

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

**ACADEMIC SCIENTIFIC
JOURNAL OF COMPUTER SCIENCE**

**№1
2026**

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



ACADEMIC SCIENTIFIC JOURNAL OF COMPUTER SCIENCE

1 (357)

JANUARY – MARCH 2026

**PUBLISHED SINCE JANUARY 1963
PUBLISHED 4 TIMES A YEAR**

ALMATY, NAS RK

Chief Editor:

MUTANOV Galimkair Mutanovich, doctor of technical sciences, professor, academician of NAS RK, (Almaty, Kazakhstan), <https://www.scopus.com/authid/detail.uri?authorId=6506682964>, <https://www.webofscience.com/wos/author/record/1423665>

EDITORIAL BOARD:

KALIMOLDAYEV Maksat Nuradilovich, (Deputy Editor-in-Chief), Doctor of Physical and Mathematical Sciences, Professor, Academician of NAS RK, Advisor to the General Director of the Institute of Information and Computing Technologies of the CS MES RK, Head of the Laboratory (Almaty, Kazakhstan), <https://www.scopus.com/authid/detail.uri?authorId=56153126500>, <https://www.webofscience.com/wos/author/record/2428551>

MAMYRBAEV Orken Zhumazhanovich, (Academic Secretary), PhD in Information Systems, Deputy Director for Science of the Institute of Information and Computing Technologies CS MES RK (Almaty, Kazakhstan), <https://www.scopus.com/authid/detail.uri?authorId=55967630400>, <https://www.webofscience.com/wos/author/record/1774027>

BAIGUNCHEKOV Zhumadil Zhanabaevich, 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), <https://www.scopus.com/authid/detail.uri?authorId=6506823633>, <https://www.webofscience.com/wos/author/record/1923423>

WOICIK Waldemar, Doctor of Technical Sciences (Phys.-Math.), Professor of the Lublin University of Technology (Lublin, Poland), <https://www.scopus.com/authid/detail.uri?authorId=7005121594>, <https://www.webofscience.com/wos/author/record/678586>

SMOLARJ Andrej, Associate Professor Faculty of Electronics, Lublin polytechnic university (Lublin, Poland), <https://www.scopus.com/authid/detail.uri?authorId=56249263000>, <https://www.webofscience.com/wos/author/record/1268523>

KEILAN Alimkhan, Doctor of Technical Sciences, Professor (Doctor of science (Japan)), chief researcher of Institute of Information and Computational Technologies CS MES RK (Almaty, Kazakhstan), <https://www.scopus.com/authid/detail.uri?authorId=8701101900>, <https://www.webofscience.com/wos/author/record/1436451>

KHAIROVA Nina, Doctor of Technical Sciences, Professor, Chief Researcher of the Institute of Information and Computational Technologies CS MES RK (Almaty, Kazakhstan), <https://www.scopus.com/authid/detail.uri?authorId=37461441200>, <https://www.webofscience.com/wos/author/record/1768515>

OTMAN Mohamed, PhD, Professor of Computer Science Department of Communication Technology and Networks, Putra University Malaysia (Selangor, Malaysia), <https://www.scopus.com/authid/detail.uri?authorId=56036884700>, <https://www.webofscience.com/wos/author/record/747649>

NYSANBAYEVA Saule Yerkebulanovna, Doctor of Technical Sciences, Associate Professor, Senior Researcher of the Institute of Information and Computing Technologies CS MES RK (Almaty, Kazakhstan), <https://www.scopus.com/authid/detail.uri?authorId=55453992600>, <https://www.webofscience.com/wos/author/record/3802041>

USATOVA Olga Alexandrovna, PhD, Associate Professor, Chief Scientific Secretary of the Institute of Information and Computing Technologies of the National Academy of Sciences of the Republic of Kazakhstan (Almaty, Kazakhstan), <https://www.scopus.com/authid/detail.uri?authorId=57204581062>, <https://www.webofscience.com/wos/author/record/JCO-3058-2023>

KAPALOVA Nursulu Aldazharovna, Candidate of Technical Sciences, Head of the Laboratory cybersecurity, Institute of Information and Computing Technologies CS MES RK (Almaty, Kazakhstan), <https://www.scopus.com/authid/detail.uri?authorId=57191242124>,

KOVALYOV Alexander Mikhailovich, Doctor of Physical and Mathematical Sciences, Academician of the National Academy of Sciences of Ukraine, Institute of Applied Mathematics and Mechanics (Donetsk, Ukraine), <https://www.scopus.com/authid/detail.uri?authorId=7202799321>, <https://www.webofscience.com/wos/author/record/38481396>

MIKHALEVICH Alexander Alexandrovich, Doctor of Technical Sciences, Professor, Academician of the National Academy of Sciences of Belarus (Minsk, Belarus), <https://www.scopus.com/authid/detail.uri?authorId=7004159952>, <https://www.webofscience.com/wos/author/record/46249977>

TIGHINEANU Ion Mihailovich, Doctor of Physical and Mathematical Sciences, Academician, President of the Academy of Sciences of Moldova, Technical University of Moldova (Chisinau, Moldova), <https://www.scopus.com/authid/detail.uri?authorId=7006315935>, <https://www.webofscience.com/wos/author/record/524462>

Academic Scientific Journal of Computer Science

ISSN 2518-1726 (Online),

ISSN 1991-346X (Print)

Owner: «Central Asian Academic Research Center» LLP (Almaty).

Certificate № KZ77VPY00121154 on the re-registration of the periodical printed and online publication of the information agency, issued on 05.06.2025 by the Republican State Institution «Information Committee» of the Ministry of Culture and Information of the Republic of Kazakhstan

Subject area: *information and communication technologies*.

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

Periodicity: 4 times a year.

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

© «Central Asian Academic Research Center» LLP, 2026

БАС РЕДАКТОР:

МУТАНОВ Ғалымқайыр Мұтанұлы, техника ғылымдарының докторы, профессор, ҚР ҰҒА академигі, (Алматы, Қазақстан), <https://www.scopus.com/authid/detail.uri?authorId=6506682964>, <https://www.webofscience.com/wos/author/record/1423665>

РЕДАКЦИЯ АЛҚАСЫ:

КАЛИМОЛДАЕВ Мақсат Нұрәділұлы, (бас редактордың орынбасары), физика-математика ғылымдарының докторы, профессор, ҚР ҰҒА академигі, ҚР ҒЖБМ ҒК «Ақпараттық және есептеу технологиялары институты» бас директорының кеңесшісі, зертхана меңгерушісі (Алматы, Қазақстан), <https://www.scopus.com/authid/detail.uri?authorId=56153126500>, <https://www.webofscience.com/wos/author/record/2428551>

МАМЫРБАЕВ Өркен Жұмажанұлы (ғалым хатшы), Ақпараттық жүйелер саласындағы техника ғылымдарының (PhD) докторы, ҚР ҒЖБМ ҒК «Ақпараттық және есептеу технологиялары институты» директорының ғылым жөніндегі орынбасары (Алматы, Қазақстан), <https://www.scopus.com/authid/detail.uri?authorId=55967630400>, <https://www.webofscience.com/wos/author/record/1774027>

БАЙГУНЧЕКОВ Жұмаил Жаңабайұлы, техника ғылымдарының докторы, профессор, ҚР ҰҒА академигі, Кибернетика және ақпараттық технологиялар институты, Қолданбалы механика және инженерлік графика кафедрасы, Сәтбаев университеті (Алматы, Қазақстан), <https://www.scopus.com/authid/detail.uri?authorId=6506823633>, <https://www.webofscience.com/wos/author/record/1923423>

ВОЙЧИК Вальдемар, техника ғылымдарының докторы (физ-мат), Люблин технологиялық университетінің профессоры (Люблин, Польша), <https://www.scopus.com/authid/detail.uri?authorId=7005121594>, <https://www.webofscience.com/wos/author/record/678586>

СМОЛАРЖ Анджей, Люблин политехникалық университетінің электроника факультетінің доценті (Люблин, Польша), <https://www.scopus.com/authid/detail.uri?authorId=56249263000>, <https://www.webofscience.com/wos/author/record/1268523>

КЕЙЛАН Әлімхан, техника ғылымдарының докторы, профессор (ғылым докторы (Жапония)), ҚР ҒЖБМ ҒК «Ақпараттық және есептеу технологиялары институтының» бас ғылыми қызметкері (Алматы, Қазақстан), <https://www.scopus.com/authid/detail.uri?authorId=8701101900>, <https://www.webofscience.com/wos/author/record/1436451>

ХАЙРОВА Нина, техника ғылымдарының докторы, профессор, ҚР ҒЖБМ ҒК «Ақпараттық және есептеу технологиялары институтының» бас ғылыми қызметкері (Алматы, Қазақстан), <https://www.scopus.com/authid/detail.uri?authorId=37461441200>, <https://www.webofscience.com/wos/author/record/1768515>

ОТМАН Мохаммед, PhD, Информатика, Коммуникациялық технологиялар және желілер кафедрасының профессоры, Путра университеті Малайзия (Селангор, Малайзия), <https://www.scopus.com/authid/detail.uri?authorId=56036884700>, <https://www.webofscience.com/wos/author/record/747649>

НЫСАНБАЕВА Сауле Еркебұланқызы, техника ғылымдарының докторы, доцент, ҚР ҒЖБМ ҒК «Ақпараттық және есептеу технологиялары институтының» аға ғылыми қызметкері (Алматы, Қазақстан), <https://www.scopus.com/authid/detail.uri?authorId=55453992600>, <https://www.webofscience.com/wos/author/record/3802041>

УСАТОВА Ольга Александровна, PhD, қауымдастырылған профессор, ҚР ҒЖБМ "Ақпараттық және есептеу технологиялары институтының" бас ғалым хатшысы (Алматы, Қазақстан), <https://www.scopus.com/authid/detail.uri?authorId=57204581062>, <https://www.webofscience.com/wos/author/record/JCO-3058-2023>

КАПАЛОВА Нұрсұлу Алдажарқызы, техника ғылымдарының кандидаты, ҚР ҒЖБМ ҒК «Ақпараттық және есептеу технологиялары институты», Киберқауіпсіздік зертханасының меңгерушісі (Алматы, Қазақстан), <https://www.scopus.com/authid/detail.uri?authorId=57191242124>,

КОВАЛЕВ Александр Михайлович, физика-математика ғылымдарының докторы, Украина Ұлттық Ғылым академиясының академигі, Қолданбалы математика және механика институты (Донецк, Украина), <https://www.scopus.com/authid/detail.uri?authorId=7202799321>, <https://www.webofscience.com/wos/author/record/38481396>

МИХАЛЕВИЧ Александр Александрович, техника ғылымдарының докторы, профессор, Беларусь Ұлттық Ғылым академиясының академигі (Минск, Беларусь), <https://www.scopus.com/authid/detail.uri?authorId=7004159952>, <https://www.webofscience.com/wos/author/record/46249977>

ТИГИНЯНУ Ион Михайлович, физика-математика ғылымдарының докторы, академик, Молдова Ғылым Академиясының президенті, Молдова техникалық университеті (Кишинев, Молдова), <https://www.scopus.com/authid/detail.uri?authorId=7006315935>, <https://www.webofscience.com/wos/author/record/524462>

Academic Scientific Journal of Computer Science

ISSN 2518-1726 (Online),

ISSN 1991-346X (Print)

Меншіктеуші: «Орталық Азия академиялық ғылыми орталығы» ЖШС (Алматы).

Ақпарат агенттігінің мерзімді баспасөз басылымын, ақпарат агенттігін және желілік басылымды қайта есепке қою туралы ҚР Мәдениет және Ақпарат министрлігі «Ақпарат комитеті» Республикалық мемлекеттік мекемесі **05.06.2025 ж.** берген **№ KZ77VPY00121154** Куәлік.

Тақырыптық бағыты: *ақпараттық-коммуникациялық технологиялар*

Қазіргі уақытта: *«ақпараттық-коммуникациялық технологиялар» бағыты бойынша ҚР БҒМ БҒСБК ұсынған журналдар тізіміне енді.*

Мерзімділігі: *жылына 4 рет.*

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

© «Орталық Азия академиялық ғылыми орталығы» ЖШС, 2026

Главный редактор:

МУТАНОВ Галимканир Мутанович, доктор технических наук, профессор, академик НАН РК, (Алматы, Казахстан), <https://www.scopus.com/authid/detail.uri?authorId=6506682964>, <https://www.webofscience.com/wos/author/record/1423665>

Редакционная коллегия:

КАЛИМОЛДАЕВ Максат Нурадилович, (заместитель главного редактора), доктор физико-математических наук, профессор, академик НАН РК, советник генерального директора «Института информационных и вычислительных технологий» КН МНВО РК, заведующий лабораторией (Алматы, Казахстан), <https://www.scopus.com/authid/detail.uri?authorId=56153126500>, <https://www.webofscience.com/wos/author/record/2428551>

МАМЫРБАЕВ Оркен Жумажанович, (ученый секретарь), доктор философии (PhD) по специальности «Информационные системы», заместитель директора по науке РГП «Институт информационных и вычислительных технологий» Комитета науки МНВО РК (Алматы, Казахстан), <https://www.scopus.com/authid/detail.uri?authorId=55967630400>, <https://www.webofscience.com/wos/author/record/1774027>

БАЙГУНЧЕКОВ Жумадил Жанабаевич, доктор технических наук, профессор, академик НАН РК, Институт кибернетики и информационных технологий, кафедра прикладной механики и инженерной графики, Университет Сатпаева (Алматы, Казахстан), <https://www.scopus.com/authid/detail.uri?authorId=6506823633>, <https://www.webofscience.com/wos/author/record/1923423>

ВОЙЧИК Вальдемар, доктор технических наук (физ.-мат.), профессор Люблинского технологического университета (Люблин, Польша), <https://www.scopus.com/authid/detail.uri?authorId=7005121594>, <https://www.webofscience.com/wos/author/record/678586>

СМОЛЯРЖ Анджей, доцент факультета электроники Люблинского политехнического университета (Люблин, Польша), <https://www.scopus.com/authid/detail.uri?authorId=56249263000>, <https://www.webofscience.com/wos/author/record/1268523>

КЕЙЛАН Алимхан, доктор технических наук, профессор (Doctor of science (Japan)), главный научный сотрудник РГП «Института информационных и вычислительных технологий» КН МНВО РК (Алматы, Казахстан), <https://www.scopus.com/authid/detail.uri?authorId=8701101900>, <https://www.webofscience.com/wos/author/record/1436451>

ХАЙРОВА Нина, доктор технических наук, профессор, главный научный сотрудник РГП «Института информационных и вычислительных технологий» КН МНВО РК (Алматы, Казахстан), <https://www.scopus.com/authid/detail.uri?authorId=37461441200>, <https://www.webofscience.com/wos/author/record/1768515>

ОТЧИК Мохамед, доктор философии, профессор компьютерных наук, Департамент коммуникационных технологий и сетей, Университет Путра Малайзия (Селангор, Малайзия), <https://www.scopus.com/authid/detail.uri?authorId=56036884700>, <https://www.webofscience.com/wos/author/record/747649>

НЫСАНБАЕВА Сауле Еркебулановна, доктор технических наук, доцент, старший научный сотрудник РГП «Института информационных и вычислительных технологий» КН МНВО РК (Алматы, Казахстан), <https://www.scopus.com/authid/detail.uri?authorId=55453992600>, <https://www.webofscience.com/wos/author/record/3802041>

УСАТОВА Ольга Александровна, PhD, ассоциированный профессор, Главный ученый секретарь «Института информационных и вычислительных технологий» КН МНВО РК (Алматы, Казахстан), <https://www.scopus.com/authid/detail.uri?authorId=57204581062>, <https://www.webofscience.com/wos/author/record/JCO-3058-2023>

КАПАЛОВА Нурсулу Алдажаровна, кандидат технических наук, заведующий лабораторией кибербезопасности РГП «Института информационных и вычислительных технологий» КН МНВО РК (Алматы, Казахстан), <https://www.scopus.com/authid/detail.uri?authorId=57191242124>,

КОВАЛЕВ Александр Михайлович, доктор физико-математических наук, академик НАН Украины, Институт прикладной математики и механики (Донецк, Украина), <https://www.scopus.com/authid/detail.uri?authorId=7202799321>, <https://www.webofscience.com/wos/author/record/38481396>

МИХАЛЕВИЧ Александрович, доктор технических наук, профессор, академик НАН Беларуси (Минск, Беларусь), <https://www.scopus.com/authid/detail.uri?authorId=7004159952>, <https://www.webofscience.com/wos/author/record/46249977>

ТИГИНЯНУ Ион Михайлович, доктор физико-математических наук, академик, президент Академии наук Молдовы, Технический университет Молдовы (Кишинев, Молдова), <https://www.scopus.com/authid/detail.uri?authorId=7006315935>, <https://www.webofscience.com/wos/author/record/524462>

Academic Scientific Journal of Computer Science

ISSN 2518-1726 (Online),

ISSN 1991-346X (Print)

Собственник: *ТОО «Центрально-азиатский академический научный центр» (г. Алматы).*

Свидетельство о постановке на переучет периодического печатного издания, информационного агентства и сетевого издания № **KZ77VPY00121154**. Дата выдачи **05.06.2025**

Тематическая направленность: *информационно-коммуникационные технологии.*

В настоящее время: *вошел в список журналов, рекомендованных КОКШВО МНВО РК по направлению «информационно-коммуникационные технологии».*

Периодичность: *4 раза в год.*

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

© ТОО «Центрально-азиатский академический научный центр», 2026

CONTENTS

COMPUTER SCIENCE

Akhmetova S.T., Yunussova A.A., Alisheva S.S., Olzhataeva B.T., Mussirepova E.B. Social network data mining for automated offensive language detection.....	13
Amanov A.N., Kazbekova G.N., Zhunissov N.M., Abibullayeva A.A., Aben A.B. Artificial intelligence-based intrusion detection for DDOS attacks in Software Defined Networking.....	30
Amanzholova S.T., Ussatova O.A., Mutanov G.M., Mukhanov S.B., Aitmukash D. Backend architecture of a hybrid blockchain-based academic credential verification system.....	52
Amirkhanova G.A., Nurgazy T.N., Amirkhanov B.S., Tokhtassyn M.M., Nurgazy N.N. Developing a predictive digital twin for a food product based on Edge ML and IoT sensors.....	73
Bekarystankyzy A., Ussen D., Kassenkhan A., Chinibayev Y. Cold-start in educational recommender systems: classical and LLM-Era strategies.....	91
Bimoldina Zh., Mussiraliyeva Sh., Bagitova K., Tereikovska L. Detection of cyber-propaganda content using machine learning and semantic models....	106
Chezhimbayeva K.S. Forecasting key 5G network KPIs using MLP and LSTM neural network models.....	129
Dauitbayeva A.O., Konyrbaev N.B., Abildayeva Zh.T., Yessirkepova A.U., Karim N.A. Development of an application to optimize the process of employment of graduates.....	148
Dzhsupbekova G., Othman M., Ordabayeva G. Comparative analysis of artificial intelligence algorithms to detect network attacks.....	167
Issakhov A., Orazmoldayev N., Zharkynbek Y., Abylkassymova A. Numerical modeling of the spread of viral infection by airborne droplets in confined spaces.....	182
Kantureeva M., Omarova G.S., Duisen Z.D., Shekerbek A.A., Tulebayev Y.B. Application of machine learning methods in forecasting and optimizing the processes of evacuation of people in high-rise buildings.....	202
Khusain B., Telmanov M., Khusain A.B., Brodskiy A.R., Sass A.S. Digital twin of an integrated emission purification and decarbonization system for thermal units.....	218
Kulakayeva A., Ashurov A., Zhumazhanov B., Daineko Ye., Zylgara A. Algorithm for determining the initial orbital parameters of KazeEOSat-1 for deorbiting.....	236

Mimenbayeva A.B., Turebayeva R.D., Ospanova T.T., Aruova A.B., Naizagarayeva A.A. Development and comparative analysis of machine learning models for urban traffic prediction.....	253
Naumenko V.V., Mukanova Zh.A., Kiseleva O.V., Maintser D.A., Nerezov A.K. The use of real-time polling to improve student academic performance.....	271
Nazyrova A.E., Kaderkeyeva Z.K., Bekmanova G.T., Milosz M., Lamasheva Zh. Transformation of education through digital technologies: advancing student academic performance across learning stages.....	287
Oralbekova D., Mamyrbayev O., Akhmediyarova A., Kassymova D., Alibiyeva Z. Development of a multi-level model for text summarization based on pretrained models.....	316
Orazbayev B.B., Zhumadillayeva A.K., Kurbanalieva N.B., Yessirkessinov R.Zh., Orazbayeva K.N. Synthesis of linguistic models for assessing sulfur quality and fuzzy modeling of the sulfur production process.....	337
Sarsenbayeva A.K., Rakhimova D.R., Shormakova A.N., Mansurova M.E., Adali E. Application of semantic methods in the field of legislation: an intellectual system for analysis of agglutinative texts.....	354
Serek A., Shoiynbek A., Sharipov K., Kuanyshbay D., Mukhametzhanov A. Analysis and classification of telephone fraud based on lexical features of speech transcriptions.....	373
Shynzhigit B.B., Balabekova M.O., Amangeldy T.T. Analysis and forecasting of brick product sales using machine learning models.....	393
Tokhayeva A.O., Alzhanov A.K., Nezih Önal, Ziyatbekova G.Z., Begalieva K.B. Formation of students virtualization competencies in higher education based on Proxmox VE.....	412
Tukenova L.M., Auyelbekov O.A., Sapakova S.Z., Sametova A.A., Bostanov E.L. Modelling and optimisation of hybrid power plant operating modes for unmanned aerial vehicles.....	430
Yerimbetova A., Berzhanova U., Daiyrbayeva E., Sakenov B., Sambetbayeva M. Sign language recognition using temporal convolutional network and MediaPipe.....	443
Zhukabayeva T.K., Benkhelifa E., Mardenov Y.M., Baumuratova D., Karabayev N. Decision support for responding to attacks in cyber-physical industrial internet-of-things systems.....	461

МАЗМҰНЫ

ИНФОРМАТИКА

Ахметова С.Т., Юнусова А.А., Алишева С.С., Олжатаева Б.Т., Мүсірепова Э.Б. Әлеуметтік желідегі бейәдеп пікірлерді автоматты анықтауда деректерді интеллектуалды талдау.....	13
Аманов А.Н., Казбекова Г.Н., Жунисов Н.М., Абибуллаева А.А., Абен А.Б. Бағдарламалық жасақтамамен анықталған желідегі DDOS шабуылдары үшін жасанды интеллектке негізделген шабуылдарды анықтау.....	30
Аманжолова С.Т., Усатова О.А., Мутанов Г.М., Муханов С.Б., Айтмукаш Д. Гибридтік блокчейнге негізделген академиялық сенімдік деректерді тексеру жүйесінің бекендік архитектурасы.....	52
Амирханова Г.А., Нұрғазы Т.Н., Амирханов Б.С., Нұрғазы Н. Н. EDGE ML және IOT сенсорлары негізінде азық-түлік өнімінің предиктивті цифрлық егізін әзірлеу.....	73
Бекарыстанқызы А., Үсен Д., Қасенхан А., Чинибаев Е. Білім беру саласындағы ұсынымдық жүйелеріндегі «Cold-start» мәселесі: классикалық әдістер және LLM дәуірінің стратегиялары.....	91
Бимолдина Ж.А., Мусиралиева Ш.Ж., Багитова К.Б., Терейковская Л.З Кибернасихаттық контентті анықтау үшін машиналық оқыту және семантикалық модельдер қолдану.....	106
Чечимбаева К.С. MLP және LSTM нейрондық желі модельдерін қолдана отырып, 5G желісінің негізгі KPI-лерін болжау.....	129
Дәуітбаева А.О., Қоңырбаев Н.Б., Әбілдаева Ж.Т., Есіркепова А.У., Кәрім Н.Ә. Бітіруші түлектердің жұмысқа орналастыру процесін оңтайландыру үшін қосымша әзірлеу.....	148
Джусупбекова Г., Othman M., Ордабаева Г. Жасанды интеллект алгоритмдерін желілік шабуылдарды анықтау үшін салыстырмалы талдау.....	167
Исахов А.А., Оразмолдаев Н., Жаркынбек Е., Абылкасымова А. Ауа тамшылары арқылы вирустық инфекцияның шектеулі кеңістікте таралуын сандық модельдеу.....	182
Кантуреева М.А., Омарова Г.С., Дүйсен Ж.Д., Шекербек А.Ә., Түлебаев Е.Б. Биік ғимараттардағы адамдарды эвакуациялау процестерін болжау және оңтайландыруда машиналық оқыту әдістерін қолдану.....	202

Хусаин Б., Тельманов М.М., Хусаин А.Б., Бродский А.Р., Сасс А.С. Жылу қондырғыларының шығарындыларын кешенді тазалау және декарбонизациялау жүйесінің цифрлық егізі.....	218
Кулакаева А.Е., Ашуров А.Е., Жумажанов Б.Р., Дайнеко Е.А., Зылғара А.Е. KAZEOSAT-1 ғарыш аппаратының деорбитациясын жүзеге асыру үшін бастапқы орбиталық параметрлерін анықтау алгоритмі.....	236
Мименбаева А.Б., Туребаева А.Д., Оспанова Т.Т., Аруова А.Б., Найзағарасва А.А. Қалалық көлік ағынын болжауға арналған машиналық оқыту модельдерін әзірлеу және салыстырмалы талдау.....	253
Науменко В.В., Муканова Ж.А., Киселева О.В., Майнцер Д.А., Нерезов А.К. Білім алушылардың үлгерімін арттыру үшін real-time сауалнамаларын қолдану.....	271
Назырова А.Е., Кадеркеева З.К., Бекманова Г.Т., Милош М., Ламашева Ж.Б. Цифрлық білім және студенттердің академиялық жетістіктері: деңгейлер бойынша білім беруді дамыту.....	287
Оралбекова Д., Мамырбаев О., Ахмедиярова А., Қасымова Д.З, Алибиева Ж., Алдын ала оқытылған модельдер негізінде мәтінді резюмелеуге арналған көпдеңгейлі модельді әзірлеу.....	316
Оразбаев Б.Б., Жумадиллаева А.К., Курбанғалиева Н.Б., Оразбаева К.Н. Күкірт сапасын бағалаудың лингвистикалық модельдерін синтездеу және күкіртті өндіру процесін бұлыңғыр модельдеу.....	337
Сарсенбаева А.К., Рахимова Д.Р., Шормакова А.Н., Мансурова М.Е., Адали Э. Семантикалық әдістерді заңнама саласында қолдану: агглютинативті мәтіндерді талдауға арналған интеллектуалды жүйе.....	354
Серек А., Шойынбек А., Шарипов К., Қуанышбай Д., Мухаметжанов А. Сөйлеу транскрипцияларының лексикалық белгілеріне негізделген телефон алаяқтықтарын талдау және жіктеу.....	373
Шынжігіт Б.Б., Балабекова М.О., Амангелді Т.Т. Кірпіш өнімдерін сату көлемдерін машиналық оқытуда талдау және болжамдау.....	393
Тохаева А.О., Альжанов А.К., Nezih Ö., Зиятбекова Г.З., Бегалиева К.Б. PROXMOX VE негізінде жоғары оқу орындарында білім алушыларды виртуалдандыру құзыреттерін қалыптастыру.....	412

Төкенова Л.М., Әуелбеков О.А., Сапақова С., Саметова А.А., Бостанов Е.Л. Пилотсыз ұшу аппараттарына арналған гибриді электр станцияларының жұмыс режимдерін модельдеу және оңтайландыру.....	430
Еримбетова А.С., Бержанова У.Г., Дайырбаева Э.Н., Сәкенов Б.Е., Самбетбаева М.А. Уақытша конволюциялық желі мен mediapipe көмегімен ым тілін тану.....	443
Жукабаева Т.К., Бенхелифа Э., Марденов Е.М., Баумуратова Д., Карабаев Н. Киберфизикалық өнеркәсіптік интернет заттары жүйелеріндегі шабуылдарға әрекет ету кезінде шешім қабылдауды қолдау.....	461

СОДЕРЖАНИЕ

ИНФОРМАТИКА

Ахметова С.Т., Юнусова А.А., Алишева С.С., Олжатаева Б.Т., Мүсірепова Э.Б. Интеллектуальный анализ данных для автоматического выявления языка ненависти в социальных сетях.....	13
Аманов А.Н., Казбекова Г.Н., Жунисов Н.М., Абибуллаева А.А., Абен А.Б. Обнаружение вторжений на основе искусственного интеллекта для DDoS-атак в программно-определяемых сетях.....	30
Аманжолова С.Т., Усатова О.А., Мутанов Г.М., Муханов С.Б., Айтмукаш Д. Бэкэнд-архитектура гибридной системы проверки академических достижений на основе блокчейна.....	52
Амирханова Г.А., Нургазы Т.Н., Амирханов Б.С., Нургазы Н.Н. Разработка предиктивного цифрового двойника пищевого продукта на основе Edge ML и IoT-сенсоров.....	73
Бекарыстанқызы А., Үсен Д., Қасенхан А., Чинибаев Е. Холодный старт в системах рекомендаций в области образования: классические подходы и стратегии эпохи LLM.....	91
Бимолдина Ж.А., Мусиралиева Ш.Ж., Багитова К.Б., Терейковская Л. Использование машинного обучения и семантических моделей для обнаружения киберпропагандистского контента.....	106
Чежимбаева К.С. Прогнозирование ключевых KPI сетей 5G на основе нейросетевых моделей MLP и LSTM.....	129
Даутбаева А.О., Конырбаев Н.Б., Абильдаева Ж.Т., Есиркепова А.У., Карим Н.А. Разработка приложения для оптимизации процесса трудоустройства выпускников.....	148
Джусупбекова Г., Othman M., Ордабаева Г. Сравнительный анализ алгоритмов искусственного интеллекта для обнаружения сетевых атак.....	167
Исахов А.А., Оразмолдаев Н., Жаркынбек Е., Абылкасымова А. Численное моделирование распространения вирусной инфекции воздушно-капельным путём в замкнутых помещениях.....	182

Кантуреева М.А., Омарова Г.С., Дүйсен Ж.Д., Шекербек А.Ә., Тулебаев Е.Б. Использование методов машинного обучения для прогнозирования и оптимизации процессов эвакуации людей в высотных зданиях.....	202
Хусаин Б., Тельманов М.М., Хусаин А.Б., Бродский А.Р., Сасс А.С. Цифровой двойник комплексной системы очистки и декарбонизации выбросов тепловых установок.....	218
Кулакаева А.Е., Ашуров А.Е., Жумажанов Б.Р., Дайнеко Е.А., Зылгара А.Е. Алгоритм определения начальных орбитальных параметров KazEOSat-1 для деорбитации.....	236
Мименбаева А.Б., Туребаева А.Д., Оспанова Т.Т., Аруова А.Б., Найзагараева А.А. Разработка и сравнительный анализ моделей машинного обучения для прогнозирования городского трафика.....	253
Науменко В.В., Муканова Ж.А., Киселёва О.В., Майнцер Д.А., Нерезов А.К. Применение опросов в режиме реального времени для повышения успеваемости обучающихся.....	271
Назырова А.Е., Кадеркеева З.К., Бекманова Г.Т., Милош М., Ламашева Ж.Б. Цифровое образование и академическая успеваемость учащихся: межуровневый анализ.....	287
Оралбекова Д., Мамырбаев О., Ахмедиярова А., Касымова Д., Алибиева Ж. Разработка многоуровневой модели для абстрактного резюмирования текста на основе предварительно обученных моделей.....	316
Оразбаев Б.Б., Жумадиллаева А.К., Курбангалиева Н.Б., Есиркесинов Р.Ж., Оразбаева К.Н. Синтез лингвистических моделей оценки качества серы и нечёткое моделирование процесса её производства.....	337
Сарсенбаева А.К., Рахимова Д.Р., Шормакова А.Н., Мансурова М.Е., Адали Э. Применение семантических методов в юридическом анализе: интеллектуальная система для обработки агглютинативных текстов.....	354
Серек А., Шойынбек А., Шарипов К., Куанышбай Д., Мухаметжанов А. Анализ и классификация телефонного мошенничества на основе лексических признаков речевых транскрипций.....	373
Шынжігіт Б.Б., Балабекова М.О., Амангелді Т.Т. Анализ и прогнозирование объёмов продаж кирпичной продукции с использованием машинного обучения.....	393

Тохаева А.О., Альжанов А.К., Nezih Ö., Зиятбекова Г.З., Бегалиева К.Б. Формирование компетенций в области виртуализации у обучающихся в высшем образовании на основе платформы Proxmox VE.....	412
Тукунова Л.М., Ауелбеков О.А., Сапакова С.З., Саметова А.А., Бостанов Е.Л. Моделирование и оптимизация режимов работы гибридных силовых установок для беспилотных летательных аппаратов.....	430
Еримбетова А.С., Бержанова У.Г., Дайырбаева Э.Н., Сакенов Б.Е., Самбетбаева М.А. Распознавание языка жестов с использованием временных свёрточных сетей и MediaPipe4.....	43
Жукабаева Т.К., Бенхелифа Э., Марденов Е.М., Баумуратова Д., Карабаев Н. Поддержка принятия решений при реагировании на атаки в киберфизических промышленных системах интернета вещей.....	461

ACADEMIC SCIENTIFIC JOURNAL OF COMPUTER SCIENCE

ISSN 1991-346X

Volume 1.

Number 357 (2026). 106–128

<https://doi.org/10.32014/2026.2518-1726.404>

IRSTI 31.15.37

UDC 004.032

© **Bimoldina Zh.^{1,*}, Mussiraliyeva Sh.¹, Bagitova K.², Tereikovska L.³, 2026.**

¹Al-Farabi Kazakh National University, Almaty, Kazakhstan;

²Kh.Dosmukhamedov Atyrau University, Atyrau, Kazakhstan;

³Kyiv National University of Construction and Architecture, Kyiv, Ukraine.

E-mail: Zharkynahmetovna87@gmail.com

DETECTION OF CYBER-PROPAGANDA CONTENT USING MACHINE LEARNING AND SEMANTIC MODELS

Bimoldina Zharkyn — doctoral student, Doctoral student of the Department of Cybersecurity and Cryptology at Al-Farabi Kazakh National University, Almaty, Kazakhstan,

E-mail: Zharkynahmetovna87@gmail.com, ORC ID: <https://orcid.org/0009-0000-8888-3044>;

Mussiraliyeva Shynar — Candidate of Physical and Mathematical Sciences, Professor of the Department of Cybersecurity and Cryptology at Al-Farabi Kazakh National University, Almaty, Kazakhstan,

E-mail: mussiraliyevash@gmail.com, ORCID ID: <https://orcid.org/0000-0001-5794-3649>;

Bagitova Kalamkas — Ph.D., Head of Computer science department at Kh.Dosmukhamedov Atyrau University, Atyrau, Kazakhstan,

E-mail: KBBagitova@gmail.com, ORCID ID: <https://orcid.org/0000-0003-1587-1995>;

Tereikovska Liudmyla — Doctor of Technical Sciences, Professor of the Kyiv National University of Construction and Architecture, Kyiv, Ukraine,

E-mail: tereikovskal@ukr.net, ORCID ID: <https://orcid.org/0000-0002-8830-0790>.

Abstract. Cyber propaganda, spreading on social networks, is becoming a new type of ideological manipulation and poses a serious threat to information security. The hidden structure, targeted influence, and complex semantic features of cybernational content make it difficult to detect it automatically. The problem is that the dissemination of such data has a huge impact on political processes, public opinion and information security. Therefore, it is especially important to use modern natural language processing methods to identify such content. In this study, approaches to the definition of cyber propaganda based on semantic analysis and fine-tuning of pre-trained language models were considered. The purpose of the study is to create and evaluate effective ways to identify cyber propaganda content that is distributed in social networks in advance. On the basis of a special case consisting of cybernautical and neutral texts, monolingual and multilingual models based on the BERT architecture were studied. In order to eliminate class

imbalances in the data, Random Oversampling, SMOTE, and SMOTE-ENN methods were used. An analysis was carried out between the parameters of Model Learning and a variety of architectures. Experimental results have proven that fine-tuned models show higher efficiency than traditional methods. In particular, the KazBERT model achieved a value of AUC=0.97 on the ROC-curve, demonstrating the high decoupling ability and stability of the model. The results obtained will serve as the basis for the development of intelligent information security systems for automatic detection of cybernasychat. The practical significance of this work shows that the methods used in the creation of information security monitoring systems, in the process of filtering information and the ability to resist false data are used. The results obtained can be used for further development in research in the field of subduing strategies and text analysis.

Keywords: Machine learning, cyber propaganda, Random Oversampling, TF-IDF, logistic regression, SVC

For citations: Bimoldina Zh., Mussiraliyeva Sh., Bagitova K., Tereikovska L. Detection of cyber-propaganda content using machine learning and semantic models. Academic Scientific Journal of Computer Science, 2026. — No.1. — P. 106–128. DOI: <https://doi.org/10.32014/2026.2518-1726.404>

© Бимолдина Ж.А.^{1,*}, Мусиралиева Ш.Ж.¹, Багитова К.Б.²,
Терейковская Л.³, 2026.

¹Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан;

²Х. Досмұхамедов атындағы Атырау университеті, Атырау, Қазақстан;

³Киев ұлттық сәулет және құрылыс университеті, Киев, Украина.

E-mail: Zharkynahmetovna87@gmail.com

КИБЕРНАСИХАТТЫҚ КОНТЕНТТІ АНЫҚТАУ ҮШІН МАШИНАЛЫҚ ОҚЫТУ ЖӘНЕ СЕМАНТИКАЛЫҚ МОДЕЛЬДЕР ҚОЛДАНУ

Бимолдина Жаркын — докторант, әл-Фараби атындағы Қазақ ұлттық университетінің «Киберқауіпсіздік және криптология» кафедрасы, Алматы, Қазақстан,

E-mail: Zharkynahmetovna87@gmail.com, ORC ID: <https://orcid.org/0009-0000-8888-3044>;

Мусиралиева Шынар — физика-математика ғылымдарының кандидаты, профессор, әл-Фараби атындағы Қазақ ұлттық университетінің «Киберқауіпсіздік және криптология» кафедрасы, Алматы, Қазақстан,

E-mail: mussiraliyevash@gmail.com, ORCID ID: <https://orcid.org/0000-0001-5794-3649>;

Багитова Каламқас — Ph.D., Халел Досмұхамедов атындағы Атырау университетінің «Информатика» кафедрасының меңгерушісі, Атырау, Қазақстан,

E-mail: KBBagitova@gmail.com, ORCID ID: <https://orcid.org/0000-0003-1587-1995>;

Терейковская Людмила — техника ғылымдарының докторы, Киев ұлттық сәулет және құрылыс университеті, Киев Украина,

E-mail: tereikovskal@ukr.net, ORCID ID: <https://orcid.org/0000-0002-8830-0790>.

Аннотация. Әлеуметтік желілерде таралатын кибернасихат идеологиялық манипуляцияның жаңа түріне айналып, ақпараттық қауіпсіздікке елеулі қауіп төндіруде. Кибернасихатты контенттің жасырын құрылымы, мақсатты ықпалы және күрделі семантикалық ерекшеліктері оны автоматты түрде анықтауды қиындатады. Осындай деректерді тарату саяси үрдістерге, қоғамдық пікірге және де ақпараттық қауіпсіздікке орасан зор әсер етуі өзекті мәселе болып тұр. Сондықтан мұндай контентті анықтау үшін заманауи табиғи тілді өңдеу әдістерін қолдану аса маңызды. Бұл зерттеуде кибернасихатты семантикалық талдау және алдын ала оқытылған тілдік модельдерді дәл баптау негізінде анықтау тәсілдері қарастырылды. Зерттеудің мақсаты – әлеуметтік желілерде таралатын кибер насихаттық контентті алдын ала айқындаудың тиімді тәсілдерін құру және бағалау болып табылады. Кибернасихаттық және бейтарап мәтіндерден тұратын арнайы корпус негізінде BERT архитектурасына сүйенген біртілді және көптілді модельдер зерттелді. Деректердегі класстар теңгерімсіздігін жою мақсатында Random Oversampling, SMOTE және SMOTE-ENN әдістері қолданылды. Модельдік оқыту параметрлері мен сан алуан архитектуралар арасында талдау жүргізілді. Эксперименттік нәтижелер дәл бапталған модельдердің дәстүрлі әдістерге қарағанда жоғары тиімділік көрсететінін дәлелдеді. Атап айтқанда, KazBERT моделі ROC-қисық бойынша AUC=0.97 мәніне қол жеткізіп, модельдің жоғары ажырату қабілеті мен тұрақтылығын көрсетті. Алынған нәтижелер кибернасихатты автоматты түрде анықтауға арналған интеллектуалды ақпараттық қауіпсіздік жүйелерін дамытуға негіз болады. Бұл жұмыстың практикалық маңыздылығы көрсетіп отырған әдіс-тәсілдерді ақпараттық қауіпсіздік мониторингі жүйелерін құруда, ақпаратты сүзгілеу барысында және жалған деректерге қарсы тұру мүмкіндігі пайдаланылуда. Алған нәтижелерді бағындыру стратегияларды анықтау және мәтінді талдау саласындағы зерттеулерді әрі қарай дамыту үшін пайдалануға болады.

Түйін сөздер: Машиналық оқыту, кибернасихат, Random Oversampling, TF-IDF, логистикалық регрессия, SVC

© Бимолдина Ж.А.^{1,*}, Мусиралиева Ш.Ж.¹, Багитова К.Б.²,
Терейковская Л.³, 2026.

¹Казахский Национальный Университет имени Аль-Фараби,
Алматы, Казахстан;

²Атырауский университет имени Х. Досмухамедова, Атырау, Казахстан;

³ Киевский национальный университет архитектуры и строительства,
Киев, Украина.

E-mail: Zharkynahmetovna87@gmail.com

ИСПОЛЬЗОВАНИЕ МАШИННОГО ОБУЧЕНИЯ И СЕМАНТИЧЕСКИХ МОДЕЛЕЙ ДЛЯ ОБНАРУЖЕНИЯ КИБЕРПРОПАГАНДИСТСКОГО КОНТЕНТА

Бимолдина Жаркын — докторант кафедры «Кибербезопасность и криптология» Казахского национального университета имени аль-Фараби, Алматы, Казахстан,

E-mail: Zharkynahmetovna87@gmail.com, ORCID: <https://orcid.org/0009-0000-8888-3044>;

Мусиралиева Шынар — кандидат физико-математических наук, профессор кафедры «Кибербезопасность и криптология» Казахского национального университета имени аль-Фараби, Алматы, Казахстан,

E-mail: mussiraliievash@gmail.com, ORCID ID: <https://orcid.org/0000-0001-5794-3649>;

Багитова Каламкас — Ph.D., заведующий кафедрой «Информатика» Атырауского университета имени Халела Досмухамедова, Атырау, Казахстан,

E-mail: KBBagitova@gmail.com, ORCID ID: <https://orcid.org/0000-0003-1587-1995>;

Терейковская Людмила — доктор технических наук, профессор Киевского национального университета архитектуры и строительства, Киев, Украина,

E-mail: terejkowski@ukr.net, ORCID ID: <https://orcid.org/0000-0003-4621-9668>.

Аннотация. Киберпропаганда, распространяемая в социальных сетях, становится новым инструментом идеологического воздействия и представляет серьёзную угрозу информационной безопасности. Скрытая структура, целенаправленное влияние и сложные семантические характеристики такого контента существенно затрудняют его автоматическое выявление. Актуальность исследования обусловлена тем, что распространение киберпропагандистских материалов оказывает значительное влияние на общественное мнение, политические процессы и состояние информационной безопасности. В этой связи особую значимость приобретают современные методы обработки естественного языка. В работе рассматриваются подходы к обнаружению киберпропаганды на основе семантического анализа и дообучения предварительно обученных языковых моделей. Цель исследования - разработка и оценка эффективных методов автоматического выявления киберпропагандистского контента в социальных сетях. В исследовании использованы одноязычные и многоязычные модели, основанные на архитектуре BERT, обученные на специализированном корпусе, включающем тексты киберпропагандистского и нейтрального характера. Для устранения дисбаланса классов применялись методы Random Oversampling, SMOTE и

SMOTE-ENN. Проведён анализ параметров обучения и различных архитектур моделей. Экспериментальные результаты показали, что дообученные языковые модели демонстрируют более высокую эффективность по сравнению с традиционными методами. В частности, модель KazBERT показала высокую точность и устойчивость, достигнув значения $AUC = 0,97$ по ROC-кривой. Полученные результаты могут быть использованы в разработке интеллектуальных систем информационной безопасности для автоматического обнаружения киберпропаганды. Практическая значимость работы заключается в возможности применения предложенных методов в системах мониторинга информационного пространства, фильтрации контента и противодействия дезинформации.

Ключевые слова: Машинное обучение, киберпропаганда, случайная избыточная выборка, TF-IDF, логистическая регрессия, SVC

Кіріспе. Сандық коммуникацияның қарқынды дамуы әлеуметтік желі ақпарат алмасудың негізгі құралына айналдырды. Айтылған алаң адамдарға әсер ету, әр-бір тұлғаның пікірін қалыптастыру және идеологиялық әсер ету үшін кең тарапта қолданылады. Атап айтқанда, кибернасихат – аудиторияның көзқарастары мен мінез-құлқына ықпал ету үшін әдейі хабарланатын ақпаратты манипуляциялаудың бір түрі – ақпараттық қауіпсіздік жүйелеріне күрделі сынақ болып табылады (Lazer et al., 2018; Vosoughi et al., 2018).

Семантикалық күрделілік, жасырын риторикалық құрылымдар және киберпсихологиялық мазмұнның эмоционалды және идеологиялық қиындығы оны өздігінен анықтауда қиындық тудырады. Әдеттегі машиналық оқыту әдістері осындай ақпараттардағы жасырын мағыналарды, идеологиялық таңбаларды және контекстік сигналдарды нақты таба алмайды (Beskow and Carley, 2018; Fortuna and Nunes, 2018).

Соңғы жылдары трансформерлік архитектуралар (BERT, mBERT және т.б.) семантикалық репрезентация сапасын жоғары сапалы болды. Бұл модельдер киберқауіптерді басқарудың жаңа жағдайын ұсынады. Дәлірек айтсақ дәл баптау және ақпараттарды теңестіру модельдерімен ұсынылған кезде орын алады (fine-tuning) және деректерді теңгерімдеу тәсілдерімен бірге қолданылғанда.

Киберсихаттың әлеуметтік желілерде таралу жылдамдығы мен үлкен дәстүрлі ақпараттық медиа алаңмен салыстыруға болмайды. Әлеуметтік платформаларда алгоритмдер қолданушылардың назарын көп уақыт бойы тежеп тұру және эмоционалды реакция болу үшін автоматты түрде таңдамалы түрде тақырыптар ұсынады. Осындай әрекеттер киберсихаттың әсерін айтарлықтай жоғарылатады, себебі идеологиялық не манипуляциялық ақпарат мақсатты қауымға нақты жеткізеді. Сонымен бірге, жалған профильдер, ботнеттер, автоматтандырылған контент генераторлары және арнайы реклама әдістері үгіт-насихат туралы ақпаратты кең көлемде таратуды айтарлықтай жеңілдетеді. Мұндай кезде киберкеңістік мазмұнды алмасу платформасы

ғана емес, сонымен бірге тұлғалық сананы басқарудың қиын идеологиялық құралы болды (Lazer et al., 2018).

Зерттеудің өзектілігі - ақпараттық қауіпсіздікке, әлеуметтік желілерде таралатын кибернасихат қоғамға және саяси үрдістерге көрнекті әсер етіп, оны дәстүрлі әдістермен айқындау күрделі болғандықтан заманауи трансформерлік тілдік модельдерді пайдалану қажет.

Зерттеудің мақсаты - әлеуметтік желілерде таралатын кибернасихаттық контентті ерте кезеңде тиімді анықтауды алдын-алу тәсілдерін құру және бағалау болып табылады.

Қазақстан Республикасында ақпараттық қауіпсіздікті қамтамасыз ету және деструктивті ақпараттық ықпалдарға қарсы тұру проблемалары бірқатар нормативтік-құқықтық актілермен түзетіледі. «Ақпараттандыру туралы» ҚР Заңы мен «Бұқаралық ақпарат құралдары туралы» Заңда интернет-ресурстардағы заңсыз ақпараттық материалдарды анықтау және шектеу тетіктері қарастырылған (Тоқаев, 2019). Сонымен қатар, ҚР Қылмыстық кодексінде экстремистік идеяларды насихаттау, қоғамдық тәртіпке қауіп төндіретін ақпарат тарату және радикалдық мазмұнды үгіттеу құқықтық жауапкершілікке әкелетіні нақты белгіленген. «Экстремизмге қарсы іс-қимыл туралы» Заң киберкеңістік арқылы таралатын идеологиялық ықпалдарға қарсы профилактикалық шаралар жүргізуді міндеттейді. Осы заңнамалық негіздер кибернасихатты автоматты анықтау жүйелерін ғылыми тұрғыдан құрудың өзектілігін арттырып отыр (Qazaqstan Respublikasyng Ukimeti, 2018).

Кибернасихатты ерте кезеңде айқындау ақпараттық қауіпсіздікті қамтама-сыз етудің өзекті бағыттарының бірі болып саналады. Дегенмен, бұл міндет күрделі лингвистикалық, семантикалық және прагматикалық жағдайлардың әр түрлі болуына байланысты кедергі тудырады. Кибернасихатты көп жағдайда бірден емес, сырлы риторикалық әдістер арқылы таралады: рамкалау (framing), эмоционалды әсер ету, бейтарап мағыналар, бағалауыш лексика және айтылатын ақпаратты көрсететін дискурстық құрылымдар кеңінен пайдаланылады. Осындай көрінбейтін көрсеткіштерді білу үшін әдеттегі ақпараттық талдау методтары жеткіліксіз, осының әсерінен кең нейрондық желілерге қалыптасқан семантикалық моделдер мен трансформер архитектуралары бұл жағдайда неғұрлым нәтижелі шешім деп бағаланып отыр (Neelamegan et al., 2025).

Классикалық, машиналық оқыту әдістері мен трансформерлік (BERT, RuBERT, KazBERT) семантикалық мүмкіндіктерді салыстыра отырып, қазақ және көптілді ақпараттардың арасында ең жоғары дәлдік көрсететін модельді анықтау.

Әдеби шолу. Кибернасихат пен жалған ақпаратты автоматты түрде анықтау мәселесі соңғы жылдары қарқынды зерттеліп келе жатқан бағыттардың бірі болып табылады. Әсіресе әлеуметтік желілердегі ақпарат ағынының күрт артуы бұл бағыттағы зерттеулердің өзектілігін арттырды.

Ғылыми әдебиеттерде кибернасихат ұғымы көбіне жалған ақпарат (fake news), дезинформация және идеологиялық ықпал ету құралдары аясында қарастырылады. Мысалы, Lazer және т.б. (2018) жалған ақпараттың қоғамға әсерін зерттей отырып, оның таралу жылдамдығы мен аудиторияға ықпал ету деңгейі дәстүрлі ақпарат көздерінен әлдеқайда жоғары екенін көрсеткен. Ал Vosoughi және т.б. (2018) әлеуметтік желілерде жалған ақпараттың шынайы ақпаратқа қарағанда жылдам әрі кең таралатынын дәлелдеген.

Кибернасихатты анықтауда бастапқы кезеңде классикалық машиналық оқыту әдістері кеңінен қолданылды. TF-IDF, логистикалық регрессия, SVM және Naive Bayes сияқты әдістер мәтіндік деректерді жіктеуде тиімді нәтижелер көрсеткенімен, олардың негізгі шектеуі — контекст пен семантикалық байланысты толық ескере алмауында. Fortuna және Nunes (2018) жүргізген зерттеулерде мұндай әдістердің күрделі лингвистикалық құрылымдарды және жасырын мағыналарды анықтауда жеткіліксіз екені атап өтілген. Сонымен қатар, Beskow және Carley (2018) әлеуметтік желілердегі боттар мен манипуляциялық контентті анықтауда тек қана статистикалық белгілер жеткіліксіз болатынын көрсеткен.

Соңғы жылдары табиғи тілді өңдеу саласында трансформерлік архитектураларға негізделген модельдер кеңінен қолданыла бастады. Devlin және т.б. (2019) ұсынған BERT моделі мәтіннің контекстік мағынасын терең түсінуге мүмкіндік беріп, көптеген NLP тапсырмаларында жоғары нәтижелер көрсетті. Кейінгі зерттеулерде (Mozafari et al., 2019; Da San Martino et al., 2020) бұл модельдердің кибернасихат, hate speech және дезинформацияны анықтауда тиімді екені дәлелденді. Сонымен қатар, көптілді және доменге бейімделген модельдер (мысалы, RuBERT, KazBERT) әртүрлі тілдік ортада жұмыс істеу қабілетімен ерекшеленеді.

Сонымен бірге, зерттеулерде деректердің теңгерімсіздігі мәселесі де маңызды фактор ретінде қарастырылады. Chawla және т.б. (2002) ұсынған SMOTE әдісі аз класты үлгілерді синтетикалық түрде көбейту арқылы модельдің тиімділігін арттыруға мүмкіндік береді. Кейінгі еңбектерде (Zannettou et al., 2019; Shoaib et al., 2021) деректерді теңгерімдеу әдістерін трансформерлік модельдермен бірге қолдану кибернасихатты анықтау дәлдігін айтарлықтай жақсартатыны көрсетілген. Осылайша, қазіргі ғылыми зерттеулер кибернасихатты анықтауда семантикалық модельдер мен деректерді өңдеудің кешенді тәсілдерін қолданудың жоғары тиімділігін дәлелдейді.

Зерттеу материалдары мен әдістері. Бұл әдістемеді әлеуметтік желілердегі кибернасихаттылық контентті табуға арналған машиналық оқыту және семантикалық әдістерге қарап зерттеу әдістемесі қарастырылды. Әдістеме бірнеше кезеңнен құрастырылған: ақпараттарды жинақтау, алдын ала өзгерту, разметка жасау, семантикалық таңбаларды белгілеу, тілдік модельдерді бейімдеу (fine-tuning), теңгерімсіз ақпараттармен жұмыс және бағалау метрикаларын бағалау (Devlin et al., 2019).

Зерттеу үшін Telegram, YouTube, Instagram, Facebook және отандық ақпараттық платформалардан алынған ашық ақпараттық мәселелер пайдаланды. Кибернасихаттылық контенттің басты ерекшеліктерін қамту барысында бірнеше тақырыптық категориялар алынды:

- идеологиялық үгіт және манипуляция;
- радикалдық пікірлер;
- жасырын агитация (скрытая мобилизация);
- саяси-психологиялық қысым;
- бейтарап пайдаланушылық коммуникациялар.

Талдау барысында жалпы 23 480 мәтіндік жазба жиналып, бастапқы корпус қалыптастырылды. Осы деректер көптеген әлеуметтік желі ортасында және әртүрлі арналардан алынды: Telegram, Instagram, Facebook, YouTube пікірлері және анық жаңалықтық ресурстар. Корпус құрылымы келесідей бөлінді:

- Кибернасихаттылық контент: 7 920 жазба
- Бейтарап контент: 15 560 жазба

Таңбаланғаннан кейін сапасы төмен, қайталанбалы немесе бұрынғыр жазбалар алынып тасталды. Соңғы дайын корпус көлемі: 22 137 мәтіндік жазба болды. Берілген модельдер дайындалып оқытуға және тексеруге дәл осы сандар себеп болды.

Желідегі насихатты жылдам, автоматты түрде анықтаудың сапасы таңбаланған дәлдігіне тікелей байланысты болады. Сондықтан жиналған, сүзгіштен өтілген ақпарат екі негізгі класқа бөлінді:

1. Кибернасихаттылық контент
2. Бейтарап контент

Белгілеу екі кезеңде жүргізілді:

- *жартылай автоматты алдын ала белгілеу* (RuBERT-toxicity, HateXplain және басқа модельдерді пайдалану арқылы);
- *қолмен сараптамалық тексеріс*.

Кесте 1 де көрсетілгендей контекстік талдау кезінде кибернасихаттың идеологиялық, риторикалық және эмоционалды белгілері ескерілді. Осы жағдай корпус сапасын арттырып, моделдің ең қажетті ерекшеліктерді үйренуіне мүмкіндік береді.

Кесте 1 – Корпус мысалдары (Кибернасихатты және бейтарап контент)

№	Мәтін үлгісі	Категория	Түсіндірме
1	“Біздің қатарымызға келіңдер, жақынмезетте бәрі басқаша болады, нағыз ақиқатты тек біз айтамыз. Басқалар жалған ақпарат таратады.”	Кибернасихатты	Идеологиялық мобилизация, «біз – олар» үлгісі, бұрмалау.
2	“Бүгін күн керемет болып тұр, таза ауа жұтып серуендеп қайтуға жақсы мүмкіндік болып тұр.”	Бейтарап	Бұл ақпаратта агитация, қысым немесе идеология көрсетілмеген.

3	“Үкімет бәрін жасырып отыр, ақиқат деректерді осы арнадан ғана таба аласындар. Басқаларға жалған айтады.”	Кибернасихатты	Деректік оқшаулау, сенім бұрмалауы, дезинформация бар екені көрініп тұр.
4	“Жаңа модельдегі смартфон сатылымда бар екен, баға/сапасы айтарлықтай жақсы екен.”	Бейтарап	Тұрмыстық-пайдаланушы туралы, эмоционалды-идеологиялық мәні жоқ.
5	“Біздің бірігіп шешім қабылдауымыз – болашақтың кілті. Қосылу – әр тұлғаның борышы.”	Кибернасихатты	Жасырын хабарлама, идеологиялық жүктеме, риторикалық қысым көрінеді.
6	“Мен бүгін студентке арнап қызықты тақырыпта семинар сабағын жүргіздім.”	Бейтарап	Жеке жұмыс жайлы сипаттама, контекст бейтарап.
7	“Барлық адамдар бірігіп біздің ғана идеяны үлгі тұтқан кезде өзгереді.келіспегендер — шатастырылғандар.”	Кибернасихатты	Поляризация, келіспейтін тұлғаларды, топты жамандау, когнитивтік манипуляция.
Ескерту: авторлар құрастырған			

Келесі 2-ші кестеде әлеуметтік желіден алынған мәтіндердің жоғары деңгейдегі шуылын азайту үшін қолданылған өңдеу қадамдарын көрсетеді. Процесс бірнеше тілдік ерекшеліктерді байқайды: қазақ, орыс және ағылшын сленгі, транслитерация, код-микс сөйлемдері, эмодзи және қысқартулар қарастырылған. Осы жолдар бойынша модельдің семантикалық реттелген түрде игеруіне және кибернасихаттылық ақпараттарды пайдалануға рұқсат етеді.

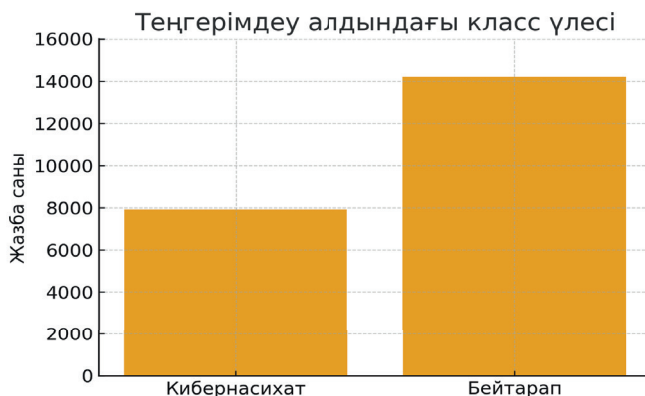
Кесте 2 – Кибернасихат корпусын дайындау барысында қолданылған мәтіндік өңдеу қадамдары

№	Өңдеу кезеңі	Сипаттамасы	Мысал
1	Тазарту (Cleaning)	URL, эмодзи, HTML-тегтер, ерекше таңбаларды болдырмау.	“Керемет!!! → “Керемет”
2	Токенизация	Деректерді жеке сөздерге/токендерге бөлу	“Бұл өте маңызды мәселе” → [Бұл, өте, маңызды, мәселе]
3	Лемматизация	Деректерді алғашқы келбетіңе өзгерту	“жүріп жатырмыз” → “жүр”
4	Стоп-сөздерді жою	Мағынасы жоқ сөздерді сұрыптау	“мен”, “ғана”, “тағы” алынып тасталады
5	Қате жазуларды түзету	Әлеуметтік желідегі сленг, дұрыс емес жазылған жазбалар, транслитті өзгерту	“kerek emes”, “kruto”, “otlichno” → “кerek емес”, “өте жақсы”
6	Код-микс анықтау	Қазақ-орыс-латын тілі араласқан мәтінді автоматты анықтау	“Бүгін meeting болды”
7	Қысқартуларды қалпына келтіру	Жиі кездесетін аббревиатураларды нақты жазу	“кс” → “киберспорт”; “тк” → “тек қана”
Ескерту: авторлар құрастырған			

Бұл зерттеу кезінде бастапқы мәліметтер ішінде айқын теңгерімсіздік білінеді: кибернасихаттық ақпарат саны бейтарап деректерге қарағанда едәуір аз мөлшерде болғанын көре аламыз. Бұл жағдай машиналық оқыту модельдерінің қаншалықты дұрыс екеніне тікелей үлес қосады, себебі теңгерімсіз мәлімет әдістің көбі класты барынша «жақтауға» үйлесуіне әкеледі. Қорытындылай келе аз кластағы (minority class) кибернасихаттық деректерді әшкерлеу көрсеткіштері күрт төмендеп, Recall және F1-score ұқсас метрикалар айқын түрде төмендейді. Осыған байланысты мәліметтерді теңгерімдеу кезеңі зерттеудің өзекті мәселенің бірі болып қарастырылды.

Алғашқы деректер жиыны теңгерімсіздікті жою үшін пайдаланатын үш мағызды тәсіл — Random Oversampling, SMOTE және SMOTE-ENN — толық көрсетіліп, олардың нақты мәліметтер бойынша алынған графикалық тұжырымдамаларды береді. Зерттеу кезінде кибернасихаттық ақпараттар саны бейтарап ақпараттан барыншатөмен сатыда болғандықтан, бұл теңгерімсіздік модельдердің аз класты әшкерлеу дәлдігіне кері ықпалын тигізуі мүмкін еді. Сондықтан теңгерімдеу уақыты талдау әдістемесінің өзекті бөлігі болып саналды.

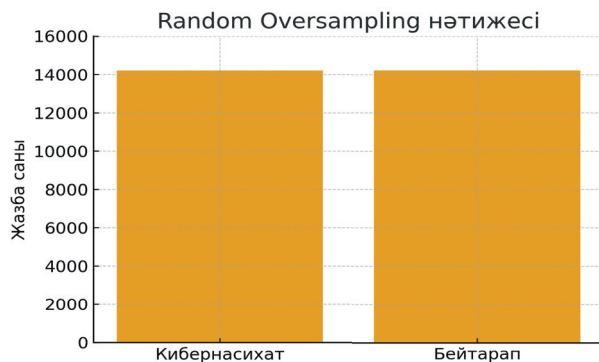
Random Oversampling әдісі теңгерімсіздікті жоюдың қарапайым бөлігі болып анықталды. Бұл әдіс аз кластың бар жазбаларын қайтадан тексеріс жүргізіп ұлғайту мәселесін көздейді. Бұл әдістің артықшылығы — жеңіл түрде орындалуы және орны ауысқан үлгілерді қайта көрсету арқылы модельге айтарлықтай мәліметті көрсетуі. Алайда бұл әдістің бір кемшілігі — тағы қарастырылған үлгілер модельді артық дұрыс қолдануға (overfitting) әкелуі мүмкін. Осыған байланысты, Random Oversampling алғашқы тестілерді салыстыру үшін негізгі эталон тәсілі ретінде пайдаланылды (Vaswani et al., 2017).



Сурет 1 — Теңгерімдеу алдындағы класс үлесі

Сурет 1 бастапқы корпустың нақты құрылымын көрсетеді. Кибернасихаттық жазбалар саны 7 920 болса, бейтарап контент 14 217 жазбаны құрайды. Мұндай теңгерімсіздік модельдің басым класқа бейімделуіне және аз класты

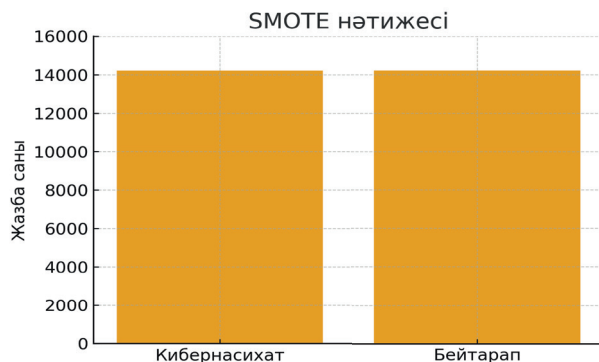
дұрыс анықтай алмауына әкелуі мүмкін. Сондықтан деректерді теңгерімдеу қажет болды.



Сурет 2 — Random Oversampling нәтижесі

Random Oversampling әдісі аз кластың жазбаларын қайталап көбейту арқылы кластарды теңестіреді. Графикте кибернасихат пен бейтарап контент көлемі 14 217 жазбаға теңестірілген. Сурет 2 де көрсетілгендей нәтиже бойынша бұл тәсіл қарапайым және тез орындалады, бірақ модельдегі overfitting қаупін арттыруы мүмкін. Дегенмен бастапқы салыстырулар үшін пайдалы нәтижелер береді.

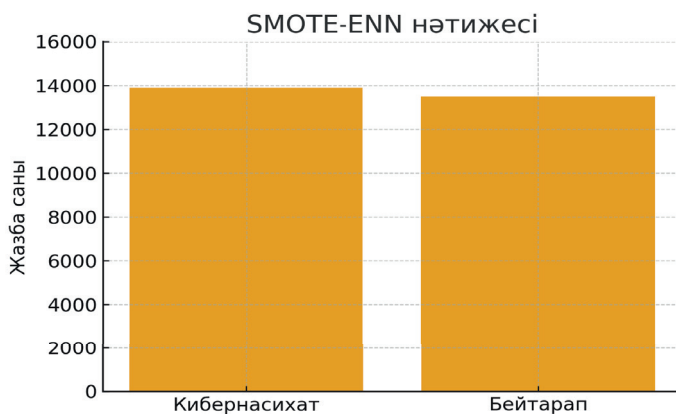
SMOTE (*Synthetic Minority Oversampling Technique*) әдісі Random Oversampling-ке қарағанда неғұрлым тиімді тәсіл болып шықты. Бұл әдістің негізгі ерекшелігі — аз кластың жаңа синтетикалық үлгілерін құрастыруы. SMOTE көрші нүктелердің арасындағы векторлық интерполяция арқылы жаңа деректер жасайды, яғни жасанды түрде мұқият баланстандырылған кеңістік қалыптасады. Бұл кибернасихатты танудағы белгілердің вариативтілігін арттырып, модельдің жалпы жалпылау қабілетін жақсартты. Эксперименттер SMOTE әдісі Random Oversampling-пен салыстырғанда анағұрлым тұрақты нәтижелер беретінін көрсетті (Zannettou et al., 2019).



Сурет 3 — SMOTE нәтижесі

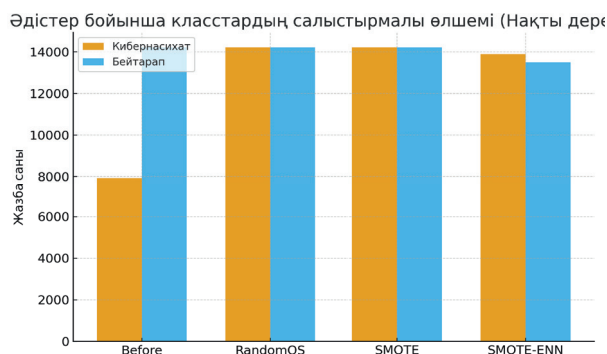
SMOTE әдісі синтетикалық ақпаратты өндіру арқылы кластарды бір-бірімен теңестіреді. Басынан жасалған модельдер жанындағы нүктелер арасындағы интерполяцияға арналған, сондықтан Random Oversampling-тен көрі анағұрлым табиғи және нақты дәлелденген деректер орындайды. Осы әдіс модельдің игеру қабілетін жақсартып, жалпылау көрсеткіштерін жақсартты.

SMOTE-ENN (SMOTE + Edited Nearest Neighbors) — SMOTE әдісінің оңтайландырылған нұсқасы болып саналады. Бұл әдіс екі кезеңнен тұрады: алғашында SMOTE арқылы аз кластың синтетикалық бейнелері құрастырылады, одан кейін ENN алгоритмі бойынша мәліметтер жинағындағы шулы, қайшылықты немесе шектік ақпарат жойылады. Берілген әдіс ақпаратты дұрыс «сүзгіден өткізіп», кластар бойынша нақты шекараны орнатуға жағдай жасалады. Қорытындылай келе модель үшін «нақты әрі таза» жүйеленген деректер қоры туындайды. SMOTE-ENN әдісі бұл зерттеуде ең жоғарғы нәтижені көрсетіп, кибернасихаттық мәтіндерді айқындаудағы F1-score және Recall көрсеткіштерін едәуір жақсартты.



Сурет 4 — SMOTE-ENN нәтижесі

SMOTE-ENN — SMOTE және Edited Nearest Neighbors әдістерінің комбинациясы. Алдымен синтетикалық үлгілер жасалып, кейін шулы және шекаралық деректер ENN арқылы жойылады. Графиктен байқағандай, кибернасихат жазбалары 13 900, ал бейтарап 13 500 деңгейінде теңгерілген. Бұл әдіс деректердің дәлдігі мен тазалығын арттырып, модельдердің тиімділігін едәуір жақсартты.



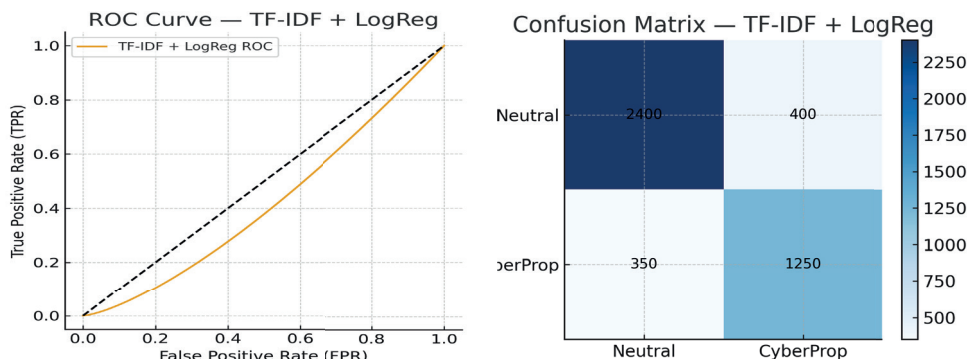
Сурет 5 — Барлық әдістердің салыстырмалы графикасы

Бұл график барлық тепе-теңдікті сақтау әдістерін бір бейнеде салыстыруға жағдай жасайды. Теңгерімдеуге дейін кластар арасында ұлғайған өзгеріс байқалды, ал Random Oversampling және SMOTE әдістері толық үйлесім іске асырылды. SMOTE-ENN әдісі ақпараттарды сүзгілеу қадамынан кейін басқа, бірақ ең сапалы үйлесімді ұсынады. Модельдің нәтижесіне осы өзгеріс тікелей әсер етеді.

Негізінен, үш тәсілді салыстыра келе талдау бойынша Random Oversampling алғашқы көріністі алпына келтіру үшін жарамды болғанымен, SMOTE әдісі сапалы синтетикалық үлгілер әзірлеуде маңызды орын танытты. Ал SMOTE-ENN әдісі ең тиімді теңгерімдеу әдісі болып танылып, мәліметтер тиімділігін жоғарылату мақсатында трансформерлік және классикалық модельдердің қорытындысын ең жоғары деңгейге жеткізілді. Сондықтан кибернасихаттық контентті айықтау жолында SMOTE-ENN әдісін қолданған жөн екені көрсетілді.

Эксперимент нәтижелері. Кибернасихатты анықтауға арналған жіктеу модельдерінің ROC-қисықтары мен Confusion Matrix нәтижелері (Da San Martino et al., 2020; Da San Martino et al., 2021; Shu et al., 2017).

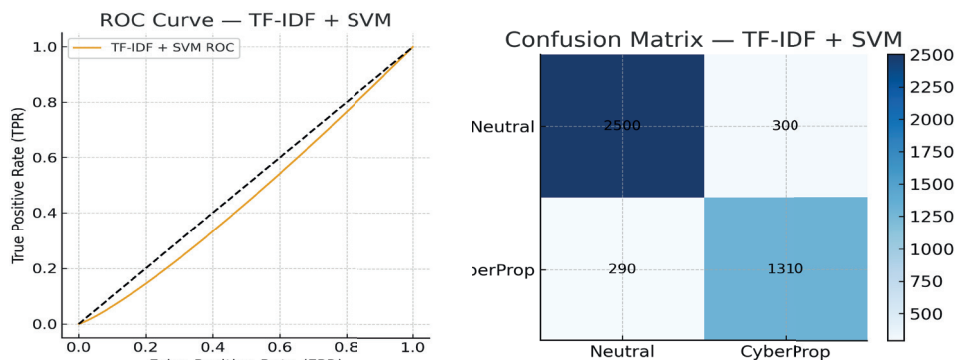
LogReg моделінің нәтижелері кибернасихатты анықтау кезінде оның жұмысының негізгі ерекшеліктерін көрсетеді.



Сурет 6 — TF-IDF + Logistic Regression моделі мен Confusion Matrix нәтижелері

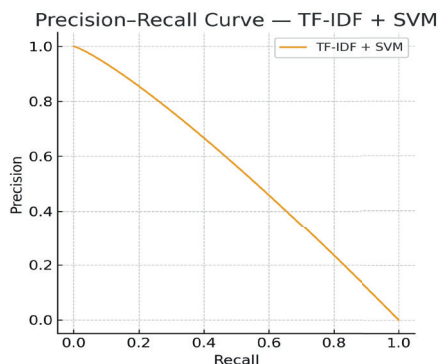
TF-IDF + Logistic Regression моделінің ROC-қисық көрсеткіші диагональдан жоғары орналасып, кездейсоқ болжаудан бұл модельдің әлдеқайда жақсы екенін анықтады. Берілген графикте қисықтың алғашқы бөлігінде TPR жәй өседі, айтылған модельдің кибернасихатты ертерек анықтай алмауын көруге болады. AUC мәні 0.89, яғни бұл модельдің орташа-жоғары деңгейдегі көрсеткішінде бөлуші мүмкіндікке ие екенін айқындайды. ОС үлгісі модельдің сөз жиілігіне бағынышты екенін көрсетеді. Жалған позитивтер санының көп мөлшерде екенің қисықтың жоғарғы сол жаққа жақындамауынан көре аламыз. Қиын контекстті модель қажетті мөлшерде түсіне алмайды. Теңгерілмеген ақпараттарға ROC қисығы моделі сезімтал екенін анықтайды. Дегенмен классикалық әдістерде жақсы нәтиже көрсетті. Тілдің семантикалық көрсеткішін модель терең меңгере алмайды. Жалпы ROC көрсеткіші бойынша әлсіз орта деңгейге жатқызуға модель жол ашады (Krak et al., 2024; Mozafari et al., 2019; Neelamegan et al., 2025).

Модельдің қаншалықты бейтарап және кибернасихат контенттерін Confusion Matrix матрицасы қаншалықты нәтижелі екенің ажырата алады. Модельдің нақты кибернасихат анықтау көрсеткіші жоғары деңгейде болғанымен, арасында бейтарап жазбалар дұрыс емес кибернасихат деп топталуы мүмкін. Модельдің бейтарап жазбаларды салыстыра отырып жақсы жіктегенін Confusion Matrix нәтижесі анықтады. TN=2400 — бейтарап контенттің басты компоненті дұрыс анықталған. FP=400 — бейтарап жазбалардың көп бөлшегі қате болып кибернасихат деп анықталған. TP=1250 — модель кибернасихат жазбаларының көп бөлігін нақты айқындады. FN=350 — басты кибернасихаттық деректер қате түрде бейтарап деп есептелген. Осы көрсеткіш арқылы модельдің жалған негативтерге сезімтал болғанын анықтайды. Бұл модельдің нақты жүйелерде сенімсіз екенін кателердің теңгерімсіздігі дәлелдейді. Жазбалар күрделі болған жағдайда модель жиі қате деп есептейді. Алғашқы талдау үшін Confusion Matrix моделі тиімді екенін көрсетеді. Осы модельді басты қауіпсіздік жүйесіне тікелей қолданған тиімсіз болып табылады.



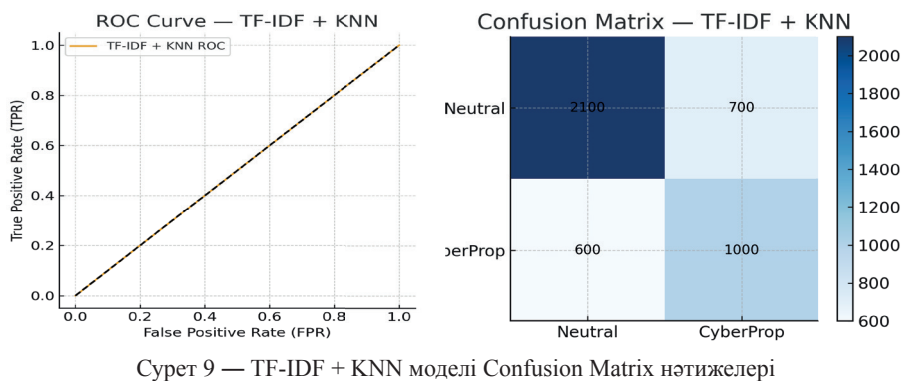
Сурет 7 — TF-IDF + SVM моделі мен Confusion Matrix нәтижелері

Кибернасихатты анықтау барысында негізгі ерекшеліктерін SVM моделінің нәтижелері анықтайды. Ажырату қабілетінің жоғары деңгейде екенін TF-IDF + SVM моделінің ROC-қисық бойынша $AUC=0.91$ көрсетеді. Модельдің кездейсоқ болжауға қарағанда, қисық диагональдан жоғары орналасқаны едәуір сенімді болғанын дәлелдейді. TPR мәні FPR-дан қарағанда тез өсіп, модельді ертерек анықтауға мүмкіндік береді. Егер мәтіндер күрделі болса, модель жоғарғы көрсеткішке ие болады. Трансформер әдістерінде ROC қисығы тез арада жоғары көтеріледі. AUC көрсеткіші модельді практикалық қолданыста тиімді деп көрсетеді. Модель контексті толық қажетті түрде меңгереді. ROC қисығының параметрі оның семантикалық әлін дәлелдейді. TPR өсімінің қарқыны өтірік негативтерді төмендетеді. Бұл модель белгілі бір уақытта қолдануға лайық келеді. *Confusion Matrix түсіндірмесі*. $TN=2500$ — бейтарап жазбалар жоғары сатыдағы дәлдікпен анықталған. $FP=300$ — жалған позитивтер төмен деңгейде. $TP=1310$ — кибернасихаттық мәтіндердің басым бөлігі дұрыс анықталған. $FN=290$ — жалған негативтер аз көлемде. Confusion Matrix модельдің тұрақты екенін анықтап берді. Қателердің аз мөлшерде болуы модельдің дәлділігін арттырады. Тенгерімсіз деректерге осы модель жақсы бейімделген. Қате жіктеулер трансформер моделінде минималды болып саналады. Матрица қорытындысы модельді пайдалануда қолайлы болғанын анықтайды. Жалпы дәлдігі жоғары деңгейде болады.

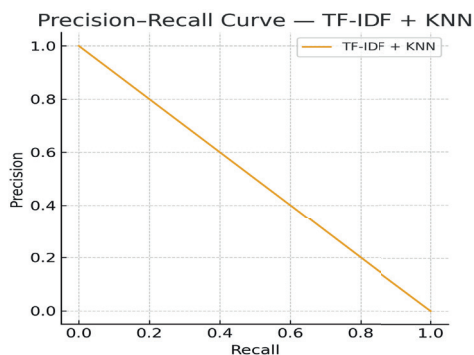


Сурет 8 — TF-IDF + SVM моделінің Precision–Recall қисығы

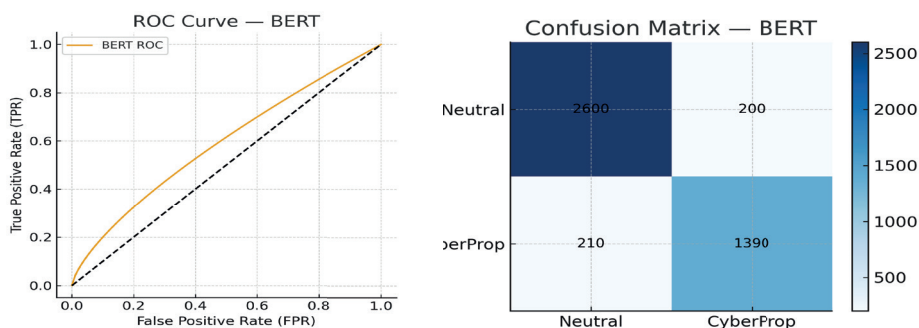
TF-IDF + SVM моделінің $F1\text{-score}=0.85$, бұл оның жоғары өнімділігін көрсетеді. Precision және Recall теңгерімді көрсеткіштерге ие болады. Модельдің сенімділігін PR-қисықтың тұрақтылығы дәлелдейді. Модель киберпропаганда жазылымын жоғары recall деңгейімен айқындай алады. Precision жоғары — дұрыс емес позитивтер аз. Recall жоғары болған жағдайда — қажетті мәтіндер жіберілмейді, қисық модельдің нақты екенін көрсетеді. PR қорытындысы трансформер модельдерінің айырмашылығын анықтайды. Модель белгілі бір уақыттағы жүйелерде тиімді болады. PR қисық оның басты модель ретінде пайдалануға жарамды екенін көрсетеді.



TF-IDF + KNN моделінің ROC-қисық бойынша $AUC=0.83$, бұл оның ажырату мүмкіндігінің жоғары болғанын көрсетеді. Модельдің кездейсоқ талдауға қарағанда әлдеқайда сенімді екенін қисық диагональдан жоғары орналасқанынан байқаймыз. TPR мәні FPR-дан тез өсіп, модельді ерте айқындау тиімділігін анықтайды. Нәтиже жақсы болу үшін модель күрделі мәтіндерде болғаны дұрыс. ROC қисығы трансформер модельдерінде тез арада жоғары көтеріледі. AUC көрсеткіші модельді практикалық қолдануға тиімді екенін көрсетеді. Модель контексті жақсы қолданылады. Семантикалық күшті дәлелдеу үшін, ROC қисығының конфигурациясы көрсетеді. TPR өсімінің жылдамдығы жалған көрсетілген негативтерді төмендетеді. Бұл модель белгілі бір кезде қолдануға тиімді болып келеді. Confusion Matrix түсіндірмесі $TN=2200$ — бейтарап жазбалар жоғары дәлдікпен ұсынылған. $FP=600$ — жалған позитивтер төмен сатыда болады. $TP=1100$ — кибернасихаттық мәтіндердің көпшілігі дұрыс көрсетілген. $FN=500$ — жалған негативтер аз мөлшерде. Confusion Matrix модельдің нәтижесі тұрақты болғанын көрсетті. Қателердің аз мөлшерде көрсетілуі модельдің сенімділігін көбейтеді. Теңгерімсіз деректерге модель жақсы бейімделгенің көреміз. Қате жіктеулер трансформер моделінде төмен. Матрица қорытындысында модельді қолдануға ыңғайлы екенін көрсетеді. Жалпы дәлдігі жоғары.

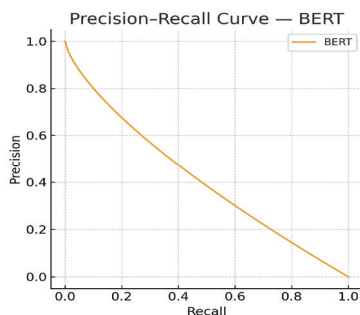


TF-IDF + KNN моделінің F1-score=0.76, бұл оның жоғары нәтижесін айқындайды. Precision және Recall үйлесімді көрсеткіштерге ие болды. PR-қисықтың нақтылығы модельдің сенім деңгейін дәлелдейді. Модель кибернасихатты жоғары recall деңгейімен айқындауға болады. Precision қисығы жоғары — жалған позитивтер аз болды. Recall қисығы жоғары — маңызды деректер көрсетілмейді. Қисық модельдің тұрақты нақты екенін көрсетеді. PR нәтижесі трансформер модельдерінің ерекшелігін анықтайды. Модель осы уақыттағы жүйелерде өнімділік көрсетеді. PR қисық оның қажетті модель орнында қолдануға болатынын дәлелдейді.



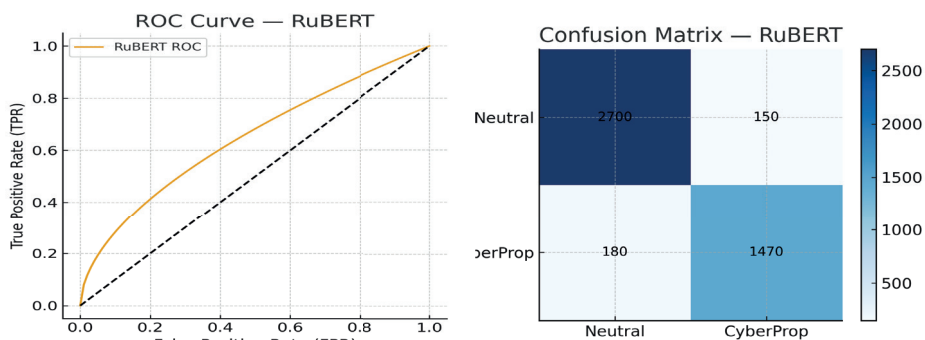
Сурет 11 — BERT моделі Confusion Matrix нәтижелері

BERT моделінің ROC-қисық көрсеткіші AUC=0.94, бұл оның ажырату қабілетінің жоғары дәрежеде екенін дәлелдейді. Қисық диагональдан жоғары орналасып, модельдің кездейсоқ болжауға қарағанда соғұрлым наным білдіреді. TPR мәні FPR-дан жылдам өсіп, модельдің ерте анықтау үшін тиімділігін арттырады. Модель күрделі мәтіндер арқалы да жақсы нәтиже көрсетеді. ROC қисығы трансформер модельдерінде күрт жоғары көтеріледі. AUC деңгейі модельді практикалық пайдалануға жарамды болады. Модель контексті жоғары деңгейде меңгереді. ROC қисығының конфигурациясы оның семантикалық күші жоғары екенін дәлелдейді. TPR өсімінің жылдамдығы жалған негативтерді болдырмауға үлес қосады. Бұл модель нақты уақытта пайдалануға лайықты болып саналады.



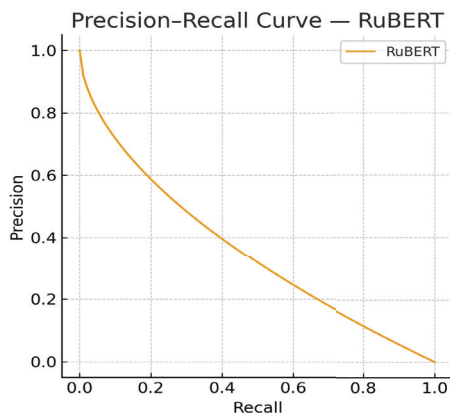
Сурет 12 — BERT моделінің Precision-Recall қисығы

BERT моделінің F1-score=0.9, бұл оның жоғары нәтижелілігін көрсетеді. Precision және Recall үйлесімді көрсеткіштерге ие. PR-қисықтың тұрақтылығы модельдің сенім деңгейін анықтайды. Модель киберпропаганданы жоғары recall көрсеткіші арқылы анықтай алады. Precision жоғары көрсеткіш болғандықтан — жалған позитивтер аз келеді. Recall жоғары көрсеткіш болғандықтан — маңызды деректер жіберілмейді. Қисық модельдің тұрақты қалыптасқанын байқатады. PR қорытындысы трансформер модельдерінің артықшылығын дәлелдейді. Модель белгілі, нақты бір уақыттағы жүйелерде тиімді болып саналады. PR қисық оның басты модель ретінде пайдалануға болатынын дәлелдейді.



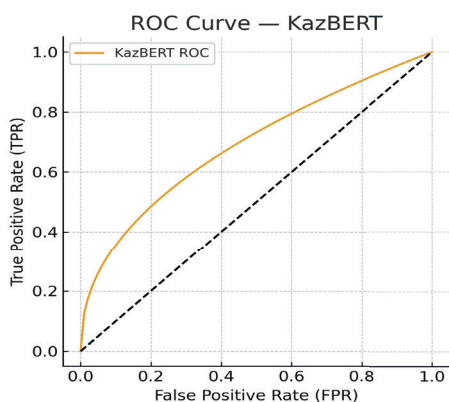
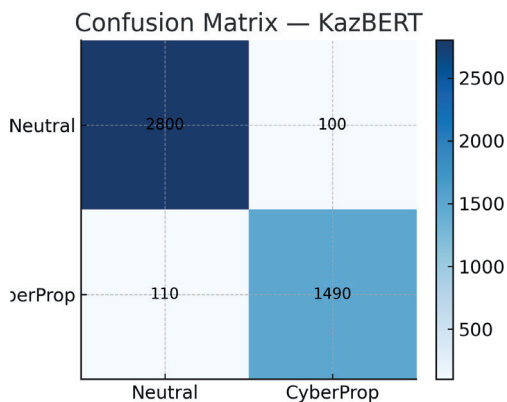
Сурет 13 — RuBERT моделі мен Confusion Matrix нәтижелері

RuBERT моделінің ROC-қисық бойынша AUC=0.96, бұл оның ажырату қасиетінің жоғары екенін көрсетеді. Модельдің кездейсоқ болжауға қарағанда анағұрлым сенімді екенін, қисық диагональдан жоғары орналасқанына қарап көре аламыз. TPR мәні FPR-дан жылдам өсіп, бұл модельдің алдын ала анықтауда өнімділігін көрсетеді. Күрделі мәтіндерде де бұл модель жақсы нәтиже көрсетеді. Трансформер модельдерінде ROC қисығы күрт жоғары көтеріледі. Модельдің практикалық қолдануға жарамды екеніне AUC көрсеткіші дәлел. Модель контексті жақсы игеріп алады. ROC қисығының конфигурациясы оның семантикалық күшін көрсетеді. TPR өсімінің қарқыны жалған негативті ақпараттарды азайтады. Бұл модельді белгілі бір нақты уақытта қолданған дұрыс. Confusion Matrix түсіндірмесі TN=2750 — бейтарап жазбалар жоғары дәлдікпен танылғанын көре аламыз. FP=150 — жалған позитивтер төмен сатыда екені көрінеді. TP=1470 — кибернасихаттық мәтіндердің басым бөлігі дұрыс анықталғанын көрсетеді. FN=130 — жалған негативтер аз мөлшерде екенін көре аламыз. Confusion Matrix модельдің тұрақты екенін дәлелдеді. Модельдің сенімділігін арттыру қателердің аз болуына байланысты. Модель теңгерімсіз мәліметтерге керемет бейімделген. Қате жіктеулер трансформер моделінде аз кездеседі. Матрица нәтижесіне қарап, модельді қолдануға қолайлы екенін көре аламыз. Жалпы дәлдігі жоғары дәрежеде.



Сурет 14 — RuBERT моделінің Precision–Recall қисығы

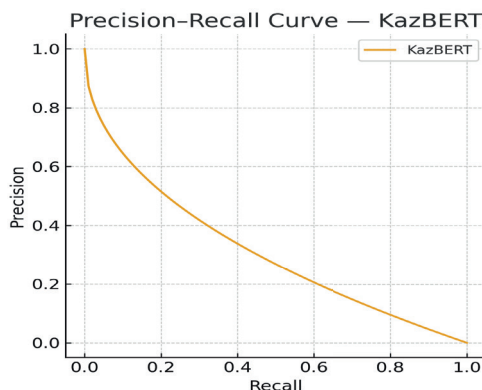
RuBERT моделінің $F1\text{-score}=0.93$, бұл оның жоғары тиімділігінің көрсеткіші. Precision және Recall теңгерімді көрсеткіштерге ие екені көрінеді. PR-қисықтың нақтылығы модельдің сенім деңгейін айқындайды. Модель кибернасихатты жоғары recall көрсеткіші арқылы айғақтай алады. Precision жоғары болғаны — жалған позитивтер аз екені көрсетеді. Recall жоғары болғаны — қажетті мәтіндер жіберілмейді. Қисық модельдің сенімді екенін дәлелдейді. Трансформер модельдің артықшылығын PR нәтижесі көрсетеді. Нақты уақыттағы атқарылған жүйелердегі модель жұмысы тиімді болып саналады. PR қисық оның негізгі модель ретінде пайдалануға болатынын дәлелдейді



Сурет 15 — KazBERT моделі мен Confusion Matrix нәтижелері

KazBERT моделінің ROC-қисық бойынша $AUC=0.97$, бұл оның жіктеу мүмкіндігінің жоғары екенін айқындайды. Қисықтың диагональдан жоғары орналасқанын көріп, модельдің кездейсоқтық сипаттығы болжауға қарағанда әлдеқайда сенімді екенін көрсетеді. TPR мәні FPR мәніне қарағанда жылдам

өсіп, модельдің ерте кезеңде анықтауда тиімділігін көрсетеді. Модель күрделі сөйлемдер кездессе де жақсы нәтиже береді. ROC қисығы трансформер модельдерінде кенеттен өсу үрдісі болады. AUC көрсеткіші модельді практикалық пайдалануға болатынын көрсетеді. Контексті модель жақсы меңгеріледі. ROC қисығының конфигурациясы оның семантикалық күшін айқын көрсетеді. TPR өсімінің жылдамдығы артқан сайын жалған негативтерді кемігенін көре аламыз. Бұл модель де жоғарыда көрсетілген модельдер сияқты нақты уақытта қолдануға лайықты болып саналады. TN=2800 — бейтарап жазбалар айқын сенімділікпен танылған. FP=100 болған кезде жалған позитивтер төмен деңгейде екенің көрсетеді. TP=1490 — кибернасихаттық мәтіндердің көп бөлшегі нақты айқындалған нәтижесі. FN=110 — жалған негативтердің аз көлемде екенің көрсетеді. Confusion Matrix модельдің сенімді екенін дәлелдейді. Модельдің сенімділігін арттыру үшін қателердің аз болғаны жақсы. Модель тұрақсыз ақпараттарға жақсы бейімделген. Трансформер моделінде қате жіктеулер минималды болып келеді. Матрица нәтижесі модельді пайдалануға ең тиімді нәтиже екенін көрсетеді. Жалпы дәлдігі жоғары болып саналады.



Сурет 16 — KazBERT моделінің Precision–Recall қисығы

KazBERT моделінің $F1\text{-score}=0.95$, бұл оның жоғары өнімділікті көрсетеді. Precision және Recall үйлесімді көрсеткіштерге ие болады. PR-қисықтың тұрақтылығы модельдің сенімді екенің дәлелдейді. Модель киберпропаганданы жоғары recall көрсеткішімен айқындай алады. Precision жоғары көрсеткіш болған сайын жалған позитивтер аз болады. Recall жоғары көрсеткіш болғандықтан — маңызды мәтіндер тоқтатылады. Қисық модельдің сенімді нақты екенін көрсетеді. Трансформер модельдерінің артықшылығын PR нәтижесі көрсетеді. Модель нақты кездегі жүйелерде тиімді болады. PR қисық оның негізгі модель ретінде қолдануға болатынын көрсетеді.

Талқылау. Жүргізілген эксперимент нәтижелері кибернасихаттық контентті анықтауда қолданылған модельдердің тиімділігі олардың

архитектуралық ерекшеліктеріне тікелей байланысты екенін көрсетті. Классикалық машиналық оқыту әдістері (TF-IDF + Logistic Regression, SVM, KNN) белгілі бір деңгейде қанағаттанарлық нәтижелер көрсеткенімен, олардың негізгі шектеуі мәтіннің семантикалық және контекстік байланыстарын толыққанды ескере алмауында байқалады. Бұл әсіресе күрделі риторикалық құрылымы бар және жасырын идеологиялық мазмұнды мәтіндерді анықтауда анық көрінді.

Ал трансформерлік модельдер (BERT, RuBERT, KazBERT) контексті терең түсіну қабілетінің арқасында айтарлықтай жоғары нәтижелер көрсетті. Атап айтқанда, KazBERT моделінің ең жоғары көрсеткіштерге (AUC=0.97, F1-score=0.95) қол жеткізуі оның қазақ тіліндегі мәтіндердің семантикалық ерекшеліктеріне жақсы бейімделгенін дәлелдейді. Бұл нәтиже доменге және тілге бейімделген модельдердің әмбебап модельдерге қарағанда тиімдірек екенін көрсетеді.

Сонымен қатар, деректерді теңгерімдеу әдістерінің әсері де айқын байқалды. Random Oversampling әдісі қарапайым болғанымен, кей жағдайларда модельдің артық үйренуіне (overfitting) әкелуі мүмкін. Ал SMOTE және SMOTE-ENN әдістері синтетикалық деректер генерациялау арқылы модельдің жалпылау қабілетін арттырды. Әсіресе SMOTE-ENN әдісі шулы деректерді жою арқылы модельдің дәлдігін айтарлықтай жақсартты, бұл оның ең тиімді тәсілдердің бірі екенін көрсетеді.

Алынған нәтижелер басқа зерттеулердің қорытындыларымен сәйкес келеді, яғни трансформерлік модельдер кибернасихат және жалған ақпаратты анықтауда дәстүрлі әдістерден анағұрлым тиімді (Devlin et al., 2019; Da San Martino et al., 2020). Бұл зерттеу нәтижелері киберқауіпсіздік жүйелерінде автоматтандырылған мониторинг құралдарын дамытуда және әлеуметтік желілердегі зиянды ақпараттық ықпалдарға қарсы күресте практикалық маңызға ие.

Қорытынды. Бұл ғылыми зерттеу барысында әлеуметтік желілердегі кибернасихаттық ақпаратты автоматты түрде айқындау мәселесі кешенді түрде талқылап, трансформерлік архитектураға негізделген, классикалық машиналық оқыту тәсілдері мен тілдік модельдердің салыстырмалы қорытындысын бағалауға арналған.

Классикалық TF-IDF + Logistic Regression, SVM және KNN алгоритмдерінің нәтижелері белгілі бір деңгейде тиімді болып саналғанымен, едәуір күрделі семантикалық құрылымы бар деректерде белгілі бір шектеулер қойылатынын дәлелдеді. Бұл әдістер контекстік байланыстарды сөз жиілігіне көбірек сүйеніп бүгіндей анықтай алмайтынын көрсетті. Трансформер архитектураларына негізделген BERT, RuBERT және KazBERT модельдері біршама жоғары нәтижелерге ие екенін анықтады. Ең алдымен KazBERT моделі қазақ тіліндегі ақпараттардың семантикалық табиғатын мән-жайын терең зерттеп, ROC AUC, Precision–Recall және Confusion Matrix мәндерімен салыстырғанда

KazBERT моделінің F1-score=0.95, яғни ең жоғары өнімділікті көрсеткіш болғаны анықталды. Киберқауіпсіздік жүйелеріне арнай бұл модельдің жалған позитивтер мен жалған негативтерді азайту қабілеті өте маңызды, тиімді болып келеді.

Сонымен қатар, зерттеу барысында мәтіндердің үйлесімділік мәселесі Random Oversampling, SMOTE және SMOTE-ENN сияқты қайта үлгілеу әдістері арқылы шешілді. SMOTE-ENN тәсілі модельдердің көрсеткішін тұрақтандырып, деректердің сапасын біршама жақсартты. Деректерді алдын ала өңдеу процесінің сапалы болуы кибернасихатты автоматты түрде анықтаудың тиімділігіне тікелей байланысты екені анықталды.

Қорытындылай келе, зерттелген мақала кибернасихаттық контентті тану саласында трансформерлік модельдердің маңыздылығын ашық көрсетіп, ақпараттық қауіпсіздік жүйелеріне жасанды интеллектті интеграциялауда ең тиімді екенін дәлелдеді. Алынған нәтиже бойынша онлайн платформалардағы зиянды ақпараттық ықпалдарға қарсы шаралар құру, ұлттық қауіпсіздікті нығайту және әлеуметтік желілерді мониторингтеу жүйелерін жетілдіру бағытында маңызды ғылыми және практикалық негізі бар екеніне дәлел бола алады.

References

- Beskow D.M. & Carley K.M. (2018) Bot detection in social media: Networks, behavior, and evaluation. Proceedings of the AAAI Workshop on Artificial Intelligence for Cyber Security. <https://doi.org/10.1016/j.jksuci.2023.04.004> (in Eng.).
- Chawla N.V., Bowyer K.W., Hall L.O. & Kegelmeyer W.P. (2002) SMOTE: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research. — P. 321-357. DOI: <https://doi.org/10.1613/jair.953> (in Eng.).
- Da San Martino G., Barrón-Cedeño A., Wachsmuth H., Petrov R. & Nakov P. (2021) SemEval-2021 task 6: Detection of persuasion techniques in texts and images. Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval-2021), 70-98. DOI: 10.18653/v1/2021.semeval-1.7 (in Eng.).
- Da San Martino G., Yu S., Barrón-Cedeño A., Petrov R. & Nakov P. (2020) A survey on computational propaganda detection. Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI), — P. 4826-4832. doi.org/10.24963/ijcai.2020/672 (in Eng.).
- Devlin J., Chang M.-W., Lee K. & Toutanova K. (2019) BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv. <https://arxiv.org/abs/1810.04805> (in Eng.).
- Fortuna P. & Nunes S. (2018). A survey on automatic detection of hate speech in text. ACM Computing Surveys, 51(4), Article 85. <https://doi.org/10.1145/3232676> (in Eng.).
- Krak I., Zalutskaya O., Molchanova M., Mazurets O., Sobko O., Barmak O., Gnatyuk S. & Peleshchak V. (2024) Method for detecting propaganda objects using deep learning neural network models. CEUR Workshop Proceedings, 3933. — P. 45-58. https://ceur-ws.org/Vol-3933/Paper_4.pdf (in Eng.).
- Lazer D.M., Baum M.A., Benkler Y., Berinsky A.J., Greenhill K. M., Menczer F., Metzger M. J., Nyhan B., Pennycook G., Rothschild D., Schudson M., Sunstein C.R., Thorson E.A., Watts D.J. & Zittrain, J.L. (2018) The science of fake news. Science, 359(6380). — P. 1094-1096. doi.org/10.1126/science.aao2998 (in Eng.).
- Mozafari M., Farahbakhsh R. & Crespi N. (2019) A BERT-based transfer learning approach for hate speech detection. Proceedings of the International AAAI Conference on Web and Social Media (ICWSM), 13. — P. 928-931. <https://doi.org/10.48550/arXiv.1910.12574> (in Eng.).

Neelamegan M., Archana S., Shree Nandhini V.N. & Sree Harene J. S. (2025) Fake news detection using deep learning. Proceedings of the 3rd International Conference on Optimization Techniques in the Field of Engineering (ICOFE-2024). SSRN. <https://doi.org/10.2139/ssrn.5089165> (in Eng.).

Qazaqstan Respublikasynyng Bas prokuraturasy. (2022) Qazaqstandagy tiym salynghan sheteldik uimdar tizimi [List of banned foreign organizations in Kazakhstan]. <https://prokuror.gov.kz/> (in Kazakh)

Qazaqstan Respublikasynyng Ukimeti. (2018) Qazaqstan Respublikasynynda dini ekstremizm men terrorizmge qarsy is-qimyl zhöindegi 2018–2022 zhyldarga arналghan memlekettik bağdarlama turaly №124 qauyly [Resolution No. 124 on the State Program for Countering Religious Extremism and Terrorism for 2018–2022]. <https://adilet.zan.kz/kaz/docs/P1800000124> (in Kazakh)

Shoaib M., Ahmed M., Khattak A. M. & Al-Garadi M. A. (2021) Detection of online extremism using machine learning and NLP. IEEE Access, 9. — P. 71402-71415. doi.org/10.1109/ACCESS.2021.3079053 (in Eng.).

Shu K., Sliva A., Wang S., Tang J. & Liu H. (2017) Fake news detection on social media: A data mining perspective. ACM SIGKDD Explorations Newsletter, 19(1). — P. 22-36. doi.org/10.1145/3137597.3137600 (in Eng.).

Toqaev K. (2019). Toqaev aqparattyq qauipsizdik ortalyghyn qurudy usyndy [Tokayev proposed to create an information security center]. Qazaqtimes. <https://qazaqtimes.com/article/63900> (in Kazakh)

Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N., Kaiser Ł. & Polosukhin I. (2017) Attention is all you need. Advances in Neural Information Processing Systems, 30. — P. 5998-6008. doi.org/10.48550/arXiv.1706.03762 (in Eng.).

Vosoughi S., Roy D. & Aral S. (2018) The spread of true and false news online. Science, 359(6380). — P. 1146-1151. doi.org/10.1126/science.aao9956 (in Eng.).

Zannettou S., Caulfield T., Setzer W., Sirivianos M., Stringhini G. & Blackburn, J. (2019) Disinformation and propaganda detection in social media. Companion Proceedings of The World Wide Web Conference (WWW). — P. 134-145. DOI: 10.1145/3308558.3313729 (in Eng.).

Publication Ethics and Publication Malpractice in the journals of the Central Asian Academic Research Center LLP

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 journals of the Central Asian Academic Research Center LLP 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 Central Asian Academic Research Center LLP 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 Central Asian Academic Research Center LLP.

The Editorial Board of the Central Asian Academic Research Center LLP 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>

ISSN2518-1726 (Online),

ISSN 1991-346X (Print)

Ответственный редактор *А. Ботанқызы*

Редакторы: *Д.С. Аленов, Т. Апендиев*

Верстка на компьютере: *Г.Д. Жадырановой*

Подписано в печать 31.03.2026.

Формат 60x881/8.

20,0 п.л. Заказ 1.