КОМПЬЮТЕРНЫЕ НАУКИ, ПРИБОСТРОЕНИЕ И АВТОМАТИЗАЦИЯ
КОМПЬЮТЕРЛІК ҒЫЛЫМДАР, АСПАП ЖАСАУ ЖӘНЕ АВТОМАТТАНДЫРУ
COMPUTER SCIENCE, INSTRUMENTATION AND AUTOMATION

ӘОЖ 004.056
ҒТАХА 20.23.17
https://doi.org/10.53364/24138614_2025_37_2_13

М.Ж. Жарылқапова[1], Д.Б. Темірбек[1], С.Т. Мамбетов[2], О.К. Джолдасбаев[4], Е.Е. Өксікбаев[1]

[1]Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан
[2]Тұран университеті, Алматы, Қазақстан
[3]Алматы технологиялық университеті, Алматы, Қазақстан
[4]ҚР Президентінің жанындағы Мемлекеттік басқару академиясы Алматы қаласы бойынша филиалы, Алматы, Қазақстан

[1]*E-mail: makpal_zharylkapova@mail.ru**

КИБЕРПРОПАГАНДАНЫ БАҚЫЛАУ ӘДІСТЕРІНЕ ШОЛУ: ГИБРИДТІ МОДЕЛДІ ҚОЛДАНУ

Аңдатпа. *Қазіргі уақытта әлеуметтік желілер мен мессенджерлерде ақпараттық манипуляциялардың және киберпропагандаңың таралуы қоғам үшін үлкен қауіп тудырып отыр. Әлеуметтік желілерде түрлі көзқарастарды қалыптастыру және қоғамды басқару мақсатында киберпропаганда кеңінен қолданылуда. Бұл құбылыс саяси жағдайларға, экономикалық тұрақтылыққа, сонымен қатар, қоғамдық пікірдің қалыптасуына әсер етеді. Осы жұмыста киберпропагандаңы анықтау және бақылау үшін қолданылатын әдістер талқыланады. Гибридті модельдер бірнеше алгоритмдердің нәтижелерін біріктіріп, мәтіннің контекстін тереңірек түсінуге мүмкіндік береді. Бұл әдіс әсіресе көп тілде және үлкен көлемдегі деректермен жұмыс істегенде тиімдірек.*

Мақалада киберпропагандаңы анықтау процессі бірнеше модельдер көмегімен қарастырылады. Сонымен қатар, CNN, Random Forest және гибридті модельдер бойынша алынған нәтижелер салыстырылып, олардың тиімділігі көрсетіледі. Нәтижелер жоғары дәлдікке ие екендігі анықталады, бірақ CNN моделінің жоғары есептеу шығындарымен және деректермен жұмыс жасаудағы қиындықтары да бар. Бұл зерттеу гибридті модельдердің артықшылықтары мен кемшіліктерін анықтап, олардың нақты жағдайларда тиімділігін талқылайды.

Түйін сөздер: *киберпропаганда, гибридті модель, CNN, Random Forest, әлеуметтік желілер.*

Кіріспе.

Қазіргі заманның ақпараттық кеңістігі жылдам қарқынмен дамып келеді. Әлеуметтік желілер мен мессенджерлер адамдардың қарым-қатынас жасау, ақпарат алмасу және ақпарат тарату тәсілдерін түбегейлі өзгертті. Киберпропаганда, яғни ақпаратты манипуляциялау және жалған ақпарат тарату, қазіргі қоғамның маңызды мәселелерінің бірі ретінде көрініс табуда.

Киберпропаганда әлеуметтік желілер мен мессенджерлер арқылы түрлі форматтарда таралуда. Оның ішінде жалған жаңалықтар, манипуляциялық мақалалар, контексті бұрмалау, бейнемазмұндар мен пікірлерді бағыттап қолдану бар [1]. Бұл әдістер қоғамдық пікірді өзгерту, саяси мақсаттарға жету немесе адамдардың мінез-құлқын басқару үшін

қолданылады. Соңғы жылдардағы халықаралық оқиғалар, соның ішінде саяси сайлаулар мен әлеуметтік қозғалыстар, киберпропаганданың ықпалын айқын көрсетті [2].

Мәтіндерге негізделген дәстүрлі әдістер, мысалы, Random Forest сияқты ансамбльдік тәсілдер және CNN (Convolutional Neural Networks) сияқты нейрондық желілер, киберпропаганданы анықтауда кеңінен қолданылып жүр. Сонымен қатар, бұл әдістерді біріктіретін гибридті модельдер әр түрлі ақпаратты талдауда жоғары тиімділік көрсетеді.

Гибридті модельдер бірнеше алгоритмдерді біріктіре отырып, деректердің контекстін жақсырақ түсінуге мүмкіндік береді. Әлеуметтік желілер мен мессенджерлерден алынған мәтіндік деректерде түрлі контекстік ерекшеліктер бар болғандықтан, мұндай модельдер мәселені шешуде өте тиімді.

Бұл тәсілдердің үйлесімі деректерді талдаудың дәлірек нәтижелерін алуға мүмкіндік береді. Сонымен қатар, гибридті модельдер жеке алгоритмдердің шектеулерін азайтып, олардың артықшылықтарын біріктіреді.

Қазіргі кезде киберпропаганданы анықтау үшін тек техникалық әдістерді қолдану жеткіліксіз. Бұл мәселеге әлеуметтік және саяси тұрғыдан да назар аудару қажет. Киберқауіпсіздікті қамтамасыз ету және қоғамдық тұрақтылықты сақтау үшін заманауи технологияларды қолдану - бұл бағыттағы маңызды қадам [3-4].

Жұмыстың өзектілігі киберпропаганданың кең таралуы, әсіресе әлеуметтік желілерде, көптеген елдерде ақапараттық қауіпсіздікке қатысты жаңа сын-қатерлер туғызды. Бұл зерттеу осы мәселеге арналған заманауи тәсілдерді, оның ішінде гибридті модельдерді қолдану арқылы тиімді шешімдер ұсынуға бағытталған.

*Әдебиеттерге шолу***.** Әлеуметтік желілердегі киберпропаганда - автоматтандырылған әдістермен зерттелетін өзекті тақырыптардың бірі. Соңғы жылдардағы зерттеулер киберпропаганданы анықтау және оған қарсы күресу үшін көптеген әдістер мен модельдерді ұсынды. Осы зерттеулерден алынған мағлұматтар мақала зерттеуінің әдістемелік негізін құрайды.

Әлеуметтік желілерде эмоционалды инфекциялардың таралуын зерттеу киберпропаганданың қоғамдағы әсерін тереңірек түсінуге мүмкіндік берді [1]. Пайдаланушылардың мінез-құлқын талдай отырып, дезинформация мен пропаганданың таралуын зерттеу әлеуметтік желілерде таралатын ақпараттың ықпалын бағалау үшін маңызды әдіс болып табылады [2]. Дезинформацияны анықтау үшін машиналық оқыту әдістері ұсынылып, киберпропаганданы автоматтандырылған жүйелер арқылы тиімді анықтау мүмкіндігін ашты [3]. Гибридті модельдер арқылы саяси пропаганданы анықтау тиімділігін көрсеткен жұмыстар әртүрлі мәтін өңдеу әдістерін біріктіру үшін маңызды құрал болды [4]. Пропаганданы анықтаудың дәлдігін арттыру тәсілдері дәстүрлі және нейрондық желілерді біріктіру арқылы ұсынылды [5]. Сөздік векторлық көріністердің пропагандалық контентті жіктеудегі тиімділігі зерттелді, бұл мәтіндерді векторлық көрініске келтіру үшін негіз болды [6-8]. Терең оқыту әдістері саяси пропаганданы анықтауда қолданылып, зерттеуге маңызды үлес қосты [9]. Трансформерлердің манипуляциялық контентті анықтаудағы күшті құрал ретінде қолданылатыны көрсетілді [10]. Фейк жаңалықтар мен саяси пропаганданың тілін талдай отырып, контексттік сараптама жасауға көмектесті [11]. Баспасөз тақырыптары мен мақалалар арасындағы байланысты зерттеу фейк жаңалықтарды анықтау тәсілдерін ұсынды [12]. Терең оқыту әдістерін пайдалана отырып, пропаганданы анықтау тәсілдері жетілдірілді, бұл алгоритмдер мен модельдерді жақсартуға мүмкіндік берді [13]. Мемдер арқылы пропаганданы анықтаудың көпқабатты тәсілдері ұсынылды, бұл мемдер мен визуалды контентті біріктіріп, пропаганданы дәл анықтау жолдарын ұсынды [14]. Жасанды интеллектіні қолдану арқылы дезинформациямен күресудің артықшылықтары мен кемшіліктері қарастырылып, киберпропаганданың анықталуында жаңа әдістерді қолдануға мүмкіндік берді [15].

Жаңа зерттеулер киберпропагандаңы анықтау әдістерінің дамуына айтарлықтай үлес қосып келеді. Қазіргі зерттеулер терең оқыту мен гибридті модельдер арқылы әлеуметтік желілердегі манипуляциялық контентті анықтаудың тиімділігін көрсетті [16, 17]. Бұл әдістер әлеуметтік желілерде ақпараттың әсерін дұрыс бағалауға мүмкіндік беріп, контекстті терең талдауды жүзеге асыруға болады [18]. Киберпропагандаңы онлайн қауымдастықтарда анықтауға арналған жаңа әдістер киберқауіпсіздік саласында маңызды құрал болып табылады [19]. Сонымен қатар, фейк жаңалықтар мен киберпропагандаңы анықтау әдістерінің жүйелі шолуы түрлі әдістердің кемшіліктері мен артықшылықтарын бағалауға мүмкіндік береді [20].

Жоғарыда аталған зерттеулер жұмыста қолданылатын әдістер мен модельдердің негізін қалады. Әсіресе, машиналық оқыту мен терең оқыту әдістерін қолдану, мәтін мен контексттің өзара байланысын зерттеу, және мемдер мен визуалды контентті біріктіру арқылы пропагандаңы анықтау тәсілдері зерттеудің маңызды құрамдас бөліктері болып табылады. Онлайн платформадағы эмоционалды инфекцияны зерттеп, пропагандаңың таралу механизмдерін жақсырақ түсінуге мүмкіндік берді.

Материалдар мен тәсілдер.

Зерттеу үшін негізгі дереккөз ретінде Twitter әлеуметтік желісі таңдалды. Бұл платформаның пайдаланушылары әртүрлі тақырыптарда өз ойларын жиі жариялап, шынайы пікір алмасады, сондықтан ол киберпропаганда белгілерін анықтауға қолайлы дереккөз болып табылады. Деректер Twitter API арқылы Python бағдарламалау тілінде жинақталды. API пайдалануға мүмкіндік беретін арнайы сұранымдар жазылып, олар деректерді нақты уақыт режимінде жинауға арналған [5].

Барлығы шамамен 12 000 жазба жиналды. Олар екі негізгі категорияларға бөлінді:

- Пропаганда: 6 000 жазба. Бұл топқа белгілі бір идеологияны, саяси бағытты немесе қоғамдық көзқарасты насихаттайтын жазбалар енгізілді.
- Пропаганда емес: 6 000 жазба. Бұл топқа бейтарап немесе пропагандалық мазмұнға жатпайтын жазбалар енгізілді.

Жазбалар ағылшын тілінде және әртүрлі форматтарда және стильдерде болды. Оларды өңдеу үшін арнайы алдын ала өңдеу әдістері қолданылды.

Кесте 1 - Деректер жиынының атрибуттары

Атрибут	Тип	Сипаттама
Text	Мәтін	Твиттің мәтіні
is_propaganda	Логикалық мән	Твиттің пропаганда ма, жоқ па екені (1 - True, 0 - False)

1-кестеде көріп отырғанымыздай, деректер жиынының әрбір атрибуты сипатталып, оның типі мен мәні түсіндірілген.

```
                                                text  \
0  Woman who held up poster of Marine Le Pen and ...
1   Zelensky: Around 150,000 people trapped in M...
2  RT @natomission_ru: RU#Russia Deputy FM Sergey...
3  #Azovstal fully liberated - Russian military\n...
4  RT @BloombergUK: "He was almost foaming at the...

                              cleaned_text_no_emojis
0  woman who held up poster of marine le pen and ...
1  zelensky around 150,000 people trapped in mari...
2  rt natomission_ru russia deputy fm sergey ryab...
3  azovstal fully liberated  russian military\n\n...
4  rt bloomberguk he was almost foaming at the mo...
```

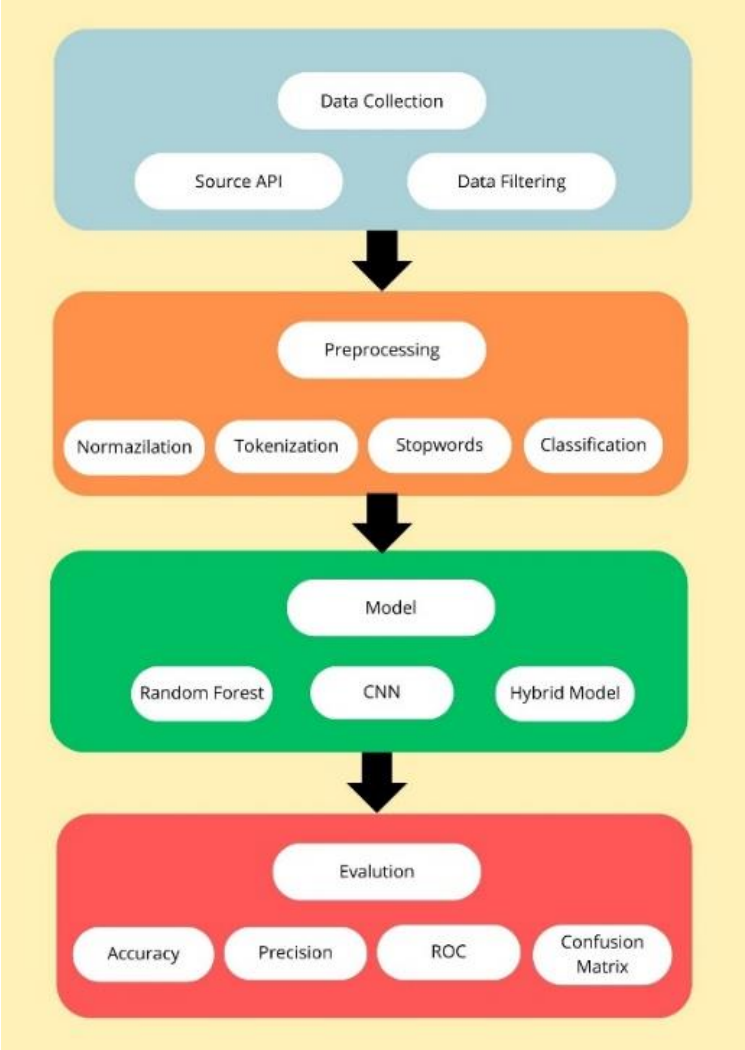
Сурет 1 – Мәтіндік деректерді өңдеу

1-суретте бейнелегендей екі негізгі баған бар: text және cleaned_text_no_emojis.

- text бағаны бастапқы мәтінді қамтиды. Бұл бағанда әртүрлі әлеуметтік желілерден алынған хабарламалар мен посттар бар. Олар түрлі формада болуы мүмкін: эмодзи, хэштегтер, сілтемелер немесе басқадай артық ақпараттармен толтырылған мәтіндер.
- cleaned_text_no_emojis бағанында бастапқы мәтіндерден эмодзи, артық белгілер және басқа қажетсіз элементтер жойылған. Мұнда тек мәтіннің мазмұны ғана сақталады, бұл мәтіннің негізгі мағынасын нақтырақ және анализ жасауға ыңғайлы етеді.

Бұл өңдеу әдісі әсіресе әлеуметтік желілердегі контенттерді зерттегенде, мысалы, пропаганда немесе жалған ақпаратты анықтағанда пайдалы болуы мүмкін. Жазбалардың өңделген түрі жасанды интеллект модельдері арқылы мәтінді класификациялауға немесе басқа да аналитикалық тапсырмаларға қолдануға дайын болады.

Деректерді алдын ала өңдеу - зерттеудің ең маңызды кезеңдерінің бірі. Әлеуметтік желілерден алынған мәтіндер әрдайым біркелкі емес және көбінесе артық символдармен немесе қажетсіз ақпаратпен толықтырылады [6-7]. Сондықтан деректерді талдау үшін оларды стандартты форматқа келтіру қажет болды.



Сурет 2 – Әдіснамалық тәсіл

Алдын ала өңдеу кезеңдері 2-суретте көрсетілгендей ретпен орындалады. Олар:

- **Нормализация.**

Жазбалардан сілтемелер, эмодзилер, арнайы таңбалар, артық пунктуациялар және басқа да қажетсіз элементтер алынып тасталды. Бұл процесс мәтіннің сапасын жақсартып,

талдау үшін маңызды ақпаратты қалдыруға мүмкіндік берді.

- **Токенизация.**

Лемматизация сөздерді олардың түбіне келтіріп, жалпы формасын анықтауға көмектесті (мысалы, «жүріп» сөзі «жүру» түріне келтірілді). Токенизация мәтіндерді жеке сөздерге немесе сөйлемдерге бөліп, олардың әрқайсысын модель үшін дерек ретінде пайдаланды.

- **Стоп сөздерді жою.**

Стоп сөздер - мәтіннің жалпы құрылымына әсер етпейтін, бірақ жиі кездесетін сөздер. Мысалы, «және», «бірақ», «мен». Бұл сөздерді жою арқылы деректерді талдаудың сапасы жақсарды.

- **Классификация.**

Жазбалар «пропаганда» және «пропаганда емес» деп бөлініп, әрқайсысына сәйкес белгілер (labels) берілді [11-12].

Ал модельдер жиынтығы осылай сипатталды:

- **Random Forest моделі.**

Random Forest - бұл ансамбльді әдіс, ол бірнеше шешім ағаштарынан тұрады. Әрбір шешім ағашы деректердің бір бөлігін талдап, болжам жасайды. Барлық ағаштардың нәтижелері біріктіріліп, соңғы шешім қабылданады [6].

- **CNN (Convolutional Neural Networks) моделі.**

CNN моделі әдетте суреттерді өңдеуге арналаған, бірақ мәтіндік деректермен жұмыс істегенде де тиімді. Бұл модель мәтіннің құрылымдық ерекшеліктерін анықтап, оларды талдауға қабілетті [9-10].

- **Гибридті модель.**

Гибридті модель зерттеуде қолданылған Random Forest, CNN модельдерін біріктіру арқылы жасалды. Әрбір модельдің нәтижелері біріктіріліп, соңғы шешім қабылданды [13-15]. Бұл модельдің артықшылықтары: бір модельдің кемшілігін екіншісі толықтыра алады және нақты нәтиже көрсеткіштері жоғарылауынан тұрады [8].

Модельді бағалау параметрлері:

- **Accuracy.**

Бұл барлық дұрыс болжамдардың жалпы санға қатынасы. Яғни, бұл модельдің қаншалықты дұрыс жұмыс істейтінін көрсетеді.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

мұндағы:

TP - дұрыс позитивті нәтижелер (True Positives);

TN - теріс нәтижелер (True Negatives);

FP - жалған позитивті нәтижелер (False Positives);

FN - жалған теріс нәтижелер (False Negatives).

- **Precision.**

Модель позитивті болжам жасағанда оның қаншалықты дұрыс болатынын көрсетеді. Бұл метрика теріс класты дұрыс анықтамаған жағдайда пайда болатын қателіктерді болдырмауға көмектеседі.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

мұндағы:

TP - дұрыс позитивті нәтижелер (True Positives);

FP - жалған позитивті нәтижелер (False Positives).

- **ROC.**

ROC (Receiver Operating Characteristic) қисығы классификатордың тиімділігін

көрсетеді. Бұл қисықтың астындағы аудан (AUC) классификатордың жалпы тиімділігін бағалайды. ROC қисығының екі басты көрсеткіші:

$$True\ Positive\ Rate \text{ немесе } Recall = \frac{TP}{TP\ +\ FP} \quad (3)$$

$$False\ Positive\ Rate = \frac{FP}{FP\ +\ TN} \quad (4)$$

ОС қисығының көмегімен классификатордың әртүрлі шекті мәндерде қалай жұмыс істейтінін көруге болады. Егер AUC мәні 1-ге жақын болса, модель жақсы жұмыс істейді.

- **Confusion Matrix.**

Бұл классификатордың нәтиже кластарын нақты класстармен салыстырып көрсететін кесте. Бұл матрица модельдің қандай қателіктер жасағанын анық көруге мүмкіндік береді.

	Predicted no Propaganda	Predicted no Propaganda
Actual no Propaganda	*TN*	*FP*
Actual Propaganda		*TP*

мұндағы:
TN (True Negatives) - дұрыс теріс болжамдар;
FP (False Positives) - теріс жағдайларды позитив деп болжаған дұрыс емес болжамдар;
FN (False Negatives) - позитив жағдайларды теріс деп болжаған дұрыс емес болжамдар;
TP (True Positives) - дұрыс позитивті болжамдар.

Confusion Matrix арқылы барлық негізгі көрсеткіштерді - Precision, Recall, F1-Score, Accuracy және басқаларын есептеуге болады.

Нәтижелер мен талқылау.

Бұл бөлімде зерттеу нәтижелері мен олардың талқылауы ұсынылады.
Random Forest архитектурасы.

Random Forest - бұл шешім ағаштарынан құралған ансамбльдік әдіс, және оның негізінде әрбір ағаш өзінің жеке классификациясын шығарады, ал соңғы нәтиже - барлық ағаштардың шешімдері бойынша көпшілік дауыспен анықталады [16-18]. Модельдің нақты формуласын қарастырайық:

$$\hat{y} = mode(f_1(x), f_2(x), \dots f_T(x)) \quad (5)$$

мұндағы:
$\hat{y}$ - болжам жасалған нәтиже (пропаганда немесе пропаганда емес);
$f_T(x)$) - t-шы шешім ағашының нәтижесі, яғни x деректері бойынша шыққан шешім;
T - ағаштардың саны (Random Forest ансамблінде ағаштардың саны T деп белгіленеді).

Шешім ағашының формуласы: Әрбір шешім ағашы x деректерін алып, оларды бірнеше сұрақтарға жауап беру арқылы екі классқа бөледі:

$$y = if\left(x_{feature} > threshold\right)\ then\ Class\ A\ else\ Class\ B \quad (6)$$

$x_{feature}$ - берілген деректің бір белгісі (мысалы, текстің бір ерекшелігі, сөз жиілігі);
$threshold$ - шешім ағашында белгіленген шекті мән.

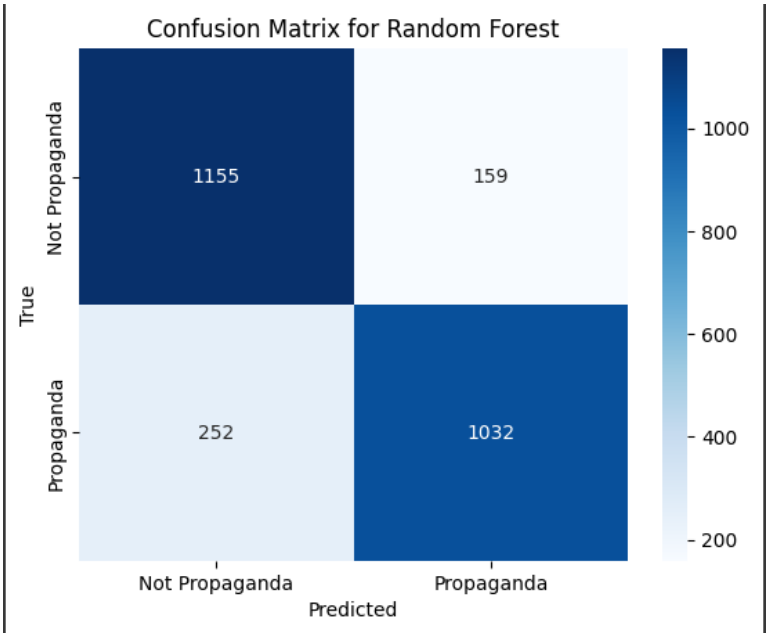
Ең көп дауыс алған ағаштың шешімі Random Forest нәтижесін береді. Random Forest моделінің нәтижелері бойынша пропаганда емес жіне пропаганда кластарын талдау 2 кестеде көрсетілгендей пропаганда емес класы үшін accuracy 0.84, precision 0.88 және recall 0.85, бұл модельдің пропаганда емес контентті дұрыс анықтауда жоғары дәлдікпен жұмыс істегенін көрсетеді. Ал, пропаганда класы үшін accuracy 0.87, precision 0.82 және recall 0.83,

бұл модельдің пропаганда контентін анықтауда да жеткілікті нәтижелер көрсеткенін білдіреді.

Кесте 2 - Random Forest моделінің классификация нәтижелері

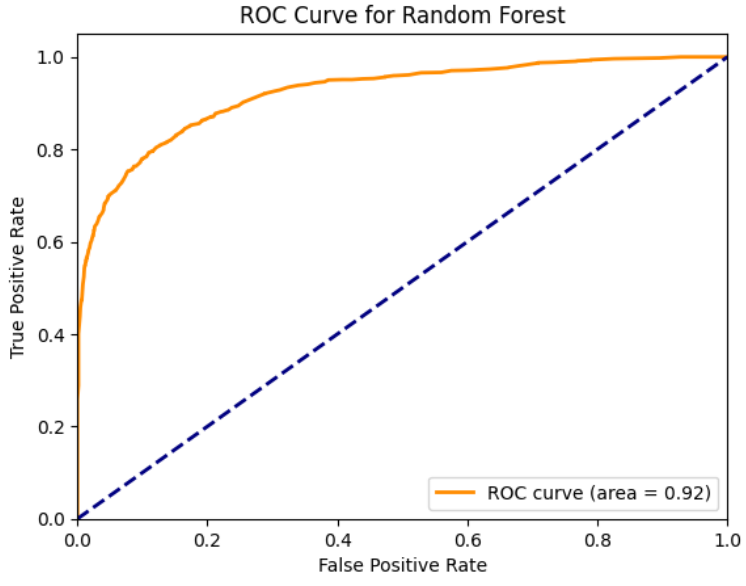
Класс	Accuracy	Precision	Recall	F1-score
Пропаганда емес	0.84	0.88	0.85	0.83
Пропаганда	0.87	0.82	0.84	0.84

Random Forest моделі екі класты да тиімді анықтады. Пропаганда емес контенттің жоғары дәлдігі мен толықтығын көрсетіп, пропаганда класы бойынша жақсы нәтижелерге жетті. Мұндай нәтижелер модельдің жалпы тиімділігін және дәлдігін дәлелдейді.



Сурет 3 – Random Forest моделінің Confusion Matrix көрінісі

Confusion Matrix-тан алынған бұл нәтижелер 3-суретте көрсетілгендей модельдің «пропаганда» және «пропаганда емес» деректерін дұрыс классификациялау қабілетінің жақсы екенін көрсетеді.



Сурет 4 – Random Forest моделінің ROC қисығы

Random Forest моделі «пропаганда» және «пропаганда емес» класстарын ажыратуда жоғары көрсеткіштерді көрсетеді. ROC AUC 4-суретте көрсетілгендей көрсеткішті 0.92 модельдің екі классты тиімді ажырату қабілетін білдіреді, бұл оның дәлдігін жоғары деңгейде сақтап отырғанын көрсетеді. CNN (Convolutional Neural Network) архитектурасы.

CNN - бұл көп қабатты нейрондық желі, оның ішінде негізгі үш компонент бар:

1.Конволюциялық қабат:

CNN-нің негізгі қабаты болып табылатын конволюциялық операция үшін формула келесі түрде болады:

$$y = \sum_{i=1}^{n} \omega_i \cdot x_i + b \quad (7)$$

мұндағы:

y - шығыс (қабаттың нәтижесі);

x_i - кіріс деректерінің мәндері (немесе алынған белгілер);ω_i - салмақтар (немесе фильтрлер);

b - биас (bias);

n - кіріс деректерінің саны.

Бұл операция көптеген фильтрлер мен ядроларды қолдана отырып, деректерді сызықты түрде өңдейді.

2. Белсендіру функциясы (Activation Function):

Көбінесе ReLU (Rectified Linear Unit) функциясы пайдаланылады:

$$f(x) = max(0, x) \quad (8)$$

Бұл функция сызықтық емес түрлендіруді жүзеге асырады, бұл модельдің күрделі деректерге жауап беру қабілетін арттырады.

3. Толық байланысқан қабат (Fully Connected Layer):

Толық байланысқан қабатта әрбір нейрон алдыңғы қабаттағы барлық нейрондармен байланысады. Бұл қабаттың функциясы:

$$y = W \cdot x + b \quad (9)$$

мұндағы:

W - салмақтар;

x - кіріс;

b - биас.

Бұл қабат арқылы CNN ішіндегі барлық белгілер біріктіріліп, соңғы шешім қабылданады.

Көбінесе CNN архитектурасының соңғы нәтижесі softmax функциясы арқылы шығарылады, бұл нәтижелердің ықтималдығын береді:

$$P(y = \kappa|x) = \frac{e^{z_k}}{\Sigma_i e^{z_k}} \quad (10)$$

мұндағы:

$P(y = \kappa|x)$ - x кірісі үшін κ-классының ықтималдығы;

z_k - κ-классының логиттер мәні (шешім қабылдау үшін сызықтық функция);

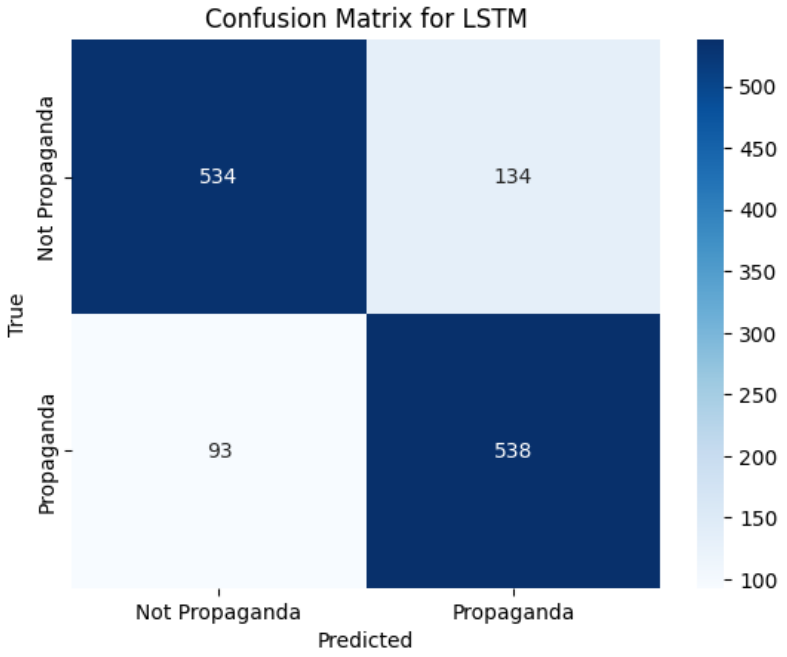
$\Sigma_i e^{z_k}$ - барлық мүмкін класстардың экспоненциалды жиынтығы.

3-кестеде көрсетілгендей пропаганда емес классы үшін accuracy 0.90, precision 0.92 және recall 0.91, бұл модельдің «пропаганда емес» контентті анықтауда жоғары дәлдікпен жұмыс істегенін білдіреді. Ал, пропаганда класы үшін accuracy 0.91, precision 0.90 және recall 0.91, бұл модельдің «пропаганда» контентін анықтауда тиімді жұмыс істегенін көрсетеді.

Кесте 3 - CNN моделінің классификация нәтижелері

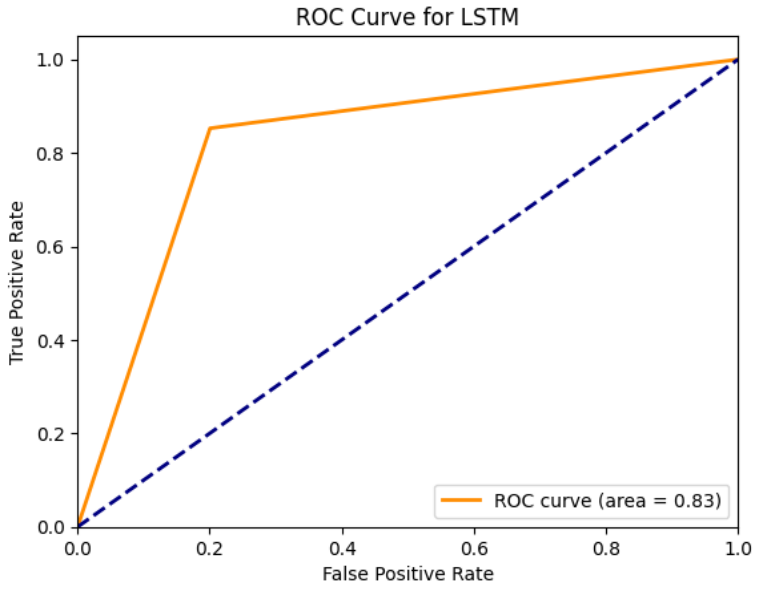
Класс	Accuracy	Precision	Recall	F1-score
Пропаганда емес	0.90	0.92	0.91	0.91
Пропаганда	0.91	0.90	0.91	0.90

Бұл көрсеткіштер жүйенің екі класс арасында тиімді классификациялау жасағанын көрсетеді.



Сурет 5 - CNN моделінің Confusion Matrix көрінісі

Confusion Matrix 5-суретте көрсетілгендей көрсеткіштер модельдің болжам жасау мүмкіндігін анықтауға көмектеседі. Алынған көрсеткіштер модельдің жақсы жұмыс істеп жатқанын білдіреді.



Сурет 6 - CNN моделінің ROC қисығы

CNN моделі үшін ROC AUC көрсеткіші 6-суретте көрсетілгендей 0.83 құрайды. Бұл көрсеткіш модельдің «пропаганда» мен «пропаганда емес» контентін ажырату қабілетінің орташа деңгейін білдіреді.

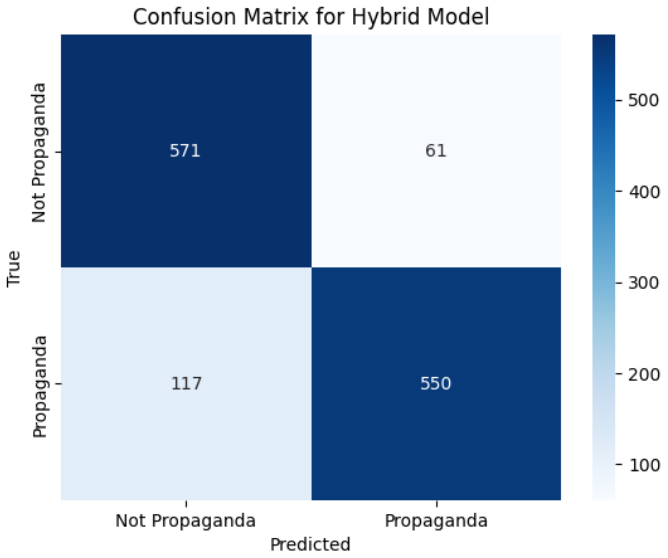
Гибридті модель архитектурасы (Random Forest + CNN)

Гибридті модельдің нәтижелері екі классты классификациялау үшін 4-кестеде көрсетілгендей жоғары дәлдік көрсетеді.

Кесте 4 - Гибридті модельдің классификация нәтижелері

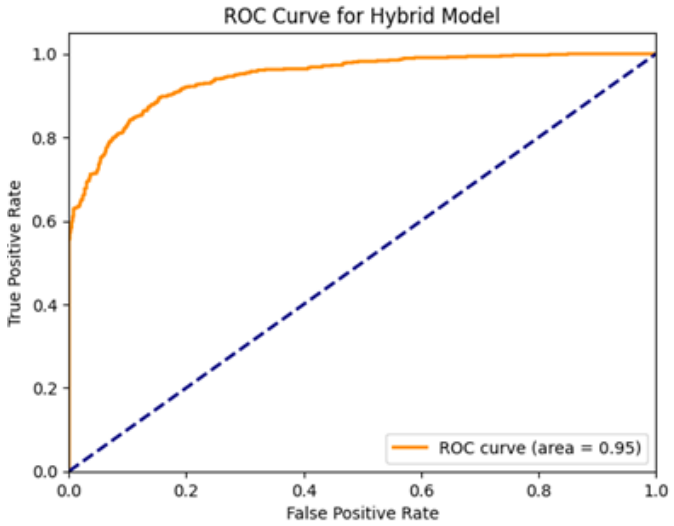
Класс	Accuracy	Precision	Recall	F1-score
Пропаганда емес	0.92	0.90	0.91	0.95
Пропаганда	0.94	0.96	0.94	0.96

Гибридті модель «пропаганда» және «пропаганда емес» контентін классификациялау кезінде жоғары дәлдікке қол жеткізді.



Сурет 7 - Гибридті модельдің Confusion Matrix көрінісі

7-суреттегі Confusion Matrix нәтижелері гибридті модельдің деректердің дәл анықтау деңгейінің жоғары екенін көрсетеді.



Сурет 8 - Гибридті модельдің ROC қисығы

ROC AUC 8-суретте көрсетілгендей 0.95 модельдің өте жақсы қабілеттерін көрсетеді және бұл көрсеткіш модельдің қате болжамдарын минимизациялап, екі классты нақты ажырататындығын дәлелдейді.

Әртүрлі машиналық оқыту әдістері киберпропаганданы анықтауда тиімді болғанымен, олардың әрқайсысының өзіндік шектеулері бар. Төмендегі кестеде осы әдістердің негізгі кемшіліктері мен оларды жою жолдары көрсетілген. Бұл шектеулерді жою арқылы модельдердің тиімділігін арттыруға болады.

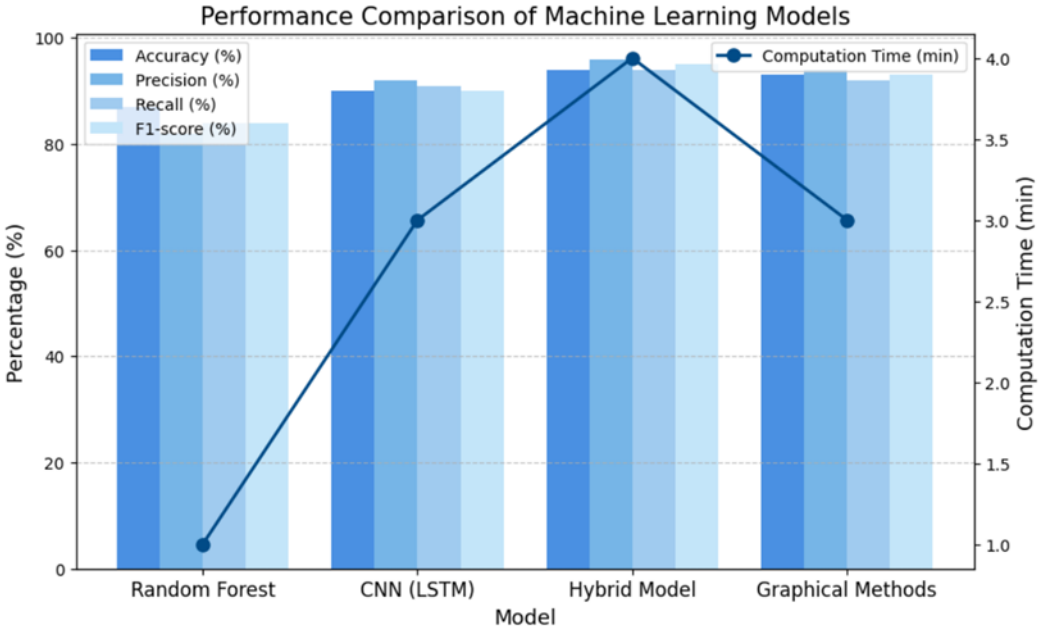
Кесте 5 – Әдістердің кемшіліктері мен оларды жою жолдары

Әдіс	Кемшіліктері	Оларды жою жолдары
Random Forest	Үлкен деректерді өңдеу кезінде өнімділіктің төмендеуі	Параллель есептеу технологияларын қолдану
CNN	Есептеу ресурстарына жоғары сұраныс	GPU жеделдеткіштерін пайдалану
Гибридті модель	- Нәтижелерді біріктіру кезінде интерпретациялау қиындығы; - Екі модельді біріктіру кезінде есептеу шығындарының артуы; - Нақты уақыт режимінде өнімділіктің төмендеуі	- Салмақты оңтайландыру және Explainable AI (XAI) қолдану; - Архитектураны жеңілдету және параллель есептеу технологияларын қолдану; - GPU жеделдеткіштерін пайдалану арқылы жылдамдықты арттыру

Үш модельді салыстырып қарағанда 6-кестеде көрсетілгендей, гибридті модель жоғары дәлдік пен тиімділікті көрсетіп, пропаганда мен пропаганда емес контентті анықтауда ең жақсы нәтижеге жетті. Random Forest және CNN модельдері де тиімді жұмыс істеді, бірақ олардың нәтижелері гибридті модельден сәл төмен болды. Мысалы, Random Forest моделі дәлдік пен ROC көрсеткіштерінде CNN моделінен озық болса да, гибридті модельдің көрсеткіштері айтарлықтай жоғары екені байқалды. Сондай-ақ, зерттеуде графикалық әдістер (Node2Vec + GNN) сияқты балама заманауи машиналық оқыту әдістері қолданылып, олардың тиімділігі де бағаланды. Нәтижелер көрсеткендей, бұл әдістер жоғары дәлдік пен өнімділікке қол жеткізді, әсіресе деректер арасындағы байланысты анықтауда және көпөлшемді кеңістікте пропагандалық үлгілерді тануда тиімді болды. Дегенмен, гибридті модель барлық негізгі метрикалар бойынша ең жоғары нәтижелерге қол жеткізіп, өз артықшылығын дәлелдеді.

Кесте 6 – Машиналық оқыту модельдерін қолдана отырып, мәтінді пропаганда және пропаганда емес категорияларға жіктеу нәтижелері

Машиналық оқыту модельдері	**Accuracy**	**Precision**	**Recall**	**F1-score**	**Есептеу шығындары, мин**
Random Forest	0.87	0.82	0.92	0.84	1
CNN (LSTM)	0.90	0.92	0.83	0.9	3
Гибридті модель	0.94	0.96	0.95	0.96	4
Графикалық әдіс (Node2Vec + GNN)	0.93	0.94	0.92	0.93	3



Сурет 9 - Random Forest, CNN (LSTM), гибридті модель және графикалық әдістерді салыстыру

Модельдер 7-кестеде көрсетілгендей радикализацияға бағытталған қауіпті контентті, сонымен қатар киберпропагандалық материалдарды дұрыс классификациялау мен анықтауда жоғары нәтижелер көрсетті.

Кесте 6 - Гибридті модельдің нәтижелері мен болжамдары

Мәтін	**Болжам**
The government's actions are a clear indication of their authoritarian agenda, and they will stop at nothing to achieve their goals.	Пропаганда емес
The opposition party is planning to unite and challenge the current regime, pushing for democratic reforms.	Пропаганда емес
A radical faction is gaining support by promising to restore order and protect the country from foreign influence.	Пропаганда
The new law aims to protect citizens from the increasing threats posed by cybercrime and terrorism.	Пропаганда емес
This movement is calling for the overthrow of the current political system and the establishment of a new order.	Пропаганда

Бұл әдістердің комбинациясы пропагандаңы анықтауда тиімді жұмыс істеді, қауіпті идеологияларды, радикализацияны және киберпропагандаңы анықтауда мүмкіндік берді және қоғамды қорғауға бағытталған маңызды құрал болып табылады.

Қорытынды.

Бұл зерттеу әлеуметтік желілер мен мессенджерлердегі мәтіндік контенттен киберпропагандаңы анықтау мақсатында Random Forest, CNN (Convolutional Neural Network) және гибридті модель әдістерінің тиімділігін талдауға бағытталған. Әр модельдің нәтижелері Confusion Matrix және ROC қисығы арқылы бағаланды. Барлық модельдер бойынша нәтижелер айтарлықтай жоғары дәлдікке қол жеткізді. Жүргізілген салыстыру нәтижелері көрсеткендей, барлық метрикалар бойынша ең жоғары нәтижелерге гибридті модель қол жеткізді. CNN модельдің тиімділігі әсіресе күрделі мәтіндер мен көп мағыналы сөздер үшін жоғары болды, себебі ол мәтіндердің құрылымын жақсы түсінуге мүмкіндік береді. Ал Random Forest деректердің үлкен көлемімен жұмыс істегенде тиімдірек болды,

себебі ол бірнеше шешім ағаштарының нәтижелерін біріктіреді. Сонымен қатар, графикалық әдістермен (Node2Vec + GNN) салыстырғанда да, гибридті модельдің дәлдігі мен тиімділігі жоғары болып, киберпропаганданы анықтауда анағұрлым нәтижелі екенін дәлелдеді. Модельдердің әрқайсысы мәтіндерге қатысты өзіндік талдау жүргізеді және олардың нәтижелері біріктіріліп, ең жақсы шешім қабылданды. Бұл әдіс киберпропаганданы анықтау процесін өте тиімді етеді.

Әдебиеттер тізімі

1. Zhang, D.Y., Wang, D., және Zhang, Y. (2017). Үлкен деректердегі әлеуметтік медиа сенсорында шындықты анықтаудың шектеулерді ескеретін динамикалық әдісі. IEEE International Conference on Big Data, 57-66. https://doi.org/10.1109/BigData.2017.8257911.

2. Chaudhari, D.D., және Pawar, A.V. (2021). Әлеуметтік медиадағы пропаганданы талдау: библиометриялық шолу. Information Discovery and Delivery, 57-70. https://doi.org/10.1108/IDD-06-2020-0065.

3. Barrón-Cedeno, A., Jaradat, I., Da San Martino, G., және Nakov, P. (2019). Proppy: жаңалықтарды олардың пропагандалық мазмұны бойынша ұйымдастыру. Information Processing & Management, 1849-1864. https://doi.org/10.1016/j.ipm.2019.03.005.

4. Hristakieva, K., Cresci, S., Torregrossa, J., және Nakov, P. (2021). Координацияланған қауымдастықтардың әлеуметтік желілерде пропаганданы таратуы. WebSci'22: Proceedings of the 14th ACM Web Science Conference, 109-201. https://doi.org/10.1145/3501247.3531543.

5. Dilley, L., Welna, W., және Foster, F. (2022). Twitter-дегі QAnon пропагандасы: ақпараттық соғыс, ықпал етушілер, желілер және нарративтер. https://doi.org/10.48550/arXiv.2207.05118. Liang, F., Zhu, Q., және Li, G.M. (2022). Пропагандалық ақпарат көздерін белгілеудің жаңалықтарды бөлісуге әсері: Twitter-дегі жартылай эксперименттік зерттеу. The International Journal of Press/Politics, 28(4). https://doi.org/10.1177/19401612221086905.

6. Mynzar, B., Stetsenko, I. V., Gordienko, Y., және Stirenko, S. (2024). Twitter мәтіндеріндегі пропаганданы анықтау үшін машиналық оқыту әдісі. Lecture Notes in Networks and Systems, 200-212. https://doi.org/10.1007/978-3-031-67348-1_15.

7. Shah, P., Pahari, S., Bhavsar, R., және Kwon, J.S.-I. (2024). Бірінші принциптер мен машиналық оқытудың гибридті моделін қолдану: қадамдық оқу әдістемелік шолуы. Computer & Chemical Engineering, 194, 108926. https://doi.org/10.1016/j.compchemeng.2024.108926.

8. Sornsuwit, P., & Jaiyen, S. (2019). Киберқауіпсіздік қауіптерін анықтау үшін адаптивті күшейтуге негізделген жаңа гибридті машиналық оқыту әдісі. Applied Artificial Intelligence, 462-482. https://doi.org/10.1080/08839514.2019.1582861.

9. Sivasankari, S., & Vadivu, G. (2021). Фейк жаңалықтардың таралу жолын әлеуметтік желілерді талдау арқылы бақылау. Soft Comput, 12883-12891. https://doi.org/10.1007/s00500-021-06043-2.

10. Azzimonti, M., & Fernandes, M. (2022). Әлеуметтік желілер, фейк жаңалықтар және поляризация. European Journal of Political Economy, 74, 102256. https://doi.org/10.1016/j.ejpoleco.2022.102256.

11. Olan, F., Jayawickrama, U., Arakpogun, E.O., Suklan J., және Liu S. (2024). Әлеуметтік медиадағы фейк жаңалықтар: қоғамға әсері. Inf Syst Front, 443-458. https://doi.org/10.1007/s10796-022-10242-z.

12. Malik, M.S.I., Imran, T., және Mamdouh, J.M. (2023). Әлеуметтік медиадағы пропаганданы қалай анықтауға болады? Семантикалық және дәл бапталған тілдік модельдерді пайдалану. PeerJ Computer Science, 9, e1248. https://doi.org/10.7717/peerj-cs.1248.

13. Syed, L., Alsaeedi, A., Alhuri, L.A., және Aljohani, H.R. (2023). Киберпропагандадан фейк жаңалықтарды анықтау үшін әлсіз қадағаланатын гибридті оқыту әдісі. Array, 19,

100309. https://doi.org/10.1016/j.array.2023.100309.

14. Pandey, R., Pandey, M., және Nazarov, A. (2022). Ақпараттық соғыстағы пропагандаңы терең оқыту арқылы анықтау. Proceedings of the 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N). https://doi.org/10.1109/ICAC3N56670.2022.10074449.

15. Ahmad, P.N., Yuanchao, L., Aurangzeb, K., Anwar M.A., және ul Haq Q.M. (2024). Әлеуметтік медиадағы пропагандалық мәтіндерді семантикалық веб негізінде анықтау: мета-оқыту әдісі. SOCA. https://doi.org/10.1007/s11761-024-00422-x.

16. Sprenkamp, K., Jones, D. G., және Zavolokina, L. (2023). Пропагандаңы анықтау үшін үлкен тілдік модельдер. Computation and language, 2310.06422. https://doi.org/10.48550/arXiv.2310.06422.

17. Khanday, A.M.U.D., Wani, M.A., Rabani, S.T., және Khan, Q.R. (2023). Әлеуметтік желілердегі пропагандистік қауымдастық пен негізгі түйінді анықтаудың гибридті әдісі. Sustainability, 15(2), 1249. https://doi.org/10.3390/su15021249.

18. Chaudhari, D.D., Pawar, A.V. (2022). Әлеуметтік медиадағы үгіт-насихатты анықтау үшін машиналық оқыту мен NLP әдістерінің жүйелі салыстырмалы талдауы. Journal of Information Technology Research, 15(1). https://doi.org/10.4018/JITR.299384

Omar, M.R., және Abdulazeez, A. (2024). Әлеуметтік желідегі фейк жаңалықтар: кешенді шолу. The Indonesian Journal of Computer Science, 13(3). https://doi.org/10.33022/ijcs.v13i3.3945

References

1. Zhang, D.Y., Wang, D., and Zhang, Y. (2017). Constraint-aware dynamic truth discovery in big data social media sensing. IEEE International Conference on Big Data, 57-66. https://doi.org/10.1109/BigData.2017.8257911

2. Chaudhari, D.D. and Pawar, A.V. (2021), Propaganda analysis in social media: a bibliometric review. Information Discovery and Delivery, 57-70. https://doi.org/10.1108/IDD-06-2020-0065

3. Barrón-Cedeno, A., Jaradat, I., Martino, G.D.S., and Nakov, P. (2019). Proppy: Organizing the news based on their propagandistic content. Information Processing & Management, 1849-1864. https://doi.org/10.1016/j.ipm.2019.03.005

4. Hristakieva, K., Cresci, S., Torregrossa, J., and Nakov, P. (2021). The spread of propaganda by coordinated communities on social media. WebSci'22: Proceedings of the 14th ACM Web Science Conference, 109-201. https://doi.org/10.1145/3501247.3531543

5. Dilley, L., Welna, W., and Foster, F. (2022). QAnon Propaganda on Twitter as Information Warfare: Influencers, Networks, and Narratives. https://doi.org/10.48550/arXiv.2207.05118

6. Liang, F., Zhu, Q., and Li, G.M. (2022). The effects of flagging propaganda sources on news sharing: Quasi-experimental evidence from Twitter. The International Journal of Press/Politics, 28(4), https://doi.org/10.1177/19401612221086905

7. Mynzar, B., Stetsenko, I. V., Gordienko, Y., and Stirenko, S. (2024). Machine learning method for detecting propaganda in Twitter texts. Lecture Notes in Networks and Systems, 200-212. https://doi.org/10.1007/978-3-031-67348-1_15

8. Shah, P., Pahari, S., Bhavsar, R., and Kwon, J.S.-I. (2024). Hybrid modeling of first-principles and machine learning: A step-by-step tutorial review for practical implementation. Computer & Chemical Engineering, 194, 108926, https://doi.org/10.1016/j.compchemeng.2024.108926

9. Sornsuwit, P., & Jaiyen, S. (2019). A new hybrid machine learning for cybersecurity threat detection based on adaptive boosting, Applied Artificial Intelligence, 462-482. https://doi.org/10.1080/08839514.2019.1582861

10. Sivasankari, S., Vadivu, G. Tracing the fake news propagation path using social network

analysis. Soft Comput, 12883-12891. https://doi.org/10.1007/s00500-021-06043-2

11. Azzimonti, M., & Fernandes, M. (2022). Social media networks, fake news, and polarization. European Journal of Political Economy, 74, 102256, https://doi.org/10.1016/j.ejpoleco.2022.102256

12. Olan, F., Jayawickrama, U., Arakpogun, E.O., Suklan J. and Liu S. (2024). Fake news on Social Media: the Impact on Society. Inf Syst Front, 443-458. https://doi.org/10.1007/s10796-022-10242-z

13. Malik, M.S.I., Imran, T., and Mamdouh, J.M. (2023). How to detect propaganda from social media? Exploitation of semantic and fine-tuned language models. PeerJ Computer Science, 9, e1248. https://doi.org/10.7717/peerj-cs.1248

14. Syed, L., Alsaeedi, A., Alhuri, L.A., and Aljohani, H.R. (2023). Hybrid weakly supervised learning with deep learning technique for detection of fake news from cyber propaganda. Array, 19, 100309. https://doi.org/10.1016/j.array.2023.100309

15. Pandey, R., Pandey, M., and Nazarov, A. (2022). Detection of Propaganda in Information Warfare using Deep Learning. Proceedings of the 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N). https://doi.org/10.1109/ICAC3N56670.2022.10074449

16. Ahmad, P.N., Yuanchao, L., Aurangzeb, K., Anwar M.A., and ul Haq Q.M. (2024). Semantic web-based propaganda text detection from social media using meta-learning. SOCA. https://doi.org/10.1007/s11761-024-00422-x

17. Sprenkamp, K., Jones, D. G., and Zavolokina, L. (2023). Large Language Models for Propaganda Detection. Computation and language, 2310.06422. https://doi.org/10.48550/arXiv.2310.06422

18. Khanday, A.M.U.D., Wani, M.A., Rabani, S.T., and Khan, Q.R. (2023). Hybrid approach for detecting propagandistic community and core node on social networks. Sustainability, 15(2), 1249. https://doi.org/10.3390/su15021249

19. Chaudhari, D.D., Pawar, A.V. (2022). A systematic comparison of machine learning and NLP techniques to unveil propaganda in social media. Journal of Information TechnologyResearch, 15(1). https://doi.org/10.4018/JITR.299384

20. Omar, M.R., Abdulazeez, A. (2024). Fake News in Social Network: A Comprehensive Review. The Indonesian Journal of Computer Science, 13(3). https://doi.org/10.33022/ijcs.v13i3.3945

ОБЗОР МЕТОДОВ КОНТРОЛЯ КИБЕРПРОПАГАНДЫ: ИСПОЛЬЗОВАНИЕ ГИБРИДНОЙ МОДЕЛИ

Аннотация. *В настоящее время распространение информационных манипуляций и киберпропаганды в социальных сетях и мессенджерах представляет серьезную угрозу для общества. Киберпропаганда широко используется с целью формирования различных точек зрения в социальных сетях и управления обществом. Это явление влияет на политические условия, экономическую стабильность, а также на формирование общественного мнения. В этой статье обсуждаются методы, используемые для обнаружения и мониторинга киберпропаганды. Этот метод особенно эффективен при работе на нескольких языках и с большими объемами данных.*

В статье рассматривается процесс определения киберпропаганды с помощью нескольких моделей. Кроме того, результаты, полученные на CNN, Random Forest и гибридных моделях, сравниваются и показывают их эффективность. Результаты показали высокую точность, но модель CNN также имеет высокие вычислительные затраты и трудности с обработкой данных. Это исследование выявляет преимущества и недостатки гибридных моделей и обсуждает их эффективность в реальных условиях.

***Ключевые слова**: киберпропаганда, гибридная модель, CNN, Random Forest, социальные сети.*

REVIEW OF CYBERPROPAGANDA CONTROL METHODS: USING A HYBRID MODEL

***Abstract.** Currently, the spread of information manipulation and cyber propaganda on social networks and messengers poses a serious threat to society. Cyber propaganda is widely used to form different points of view on social media and to manage society. This phenomenon affects political conditions, economic stability, as well as the formation of public opinion. This article discusses the methods used to monitor cyber propaganda. Hybrid models combine the results of several algorithms for a deeper understanding of the text context. This method is especially effective when working in multiple languages and with large amounts of data.*

The article examines the process of defining cyber propaganda using several models. In addition, the results obtained on CNN, Random Forest and hybrid models are compared and show their effectiveness. The results showed high accuracy, but the CNN model also has high computational costs and difficulties with data processing. This study identifies the advantages and disadvantages of hybrid models and discusses their effectiveness in real-world conditions.

***Keywords:** cyber propaganda, hybrid model, CNN, Random Forest, social networks.*

Авторлар туралы мәлімет

Жарылқапова Мақпал Саматқызы	әл-Фараби атындағы Қазақ Ұлттық Университетінің магистранты, Алматы, Қазақстан, E-mail: makpal_zharylkapova@mail.ru
Темірбек Дильназ Бекқызы	әл-Фараби атындағы Қазақ Ұлттық Университетінің магистранты, Алматы, Қазақстан, E-mail: dilnaz.temirbek11@gmail.com
Мамбетов Сәкен Төлегенұлы	Магистр, Тұран университетінің «Ақпараттық технологиялар» жоғары мектебі директоры, Алматы қ., Қазақстан E-mail: s.mambetov@turan-edu.kz
Джолдасбаев Орынбасар Капарович	PhD, Қазақстан Республикасы Президенті жанындағы мемлекеттік басқару академиясының Алматы қаласы бойынша филиалының аға оқытушысы, Алматы қ., Қазақстан e-mail: orynbassarjoldasbayev@gmail.com
Өксікбаев Ернұр Ерболұлы	әл-Фараби атындағы Қазақ ұлттық университеті, «Киберқауіпсіздік және криптология» кафедрасының аға оқытушы, Алматы, Қазақстан E-mail: ernur.oksukbaev@gmail.com

Сведение об авторах

Жарылқапова Мақпал Саматқызы	Магистрант, Казахский национальный университет им. Аль-Фараби, Алматы, Казахстан E-mail: makpal_zharylkapova@mail.ru
Темірбек Дильназ Бекқызы	Магистрант, Казахский национальный университет им. Аль-Фараби, Алматы, Казахстан E-mail: dilnaz.temirbek11@gmail.com
Мамбетов Сәкен Төлегенұлы	Магистр, Директор Высшей школы «Информационных технологий» университета Туран, г. Алматы, Казахстан E-mail: s.mambetov@turan-edu.kz
Джолдасбаев Орынбасар Капарович	PhD по управлению проектами, старший преподаватель филиала Академии государственного управления при Президенте РК по городу Алматы, г. Алматы, Казахстан

	e-mail: orynbassarjoldasbayev@gmail.com
Өксікбаев Ернұр Ерболұлы	старший преподаватель кафедры «Кибербезопасность и криптологии», Казахский национальный университет имени аль-Фараби, Алматы, Казахстан E-mail: ernur.oksukbaev@gmail.com

Information about the authors

Makpal Zharylkapova	Master's student, Al-Farabi Kazakh National University, Almaty, Kazakhstan, E-mail: makpal_zharylkapova@mail.ru
Dilnaz Temirbek	Master's student, Al-Farabi Kazakh National University, Almaty, Kazakhstan, E-mail: dilnaz.temirbek11@gmail.com
Saken Mambetov	MSc, Director of the Higher School of Information Technology, Turan University, Almaty, Kazakhstan E-mail: s.mambetov@turan-edu.kz
Orynbassar K. Joldasbayev	PhD in Project Management, Senior Lecturer of the branch of the Academy of Public Administration under the president of the Republic of Kazakhstan in Almaty, Almaty, Kazakhstan e-mail: orynbassarjoldasbayev@gmail.com
Yernur Oxikbayev	Senior Lecturer at the Department of «Cybersecurity and Cryptology», Kazakh National University named after Al-Farabi, Almaty, Kazakhstan E-mail: ernur.oksukbaev@gmail.com