

КИБЕРБУЛЛИНГТІ АНЫҚТАУДА ТЕРЕҢ ОҚЫТУ ӘДІСТЕРІНІҢ ТИІМДІЛІГІН ТАЛДАУ

АБДРАХМАНОВ Р.Б.¹, КЕРІМХАН Б.Т.^{2*}, КАСЫМБЕКОВ А.С.¹,
ИСКАКОВ Т.Б.¹, УКСИКБАЕВ Е.З.³

Абдрахманов Рустам Бахтиёрулы¹ – Техника ғылымдарының кандидаты, доцент, Халықаралық туризм және меймандостық университеті, Түркістан қ., Қазақстан

E-mail: abdrakhmanov.rustam@iuth.edu.kz, <https://orcid.org/0000-0002-5508-389X>

*Керімхан Бекжан Темірханұлы² – PhD, «Компьютерлік және программалық инженерия» кафедрасының аға оқытушысы, Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана қ., Қазақстан

E-mail: bek_zhan_16@mail.ru, <https://orcid.org/0000-0002-8480-7316>

Касымбеков Аббазбек Сугирович¹ – Техника ғылымдарының кандидаты, доцент, Халықаралық туризм және меймандостық университеті, Түркістан қ., Қазақстан

E-mail: abbazbek@mail.ru, <https://orcid.org/0000-0001-8022-5637>

Искаков Тайыржан Бахытбаевич¹ – PhD, доцент, Халықаралық туризм және меймандостық университеті, Түркістан қ., Қазақстан

E-mail: t.iskakov@iuth.edu.kz, <https://orcid.org/0000-0002-7499-9297>

Уксикбаев Еркебулан Зейнуллаұлы³ – Магистр, Қ.А.Ясауи атындағы Халықаралық қазақ-түрік университеті, Түркістан қ., Қазақстан

E-mail: erkebulan.uxikbayev@ayu.edu.kz, <https://orcid.org/0009-0003-3868-6773>

Андатпа. Зерттеудің мақсаты – цифрлық коммуникациялардағы мәтіндік деректерден кибербуллинг белгілерін автоматты түрде анықтайтын терең оқыту әдістерінің тиімділігін салыстыру және неғұрлым нәтижелі нейрондық архитектураны айқындау. Кибербуллинг ақпараттық қоғамдағы өзекті әлеуметтік-психологиялық мәселелердің бірі болып табылады және әсіресе жасөспірімдер мен жастардың психоэмоциялық жағдайына кері әсер етеді. Әлеуметтік желілер мен онлайн-платформалардағы мәтіндік контент көлемінің өсуі қорлау, агрессия және қоқан-лоққы сипатындағы хабарламаларды ерте анықтайтын автоматтандырылған жүйелерді әзірлеуді талап етеді.

Зерттеу барысында таңбаланған мәтіндер корпусы негізінде көпқабатты персептрон (MLP), конволюциялық нейрондық желі (CNN), рекуррентті нейрондық желі (RNN), қысқа және ұзақ мерзімді жад желісі (LSTM), сондай-ақ назар аудару механизмі бар терең нейрондық модель зерттелді. Модельдердің сапасы дәлдік, толықтық, оң болжамдардың дәлдігі және F1 өлшемі бойынша бағаланды. Нәтижелер MLP моделінің күрделі контекстік байланыстарды ескеруде шектеулі екенін, CNN архитектурасының жергілікті тілдік үлгілерді тиімді анықтайтынын, ал RNN және LSTM модельдерінің сөздер реттілігі мен ұзақ мерзімді тәуелділіктерді жақсы өңдейтінін көрсетті. Ең жоғары жіктеу сапасын назар аудару механизмі бар модель көрсетті, себебі ол агрессия, қорқыту және қорлау белгілері бар маңызды мәтін фрагменттеріне басымдық береді.

Зерттеу нәтижелері модельдердің тиімділігі таңбаланған деректердің сапасы мен теңгерімділігіне тәуелді екенін көрсетті. Мәтінді алдын ала өңдеу, сөздерді біріздендіру және токендеу оқыту сапасын арттырды. Ұсынылған модельдерді нақты онлайн ортада қолдану үшін оларды жаңа деректер негізінде жүйелі түрде қайта оқыту қажет. Алгоритмдерді қазақ тілінің морфологиялық және семантикалық ерекшеліктеріне бейімдеу отандық цифрлық кеңістікті мониторингтеу сапасын жоғарылатады. Болашақта трансформерлік архитектураларды, көптілді деректер жиындарын және түсіндірілетін жасанды интеллект әдістерін пайдалану жоспарланады. Алынған нәтижелер кибербуллингі ерте анықтау және алдын алу жүйелерін жетілдіруде қолданылуы мүмкін.

Түйін сөздер: кибербуллинг, терең оқыту, нейрондық желілер, табиғи тілді өңдеу, мәтінді жіктеу, назар аудару механизмі.

Кіріспе

Кибербуллинг жәбірленушілер үшін де, құқық бұзушылар үшін де елеулі салдарға әкелуі мүмкін өсіп келе жатқан мәселе болып табылады. Бұл интернет платформалар, әлеуметтік желілер және мобильді телефондар сияқты цифрлық құралдарды адамдарды қудалау, қорқыту немесе өзге де жолмен зиян келтіру мақсатында пайдалануды білдіреді. Кибербуллинг қорлайтын немесе қоқан-лоққы сипатындағы хабарламаларды жариялау, қауесет тарату, ұятқа қалдыратын

фото және бейнематериалдарды бөлісу немесе біреуді онлайн әлеуметтік топтардан шеттетуге сияқты әртүрлі түрде көрінуі мүмкін.

Кибербуллингтің салдары айтарлықтай ауыр болуы мүмкін. Зардап шеккен адамдарда мазасыздық, депрессия, өзін-өзі бағалаудың төмендеуі және суицидтік ойлар байқалуы мүмкін. ESET антивирустық компаниясы жазғы демалыс қарсаңында қазақстандық интернет пайдаланушылар арасында сауалнама жүргізіп, Қазақстандағы балалардың 75 пайызы интернеттегі әртүрлі қауіп-қатерлерге тап болғанын анықтады. Бұдан бөлек, олар ұйқы, тәбет, зейін қою, сондай-ақ оқу және тұлғааралық қарым-қатынасқа байланысты қиындықтарға кезігуі мүмкін. Кейбір жағдайларда кибербуллинг физикалық салдарға да әкеледі, өйткені нысанаға алынған адам оқшаулануды сезінеді немесе өзін қауіп ішінде қабылдайды. 2020 жылы интернет кеңістігіндегі қудалау салдарынан 143 жасөспірім суицид жасаған, ал 306 кәмелетке толмаған жасөспірім суицид жасауға әрекеттенген.

Кибербуллинг жасайтын адамдар да жағымсыз салдарға ұшырауы мүмкін. Кибербуллинг дәстүрлі буллингке қарағанда сирек кездеседі. Жалпы алғанда, жасөспірімдердің 5 пайызы айына екі-үш рет немесе одан да жиі кибербуллингтің құрбаны болған немесе өздері басқа адамдарға қатысты кибербуллингке қатысқан. Кем дегенде бір рет кибербуллингке 12 пайыз жасөспірім ұшыраған. Мұндай адамдардың әрекеті қылмыс ретінде танылса, олар құқықтық салдарға, сондай-ақ беделінің төмендеуі, достар табуда немесе жұмысқа орналасуда қиындықтар сияқты әлеуметтік және кәсіби мәселелерге тап болуы мүмкін. Статистика бойынша, кибербуллингке ұшыраған респонденттердің үштен бірінен астамы жеке хабарламаларда қорлау сөздерін алған, 24 пайызы өздері туралы жағымсыз жазбаларды көрген, ал шамамен үштен бірі, яғни 31 пайызы, өз фотоларының астындағы жағымсыз пікірлерді оқыған.

Кибербуллингті азайтудың бір жолы, адамдарды олардың әрекеттерінің салдары туралы ақпараттандыру. Мұны қоғамдық ақпараттық науқандар, білім беру бағдарламалары, сондай-ақ ата-аналар мен мұғалімдерге арналған ресурстар арқылы жүзеге асыруға болады. Кибербуллингтің салдары туралы хабардарлықты арттыру арқылы адамдар мұндай әрекетке бармас бұрын көбірек ойланады.

Кибербуллингті азайтудың тағы бір жолы, жәбірленушілерге қолдау мен ресурстар ұсыну. Мұндай қолдау кеңес беруді, терапияны, сондай-ақ кибербуллинг фактілері туралы хабарлау және олардың салдарын жою тетіктерін қамтуы мүмкін. Осындай құралдардың арқасында жәбірленушілер өздерін сенімдірек сезініп, әрекет етуге және көмек сұрауға мүмкіндік алады.

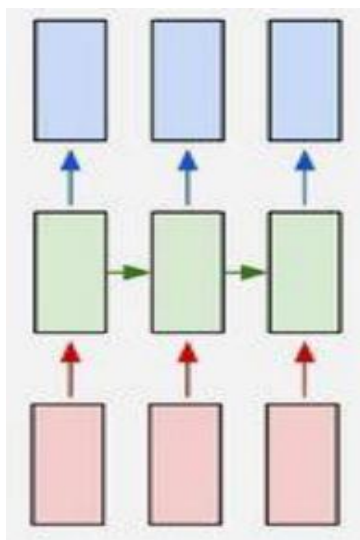
Сондай-ақ құқық бұзушыларды өз әрекеттері үшін жауапкершілікке тарту маңызды. Бұл кибербуллингпен айналысатын оқушыларға тәртіптік шара қолдануды немесе мұндай әрекет қылмыс деп танылған жағдайда сот ісін қозғауды қамтуы мүмкін. Кінәлілерді жауапкершілікке тарту арқылы кибербуллингтің жол берілмейтін әрекет екені және оның ауыр салдарға соқтыратыны көрсетіледі. Соңында, қоғамда құрмет пен ізгілік мәдениетін қалыптастыру маңызды. Бұған мейірімділік пен жанашырлық сияқты жағымды мінез-құлықты ынталандыру және сол мінез-құлықты жеке үлгімен көрсету арқылы қол жеткізуге болады. Осындай мәдениетті қалыптастыру арқылы біз баршаға қолайлы әрі инклюзивті орта құра аламыз.

Машиналық оқыту тәсілдері зерттеулерде пайдалы құрал болғанымен, олардың да кемшіліктері бар. Мысалы, модельдің болжау сапасы оны оқытуға пайдаланылған деректердің сапасы мен көлеміне, сондай-ақ модельдің күрделілігіне тәуелді екенін түсіну маңызды. Егер деректер жеткіліксіз болса немесе олардың құрамында қателер мен бос мәндер болса, онда модель дұрыс жұмыс істемеуі немесе қате нәтиже беруі мүмкін. Бұдан басқа, машиналық оқыту өзінің қарапайымдылығына байланысты деректер арасындағы жасырын тәуелділіктерді әрдайым таба бермейді [1]. Нейрондық желілер, керісінше, мәтіндерді жіктеу міндетінде елеулі артықшылықтарға ие. Олар мәтіндік деректердің үлкен көлемімен жұмыс істеуге және сөздер мен сөз тіркестері арасындағы жасырын тәуелділіктерді анықтауға мүмкіндік береді. Бұл әсіресе

сөздердің контексті мен мағынасын ескеру қажет болатын тоналдылықты талдау немесе мәтінді жіктеу сияқты міндеттер үшін маңызды [2]. Сонымен қатар, нейрондық желілерді жасырын заңдылықтарды табу қажет болатын қадағалаусыз оқытуда да пайдалануға болады. Мысалы, мәтіндік деректердің жасырын белгілерін табу үшін автоэнкодерлерді қолдануға болады. Біз бұл тәсілді осы тараудың соңында ұсынылған назар аудару механизмі бар нейрондық желіні әзірлеу кезінде пайдаландық. Сонымен қатар, нейрондық желілерді қолдану бізге көптеген мүмкіндіктер ашады: шағын да, көлемді де нейрондық желілер құруға, оқыту сатысында ондаған миллион параметрлерді пайдаланатын модельдер жасауға, әртүрлі архитектураларды біріктіруге және модель сапасын жақсарту үшін сан алуан оңтайландыру тәсілдерін қолдануға болады. Машиналық оқыту алгоритмдерімен жұмыс істегенде ресурстар мен есептеу қуаты шектеулі болды. Машиналық және терең оқытуды салыстырғанда, машиналық оқытуды бір ғана жасырын қабаты бар нейрондық желімен салыстыруға болады, ал терең нейрондық желілер көптеген қабаттардан тұрады, кей жағдайда олардың саны ондаған мыңға дейін жетуі мүмкін. Зерттеу аясында рекуррентті нейрондық желілермен, конволюциялық нейрондық желілермен, қысқа және ұзақ мерзімді жад желілерімен, көпқабатты персептрондармен, қарапайым бапталатын терең нейрондық желілермен және ең нәтижелі модель ретінде назар аудару механизмін пайдаланатын терең нейрондық желімен тәжірибелер жүргізілді.

Зерттеу материалдары мен әдістері

Рекуррентті нейрондық желілерді кибербуллингті анықтау міндетінде қолдану. Рекуррентті нейрондық желілердің құрылымы туралы айтпас бұрын, олардың тарихына аздап тоқталайық. Рекуррентті нейрондық желілер классикалық нейрондық желілердің дамуы ретінде 1980 жылдардың соңы мен 1990 жылдардың басында пайда болды [3]. Олардың негізгі идеясы желі элементтері арасындағы рекуррентті байланыстарды пайдалану болып табылады, бұл желіге сөйлемдегі сөздер тізбегі немесе уақыттық қатарлар тізбегі сияқты ретті деректерді өңдеуге мүмкіндік береді. Басқаша айтқанда, рекуррентті нейрондық желілер циклдары бар желілер болып саналады және олар Seq-to-Seq тәрізді реттіліктерді өңдеу міндеттеріне жақсы бейімделген. Рекуррентті нейрондық желілерді қолданудың мәні, біз оларды деректердің реттілігі маңызды болғанда, яғни кіріс деректерінің бірізділігін сақтау аса қажет кезде пайдалана аламыз. Ал мәтінді өңдеу міндеттерінде сөздердің орналасу реті ерекше рөл атқарады, өйткені сөз тәртібінің өзгеруі мағынаны түбегейлі өзгертуі мүмкін. RNN төрт түрге бөлінеді: бірден бірге, бірден көпке, көптен бірге және көптен көпке. Біздің зерттеу мақсатымыз үшін көптен көпке түрі қолайлы. Төменде осындай желінің құрылымын көрсететін сурет берілген (1 сурет).



Сурет 1. Рекуррентті нейрондық желінің құрылымы.

$$X = (x_1, x_2, \dots, x_t),$$

түрінде қарастырайық, мұнда x_t , t уақыт қадамындағы кіріс деректер болып табылады. RNN әрбір уақыт қадамында h_t жасырын күйін сақтайды, ол алдыңғы жасырын күй $h_{(t-1)}$ және ағымдағы кіріс x_t арқылы жаңартылады. Кіріс деректерін жасырын күйге түрлендіру мына түрде жазылады:

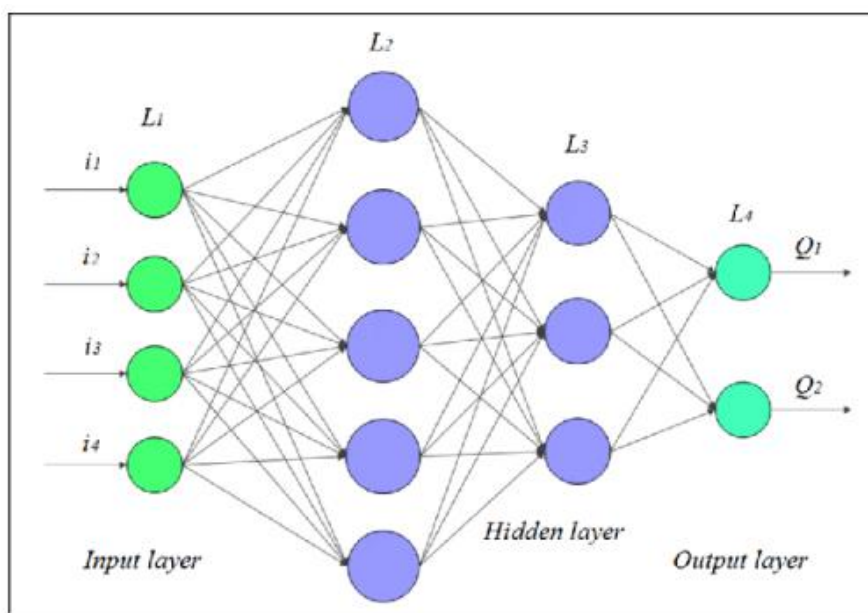
$$h_t = \tanh(W_{hh} * h_{(t-1)} + W_{xh} * x_t + b_h)$$

Жасырын шығысты есептеу:

$$o_t = W_{ho} * h_t + b_o$$

мұнда W_{hh} , W_{xh} , W_{ho} – матрицалары салмақ коэффициенттері, b_h , b_o – ығысу мүшелері, ал $\tanh$ – белсендіру функциясы болып табылады. Сонымен бірге басқа белсендіру функциялары да қолданылуы мүмкін.

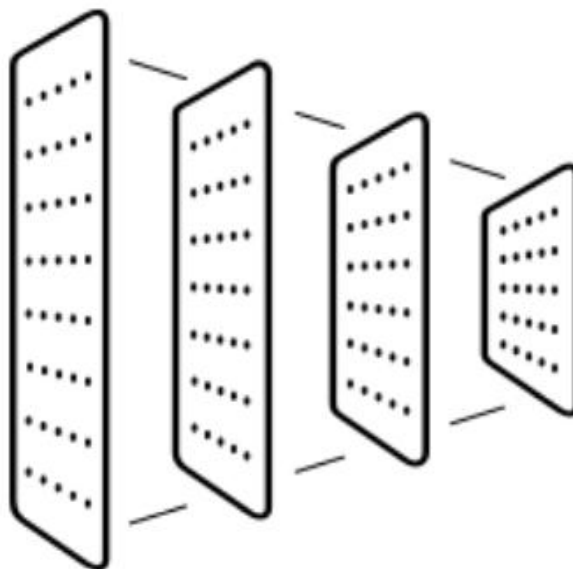
Қарапайым терең нейрондық желілерді кибербуллингті анықтау міндетінде қолдану. Анықтама. Терең оқыту, бұл бастапқы өңделмеген кіріс деректерден жоғары деңгейлі белгілерді біртіндеп бөліп алу үшін бірнеше қабатты пайдаланатын машиналық оқыту алгоритмдерінің класы. Мысалы, кескіндерді өңдеу кезінде төменгі қабаттар жиектерді анықтаса, жоғары қабаттар цифрлар, әріптер немесе бет-әлпет сияқты адамға қатысты ұғымдарды тануы мүмкін [4]. Біздің жағдайда нейрондық желілер де осыған ұқсас принциппен жұмыс істейді: бір қабат кіріс мәтіндегі сөздердің орналасу ретін анықтап, бастапқы ақпарат береді, келесі қабат сөздер арасындағы тәуелділіктерді ажыратады және солай жалғаса береді. Жасалған нейрондық желінің теңдеуі әрбір нейрон үшін тәуелсіз айнымалылар мен олардың сәйкес салмақтарының сызықтық комбинациясын білдіреді. Нейрондық желінің теңдеуін жалпы түрде былай жазуға болады (2 сурет):



Сурет 2. Қарапайым терең нейрондық желінің құрылымы.

Конволюциялық нейрондық желілерді мәтіндерді жіктеу міндетінде қолдану. Анықтама.

Конволюциялық нейрондық желілер, бұл негізінен кескіндерді өңдеуде және компьютерлік көру міндеттерінде қолданылатын нейрондық желі архитектурасының түрі. Дегенмен, олар жергілікті және кеңістіктік ерекшеліктерді тану қабілетінің арқасында мәтіндерді жіктеу міндеттерінде де табысты пайдаланылды. Мәтінді жіктеу контекстінде конволюциялық нейрондық желілер мәтін ішіндегі n -граммалар сияқты маңызды үлгілерді анықтау үшін қолданылады, бұл мәтінді әртүрлі сыныптарға бөлуге көмектеседі [5]. Ол үшін мәтіндік құжаттар word2vec немесе fastText сияқты векторлық модельдер арқылы сандық көрініске айналдырылады. Осыдан кейін мәтіндердің сандық көріністері конволюциялық нейрондық желіге кіріс деректер ретінде беріледі. Желі бірнеше конволюциялық қабаттан тұруы мүмкін. Бұл қабаттар мәтіндік құжаттардан маңызды белгілерді шығарады, ал толық байланысқан қабаттар осы белгілер негізінде мәтіндерді жіктеу үшін пайдаланылады (3 сурет).



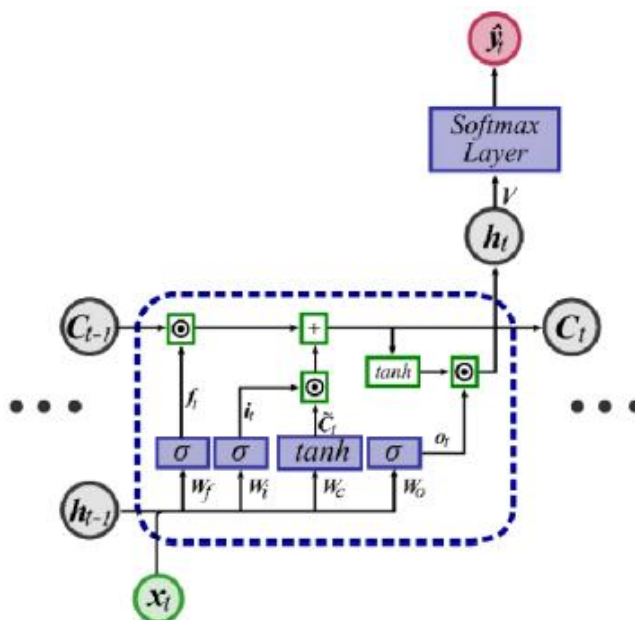
Сурет 3. Конволюциялық нейрондық желінің көрнекі бейнеленуі.

LSTM (ұзақ және қысқа мерзімді жад желілерін) кибербуллингті анықтау үшін қолдану.

Анықтама. LSTM (Long Short-Term Memory) немесе ұзақ және қысқа мерзімді жад желілері, дәстүрлі рекуррентті нейрондық желілерде пайда болатын градиенттің жоғалу мәселесін шешуге арналған рекуррентті нейрондық желі архитектурасының түрі болып табылады [6]. Техникалық тұрғыдан алғанда, ол рекуррентті нейрондық желілердің логикалық модификациясы, бірақ бір маңызды айырмашылығы бар. Қарапайым терең нейрондық желісінің бірегей архитектурасы оған Seq форматындағы ұзын тізбектер бойында ақпаратты сақтауға және оны өңдеуге мүмкіндік береді [7]. Осы айырмашылықтың арқасында қысқа және ұзақ мерзімді жад желілері табиғи тілді өңдеу әдістерін едәуір жақсартты.

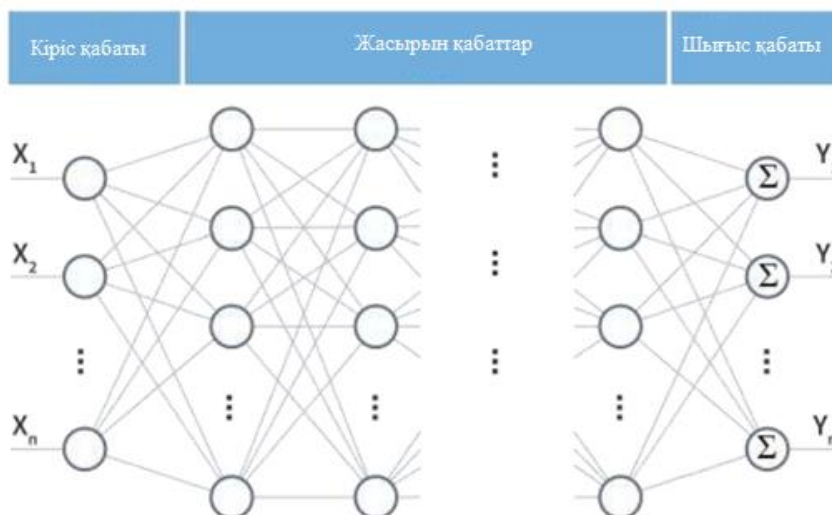
Бұл желінің негізгі компоненттері мыналар. Ұяшық күйі, бұл кіріс тізбегінің ағымдағы жай-күйі туралы ақпаратты сақтайтын және әр қадам сайын кіріс деректерге, шығыс деректерге және желінің алдыңғы күйіне сүйене отырып жаңартылатын элемент. Кіріс қақпасы, бұл келіп түскен ақпараттың қай бөлігі ұяшық күйінде сақталуы керек екенін анықтайтын элемент. Ол 0 мен 1 аралығындағы мәнді алу үшін sigmoid белсендіру функциясын пайдаланады. Мұнда 0, ешқандай ақпарат сақталмайтынын, ал 1, кіріс ақпараттың толық сақталатынын білдіреді. Шығыс қақпасы, бұл ұяшық күйінің қай бөлігі келесі қабатқа ашылатынын анықтайтын элемент. Ол да дәл осындай қағидамен sigmoid белсендіру функциясын қолданады. Ұмыту қақпасы, бұл алдыңғы

ұяшық күйінің қай бөлігі sigmoid функциясының нәтижесіне қарай жойылуы керектігін көрсететін элемент. Мұнда 0, бүкіл күйдің алынып тасталатынын, ал 1, оның толық сақталып, келесі қадамға берілетінін білдіреді. Осылайша, қарапайым терең нейрондық желінің негізгі ерекшелігі, ұзын тізбектердегі ақпаратты сақтай алатын жад ұяшығын енгізу болып табылады. Бұл жад ұяшығы кіріс, шығыс және ұмыту қақпалары арқылы реттеледі. Аталған қақпалар жад ұяшығына ақпараттың енуін, одан шығуын және оның ішкі айналымын басқарады (4 сурет).



Сурет 4. LSTM желісінің құрылымы.

Көпқабатты персептронды мәтіндік деректерді жіктеу міндетінде қолдану. Анықтама. Көпқабатты персептрон, бұл өзара байланысқан нейрондардың бірнеше қабатынан тұратын жасанды нейрондық желінің түрі. Бұл тікелей байланысы бар желі, яғни ақпарат кіріс қабатынан жасырын қабаттар арқылы шығыс қабатына ешқандай циклсіз өтеді [8]. Көпқабатты персептрондар жіктеу, регрессия және функцияларды жуықтау сияқты әртүрлі міндеттерде қолданылады. Әдетте көпқабатты персептрон келесі бөліктерден тұрады: кіріс қабаты, ол кіріс деректерді білдіретін нейрондардан құралады, әрбір нейрон белгілі бір белгіге сәйкес келеді және олардың саны кіріс деректердің өлшеміне байланысты болады. Жасырын қабаттар, олар кіріс деректерге бейсызық түрлендірулер орындайтын нейрондардан тұрады. Жасырын қабаттар желіге күрделі үлгілер мен көріністерді меңгеруге мүмкіндік береді. Жасырын қабаттардың саны мен әр қабаттағы нейрондар саны міндеттің күрделілігі мен архитектура дизайнына байланысты өзгереді. Шығыс қабаты, бұл соңғы болжамды немесе жіктеу нәтижесін шығаратын нейрондар жиынтығы. Екілік жіктеу үшін әдетте бір ғана шығыс нейроны қолданылады, ал көпсыныпты жіктеуде нейрондар саны сыныптар санына тең болады. Регрессия міндеттерінде шығыс нейрондарының саны мақсатты айнымалының өлшемімен анықталады. Кері тарату, көпқабатты персептрондар әдетте мұғаліммен оқытуға негізделген кері тарату алгоритмі арқылы оқытылады. Бұл алгоритм шығын функциясының әрбір салмаққа қатысты градиентін тізбек ережесі бойынша есептеуді, одан кейін градиенттік түсу немесе оның нұсқалары сияқты оңтайландыру алгоритмдері арқылы салмақтарды жаңартуды қамтиды (5 сурет).



Сурет 5. Көпқабатты персептронның графикалық көрінісі.

Салыстырмалы қарапайымдылығына қарамастан, көпқабатты персептронмен жұмыс істегенде ескеру қажет бірқатар жайттар бар. Өз құрылымына байланысты ол мәтінді өңдеудің сәл өзгеше тәсілін және гиперпараметрлерді бөлек баптауды талап етеді [8]. Бірінші маңызды жайт, тығыз қабаттар сөздер арасындағы тәуелділіктерді әрқашан дәл ұстай бермейді, сондықтан деректерді алдын ала өңдеу де өзгеше жүргізілуі керек. Егер алдыңғы модельдер үшін TF-IDF арқылы векторлау жеткілікті болса, MLP үшін Word2Vec немесе GloVe тәрізді неғұрлым күрделі векторлау әдістері қажет. Екі алгоритм талданғаннан кейін Word2Vec қолдану туралы шешім қабылданды [15]. Word2Vec көбіне басқа векторлау әдістеріне қарағанда жақсырақ деп есептеледі, өйткені ол сөздер арасындағы семантикалық және синтаксистік байланыстарды тиімді түрде ұстай алады. Дегенмен, басқа векторлау әдістерінің де өз артықшылықтары бар және олар нақты қолдану жағдайына байланысты анағұрлым қолайлы болуы мүмкін. Көп жағдайда Word2Vec өзінің құрылымында CBOW немесе Skip-gram тәсілдерін қолданатындықтан басымдыққа ие.

Назар аудару механизмі. Терең нейрондық желілердегі назар аудару механизмі, бұл модельге болжам жасау кезінде кіріс деректердің белгілі бір бөліктеріне шоғырлануға мүмкіндік беретін тетік [9]. Ол әдетте машиналық аударма, мәтінді қысқаша мазмұндау және көңіл күйді талдау сияқты табиғи тілді өңдеу міндеттерінде қолданылады. Бұдан бөлек, ол екілік және көпсыныпты жіктеу міндеттеріне де жақсы бейімделген. Назар аудару қабатының архитектурасы үш құрамдас бөліктен тұрады.

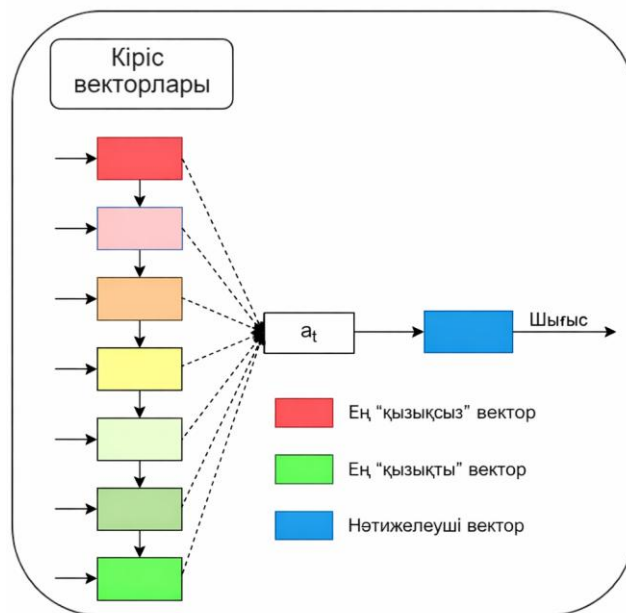
1. Сұрау, Query, бұл модельдің ағымдағы күйін немесе алдыңғы қабаттың шығыс деректерін білдіретін вектор. 2. Кілт, Key, бұл кіріс деректер туралы ақпаратты сақтайтын матрица. 3. Мән, Value, бұл кіріс деректерді ұсынатын матрица.

Назар аудару механизмі сұрау мен кілт арасындағы назар бағаларының жиынтығын есептейді, содан кейін бұл бағалар мәндер матрицасының салмақталған қосындысын есептеу үшін қолданылады. Назар қабатының нәтижесі, кіріс деректердегі маңызды ақпаратты қамтитын контекстік вектор. Назар аудару механизмінің математикалық көрінісі төмендегідей [10]:

$\text{Attention}(\text{Query}, \text{Key}, \text{Value}) = \text{softmax}(\text{Query} \times \text{Key}^T) \times \text{Value}$, мұнда softmax , назар көрсеткіштерін қалыпқа келтіретін функция, $\times$, нүктелік көбейту операциясы, ал T , транспозицияланған матрицаны білдіреді.

Іс жүзінде назар аудару механизмдерінің бірнеше басы болуы мүмкін, бұл модельге кіріс деректердің әртүрлі бөліктерін қатар өңдеуге мүмкіндік береді. Бірнеше бастың шығыстары біріктіріліп, механизмнің қорытынды шығыс сигналы қалыптастырылады [16]. Назар аудару

механизмі мүлде басқа принциппен жұмыс істейтіндіктен, бүкіл оқыту моделін толық түсіну үшін оның құрылымын жеке сипаттау қажет. Ең алдымен, неге назар аудару механизмі машиналық аударманы және табиғи тілді өңдеудің басқа міндеттерін жеңілдеткенін түсіну маңызды. Бұрын машиналық аударма кодтаушы-декодтаушы модельдеріне сүйенетін. Олар екі желіден, яғни кодтаушыдан және декодтаушыдан тұрады. Кодтаушы модельдің рөлі, кіріс тізбегін сөйлем туралы барлық ақпаратты жинақтайтын векторға түрлендіру [12]. Содан кейін бұл вектор қажетті шығыс сигналын генерациялайтын декодтерге беріледі. Бұл қабат кіріс ретінде векторларға кодталған деректерді қабылдайды және олардың мәндері бойынша әрбір вектордың маңыздылығын есептейді. Ол анағұрлым релевантты векторларды табады және басқа векторларға қарағанда көбірек ақпарат беретін маңызды векторлармен жұмыс істейді (6 сурет) [11, 13].



Сурет 6. Назар аудару механизмінің құрылымы.

Бұл деңгейдегі назарды есептеу формуласы жоғарыда келтірілді. a_t ұяшығы өз ішінде маңызды векторлар туралы ақпаратты сақтайды және оларды назар ұяшығының қорытынды шығыс векторы болып табылатын контекстік векторға береді [12].

Нәтижелер және оларды талқылау

Осы зерттеу аясында мәтіндік деректердегі кибербуллингті анықтау міндеті үшін терең оқытудың әртүрлі архитектураларының тиімділігіне салыстырмалы баға берілді. Эксперименттер құрамында бейтарап хабарламалар да, кибербуллинг белгілері бар хабарламалар да қамтылған таңбаланған мәтіндер корпусында орындалды. Модельдердің сапасын бағалау үшін жіктеудің стандартты метрикалары қолданылды: дәлдік, толықтық, оң болжамдардың дәлдігі және F1 өлшемі.

Эксперименттер барысында келесі модельдер жүзеге асырылып, оқытылды: көпқабатты персептрон, конволюциялық нейрондық желі, рекуррентті нейрондық желі, қысқа және ұзақ мерзімді жад желісі, сондай-ақ назар аудару механизмі бар терең нейрондық желі. Барлық модельдер оқытудың салыстырмалы параметрлерін пайдалана отырып, бірдей деректер жиынында оқытылды. Бұл салыстырмалы талдаудың дұрыстығын қамтамасыз етуге мүмкіндік берді.

Эксперимент нәтижелері көпқабатты персептрон сияқты базалық модельдердің дәлдік көрсеткіштері қанағаттанарлық екенін, алайда сөздер арасындағы контекстік және семантикалық

тәуелділіктерді ескеру қабілеті бойынша күрделірек архитектуралардан төмен екенін көрсетті. Конволюциялық нейрондық желілер жергілікті мәтіндік үлгілерді тиімді анықтау қабілетінің арқасында MLP моделіне қарағанда жақсы нәтижелер көрсетті, бірақ ұзақ мерзімді тәуелділіктерді модельдеудегі шектеулері кибербуллингті анықтаудың жалпы толықтығын төмендетеді.

Рекуррентті нейрондық желілер мен LSTM модельдері толықтық пен F1 өлшемі бойынша жоғары мәндер көрсетті. Бұл олардың ретті деректерді өңдеуде және сөздердің тәртібін ескеруде тиімді екенін айғақтайды. Жіктеу сапасының ең жоғары көрсеткіштері назар аудару механизмі бар модельді қолданғанда алынды. Бұл архитектура модельге агрессия, қорқыту немесе қорлау белгілері бар мәтіннің неғұрлым маңызды фрагменттеріне шоғырлануға мүмкіндік берді, соның нәтижесінде кибербуллингті анықтау дәлдігі едәуір артты.

Осылайша, эксперименттік нәтижелер назар аудару механизмі бар терең нейрондық желілерді пайдалану классикалық және базалық терең оқыту архитектураларымен салыстырғанда мәтіндік деректердегі кибербуллингті автоматтандырылған түрде анықтаудың ең перспективалы тәсілі екенін растайды.

Қорытынды

Бұл жұмыста цифрлық коммуникациялардағы кибербуллингті анықтау міндетіне арналған терең оқытудың заманауи тәсілдері қарастырылып, талданды. Зерттеу барысында рекуррентті, конволюциялық және толық байланысқан модельдерді, сондай-ақ қысқа және ұзақ мерзімді жад пен назар аудару механизмі бар модельдерді қолдану ерекшеліктері зерттелді.

Жүргізілген эксперименттік зерттеулер дәстүрлі машиналық оқыту әдістері мен қарапайым нейрондық архитектуралардың күрделі семантикалық және контекстік тәуелділіктері бар мәтіндерді талдау міндеттерінде мүмкіндіктері шектеулі екенін көрсетті. Ал терең оқыту әдістері ақпараттық белгілерді автоматты түрде бөліп алу және тілдік деректердің құрылымын ескеру қабілетінің арқасында айтарлықтай жоғары тиімділік көрсетті.

Ең жақсы нәтижелер назар аудару механизмі бар терең нейрондық желіні қолданғанда алынды. Бұл аталған тәсілді кибербуллингті анықтауға байланысты мәтіндерді жіктеу міндеттерінде қолданудың орынды екенін растайды. Назар аудару механизмі модельдің интерпретациялануын арттыруға және кіріс мәтіннің ең маңызды элементтеріне назар шоғырландыру арқылы жіктеу сапасын жақсартуға мүмкіндік береді.

Ғылыми жаңалығы: зерттеуде MLP, CNN, RNN, LSTM және Attention негізіндегі модельдер бірдей деректер жиынында салыстырылып, назар аудару механизмі бар архитектуралардың кибербуллингті анықтауда жоғары тиімділік көрсететіні дәлелденді.

Алынған нәтижелерді әлеуметтік желілерге, онлайн платформаларға және білім беру цифрлық орталарына мониторинг жүргізетін зияткерлік жүйелерді әзірлеуде кибербуллингті ерте анықтау және алдын алу мақсатында пайдалануға болады. Болашақта зерттеуді трансформер архитектураларын, көптілді деректер корпустарын қолдану, сондай-ақ терең оқыту модельдерін цифрлық қауіпсіздікті қамтамасыз етудің нақты қолданбалы жүйелеріне біріктіру есебінен кеңейту орынды.

Алғыс айту. Бұл зерттеу Қазақстан Республикасы Ғылым және жоғары білім министрлігі Ғылым комитеті тарапынан қаржыландырылды, грант № AP23488900, «Жасанды интеллектіні пайдалана отырып, әлеуметтік желілердегі жастар арасындағы кибербуллингті автоматты түрде анықтау».

Әдебиеттер тізімі

1. Alsubait T., Alfageh D. Comparison of machine learning techniques for cyberbullying detection on YouTube Arabic comments // International Journal of Computer Science and Network Security, —

2021. — Vol. 21, no. 1. — P. 1–5. — URL: <https://koreascience.kr/article/JAKO202121055656002.page>
2. Dewani M., Memon M., Bhatti, S. Cyberbullying detection: Advanced preprocessing techniques & deep learning architecture for Roman Urdu data. *Journal of Big Data*, — 2021. — Vol. 8. — Art. 160. — DOI: [10.1186/s40537-021-00550-7](https://doi.org/10.1186/s40537-021-00550-7).
3. Hall D., Silva Y., Wheeler Y., Cheng L., Baumel, K. Harnessing the power of interdisciplinary research with psychology-informed cyberbullying detection models. // *International Journal of Bullying Prevention*. — 2022. — Vol. 4, no. 1. — P. 47–54. — DOI: [10.1007/s42380-021-00107-5](https://doi.org/10.1007/s42380-021-00107-5).
4. Arce-Ruelas K. I., Alvarez-Xochihua O., Pellegrin L., Cardoza-Avendaño L., González-Fraga J. Á. // Automatic cyberbullying detection: A Mexican case in high school and higher education students. *IEEE Latin America Transactions*. — 2022. — Vol. 20, no. 5. — P. 770–779. — URL: <https://latam.ieee9.org/index.php/transactions/article/view/5934>
5. Ahmed T., Rahman M., Nur S., Islam, A., Das, D. Natural language processing and machine learning based cyberbullying detection for Bangla and romanized Bangla texts. // *TELKOMNIKA (Telecommunication Computing Electronics and Control)*. — 2022. — Vol. 20, no. 1. — P. 89–97. — DOI: [10.12928/TELKOMNIKA.v20i1.18630](https://doi.org/10.12928/TELKOMNIKA.v20i1.18630).
6. Omarov A., Tursynova A., Postolache O., Gamry K., Batyrbekov, A., et al. Modified UNet model for brain stroke lesion segmentation on computed tomography images. // *Computers, Materials & Continua*. — 2022. — Vol. 71, no. 3. — P. 4701–4717. — DOI: [10.32604/cmc.2022.020998](https://doi.org/10.32604/cmc.2022.020998).
7. Al-Marghilani. Artificial intelligence-enabled cyberbullying-free online social networks in smart cities. // *International Journal of Computational Intelligence Systems*. — 2022. — Vol. 15. — Art. 9. — DOI: [10.1007/s44196-022-00063-y](https://doi.org/10.1007/s44196-022-00063-y).
8. Theng C., Othman N., Abdullah R., Anawar S., Ayop Z., et al. Cyberbullying detection in Twitter using sentiment analysis. // *International Journal of Computer Science and Network Security*. — 2021. — Vol. 21, no. 11. — P. 1–10. — URL: <https://koreascience.kr/article/JAKO202105254945209.do>
9. Sadiq S., Mehmood A., Ullah S., Ahmad M., Choi G., et al. Aggression detection through deep neural model on Twitter. // *Future Generation Computer Systems*. — 2021. — Vol. 114. — P. 120–129.
10. Sarac Essiz E., Oturakci M. Artificial bee colony-based feature selection algorithm for cyberbullying. // *The Computer Journal*. — 2021. — Vol. 64, no. 3. — P. 305–313. — DOI: [10.1093/comjnl/bxaa066](https://doi.org/10.1093/comjnl/bxaa066).
11. Gomez C. E., Sztainberg M. O., Trana R. E. Curating cyberbullying datasets: A human-AI collaborative approach. // *International Journal of Bullying Prevention*. — 2022. — Vol. 4, no. 1. — P. 35–46. — DOI: [10.1007/s42380-021-00114-6](https://doi.org/10.1007/s42380-021-00114-6).
12. Salawu S., Lumsden J., He Y. A mobile-based system for preventing online abuse and cyberbullying. // *International Journal of Bullying Prevention*. — 2022. — Vol. 4, no. 1. — P. 66–88. — DOI: [10.1007/s42380-021-00115-5](https://doi.org/10.1007/s42380-021-00115-5).
13. Mahajan E., Mahajan H., Kumar S. EnsMulHateCyb: Multilingual hate speech and cyberbully detection in online social media. // *Expert Systems with Applications*. — 2024. — Vol. 236. — Art. 121228. — DOI: [10.1016/j.eswa.2023.121228](https://doi.org/10.1016/j.eswa.2023.121228).
14. Jamjoom A. A., Karamti H., Umer M., Alsubai S., Kim T. H., Ashraf, I. RobertaNet: Enhanced roberta transformer based model for cyberbullying detection with glove features. // *IEEE Access*. — 2024. — Vol. 12. — P. 58950–58959. — DOI: [10.1109/ACCESS.2024.3386637](https://doi.org/10.1109/ACCESS.2024.3386637).
15. Zhang J., Wu Q., Xu Y., Cao C., Du Z., Psounis, K. Efficient toxic content detection by bootstrapping and distilling large language models. // *Proceedings of the AAAI Conference on Artificial Intelligence*. — 2024. — Vol. 38, no. 19. — P. 21779–21787. — DOI: [10.1609/aaai.v38i19.30178](https://doi.org/10.1609/aaai.v38i19.30178).
16. Xu M., Liu S. Rb_bg_mha: A roberta-based model with bi-gru and multi-head attention for chinese offensive language detection in social media. // *Applied Sciences*. — 2023. — Vol. 13, no. 19. — Art. 11000. — DOI: [10.3390/app131911000](https://doi.org/10.3390/app131911000).

References

1. Alsubait T., Alfageh D. Comparison of machine learning techniques for cyberbullying detection on YouTube Arabic comments // International Journal of Computer Science and Network Security, — 2021. — Vol. 21, no. 1. — P. 1–5. — URL: <https://koreascience.kr/article/JAKO202121055656002.page>
2. Dewani M., Memon M., Bhatti, S. Cyberbullying detection: Advanced preprocessing techniques & deep learning architecture for Roman Urdu data. Journal of Big Data, — 2021. — Vol. 8. — Art. 160. — DOI: [10.1186/s40537-021-00550-7](https://doi.org/10.1186/s40537-021-00550-7).
3. Hall D., Silva Y., Wheeler Y., Cheng L., Baumel, K. Harnessing the power of interdisciplinary research with psychology-informed cyberbullying detection models. // International Journal of Bullying Prevention. — 2022. — Vol. 4, no. 1. — P. 47–54. — DOI: [10.1007/s42380-021-00107-5](https://doi.org/10.1007/s42380-021-00107-5).
4. Arce-Ruelas K. I., Alvarez-Xochihua O., Pellegrin L., Cardoza-Avendaño L., González-Fraga J. Á. // Automatic cyberbullying detection: A Mexican case in high school and higher education students. IEEE Latin America Transactions. — 2022. — Vol. 20, no. 5. — P. 770–779. — URL: <https://latam.ieee9.org/index.php/transactions/article/view/5934>
5. Ahmed T., Rahman M., Nur S., Islam, A., Das, D. Natural language processing and machine learning based cyberbullying detection for Bangla and romanized Bangla texts. // TELKOMNIKA (Telecommunication Computing Electronics and Control). — 2022. — Vol. 20, no. 1. — P. 89–97. — DOI: [10.12928/TELKOMNIKA.v20i1.18630](https://doi.org/10.12928/TELKOMNIKA.v20i1.18630).
6. Omarov A., Tursynova A., Postolache O., Gamry K., Batyrbekov, A., et al. Modified UNet model for brain stroke lesion segmentation on computed tomography images. // Computers, Materials & Continua. — 2022. — Vol. 71, no. 3. — P. 4701–4717. — DOI: [10.32604/cmc.2022.020998](https://doi.org/10.32604/cmc.2022.020998).
7. Al-Marghilani. Artificial intelligence-enabled cyberbullying-free online social networks in smart cities. // International Journal of Computational Intelligence Systems. — 2022. — Vol. 15. — Art. 9. — DOI: [10.1007/s44196-022-00063-y](https://doi.org/10.1007/s44196-022-00063-y).
8. Theng C., Othman N., Abdullah R., Anawar S., Ayop Z., et al. Cyberbullying detection in Twitter using sentiment analysis. // International Journal of Computer Science and Network Security. — 2021. — Vol. 21, no. 11. — P. 1–10. — URL: <https://koreascience.kr/article/JAKO202105254945209.do>
9. Sadiq S., Mehmood A., Ullah S., Ahmad M., Choi G., et al. Aggression detection through deep neural model on Twitter. // Future Generation Computer Systems. — 2021. — Vol. 114. — P. 120–129.
10. Sarac Essiz E., Oturakci M. Artificial bee colony-based feature selection algorithm for cyberbullying. // The Computer Journal. — 2021. — Vol. 64, no. 3. — P. 305–313. — DOI: [10.1093/comjnl/bxaa066](https://doi.org/10.1093/comjnl/bxaa066).
11. Gomez C. E., Sztainberg M. O., Trana R. E. Curating cyberbullying datasets: A human-AI collaborative approach. // International Journal of Bullying Prevention. — 2022. — Vol. 4, no. 1. — P. 35–46. — DOI: [10.1007/s42380-021-00114-6](https://doi.org/10.1007/s42380-021-00114-6).
12. Salawu S., Lumsden J., He Y. A mobile-based system for preventing online abuse and cyberbullying. // International Journal of Bullying Prevention. — 2022. — Vol. 4, no. 1. — P. 66–88. — DOI: [10.1007/s42380-021-00115-5](https://doi.org/10.1007/s42380-021-00115-5).
13. Mahajan E., Mahajan H., Kumar S. EnsMulHateCyb: Multilingual hate speech and cyberbully detection in online social media. // Expert Systems with Applications. — 2024. — Vol. 236. — Art. 121228. — DOI: [10.1016/j.eswa.2023.121228](https://doi.org/10.1016/j.eswa.2023.121228).
14. Jamjoom A. A., Karamti H., Umer M., Alsubai S., Kim T. H., Ashraf, I. RobertaNet: Enhanced roberta transformer based model for cyberbullying detection with glove features. // IEEE Access. — 2024. — Vol. 12. — P. 58950–58959. — DOI: [10.1109/ACCESS.2024.3386637](https://doi.org/10.1109/ACCESS.2024.3386637).
15. Zhang J., Wu Q., Xu Y., Cao C., Du Z., Psounis, K. Efficient toxic content detection by bootstrapping and distilling large language models. // Proceedings of the AAAI Conference on Artificial Intelligence. — 2024. — Vol. 38, no. 19. — P. 21779–21787. — DOI: [10.1609/aaai.v38i19.30178](https://doi.org/10.1609/aaai.v38i19.30178).
16. Xu M., Liu S. Rb_bg_mha: A roberta-based model with bi-gru and multi-head attention for

АНАЛИЗ ЭФФЕКТИВНОСТИ МЕТОДОВ ГЛУБОКОГО ОБУЧЕНИЯ ПРИ ВЫЯВЛЕНИИ КИБЕРБУЛЛИНГА

АБДРАХМАНОВ Р.Б.¹, КЕРІМХАН Б.Т.^{2*}, КАСЫМБЕКОВ А.С.¹,
ИСКАКОВ Т.Б.¹, УКСИКБАЕВ Е.З.³

Абдрахманов Рустам Бахтиёрұлы¹ — Кандидат технических наук, доцент, Международный университет туризма и гостеприимства, г. Туркестан, Казахстан

E-mail: abdrakhmanov.rustam@iuth.edu.kz, <https://orcid.org/0000-0002-5508-389X>

*Керімхан Бекжан Темірханұлы² — PhD, старший преподаватель кафедры компьютерной и программной инженерии, Евразийский национальный университет имени Л.Н. Гумилёва, г. Астана, Казахстан

E-mail: bek_zhan_16@mail.ru, <https://orcid.org/0000-0002-8480-7316>

Касымбеков Аббазбек Сугирович¹ — Кандидат технических наук, доцент, Международный университет туризма и гостеприимства, г. Туркестан, Казахстан

E-mail: abbazbek@mail.ru, <https://orcid.org/0000-0001-8022-5637>

Искаков Тайыржан Бахытбаевич¹ — PhD, доцент, Международный университет туризма и гостеприимства, г. Туркестан, Казахстан

E-mail: t.iskakov@iuth.edu.kz, <https://orcid.org/0000-0002-7499-9297>

Уксикбаев Еркебулан Зейнуллаұлы³ — Магистр, Международный казахско-турецкий университет имени Ходжи Ахмеда Ясави, г. Туркестан, Казахстан

E-mail: erkebulan.uxikbayev@ayu.edu.kz, <https://orcid.org/0009-0003-3868-6773>

Аннотация. Цель исследования — сравнить эффективность методов глубокого обучения, предназначенных для автоматического выявления признаков кибербуллинга в текстовых данных цифровых коммуникаций, и определить наиболее эффективную нейронную архитектуру. Кибербуллинг является одной из актуальных социально-психологических проблем информационного общества и оказывает негативное влияние на психозмоциональное состояние пользователей, особенно подростков и молодежи. Увеличение объема текстового контента в социальных сетях и на онлайн-платформах требует разработки автоматизированных систем, способных своевременно выявлять сообщения, содержащие оскорбления, агрессию и угрозы.

В ходе исследования на основе корпуса размеченных текстов были рассмотрены многослойный персептрон (MLP), сверточная нейронная сеть (CNN), рекуррентная нейронная сеть (RNN), сеть долгой краткосрочной памяти (LSTM), а также глубокая нейронная модель с механизмом внимания. Качество моделей оценивалось по показателям точности, полноты, точности положительных прогнозов и F1-мере. Результаты показали, что модель MLP обладает ограниченными возможностями при учете сложных контекстных связей, архитектура CNN эффективно выявляет локальные языковые шаблоны, а модели RNN и LSTM лучше обрабатывают последовательность слов и долгосрочные зависимости. Наивысшее качество классификации продемонстрировала модель с механизмом внимания, поскольку она придает больший вес значимым фрагментам текста, содержащим признаки агрессии, запугивания и оскорбления.

Результаты исследования показали, что эффективность моделей зависит от качества и сбалансированности размеченных данных. Предварительная обработка текста, нормализация слов и токенизация способствовали повышению качества обучения. Для применения предложенных моделей в реальной онлайн-среде необходимо регулярно переобучать их на новых данных. Адаптация алгоритмов к морфологическим и семантическим особенностям казахского языка позволит повысить качество мониторинга отечественного цифрового пространства. В дальнейшем планируется использовать трансформерные архитектуры, многоязычные наборы данных и методы объяснимого искусственного интеллекта. Полученные результаты могут применяться для совершенствования систем раннего выявления и предотвращения кибербуллинга.

Ключевые слова: кибербуллинг, глубокое обучение, нейронные сети, обработка естественного языка, классификация текста, механизм внимания.

Қ.Жұбанов атындағы Ақтөбе өңірлік университетінің хабаршысы, №3 (85), қыркүйек 2026
Компьютерлік ғылымдар-Компьютерные науки-Computer science
**ANALYSIS OF THE EFFECTIVENESS OF DEEP LEARNING METHODS IN DETECTING
CYBERBULLYING**

**ABDRAKHMANOV R.B.¹, KERIMKHAN B.T.^{2*}, KASYMBEKOV A.S.¹,
ISKAKOV T.B.¹, UKSIKBAYEV E.Z.³**

Abdrakhmanov Rustam Bakhtiyoruly¹ - Candidate of technical sciences, associate professor, International university of tourism and hospitality, Turkistan, Kazakhstan

E-mail: abdrakhmanov.rustam@iuth.edu.kz, <https://orcid.org/0000-0002-5508-389X>

***Kerimkhan Bekzhan Temirkhanuly²** - PhD, senior lecturer at the department of computer and software engineering, L.N. Gumilyov Eurasian national university, Astana, Kazakhstan

E-mail: bek_zhan_16@mail.ru, <https://orcid.org/0000-0002-8480-7316>

Kasymbekov Abbazbek Sugirovich¹ - Candidate of technical sciences, associate professor, International university of tourism and hospitality, Turkistan, Kazakhstan

E-mail: abbazbek@mail.ru, <https://orcid.org/0000-0001-8022-5637>

Iskakov Taiyrzhan Bakhytbayevich¹ - PhD, associate professor, International university of tourism and hospitality, Turkistan, Kazakhstan

E-mail: t.iskakov@iuth.edu.kz, <https://orcid.org/0000-0002-7499-9297>

Uksikbayev Yerkebulan Zeynullauly³ - Master's degree holder, Khoja Akhmet Yassawi international kazakh-turkish university, Turkistan, Kazakhstan

E-mail: erkebulan.uxikbayev@ayu.edu.kz, <https://orcid.org/0009-0003-3868-6773>

Abstract. The purpose of this study is to compare the effectiveness of deep learning methods designed to automatically detect signs of cyberbullying in textual data from digital communications and to identify the most effective neural network architecture. Cyberbullying is one of the most pressing socio-psychological problems of the information society and negatively affects users' psychological and emotional well-being, particularly that of adolescents and young people. The increasing volume of textual content on social networks and online platforms requires the development of automated systems capable of promptly detecting messages containing insults, aggression, and threats.

The study examined a multilayer perceptron (MLP), a convolutional neural network (CNN), a recurrent neural network (RNN), a long short-term memory network (LSTM), and a deep neural model with an attention mechanism using a corpus of labeled texts. Model performance was evaluated using accuracy, recall, precision, and the F1-score. The results showed that the MLP model has limited ability to capture complex contextual relationships, the CNN architecture effectively identifies local linguistic patterns, and the RNN and LSTM models are better at processing word sequences and long-term dependencies. The attention-based model demonstrated the highest classification performance because it assigns greater importance to significant text fragments containing signs of aggression, intimidation, and abuse.

The findings showed that model effectiveness depends on the quality and balance of labeled data. Text preprocessing, word normalization, and tokenization contributed to improved training performance. To apply the proposed models in real-world online environments, they must be regularly retrained using new data. Adapting the algorithms to the morphological and semantic features of the Kazakh language will improve the monitoring of Kazakhstan's digital environment. Future research will employ transformer-based architectures, multilingual datasets, and explainable artificial intelligence methods. The findings can be used to improve systems for the early detection and prevention of cyberbullying.

Key words: cyberbullying, deep learning, neural networks, natural language processing, text classification, attention mechanism.