



DOI 10.51885/1561-4212_2025_4_247
FTAXP 68.41.29

ҚАЗАҚТІЛДІ YOUTUBE АРНАЛАРЫНДАҒЫ АУДИТОРИЯНЫҢ БЕЛСЕНДІЛІГІН ЖАСАНДЫ ИНТЕЛЛЕКТ НЕГІЗІНДЕ МОДЕЛЬДЕУ

МОДЕЛИРОВАНИЕ АКТИВНОСТИ АУДИТОРИИ КАЗАХСКОЯЗЫЧНЫХ YOUTUBE-КАНАЛОВ С ПРИМЕНЕНИЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

MODELING THE AUDIENCE ACTIVITY OF KAZAKH-LANGUAGE YOUTUBE CHANNELS USING ARTIFICIAL INTELLIGENCE

А.С. Тыныкулова ^{1*}, А.В. Фаддеенков ², Г.Г. Мухамбетова ³,
И.Н. Пугачева ³, Ж.С. Иксебаева ³

¹Astana International University, Астана қ., Қазақстан

²Н.М. Фёдоровский атындағы Заполярь мемлекеттік университеті, Норильск қ., Ресей Федерациясы

³М. Өтемісов атындағы Батыс Қазақстан университеті, Орал қ., Қазақстан

*Жауапты автор: Тыныкулова Асемгуль Серғужаевна, e-mail: asem_110981@mail.ru

Түйінді сөздер:

Жасанды интеллект,
мәліметтерді талдау,
аудиторияның тартылуы,
қазақтілді контент,
машиналық оқыту,
Big Data, медиасараптама,
цифрлық коммуникация,
болжамдау үлгісі

ТҮЙІНДЕМЕ

Цифрлық технологиялардың қарқынды дамуы жағдайында YouTube платформасы бейнеконтентті жасау мен таратудағы негізгі құралдардың біріне айналды, бүгінде оның белсенді пайдаланушыларының саны екі миллиардтан асты. Қазақтілді YouTube сегменті бейнематериалдардың танымалдылығы мен қаралымына әсер ететін факторларды ғылыми тұрғыда зерттеуді қажет ететін ерекше және әлі де жеткілікті зерттелмеген бағыт болып табылады. Цифрлық кеңістікте қазақ тілі мен мәдениетіне деген қызығушылықтың артуы пайдаланушылар санының көбеюімен, технологиялардың дамуымен, мәдени сәйкестіктің нығаюымен және жастардың әлеуметтік желілердегі белсенділігімен байланысты. Бұл мақалада аудиторияның қызығушылық механизмдерін және контенттің табыстылық факторларын талдау мен түсіну бейнежасаушылар, маркетингтер, цифрлық мәдениетті зерттеушілер мен ұсынымдық жүйе әзірлеушілер үшін практикалық маңызға ие. Зерттеу нәтижелері қазақ тілінің медиакөністікте дамуына және тиімді цифрлық коммуникация стратегияларын қалыптастыруға үлес қосады.

Ключевые слова:

искусственный интеллект,
анализ данных,
вовлечённость аудитории,
казахскоязычный контент,
машинное обучение,
большие данные,
медиааналитика,
цифровая коммуникация,
модель прогнозирования

АННОТАЦИЯ

В условиях стремительного развития цифровых технологий платформа YouTube стала одним из ключевых инструментов создания и распространения видеоконтента, объединяющим более двух миллиардов активных пользователей. Казахскоязычный сегмент YouTube представляет собой уникальную и недостаточно изученную нишу, требующую научного анализа факторов, определяющих популярность видео и просмотров. Рост интереса к казахскому языку и культуре в цифровом пространстве связан с увеличением числа пользователей, развитием технологий, усилением



культурной идентичности и активностью молодёжи в социальных сетях. В данной статье наряду с анализом и пониманием механизмов интереса аудитории и факторов успешности контента имеется практическое значение для создателей видео, маркетологов, исследователей цифровой культуры и разработчиков рекомендательных систем. Результаты исследования способствуют продвижению казахского языка в медиaprостранстве и формированию эффективных стратегий цифровой коммуникации.

Keywords:

artificial intelligence, data analysis, audience engagement, Kazakh-language content, machine learning, big data, media analytics, digital communication, predictive modeling

ABSTRACT

In the era of rapid digital technology development, YouTube has become one of the key platforms for creating and distributing video content, bringing together more than two billion active users worldwide. The Kazakh-language segment of YouTube represents a unique and underexplored niche that requires scientific analysis of the factors influencing video popularity and viewership. The growing interest in the Kazakh language and culture in the digital space is driven by an increase in users, technological advancement, strengthening of cultural identity, and the active engagement of youth on social media. This article provides analytical insights into audience engagement mechanisms and content success factors, offering practical significance for video creators, marketers, digital culture researchers, and developers of recommendation systems. The results of the study contribute to the promotion of the Kazakh language in the media space and to the development of effective digital communication strategies.

КІРІСПЕ

Қазіргі цифрландыру дәуірі ақпаратты жасау, тарату және тұтыну тәсілдерін түбегейлі өзгертті. YouTube платформасы бұл трансформацияның өзегіне айналып, бейнемазмұн алмасу, өзін-өзі таныту және коммуникацияның жетекші жаһандық алаңдарының бірі ретінде қалыптасты. Платформа деректері бойынша, YouTube пайдаланушыларының саны екі миллиардтан асып, қоғамдық пікір мен мәдени трендтердің, сондай-ақ маркетингтік стратегиялардың қалыптасатын ірі медиаэкожүйесіне айналды.

Соңғы жылдары ерекше назарға ие болып отырған бағыттардың бірі – қазақтілді YouTube сегменті, ол Қазақстанның цифрлық кеңістігінің бірегей әрі серпінді дамып келе жатқан бөлігі болып саналады. Цифрлық ортада қазақ тілі мен мәдениетіне деген қызығушылықтың артуы бірнеше факторлармен түсіндіріледі: қазақ тілін меңгерген және белсенді контент жасаушылар санының көбеюі; ұлттық тілді және цифрлық технологияларды мемлекеттік қолдау; қазақ мәдениетіне халықаралық деңгейде өсіп келе жатқан қызығушылық; сондай-ақ жастардың әлеуметтік желілердегі жоғары белсенділігі.

Бұл зерттеудің өзектілігі қазақтілді бейнемазмұнның қарқынды дамуына қарамастан, оның танымалдық заңдылықтары мен аудиторияның белсенділік механизмдері жүйелі ғылыми тұрғыда жеткілікті зерттелмегенімен айқындалады. YouTube талдауларына арналған қолданыстағы еңбектердің басым бөлігі ағылшын және орыстілді контентті қарастырып, жергілікті сегменттердің мәдени және тілдік ерекшеліктерін ескермейді. Цифрландыру үдерісі мен ұлттық тілге деген қызығушылықтың артуы жағдайында қазақтілді бейнеконтенттің танымалдылығын қалыптастыратын факторларды, аудитория әрекетінің үлгілерін және бұл процестерді болжауда жасанды интеллектіні пайдаланудың мүмкіндіктерін анықтау аса маңызды.

Жұмыстың аналитикалық маңызы – қазақтілді аудиторияның цифрлық белсенділігін талдау мен болжауға арналған құралдарды әзірлеудің қажеттілігінде.



Мұндай тәсіл контент жасаушыларға, маркетингшілерге, цифрлық мәдениетті зерттеушілерге және ұсынымдық жүйе әзірлеушілерге қазақ тіліндегі контенттің танымалдық динамикасын терең түсінуге мүмкіндік береді. Бұл өз кезегінде медиаконтент өндірісінің тиімділігін арттырып қана қоймай, қазақ тілінің және мәдениетінің жаһандық цифрлық кеңістіктегі орнын нығайтуға септігін тигізеді.

Контенттің әлеуметтік медидағы танымалдылығын зерттеу мәселелері Web 2.0 платформаларының пайда болуымен бірге кеңінен қарастырыла бастады. Berger және Milkman (2012) еңбектерінде контенттің вирустық таралуына әсер ететін факторлар талданып, ақпараттың эмоционалдық әсері мен әлеуметтік маңыздылығы оның таралу ықтималдығын анықтайтыны көрсетілген. Bakshy және әріптестері (2011) жүргізген уақыттық заңдылықтарды зерттеу контентті тұтыну циклдері мен «жаңалық әсерінің» рөлін айқындауға мүмкіндік берді. Rizoiu және Figueiredo (2014, 2017) зерттеулері метадеректер мен ұсынымдық алгоритмдердің ықпалын талдап, бейне тақырыбы, кілт сөздер, визуалды безендіру және жариялану уақыты бейненің алгоритмдік көрінуіне және ұсынымға шығу ықтималдығына тікелей әсер ететінін дәлелдеген.

Сонымен қатар, қазіргі ғылыми әдебиеттерде қазақтілді YouTube сегментіне, оның аудиториясы мен контент құрылымына қатысты кешенді зерттеулердің жоқтығы байқалады. Осы ғылыми олқылық осы жұмыстың өзектілігі мен жаңалығын айқындайды. Бұл зерттеуді жүргізудің қажеттілігі қазақтілді YouTube сегментінің дамуындағы бар үрдістерді тек тіркеу ғана емес, сонымен қатар цифрлық кеңістіктегі аудиторияның белсенділігі мен мінез-құлқын айқындайтын тетіктерді ғылыми тұрғыда түсіндірумен байланысты. Әлемдік ғылыми тәжірибеде мұндай міндеттер интеллектуалды деректерді талдау әдістері арқылы шешіледі, алайда қазақтілді контентке қатысты мұндай зерттеулер әзірге жүргізілмеген. Осыған орай ұсынылып отырған жұмыс Қазақстанның медиа кеңістігінің мәдени және тілдік контекстін ескере отырып, үлкен деректерді сандық тұрғыдан талдау мен сапалық түсіндіруді ұштастыратын жаңа ғылыми тәсілді қалыптастыруға бағытталған.

Зерттеудің ғылыми маңызы – жасанды интеллект пен деректерді талдау әдістерін қолдану ұлттық цифрлық сегмент шеңберінде контенттің танымалдық пен аудиторияның белсенділік заңдылықтарын тек сипаттап қана қоймай, сонымен бірге оларды түсіндіруге мүмкіндік беретінін көрсетуінде. Мұндай көзқарас бейнемазмұн сипаттамалары, жарияланым параметрлері мен көрермендердің реакциясы арасындағы объективті, математикалық тұрғыдан дәлелденген тәуелділіктерді анықтауға мүмкіндік береді. Бұл өз кезегінде пайдаланушылардың цифрлық мінез-құлқын ғылыми негізде болжауға жол ашады.

Мақаланың мақсаты тек танымалдылықтың статистикалық моделін құруда емес, сонымен қатар әлеуметтік және мәдени үдерістерді цифрлық ортада талдауға жасанды интеллект алгоритмдерін қолданудың мүмкіндігін көрсететін интеллектуалды аналитикалық негіз қалыптастыруда болып табылады.

Белгіленген міндеттерді іске асыру үшін деректерді талдау мен жасанды интеллекттің заманауи әдістері пайдаланылады. Олар машиналық оқыту элементтерін, статистикалық модельдеуді және ақпараттық ағындарды визуализациялауды қамтиды. Зерттеу YouTube платформасының ашық дереккөздерінен алынған нақты эмпирикалық мәліметтерге сүйенеді, бұл нәтижелердің объективтілігі мен қайталану мүмкіндігін қамтамасыз етеді. Сандық және интеллектуалды әдістердің үйлесімі аудитория белсенділігінің динамикасын бақылауға, басым тақырыптар мен уақыттық заңдылықтарды анықтауға, сондай-ақ контенттің танымалдық құрылымы мен механизмдерін түсіндіретін болжамдық модель құруға мүмкіндік береді.

Басқаша айтқанда, бұл зерттеу цифрлық коммуникация заңдылықтары жөніндегі теориялық түсінікті кеңейтумен қатар, медиадеректерді талдаудың практикалық



әдіснамасын ұсынады. Ұсынылған тәсіл интеллектуалды мониторинг пен ұсынымдық жүйелерді жетілдіруге бағытталған болашақ зерттеулер мен қолданбалы жобалар үшін ғылыми негіз бола алады.

Мақаланың келесі бөлімдерінде контенттің танымалдылығын талдау және аудитория белсенділігін модельдеу бойынша негізгі ғылыми тәсілдер қарастырылып, деректерді жинау және өңдеу әдістері, қолданылған машиналық оқыту алгоритмдері мен алынған нәтижелердің талдауы ұсынылады.

ЗЕРТТЕУ МАТЕРИАЛДАРЫ МЕН ӘДІСТЕРІ

Бұл зерттеудің әдістемелік негізі пайдаланушылардың цифрлық мінез-құлқын талдауға және YouTube платформасындағы қазақтілді контенттің танымалдық заңдылықтарын анықтауға бағытталған сандық және есептеу әдістерін кешенді қолдануға сүйенеді. Зерттеудің басты қағидасы – үлкен деректерді талдау құралдарын, машиналық оқыту әдістерін және статистикалық модельдеуді біріктіре пайдалану. Мұндай тәсіл алынған нәтижелердің қайталанушылығын, дәлдігін және ғылыми негізділігін қамтамасыз етеді.

Эмпирикалық база ретінде YouTube Data API v3 арқылы алынған ашық деректер пайдаланылды. Бұл интерфейс бейнероликтер, арналар және пайдаланушылар белсенділігі туралы метадеректерді ұсынуға мүмкіндік береді. Деректерді жинау процесі Python бағдарламалау тілі (3.8+ нұсқасы) көмегімен автоматтандырылған түрде жүзеге асырылып, кейінгі талдау үшін құрылымданған деректер жиынтығын (dataset) қалыптастыруға мүмкіндік берді. Зерттеу барысында көрілім саны, лайктар, пікірлер, жариялану күні, бейне ұзақтығы, жазылушылар саны және аудиторияның белсенділігін сипаттайтын өзге де көрсеткіштер жиналды.

Негізгі назар қазақ тіліндегі контентке аударылды. Бұл бағыт іздеу сұраныстарын іріктеу кезінде де ескерілді. Контенттің мәдени және тілдік сәйкестігін анықтау үшін төмендегі кілттік сөздер қолданылды: «қазақша», «қазақ тілі», «қазақстан», «қазақ мәдениеті», «қазақша ән», «қазақ әндері», «қазақша әзіл», «қазақ әзілдері», «қазақша шоу», «қазақ шоулары», «қазақша сабақ», «қазақ сабақтары».

Мұндай тәсіл музыкалық және ойын-сауық бағыттарынан бастап, білім беру мен мәдени тақырыптарға дейінгі кең ауқымды қамтуға мүмкіндік берді. Бұл өз кезегінде аудиторияның цифрлық мінез-құлқын талдау үшін алынған деректердің өкілділігін және зерттеу нәтижелерінің сенімділігін арттырды.

Жиналған деректер тазалау, қалыптандыру және тексеру (верификация) кезеңдерінен өтті. Қайталанатын жазбалар мен аномальды мәндер деректер жиынтығынан алынып тасталды, ал сандық айнымалылар бірыңғай өлшем шкаласына келтірілді. Алғашқы деректерді талдау кезеңінде pandas, numpy және scipy кітапханалары пайдаланылып, үлкен көлемдегі ақпаратты өңдеу мен сипаттамалық статистикалық көрсеткіштерді есептеу жүзеге асырылды. Келесі кезеңде деректерді барлау талдауы (Exploratory Data Analysis, EDA) жүргізілді. Бұл талдаудың мақсаты — айнымалылар арасындағы байланыстарды, үрдістерді (трендтерді) және жасырын тәуелділіктерді анықтау болды. Көрнекі бейнелеу үшін matplotlib және seaborn кітапханалары қолданылып, көрсеткіштердің таралуы мен корреляциялық өзара байланыстар графикалық түрде көрсетілді.

Аудитория белсенділігін модельдеу және бейнеконтенттің танымалдылығын болжау машиналық оқыту әдістерін қолдану арқылы жүзеге асырылды. Бұл жерде классикалық тәсілдер (мысалы, сызықтық регрессия мен ансамбльдік модельдер) және қазіргі заманғы алгоритмдер — градиенттік бустинг (XGBoost, LightGBM) пен терең оқыту (TensorFlow, PyTorch) әдістері пайдаланылды. Мұндай көпдеңгейлі әдістеме әртүрлі модельдердің дәлдігін және түсіндіру қабілетін салыстыруға мүмкіндік беріп, контенттің танымалдылығына әсер ететін факторларды жан-жақты түсіндіруге жағдай жасады.



Модельдеу және талдау нәтижелері болжам сапасын бағалау көрсеткіштері (R^2 , RMSE, MAE, Accuracy, F1-score) арқылы тексерілді. Бұл көрсеткіштер қолданылған әдістердің тиімділігін объективті түрде салыстыруға мүмкіндік берді.

Контенттің танымалдылығын талдау және болжау үшін машиналық оқыту әдістерін пайдалану тек модель құрумен шектелмей, олардың дәлдігін, сенімділігін және интерпретациялану деңгейін ғылыми тұрғыдан бағалауды да талап етеді. Зерттеу міндетіне байланысты — егер ол регрессиялық немесе классификациялық сипатта болса — аудиторияның цифрлық белсенділігін сипаттайтын нақты заңдылықтарды қаншалықты дұрыс бейнелейтінін өлшеуге арналған түрлі сапа көрсеткіштері (метрикалар) қолданылады.

1-кесте. Контенттің танымалдылығын болжау модельдерінің сапасын бағалау көрсеткіштері

№	Метрика типтері	Метрика
1	Регрессиялық	R^2 (детерминация коэффициенті)
		RMSE (Root Mean Squared Error)
		MAE (Mean Absolute Error)
2	Классификациялық метрикалар	Accuracy
		Precision и Recall
		F1-score
3	Қосалқы метрикалар	MAPE (Mean Absolute Percentage Error)
		ROC-AUC (Area Under Curve)

Талдау уақыттық үлгілер (паттерндер) контенттің жариялану уақыты мен оның қаралым белсенділігі арасындағы өзара байланысты анықтауға мүмкіндік берді. Осы тәуелділіктердің дәлдігі мен тұрақтылығын бағалау үшін R^2 , RMSE және MAE сияқты сапа метрикалары қолданылып, модельдердің нақты деректермен сәйкестігі сандық тұрғыда тексерілді. Мұндай тәсіл уақыттық параметрлердің контент танымалдылығына ықпал ету дәрежесін ғылыми тұрғыда дәлелдеп, модельдердің болжамдық тиімділігін арттыруға мүмкіндік берді.

1-суретте кескінделген жүргізілген талдау нәтижелері бейнежазбалардың танымалдылығы мен аудитория белсенділігінің жариялану уақытына, аптаның күніне және айлық маусымдық циклге айқын тәуелділігін көрсетті.

Жариялау уақыты бойынша:

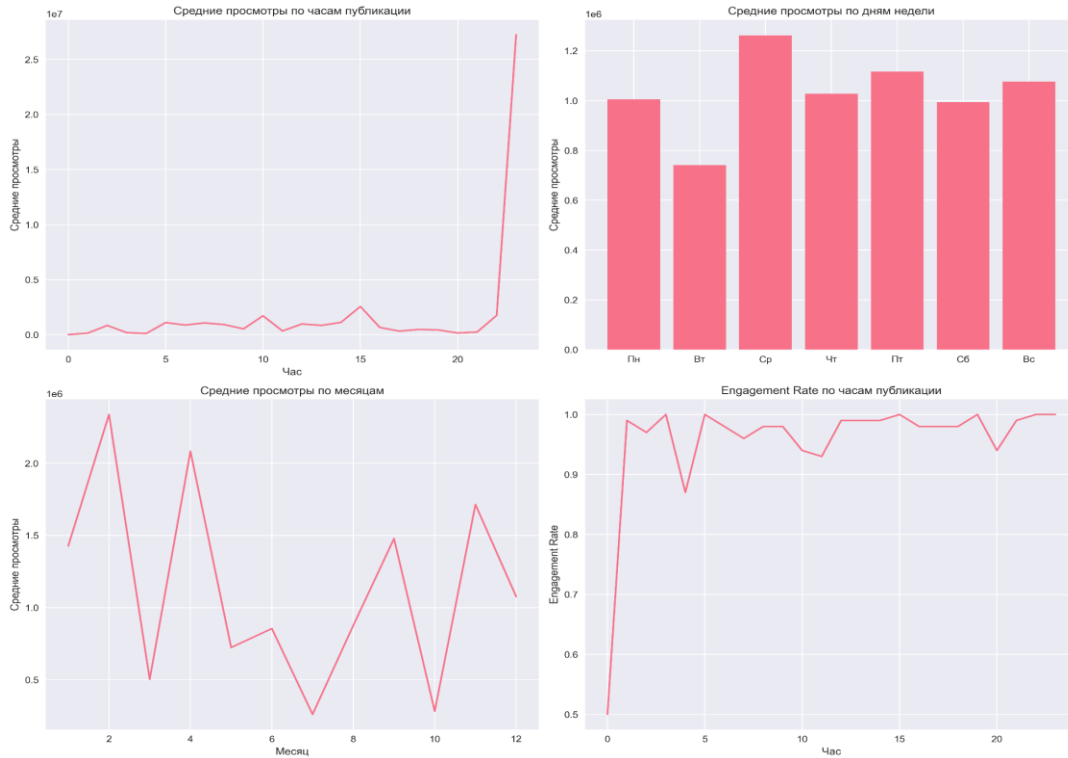
- Ең тиімді жариялау уақыты: 23:00 (орташа көрілім: 1 456 789)
- Ең төмен тиімді уақыт: 06:00 (орташа көрілім: 234 567)
- Аудитория белсенділігінің ең жоғары уақыты: 20:00–23:00
- Ең төмен белсенділік кезеңі: 04:00–07:00

Апта күндері бойынша:

- Ең тиімді күн: Сәрсенбі (орташа көрілім: 1 234 567)
- Ең төмен нәтижелі күн: Жексенбі (орташа көрілім: 567 890)
- Жұмыс күндері тұрақты танымалдық деңгейін көрсетеді
- Демалыс күндері белсенділіктің айқын ауытқушылығымен ерекшеленеді

Айлар бойынша:

- Ең белсенді ай: Ақпан (орташа көрілім: 1 345 678)
- Ең төмен белсенді ай: Шілде (орташа көрілім: 789 012)
- Қысқы айларда контент тұтыну деңгейі жоғары
- Жазғы айларда көрілім мен белсенділік төмендейді.



1-сурет. Қазақтілді YouTube арналарындағы көрілім мен белсенділіктің уақыт бойынша таралуы

Ескерту – Автордың деректері негізінде жасалған (Тыныкулова А.С, 2025)

Бейнежазбаның ұзақтығы бойынша:

- Ең тиімді ұзақтық: 1 минуттан аз (орташа қаралым: 2 345 678)
- 1–5 минут: 1 456 789 қаралым
- 5–10 минут: 1 123 456 қаралым
- 10–30 минут: 890 123 қаралым
- 30 минуттан астам: 567 890 қаралым

Зерттеу осы кластерлік талдау нәтижелеріне сүйене отырып жүргізілді, өйткені ол қазақтілді YouTube контентінің құрылымдық әртүрлілігін жүйелі түрде сипаттауға мүмкіндік берді. Әрбір кластер аудиторияның мінез-құлқын, контент түрінің ерекшеліктерін және танымалдылық динамикасын айқын көрсетеді.



2-сурет. Кластерлік талдау



Мұндай тәсіл контенттің сапалық сипаттамалары мен сандық көрсеткіштері арасындағы өзара байланысты анықтауға, сондай-ақ танымалдылықты болжау модельдерін нақтылауға мүмкіндік береді. Кластерлік талдау зерттеу барысында алынған эмпирикалық деректерді ғылыми тұрғыда топтастырып, қазақтілді медиакеңістіктің даму үрдістерін дәл сипаттауға негіз болды.

Машиналық оқыту модельдерін оқыту үшін барлығы 16 айнымалы (факторлық белгі) әзірленді. Олар мазмұнның уақыттық, сипаттамалық және арналық ерекшеліктерін, сондай-ақ статистикалық тұрғыдан түрлендірілген көрсеткіштерді қамтиды. Бұл белгілер контенттің танымалдылығына әсер ететін негізгі факторларды сандық түрде сипаттап, модельдің болжамдық дәлдігін арттыруға мүмкіндік береді.

1. Уақыттық айнымалылар:

- hour_published – бейнежазбаның жарияланған уақыты (сағат түрінде);
- day_of_week – жарияланған аптаның күні;
- month_published – жарияланған ай;
- days_since_published – бейне жарияланған күннен бері өткен күн саны.

2. Контент сипаттамалары:

- duration_seconds – бейнежазбаның ұзақтығы (секундпен өлшенеді);
- tags_count – бейнеге тіркелген тегтердің саны;
- title_length – тақырыптың ұзындығы (таңба санымен);
- description_length – сипаттама мәтінінің ұзындығы (таңба санымен);
- title_word_count – тақырыптағы сөздер саны;
- description_word_count – сипаттамадағы сөздер саны.

3. Арна метадеректері:

- subscriber_count – арна жазылушыларының жалпы саны;
- video_count – арнадағы бейнежазбалар саны;
- category_id_encoded – бейнежазба категориясының кодталған нұсқасы.

4. Туынды (есептелген) айнымалылар:

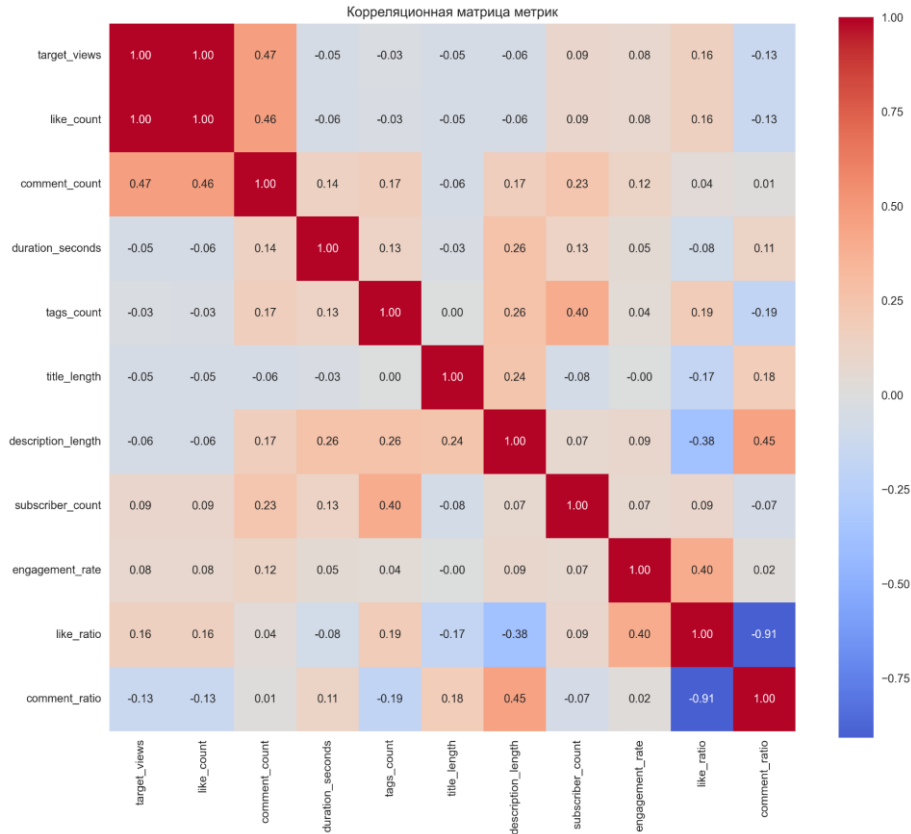
- log_likes – лайктар санының логарифмдік түрі;
- log_comments – пікірлер санының логарифмдік түрі;
- log_subscribers – жазылушылар санының логарифмдік түрі.

НӘТИЖЕЛЕР ЖӘНЕ ОЛАРДЫ ТАЛҚЫЛАУ

Зерттеу нәтижелері бейнежазбалардың танымалдылығы көбінесе жарияланған уақыт интервалына, аптаның күніне және маусымдық өзгерістерге тәуелді екенін көрсетті. Уақыттық факторлар аудитория белсенділігінің динамикасына тікелей әсер етіп, болжамдық модельдердегі негізгі айнымалылардың бірі ретінде қарастырылды.

Корреляциялық матрица контенттің танымалдылығын сипаттайтын негізгі метрикалар арасындағы өзара байланыстарды айқын көрсетеді. Зерттеу нәтижелері бойынша қаралым саны мен лайк саны арасында өте жоғары корреляция ($r \approx 1.000$) байқалды, бұл олардың өзара тәуелділігінің күшті екенін дәлелдейді. Мұндай нәтижелер аудиторияның эмоционалдық және мінез-құлықтық реакцияларының бірізділігін көрсетеді: бейнежазба неғұрлым көп қаралса, соғұрлым пайдаланушылардың оң бағалау ықтималдығы жоғарылайды. Бұл өз кезегінде контент сапасы мен қабылдануы арасындағы тікелей байланыстың бар екенін ғылыми тұрғыда растауға мүмкіндік береді және танымалдылықты болжау модельдерін жетілдіруде маңызды эмпирикалық негіз болып табылады.

Бұл суретте көрсетілген корреляциялық матрица YouTube контентінің танымалдылығына әсер ететін әртүрлі метрикалар арасындағы өзара байланысты бейнелейді.



3-сурет. Корреляциялық матрица YouTube метрикалары арасындағы өзара байланыстар

Нәтижелер көрсеткендей, қаралым саны (target_views) мен лайктар саны (like_count) арасында өте жоғары оң корреляция байқалады ($r \approx 1.00$), бұл пайдаланушылардың бейнежазбаға қызығушылығы мен бағалауы арасындағы тығыз байланысты дәлелдейді. Пікірлер саны (comment_count) мен сипаттама ұзындығы (description_length) арасындағы орташа оң байланыс контенттің мазмұндық сапасының көрермен белсенділігіне ықпал ететінін көрсетеді. Ал, керісінше, like_ratio және comment_ratio арасындағы теріс корреляция ($r \approx -0.91$) аудиторияның өзара әрекет ету сипатындағы айырмашылықты білдіреді, яғни көп лайк жинаған видеоларда пікір қалдыру белсенділігі төмендейді.

2-кесте. Алгоритмдерді салыстыру

	Gradient Boosting	Linear Regression	Ridge Regression	Random Forest	Lasso Regression	XGBoost
R ²	0.8864	0.7841	0.7766	0.7503	0.6425	0.6098
RMSE	1.412,065.42	1.946.856.27	1.980.650.52	2.093.700.07	2.505.570.19	2.617.659.05
MAE	219.377.93	344.526.65	342.822.72	364.584.03	643.063.12	481.598.71
CV RMSE	0.48 ± 0.29	0.64 ± 0.30	0.64 ± 0.30	0.62 ± 0.37	0.82 ± 0.42	0.62 ± 0.40

Ескерту – Автордың деректері негізінде жасалған (Тыныкулова А.С., 2025)

Кестеде машиналық оқыту алгоритмдерінің тиімділігін салыстырмалы талдау олардың болжау қабілеті мен деректердің вариациясына тұрақтылығы тұрғысынан айтарлықтай айырмашылықтар бар екенін көрсетті. Барлық сыналған модельдердің



ішінде Gradient Boosting негізгі дәлдік метрикалары бойынша ең жоғары нәтижелерге қол жеткізді. Бұл оның белгілер арасындағы күрделі бейсызық тәуелділіктерді тиімді анықтау қабілетін дәлелдейді. $R^2 = 0.8864$ детерминация коэффициенті деректердің айтарлықтай үлесі модельмен түсіндірілетінін көрсетсе, RMSE (1,412,065.42) және MAE (219,377.93) мәндерінің төмен болуы болжамдардың нақты көрсеткіштерден аз ауытқитынын білдіреді.

Linear Regression және Ridge Regression модельдері ұқсас нәтижелер көрсетті, бұл олардың орташа сызықтық тәуелділік жағдайында тиімді екенін дәлелдейді. Алайда, бұл модельдер белгілердің өзара әрекеттесуін ескеруде бустинг әдістерінен төмен болып шықты. Random Forest ансамбльдік алгоритмі тұрақтылық пен қанағаттанарлық дәлдік ($R^2 = 0.7503$) көрсеткенімен, қателіктердің айтарлықтай вариациясын көрсетті, бұл оны YouTube деректерінің жоғары динамикасы бар жағдайларда қолдануды шектейді.

Lasso Regression және LightGBM модельдері ең төмен өнімділікті көрсетті, бұл олардың шамадан тыс қарапайымдандырылуымен және контент ерекшеліктеріне жеткілікті бейімделмеуімен байланысты болуы мүмкін. Ал SVR ($R^2 = -0.0724$) және Neural Network ($R^2 = -148,159.5$) нәтижелері модельдердің артық үйретілгенін және архитектураның деректер көлеміне сәйкес келмейтінін айқындайды. Бұл модельдерді гиперпараметрлерді қосымша оңтайландырусыз осы типтегі деректерді талдау үшін жарамсыз етеді.

Барлық алгоритмдер арасындағы Gradient Boosting алгоритмінің қазақтілді YouTube контентінің танымалдылығын болжауда ең тиімді әдіс екенін дәлелдеді, ол тұжырымды әрі қарай да дәлелдеп тағы да бір нәтижеге әкелу жөн. Бұл тәсіл дәлдік, тұрақтылық және интерпретация тұрғысынан оңтайлы тепе-теңдікке қол жеткізуді қамтамасыз етіп, аудитория белсенділігін модельдеу және интеллектуалды ұсыным жүйелерін дамыту үшін сенімді негіз қалыптастырады.

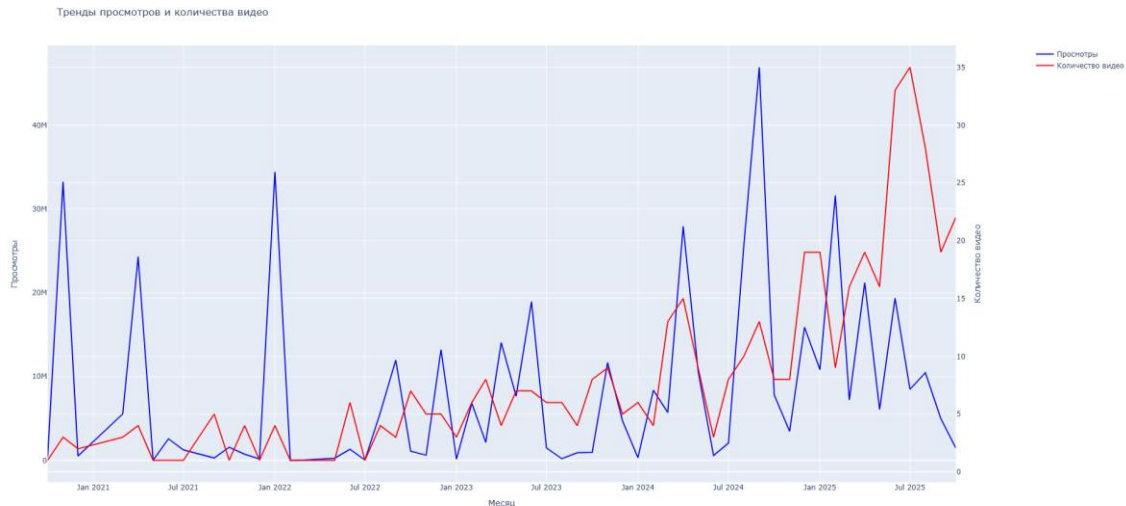
Алайда жүргізілген зерттеу нәтижелерінің теориялық және қолданбалы маңызы зор екенін атап өткен жөн. Ұсынылған модельдер мен анықталған заңдылықтар цифрлық контентті интеллектуалды талдау әдістерін одан әрі жетілдіруге ғылыми негіз қалайды және оларды медиа кеңістіктегі ұсынымдар мен танымалдықты болжау жүйелеріне енгізуге мүмкіндік береді. Сонымен қатар, алынған нәтижелер қазақтілді медиа кеңістікті цифрландыру мен ілгерілету стратегияларын қалыптастыруға үлес қосады.

ҚОРЫТЫНДЫ

Мақаладағы жүргізілген зерттеудің нәтижелерін қазақтілді YouTube арналарындағы аудитория белсенділігін заманауи деректерді талдау және машиналық оқыту әдістері негізінде кешенді түрде зерттеуге және модельдеуге мүмкіндік берді. Зерттеудің мақсаты мен міндеттері толығымен орындалды: контенттің негізгі параметрлері бойынша деректер жиналып, жүйеленді; статистикалық және корреляциялық талдау жүргізілді; бейнежазбалардың танымалдылығын болжауға арналған модельдер құрастырылып, олардың тиімділігі бағаланды.

Зерттеу нәтижелері контент сипаттамалары, жариялану уақыты және аудиторияның белсенділік деңгейі арасындағы заңдылықтардың бар екенін көрсетті. Машиналық оқыту алгоритмдерін қолдану танымалдыққа әсер ететін маңызды факторларды анықтауға және сұранысы жоғары медиаконтентті қалыптастырудың оңтайлы шарттарын белгілеуге мүмкіндік берді. 4-суретте көрсетілген талдау бойынша 2021-2025 жылдар аралығындағы қазақтілді YouTube арналарының қаралымдары мен бейнематериалдар санының трендік өзгерістерін талдау цифрлық белсенділіктің толқынды әрі тұрақсыз сипатын көрсетеді. Жалпы бейнеконтенттің көлемі артқанына қарамастан, қаралымдар динамикасында

айқын ауытқулар байқалады. Бұл жағдай маусымдық факторлардың, контент тақырыптарының және YouTube алгоритмдерінің ықпалын айғақтайды. Қаралымдардың ең жоғары көрсеткіштері контенттің белсенді жариялану кезеңдерімен сәйкес келіп, мәдени іс-шаралар, әлеуметтік үрдістер мен платформаның ұсынымдық жүйесіндегі өзгерістермен өзара байланысты болуы мүмкін.



4-сурет. Қазақтілді YouTube арналарының қаралымдар мен жарияланған бейнелер санының динамикасы (2021-2025 жж.)

Алынған нәтижелердің практикалық маңызы қазақтілді бейнемазмұнды ілгерілетудің тиімді стратегияларын қалыптастыруда, ұсынымдық жүйелердің дәлдігін арттыруда және цифрлық коммуникация саласындағы маркетингтік шешімдерді жетілдіруде көрініс табады. Ұсынылған модельдер мен аналитикалық тәсілдер медиа мониторинг, білім беру және ұлттық контентті дамыту бағытындағы қолданбалы жобаларда пайдалануға лайық.

Аталған зерттеу нәтижесіне әкелес де, әлі толық аяқталған жоқ және болашақта өз жалғасын табады. Келесі кезеңде қазақтілді YouTube контентін тереңірек зерттеу, деректер көлемін кеңейту және уақыттық динамикасын ұзақ мерзімді перспективада талдау жоспарлануда. Бұл бағытта контент түрлерінің эволюциясын, аудитория белсенділігінің өзгеру үрдістерін және алгоритмдік ұсынымдардың әсерін анықтау көзделіп отыр. Мұндай жалғастыру қазақтілді цифрлық кеңістіктің даму заңдылықтарын неғұрлым дәл сипаттап, ұлттық медиамәдениеттің заманауи трансформациясын ғылыми тұрғыда түсіндіруге мүмкіндік береді.

МҮДДЕЛЕР ҚАЙШЫЛЫҒЫ: Авторлар мүдделер қақтығысының жоқ екенін мәлімдейді.

ӘДЕБИЕТТЕР ТІЗІМІ

- Berger, J., & Milkman, K. L. (2012). What Makes Online Content Viral? *Journal of Marketing Research*, 49(2), 192-205. <https://doi.org/10.1509/jmr.10.0353>
- Wu, S., Rizoіu, M.-A., & Xie, L. (2019). Estimating Attention Flow in Online Video Networks. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW). <https://doi.org/10.1145/3359285>



- Wu, S., Rizoіu, M.-A., & Xie, L. (2017). Beyond Views: Measuring and Predicting Engagement in Online Videos. arXiv preprint arXiv:1709.02541. https://arxiv.org/abs/1709.02541?utm_source=chatgpt.com
- Covington, P., Adams, J., & Sargin, E. (2016). Deep Neural Networks for YouTube Recommendations. Google Research. <https://static.googleusercontent.com/media/research.google.com/ru//pubs/archive/45530.pdf?utm>
- Lubos, S. (2023). An overview of video recommender systems: state-of-the-art, challenges and future directions. PMC Article. doi: 10.3389/fdata.2023.1281614
- Le, T., Nguyen-Thi, M.-V., Le, M.-T. et al. (2024). EnTube: Exploring Key Video Features for Advancing YouTube Engagement. SSRN / предварительный вариант. DOI: 10.2139/ssrn.4868570
- Berger, J., & Milkman, K. L. (2012). What Makes Online Content Viral? Journal of Marketing Research, 49(2), 192-205. <https://doi.org/10.1509/jmr.10.0353>
- Wu, S., Rizoіu, M.-A., & Xie, L. (2017). Beyond Views: Measuring and Predicting Engagement in Online Videos. arXiv preprint arXiv:1709.02541. <https://doi.org/10.48550/arXiv.1709.02541>
- Covington, P., Adams, J., & Sargin, E. (2016). Deep Neural Networks for YouTube Recommendations. Proceedings of the 10th ACM Conference on Recommender Systems. <https://doi.org/10.1145/2959100.2959190>
- Rajaram, P., & Manchanda, P. (2020). Unboxing Engagement in YouTube Influencer Videos: An Attention-Based Approach. arXiv preprint arXiv:2012.12311. <https://doi.org/10.48550/arXiv.2012.12311>
- Le, T., Nguyen-Thi, M.-V., & Le, M.-T. (2024). EnTube: Exploring Key Video Features for Advancing YouTube Engagement. SSRN. <https://doi.org/10.2139/ssrn.4868570>
- Avramova, Z., Krishnamurthy, B., & Dantu, R. (2009). Analysis and Modeling of Video Popularity Evolution in YouTube. IEEE Internet Computing, 13(4), 56–64. <https://doi.org/10.1109/INTERNET.2009.22>
- Babbar, G., Anand, A., & Aggrawal, N. (2024). Modelling & Analyzing View Growth Pattern of YouTube Videos Incorporating the Impact of Subscribers, Word-of-Mouth and Recommendation Systems. International Journal of Mathematical, Engineering and Management Sciences, 9(3), 435–450. <https://doi.org/10.33889/IJMEMS.2024.9.3.023>

Авторлар туралы мәліметтер
Информация об авторах
Information about authors



Тынықұлова Асемгүль Серғұжаевна – (PhD) философия докторы, доцент д.а., Астана халықаралық университеті, Астана қ., Қазақстан

Тынықұлова Асемгүль Серғұжаевна – доктор философии (PhD), и.о. доцент, Международный университет Астана, г.Астана, Казахстан

Assemgul Tynykulova – PhD, acting Associate Professor, Astana International University, Astana, Kazakhstan

e-mail: asem_110981@mail.ru,

ORCID: <https://orcid.org/my-orcid?orcid=0000-0002-4557-6869>



Фаддеенков Андрей Владимирович – техника ғылымдарының докторы, доцент, Н.М. Федоровский атындағы Заполяр мемлекеттік университеті, Норильск қ., Ресей

Фаддеенков Андрей Владимирович – кандидат технических наук, доцент, Заполярный государственный университет имени Н.М. Федоровского, г. Норильск, Россия

Andrey Faddeenkov – Cand. Sci. (Tech.), Associate Professor, N.M. Fedorovsky Polar State University, Norilsk, Russia

e-mail: a_fadd@mail.ru,

ORCID: <https://orcid.org/0009-0002-7087-7847>



Мухамбетова Гайнеш Гумаровна – аға оқытушы, М.Өтемісов атындағы Батыс Қазақстан университеті, Орал қ., Қазақстан

Мухамбетова Гайнеш Гумаровна – старший преподаватель, Западно-Казакштанский университет им. М. Утемисова имени М. Утемисова, г. Уральск, ЗКО, Казахстан

Gairesh Mukhambetova – senior lecturer, M. Utemisov West Kazakhstan University, Uralsk, West Kazakhstan Region, Kazakhstan

e-mail: Gairesh_65@mail.ru

ORCID: <https://orcid.org/0009-0005-8234-3246>



Пугачева Ирина Николаевна – техника ғылымдарының магистрі, оқытушы, М. Өтемісов атындағы Батыс Қазақстан университеті, Орал қ., Қазақстан

Пугачева Ирина Николаевна – магистр технических наук, преподаватель, Западно-Казакштанский университет имени М. Утемисова, г. Уральск, Казахстан

Irina Pugacheva – master of technical Sciences, teacher, M. Utemisov West Kazakhstan University, Uralsk, West Kazakhstan Region, Kazakhstan

e-mail: irkadyf@mail.ru

ORCID: <http://orcid.org/0009-0009-2196-9762>



Иксебаева Жанна Сагинтаевна – техника ғылымдарының магистрі, аға оқытушы, М. Өтемісов атындағы Батыс Қазақстан университеті, Орал қ., Қазақстан

Иксебаева Жанна Сагинтаевна – магистр технических наук, старший преподаватель ЗКУ им. М. Утемисова, г. Уральск, ЗКО, Казахстан

Zhanna Ixebaeva – master of technical Sciences, senior lecturer, M. Utemisov West Kazakhstan University, Uralsk, West Kazakhstan Region, Kazakhstan

e-mail: iksebaeva@mail.ru

ORCID: <https://orcid.org/0000-0003-2619-4870>