

ISSN 2518-1726 (Online),
ISSN 1991-346X (Print)



«ҚАЗАҚСТАН РЕСПУБЛИКАСЫ
ҰЛТТЫҚ ҒЫЛЫМ АКАДЕМИЯСЫ» РҚБ

Х А Б А Р Л А Р Ы

ИЗВЕСТИЯ

РОО «НАЦИОНАЛЬНОЙ
АКАДЕМИИ НАУК РЕСПУБЛИКИ
КАЗАХСТАН»

N E W S

OF THE ACADEMY OF SCIENCES
OF THE REPUBLIC OF
KAZAKHSTAN

PHYSICO-MATHEMATICAL SERIES

4 (352)

OCTOBER – DECEMBER 2024

PUBLISHED SINCE JANUARY 1963

PUBLISHED 4 TIMES A YEAR

ALMATY, NAS RK

БАС РЕДАКТОР:

МУТАНОВ Ғалымқайыр Мұтанұлы, техника ғылымдарының докторы, профессор, ҚР ҰҒА академигі, ҚР БҒМ ҒК «Ақпараттық және есептеу технологиялары институты» бас директорының м.а. (Алматы, Қазақстан), **Н=5**

БАС РЕДАКТОРДЫҢ ОРЫНБАСАРЫ:

МАМЫРБАЕВ Өркен Жұмажанұлы, ақпараттық жүйелер мамандығы бойынша философия докторы (Ph.D), ҚР БҒМ Ғылым комитеті «Ақпараттық және есептеуші технологиялар институты» РМК жауапты хатшысы (Алматы, Қазақстан), **Н=5**

РЕДАКЦИЯ АЛҚАСЫ:

ҚАЛИМОЛДАЕВ Мақсат Нұрәділұлы, физика-математика ғылымдарының докторы, профессор, ҚР ҰҒА академигі (Алматы, Қазақстан), **Н=7**

БАЙГУНЧЕКОВ Жұмаділ Жанабайұлы, техника ғылымдарының докторы, профессор, ҚР ҰҒА академигі, Кибернетика және ақпараттық технологиялар институты, Сатпаев университетінің Қолданбалы механика және инженерлік графика кафедрасы, (Алматы, Қазақстан), **Н=3**

ВОЙЧИК Вальдемар, техника ғылымдарының докторы (физика), Люблин технологиялық университетінің профессоры (Люблин, Польша), **Н=23**

БОШКАЕВ Қуантай Авғазыұлы, Ph.D. Теориялық және ядролық физика кафедрасының доценті, әл-Фараби атындағы Қазақ ұлттық университеті (Алматы, Қазақстан), **Н=10**

QUEVEDO Nemando, профессор, Ядролық ғылымдар институты (Мехико, Мексика), **Н=28**

ЖҮСІПОВ Марат Абжанұлы, физика-математика ғылымдарының докторы, теориялық және ядролық физика кафедрасының профессоры, әл-Фараби атындағы Қазақ ұлттық университеті (Алматы, Қазақстан), **Н=7**

КОВАЛЕВ Александр Михайлович, физика-математика ғылымдарының докторы, Украина ҰҒА академигі, Қолданбалы математика және механика институты (Донецк, Украина), **Н=5**

РАМАЗАНОВ Тілекқабұл Сәбитұлы, физика-математика ғылымдарының докторы, профессор, ҚР ҰҒА академигі, әл-Фараби атындағы Қазақ ұлттық университетінің ғылыми-инновациялық қызмет жөніндегі проректоры, (Алматы, Қазақстан), **Н=26**

ТАКИБАЕВ Нұрғали Жабағаұлы, физика-математика ғылымдарының докторы, профессор, ҚР ҰҒА академигі, әл-Фараби атындағы Қазақ ұлттық университеті (Алматы, Қазақстан), **Н=5**

ТИГИНЯНУ Ион Михайлович, физика-математика ғылымдарының докторы, академик, Молдова Ғылым Академиясының президенті, Молдова техникалық университеті (Кишинев, Молдова), **Н=42**

ХАРИН Станислав Николаевич, физика-математика ғылымдарының докторы, профессор, ҚР ҰҒА академигі, Қазақстан-Британ техникалық университеті (Алматы, Қазақстан), **Н=10**

ДАВЛЕТОВ Асқар Ербуланович, физика-математика ғылымдарының докторы, профессор, әл-Фараби атындағы Қазақ ұлттық университеті (Алматы, Қазақстан), **Н=12**

КАЛАНДРА Пьетро, Ph.D (физика), Нанокұрылымды материалдарды зерттеу институтының профессоры (Рим, Италия), **Н=26**

«ҚР ҰҒА Хабарлары. Физика және информатика сериясы».

ISSN 2518-1726 (Online),

ISSN 1991-346X (Print)

Меншіктеуші: «Қазақстан Республикасының Ұлттық ғылым академиясы» РҚБ (Алматы қ.). Қазақстан Республикасының Ақпарат және қоғамдық даму министрлігінің Ақпарат комитетінде 14.02.2018 ж. берілген **№ 16906-Ж** мерзімдік басылым тіркеуіне қойылу туралы куәлік.

Тақырыптық бағыты: *физика және ақпараттық коммуникациялық технологиялар сериясы*. Қазіргі уақытта: *«ақпараттық технологиялар» бағыты бойынша ҚР БҒМ БҒСБК ұсынған журналдар тізіміне енді.*

Мерзімділігі: *жылына 4 рет.*

Тиражы: *300 дана.*

Редакцияның мекен-жайы: *050010, Алматы қ., Шевченко көш., 28, 219 бөл., тел.: 272-13-19*
<http://www.physico-mathematical.kz/index.php/en/>

ГЛАВНЫЙ РЕДАКТОР:

МУТАНОВ Галимкаир Мутанович, доктор технических наук, профессор, академик НАН РК, и.о. генерального директора «Института информационных и вычислительных технологий» КН МОН РК (Алматы, Казахстан), **H=5**

ЗАМЕСТИТЕЛЬ ГЛАВНОГО РЕДАКТОРА:

МАМЫРБАЕВ Оркен Жумажанович, доктор философии (PhD) по специальности Информационные системы, ответственный секретарь РГП «Института информационных и вычислительных технологий» Комитета науки МОН РК (Алматы, Казахстан), **H=5**

РЕДАКЦИОННАЯ КОЛЛЕГИЯ:

КАЛИМОЛДАЕВ Максат Нурадилович, доктор физико-математических наук, профессор, академик НАН РК (Алматы, Казахстан), **H=7**

БАЙГУНЧЕКОВ Жумадил Жанабаевич, доктор технических наук, профессор, академик НАН РК, Институт кибернетики и информационных технологий, кафедра прикладной механики и инженерной графики, Университет Сагпаева (Алматы, Казахстан), **H=3**

ВОЙЧИК Вальдемар, доктор технических наук (физ.-мат.), профессор Люблинского технологического университета (Люблин, Польша), **H=23**

БОШКАЕВ Куантай Авгазыевич, доктор Ph.D, преподаватель, доцент кафедры теоретической и ядерной физики, Казахский национальный университет им. аль-Фараби (Алматы, Казахстан), **H=10**

QUEVEDO Hemando, профессор, Национальный автономный университет Мексики (UNAM), Институт ядерных наук (Мехико, Мексика), **H=28**

ЖУСУПОВ Марат Абжанович, доктор физико-математических наук, профессор кафедры теоретической и ядерной физики, Казахский национальный университет им. аль-Фараби (Алматы, Казахстан), **H=7**

КОВАЛЕВ Александр Михайлович, доктор физико-математических наук, академик НАН Украины, Институт прикладной математики и механики (Донецк, Украина), **H=5**

РАМАЗАНОВ Тлексабул Сабитович, доктор физико-математических наук, профессор, академик НАН РК, проректор по научно-инновационной деятельности, Казахский национальный университет им. аль-Фараби (Алматы, Казахстан), **H=26**

ТАКИБАЕВ Нургали Жабатаевич, доктор физико-математических наук, профессор, академик НАН РК, Казахский национальный университет им. аль-Фараби (Алматы, Казахстан), **H=5**

ТИГИНЯНУ Ион Михайлович, доктор физико-математических наук, академик, президент Академии наук Молдовы, Технический университет Молдовы (Кишинев, Молдова), **H=42**

ХАРИН Станислав Николаевич, доктор физико-математических наук, профессор, академик НАН РК, Казахстанско-Британский технический университет (Алматы, Казахстан), **H=10**

ДАВЛЕТОВ Аскар Ербуланович, доктор физико-математических наук, профессор, Казахский национальный университет им. аль-Фараби (Алматы, Казахстан), **H=12**

КАЛАНДРА Пьетро, доктор философии (Ph.D, физика), профессор Института по изучению наноструктурированных материалов (Рим, Италия), **H=26**

«Известия НАН РК. Серия физика и информатики».

ISSN 2518-1726 (Online),

ISSN 1991-346X (Print)

Собственник: *Республиканское общественное объединение «Национальная академия наук Республики Казахстан» (г. Алматы).*

Свидетельство о постановке на учет периодического печатного издания в Комитете информации Министерства информации и общественного развития Республики Казахстан **№ 16906-Ж** выданное 14.02.2018 г.

Тематическая направленность: *серия физика и информационные коммуникационные технологии.* В настоящее время: *вошел в список журналов, рекомендованных ККСОН МОН РК по направлению «информационные коммуникационные технологии».*

Периодичность: *4 раз в год.*

Тираж: *300 экземпляров.*

Адрес редакции: *050010, г. Алматы, ул. Шевченко, 28, оф. 219, тел.: 272-13-19*

<http://www.physico-mathematical.kz/index.php/en/>

EDITOR IN CHIEF:

MUTANOV Galimkair Mutanovich, doctor of technical Sciences, Professor, Academician of NAS RK, acting director of the Institute of Information and Computing Technologies of SC MES RK (Almaty, Kazakhstan), **H=5**

DEPUTY EDITOR-IN-CHIEF

MAMYRBAYEV Orken Zhumazhanovich, Ph.D. in the specialty "Information systems, executive secretary of the RSE "Institute of Information and Computational Technologies", Committee of Science MES RK (Almaty, Kazakhstan) **H=5**

EDITORIAL BOARD:

KALIMOLDAYEV Maksat Nuradilovich, doctor in Physics and Mathematics, Professor, Academician of NAS RK (Almaty, Kazakhstan), **H=7**

BAYGUNCHEKOV Zhumadil Zhanabayevich, doctor of Technical Sciences, Professor, Academician of NAS RK, Institute of Cybernetics and Information Technologies, Department of Applied Mechanics and Engineering Graphics, Satbayev University (Almaty, Kazakhstan), **H=3**

WOICIK Waldemar, Doctor of Phys.-Math. Sciences, Professor, Lublin University of Technology (Lublin, Poland), **H=23**

BOSHKAYEV Kuantai Avgazievich, PhD, Lecturer, Associate Professor of the Department of Theoretical and Nuclear Physics, Al-Farabi Kazakh National University (Almaty, Kazakhstan), **H=10**

QUEVEDO Hemando, Professor, National Autonomous University of Mexico (UNAM), Institute of Nuclear Sciences (Mexico City, Mexico), **H=28**

ZHUSSUPOV Marat Abzhanovich, Doctor in Physics and Mathematics, Professor of the Department of Theoretical and Nuclear Physics, al-Farabi Kazakh National University (Almaty, Kazakhstan), **H=7**

KOVALEV Alexander Mikhailovich, Doctor in Physics and Mathematics, Academician of NAS of Ukraine, Director of the State Institution «Institute of Applied Mathematics and Mechanics» DPR (Donetsk, Ukraine), **H=5**

RAMAZANOV Tlekkabul Sabitovich, Doctor in Physics and Mathematics, Professor, Academician of NAS RK, Vice-Rector for Scientific and Innovative Activity, al-Farabi Kazakh National University (Almaty, Kazakhstan), **H=26**

TAKIBAYEV Nurgali Zhabagaevich, Doctor in Physics and Mathematics, Professor, Academician of NAS RK, al-Farabi Kazakh National University (Almaty, Kazakhstan), **H=5**

TIGHINEANU Ion Mikhailovich, Doctor in Physics and Mathematics, Academician, Full Member of the Academy of Sciences of Moldova, President of the AS of Moldova, Technical University of Moldova (Chisinau, Moldova), **H=42**

KHARIN Stanislav Nikolayevich, Doctor in Physics and Mathematics, Professor, Academician of NAS RK, Kazakh-British Technical University (Almaty, Kazakhstan), **H=10**

DAVLETOV Askar Erbulanovich, Doctor in Physics and Mathematics, Professor, al-Farabi Kazakh National University (Almaty, Kazakhstan), **H=12**

CALANDRA Pietro, PhD in Physics, Professor at the Institute of Nanostructured Materials (Monterotondo Station Rome, Italy), **H=26**

News of the National Academy of Sciences of the Republic of Kazakhstan.

Series of physics and informatics.

ISSN 2518-1726 (Online),

ISSN 1991-346X (Print)

Owner: RPA «National Academy of Sciences of the Republic of Kazakhstan» (Almaty). The certificate of registration of a periodical printed publication in the Committee of information of the Ministry of Information and Social Development of the Republic of Kazakhstan **No. 16906-ЖК**, issued 14.02.2018
Thematic scope: *series physics and information technology.*

Currently: *included in the list of journals recommended by the CCSES MES RK in the direction of «information and communication technologies».*

Periodicity: *4 times a year.*

Circulation: *300 copies.*

Editorial address: *28, Shevchenko str., of. 219, Almaty, 050010, tel. 272-13-19*

<http://www.physico-mathematical.kz/index.php/en/>

NEWS OF THE NATIONAL ACADEMY OF SCIENCES OF THE REPUBLIC OF KAZAKHSTAN
PHYSICO-MATHEMATICAL SERIES

ISSN 1991-346X

Volume 4. Number 352 (2024). 73–88

<https://doi.org/10.32014/2024.2518-1726.308>

ӨЖ 004.931

©A.S. Baegizova¹, G.I. Mukhamedrakhimova^{1*}, I. Bapiyev²,
M.Zh. Bazarova³, U.M. Smailova⁴, 2024.

¹L.N. Gumilyov Eurasian National University, Astana, Kazakhstan;

²Zhangir khan West Kazakhstan Agrarian-Technical University, Uralsk, Kazakhstan;

³Sarsen Amanzholov East Kazakhstan University, Ust-Kamenogorsk, Kazakhstan;

⁴Center of Excellence AEO «Nazarbayev Intellectual Schools», Astana, Kazakhstan.

E-mail: isatai-07@mail.ru

EVALUATING THE EFFECTIVENESS OF MACHINE LEARNING METHODS FOR KEYWORD COVERAGE

Baegizova Aigulim – senior lecturer at the Department of Radio Engineering, Electronics and Telecommunications, Eurasian National University named after L.N. Gumilyov, Astana, Kazakhstan, E-mail: baegiz_a@mail.ru, <https://orcid.org/0000-0003-2293-2143>;

Mukhamedrakhimova Galiya – senior lecturer, Department of Radio Engineering, Electronics and Telecommunications, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan, E-mail: isatai-07@mail.ru, <https://orcid.org/0000-0002-9951-6263>;

Bapiyev Ideyat – «Zhangir khan West Kazakhstan agrarian Technical University», associate professor, doctor of philosophy (Ph.D), Uralsk, Kazakhstan. E-mail: bapiyev@mail.ru, <https://orcid.org/0000-0001-8468-8938>;

Bazarova Madina - Sarsen Amanzholov East Kazakhstan university, associate professor of the Department of computer modeling and information technology, PhD, Ust-Kamenogorsk, Kazakhstan. E-mail: madina_vkgtu@mail.ru, <https://orcid.org/0000-0003-2580-6580>;

Smailova Ulmeken - Center of Excellence AEO «Nazarbayev Intellectual Schools», Astana, Kazakhstan samilova_tarsu@mail.ru.

Abstract. This paper provides a thorough comparative analysis of two modern hybrid machine learning approaches, namely, Bidirectional Encoder Representations from Transformers (BERT) combined with an autoencoder (AE) and Term Frequency-Inverse Document Frequency (TF-IDF) combined with an autoencoder. The study focuses on the task of keyword extraction using semantic analysis methods of text data. The main goal of the work is to evaluate the effectiveness of these methods in ensuring adequate keyword coverage in large text corpora covering various subject areas. The authors study in detail the architecture and operating principles of each of the considered methods. Particular attention is paid to the features of integrating these methods with autoencoders, which allows to significantly improve the semantic integrity and relevance of the extracted keywords. The experimental part of the study includes a detailed analysis of the

performance of both methods on various text datasets, demonstrating how the structure and semantic richness of the original data affect the performance of each method. The paper also describes in detail the applied methodology for assessing the quality of keyword extraction, including such metrics as precision, recall, and F1 score. The advantages and disadvantages of each approach, as well as their suitability for specific types of text tasks, are analyzed. The results of the study provide valuable data for the scientific community and can be used to select the most appropriate text processing method in various applications that require a deep understanding of semantic content and high accuracy of information extraction.

Key words: Machine learning, keywords, semantic analysis, BERT, Autoencoder, TF-IDF, hybrid approaches, information extraction.

©**А.С. Баегизова¹, Г.И. Мухамедрахимова^{1*}, И.М. Бапиев²,
М.Ж. Базарова³, У.М. Смайлова⁴, 2024.**

¹Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана, Қазақстан;

²Жәңгір хан атындағы Батыс Қазақстан аграрлық-техникалық университеті,
Орал, Қазақстан;

³С. Аманжолов атындағы Шығыс Қазақстан университеті, Өскемен, Қазақстан;

⁴«Назарбаев зияткерлік мектептері» ДББҰ Педагогикалық шеберлік
орталығы, Астана, Қазақстан.

E-mail: isatai-07@mail.ru

ТҮЙІН СӨЗДЕРДІ ҚАМТУ ҮШІН МАШИНАЛЫҚ ОҚЫТУ ӘДІСТЕРІНІҢ ТИІМДІЛІГІН БАҒАЛАУ

Баегизова Айгулим Сейсенбековна – Л.Н. Гумилев атындағы Еуразия ұлттық университетінің Радиотехника, электроника және телекоммуникация кафедрасының аға оқытушысы, Астана, Қазақстан, E-mail: baegiz_a@mail.ru, <https://orcid.org/0000-0003-2293-2143>;

Мухамедрахимова Галия Исатаевна – Л.Н. Гумилев атындағы Еуразия ұлттық университетінің Радиотехника, электроника және телекоммуникация кафедрасының аға оқытушысы, Астана, Қазақстан, E-mail: isatai-07@mail.ru, <https://orcid.org/0000-0002-9951-6263>;

Бапиев Идеят Мэлсович – «Жәңгір хан атындағы Батыс Қазақстан аграрлық- техникалық университеті», доцент м.а., философия докторы (Ph.D), Орал, Қазақстан. E-mail: bapiev@mail.ru, <https://orcid.org/0000-0001-8468-8938>;

Базарова Мадина Жомартовна – С. Аманжолов атындағы Шығыс Қазақстан университеті, «Компьютерлік үлгілеу және ақпараттық технологиялар» кафедрасының қауымдастырылған профессоры, PhD, Өскемен, Қазақстан. E-mail: madina_vkgtu@mail.ru, <https://orcid.org/0000-0003-2580-658>;

Смайлова Улмекен Мухитовна – «Назарбаев зияткерлік мектептері» ДББҰ Педагогикалық шеберлік орталығы, ф.-м.ғ.к., доцент, Астана, Қазақстан. E-mail: samilova_tarsu@mail.ru, <https://orcid.org/0009-0003-7696-4615>.

Аннотация. Бұл мақалада автоматты кодтаушымен (AE) біріктірілген Bidirectional Encoder Representations from Transformers (BERT)) және Term Frequency-Inverse Document Frequency, TF-IDF сияқты машиналық оқытудағы екі заманауи гибридтік тәсілдердің толық салыстырмалы талдауы берілген.

Зерттеу мәтіндік деректерге семантикалық талдау әдістерін қолдана отырып, кілт сөздерді шығару міндетіне назар аударады. Жұмыстың негізгі мақсаты – әртүрлі тақырыптық аумақтарды қамтитын үлкен мәтіндерге сәйкес кілт сөздерді табуды қамтамасыз етуде осы әдістердің тиімділігін бағалау. Авторлар қарастырылатын әдістердің әрқайсысының архитектурасы мен жұмыс істеу принциптерін толық зерттейді. Бұл әдістерді таңдалған түйінді сөздердің мағыналық тұтастығы мен өзектілігін айтарлықтай жақсартуға мүмкіндік беретін автокодерлермен біріктіру ерекшеліктеріне ерекше назар аударылады. Зерттеудің эксперименттік бөлімі бастапқы деректердің құрылымы мен семантикалық байлығының әрбір әдістің нәтижелеріне қалай әсер ететінін көрсететін әртүрлі мәтіндік деректер жиындарында екі әдістің де тиімділігін егжей-тегжейлі талдауды қамтиды. Сондай-ақ мақалада дәлдік, F1 сияқты көрсеткіштерді қоса, кілт сөзді шығару сапасын бағалау үшін қолданылатын әдістеме егжей-тегжейлі сипатталған. Әрбір тәсілдің артықшылықтары мен кемшіліктері, сондай-ақ олардың сөздік есептердің нақты түрлеріне сәйкестігі талданады. Зерттеу нәтижелері ғылыми қоғамдастық үшін құнды деректер береді және семантикалық мазмұнды терең түсіну және ақпаратты алудың жоғары дәлдігі қажет болатын әртүрлі қолданбаларда мәтінді өңдеудің ең қолайлы әдісін таңдау үшін пайдаланылуы мүмкін.

Түйін сөздер: Машиналық оқыту, кілт сөздер, семантикалық талдау, BERT, автокодер, TF-IDF, гибриді тәсілдер, ақпаратты алу.

©А.С. Баегизова¹, Г.И. Мухамедрахимова^{1*}, И.М. Бапиев²,
М.Ж. Базарова³, У.М. Смайлова⁴, 2024.

¹Евразийский национальный университет имени Л.Н. Гумилева,
Астана, Казахстан;

²Западно-Казахстанский аграрно-технический университет им. Жангир хана,
Уральск, Казахстан;

³Восточно-Казахстанский университет имени С. Аманжолова,
Усть-Каменогорск, Казахстан;

⁴Центр педагогического мастерства АОО «Назарбаев Интеллектуальные
школы», Астана, Казахстан.

E-mail: isatai-07@mail.ru

ОЦЕНКА ЭФФЕКТИВНОСТИ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ОХВАТА КЛЮЧЕВЫХ СЛОВ

Баегизова Айгулим Сейсенбековна – старший преподаватель кафедры Радиотехники, электроники и телекоммуникаций Евразийского национального университета имени Л.Н. Гумилева, Астана, Казахстан, E-mail: baegiz_a@mail.ru, <https://orcid.org/0000-0003-2293-2143>;

Мухамедрахимова Галия Исатаевна – старший преподаватель кафедры Радиотехники, электроники и телекоммуникаций Евразийского национального университета имени Л.Н. Гумилева, Астана, Казахстан, E-mail: isatai-07@mail.ru, <https://orcid.org/0000-0002-9951-6263>;

Бапиев Идеят Мэлсович – Западно-Казахстанский аграрно-технический университет имени

Жангир хана, и.о. доцента, PhD, Уральск, Казахстан, E-mail: bapiev@mail.ru, <https://orcid.org/0000-0001-8468-8938>;

Базарова Мадина Жомартовна – Восточно-Казахстанский университет имени С. Аманжолова, ассоциированный профессор кафедры «Компьютерное моделирование и информационные технологии», PhD, Усть-Каменогорск, Казахстан, E-mail: madina_vkgtu@mail.ru, <https://orcid.org/0000-0003-2580-6580>;

Смайлова Улмекен Мухитовна – Центр педагогического мастерства АОО «Назарбаев Интеллектуальные школы», Астана, Казахстан, E-mail: samilova_tarsu@mail.ru, <https://orcid.org/0009-0003-7696-4615>.

Аннотация. В данной статье осуществляется тщательный сравнительный анализ двух современных гибридных подходов в машинном обучении, такие как Bidirectional Encoder Representations from Transformers (BERT) в сочетании с автокодировщиком (Autoencoder, AE) и Термино-Частотное Обратное Документное Частотное (Term Frequency-Inverse Document Frequency, TF-IDF) в сочетании с автокодировщиком. Исследование фокусируется на задаче извлечения ключевых слов с применением методов семантического анализа текстовых данных. Основная цель работы заключается в оценке эффективности данных методов для обеспечения адекватного охвата ключевых слов в больших текстовых корпусах, охватывающих различные тематические области. Авторы подробно изучают архитектуру и принципы работы каждого из рассматриваемых методов. Особое внимание уделяется особенностям интеграции этих методов с автоэнкодерами, что позволяет значительно улучшить семантическую целостность и релевантность выделенных ключевых слов. Экспериментальная часть исследования включает в себя детальный анализ эффективности обоих методов на различных наборах текстовых данных, демонстрируя, как структура и семантическая насыщенность исходных данных влияют на результаты работы каждого из методов. В работе также подробно описывается примененная методология оценки качества извлечения ключевых слов, включая такие показатели, как точность, полнота и мера F1. Анализируются преимущества и недостатки каждого подхода, а также их пригодность для конкретных типов текстовых задач. Результаты исследования предоставляют ценные данные для научного сообщества и могут быть использованы для выбора наиболее подходящего метода обработки текстов в различных приложениях, где требуется глубокое понимание семантического содержания и высокая точность извлечения информации.

Ключевые слова: машинное обучение, ключевые слова, семантический анализ, BERT, Autoencoder, TF-IDF, гибридные подходы, извлечение информации.

Кіріспе

Қазіргі мәтінді өңдеуде (Садирмекова, 2023; Поло-Бланко, et al, 2024) кілт сөздерді шығару міндеті (Сузуки, et al, 2023; Тивари, et al, 2023) барған сайын маңызды бола бастады, ақпаратты ұйымдастыруда, іздеуде және талдауда шешуші рөл атқарады (Хунг, et. al, 2023; Шай, et al, 2023). Тиімді кілт сөздерді

шығару (Чигбу, et al, 2023; Хикман, et al, 2022) іздеу жүйелерінің, ұсыныстар жүйелерінің, сондай-ақ аналитикалық және білім беру құралдарының өнімділігін арттырады. Сондықтан осы процесті автоматтандыруға және оңтайландыруға қабілетті машиналық оқыту әдістерін әзірлеу және салыстыру өзекті зерттеу міндеті болып табылады. Бұл жұмыс екі гибриді тәсілдің терең талдауын жүргізеді: Трансформаторлардан Bidirectional Encoder Representations from Transformers (BERT) + Автокодер (AE) және Term Frequency-Inverse Document Frequency (TF-IDF) + автокодер. Екі әдіс те мәтіндік деректердің үлкен көлемін өңдеу үшін машиналық оқытудағы (Хассани, et al, 2020) және табиғи тілді өңдеудегі заманауи жетістіктерді пайдаланады (Таха, et al, 2023). Google әзірлеген BERT моделі (Мюррей, et al, 2021) мәтіндегі сөздердің семантикасын түсінудің озық технологиясын білдіреді, ал TF-IDF (Байер, et al, 2023) сөздің маңыздылығын бағалаудың дәстүрлі статистикалық әдісі болып табылады.

Біздің зерттеуіміздің мақсаты - әртүрлі мәтіндерден кілт сөздерді жеткілікті түрде қамту қабілетін бағалау үшін кілт сөздерді шығарудағы осы екі әдісті салыстыру. Біз әрбір әдістің архитектурасын, жұмыс істеу принциптерін және автокодерлермен біріктіру мүмкіндіктерін талдаймыз және бірнеше мәтіндік деректер жиынында эксперименттік салыстырулар жүргіземіз. Бұл бастапқы деректердің құрылымы мен семантикасына байланысты олардың тиімділігіндегі айырмашылықтарды анықтауға мүмкіндік береді. Дәлдік, F1 өлшемін қоса алғанда, кілт сөздерді шығару сапасын бағалау әдістемесіне ерекше назар аударылады. Біздің зерттеуіміздің нәтижелері ғылыми қауымдастыққа маңызды ақпарат береді және терең семантикалық талдауды және мәтінді өңдеуде жоғары дәлдікті қажет ететін нақты қолданбалар үшін ең қолайлы әдісті таңдауға көмектеседі.

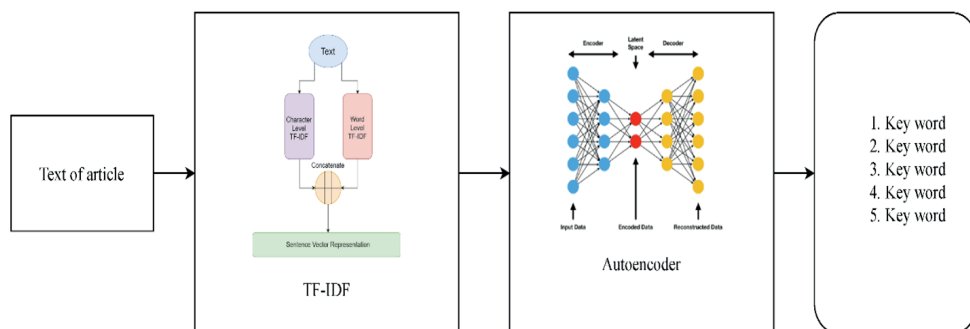
Әдістер мен материалдар.

Бұл зерттеуде біз екі гибриді машиналық оқыту әдісіне назар аудардық: BERT+Autoencoder және TF-IDF+Autoencoder. Олардың семантикалық деректерді іздеу арқылы үлкен мәтіндік корпустан кілт сөздерді шығару тапсырмасындағы тиімділігін бағалау. Екі әдіс те терең оқыту және автокодерлер концепцияларын біріктіреді (Тапех, et al, 2023; Бенитес-Андрасес, et al, 2022). Бұл мәтіндерді беткі деңгейде талдауға ғана емес, сонымен қатар сөздер мен сөз тіркестерінің арасындағы терең семантикалық және контекстік қатынастарға енуге мүмкіндік береді және кілт сөздерді шығару процесін байытады.

Біздің зерттеу жұмысымыздың алғашқы қадамы TF-IDF әдісін қолдану болды, бұл құжаттағы әрбір сөздің бүкіл мәтіндік корпусқа қатысты маңыздылығын бағалауға мүмкіндік береді. TF-IDF сөз салмағын олардың құжаттағы жиілігі және олар орын алған корпустағы құжаттардың кері жиілігі негізінде есептейді. Бұл тәсіл белгілі бір мәтіндерге ғана тән сөздерді бөлектеу арқылы жиі қолданылады, бірақ ақпаратсыз сөздердің әсерін

азайтуға көмектеседі. Алынған сөздердің векторлық көріністері автокодер өңдеудің келесі кезеңі үшін кіріс деректері ретінде қызмет етті. Бұл жұмыста қолданылатын автокодер екі негізгі компоненттен тұрады: кодтаушы және дешифратор. Кодер TF-IDF арқылы алынған мәтіннің векторлық көрінісін тығыздық және ақпаратты ішкі көрініске қысады. Содан кейін дешифратор ақпараттың жоғалуын азайтуға тырысып, осы сығылған көріністен бастапқы векторды қайта құру үшін жұмыс істейді. Бұл кезеңнің мақсаты - алынған кілт сөздердің сапасы мен дәлдігін жақсартатын бастапқы мәтіннен ең маңызды семантикалық мүмкіндіктерді шығарып, сақтай алатындай модельді үйрету.

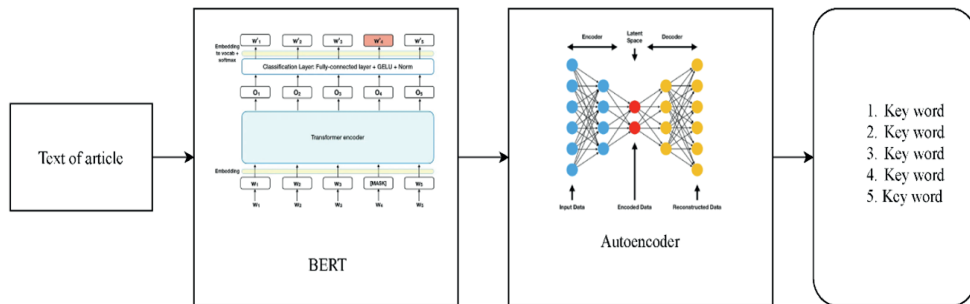
Автокодерлерді TF-IDF комбинациясында қолдану таңдалған түйінді сөздердің мағыналық тұтастығын айтарлықтай жақсартты. Бұл әдіс мәтіндік деректерді тереңірек талдауды, жасырын байланыстарды және сөздердің мағынасын кең мағынада ашуды қамтамасыз етеді. Осылайша, осы тәсілдерді біріктіру мәтінді іздеудің неғұрлым қуатты және дәл құралдарын жасауға ықпал етеді. Бұл кілт сөздерді алуды автоматтандыруға және ақпаратты іздеу мен үлкен деректерді өңдеуді жақсартуға маңызды қадам болып табылады (Сур. 1).



Сур. 1. TF-IDF + Автокодер гибриді әдісінің архитектурасы
(Fig. 1. Architecture of the hybrid TF-IDF + Autoencoder method)

Мәтіндік деректерді талдауды жақсарту және кілт сөздерді дәлірек шығару үшін бұл зерттеу барысында BERT үлгісін пайдаланылды. BERT – мәтіндерді талдау үшін трансформаторлық назар аудару механизмдерін қолданатын табиғи тілді өңдеу саласындағы алдыңғы қатарлы әдістердің бірі. BERT ерекшелігі оның мәтінді екі жақты өңдеу мүмкіндігі болып табылады, бұл модельге солдан оңға да, оңнан солға да сөздердің мағынасын бір уақытта талдауға мүмкіндік береді. Бұл екі жақты мәтінді түсіну модельдің сөздер арасындағы семантикалық қарым-қатынастарды түсіру қабілетін айтарлықтай жақсартады, контекстік нюанстарды терең түсіруді қамтамасыз етеді және кілт сөзді шығару сапасын жақсартады. BERT көмегімен мәтінді өңдегеннен кейін келесі қадам нәтижесінде векторлық көріністерді қысу үшін автокодерді пайдалану болды. Кодер мен дешифратордан тұратын автокодерлер бастапқы

вектордан қысылған бейнелеуге және кері ауысқанда ақпараттың жоғалуын азайту принципі бойынша жұмыс істейді. Біздің жағдайда кодтаушы көп өлшемді BERT векторларын тығызырақ векторларға сығымдады, содан кейін декодер оларды қайта құруға әрекет жасады. Бұл процедураның мақсаты мәтіндегі ең маңызды ақпаратты бөлектеу және сақтау, осылайша кілт сөздерді шығаруда жоғары дәлдік пен сапаны қамтамасыз ету болды. Бұл қысу деректердің үлкен көлемін басқару және талдауды жақсартуға көмектесті, кейінірек әртүрлі ақпаратты өңдеу қолданбаларында пайдалану үшін мәтіннің ең маңызды элементтерін анықтады (Сур. 2).



Сур. 2. BERT + Autoencoder гибриді әдісінің архитектурасы
(Fig. 2. Architecture of the hybrid BERT + Autoencoder method)

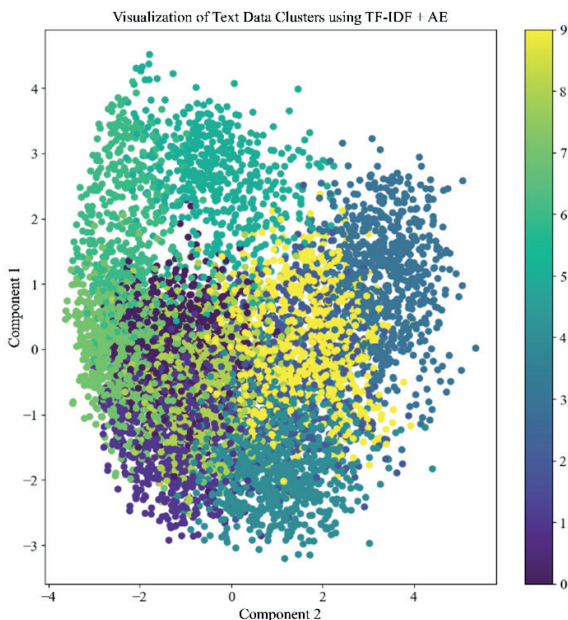
Біздің зерттеуімізде автокодерлермен біріктірілген BERT және TF-IDF әдістерін пайдалану кілт сөздерді шығару үшін мәтінмәндік өңдеудің маңыздылығын толық түсінуге мүмкіндік берді. BERT беретін мағынаны екі жақты түсіну ерекше маңызды болды. Бұл мәтіннің реттілігін талдауға ғана емес, сонымен қатар мәтінді қабылдауды айтарлықтай байытатын көптеген контекстік тәуелділіктерді ескеруге мүмкіндік береді. Бұл технологияларды ақпаратты тиімді қысатын автокодерлермен біріктіру бұл әсерді күшейтеді, ең маңызды семантикалық атрибуттарды ерекшелейді және алынған кілт сөздердің сапасын жақсартады. Эксперименттік нәтижелер әдістердің мұндай кешенді қолданылуы кілт сөздерді анықтау дәлдігін арттырып қана қоймай, мәтіндерді тереңірек талдауға ықпал ететінін растады. Ақпаратты қысу үшін автокодерлерді пайдалану деректердің шамадан тыс жүктелуін болдырмауға көмектеседі және деректерді өңдеуді басқарылатын және тиімді етеді. Бұл әрбір сөз бен оның мағынасы талдау нәтижесіне айтарлықтай әсер ететін үлкен мәтіндік корпуспен жұмыс істегенде өте маңызды. Нәтижесінде бұл тәсіл кілт сөздерді алудың дәлдігін жақсартып қана қоймайды, сонымен қатар көптеген салаларда, соның ішінде ғылыми зерттеулерде, білімді басқаруда және ақпаратты іздеуде маңызды болып табылатын мәтін құрылымы мен мағынасын түсінуді байытады.

Нәтижелер және оларды талқылау

Зерттеу үшін әртүрлі ғылыми салаларды қамтитын 182 ғылыми мақала

пайдаланылды. Бұл салаларға биология, информатика, физика, химия, психология және лингвистика кіреді. Әрбір санат белгілі бір білім саласын білдіреді және шамамен 20-25 құжатты қамтиды. Санаттардың бұл әртүрлілігі жан-жақты талдауды қамтамасыз етеді және әртүрлі ғылыми пәндердегі кілт сөздерді алу әдістерінің тиімділігін бағалауға мүмкіндік береді. Мақалалар әрбір ғылыми бағытты біркелкі көрсететіндей етіп таңдалды, бұл таңдалған әдістердің әртүрлі салаларда қолданылуы туралы жалпыланған қорытынды жасауға мүмкіндік береді. Эксперименттерді жүргізу үшін біз әртүрлі тақырыптар мен жазу стильдерін қамтитын әртүрлі мәтіндік деректер жиынын қолдандық. Бағалау әдістеріне кілт сөзді шығару сапасын бағалау үшін дәлдік және F1 өлшемдерін талдау кіреді. Екі әдісті де қолданып алынған түйінді сөздер векторлық кеңістіктегі сөздердің кластерленуін және салыстырмалы орындарын талдау үшін KMeans және PCA әдістерінің көмегімен көрнекі түрде көрсетілді. Бұл әдістер негізгі сөздердің семантикалық жақындығы мен тақырыпқа қатыстылығы негізінде топтастырылу жолын бағалауға көмектесті.

3-суретте нүктелер арасындағы қатынасты семантикалық немесе түйінді сөздер арасындағы жақындық дәрежесі ретінде түсіндіруге болады. Нүктелері тығыз топтастырылған кластерлер деректердегі нақтырақ анықталған тақырыптарды немесе ұғымдарды көрсетуі мүмкін. Оң жақтағы түс шкаласы кілт сөздермен байланысты салмақты немесе метриkanı көрсетеді, бірақ оның табиғатын қосымша сөздерсіз анықтау қиын. Бұл деректер жиынындағы кілт сөздердің маңыздылығы немесе жиілігі болуы мүмкін. Түйінді сөздерді кластерлеу KMeans және PCA (Principal Component Analysis) әдістерін қолдану арқылы екі графикте берілген. BERT үлгісінің визуализациясы сөздердің неғұрлым мәнерлі және сараланған векторлық көрінісін көрсете отырып, кластерлердің айқынырақ және айқынырақ таралуын көрсетеді. Бұл BERT туынды векторлары кілт сөздер арасындағы семантикалық ұқсастықтар мен айырмашылықтарды жақсырақ көрсететінін көрсетеді, бұл жақсы кластерлеу үшін маңызды. Ұсынылған график TF-IDF және автокодерлердің тіркесімін пайдалана отырып, кілт сөздерді кластерлеу нәтижелерін көрсетеді. Өлшемді азайту және кейінгі визуализация үшін KMeans және PCA әдістері қолданылады. Нәтижелер екі өлшемді кеңістікте концентрациясы мен дисперсиясының әртүрлі дәрежелері бар кластерлердің қалыптасуын көрсетеді, бұл сынақ деректер жинағындағы түйінді сөздердің семантикалық жақындығының негізгі тенденцияларын көрсетеді. Түс шкаласы әрбір кілт сөзбен байланысты қосымша көрсеткіштерді көрсете алады.



Сур. 3. TF-IDF + Autoencoder әдісі арқылы түйінді сөздерді кластерлеу нәтижесі
 (Fig. 3. Result of keyword clustering using TF-IDF + Autoencoder method)

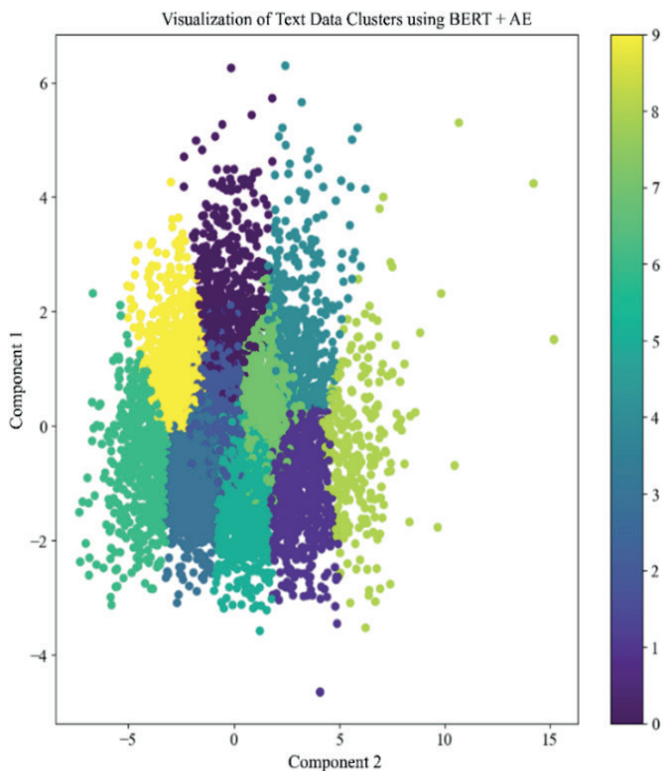
4-суретте түйінді сөздер тізімі және олардың сәйкес пайыздары бар мәгін нәтижесі көрсетілген. Бұл деректер TF-IDF әдісін қолданып, сипатталғандай автокодермен бірге ғылыми мақалалар мәтінінен алынған. «Құжаттың кері мерзімді жиілігі» дегенді білдіретін TF-IDF жинақтың немесе корпусстың бөлігі болып табылатын құжаттағы сөздің маңыздылығын бағалау үшін пайдаланылатын статистикалық өлшем болып табылады. Автокодерлерді деректер өлшемін азайту үшін пайдалануға болады, бұл ең маңызды мәліметтерді анықтауға көмектеседі.

```
1/1 [-----] - 0s 112ms/step
Top keywords:
layer : 0.389
deep : 0.344
network : 0.311
convolutional : 0.292
lstm : 0.255
neural network : 0.219
training : 0.21
architecture : 0.204
recurrent : 0.201
rnn : 0.193
```

Сур. 4. TF-IDF + Autoencoder әдісі арқылы түйінді сөздерді жіктеу нәтижесі
 (Fig. 4. The result of keyword classification by TF-IDF + Autoencoder method)

Осы тізімге сүйене отырып, бұл түйінді сөздер алынған ғылыми мақаланың мәтіні Machine Learning саласына, атап айтқанда терең оқыту (терең оқыту) және нейрондық желілер (нейрондық желілер) саласына жатады деп болжауға болады. «Деректер», «маңызды», «аудан», «машина», «өңдеу» және «есептеу» сөздері талқылауға деректерді өңдеуге және машиналық оқытудың есептеу аспектілеріне қатысты тақырыптар болуы мүмкін екенін мойындайды. «Оқу» және «терең» ең жоғары пайызды иелену фактісі бұл ұғымдардың қарастырылып отырған мақалада орталық екенін көрсетуі мүмкін. Сонымен қатар, «машина», «өңдеу» және «есептеу» сияқты сөздердің тең пайызы олардың зерттелетін тақырыптағы байланысын көрсете алады.

5-суретте автокодермен бейімделген Burt үлгісін пайдалану арқылы сынақ деректер жинағынан алынған түйінді сөздерді кластерлеу нәтижелері көрсетілген. Нәтижелерді екі өлшемді кеңістікте визуализациялау үшін KMeans және PCA әдістері қолданылды. Графикте әрбір нүкте кілт сөзге сәйкес келеді, нүктелердің түстері кілт сөз жататын кластерді көрсетеді және олардың орны PCA көмегімен алынған алғашқы екі негізгі компоненттің мәндерімен анықталады.



Сур. 5. BERT+Autoencoder әдісі арқылы түйінді сөздерді кластерлеу нәтижесі
(Fig. 5. Result of keyword clustering using BERT+Autoencoder method)

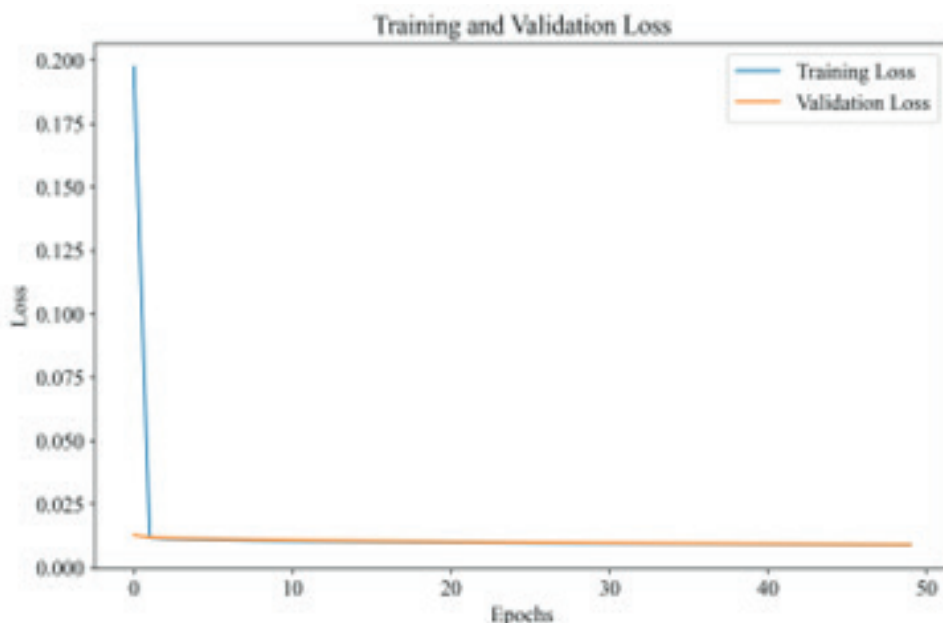
6-суретте ғылыми мақала мәтінінен Берт үлгісі мен автокодер тіркесімін пайдаланып алынған түйінді сөздердің тізімі көрсетілген, мұнда әрбір түйінді сөзге белгілі бір пайыз беріледі. Ол өңделген мәтіндегі әрбір кілт сөздің үлесін көрсетеді. Машиналық оқыту мен жасанды интеллектке қатысты сөздер тізімнің басында. Бірінші орында 5,78% үлеспен «оқу» сөзінің, одан кейін 4,33% үлеспен «терең» сөзінің болуы талданатын мәтінде терең оқыту тақырыбына ерекше көңіл бөлінгенін көрсетеді. Келесі «желі», «деректер» және «нейрондық» терминдері нейрондық желілер мен деректерге назар аудару арқылы мұны қолдайды. Мақаланың мазмұнында «маңызды», «аудан», «машина», «өңдеу» және «есептеу» түйінді сөздері де маңызды рөл атқара алады, бірақ негізгі тақырыппен салыстырғанда азырақ қосымша аспектілерді немесе сипаттарды көрсетеді.

```
Top keywords with a certain percentage:  
learning: 5.78%  
deep: 4.33%  
network: 3.61%  
data: 2.53%  
neural: 2.17%  
significant: 1.81%  
area: 1.44%  
machine: 1.08%  
processing: 1.08%  
computing: 1.08%
```

Сур. 6. BERT+Автокодер әдісі арқылы жіктеу нәтижесі
(Fig. 6. Classification result using BERT+Autoencoder method)

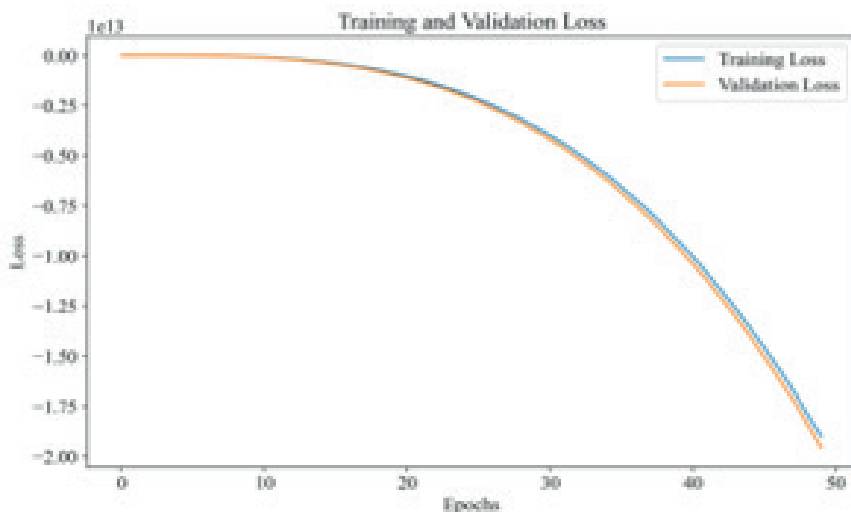
Зерттеу нәтижелері Автокодермен біріктірілген BERT әдісі кілт сөзді шығару тапсырмасында Автокодермен біріктірілген TF-IDF әдісімен салыстырғанда жақсы нәтижелер көрсететінін көрсетеді. Бұл BERT+Autoencoder әдісінің табиғи тілді тереңірек түсіну мүмкіндігін көрсетеді. Ол сондай-ақ күрделі мәтіндік деректерге жақсырақ бейімделеді. Модельдердің оқыту тарихы оқыту және тестілеу кезіндегі қателіктердің өзгеруін көрсететін графиктермен ұсынылған. Үлгілік сюжеттер оқыту және тестілеу кезеңдеріндегі қателіктің барлық дәуірлерде тұрақты және тұрақты түрде төмендегенін көрсетеді. Бұл BERT моделінің жоғары жалпылау қабілетін көрсетеді, ол ақпаратты артық орнатудың айқын белгілерінсіз сенімді түрде меңгеруге қабілетті. Керісінше, TF-IDF үлгісі туралы деректер ұсынылмайды, бірақ машиналық оқытудағы жалпы тенденцияларға сүйене отырып, TF-IDF деректердің белгілі бір түрлерінде жақсы жұмыс істей алады, бірақ кешенде BERT үлгісінен төмен болуы мүмкін деп болжауға болады. табиғи тілді өңдеу тапсырмалары.

7-сурет TF-IDF және автокодерлердің тіркесімін пайдалана отырып, ғылыми мақалалардан түйінді сөздерді алудың екі түрлі әдісінің салыстырмалы нәтижелерін көрсететін график. График екі жолды көрсетеді: «оқу қатесі» және «оқу қатесін тексеру», мұнда x осі деректер жинағы арқылы оқу алгоритмінің қайталану немесе өту санын көрсететін «дәуірлер» деп белгіленеді. График нақты кезеңдерде өлшенген және жазылған жаттығу қатесін және тексеру қатесін көрсетеді. Кестеге сүйенсек, жаттығу қатесі бастапқы кезеңде күрт төмендейді, бұл модель оқу деректерінен мақсатты кілт сөздерді шығаруда жылдам жақсаратынын көрсетеді. Дегенмен, валидация қатесі бүкіл процесс барысында салыстырмалы түрде тұрақты болып қалады, бұл шамадан тыс орнатуды немесе модельдің жаңа, белгісіз деректерде өнімділікті жақсартуға қабілетсіздігін көрсетуі мүмкін. Бұл екі қатені де бақылаудың маңыздылығын көрсетеді, өйткені тестілеу қателігін жақсартусыз оқыту қателігін айтарлықтай жақсарту оның жалпылануын жақсарту және тестілеу үшін басқа деректер жиынын реттеу немесе алу үшін үлгіні нақтылау қажеттілігін көрсетуі мүмкін. Сондай-ақ, график қате мәндері туралы ақпарат бермейтінін ескеру маңызды, өйткені Y осі «сандық қате» деп белгіленген және абсолютті қате мәндерін көрсететін қосымша белгілерсіз 0-ден 0,2-ге дейінгі шкалаға ие, бұл дәл түсіндіруді білдіреді. қателер мен өнімділік әдістері қосымша мәтінді немесе деректерді қажет етеді.



Сур. 7. Жаттығу кезінде қате динамикасын көрсететін TF-IDF + Автокодер үшін жаттығу және валидация графигі
(Fig. 7. Training and validation graph for TF-IDF + Autoencoder showing the error dynamics during training)

Графикте тік ось (Y) «катализатор массасы» деп белгіленеді және жоғарғы сол жақ бұрыштағы «1e13» префиксі арқылы көрінетін теріс логарифмдік шкалада көрсетілген қате мәнін білдіреді, бұл қатенің таралуын көрнекі етеді, әсіресе мәндердің үлкен диапазонында. Екі қисық бірдей пішінге ие, бұл жаттығу қателігі мен тестілеу қателігінің жаттығу процесі кезінде тығыз байланыста төмендейтінін көрсетеді. Бұл модель оқу деректерін жақсы меңгеріп қана қоймай, сонымен қатар алынған білімді оқу жиынына кірмейтін деректерге адекватты түрде жалпылайтынының оң белгісі. Бұл қисықтардың траекторияларының ұқсастығы модельді оқыту үдерісінің тұрақты болғанын және артық сәйкестік болмағанын көрсетуі мүмкін, бұл терең оқытуда жиі кездесетін мәселе, мұнда модель оқу деректерінде жақсы, бірақ бұрын-соңды болмаған деректерде нашар жұмыс істейді.



Сур. 8. BERT + Autoencoder үшін жаттығу графигі жаттығу барысында үлгі сапасының жақсаруын көрсетеді
(Fig.8. Training graph for BERT + Autoencoder showing the improvement in model quality as training progresses)

Екі әдістен де кілт сөзді шығару нәтижелері екі үлгі де бірдей терминдерді мағыналы деп анықтайтынын көрсетеді. Дегенмен, сөз семантикасын тереңірек түсіну арқылы BERT неғұрлым нюансты және контекстік тұрғыдан сәйкес кілт сөздерді анықтай алады, бұл оны күрделі ғылыми мәтіндерден кілт сөздерді шығару тапсырмаларына қолайлы етеді. BERT мәтіндерді талдау қабілетімен жақсы белгілі, бұл оны кілт сөздерді шығаруда өте тиімді етеді. Әрбір сөзді жеке қарастыратын TF-IDF-тен айырмашылығы, BERT сөздер арасындағы екі жақты қатынастарды ескереді, бұл олардың мәтіндегі мағынасын дәлірек түсінуге мүмкіндік береді. Бұл ғылыми мақалаларды талдау кезінде өте маңызды, мұнда бір терминдер әртүрлі категорияларда

әртүрлі мағынаға ие болуы мүмкін. Қорытындылай келе, талдау негізінде, ғылыми мәтіндерден түйінді сөздерді шығару тапсырмасында автокодермен BERT әдісі автокодермен TF-IDF-тен асып түседі деген қорытынды жасауға болады. BERT жақсы кластерлеуді және кілт сөздерді шығаруды қамтамасыз етіп қана қоймайды, сонымен қатар оқыту және тестілеу кезінде қателердің тұрақты төмендеуін көрсетеді. Күшті сөздерді ұсыну және терең семантикалық талдау мүмкіндіктері бар BERT күрделі ғылыми материалдарды өңдеу үшін таңдаулы таңдау болып табылады, мұнда әрбір терминнің нақты мағынасы маңызды.

Қорытынды

Бұл мақала ғылыми мәтіндерден кілт сөздерді алу үшін машиналық оқытудағы екі заманауи гибриді тәсілдердің мұқият салыстырмалы талдауын ұсынады: BERT (Трансформаторлардан екі бағытты кодтаушы өкілдіктері) автокодермен және TF-IDF (термдік жиілік-кері құжат жиілігі) комбинациясында. автокодермен. Жүргізілген зерттеулер негізінде мынадай қорытынды жасауға болады.

Біріншіден, нәтижелер BERT+Autoencoder кілт сөзді шығару тапсырмасында TF-IDF+Autoencoder қарағанда жақсырақ орындайтынын көрсетеді. Бұл BERT үлгісінің табиғи тілді жақсы түсіну және күрделі мәтіндік деректерге жақсырақ бейімделу қабілетін көрсетеді. BERT-тің екі жақты мәтінді өңдеуі модельге сөздер арасындағы қарым-қатынастарды ескеруге мүмкіндік береді, бұл алынған кілт сөздердің дәлдігі мен өзектілігін айтарлықтай жақсартады.

Екіншіден, графиктер арқылы берілген үлгілерді оқыту тарихы, BERT моделі үшін оқыту және сынау фазаларындағы қателік барлық дәуірлерде тұрақты түрде азайғанын көрсетеді, бұл оның жоғары жалпылау қабілетін шамадан тыс орнату белгілерінсіз көрсетеді. TF-IDF үлгісі жағдайында ұқсас тенденция анықталмады, бұл оның күрделі мәтіндерді өңдеудегі тиімділігінің төмендігін көрсетуі мүмкін.

Үшіншіден, KMeans және PCA әдістерін қолдану арқылы түйінді сөздерді кластерлеу нәтижелерін визуализациялау BERT моделінің артықшылығын растады. BERT көмегімен алынған сөздердің векторлық бейнелері түйінді сөздер арасындағы семантикалық ұқсастықтар мен айырмашылықтарды жақсырақ көрсетеді, бұл дәлірек кластерлеуді және мәтінді тереңдетуді жеңілдетеді. Осылайша, ғылыми мәтіндерден түйінді сөздерді алу міндеті бойынша автокодермен BERT әдісі автокодермен TF-IDF әдісінен жоғарырақ деген қорытынды жасауға болады. BERT жақсы кластерлеуді және кілт сөздерді шығаруды қамтамасыз етіп қана қоймайды, сонымен қатар оқыту және тестілеу кезінде қателердің тұрақты төмендеуін көрсетеді. Күшті сөздерді ұсыну және терең семантикалық талдау мүмкіндіктері бар BERT күрделі ғылыми материалдарды өңдеу үшін таңдаулы таңдау болып табылады, мұнда әрбір терминнің нақты мағынасы маңызды.

Әдебиеттер

Байер, М., Кауфхольд, М. А., Буххольд, Б., Келлер, М., Даллмейер, Дж., Және Ройтер, С. (2023). Табиғи тілді өңдеудегі деректерді көбейту: ұзақ және қысқа мәтіндік жіктеуіштерге арналған мәтін құрудың жаңа тәсілі. Халықаралық машиналық оқыту және кибернетика журналы, 14 (1), 135-150.

Бенитес-Андрасес, Дж.А., Алия-Перес, Дж. М., Видал, М. Е., Пастор-Варгас, Р., Және Гарсия-Ордас, М. Т. (2022). Машиналық оқытудың дәстүрлі модельдері және тамактанудың бұзылуы туралы твиттерді трансформаторға (BERT) негізделген автоматты түрде жіктеудің екі бағытты кодтаушы көріністері: Алгоритмдерді әзірлеу және валидацияны зерттеу. JMIR медициналық информатика, 10 (2), e34492.

Мюррей, Д.Г., Симса, Дж., Климович, А., Және Индик, И. (2021). тф. деректер: машиналық оқыту деректерін өңдеу жүйесі. arXiv алдын ала басып шығару arXiv:2101.12127.

Поло-Бланко, И., Гонсалес Лопес, М. Дж., Бруно, А., Және Гонсалес-Санчес, Дж. (2024). Жеңіл интеллектуалды кемістігі бар студенттерге схемаға негізделген оқытуды қолдана отырып, сөздік есептерді шешуге үйрету. Оқу Қабілетсіздігі Тоқсан Сайын, 47(1), 3-15.

Садирмекова, З., Тусупов, Ж., Мурзахметов, А., Жидекулова, Г., Тунгатарова, А., Төленбаев, М.,... & Боранқұлова, Г. (2023). Мәтінді өңдеудің автоматты әдістерінің онтологиялық инженериясы. Халықаралық Электротехника Және Есептеу Техникасы Журналы(IJECSE), 13 (6), 6620-6628.

Сузуки, Ю., Чжон, Х., Цуй, Х., Окамото, К., Кавашима, Р. Және Сугиура, М. (2023). Жасырын білімнің өлшемі ретінде сөздерді бақылау тапсырмасын fMRI валидациясын зерттеу: мінез-құлық пен нейрондық өңдеудегі айқын және жасырын бейімділіктердің ролін Зерттеу. Екінші Тілді Меңгеру бойынша зерттеулер, 45 (1), 109-136.

Тапех, А.Т. Г. Және Насер, М. З. (2023). Жасанды интеллект, машиналық оқыту және құрылымдық инженериядағы терең оқыту: тенденциялар мен озық тәжірибелерге ғылыми-метрикалық шолу. Инженериядағы Есептеу Әдістерінің мұрағаты, 30 (1), 115-159.

Таха, А.М., Ариффин, Д. С. Б. Б. Және Абу-Насер, С. С. (2023). Ми Ісігі Мен Мета-Анализдегі Терен Және Машиналық Оқыту Алгоритмдеріне Жүйелі Әдеби Шолу. Теориялық Және Қолданбалы Ақпараттық Технологиялар журналы, 101 (1), 21-36.

Тивари, П., Чаудхари, С., Мажи, Д., Және Мукерджи, Б. (2023). Автор ұсынған кілт сөздерді машинадан алынған терминдермен зерттеу тенденцияларын салыстыру: неврологиялық бұзылулар туралы жарияланымдар деректерін пайдалана отырып, ML алгоритмінің тәсілі. Ибероамерикандық Ғылыми Өлшеу Және Коммуникация Журналы, 3 (1), 2.

Хассани, Х., Бенеки, К., Унгер, С., Мазинани, М. Т. Және Еганеги, М. Р. (2020). Үлкен деректер аналитикасындағы мәтінді іздеу. Үлкен Деректер Және Когнитивті Есептеу, 4 (1), 1.

Хикман, Л., Тапа, С., Тэй, Л., Цао, М., Және Сринивасан, П. (2022). Ұйымдастырушылық зерттеулерде мәтінді өңдеуге арналған мәтінді алдын-ала өңдеу: Шолу және ұсыныстар. Ұйымдастырушылық Зерттеу Әдістері, 25 (1), 114-146.

Хунг, Л.П. Және Бүркеншік Ат, С. (2023). Көңіл-күйді талдаудан тыс: мәтінге негізделген көңіл-күйді талдау мен эмоцияны анықтаудағы соңғы тенденцияларға шолу. Жетілдірілген Есептеу Интеллекісі және Интеллектуалды Информатика журналы, 27 (1), 84-95.

Чигбу, Ұлыбритания, Атику, Оңтүстік Каролина Және Ду Плесси, Колумбия Округі (2023). Әдебиет Ғылымына Шолу: Іздеу, Анықтау, Таңдау және Синтездеу. Жарияланымдар, 11 (1), 2.] [Стоун, П. Дж. (2020). Тақырыптық мәтінді талдау: мәтіндік мазмұнды талдаудың Жаңа күн тәртібі. Әлеуметтік ғылымдарға арналған мәтіндік талдау, 35-54.

Шай, С. Р. (2023). Мәтінді алдын-ала өңдеу әдістерін салыстыру. Табиғи Тіл Инженериясы, 29 (3), 509-553.

References

Bayer, M., Kaufhold, M. A., Buchhold, B., Keller, M., Dallmeyer, J., & Reuter, C. (2023). Data augmentation in natural language processing: a novel text generation approach for long and short text classifiers. International journal of machine learning and cybernetics, 14(1), 135-150.

Benítez-Andrades, J. A., Alija-Pérez, J. M., Vidal, M. E., Pastor-Vargas, R., & García-Ordás, M. T. (2022). Traditional machine learning models and bidirectional encoder representations from transformer (BERT)-based automatic classification of tweets about eating disorders: Algorithm development and validation study. *JMIR medical informatics*, 10(2), e34492.

Chai, C. P. (2023). Comparison of text preprocessing methods. *Natural Language Engineering*, 29(3), 509-553.

Chigbu, U. E., Atiku, S. O., & Du Plessis, C. C. (2023). The Science of Literature Reviews: Searching, Identifying, Selecting, and Synthesising. *Publications*, 11(1), 2.] [Stone, P. J. (2020). Thematic text analysis: New agendas for analyzing text content. *Text analysis for the social sciences*, 35-54.

Hassani, H., Beneki, C., Unger, S., Mazinani, M. T., & Yeganegi, M. R. (2020). Text mining in big data analytics. *Big Data and Cognitive Computing*, 4(1), 1.

Hickman, L., Thapa, S., Tay, L., Cao, M., & Srinivasan, P. (2022). Text preprocessing for text mining in organizational research: Review and recommendations. *Organizational Research Methods*, 25(1), 114-146.

Hickman, L., Thapa, S., Tay, L., Cao, M., & Srinivasan, P. (2022). Text preprocessing for text mining in organizational research: Review and recommendations. *Organizational Research Methods*, 25(1), 114-146.

Hung, L. P., & Alias, S. (2023). Beyond sentiment analysis: A review of recent trends in text based sentiment analysis and emotion detection. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, 27(1), 84-95.

Murray, D. G., Simsa, J., Klimovic, A., & Indyk, I. (2021). tf. data: A machine learning data processing framework. *arXiv preprint arXiv:2101.12127*.

Polo-Blanco, I., González López, M. J., Bruno, A., & González-Sánchez, J. (2024). Teaching students with mild intellectual disability to solve word problems using schema-based instruction. *Learning Disability Quarterly*, 47(1), 3-15.

Sadirmekova, Z., Tussupov, J., Murzakhmetov, A., Zhidekulova, G., Tungatarova, A., Tulenbayev, M., ... & Borankulova, G. (2023). Ontology engineering of automatic text processing methods. *International Journal of Electrical and Computer Engineering (IJECE)*, 13(6), 6620-6628.

Suzuki, Y., Jeong, H., Cui, H., Okamoto, K., Kawashima, R., & Sugiura, M. (2023). An fMRI validation study of the word-monitoring task as a measure of implicit knowledge: Exploring the role of explicit and implicit aptitudes in behavioral and neural processing. *Studies in Second Language Acquisition*, 45(1), 109-136.

Taha, A. M., Ariffin, D. S. B. B., & Abu-Naser, S. S. (2023). A Systematic Literature Review of Deep and Machine Learning Algorithms in Brain Tumor and Meta-Analysis. *Journal of Theoretical and Applied Information Technology*, 101(1), 21-36.

Tapeh, A. T. G., & Naser, M. Z. (2023). Artificial intelligence, machine learning, and deep learning in structural engineering: a scientometrics review of trends and best practices. *Archives of Computational Methods in Engineering*, 30(1), 115-159.

Tiwari, P., Chaudhary, S., Majhi, D., & Mukherjee, B. (2023). Comparing research trends through author-provided keywords with machine extracted terms: A ML algorithm approach using publications data on neurological disorders. *Iberoamerican Journal of Science Measurement and Communication*, 3(1), 2.

CONTENTS

INFORMATION AND COMMUNICATION TECHNOLOGIES

M. Aitimov, R.U Almenayeva, K.K. Makulov, A.B. Ostayeva, R. Muratkhan
APPLICATION OF MACHINE LEARNING METHOD TO ANALYZE AND
EXTRACT SEMANTIC STRUCTURES FROM SCIENTIFIC TEXTS.....5

A.K. Aitim, G.K. Sembina
MODELING OF HUMAN BEHAVIOR FOR SMARTPHONE WITH USING
MACHINE LEARNING ALGORITHM.....17

G. Aksholak, A. Bedelbayev, R. Magazov
ANALYSIS AND COMPARISON OF MACHINE LEARNING METHODS
FOR MALWARE DETECTION.....29

A.L. Alexeyeva
SUBSONIC VIBROTRANSPORT SOLUTIONS OF THE WAVE EQUATION
IN SPACES OF DIMENSION $N=1,2,3$42

K. Bagitova, Sh. Mussiraliyeva, K. Azanbai
ANALYSIS OF SYSTEMS FOR RECOGNIZING POLITICAL EXTREMISM
IN ONLINE SOCIAL NETWORKS.....60

**A.S. Baegizova, G.I. Mukhamedrakhimova, I. Bapiyev, M.Zh. Bazarova,
U.M. Smailova**
EVALUATING THE EFFECTIVENESS OF MACHINE LEARNING
METHODS FOR KEYWORD COVERAGE.....73

G. Bekmanova, B. Yergesh, G. Yelibayeva, A. Omarbekova, M. Strecker
MODELING THE RULES AND CONDITIONS FOR CONDUCTING
PRE-ELECTION DEBATES.....89

M. Bolatbek, M. Sagynay, Sh. Mussiraliyeva
USING MACHINE LEARNING METHODS FOR DETECTING
DESTRUCTIVE WEB CONTENT IN KAZAKH LANGUAGE.....99

Y. Golenko, A. Ismailova, K. Kadirkulov, R. Kalendar
DEVELOPMENT OF AN ONLINE PLATFORM FOR SEARCHING FOR
TANDEM REPEATS USING WHOLE GENOME SEQUENCING.....112

T. Zhukabayeva, L. Zholshiyeva, N. Karabayev, Sh. Akhmetzhanova A BIBLIOMETRIC ANALYSIS OF EDGE COMPUTING IN INDUSTRIAL INTERNET OF THINGS (IIoT) CYBER-PHYSICAL SYSTEMS.....	123
S.S. Koishybay, N. Meirambekuly, A.E. Kulakaeva, B.A. Kozhakhmetova, A.A. Bulin DEVELOPMENT OF THE DESIGN OF A MULTI-BAND DISCONE ANTENNA.....	138
A. Kydyrbekova, D. Oralbekova SPEAKER IDENTIFICATION USING DISTRIBUTION-PRESERVING X-VECTOR GENERATION.....	152
B. Medetov, A. Nurlankyzy, A. Akhmediyarova, A. Zhetpisbayeva, D. Zhexebay COMPARATIVE ANALYSIS OF THE EFFECTIVENESS OF NEURAL NETWORKS WITHIN THE LOW SNR.....	163
A.A Myrzatay, L.G. Rzaeva, B. Zhumadilla, A.A. Mukhanova, G.A. Uskenbayeva DOUBLE EXPONENTIAL SMOOTHING AND TIME WINDOW METHODS FOR PREDICTIVE LAN MONITORING: ANALYSIS, COMPARISON AND APPLICATION.....	174
L. Naizabayeva, M.N. Satymbekov PREDICTING URBAN SOIL POLLUTION USING MACHINE LEARNING ALGORITHMS.....	194
A.U. Mukhiyadin, U.T. Makhazhanova, A.Z. Alimagambetova, A.A. Mukhanova, A.I. Akmoldina PREDICTING STUDENT LEARNING ENGAGEMENT USING MACHINE LEARNING TECHNIQUES: ANALYSIS OF EDUCATION DATA IN KAZAKHSTAN.....	204
Zh. Tashenova, Zh. Abdugulova, Sh. Amanzholova, E. Nurlybaeva PENETRATION TESTING APPROACHES EMPLOYING THE OPENVAS VULNERABILITY MANAGEMENT UTILITY.....	218
D.B. Tyulemissova, A.K. Shaikhanova, V. Martsenyuk, G.A. Uskenbayeva MODERN APPROACHES TO STUDYING THE DYNAMICS OF INFORMATION FLOW IN SOCIAL MEDIA BASED ON MACHINE LEARNING METHODS.....	231

МАЗМҰНЫ

АҚПАРАТТЫҚ-КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАР

М. Айтимов, Р.У Альменаева, К.К. Макулов, А.Б. Остаева, Р. Муратхан
ҒЫЛЫМИ МӘТІНДЕРДЕН СЕМАНТИКАЛЫҚ ҚҰРЫЛЫМДАРДЫ
ТАЛДАУ ЖӘНЕ АЛУ ҮШІН МАШИНАЛЫҚ ОҚЫТУ ӘДІСІН
ҚОЛДАНУ.....5

Ә.Қ. Әйтiм, Г.К. Сембина
МАШИНАЛЫҚ ОҚУ АЛГОРИТМІН ПАЙДАЛАНЫП СМАРТФОН
ҮШІН АДАМ МІНЕЗІН МОДЕЛДЕУ.....17

Г.И. Ақшолақ, А.А. Бедельбаев, Р.С. Мағазов
ЗИЯНДЫ БАҒДАРЛАМАЛАРДЫ АНЫҚТАУҒА АРНАЛҒАН
МАШИНАЛЫҚ ОҚЫТУ ӘДІСТЕРІН ТАЛДАУ ЖӘНЕ САЛЫСТЫРУ.....29

А.Л. Алексеева
N=1,2,3 ӨЛШЕМДІ КЕҢІСТІГІНДЕГІ ТОЛҚЫНДЫҚ ТЕҢДЕУДІҢ
ДЫБЫСҚА ДЕЙІНГІ ДІРІЛКӨЛІКТІК ШЕШІМДЕРІ.....42

Қ.Б. Бағитова, Ш.Ж. Мусиралиева, Қ. Азанбай
ӘЛЕУМЕТТІК ЖЕЛІЛЕРДЕГІ САЯСИ ЭКСТРЕМИЗМДІ ОНЛАЙН ТАҢУ
ЖҮЙЕЛЕРІН ТАЛДАУ.....60

**А.С. Баегизова, Г.И. Мухамедрахимова, И.М. Бапиев, М.Ж. Базарова,
У.М. Смайлова**
ТҮЙІН СӨЗДЕРДІ ҚАМТУ ҮШІН МАШИНАЛЫҚ ОҚЫТУ ӘДІСТЕРІНІҢ
ТИІМДІЛІГІН БАҒАЛАУ.....73

**Г.Т. Бекманова, Б.Ж. Ергеш, Г.К. Елибаева, А.С. Омарбекова,
М. Strecker**
САЙЛАУ АЛДЫНДАҒЫ ПІКІРТАЛАСТАРДЫ ӨТКІЗУ ЕРЕЖЕЛЕРІ
МЕН ШАРТТАРЫН МОДЕЛЬДЕУ.....89

М.А. Болатбек, М.Сағынай, Ш.Ж. Мусиралиева
ҚАЗАҚ ТІЛІНДЕГІ ДЕСТРУКТИВТІ ВЕБ-КОНТЕНТТІ АНЫҚТАУ ҮШІН
МАШИНАЛЫҚ ОҚЫТУ ӘДІСТЕРІН ҚОЛДАНУ.....99

Е.С. Голенко, А.А. Исмаилова, К.К. Кадиркулов, Р.Н. Календарь
ТОЛЫҚ ГЕНОМДЫҚ СЕКВЕНИРЛЕУДЕ ТАНДЕМДІК
ҚАЙТАЛАНУЛАРДЫ ІЗДЕУ ҮШІН ОНЛАЙН ПЛАТФОРМАСЫН
ӘЗІРЛЕУ.....112

- Т. Жукабаева, Л. Жолшиева, Н. Карабаев, Ш. Ахметжанова**
ӨНДІРІСТІК ЗАТТАР ИНТЕРНЕТІ (IoT) КИБЕРФИЗИКАЛЫҚ
ЖҮЙЕЛЕРІНДЕ ШЕТКІ ЕСЕПТЕУЛЕРДІ ҚОЛДАНУҒА
БИБЛИОМЕТРИЯЛЫҚ ТАЛДАУ.....123
- С.С. Қойшыбай, Н. Мейрамбекұлы, А.Е. Кулакаева, Б.А. Кожаметова,
А.А. Булин**
КӨПДИАПАЗОНДЫДИСКОНУСТЫҚАНТЕННАКОНСТРУКЦИЯСЫН
ӘЗІРЛЕУ.....138
- А.С. Кыдырбекова, Д.О. Оралбекова**
ТАРАТУДЫ САҚТАЙТЫН Х-ВЕКТОРЛАР ГЕНЕРАЦИЯСЫН
ПАЙДАЛАНЫП ДАУЫСТЫ ИДЕНТИФИКАЦИЯЛАУ.....152
- Б. Медетов, А. Нурланқызы, А. Ахмедиярова, А. Жетписбаева, Д. Жексебай**
СИГНАЛШУЫЛ ҚАТЫНАСЫ ТӨМЕН ЖАҒДАЙДА НЕЙРОНДЫҚ
ЖЕЛЛЕРДІҢ ТИІМДІЛІГІНЕ САЛЫСТЫРМАЛЫ ТАЛДАУ ЖАСАУ.....163
- А.А. Мырзатай, Л.Г. Рзаева, Б. Жұмаділла, А.А. Муханова,
Г.А. Ускенбаева**
ЖЕРГІЛІКТІ ЖЕЛІНІ БОЛЖАМДЫ БАҚЫЛАУҒА АРНАЛҒАН ҚОС
ЭКСПОНЕНЦИАЛДЫ ТЕГІСТЕУ ЖӘНЕ УАҚЫТ ТЕРЕЗЕЛЕРІНІҢ
ӘДІСТЕРІ: ТАЛДАУ, САЛЫСТЫРУ ЖӘНЕ ҚОЛДАНУ.....174
- Л. Найзабаева, М.Н. Сатымбеков**
МАШИНАЛЫҚ ОҚЫТУ АЛГОРИТМДЕРІН ПАЙДАЛАНУ АРҚЫЛЫ
ҚАЛА ТОПЫРАҒЫНЫҢ ЛАСТАНУЫН БОЛЖАУ.....194
- А.Ұ. Мұхиядин, У.Т. Махажанова, А.З. Алимагамбетова, А.А.Муханова,
А.И. Акмолдина**
МАШИНАЛЫҚ ОҚЫТУ ӘДІСТЕРІН ПАЙДАЛАНА ОТЫРЫП,
ОҚУШЫЛАРДЫҢ БІЛІМ АЛУҒА ЫНТАСЫН БОЛЖАУ:
ҚАЗАҚСТАНДАҒЫ БІЛІМ БЕРУ ДЕРЕКТЕРІН ТАЛДАУ.....204
- Ж.М. Ташенова, Ж.К. Абдугулова, Ш.А. Аманжолова, Э. Нурлыбаева**
OPENVAS ОСАЛДЫҒЫН БАСҚАРУ УТИЛИТАСЫН ҚОЛДАНА
ОТЫРЫП, ЕНУДІ ТЕСТІЛЕУ ТӘСІЛДЕРІ.....218
- Д.Б. Тюлемисова, А.К. Шайханова, В.П. Мартценюк, Г.А. Ускенбаева,
Г.В. Бекешева**
МАШИНАЛЫҚ ОҚЫТУ ӘДІСТЕРІНЕ НЕГІЗДЕЛГЕН ӘЛЕУМЕТТІК
ЖЕЛЛЕРДЕГІ АҚПАРАТ АҒЫНЫНЫҢ ДИНАМИКАСЫН ЗЕРТТЕУДІҢ
ЗАМАНАУИ ТӘСІЛДЕРІ.....231

СОДЕРЖАНИЕ

ИНФОРМАЦИОННО-КОММУНИКАЦИОННЫЕ ТЕХНОЛОГИИ

М. Айтимов, Р.У Альменаева, К.К. Макулов, А.Б. Остаева, Р. Муратхан ПРИМЕНЕНИЕ МЕТОДА МАШИННОГО ОБУЧЕНИЯ ДЛЯ АНАЛИЗА И ИЗВЛЕЧЕНИЯ СЕМАНТИЧЕСКИХ СТРУКТУР ИЗ НАУЧНЫХ ТЕКСТОВ.....	5
А.К. Айтим, Г.К. Сембина МОДЕЛИРОВАНИЕ ЧЕЛОВЕЧЕСКОГО ПОВЕДЕНИЯ ДЛЯ СМАРТФОНА С ИСПОЛЬЗОВАНИЕМ АЛГОРИТМА МАШИННОГО ОБУЧЕНИЯ.....	17
Г.И. Акшолок, А.А. Бедельбаев, Р.С. Магазов АНАЛИЗ И СРАВНЕНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ОБНАРУЖЕНИЯ ВРЕДОНОСНОГО ПО.....	29
Л.А. Алексеева ДОЗВУКОВЫЕ ВИБРОТРАНСПОРТНЫЕ РЕШЕНИЯ ВОЛНОВОГО УРАВНЕНИЯ В ПРОСТРАНСТВАХ РАЗМЕРНОСТИ $N=1,2,3$	42
К.Б. Багитова, Ш.Ж. Мусиралиева, К. Азанбай АНАЛИЗ СИСТЕМ РАСПОЗНАВАНИЯ ПОЛИТИЧЕСКОГО ЭКСТРЕМИЗМА В СОЦИАЛЬНЫХ СЕТЯХ ОНЛАЙН.....	60
А.С. Баегизова, Г.И. Мухамедрахимова, И.М. Бапиев, М.Ж. Базарова, У.М. Смайлова ОЦЕНКА ЭФФЕКТИВНОСТИ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ОХВАТА КЛЮЧЕВЫХ СЛОВ.....	73
Г.Т. Бекманова, Б.Ж. Ергеш, Г.К. Елибаева, А.С. Омарбекова, М. Strecker МОДЕЛИРОВАНИЕ ПРАВИЛ И УСЛОВИЙ ПРОВЕДЕНИЯ ПРЕДВЫБОРНЫХ ДЕБАТОВ.....	89
М.А. Болатбек, М. Сагынай, Ш.Ж. Мусиралиева ИСПОЛЬЗОВАНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ОБНАРУЖЕНИЯ ДЕСТРУКТИВНОГО ВЕБ-КОНТЕНТА НА КАЗАХСКОМ ЯЗЫКЕ.....	99
Е.С. Голенко, А.А. Исмаилова, К.К. Кадиркулов, Р.Н. Календарь РАЗРАБОТКА ОНЛАЙН-ПЛАТФОРМЫ ДЛЯ ПОИСКА ТАНДЕМНЫХ ПОВТОРОВ ПРИ ПОЛНОГЕНОМНОМ СЕКВЕНИРОВАНИИ.....	112

Т. Жукабаева, Л. Жолшиева, Н. Карабаев, Ш. Ахметжанова БИБЛИОМЕТРИЧЕСКИЙ АНАЛИЗ ПРИМЕНЕНИЯ ГРАНИЧНЫХ ВЫЧИСЛЕНИЙ В КИБЕРФИЗИЧЕСКИХ СИСТЕМАХ ПРОМЫШЛЕННОГО ИНТЕРНЕТА ВЕЩЕЙ (IIoT).....	123
С.С. Койшыбай, Н. Мейрамбекұлы, А.Е. Кулакаева, Б.А. Кожаметова, А.А. Булин РАЗРАБОТКА КОНСТРУКЦИИ МНОГОДИАПАЗОННОЙ ДИСКОНУСНОЙ АНТЕННЫ.....	138
А.С. Кыдырбекова, Д.О. Оралбекова ИДЕНТИФИКАЦИЯ ГОВОРЯЩЕГО С ИСПОЛЬЗОВАНИЕМ ГЕНЕРАЦИИ X-ВЕКТОРОВ С СОХРАНЕНИЕМ РАСПРЕДЕЛЕНИЯ...152	152
Б. Медетов, А. Нурланкызы, А. Ахмедиярова, А. Жетписбаева, Д. Жексебай СРАВНИТЕЛЬНЫЙ АНАЛИЗ ЭФФЕКТИВНОСТИ НЕЙРОННЫХ СЕТЕЙ ПРИ НИЗКОМ ЗНАЧЕНИИ ОТНОШЕНИЯ С/Ш.....	163
А.А. Мырзатай, Л.Г. Рзаева, Б. Жұмаділла, А.А. Муханова, Г.А. Ускенбаева МЕТОДЫ ДВОЙНОГО ЭКСПОНЕНЦИАЛЬНОГО СГЛАЖИВАНИЯ И ВРЕМЕННЫХ ОКОН ДЛЯ ПРЕДИКТИВНОГО МОНИТОРИНГА ЛВС: АНАЛИЗ, СРАВНЕНИЕ И ПРИМЕНЕНИЕ.....	174
Л. Найзабаева, М.Н. Сатымбеков ПРОГНОЗИРОВАНИЕ ЗАГРЯЗНЕНИЯ ГОРОДСКОЙ ПОЧВЫ С ИСПОЛЬЗОВАНИЕМ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ.....	194
А.У. Мухиядин, У.Т. Махажанов, А.З. Алимагамбетова, А.А. Муханова, А.И. Акмолдина ПРОГНОЗИРОВАНИЕ МОТИВАЦИИ УЧАЩИХСЯ К ОБУЧЕНИЮ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ: АНАЛИЗ ДАННЫХ ОБ ОБРАЗОВАНИИ В КАЗАХСТАНЕ.....	204
Ж.М. Ташенова, Ж.К. Абдугулова, Ш.А. Аманжолова, Э. Нурлыбаева ПОДХОДЫ К ТЕСТИРОВАНИЮ НА ПРОНИКНОВЕНИЕ С ИСПОЛЬЗОВАНИЕМ УТИЛИТЫ УПРАВЛЕНИЯ Уязвимостями OPENVAS.....	218
Д.Б. Тюлемисова, А.К. Шайханова, В. Мартценюк, Г.А. Ускенбаева, Г.В. Бекешева СОВРЕМЕННЫЕ ПОДХОДЫ К ИЗУЧЕНИЮ ДИНАМИКИ ИНФОРМАЦИОННОГО ПОТОКА В СОЦИАЛЬНЫХ МЕДИА НА ОСНОВЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ.....	231

**Publication Ethics and Publication Malpractice
the journals of the National Academy of Sciences of the Republic of Kazakhstan**

For information on Ethics in publishing and Ethical guidelines for journal publication see <http://www.elsevier.com/publishingethics> and <http://www.elsevier.com/journal-authors/ethics>.

Submission of an article to the National Academy of Sciences of the Republic of Kazakhstan implies that the described work has not been published previously (except in the form of an abstract or as part of a published lecture or academic thesis or as an electronic preprint, see <http://www.elsevier.com/postingpolicy>), that it is not under consideration for publication elsewhere, that its publication is approved by all authors and tacitly or explicitly by the responsible authorities where the work was carried out, and that, if accepted, it will not be published elsewhere in the same form, in English or in any other language, including electronically without the written consent of the copyright-holder. In particular, translations into English of papers already published in another language are not accepted.

No other forms of scientific misconduct are allowed, such as plagiarism, falsification, fraudulent data, incorrect interpretation of other works, incorrect citations, etc. The National Academy of Sciences of the Republic of Kazakhstan follows the Code of Conduct of the Committee on Publication Ethics (COPE), and follows the COPE Flowcharts for Resolving Cases of Suspected Misconduct (http://publicationethics.org/files/u2/New_Code.pdf). To verify originality, your article may be checked by the Cross Check originality detection service <http://www.elsevier.com/editors/plagdetect>.

The authors are obliged to participate in peer review process and be ready to provide corrections, clarifications, retractions and apologies when needed. All authors of a paper should have significantly contributed to the research.

The reviewers should provide objective judgments and should point out relevant published works which are not yet cited. Reviewed articles should be treated confidentially. The reviewers will be chosen in such a way that there is no conflict of interests with respect to the research, the authors and/or the research funders.

The editors have complete responsibility and authority to reject or accept a paper, and they will only accept a paper when reasonably certain. They will preserve anonymity of reviewers and promote publication of corrections, clarifications, retractions and apologies when needed. The acceptance of a paper automatically implies the copyright transfer to the National Academy of Sciences of the Republic of Kazakhstan.

The Editorial Board of the National Academy of Sciences of the Republic of Kazakhstan will monitor and safeguard publishing ethics.

Правила оформления статьи для публикации в журнале смотреть на сайтах:

www.nauka-nanrk.kz

<http://physics-mathematics.kz/index.php/en/archive>

ISSN 2518-1726 (Online),

ISSN 1991-346X (Print)

Директор отдела издания научных журналов НАН РК *А. Ботанқызы*

Редакторы: *Д.С. Аленов, Ж.Ш. Әден*

Верстка на компьютере *Г.Д. Жадыранова*

Подписано в печать 2.12.2024.

Формат 60x881/8. Бумага офсетная. Печать – ризограф.

16,0 п.л. Тираж 300. Заказ 4.