

УДК 004.85:811'271.14-048.445

DOI <https://doi.org/10.24919/2308-4863/76-3-15>

Ірина ЛАКТИОНОВА,

orcid.org/0009-0005-6340-8761

*викладач кафедри загального мовознавства та іноземних мов
Національного університету «Полтавська політехніка імені Юрія Кондратюка»
(Полтава, Україна) iryna.maslii@ukr.net*

Микола МОВІН,

orcid.org/0000-0002-7470-6482

*аспірант кафедри автоматичної, електроніки та телекомунікацій
Національного університету «Полтавська політехніка імені Юрія Кондратюка»
(Полтава, Україна) mykolatomvin@gmail.com*

Олексій МИХАЙЛІЧЕНКО,

orcid.org/0009-0009-3512-0030

*аспірант кафедри автоматичної, електроніки та телекомунікацій
Національного університету «Полтавська політехніка імені Юрія Кондратюка»
(Полтава, Україна) aleksejmikhajlichenko@gmail.com*

РОЗРОБКА КЛАСИФІКАТОРА ГРАМАТИЧНИХ ПОМИЛОК ІНОЗЕМНОЇ МОВИ ЗАСОБАМИ МАШИННОГО НАВЧАННЯ

Розглянуто процес розробки класифікатора граматичних помилок іноземної мови засобами машинного навчання, зокрема інструментами класифікації. Актуальність дослідження підтверджується у зв'язку з популярністю англійської мови у світі та Україні. Не існує єдиних інструментів щодо розробки моделей для виявлення вірних і хибних варіантів побудови речень англійською мовою. Використання класифікатора, котрий розрізняє вірно побудовані речення англійською і хибно є елементом неформальних цифрових методів вивчення англійської мови.

Мето дослідження є підвищенні ефективності вивчення іноземної мови шляхом використання класифікатора граматичних помилок. Відповідно до методології дослідження процес аналізу й побудови моделей машинного навчання реалізовано засобами мови програмування Python й бібліотеки sklearn, де досліджувалося завдання багатокласової класифікації.

Дослідження проведено на основі відомого датасету Grammar Correction, де зафіксовано набір типів граматичних помилок та вірні приклади вживання речень англійською мовою. Особливістю підготовки даних для моделей машинного навчання є перетворення текстових змінних у числовий (векторний) формат.

Для оптимального підбору гіперпараметрів моделей й уникнення процесу перенавчання моделей Support Vector Classification, K Neighbors Classifier, Random Forest Classifier використовувався інструмент GridSearchCV.

Ефективність побудованих класифікаторів вивчалася методами точності (окремо для навчального й тестового наборів), матрицею плутанини, дисперсією й зміщенням.

Практичне використання створених моделей можливе для опанування англійської мови, за певних обмежень, зокрема при вивченні типу граматичних помилок Subject-Verb Agreement. Для мінімізації впливу перенавчання створену модель пропонується оновлювати за рахунок збільшення обсягу досліджуваних типів помилок англійської мови.

Ключові слова: машинне навчання, завдання класифікації, граматичні помилки, англійська мова, педагогічний інструмент.

Iryna LAKTIONOVA,

orcid.org/0009-0005-6340-8761

Lecturer at the Department of General Linguistics and Foreign Languages
National University "Yuri Kondratyuk Poltava Polytechnic"
(Poltava, Ukraine) iryna.maslii@ukr.net

Mykola MOVIN,

orcid.org/0000-0002-7470-6482

Ph.D. student at the Department of Automation, Electronics, and Telecommunications
National University "Yuri Kondratyuk Poltava Polytechnic"
(Poltava, Ukraine) mykolamovin@gmail.com

Oleksiy MYKHAYLICHENKO,

orcid.org/0009-0009-3512-0030

Ph.D. student at the Department of Automation, Electronics, and Telecommunications
National University "Yuri Kondratyuk Poltava Polytechnic"
(Poltava, Ukraine) aleksejmikhajlichenko@gmail.com

DEVELOPMENT OF A CLASSIFIER FOR GRAMMATICAL ERRORS IN A FOREIGN LANGUAGE USING MACHINE LEARNING

The process of developing a classifier for grammatical errors in a foreign language using machine learning techniques, particularly classification tools, is examined. The relevance of the research is confirmed due to the global and Ukrainian popularization of the English language. There are no universal tools for developing models to detect correct and incorrect sentence constructions in English. The use of a classifier that distinguishes correctly constructed English sentences from incorrect ones is an element of informal digital methods for learning English.

The aim of the research is to enhance the efficiency of foreign language learning by employing a classifier for grammatical errors. According to the research methodology, the process of analyzing and building machine learning models was implemented using Python programming language and the sklearn library, focusing on the task of multiclass classification.

The research was conducted based on the well-known Grammar Correction dataset, which includes a set of types of grammatical errors and correct examples of English sentence usage. The peculiarity of preparing data for machine learning models is transforming textual variables into numerical (vector) format.

For optimal selection of model hyperparameters and avoiding overfitting, the GridSearchCV tool was utilized with Support Vector Classification, K Neighbors Classifier, and Random Forest Classifier models.

The effectiveness of the constructed classifiers was evaluated using accuracy methods (separately for training and testing sets), confusion matrix, dispersion, and bias.

The practical use of the created models is possible for mastering English, under certain limitations, particularly in learning the type of grammatical errors like Subject-Verb Agreement. To minimize the impact of overfitting, it is suggested to update the created model by increasing the volume of researched types of English language errors.

Key words: machine learning, classification task, grammatical errors, English language, pedagogical tool.

Постановка проблеми. Використання англійської мови в Україні розглядається на законодавчому рівні, зокрема це тлумачить схвалений Законопроект №9432 про застосування англійської мови в Україні. У зв'язку з цим фактом постає необхідність розробки традиційних, нетрадиційних педагогічних інструментів щодо опанування мови (Banno, Matassoni, 2023:103025). Одними з таких представників є методи штучного інтелекту, точність використання котрих зростає й сягає понад 90% (Li, 2024:200328).

Крім елементів штучного інтелекту, вивчається проблематика залучення вчителів для інтеграції неформального цифрового вивчення англійської мови (Liu et al., 2023) та учнів початкової англійської мови за допомогою автоматичного зворотного зв'язку (Wilson J. et al., 2024).

Проблематика побудови моделей штучного інтелекту для розпізнавання помилок пов'язана з певними особливостями побудови моделей, де основним принципом, на відміну від (Shefer et al., 2024:86), є перетворення текстової форми даних у векторну. При цьому не існує єдиного підходу, що пов'язано з використанням різних бібліотек, баз даних тощо. У зв'язку з цим актуалізується необхідність розробки класифікатора граматичних помилок іноземної мови засобами машинного навчання, котрий стане у нагоді вчителю.

Аналіз досліджень. Існуючі наукові дослідження вказують на різноманітні інноваційні підходи щодо вивчення англійської мови у світі. Перевагами використання штучного інтелекту для вивчення англійської мови є використання принципу зворотного

зв'язку (O'Shaughnessy, 2023:101538), що дозволяє у процесі життєвого циклу моделей машинного навчання поліпшувати їх якість. Це досягається за рахунок оновлення даних, підбору нових моделей, використання інструментів, котрі точніше опрацюють вхідні дані.

Особливістю процесу створення моделей машинного навчання, для класифікації граматичних помилок, є перетворення вихідного тексту у векторну форму. При цьому попередня обробка тексту впливає на продуктивність моделей (Siino et al., 2024), де вихідний текст може буди задано з використанням синтезатора або природньою мовою.

Особливістю обробки природньої мови є використання спеціалізованого обладнання. У (Rezaul, et al., 2024) запропоновано нормування вхідного тексту для синтезатора мовлення, де точність моделей XGBClassifier сягає 99,997%. При цьому запропоноване рішення не використовувалося для англійської мови.

Науковий доробок (Olayiwola et al., 2023) присвячений розробці автоматичної перевірки граматики для обробки текстів інструментарієм Natural Language Tool kit. Розроблені моделі сягають точності 86,67%, 88,89% для простих і складних речень відповідно, що є вагомим результатом. Запропоновані у роботах підходи є точними, де точність сягає понад 99,9%, але однією з принципових проблем лишається обмеження наборів вихідних даних, зокрема використання обмеженої кількості мов (Putra et al., 2023).

З іншого боку, глобальною проблемою розпізнавання людської мови є чіткість вимови, де при проблемах з мовленням складно зафіксувати й опрацювати дані. Проблематикою такого типу займалися у роботі (Lukic et al., 2024). Утім, використання моделей для класифікації хибних відповідей вивчалось не у повній мірі.

Ідеї дослідження (Lukic et al., 2024) ширше розкриваються у (Thaher et al., 2024), де пропонується процес удосконалення метаевристичного алгоритму пошуку за допомогою механізму структурованої сукупності. Ефективність отриманої моделі перевірено шляхом порівняльного аналізу із сімнадцятьма відомими підходами.

Існуючі дослідження демонструють процеси розробки моделей класифікації, котрі демонструють показники точності на рівні 90% і вище. Принциповими обмеженнями існуючих досліджень є формування вхідних даних у текстовій формі, що не є універсальним способом. Крім того, вихідні дані можуть бути й числовими як у роботі (Laktionov et al., 2023). Не у повній мірі здійсню-

ється розробка моделей класифікації граматичних помилок іноземної мови. У більшості випадків це викликає проблеми у людей з особливими можливостями. Тому актуалізується проблема розробки класифікатора граматичних помилок англійської мови, котрий дозволить підвищити ефективність набуття компетенцій стосовно англійської мови.

Мета дослідження полягає у підвищенні ефективності вивчення іноземної мови шляхом використання класифікатора граматичних помилок.

Виклад основного матеріалу. Методологія дослідження розробки класифікатора граматичних помилок іноземної мови. Відповідно до програми дослідження (грудень 2023 року – червень 2024 року) передбачалося вивчення процесу побудови класифікаційних моделей граматичних помилок іноземною мовою. Перший етап дослідження (листопад 2023 року – березень 2024 року) мав на меті розробку й експериментальну верифікацію моделей класифікації для з'ясування їх точності. Другий етап дослідження (квітень 2024 року – червень 2024 року) зорієнтований на пошук шляхів удосконалення моделей класифікації граматичних помилок для досягнення точності класифікатора 80% і більше.

Диференціювання граматичних помилок іноземної мови здійснено з використання відомого датасету Grammar Correction, що включає N=2018 записів (Grammar Correction). Датасет нараховує 36 різних типів граматичних помилок та має наступні змінні, зокрема тип помилки, неграматичне твердження, стандартна англійська.

Кількість елементів у типах граматичних помилок варіюється від 20 до 103. Перший етап дослідження (листопад 2023 року – березень 2024 року) передбачав вивчення наступних типів граматичних помилок Sentence Structure Errors, Verb Tense Errors, Subject-Verb Agreement, Article Usage, Spelling Mistakes, Preposition Usage. Кожен з цих типів нараховує 103, 100, 100, 100, 100, 95 записів відповідно. Умовно, вказані типи граматичних помилок віднесено до першого кластеру помилок, оскільки інші типи граматичних помилок нараховують 60 і менше записів – другий кластер. Використання всіх записів одночасно приводить до незбалансованих класів.

Відмінністю вказаного дослідження від існуючих є використання типів граматичних помилок у текстовій формі, що потребує відповідних перетворень у числовий формат. Процес фільтрації необхідного обсягу типів граматичних помилок здійснено за класичною методикою побудови моделей класифікації (Müller A. C., Guido S., 2018) мовою програмування Python й бібліотекою sklearn.

Перш за все виділено типи граматичних помилок Sentence Structure Errors, Verb Tense Errors, Subject-Verb Agreement, Article Usage, Spelling Mistakes, Preposition Usage, котрим присвоєно мітки класів 0, 1, 2, 3, 4, 5. Для виділення досліджуваної вибірки здійснено фільтрацію й перевірку на існування пропусків.

Для перетворення стовпців Ungrammatical Statement, Standard English у вектори використано метод CountVectorizer з бібліотеки sklearn.

Класифікація типів граматичних помилок у іноземній мові передбачала використання методів багатокласової класифікації Support Vector Classification, K Neighbors Classifier, Random Forest Classifier, де пошук гіперпараметрів здійснювався інструментом GridSearchCV. Розподіл даних на тренувальний і тестовий набори здійснювався у пропорції test_size=0.35 з перемішуванням, де shuffle=True.

Ефективність побудованих класифікаторів вивчалася методами точності (окремо для навчального й тестового наборів), матрицею плутанини, дисперсією й зміщенням.

Результати дослідження. Формально постановку завдання розробки класифікатора граматичних тестів можна описати як пошук функції (1):

$$f: X_i \rightarrow Y, \quad (1)$$

де X_i – досліджувані змінні щодо неграматичного твердження, стандартної англійської;
 Y – тип граматичної помилки.

Вихідним датасетом дослідження типів граматичних помилок у іноземній мові є Grammar Correction, що має наступну структуру, таблиця 1.

Відповідно до методології дослідження першочергово здійснено виявлення пропусків у рядках, котрі видалялися. Два пропуски даних двох рядків зафіксовані у типі граматичної помилки

Sentence Structure Errors, де були відсутні записи для Standard English. Тому досліджуваний датасет становив 595 записів типів граматичних помилок іноземної мови.

Векторизувавши стовпці Ungrammatical Statement, Standard English й присвоївши унікальні значення типу граматичної помилки дані підготовлено до використання у моделях машинного навчання, котрі першочергово вивчалися без налаштувань гіперпараметрів, таблиця 2.

Як видно з таблиці 2, всі досліджувані моделі мають ознаки перенавчання, про що свідчить різниця точності на навчальному й тестовому наборах. При цьому, точність моделі K Neighbors Classifier на тестовому наборі даних є низькою й становить 0,39, тому надалі розглядаємо лише моделі Support Vector Classification і Random Forest Classifier.

За результатами дослідження моделі Support Vector Classification за допомогою GridSearchCV найліпшими гіперпараметрами є {'C': 0.9, 'gamma': 0.1, 'kernel': 'rbf'}, що забезпечують точність на навчальному й тестовому наборах 0,59/0,66 відповідно. Для вказаної моделі це оптимальний результат, хоча спостерігається незначне перенавчання.

Як видно з таблиці 3, Support Vector Classification демонструє найліпші показники вірно класифікованих типів граматичних помилок для класів 2, 3, 4, що становить 0,8; 0,77; 0,63 відповідно. При цьому показники збалансованості сягають 0,8; 0,81; 0,72, що відповідає очікуванням.

Модель Random Forest Classifier демонструє найліпші гіперпараметри {'max_depth': 22, 'min_samples_leaf': 1, 'min_samples_split': 5, 'n_estimators': 17} при точності на навчальному/тестовому наборах даних 0,52/0,48 відповідно, що теж вказує на ознаки перенавчання. Проте це є найліпшим зафіксованим результатом для досліджуваного набору даних при дисперсії 2,054 й змі-

Таблиця 1

Вихідні дані дослідження граматичних помилок іноземної мови

Serial Number	Error Type	Ungrammatical Statement	Standard English
1	Verb Tense Errors	I goes to the store everyday.	I go to the store everyday.
2	Verb Tense Errors	They was playing soccer last night.	They were playing soccer last night.
...
597	Sentence Structure Errors	The company is celebrating it's 10th anniversary this year.	The company is celebrating its 10th anniversary this year.

Таблиця 2

Результати дослідження моделей класифікації граматичних помилок іноземної мови

№	Назва моделі класифікації	Точність навчального/тестового наборів даних
1	Support Vector Classification	1,0/0,55
2	K Neighbors Classifier	1,0/0,39
3	Random Forest Classifier	1,0/0,51

Таблиця 3

**Результати класифікації типів граматичних помилок в іноземній мові
за допомогою Support Vector Classification**

Досліджуваний тип граматичної помилки	Precision	Recall	F1-score	Support
0	0,56	0,48	0,51	21
1	0,52	0,6	0,56	20
2	0,8	0,8	0,8	20
3	0,77	0,85	0,81	20
4	0,63	0,85	0,72	20
5	0,7	0,37	0,48	19
Точність класифікатора 0,66				120

Таблиця 4

**Результати класифікації типів граматичних помилок в іноземній мові
за допомогою Random Forest Classifier**

Досліджуваний тип граматичної помилки	Precision	Recall	F1-score	Support
0	0,38	0,43	0,4	21
1	0,39	0,35	0,37	20
2	0,67	0,8	0,73	20
3	0,58	0,75	0,65	20
4	0,5	0,35	0,41	20
5	0,29	0,21	0,24	19
Точність класифікатора 0,48				120

щенні 2,664. Збільшення обсягу тестової вибірки породжує збільшення точності на навчальному наборі даних і зменшення на тестовому, таблиця 4.

Як видно з таблиці 4, Random Forest Classifier демонструє найліпші показники вірно класифікованих типів граматичних помилок для класів 2, 3, 4, що становить 0,67; 0,58; 0,5 відповідно. При цьому показники збалансованості сягають 0,73; 0,65; 0,41, що не у повній мірі відповідає очікуванням. Інші класи 0, 1, 5 продемонстрували значно нижчі результати класифікації, що пов'язано з низькою якістю вхідних даних.

Таким чином, за результатами діагностики моделей машинного навчання щодо класифікації типів граматичних помилок іноземної мови рекомендовано використовувати модель Support Vector Classification з гіперпараметрами {'C': 0.9, 'gamma': 0.1, 'kernel': 'rbf'}, котра продемонструвала точність 0,66. При цьому варто врахувати, що модель характеризується ознаками перенавчання.

З практичної точки зору вона може бути корисною для опанування азів англійської мови, зокрема при вивченні типу граматичних помилок Subject-

Verb Agreement, що відноситься до класу 2. Побудована модель також розглядається як основа для побудови точнішого класифікатора, за наявності нових даних типів граматичних помилок іноземної мови.

Висновки. Запропоновано педагогічний інструмент підвищення ефективності вивчення англійської мови. Завдання розробки класифікатора граматичних помилок іноземної мови вирішується шляхом використання ідей машинного навчання, зокрема задачі багатокласової класифікації.

Побудована модель класифікації Support Vector Classification з гіперпараметрами {'C': 0.9, 'gamma': 0.1, 'kernel': 'rbf'} демонструє точність класифікації на рівні 0,66. Вказана модель якісно класифікує граматичну помилку типу Subject-Verb Agreement на рівні точності 0,8. Враховуючи обмеження, модель запропоновано до використання для класифікації текстів на предмет аналізу граматичних помилок. З іншої сторони, життєвий цикл моделі передбачає оновлення типів граматичних помилок з метою збільшення обсягу набору даних.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Bannò S., Matassoni M. Back to grammar: Using grammatical error correction to automatically assess L2 speaking proficiency. *Speech Communication*. 2023. P. 103025. URL: <https://doi.org/10.1016/j.specom.2023.103025> (date of access: 15.03.2024).
2. Li B. An improved method for extracting essay tangency features in intelligent scoring of English essays. *Intelligent Systems with Applications*. 2024. P. 200328. URL: <https://doi.org/10.1016/j.iswa.2024.200328> (date of access: 15.03.2024).
3. Liu G., Wang Y. Modeling EFL teachers' intention to integrate informal digital learning of English (IDLE) into the classroom using the theory of planned behavior. *System*. 2023. P. 103193. URL: <https://doi.org/10.1016/j.system.2023.103193> (date of access: 15.03.2024).

4. Wilson J., Palermo C., Wibowo A. Elementary English learners' engagement with automated feedback. *Learning and Instruction*. 2024. Vol. 91. P. 101890. URL: <https://doi.org/10.1016/j.learninstruc.2024.101890> (date of access: 15.03.2024).
5. Practical principles of integrating artificial intelligence into the technology of regional security predicting / O. Shefer et al. *Advanced Information Systems*. 2024. Vol. 8, no. 1. P. 86–93. URL: <https://doi.org/10.20998/2522-9052.2024.1.11> (date of access: 15.03.2024).
6. O'Shaughnessy D. Understanding Automatic Speech Recognition. *Computer Speech & Language*. 2023. P. 101538. URL: <https://doi.org/10.1016/j.csl.2023.101538> (date of access: 15.03.2024).
7. Sino M., Tinnirello I., La Cascia M. Is text preprocessing still worth the time? A comparative survey on the influence of popular preprocessing methods on Transformers and traditional classifiers. *Information Systems*. 2024. Vol. 121. P. 102342. URL: <https://doi.org/10.1016/j.is.2023.102342> (date of access: 15.03.2024).
8. Rezaul Islam M., Ahmad A., Shahidur Rahman M. Bangla Text Normalization for Text-to-speech Synthesizer Using Machine Learning Algorithms. *Journal of King Saud University – Computer and Information Sciences*. 2023. P. 101807. URL: <https://doi.org/10.1016/j.jksuci.2023.101807> (date of access: 15.03.2024).
9. Olayiwola A., Olayiwola D., Oyedeji A. Development of an Automatic Grammar Checker for Yorùbá Word Processing using Government and Binding Theory. *Expert Systems with Applications*. 2023. P. 121351. URL: <https://doi.org/10.1016/j.eswa.2023.121351> (date of access: 15.03.2024).
10. Putra I. M. S., Siahaan D., Saikhu A. Recognizing textual entailment: A review of resources, approaches, applications, and challenges. *ICT Express*. 2023. URL: <https://doi.org/10.1016/j.ict.2023.08.012> (date of access: 15.03.2024).
11. Discriminating nonfluent/agrammatic and logopenic PPA variants with automatically extracted morphosyntactic measures from connected speech / S. Lukic et al. *Cortex*. 2024. URL: <https://doi.org/10.1016/j.cortex.2023.12.013> (date of access: 15.03.2024).
12. Enhanced variants of crow search algorithm boosted with cooperative based island model for global optimization / T. Thaher et al. *Expert Systems with Applications*. 2023. P. 121712. URL: <https://doi.org/10.1016/j.eswa.2023.121712> (date of access: 15.03.2024).
13. Investigation of combined ensemble methods for diagnostics of the quality of interaction of human-machine systems / O. Laktionov et al. *Naukovyi Visnyk Natsionalnoho Hirnychoho Universytetu*. 2023. No. 4. P. 138–143. URL: <https://doi.org/10.33271/nvngu/2023-4/138>
14. Grammar Correction. *Kaggle: Your Machine Learning and Data Science Community*. URL: <https://www.kaggle.com/datasets/satishgunjal/grammar-correction/data> (date of access: 15.03.2024).
15. Müller A. C., Guido S. Introduction to Machine Learning with Python: A Guide for Data Scientists. O'Reilly Media, Incorporated, 2018.

REFERENCES

1. Bannò S., Matassoni M. (2023) Back to grammar: Using grammatical error correction to automatically assess L2 speaking proficiency. *Speech Communication*. URL: <https://doi.org/10.1016/j.specom.2023.103025> (date of access: 15.03.2024).
2. Li B. (2024) An improved method for extracting essay tangency features in intelligent scoring of English essays. *Intelligent Systems with Applications*. URL: <https://doi.org/10.1016/j.iswa.2024.200328> (date of access: 15.03.2024).
3. Liu G., Wang Y. (2023) Modeling EFL teachers' intention to integrate informal digital learning of English (IDLE) into the classroom using the theory of planned behavior. *System*. URL: <https://doi.org/10.1016/j.system.2023.103193> (date of access: 15.03.2024).
4. Wilson J., Palermo C., Wibowo A. (2024) Elementary English learners' engagement with automated feedback. *Learning and Instruction*, vol. 91. URL: <https://doi.org/10.1016/j.learninstruc.2024.101890> (date of access: 15.03.2024).
5. Practical principles of integrating artificial intelligence into the technology of regional security predicting / O. Shefer et al. (2024) *Advanced Information Systems*, vol. 8, no. 1. 86–93. URL: <https://doi.org/10.20998/2522-9052.2024.1.11> (date of access: 15.03.2024).
6. O'Shaughnessy D. (2023) Understanding Automatic Speech Recognition. *Computer Speech & Language*. URL: <https://doi.org/10.1016/j.csl.2023.101538> (date of access: 15.03.2024).
7. Siino M., Tinnirello I., La Cascia M. (2024) Is text preprocessing still worth the time? A comparative survey on the influence of popular preprocessing methods on Transformers and traditional classifiers. *Information Systems*, vol. 121. URL: <https://doi.org/10.1016/j.is.2023.102342> (date of access: 15.03.2024).
8. Rezaul Islam M., Ahmad A., Shahidur Rahman M. (2023) Bangla Text Normalization for Text-to-speech Synthesizer Using Machine Learning Algorithms. *Journal of King Saud University – Computer and Information Sciences*. URL: <https://doi.org/10.1016/j.jksuci.2023.101807> (date of access: 15.03.2024).
9. Olayiwola A., Olayiwola D., Oyedeji A. (2023) Development of an Automatic Grammar Checker for Yorùbá Word Processing using Government and Binding Theory. *Expert Systems with Applications*. URL: <https://doi.org/10.1016/j.eswa.2023.121351> (date of access: 15.03.2024).
10. Putra I. M. S., Siahaan D., Saikhu A. (2023) Recognizing textual entailment: A review of resources, approaches, applications, and challenges. *ICT Express*. URL: <https://doi.org/10.1016/j.ict.2023.08.012> (date of access: 15.03.2024).
11. Discriminating nonfluent/agrammatic and logopenic PPA variants with automatically extracted morphosyntactic measures from connected speech / S. Lukic et al. (2024) *Cortex*. URL: <https://doi.org/10.1016/j.cortex.2023.12.013> (date of access: 15.03.2024).
12. Enhanced variants of crow search algorithm boosted with cooperative based island model for global optimization / T. Thaher et al. (2023) *Expert Systems with Applications*. URL: <https://doi.org/10.1016/j.eswa.2023.121712> (date of access: 15.03.2024).
13. Investigation of combined ensemble methods for diagnostics of the quality of interaction of human-machine systems / O. Laktionov et al. (2023) *Naukovyi Visnyk Natsionalnoho Hirnychoho Universytetu*, 4. 138–143. URL: <https://doi.org/10.33271/nvngu/2023-4/138>
14. Grammar Correction. *Kaggle: Your Machine Learning and Data Science Community*. URL: <https://www.kaggle.com/datasets/satishgunjal/grammar-correction/data> (date of access: 15.03.2024).
15. Müller A. C., Guido S. (2018) Introduction to Machine Learning with Python: A Guide for Data Scientists. O'Reilly Media, Incorporated.