ҒТАХР 28.23.37

10.51889/2959-5894.2023.84.4.017

Б.С. Омаров[1], А.Б. Тоқтарова[2,4], Ж.Ж. Ажибекова [3],*
Г.С. Рахимбаева[3], Г.И. Бейсенова[4]

[1]Әл-Фараби атындағы Қазақ Ұлттық университеті Алматы қ., Қазақстан
[2]Қожа Ахмет Ясауи атындағы Халықаралық Қазақ -Түрік Университеті, Түркістан қ., Қазақстан
[3]С. Ж. Асфендияров атындағы Қазақ ұлттық медицина университет, Алматы қ., Қазақстан
[4]М. Ауезов атындағы Оңтүстік Қазақстан университеті, Шымкент қ., Қазақстан
[*]e-mail: toktar.aigerim@list.ru

ҒАЙБАТ ПІКІРЛЕРДІ АНЫҚТАУДА ЕКІ БАҒЫТТЫ ҰЗАҚ-ҚЫСҚА МЕРЗІМДІ ЖАД ЖЕЛІСІН ҚОЛДАНУ

Аңдатпа

Онлайн контенттегі ғадауат тілді сөздерді анықтау бүгінгі цифрлық дәуірде мазмұнды модерациялаудың тиімді жүйелерін дамытуға мүмкіндік беретін маңызды міндет болып табылады. Дегенмен, дайын бейәдеп тілден құралған мәліметтер қоры аз тілдерде оларды анықтау біршама қиындықтар тудыратынын байқауға болады. Бұл зерттеу жұмысында ғадауат тілді сөздерді анықтау бойынша мәліметтер ресурсы төмен тіл ретінде қазақ тілді контекстерді шешуге бағытталады. Ұсынылған тапсырманы шешу үшін біз табиғи тілді өңдеу алгоритмдерінде жоғары нәтиже көрсеткен екі бағытты ұзақ - қысқа мерзімді жады (BiLSTM) желілеріне негізделген жаңа тәсілді ұсынамыз. BiLSTM архитектурасының екі жақты сипатын пайдалана отырып, біз бейәдеп тілдің дәлірек сәйкестендірілуін қамтамасыз ететін кіріс мәтініндегі контекстік және ұзақ мерзімді тәуелділіктерді аламыз. Біз ұсынып отырған әдісте сонымен қатар трансферлі оқыту әдістерін ресустары аз тілдердің деректердің жетіспеушілігін азайту үшін де қолдануға болады. Қазақ тілінің ғадауат тілді сөздер деректер жинағымен бірқатар эксперимент жүргізу арқылы біз ресурсы төмен қазақ тіліндегі ғадауат тілді сөздерді анықтауда соңғы нәтижелерге қол жеткізе отырып, ұсынылған әдісіміздің тиімділігін көрсетеміз. Сонымен қатар, біз әртүрлі үлгі конфигурацияларының және оқыту стратегияларының біздің әдісте жұмыс істеу тиімділігін талдаймыз. Біздің зерттеуіміздің нәтижелері ресурсы аз тілдердегі ғадауат тілді сөздерді анықтау әдістері туралы мәліметтер ұсына алады және нақты тілдік контексттерге бейімделген мазмұнды модерациялау жүйелеріне жол ашады.

Түйін сөздер: бейәдеп тілді сөздер, аз ресурсты тіл, BiLSTM, машиналық оқыту алгоритмдері.

Б.С. Омаров[1], А.Б. Токтарова[2,4], Ж.Ж. Ажибекова [3], Г.С. Рахимбаева[3], Г.И. Бейсенова[4]
[1]Казахский Национальный университет им. аль- Фараби, г.Алматы, Казахстан
[2]Международный казахско - турецкий университет им. Ходжа Ахмет Ясауи, г. Түркестан, Казахстан
[3]Казахский Национальный медицинский университет имени С. Д. Асфендиярова, г. Алматы, Казахстан
[4]Южно Казахстанский Университет имени М.Ауезова, г.Шымкент, Казахстан

ИДЕНТИФИКАЦИЯ НЕЦЕНЗУРНЫХ КОМЕНТАРИЕВ С ИСПОЛЬЗОВАНИЕМ ДВУНАПРАВЛЕННОЙ ДОЛГОВРЕМЕННОЙ КРАТКОВРЕМЕННОЙ ПАМЯТИ

Аннотация

Выявление ненормативной лексики является важнейшей задачей в нынешнюю цифровую эпоху, что позволяет создавать эффективные системы модерации контента. Тем не менее, это создает проблемы в языках с ограниченными ресурсами, где доступны небольшие количества аннотированных данных. Эта исследовательская работа пытается решить проблему определения оскорбительного языка на малоресурсном, казахском языке. Мы предлагаем новый подход, основанный на сетях двунаправленной долговременной кратковременной памяти (BiLSTM), который продемонстрировал

высокую эффективность в задачах обработки естественного языка, где этот подход решает эту проблему. Мы можем более точно идентифицировать оскорбительный язык во входном тексте, фиксируя как долгосрочные, так и контекстные зависимости, используя двунаправленный характер архитектуры BiLSTM. Чтобы уменьшить нехватку аннотированных данных при ограниченных ресурсах, наш метод также использует методы трансферного обучения. После проведения обширных экспериментов с набором данных оскорбительных языков в казахском языке мы демонстрируем эффективность предложенного нами метода. Эти эксперименты показывают самые современные результаты в определении оскорбительных языков в низкоресурсном казахском языке. Кроме того, мы рассматриваем, как различные конфигурации модели и методы обучения влияют на эффективность нашего метода. Наше исследование дает полезную информацию о способах обнаружения оскорбительного языка в языках с низким уровнем ресурсов. Кроме того, они прокладывают путь к более надежным системам модерации контента, которые подходят для определенных языковых контекстов.

Ключевые слова: нецензурная речь, малоресурсный язык, BiLSTM, алгоритмы машинного обучение.

B. Omarov [1], A. Toktarova [2,4], Zh. Azhibekova [3], G. Rakhimbayeva [3], G. Beissenova [4]
[1]AL-Farabi Kazakh National University, Almaty, Kazakhstan
[2]Khoja Akhmet Yassawi International Kazakh - Turkish University, Turkistan, Kazakhstan
[3]Asfendiyarov Kazakh National Medical University, Almaty, Kazakhstan
[4]South Kazakhstan University named after M.Auezov, Shymkent, Kazakhstan

IDENTIFICATION OFFENSIVE COMMENTS BY USING BIDIRECTIONAL LONG-SHORT-TERM MEMORY

Abstract

The detection of profanity is a critical task in the current digital age, which allows you to create effective content moderation systems. However, this creates problems in resource-constrained languages where small amounts of annotated data are available. This research work attempts to solve the problem of defining offensive language in a low-resource language, Kazakh. We propose a new approach based on Bidirectional Long Short Term Memory (BiLSTM) networks, high performance in natural language processing tasks, this approach solves this problem. We can more accurately identify the offending language in the input text by capturing both long term and context dependencies using the bidirectional nature of the BiLSTM architecture. To reduce the shortage of annotated data with limited resources, our method also uses transfer learning methods. After conducting extensive experiments with a data set of offensive languages in the Kazakh language, we demonstrate the effectiveness of our proposed method. These experiments show the most up-to-date results in identifying offensive languages in low-resource Kazakh. In addition, we consider how different model configurations and training methods affect the performance of our method. Our research provides useful information on how offensive language detected in low-resource languages. In addition, they pave the way for more robust content moderation systems that are appropriate for certain language contexts.

Keywords: obscene language, low-resource language, BiLSTM, machine learning algorithms.

Кіріспе

Соңғы жылдары әлеуметтік медиа платформалары мен онлайн байланыс арналарының көбеюі жаһандық ауқымда ақпарат пен идеялардың жылдам алмасуына ықпал етті. Кескінді өңдеу, автоматтандыру, мәтінді өңдеу және т.б. үшін машиналық оқытуды қолданатын көптеген қолданбалы бағдарламалар бар [1-3]. Бұл байланыс көптеген артықшылықтар әкелгенімен, ол сондай-ақ күрделі проблеманы тудырды: желідегі бейәдеп тілді сөздер мен ғадауат тілді сөздердің таралуы. Кемсіту немесе қорлау көртесу белгілері бар сөздер жеке адамдар мен қауымдастықтарға зиянын тигізіп қана қоймайды, сонымен қатар онлайн платформаларды қолдануда желі пайдаланушысына зиян келтіруі мүмкін [3]. Сондықтан ғадауат тілді және сөздер мазмұнын ажыратуды анықтаудың сенімді және тиімді жүйелерін әзірлеу қажеттілікке ие болып отыр. Ғадауат тілді анықтау бойынша зерттеулер негізінен ағылшын, испан және француз тілдері сияқты ресурсы жеткілікті мөлшердегі тілдерге бағытталған. Бұл тілдер ғадауат тілді сөздерді анықтауда жоғары дәлдікті қамтамасыз ететін

күрделі машиналық оқыту үлгілерін қолдануға мүмкіндік беретін таңбаланған деректердің үлкен көлемі зерттеу жұмысында жақсы нәтиже беріп отыр [3]. Дегенмен, бейәдеп сөздерден құралған деректердің жоқтығы үлкен кемшілік болып табылатын ресурсы аз тілдер үшін де дәл солай болжамдау қиындық туғызып отыр. Ресурсы аз тілдер әдетте жағымсыз сөздер деректер жиынын, тіл үлгілерін және алдын ала дайындалған ендірулерді жүзеге асыруда шектеулі тілдік ресурстармен сипатталады [4]. Бұл тапшылық осы тілдердің лингвистикалық сипаттмалары мен мәдени мазмұнды сөздерге бейімделген тілдің анықтау жүйелерін дамытуға кедергі келтіреді.

Бұл зерттеу жұмысымызда қазақ тіліне ерекше назар аудара отырып, ресурсы төмен тілдердегі ғадауат сөздерді анықтау мәселесін арнайы қарастырдық. Қазақ тілі негізінен Қазақстанда және көршілес аймақтарда сөйлейтін түркі тілі болып табылады және тегтелген деректер мен тілдік ресурстардың шектеулі болуына байланысты ресурсы төмен тіл санатына жатады [5]. Біздің мақсатымыз – қазақ тілінің бірегей сипаттамаларын тиімді өңдей алатын сенімді және дәл ғадауат тілді анықтау үлгісін жасау. Осы мақсатқа жету үшін біз табиғи тілді өңдеудің әртүрлі есептеулерінде жақсы нәтиже көрсеткен екі бағытты ұзақ қысқа мерзімді жады (BiLSTM) желілеріне негізделген жаңа тәсілді ұсынамыз [6]. BiLSTM архитектурасы негізгі семантиканы толық түсінуді қамтамасыз ете отырып, кіріс мәтініндегі тікелей және кері контексттік тәуелділіктерді анықтайды [7]. Осы екі жақты модельдеуді пайдалана отырып, біздің әдістемеміз ресурсы төмен қазақ тілінде ғадауат тілді анықтау жұмысын жақсартуға бағытталған.

Дегенмен, ресурсы төмен тілдерде ерікті түрдегі аннататорлар көмегімен жинақталған деректердің тапшылығы модельді оқыту үшін айтарлықтай қиындық тудырады. Бұл мәселені жеңілдету үшін біз жоғары ресурс тілдеріндегі ауқымды деректер жиынында оқытылатын алдын ала дайындалған тіл үлгілерін пайдалана отырып, оқытудың трансферттік әдістерін қолданамыз [8]. Бұл модельдерді шектеулі қазақ тіліндегі ғадауат тілді деректер жинағында дәл анықтау арқылы біз ресурсы төмен қазақ тіліндегі бейәдеп тілді анықтау моделінің өнімділігін жақсарту үшін жоғары ресурсты тілдерден үйренген білімді беруді мақсат етеміз. Бұл зерттеу жұмысында, біз қазақ тілінің бейәдеп сөзді тілдер деректер жиынтығына қатысты біздің ұсынылған тәсілдің жан-жақты жұмыс істеу принцптерін ұсынамыз. Біз әртүрлі үлгі конфигурацияларының, оқыту стратегияларының және трансферттік оқыту тәсілдерінің ғадауат тілді анықтау өнімділігіне әсерін бағалау үшін біршама эксперименттер жүргіземіз. Сонымен қатар, қолданыстағы әдістермен салыстырып, оның жоғары нәтижеге жету өнімділігін көрсетеміз, ресурсы төмен қазақ тіліне ғадауат тілді анықтау тапсырмасында заманауи нәтижелерге қол жеткіземіз.

Бұл зерттеу жұмысының зерттеуге қосқан үлесін төмендегідей қорытындылауға болады: (1) Біз қазақ тіліне назар аудара отырып, ресурсы төмен тілдердегі бейәдеп тілді анықтау үшін BiLSTM желілеріне негізделген жаңа тәсілді ұсынамыз. (2) Біз ресурсы жоғары тілдерді пайдалана отырып, алдын ала дайындалған үлгілерді пайдалану және ресурсы төмен параметрлерде ғадауат тілді сөздерді анықтау жұмысын жақсарту үшін трансферлі оқыту әдістерін қолданамыз. Біз ұсынылған әдістің тиімділігін қазақ тіліндегі ғадауат тілді сөздер деректер жинағындағы заманауи нәтижеге жету арқылы көрсетеміз.

Ұсынылып отырған ғылыми зерттеу жұмысы келесідей ұйымдастырылған: II бөлімде ғадауат тілді анықтаудағы тиісті жұмыстарға шолу жасалады және ресурсы төмен тілдерге тән мәселелерді атап өтеді. III бөлімде ұсынылған BiLSTM негізіндегі тәсілді және қолданылатын оқыту әдістерін трансферлі әдістерді сипаттай отырып, әдістеме ұсынылады. IV бөлімде бағалау көрсеткіштері талқыланады. V бөлімде эксперимент нәтижелері берілген. VI бөлімде біздің зерттеуіміздің нәтижелері талқыланады, ресурсы аз тілдердегі ғадауат тілді анықтау туралы түсінік беріледі және осы саладағы болашақ зерттеу бағыттары талқыланады. Соңында VI бөлім қортындылап, ғылыми зерттеу мақаласын аяқтайды.

Әдебиеттік шолу

Желідегі өшпенділік сөзіне және оның жеке адамдар мен қауымдастықтарға ықтимал теріс әсері туралы мәселелердің артуына байланысты соңғы жылдары ғадауат тілді сөздерді анықтауға үлкен көңіл бөлінуде [8]. Бірнеше зерттеулер ең алдымен ағылшын, испан және француз сияқты жақсы ресурсы бар тілдерде тиімді ғадауат тілді анықтау үлгілерін жасауға бағытталған [9].

Дегенмен, ресурсы аз тілдерде қорлау тілін анықтаумен байланысты мәселелер салыстырмалы түрде зерттелмеген. Бұл әдебиетті шолуда біз ғадауат тілді сөздерді анықтау үшін қолданылған бар зерттеулер мен әдістемелерді талқылаймыз, ресурсы төмен тілдерге назар аударамыз. Бұған қоса, біз екі бағытты ұзақ - қысқа мерзімді жад (BiLSTM) желісінің маңыздылығын және оның мұндай тілдердегі бейәдеп тілдерді анықтау мәселесін шешудегі маңыздылығын атап өтеміз. Көптеген зерттеулер ғадауат тілді анықтау үшін машиналық оқыту әдістерін пайдаланды. Wulczyn және басқалар (2017) ағылшынша Wikipedia түсініктемелеріндегі жеке адамдарға сөз арқылы келетін шабуылдарды анықтауға бағытталған Wikipedia Detox жобасын ұсынды[10]. Олар логистикалық регрессия, градиентті күшейту және терең нейрондық желілерді қоса алғанда, перспективалық нәтижелер беретін әртүрлі бақыланатын оқыту алгоритмдерін пайдаланды. Сол сияқты, Djuric және т.б. (2015) әлеуметтік медиа мәтіндеріндегі ғадауат тілді сөздердді анықтау үшін n-grams және синтаксистік үлгілерді қолданатын мүмкіндіктерге негізделген тәсілді зерттеді [11].

Төмен ресурсты тілдерге келетін болсақ, ерікті түрде аннотатор көмегімен жинақталған ғадауат тілді деректер қорының болмауы басты мәселе болып табылады. Бұл ретте бейәдеп тілдер идентификациясын арнайы қарастырған зерттеулер аз. Дегенмен, оқытудың трансферттік әдістері деректер тапшылығы мәселесін жеңілдетуге болатындығы көрсетілген. Мысалы, Fortuna және Nunes (2018) ресурсы төмен Галисия тіліндегі бейәдеп мазмұнды сөздерді анықтау үшін жоғары ресурс тілінен, португал тілінен алдын ала дайындалған кірістірулерді пайдалана отырып, трансферттік оқытуды пайдаланды [10]. Олардың тәсілі дәстүрлі әдістерге қарағанда жақсартылған нәтиже өнімділігін көрсеткен. Ғадауат тілді сөздерді анықтау саласында күрделі тілдік үлгілер мен контекстік тәуелділіктерді түсіру қабілетіне байланысты терең оқыту үлгілері айтарлықтай назар аударды. Бұл салада конволюционды нейрондық желілер (CNN) кеңінен қолданылады. Park және басқалар (2017) CNN-ді ағылшын тіліндегі твиттердегі бейәдеп сөздерін анықтау үшін қолданып, бәсекеге қабілетті нәтижелерге қол жеткізді [12]. Олардың моделі лингвистикалық ақпараттың әртүрлі деңгейлерін түсіру үшін әртүрлі өлшемдегі бірнеше конволюционды сүзгілерді пайдаланды.

Төменде көрсетілген кестеде ғылыми зерттеу жұмыстарына талдау жүргізілген, яғни зерттеу жұмысының тілі, зерттеуге қолданған машиналық әдістер сонымен қатар олардан алынған нәтижелерді бағалау көрсеткіштері мен алынған деректер қорының ашық дерек көздері көрсетілген (Кесте 1).

Бағалау өлшемдері тұрғысынан жиі қолданылатын өлшемдерге accuracy (дәлдік), precision (дәлме - дәлдік), recall (еске түсіру) және F1-score (F1 көрсеткіші) кіреді. Accuracy (дәлдік) модель болжамдарының жалпы дұрыстығын білдіреді, precision (дәлме - дәлдік) ғадауат тілді мазмұнның барлық болжанған сөздерінің арасында дұрыс анықталған бейәдеп сөздердің үлесін өлшейді. Recall (еске түсіру) сезімталдық ретінде белгілі, барлық нақты ғадауат тілді сөздердің ішіндегі дұрыс анықталған бейәдеп тілді сөздердің пайызын білдіреді. F1-score (F1 көрсеткіші) үлгі өнімділігі үшін теңдестірілген балл беру үшін precision пен recall – ді біріктіреді.

Кесте 1. Ғылыми зерттеу жұмыстарына талдау

Ғылыми зерттеу жұмысы	*Тілі*	*Қолданылған әдістер*	*Мәліметтер қоры*	*Бағалау*
Wulczyn (2017) және басқалар	*Ағылшын*	*логистикалық регрессия, градиентті күшейту, Терең нейрондық желілер*	*википедиядан алынған пікірлер*	*Accuracy, Precision, Recall, F1- Score*
Djuric және басқалары (2015)	*Ағылшын*	*Feature-based approach (n грамм, синтаксистік үлгілер)*	*Әлеуметтік желілерден алынған пікірлер*	*Accuracy, Precision, Recall, F1- Score*
Fortuna және Nunes (2018)	*Галисия тілі*	*Transfer Learning, Pre trained Embeddings*	*Әлеуметтік желілерден алынған пікірлер*	*Accuracy, Precision, Recall, F1- Score*
Chen және басқалары (2018)	*Қытай*	*BiLSTM*	*Әлеуметтік желілерден алынған пікірлер*	*Accuracy, Precision, Recall, F1- Score*
Nobata және басқалары (2016)	*Ағылшын*	*Attention based BiLSTM Networks*	*Интернет-қауымдастық*	*Accuracy, Precision, Recall, F1- Score*
Hassan және басқалары (2019)	*Араб тілі*	*Терең оқыту алгоритмдері*	*Әлеуметтік желілерден алынған пікірлер*	*Accuracy, Precision, Recall, F1- Score*
Imran және басқалары (2018)	*Урду тілі*	*Ерекшеліктерге негізделген тәсіл, , SVM*	*Твиттерден алынған пікірлер*	*Accuracy, Precision, Recall, F1- Score*
Choubey және басқалары (2019)	*Хинди*	*Терең оқыту алгоритмдері*	*Әлеуметтік желілерден алынған пікірлер*	*Accuracy, Precision, Recall, F1- Score*
Jha және басқалары (2018)	*Бенгал тілі*	*LSTM, Embeddings сөздер*	*Әлеуметтік желілерден алынған пікірлер*	*Accuracy, Precision, Recall, F1- Score*

Зерттеу әдіснамасы

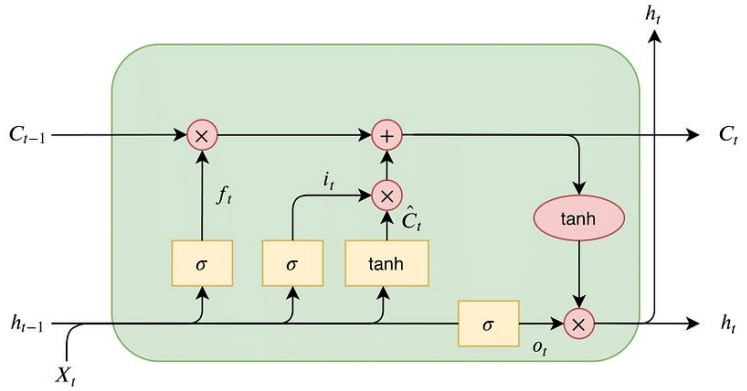
Бұл зерттеу жұмысы мәтіндік деректерде бейәдеп тілді анықтау үшін екі бағытты ұзақ - қысқа мерзімді жад желілерін (BiLSTM) қолдануды зерттейді. Ол онлайн платформалардағы кемсіту мақсатында қолданылған тілдің кең таралған мәселесін қарастырады және терең оқытудың күшін пайдаланатын автоматтандырылған шешімді ұсынады[12]. BiLSTM үлгісі, дәстүрлі LSTM құрылымының кеңейтімі, оның алға және кері бағыттағы уақытша тәуелділіктерді түсіру қабілеті үшін таңдалған, бұл тілдер контекстін түсінуде тиімді болып келеді. Екі бағытты LSTM деп аталатын ретті өңдеу әдісі екі LSTM-ден тұрады, олардың біреуі тікелей бағытта, екіншісі кері бағытта енгізуді қабылдайды. Жалпы, e-BiLSTM ұзақ тәуелді енгізу тізбегі туралы ақпараттан басқа, енгізу мүмкіндіктері мен мақсат арасындағы жасырын қатынасты шығару үшін қолданылады [13]. Ұзақ мерзімді тарихи деректерді сақтау үшін жад ұяшықтарын пайдалану және оны есік механизмімен басқару мұнда қарастырылатын ең маңызды екі аспект болып табылады. Есік құрылымы ешқандай ақпаратты тасымалдамайды, керісінше, ол қол жеткізуге болатын деректер көлемін шектейтін тосқауыл ретінде әрекет етеді. Шын мәнінде, қақпаны басқару механизмін жүзеге асыру көп деңгейлі мүмкіндіктерді таңдау әдісі болып табылады. LSTM – уақыттық қатарлар деректерін талдау және болжау кезінде бірнеше артықшылықтар беретін пайдалы құрал. RNN және LSTM екеуі де өздерінің архитектураларында тізбекті желі модуліне ие. RNN-де модуль бір нейроннан құрастырылған, ал LSTM-де ол әрқайсысында үш қақпасы бар ұяшықтардан құрастырылған. Шығару қақпасы,

кіріс қақпасы және ұмыту қақпасы функцияны таңдау процесінде ұяшық пайдаланатын сурет 1-де көрсетілгендей үш қақпа болып табылады [14].

Сурет 1-де негізінен шығу, кіру және ұмыту қақпаларынан тұратын ұяшықтың құрылымын көрсетеді. Төменде осы үш түрлі қақпа түрлерімен пайдалануға болатын есептеу әдістерінің кейбір мысалдары берілген:

$$input\ (t) = \sigma(W_i x(t) + V_i\, h(t-1) + b_i) \tag{1}$$

Теңдеу (2) ұяшықты ұмыту қақпасы пайдаланатын есептеу механизмінің сипаттамасын береді. Теңдеудегі Wf және Vf қақпаның ұмытылған салмақтары болып табылады және бұл қақпа ұяшықтағы қандай деректерді жою керектігін анықтайды. Басқаша айтқанда, Wf және Vf - ұмытылған қақпа салмақтары.



Сурет 1. BiLSTM желісі

$$forget\ (t) = \sigma(W_f x(t) + V_f\, h(t-1) + b_f) \tag{2}$$

Ұяшықтағы кіріс элементін есептеу процедурасы (1) теңдеумен сипатталады, мұнда h(t-1) - алдыңғы ұяшықтың шығысы, x(t) - ағымдағы ұяшықтың кірісі, σ - сигма тәрізді функцияны білдіреді. Ал, Wi және Vi - кіріс қақпасының салмақтары.

$$\tilde{C}\ (t) = tanh(W_c x(t) + V_c\, h(t-1) + b_c) \tag{3}$$

$$C\ (t) = forget\ (t) * C(t-1) + input(t) * \tilde{C}(t) \tag{4}$$

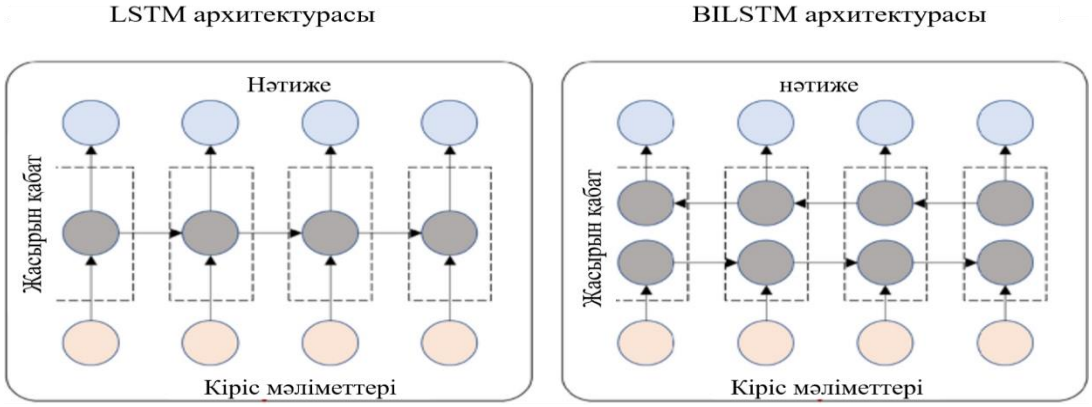
Жаңарту процедуралары (3) және (4) теңдеулер арқылы сипатталған, мұнда (3) баламалы жаңарту деректерін жасайтын үміткер жад блогын білдіреді және (4) ұяшық күйін жаңарту процесін білдіреді. Жаңарту деректері жаңа күйді жасау үшін ұмыту қақпасындағы ақпаратпен біріктіріледі, мұнда Wc және Vc балама жаңа күйдің салмақтарын білдіреді (және * Hadamard өнімін білдіреді).

$$output\ (t) = \sigma(W_0 x(t) + V_0\, h(t-1) + b_0) \tag{5}$$

$$h\ (t) = output\ (t) * \tanh(C(t)) \tag{6}$$

Шығару қақпасын есептеу тәртібі сәйкесінше (5) және (6) теңдеулерімен сипатталған. Бірінші қадамда ұяшықтың шығыс күйінде екенін анықтау үшін сигма тәрізді қабат қолданылады. Екінші қадам жаңартылған ұяшық күйіне *tanh* функциясын қолдануды қамтиды**.** Үшінші және соңғы қадам *h(t)* алу үшін ұяшықтың ағымдағы күйін шығыс нәтижеге (t) көбейтуді қамтиды. *Vo* шығыс қақпасының салмағын білдіреді. Жоғарыда көрсетілген ұяшық LSTM нейрондық желісінің орталығы ретінде қызмет етеді. Бұл топология екі бағытты

LSTM желісін құру үшін негіз ретінде пайдаланылады, содан кейін деректер қасиеттерін шығару үшін пайдаланылады. Дәстүрлі LSTM екі бағытты LSTM-ден өзі шығара алатын контекстік деректердің көлемі жағынан жоғары. Алға және кері уақыт қатарлары желіге уақыт қатарларын дәлірек болжауға мүмкіндік беретін өткен және болашақ уақыт белгілері туралы ақпарат беру үшін пайдаланылады. Алдыңғы және артқы қабаттар арасында тікелей байланыс болмағандықтан, құрылымды ациклді деп сипаттауға болады. Енгізу деңгейінде деректер болған жағдайда, кері және тікелей қабаттардың нәтижелері шығыс деректерін қалыптастыру үшін шығыс деңгейінде біріктіріледі. Әрбір мүмкіндік екі бағытты LSTM арқылы өңделіп, толық қосылған қабат арқылы өткеннен кейін, біріктірілген қабат арқылы барлық мүмкіндіктер біріктіріледі. Сурет 2- де екі бағытты LSTM (BiLSTM) және LSTM нейрондық желісінің негізгі архитектурасын бейнелейді.



Сурет 2. LSTM және BILSTM архитектурасы

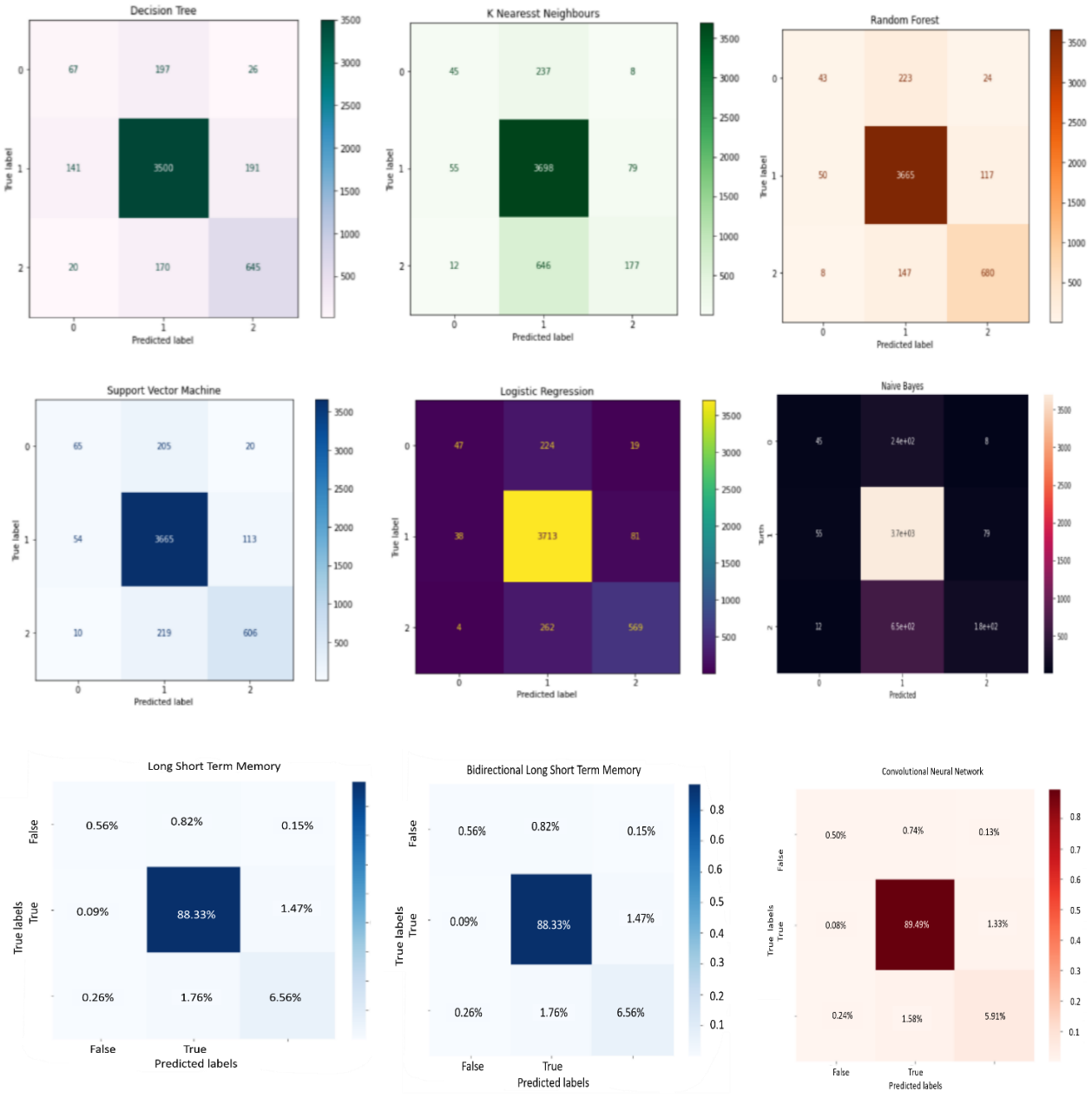
Сурет 2 -де BiLSTM алгоритмі LSTM-тің өзге қабаттарын қалай қосатыны көрсетілген, ол өз кезегінде ақпарат ағынының бағытын өзгертеді. Бұл қарапайым тілмен айтқанда, енгізу тізбегі қосымша LSTM қабатында кері тәртіпте орындалады дегенді білдіреді. Одан кейін екі LSTM қабатының нәтижелері қосу, орташалау, біріктіру және нәтижелерді көбейту сияқты бірнеше түрлі операциялардың көмегімен біріктіріледі. Осыған байланысты желі қол жеткізе алатын ақпарат көлемі артады және алгоритм беретін контекст дәлірек болады [14]. Әдеттегі LSTM-ден айырмашылығы, кірістер екі бағытта да жүре алады және ол кез келген бағыттағы ақпаратты пайдалана алады.

Зерттеу нәтижелері мен талқылаулар

BiLSTM-ді ғадауат тілді сөздерді анықтауға қолдану көптеген салаларда, әсіресе әлеуметтік медиа модерациясында және цифрлық қауымдастықты басқаруда кең ауқымды әсер етеді. Бұл платформалардың алдында тұрған маңызды қиындықтардың бірі кемсітушілік белгілері бар сөздер, ғайбат сөздерді қамтитын желі пайдаланушы пайдаланатын пікірлердің үлкен көлемін басқару болып табылады. Мұндай пікірлер тізбегін қолмен жіктеп бөлу уақытты қажет етеді, BiLSTM негізіндегі модельді енгізу осы модерация процестерінің тиімділігін айтарлықтай арттыра алады, өйткені ол мәтіндерді ғадауат тілді сөздерді автоматты түрде және үздіксіз тексере алады. Бұл мұндай пікірлерді ерте анықтауға және жоюға көмектеседі, осылайша қауіпсіз және инклюзивті онлайн ортаны жасауға мүмкіндік туады. Сонымен қатар, бұл модель пайдаланушы шолулары мен пікірлері жиі тексерілмей қалатын жаңалықтар порталдары және электрондық коммерция веб-сайттары сияқты басқа сандық платформалар үшін де пайдалы болуы мүмкін.

Сурет 3-де дөрекі тіл, позитивті тіл және бейтарап тіл ретінде дөрекі тілді анықтаудың үш класында, яғни ғадауат тілді пікірлер, кемсіту белгілері бар сөз тіркестер және бейтарап пікірлер бойынша машиналық оқытудың әртүрлі әдістерін қолдану арқылы алынған анықтау матрицаларының нәтижелері көрсетілген.

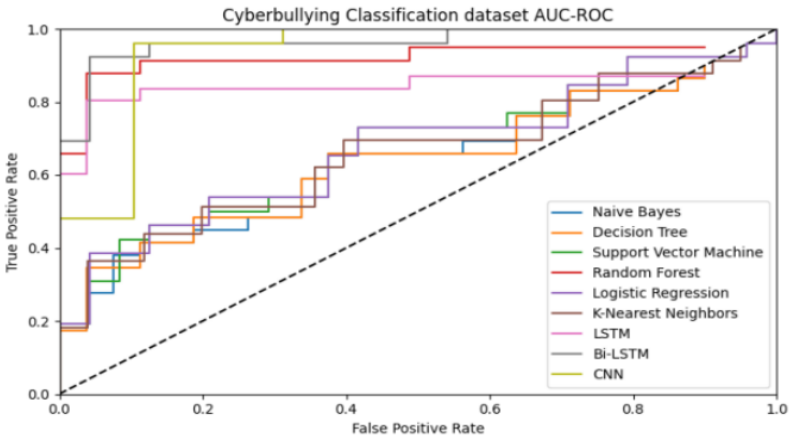
Сурет 4-де әртүрлі машиналық оқыту алгоритмдерінің AUC-ROC қисықтары салыстырылады, оның ішінде екі жақты ұзақ қысқа мерзімді жад желісін қорлайтын тілдің екілік классификациясында. Нәтижелер зерттелген BiLSTM желісі алғашқы оқу дәуірінен жоғары нәтиже беретінін көрсетеді.



Сурет 3. Басқа үлгілерді қолданатын шатастыру матрицасы

BiLSTM моделі ғадауат тілді анықтау мәселесінде бірнеше артықшылықтарды ұсынады. Оның негізгі күші деректердегі күрделі үлгілер мен тәуелділіктерді шығаруға мүмкіндік беретін деректер тізбегін алға және кері өңдеу мүмкіндігіне ие.

Бұл екі бағытты тәсіл модельге тілді қолданудың кең контекстін түсіруге мүмкіндік береді, ол тіл қолдану контекстіне және нәзіктіктеріне өте тәуелді болуы мүмкін қорлайтын тілді дәл анықтау үшін өте маңызды.



Сурет 4. Гайбат сөзді тілді анықтаудағы AUC-ROC қисығы

Қолмен жасалған мүмкіндіктерге сүйенетін дәстүрлі машиналық оқыту үлгілерінен айырмашылығы, BiLSTM деректерден сәйкес мүмкіндіктерді автоматты түрде игере алады (Кесте 2), бұл үлкен көлемді деректерді жобалау қажеттілігін азайтады.

Кесте 2. Алынған нәтижелерді салыстыру

Мәліметтер қоры	*Қолданылған тәсілдер*	*Моделдер*	*Нақтылық*	*Дәлдік*	*Қайта шақыру*	*F-score*	*ROC*
Бейәдеп тілді сөздерден құралған пікірлер	*Машиналық оқыту моделдері*	*SVM*	*0.873*	*0.852*	*0.862*	*0.851*	*0.78*
		KNN	*0.856*	*0.839*	*0.831*	*0.837*	*0.92*
		NB	*0.874*	*0.832*	*0.863*	*0.851*	*0.80*
		DT	*0.602*	*0.524*	*0.585*	*0.642*	*0.65*
		RF	*0.851*	*0.854*	*0.822*	*0.856*	*0.77*
		LR	*0.862*	*0.853*	*0.837*	*0.858*	*0.78*
	Терең оқыту моделдері	*CNN*	*0.892*	*0.895*	*0.898*	*0.896*	*0.93*
		LSTM	*0.901*	*0.896*	*0.91*	*0.898*	*0.93*
		BiLSTM	*0.902*	*0.916*	*0.904*	*0.899*	*0.94*

Ұсынылған BiLSTM моделінің ғадауат тілді сөздерді анықтауға арналған көптеген артықшылықтарына қарамастан, ескеретін біршам шектеулері бар екенін де ескерген дұрыс. Біріншіден, екі бағытты модель ескі және жаңа контексттерді қамтығанымен, модельдің күрделілігі мен есептеулер жеткілікті мөлшерде уақытты бөлуді талап етеді. Бұл жылдамдық маңызды рөл атқаратын қолданбаларда пайдаланғанда қиындықтар тудыруы мүмкін. Екіншіден, модельдің өнімділігі көбінесе оқыту деректерінің сапасына байланысты, яғни, жаттығу деректері ғадауат тілді сөздің әртүрлілігін жеткілікті түрде көрсетпесе, үлгі көрінбейтін деректерді жалпы бір ортаға келтіре алмауы мүмкін. Сонымен қатар, модель шығысы гиперпараметрлерге жіті көңіл аударуды талап етеді, жоғары өнімділікті нәтижеге жету үшін ауқымды түрде реттеуді қажет етеді. Соңында, BiLSTM моделі ұзақ тізбекті мәтіндерді өңдей алатынына қарамастан, ол өзінің бекітілген өлшемді жасырын күйіне байланысты өте ұзақ мәтіндермен жасырын түрде жұмыс жүргізіліп жатуы мүмкін.

Қорытынды

Қорытындылай келе, қиындықтар мен одан әрі жетілдіру мүмкіндіктері сақталғанымен, ұсынылған BiLSTM моделі сандық платформалардағы кең таралған ғадауат тілді сөздерден құралған пікірлер мәселесін шешуде айтарлықтай жетістіктер мен нәтижелер көрсетті. Ол дәстүрлі әдістермен тиімді шешуге болмайтын мәселелерге автоматтандырылған шешімдерді

ұсыну арқылы адам тілінің күрделілігін түсінуде терең оқыту әдістерінің тиімділігін көрсетті. Бұл ғылыми зерттеу жұмысы қауіпсіз және инклюзивті цифрлық коммуникация платформаларын жасау үшін жасанды интеллект күшін пайдалану жолындағы маңызды қадамды білдіреді. Бұл саладағы болашақ жетістіктер тек сенімді және тиімді үлгілерді жасап қана қоймайды, сонымен қатар цифрлық медиада тілді түсіну және модельдеу туралы жаңа түсініктер береді деп күтілуде.

Пайдаланылған әдебиеттер тізімі:

1 Omarov B., Suliman A., Tsoy A. Parallel backpropagation neural network training for face recognition // Far East Journal of Electronics and Communications. – 2016. – Т. 16. – №. 4. – С. 801-808.

2 Toktarova A. et al. Hate Speech Detection in Social Networks using Machine Learning and Deep Learning Methods //International Journal of Advanced Computer Science and Applications. – 2023. – Т. 14. – №. 5.

3 Govers J. et al. Down the Rabbit Hole: Detecting Online Extremism, Radicalisation, and Politicised Hate Speech //ACM Computing Surveys. – 2023.

4 Govers J. et al. Down the Rabbit Hole: Detecting Online Extremism, Radicalisation, and Politicised Hate Speech //ACM Computing Surveys. – 2023.

5 Ali M. et al. Social media content classification and community detection using deep learning and graph analytics //Technological Forecasting and Social Change. – 2023. – Т. 188. – С. 122252.

6 Husain F., Uzuner O. A survey of offensive language detection for the Arabic language //ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP). – 2021. – Т. 20. – №. 1. – С. 1-44.

7 Babu N. V., Kanaga E. G. M. Sentiment analysis in social media data for depression detection using artificial intelligence: a review //SN Computer Science. – 2022. – Т. 3. – С. 1-20.

8 Asghar M. Z. et al. Exploring deep neural networks for rumor detection //Journal of Ambient Intelligence and Humanized Computing. – 2021. – Т. 12. – С. 4315-4333.

9 Ullah F. et al. IDS-INT: Intrusion detection system using transformer-based transfer learning for imbalanced network traffic //Digital Communications and Networks. – 2023.

10 Azzi S. A., Zribi C. B. O. From machine learning to deep learning for detecting abusive messages in arabic social media: survey and challenges //International Conference on Intelligent Systems Design and Applications. – Cham : Springer International Publishing, 2020. – С. 411-424.

11 Ghosal S., Jain A. HateCircle and Unsupervised Hate Speech Detection Incorporating Emotion and Contextual Semantics //ACM Transactions on Asian and Low-Resource Language Information Processing. – 2023. – Т. 22. – №. 4. – С. 1-28.

12 Yadav D. et al. Age group prediction on textual data using sentiment analysis //Proceedings of the 9th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion. – 2020. – С. 61-65.

13 Machová K., Mach M., Porezaný M. Deep Learning in the Detection of Disinformation about COVID-19 in Online Space //Sensors. – 2022. – Т. 22. – №. 23. – С. 9319.

14 Singh J. P. et al. Attention-based LSTM network for rumor veracity estimation of tweets //Information Systems Frontiers. – 2020. – С. 1-16.