

**А. Б. ТОҚТАРОВА<sup>1\*</sup>, Б. С. ОМАРОВ<sup>2</sup>, Ж. Ж. АЖИБЕКОВА<sup>3</sup>,  
А. Н. ЖАКСАНОВА<sup>4</sup>, Р. Б. АБДРАХМАНОВ<sup>5</sup>**

<sup>1</sup>Қожа Ахмет Ясауи атындағы Халықаралық Қазақ – Түрік Университеті,  
Түркістан, Қазақстан

<sup>2</sup>аль-Фараби атындағы Қазақ Ұлттық университеті, Алматы, Қазақстан

<sup>3</sup>С. Ж. Асфендияров атындағы Қазақ ұлттық медицина университеті,  
Алматы, Қазақстан

<sup>4</sup>М.Ауезов атындағы Оңтүстік Қазақстан Университеті, Шымкент, Қазақстан

<sup>5</sup>Халықаралық туризм және меймандостық университеті,  
Түркістан, Қазақстан

## **ЖЕЛІ ҚОЛДАНУШЫЛАРЫНЫҢ «ЭМОЦИЯЛЫ» ПІКІРІ АРҚЫЛЫ АВТОМАТТАНДЫРЫЛҒАН ҒАДАУАТ СӨЗДЕР КЛАССИФИКАЦИЯСЫ**

Кез келген онлайн-пікірді әлеуметтік желіде оңай енгізе алудың нәтижесінде соңғы жылдары желіде өшпенділік білдіру кеңінен тарады. Бұл тенденция негізінен мобильді құрылғылар, компьютерлер мен Интернеттің қарқынды өсуіне байланысты. Ғадауат сөздерін автоматты түрде анықтау бойынша зерттеулерге үлкен қызығушылық туып отыр. Дегенмен, әлеуметтік желі сайттары бейәдеп пікірлерден туатын қылмыстарға және соларға қатысты қауымдастықтарды анықтауға көмектесетіні туралы ғылыми зерттеулер жұмыстарының аздығы байқалады. Ғадауат пікірлер кез келген адамның психикалық деңгейіне кері әсер береді. Дегенмен кейбір адамдар басқаларға қарағанда бейәдеп сөздерден келетін эмоциональдық тұрақсыздыққа бейім келеді. Сол себепті, онлайн контенттегі қорлау сөздер корпусы және оны анықтауда ағылшын, француз және т.б. қолданыс деңгейі жоғары тілдер үшін зерттеулер жетерлік екенін байқауға болады. Ал, қолдану тілі танымал емес нәсілдік немесе этникалық топтар үшін Интернетте жасалған мәтіндерді автоматты түрде жинау және бағалау қиын. Бұл мақалада мәтінді өңдеуде және табиғи тілді өңдеуде жасанды интеллект әдістерін қолдану арқылы әлеуметтік желілерде ғадауат сөздерді анықтау үшін, алдымен, әртүрлі кілттік сөздерді пайдаланып деректер қорын жиналды. Екіншіден, мәтіндерді бірнеше санаттарға жіктеу үшін машиналық оқыту алгоритмдері қолданылды.

**Түйін сөздер:** ғадауат сөздер, машиналық оқыту, терең оқыту, алгоритм, корпус, аккаунт

**Кіріспе.** Интернеттің арқасында ақпаратқа қол жеткізу және оны тарату бұрын-соңды болмаған жылдамдықпен жүріп жатыр. Бұл сондай-ақ онлайн контентте тұлға өзін танымал етуге мүмкіндік алумен қатар, сонымен қатар, желі қолданушыға зиян келтіретін немесе қорлау, кемсіту мазмұндас пікірлерді де тарату үшін жақсы орта болып отыр [1]. Осыған байланысты біршама ғылыми зерттеушілер бұл құбылысты автоматты түрде анықтаудың қиындығын зерттеді [1], сонымен қатар осымен салыстырылатын мәселелерді шешуге арналған іс – әрекеттерді біршама класстары бар екендігі анықталды (мысалы, агрессия [2] немесе өшпенділік/қорлау мазмұндас пікірлер).

Негізгі үлгілерді немесе кілт сөздерді пайдалану бұл мәселені шешуде іске асыруға болатын стратегиялардың бірі болып табылады. Нәтижеге жету іс - әрекеттерін

---

\* E-mail корреспондирующего автора: [toktar.aigerim@list.ru](mailto:toktar.aigerim@list.ru)

классикалық жасанды интеллект тәсілдерімен біріктіру тұжырымдамасы кең таралған тәсілге айналуға. Сөз қапшығы (Bag of Words, BoW) – Kwok пен Wang қолданған әдіс және олар оны қажетті нәтижелерге қол жеткізу үшін аңғал Байес (Naive Bayes) классификаторымен біріктірді. [3]

Басқа зерттеушілер керісінше, машиналық оқыту алгоритмдері үшін функцияларды ұсынудың күрделі әдістерін қолданды (мысалы SVM, аңғал Байес және логистикалық регрессия) [4]. Бұл жалған нәтижелердің жоғары пайызы болғандықтан жасалды, деген қорытындыға әкелді. Эксперименттерді зерттеушілер кемсіту және кемсітушілік санатқа жатпайтын басқа көптеген түрлі классификацияларды қолдана отырып, (мысалы, айыптау, қорлау және т.б.) сызықтық ядросы бар тірек векторлық машинаны (SVM) пайдалануды ұсынды.

Әлеуметтік желі сайттарының трансформациялық потенциалы өзімен бірге көптеген қиындықтарды әкеледі, олардың кейбіреулері адамдар және белгілі бір топтар үшін зиянды болуы мүмкін [5]. Бұл қиын жағдай өзін жан-жақты көрсетеді, соның бірі – интернетте өшпенділік сөздердің таралуы үлкен алаңдаушылық пен арандатушылық туғызатынын байқауға болады. Тұрғылықты жері, ұстанымы, діни сенімі және экономикалық жағдайына қарамастан кез - келген адам кемсітушіліктің бірнеше түріне ұшырауы мүмкін. [1]

Сонымен қатар, ғадауат тілді сөздердің не екенін анықтау үшін бірқатар зерттеу жұмыстары жүргізілді және баспа беттеріне жарияланды. Тиісті корпустар оңай қол жетімді болғандықтан, ғадауат тілді сөздерді зерттеудің басым көпшілігі ағылшын тіліндегі деректерді пайдалана отырып жүргізілді. Дегенмен, соңғы жылдары қазақ, голланд және индонезия тілдеріндегі өшпенділік сөздерін анықтау үшін бірнеше зерттеулер жүргізілді.

Кейбір жағдайларда кілт сөзге негізделген әдістерді пайдаланушылардан немесе жеккөрінішті деп саналатын беттерден, яғни қорлау немесе кемсіту элементтерді табу мүмкіндігі болған кезде немесе даулы тақырыптардағы пікірталас форумдары беттерінен ғадауат тілді мазмұнды сөз тіркестерін табу ықтималдығы жоғары деп берілген.

Басқаша айтқанда, ғадауат тілді мазмұнды пікірлерді табу ықтималдығы артады. Бұл әртүрлі адамдар тобынан материал жинаумен байланысты шектеулерді ескере отырып жасалады. Зерттеушілер қорлайтын және кемсіту сөз тіркестері кездесетін пікірлерді жинау үшін біріктірілген стратегияны қолданды. Бұл әдіс ғадауат сөздер жазатын аккаунттарды бақылауды, анықталған аккаунт иелерінің тарихын жүктеп алуды және онлайн контенттегі посттарды кілт сөздермен сүзуді қамтиды. Табылған пікірлер осы тәсіл арқылы жиналды. Кейбір жағдайларда, бір аккаунттан барлық мәліметтер қоры жиналды деген болжам келтірген. Мұндай санаттау пікірлерге белгілі бір дереккөзден жиналған барлық элементтерді ғадауат тілді хабарламалар деп санауға болады деген болжамға негізделген. [6]

**Зерттеу әдістері.** Веб-ресурстардағы ғадауат тілді пікірлерді анықтау үшін ең алдымен машиналық әдістерді оқыту және сынау үшін қазақ тілінде корпус құру.

Зерттеудің бастапқы кезеңінде ғадауат тілді мәтіндермен қатар өшпенділік белгісі жоқ мәтіндерді корпустық жинақтау жоспарланған. Ғадауат тілді мазмұндағы мәтінге *дискриминация* (адам құқығын шамадан тыс бейәдеп сөздер арқылы шектеу), *қорлау*

тілі (жағымсыз сөздер арқылы кемсіту, қорлық көрсету), кибербуллинг (интернет желісінде адамды қорлау немесе қудалау), экстремизм (саяси және діни бағыттағы шектен шыққан жағымсыз пікірмен бөлісу) және радикализм (өмір сүріп отырған саяси органы сөзбен сынға алу) кіретінін көруге болады.

Өшпенділік белгісі кездеспейтін санатқа жататын мәтіндерге түрлі құттықтаулар, жан – жануарлар атауын жануарларға қатысты қолдану, қорқыту немесе кемсіту белгілері кездеспейтін пікірлер мен хабарламалар тізбегі, түрлі жарнамалық спамдар және т.б.

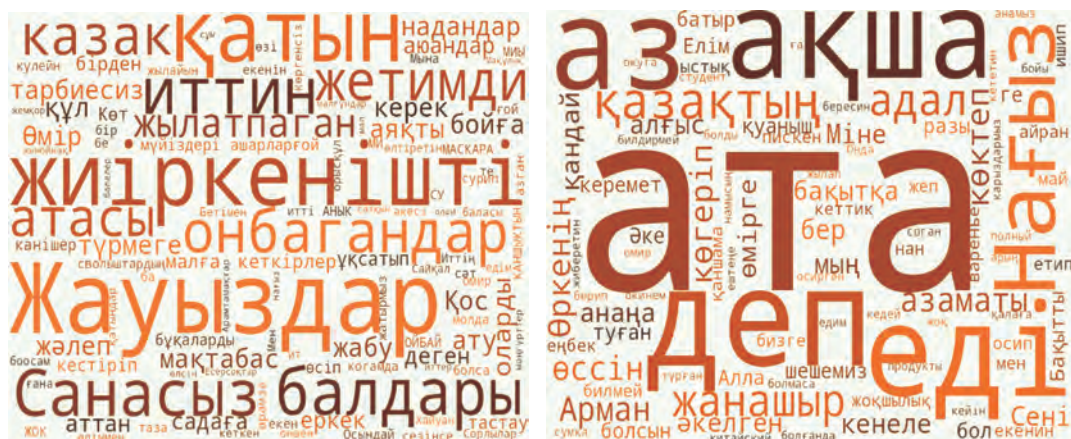
Зерттеудің басында онлайн ресурстардан ғадауат тілді мәтіндерді іздеу жұмыстары жүргізілді. Қазақ тіліндегі бейәдеп тілді пікірлер тізбегі Instagram әлеуметтік желісіндегі қазақ тілді аудиториясы бар блогерлер, әншілер, танымал тұлғалар аккаунттарынан ерікті аннотатор яғни, нақты бір мақсатта онлайн – іздеуші қазақ тілді ғадауат сөздерден тұратын мәтіндер жинады. Дегенмен, жиналған деректер қоры тек бір немесе екі – үш сөзден тұратын мәтіндер тізбегі ретінде жинақталған. Мысалы, «мал», «шошқа сияқты», «ешак екен» деген сөз тіркестері машиналық оқыту жүйесін үйрету жеткіліксіз болды. Мысалы, «мал» сөзінің мағынасы әр түрлі мазмұнда қолданылуы мүмкін, ол «мал азығы қораға жеткізілді» және «мал екені айтпай –ақ көрініп тұр» деген сөйлемдерде «мал» сөзін ғадауат сөздер қатарына косуда немесе қоспай қалдыруда машиналық оқытуда қаттелікке ұрынады. Сондықтан, онлайн парсинг және ерікті аннотатор арқылы Facebook- тегі қазақ тілді қауымдастықтар мен аккаунттар және Youtube платформаларынан үш немесе одан да көп сөздерден тұратын бейәдеп мазмұнды 10224 сөйлемдер жинақталды.

data hate speech -kazakh 10k.xlsx		A	B
		Қалмақ шалды пір тұтып жүргендер кезінде шалдың арқасында билікке, байлыққа қол жеткізгендер. Олардың : „Бізге егемендікті алып берген Назарбаев” дегенін естігенде есіңнен танып қала	
10027	жаздайсың. Құлдық сана оңайшылықпен өзгермейді екен.	Бұл адамның дінге ешқандай қатысы жоқ, мұсылмандар қауымдастығы деген қоғамды өзі ашып алған замованец, немен айналысатынын ешкім білмейді, әйтеуір ұсталған екен оңбаған, тексеру	1
10028	керек қай жердің жансызы екенін	Мына малдарға заң жоқпа күннен күнге өршіп бәрәтәр,билік қайда қарап отыр бұларды ату жазасына берсе көрер едік ойларына келгенді жасағандарын.Сәбилерге өмір сүруге мүмкіндік	1
10029	бермеген қайюандар.	Кішкентай қыздың өмірін құртқан педофилдерге ату жазасын беру керек.Неге Заң қатаң	1
10030	Болмайды.Ату жазасын берсін.Өлтіру керек педофилдерді.Сонда педофил дегендер болмайды.		1
10031	Обалай ату керек ондаймағұлынды		1
10032	Жанболат бостандыққа шықсын . Назарбаевпен Тоқаев қамалсын !	Шахмат деген ақыл ойдың ойыны емеспе, оған бәлендей жатықтырушыныңда қажеті шамалы сияқты, бейшара мен Мәскеуде тұрам дегенді дәреже көріп жүргенде, қазіргі саяси жағдайда Мәскеуіңе итте саруға жиіркенеді, орыстың қазаққа деген көзқарасыда жақсы емес, скинхэдтер былай тұрсын жәй орыстар оңаша кезігіп қалса айыратыны айдан анық дәлдұлың шығып	1
10033	жезөкшелікке түсіпкетпесең болар әйтеуір.	Кілең ұрылар отыр билікте,яғни бұндай қаражатты қайдан алды???.....	1
10034	Назарбаевтың ұры жүйесі қарапайым халықтың жүйкесін жұқартты	Керісінше бұл сатқын қазақтың төлқұжатын өткізсін. Дені сау болса, алдымен неге қашып кетті? Қара басын қорғаштап қашты. Одан әрі құрысын. Қазақты ойламаған, Отанын сатқындардың жазасы	1
10035	ауыр		1
10036	Қазақтар сігіп жіберер ашуна тисек деген шығар 🤔		1
10037	Провакатор группа сияқтығой бұл тупой админ нацист		1
10038	Сеннин ебальнигинен пил тепкен мени бузау тепсе	Өңкей еріншек, жалқаулар ғой мына тізім жасаған прокурорлар. Жөнді тізім жарты жылда жасауға	1
10039	да жарымады		1
10040	Тоқаев кыра салыпты ғой. Кудайынан қорыспаған бәле		1

Сурет 1 – Ғадауат тілді пікірлер үлгісі

Әзірленген корпуста ғадауат тілді мәтіндер бар құжат 3 бағаннан тұрады: хабарламаның немесе пікірдің реттік нөмірі, хабарлама немесе пікірдің өзі және хабарламаның немесе пікірдің ғадауат тілді екенін білдіретін «1» атрибуты Сурет 1-де көрсетілген. Алдағы уақытта корпусты ғадауат тілді мәтіндермен әрі қарай толтыру жоспарлануда.

*Сөздік бұлт.* Деректерді көрнекі түрде көрсету үшін сөздік бұлттар пайдаланылды. Ықтимал кемсіту немесе өшпенділік идеялары бар пайдаланушы жазбалары Сурет 2 –де бөлек көрсетілген. Көріп отырғанымыздай, ғадауат жазбалар пайдаланушылардың кемсітіп, қорлау немесе қорқыту мақсатында екенін тікелей көрсететін «жәлеп», «мақтабас құл», «атып тастау керек» сияқты сөздерді жиі пайдаланады.



а) ғадауат сөздер сөздік бұлтты

б) қалыпты сөздер сөздік бұлтты

Сурет 2 – Ғадауат сөздер және қалыпты сөздер сөздік бұлтты

#### *Ғадауат сөздер ерекшеліктері:*

Қазақ тіліндегі ғадауат тілді мәтіндерді талдау олардың өзіне тән ерекшеліктерге ие екендігін көрсетті:

- 1) қазақ тілінің әріптерін кириллица әріптерімен ауыстыру (мысалы, «сұмырай» - «сумырай», «өліп қал» - «олип қал» және «оңбаған» - «онбаған»),
- 2) қазақ тілінің әріптерін латын әріптерімен ауыстыру (мысалы, «шошқа» - «wowka», «қаншық» - «kanwyk» және «хайуан» - «хайуан»),
- 3) теңеу сөздер арқылы немесе метафра арқылы кемсіту (мысалы, «мақтабас», «орысқұл» және «одақбас»),
- 4) тұрғылықты жеріне байланысты ғадауат кемсіту сөздерін қолдану (мысалы, «южандар», «тexas», «орыстар», «картопбас»),
- 5) орфографиялық қателер немесе әріп орнына символ қолданып жазу (мысалы, «к.т екенсің», «шешен @мы»),
- 6) орыс тілінен енген сөздерді қосымша жұрнақтар арқылы қазақшалау (мысалы, «далбандар», «сукасың», «черттар»).

Ғадауат сөздерден тұратын мәліметтер қоры төмендегі типтегі мазмұнды тексттерді қамтиды:

1) *адамның бойы бойынша*, демек бойының ұзындығын немесе қысқалығын басшылыққа ала «отырып, сөз арқылы жәбір көрсету, мысалы, «мыртыған қортық», «бойы сорайған керік» деген сияқты сөз тіркестері кездеседі.

2) *жасы бойынша (эйджизм)* – адамның жасына байланысты алалау, егде жастағы адамдарды, кішкентай балаларды кемсітуде жиі қолданылатынын байқауға болады. Мысалы, «кәрі кемпір», «кәрі қарға», «алжыған қақпас», «қаршадай болып бәлесін қарашы» деген сияқты сөз тіркестері кездеседі.

3) *касталық дискриминация* – белгілі бір тұрғылықты жердегі адамдарды, шығу тегіне немесе санының аз болуына қарай қоғамдық топтарды құқығын шектеу немесе бөлу, мысалы, солтүстік тұрғындарын «орыстар», «мәнгүрттер», оңтүстік тұрғындарын «ежиктер» деген сияқты сөз тіркестерін кездестіруге болады.

4) *мамандығына немесе жұмыс сферасы бойынша дискриминация*, қандай да бір мамандық иесін кемсіту, мысалы әншілерді «секеңдеген ешкі» немесе дәрігерлерді «қолына ақша бермесен өлгеніңмен жұмысы жоқ», сонымен қатар «топастар ғой, дипломын сатып алған» деген мағынадағы сөз тіркестерін кездестіруге болады.

5) *эйлизм немесе дисаблизм* – адамдарды мүгедектік сипатына немесе физикалық ерекшеліктері бойынша алалауды жатқызуға болады, мысалы «кеміс», «жынұрған», «топас», «құбыжық» деген сөздер топтамасын осы бөлімге жатқызуға болады.

6) *расизм* – этникалық шығу тегіне және нәсіліне қарай адамдарды бөліп қарау. Нәсілшілдер адамның түр – түсіне, өзін-өзі ұстау ерекшелігі мен жүріп-тұру әдебіне қарай оны өзгелерден бөліп көрсетуге, басқаша екенін кемсіту арқылы дәлелдеуге тырысады, мысалы «өзбек», «сарт», «кәпір», «орысқұл», «шошқақұл», «арапқұл» деген сөз тіркестерінен тұратын сөйлемдерді кездестіруге болады.

7) *діни сенімі бойынша дискриминацияда* адам өзі сенетін және ұстанатын дінін басқалардан жоғары қояды, яғни өзге дінді ұстанатындарды төмендетіп, алалауға тырысады, «исламистер», «арапқұлдар», «кәпірлер» деген сөз тіркестерінен байқауға болады.

8) *сексизм* – адамды жынысына байланысты бөліп, бірінен екіншісін артық қою. Гендер теңсіздігімен байланыстыруға болады. Сексизм — әйел мен ер адамдарды жынысына қарай алалау, күші мен құқығын тең етіп қабылдау, мысалы «гейлер», «лесбилер», «қатын екенсің» деген ғадауат тілді сөздер бар сөз тіркестерін байқауға болады.

9) *гетеросексизм немесе гомофобия* – жыныс ориентациясына қарай қоғамнан қорлану, мазаққа ұшырау және құқығы шектелуі.

10) *сөйлеу тілі бойынша кемсіту* бір тілде ғана сөйлейтін немесе қандай да бір тілде сөйлей алмайтын адамды кемсіту мен алалау. [7]

**Зерттеу нәтижелері.** Жоғарыдағы жұмыстың нәтижесінде а) онлайн парсингтер көмегімен және аннотатор көмегімен «Instagram» әлеуметтік желісінен, «Youtube» және Facebook әлеуметтік желілердегі қазақ тілді желі қолданушылар парақшалары мен топтарынан мәтіндік деректер жиналды; б) мәтіндерді «ғадауат тілді» және «нейтральды тексттер» деп екі сыныпқа қолмен жіктеу жасалды; в) модельді одан әрі оқыту үшін корпус құрылды; г) ғадауат тілді және ешқандай өшпенділік сипаты кездеспейтін мәтіндердің, сөздердің бұлттарының таралуын пайдалана отырып, корпусқа талдау жүргізіліп, ғадауат тілді мазмұндағы мәтіндердің ерекшеліктері анықталды.

*Ғадауат тілді пікірлерді анықтауда терең оқытуды қолдану.* Терең оқыту эмоцияларды талдау, мәтінді жіктеу және автоматты аударма сияқты табиғи тілді өңдеу (NLP) тапсырмалары үшін өте қолайлы. Мәтінді жіктеу үшін біз терең оқытудың келесі әдістерін қолданамыз: Конволюционды нейрондық желі (CNN) және LSTM қайталанатын нейрондық желі. Терең оқыту эмоцияны талдау, мәтінді жіктеу және машиналық аударма сияқты табиғи тілді өңдеу (NLP) тапсырмалары үшін өте қолайлы. Мәтінді жіктеу үшін біз терең оқытудың келесі әдістерін қолданамыз: *Конволюционды нейрондық желі (CNN) және LSTM қайталанатын нейрондық желі.*

Жинақталған деректер қорын сүзу үшін келесі алгоритм қолданылды:

• барлық таңбаларды бір регистрге келтіру және қосымша таңбаларды алып тастау,

• тоқтату сөздерді алып тастау (стоп - слова),

• штрихтау мен лемматизацияны орындау;

• лексемалау үлгісін (мәтінді сөздерге – лексемаларға бөлу) және сөздердің n-граммалық үлгісін (лексемадағы мүмкін сөздердің санын) көрсету.

Нейрондық желілер тек сандық деректердегі үлгілерді табуды үйрене алады, сондықтан нейрондық желіге енгізу ретінде мәтінді енгізбес бұрын, біз әрбір сөзді сандық мәнге түрлендірдік. Бұл процесс сөзді кодтау немесе токенизация деп аталады.

Токенизация үшін сөздерді ендіруді қолданылды, бұл әдіс сөздерді тығыз сөз векторлары ретінде көрсетеді (сондай-ақ сөзді ендіру деп аталады). Бұл «енгізу» сөзі көп ақпаратты аз өлшемдерге жинайтынын білдіреді. Олардың мақсаты – геометриялық кеңістікте семантикалық мағынаны көрсету. Бұл геометриялық кеңістік кірістіру кеңістігі деп аталады. Бұл сандар немесе түстер сияқты ұяшық кеңістігіне жақын мағыналық ұқсас сөздерді көрсетеді.

Деректерді таңбалау үшін Керастағы Tokenizer утилитасы қолданылды, ол мәтіндік корпусты бүтін сандар тізіміне векторлай алады.

```
from keras.preprocessing.text import Tokenizer
from keras.preprocessing.sequence import pad_sequences
from keras.preprocessing import text, sequence

tokenizer = Tokenizer(num_words=20000)
tokenizer.fit_on_texts(list(X_train))

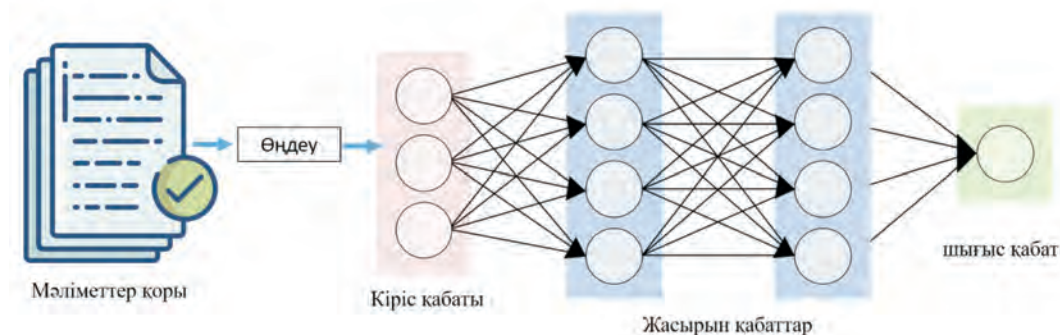
X_train = tokenizer.texts_to_sequences(X_train)
X_test = tokenizer.texts_to_sequences(X_test)

X_train = sequence.pad_sequences(X_train, maxlen=200)
X_test = sequence.pad_sequences(X_test, maxlen=200)

print('X_train shape:', X_train.shape)
print('X_test shape:', X_test.shape)
```

Токенизаторда екі параметр қолданылды: сөздік өлшемін орнатуға жауап беретін `num_words` және сөздер тізбегін нөлмен толтыратын `pad_sequence()`, ол параметрі мәтін тізбегіндегі әртүрлі ұзындықтағы сөздердің мәселесін шешу үшін қолданылады. Сөздік өлшемін орнатуға жауапты `num_words` параметрі қосылды. Сөз\_санының мәнін 2000 мәніне орнатылды. Зерттеу мәселесінің бірі әрбір мәтін тізбегінің көп жағдайда сөз ұзындығының әртүрлі болуы. Ал реттіліктердің қаншалықты ұзақ болуы керектігін көрсету үшін `maxlen` параметрі қосылды. Бұл осы саннан асатын тізбектерді қысқартады. [8]

*Мәтінді жіктеу үшін терең оқыту моделін құру.* Нейрондық желіде кіріс қабаты, жасырын қабат және шығыс қабаты сияқты бірнеше терминдерді кездестіруге болады. Сонымен, терең оқыту архитектурасы мен нейрондық желінің архитектурасы арасындағы айырмашылық жасырын қабаттардың саны болып табылады. Қарапайым нейрондық желіде тек 1 жасырын қабат болса, Deep Learning-те 1-ден көп жасырын қабат бар екендігін төмендегі Сурет 3 – тен байқауға болады.



Сурет 3 – Терең оқыту моделінің архитектурасы

Ең алдымен кіріс нейрондарының қабатынан басталады, онда функция векторларын енгізіп, содан кейін мәндер жасырын қабаттарға ауыстырылады. Әрбір қосылымда мән алға жіберіледі ал мән салмаққа көбейтіледі және мәнге ығысу қосылады. Бұл әрбір қосылымда орын алады және соңында шығыс қабатының мәні алынады. Шығыс деңгейі бір немесе бірнеше шығыс түйіндерінен тұрады. Біздің жағдайда, бір түйін, себебі екілік классификациясы орындалады.

Нейрондық желі формуласы: әрбір шығыс түйінінің мәндерін есептеу үшін біз әрбір кіріс  $p$  түйінін  $W$  салмағына көбейтіп,  $b$  қиғаштығын қосу қажет. Содан кейін мұның барлығын қорытындылап,  $f$  функциясына беру керек. Бұл функция белсендіру функциясы болып саналады және әртүрлі функциялар бар. Әдетте жасырын қабаттар үшін түзетілген сызықтық блок (ReLU), екілік классификация есепіндегі шығыс қабат үшін сигма тәрізді функция немесе көп класты жіктеу есептерінің шығыс қабаты үшін `softmax` функциясы қолданылады. Сигма тәрізді функцияны қолданамыз. Алгоритм салмақтарды кездейсоқ мәндермен инициализациялаудан басталады, содан кейін олар кері таралу деп аталатын әдіспен оқытылады. Бұл есептелген және қажетті нәтиже (сонымен бірге мақсатты нәтиже деп аталады) арасындағы қатені азайту үшін

градиенттің түсуі сияқты оңтайландыру әдістері (оңтайландырушы деп те аталады) арқылы орындалады. Қате жоғалту функциясымен анықталады, оның жоғалуын келесі зерттеулерде оптимизатордың көмегімен азайтуға болатындығын көрсетуге болады. [9]

**Талқылаулар.** Осылайша, бұл зерттеуде әлеуметтік желілердегі ғадауат сөздерге қатысты мәтінді анықтаудың ұсынылатын әдісі түсіндірілді, онда деректерді алу, n-грамды анықтау, мәтінді токенизациялау, тоқтау сөздерді жою сияқты әртүрлі әдістерді қолдану арқылы деректерді тазалау сияқты әртүрлі қадамдарды пайдалана отырып, нәтижесінде алдын ала өңделген мәтіндерді алу.

Келесі қадам ретінде, ғадауат тілді қазақша сөйлемдер корпусын оқу деректер жинағы және сынақ (тест) жинағы ретінде екі бөлікке бөлінеді. Ұсынылған деректер жинағы 10224 сөйлемнен тұрады. Олардың 7683-і (75%) үлгілерді оқытуға, 2556-і (25%) үлгіні тексеруге арналған.

Сонымен қатар, өз алдына зияны жоқ көптеген идиомалар бар; дегенмен, егер оларды тиісті контексте қолданылса, олар қорлау, кемсіту тілдер қатарына жатуы мүмкін [10]. Дегенмен, ғадауат тілдерді қолдану жағдайында да әртүрлі желі қолданушылардан әр түрлі дәрежедегі құқық бұзушылықтары болмауы мүмкін емес [11], сонымен бірге іс – әрекеттер, сондай-ақ бір терминнің түрлі деңгейде қолданылуына байланысты әр уақытта өзгеруі мүмкін.

**Қорытындылай келе,** ғадауат сөздерін коммуникацияның басқа түрлерінен ерекшелендіретін санаттау үлгісін әзірлеуге қол жеткіздік. Бұған қоса, классификациялық модельдерімізді оқытуға енгізілген мәтіндік мүмкіндіктерді жасадық. Біздің соңғы жіктеу үлгіміз әлеуметтік медиа сайттарында ғадауат сөздерін анықтау үшін машиналық оқытуды қолдануға бола ма деген сұраққа жауап берді. Қолданыстағы зерттеулерге қатысты белгілі бір терминдердің ғадауат тілді сөздерді мен бейәдеп мағынасы жоқ сөйлеу коммуникацияларын ажырату үшін маңызды екендігі танылды. Кемсіту хэштегімен белгіленген твиттерде гомофобиялық, нәсілшілдік және сексистік пікірлер бар. Бұл кемсіту мен қорлайтын хабарламаларды анықтауды жеңілдетсе де, егер олар стандартты жек көретін сөздерді қамтымаса, алгоритм фразаларды қате жіктеуі мүмкін. Осы деректерден шығатын тағы бір қорытынды: ғадауат тілді сөздер белгілі бір адамға, белгілі бір топқа бағытталған болуы мүмкін немесе ешқандай мақсатсыз қолданылуы мүмкін.

## ӘДЕБИЕТ

1 Gröndahl, T., Pajola, L., Juuti, M., Conti, M., & Asokan, N. (2018, January). All you need is» love» evading hate speech detection. In Proceedings of the 11th ACM workshop on artificial intelligence and security (pp. 2-12).

2 Rohmawati, U. A. N., Sihwi, S. W., & Cahyani, D. E. (2018, November). SEMAR: An interface for Indonesian hate speech detection using machine learning. In 2018 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI) (pp. 646-651). IEEE.

3 TOKTAROVA, A., IZTAEV, Z., KOZHABEKOVA, P., SUIEUOVA, N., OPONDO, R. O., KERIMBEKOV, M., & ZHUNISBEKOVA, Z. (2022). AUTOMATED HATE SPEECH CLASSIFICATION USING EMOTION ANALYSIS IN SOCIAL MEDIA USER GENERATED TEXTS. Journal of Theoretical and Applied Information Technology, 100(22).



4 Omar, A., Mahmoud, T. M., & Abd-El-Hafeez, T. (2020, April). Comparative performance of machine learning and deep learning algorithms for Arabic hate speech detection in osns. In *The International Conference on Artificial Intelligence and Computer Vision* (pp. 247-257). Springer, Cham.

5 Arango, A., Pérez, J., & Poblete, B. (2020). Hate speech detection is not as easy as you may think: A closer look at model validation (extended version). *Information Systems*, 101584.

6 Sultan, D., Suliman, A., Toktarova, A., Omarov, B., Mamikov, S., & Beissenova, G. (2021, January). Cyberbullying Detection and Prevention: Data Mining in Social Media. In *2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence)* (pp. 338-342). IEEE.

7 Nayel, H. A., & Shashirekha, H. L. (2019, December). DEEP at HASOC2019: A Machine Learning Framework for Hate Speech and Offensive Language Detection. In *FIRE (Working Notes)* (pp. 336-343).

8 Sultan, D., Toktarova, A., Zhumadillayeva, A., Aldeshov, S., Mussiraliyeva, S., Beissenova, G., ... & Imanbayeva, A. (2023). Cyberbullying-related hate speech detection using shallow-to-deep learning. *CMC-COMPUTERS MATERIALS & CONTINUA*, 74(1), 2115-2131.

9 Ibrohim, M. O., & Budi, I. (2019, August). Multi-label hate speech and abusive language detection in Indonesian twitter. In *Proceedings of the Third Workshop on Abusive Language Online* (pp. 46-57).

10 Toktarova, A., Beissenova, G., Kozhabekova, P., Makhanova, Z., Tulegenova, B., Rakhymbek, N., ... & Azhibekova, Z. (2021). Automatic offensive language detection in online user generated contents. *Journal of Theoretical and Applied Information Technology*, 99(9), 2054-2067.

11 K. Kumari, J. P. Singh, Y. K. Dwivedi and N. P. Rana, "Bilingual cyber-aggression detection on social media using LSTM autoencoder," *Soft Computing*, vol. 25, no. 14, pp. 8999–9012, 2021

## REFERENCES

1 Gröndahl, T., Pajola, L., Juuti, M., Conti, M., & Asokan, N. (2018, January). All you need is" love" evading hate speech detection. In *Proceedings of the 11th ACM workshop on artificial intelligence and security* (pp. 2-12).

2 Rohmawati, U. A. N., Sihwi, S. W., & Cahyani, D. E. (2018, November). SEMAR: An interface for Indonesian hate speech detection using machine learning. In *2018 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)* (pp. 646-651). IEEE.

3 TOKTAROVA, A., IZTAEV, Z., KOZHABEKOVA, P., SUIEUOVA, N., OPONDO, R. O., KERIMBEKOV, M., & ZHUNISBEKOVA, Z. (2022). AUTOMATED HATE SPEECH CLASSIFICATION USING EMOTION ANALYSIS IN SOCIAL MEDIA USER GENERATED TEXTS. *Journal of Theoretical and Applied Information Technology*, 100(22).

4 Omar, A., Mahmoud, T. M., & Abd-El-Hafeez, T. (2020, April). Comparative performance of machine learning and deep learning algorithms for Arabic hate speech detection in osns. In *The International Conference on Artificial Intelligence and Computer Vision* (pp. 247-257). Springer, Cham.

5 Arango, A., Pérez, J., & Poblete, B. (2020). Hate speech detection is not as easy as you may think: A closer look at model validation (extended version). *Information Systems*, 101584.

6 Sultan, D., Suliman, A., Toktarova, A., Omarov, B., Mamikov, S., & Beissenova, G. (2021, January). Cyberbullying Detection and Prevention: Data Mining in Social Media. In *2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence)* (pp. 338-342). IEEE.

7 Nayel, H. A., & Shashirekha, H. L. (2019, December). DEEP at HASOC2019: A Machine Learning Framework for Hate Speech and Offensive Language Detection. In *FIRE (Working Notes)* (pp. 336-343).

8 Sultan, D., Toktarova, A., Zhumadillayeva, A., Aldeshov, S., Mussiraliyeva, S., Beissenova, G., ... & Imanbayeva, A. (2023). Cyberbullying-related hate speech detection using shallow-to-deep learning. CMC-COMPUTERS MATERIALS & CONTINUA, 74(1), 2115-2131.

9 Ibrohim, M. O., & Budi, I. (2019, August). Multi-label hate speech and abusive language detection in Indonesian twitter. In Proceedings of the Third Workshop on Abusive Language Online (pp. 46-57).

10 Toktarova, A., Beissenova, G., Kozhabekova, P., Makhanova, Z., Tulegenova, B., Rakhymbek, N., ... & Azhibekova, Z. (2021). Automatic offensive language detection in online user generated contents. Journal of Theoretical and Applied Information Technology, 99(9), 2054-2067.

11 K. Kumari, J. P. Singh, Y. K. Dwivedi and N. P. Rana, "Bilingual cyber-aggression detection on social media using LSTM autoencoder," Soft Computing, vol. 25, no. 14, pp. 8999–9012, 2021

**А. Б. ТОКТАРОВА<sup>1</sup>, Б. С. ОМАРОВ<sup>2</sup>, Ж. Ж. АЖИБЕКОВА<sup>3</sup>,  
А. Н. ЖАКСАНОВА<sup>4</sup>, Р. Б. АБДРАХМАНОВ<sup>5</sup>**

<sup>1</sup>Международный казахско-турецкий университет им. Ходжа Ахмета Ясауи,  
Туркестан, Казахстан

<sup>2</sup>Казахский национальный университет им. аль-Фараби, Алматы, Казахстан

<sup>3</sup>Казахский национальный медицинский университет имени С. Д. Асфендиярова,  
Алматы, Казахстан

<sup>4</sup>Южно-Казахстанский университет имени Мухтара Ауезова,  
Шымкент, Казахстан

<sup>5</sup>Международный университет туризма и гостеприимства, Туркестан, Казахстан

### **АВТОМАТИЗИРОВАННАЯ КЛАССИФИКАЦИЯ ОСКОРБИТЕЛЬНЫХ СЛОВ ПО «ЭМОЦИОНАЛЬНЫМ» КОММЕНТАРИЯМ ПОЛЬЗОВАТЕЛЕЙ СЕТИ**

*В последние годы язык ненависти в Интернете получил широкое распространение в результате легкости, с которой любое онлайн-мнение может быть представлено в социальных сетях. Эта тенденция в основном связана с быстрым ростом мобильных устройств, компьютеров и Интернета. Существует большой интерес к исследованиям по автоматическому обнаружению ненормативной лексики. Тем не менее, существует еще мало исследований о том, могут ли сайты социальных сетей помочь в выявлении преступлений на почве ненависти и связанных с ними сообществ. Плохие мнения негативно сказываются на психическом уровне любого человека. Однако некоторые люди более склонны к эмоциональной нестабильности, чем другие. Расовым или этническим группам, язык которых не пользуется популярностью, сложно автоматически собирать и оценивать тексты, созданные в Интернете. В этой статье с использованием методов искусственного интеллекта в обработке текста и обработки естественного языка впервые была собрана база данных с использованием различных ключевых слов для выявления оскорбительных слов в социальных сетях. Во-вторых, алгоритмы машинного обучения использовались для классификации текстов по нескольким категориям.*

**Ключевые слова:** оскорбительные слова, машина обучение, глубокое обучение, алгоритм, корпус, аккаунт.

**A. B. TOKTAROVA<sup>1</sup>, B. S. OMAROV<sup>2</sup>, ZH. ZH. AZHIBEKOVA<sup>3</sup>,  
A. N. ZHAXANOVA<sup>4</sup>, R. B. ABDRAKHMANOV<sup>5</sup>**

<sup>1</sup>*Khoja Akhmet Yassawi International Kazakh – Turkish University, Turkistan, Kazakhstan*

<sup>2</sup>*AL-Farabi Kazakh National University, Almaty, Kazakhstan*

<sup>3</sup>*Asfendiyarov Kazakh National Medical University, Almaty, Republic of Kazakhstan*

<sup>4</sup>*Mukhtar Auezov South Kazakhstan University, Shymkent, Kazakhstan*

<sup>5</sup>*International University Of Tourism And Hospitality  
Turkestan, Kazakhstan*

### **AUTOMATED OFFENSIVE LANGUAGE CLASSIFICATION THROUGH “EMOTIONAL” COMMENTS FROM NETWORK USERS**

*In recent years, online hate speech has become widespread as a result of the ease with which any online opinion can be presented on social media. This trend is mainly due to the rapid growth of mobile devices, computers and the Internet. There is a lot of interest in research on automatic hate speech detection. However, there is not a lot research on whether social networking sites can help identify hate crimes and related communities. Hate speech has a negative impact on the mental level of any person. However, some people are more prone to emotional instability than others. It is difficult for racial or ethnic groups whose language is not popular automatically collecting and evaluate texts created on the Internet. In this article, using artificial intelligence techniques in text processing and natural language processing, for the first time, a database was collected using various keywords to identify offensive words in social networks. Second, machine learning algorithms have been used to classify texts into multiple categories*

**Key words:** *offensive language, machine learning, deep learning, algorithm, corpus, account*