

ҚАЗАҚСТАН РЕСПУБЛИКАСЫНЫҢ ҒЫЛЫМ ЖӘНЕ ЖОҒАРЫ БІЛІМ МИНИСТРЛІГІ
МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РЕСПУБЛИКИ КАЗАХСТАН
MINISTRY OF SCIENCE AND HIGHER EDUCATION OF THE REPUBLIC OF KAZAKHSTAN



**ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ ЖӘНЕ
КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАР
ЖУРНАЛЫ**

**МЕЖДУНАРОДНЫЙ ЖУРНАЛ
ИНФОРМАЦИОННЫХ И
КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ**

**INTERNATIONAL JOURNAL OF INFORMATION
AND COMMUNICATION TECHNOLOGIES**

2025 (22) 2

сәуір - маусым

ISSN 2708–2032 (print)
ISSN 2708–2040 (online)

БАС РЕДАКТОР:

Исахов Асылбек Абдишимович — есептеу теориясы саласында математика бойынша PhD доктор, "Компьютерлік ғылымдар және информатика" бағыты бойынша қауымдастырылған профессор, Халықаралық ақпараттық технологиялар университетінің Басқарма Төрағасы – Ректор (Қазақстан)

БАС РЕДАКТОРДЫҢ ОРЫНБАСАРЫ:

Колесникова Катерина Викторовна — техника ғылымдарының докторы, профессор, Халықаралық ақпараттық технологиялар университетінің ғылыми-зерттеу қызметі жөніндегі проректор (Қазақстан)

ҒАЛЫМ ХАТШЫ:

Ипалакова Мадина Түлегеновна — техника ғылымдарының кандидаты, қауымдастырылған профессор, Халықаралық ақпараттық технологиялар университетінің ғылыми-зерттеу қызметі жөніндегі департамент директоры (Қазақстан)

РЕДАКЦИЯЛЫҚ АЛҚА:

Разак Абдул — PhD, Халықаралық ақпараттық технологиялар университеті киберқауіпсіздік кафедрасының профессоры (Қазақстан)
Лучино Томмазо де Паолис — Саленто Университеті (Италия) инновация және технологиялық инжиниринг департаменті AVR зертханасының зерттеу және әзірлеу бөлімінің директоры

Лиз Бэкон — профессор, Абертей Университеті (Ұлыбритания) вице-канцлерінің орынбасары

Микеле Пагано — PhD, Пиза Университетінің (Италия) профессоры

Өтелбаев Мұхтарбай Өтелбайұлы — физика-математика ғылымдарының докторы, профессор, КР ҰҒА академигі, Халықаралық ақпараттық технологиялар университеті математика және компьютерлік модельдеу кафедрасының профессоры (Қазақстан)

Рысбайұлы Болатбек — физика-математика ғылымдарының докторы, профессор, Есептеу және деректер ғылымдары департаментінің профессоры, Astana IT University (Қазақстан)

Дайнеко Евгения Александровна — PhD, Халықаралық ақпараттық технологиялар университеті ақпараттық жүйелер кафедрасының профессор-зерттеушісі (Қазақстан)

Дузаев Нуржан Токсужаевич — PhD, қауымдастырылған профессор, Халықаралық ақпараттық технологиялар университеті цифрландыру және инновациялар жөніндегі проректор (Қазақстан)

Синчев Бахтгерей Куспанович — техника ғылымдарының докторы, профессор, Халықаралық ақпараттық технологиялар университеті ақпараттық жүйелер кафедрасының профессоры (Қазақстан)

Сейлова Нургуль Абдуллаевна — техника ғылымдарының докторы, Халықаралық ақпараттық технологиялар университеті компьютерлік технологиялар және киберқауіпсіздік факультетінің деканы (Қазақстан)

Мұхамедиева Ардак Габитовна — экономика ғылымдарының кандидаты, Халықаралық ақпараттық технологиялар университеті бизнес медиа және басқару факультетінің деканы (Қазақстан)

Абдикаликова Замира Тұрсынбаевна — PhD, қауымдастырылған профессор, Халықаралық ақпараттық технологиялар университеті математика және компьютерлік модельдеу кафедрасының меңгерушісі (Қазақстан)

Шильдибеков Ерлан Жаржанович — PhD, қауымдастырылған профессор, Халықаралық ақпараттық технологиялар университеті экономика және бизнес кафедрасының меңгерушісі (Қазақстан)

Дамелия Максумовна Ескендірова — техника ғылымдарының кандидаты, қауымдастырылған профессор, Халықаралық ақпараттық технологиялар университеті киберқауіпсіздік кафедрасының меңгерушісі (Қазақстан)

Ниязгулова Айгуль Аскарбековна — филология ғылымдарының кандидаты, доцент, профессор, Халықаралық ақпараттық технологиялар университеті медиакоммуникация және Қазақстан тарихы кафедрасының меңгерушісі (Қазақстан)

Айтмағамбетов Алтай Зуфарович — техника ғылымдарының кандидаты, Халықаралық ақпараттық технологиялар университеті радиотехника, электроника және телекоммуникация кафедрасының профессоры (Қазақстан)

Бахтиярова Елена Азизбековна — техника ғылымдарының кандидаты, қауымдастырылған профессор, Халықаралық ақпараттық технологиялар университеті радиотехника, электроника және телекоммуникация кафедрасының меңгерушісі (Қазақстан)

Канибек Сансызбай — PhD, қауымдастырылған профессор, Халықаралық ақпараттық технологиялар университеті киберқауіпсіздік кафедрасының профессор-зерттеушісі (Қазақстан)

Тынымбаев Сахиябай — техника ғылымдарының кандидаты, профессор, Халықаралық ақпараттық технологиялар университеті компьютерлік инженерия кафедрасының профессор-зерттеушісі (Қазақстан)

Алимереб Али Абд — PhD, Халықаралық ақпараттық технологиялар университеті киберқауіпсіздік кафедрасының қауымдастырылған профессоры (Қазақстан)

Мохамед Ахмед Хамада — PhD, Халықаралық ақпараттық технологиялар университеті ақпараттық жүйелер кафедрасының қауымдастырылған профессоры (Қазақстан)

Янг Им Чу — PhD, Гачон университетінің профессоры (Оңтүстік Корея)

Талеуш Валдас — PhD, Адам Мицкевич атындағы (Польша) университеттің проректоры

Мамырбаев Оркен Жұмажанович — PhD, КР ҒЖБМ Ғылым комитеті ақпараттық және есептеу технологиялары институты ӨМК директорының ғылым жөніндегі орынбасары (Қазақстан)

Бушуев Сергей Дмитриевич — техника ғылымдарының докторы, профессор, Украинаның "УКРНЕТ" жобаларды басқару қауымдастығының директоры, Киев ұлттық құрылыс және сәулет университеті жобаларды басқару кафедрасының меңгерушісі (Украина)

Белошицкая Светлана Васильевна — техника ғылымдарының докторы, доцент, Astana IT University есептеу және деректер ғылымы кафедрасының профессоры (Қазақстан)

РЕДАКТОР:

Мрзабаева Раушан Жалиевна — магистр, Халықаралық ақпараттық технологиялар университетінің редакторы (Қазақстан)

Халықаралық ақпараттық және коммуникациялық технологиялар журналы

ISSN 2708–2032 (print)

ISSN 2708–2040 (online)

Меншік иесі: АҚ «Халықаралық ақпараттық технологиялар университеті» (Алматы қ.).

Қазақстан Республикасы Ақпарат және қоғамдық даму министрлігіне мерзімді баспасөз басылымын есепке қою туралы куәлік № KZ82VPY00020475, 20.02.2020 ж. берілген

Тақырып бағыты: ақпараттық технологиялар, ақпараттық қауіпсіздік және коммуникациялық технологиялар, әлеуметтік-экономикалық жүйелерді дамытудағы цифрлық технология.

Мерзімділігі: жылына 4 рет.

Тираж: 100 дана.

Редакция мекенжайы: 050040 Алматы қ., Манас к., 34/1, каб. 709, тел: +7 (727) 244-51-09.

E-mail: ijict@iitu.edu.kz

Журнал сайты: <https://journal.iitu.edu.kz>

© Халықаралық ақпараттық технологиялар университеті АҚ, 2025

Журнал сайты: <https://journal.iitu.edu.kz> © Авторлар ұжымы, 2025

ГЛАВНЫЙ РЕДАКТОР

Исахов Асылбек Абдиашимович — доктор PhD по математике в области теории вычислимости, ассоциированный профессор по направлению "Компьютерные науки и информатика", Председатель Правления – Ректор Международного университета информационных технологий (Казахстан)

ЗАМЕСТИТЕЛЬ ГЛАВНОГО РЕДАКТОРА:

Колесникова Катерина Викторовна — доктор технических наук, профессор, проректор по научно-исследовательской деятельности Международного университета информационных технологий (Казахстан)

УЧЕНЫЙ СЕКРЕТАРЬ:

Ипалакова Мадина Тулегеновна — кандидат технических наук, ассоциированный профессор, директор департамента по научно-исследовательской деятельности Международного университета информационных технологий (Казахстан)

РЕДАКЦИОННАЯ КОЛЛЕГИЯ:

Разак Абдул — PhD, профессор кафедры кибербезопасности Международного университета информационных технологий (Казахстан)

Лучио Томмазо де Паолис — директор отдела исследований и разработок лаборатории AVR департамента инноваций и технологического инжиниринга Университета Саленто (Италия)

Лиз Бэкон — профессор, заместитель вице-канцлера Университета Абертей (Великобритания)

Микеле Пагано — PhD, профессор Университета Пизы (Италия)

Отелбаев Мухтарбай Отелбайұлы — доктор физико-математических наук, профессор, академик НАН РК, профессор кафедры математического и компьютерного моделирования Международного университета информационных технологий (Казахстан)

Рысбайұлы Болатбек — доктор физико-математических наук, профессор, профессор Astana IT University (Казахстан)

Дайнеко Евгения Александровна — PhD, профессор-исследователь кафедры информационных систем Международного университета информационных технологий (Казахстан)

Дузбаев Нуржан Токкужаевич — PhD, ассоциированный профессор, проректор по цифровизации и инновациям Международного университета информационных технологий (Казахстан)

Синчев Бахтгерей Куспанович — доктор технических наук, профессор, профессор кафедры информационных систем Международного университета информационных технологий (Казахстан)

Сейлова Нургуль Абадуллаевна — кандидат технических наук, декан факультета компьютерных технологий и кибербезопасности Международного университета информационных технологий (Казахстан)

Мухамедиева Ардак Габитовна — кандидат экономических наук, декан факультета бизнеса медиа и управления Международного университета информационных технологий (Казахстан)

Абдикаликова Замира Турсынбаевна — PhD, ассоциированный профессор, заведующая кафедрой математического и компьютерного моделирования Международного университета информационных технологий (Казахстан)

Шильдибеков Ерлан Жаржанович — PhD, ассоциированный профессор, заведующий кафедрой экономики и бизнеса Международного университета информационных технологий (Казахстан)

Дамеля Максумовна Ескендирова — кандидат технических наук, ассоциированный профессор, заведующая кафедрой кибербезопасности Международного университета информационных технологий (Казахстан)

Ниязгулова Айгуль Аскарбековна — кандидат филологических наук, доцент, профессор, заведующая кафедрой медиакоммуникации и истории Казахстана Международного университета информационных технологий (Казахстан)

Айтмагамбетов Алтай Зуфарович — кандидат технических наук, профессор кафедры радиотехники, электроники и телекоммуникаций Международного университета информационных технологий (Казахстан)

Бахтиярова Елена Ажибековна — кандидат технических наук, ассоциированный профессор, заведующая кафедрой радиотехники, электроники и телекоммуникаций Международного университета информационных технологий (Казахстан)

Канибек Сансызбай — PhD, ассоциированный профессор, профессор-исследователь кафедры кибербезопасности, Международного университета информационных технологий (Казахстан)

Тынымбаев Сахиябай — кандидат технических наук, профессор, профессор-исследователь кафедры компьютерной инженерии, Международного университета информационных технологий (Казахстан)

Алмисреб Али Абд — PhD, ассоциированный профессор кафедры кибербезопасности Международного университета информационных технологий (Казахстан)

Мохамед Ахмед Хамада — PhD, ассоциированный профессор кафедры информационных систем Международного университета информационных технологий (Казахстан)

Янг Им Чу — PhD, профессор университета Гачон (Южная Корея)

Талеуш Валлас — PhD, проректор университета имен Адама Мицкевича (Польша)

Мамырбаев Оркен Жумажанович — PhD, заместитель директора по науке РГП Института информационных и вычислительных технологий Комитета науки МНВО РК (Казахстан)

Бушуев Сергей Дмитриевич — доктор технических наук, профессор, директор Украинской ассоциации управления проектами «УКРНЕТ», заведующий кафедрой управления проектами Киевского национального университета строительства и архитектуры (Украина)

Белошницкая Светлана Васильевна — доктор технических наук, доцент, профессор кафедры вычислений и науки о данных Astana IT University (Казахстан)

РЕДАКТОР:

Мрзабаева Раушан Жалиевна — магистр, редактор Международного университета информационных технологий (Казахстан)

Международный журнал информационных и коммуникационных технологий

ISSN 2708–2032 (print)

ISSN 2708–2040 (online)

Собственник: АО «Международный университет информационных технологий» (г. Алматы).

Свидетельство о постановке на учет периодического печатного издания в Министерство информации и общественного развития Республики Казахстан № KZ82VPY00020475, выданное от 20.02.2020 г.

Тематическая направленность: информационные технологии, информационная безопасность и коммуникационные технологии, цифровые технологии в развитии социально-экономических систем.

Периодичность: 4 раза в год.

Тираж: 100 экземпляров.

Адрес редакции: 050040 г. Алматы, ул. Манаса 34/1, каб. 709, тел: +7 (727) 244-51-09.

E-mail: ijict@iitu.edu.kz

Сайт журнала: <https://journal.iitu.edu.kz>

© АО Международный университет информационных технологий, 2025

© Коллектив авторов, 2025

EDITOR-IN-CHIEF

Assylbek Issakhov — PhD in Mathematics in Computability Theory, associate professor in “Computer Science and Informatics,” Chairman of the Board – Rector of the International Information Technology University (Kazakhstan)

DEPUTY EDITOR-IN-CHIEF

Kateryna Kolesnikova — Doctor of Technical Sciences, professor, Vice-Rector for Research, International Information Technology University (Kazakhstan)

ACADEMIC SECRETARY

Madina Ipalakova — Candidate of Technical Sciences, associate professor, Director of the Research Department, International Information Technology University (Kazakhstan)

EDITORIAL BOARD

Abdul Razak — PhD, professor, Department of Cybersecurity, International Information Technology University (Kazakhstan)

Lucio Tommaso De Paolis — Director of the R&D Department of the AVR Laboratory, Department of Engineering for Innovation, University of Salento (Italy)

Liz Bacon — Professor, Deputy Vice-Chancellor, Abertay University (United Kingdom)

Michele Pagano — PhD, Professor, University of Pisa (Italy)

Mukhtarbay Otelbayev — Doctor of Physical and Mathematical Sciences, professor, academician of the National Academy of Sciences of the Republic of Kazakhstan, professor of the Department of Mathematical and Computer Modeling, International Information Technology University (Kazakhstan)

Bolatbek Rysbauly — Doctor of Physical and Mathematical Sciences, professor, professor of the Department of Computing and Data Science, Astana IT University (Kazakhstan)

Yevgeniya Daineko — PhD, research professor, Department of Information Systems, International Information Technology University (Kazakhstan)

Nurzhan Duzbayev — PhD, associate professor, Vice-Rector for Digitalization and Innovation, International Information Technology University (Kazakhstan)

Bakhtgerai Sinchev — Doctor of Technical Sciences, professor, Department of Information Systems, International Information Technology University (Kazakhstan)

Nurgul Seilova — Candidate of Technical Sciences, Dean of the Faculty of Computer Technologies and Cybersecurity, International Information Technology University (Kazakhstan)

Ardak Mukhamediyeva — Candidate of Economic Sciences, Dean of the Faculty of Business, Media and Management, International Information Technology University (Kazakhstan)

Zamira Abdikalikova — PhD, associate professor, Head of the Department of Mathematical and Computer Modeling, International Information Technology University (Kazakhstan)

Yerlan Shildibekov — PhD, associate professor, Head of the Department of Economics and Business, International Information Technology University (Kazakhstan)

Damilya Yeskendiroya — Candidate of Technical Sciences, associate professor, Head of the Department of Cybersecurity, International Information Technology University (Kazakhstan)

Aigul Niyazgulova — Candidate of Philological Sciences, Professor, Head of the Department of Media Communications and History of Kazakhstan, International Information Technology University (Kazakhstan)

Altai Aitmagambetov — Candidate of Technical Sciences, Professor, Department of Radio Engineering, Electronics and Telecommunications, International Information Technology University (Kazakhstan)

Yelena Bakhtiyarova — Candidate of Technical Sciences, associate professor, Head of the Department of Radio Engineering, Electronics and Telecommunications, International Information Technology University (Kazakhstan)

Kanibek Sansyzbay — PhD, research professor, Department of Cybersecurity, International Information Technology University (Kazakhstan)

Sakhybay Tynymbayev — Candidate of Technical Sciences, Professor, Research Professor, Department of Computer Engineering, International Information Technology University (Kazakhstan)

Ali Abd Almisreb — PhD, associate professor, Department of Cybersecurity, International Information Technology University (Kazakhstan)

Mohamed Ahmed Hamada — PhD, associate professor, Department of Information Systems, International Information Technology University (Kazakhstan)

Yang Im Chu — PhD, Professor, Gachon University (South Korea)

Tadeusz Wallas — PhD, Vice-Rector, Adam Mickiewicz University (Poland)

Orken Mamyrbayev — PhD, Deputy Director for Science, RSE Institute of Information and Computational Technologies, Committee for Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Kazakhstan)

Sergey Bushuyev — Doctor of Technical Sciences, professor, Director of the Ukrainian Project Management Association “UKRNET,” Head of the Department of Project Management, Kyiv National University of Construction and Architecture (Ukraine)

Svetlana Beloshitskaya — Doctor of Technical Sciences, professor, Department of Computing and Data Science, Astana IT University (Kazakhstan)

EDITOR

Raushan Mrzabayeva — Master of Science, editor, International Information Technology University (Kazakhstan)

«International Journal of Information and Communication Technologies»

ISSN 2708–2032 (print)

ISSN 2708–2040 (online)

Owner: International Information Technology University JSC (Almaty).

The certificate of registration of a periodical printed publication in the Ministry of Information and Social Development of the Republic of Kazakhstan, Information Committee No. KZ8ZVPY00020475, issued on 20.02.2020.

Thematic focus: information technology, digital technologies in the development of socio-economic systems, information security and communication technologies

Periodicity: 4 times a year.

Circulation: 100 copies.

Editorial address: 050040. Manas st. 34/1, Almaty. +7 (727) 244-51-09. E-mail: ijict@iitu.edu.kz

Journal website: <https://journal.iitu.edu.kz>

© International Information Technology University JSC, 2025

© Group of authors, 2025

МАЗМҰНЫ

М.Ж. Айтимов, К.К. Макулов, А.Б. Остаева, Г.Ш. Мусагулова, Я. Култан АЙМАҚТЫҚ ДАМУҒА ӘЛЕУМЕТТІК-ЭКОНОМИКАЛЫҚ ФАКТОРЛАРДЫҢ ӨСЕРІН ЗЕРТТЕУ ҮШІН ГИБРИДТІ МОДЕЛЬДІ ҚОЛДАНУ.....	8
А.С. Баегизова, Г.Е. Мырзабекова, А.З. Алимагамбетова, Г.И. Мухамедрахимова, М. Кассим ИНТЕЛЛЕКТУАЛДЫ КЛАСТЕРЛЕУ ӘДІСТЕРІН ҚОЛДАНА ОТЫРЫП, ҚЫСҚА МӘТІНДЕРДІ ТАЛ ДАУ.....	23
Г.У. Бектемісова, А.А. Быков, Б.А. Нурашихин, А.А. Кереш, М.О. Даулетбек КОМПЬЮТЕРЛІК КӨРУ ЖӘНЕ МАШИНАЛЫҚ ОҚЫТУ НЕГІЗІНДЕ ҚҰРЫЛЫС ПРОЦЕСТЕРІН БАҚЫЛАУДЫҢ ИНТЕЛЛЕКТУАЛДЫ ЖҮЙЕСІН ӨЗІРЛЕУ ЖӘНЕ СЫНАУ	37
Д. Еділхан, Н. Хаймульдин, Д. Оспанова, Б. Амиргалиев БАРЛЫҚ КӨШЕЛЕРДІ ҚОЗҒАЛЫС ШЕКТЕУЛЕРІМЕН ҚАМТУҒА АРНАЛҒАН ОҢТАЙЛЫ МАРШРУТТЫ ТАБУДЫҢ ИНТЕЛЛЕКТУАЛДЫ МОДЕЛІ.....	57
А. Игнатович, А.С. Есенгельдина, Л.Т. Курмангазиева, С.Р. Шармұханбет, М.У. Худойберганов МАШИНАЛЫҚ ОҚЫТУ КӨМЕГІМЕН ҚАЗАҚСТАНДАҒЫ СОТ ШЕШІМДЕРІН БОЛЖАУҒА АРНАЛҒАН ДЕРЕКТЕРДІ ҚҰРУ.....	75
А.Х. Касымова, Г.Б. Абдикеримова, А.К. Асабай, А.Б. Серикбаева, С.М. Ахмедова БЕЙНЕ АҒЫНДАРЫНДА ҚОЗҒАЛАТЫН ОБЪЕКТІЛЕРДІ АНЫҚТАУ ҮШІН НАҚТЫ УАҚЫТТАҒЫ МАТЕМАТИКАЛЫҚ МОДЕЛЬДЕУ.....	91
С.К. Кумаргажанова, Д. Нұрлыбекова, А.С. Тлебалдинова, М.А. Карменова, С.А. Смаилова ШЫҒЫС ҚАЗАҚСТАН ОБЛЫСЫНЫҢ ҚАР КӨШКІНІ АЙМАҚТАРЫН МАШИНАЛЫҚ ОҚЫТУ ӘДІСІ АРҚЫЛЫ АНЫҚТАУ.....	115
С.Б. Муханов, К. Бекболат, Н.А. Сейлова, Ж.М. Бекаулова, С.Ж. Жакыпбеков САУДА ЖҮПТАРЫНДАҒЫ КРИПТОВАЛЮТА БАҒАСЫНЫҢ ҚҰБЫЛМАЛЫЛЫҒЫН БОЛЖАУҒА АРНАЛҒАН НЕЙРОНДЫҚ ЖЕЛІ МОДЕЛЬДЕРІН САЛЫСТЫРУ.....	130
Н.А. Оданова, К.Б. Багитова, Э.Э. Эльдарова, Ж.Т. Қабылхамит, Л.Терейковская АҚПАРАТТЫҚ ҚАУІПСІЗДІК САЛАСЫНДА ПЕРНЕТАҚТАДА ТЕРУ ҮЛГІСІН ТАЛДАУҒА АРНАЛҒАН НЕЙРОЖЕЛІЛІК МОДЕЛЬДЕРДІ ҚОЛДАНУ.....	142
А.С. Сраж, М. Нұртас, А.А. Алтайбек, З.Т. Абдикаликова НЕЙРОНДЫҚ ЖЕЛІЛЕР МАТЕМАТИКАЛЫҚ ЕСЕПТЕУ ҚҰРАЛЫ РЕТІНДЕ.....	154
А.М. Темірхан, Г.Н. Пашенко БАНКТЫҚ АЛАЯҚТЫҚТЫ АНЫҚТАУДАҒЫ ГРАФТЫҚ НЕЙРОНДЫҚ ЖЕЛІЛЕР: ӘДІСТЕРДІ САЛЫСТЫРМАЛЫ ТАЛДАУ	174
Е.Б. Тулебаев, А.А. Муханова, Н. Горанин МУНИЦИПАЛДЫҚ АКТИВТЕРДІҢ ТӘУЕКЕЛДЕРІН БОЛЖАУ: МАШИНАЛЫҚ ОҚЫТУДЫ ПАЙДАЛАНУ АРҚЫЛЫ ЖАС, ТОЗУ ЖӘНЕ ПАЙДАЛАНУ ҚАРҚЫНДЫЛЫҒЫ НЕГІЗІНДЕ ТЕХНИКАЛЫҚ ҚЫЗМЕТ КӨРСЕТУ ШЕШІМДЕРІН МОДЕЛЬДЕУ.....	186
С. Тынымбаев, Ж.Е. Темирбекова, Р.Ш. Бердибаев, С.Е. Маманова ТӨРТ РАЗРЯДТЫ ЕКІЛІК ПАРАЛЛЕЛЬ ТАСЫМАЛДАҒЫШЫ БАР ҚОСЫНДЫЛАРДЫ САЛЫСТЫРМАЛЫ ТАЛДАУ.....	200
Б.М. Укибасов, С.Б. Рахметулаева, Ж.О. Жанабеков, Ансар-уль-Хак Ясар ЭХОКАРДИОГРАММАЛАРДАҒЫ АҚАУЛАРДЫ ДИАГНОСТИКАЛАУ ҮШІН БЕТТІ АНЫҚТАУ ЖӘНЕ ТАНУ ҮЛГІЛЕРІН БЕЙІМДЕУ.....	214
Д.И. Усипбекова, М.Б. Есенова, Н. Ұзаққызы, Р. Муратхан, П. Шмидт МАММОГРАФИЯЛЫҚ СУРЕТТЕРДЕ СҮТ БЕЗІ ҚАТЕРЛІ ІСІГІН ДӨЛ АНЫҚТАУҒА АРНАЛҒАН БЕЛГІЛЕР ВЕКТОРЫН ҚҰРУ.....	235

СОДЕРЖАНИЕ

М.Ж. Айтимов, К.К. Макулов, А.Б. Остаева, Г.Ш. Мусагулова, Я. Култан ПРИМЕНЕНИЕ ГИБРИДНОЙ МОДЕЛИ ДЛЯ ИЗУЧЕНИЯ ВЛИЯНИЯ СОЦИАЛЬНО- ЭКОНОМИЧЕСКИХ ФАКТОРОВ НА РЕГИОНАЛЬНОЕ РАЗВИТИЕ.....	8
А.С. Баегизова, Г.Е. Мырзабекова, А.З. Алимагамбетова, Г.И. Мухамедрахимова, М. Кассим АНАЛИЗ КОРОТКИХ ТЕКСТОВ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ ИНТЕЛЛЕКТУАЛЬНОЙ КЛАС- ТЕРИЗАЦИИ.....	23
Г.У. Бектемысова, А.А. Быков, Б.А. Нурашихин, А.А. Кереш, М.О. Даулетбек РАЗРАБОТКА И ТЕСТИРОВАНИЕ ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ МОНИТОРИНГА СТРОИТЕЛЬНЫХ ПРОЦЕССОВ НА ОСНОВЕ КОМПЬЮТЕРНОГО ЗРЕНИЯ И МАШИННОГО ОБУЧ- ЕНИЯ.....	37
Д. Едилхан, Н. Хаймульдин, Д. Оспанова, Б. Амиргалиев ИНТЕЛЛЕКТУАЛЬНАЯ МОДЕЛЬ ПОИСКА ОПТИМАЛЬНОГО МАРШРУТА ПОКРЫТИЯ ВСЕХ УЛИЦ С ОГРАНИЧЕНИЯМИ ПО ПЕРЕМЕЩЕНИЯМ.....	57
А. Игнатович, А.С. Есенгельдина, Л.Т. Курмангазиева, С.Р. Шармуханбет, А.М. Недзьведь ГЕНЕРАЦИЯ ДАННЫХ ДЛЯ ПРОГНОЗИРОВАНИЯ СУДЕБНЫХ РЕШЕНИЙ В КАЗАХСТАНЕ С ИСПОЛЬЗОВАНИЕМ МАШИННОГО ОБУЧЕНИЯ.....	75
А.Х. Касымова, Г.Б. Абдикеримова, А.К. Асабай, А.Б. Серикбаева, С.М. Ахмедова МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ В РЕАЛЬНОМ ВРЕМЕНИ ДЛЯ ОБНАРУЖЕНИЯ ДВИЖУЩИХСЯ ОБЪЕКТОВ В ВИДЕОПОТОКАХ.....	91
С.К. Кумаргажанова, Д. Нурлыбекова, А.С. Тлебалдинова, М.А. Карменова, С.А. Смаилова ВЫЯЛЕНИЕ ЛАВИНООПАСНЫХ ЗОН В ВОСТОЧНОМ КАЗАХСТАНЕ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ.....	115
С.Б. Муханов, К. Бекболат, Н.А. Сейлова, Ж.М. Бекаулова, С. Ж. Жакыпбеков СРАВНЕНИЕ МОДЕЛЕЙ НЕЙРОННЫХ СЕТЕЙ ДЛЯ ПРОГНОЗИРОВАНИЯ ВОЛАТИЛЬНОСТИ ЦЕН КРИПТОВАЛЮТ В ТОРГОВЫХ ПАРАХ.....	130
Н.А. Оданова, К.Б. Багитова, Э.Э. Эльдарова, Ж.Т. Кабылхамит, Л. Терейковская ПРИМЕНЕНИЕ НЕЙРОСЕТЕВЫХ МОДЕЛЕЙ ДЛЯ АНАЛИЗА КЛАВИАТУРНОГО ПОЧЕРКА В ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ.....	142
А.С. Сраж, М. Нуртас, А.А. Алтайбек, З.Т. Абдикаликова НЕЙРОННЫЕ СЕТИ КАК ИНСТРУМЕНТ МАТЕМАТИЧЕСКИХ ВЫЧИСЛЕНИЙ.....	154
А.М. Темирхан, Г.Н. Пашенко ГРАФОВЫЕ НЕЙРОННЫЕ СЕТИ ДЛЯ ОБНАРУЖЕНИЯ МОШЕННИЧЕСТВА: СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ.....	174
Е.Б. Тулебаев, А.А. Муханова, Н. Горанин ПРОГНОЗНАЯ ОЦЕНКА РИСКОВ МУНИЦИПАЛЬНЫХ АКТИВОВ С ИСПОЛЬЗОВАНИЕМ МАШИННОГО ОБУЧЕНИЯ: МОДЕЛИРОВАНИЕ ВОЗРАСТА, ИЗНОСА И ИСПОЛЬЗОВАНИЯ ДЛЯ ПРИНЯТИЯ РЕШЕНИЙ ПО ОБСЛУЖИВАНИЮ.....	186
С. Тынымбаев, Ж.Е. Темирбекова, Р.Ш. Бердибаев, С.Е. Маманова СРАВНИТЕЛЬНЫЙ АНАЛИЗ ЧЕТЫРЕХРАЗЯДНЫХ ДВОИЧНЫХ СУММАТОРОВ С ПАРАЛЛЕЛЬНЫМИ ПЕРЕНОСАМИ.....	200
Б.М. Укибасов, С.Б. Рахметулаева, Ж.О. Жанабеков, Ансар-уль-Хак Ясар АДАПТАЦИЯ МОДЕЛЕЙ ОБНАРУЖЕНИЯ И РАСПОЗНАВАНИЯ ЛИЦ ДЛЯ ДИАГНОСТИКИ ДЕФЕКТОВ НА ЭХОКАРДИОГРАММАХ.....	214
Д.И. Усипбекова, М.Б. Есенова, Н. Узаккызы, Р. Мурадхан, П. Шмидт ОПРЕДЕЛЕНИЕ ВЕКТОРА ПРИЗНАКОВ ДЛЯ ТОЧНОГО ВЫЯВЛЕНИЯ РАКА МОЛОЧНОЙ ЖЕЛЕЗЫ НА ИЗОБРАЖЕНИЯХ МАММОГРАММЫ.....	235

CONTENTS

M. Aitimov, K. Makulov, A. Ostayeva, G. Mussagulova, J. Kultan APPLICATION OF A HYBRID MODEL TO STUDY THE IMPACT OF SOCIO-ECONOMIC FACTORS ON REGIONAL DEVELOPMENT.....	8
A. Bayegizova, G. Murzabekova, A. Alimagambetova, G. Mukhamedrakhimova, M. Kassim INTELLIGENT CLUSTERING METHODS FOR PROCESSING AND ANALYZING SHORT TEXTS.....	23
G.U. Bektemysova, A.A. Bykov, B.A. Norashikin, A.A. Keresh, M.O. Dauletbek DEVELOPMENT AND TESTING OF AN INTELLIGENT MONITORING SYSTEM FOR CONSTRUCTION PROCESSES BASED ON COMPUTER VISION AND MACHINE LEARNING.....	37
D. Yedilkhan, N. Khaimuldin, D. Ospanova, B. Amirgaliyev INTELLIGENT MODEL FOR FINDING THE OPTIMAL ROUTE TO COVER ALL STREETS WITH MOVEMENT CONSTRAINTS.....	57
A. Ignatovich, A.S. Yesengeldina, L.T. Kurmangazyeva, S.R. Sharmukhanbet, A.M. Nedzved SIGNIFICANT FACTORS ANALYSIS USING SHAP AND PCA FOR OPTIMIZING AGRICULTURAL RESOURCE MANAGEMENT.....	75
A. Kassymova, G. Abdikerimova, A. Assabay, A. Serikbayeva, S. Akhmedova REAL-TIME MATHEMATICAL MODELING FOR MOVING OBJECT DETECTION IN VIDEO STREAMS.....	91
S.K. Kumargazhanova, D. Nurlybekova, A.S. Tlebaldinova, M.A. Karmenova, S.S. Smailova FINDING AVALANCHE AREAS IN EAST KAZAKHSTAN USING MACHINE LEARNING.....	115
S.B. Mukhanov, K. Bekbolat, N.A. Seilova, Zh.M. Bekaulova, S.Zh. Zhakypbekov COMPARISON OF NEURAL NETWORK MODELS FOR PREDICTION CRYPTOCURRENCY PRICE VOLATILITY IN TRADING PAIRS.....	130
N. Odanova, K. Bagitova, E. Eldarova, Zh. Kabylkhamit, L. Tereikovska APPLICATION OF NEURAL NETWORK MODELS FOR KEYSTROKE DYNAMICS ANALYSIS IN INFORMATION SECURITY.....	142
A.S. Srazh, M. Nurtas, A.A. Altaibek, Z.T. Abdikalikova EURAL NETWORKS AS A TOOL FOR MATHEMATICAL COMPUTATION.....	154
A.M. Temirkhan, G.N. Pachshenko GRAPH NEURAL NETWORKS FOR FRAUD DETECTION: COMPARATIVE ANALYSIS OF METHODS.....	174
Y.B. Tulebayev, A.A. Mukhanova, N. Goranin PREDICTIVE RISK ASSESSMENT OF MUNICIPAL ASSETS USING MACHINE LEARNING: MODELING AGE, DEPRECIATION, AND USAGE FOR MAINTENANCE DECISIONS.....	186
S. Tynymbayev, Zh. E. Temirbekova, R. Berdibayev, E.S. Mamanova COMPARATIVE ANALYSIS OF FOUR-BIT BINARY ADDERS WITH PARALLEL CARRIES.....	200
B.M. Ukibassov, S.B. Rakhmetulayeva, Zh.O. Zhanabekov, Ansar-Ul-Haque Yasar ADAPTING FACE DETECTION AND RECOGNITION MODELS FOR ECHOCARDIOGRAPHIC DEFECT DIAGNOSIS.....	214
D. Ussipbekova, M. Yessenova, N. Uzakkyzy, R. Muratkhan, P. Schmidt DETERMINATION OF SIGNS VECTOR FOR ACCURATE DETECTION OF BREAST CANCER IN MAMMOGRAM IMAGES.....	235

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES

ISSN 2708–2032 (print)

ISSN 2708–2040 (online)

Vol. 6. Is. 1. Number 22 (2025). Pp. 23–36

Journal homepage: <https://journal.iitu.edu.kz><https://doi.org/10.54309/IJICT.2025.22.2.002>

УДК 004.931

INTELLIGENT CLUSTERING METHODS FOR PROCESSING AND ANALYZING SHORT TEXTS

*A. Bayegizova¹, G. Murzabekova² *, A. Alimagambetova¹, G. Mukhamedrakhimova¹, M. Kassim³*

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

²S. Seifullin Kazakh Agrotechnical Research University, Astana, Kazakhstan;

³MARA University of Technology, Malaysia.

E-mail: murzabekova.gulden@mail.ru

Bayegizova A. — senior lecturer, «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>;

Murzabekova G. — associate professor, «The Department of Computer Science», Candidate of Physical and Mathematical Sciences, Kazakh Agrotechnical Research University named after S. Seifullin, Astana, Kazakhstan

E-mail: murzabekova.gulden@mail.ru, <https://orcid.org/0000-0001-9807-5200>;

Alimagambetova A. — senior lecturer, «The Department of Information Systems», Candidate of Physical and Mathematical Sciences, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan

E-mail: ainash_777@mail.ru, <https://orcid.org/0000-0002-9859-2029>;

Mukhamedrakhimova G. — senior lecturer, «The 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>;

Murizah K. — PhD, associate professor, MARA University of Technology, Malaysia

E-mail: Kassim@gmail.com, <https://orcid.org/0000-0002-8494-4783>.

© A. Bayegizova, G. Murzabekova, A. Alimagambetova, G. Mukhamedrakhimova, M. Kassim, 2025

Abstract. This study presents an in-depth analysis of short text clustering using advanced methods: Bidirectional Encoder Representations from Transformers (BERT), Term Frequency-Inverse Document Frequency (TF-IDF), and the innovative hybrid method Latent Dirichlet Allocation+BERT+Autoencoder (LDA+BERT+AE). The experimental section compares the effectiveness of these methods for short text clustering, with particular attention paid to the hybrid approach LDA+BERT+AE. During training and validation of the LDA-BERT model for 200 epochs, the loss values are initially higher than 1.2, rapidly decrease during the first 25 epochs, reach about 0.8, and finally stabilize at about 0.4. The research results show that the hybrid method LDA+BERT+AE significantly improves the clustering quality compared to the methods used individually. Based on the results obtained, recommendations are



given on the selection and combination of optimal methods for various types of short text clustering and natural language processing (NLP) tasks. The study highlights the importance of a comprehensive analysis method for short texts, its necessity for increasing deep semantic understanding and better information retrieval.

Key words: clustering methods, semantic extraction, categorization of text data, hybrid method, NLP systems

For citation: A. Bayegizova, G. Murzabekova, A. Alimagambetova, G. Mukhamedрахимова, M. Kassim. Intelligent clustering methods for processing and analyzing short texts//International journal of information and communication technologies. 2025. Vol. 6. No. 22. Pp. 23-36. (in Kaz). <https://doi.org/10.54309/IJICT.2025.22.2.002>.

Conflict of interest: The authors declare that there is no conflict of interest.

ИНТЕЛЛЕКТУАЛДЫ КЛАСТЕРЛЕУ ӘДІСТЕРІН ҚОЛДАНА ОТЫРЫП, ҚЫСҚА МӘТІНДЕРДІ ТАЛДАУ

А.С. Баегизова¹, Г.Е. Мырзабекова^{2}, А.З. Алимагамбетова¹,
Г.И. Мухамедрахимова¹, М. Кассим³*

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

²С.Сейфуллин атындағы Қазақ агротехникалық зерттеу университеті, Астана, Қазақстан;

³MARA технологиялар университеті, Малайзия.

E-mail: murzabekova.gulden@mail.ru

Баегизова А. — 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>;

Мырзабекова Г. — С.Сейфуллин атындағы Қазақ агротехникалық зерттеу университеті, «Компьютерлік ғылымдар» кафедрасының доценті, физика-математика ғылымдарының кандидаты, Астана, Қазақстан
E-mail: murzabekova.gulden@mail.ru, <https://orcid.org/0000-0001-9807-5200>;

Алимагамбетова А. — Л.Н. Гумилев атындағы Еуразия ұлттық университеті, «Ақпараттық жүйелер» кафедрасының аға оқытушысы, ф.-м.ғ.к., Астана, Қазақстан
E-mail: ainash_777@mail.ru, <https://orcid.org/0000-0002-9859-2029>;

Мухамедрахимова Г. — .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>;

Муризах К. — PhD, қауымдастырылған профессор, MARA технологиялар университеті, Малайзия
E-mail: Kassim@gmail.com, <https://orcid.org/0000-0002-8494-4783>.

© А.С. Баегизова, Г.Е. Мырзабекова А.З. Алимагамбетова, Г.И. Мухамедрахимова, М. Кассим, 2025

Аннотация. Бұл зерттеу қысқа мәтіндерді кластерлеудің терең талдауын ұсынады, ол озық әдістерді қолдану арқылы жүзеге асырылады: Bidirectional Encoder Representations from Transformers (BERT), Term Frequency-Inverse Document Frequency (TF-IDF) және инновациялық гибридті әдіс Latent Dirichlet

Allocation+BERT+Autoencoder (LDA+BERT+AE). Эксперименттік бөлімде осы әдістердің қысқа мәтіндерді кластерлеу тиімділігі салыстырылады, әсіресе LDA+BERT+AE гибриді тәсіліне ерекше назар аударылады. LDA-BERT моделін 200 эпох бойы оқыту және тексеру барысында жоғалту функциясының мәндері (loss values) бастапқыда 1.2-ден жоғары болып, алғашқы 25 эпох ішінде тез төмендеп, шамамен 0.8 деңгейіне жетеді және ақырында 0.4 шамасында тұрақталады. Зерттеу нәтижелері LDA+BERT+AE гибриді әдісі жеке қолданылған әдістерге қарағанда кластерлеу сапасын едәуір жақсартатынын көрсетеді. Алынған нәтижелер негізінде қысқа мәтіндерді кластерлеудің әртүрлі түрлері мен табиғи тілдік өңдеу (NLP) тапсырмалары үшін оңтайлы әдістерді таңдау және үйлестіру бойынша ұсыныстар беріледі. Зерттеу қысқа мәтіндерді кешенді талдау әдісінің маңыздылығын, оның терең семантикалық түсінуді арттыру және ақпаратты жақсырақ алу үшін қажеттілігін ерекше атап көрсетеді.

Түйін сөздер: Кластерлеу әдістері, семантикалық шығару, мәтіндік деректерді санаттау, гибриді әдіс, NLP жүйелері

Дәйексөздер үшін: А.С. Баегизова, Г.Е. Мырзабекова А.З. Алимагамбетова, Г.И. Мухамедрахимова, М. Кассим. Интеллектуалды кластерлеу әдістерін қолдана отырып, қысқа мәтіндерді талдау//Халықаралық ақпараттық және коммуникалық технологиялар журналы. 2025. Т. 6. No. 22. 23–36 бет. (қазақ тілінде). <https://doi.org/10.54309/IJICT.2025.22.2.002>.

Мүдделер қақтығысы: Авторлар осы мақалада мүдделер қақтығысы жоқ деп мәлімдейді.

АНАЛИЗ КОРОТКИХ ТЕКСТОВ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ ИНТЕЛЛЕКТУАЛЬНОЙ КЛАСТЕРИЗАЦИИ

А.С. Баегизова¹, Г.Е. Мырзабекова^{2,}, А.З. Алимагамбетова¹,
Г.И. Мухамедрахимова¹, М. Кассим³*

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

²Казахский агротехнический исследовательский университет имени С. Сейфуллина, Астана, Казахстан;

³Технологический университет МАРА, Малайзия.

E-mail: murzabekova.gulden@mail.ru

Баегизова А. — старший преподаватель, «Кафедра радиотехники, электроники и телекоммуникаций», Евразийский национальный университет имени Л.Н. Гумилёва, Астана, Казахстан
E-mail: baegiz_a@mail.ru, <https://orcid.org/0000-0003-2293-2143>;

Мырзабекова Г. — кандидат физико-математических наук, доцент, «Кафедра компьютерных наук», Казахский агротехнический исследовательский университет имени С. Сейфуллина, Астана, Казахстан
E-mail: murzabekova.gulden@mail.ru, <https://orcid.org/0000-0001-9807-5200>;

Алимагамбетова А. — кандидат физико-математических наук, старший преподаватель, «Кафедра информационных систем», Евразийский национальный университет имени Л.Н. Гумилёва, Астана, Казахстан
E-mail: ainash_777@mail.ru, <https://orcid.org/0000-0002-9859-2029>;

Мухамедрахимова Г. — старший преподаватель, «Кафедра радиотехники, электроники и телекоммуникаций», Евразийский национальный университет имени Л.Н. Гумилёва, Астана, Казахстан

E-mail: isatai-07@mail.ru, <https://orcid.org/0000-0002-9951-6263>;

Муризах К. — доктор философии, доцент, Технологический университет MAPA, Малайзия

E-mail: Kassim@gmail.com, <https://orcid.org/0000-0002-8494-4783>.

© А.С. Баегизова, Г.Е. Мырзабекова А.З. Алимагамбетова, Г.И. Мухамедрахимова, М. Кассим, 2025

Аннотация. В статье представлен углубленный анализ кластеризации коротких текстов с использованием передовых методов: двунаправленные представления кодировщика из преобразователей (BERT), частота терминов-обратная частота документа (TF-IDF) и инновационный гибридный метод Latent Dirichlet Allocation+BERT+Autoencoder(LDA+BERT+AE). В экспериментальном разделе сравнивается эффективность этих методов для кластеризации коротких текстов, при этом особое внимание уделяется гибриднему подходу LDA+BERT+AE. Во время обучения и проверки модели LDA-BERT для 200 эпох значения потерь изначально превышают 1,2, быстро уменьшаются в течение первых 25 эпох, достигают около 0,8 и, наконец, стабилизируются на уровне около 0,4. Результаты исследования показывают, что гибридный метод LDA+BERT+AE значительно улучшает качество кластеризации по сравнению с методами, используемыми по отдельности. На основе полученных результатов даны рекомендации по выбору и комбинированию оптимальных методов для различных типов задач кластеризации коротких текстов и обработки естественного языка (NLP). В исследовании подчеркивается важность метода комплексного анализа для коротких текстов, его необходимость для повышения глубины семантического понимания и лучшего извлечения информации.

Ключевые слова: методы кластеризации, семантическое извлечение, категоризация текстовых данных, гибридный метод, системы NLP

Для цитирования: А.С. Баегизова, Г.Е. Мырзабекова А.З. Алимагамбетова, Г.И. Мухамедрахимова, М. Кассим. Анализ коротких текстов с использованием методов интеллектуальной кластеризации//Международный журнал информационных и коммуникационных технологий. 2025. Т. 6. №. 22. Стр. 23–36. (На каз.). <https://doi.org/10.54309/IJICT.2025.22.2.002>.

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Кіріспе

Мәтіндік деректерді практикалық түсіну және санаттау (Manias et al., 2023) әлеуметтік медиа талдауы (Fu et al., 2023), ұсыныс жүйелері және автоматтандырылған контент модерациясы (Edara et al., 2023) сияқты көптеген қолданбалар үшін маңызды. Соңғы онжылдықта шектеулі және сығылған ақпаратпен сипатталатын қысқа мәтіндерді өңдеу мен мағынаны шығару әдістерін дамытуға айтарлықтай күш салынды. Бұл саладағы ең қуатты әдістердің қатарына BERT, TF+IDF және біріктірілген LDA+BERT+AE әдісі

жатады. BERT (Bidirectional Encoder Representations from Transformers) – бұл мәтіндегі контекстік байланыстарды жақсырақ түсінуге мүмкіндік беретін терең оқытудағы инновациялық әдіс. TF+IDF, классикалық статистикалық тәсіл ретінде, құжаттағы сөздердің маңыздылығын бағалау үшін қолданылады, бұл оны ақпаратты сүзу және негізгі тақырыптарды анықтау үшін таптырмас құрал етеді. LDA+BERT+AE әдісі мәтіндік түсінікті терең және кеңінен синтездеуге мүмкіндік береді: LDA (Latent Dirichlet Allocation) тақырыптық модельдеу үшін, BERT контекстік талдау үшін, ал AE (Autoencoder) өлшемділікті азайту және деректерді оңтайландыру үшін пайдаланылады.

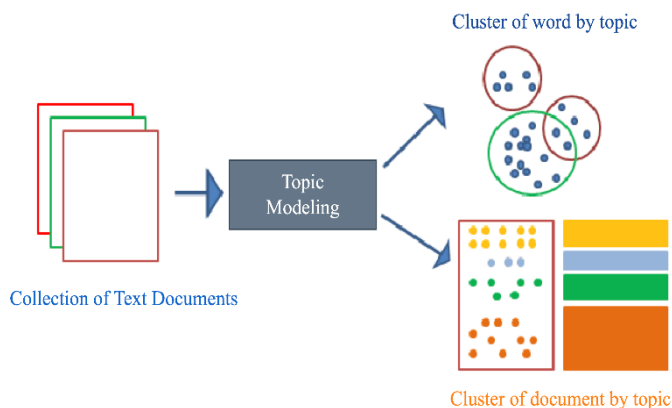
Қысқа мәтіндерді интеллектуалды кластерлеу әдістерін қолдана отырып талдаудың маңыздылығы (Balaji et al., 2021) қазіргі цифрлық коммуникацияларға тән үлкен көлемдегі қысқа мәтіндік ақпаратты тиімді өңдеу және түсіндіру қажеттілігінің артуынан туындайды. Әлеуметтік желілердің, онлайн платформалардың және мобильді қосымшалардың өсуімен қатар твиттер, жазбалар және пікірлер сияқты қысқа мәтіндердің саны да артып келеді, бұл табиғи тілді өңдеу жүйелерінің талдау дәлдігі мен жылдамдығын жақсарту қажеттілігін тудырады. Мұндай мәтіндер көбінесе қоғамдық пікір, пайдаланушы қалаулары және ағымдағы трендтер туралы маңызды ақпаратты қамтиды, сондықтан оларды зерттеу маркетинг, брендті басқару, саяси талдау және басқа да көптеген салалар үшін құнды болып табылады. Осылайша, BERT, TF+IDF және LDA+BERT+AE сияқты әдістерді дамыту және оңтайландыру қысқа мәтіндерді дәл және тиімді өңдеу үшін жасанды интеллект және NLP салаларында маңызды рөл атқарады (Abbas et al., 2022; McKittrick et al., 2023). Бұл жұмыстың мақсаты – интеллектуалды кластерлеу әдістерін қолдану арқылы мәтіндік деректерді кластерлеу сапасын арттыру және процесті жылдамдату, осылайша пайдалы ақпаратты дәлірек және тиімдірек шығару. Ерекше назар бұл әдістердің бейімделгіштігі мен әртүрлі қолдану жағдайларындағы қолданылуына аударылады, өйткені дәстүрлі тәсілдер кейде деректер көлемінің шектеулігі немесе әртүрлілігі салдарынан қиындықтарға тап болуы мүмкін.

Бұл зерттеудің нәтижелері қысқа мәтіндерді талдау саласында айтарлықтай жетілдірулер ұсынады, мәтін мазмұнын тереңірек және дәлірек түсінуге мүмкіндік береді және тиімдірек NLP жүйелерін жасауға ықпал етеді. Цифрлық дәуірде деректер көлемінің қарқынды өсуі жағдайында бұл тақырыптың маңыздылығы арта түседі. Твиттер жазбалары, пайдаланушы пікірлері және әлеуметтік желілердегі комментарийлер сияқты қысқа мәтіндер ақпараттың маңызды бөлігін құрайды. Дәстүрлі талдау әдістері (Becken et al., 2021; Chakraborty et al., 2020) мұндай форматтармен жұмыс істеуде көбінесе тиімсіз, өйткені олар аз сөзбен жоғары мағына концентрациясын қамтиды. Бұл зерттеушілерге қолайлы деректерді өңдеу құралдарын таңдауда және оларды қысқа мәтін форматтарының ерекше талаптарына бейімдеуде қиындықтар тудырады. Кластерлеу сапасы мен дәлдігіне контекстік талдаудың әсері де зерттеледі. BERT (Horta et al., 2023; He et al., 2021) сияқты заманауи

технологиялар тілді түсінуге төңкеріс жасайтын тәсілді ұсынады, бұл жүйелерге табиғи тілдің көп мағыналылығы мен күрделілігін жақсы өңдеуге мүмкіндік береді. BERT-ті LDA және AE (Fasel et al., 2024; Saranya et al., 2023) әдістерімен біріктіру талдау дәлдігін арттыру үшін мәтін құрылымдары мен семантикасын тереңірек түсінуге жол ашады. Бұл жұмыс осындай біріктірілген тәсілдердің қысқа мәтіндерді кластерлеу және жіктеу процестерін жетілдіру үшін қалай қолданылатынын көрсетуге бағытталған, олардың бірегей сипаттамалары мен қажеттіліктерін ескере отырып. Осылайша, зерттеу қысқа мәтіндерді талдау мәселесіне кешенді көзқарас ұсынып, деректерді талдау негізінде кластерлеу технологияларын дамытудың перспективалық бағыттарын зерттейді. Алынған нәтижелер табиғи тілді өңдеу (NLP) саласына (Hupkes et al., 2023; Chung et al., 2023) маңызды үлес қосатынын көрсетеді және қазіргі ақпараттық әлемдегі кең ауқымды тапсырмаларды орындауға қабілетті тиімді және бейімделгіш NLP жүйелерін құруда қолданылуы мүмкін (Xin et al., 2023; Işıkdemir et al., 2024).

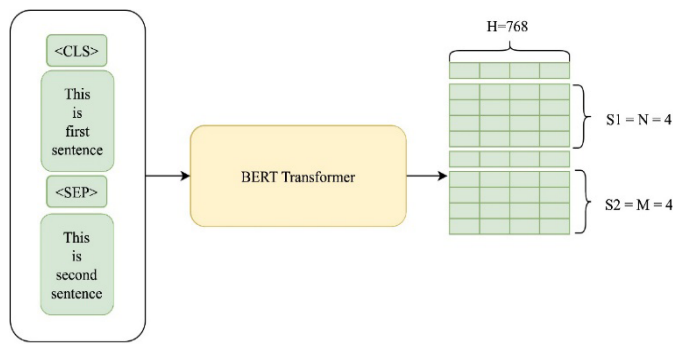
Әдістер мен материалдар

Бұл жұмыс мәтіндік деректерді талдау және кластерлеудің толық процесін қарастырады, ол екі озық табиғи тілді өңдеу (NLP) технологиясын біріктіреді: LDA (Жасырын Дирихле бөлуі) және BERT (Трансформерлерден алынған екі бағытты кодтау). Бұл «тәуелді ендіру» (dependent embedding) деп аталатын гибридті әдіс, әр әдісті жеке қолданған кезде туындайтын шектеулерді еңсеруге бағытталған. Осы әдістерден бөлек, зерттеу мәтіндегі сөздердің маңыздылығын бағалауды жақсарту үшін TF+IDF әдісін де пайдаланады. LDA әдісі (1-сурет) – тақырыптық модельдеуге арналған статистикалық тәсіл ретінде үлкен мәтіндік массивтердегі ортақ тақырыптарды анықтауға мүмкіндік береді. Алайда, LDA-ның басты кемшілігі – ол сөздердің ретін және мағынасын ескермейді, бұл мағынасы маңызды болатын мәтіндерді талдау кезінде оның қолдану аясын едәуір шектеуі мүмкін.



Сур. 1. LDA әдісінің архитектурасы

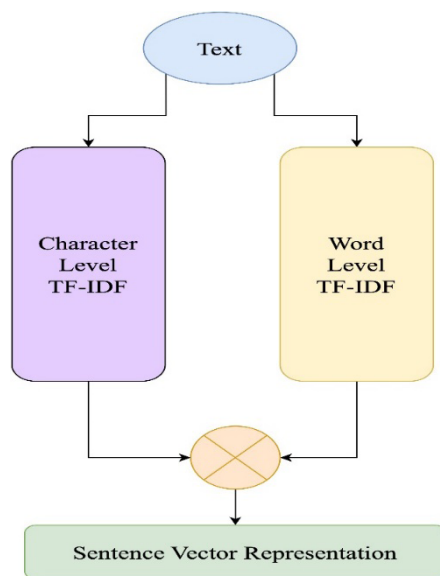
Осы шектеуді еңсеру үшін BERT моделі енгізіледі – трансформер архитектурасына негізделген терең оқыту моделі, ол сөйлемдегі әрбір сөзді басқа барлық сөздермен байланыстыра отырып талдай алады, осылайша тілдің контекстік ерекшеліктерін түсінуге мүмкіндік береді. Талдау деректерді алдын ала өңдеуден басталады, оған мәтіндердегі артық шуды жою (арнайы таңбалар мен стоп-сөздерді алып тастау), сөздерді токенизациялау және лемматизациялау жатады. Осыдан кейін LDA әдісі мәтіндер жинағындағы кең ауқымды тақырыптық кластерлерді анықтау үшін қолданылады, бұл мәтіндер арасындағы ортақ тақырыптарды табуға көмектеседі. Кейін BERT моделі әрбір сөздің векторлық көрінісін жасау үшін қолданылады, бұл мәтіндердегі семантикалық байланыстар мен мағыналық ерекшеліктерді тереңірек түсінуге мүмкіндік береді (2-сурет).



Сур. 2. BERT әдісінің архитектурасы

Бұл зерттеуде TF+IDF (Term Frequency-Inverse Document Frequency) әдісі талдау мен кластерлеу сапасын одан әрі жақсарту үшін қолданылады. Бұл әдіс сөздің маңыздылығын бағалау үшін оның белгілі бір құжатта кездесу жиілігін барлық құжаттардағы жалпы жиілігімен салыстырады. Мұндай терминдерді өлшеу тәсілі нақты мәтіндер үшін ең маңызды кілт сөздерді анықтауға мүмкіндік береді. Нәтижесінде, бұл әдіс мәтіндерді кластерлеудің келесі кезеңдеріне неғұрлым нақты және мазмұнды көрінісін қамтамасыз етеді (3-сурет).

Кластерлеу сапасын бағалау үшін қолданылатын көрсеткіштерге мыналар кіреді: силуэт — кластерлеу кезінде нысандардың бір-бірінен қалай бөлінгенін көрсетеді (неғұрлым жоғары болса, соғұрлым жақсы); calinski_harabasz индексі — кластер ішілік және кластер аралық вариацияны бағалайды (неғұрлым жоғары болса, соғұрлым жақсы); davies_bouldin — кластерлік «ұқсастық» орташа көрсеткіші (төменірек, соғұрлым жақсы); реттелген_ранды (ARI) — болжамды кластерлерді негізгі шындық белгілерімен салыстырады (неғұрлым жоғары болса, соғұрлым жақсы); гомогенділік — әрбір кластер тек бір нақты санаттағы объектілерді қаншалықты жақсы қамтитынын көрсетеді; толықтық — бір санаттағы объектілердің бір кластерге қаншалықты сәйкес келетінін көрсетеді; v_measure — біртектілік пен толықтық арасындағы гармоникалық орта (неғұрлым жоғары болса, соғұрлым жақсы).



Сур. 3. $TF+IDF$ әдісінің архитектурасы

Нәтижелер және оларды талқылау

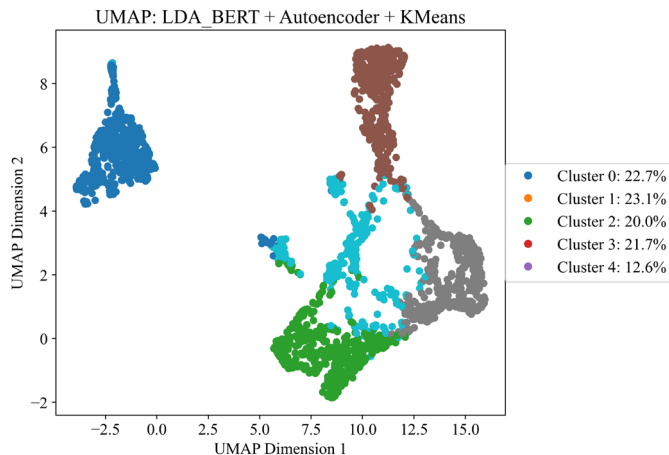
Біздің зерттеуімізде қолданылған оқыту деректер жинағы Kaggle платформасында ұсынылған және 2225 жазбадан тұрады, олар бес санатқа бөлінген: спорт, технология, бизнес, ойын-сауық және саясат. Жаңалықтар мақалаларынан алынған деректер теңгерімді және әртүрлі жиынтықты білдіреді, бұл оны әртүрлі кластерлеу әдістерін зерттеу және салыстыру үшін тамаша негізге айналдырады. Мәліметтер жинағының құрылымы машиналық оқыту және мәтіндік деректерді түсіну мен алдын ала белгіленген санаттарға жіктеуге бағытталған жобалар үшін қолайлы болып табылады. Бұл деректер жинағы екі негізгі бағанға бөлінген: «Category» (Санат) бағаны әрбір мақаланы жоғарыда аталған санаттардың біріне жатқызады, мұнда «спорт» санаты ең жиі кездеседі. «Text» (Мәтін) бағаны жаңалық мақаласының толық мәтінін қамтиды. Мақалалардың мәтіндері ұзындығы мен мазмұны бойынша әртүрлі, олардың әрқайсысына бірегейлік беретін санат ішіндегі әртүрлі тақырыптарды қамтиды. Алайда, бірдей мәтіні бар қайталанатын деректер де кездеседі. Бұл деректер жинағы мәтінді жіктеу модельдерін әзірлеу және тестілеу үшін құнды мүмкіндіктер ұсынады, алгоритмдерге мәтін мазмұнына негізделген санаттарды анықтауды үйренуге мүмкіндік береді. Атап айтқанда, бұл табиғи тілді өңдеу (NLP) саласындағы қолданбалар үшін өзекті болып табылады, мысалы: тақырыптық модельдеу, кілт сөздерді талдау, жаңалық мақалаларын веб-сайттың тиісті бөлімдеріне автоматты түрде сұрыптау жүйелерін әзірлеу. Деректер жинағының құрылымы мен мазмұны әртүрлі жаңалық мақалаларында қолданылатын тілдік ерекшеліктерді зерттеуге, сондай-ақ мәтіндік деректерді жіктеу мен талдаудың тиімді әдістерін әзірлеуге бағытталған ғылыми жобалар

үшін оңтайлы негіз бола алады.

Бірінші кезекте, әртүрлі векторизация әдістерінің (TF-IDF, BERT, LDA_BERT) Uniform Manifold Approximation and Projection (UMAP) әдісі арқылы құжаттарды кластерлеуге қалай әсер ететінін көрсетеміз. UMAP әдісі мәтіндік деректерді екі өлшемді кеңістікке түрлендіру үшін қолданылды, мұнда белгілер TF-IDF, BERT немесе гибриді LDA+BERT+AE әдістерімен алдын ала алынды. UMAP өлшемділікті бейсызықты қысқартуды жүзеге асырады және деректердің топологиялық құрылымын сақтауға мүмкіндік береді. UMAP әдісінің дәлдігі тікелей реттеу параметрлеріне (`n_neighbors`, `min_dist` және т.б.) және деректердің күрделілігіне байланысты. Бұл жағдайда UMAP әдісін қолданудың негізгі мақсаты – кластерлеу нәтижелерін визуализациялау, ал жіктеу дәлдігін бағалау емес. Екі өлшемді кеңістікте әрбір құжат нүкте ретінде бейнеленеді, ал оның түсі тиісті кластерді (K-Means әдісі арқылы алынған) білдіреді. Бұл визуализациялар объектілердің белгілі бір векторизация әдісі бойынша (TF-IDF, таза BERT эмбедингтері немесе LDA мен BERT-тің гибриді үйлесімі) қаншалықты топтасатынын немесе араласып кететінін көрсетеді.

Алайда, тек LDA мен BERT-ті біріктіру векторлық кеңістіктің ең оңтайлы құрылымын қамтамасыз етпеуі мүмкін. Осы мәселені шешу үшін автоэнкодер (AE) енгізілді – бұл өзіндік оқытылатын нейрондық желі, ол LDA+BERT үйлесімінің векторын неғұрлым ықшам латенттік көрініске қысу (кодтау) және оны қайтадан бастапқы түрге шығару (декодтау) арқылы деректерді оңтайландырады. Бұл тәсіл артық ақпаратты жою және ең маңызды вариациялық факторларды анықтау үшін қолданылады. Түрлендірілген құжаттарды кластерлеу үшін кластерлер саны k болатын K-Means алгоритмі қолданылады. K-Means алгоритмі кластерлердің орталықтары мен нүктелер арасындағы квадраттық қашықтықтардың қосындысын минимизациялау арқылы итеративті түрде жұмыс істейді, осылайша векторлық кеңістікте ұқсас құжаттардан құралған топтарды қалыптастырады. K-Means әдісі оның қарапайымдылығымен, кең таралғандығымен және әртүрлі векторизация әдістерінің тиімділігін бастапқы бағалау үшін жеткілікті болуымен таңдалды (TF-IDF, BERT, LDA_BERT).

4-суретте екі қуатты мәтінді талдау әдісін – LDA (Latent Dirichlet Allocation) және BERT (Bidirectional Encoder Representations from Transformers) біріктіретін «контекст-тақырыптық бекіту» деп аталатын синтезделген тәсілді қолдану арқылы жүзеге асырылған кластерлеу нәтижелері көрсетілген. Бұл гибриді әдістің басты мақсаты – әр әдісті жеке қолданудың негізгі шектеулерін еңсеру, яғни LDA-ның статистикалық тақырыптық модельдеуі мен BERT-тің терең контекстік тілдік түсінігін біріктіру. Мұндай интеграция мәтіндік деректердің семантикалық және контекстік аспектілерін тереңірек зерттеуге мүмкіндік береді, бұл өз кезегінде мазмұнды дәлірек түсінуді қамтамасыз етеді. Бұл кластерлеу мен кейінгі талдаудың тиімділігін арттыру үшін шешуші рөл атқарады.

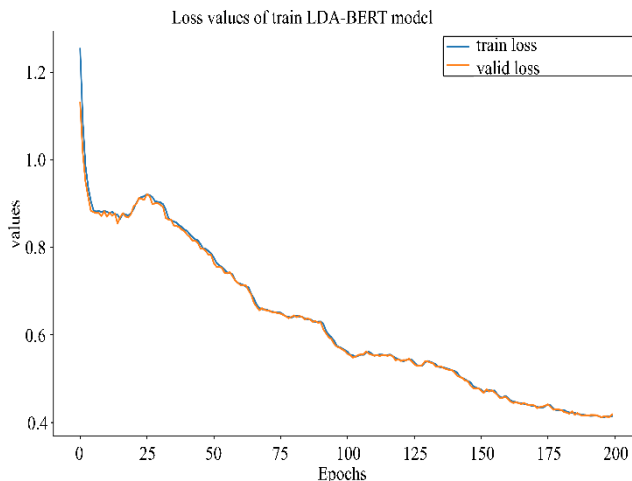


Сур. 4. LDA+BERT+AE әдісі арқылы кластерлеу нәтижесі

5-суретте LDA+BERT+AE гибриді моделінің 200 эпох бойы оқытылу процесі көрсетілген. Оқыту барысында модельдің шығын функциясы (loss) айтарлықтай төмендейді, бұл оның тиімді түрде үйрену және бейімделу қабілетін көрсетеді. Абсцисса осінде оқыту эпохтарының саны, ал ордината осінде шығын функциясының мәндері бейнеленген. Оқытудың бастапқы кезеңінде нөлінші эпохада оқыту және тексеру жиынтықтарындағы шығын мәндері шамамен 1.25 деңгейінде. Бұл оқытудың бастапқы кезеңдеріне тән қалыпты көрсеткіш, өйткені модель әлі оңтайландырылмаған және оның параметрлері жаңадан реттелуде. Алайда, алғашқы 25 эпох ішінде модель айтарлықтай жылдам төмендеу динамикасын көрсетіп, шамамен 0.9 мәніне жетеді. Бұл кезеңде модель өз салмақтары мен параметрлерін белсенді түрде бейімдеп, жаттығуды жедел жүргізеді.

25-тен 75-ші эпохаға дейінгі аралықта оқыту және тексеру жиынтықтарының шығын қисықтары біртіндеп тегістеліп, тұрақты бола бастайды, аздаған ауытқулар байқалады. Мұндай ауытқулар модельдің параметрлерін нақты баптау процесін білдіреді, бұл оның деректердің ерекшеліктеріне бейімделуіне және артық үйренуден (overfitting) аулақ болуына ықпал етеді. 200-ші эпоханың соңына қарай екі қисық шамамен 0.4 мәніне тұрақтанады, бұл оқыту процесінің сәтті аяқталғанын көрсетеді. Оқыту және тексеру қисықтарының ұқсас болуы LDA+BERT+AE моделінің жоғары жалпылау қабілетін көрсетеді және оның жаңа деректермен тиімді жұмыс істей алатынын дәлелдейді. Модельдің тиімділігі оның қолдану нәтижелерінде де расталды. LDA+BERT+AE гибриді моделін қолдана отырып, мәтіндерді кластерлеу мысалдары жоғары дәлдік көрсеткіштерін көрсетті, нәтижелері іс жүзінде шынайы мәндерге сәйкес келеді. Мысалы, бір жағдайда модель «спорт» тақырыбын 98.20 % дәлдікпен дұрыс анықтады, бұл шын мәніндегі тақырыппен толық сәйкес келді. Сол сияқты, модель «бизнес» және «саясат» сияқты басқа санаттарға да мәтіндерді жоғары

дәлдікпен жіктей алды, бұл оның кластерлеу мен мәтіндік деректерді талдау міндеттерінде жоғары тиімділік пен сенімділікке ие екенін растайды. Осылайша, LDA+BERT+AE гибриді моделі оқыту процесін сәтті аяқтап, тәжірибелік міндеттерде үздік нәтижелер көрсетті, бұл оны қысқа мәтіндерді талдау және табиғи тілді өңдеу (NLP) саласындағы басқа да міндеттер үшін аса пайдалы құралға айналдырады.



Сур. 5. Гибриді LDA+BERT+AE үлгісін оқыту кезінде жоғалту мәнінің өзгеруі

Нәтижелер бақыланатын классификация жоғары ARI/біртектілік мәндеріне қол жеткізетінін көрсетті. Дегенмен, бақыланбайтын кластер салыстырмалы нәтиже береді. Бұл алынған векторларда (әсіресе LDA_BERT) деректер құрылымы туралы сапалы ақпарат бар екенін көрсетеді. Зерттеудің қайталануы мен ашықтығын қамтамасыз ету үшін гибриді модельдің бастапқы коды және талдау әдістері қоғамдық игілікке орналастырылды. Код келесі сілтеме бойынша репозиторийде қолжетімді: <https://github.com/JamalbekTussupov01/Text-clustering/tree/main>. Жарияланған код деректерді алдын ала өңдеуді, маңызды мүмкіндіктерді таңдауды, параметрді талдауды және ауытқуды болжауды қоса алғанда, модельді іске асырудың барлық кезеңдерін қамтиды. Бұл зерттеушілер мен практиктерге ұсынылған тәсілді өз міндеттері үшін пайдалануға және қажет болған жағдайда жақсартулар жасауға және әдістемені әртүрлі жағдайларға бейімдеуге мүмкіндік береді. Бұл тәсіл ғылыми жұмыстың ашықтығын арттыруға көмектеседі және ашық ғылыми талқылауды қолдайды.

Қорытынды

Бұл зерттеу жұмысында BERT, TF+IDF және гибриді тәсіл LDA+BERT+AE сияқты қысқа мәтіндерді интеллектуалды кластерлеу әдістерінің егжей-тегжейлі талдауы ұсынылған. Зерттеу нәтижелері қысқа мәтіндермен жұмыс істеу кезінде әрбір әдістің артықшылықтары мен шектеулері бар екенін көрсетеді. Атап айтқанда, BERT контекстік тәуелділіктерді

түсінудің мықты қабілетін көрсетеді, ал TF+IDF құжаттардағы терминдердің маңыздылығын бағалайды. Дегенмен, LDA генеративті модельдеу, BERT терең контекстік талдау және автокодерлерді пайдалана отырып өлшемділікті азайту мүмкіндігін біріктіретін LDA+BERT+AE гибриді әдісі қысқа мәтіндерді кластерлеуде ең тамаша тиімділікті көрсетті. Нәтижелер гибридік әдістің жоғары тиімділігін растады, әсіресе мәтіндерге терең мағыналық талдауды қажет ететін тапсырмаларда. Атап айтқанда, 200 дәуірден астам LDA-BERT моделін оқыту және тексеру кезінде жоғалту мәндері 1,2-ден жоғары басталып, алғашқы 25 дәуірде тез төмендеп, шамамен 0,8-ге жетті. Жоғалту мәндері жаттығу барысында төмендеп, 200-ші дәуірде шамамен 0,4 тұрақтанды. Оқыту және валидация жоғалту қисықтарының тығыз үйлесуі модельдің практикалық оқуын және жақсы жалпылау мүмкіндіктерін көрсетті, сонымен қатар шектен шығудың минималды белгілері бар. Бұл зерттеудің практикалық салдары NLP жүйелерін оңтайландыруда жатыр, ол маркетинг, әлеуметтік медиа және автоматтандырылған мазмұнды модерациялау сияқты салаларда мәтіндік деректерді талдау сапасын айтарлықтай жақсарты алады. Сонымен қатар, әдіс көптілді мәтіндік талдауға бейімделу мүмкіндігін көрсетеді, бұл әсіресе жаһандық платформалар мен трансұлттық компаниялар үшін өзекті. Дегенмен, әдістің шектеулері бар, оның ішінде еңбек пен ресурстарды көп қажет ететін табиғаты шағын бизнес пен стартаптар үшін қиындықтар тудыруы мүмкін. Бұған қоса, әдіс өнімділігі деректер сапасына қатты байланысты, өйткені нашар немесе толық емес деректерді алдын ала өңдеу кластерлеу нәтижелеріне кері әсер етуі мүмкін.

Болашақ зерттеу бағыттары есептеу шығындарын азайту үшін алгоритмдерді оңтайландыруды, деректерді өңдеу жылдамдығын арттыруды және әдісті биомедицина, құқық және экономика сияқты басқа салаларда қолдану үшін бейімдеуді қамтиды. Сондай-ақ деректерді талдау сапасы мен жылдамдығын жақсарту үшін бұл әдісті GPT немесе трансформаторлар сияқты басқа үлгілермен одан әрі интеграциялау мүмкіндігі бар. Қорытындылай келе, бұл зерттеу LDA+BERT+AE сияқты гибриді интелектуалды кластерлеу әдістері қысқа мәтінді өңдеудің дәлдігі мен тиімділігін арттырудың перспективалы перспективаларын ұсынатынын растайды. Нәтижелер қазіргі ақпаратқа негізделген әлемде көптеген тапсырмаларды шешуге қабілетті неғұрлым жетілдірілген және бейімделген NLP жүйелерін дамытуға жол ашады.

ӘДЕБИЕТТЕР

Маниас Г., Маврогиоргу А., Киуртис А., Симвулидис С. және Кириазис Д. (2023). Мәтінді көптілді санаттарға бөлу және көңіл-күйді талдау: twitter деректерін жіктеудің көптілді тәсілдерін салыстырмалы талдау. *Нейрондық Есептеу Және Қолдану*. — 35(29). — Б. 21415–21431, doi: 10.1007/s00521-023-08629-3

Фу, Г., Ли, Б., Янг, Ю. Және Ли, С. (2023). Мәтінді санаттарға бөлу үшін көп сатылы агрегациясы бар ансамбльді ҚАЙТА бағалау және ТОПСИС негізінде таңдау. *Үлгіні Тану Хаттары*. — 168. — Б. 47–56, doi:10.1016/Дж.патрек.2023.02.027.

Эдара, Колумбия Округі, Ванукури, Лос-Анджелес, Систла В. және Колли В.К.К. (2023). *Көңіл-күйді*



талдау және lstm көмегімен онкологиялық аурулардың медициналық жазбаларын мәтіндік санаттарға бөлу. Қоршаған Ортаны Барлау және Ізгілендірілген Есептеу журналы. — 14(5). — 5309-5325, doi: 10.1007/s12652-019-01399-8 б.

Баладжи Т.К., Аннаварапу К.С.Р. және Баблани А. (2021). Әлеуметтік медианы талдауға арналған машиналық оқыту алгоритмдері: сауалнама. Информатикаға Шолу. — 40. — 100395, doi: 10.1016/Дж.косрев.2021.100395.

Аббас А.Ф., Джусох А., Мас' од, А., Алшариф А.Х., және Али, Дж. (2022). Әлеуметтік желілердегі ақпарат алмасудың библиометриялық талдауы. Сенімді Бизнес Және Менеджмент. — 9 (1). — 2016556. doi: 10.1080/23311975.2021.2016556.

Маккитрик, М. К., Шуурман, Н., & Алаяқтар, В.А. (2023). Орналасқан жеріне негізделген әлеуметтік медиа деректерін жинау, талдау және визуализациялау: ГАЗ-дағы әдістерге шолу-әлеуметтік медианы талдау. Геожурнал. — 88(1). — Б. 1035–1057, doi: 10.1007/s10708-022-10584- 6.

Бекен, П., Фридл, Х., Стэнтик, Б., Коннолли, Р.М. Және Чен, Дж. (2021). Климаттық дағдарыс және ұшу: Әлеуметтік медианы талдау «ұшу ұятының»өсуін қадағалайды. Тұрақты Туризм журналы. — 29 (9). —Б. 1450–1469, doi: 10.1080/09669582.2020.1851699.

Чакабортти К., Бхаттачария П., және Баг Р. (2020). Әлеуметтік медиа деректерінен көңіл-күйді талдау сауалнамасы. Есептеу Әлеуметтік Жүйелеріндегі IEEE Транзакциялары. — 7 (2). — 450–464, doi:10.1109/TCSS.2019.2956957.

Хорта Рибейро, М., Ченг, Дж., Және Уэст, Р. (2023, Сәуір). Мазмұнды автоматтандырылған модерациялау қауымдастық нұсқауларын сақтауды арттырады. АСМ 2023 веб-конференциясының Материалдарында. — 2666–2676 беттер. doi: 10.1145/3543507.3583275.

Ух К., Хонг Ю. және Рагу Т.С. (2021). Машиналық платформаны басқарудың әсері: мазмұнды модерациялауды эмпирикалық зерттеу. SSRN 3767680 мекен-жайы бойынша қол жетімді.

Фазель М. & Weerts P. (2024). Facebook қауымдастығының стандарттары заңды сенімділікке сәйкес келе ала ма? Сандық Қызметтер Туралы Заңның қысымымен мазмұнды модерациялауды басқару. Саясат Және Интернет, doi: 10.1002 / roi3.391.

Сарания П., және Уша Г. (2023). Twitter-дегі Көңіл-күйді Талдау Үшін IntelligentWordNet Lemmatize Көмегімен Машиналық оқытуға негізделген Әдіс. Интеллектуалды Автоматтандыру Және Жұмсақ Есептеу. — 36 (1).

Хапкс Д., Джулианелли М., Данкерс В., Артеткс М., Элазар Ю., Пиментель Т. & Джин З. (2023). Nlp-дегі жалпылама зерттеулердің таксономиясы және шолуы. Табиғат Машинасының Интеллектісі. — 5(10). — 1161–1174. doi: 10.1038/s42256-023-00729- ы.

Чанг П., Мун П., Кім Дж., Кім Дж., Лим П., және Чи П. (2023). Құрылыстағы және информатикадағы табиғи тілді өңдеу (NLP) қолданбаларын жүйелі шолулар (PRISMA) үшін таңдаулы есеп беру элементтерін пайдалана отырып салыстыру. Құрылыстағы автоматтандыру. — 154. — 105020, doi: 10.1016/j.aut-con.2023.105020.

Син К., Ол Ю., Пан Ю., Ван Ю. және Ду П. (2023). NLP алгоритміне негізделген ЖАСАНДЫ ИНТЕЛЛЕКТКЕ негізделген жарнамалық итеру жүйесін енгізу. — *Халықаралық Информатика Және Ақпараттық Технологиялар Журналы*. — 1 (1). — 30-37, doi: 10.62051/ijcsit.v1n1.05.

Ишикдемир, Ю.Е. (2024). NLP ТРАНСФОРМАТОРЛАРЫ: LLM-ДІ ТАЛДАУ ЖӘНЕ МӘТІНДІ КЕҢЕЙТІЛГЕН ҚОРЫТЫНДЫЛАУДЫҢ ДӘСТҮРЛІ ТӘСІЛДЕРІ. *Эскишехир Османгази Университетінің Инженерлік-Сәулет Факультетінің Журналы*. — 32(1). — Рр. 1140–1151, doi: 10.31796/ogummf.1303569.

REFERENCES

Manias G., Mavrogiorgou A., Kiourtis A., Symvoulidis C. & Kyriazis D. (2023). Multilingual text categorization and sentiment analysis: a comparative analysis of multilingual approaches for classifying twitter data. *Neural Computing and Applications*. — 35(29). — 21415–21431, doi: 10.1007/s00521-023-08629-3. (in Eng.)

Fu G., Li B., Yang Y. & Li C. (2023). Re-ranking and TOPSIS-based ensemble feature selection with multi-stage aggregation for text categorization. *Pattern Recognition Letters*. — 168. — Pp. 47–56, doi: 10.1016/j.patrec.2023.02.027. (in Eng.)

Edara D.C., Vanukuri L.P., Sistla V. & Kolli V.K.K. (2023). Sentiment analysis and text categorization of cancer medical records with LSTM. *Journal of Ambient Intelligence and Humanized Computing*. — 14(5). — Pp. 5309–5325, doi: 10.1007/s12652-019-01399-8. (in Eng.)

Balaji T.K., Annavarapu C.S.R. & Bablani A. (2021). Machine learning algorithms for social media analysis: A survey. *Computer Science Review*. — 40. — 100395, doi: 10.1016/j.cosrev.2021.100395. (in Eng.)

Abbas A.F., Jusoh A., Mas' od A., Alsharif A.H. & Ali J. (2022). Bibliometrix analysis of information sharing in social media. *Cogent Business & Management*. — 9(1). — 2016556, doi: 10.1080/23311975.2021.2016556. (in Eng.)



McKittrick M.K., Schuurman N. & Crooks V.A. (2023). Collecting, analyzing, and visualizing location-based social media data: review of methods in GIS-social media analysis. *GeoJournal*. — 88(1). — Pp. 1035–1057, doi: 10.1080/s10708-022-10584-w. (in Eng.)

Becken S., Friedl H., Stantic B., Connolly R.M. & Chen J. (2021). Climate crisis and flying: Social media analysis traces the rise of “flightshame”. *Journal of Sustainable Tourism*. — 29(9). — 1450–1469, doi: 10.1080/09669582.2020.1851699. (in Eng.)

Chakraborty K., Bhattacharyya S. & Bag R. (2020). A survey of sentiment analysis from social media data. *IEEE Transactions on Computational Social Systems*. — 7(2). — Pp. 450–464. doi: 10.1109/TCSS.2019.2956957. (in Eng.)

Horta Ribeiro M., Cheng J. & West R. (2023, April). Automated content moderation increases adherence to community guidelines. In *Proceedings of the ACM web conference 2023*. — Pp. 2666–2676, doi: 10.1145/3543507.3583275. (in Eng.)

He, Q., Hong Y. & Raghu T.S. (2021). The effects of machine-powered platform governance: An empirical study of content moderation. — Available at SSRN 3767680. (in Eng.)

Fasel M. & Weerts S. (2024). Can Facebook’s community standards keep up with legal certainty? Content moderation governance under the pressure of the Digital Services Act. — *Policy & Internet*, doi: 10.1002/poi3.391. (in Eng.)

Saranya, S., & Usha, G. (2023). A Machine Learning-Based Technique with Intelligent WordNet Lemmatize for Twitter Sentiment Analysis. *Intelligent Automation & Soft Computing*. — 36(1). (in Eng.)

Hupkes D., Giulianelli M., Dankers V., Artetxe M., Elazar Y., Pimentel T. & Jin Z. (2023). A taxonomy and review of generalization research in NLP. *Nature Machine Intelligence*. — 5(10). — Pp. 1161–1174, doi: 10.1038/s42256-023-00729-y. (in Eng.)

Chung S., Moon S., Kim J., Kim J., Lim S. & Chi S. (2023). Comparing natural language processing (NLP) applications in construction and computer science using preferred reporting items for systematic reviews (PRISMA). *Automation in Construction*. — 154. — 105020, doi: 10.1016/j.autcon.2023.105020. (in Eng.)

Xin, Q., He, Y., Pan, Y., Wang, Y., & Du, S. (2023). The implementation of an AI-driven advertising push system based on a NLP algorithm. *International Journal of Computer Science and Information Technology*. — 1(1). — Pp. 30–37, doi: 10.62051/ijcsit.v1n1.05. (in Eng.)

Işıkdemir Y.E. (2024). NLP TRANSFORMERS: ANALYSIS OF LLMS AND TRADITIONAL APPROACHES FOR ENHANCED TEXT SUMMARIZATION. *Eskişehir Osmangazi Üniversitesi Mühendislik ve Mimarlık Fakültesi Dergisi*. — 32(1). — Pp.1140-1151, doi: 10.31796/ogummf.1303569. (in Eng.)

**ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ ЖӘНЕ
КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАР ЖУРНАЛЫ**

**МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИНФОРМАЦИОННЫХ И
КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ**

**INTERNATIONAL JOURNAL OF INFORMATION AND
COMMUNICATION TECHNOLOGIES**

Правила оформления статьи для публикации в журнале на сайте:

<https://journal.iitu.edu.kz>

ISSN 2708–2032 (print)

ISSN 2708–2040 (online)

Собственник: АО «Международный университет
информационных технологий» (Казахстан, Алматы)

ОТВЕТСТВЕННЫЙ РЕДАКТОР
Мрзабаева Раушан Жалиқызы

НАУЧНЫЙ РЕДАКТОР
Ермакова Вера Александровна

ТЕХНИЧЕСКИЙ РЕДАКТОР
Рашидинов Дамир Рашидинович

КОМПЬЮТЕРНАЯ ВЕРСТКА
Асанова Жадыра

Подписано в печать 15.06.2025.

Формат 60x881/8. Бумага офсетная. Печать - ризограф. 9,0 п.л. Тираж 100
050040 г. Алматы, ул. Манаса 34/1, каб. 709, тел: +7 (727) 244-51-09).

Издание Международного университета информационных технологий
Издательский центр КБТУ, Алматы, ул. Толе би, 59