МРНТИ 81.93.29 **https://doi.org/10.58805/kazutb.v.4.29-1243**

МӘТІНДІК ДЕРЕКТЕРДЕ КРИМИНАЛИСТИКАЛЫҚ ТАЛДАУ ЖҮРГІЗУГЕ АРНАЛҒАН ТЕРЕҢ НЕЙРОНДЫҚ ЖЕЛІ МОДЕЛІН ҚҰРУ

[1]А.А. Құттыбек, [1]Н.М. Казиева, [2]А.С. Амирова, [3]Д.А. Рюмин

[1]Л.Н. Гумилев атындағы Еуразия Ұлттық Университеті, Астана, Қазақстан,
[2]Astana IT University, Астана, Қазақстан,
[3]Ресей Ғылыми Академиясы, Федералды ғылыми - зерттеу орталығы, Санкт-Петербург, Ресей

Корреспондент-автор: azhar.kuttybek@astanait.edu.kz

Осы мақалада криминалистика аясында мәтіндік деректерді интеллектуалды талдау мен автоматты зерттеу үшін жасанды интеллект және терең нейрондық желілерді пайдалану мәселесі қарастырылды. Бүгінгі таңда ғаламтордағы деректердің шамадан тыс көбеюі мен әлеуметтік желілерде пайда болатын ақпараттардың тез таралуы құқық қорғау органдары үшін көптеген міндеттер тудырып отыр. Сот сараптамалары мен тергеулер, жұмыс барысында дәлелдер құрамында мәтіндік материалдар көп кездеседі, сол себепті деректерді автоматты өңдеу және сот сараптамасы ретінде бағалау, өзекті мәселелердің бірі болып саналады.

Ғылыми мақаланың мақсаты - мәтіндік деректерден агрессивті және құқық бұзушылық белгілерін, экстремистік негізіндегі мазмұндарды анықтай алатын терең нейрондық желі моделін құру. Мақалада Attention және LSTM - (Ұзақ мерзімді жад) архитектурасының негізінде құрылған үлгі ұсынылады. Осы тәсіл қазақ тіліндегі мәтіндердің семантикалық байланысын, контекстік байланыстарын және эмоционалды реңкін тиімді сараптауға мүмкіндік береді. Ақпараттарды өңдеу барысында мәтіндер токенизацияланып, артық белгілердің тазартылады. Сонымен қатар, TF-IDF Word2Vec әдістерінің көмегімен векторлық түрлендіру жүргізіледі. Осындай көрсеткіш классикалық әдістерден жоғары екенін көрсетеді.

Криминалистикалық сараптама барысында, ұсынылған әдістеме деректерді тез өңдеуге және дәл өңдеуге мүмкіндік береді. Қылмыстық істерді тергеу процессін автоматтандыруға, сот сарапшыларының жұмыстарын оңайлатуға, сондай-ақ адам факторына тәуелділікті азайтуға бағытталған, маңызды қадамдардың бірі болып табылады. Осы мақала нәтижелері болашақта жасанды интеллектуалды сараптау системасын дамытуға, қазақ тіліндегі криминалистикалық ақпараттар қорын ұлғайтуға негіз болады.

Түйін сөздер: мәтіндік деректер, криминалистика, жасанды интеллект, терең оқыту, LSTM, Attention, лингивистикалық сараптама.

РАЗРАБОТКА МОДЕЛИ ГЛУБОКОЙ НЕЙРОННОЙ СЕТИ ДЛЯ ПРОВЕДЕНИЯ КРИМИНАЛИСТИЧЕСКОГО АНАЛИЗА ТЕКСТОВЫХ ДАННЫХ

[1]А.А. Құттыбек, [1]Н.М. Казиева, [2]А.С. Амирова, [3]Д.А. Рюмин

[1]Евразийский национальный университет им. Л. Н. Гумилева, Астана, Казахстан,
[2]Astana IT University, Астана, Казахстан,
[3]Российская академия наук, Федеральный научно-исследовательский центр, Санкт-Петербург, Россия,
e-mail: azhar.kuttybek@astanait.edu.kz

В данной статье рассматривается применение искусственного интеллекта и глубоких нейронных сетей для интеллектуального анализа и автоматического изучения текстовых данных в области криминалистики. Сегодня чрезмерный рост объёма данных в интернете и быстрое распространение информации в социальных сетях создают множество проблем для правоохранительных органов. В криминалистических экспертизах и расследованиях доказательствами зачастую являются текстовые материалы, поэтому автоматическая обработка и оценка данных в рамках криминалистического анализа являются одной из наиболее актуальных задач.

Цель научной статьи – создание модели глубокой нейронной сети, способной выявлять агрессивные и криминальные признаки, а также экстремистский контент в текстовых данных. В статье представлена модель, основанная на архитектурах Attention и LSTM. Данный подход позволяет эффективно анализировать семантические связи, контекстные связи и эмоциональную окраску текстов на казахском языке. В процессе обработки информации тексты тонизируются и очищаются от избыточных признаков. Кроме того, выполняется векторное преобразование с использованием методов TF-IDF Word2Vec. Этот показатель свидетельствует о её превосходстве над классическими методами.

В области криминалистического анализа предлагаемая методология позволяет осуществлять быструю и точную обработку данных. Это один из важных шагов, направленных на автоматизацию процесса расследования уголовных дел, упрощение работы экспертов-криминалистов и снижение зависимости от человеческого фактора. Результаты данной статьи послужат основой для дальнейшего развития систем анализа на основе искусственного интеллекта и увеличения объема криминалистической информации на казахском языке.

Ключевые слова: текстовые данные, криминалистика, искусственный интеллект, глубокое обучение, LSTM, внимание, лингвистический анализ.

DEVELOPMENT OF A DEEP NEURAL NETWORK MODEL FOR FORENSIC ANALYSIS ON TEXTUAL DATA

[1]A.A. Kuttybek, [1]N.M. Kaziyeva, [2]A.S. Amirova, [3]D.A. Ryumin

[1]L.N. Gumilyov Eurasian National University, Astana, Kazakhstan,
[2]Astana IT University, Astana, Kazakhstan,
[3]Russian Academy of Sciences, Federal Research Center, St. Petersburg, Russia,
e-mail: azhar.kuttybek@astanait.edu.kz

This article considers the use of artificial intelligence and deep neural networks for intelligent analysis and automatic study of text data in the field of forensics. Today, the excessive growth of data on the Internet and the rapid spread of information appearing in social networks pose many challenges for law enforcement agencies. In forensic examinations and investigations, evidence often includes text materials, so automatic data processing and evaluation as a forensic analysis are one of the most pressing issues.

The purpose of the scientific article is to create a deep neural network model that can identify aggressive and criminal signs, extremist content from text data. The article presents a model based on the Attention and LSTM architectures. This approach allows for effective analysis of semantic connections, contextual connections and emotional tone of texts in the Kazakh language. During information processing, texts are tokenized and cleared of redundant features. In addition, vector transformation is performed using the TF-IDF and Word2Vec methods. This indicator indicates that it is superior to classical methods.

In the field of forensic analysis, the proposed methodology allows for fast and accurate data processing. It is one of the important steps aimed at automating the process of investigating criminal cases, simplifying the work of forensic experts, and reducing dependence on the human factor. The results of this article will serve as a basis for the future development of artificial intelligence analysis systems and increasing the stock of forensic information in the Kazakh language.

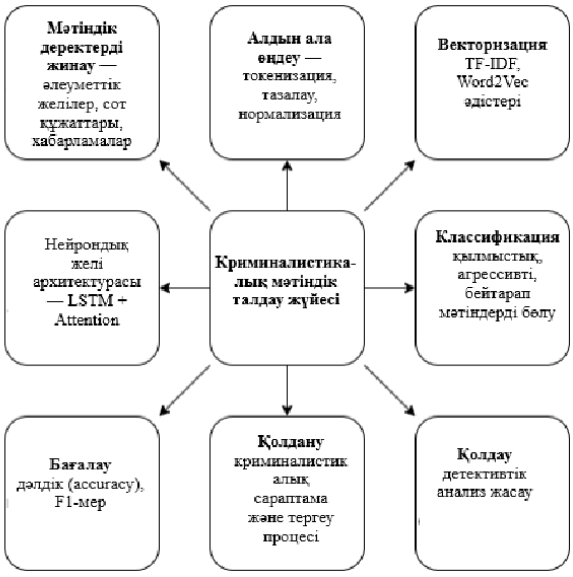
Keywords: text data, forensics, artificial intelligence, deep learning, LSTM, Attention, linguistic analysis.

Кіріспе. Криминалистикалық мәтіндік талдау - құқықтық сипаттағы деректерді интеллектуалды өңдеу мен жасырын заң бұзушылық белгілерін анықтауға бағытталған зерттеу саласы. Осындай заң бұзушылық мәтіндегі материалдарға: тергеу хаттамалары, әлеуметтік желілер-

ге жазған жазбалары мен куәгердің хабарламаларын жатқызуға болады. Терең нейрондық желілер арқылы мәтіндік деректерге талдау жүйелері, деректердің мағынасын семантикалық негізде түсініп, агрессивті, қорқыту немесе экстремисстік сипаттағы сөз тіркестерін автоматты түрде анықтай алады.

Криминалистикалық тұрғыдан қарағанда, жасырын қылмыстық ниетті мәтіндерді, қорқыту белгісінің болуы мен жалған ақпараттарды анықтау. Мысалы, кез келген мәтінде пайдаланылған лексикалық құрылым мен синтаксистік байланыстар, заң бұзушылық әрекеттің басталатынын және қылмыстық сипаты бар іс-әрекетке алып келетінін көрсетуі мүмкін. Осындай жағдайларда, терең оқыту үлгісі, дәлірек айтқанда LSTM және Attention механизмдері, мәтінді анализ жасау арқылы, қауіпті мәтінді дәл анықтай алады [1].

Осындай әдістердің тиімділігі олардың ауқымды көлемдегі мәтіндерден үйрену қабілетінде. Мәтіндерді талдай отырып, жүйе ескі деректерден алынған, жаңа хабарламалардағы ұқсас белгілерді автоматты түрде анықтайды. Криминалистикалық сарапшыларға осы жүйе, уақыт үнемдеуге, сонымен қатар материалдарды объективті бағалауға мүмкіндік береді [2]. Терең нейрондық желілерге негізделген үлгі, мәтіндік деректерді интелектуалды сараптаудың ең тиімді құралы болып табылады, сонымен қатар, заманауи криминалистикада қылмыстық сипаттағы мінез – құлықтардың белгілерін автоматты анықтауда маңызды рөл атқарады [3-5].



1-сурет. Криминалистикалық мәтінді талдау диаграммасы

Контексттік деректерді қылмыстық типте анализ жасау барысында, ***1-суретте*** ең маңызды параметрлердің бірі - ғаламтор кеңістігінде заң бұзушылық сипатындағы хабарламаларды анықтау, сонымен қатар мәтіндердің маңыздылығын анықтау болып табылады. Осындай деректерге жалған ақпараттарды тарату, абыройға нұқсан келтіру, бопсалау, қорқыту немесе жалған айыппұл салу сынды ақпараттар жатады [6].

Көптеген ситуацияларда криминалистикалық сипаттағы мәтіндер ашық интернет көздерінде, дәлірек айтқанда әлеуметтік желілер мен форумдарда буркемеленген түрде таралады. Осындай контентті анықтау үшін терең оқыту модельдері -LSTM және Attention архитектуралары - деректердің семантикалық құрылымын анализ жасап, мәтіндік байланыстар арқылы күмәнді деректерді таба алады.

Заң бұзушылық немесе құқық бұзушылық белгісі бар деректерде жиі табылатын ерекшеліктер - агрессивті тілді құрылымдар, бопсалау мәніндегі тіркестер, субъективті лексика және эмоция болып табылады. Осы белгілерді динамикалық түрде анықтау үшін модель (үлгі) алдын-ала үйретілген көп көлемдегі мәтіндер жиынтындағы (corpus) мысалдарды пайдаланады.

Сонымен қатар, қылмыстық талдауда маңызды назар аудартатын аспектілердің бірі - өз атын жасыратын авторлардың (анонимді) жазбасы және жалған аккаунттар арқылы жарияланған мәтіндерді атап айтуға болады.

Терең нейрондық желілер мұндай контенттің лингвистикалық маңыздылығын сараптау арқылы авторлық стильді талдауға және мәтіннің түп нұсқалығын бағалауға мүмкіндік береді [7-8].

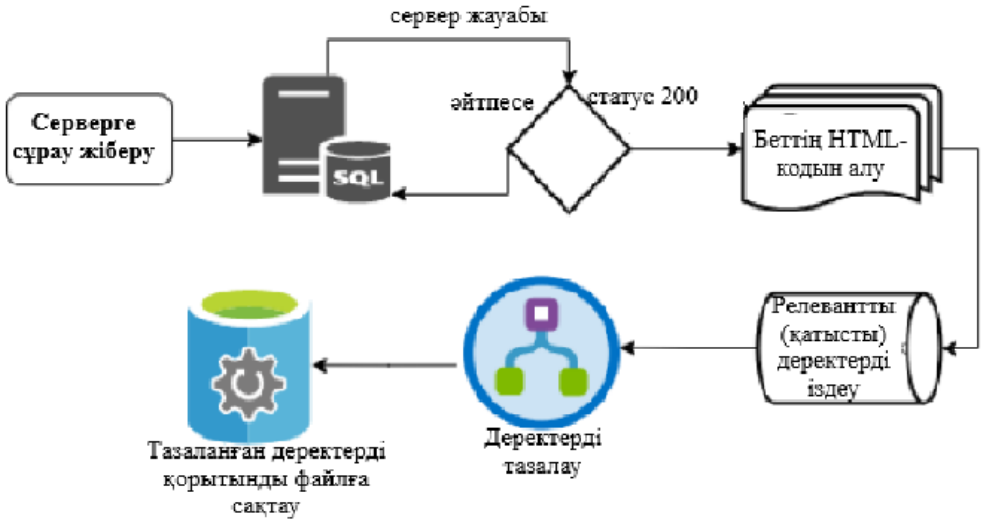
Жоғарыдағы көрсетілген диаграмма арқылы құқық қорғау органдыраның мүшелері ғаламтордағы ақпараттардың арасынан нақты қауіп төндіретін немесе заң бұзушылыққа итермелейтін контентті қысқа мерзім уақыт аралығында анықтай алады. Осындай жүйе дәстүрлі талдау әдістерімен салыстырғанда тез, әрі нақты және үлкен көлемдегі деректерді қамти алады.

Материалдар мен әдістер. Контекстік деректердің үлкен және шексіз көлемі және оның динамикалық сипаты дәстүрлі талдау әдістерін тиімсіз етеді. Сондықтан терең оқыту және табиғи тілді өңдеу (NLP) әдістеріне негізделген автоматтандырылған жүйелерді пайдалану барған сайын маңызды болып келеді [9].

Терең оқыту желілері мәтін мазмұнын талдау арқылы жасырын агрессивті, шантаж немесе экстремистік мазмұны бар сөз тіркестерін анықтауға мүмкіндік береді. Бұл тәсіл адамның эмоционалдық жағдайын немесе ықтимал қауіпті ниеттерін автоматты түрде анықтауға бағытталған. Бұл бөлімде криминалистикалық мәтін деректерін өңдеудің негізгі кезеңдері талқыланады: деректерді жинау, өңдеу және ерекшеліктерді анықтау **(2-сурет).**

Деректерді жинау. Жоғары сапалы және теңгерімді деректер жиынтығын жасау криминалистикалық мәтінді талдауда маңызды рөл атқарады. Бұл деректер жиынтығы сот-медициналық сараптама материалдарынан, әлеуметтік желілерден, хабарламалардан және ашық интернет көздерінен жиналады [10-12].

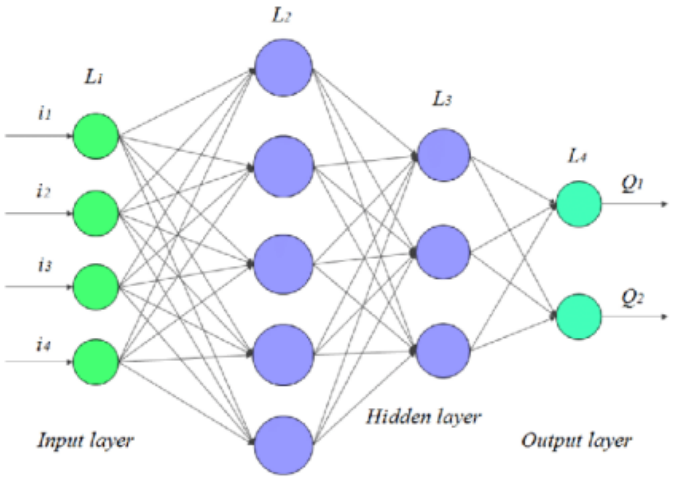
Жиналған деректер алдын ала сүзіледі және тек криминалистикалық маңызы бар жазбалар таңдалады. Мысалы, қылмысты ынталандыратын сөз тіркестері, қауіпті мағынасы бар сөздер, жалған ақпарат немесе агрессивті пікірлер. Сонымен қатар, мәтінді топтастыру үшін TF-IDF, Word2Vec немесе FastText әдістері қолданылады.



2-сурет. Мәтіндік деректерді жинау және алдын ала өңдеу сызбасы

Бұл әдістер мәтін мазмұнын сандық векторларға түрлендіруге және оларды нейрондық желі моделіне беруге мүмкіндік береді. Деректерді жинау кезеңінің сапасы модельдің жалпы дәлдігіне тікелей әсер етеді [13].

Құқық бұзушылық деректерде жиі кездесетін ерекшелік - агрессия және бопсалау, қорқыту немесе манипуляция сипатындағы сөз тіркестері мен контекстік мағыналар. Терең нейрондық желілер (Deep Neural Networks - DNN) осындай мәтіндерді көп деңгейлі өңдеу арқылы анализ жасап, мағыналық ұқсастығын немесе қауіптілігін бағалай алады. Терең нейрондық желілер – бұл бірнеше қабаттан тұратын машиналық оқыту алгоритмдері, олар күрделі және абстрактілі заңдылықтарды анықтау үшін мәліметтерді көпдеңгейлі өңдеуге мүмкіндік береді [14-15]. Көп деңгейілі желілер мәтіндік деректердің семантикалық және синтаксистік құрылымдарын үйренуге және қылмыстық сипаттағы жасырын байланыстарды іздеуге қабілетті. Қарапайым терең нейрондық желі архитектурасы үш негізгі қабаттан тұрады (***3-сурет***): кіріс қабаты (Input Layer), жасырын қабаттар (Hidden Layers) және шығыс қабаты (Output Layer). Кіріс қабатында дерек вектор түрінде беріледі, жасырын қабаттарда белсендіру функциялары (ReLU, tanh) арқылы ақпарат түрленеді, ал шығыс қабаты нәтиже ретінде мәтіннің криминалистикалық категориясына мысалы, “агрессивті”, “бейтарап”, “қорқыту” сынды сөз тіркестерін кіріктіруге болады.



3-сурет. Қарапайым терең нейрондық желі архитектурасының құрылымы

Мәтіндік деректерді жіктеу үшін LSTM нейрондық желілерін пайдалану: LSTM - тізбектерді танитын және уақытқа тәуелді деректерді, яғни сөйлемдер мен мәтіндер арасындағы логикалық қатынастарды сақтауға арналған қайталанатын нейрондық желінің (RNN) бір түрі [16].

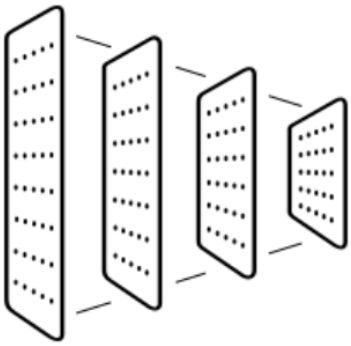
LSTM моделінің басты артықшылығы - ол контекстті ұзақ уақыт бойы есте сақтай алады және кейінгі талдауда алдыңғы сөздердің мағынасын пайдалана алады. Бұл қасиет криминалистикалық мәтінді талдауда әсіресе маңызды, себебі қылмыс белгілері көбінесе мәтіннің жалпы мағынасына байланысты.

Мысалды қолданыңыз: егер мәтінде «Мен сені табамын, сен өкінесің» сияқты сөз тіркестері болса, LSTM желісі мұндай тізбекті «қорқыту» деп жіктей алады. Бұл модель тек жеке сөздерді ғана емес, сонымен қатар олардың арасындағы уақытша қатынастарды да талдайды.

Модельдің соңғы қабаты мәтінді бірнеше класқа бөлетін softmax функциясын пайдаланады: бейтарап, агрессивті, қылмыстық және ұқсас.

4-суретте көрсетілгендей, әрбір келесі қабаттағы нейрондар саны біртіндеп азаяды, ал ақпа-

рат жоғары деңгейлі семантикалық ерекшеліктерге айналады. Бұл процесс модельге күрделі тілдік құрылымдарды тануға және жасырын мағыналарды анықтауға мүмкіндік береді. Бұл тәсіл криминалистикалық мәтін талдауын дәстүрлі әдістерге қарағанда дәлірек және тиімдірек жүргізуге мүмкіндік береді.



4-сурет. LSTM негізіндегі мәтін классификациялау желісінің визуалды құрылымы

Осы мақалада ұсынылған терең нейрондық желі (DNN) қылмыстық процесстерде мәтін талдауы үшін қолданылады. Бұл желінің көп қабатты архитектурасынан тұрады, әрбір қабат белгілі бір деңгейде семантикалық ақпаратты өңдеуге арналған.

Модельдің құрылымы келесі принципке негізделген: әрбір келесі қабаттағы нейрондар саны алдыңғысына қарағанда аз, яғни желі иерархиялық деңгейлерді пайдаланып маңызды деректер ерекшеліктерін алады. Бұл тәсіл артық ақпаратты азайтады және тиімді оқытуды қамтамасыз етеді.

Бірінші қабат вектормен түрлендірілген мәтінмен беріледі. Мысалы, әрбір сөйлем Word2Vec немесе TF-IDF әдістерін қолдана отырып, 784 өлшемді вектор ретінде ұсынылады. Бұл векторлар кіріс қабатына беріледі, ал кейінгі қабаттар - тығыз қабаттар - деректер арасындағы күрделі қатынастарды үйренеді.

Нәтижелер және талқылау. Модель келесідей жұмыс істейді: бірінші қабат негізгі ерекшеліктерді (лексикалық және синтаксистік құрылымдарды) бөліп алады. Ортаңғы қабаттар семантикалық және контекстік тәуелділіктерді өңдейді. Соңғы қабат мәтінді белгілі бір санатқа жіктейді: агрессивті, қауіпті немесе бейтарап [16-17]. Бұл архитектураны пайдалану есептеу шығындарын азайтады, оқыту жылдамдығын арттырады және модельдің дәлдігін тұрақтандырады. Сонымен қатар, төменгі қабаттар қарапайым ерекшеліктерді үйренсе, жоғарғы қабаттар оларды біріктіріп, күрделі семантикалық заңдылықтарды анықтайды.

Базаға енгізілетін мәтіннің математикалық сипаттамасы төмендегідей формуламен есептеліп төмендегі жүйеге енгізілетін болады:

$$Z = w_1 x_1 + w_2 x_2 + \ldots + w_n x_n \qquad (1)$$

Модельдің бір қабатындағы есептеу формуласы:

мұндағы:

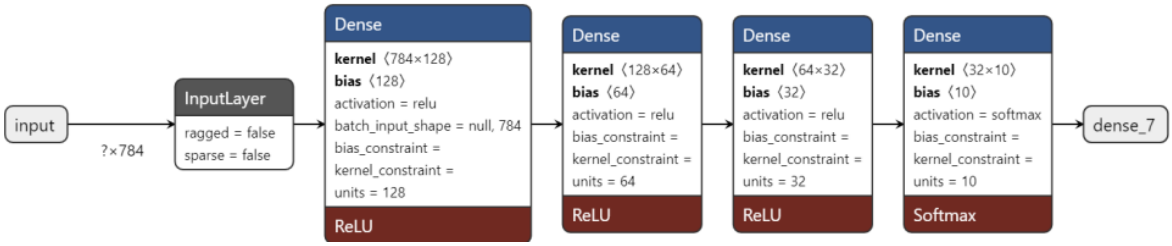
Z - нәтижелік вектор,

w - кіріс мәтіндерінің жиыны,

x - енгізілетін деректер.

5-суретте ендірілген нейрондық желінің графикалық көрінісі көрсетілген. Кіріс қабаты (Input Layer) input_length = 784 өлшемді векторлар жиынтығын жасайды және оны бірінші толық қосылған қабатқа (Dense Layer) береді. Әрбір келесі қабатта нейрондар саны азаяды, бұл желіге маңызды ерекшеліктерді алуға және шығысты соңғы қабатқа беруге мүмкіндік береді. Соңғы қабат бір нейроннан тұрады және softmax функциясын пайдаланып болжам жасайды: егер нейрон белсенді болса, мәтін бірінші санатқа жа-

тады (мысалы, қылмыстық мәтін), ал егер ол белсенді емес болса, мәтін екінші санатқа жатады (бейтарап мәтін).



5-сурет. Қарапайым терең нейрондық желінің визуалды моделі

Бұл құрылым терең оқытудың тиімділігін көрсетеді, себебі ол мәтіндік деректердегі күрделі семантикалық қатынастарды зерттеуге және сотмедициналық талдауда дәл жіктеуге мүмкіндік береді.

Қорытынды. Бұл мақала криминалистикалық контексте интеллектуалды мәтінді талдау үшін терең нейрондық желі моделін жасаудың ғылыми және практикалық негізін жасады. Мақсатқа - мәтін мазмұнындағы жасырын қылмыстық ниетті, қорқыту белгілерін және агрессияны автоматты түрде анықтауға - сәтті қол жеткізілді.

Жұмыстың теориялық негізі терең оқыту және табиғи тілді өңдеу (NLP) әдістерімен қалыптасты. LSTM және Attention архитектураларының үйлесімі жүйеге мәтіннің семантикалық қатынастарын, контекстік мағынасын және эмоционалды реңкін тиімді талдауға мүмкіндік берді. Бұл тәсіл дәстүрлі статистикалық модельдермен салыстырғанда дәлдікті арттырды және жалған немесе бейтарап мәтіндерді қылмыстық мазмұннан сенімді түрде ажыратты.

Зерттеу барысында алынған нәтижелер нейрондық желі архитектурасының (Input–Hidden-Output құрылымы) криминалистикалық мәтіндерді автоматты түрде жіктеуде жоғары өнімділік көрсетті. Word2Vec және TF-IDF көмегімен мәтінді векторлау жүйені оқыту процесін үйренуге және контекстті тереңірек түсінуге мүмкіндік береді. Модельді тестілеу нәтижелері ұсынылған жүйе мәтіндерді үш негізгі санатқа - агрессивті, қауіп төндіретін және бейтарап – мазмұндары дәлдікпен жіктей алатынын көрсетеді. Бұл дәстүрлі әдістерден (логистикалық регрессия, SVM) айтарлықтай асып түседі.

Практикалық тұрғыдан алғанда, әзірленген модель криминалистикалық талдау мен құқық қорғау саласында тамаша көмекші бола алады. Бұл әлеуметтік желілердегі және ашық интернет көздеріндегі қауіпті мазмұнды уақтылы анықтауға, сондай-ақ криминалистикалық сараптамаларға дәлелдемелерді дайындауды автоматтандыруға мүмкіндік береді.

Жалпы алғанда, жұмыс криминалистикалық ғылымдағы терең нейрондық желілердің әлеуетін айқын көрсетті. Ұсынылған әдіс ақпараттық қауіпсіздікті, қылмыстың алдын алуды және цифрлық дәлелдемелерді талдау сапасын жақсартуға айтарлықтай үлес қоса алады.

Литература

1. F. Bozyiğit, O. Doğan and D. Kilinç. Categorization of customer complaints in food industry using machine learning approaches// Journal of Intelligent Systems Theory and Applications. -2022.-Vol.5(1). - P.85-91. DOI 10.38016/jista.954098.

2. B. Omarov, S. Narynov, Z. Zhumanov, A. Gumar and M. Khassanova. A skeleton-based approach for campus violence detection// Computers, Materials & Continua. - 2022.-Vol.72(1). - P.315-331. DOI 10.32604/cmc.2022.024566.

3. Бектұров Ш.К. Қазақ тілі: лексика, фонетика,морфология, синтаксис.-Алматы: Атамұра, 2006-336. ISBN 9965-34-371-3.

4. A. Fedorenko, J. Tussupov, M. Sametbayeva, I. Idrissova, and A. Yerimbetova. Development and implementation of a morphological model of Kazakh language// Eurasian Journal of Mathematical and Computer Applications. - 2015.- Vol.3(3). - P.69-79.

5. M. Borodin, E. Lisin and W. Strielkowski. Data algorithms for processing and analysis of unstructured text documents// Applied mathematical sciences. - 2014. - Vol.8(25). - P.1213–1222. DOI 10.12988/ams.2014.4125.

6. M. Mowafy, A. Rezk, and H. El-Bakry. An efficient classification model for unstructured text document// American Journal of Computer Science and Information Technology. - 2018.- Vol.6(1):16. DOI 10.21767/2349-3917.100016.

7. Аксенов А.А., Рюмина Е.В., Рюмин Д.А. Метод генерации анимации цифрового аваиара с речевой и невербальной синхронизацией на основе бимодальных данных// Научно-технический вестник Информационных технологий, механики и оптики. -2025-Т.25(4). -С.651-662. DOI 10.17586/2226-1494-2025-25-4-651-662

8. Иванько Д.В., Рюмин Д.А. Автоматический сурдоперевод: обзор нейросеиевых методов распознавания и синтеза звучащей и жестовой речи// Научно-технический вестник Информационных технологий, механики и оптики. - 2025- Т.24(5)-С.669-686. DOI 10.17586/2226-1494-2024-24-5-669-686

9. Kaziyeva N., Madiev A., Aitzhanov S., Kaliyev A., Kuttybek A. Animated biometric QR-codes as an innovative solution in information systems//2025 IEEE 5th International Conference on Smart Information Systems and Technologies (SIST). -2025. DOI 10.1109/SIST61657.2025.11139180.

10. J. L. Elman. Finding structure in time// Cognitive Science. -1990. - Vol.14(2). - P.179–211. DOI 10.1207/s15516709cog1402_1.

11. H. Schmid. Probabilistic part-of-speech tagging using decision trees// Proceedings of International Conference on New Methods in Language Processing, Manchester, UK. - 1994. - P.44-49. URL: https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/tree-tagger1.pdf. (accessed 07.11.2025).

12. T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.- 2013. DOI 10.48550/arXiv.1301.3781.

13. J. Pennington, R. Socher, and C. D. Manning. GloVe: Global vectors for word representation// Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). -2014.- P.1532–1543. DOI 10.3115/v1/D14-1162.

14. C. Bishop, Pattern Recognition and Machine Learning. Berlin, Heidelberg: Springer, 2006.-737 s. - P.137–145. ISBN-10: 0-387-31073-8. ISBN-13: 978-0387-31073-2/

15. Sammut, Claude et al. Encyclopedia of Machine Learning and Data Mining. - Springer-2017. DOI 10.1007/978-1-4899-7687-1, ISBN978-1-4899-7685-7

16. D. Hosmer and S. Lemeshow, Applied Logistic Regression: A Self-Learning Text//New York, NY: Springer-Verlag. - 528 s.2013. ISBN 9780470582473

17. T. Hastie, R. Tibshirani, and J. Friedman. The Elements of Statistical Learning. -2017.-Springer. DOI 10.1007/978-0-387-21606-5.

References

1. F. Bozyiğit, O. Doğan and D. Kilinç. Categorization of customer complaints in food industry using

machine learning approaches// Journal of Intelligent Systems Theory and Applications. -2022.-Vol.5(1). - P.85-91. DOI 10.38016/jista.954098.

2. B. Omarov, S. Narynov, Z. Zhumanov, A. Gumar and M. Khassanova. A skeleton-based approach for campus violence detection// Computers, Materials & Continua. - 2022.-Vol.72(1). - P.315-331. DOI 10.32604/cmc.2022.024566.

3. Bektүrov Sh.K. Қазақ tili: leksika, fonetika,morfologija, sintaksis.-Almaty: Atamүra, 2006-336. ISBN 9965-34-371-3. [in Kazakh]

4. Fedorenko A., Tussupov J., Sametbayeva M., Idrissova I., Yerimbetova A.. Development and implementation of a morphological model of Kazakh language// Eurasian Journal of Mathematical and Computer Applications. - 2015. - Vol.3(3). - P.69-79.

5. Borodin M., Lisin E. and W. Strielkowski. Data algorithms for processing and analysis of unstructured text documents// Applied mathematical sciences. - 2014. - Vol.8(25). - P.1213–1222. DOI 10.12988/ams.2014.4125.

6. Mowafy M., Rezk A., and H. El-Bakry. An efficient classification model for unstructured text document// American Journal of Computer Science and Information Technology. - 2018.- Vol.6(1):16. DOI 10.21767/2349-3917.100016.

7. Aksenov A.A., Rjumina E.V., Rjumin D.A. Metod generacii animacii cifrovogo avaiara s rechevoj i neverbal’ noj sinhronizaciej na osnove bimodal’ nyh dannyh// Nauchno-tehnicheskij vestnik Informacionnyh tehnologij, mehaniki i optiki. -2025.-T.25(4). -S.651-662. DOI 10.17586/2226-1494-2025-25-4-651-662. [in Russian]

8. Ivan’ ko D.V., Rjumin D.A. Avtomaticheskij surdoperevod: obzor nejroseievyh metodov raspoznavanija i sinteza zvuchashhej i zhestovoj rechi// Nauchno-tehnicheskij vestnik Informacionnyh tehnologij, mehaniki i optiki. - 2025.- T.24(5).- S.669-686. DOI 10.17586/2226-1494-2024-24-5-669-686. [in Russian]

9. Kaziyeva N., Madiev A., Aitzhanov S., Kaliyev A., Kuttybek A. Animated biometric QR-codes as an innovative solution in information systems//2025 IEEE 5th International Conference on Smart Information Systems and Technologies (SIST). -2025. DOI 10.1109/SIST61657.2025.11139180.

10. J. L. Elman. Finding structure in time// Cognitive Science. -1990. - Vol.14(2). - P.179–211. DOI 10.1207/s15516709cog1402_1.

11. H. Schmid. Probabilistic part-of-speech tagging using decision trees// Proceedings of International Conference on New Methods in Language Processing, Manchester, UK. - 1994. - P.44-49. URL: https: //www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/tree-tagger1.pdf. (accessed 07.11.2025).

12. T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.- 2013. DOI 10.48550/arXiv.1301.3781.

13. J. Pennington, R. Socher, and C. D. Manning. GloVe: Global vectors for word representation// Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). -2014.- P. 1532–1543. DOI 10.3115/v1/D14-1162.

14. C. Bishop, Pattern Recognition and Machine Learning. Berlin, Heidelberg: Springer, 2006.-737 s. - P.137–145. ISBN-10: 0-387-31073-8. ISBN -13: 978-0387-31073-2/

15. Sammut, Claude et al. Encyclopedia of Machine Learning and Data Mining.- Springer-2017. DOI 10.1007/978-1-4899-7687-1, ISBN978-1-4899-7685-7

16. D. Hosmer and S. Lemeshow, Applied Logistic Regression: A Self-Learning Text//New York, NY: Springer-Verlag. - 528 s.2013. ISBN 9780470582473

17. T. Hastie, R. Tibshirani, and J. Friedman. The Elements of Statistical Learning. -2017.-Springer. DOI 10.1007/978-0-387-21606-5.

Авторлар туралы мәліметтер

Құттыбек А.А. - докторант 1-ші курс, Л.Н. Гумилев атындағы Еуразия Ұлттық Университеті, Астана, Қазақстан, e-mail: azhara9322@gmail.com;

Казиева Н.М. - PhD докторы, доцент, Л.Н. Гумилев атындағы Еуразия Ұлттық Университеті, Астана, Қазақстан, e-mail: kaziyevanm@gmail.com;

Амирова А.С. - PhD докторы, ассистент - профессор, Astana IT University, Астана, Қазақстан, e-mail: akzhibek.amirova@astanait.edu.kz;

Рюмин Д.А.- кандидат физика-математикалық ғылымдары, доцент Ресей Ғылым Академиясының Федеральды Ғылыми-зерттеу Орталығының ғылыми қызметкері, Санкт-Петербург,Ресей, - e-mail: neweraairesearch@gmail.com.

Information about the Authors

Kuttybek A.A. - 1st-year PhD student, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan, e-mail: azhara9322@gmail.com;

Kaziyeva N.M. PhD, Associate Professor, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan, e-mail: kaziyevanm@gmail.com;

Amirova A.S. - PhD, Assistant Professor, Astana IT University, Astana, Kazakhstan, e-mail: akzhibek.amirova@astanait.edu.kz;

Ryumin D.A. - Candidate of Physical and Mathematical Sciences, Associate Professor, Researcher, Federal Scientific Research Center of the Russian Academy of Sciences, St. Petersburg, Russia, e-mail: neweraairesearch@gmail.com.