

УДК 004.9, 004.89, 004.021

## ОЦІНЮВАННЯ ШАНСІВ ВСТУПУ АБІТУРІЄНТОМ ЗВО НА ОСНОВІ МОДЕЛІ СТЕКІНГОВОГО ОБ'ЄДНАННЯ МАШИНИ ОПОРНИХ ВЕКТОРІВ

Х. В. Зуб

*Національний університет «Львівська політехніка»,  
вул. С. Бандери, 12, Львів, 79013, Україна*

*Розглянуто оцінювання успішності вступу абітурієнта як один із критично важливих етапів вступної кампанії закладів вищої освіти. Враховуючи велику кількість чинників, що впливають на вибір абітурієнта та складність їх аналізу, розглянуто потребу створення системи підтримки прийняття рішень абітурієнтів закладів вищої освіти. З огляду на тенденції сучасних досліджень, засобом вирішення задачі прогнозування вступу було обрано методи машинного навчання. Враховуючи результати попередніх досліджень, підвищення точності оцінювання шансів вступу було здійснено шляхом застосування гетерогенного стекінгового об'єднання машини опорних векторів з різними ядрами. Модель полягає у застосуванні ансамблю чотирьох методів машини опорних векторів з такими чотирма ядрами, як лінійне, поліноміальне, радіально-базисне, сигмоподібне та логістичної регресії як метаалгоритму. Отримані результати свідчать про підвищення точності роботи застосованої моделі.*

**Ключові слова:** *вступ, абітурієнт, заклад вищої освіти (ЗВО), прогнозування, машинне навчання, ансамблеве навчання, машина опорних векторів, логістична регресія.*

**Постановка проблеми.** Одним із пріоритетних завдань закладів вищої освіти (ЗВО) є якісний, надійний та ефективний інформаційний супровід усіх зацікавлених у вступі. Сьогодні ЗВО у великому обсязі та доступній формі забезпечують цільову аудиторію актуальною та достовірною інформацією на офіційних вебсайтах. Існують також відкриті онлайн-сервіси, що додатково інформують про рейтингові списки та перебіг вступної кампанії. Проте, окрім інформаційного опису освітніх пропозицій та правил прийому, існує чимало факторів, які студенти повинні врахувати при виборі спеціальності для вступу [1]. Водночас постійна зміна тенденцій ринку праці, недостовірність та низька якість джерел (поради батьків чи друзів), великий обсяг правил прийому можуть ускладнити процес вибору для абітурієнта. Як наслідок, з'являється ризик того, що зробити правильний вибір та якісну оцінку шансів на вступ абітурієнт не зможе.

Вибір спеціальності для вступу є критично важливим моментом, оскільки прямо впливає на життєву траєкторію студента та на якість освіти в цілому. Одним із факторів впливу на кінцеве рішення є оцінка шансів на вступ з точки зору

абітурієнта. Тож можна припустити, що існування системи, яка б забезпечила належну підтримку у прийнятті рішень, було б корисним для абітурієнтів. Враховуючи стрімкий розвиток інформаційних технологій, доречно детальніше розглянути можливість застосування методів, які б вирішили задачу прогнозування успішності вступу.

**Аналіз останніх досліджень та публікацій.** Серед вітчизняних досліджень є лише незначна частина праць, що стосується підтримки прийняття рішень абітурієнтів. В. В. Осадчий у своїх працях описав застосування нейронних мереж з метою прогнозування вступу [2, 3]. Критично важливим для отримання точних результатів автор визначив використання великого набору даних та якісної навчальної вибірки. Запропонований програмний засіб дав достатньо точні результати.

Значно ширше питання прогнозування результатів вступу відображено в сучасних дослідженнях західних науковців. Отримані результати свідчать, що достатньо ефективним засобом вирішення задачі прогнозування є застосування методів машинного навчання. Зокрема, у роботі [4] було запропоновано стекінгову модель ансамблю, яка передбачає шанси вступу студента до певного університету. Автори наголосили на необхідності врахування різних факторів, пов'язаних із профілем студента. Запропонована система була оцінена на основі різних алгоритмів машинного навчання та показала високу точність результатів. Ще в одній роботі [5] було запропоновано модель, що використовує алгоритми лінійної регресії та випадкових лісів, але найвищу точність показав один із алгоритмів градієнтного підсилення. У іншій роботі авторами було розроблено підхід машинного навчання для автоматичного передбачення можливості вступу до аспірантури, щоб допомогти випускникам визнати та орієнтуватися на університети, які найкраще підходять для їх профілю [6]. За допомогою керованого навчання в іншій роботі було представлено модель, що прогнозує результати вступу на основі даних, введених користувачами [7]. Варто зазначити, що в деяких випадках дослідження спрямовані на те, щоб забезпечити й задовільнити цілі ЗВО, наприклад, передбачити кількість майбутніх студентів тощо [8].

Джерелами даних, що були використані в дослідженнях, є результати опитувань, інформація про успішність студентів та (або) історичні дані про абітурієнтів попередніх років вступу. Проте залишається проблема неврахування важливих параметрів набору даних при застосуванні тих чи інших методів. Іншими недоліками застосованих моделей були недостатньо висока точність результатів та відсутність порівняння результатів з іншими існуючими методами. Водночас праці вітчизняних фахівців недостатньо широко та повно досліджують задачі прогнозування успішності вступу, тож цей напрям потребує подальшого дослідження. Залишається актуальним завдання відбору, застосування та оцінки моделей, методів та засобів, які б забезпечили найвищу точність прогнозування шансів вступу до певного ЗВО для обраної спеціальності.

**Мета статті** – підвищення точності оцінювання шансів вступу до ЗВО абітурієнтом на основі стекінгового об'єднання машини опорних векторів (МОВ) з чотирма різними ядрами.

**Виклад основного матеріалу дослідження.** Попередньо проведені дослідження щодо визначення ефективності низки існуючих регресійних методів на основі машинного навчання показали доцільність їх застосування для прогнозування успіху вступу до ЗВО [9]. Проте, з огляду на отримані результати, все ж існує необхідність підвищення точності прогнозування шансів вступу. Це дослідження полягає в оцінюванні ефективності застосування ансамблевого підходу до розв'язання поставленої задачі, зокрема запропоновано використання моделі стекінгового об'єднання МОВ з різними ядрами.

МОВ є одним із поширених та ефективних методів вирішення задач бінарної класифікації (число класів обмежено двома) чи регресійного аналізу. Він перетворює вхідні дані поставленої задачі у простір вищої розмірності для побудови оптимальної роздільної гіперповерхні. Основна ідея полягає у тому, аби знайти лінію між точками на площині, яка забезпечить найбільш точну процедуру віднесення точок до відповідних класів. Даний метод характеризується високою швидкістю роботи та водночас можливістю обробки великих обсягів даних [10].

Однією з характеристик, що відрізняє МОВ від інших методів, є можливість використання функцій ядра. Ядро відіграє важливу роль у класифікації і використовується для аналізу деяких закономірностей у певному наборі даних. Різні алгоритмічні реалізації МОВ можуть використовувати різні види функцій ядра. Серед функцій найбільш популярними є лінійне (linear), сигмоподібне (sigmoid), поліноміальне (polynomial), радіально-базисне (radial basis function, RBF), ядро Гауса (Gaussian Kernel), ядро функції Бесселя (Bessel function kernel). Варто зазначити, що точність класифікатора певною мірою залежить і від вибору ядра [11].

З метою отримання більш точних результатів при застосуванні методів машинного навчання використовують алгоритм їх ансамблів, що полягає в одночасному застосуванні декількох базових алгоритмів. При їх використанні алгоритми навчаються одночасно та можуть виправляти помилки одне одного. Ансамбль сам по собі є алгоритмом навчання з учителем, оскільки він може бути тренований і потім використаний для здійснення прогнозування, тому тренований ансамбль представляє одну гіпотезу. Ця гіпотеза, однак, не обов'язково лежить в просторі гіпотез моделей, з яких вона побудована. Таким чином, ансамблі можуть мати більшу гнучкість у функціях, які вони представляють.

Існує кілька підходів до побудови ансамблевих алгоритмів, зокрема стекінг (stacking), бустинг (boosting) та бегінг (bagging). Бустинг полягає у поступовому застосуванні кожної з моделей, при цьому кожна наступна виправляє помилки попередньої. При бегінгу базові алгоритми навчаються паралельно, а кінцеві результати агрегуються. Під час стекінгу навчаються декілька різних алгоритмів, а на їх виходах метамодель видає вже кінцевий результат.

З метою підвищення точності результатів прогнозування у цій роботі для проведення досліджень було обрано ансамблевий метод гетерогенного стекінгового об'єднання МОВ. Базовими алгоритмами є МОВ з відмінними ядрами та металгоритмом – логістична регресія. Схема запропонованого ансамблю показана на рис. 1.

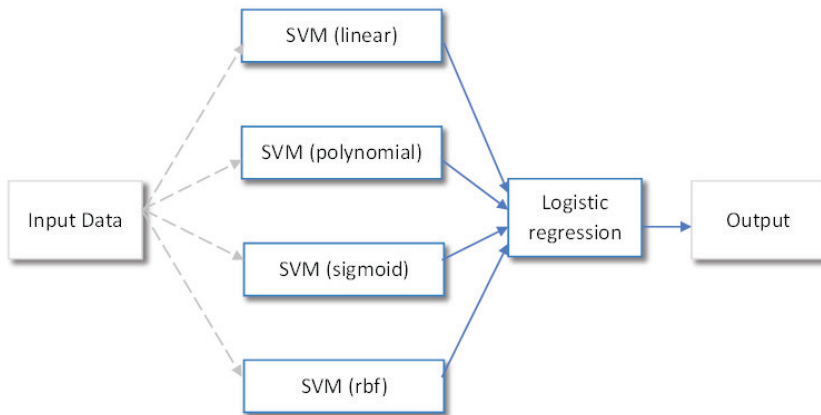


Рис. 1. Модель стекингового об'єднання МОВ

*Моделювання.* Нами було досліджено ефективність застосування стекингового об'єднання МОВ для розв'язання задачі прогнозування успішності вступу абітурієнта до ЗВО. Моделювання проводилося за допомогою програмного забезпечення Orange [12] – це онлайн-інструмент для візуалізації даних, машинного навчання та аналізу даних з відкритим кодом. Він оснащений візуальним інтерфейсом програмування для швидкого якісного аналізу та інтерактивної візуалізації даних.

*Дані.* Нами було досліджено задачу бінарної класифікації, що полягає у тому, аби визначити – вступить абітурієнт до ЗВО чи ні. Оцінка ефективності обраного методу проводилася на основі реального набору даних про вступ до ЗВО Сполучених Штатів Америки (США) [13]. Автори зібрали інформацію про вступ до аспірантури в 29-х університетах США. Набір даних містить 1653 записи, дев'ять атрибутів. Було розглянуто наступні властивості набору даних: оцінка з англійської мови, оцінка іспитів для випускників, кількісні міркування (Gre оцінка Quat), словесна аргументація (оцінка Gre Grebal Verbal), опублікована наукова робота, рейтинг, бал бакалаврату, досвід роботи.

*Модель.* Модель полягає у застосуванні ансамблю чотирьох методів МОВ з різними ядрами: лінійним, поліноміальним, радіально-базисним та сигмоподібним. Схему застосування обраної моделі в застосунку Orange подано на рис. 2.

З огляду на достатній розмір вибірки даних було здійснено поділ на навчальну та тестову вибірки у співвідношенні 80 % до 20 %. Оцінки точності класифікації проводились за допомогою 10-кратної крос-перевірки (Cross-validation), що полягає у порівнянні результатів точності класифікації тестової та навчальної множин. Вважається, що досліджувана модель пройшла перевірку за умови, якщо класифікація тестової множини дає приблизно такі ж результати за точністю, як і класифікація навчальної.

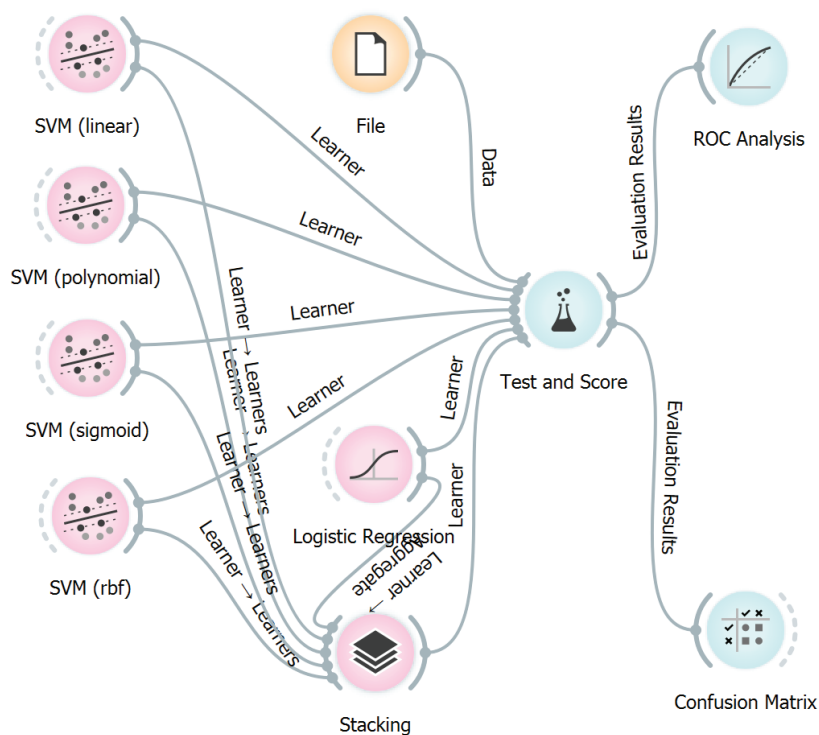


Рис. 2. Моделювання стекінгового об'єднання МОВ в застосунку Orange

*Результати дослідження.* Показники ефективності класифікаторів узагальнені у таблиці 1. Для аналізу результатів було використано оцінки точності класифікації (Classification accuracy, AUC) як частку правильно класифікованих прикладів. Відповідно до отриманих результатів та у порівнянні з попередньо отриманими результатами роботи різних класифікаторів стекінговий ансамбль показав вищу точність, тому можна вважати, що застосування обраного алгоритму для вирішення завдання бінарної класифікації в нашому випадку повністю себе виправдало.

Таблиця 1

## Результати дослідження

Модель	AUC	CA	F1	Precision	Recall
Stacking	0.714	0.700	0.675	0.682	0.700
SVM (rbf)	0.708	0.700	0.674	0.638	0.700
SVM (linear)	0.703	0.680	0.628	0.657	0.680
Logistic Regression	0.701	0.693	0.661	0.674	0.693
SVM (polynomial)	0.668	0.663	0.590	0.625	0.663
SVM (sigmoid)	0.598	0.633	0.632	0.630	0.633

Отримані числові значення метрик підтверджено і візуальним аналізом. Для візуалізації результатів дослідження використано криву помилок, так звану ROC-криву (ROC-curve) (рис. 3). Це один з найчастіше використовуваних методів демонстрації результатів бінарної класифікації. ROC-крива показує залежність кількості істинно-позитивних значень від кількості хибно-позитивних значень для кожного окремого класу. Відповідно, досліджуваний класифікатор, ROC-крива якого знаходиться зверху та ліворуч від графіка, демонструє більшу точність.

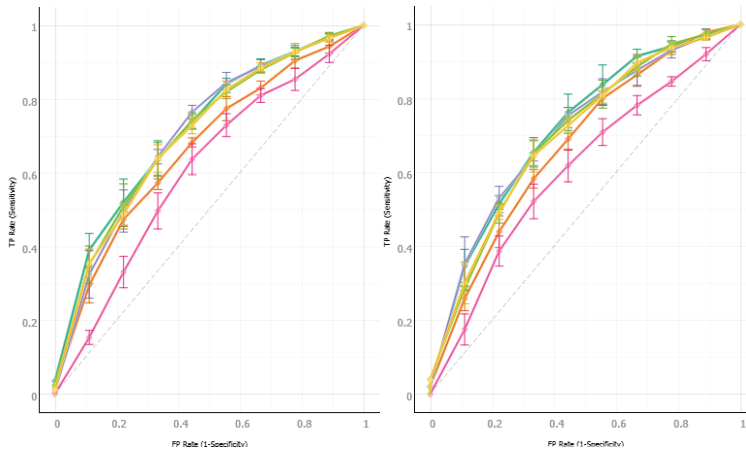


Рис. 3. ROC-криві: а) для першого класу, б) для другого класу

Як видно з рис. 3, застосування стекінгового об'єднання машини опорних векторів забезпечило підвищення точності у порівнянні з окремими моделями, які формують стекінг.

**Висновки.** Розглянуто завдання підтримки прийняття рішень абітурієнтів ЗВО під час вибору спеціальності для вступу. За допомогою методів машинного навчання досліджено задачу бінарної класифікації щодо оцінки шансів абітурієнта на вступ. Отримані результати свідчать про підвищення точності роботи обраної моделі гетерогенного стекінгу машини опорних векторів з чотирма різними ядрами та логістичної регресії як метаалгоритму.

#### СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Zhezhnych P., Berezko O., Zub K., Demydov I. Analysis of Features and Abilities of Online Systems and Tools Meeting Information Needs of HEIs. Entrants, CEUR-WS.org, 2020. Vol. 2616. Pp. 76–85.
2. Осадчий В. В., Круглик В. С., Осадча К. П., Сердюк І. М., Букреєв 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. 30 (69). № 1. 1.
3. 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. Vol. 6. № 3. Art. no. 3.

4. Sridhar S., Mootha S., Kolagati S. A University Admission Prediction System using Stacked Ensemble Learning. 2020 Advanced Computing and Communication Technologies for High Performance Applications (ACCTHPA), 2020, Pp. 162–167. doi: 10.1109/ACCTHPA49271.2020.9213205.
5. Sivasangari A., Shivani V., Bindhu Y., Deepa D., Vignesh R. Prediction Probability of Getting an Admission into a University using Machine Learning, in 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), Apr. 2021, Pp. 1706–1709. doi: 10.1109/ICCMC51019.2021.9418279.
6. AlGhamdi A., Barsheed A., AlMshjary H., AlGhamdi H. A Machine Learning Approach for Graduate Admission Prediction, in Proceedings of the 2020 2nd International Conference on Image, Video and Signal Processing, New York, NY, USA, Mar. 2020, Pp. 155–158. doi: 10.1145/3388818.3393716.
7. Sujay, S. Supervised Machine Learning Modelling & Analysis For Graduate Admission Prediction. *Published in International Journal of Trend in Research and Development (IJTRD)*, ISSN: 2394-9333, Vol.-7 | Issue-4 , August 2020.
8. Devarapalli D. J. Classification Method to Predict Chances of Students' Admission in a Particular College, in Proceedings of International Conference on Recent Trends in Machine Learning, IoT, Smart Cities and Applications, Vol. 1245, V. K. Gunjan and J. M. Zurada, Eds. Singapore: Springer Singapore, 2021, Pp. 225–238. doi: 10.1007/978-981-15-7234-0\_19.
9. Zhezhnych P., Zub K., Berezko O., Shilinh A. A Comparison of Machine Learning Algorithms for Prediction Higher Education Institution's Entrants Admissions, *Lecture Notes on Data Engineering and Communications Technologies*, Vol. 82, Pp. 171–179, 2021. doi: 10.1007/978-3-030-80475-6\_17.
10. Izonin I., Trostianchyn A., Duriagina Z., Tkachenko R., Tepla T., Lotoshynska N. The Combined Use of the Wiener Polynomial and SVM for Material Classification Task in Medical Implants Production. *International Journal of Intelligent Systems and Applications*, Vol. 10, Pp. 40–47, Sep. 2018. doi: 10.5815/ijisa.2018.09.05.
11. Awasthi S. Seven Most Popular SVM Kernels. *Dataaspirant*, Dec. 17, 2020. URL: <https://dataaspirant.com/svm-kernels/> (accessed Oct. 13, 2021).
12. B. L. Ljubljana University of, Data Mining. URL: <https://orangedatamining.com/> (accessed Oct. 13, 2021).
13. Graduate Admission 2. URL: <https://kaggle.com/mohansacharya/graduate-admissions> (accessed Oct. 13, 2021).

## REFERENCES

1. Zhezhnych, P., Berezko, O., Zub, K., & Demydov, I. (2020). Analysis of Features and Abilities of Online Systems and Tools Meeting Information Needs of HEIs. *Entrants*, CEUR-WS.org, 2616, 76–85 (in English).
2. Osadchyi, V. V., Kruhlyk, V. S., Osadcha, K. P., Serdiuk, I. M., & Bukreiev D. O. (2019). Osoblyvosti rozrobky prohramnoho zasobu dlia prohnozuvannia vstupu abiturientiv do zakladiv vyshchoi osvity: Vcheni zapysky Tavriiskoho Natsionalnoho universytetu imeni V. I. Vernadskoho. *Seriia: Tekhnichni nauky*, 30 (69), 1, 1 (in Ukrainian).

3. Osadchyi, V. V., Kruhlyk, V. S., & Bukreiev, D. O. (2018). Rozrobka prohramnoho zasobu dlia prohozuvannia vstupu abiturientiv do zakladiv vyshchoi osvity: Ukrainian Journal of Educational Studies and Information Technology, 6, 3. Art. no. 3 (in Ukrainian).
4. Sridhar, S., Mootha, S., & Kolagati, S. (2020). A University Admission Prediction System using Stacked Ensemble Learning. 2020 Advanced Computing and Communication Technologies for High Performance Applications (ACCTHPA), 162–167. doi: 10.1109/ACCTHPA49271.2020.9213205 (in English).
5. Sivasangari, A., Shivani, V., Bindhu, Y., Deepa, D., & Vignesh, R. (Apr. 2021). Prediction Probability of Getting an Admission into a University using Machine Learning, in 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), 1706–1709. doi: 10.1109/ICCMC51019.2021.9418279 (in English).
6. AlGhamdi, A., Barsheed, A., AlMshjary, H., & AlGhamdi, H. (Mar. 2020). A Machine Learning Approach for Graduate Admission Prediction, in Proceedings of the 2020 2nd International Conference on Image, Video and Signal Processing, New York, NY, USA, 155–158. doi: 10.1145/3388818.3393716 (in English).
7. Sujay, S. Supervised Machine Learning Modelling & Analysis For Graduate Admission Prediction: Published in International Journal of Trend in Research and Development (IJTRD), ISSN: 2394-9333, 7 | Issue-4 , August 2020 (in English).
8. Devarapalli, D. J. (2021). Classification Method to Predict Chances of Students' Admission in a Particular College, in Proceedings of International Conference on Recent Trends in Machine Learning, IoT, Smart Cities and Applications, 1245. V. K. Gunjan and J. M. Zurada, Eds. Singapore: Springer Singapore, 225–238. doi: 10.1007/978-981-15-7234-0\_19 (in English).
9. Zhezhnych, P., Zub, K., Berezko, O., & Shilinh, A. (2021). A Comparison of Machine Learning Algorithms for Prediction Higher Education Institution's Entrants Admissions, Lecture Notes on Data Engineering and Communications Technologies, 82, 171–179. doi: 10.1007/978-3-030-80475-6\_17 (in English).
10. Izonin, I., Trostianchyn, A., Duriagina, Z., Tkachenko, R., Tepla, T., Lotoshynska, N. (Sep. 2018). The Combined Use of the Wiener Polynomial and SVM for Material Classification Task in Medical Implants Production: International Journal of Intelligent Systems and Applications, 10, 40–47. doi: 10.5815/ijisa.2018.09.05 (in English).
11. Awasthi, S. (Dec. 17, 2020). Seven Most Popular SVM Kernels. Dataaspirant. Retrieved from <https://dataaspirant.com/svm-kernels/> (accessed Oct. 13, 2021) (in English).
12. B. L. Ljubljana University of, Data Mining. Retrieved from <https://orangedatamining.com/> (accessed Oct. 13, 2021) (in English).
13. Graduate Admission 2. Retrieved from <https://kaggle.com/mohansacharya/graduate-admissions> (accessed Oct. 13, 2021) (in English).

doi: 10.32403/1998-6912-2021-2-63-168-176

**THE EVALUATION OF THE HEI'S ENTRANTS ADMISSION CHANCES  
BASED ON THE STACKING MODEL OF THE SUPPORT  
VECTORS MACHINE**



K. V. Zub

*Lviv Polytechnic National University,  
12, St. Bandera St., Lviv, 79013, Ukraine  
khrystyna.v.zub@lpnu.ua*

*The choice of specialty for admission from the entrant's point of view as one of the critical stages of the admission campaign is considered. Taking into account the large number of factors influencing the choice of entrants and the complexity of their analysis, the need to create a decision support system for entrants to higher education institutions is described. The analysis of modern domestic and foreign research concerning support of decision-making of entrants is carried out. Taking into account modern research trends, machine learning methods to predict admission to higher education institutions are chosen. To effectively solve the task of prediction, namely the high accuracy of the result, an ensemble model of machine learning methods is applied. Taking into account the results of previous authors' research, improving the accuracy of admission to higher education institutions, chances assessment is achieved through the use of heterogeneous stacking of the Support Vectors Machine. The model is constructed based on an ensemble of four Support Vector Machine methods with different kernels: linear, polynomial, radial basis, sigmoid and logistic regression as a meta-algorithm. Classification accuracy estimates are used to analyse the results as a proportion of correctly classified examples. The obtained results indicate an increase in the accuracy of the model. The highest accuracy in all performance indicators is obtained by stacking algorithm. The obtained numerical values of the metrics are confirmed by visual analysis using ROC-curve. As a result, one can assume that the use of the chosen algorithm to solve the task of binary classification in our case has fully justified itself. The obtained numerical values of the metric are confirmed by visual analysis. A ROC curve is used to visualize the results of the study.*

**Keywords:** *admission, entrant, higher education institution, HEI, prediction, machine learning, ensemble learning, support vectors machine, logistic regression.*

*Стаття надійшла до редакції 25.10.2021.*

*Received 25.10.2021.*