, що також покращує рівень ознайомлення вступниками з освітніми програмами ЗВО (Рис. 4.10).

Галузь знань	Освітня програма	Ключове слово	Кластер
Інформаційні технології	Інженерія програмного забезпечення	набору даних	даних
Інформаційні технології	Комп'ютерна інженерія	адміністратор бази даних	даних
Інформаційні технології	Комп'ютерна інженерія	аналітик банку даних	даних
Математика та статистика	Прикладна математика та інформатика	адміністратор бази даних	даних
Математика та статистика	Прикладна математика та інформатика	баз даних	даних
Математика та статистика	Прикладна математика та інформатика	даних	даних
Інформаційні технології	Системний аналіз	оцінювати вихідні дані	даних
Інформаційні технології	Комп'ютерні науки (Обчислювальний інтелект смарт-систем)	аналітика даних	даних
Інформаційні технології	Комп'ютерні науки (Проектування і програмування інтелектуальних систем та пристроїв)	адміністратора баз даних	даних
Інформаційні технології	Комп'ютерні науки (Системи штучного інтелекту)	аналітики даних	даних
Інформаційні технології	Розподілені інформаційні системи та технології	сховищами даних	даних
Інформаційні технології	Системна інженерія (Інтернет речей)	адміністраторів баз даних	даних
Інформаційні технології	Кібербезпека та захист інформації	адміністратор баз даних	даних
Архітектура та будівництво	Геодезія та землеустрій	геопросторових даних	даних
Математика та статистика	Фінансовий інжиніринг	збором, обробкою та аналізом даних	даних
Математика та статистика	Фінансовий інжиніринг	аналізом даних	даних
Міжнародні відносини	Міжнародна інформація	бази даних	даних

Рис.4.10 Перелік освітніх програм за кластером «даних»

Отримані результати забезпечують впорядкування даних й спрощення аналізу можливих шляхів вибору освітньої програми для вступників. Також, сприяють підвищенню компетентності знань вступників щодо цього рішення

засобами інформаційних технологій. Так, можна здійснити вибір ОП на перетині кількох галузей знань. Оскільки можливий той випадок, коли вступник може неправильно асоціювати галузь знань із конкретною освітньою програмою, такий сервіс є доречним.

4.3. Оцінювання результатів роботи двоетапного методу класифікації даних

4.3.1. Результати попереднього опрацювання набору даних для удосконаленого класифікатора

Для оцінювання впливу результатів роботи кожного із методів вибору важливих ознак, які описано у третьому розділі роботи, на точність роботи класифікатора, у цьому підрозділі виконано ряд експериментальних досліджень. Моделювання виконувалося для очищеного набору даних, який містив 15921 спостереження. Дані було розділено на навчальну та тестову вибірки у співвідношенні 80% до 20% відповідно та нормалізувалися із використанням `MaxAbsScaler`. Процедура класифікації із визначеними важливими ознаками кожним із досліджуваних методів виконувалася із використанням алгоритму МОВ з рбф ядром. Оптимальні параметри цього алгоритму визначалися методом пошуку на сітці.

Результати таких експериментальних досліджень на основі показника F1-міра подано в Таблиці 4.2. Слід зазначити, у Таблиці А.1. Додатку А подано значно більше результатів на основі різноманітних метрик для оцінювання ефективності роботи класифікатора при використанні усіх вищевказаних методів вибору важливих ознак у заданому наборі даних.

Таблиця 4.2. Значення F1-міри під час класифікації даних при використанні МОВ з рбф ядром та різних методів вибору важливих ознак із заданого набору даних

<i>Метод</i>	<i>Використовуються усі ознаки</i>	<i>Поріг дисперсії при значенні 0.05</i>	<i>Фільтр кореляції Пірсона</i>	<i>Фільтр взаємної інформації при порозі 0.1</i>	<i>Фільтр Anova F-score при порозі 10</i>
Загальна точність класифікатора у режимі навчання	0,935	0,897	0,876	0,900	0,927
Загальна точність класифікатора у режимі застосування	0,934	0,894	0,845	0,843	0,932
Оптимальна кількість ознак, визначених методом	71	17	15	6	58
Різниця між точностями режимів навчання і застосування	0,001	0,004	0,031	0,057	-0,005

З Таблиці 4.2. та Таблиці А.1. Додатку А можна зробити наступні висновки:

- нелінійний класифікатор на основі МОВ з рбф ядром демонструє задовільну точність класифікації на основі зібраного набору даних при використанні різної кількості ознак;

- найменші точності класифікації отримано при використанні 6 та 15 ознак (замість 71) які були визначені *Фільтром взаємної інформації при порозі 0.1* та *Фільтром кореляції Пірсона відповідно*;
- використання 58 під час класифікації, які визначено *Фільтром Anova F-score при порозі 10*, демонструє високу точність класифікації у режимі застосування. Проте цей метод показує низьку точність у режимі навчання методу, що свідчить про погану генералізацію класифікатора при використанні визначених ознак;
- використання 17 ознак, які визначено із використанням *Порогу дисперсії при значенні 0,05* демонструє високу точність класифікатора, проте вона не є достатньою для практичного використання класифікатора;
- найкращі результати класифікації отримано при використанні усіх 71 ознаки, визначених у цьому дослідженні.

Останній висновок є дуже важливими, оскільки свідчить про вірний відбір незалежних ознак для розв'язання задачі передбачення успішності вступу до ЗВО. Слід зазначити, що результати роботи класифікатора при використанні усіх 71 незалежних ознак показують не лише найвищу точність в режимах навчання і застосування, а й демонструють найвищі генералізаційні властивості класифікатора серед усіх розглянутих у таблиці 4.2 випадків. Це підтверджується найменшим значенням різниці обох цих точностей, серед усіх отриманих результатів. Саме тому, подальші дослідження ґрунтуватимуться на аналізі усіх 71 незалежних ознаки визначених у роботі.

4.3.2. Результати роботи класифікаторів на основі машинного навчання

Моделювання роботи описаних у попередньому розділі методів машинного навчання з різних класів виконувалося із використанням програмного забезпечення на мові Python. Зокрема, реалізацію існуючих класифікаторів на основі машинного навчання взято із бібліотеки [116]. Для каскадного методу з [96] розроблено власну програмну реалізацію.

Експериментальні дослідження виконувалися на зібраному та очищеному наборі даних. Перед виконанням процедури навчання обраного методу, дані нормалізувалися шляхом ділення кожної ознаки окремого стовпця на максимальне значення (по модулю) ознаки в цьому стовпці MaxAbsScaler [117]. Окрім цього, моделювання кожного досліджуваного методу виконувалося із використанням процедури 5-кратної кросвалідації. Такий підхід дозволив отримати результати кожного із досліджуваних методів які є незалежними від того які саме спостереження попали в навчальну вибірку, а які – в тестову. Таким чином, кожен із досліджуваних методів при кожному запуску отримував близько 12736 спостережень для навчання та 3185 спостережень для застосування моделі. Такий процес відбувався 5 разів після чого результати показників ефективності усереднювалися.

Підбір оптимальних параметрів кожного із досліджуваних методів виконувався із використанням методу пошуку на сітці. Оптимальні значення усіх параметрів для усіх досліджуваних методів під час аналізу досліджуваного набору даних наведено у Таблиці Б.1. Додатку Б.

Результати моделювання після виконання 5-кратної кросвалідації на основі показників ефективності (2.1)-(2.4) зведено у Таблиці 4.3.

Таблиця 4.3. Показники ефективності досліджуваних методів у режимі застосування під час розв'язання поставленої у роботі задачі.

<i>Метод/ показник ефективності</i>	<i>Влуч- ність</i>	<i>Повно- та</i>	<i>F1- міра</i>	<i>Коефіцієнт кореляції Метьюса</i>
<i>Лінійні методи</i>				
Класифікатор на основі стохастичного градієнтного спуску	0.895	0.895	0.895	0.779
Класифікатор на основі машини опорних векторів з лінійним ядром	0.899	0.9	0.899	0.79
<i>Нелінійні методи</i>				

Класифікатор на основі Дерева рішень	0.86	0.873	0.865	0.718
Класифікатор на основі машини опорних векторів з сигмоїдним ядром	0.93	0.928	0.929	0.873
Класифікатор на основі машини опорних векторів з поліноміальним ядром	0.908	0.905	0.906	0.825
Класифікатор на основі машини опорних векторів з рбф ядром	0.935	0.934	0.934	0.885
<i>Нейромережеві методи</i>				
Класифікатор на основі багатошарового перцептронну	0.942	0.943	0.942	0.897
Класифікатор на основі ймовірнісної нейронної мережі	0.936	0.952	0.946	0.9
<i>Ансамблеві методи</i>				
Класифікатор на основі алгоритму AdaBoost	0.943	0.944	0.944	0.926
Класифікатор на основі алгоритму Випадкового лісу	0.922	0.921	0.922	0.881
Класифікатор на основі стекінгового об'єднання МОВ з 4 ядрами	0.938	0.94	0.94	0.897
Класифікатор на основі сумісного використання ЙМН та МОВ з рбф ядром (каскад)	0.954	0.953	0.953	0.902

Як видно з Таблиці 4.3. усі методи забезпечують достатньо високу точність класифікації. Це ще раз свідчить про правильний підбір незалежних ознак у

зібраному наборі даних. Проте, найменшу точність як і очікувалося показали лінійні методи машинного навчання. Деяке підвищення точності отримано із використанням нелінійних методів, серед яких чітко виділяється МОВ з різними ядрами. Серед нейромережевих методів найкращі результати показала ЙНМ, яка характеризується високими властивостями до узагальнення. Проте найкращі результати отримано із використанням ансамблевих методів, зокрема каскадного ансамблю з [96]. Незважаючи на це, похибка роботи цього методу у 5% може бути суттєво відчутною під час практичного використання методу під час розв'язання поставленої задачі. Тому виникає необхідність у її зменшенні

4.3.3. Результати роботи удосконаленого двоетапного методу класифікації даних

Для моделювання роботи удосконаленого двоетапного методу класифікації розроблено власне програмне забезпечення на мові Python. Експериментальні дослідження виконувалися із використанням очищеного набору даних. Перед застосуванням, дані нормалізувалися із використанням MaxAbsScaler [117].

В основі роботи першого етапу удосконаленого двоетапного методу класифікації даних покладено використання ЙНМ, єдиним параметром якої є розмах гаусівської функції. Підбір цього параметра відбувався на проміжку [0.1, 1] з кроком 0.1. Оптимальне значення цього параметру для досліджуваного набору даних становило 0.9.

На другому етапі роботи удосконаленого двоетапного методу класифікації даних використовується стекінгове об'єднання чотирьох МОВ з різними ядрами. Для підбору параметрів кожної з них використовувався метод пошуку по сітці. Значення знайдених оптимальних параметрів цих 4 методів наведено в Таблиці Б.2. Додатку Б.

В Таблиці 4.4. наведено результати моделювання як для удосконаленого методу так і для його основних компонентів на основі (3.1)-(3-4) при виконанні 5-кратної кросвалідації.

Таблиця 4.4. Результати роботи удосконаленого методу та його основних компонентів в режимі застосування

<i>Метод/показник ефективності</i>	<i>Влучність</i>	<i>Повнота</i>	<i>F1-міра</i>	<i>Коефіцієнт кореляції Метьюса</i>
Класифікатор на основі сумісного використання ЙМН та машини опорних векторів з лінійним ядром	0.933	0.9	0.931	0.873
Класифікатор на основі сумісного використання ЙМН та машини опорних векторів з сигмоїдним ядром	0.955	0.956	0.954	0.905
Класифікатор на основі сумісного використання ЙМН та машини опорних векторів з поліноміальним ядром	0.942	0.943	0.942	0.879
Класифікатор на основі сумісного використання ЙМН та машини опорних векторів з рбф ядром	0.953	0.954	0.953	0.902
Удосконалений двоетапний метод класифікації даних	0.981	0.984	0.983	0.969

Як видно з результатів поданих у таблиці 4.4, удосконалений метод демонструє суттєве підвищення точності в порівнянні із його окремими складовими (класифікаторами на основі МОВ з різними ядрами)

Окрім цього, якщо порівняти результати у цій та попередній таблицях, то відразу видно, що використання ЙМН як засобу попереднього опрацювання даних забезпечило суттєве підвищення кожної із чотирьох МОВ. Це пояснюється нелінійним розширення простору вхідних даних задачі виходами шару

сумування ЙМН. Проте для більш точної оцінки ефективності удосконаленого методу, у роботі також проведено порівняння точності його роботи з точністю за існуючими методами.

4.3.4. Порівняння ефективності роботи удосконаленого методу та інтерпретація результатів його роботи

Для оцінювання ефективності роботи удосконаленого двоетапного методу класифікації, результати його роботи порівнювалися з такими методами:

- базовими класифікаторами які його формують (ймовірнісна нейронна мережа, машина опорних векторів з чотирьома різними ядрами: рбф, сигмоїдне, поліноміальне та лінійне). Слід зазначити, що базові МОВ в цьому випадку використовують розширений після використання ЙМН набір даних. Саме тому, класифікатор на основі сумісного використання ймовірнісної мережі та машини опорних векторів з рбф ядром являється також і базовим методом з [85] для порівняння;
- існуючим класифікатором на основі стекінгового об'єднання машини опорних векторів з чотирьома різними ядрами: рбф, сигмоїдне, поліноміальне та лінійне.

За основу для порівняння вищевказаних методів, серед метрик (3.1)-(3.4) обрано F1- міру, оскільки її найчастіше використовують в таких цілях. Результати порівняння зведено на Рис.4.11

Як видно з Рис. 4.11, класифікатор на основі сумісного використання ЙМН та МОВ з лінійним ядром показує найменшу точність під час розв'язання поставленої задачі. Це пояснюється використанням власне лінійного ядра у методі МОВ, яке не завжди забезпечує хороші результати. Дещо вищу точність показав класифікатор на основі стекінгового об'єднання МОВ з чотирьома різними ядрами. Оскільки він працював з початковим набором даних, тобто не розширеним ймовірностями належності до обох класів задачі, результат його роботи є задовільним.

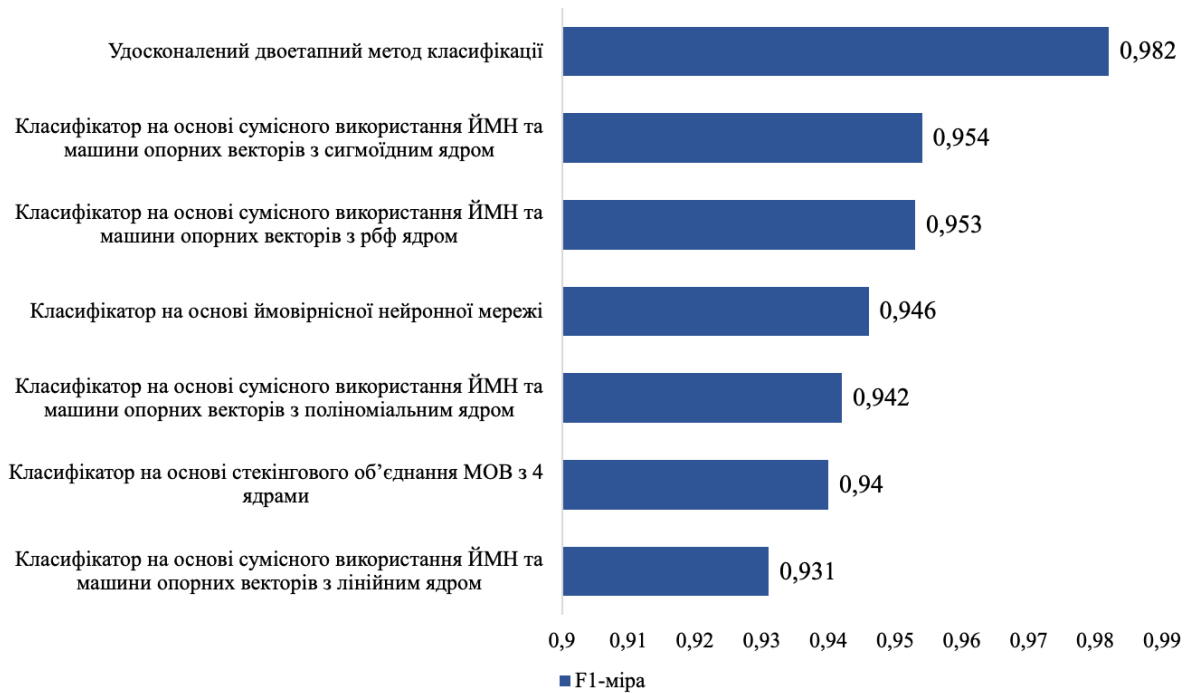


Рис.4.11 Порівняння ефективності роботи досліджуваних методів на основі F1-міри

Незначене покращення обох вищеописаних методів вдалося отримати із використанням класифікатора на основі сумісного використання ЙМН та МОВ з поліноміальним ядром та класифікатора на основі використання МОВ. Далі, по списку зростання точності, йдуть два методи: базовий метод який удосконалювався у цій роботі (класифікатор на основі сумісного використання ЙМН та МОВ з рбф ядром) та такий же класифікатор лише з сигмоїдним ядром. Слід зауважити, що останній показав дещо кращий результат в порівнянні з базовим.

Суттєве покращення точності показав удосконалений у цій роботі метод. Зокрема, використання цього методу на основі F1-міри показує на 2.9% кращий результат за базовий метод та на 4.2% вищу точність в порівнянні із існуючим класифікатором на основі стекінгового об'єднання МОВ з чотирьома різними ядрами.

Для інтерпретації отриманих результатів, розглянемо розмір тестового набору та отримані відсоткові результати приросту точності класифікації більш

детально. Тестовий набір даних для якого обчислювалися значення F1-міри при використанні 5-кратної кросвалідації містить близько 3185 спостережень при кожному запуску. Якщо припустити що алгоритм машинного навчання абсолютно правильно класифікує усі ці спостереження, то значення F1-міри мало б бути рівне 1 (тобто 100%). Враховуючи це, 1% правильно класифікованих зразків це 32 спостереження. Удосконалений метод демонструє на 2.9% кращий результат в точності в порівнянні з точністю за базовим методом. Це означає, щодо його використання супроводжується правильною класифікацією для ще близько 92 додаткових спостережень. Оскільки під одним спостереженням розуміється інформація про одного вступника, удосконалений у цій роботі метод класифікації даних забезпечує коректне передбачення успішності вступу до ЗНО ще для 92 вступників (в порівнянні із використанням базового методу класифікації). Враховуючи це, а також те, що похибка роботи методу на основі F1-міри становить всього 1,8%, удосконалений метод варто використати для реалізації інтелектуалізованої компоненти передбачення успішності вступу вступника ЗВО.

Висновки до розділу

У четвертому розділі дисертаційної роботи розроблено архітектуру інформаційної системи підтримки прийняття рішень вступників ЗВО та виконано оцінювання результатів моделювання її компонентів. Серед вагомих отриманих результатів слід виділити наступні:

- розроблено архітектуру інформаційної системи підтримки прийняття рішень, в основі якої покладено застосування агентів попереднього опрацювання даних та використання компонент надання рекомендацій і передбачення успішності вступу до ЗВО України;
- розроблена технологія сумісного використання фільтрації на основі змісту та колаборативної фільтрації забезпечує надання розширених рекомендацій вступнику, що підвищує рівень його обізнаності щодо

пропонованих ЗВО освітніх програм та може сприяти вибору однієї із них для навчання у ЗВО;

- виконано експериментальне порівняння ефективності роботи існуючого інструментарію машинного навчання, що продемонструвало недостатню точність його роботи та зумовило необхідність у вдосконаленні існуючого двоетапного методу машинного навчання для розв'язання поставленої задачі;

- проведено моделювання роботи удосконаленого двоетапного методу класифікації даних і шляхом порівняння з рядом існуючих методів з різних класів встановлено підвищення точності його роботи на основі F1-міри на 2.9% під час розв'язання задачі передбачення успішності вступу вступника ЗВО;

- на основі низької похибки роботи удосконаленого двоетапного методу класифікації даних у 1,8% (на основі F1-міри), обґрунтовано можливість його інтеграції у склад великих інформаційних систем підтримки та супроводу вступної кампанії як інтелектуалізованої компоненти передбачення успішності вступу вступника ЗВО.

Висновки

У дисертаційній роботі розв'язано актуальну наукову задачу підвищення рівня обізнаності вступників ЗВО України щодо вибору освітніх програм відповідно до їхніх інтересів на основі побудови нових і вдосконалення існуючих моделей, методів та інформаційних технологій підтримки прийняття рішень вибору освітніх програм з урахуванням особливостей процесу вступу до ЗВО України. При цьому отримано такі основні науково-практичні результати:

1. Визначено особливості та принципи процесу прийняття рішень вступників ЗВО в різні періоди вступної кампанії та проаналізовано сучасний рівень забезпечення інформаційних технологій для вступників ЗВО, що дало змогу виокремити прогалини у подібних системах вітчизняних та іноземних ЗВО.

2. Вперше розроблено модель інформаційної системи підтримки прийняття рішення вступників ЗВО, яка, на відміну від вже існуючих, використовує поєднання методу надання рекомендацій та вдосконаленого методу передбачення успішності вступу, що підвищує рівень обізнаності вступників ЗВО України щодо вибору освітніх програм відповідно до їхніх інтересів.

3. Розроблено технологію збору та попереднього опрацювання даних, яка уможливила подальше проведення інтелектуального аналізу на їхньому фундаменті. На основі використання чотирьох методів визначення важливих для аналізу атрибутів встановлено, що відбір 71 незалежних ознак для розв'язання задачі передбачення успішності вступу вступника ЗВО є коректним, і забезпечує найвищу точність аналізу таких даних методами машинного навчання.

4. Набуло подальшого розвитку використання технології надання рекомендацій вступнику на основі сумісного застосування колаборативної фільтрації із використанням архівних даних вступної кампанії та освітніх даних, а також фільтрації на основі вмісту з використанням ключових слів і кластерів

освітніх програм, що уможливило розширення переліку альтернативних освітніх програм, які відповідають інтересам вступника.

5. Удосконалено двоетапний метод класифікації даних за рахунок заміни класифікатора на основі машини опорних векторів на другому етапі методу на стекінгове об'єднання чотирьох класифікаторів на основі машини опорних векторів із різними ядрами й агрегуванням результатів їхньої роботи логістичною регресією, що дало змогу підвищити точність на 2,9 % на тестових даних під час розв'язання задачі передбачення успішності вступу до ЗВО.

6. Розроблено архітектуру інформаційної системи підтримки прийняття рішень, в основу якої покладено застосування агентів попереднього опрацювання даних та використання компонент надання рекомендацій і передбачення успішності вступу до ЗВО України.

Література

1. Порядок прийому на навчання для здобуття вищої освіти в 2023 році : наказ Міністерства освіти і науки України від 27 квітня 2022 року № 392. Дата оновлення : 02.05.2022. URL: <https://zakon.rada.gov.ua/laws/show/z0487-22#Text> (дата звернення: 09.10.2023)
2. Про вищу освіту: Закон України від 01.07.2014 № 1556-VII/21 № 3482-IX. Дата оновлення: 24.03.2024. URL: <https://zakon.rada.gov.ua/laws/show/1556-18#Text> (дата звернення: 25.03.2024)
3. Шілінг А. Ю., Жежнич П. І. Оцінювання якості планування надання освітніх послуг закладам вищої освіти. Стандартизація. Сертифікація. Якість. 2019. № 2. с. 61-72. URL: http://nbuv.gov.ua/UJRN/ssia_2019_2_7
4. Зуб Х. В., Жежнич П. І. Аналіз ефективності вступної кампанії закладів вищої освіти України та способів її підвищення шляхом впровадження інформаційних технологій. Вісник Вінницького Політехнічного Інституту, 2022. № 3. С. 52–59. DOI: <https://DOI.org/10.31891/csit-2023-1-11>
5. Каричковський В. Д. Аналіз відносних показників вступних кампаній у вищих аграрних навчальних закладах України. Педагогіка формування творчої особистості у вищій і загальноосвітній школах, 2017. № 53 (106). с. 467-474. URL: <http://www.pedagogy-journal.kpu.zp.ua/archive/2017/53/60.pdf>
6. Бідюк П.І., Тимощук О.Л., Коваленко А.Є. Коршевніук Л.О. Системи і методи підтримки прийняття рішень : навч. посіб. для здобувачів ступеня магістра за освітніми програмами «Системний аналіз та управління», «Системний аналіз фінансового ринку» спеціальності 124 «Системний аналіз», Київ : КПІ ім. Ігоря Сікорського, 2020. 259 с. URL: <https://ela.kpi.ua/server/api/core/bitstreams/73de1d73-bd75-41b3-b775-a5bcfa95a454/content> (дата звернення: 09.09.2023)
7. Пономарьова Н.О. Відбір абітурієнтів на ІТ-спеціальності в Україні: стан і проблеми. Наукові записки. Серія: проблеми методики фізико-математичної і технологічної освіти, 2016. Т.10, № 3. С.131-137. URL:

- https://cusu.edu.ua/images/conf-2016-10/s5/%D0%9F%D0%BE%D0%BD%D0%BE%D0%BC%D0%B0%D1%80%D1%8C%D0%BE%D0%B2%D0%B0_%D1%81%D1%82%D0%B0%D1%82%D1%82%D1%8F.pdf
8. Нестеренко О.В., Савенков О.І., Фаловський О.О. Інтелектуальні системи підтримки прийняття рішень: Навч. посіб. За ред. Бідюка П.І. Київ : Національна академія управління, 2016. 188 с.
 9. Коломієць М. Б., Мирний Р. Ф. Вступна кампанія закладу вищої освіти як система. Вісник Глухівського національного педагогічного університету імені Олександра Довженка. Педагогічні науки, 2017. Вип. 3. С. 105-111. URL: http://nbuv.gov.ua/UJRN/vgnpu_2017_3_16 (дата звернення: 12.10.2023)
 10. Про затвердження плану заходів з реалізації Концепції державної системи професійної орієнтації населення : Розпорядження Кабінету Міністрів України від 04.07.2018 р. № 469-р : станом на 21 жовт. 2020 р. URL: <https://zakon.rada.gov.ua/laws/show/469-2018-p#Text> (дата звернення: 09.12.2023)
 11. Zhezhnych, P., Berezko, O., Zub, K., & Demydov, I. Analysis of Features and Abilities of Online Systems and Tools Meeting Information Needs of HEIs' Entrants. International Workshop on Control, Optimisation and Analytical Processing of Social Networks, 2020. Vol. 2616. P. 76–85. URL: <https://ceur-ws.org/Vol-2616/paper7.pdf> (Last accessed: 02.03.2024)
 12. World. Ranking Web of Universities: Webometrics ranks 30000 institutions : веб-сайт. URL: <https://www.webometrics.info/en/world> (Last access: 09.02.2024)
 13. Консолідований рейтинг вишів України 2023 року. Освіта.UA : веб-сайт. URL: https://osvita.ua/vnz/rating/51741/#google_vignette (дата звернення: 09.02.2024)
 14. Зуб Х. Зуб Х. Аналіз освітніх веб-ресурсів задоволення інформаційних потреб абітурієнта ЗВО України. Інформація, комунікація, суспільство

- 2021: матеріали 10-ої Міжнародної наукової конференції ІКС-2021, 20–22 травня 2021 року, Україна, Львів : Видавництво Львівської політехніки, 2021. С. 15–16. URL: http://ics.skid-lp.info/ics_2021.pdf
15. Бідюк П.І., Тимошук О.Л., Коваленко А.Є. Коршевнюк Л.О. Системи і методи підтримки прийняття рішень : підручник, Київ : КПІ ім. Ігоря Сікорського, 2022. 610 с. URL: <https://ela.kpi.ua/server/api/core/bitstreams/6958f683-fbac-4506-9c85-5115c8f8b4c6/content> (дата звернення: 12.06.2023)
 16. Токар В., Дубиківський С., Палагута К., Самойленко Ю., Пашорін В. Архітектура інтелектуальних систем прийняття рішень інформаційної інфраструктури ЗВО з урахуванням вимог ЄС. Інформаційні технології та суспільство, 2023. Вип. 1 (7). с. 80-87. URL: <https://dspace.nuft.edu.ua/server/api/core/bitstreams/70b8527f-6c81-4070-a093-ef47d7b769f0/content> DOI: <https://DOI.org/10.32689/maup.it.2023.1.11>
 17. Інформаційно-аналітична система підтримки освітньої діяльності структурних підрозділів закладів вищої освіти / Ю. В. Триус та ін. Вісник Черкаського державного технологічного університету. 2021. № 4. С. 27–38. URL: https://er.chdtu.edu.ua/bitstream/ChSTU/1882/3/27-38_%D0%A2%D1%80%D0%B8%D1%83%D1%81.pdf DOI: 10.24025/2306-4412.4.2020.219482 (дата звернення: 10.05.2023)
 18. Novorushchenko T., Izonin I., Kutucu H. Advancements in AI-Based Information Technologies: Solutions for Quality and Security. Systems. 2024. Vol. 12, no. 2. P. 58. URL: <https://doi.org/10.3390/systems12020058>
 19. Бідюк П.І., Коршевнюк Л.О. Проектування комп'ютерних інформаційних систем підтримки прийняття рішень : навчальний посібник, Київ : ННК «ІПСА» НТУУ «КПІ», 2010. 340 с. URL: <http://mmsa.kpi.ua/sites/default/files/publications/%D0%91%D1%96%D0%B4%D1%8E%D0%BA%20%D0%9F%D0%B5%D1%82%D1%80%D0%BE%20%D0%86%D0%B2%D0%B0%D0%BD%D0%BE%D0%B2%D0%B8%D1%8>

- 7/bidyuk-p-i-korshevnyuk-l-o-proektuvannya-kompyuternih-informacii-nih-sistem-pidtrimki-priinyattya-rishen-navchalnii-posibnik.pdf (дата звернення: 11.08.2023)
20. Pankratova N., Bidyuk P., Golinko I. Decision support system for microclimate control at large industrial enterprises. *Computer Modeling and Intelligent Systems : CEUR Workshop Proceedings*. 2020. Vol. 2608. pp. 489–498. URL: <https://ceur-ws.org/Vol-2608/paper37.pdf>
 21. Bidiuk P., Prosyankina-Zharova T., Diakon V., Diakon D. The improvement of the intelligent decision support system for forecasting non-linear non-stationary processes. *Technology audit and production reserves*. 2023. Vol. 4, no. 2(72). P. 37–46. URL: <https://doi.org/10.15587/2706-5448.2023.286516>
 22. Hovorushchenko T., Hnatchuk Y., Hnatchuk A., Fehyr O. Decision Support System For The Coaching Staff in the Process of Preliminary Selection of Players. *2023 IEEE 18th International Conference on Computer Science and Information Technologies (CSIT), Lviv, Ukraine, 19–21 October 2023*. 2023. P. 1-4 URL: <https://doi.org/10.1109/csit61576.2023.10324119>
 23. Hnatchuk Y., Hovorushchenko T., Pavlova O. Methodology for the development and application of clinical decisions support information technologies with consideration of civil-legal grounds. *Radioelectronic and Computer Systems*. 2023. No. 1. P. 33–44. URL: <https://doi.org/10.32620/reks.2023.1.03>
 24. Hnatchuk Y., Hovorushchenko T., Drapak G., Kysil T. Technology of Decision-Making Support Regarding the Possibility of Donation and Transplantation Considering Civil Law. *International journal of computer science and network security : IJCSNS*, 2022. Vol. 22, no. 9. P. 307-315. URL: <https://doi.org/10.22937/ijcsns.2022.22.9.41>
 25. Decision support system for assessing the economic development potential of a territorial community / T. Tereshchenko et al. *CEUR Workshop Proceedings*. 2023, Volume 3675, P. 100–117. URL: <https://ceur-ws.org/Vol-3675/paper8.pdf>

26. Довбиш А., Васильєв А., Любчак В. Інтелектуальні інформаційні технології в електронному навчанні : монографія. Суми : СумДУ, 2013. 177 с.
27. Бережна С. В. Соціальні мережі як інструмент профорієнтаційної роботи з абітурієнтами ЗВО. Абітурієнтське середовище періоду глокалізації: шляхи формування та тенденції розвитку : матеріали XIX Міжнар. наук.-практ. конф., Харків, 12 лют. 2021 р. Харків: НУА, 2021. С. 41–44.
28. Жарська, І. О. Принципи організації маркетингових комунікацій ЗВО з абітурієнтами у мережі інтернет. Науковий вісник Херсонського державного університету. Серія: Економічні науки. Херсон, 2020. Вип. 37. С. 34-39. URL: http://nbuv.gov.ua/UJRN/Nvkhdu_en_2020_37_8
29. Жегус, О. В. Формування стратегії маркетингу в соціальних мережах закладу вищої освіти. Маркетинг і цифрові технології, 2018. Т. 2, № 2. С. 58-75. URL: http://nbuv.gov.ua/UJRN/mardigt_2018_2_2_8
30. Зінчук Н. Маркетингові комунікації як фактор успіху навчального закладу на ринку освітніх послуг. Збірник матеріалів Всеукраїнської науково-практичної конференції «Наукові підходи в управлінні навчальними закладами». с. 145-151. URL: <http://eprints.zu.edu.ua/17438/1/%D0%97%D1%96%D0%BD%D1%87%D1%83%D0%BA.pdf> (дата звернення: 15.02.2020)
31. Zhezhnych P., Shilinh A., Melnyk V. Linguistic Analysis of User Motivations of Information Content for University Entrant's Web-forum. International Journal of Computing, 2019. P. 67–74. DOI: <https://DOI.org/10.47839/ijc.18.1.1275>
32. Шілінг А. Ю., Жежнич П. І. Модель поведінки закладу вищої освіти у процесі планування надання освітніх послуг на основі лінгвістичного аналізу комунікативної активності в освітніх веб-спільнотах. Вісник Хмельницького національного університету. Технічні науки, 2019. № 2. С. 145-149. URL: http://nbuv.gov.ua/UJRN/Vchnu_tekh_2019_2_29
33. Шілінг А., Жежнич П. Побудова системи прогнозування контингенту студентів ЗВО на основі вибору пріоритету. X Науково-практична

- конференція «Інноваційні комп'ютерні технології у вищій школі», 21-23 листопада 2018 р. Львів, 2018. С. 178-182.
34. Сайт Приймальної комісії КПІ ім. Ігоря Сікорського як основне джерело інформування вступників / В. М. Можаровський та ін. Вступна кампанія до закладів вищої освіти України: проблеми та перспективи 2020 : Збірник матеріалів II Всеукраїнської науково-практичної конференції. Київ, 2019. С. 12-14. URL: <https://openarchive.nure.ua/server/api/core/bitstreams/8a87d994-f510-4f25-8ccf-d68f8f72a090/content>
 35. Особливості підходів до організації сайтів структурних підрозділів та їхнє значення в профорієнтаційній роботі в умовах інформаційного суспільства / В.М. Можаровський та ін. Проблеми та перспективи : 3-а Всеукраїнська науково-практична конференція. Київ, 2020. с 8-11; URL: https://pk.kpi.ua/wp-content/uploads/2020/07/Zbirnyk_2020.pdf
 36. Карашецький В., Яркун, В. Автоматизована система моніторингу вступної кампанії закладу вищої освіти. Інформаційні технології та комп'ютерна інженерія, 2021. Вип. 51, №. 2. С. 12–16. URL: <https://DOI.org/10.31649/1999-9941-2021-51-2-12-16> (дата звернення: 10.07.2023).
 37. Макоедова В. Інформатизація процесів вступної кампанії в закладах вищої освіти. Технічні науки та технології, 2023. № 1(31). С. 90–97. URL: <http://tst.stu.cn.ua/article/view/278890/273541> (дата звернення: 1.07.2023).
 38. Красильникова, Г., Красильников, С. Використання інформаційно-комунікаційних технологій для профорієнтаційної роботи з абітурієнтами ЗВО. Наукові записки ВДПУ імені Михайла Коцюбинського. Серія: педагогіка і психологія, 2022. Вип. 69. С. 73-79 URL: <http://dspace.vspu.edu.ua/handle/123456789/10462>
 39. Осадчий В. В., Круглик В. С., Букреєв Д. О. Розробка програмного засобу для прогнозування вступу абітурієнтів до закладів вищої освіти. Ukrainian Journal of Educational Studies and Information Technology, 2018. Т. 6 № 3. с. 55-69. URL: <http://ojs.mdpu.org.ua/index.php/itse>

40. Осадчий В. В., Круглик В. С., Осадча К. П., Сердюк І. М., Букреев Д. О. Особливості розробки програмного засобу для прогнозування вступу абітурієнтів до закладів вищої освіти. Вчені записки Таврійського національного університету імені В.І. Вернадського. Серія Технічні науки, 2019. Т.30 (69). №1. С.110-114
41. Зуб Х. В., Жежнич П.І. Огляд сучасного стану систем підтримки прийняття рішень абітурієнтів закладів вищої освіти. Міжнародний науково-технічний журнал «Інформаційні технології та комп'ютерна інженерія», 2022. № 53(1), С. 28–36. URL: <https://DOI.org/10.31649/1999-9941-2022-53-1-28-36>.
42. Sridhar S., Mootha S., Kolagati S. A University Admission Prediction System using Stacked Ensemble Learning». Advanced Computing and Communication Technologies for High Performance Applications (ACCTHPA), India. 2020. P. 162–167. URL: <https://doi.org/10.1109/ACCTHPA49271.2020.9213205>
43. Chakrabarty, N., Chowdhury, S., Rana, S. A Statistical Approach to Graduate Admissions' Chance Prediction. Innovations in Computer Science and Engineering. 2020. Vol. 103, P. 333–340. URL: http://dx.doi.org/10.1007/978-981-15-2043-3_38
44. Singhal S., Sharma iA. Prediction of Admission Process for Gradational Studies using AI Algorithm. European Journal of Molecular & Clinical Medicine (EJMCM). 2020. Vol. 7, no 4. P. 116–120. URL: <https://ejmcm.com/uploads/paper/60fdddde94c92f76875987ceda3c3899.pdf> (date of access: 06.05.2020)
45. Prediction for University Admission using Machine Learning / Chithra Apoorva D.A. etc. International Journal of Recent Technology and Engineering (IJRTE). 2020. Vol. 8, no. 6. P. 4922–4926. URL: <https://www.ijrte.org/portfolio-item/f9043038620>
46. Khan M. A., Dixit M., Dixit A. Demystifying and Anticipating Graduate School Admissions using Machine Learning Algorithms. IEEE 9th International Conference on Communication Systems and Network Technologies (CSNT)

- (Gwalior, India, 10-12 April 2020). India, 2020. P. 19–25. URL: <https://ieeexplore.ieee.org/document/9115788>
47. Qamhie M., Sammaneh H., Demaidi M. N. PCRS: Personalized Career-Path Recommender System for Engineering Students. *IEEE Access*. 2020. Vol. 8. P. 214039–214049. URL: <https://doi.org/10.1109/access.2020.3040338>
 48. Question Answering System to Support University Students' Orientation, Recruitment and Retention / W. A. Elnozahy et al. *Procedia Computer Science*. 2019. Vol. 164. P. 56–63. URL: <https://doi.org/10.1016/j.procs.2019.12.154>
 49. Alghamdi S., Alzhrani N., Algethami H. Fuzzy-Based Recommendation System for University Major Selection. *Proceedings of the 11th International Joint Conference on Computational Intelligence*. Vienna, Austria, 17–19 September 2019. Vienna. 2019. P. 317–324. URL: <https://doi.org/10.5220/0008071803170324>
 50. Dhanashri, D. J. College Recommendation System For Admission. *International Research Journal of Engineering and Technology (IRJET)*, 2018. Vol. 5, no. 3. P. 1269–1272.
 51. Ahlawat, R., Sahay, S., Sabitha, S., & Bansal, A. Analysis of factors affecting enrollment pattern in Indian universities using k-means clustering. *International Conference on Information Technology (InCITE)-The Next Generation IT Summit on the Theme -Internet of Things: Connect your Worlds*. Noida, 2016. P. 321–326. DOI: 10.1109/INCITE.2016.7857639
 52. Farid Shamsudin, M., Mohd Ali, A., Ab Wahid, R., Saidun, Z. Factors Influence Undergraduate Students' Decision Making To Enroll And Social Media Application As An External Factor. *Humanities and Social Sciences Review*, 2019. Vol. 7, no. 1. P. 126–136. DOI: 10.18510/hssr.2019.7116.
 53. Patel H.I. Assessment of Affecting Factors for Higher Education Admission Process. *International Journal of Engineering and Advanced Technology*, 2019. Vol. 9, no. 1. P. 63–67. DOI: 10.35940/ijeat.A1042.109119.

54. Devarapalli D. J. Classification Method to Predict Chances of Students' Admission in a Particular College. Proceedings of International Conference on Recent Trends in Machine Learning, IoT, Smart Cities and Applications, 2017. Vol. 1245. P. 225–238. DOI: 10.1007/978-981-15-7234-0_19.
55. Gupta N., Sawhney, A., Roth, D. Will I Get in? Modeling the Graduate Admission Process for American Universities. Proceedings of IEEE 16th International Conference on Data Mining Workshops (ICDMW), Barcelona, Spain, 2016. P. 631–638. DOI: 10.1109/ICDMW.2016.0095.
56. Wu B., Ke Z., Fu M., Xia Y. SOUA: Towards Intelligent Recommendation for Applying for Overseas Universities. International Conference on Intelligent Computing, Automation and Systems (ICICAS). China, 2019. P. 124–128. DOI: 10.1109/ICICAS48597.2019.00033
57. Park T., Kim C. Predicting the Variables That Determine University (Re-)Entrance as a Career Development Using Support Vector Machines with Recursive Feature Elimination: The Case of South Korea. Sustainability, 2020. Vol. 12, no 18. P. 7365. DOI: 10.3390/su12187365.
58. Protikuzzaman, Md., Kanti, M., Kumar, M., & Chandra, B. Predicting Undergraduate Admission: A Case Study in Bangabandhu Sheikh Mujibur Rahman Science and Technology University, Bangladesh. International Journal of Advanced Computer Science and Applications (IJACSA). 2020. Vol. 11, no 12. DOI: 10.14569/IJACSA.2020.0111217.
59. Aarthi, S., Sarvathanayan, M., Kumar, B. P. Post-Graduate College Admission Recommender Using Data Analytics. International Journal of Innovative Technology and Exploring Engineering. 2019. Vol. 8, no. 6.
60. Li, C., Ma, Z., Zhang, H., Liu, Y. The Prediction Model For College Admission Score Based On Support Vector Machine. ICIC Express Letters, Part B: Applications An International Journal of Research and Surveys. 2017. Vol. 8. P. 889–893.

61. Acharya, M. S., Armaan, A., Antony, A. S. A Comparison of Regression Models for Prediction of Graduate Admissions. International Conference on Computational Intelligence in Data Science (ICCIDS). 2019. P. 1-5. DOI: 10.1109/ICCIDS.2019.8862140
62. Yazdipour S., Taherian N. Data Driven Decision Support to Fund Graduate Studies in Abroad Universities. International Conference on Machine Learning and Data Science (MLDS). Noida, 2017. P. 44-50. DOI: 10.1109/MLDS.2017.17.
63. Subba Reddy Y., Govindarajulu P. College Recommender system using student' preferences/voting: A system development with empirical study. IJCSNS International Journal of Computer Science and Network Security. 2018. Vol. 18, no 1.
64. Khairina D. M., Ramadhani F., Maharani S., Hatta H. R. Department recommendations for prospective students Vocational High School of information technology with Naïve Bayes method. 2nd International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE). Indonesia, 2015. P. 92–96. DOI: 10.1109/ICITACEE.2015.7437777.
65. Baskota A., Ng Y.-K. A Graduate School Recommendation System Using the Multi-Class Support Vector Machine and KNN Approaches. IEEE International Conference on Information Reuse and Integration (IRI). Salt Lake City, UT, 2018. P. 277–284. DOI: 10.1109/IRI.2018.00050.
66. Sharma V., Trehan T., Chanana R., Dawn S. StudieMe: College Recommendation System. 3rd International Conference on Recent Developments in Control, Automation & Power Engineering (RDCAPE)., India, 2019. P. 227–232. DOI: 10.1109/RDCAPE47089.2019.8979030
67. Comparison of Conventional Statistical Methods with Machine Learning in Medicine: Diagnosis, Drug Development, and Treatment / H. S. R. Rajula et al. *Medicina*. 2020. Vol. 56, no. 9. P. 455. URL: <https://doi.org/10.3390/medicina56090455> (date of access: 12.01.2024).

68. The Architecture of Mobile Information System for Providing Safety Recommendations During the Trip / V. Savchuk et al. *Advances in Computer Science for Engineering and Education II*. Cham, 2019. Vol. 938. P. 493–502. URL: https://doi.org/10.1007/978-3-030-16621-2_46
69. An Intelligent System for Generating End-User Symptom Recommendations Based on Machine Learning Technology / S. Makara et al. *CEUR Workshop Proceedings*. 2020. Vol. 2604. P. 844-883. URL: <https://ceur-ws.org/Vol-2604/paper57.pdf>
70. Баран М., Висоцька В., Голощук Р. Програмна реалізація інтелектуальної системи для вирішення проблеми «Холодного старту». *Вісник Національного університету «Львівська політехніка» «Інформаційні системи та мережі»*. Львів : Видавництво Львівської політехніки, 2023. Вип. 13. С. 274-299. URL: <https://doi.org/10.23939/sisn2023.13.274>
71. Fedushko S., Ustyianovych T., Syerov Y. Intelligent Academic Specialties Selection in Higher Education for Ukrainian Entrants: A Recommendation System. *Journal of Intelligence*. 2022. Vol. 10, no. 2. P. 32. URL: <https://doi.org/10.3390/jintelligence10020032>
72. Зуб Х. Рекомендаційна система для абітурієнтів закладів вищої освіти з використанням освітніх даних. *Інформація, комунікація, суспільство 2023 : матеріали 12-ї Міжнародної наукової конференції ICS-2023, 18-23 травня 2023, Львів : Видавництво Львівської політехніки, 2023. С. 172–173.*
73. Fedushko S., Molodetska K., Syerov Y. Analytical method to improve the decision-making criteria approach in managing digital social channels. *Heliyon*. 2023. P. e16828. URL: <https://doi.org/10.1016/j.heliyon.2023.e16828>
74. Fedushko S., Molodetska K., Syerov Y. Decision-making approaches in the antagonistic digital communication of the online communities users. *Social Network Analysis and Mining*. 2023. Vol. 13, no. 1. URL: <https://doi.org/10.1007/s13278-022-01021-4>

75. Roy D., Dutta M. A systematic review and research perspective on recommender systems. *Journal of Big Data*. 2022. Vol. 9, no. 1. URL: <https://doi.org/10.1186/s40537-022-00592-5>.
76. Зуб Х. Рекомендаційні інформаційні системи як засіб підтримки прийняття рішень вступників ЗВО України. *Інформація, комунікація, суспільство 2022 : матеріали 11-ї Міжнародної наукової конференції ICS-2022, 19-21 травня 2022, Україна, Львів : Видавництво Львівської політехніки, 2022. С. 49–50.* URL: http://skid.lpnu.ua/wp-content/uploads/2022/05/ICS2022_Proceedings.pdf
77. Classification System Based on Ensemble Methods for Solving Machine Learning Tasks / P. Bidyuk et al. *CEUR Workshop Proceedings*. 2023, Vol. 3426, pp. 1–11. URL: <https://ceur-ws.org/Vol-3426/paper1.pdf>
78. Glossary: Definition of Enterprace AI and data Science Terms. C3 AI. URL: <https://c3.ai/glossary/data-science/recall/> (date of access: 14.01.2023).
79. Zhezhnych P., Zub K., Berezko O., Shilinh A. A Comparison of Machine Learning Algorithms for Prediction Higher Education Institution's Entrants Admissions. *Advances in Artificial Systems for Logistics Engineering / Z. Hu, Q. Zhang, S. Petoukhov, and M. He, Eds., in Lecture Notes on Data Engineering and Communications Technologies*. Cham : Springer International Publishing, 2021. Vol. 82. P. 171–179. DOI: 10.1007/978-3-030-80475-6_17.
80. Zub K. Application of support vector machine method for prediction HEI entrants' admission success. 10th International Youth Science Forum 'Litteris et Artibus'. 2021. URL: <https://nauka.international/lea-2021/application-support-vector-machine-method-prediction-hei-entrants-admission-success>
81. SGD-Based Cascade Scheme for Higher Degrees Wiener Polynomial Approximation of Large Biomedical Datasets / I. Izonin et al. *Machine Learning and Knowledge Extraction*. 2022. Vol. 4, no. 4. P. 1088–1106. URL: <https://doi.org/10.3390/make4040055>

82. A SVM Regression Based Approach to Filling in Missing Values / F. Honghai et al. Lecture Notes in Computer Science. Berlin, Heidelberg, 2005. P. 581–587. URL: https://doi.org/10.1007/11553939_83
83. The Combined Use of the Wiener Polynomial and SVM for Material Classification Task in Medical Implants Production / I. Izonin et al. International Journal of Intelligent Systems and Applications. 2018. Vol. 10, no. 9. P. 40–47. URL: <https://doi.org/10.5815/ijisa.2018.09.05>
84. Neural Network Approach for Semantic Coding of Words / V. Golovko et al. Advances in Intelligent Systems and Computing. Cham, 2019. C. 647–658. URL: https://doi.org/10.1007/978-3-030-26474-1_45
85. Kit I., Lipyanina-Goncharenko H., Lendyuk T., Sachenko A., Komar M. Neural Network Method of Items Catalog Forming for Online Store.) Advances in Artificial Systems for Logistics Engineering. ICAILE 2022. Lecture Notes on Data Engineering and Communications Technologies. Cham : Springer, 2022. Vol. 135. P. 157-169 DOI: 10.1007/978-3-031-04809-8_14
86. Deep Multilayer Neural Network for Predicting the Winner of Football Matches / S. Anfilets et al. International Journal of Computing. 2020. Vol. 19, no. 1. P. 70–77. URL: <https://doi.org/10.31891/1727-6209/2020/19/1-70-77>
87. Ratnovsky A., Rozenes S., Bloch E., Halpern P. Statistical learning methodologies and admission prediction in an emergency department. Australasian Emergency Care. 2021. URL: <https://doi.org/10.1016/j.auec.2020.11.004>
88. Izonin I., Tkachenko R., Greguš M. I-PNN: An Improved Probabilistic Neural Network for Binary Classification of Imbalanced Medical Data. Lecture Notes in Computer Science. Cham : Springer International Publishing, 2022. P. 147–157. URL: https://doi.org/10.1007/978-3-031-12426-6_12
89. A Study of Multilayer Perceptron Networks Applied to Classification of Ceramic Insulators Using Ultrasound / N. F. Sopelsa Neto et al. Applied Sciences. 2021. Vol. 11, no. 4. P. 1592. URL: <https://doi.org/10.3390/app11041592>

90. Ткаченко Р.О., Ткаченко П.Р., Ізонін І.В. Нейромережеві засоби штучного інтелекту : навч. посібник. Львів : Видавництво Львівської політехніки, 2017. 206 с.
91. Yakovyna V., Shakhovska N., Szpakowska A. A novel hybrid supervised and unsupervised hierarchical ensemble for COVID-19 cases and mortality prediction. *Scientific Reports*. 2024. Vol. 14, no. 1. URL: <https://doi.org/10.1038/s41598-024-60637-y>
92. Shakhovska N., Yakovyna V., Chopyak V. A new hybrid ensemble machine-learning model for severity risk assessment and post-COVID prediction system. *Mathematical Biosciences and Engineering*. 2022. Vol. 19, no. 6. P. 6102–6123. URL: <https://doi.org/10.3934/mbe.2022285>
93. Зуб Х., Жежнич П. Бустингові методи машинного навчання для прогнозування успішності вступу абітурієнтів ЗВО України. *Computer systems and information technologies*. 2023. № 1. С. 84–90. URL: <https://doi.org/10.31891/csit-2023-1-11>
94. Zub K., Zhezhnych P. Performance Evaluation of ML-based Classifiers for HEI Graduate Entrants. *CEUR-WS.org*, 2021. Vol. 3003. P. 92–99.
95. Зуб Х. Оцінювання шансів вступу вступником ЗВО на основі моделі стекінгового об'єднання машини опорних векторів. *Наукові записки. Українська академія друкарства*. 2021. Вип. 2, №. 63. С. 168–176. URL: <https://doi.org/10.32403/1998-6912-2021-2-63-168-1>
96. PNN-SVM Approach of Ti-Based Powder's Properties Evaluation for Biomedical Implants Production / I. Izonin et al. *Computers, Materials & Continua*. 2022. Vol. 71, no. 3. P. 5933–5947. URL: <https://doi.org/10.32604/cmcc.2022.022582>
97. Fawagreh K., Gaber M. M., Elyan E. Random forests: from early developments to recent advancements. *Systems Science & Control Engineering*. 2014. Vol. 2, no. 1. P. 602–609. URL: <https://doi.org/10.1080/21642583.2014.956265>
98. Stacking-based GRNN-SGTM Ensemble Model for Prediction Tasks / I. Izonin et al. 2020 International Conference on Decision Aid Sciences and Application

- (DASA), Sakheer, Bahrain, 8–9 November 2020. Bahrain : IEEE, 2020. P. 326–330. URL: <https://doi.org/10.1109/dasa51403.2020.9317124>
99. Parallel Support Vector Machines: The Cascade SVM / H. Graf et al. *Advances in Neural Information Processing Systems 17 (NIPS 2004)*. MIT Press, 2004
URL:
<https://papers.nips.cc/paper/2004/hash/d756d3d2b9dac72449a6a6926534558a-Abstract.html>
100. Dudzik, W., Nalepa, J., & Kawulok, M. Ensembles of evolutionarily-constructed support vector machine cascades. *Knowledge-Based Systems*. 2024. Vol. 288, P. 111490. DOI: 10.1016/j.knosys.2024.111490.
101. Comparison of ensemble hybrid sampling with bagging and boosting machine learning approach for imbalanced data / N. H. A. Malek et al. *Indonesian Journal of Electrical Engineering and Computer Science*. 2022. Vol. 29, no. 1. P. 598. URL: <https://doi.org/10.11591/ijeecs.v29.i1.pp598-608>
102. Правила прийому на навчання до Національного університету «Львівська політехніка» у 2024 році. Національний університет «Львівська політехніка». URL: <https://lpnu.ua/pryimalna-komisiia/pravy-la-pryiomu> (дата звернення: 6.03.2024)
103. Положення про гарантів освітніх програм у Національному університеті «Львівська політехніка». Національний університет «Львівська політехніка». URL: <https://lpnu.ua/polozhennia-pro-garantiv-osvitnikh-program> (дата звернення: 10.02.2023)
104. Комар М. Інформаційна технологія інтелектуальної обробки та аналізу великих даних. *Вісник Хмельницького національного університету. Технічні науки*, 2020. № 5. С. 125–130. URL: <https://www.scopus.com/record/display.uri?eid=2-s2.0-85138786653&origin=resultslist>
105. Papagiannopoulou E., Tsoumakas G. A Review of Keyphrase Extraction.

- arXiv:1905.05044v2 URL: <https://arxiv.org/abs/1905.05044v2> (Last accessed: 02.03.2024)
106. A Review of Unsupervised Keyphrase Extraction Methods Using Within-Collection Resources / C. Sun et al. *Symmetry*. 2020. Vol. 12, no. 11. P. 1864. URL: <https://doi.org/10.3390/sym12111864>
 107. YAKE! Keyword extraction from single documents using multiple local features / R. Campos et al. *Information Sciences*. 2020. Vol. 509. P. 257–289. URL: <https://doi.org/10.1016/j.ins.2019.09.013>
 108. K-means Clustering Algorithms: A Comprehensive Review, Variants Analysis, and Advances in the Era of Big Data / A. M. Ikotun et al. *Information Sciences*. 2022. URL: <https://doi.org/10.1016/j.ins.2022.11.139>
 109. Levenshtein Distance. ScienceDirect. URL: <https://www.sciencedirect.com/topics/computer-science/levenshtein-distance> (date of access: 21.05.2023)
 110. Cherrington M., Thabtah F., Lu J., Xu Q. Feature Selection: Filter Methods Performance Challenges. 2019 International Conference on Computer and Information Sciences (ICCIS), Sakaka, Saudi Arabia, 2019, P. 1–4. DOI: 10.1109/ICCISci.2019.8716478.
 111. Zub K., Zhezhnych P., Strauss C. Two-Stage PNN–SVM Ensemble for Higher Education Admission Prediction. *Big Data and Cognitive Computing*. 2023. Vol. 7, no. 2. P. 83. URL: <https://doi.org/10.3390/bdcc7020083>
 112. GRNN Approach Towards Missing Data Recovery Between IoT Systems / I. Izonin et al. *Advances in Intelligent Networking and Collaborative Systems*. Cham, 2019. Vol. 1035. P. 445–453. URL: https://doi.org/10.1007/978-3-030-29035-1_43
 113. Data Classification Based on the Features Reduction and Piecewise Linear Separation / I. Krak et al. *Advances in Intelligent Systems and Computing*. Cham, 2019. Vol.1072. P. 282–289. URL: https://doi.org/10.1007/978-3-030-33585-4_28

114. Shakhovska N., Shebeko A., Prykarpatsky Y. A Novel Explainable AI Model for Medical Data Analysis. *Journal of Artificial Intelligence and Soft Computing Research*. 2024. Vol. 14, no. 2. P. 121–137. URL: <https://doi.org/10.2478/jaiscr-2024-0007>
115. Malik A., Burney A., Ahmed F. A Comparative Study of Unstructured Data with SQL and NO-SQL Database Management Systems. *Journal of Computer and Communications*. 2020. Vol. 08, no. 04. P. 59–71. URL: <https://doi.org/10.4236/jcc.2020.84005>
116. Supervised learning. scikit-learn. URL: https://scikit-learn.org/stable/supervised_learning.html (date of access: 13.05.2024)
117. Singh D., Singh B. Investigating the impact of data normalization on classification performance. *Applied Soft Computing*. 2020. Vol. 97. P. 105524. URL: <https://doi.org/10.1016/j.asoc.2019.105524>

Додаток А.

Таблиця А.1. Показники ефективності розв'язання задачі класифікації при використанні МОВ з рбф ядром на основі аналізу різної кількості незалежних атрибутів відібраних чотирьома методами вибору важливих ознак із заданого набору даних (для режимів як навчання так і застосування класифікатора).

<i>Метод / показник ефективності</i>	<i>Загальна точність</i>	<i>Влучність</i>	<i>Повнота</i>	<i>F1-міра</i>	<i>Коеф. Кореляції Метьюса</i>	<i>Опт. кількість ознак отримана відповідним методом</i>
Режим навчання						
Використовуються усі ознаки	0,958	0,937	0,935	0,935	0,889	71
Поріг дисперсії (VarianceThreshold) = 0.05	0,912	0,904	0,891	0,897	0,790	17
Фільтр кореляції Пірсона (Pearson correlation filter)	0,886	0,883	0,870	0,876	0,735	15
Фільтр взаємної інформації (Mutual Information filter) = 0.1	0,914	0,897	0,904	0,900	0,795	6
Фільтр Anova F-score (Analysis of Variance F-score filter) = 10	0,947	0,930	0,925	0,927	0,866	58

<i>Режим застосування</i>						
Використовують ся усі ознаки	0,956	0,935	0,934	0,934	0,885	73
Поріг дисперсії (VarianceThreshold) = 0.05	0,907	0,902	0,886	0,894	0,780	17
Фільтр кореляції Пірсона (Pearson correlation filter)	0,850	0,870	0,820	0,845	0,662	15
Фільтр взаємної інформації (Mutual Information filter) = 0.1	0,844	0,849	0,838	0,843	0,647	6
Фільтр Anova F- score (Analysis of Variance F-score filter) = 10	0,954	0,935	0,930	0,932	0,880	58

Додаток Б.

Таблиця Б.1. Оптимальні значення параметрів усіх досліджуваних методів машинного навчання та штучних нейронних мереж під час аналізу досліджуваного набору.

<i>Метод</i>	<i>Оптимальні параметри роботи методу</i>
Класифікатор на основі стохастичного градієнтного спуску	(loss='hinge', *, penalty='l1', alpha=0.001, l1_ratio=0.15, max_iter=1000)
Класифікатор на основі машини опорних векторів з лінійним ядром	(C=1.0, kernel='linear', gamma='scale', max_iter=-1)
Класифікатор на основі Дерева рішень	(criterion='gini', splitter='best', max_depth=None, min_samples_split=2, class_weight=None)
Класифікатор на основі машини опорних векторів з сигмоїдним ядром	(C=1.0, kernel='sigmoid', gamma='scale', max_iter=-1)
Класифікатор на основі машини опорних векторів з поліноміальним ядром	(C=1.0, kernel='poly', degree=2, gamma='scale', max_iter=-1)
Класифікатор на основі машини опорних векторів з рбф ядром	(C=1.0, kernel='rbf', gamma='scale', max_iter=-1)
Класифікатор на основі багат шарового перцептрону	(73 нейрони у вхідному шарі, hidden_layer_sizes=(150), activation='relu', solver='adam', alpha=0.0001, max_iter=1000)
Класифікатор на основі ймовірнісної нейронної мережі	(73 нейрони у вхідному шарі, 1 – у вихідному, розмах гаусівської функції = 0,9);

Класифікатор на основі алгоритму AdaBoost	(estimator= DecisionTreeClassifier , n_estimators=100, learning_rate=1.0, algorithm='SAMME.R')
Класифікатор на основі алгоритму Випадкового лісу	(n_estimators=100, min_samples_split=2, min_samples_leaf=1, bootstrap=True,)
Класифікатор на основі стекінгового об'єднання МОВ з 4 ядрами	(C=1.0, kernel='linear', gamma='scale', max_iter=-1); (C=1.0, kernel=' sigmoid', gamma='scale', max_iter=-1); (C=1.0, kernel='poly', degree=2, gamma='scale', max_iter=-1); (C=1.0, kernel='rbf', gamma='scale', max_iter=-1); (penalty='l2', C=1.0, class_weight=None, solver='lbfgs', max_iter=1000)
Класифікатор на основі сумісного використання ЙМН та МОВ з рбф ядром (каскад)	(ЙМН: 73 нейрони у вхідному шарі, 1 – у вихідному, розмах гаусівської функції = 0,9); (C=1.0, kernel='rbf', gamma='scale', max_iter=-1)

Таблиця Б.1. Оптимальні значення параметрів двоетапного методу класифікації даних та його основних компонентів під час аналізу досліджуваного набору даних

<i>Метод</i>	<i>Оптимальні параметри роботи методу</i>
Класифікатор на основі ймовірнісної нейронної мережі (ЙМН)	(73 нейрони у вхідному шарі, 1 – у вихідному, розмах гаусівської функції = 0,9);

Класифікатор на основі сумісного використання ЙМН та машини опорних векторів з лінійним ядром	(C=1.0, kernel='linear', gamma='scale', max_iter=-1)
Класифікатор на основі сумісного використання ЙМН та машини опорних векторів з сигмоїдним ядром	(C=1.0, kernel='sigmoid', gamma='scale', max_iter=-1)
Класифікатор на основі сумісного використання ЙМН та машини опорних векторів з поліноміальним ядром	(C=1.0, kernel='poly', degree=2, gamma='scale', max_iter=-1)
Класифікатор на основі сумісного використання ЙМН та машини опорних векторів з рбф ядром	(C=1.0, kernel='rbf', gamma='scale', max_iter=-1)
Класифікатор на основі логістичної регресії	(penalty='l2', C=1.0, class_weight=None, solver='lbfgs', max_iter=1000)
Удосконалений двоетапний метод класифікації даних	(73 нейрони у вхідному шарі, 1 – у вихідному, розмах гаусівської функції = 0,9); (C=1.0, kernel='linear', gamma='scale', max_iter=-1); (C=1.0, kernel='sigmoid', gamma='scale', max_iter=-1); (C=1.0, kernel='poly', degree=2, gamma='scale', max_iter=-1); (C=1.0, kernel='rbf', gamma='scale', max_iter=-1); (penalty='l2', C=1.0, class_weight=None, solver='lbfgs', max_iter=1000)

Додаток В.

Акти впровадження

«ЗАТВЕРДЖУЮ»

Проректор

з науково-педагогічної роботи

та інформатизації

Дрогобицького державного

педагогічного університету

імені Івана Франка

Володимир ГАЛИК

29 травня 2024р



Акт

про використання результатів дисертаційної роботи

Зуб Христини Віталіївни

«Інформаційна технологія підтримки прийняття рішень вступників ЗВО України»

Цей акт складений про те, що теоретичні результати дисертаційної роботи Зуб Х.В. «Інформаційна технологія підтримки прийняття рішень вступників ЗВО України» у частині застосування методів надання рекомендацій освітніх програм вступникам ЗВО використані під час проведення інформаційних та маркетингових заходів щодо популяризації пропонуванних освітніх програм та залучення потенційних вступників в межах вступної кампанії 2023-2024 р.р. Дрогобицьким державним педагогічним університетом імені Івана Франка.

Використання результатів дисертаційної роботи Зуб Х.В. дозволило:

- забезпечити збір та опрацювання необхідних даних описів освітніх програм для надання рекомендацій вступникам методом фільтрації на основі вмісту;
- розширити обізнаність та зацікавленість потенційних абітурієнтів на вступ на перший курс для здобуття ступеня бакалавра за допомогою наданих рекомендацій альтернативних освітніх програм;
- розширити спектр надання інформаційних послуг під час консультування зацікавлених у вступі до Дрогобицького державного педагогічного університету імені Івана Франка.

Експерти

Адміністратор програмних засобів доступу до Єдиної державної електронної бази з питань освіти

Відповідальний секретар приймальної комісії

Роман БЛИЙ

Анна ЧЕПЕЛЮК



“ЗАТВЕРДЖУЮ”

Проректорка з науково-педагогічної роботи
та міжнародних зв'язків
Наталія ЧУХРАЙ

“18 травня” 2024 р.

Акт

про використання результатів дисертаційної роботи

Зуб Христини Віталіївни

«Інформаційна технологія підтримки прийняття рішень вступників ЗВО
України»

Цей акт складений про те, що теоретичні результати дисертаційної роботи Зуб Х.В. «Інформаційна технологія підтримки прийняття рішень вступників ЗВО України» у частині застосування технологій збору, попереднього опрацювання та аналізу даних використано під час виконання дослідження щодо академічної доброчесності та відкритої науки в межах міжнародного проекту Erasmus+ «Відкриті практики, прозорість та доброчесність для сучасної вищої школи» (ОРТІМА, 618940-EPP-1-2020-1-UA-EPPKA2-SBHE- JP).

Зокрема, було здійснено попередню обробку та аналіз наборів даних, які є результатами комплексного опитування, проведеного у 2021-му, 2022-му та 2023-му роках на національному рівні в Україні та в межах українських університетів-партнерів проекту. Аспіранткою Зуб Х.В. було виконано аналіз повноти даних, обчислення статистичних показників, статистичний аналіз - для виявлення зв'язків між результатами опитування та розуміння впливу факторів на результати опитування та візуалізацію результатів для їх інтерпретації.

Координатор проекту ОРТІМА
д.т.н., проф.
Менеджер проекту ОРТІМА
к.т.н., доц.

Павло ЖЕЖНИЧ

Олександр БЕРЕЗКО