

ӘОЖ 004.852  
ҒТАХР 28.23.25

<https://doi.org/10.55452/1998-6688-2024-21-4-91-106>

<sup>1</sup>Сапақова С.З.,

физ.-мат.ғ.к., қауымдастырылған профессор, ORCID ID: 0000-0001-6541-6806,  
e-mail: sapakovasz@gmail.com

<sup>2</sup>Қамбарова Б.,

докторант, ORCID ID: 0009-0008-6153-4934,  
e-mail: gabitqyzy96@mail.ru

<sup>1</sup>Халықаралық ақпараттық технологиялар университеті, Алматы қ., Қазақстан

<sup>2</sup>Қазақ ұлттық қыздар педагогикалық университеті, Алматы қ., Қазақстан

## МАШИНАЛЫҚ ОҚЫТУМЕН ӘЛЕУМЕТТІК ЖЕЛІДЕГІ ҚАЗАҚ ТІЛІНІҢ ДАМУЫНА ОРЫС ТІЛІНДЕГІ СӨЗДЕРДІҢ ЫҚПАЛЫН ЗЕРТТЕУ

### Аңдатпа

Бұл зерттеу қазақ тілінің шұбарлануына орыс тіліндегі сөздердің ықпалын зерттеуге арналған. Ақпараттық технологияның қарқындап дамуы әлеуметтік желі тілінің жай-күйіне әсер етеді. Біріншіден, желідегі бейберекетсіздік тілдің қолданыс аясын тежеп, шұбарлануға алып келсе, екіншіден, даму динамикасы жоғары, қазақ тіліндегі ақпараттар да жылдам тарап жатыр деуге болады. Мақаланың негізгі мақсаты қазіргі заманғы интернеттегі шетелдік сөздер, бір тілден екінші тілге ешқандай өзгеріссіз ауысатын тілдік операциялар, аралас жазбалар (шетелдік сөздердің лексикалық және семантикалық мәртебесі сақталған сөздерді қоса алғанда), жаңа қысқартулар, варваризмдердің қазақ тіліне әсерін зерттеу. Зерттеу нәтижесінде әлеуметтік желідегі сөйлеу формалары айқындалады; қарым-қатынастың ауызша түрінде ойлау мен ойды тіл арқылы жарыққа шығару сәтінің бір уақытта жүзеге асырылуына байланысты тілдің шұбарлануына ықпал ететін факторлардың ықтималдығы анықталады. Машиналық оқыту алгоритмдерінің, атап айтқанда Naive Bayes классификаторының қолданылуы деректерді талдау процесін автоматтандырып, жасырын заңдылықтарды анықтады, бұл осы кірме сөздердің қазақ тіліне цифрлық ортада қалай әсер ететінін тереңірек түсінуге мүмкіндік берді. Ақпараттық технологияның прогрессивті даму үрдісімен ерекшеленетін жаһандық платформалар аясында Facebook және Instagram желілеріндегі мәтін мазмұны сарапталды.

**Тірек сөздер:** коммуникация, прагматика, лингвостилистика, әлеуметтік желі, машиналық оқыту, алгоритмдер.

### Кіріспе

Бүгінде IT саласындағы жетістіктер, цифрландыру процестері, мемлекеттік порталдар, медиаканалдар, жасанды интеллектінің дамуы, әлеуметтік желілердің көбеюі тілдік жүйеге әсер етті. Әлеуметтік желілер қазіргі қоғамда тек ақпарат алмасудың негізгі құралдары ғана емес, әлемде болып жатқан түрлі оқиғаларға пікір білдіріп, талдайтын, тақырыпты дискурсқа айналдыратын платформаға айналды. 2024 жылдың қаңтарында Қазақстанда 14,10 миллион белсенді әлеуметтік желі қолданушысы болған. Соңғы бір жылда олардың саны 34,9%-ға өсті. Бұны Қазақстан Республикасындағы барлық әлеуметтік желіні пайдаланушылардың санына теңестіруге болады. Оның ішінде, instagram-да – 66,5% (12,1 млн) және Facebook-та – 14,3% (2,6 млн). Бейресми деректер бойынша, TikTok – елдегі ең танымал әлеуметтік желі. Қысқа бейнероликтерді көруге арналған платформада 14 миллионнан астам қазақстандық тіркелген [1]. Демек әлеуметтік желідегі қазақша жазу, қазақ тіліндегі контенттер – тілдің дамуы үшін маңызды. Бірақ ондағы жазылатын мәтіндер мен пікірлердің мазмұны әртүрлі. Мұнда адамдардың ішкі мәдениетін көрсететін, сөз бостандығын айқындайтын алаң қалыптасқан. Қазір қолданыстағы тіл ғана тірі және байиды деп есептесек, қазақ тілі өзінің тілдік нормасынан ажырап, өзге тілдердің араласуынан шұбарлану көбейді. XX ғасырдағы ауызекі

тіл мен жазба мәтіндерінде орыс сөздерінің орынсыз қолданылуы қазіргі кезде әлеуметтік желідегі жазба мәтіндерінде кездесе бастады. Бұл феномен тілдің табиғи дамуына және оның таза сақталуына әсер етуі мүмкін. Әлеуметтік желілерде орыс тіліндегі варваризмдердің кең таралуына ықпал болатын бірнеше себептер:

1. Интернеттің халықаралық әсері: әлеуметтік желілер мен интернеттің жаһандануы басқа тілдердің сөздерін кеңінен енгізуге мүмкіндік береді.

2. Сөздердің қысқартылуы мен өзгертілуі: әлеуметтік желілердегі жылдам коммуникация қажеттілігі варваризмдердің жиі қолданылуына әкеп соқтырады. Бұл сөздердің қысқартылуы мен өзгертілуі түрінде көрінеді.

3. Трендтер мен мода: жаңа сөздер мен сөз тіркестерінің трендке айналуы олардың әлеуметтік желіде көп қолданылуына әкеледі. Мысалы: жиза, охват, тема, ссылка, direct, coment, please, screen, podcast, respect, image, storis.

Тіл – ойлаудың жемісі. Адамзат тіл мен ойлаудың біртұтас екенін әлдеқашан дәлелдеген. Анар Салқынбай: «Дұрыс ойлайтын кісі дұрыс жазады. Қазақ тілінде бұрыс сөйлейтіндер әлеуметтік желіде де қате жіберуі заңдылық. Норманың бұзылуы осыдан басталмақ, бұл әлеуметтік желіні қолданушының ішкі мәдениетіне, өзін бағалауына, өзі арқылы тілін, ұлтын бағалауымен астасып жатыр», – дейді [2]. Тіл мамандары ғаламтордағы тілдің қолданылуына қатысты зерттеулерінің нәтижесінде қолданбалы лингвистикадағы бұл бағыттың қалыптасуының алғышарттарын атап қана қоймай, жеке пән ретінде қарастыра бастады. АҚШ ғалымы Н.С. Барон 1984 ж. «Computer Mediated Communication as a Force in Language Change» мақаласын жарияласа [3], 2010 ж. терең талдай келе «Always on: language in an online and mobile world» деген монографиясын шығарды [4]. Ғаламтордың ағылшын тіліне әсерін зерттеген Д. Кристал «Internet and language» еңбегінде интернет тілін жеке ғылыми бағыт деп тұжырымдаған [5]. О. Панюшкина әлеуметтік желілердегі тілдердің өзгерісін испан тілі негізінде қарастырса [6], Л.Ю. Иванов ғаламтордың тілдік ерекшеліктерін және ондағы болған өзгерістерді жіктеді [7].

Ғалымдардың әлеуметтік желі лингвистикасына деген қызығушылығының артуына қарамастан, интернет лингвистиканың көптеген аспектілері әлі де жеткілікті түрде түсініліп, сипатталған жоқ. Бірыңғай терминологиялық жүйе, зерттеу әдістемесі жасалмаған. Әсіресе, әлеуметтік желідегі қазақ тілінің жай-күйі туралы қанша айтылып, жазылып жүрсе де, тілдің шұбарлануына алып келетін себептер, жастар тіліндегі орыс және ағылшын сөздерінің қазақ тіліне ықпалы толық анықталмады. Бірақ «жастардың өз сленгі болуы да заңды. Әлеуметтік желіде жастар арасында осы сленгтер қолданысқа ие болуы да ықтимал. Бұл бұрыс қолданыс болуы мүмкін, бірақ, уақытша құбылыс» және «әлеуметтік желіні тым жиі пайдаланатын жастардың қысқарта жазуы әдеби тіл даму бағытының көрсеткіші, индикаторы бола алмайды. Әлеуметтік желінің тілі қазақ әдеби тілі дамуының, я кері кетуінің басты көрсеткіші емес, бұл тек ауызекі сөйлеу тілінің бір сипаты ғана. Бұл әлемдік үдеріс – цифрландырудың ықпалы», – дейді Анар Салқынбай [2]. Алайда шет тіл білімінде ғаламтордағы тілдің қолданылуын зерттеу ондаған жылдарға созылып келеді. Ал қазақ тіл білімінде бұған қатысты зерттеулер енді ғана басталды. Ұдайы өзгерістерге толы цифрлық әлемдегі қазіргі қазақ тілінің жай-күйін қарастыру – қазақ тілінің технологиялар ғасырындағы «өмір сүру» деңгейін анықтайды. Әлеуметтік желілерде қазіргі уақытта байқалып отырған IT трендтері мыналар:

1. Әлеуметтік сауда (Social Commerce): Әлеуметтік желілер арқылы өнімдер мен қызметтерді тікелей сатып алу мүмкіндігі кеңінен таралуда. Instagram, Facebook, TikTok және басқа платформаларда дүкендер мен өнімдерді тікелей жарнамалау мен сату үшін интеграцияланған функциялар ұсынылуда.

2. Әлеуметтік Жасанды Интеллект (AI): Әлеуметтік желілерде жасанды интеллект қолдану жеке контент ұсыныстарын, чат-боттарды және әлеуметтік медиадағы трендтерді талдауды жақсартуға мүмкіндік береді. AI алгоритмдері пайдаланушының қызығушылығына сәйкес мазмұнды автоматты түрде көрсетуге көмектеседі.

3. Қысқа видео контент: TikTok және Instagram Reels сияқты платформалар арқылы қысқа және қызықты видеолар жасау танымал болып келеді. Бұл форматтың негізгі ерекшелігі – пайдаланушыларға қысқа уақытта көп ақпаратты немесе көңіл көтеретін контентті ұсыну.

4. Әлеуметтік медиа арқылы жаһандық қоғамдастықтар: Платформалар арасында қауымдастықтар мен топтардың кеңеюі әлеуметтік желілерде белгілі бір тақырыптарға қызығушылық танытатын адамдарды біріктіруге мүмкіндік береді. Бұл, әсіресе, арнайы қызығушылықтар мен кәсіби салаларда байқалады.

5. Әлеуметтік желі брендтері және инфлюенсерлер: Инфлюенсер маркетингінің дамуы брендтерге өз өнімдерін әлеуметтік желі арқылы жылжытуға мүмкіндік береді. Инфлюенсерлер өз аудиториясына байланысты өнімдер мен қызметтерді таныстырып, жарнамалайды.

6. Құпиялылық пен деректерді қорғау: Әлеуметтік желілерде құпиялылық мәселелері күн тәртібінде тұр. Пайдаланушылар өз деректерінің қалай пайдаланылып жатқанын қадағалап, жеке ақпараттарын қорғау мүмкіндігін іздейді. Бұл тренд платформалардың деректерді қорғау саясатын жақсартуға итермелейді.

7. Мазмұнды платформалардағы платформалық интеграция: Әлеуметтік желілер мен контент жасау платформалары арасындағы интеграция артып келеді. Мысалы, YouTube видеолары тікелей Instagram немесе TikTok-та бөлісу мүмкіндігі ұсынылады, бұл пайдаланушыларға контентті әртүрлі платформаларда таратуға мүмкіндік береді.

8. Мәтіндік жасанды интеллект (AI) көмекшілері: Әлеуметтік желілерде мәтіндік AI көмекшілері мен чат-боттар арқылы пайдаланушыларға жедел жауап беру, контент жасау және басқару мүмкіндіктері кеңейтілуде.

9. Виртуалды және артқы реальдық (AR) технологиялары: AR технологиялары әлеуметтік желілерде филтрлер мен әсерлер жасау үшін кеңінен қолданылады. Бұл тренд пайдаланушылардың әлеуметтік медида өз контенттерін қызықты әрі ерекше етіп көрсетуге мүмкіндік береді.

10. Мазмұнды модерация және әдептілік: Платформаларда қолданушылардың мазмұнын модерациялау мен әдептілік сақтау маңызды болуда. Бұл контенттің сапасын жақсарту және кибербуллинг пен теріс пікірлермен күресуге көмектеседі. Әлеуметтік желілердің қарқынды дамуы мен өзгеруіне байланысты, жаңа трендтер мен технологиялар тұрақты түрде пайда болып жатыр. Бұл – тілдің қолданыс аясының кеңеюінің бірден-бір жолы. Демек әлеуметтік желіде тілді дұрыс қолдану, тілдік норманы сақтау, сауатты жазу маңызды [8].

Ресей мен Қазақстан арасындағы тарихи, мәдени және экономикалық байланыстар орыс тілінің қазақ тіліне ықпалын тигізді. Бұл ықпалдың түпкі себебі екі халықтың ғасырлар бойғы араласуында жатыр. Ресей империясының Қазақстанды басып алуымен бірге орыс тілі кеңінен қолданылып, мемлекеттік басқару, білім беру және қоғамдық өмір салаларында өз орнын тапты. Орыс тілінен енген сөздер қазақ тілінің лексикасын байытты және оның дамуына үлкен әсер етті. Алайда қазіргі таңда қазақ тілі басымдыққа ие болса да, орыс сөздерінің дені әлеуметтік желілердегі қазақша контенттерге әсерін тигізуде. Яғни ауызша және жазбаша тілдік қатынастан бөлек, әлеуметтік желіде виртуалды қатынасты құрайтын тілдік орта пайда болды. Тілдік орта – қоғам мүшелерінің тілдік санасының қалыптасуына әсер ететін, тілдің қоғамдағы болмысына, оның дамуы мен қызмет етуіне ықпал ететін экономикалық, идеологиялық, мәдени, әлеуметтік, психологиялық, этнопсихологиялық факторлардың жиынтығы [9]. Тілдік ортаны тілдегі барлық жазба мәтіндер, жазарман, оқырман, айтарман, көрермен құрайды. Әлеуметтік желідегі белсенді тілдік орталар – instagram және facebook. Мұнда жарияланатын түрлі деңгейдегі жазбаларды тіл мамандары тексеріп, қадағаламайды. Тілдік норманы білмейтін, тілдің ережелерін меңгермеген желі қолданушылары әртүрлі қателер жібереді. Мәтін иесінің немесе танымал блогердің жазбасы қате болса да, оқырман оны дұрыс деп қабылдайды. М. Балақаев: «Тіл мәдениетінің қамын жеу, сайып келгенде, тіл тазалығына, ой дәлдігіне қамқорлық жасау...» десе [10], Р. Сыздық «Тіл мәдениеті дегеніміз – сөздерді дұрыс орнымен қолдану (лексикалық), дұрыс құрастыру (синтаксистік), дұрыс қиюластыру (морфологиялық), дұрыс дыбыстау (орфоэпиялық), сауатты жазу (орфографиялық), тілді

әсерлі етіп жұмсау (лингвостилистикалық) нормаларын ұстану, орнықтыру, жетілдіру», – дейді [9]. Н. Уәли: «Тіл мәдениетіне қажетті аса маңызды сапалардың бірі – сөз тазалығы. Сөз тазалығы дегенде ойымызға оралатыны – тіл тазалығы, сөзімізде бөгде тілдік элементтердің болмауы. Бірақ әдеби тілімізде өзге тілдерден енген сөздер, тіпті сөз тудырушы қосымшалар да бар екені мәлім. Ондай құбылысты тіл-тілдің барлығынан дерлік кездестіруге болады. Өзге тілден сөз алмаған, ауыз-түйістігі жоқ «таза» тілдер кемде-кем», – дейді [11]. Десек те орыс тіліндегі варваризмдер әлеуметтік желілерде сөйлеуді бейресми ету, эмоцияларды білдіру немесе тірі қарым-қатынас елесін жасау үшін қолданылады. Белгілі бір жағдайда бұндай қолданыстың қазақ тілінің әдеби нормасының қалпын сақтамауға әсері бар. Варваризмдерді шамадан тыс қолдану мәтіннің анықтығы мен құрылымын төмендетуі мүмкін. Бұндай тілдердің өзара әрекеттесуі туралы ғылыми тұжырымдар ғасырлар бойы айтылып келеді және бұл соңы емес екені белгілі. В. Матезиус «Тіл және стиль» еңбегінде тілдердің өзара әрекеттесу бағыттарын атап көрсетті: біріншіден, бір тілдің элементтерінің екінші тіл жүйесіне енуі тілдің дифференциалану дәрежесінің әлсіреуінен деп түсіндіреді. Алайда оған қарсы тұру мүмкін емес, – дейді ғалым. Себебі халықтар арасындағы идеялардың, заттардың, атаулардың және ұғымдардың үнемі алмасуы болса, онда шетел сөздері басымырық болуы тілдің ішке ресурстарының енуіне әсер етеді [11].

Біздің зерттеу саламыз бойынша әлеуметтік желідегі тіл мәселесін прагматикалық тұрғыдан [12] қарастырған. «Pragmatic discourse of the comments on Facebook in the Kazakh language» тақырыбында Facebook әлеуметтік желісі негізінде пікірлердің прагматикасы бойынша ауызекі сөйлеу тілінің ерекшеліктері, қарапайым сөздер, алғыс айту және қарғыс айту категориялары, эмоционалды-экспрессивті сөздер, вербалды емес әрекеттер, әртүрлі графикалық сызықтар, символдық белгілердің, тілдік объектілердегі таңбалар тілдің мағынасын тарылтатындығын көрсетеді. Ал [13] «Linguistic Representation of Value Dominants in Virtual Space: The Example of Kazakh and English Materials» зерттеуінде қазақ және ағылшын тіліндегі әлеуметтік желіде орналасқан жаңалықтарды, оның ішінде, саяси мәтіндерді қарастырып, түсініктеме береді. Зерттеу нәтижелері саяси мәтіндердің лингвистикалық ерекшеліктері мен олардың коммуникациядағы маңызын анықтайды. «A Linguistic Analysis of Social Network Communication» зерттеуі әлеуметтік желінің тілдік ерекшеліктеріне анықталған [14]. Әлеуметтік желілердің тілдік ерекшеліктеріне қызығушылық уақыт өткен сайын артып келеді. Келесі бір авторлар тобы виртуалды кеңістіктегі қазақ және ағылшын тілінде мәтін құрудың алгоритмін зерттесе [13], [14] «Communicative strategies in modern linguistics (on the material of English travel blogs)» тақырыбындағы жұмысын әлеуметтік желідегі коммуникацияға арнады.

Қазақ аудиториясы жиі кіретін әлеуметтік желілердің саны өте көп. Онлайн қарым-қатынас өмірдің барлық салаларында көрініс тауып, әртүрлі дискурсты қамтитын жағдайға жетті. Белгілі бір стилі анықталмаған хаттар мен хабарламалардың халыққа жеткізілу, таралу үлгісі де әртүрлі. Сандық және сапалық айырмашылықтар қазіргі әлеуметтік желіде де, медиа саласында да байқалады. Интернет-коммуникация – бұл екі адамның ғана қарым-қатынасынан құралмайды, онда мыңдаған пікірлер, толассыз тартыстар, миллиондаған қаралымдар пен лүпілдер, репосттар болады. Әлеуметтік желідегі тілдің ең үлкен артықшылығы – сөздің әртүрлі семантикада жұмсалуды, жанрлық және стильдік нормалар, ауызекі тілдік элементтер, жаргондар мен сленгтер, паразит сөздер мен варваризмдердің араласа қолданылуы. Сөйлеудің бұл стратегиялары кез келген чатта, кез келген платформада да жүзеге асырыла береді. Мәтіндерінің көлемі шағын болғанына қарамастан, күрделі мәселелерді көтереді. Сондықтан әлеуметтік желінің қай түрі болмасын, адамзатты өзіне тәуелді етті. Лингвистикалық тұрғыдан М. Мәліктің «Әлеуметтік желілердегі сөйлеу қызметінің прагматикасы мен функциясы» атты зерттеу еңбегі әлеуметтік желі тілі мен қазіргі қазақ әдеби тілінің нормаларына арналады [13]. Бұған дейінгі зерттеулер интернетке, оның түрлеріне, кеңістігіне арналса, енді прагматикасы, коммуникациясы, лингвостилистикасы зерттелмек. Ғаламтор адам санасына тиімді тілдік әсер етудің жалғыз жолы болғандықтан, оның тілі нормаға сай, сауатты, қатесіз болуы керек.

Қазақ тіліндегі стиль түрлері әлеуметтік желіде араласа жұмсала беріледі. Яғни бір стильде әртүрлі стиль элементтерін қолдануға және көпқабатты стильдік құбылыстарды жасауға болады. Қарыштап дамыған технологиялардың әсерінен интернеттегі сөйлесу стилі де сан алуан. Әлеуметтік желідегі кез келген мәтін мен пікір сөйлеу мәнерінен жазбаша нұсқаға ауысқан. Бұрын тек ауызша таралып, ауызша қызмет етсе, қазіргі таңда ғаламторда функционалды стильдердің дамуы сөйлеу стилінің өзгеруіне және жаңа элементтердің пайда болуына әкелді. Бұл стиль ешқандай ережеге бағынбайды. Өзінің жаргондарымен, сленгтерімен, варваризмдерімен ерекшеленеді. Әлеуметтік желілердегі тілдік құралдарды қолданудың әртүрлілігі әдеби нормадан ажырап, ұлттық құндылықтардың азайып бара жатқанын көрсетеді.

### Материалдар мен әдістер

Зерттеу барысында біз Instagram және Facebook сияқты әлеуметтік желілерде жиі талқыланатын тақырыптар бойынша блогерлердің жазбаларын зерттеу үшін сапалы және сандық әдістерді қолдандық. Мазмұнды талдау және лингвистикалық деректерді іздеу әдістерін қолдана отырып, біз осы жазбаларға қатысты пікірлердің маңызды үлгісін талдадық. Бұл тәсіл белгілі бір сөздер мен сөз тіркестерінің жиі қолданылуын анықтауға мүмкіндік берді. Содан кейін біз қазақ тілінің эволюциясына әсер ететін терминдердің лексикасын құрастырдық, бұл терминдердің тілді қолданудағы өзгерістерді және шет тілдерінің қазақ тіліне әсерін қалай көрсететініне назар аудардық. Деректер әлеуметтік желілердегі дискурстың қазақ тілінің өзгеруіне қалай ықпал ететінін көрсете отырып, тілге бейімделу заңдылықтары мен тенденцияларын анықтау үшін санаттарға бөлініп, талданды.

АТ көмегімен әлеуметтік желілерден мәліметтер жинауға шолу

АТ құралдарын қолдана отырып, әлеуметтік желілерден деректерді жинау зерттеу әдістемелеріндегі айтарлықтай ілгерілеуді білдіреді. АТ құралдары мен әдістерін қолдану әртүрлі әлеуметтік медиа платформаларынан деректердің үлкен көлемін жүйелі түрде жинауға, талдауға және түсіндіруге мүмкіндік береді. Бұл тәсіл деректерді жинау тиімділігін арттырып қана қоймайды, сонымен қатар қолмен әдістерді қолдану арқылы жіберіп алуы мүмкін егжей-тегжейлі ақпаратты алуға көмектеседі.

Деректерді жинаудың автоматтандырылған құралдары Facebook, Instagram, Twitter және LinkedIn сияқты әлеуметтік медиа желілерінен үлкен көлемдегі ақпаратты ала алады. Бұл құралдар профильдерді, жазбаларды, түсініктемелерді және басқа да тиісті мазмұнды алу үшін алгоритмдерді пайдаланады, деректерді кейінгі талдау үшін құрылымдайды. Деректерді жинау құралдарының этикалық деректерді жинауды қамтамасыз ету үшін әрбір платформаның қызмет көрсету шарттарына және құпиялылық саясатына сәйкес келуі өте маңызды.

Көптеген әлеуметтік желілер зерттеушілерге деректерге құрылымдық форматта қол жеткізуге мүмкіндік беретін API интерфейстерін ұсынады. API интерфейстері пайдаланушылардың өзара әрекеттесуі, мазмұны және трендтері туралы ақпаратты алудың бақыланатын әдісін ұсынады. Мысалы, Twitter API интерфейсін твиттер мен қатысу көрсеткіштерін жинау үшін пайдалануға болады, Ал Instagram Graph API жариялау көрсеткіштері мен пайдаланушы деректеріне қол жеткізуді қамтамасыз етеді.

Әлеуметтік желілерден жиналған мәтіндік деректерді талдау үшін табиғи Тілді Өңдеу (NLP) әдістері қолданылады. Көңіл-күйді талдауды, тақырыпты модельдеуді және кілт сөздерді шығаруды қолдана отырып, зерттеушілер әлеуметтік медиа мазмұнындағы тенденцияларды, көңіл-күйлерді және тақырыптық үлгілерді анықтай алады. NLP алгоритмдері пайдаланушының тілі мен мінез-құлқы туралы құнды түсініктер ұсына отырып, пайдаланушы жасаған мазмұнның мәнмәтіні мен нюанстарын түсінуге көмектеседі.

Машиналық оқыту модельдерін үлгілерді тануға және әлеуметтік медиа деректеріне негізделген нәтижелерді болжауға үйретуге болады. Мысалы, жіктеу алгоритмдері жазбаларды көңіл-күй бойынша жіктей алады, ал кластерлеу әдістері ұқсас мазмұнды немесе пайдала-

нушы профильдерін топтастыра алады. Машиналық оқыту үлкен деректер жиынын талдау және бірден байқалмайтын жасырын үлгілерді ашу мүмкіндігін арттырады.

Әлеуметтік желілер күн сайын үлкен көлемдегі деректерді шығарады, оларды көбінесе үлкен деректер деп атайды. Hadoop немесе Spark сияқты үлкен деректер құрылымдарын пайдалану зерттеушілерге осы үлкен көлемдегі ақпаратты тиімді өңдеуге және талдауға мүмкіндік береді. Бұл технологиялар құрылымдалмаған деректерді өңдеуді қолдайды, бұл әлеуметтік медиа трендтері мен үлгілерін жан-жақты талдауға және визуализациялауға мүмкіндік береді. Деректерді визуализациялау құралдары күрделі әлеуметтік медиа деректерін түсіндіру үшін қажет. Графиктер, диаграммалар және жылу карталары сияқты көрнекі бейнелерді жасау зерттеушілерге трендтерді, корреляцияларды және ауытқуларды оңай анықтауға көмектеседі. Визуализация нәтижелерді тиімді жеткізуге және әлеуметтік медиа динамикасын жақсырақ түсінуге көмектеседі.

Әлеуметтік желілерден деректерді жинау кезінде этикалық ойлар өте маңызды. Зерттеушілер деректерді жинау әдістерінің пайдаланушылардың құпиялылығын сақтауын және құқықтық және этикалық нормаларды сақтауын қамтамасыз етуі керек. Деректерді анонимизациялау және біріктіру пайдаланушының жеке басын қорғауға және ақпаратты теріс пайдаланудың алдын алуға көмектеседі [15].

Осылайша, әлеуметтік желілерден деректерді жинау үшін АТ құралдарын пайдалану зерттеушілерге үлкен көлемдегі ақпаратпен дәлірек және тиімдірек жұмыс істеуге мүмкіндік береді. Автоматтандырылған құралдарды, API Интерфейстерін, NLP, машиналық оқытуды, үлкен деректерді талдауды және визуализация әдістерін қолдану әлеуметтік медиа құбылыстарын тереңірек түсінуге ықпал етеді және цифрлық коммуникация тенденциялары туралы түсінікті арттырады [16–17].

Кесте 1 – Қазақ тілінің дамуына және қолданушылардың қатысуына әсер ету контекстіндегі әлеуметтік желілерді салыстырмалы талдау

| Әлеуметтік желі | Жаһандық аудитория | Қазақ ауд-сы | Жасы         | Функциясы                                                                                                | Тақырыбы                                          | Қолайлылық                                                                           | Пайдаланушының ой-өрісі                                          |
|-----------------|--------------------|--------------|--------------|----------------------------------------------------------------------------------------------------------|---------------------------------------------------|--------------------------------------------------------------------------------------|------------------------------------------------------------------|
| Facebook        | 3 млрд             | 2.6 млн      | 25–34, 35–44 | Іскерлік, әлеуметтік және білім беру мазмұнын тұтынушылар; экономикалық белсенді; маркетингтік платформа | Саясат, тіл, дін, білім, экономика                | Ақпарат тез таралады; Жаңалықтар уақтылы және шынайы; Қоғамдық мәселелер талқыланады | Мазмұнды контент, ойын-сауық аз                                  |
| Instagram       | 2 млрд             | 1.5 млн      | 18–24, 25–34 | Мазмұнды көрнекі түрде бөлісу; әсер етуші маркетинг; трендке негізделген                                 | Сән, өмір салты, саяхат, тамақтану, фитнес        | Өте көрнекі; қызықты әңгімелер мен бейнелер; Трендтер тез таралады                   | Көрнекі тартымдылық пен тенденцияларға, ойын-сауыққа бағытталған |
| Twitter         | 450 млн            | 800,000      | 18–34        | Нақты уақыттағы жаңартулар; Жаңалықтар және қоғамдық дискурс; ақпаратты жылдам тарату                    | Саясат, Жаңалықтар, Технологиялар, Атақты адамдар | Нақты уақыттағы жаңартулар; трендтерге арналған хэштегтер; қысқаша жазбалар          | Нақты уақыттағы ақпаратқа назар аудару; Қоғамдық дискурс         |

1-кестенің жалғасы

|          |         |         |       |                                                                          |                                                |                                                                                            |                                                                  |
|----------|---------|---------|-------|--------------------------------------------------------------------------|------------------------------------------------|--------------------------------------------------------------------------------------------|------------------------------------------------------------------|
| TikTok   | 1 млрд  | 1.2 млн | 16–24 | Қысқа бейне мазмұны; әсер етушілерге бағытталған тенденциялар;           | Ойын-сауық, ән, би, комедия                    | Қысқа, тартымды бейнелер; трендке негізделген мазмұн; жоғары интерактивті                  | Қысқа формадағы трендтер мен ойын-сауық әзілдері                 |
| LinkedIn | 900 млн | 500,000 | 25–54 | Кәсіби желі; бизнес мазмұны; жұмыс іздеу                                 | Мансап, сала жаңалықтары, біліктілікті арттыру | Кәсіби мазмұн; желілік мүмкіндіктер; жұмыс туралы хабарландырулар                          | Кәсіби өсуге назар аудару; бизнеске бағытталған                  |
| WhatsApp | 2 млрд  | 1 млн   | 18–54 | Жедел хабар алмасу; жеке және топтық байланыс; файлдарды ортақ пайдалану | Жеке әңгімелер, топтық чаттар                  | Шифрланған хабарламалар; мәтіндерді, фотосуреттерді, бейнелерді оңай бөлісу; топтық чаттар | Жеке қарым-қатынасқа назар аудар; тікелей және жеке хабар алмасу |

Бұл кестеде олардың жаһандық және қазақстандық аудиториясына, жас ерекшеліктеріне, негізгі функцияларына, кең таралған тақырыптарына, пайдаланушылардың ыңғайлылығына және ойлауына назар аударатын, бес негізгі әлеуметтік желіні салыстыру берілген. Талдау пайдаланушылардың өзара әрекеттесуі, мазмұн теңгерімдері және қарым-қатынас мәнерлері арқылы әртүрлі платформалардың қазақ тіліне қалай әсер ететінін түсінуге бағытталған.

Әлеуметтік желілерден деректерді жинау процесі

Бұл зерттеуде лингвистикалық заңдылықтарды және олардың қазақ тіліне әсерін талдау үшін Instagram және Facebook әлеуметтік медиа платформаларынан мәліметтер жиналды. Деректерді жинау процесі жан-жақты және сенімді нәтижелерді қамтамасыз ету үшін құрылымдалған.

1. Деректерді жинаудың міндеттері. Негізгі мақсат лингвистикалық тенденцияларды, соның ішінде шетелдік сөздерді, эмоционалды өрнектерді және саяси терминдерді қолдануды анықтау үшін Instagram және Facebook-тан пайдаланушы жасаған мазмұнды жинау және талдау болды.

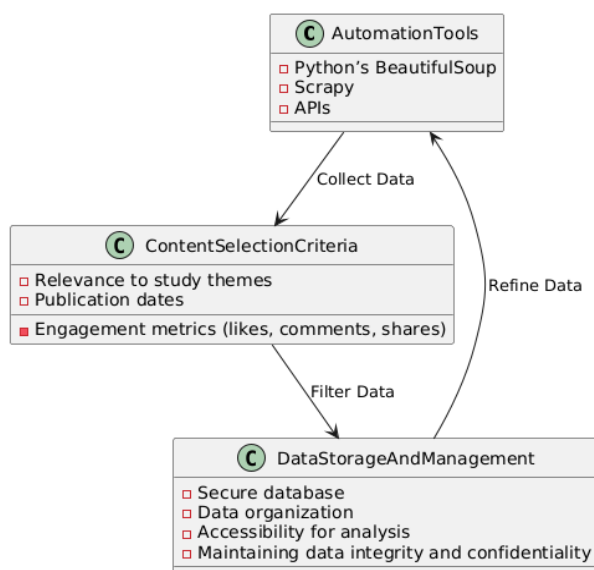
2. Әлеуметтік желілерді таңдау. Instagram және Facebook кеңінен қолданылуына және цифрлық коммуникацияда тілді қолдануға айтарлықтай әсер етуіне байланысты таңдалды. Instagram визуалды және мәтіндік деректерді ұсынады, ал Facebook әртүрлі және кең мәтіндік мазмұнды ұсынады.

3. Деректерді жинау әдістері.

♦ Автоматтандыру құралдары: деректер әлеуметтік медиа платформалары ұсынатын Python BeautifulSoup, Scrapy және API интерфейстері сияқты автоматтандырылған құралдар мен кітапханалар арқылы жиналды. Бұл құралдар жазбаларды, пікірлерді және пайдаланушылардың өзара әрекеттесуін шығаруды жеңілдетті.

♦ Мазмұнды таңдау критерийлері: мазмұн белгілі бір критерийлер негізінде таңдалды, соның ішінде зерттеу тақырыптарының өзектілігі (мысалы, саяси, эмоционалдық және шетелдік сөздер), қатысу көрсеткіштері (ұнатулар, пікірлер, бөлісулер) және жарияланған күндері.

♦ Деректерді сақтау және басқару: жиналған деректер қауіпсіз дерекқорда сақталды, бұл дұрыс ұйымдастыруды және кейінгі талдау үшін қолжетімділікті қамтамасыз етті. Процесс барысында деректердің тұтастығы мен құпиялылығы сақталды [18–19].



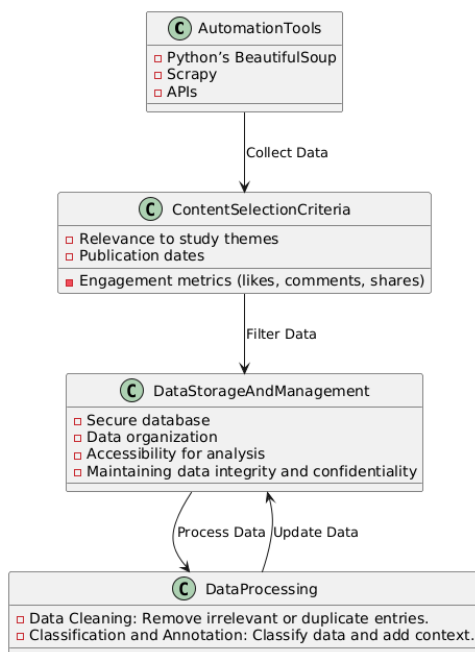
Сурет 1 – Деректер ағыны және өңдеу диаграммасы

Бұл диаграмма деректерді өңдеу жүйесінің негізгі компоненттерін және олардың өзара әрекеттесуін бейнелейді.

Деректерді өңдеу

♦ Деректерді тазалау: бастапқы деректерді тазалау деректер жиынының дәл және репрезентативті болуын қамтамасыз ете отырып, маңызды емес немесе қайталанатын жазбаларды жою үшін орындалды.

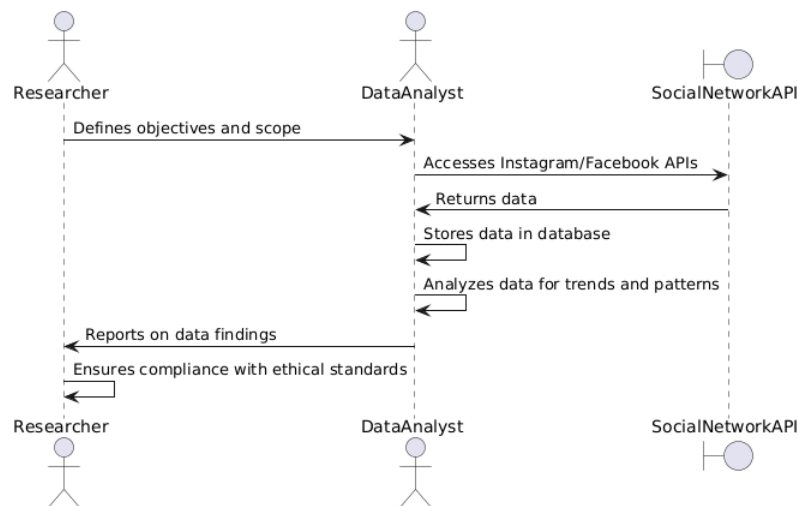
♦ Жіктеу және аннотация: Деректер жағымсыз сөздер, эмоционалды-экспрессивті сөздер, саяси терминдер және шет тіліндегі сөздер сияқты санаттарға жіктелді. Аннотация әрбір жазбаға контекст пен өзектілік қосу үшін жасалды [20–21].



Сурет 2 – Деректерді өңдеу және басқару ағыны

Бұл диаграмма 2-суретте деректерді өңдеу және басқару жүйесіндегі негізгі компоненттер арасындағы мәліметтер ағыны көрсетілген. Оған автоматтандыру құралдары, мазмұнды таңдау критерийлері, деректерді сақтау және басқару, деректерді өңдеу кезеңдері кіреді. Диаграмма жүйеде деректердің қалай жиналатынын, сүзілетінін, өңделетінін және жаңартылатынын көрсетеді.

Осы процедураларды орындай отырып, зерттеу цифрлық дәуірдегі тіл динамикасын тереңірек түсінуге ықпал ете отырып, әлеуметтік желілердегі лингвистикалық тенденцияларға сенімді талдау жасауға бағытталған [21–22].



Сурет 3 – Әлеуметтік желі деректерін талдау процесі

Бұл реттілік диаграммасы 3-суретте зерттеуге қатысушылар мен әлеуметтік желілерден деректерді жинау және талдау жүйесінің өзара әрекеттесу процесі көрсетілген.

Процесс қадамдарының сипаттамасы:

1. Мақсаттары мен көлемін анықтау: зерттеуші деректерді талдаудың мақсаттары мен көлемін белгілейді.
2. Деректерді шығару: деректер талдаушысы API арқылы әлеуметтік желілерден (Instagram/Facebook) деректерге қол жеткізеді.
3. Деректерді қабылдау: API сұралған деректерді деректер талдаушысына қайтарады.
4. Деректерді сақтау: деректер талдаушысы жиналған деректерді мәліметтер базасында сақтайды.
5. Деректерді талдау: деректерді талдаушы трендтер мен заңдылықтарды анықтай отырып, деректерді талдайды.
6. Есеп нәтижелері: деректер талдаушысы талдау нәтижелерін зерттеушіге хабарлайды.
7. Этикалық сәйкестікті қамтамасыз ету: зерттеуші деректер мен талдаудың этикалық нормаларға сәйкестігін қамтамасыз етеді.

Бұл диаграмма деректерді талдау процесінің негізгі қатысушылары арасындағы қадамдар мен өзара әрекеттесулердің реттілігін визуализациялауға көмектеседі.

Әдістеме

Бұл зерттеуде мәтіндік деректерді талдау және жіктеу үшін мәтіндік талдау мен машиналық оқыту әдістерін біріктіретін екі фазалы әдістеме қолданылды. Бұл процесс әлеуметтік медиа платформаларынан деректерді алуды, мәтінді алдын ала өңдеуді және машиналық оқытуды қамтиды.

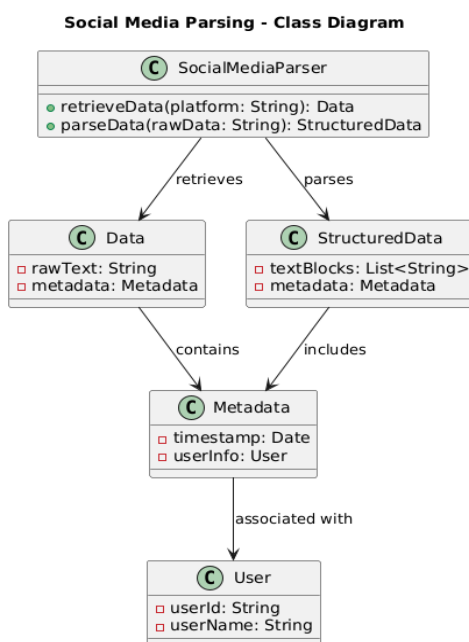
1. Әлеуметтік медианы талдау:

◆ Деректер instagram және Facebook-тан жеке талдау әдістерін қолдана отырып алынды. Бұл осы платформалардың жазбаларынан, түсініктемелерінен және басқа да тиісті бөлімдерінен мәтіндік мазмұнға бағдарламалық түрде қол жеткізуді және алуды қамтыды. Талдау процесі 4-суреттегі әлеуметтік желілерде кездесетін нақты деректер пішімдері мен құрылымдарын өңдеуге арналған.

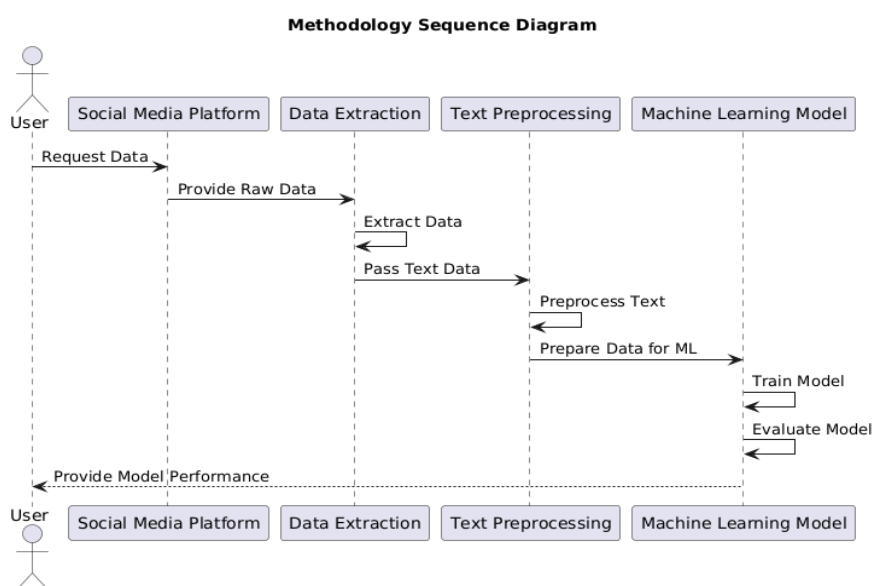
## 2. Деректерді құрылымдау:

Алынған әлеуметтік медиа деректері талдауға жарамды форматта құрылымдалған. Instagram және Facebook үшін деректер мәтіндік блоктар мен метадеректерге бөлініп, мәтіндік мазмұнды келесі қадамдарда тиімді өңдеуге және талдауға мүмкіндік берді.

Бұл әрекет диаграммасы әлеуметтік медианы талдау процесіне қатысатын жұмыс процесін көрсетеді. Ол әлеуметтік медиа платформаларынан деректерді алудан бастап өңделген ақпаратты сақтауға дейінгі әрекеттердің реттілігін сипаттайды. Диаграммада API қоңырауларын шалу және веб-скрепинг сияқты платформаға (Instagram немесе Facebook) негізделген әртүрлі әрекеттерді өңдеу үшін шешім қабылдау нүктесі бар. Содан кейін ол шикі деректерді алу, талдау, құрылымдау және соңында өңделген деректерді сақтау қадамдарын көрсетеді.



Сурет 4 –Әлеуметтік медианы талдау



Сурет 5 – Әлеуметтік медианы талдауға арналған жұмыс процесінің белсенділік диаграммасы

Мәтінде талдау

1. Мәтінді алдын ала өңдеу:

Мәтіндік деректер дәйектілікті қамтамасыз ету және оларды талдауға дайындау үшін алдын ала өңделді. Бұл регистрге сезімталдықты сақтау үшін барлық мәтінді кіші әріпке түрлендіруді және 5-суреттегі тұрақты тіркестерді қолдана отырып, мәтінді жеке сөздерге таңбалауды қамтыды. Бұл әрекет диаграммасы әлеуметтік желіні талдау процесіне қатысатын жұмыс процесін көрсетеді. Ол әлеуметтік медиа платформаларынан деректерді алудан бастап өңделген ақпаратты сақтауға дейінгі әрекеттердің реттілігін сипаттайды. Диаграммада API қоңырауларын шалу және веб-скрепинг сияқты платформаға (Instagram немесе Facebook) негізделген әртүрлі әрекеттерді өңдеу үшін шешім қабылдау нүктесі бар. Содан кейін ол шикі деректерді алу, талдау, құрылымдау және соңында өңделген деректерді сақтау қадамдарын көрсетеді [25].

2. Пайыздық сәйкестік:

Өңделген мәтін мен анықтамалық терминдер арасындағы сәйкес сөздердің пайызы есептелді. Анықтамалық терминдер жеке құжаттағы кестелерден алынды және сәйкестік пайызы формула бойынша есептелді:

$$\text{Percentage} = \left( \frac{\text{Total Matches}}{\text{Total Words}} \right) \times 100$$

Бұл көрсеткіш әлеуметтік медиа мәтінінің анықтамалық кестелерде көрсетілген нақты терминологияларға немесе тұжырымдамаларға қаншалықты сәйкес келетінін көрсетті.

Машиналық оқытудың жіктелуі

1. Деректерді дайындау:

Алдын ала өңделген мәтіндік деректер мен анықтамалық сөздер «scikit-learn» кітапханасындағы «CountVectorizer» көмегімен мүмкіндік векторларына түрлендірілді. Бұл түрлендіру мәтіндік деректерді сандық форматқа түрлендіру арқылы машиналық оқыту алгоритмдерін қолдануды жеңілдетті.

2. Модельдік оқыту:

Мәтіндік деректерді жіктеу үшін Аңғал Байес классификаторы ('MultinomialNB') қолданылды. Бұл модельді таңдау оның категориялық деректерді өңдеуге жарамдылығына және оны жүзеге асырудағы тиімділігіне негізделген. Деректер жиынтығы модельдің жеке деректер негізінде бағалануын қамтамасыз ету үшін 80-нен 20-ға дейінгі арақатынаста оқыту мен тестілеудің ішкі жиындарына бөлінді.

3. Модельді бағалау:

Жіктеуіштің өнімділігі дәлдікпен өлшенді, бұл дұрыс жіктелген даналардың үлесін көрсетеді. Модель 1,00 дәлдік көрсеткішіне қол жеткізді, бұл сынақ деректері бойынша тамаша жіктеу өнімділігін көрсетеді.

Әдістеме әлеуметтік желілердегі мәтіндерді талдауға кешенді көзқарасты қамтамасыз ету үшін деректерді алудың озық әдістерін мәтіндік талдаумен және машиналық оқытумен біріктірді. Деректер Instagram және Facebook-тан талданды, талдау үшін өңделді және Аңғал байес классификаторын үйрету үшін пайдаланылды. Бұл тәсіл әлеуметтік медиа мазмұнын дәл жіктеуге және тиімді талдауға мүмкіндік берді. Әлеуметтік желілерден деректерді алудан бастап, талдау және құрылымдау арқылы соңғы деректерді сақтауға дейінгі жұмыс процесін бейнелейді. Платформаға тән әрекеттер үшін шешім қабылдау нүктелерін қамтиды [26].

## Нәтижелер мен талқылау

Қоғамдағы барлық өзгерістер тілдің дамуына немесе тежелуіне әсер етеді. Қазіргі уақытта тіл біліміне жаңа көзқараспен тілді антропоспецификалық және экологиялық тұрғыдан зерттеу жүргізілуде. Яғни, адамның тілге қалай әсер ететінін, тілдің адамның ойлауына, сөйлеуіне, мәдениетіне қандай әсер ететінін анықтау. Әлеуметтік желі тілін прагматикалық,

коммуникативті, лингвостилистикалық тұрғыдан қарастыру тілді экологиялық тұрғыдан зерттеуге жатады.

```
Analyzing text using data from tables...
Match percentage for 'Situation': 0.01%
Match percentage for 'Agression': 0.00%
Match percentage for 'Trolling': 0.01%
Match percentage for 'EEW': 0.97%
Match percentage for 'PW': 0.13%
Match percentage for 'RW': 0.24%
Match percentage for 'EnW': 0.57%
Match percentage for 'ErW': 0.34%
Match percentage for 'Abb': 0.17%
Preparing data for machine learning...
Splitting data into training and testing sets...
Training Naive Bayes classifier...
Evaluating model accuracy...
Model accuracy: 1.00
Process completed.
```

Толық талдау нәтижелері:

1. Ситуация: 0.01%

"Ситуация" санатына сәйкес келетін мәтіннің пайызы өте төмен, бұл мәтінде ситуациялық контекстке қатысты сілтемелер немесе элементтер аз болатындығын көрсетеді.

2. Агрессия: 0.00%

Агрессивті тілдің немесе мазмұнның анықталатын жағдайлары табылған жоқ. Бұл мәтінде ешқандай агрессивті өрнектер мен тондардың жоқтығын көрсетеді.

3. Троллинг: 0,01%

Троллингке қатысты мазмұнның болуы шамалы, бұл троллингтің мінез-құлқы немесе тілі мәтінде іс жүзінде жоқ екенін көрсетеді.

4. EEW (Эмитикондарды пайдалану): 0,97%

Смайликтерді пайдалану кезінде сәйкестіктің салыстырмалы түрде жоғары пайызы байқалады – 0,97%. Бұл смайликтердің мәтінде кездейсоқ немесе бейресми реңкті білдіретін жиі қолданылатынын көрсетеді.

5. PW (Саяси сөздер): 0.13%

Саяси сөздер мәтіннің аз бөлігінде кездеседі. 0,13% деңгейінде бұл саяси сілтемелердің немесе терминологияның бар екенін, бірақ басым емес екенін көрсетеді.

6. RW (Орысша сөздер): 0,24%

Мәтінде 0,24% сәйкес келетін орыс сөздерінің қарапайым саны бар. Бұл мәтінде орыс тіліне немесе контекстке қатысты кейбір элементтер бар екенін көрсетеді.

7. EnW (Ағылшын сөздері): 0,57%

Ағылшын сөздері мәтіннің едәуір бөлігін құрайды, 0,57% сәйкес келеді. Бұл ағылшын тілінің мәтіндегі көрнекті тіл екенін көрсетеді, бірақ ол қолданылатын жалғыз тіл емес.

8. ErW (Қателер/Қате сөздер): 0.34%

Мәтіннің 0,34% құрайтын қателіктердің немесе қате сөздердің айтарлықтай болуы байқалады. Бұл емле немесе грамматикада кейбір мәселелер бар екенін көрсетеді.

9. Абб (Қысқартулар): 0.17%

Қысқартулар мәтіннің кішкене бөлігінде 0,17% сәйкестікпен пайда болады. Бұл аббревиатуралардың қолданылатынын, бірақ мәтіннің негізгі ерекшелігі емес екенін көрсетеді.

Қысқаша мазмұны:

Талдау келесі негізгі ойларды анықтайды:

♦ Эмотикондарды пайдалану (EEW) – бұл ең көп таралған функция, ол 0,97% сәйкес келеді. Бұл эмоцияларды немесе өрнектерді жеткізу үшін смайликтерді қолданатын кездейсоқ немесе бейресми стильді ұсынады.

♦ Ағылшын сөздері (EnW) 0,57% деңгейінде айтарлықтай қатысады, бұл ағылшын тілінің мәтіндегі негізгі тіл екенін көрсетеді.

♦ Орыс тіліндегі сөздер (RW) және қателер/кәте сөздер (ErW) қалыпты қатысуды көрсетеді, бұл көптілді мазмұнды немесе мәтін сапасын жақсартуды қажет ететін аймақтарды көрсетуі мүмкін.

♦ Саяси сөздер (PW), аббревиатуралар (Abb), ситуация және агрессия азырақ байқалады, бұл осы аспектілерге шектеулі назар аударуды білдіреді. Бұл егжей-тегжейлі бөлім мәтіндегі лингвистикалық және тақырыптық элементтер туралы түсінік береді, әрі қарай талдау үшін қызығушылық тудыратын және ықтимал фокусты көрсетеді.

## Қорытынды

Зерттеу мақаласында орыс тілі элементтерінің әлеуметтік желілердегі қазақ тілінің дамуына әсері қарастырылған. Әлеуметтік желілердегі пікірлерді талдай отырып, қазақ тілінің қазіргі жағдайы анықталып, интернет байланысы қазіргі кезде қысқартулар, бұрмалаулар, тыныс белгілерінің қателері, смайликтер, жаргондар және шет тілдерінің араласуы сияқты лингвокреативтіліктің әртүрлі элементтерін қамтитынын көрсетеді.

Біздің зерттеуіміз орыс сөздерінің (варваризмдердің) әлеуметтік желілердегі қазақ тілінің эволюциясына әсеріне бағытталған. Нәтижелер орыс сөздерін жиі қолдану қазақ грамматикасы мен сөздік қорына теріс әсер етіп, лингвистикалық стандарттардың әлсіреуіне әкелуі мүмкін екенін көрсетеді. Варваризмдердің осылай қосылуы қазақ тілінің сөздік қорының кедерленуіне және оның қайталанбас ерекшеліктерінің эрозиясына әкелуі мүмкін. Сонымен қатар, орыс сөздерінің таралуы қазақтың мәдени және дәстүрлі контексттерін түсіну мен бағалауды төмендетуі мүмкін.

Осы мәселелерді шешу үшін бірнеше стратегия ұсынылады. Әлеуметтік желілерде қазақ тілін дұрыс қолдануға, орфографиялық және грамматикалық ережелерді сақтауға баса назар аудару керек. Тіл білімін жетілдіруге және шетелдік ықпалға қарсы тұру үшін қазақ тілінің сөздік қорын байытуға күш салу керек. Білім беру бастамалары қазақ тілін дұрыс қолдануға және варварлықтардың әсеріне бағытталуы керек. Қазақ тілінің мәртебесін көтеретін тілдік саясат пен мәдени бағдарламаларды қолдау оның цифрлық салада тазалығы мен табиғи дамуын қамтамасыз етуге көмектеседі.

Варваризмдердің әсерін жою үшін қазақ тілінің әлеуметтік желілерде дамуын жалғастыра отырып, оның тұтастығы мен айқындылығын сақтау үшін өте маңызды.

## ӘДЕБИЕТТЕР

- 1 Интернет-ресурс. URL: <https://www.gov.kz/memleket/entities/mam/press/article/details/151678> (дата обращения: 07.10.2024).
- 2 Салкынбай А. Ашық кітапхана (сұқбат). – 2024.
- 3 Baron N.S. Computer Mediated Communication as a Force in Language Change // *Visible Language*. – 1984. – V. 18. – No. 2. – P. 118–141. URL: <https://s3-us-west-2.amazonaws.com/visiblelanguage/pdf/18.2/computer-mediated-communication-s-a-force-in-language-change.pdf> (дата обращения: 07.10.2024).
- 4 Baron N.S. *Always on: Language in an Online and Mobile World*. – Oxford University Press, 2010. <https://doi.org/10.1093/acprof:oso/9780195313055.001.0001>.
- 5 Crystal D. *Language and the Internet*. – Cambridge University Press, 2006. – P. 24.
- 6 Панюшкина О. Изменение языка под влиянием социальных сетей (на материале испанского языка) // *Альманах современной науки и образования*. – 2014. – № 3(82). – С. 133–135.
- 7 Иванов Л. *Язык в электронных средствах коммуникации* // *Культура русской речи*. – М.: Флинта, 2003. – С. 791–793.
- 8 Сыздыкова Р. Норма языка и его нормализация (кодификация) // *Сборник многотомного издания*. – Алматы: Ел-Шежире, 2014.
- 9 Балакаев М. *Казах әдеби тілі*. – Алматы, 2007. – С. 396–397.
- 10 Уали Н. *Казақ сөз мәдениетінің теориялық негіздері: филол. ғыл. д-ры... дисс.*: 02/10/02. Алматы, 2007. 328 с.

- 11 Алкебаева Д., Саткенова Ж.Б., Конурбекова Т. Изменения в казахском языке в социальных сетях // Серия Филология КазНУ им. аль-Фараби. – № 2(182). – Алматы, 2021.
- 12 Malik M., Alkebaeva D., & Ahkmetkaliyeva L. Pragmatic discourse of the comments on Facebook in the Kazakh language // *Opción*. – 2019. – V. 35. – No. 90-2. – P. 369–384. ISSN 1012-1587, ISSN 2477-9385.
- 13 Kurkimbayeva A.M., Akhatova B.A., Gumovskaya G.N., Kotenyatkina I.B., & Khuziakhmetov A.N. Communicative strategies in modern linguistics (on the material of English travel blogs) // *XLinguae Journal*. – 2020. – V. 13. – No. 2. – P. 33–49. <https://doi.org/10.18355/XL.2020.13.02.03>.
- 14 Kurmanbekova Z., Shokabayeva S.K. Linguistic Analysis of Social Network Communication // *International Journal of Society, Culture & Language*, 2022.
- 15 Пастухова Л.С. Основы культуры речи и стилистики. – Симферополь, 2010. – С. 127–244.
- 16 Жанпеисова Н.М. Адаптация заимствованных из русского языка слов в казахском интернет-дискурсе // *Медиалингвистика*. – 2023.
- 17 Mutanov G. Multi-Class Sentiment Analysis of Social Media Data with Machine Learning Algorithms // *Computers Materials & Continua*. – 2021.
- 18 Алкебаева Д. Изменения в казахском языке в социальных сетях // *Евразийский журнал филологии: наука и образование*. – 2021.
- 19 Нарынов С.С. Об одном подходе к решению задачи анализа настроений для казахского и русского языков с использованием глубокого обучения // *International Conference on Computational Collective Intelligence*. – 2016.
- 20 Иманберди А. Подходы к проектированию системы мониторинга и анализа данных социальных сетей, адаптированных к специфике казахского и русского языков // *Вестник Национальной инженерной академии Республики Казахстан*. – 2023.
- 21 Абибуллаева А.А. Извлечение ключевых слов из казахских текстов с использованием алгоритмов машинного обучения // *Вестник серии физико-математических наук*. – 2024.
- 22 Токтарова А.Б. Сбор непристойных комментариев в казахском языке с использованием методов машинного обучения // *Вестник серии физико-математических наук*. – 2023.
- 23 Зоткина А. Применение методов машинного обучения в задаче прогнозирования кибербуллинга пользователей социальных сетей // *Современные высокие технологии*. – 2022.
- 24 Абилдаев К.М. К вопросу заимствования русской лексики в деловой сфере казахского языка // *Вестник Евразийского национального университета им. Л.Н. Гумилева. Серия: филология*. – 2023.
- 25 Жапарова Т. Синергетический анализ процессов словообразования как реакция на казахстанскую реальность // *Acta Universitatis Lodziensis Folia Linguistica Rossica*. – 2023.

## REFERENCES

- 1 <https://www.gov.kz/memleket/entities/mam/press/article/details/151678>
- 2 Salkynbay A. (2024) *Ashyq qitaphana (syqbat)*.
- 3 Baron N. S. (1984) Computer Mediated Communication as a Force in Language Change. *Visible Language*, vol. 18, no. 2, pp. 118–141. URL: [<https://s3-us-west-2.amazonaws.com/visiblelanguage/pdf/18.2/computer-mediated-communication-s-a-force-in-language-change.pdf>]
- 4 Baron N. S. (2010) *Always on: Language in an Online and Mobile World*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195313055.001.0001>.
- 5 Crystal D. (2006) *Language and the Internet*. Cambridge University Press, 24 p.
- 6 Panyushkina O. (2014) Change of language under the influence of social networks (On the material of the Spanish language). *Almanac of Modern Science and Education*, vol. 3, no. 82, pp. 133–135.
- 7 Ivanov L. (2003) Language in electronic means of communication. In *Culture of Russian Speech*, Moscow: Flint, pp. 791–793.
- 8 Syzdykova R. (2014) The language norm and its normalization (codification). In *Collection of Multi-Volume Work*. Almaty: El-shezhire.
- 9 Balakaev M. (2007) *Qazaq adebi tili*, pp. 396–397.
- 10 Uali N. (2007). *Kazakh Soz Madeniyetinin Theory of Negizderi*: Philol. Gyl. D-ry... Diss: 02/10/02. Almaty, 328 p.
- 11 Alkebaeva D., Satkenova, Zh. B., & Konyrbekova, T. (2021) Changes in the Kazakh language in social. *Al-Farabi Kazakh National University, Kazakhstan, Almaty, Philology Series*, no. 2(182).

- 12 Malik M., Alkebaeva D., & Ahkmetkaliyeva, L. (2019) Pragmatic discourse of the comments on Facebook in the Kazakh language. *Opción*, vol. 35, no. 90-2, pp. 369–384. ISSN 1012-1587/ISSNe: 2477-9385.
- 13 Kurkimbayeva A.M., Akhatova B.A., Gumovskaya G.N., Koteniyatkina I.B., & Khuziakhmetov A.N. (2020) Communicative strategies in modern linguistics (on the material of English travel blogs). *XLinguae Journal*, vol. 13, no. 2, pp. 33–49. <https://doi.org/10.18355/XL.2020.13.02.03>.
- 14 Kurmanbekova Z., & Shokabayeva S. (2022) K Linguistic Analysis of Social Network Communication. *International Journal of Society, Culture & Language*.
- 15 Pastukhova L.S. (2010) Fundamentals of Speech Culture and Stylistics, pp. 127–244.
- 16 Zhanpeissova N.M. (2023) Adaptation of foreign borrowings from the Russian language in Kazakh Internet discourse. *Media Linguistics*.
- 17 Mutanov G. (2021) Multi-Class Sentiment Analysis of Social Media Data with Machine Learning Algorithms. *Computers Materials & Continua*.
- 18 Alkebaeva D. (2021) Changes in the Kazakh language in social networks. *Eurasian Journal of Philology: Science and Education*.
- 19 Narynov S.S. (2016) On One Approach of Solving Sentiment Analysis Task for Kazakh and Russian Languages Using Deep Learning. *International Conference on Computational Collective Intelligence*.
- 20 Imanberdi A. (2023) Approaches to the design of a software system for monitoring and analyzing data of social networks, adapted to the specificity of the Kazakh and Russian languages. *Bulletin of the National Engineering Academy of the Republic of Kazakhstan*.
- 21 Abibullayeva A.A. (2024) Keyword extraction from Kazakh texts using machine learning algorithms. *Bulletin Series Physical and Mathematical Sciences*.
- 22 Toktarova A.B. (2023) Collection of obscene comments in Kazakh language online content using machine learning. *Bulletin Series Physical and Mathematical Sciences*.
- 23 Zotkina A. (2022) Application of machine learning methods in the task of predicting cyberbullying of social network users. *Modern high-tech technologies*.
- 24 Abildaev K.M. (2023) On the issue of borrowing Russian vocabulary in the business sphere of the Kazakh language. *Bulletin of L.N. Gumilyov Eurasian National University. PHILOLOGY Series*.
- 25 Japarova T. (2023) Synergetic analysis of word formation processes as a reaction to Kazakhstan's reality. *Acta Universitatis Lodziensis Folia Linguistica Rossica*.

**<sup>1</sup>Sapakova S.Z.,**

PhD, Associate Professor,  
ORCID ID: 0000-0001-6541-6806  
e-mail: [sapakovasz@gmail.com](mailto:sapakovasz@gmail.com)

**<sup>2</sup>Kambarova B.**

doctoral student, Department of Philology  
ORCID ID: 0009-0008-6153-4934  
e-mail: [gabitqyzy96@mail.ru](mailto:gabitqyzy96@mail.ru)

<sup>1</sup>International University of Information Technologies

<sup>2</sup>Kazakh National Women's Pedagogical University

## **ANALYZING RUSSIAN BORROWINGS IN KAZAKH SOCIAL MEDIA USING MACHINE LEARNING**

### **Abstract**

This study explores the influence of Russian words on the development of the Kazakh language in social networks. The rapid advancement of information technology significantly impacts the language used in online communications. While the chaotic nature of online interactions can complicate language use and create confusion, it also accelerates the spread of information in Kazakh. This research examines how foreign words affect modern Kazakh internet discourse, including direct borrowings that enter the language without modification, mixed

phrases that retain the lexical and semantic properties of foreign words, the emergence of new abbreviations, and the influence of barbarisms. The study utilizes machine learning methods to analyse social media content from Instagram and Facebook. This approach enabled the processing of over 100,000 posts, revealing key linguistic shifts associated with the integration of Russian borrowings into Kazakh. The use of machine learning algorithms, such as the Naive Bayes classifier, automated the data analysis process and uncovered hidden patterns, providing a deeper understanding of how these borrowings affect the Kazakh language in the digital environment.

**Key words:** communication, pragmatics, linguistics, barbarism, Internet, comments, machine learning.

<sup>1</sup>\*Сапакова С.З.,

к.ф.-м.н., ассоц. профессор, ORCID ID: 0000-0001-6541-6806,

\*e-mail: sapakovasz@gmail.com

<sup>2</sup>Камбарова Б.,

докторант, ORCID ID: 0009-0008-6153-4934,

e-mail: gabitqyzy96@mail.ru

<sup>1</sup>Международный университет информационных технологий, г. Алматы, Казахстан

<sup>2</sup>Казахский национальный женский педагогический университет, г. Алматы, Казахстан

## ИЗУЧЕНИЕ ВЛИЯНИЯ РУССКИХ СЛОВ НА РАЗВИТИЕ КАЗАХСКОГО ЯЗЫКА В СОЦИАЛЬНЫХ СЕТЯХ С ПОМОЩЬЮ МАШИННОГО ОБУЧЕНИЯ

### Аннотация

Данное исследование посвящено изучению влияния слов в русском языке на развитие казахского языка. Интенсивное развитие информационных технологий влияет на состояние языка социальных сетей. Во-первых, хаотичность в сети тормозит сферу применения языка и приводит к его искажению, а во-вторых, динамика развития высокая, можно сказать, что информация на казахском языке распространяется быстро. Основная цель статьи – изучение влияния иностранных слов в современном Интернете, языковых операций с переходом с одного языка на другой без каких-либо изменений, смешанных записей (включая слова с сохраненным лексическим и семантическим статусом иностранных слов), новых сокращений, варварства на казахский язык. В результате исследования выявляются формы речи в социальных сетях; в вербальной форме общения определяется вероятность факторов, способствующих искажению языка в связи с одновременным осуществлением момента проецирования мышления и мысли на свет через язык. Применение алгоритмов машинного обучения, таких как классификатор Наивного Байеса, автоматизировало процесс анализа данных и выявило скрытые закономерности, что позволило глубже понять влияние этих заимствований на казахский язык в цифровой среде. В рамках глобальных платформ, отличающихся прогрессивной тенденцией развития информационных технологий, анализируется текстовое содержание в Facebook и Instagram.

**Ключевые слова:** коммуникация, прагматика, лингвистика, варварство, Интернет, комментарии, чат.

Мақаланың редакцияға түскен күні: 07.10.2024