Қожа Ахмет Ясауи атындағы Халықаралық қазақ-түрік университеті

ӘОЖ 004.056 Қолжазба құқығында

ТОҚТАРОВА АЙГЕРІМ БАСТАРБЕККЫЗЫ

Табиғи тілдерді өңдеу және машиналық оқыту әдістерін қолдана отырып онлайн контенттегі ғадауат сөздерді анықтау

8D06115 – Ақпараттық жүйелер білім беру бағдарламасы бойынша

Философия докторы (PhD) дәрежесін алу үшін
дайындалған диссертация

Ғылыми кеңесші
PhD, қауымдастырылған профессор
Омаров Б.С.

Шетелдік ғылыми кеңесші
Доктор, профессор Адали Е.
(Стамбул Техникалық
Университеті, Стамбул)

Қазақстан Республикасы
Түркістан, 2025

МАЗМҰНЫ

НОРМАТИВТІК СІЛТЕМЕЛЕР

Диссертациялық жұмыста келесідей нормативтік құжаттарға сілтемелер жасалынды:

Қазақстан Республикасы МЖМБС 5.04.034–2011. Қазақстан Республикасы Мемлекеттік жалпыға міндетті білім беру стандарты. Жоғары оқу орнынан кейінгі білім. Докторантура. Негізгі ережелер. Қазақстан Республикасы Білім және ғылым министрлігінің 2011 жылғы 17 маусымдағы № 261 бұйрығымен бекітілген.

Қазақстан Республикасы Білім және ғылым министрлігінің Жоғары аттестациялық комиссиясы. Диссертацияны безендіру нұсқаулығы. Қазақстан Республикасы Білім және ғылым министрлігінің 2004 жылғы 28 қыркүйектегі № 377–3 бұйрығымен бекітілген.

МЕСТ 7.32–2001. Ғылыми-зерттеу жұмысының есебі. Құрылымы мен ресімдеу ережелері.

МЕСТ 7.1–2003. Библиографиялық жазба. Библиографиялық сипаттама. Жалпы талаптар және құрастыру ережелері.

АНЫҚТАМАЛАР

Бұл диссертациялық жұмыста келесі терминдерге сәйкес анықтамалар қолданылған:

Аккаунт – сайтқа немесе әлеуметтік желіге тіркелгеннен кейінгі адамның жеке парағы немесе кабинеті.

Аннататор – машиналық оқыту алгоритмдерін үйрету үшін деректерді қолмен сныпқа бөлуші мамандар.

Ансамбльдік әдістер - мәселені шешу үшін бірнеше модельдер біріккен машиналық оқытудағы әдіс.

Буллер – интернет арқылы адамдарға қоқан лоққы көрсетуші адам.

Ғадауат тілді сөздер - діниһе, этникалық тегіне, жынысына немесе басқа факторларға негізделе отырып жеке адамға немесе топқа қарсы қорлау немесе кемсіту ниетінде ауызша немесе жазбаша сөз тіркестері немесе сөйлемдер.

Кибербуллинг – цифрлық құралдарды пайдалану арқылы қорқыту мен бопсалаушылық көрсету.

Конволюциялық нейрондық желілер – кескінді, бейнені, аудионы өңдеу үшін қабаттардан тұратын терең оқыту моделі.

Көпқабатты перцептрондар - кем дегенде кіріс, жасырын және шығыс сияқты үш қабаттан тұратын жасанды нейрондық желілердің класы.

Қайталанатын нейрондық желілер – алдыңғы күй туралы ақпаратты сақтайтын кері байланысы бар әр түрлі ұзындықтағы мәліметтерді өңдейді.

Қысқа мерзімді ұзақ мерзімді жады бар желілер - бұл терең оқытудағы қайталанатын нейрондық желі (RNN) архитектурасының бір түрі.

Лематизация - бұл сөздің барлық ауыспалы формаларын бір мағынаға келтіру процесі.

Машиналық оқыту – ақпараттық техникалық құралдардың адамның қатысуынсыз жаңа деректерді үйреніп, өз бетінше шешім шығара алуы.

Метрика – модельдің қаншалықты жақсы жұмыс жасайтынын анықтау жолы.

Стемминг – сөздің түбірін іздеу.

Твиттер – twitter әлеуметтік желісіндегі пікірлер атауы.

Терең нейрондық желілер – кіріс және шығыс қабаттары арасында бірнеше жасырын қабаттары бар жасанды нейрондық желі.

Токендеу – машинаның адамның тілін түсінуін жеңілдету үшін мәтінді кіші бөліктерге бөлу процесі.

stop words – мағынасы жағынан сөйлемге немесе сөздерге ешқкандай нұқсан келтірмейтін сөздер немесе сөз тіркестер жиынтығы.

БЕЛГІЛЕУЛЕР МЕН ҚЫСҚАРТУЛАР

Бұл диссертациялық жұмыста келесідей белгілеулер мен қысқартулар қолданылған:

API	- application programming interface (қолданбалы бағдарламалау интерфейсі)
AUC	- area under curve (қисық асты ауданы)
CNN	- convolutional neural networks (конволюциялық нейрондық желілер)
CPU	- central processing unit (орталық өңдеу блогы)
DL	- deep learning (терең оқыту)
DNN	- deep neural networks (терең нейрондық желілер)
DT	- decision tree (шешім ағашы)
FN	- false negative (жалған негативті)
FP	- false positive (жалған позитивті)
GPU	- graphical processing unit (графикалық өңдеу блогы)
IDF	- inverse document frequency (кері құжат жиілігі)
KNN	- k-nearest neighbors (k- ең жақын көршілер)
LR	- logistic regression (логистикалық регрессия)
LSTM	- long short-term memory (ұзақ қысқа мерзімді жады)
ML	- machine learning (машиналық оқыту)
NB	- naïve bayes (аңғал байес)
NLP	- natural language processing (табиғи тілді өңдеу)
NN	- neural networks (нейрондық желілер)
RF	- random forest (кездойсоқ орман)
RNN	- recurrent neural networks (қайталанатын нейрондық желілер)
ROC	- receiver operating characteristics (қабылдауыштың жұмыс сипаттамасы)
SA	- sentiment analysis (сезімдерді талдау)
SVM	- support vector machine (тірек векторлы машина)
TF	- term frequency (мерзімді жиілік)
TN	- true negative (шын негативті)
TP	- true positive (шын позитивті)

КІРІСПЕ

Зерттеу жұмысының өзектілігі. Бүгінгі күні онлайн әлеуметтік желілердегі ғадауат тілді сөздерді бар пікірлерді, яғни дініне, этникалық тегіне, жынысына немесе басқа факторларға негізделе отырып жеке адамға немесе топқа қарсы қорлау немесе кемсіту ниетінде ауызша немесе жазбаша сөз тіркестері немесе сөйлемдерді деректер базасы ретінде жинақтау үшін машиналық оқыту алгоритмдерін пайдалану арқылы оңай және автоматтандырылған қадамдарды жүзеге асыруға болады.

Бүгінде, интернет пайдаланушыларының саны артқан сайын, онлайн контенттегі жағымсыз пікір қалдырушылар саны қаптап кетуі мәселе туғызып отыр.

Ғадауат тілді пікірлер немесе адамның психологиялық жағдайына қауіп төндіретін сөздер және сөз тіркестер – интернет немесе әлеуметтік желі құрбандарына құқық бұзушылардың қауіп төндіруі күн сайын артып келеді.

Ғадауат тілді сөздер немесе қоқан-лоққы пиғылды пікірлерді, қауесеттерді тарату, кемсіту немесе қорлау мақсатында жасалған суреттерді, бейнелерді тарату немесе интернеттегі әлеуметтік желілерден адамдарды бұғаттау сияқты бірнеше жолмен көрінеді.

Ғадауат тілді пікірлерді жариялау, бөлісу салдары өте зиянды болуы мүмкін. Жәбірленушілерде алаңдаушылық, үмітсіздік, өзін-өзі бағалаудың төмендеуі және өзіне-өзі қол жұмсау туралы ойлардың белгілері көрінуі мүмкін. Қазақстандық жастардың 75 пайызы интернеттегі қауіптің алуан түрлерін басынан кешірген. Мұны Facebook компаниясы әлеуметтік желі қолданушылары арасында жүргізген онлайн сауалнаманың нәтижесінен байқауға болады. Сонымен қатар, оларда ұйқысыздық, тамаққа тәбеттің азаюы, өз –өзіне көңіл бөлудің жеткіліксіздігі және сабақ оқуда және қарым – қатынасс қалыптастыру өнімділіктері төмендеуі мүмкін. Кейбір жағдайларда ғадауат тілді пікірлердің көптігі физикалық зардаптарға әкелуі мүмкін, себебі жәбірленуші өз ортасынан оқшауалануды сезінуі немесе қауіптің жоғарлауын сезінуі мүмкін. 2024 жылы 165 жасөспірім өз-өзіне қол жұмсап, онлайн контенттен балағаттау немесе кемсіту мағынасына ие пікірлердің салдарынан 378 кәмелетке толмаған жасөспірім өз-өзіне қол жұмсауға әрекеттенген.

Жалпы алғанда, жасөспірімдердің 5% интернеттен ғадауат тілді пікірлерді оқу, көру арқылы келетін қауіпті бастан өткерген немесе айына 2-3 рет немесе одан да көп онлайн қауіп тууы мүмкін. Жасөспірімдердің 12% кем дегенде бір рет интернеттен келетін қауіпке ұшыраған. Егер олардың іс - әрекеті заңсыз деп жіктелсе, әлеуметтік және кәсіби салдарлары, соның ішінде беделіне нұқсан келтіру, достық қарым-қатынас орнатудағы қиындықтар немесе жұмысқа орналасуды қамтамасыз етуде өзіндік жағымсыз салдары бар екенін көруге болады. Статистика көрсеткендей, интернеттен келетін қауіпке тап болғандардың үштен бірінен астамы жеке хабарламалар арқылы қорлау, кемсіту пиғылдары кездесетін хабарлама алған, 24 пайызы өздері туралы кемсіту мақсаты айқын

жазбаларды көрген және үштен бір бөлігі (31%) фотосуреттерінің астында жағымсыз пікірлерге тап болған.

Адамдарға психологиялық тұрақтылықтың салдары туралы білім беру онлайн қауіпті азайтудың бір стратегиясы болып табылады. Білімділікті арттыруға арналған вебинарлар, нұсқаулықтар және тәрбиешілер мен ата-аналарға арналған материалдар өз кезегінде төнетін онлайн қауіптің жетегінде кетпеуге көмектесе алады.

Біз барлығы бір-біріне құрметпен және жанашырлықпен қарайтын ортаны қалыптастыру үшін қолдан келгеннің бәрін жасауымыз керек. Біз өзара сыйластық пен жайлы атмосфераны қалыптастыру арқылы әлемді барлық адамдар үшін жақсы ортаға айналдыра аламыз.

Мысалы, 2024 жылы жүргізілген әлем мемлекеттерінің интернетті қолдану деңгейі статистикасы бойынша Қазақстан мемлекеті 20 075277 миллион жалпы халықтың 18,9 миллион адамы, яғни 92,3% интернетті тұрақты түрде қолданатынын көрсетіп отыр. Соның ішінде әлеуметтік желі қолданушылар саны жалпы халықтың 14 миллионын, яғни, 71,5% құрап отыр.

Онлайн желідегі ақпараттар, жаңалықтар аптасына 7 күн, 24 сағат бойы қолжетімділікке ие, интернеттен қасақана немесе абайсызда күннің кез-келген уақытында келетін «шабуылдан» қорғана алмайсың, себебі электрондық пошта, ұялы телефондардағы мессенджерлер немесе жеке әлеуметтік жерлілерге, ақпараттық – танымдық сайттарға кез- келген уақытта келіп түсу мүмкіндігі әрдайым жоғары. Бұл желіні, интернетті онлайн пайдаланушыларға, оның ішінде жастар мен жеткіншектерге психологиялық түрде кері әсерін бірден тигізуі мүмкін. Сонымен қатар, инетернетте, әлеуметтік желілерде өз аты – жөнінін жасырып, лақап атпен желі қолданушысы бола білу мүмкіндігі бар, яғни интернеттегі «шабуылдың» кімнен келгенін білу көрсеткіші төмендеуі мүмкін.

Эмоционалдық «шабуылдың» зорлық – зомбылықтың физикалық түріне қарағанда адамды психологиялық ауытқуларға әкеліп соғу қаупі жоғары екенін психологтар да мойындап отыр.

Адамдардың психологиялық жағдайына кері әсерін беретін зиянды мәліметтерді анықтау қазіргі таңда ақпараттық техникалық құралдардың негізі міндеті деп алсақ, бұл міндетті іске асырып, өңдеуде машиналық оқыту әдістерін пайдалану арқылы автоматтандыру әлдқайда тиімді деп саналып отыр.

Ғылыми зерттеу жұмысының мақсаты - мәтіндік деректерде ғадауат тілді сөздерді автоматты түрде анықтау үшін терең нейрондық желі моделін құру.

Ғылыми зерттеу жұмысының міндеттері: Зерттеу мақсатына жету үшін келесі міндеттер шешілді:

1. Ғадауат тілді пікірлерді анықтауға арналған екілік және көп класты жіктеуге арналған машиналық оқыту алгоритмдерін талдау.

2. Ғылыми зерттеу жұмысы үшін ML және DL алгоритмдерін оқыту үшін қазақ тілінде деректерді жинау және алдын ала өңдеу.

3. Терең оқытудың архитектурасына талдау жүргізу. Терең оқыту алгоритмдері түрлеріне оқытуды жүргізу, мысалы:

а) конволюциялық нейрондық желілер үшін;

ә) терең нейронды желілерде;

б) қайталанушы нейронды желілерде;

г) қысқа және ұзақ мерзімді жадысы бар желілерде;

ғ) көпқабатты перцептрондар үшін;

д) жасанды интеллекттің механизмдерін пайдалана алатын терең нейронды желілерде;

ж) модельдің нәтижесін жақсартуда эксперименталды зерттеулерді жүргізу, салыстырмалы жұмыс жүргізу, үлгіні таңдау, гиперпараметрлерге баптау жүргізу.

Зерттеу тақырыбы: Мәтіндік деректерде ғадауат тілді пікірлерді анықтау үшін машиналық оқыту және терең оқыту алгоритмдерін қолдану.

Зерттеу нысаны: Танымал әлеуметтік танымдық желілер (Instagram,Tik-tok, Facebook, Youtube, Twitter), жаңалықтар порталдары мен сайттар (nur.kz, tengri news, serke.org, countr.kz).

Зерттеудің теориялық маңыздылығы: Мәтіндік деректердегі ғадауат тілді пікірлерді анықтау бойынша аз ресурсты тілдерді зерттеуге арналған ғылыми зерттеу жұмыстарды зерттеу, табиғи тілдің өңдеу құралдарына талдау жүргізу.

Зерттеудің практикалық маңыздылығы: терең нейрондық желілерді әзірлеу және оқыту, табиғи тілді өңдеу, әлеуметтік медиа кеңістігінде ғадауат тілді пікірлерді анықтау бойынша эксперименттер жүргізу.

Зерттеу жұмысының ғылыми жаңалығы:

- қазақ тілі үшін деректер жинағы жасақталды, алдын ала өңделді және одан әрі машиналық және терең оқыту тапсырмалары үшін қолмен сыныптарға бөлінді;
- ғадауат тілді сөздерді анықтау тапсырмасында екілік класты жіктеу тапсырмалары үшін назар аудару механизмі бар терең нейрондық желі әзірленді және оқытылды;
- ғадауат тілді анықтау тапсырмасында машиналық оқыту мен терең оқыту алгоритмдері арасында салыстырмалы талдау жүргізілді.

Докторанттың жеке қосқан үлесі.

Диссертация тақырыбына сәйкес әдеби материалдар мен патенттік дереккөздерді талдау және жалпылау; талдау және зерттеу әдістемелерін таңдау; ғылыми-зерттеу және практикалық тапсырмаларды орындаумен қатар теориялық және эксперименттік ғылыми зерттеулерді орындау.

Жұмыс тақырыбы бойынша жариялымдар. Диссертациялық жұмыс бойынша 13 ғылыми зерттеу мақалалары жарық көрді, оның ішінде SCOPUS базасына енетін журналдарда – 2 мақала, халықаралық конференцияларда -3 ғылыми зерттеу мақаласы, сонымен қатар, ҚР ҒБМ Ғылым және Жоғары Білім саласындағы сапаны қамтамасыз ету комитеті ұсынған баспаларда – 8 мақала жарық көрді:

1. Toktarova, A., Abushakhma, A., Adylbekova, E., Manapova, A., Kaldarova, B., Atayev, Y., ... & Aidarkhanova, A. (2023). Offensive language identification in low resource languages using bidirectional long-short-term memory network.

International Journal of Advanced Computer Science and Applications, 14(6). DOI: http://dx.doi.org/10.14569/IJACSA.2023.0140687

2. Toktarova, A., Syrlybay, D., Myrzakhmetova, B., Anuarbekova, G., Rakhimbayeva, G., Zhylanbaeva, B., ... & Kerimbekov, M. (2023). Hate speech detection in social networks using machine learning and deep learning methods. International Journal of Advanced Computer Science and Applications, 14(5).

3. Toktarova, A., Sultan, D., & Azhibekova, Z. (2024, May). Review of Machine Learning Models in Cyberbullying Detection Problem. In 2024 IEEE 4th International Conference on Smart Information Systems and Technologies (SIST) (pp. 233-238). IEEE.

4. Sultan, D., Suliman, A., Toktarova, A., Omarov, B., Mamikov, S., & Beissenova, G. (2021, January). Cyberbullying detection and prevention: Data mining in social media. In 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence) (pp. 338-342). IEEE.

5. Sultan, D., Mussiraliyeva, S., Toktarova, A., Nurtas, M., Iztayev, Z., Zhaidakbaeva, L., ... & Omarov, B. (2021, May). Cyberbullying and hate speech detection on kazakh-language social networks. In 2021 7th IEEE Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing,(HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS) (pp. 197-201). IEEE..

5. Toktarova A.B., et al. Cyberbullying Detection and Prevention: Data Mining in Social Media // Proceedings of the 7th IEEE Intl Conference on Big Data Security on Cloud (BigDataSecurity), High Performance and Smart Computing (HPSC), and Intelligent Data and Security (IDS). – 2021. – P. 338–342. – DOI: 10.1109/BigDataSecurityHPSCIDS52275.2021.00045.

6. Toktarova A.B., Omarov B.S., Kazbekova G.N., Mamikov S.A., Temirbekova F.E. Collecting hate speech database on social network in Kazakh language by using machine learning // Bulletin of the National Academy of Sciences of the Republic of Kazakhstan. Series of Physics and Mathematics. – 2023. – No. 1. – P. 191–203. – DOI: https://doi.org/10.32014/2023.2518-1726.177.

7. Тоқтарова А.Б., Ажибекова Ж.Ж., Султан Д.Р., Керимбеков М.А. Онлайн контенттегі қазақ тілді бейәдеп пікірлерді машиналық оқытуда жинақтау // Абай атындағы ҚазҰПУ Хабаршысы. Сериясы: Физика-математика ғылымдары. – 2023. – Т. 81, №1.

8. Тоқтарова А.Б., Омаров Б.С., Ажибекова Ж.Ж., Мамиков С.А. Онлайн контенттегі ғадауат сөздерді анықтауда жасанды интеллектің маңызы // Торайғыров атындағы университет хабаршысы. Энергия сериясы. – 2023. – №1. – Б. 311–322. – ISSN 2710-3420.

9. Тоқтарова А.Б., Омаров Б.С., Ажибекова Ж.Ж. Желі қолданушыларының «эмоциялы» пікірі арқылы автоматтандырылған ғадауат сөздер классификациясы // Қазақстан Республикасы Ұлттық ғылым академиясының Хабаршысы. – 2023. – №2 (88). – DOI: https://doi.org/10.47533/2023.1606-146X.9.

10. Тоқтарова А.Б., Омаров Б.С., Қалдарова Б.А. Бейәдеп сөздерді аз ресурсты тілдерден анықтауда BiLSTM-ді қолдану // Қазақстан Республикасы Ұлттық ғылым академиясының Хабаршысы. Сериясы: Физика және математика. – 2024. – №3. – Б. 174–189. – DOI: 10.32014/2024.2518-1726.299.

11. Toktarova A.B., Omarov B.S., Beissenova G.I., Abdrakhmanov R.B. Analysis of hate speech words in online content by using data mining // Bulletin of the National Academy of Sciences of the Republic of Kazakhstan. Series of Physics and Mathematics. – 2023. – No. 2 (346). – P. 237–251. – DOI: https://doi.org/10.32014/2023.2518-1726.196.

12. Тоқтарова А.Б., Омаров Б.С., Темірбекова Ф.С. Қазақтілді бейәдеп сөздер классификациясы және машиналық оқыту әдісін оларды анықтауға бейімдеу // Абай атындағы ҚазҰПУ Хабаршысы. Сериясы: Физика-математика ғылымдары. – 2023. – Т. 82, №2.

13. Toktarova A., Azhibekova Zh., Aliyeva A., Sarsenbieva N. Bidirectional Long Short-Term Memory in Hate Speech Detection Problem on Networks // Bulletin of KazNPU named after Abai. Series: Physical and Mathematical Sciences. – 2024. – Vol. 87, No. 3. – DOI: 10.51889/29595894.2024.87.3.010.

Ғылыми зерттеу жұмысының көлемі мен құрылымы: Диссертациялық жұмыс кіріспеден, 5 бөлімнен, талқылаулар мен қорытындыдан тұрады. Ғылыми зерттеу жұмысының толық көлемі: 123 беттік машиналық мәтіннен, оның ішінде 36 сурет, 8 кестеден, 139 пайдаланылған әдебиеттер тізімінен тұрады.

1 ОНЛАЙН КОНТЕНТТЕГІ ҒАДАУАТ СӨЗДЕРДІ АНЫҚТАУДЫҢ ЗАМАНАУИ ЖАҒДАЙЫ МЕН МӘСЕЛЕЛЕРІ

Онлайн контентте ғадауат тілді пікірлер жариялау - қоғамды бөліну мен зорлық-зомбылыққа әкелетін белгілі бір қасиеттерге негізделген жеке адамға әділетсіз қарым-қатынасты қамтитын кемсітушіліктің бір түрі. Ол кез келген қоғамдық, этникалық, діни, жыныстық, әлеуметтік немесе саяси топтың мүшелерін қорлауға және қорлауға бағытталған БАҚ немесе онлайн платформалардағы көрнекі бейнелерді, қорлайтын сөздерді, сөз тіркестері мен сөйлемдерді қамтиды. Қазақ қоғамындағы ғадауат тілді сөздерді жазылу, айтылу және қолдану ерекшеліктеріне байланысты түрлері бар екендігі анықталған. Ғадауат тілді пікірлер – мәтіндік хабарламалар мен әлеуметтік медиа платформалары арқылы қудалау мен қорқытуды қамтитын жек көрушілік сөйлеудің тағы бір түрі.

Онлпйн контенттегі жағымсыз ахуал қоғамдағы экономикалық, әлеуметтік және саяси мәселелерге айтарлықтай әсер етеді. Ол ғаламторда, әсіресе қазақ тілді әлеуметтік желіні қолданушылар арасында кеңінен таралу интернетті пайдалану деңгейі артқан сайын өсіп келеді. Психологтардың айтуынша жағымсыз пікір таратушылар психологиялық проблемалары бар, әлеуметтік теңсіздікке ұшыраған, кемсітушілік, жеке өмірінде мәселелер туындаған, отбасылық мәселелерге байланысты күйзеліске ұшыраған, балалар мен ата-аналар арасындағы қарым-қатынастардың қиындаған ойнлайн желі қолданушылары болып табылады деген тұжырымға келген.

Онлайн контенттен келетін қауіп қатер – бұл балалар мен жасөспірімдерге ғана емес, ересектерге де әсер ететін, өзіне деген сенімділігінің төмендеуіне, өзін дарынсыз сезінуге және тіпті суицидке әкелетін күрделі мәселе.

ЮНИСЕФ онлайн қорлау мен кемсітудің 10 түрін анықтады, оның ішінде желідегі әзілдер мен ғадауат тілді сөздерді ажырата білудің жолдарын көрсетті.

Онлайн контентте қорлау мен кемсітуге ұшыраған желі пайдаланушылары қысым көрсетушіні блоктауға тырысып, олардың әрекеті туралы көмек көрсету орындарына хабарлап, жағымсыз пікірлердің психологиялық тұрғыдан залал келтіруіне жол бермеу қажет. Әлеуметтік медиа платформалары пайдаланушылардың қауіпсіздігін қамтамасыз етуге міндетті және оларды онлайн контенттен келетін шабуылдардан қорғау мақсатында жеке ақпараттарды сақтауда құпиялылығы жоғары функциялар ұсынып отыр. Онлайн контенттен келетін қауіп – қатерге ұшырағанжасөспірім ді де, ересек адамды да қолдау және оны тыңдауды, оған сенімді ересек адамнан көмек сұрауға немесе кеңесші көмегіне жүгіуді ынталандыру және қолдау мен ресурстарды ұсынуды қамтиды.

1.1 Онлайн контенттегі ғадауат сөздерді анықтаудағы негізгі мәселелері

Ғадауат тілді сөздер мағынасын әркім әрқалай түсіндіреді. Сондықтан алдымен «ғадауат тілді сөздер» сөз тіркесінің концептуалды мағынасын ашып алу қажет.

Оны журналистика мектебінің мұғалімі Оразай Қыдырбаев былай түсіндіреді: «Ғадауат тілді сөздер – адамның бойында әлдебір адамға пайда болатын жек көрушілік, кемсітушлік, кейде біреуге зиян тигізу ниеті»- деп, түсіндіреді [1].

Орталық Азиядағы Бейбітшілік және медиатехнологиялар мектебінің директоры, танымал медиа-сарапшы Инга Сикорскаяның айтуынша, ол: «Ғадауат тілді сөздер – бұқаралық ақпарат құралдарында немесе интернетте қоғам өкілдеріне, кез келген этникалық, діни, жыныстық, әлеуметтік немесе саяси топқа қарсы жағымсыз және қорлау мақсатындағы визуалды бейнелер немесе өшпенді сөздер мен риторика»- деген тұжырымға келеді [2].

Ғадауат тілді сөздер туралы айтқан кезде кемсітушілікті (дискриминацияны) ұмытуға болмайды. Ғадауат тілді сөздер кемсітушіліктің көрінісі болып саналады. Ал кемсіту – белгілі бір топты немесе жеке адамдарды негізсіз шеттету, қорлау немесе теңсіздікті көрсету.

Сондай – ақ, «ғадауат» сөзі күнделікті өмірде жиі қолданыла бермейтіндіктен, оның мағынасы мен шығу тегін талдадық. А.Т. Қайдаров тың «арабша-қазақша түсіндірме сөздігінде» көрсетілгендей [3,4], «қазақ тілі үшін араб сөздері» сөздігіндегі II томын алып қарастырсақ: ««Ғадауат» сөзі – 1) дұшпандығын көрсету, қастық жасау, жауласу; 2) алауыздық таныту, келіспеушілігін білдіру, өшпенділікте болу» деп ашып көрсеткен. Ал ағылшынша-арабша сөздікке келер болсақ, бұл сөзді «a state of deep-seated ill-will» деп аударған [5,6]. Сөздік мағынасында оны «махаббатқа қарсы сезім» немесе «антимахаббат» деп түсінуге болады.

«Ғадауат» сөзі ағылшын тіліне тікелей аударылғанда «hate speech» ұғымы пайда болады. Кембридж сөздігі [7] оны «нәсілі, діні, жынысы немесе жыныстық бағдары сияқты белгілер негізінде адамдарға немесе топтарға қарсы өшпенділік білдіретін немесе зорлық-зомбылыққа шақыру мақсатында көпшілік алдында сөйлеу» деп анықтайды [8]. Қазақ тіліне аударып қарайтын болсақ, бұл ұғым адамның нәсіліне, ұстанған дініне, оның жынысына немесе ұстанатын жыныстық бағдарына негізделе отырып, адамдарға немесе топтық өкілдерге өшпенділігін, жаулығын және жек көрініш сезімін білдіретін және зорлық-зомбылыққа негіздейтін сөздер немесе сөз тіркестері ретінде қарастырған.

1.2 Ғадауат тілді сөздер және классификациясы

Ғадауат тілді сөздер және оның жіктелуі

Ғадауат тілді сөздерге *дискриминация* (адам құқығын шектейтін орынсыз сөздер арқылы кемсіту), *қорлау* (теріс сөздер арқылы кемсіту, ар-намысты қорлау), *кибербуллинг* (интернет арқылы адамды мазақ ету, қорқыту немесе қысым көрсету), *экстремизм* (діни бағытына немесе саяси жағдайда шектен тыс теріс пікір білдіру) және *радикализм* (қоғамдағы саяси жүйені сынау) жатады [9].

Дискриминация (лат. «discrīminātio» – оқшаулану, бөлектену) – белгілі бір белгілері бойынша адамды кемсіту, құқықтарынан айыру немесе бостандығын шектеу [10]. Қазіргі кездегі дискриминацияның ең көп тараған түрлерін келесідей жіктеуге болады:

1. *Бой өлшемі бойынша кемсіту* – адамды бойына қарап келеке, мазақ ету. Мысалы, «қысқа», «қортық» сияқты сөздер қатары.

2. *Жасы бойынша (эйджизм)*- адамды оның жасына бойынша басқалардан өзгеше етіп көрсетуге тырысу, егде, орта жастағы адамдарды, кішкентай балаларды онлайн контентте әлімжіттік көрсетуде жиі жазылып, жарияланатынын байқауға болады. Мысалы, «кәртиген кемпір», «ақсақ шал», «кәрі қарға», «алжыған қақпас екенсің», «қаршадай болып, бәлесі басынан асады» деген сияқты сөз кемсіту мен қорлау ниеті бар сөз тіркестері кездеседі. Сонымен қатар, жұмысқа қабылдау жағдайында, егде адамдардың жас ерекшеліктерін басшылыққа алып, жұмыс шартына сәйкес келмейтінін алға тартады.

3. *Касталық дискриминация* – белгілі бір аймақта немесе аз мөлшерде тұратын топтардың шығу тегіне байланысты кемсіту және құқықтарын шектеу.

4. *Мамандығы бойынша дискриминация* – жұмыс берушінің қызметкерді «біліксіз», «қабілетсіз» деп кемсітуі.

5. *Мүгедектікке байланысты кемсіту* (эйлизм немесе дисаблизм) – адамды физикалық ерекшеліктеріне қарап мазақ ету. Мысалы, «ақсақ», «құбыжық», «жынұрған» сияқты сөздер.

6. *Нәсілдік дискриминация (нәсілшілдік)* – адамның терісінің түсіне, ұлтына немесе сыртқы белгілеріне байланысты кемсіту. Нәсілшілдік пен ұлтшылдықты ажырата білу керек, өйткені нәсілшілдік адамды этникалық тегіне қарай кемсітеді.

7. *Діни сенімге негізделген дискриминация* – белгілі бір дін өкілі өз сенімін басқалардан жоғары қойып, басқа дін өкілдерін кемсіту.

8. *Жынысы бойынша кемсіту (сексизм)* – адамды жынысына байланысты кемсіту немесе ерлер мен әйелдерді тең құқылы деп танымауы, бірінен екенішсін артық бағалау.

9. *Гетеросексизм немесе гомофобия* – адамды жыныстық бағдарына байланысты кемсіту, мазақ ету немесе құқықтарын шектеу.

10. *Тілдік кемсітушілік* – белгілі бір тілде сөйлей алмағаны немесе басқа тілде сөйлегені үшін адамды мазақ ету немесе кемсіту.

11. *Позитивті кемсітушілік* – мемлекеттің белгілі бір топтарға ерекше қолдау көрсетуі. Мысалы, халық арасынан аз қамтылған отбасыларға жеңілдіктер немесе ай сайынғы жәрдемақы тағайындалып отыруы [7].

Кемсітудің осы түрлерінің әрқайсысы жеке адамдардың немесе топтардың құқықтарына қол сұғу арқылы қоғамда бөліну тудыруы мүмкін.

Қазақ қоғамындағы басқа халыққа ұқсамайтын өзіндік дискрминация түрі бар деп қарастыруға болады, яғни, бір – бірінен ру сұрасу арқылы туындайтын шектетулер [5]. Адамынң руын сұрау арқылы сұраушы адам ұнатпайтын, іштей жақтырмайтын ру өкілі болып шықса, оны алшақтатуға, кемсітуге, кей жағдайда

белгілі бір шектеу көрсетіп қалуға, тіпті қорлауға дейін баратын жайттарды кездестіруге болады.

Сонымен қатар, жерлес адамдар басқа елді мекендердің тұрғындарын өз қатарына қоспай, олардың ерекшеліктерін кемсіту, қорлау ниеті бар сөздер арқылы оларды бөліп тастау, қатарына қоспау пиғылы бар екенін білдіріп жататын жайттарды да кездестіруге болады [5].

Ағылшын тілінен аударғанда қысым, қорлау және қудалау деген мағынаны білдіретін кибербуллинг – әлеуметтік желідегі мессенджерлерді пайдаланып, кейде қудалау деңгейіне жететін мәтін арқылы қорқыту мен қысым көрсету. Кибербуллинг жеке адамдардың жасына, ұлтына, нәсіліне немесе әлеуметтік мәртебесіне қатысты кемсітушілік болып табылады [11].

Интернетті басқаларды әдейі ренжіту, қорлау немесе қысым көрсету үшін пайдалану кибербуллинг болып саналады. Біреу туралы жағымсыз жазбаны ұнату немесе қорлайтын хабарлама жіберу кибербуллингтің бір түрі ретінде қарастырылуы мүмкін.

Oxford Advanced Learner Dictionary [12], Merriam-Webster Dictionary [13] және Collins сөздігі [14] арқылы «ғадауат» сөзінің жіктелуі төмендегі кесте 1- де көрсетілген.

Кесте 1.1 - Әлем сөздіктерінде «ғадауат» сөзінің мағынасы

1	2
Тыйым сөздер ретіндегі сөздер тізбегі (табу) және ұятты сөздер	Бір немесе бірнеше сөзердер тізебегінен құралған сөз тіркестері жәбірленушінің әлеуметтік беделін төмендетуге және олардың кемшіліктерін басқаларға паш ету арқылы манипуляциялауға бағыттай отырып, агрессиялық құрал ретінде қарастырады. Бұл лексикалық тіл нағыз ғадауат тілді сөздерді қамтиды. Мысалы, «м*лғұн», «т*пас», «б*йбақ», «м*ңгүрт» және «ж*леп қ*тын» сияқты сөздерді кездетіруге болады [5].
Жыныстық бағдары бойынша және гендерлік кемсітушілікке байланысты терминология	«Жәбірленушілер» ретінде анықталған адамдар «гей», «лесбиян», «транс», «педофил/педик» және «сексист» сияқты терминдерді қолданып, өздерінің дәстүрлі емес жыныстық бағдарларына қатысты пікірлер жиі сол тұлғаларға бағытталып жазылады.
Зұлымдық жасау пен өлім тілеу ниеті бар сөздерден тұратын сөз тіркестері	Бұл лексикалық етістіктердің қолданылуы бұзақының жәбірленушіге деген жиіркенішінен немесе оларға деген терең дұшпандық сезімінен туындауы мүмкін. Мысалы, «*_ылып _ліп қал», «т*_ысың бітіп қалсын», «*_генің артық», «*_сейші», «*_іп қалшы», «_ені жер қалай _өтеріп жүр» және т.б. сияқты лексикалық сөз тіркестерінің қолданылуы желі арқылы буллердің жәбір көрушіден артықшылығын, одан жоғары екендігін көрсеткісі келгенде, ұлылығын, басқалардан биік тұратындығын дәлелдегісі келетінін байқатады [8].

1.1-кестенің жалғасы

1	2
Кемсіту пиғылы және қорлау ниетттегі сөздер тізбегі	«Бірнеше жауыз – бір құрбандық» бағыты бойынша онлайн контентте қудалауға ұшыратқан кезде кезде жиі қолданылатын сөздермен сөз тіркестері. Жәбірленушінің адами қасеттерін төмендетуді білдіретін ғадауат тілді сөздер немесе сөз тізбектері. Мысалы, «с_м», «с_мелек», «н_дан», «т_пас», «ақ_мақ», т.б. лексикалық сөз тіркестері. Осылайша, буллер жәбірленушден күшті екенін көрсетуге тырысады [5, б. 291].
Ұлтты кемсітуге және нәсіліне байланысты ұғымдарды біріктіретін сөздер немесе сөз тіркестері	Ұлттың, нәсілдің ерекшелігін пайдалана отырып, олдарды сөздер арқылы кемсіту. Мысалы, «негір», «өзбек», «сарт», «ит қытай» және т.б. – агрессия көрсетуші жәбірленушіге моральдық зиянын тигізуі мүмкін [5].
Адамды жануарларға теңеу арқылы кемсіту	Мұндай сөз тіркестері желі пайдаланушысын қорлау, кемсіту мақсатында қолданылады, мұндай лексикалық сөз тіркестері дәстүрлі түрде қолданылатын балағат сөздермен қатар инвективті сөздер санатына өтеді. «С_ыр», «ес_к», «ш_шқа», «қ_й», «ж_бырлаған құрт», «с_р жылан» т.б. сияқты жәбірленушіні жануарларға теңеу арқылы адамды кемсітіп, мазақ ету, мұндай сөз қатары адамның қоғамдағы мәртебесін шектен тыс төмендеуіне әкеліп соғуы мүмкін [15].
Физикалық тұрғыдан және ақыл-ой жеткіліксіздігі бар адамдарды кемсіту, мазақтау мақсатындағы сөз тіркестері	Мұндай сөз тіркестерін қолданудың негізгі мақсаты: жәбірленушінің ар-намысына тие отырып, оның қадір-қасиетін түсіру негізінде, әлеуметтік беделіне нұқсан келтіру үшін, тұлғалық қасиетін шектеу үшін қолданылады. Қорлаушы, жәбірлеуші (буллер) жәбірленушінің ақыл – ой немесе физикалық кемістігін байқағанда қолдануға тырысады. Бұл сөздер тобына «к_рең», «м_гедек», «т_пас», «ақ_мақ», «м_ңғұрт», «әлс_з әлжу_з» сөздер тізбегі кіреді.
Экстермизм	Жалпы халыққа ортақ тәртіпті мойындағысы келмейтін, өзінің ой - пікірін ғана дұрыс деп есептеп, сол бойынша іс-әрекетті жасаушы дегенді білдіреді. Экстремизм адамның надандыққа, көрсеқызарлыққа ұмтылудағы көзқарасынан пайда болады. Яғни, адам тек қана өз пікірін дұрыс деп санайды және басқа адамдардың өзінің пікірімен жүргенін қалап тұрады. Экстремизмнің тағы бір қосымша белгісі ретінде басқа адамдарды жетістігін, қабілетін көре алмаушылық, түсінбеушілік немесе түсінісіп, ортақ мәмлеге келуден бас тартатын адамдар қатарын жатқызуға болады [16].

Сонымен қатар, экстремизмді ұйымдастырушы лаңкесшілер дұрыс бағытты ұстанып, жол көрсетуші адамдардың кеңесіне мүлдем құлақ асудан бас тартады. Олар өздерін ең таза, кіршіксіз, күнәсіз ретінде көріп, басқа жандардың адасушылар қатарына жатқызып, жөн білетін адамдарды мойындағысы

келмейді. Мұның салдары әлемде түрлі төңкерістер мен қантөгістер орын алып жатқандығын байқауға болады [17].

Экстремистік әрекеттің негізгі түрлері:

1) діни экстремизм – кез келген діннің кері жағы, басқа діни идеяларды қатаң түрде теріске шығаруға, басқа конфессия өкілдеріне агрессивті көзқарас пен мінез-құлыққа, «шындықты» насихаттауға бағытталған қараңғы идеялар мен пікірлер тарату, кей жағдайда басқа конфессия өкілдерін физикалық қырып-жоюға дейін ұмтылу. Халықтың немесе бірнеше халықтардың өкілдері белгілі бір діннің әлеуетті жақтастары, ал қалғандарының бәрі оның қарсыластары болып саналады;

2) саяси экстремизм – саяси партиялар мен топтардың, сондай-ақ лауазымды адамдар мен қарапайым азаматтардың қалыптасқан мемлекеттік жүйені күштеп өзгертуге, қалыптасқан мемлекеттік құрылымдарды жоюға және тоталитарлық тәртіп диктатурасын орнатуға бағытталған, ұлттық және әлеуметтік араздықты қоздыруға бағытталған заңсыз әрекеттері мақсатында пікірлер таратып халықтың санасын улау;

3) діни-саяси экстремизм - мемлекеттік құрылысты күштеп өзгертуге немесе билікті күштеп басып алуға, мемлекеттің егемендігі мен аумақтық тұтастығын бұзуға, осы мақсаттарда діни араздықты және өшпенділікті қоздыруға бағытталған іс-әрекеттер тзбегі мен идеяларды халық арасында насихаттап, күш біріктруге шақыру;

4) ұлтшылдық экстремизм – әрқашан саяси экстремизм элементтерін және діни экстремизм белгілерін қамтиды [18].

Ал, радикализм - адамның немесе топтың (партияның) қалыптасқан әлеуметтік, саяси және мәдени жағдайды түбегейлі және ымырасыз өзгертуге ұмтылуынан тұратын ұстанымы [19].

Радикализмді оң және сол деп екі түрге бөліп қарастырамыз:

Оңшыл радикализм - қаншалықты қарама-қайшы болып көрінсе де, сол немесе басқа қоғамның қандай да бір дәстүрлі, архаикалық құрылымын қайтаруға ұмтылу болып табылады.

Солшыл радикализм - сол жақтағы радикалдар оңға тура қарама-қарсы болып келеді. Яғни, дәстүрлі жүйеге көшу арқылы қалыптасқан жүйені түбегейлі өзгертуді жақтаса, көбінесе діннің қоғамдағы рөлі үлкен (діни экстремизм), онда біріншілері өзгертуге және кейіннен түбегейлі жаңа жүйені орнатуға ұмтылады [20].

1.3 Онлайн контенттегі қазақ тілді ғадауат сөздердің тигізетін әсері және түрлері

Жалпылама ғадауат сөздің не екенін және олардың классификациясы туралы сөз қозғадық, ендігі кезекте онлайн контенттегі ғадауат сөздердің тигізетін әсері мен олардың түрлеріне тоқталамыз.

Интернетте қол жетімді әлеуметтік медиа платформалары біздің қазіргі өмірімізде үлкен рөл атқарады және біздің күнделікті іс - әрекетімізге айтарлықтай әсер етеді. Виртуалды коммуникациялық платформаларды

пайдалану адамдарға соңғы жаңалықтар туралы білім алуға, қызықты және танымдық сәттерді бөлісуге және бір-бірімен ашық сөйлесуге мүмкіндік береді. Екінші жағынан, бұл платформаларда адамдарға немесе ұйымдарға әртүрлі дәрежеде зиян келтіруі мүмкін. Яғни, дұшпандық сипаттағы зорлық-зомбылыққа ниеттелген пікірлер мен сөз тіркестері жиі кездеседі. Интернет саласында туындайтын маңызды мәселелердің бірі, мысалы, басқа адамды қасақана қорлау немесе балағат сөздерді тарату [5].

Желіде (виртуалды) жүзеге асатын коммуникацияның өзіндік ерекшеліктері бар. Байланыс оның виртуалдылығы мен анонимділігін қамтитын осы ерекшеліктермен сипатталады. Басқаша айтқанда, адамдар жалған жеке басын куәландыратын және бір уақытта көптеген «лақап аттарды» пайдалана алады. Мұндай атмосферада жұмыс істеу кезінде агрессивті әдістер мен ауызша шабуылдарды қолдану онлайн контенттегі әдеттегі тәжірибе болып табылады. Мұның бәрі қазақта «жеккөрінішті сөз» немесе «ғадауат сөздер» деп аталатын мысалдар.

Ғадауат сөздердің тигізетін әсері және оларды анықтау [21]

Онлайн қарым-қатынас саласында төрт агрессивті қарым-қатынас стратегиясы бар:

1. *Холивар (ағылшын тілінен аударғанда Holy war – қасиетті соғыс)* : Бұл бір-біріне түбегейлі сәйкес келмейтін қарама-қайшы пікірлері бар адамдар арасындағы ешқашан аяқталмайтын дау. «Шабдалы мен шабдалы шырыны» мысалын алайық: қайсысы адам денсаулығына пайдалы? тақырыбында мәңгі дауласуға болады.

2. *Троллинг* - әлеуметтік арандатудың, кемсітудің немесе қорлаудың бір түрі. Оны атақ-даңққа ие болғысы немесе назар аудартқысы келетін адамдар, сондай-ақ анонимді болып қалатын пайдаланушылар жасайтын жағдайлар да кездеседі.

3. *Флуд - (ағылшын тілінен аударғанда Flood- су тасқыны)* термині ортақ әңгіме тақырыбына мүлде байланыссыз және мазмұны жағынан ешқандай маңызы жоқ хабарламалар жиынтығын білдіреді. Басқа пайдаланушылардың назарын аудару немесе алаңдату үшін онлайн форумдар мен чаттарда «су тасқыны» термині жиі қолданылады.

4. *Флейм (ағылшын тілінен аударғанда Flame - алау)* бұл әңгіме барысында болатын «ауызша соғыс». Көп жағдайда мұндай пікірталастардың бастапқы себебі оның қорытындысынан кейін болатын талқылау үшін маңызды емес.

5. *Вербальды агрессия (сөз арқылы қарым – қатынас орнату)* осы тақырыпты зерттеуші Ю.В. Щербинина [22] ғадауат тілде сөйлеу адамдардың белгілі бір топтарға, мәдени қауымдастықтарға немесе әлеуметтік қоғамға деген өшпенділік сезімін жеткізуге арналғанын айтады. Мұндай сөздер әртүрлі контексттерде, соның ішінде ресми және кәсіби параметрлерде, сондай-ақ қарапайым әңгімелерде де кездеседі. Саясаткерлердің сөйлеген сөздері, зейнеткерлердің сөйлесулері, студенттер мен оқушылардың әңгімелері, тіпті отбасы мүшелері арасындағы қарым-қатынастарды қоса алғанда, әртүрлі контексттерде өшпенділік пиғылдағы сөз тіркестері кездесуі мүмкін [5].

Зерттеулерге сәйкес, ғадауат тілді сөздердің қалыптасуы қоғамда болып жатқан экономикалық, әлеуметтік және саяси мәселелерге қатты әсер ететіні көрсетілген. Офлайн кеңістікте айтылған жағымсыз пікірлер интернетте жылдам таралады және олардың байланыс түрі ретінде пайдалану жиілігі үнемі өсіп келеді. Мәселен, Medianet халықаралық журналистика орталығы жүргізген зерттеудің қорытындысына сәйкес, Facebook желісі қазақ тілді бөлімде адамдарды ұлтына немесе дініне қарай кемсітуге бағытталған өшпенділік сөздері басымырақ екені анықталған [23].

Қоғамның тұрақтылығына өшпенділік әсері нашар әсер етеді, бұл жеке адамдар арасындағы антагонизмнің күшеюіне ықпал етеді. Сондықтан бұл мәселеге қарсы әрекет ету үшін онлайн этикалық және құқықтық сауаттылықты дамыту өте маңызды.

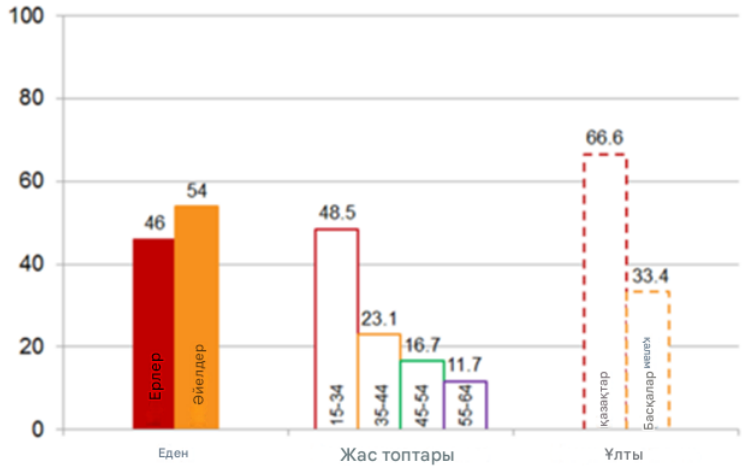
Ғаламторды пайдаланушылар арасында ғадауат сөйлеудің көріністеріне психологиялық проблемалар да, кедейлік, теңсіздік, кемсітушілік, жеке өмірдегі тәртіпсіздік, отбасындағы, балалармен және ата-аналармен қарым-қатынастағы мәселелерге байланысты стресс те әсер етуі мүмкін және қасақана кәсіби түрде жүзеге асырылады [24].

Психологтордың айтуынша жоғарыда айтылғандардың барлығы әлеуметтік желілердегі агрессия арқылы адамның өзін-өзі «қалыпқа келтіру» қажеттілігіне ықпал ете алады, онлайн контентентке жариялаған бейәдеп пікірлер арқылы ол шынайы өміріндегі психикалық стрессті жеңілдете алады. Әлеуметтік желілерде ғадауат тілді сөздерді жариялау ыңғайлы болды. Жабық және лақап есімдерді пайдалану арқылы басқа желі қолданушыларға жағымсыз сөздер арқылы шабуыл жасап, қорлық көрсету қол жетімді және ыңғайлы болып отыр. Психологтардың зерттеуінше, онлайн контенттегі ғадауат пкірлерді жариялаушылар, яғни кибербуллинг көрсетушілер өзгені мазақтап, кемсітіп, намысына тиетін сөздер арқылы қорлау арқылы өзінің эгосын тояттандырып, «психикалық азық» алып отырады [25].

Сонымен қатар, өз ойы жоқ, басқа адам не айтсада істің байыбына бармай жатып соны қолдап кететін адамдар көпшіліктің пікірне ере отырып, қандай да бір адамды даттап, қаралап жатқанын байқамай да қалады. Тіпті, ақсын төлеп боттар жалдап терс пікір жаздыруға бастама беретін ұйымдар немесе жеке адамдар да бар екені анықталған. Шабуыл объектілері әйелдер, ерлер, гейлер, басқа этникалық топ, балалар және жұлдыздар болуы мүмкін. Бұл жағдайда пайдаланушы теріс энергияны шығару үшін кез келген сылтауды пайдалана отырып басқаларды балағаттауға, өшпенділік жасауға тиімді деп санайды [26].

Өшпенділік сөздерін таратудың негізгі мақсаты – белгілі бір топтарға қарсы дұшпандық сезімін ояту және қоғамдық дискурстың ықтимал тақырыптарының ауқымын кеңейту арқылы жалпы халықтың агрессивті бейімділігін қалыптастыру. Бақылаулар көрсеткендей, мұндай ескертулер арандатушылық, кемсітушілік немесе радикалды сипаттағы жеке хабарламалардан бұқаралық коммуникацияның бір бөлігі болып табылатын мінез-құлыққа айналады.

Media Marketing Index статистикасы ұялы телефондарын пайдаланып интернетті пайдаланатын адамдар туралы ақпаратты береді [27]. 2021 жылға қарай жүз мыңнан астам халқы бар қалаларда тұратын он бес жастан асқан қазақстандықтардың шамамен алпыс пайызы интернетке қол жеткізудің ерекше құралы ретінде мобильді құрылғыларды пайдаланады. 2021 жылмен салыстырғанда бұл көрсеткіш 7,3%-ға, ал 2020 жылмен салыстырғанда 30%-дан астамға өсті. Соңғы төрт жылда интернетке тек дербес компьютер арқылы кіретін пайдаланушылардың 2018 жылғы 4,9%-дан 2023 жылы 1,1%-ға дейін төмендеді. Бұл үрдіс мұндай пайдаланушылардың санының өсіунің кепілі болып табылады. Сонымен қатар, 2023 жылы мобильді және жұмыс үстелі құрылғыларын пайдаланып интернетке кіретін адамдардың пайызы 37%-ға жетті (Сурет 1.1) [27].



Сурет 1.1 - Мобильді жұмыс үстелі құрылғыларын пайдаланушылар

Ескерту – [27] әдебиет негізінде құралған

2024 жылдың қорытындысы бойынша, Datareportal [28] сандық технологиялар бойынша жүргізген зерттеулерге сүйенсек, 2024 жылы Қазақстанда интернетті жалпы халықтың 18,9 миллионы пайдаланған, бұл ел халқының 92,3 пайызын құрайды.

«Қолжетімділігі жоғары» критериі бойынша Қазақстан халқының 14 миллион тұрғыны, яғни жалпы халықтың 71,5 пайызы әлеуметтік желіні үздіксіз пайдаланатынын көрсеткен. Тіркелген аккаунт бойынша, түрлі әлеуметтік желідгі желі қолданушылар саны төмендегі 1.2-суретте көрестілген:[29] [30].

Tik - tok (тик ток) -14.1 миллион.

Instagram (инстаграм) -12.1 миллион.

Facebook (фейсбук) -2.6 миллион.

Linkedln (линкедлн) – 1,4 миллион.

Snapchat (снапчат) – 1.07 миллион.

Twitter (қазіргі атауы X) -0,32 миллион.

Сурет 1.2 - Пайдалану жиілігі жоғары Қазақстандық топ 6 желі
Ескерту – [28] әдебиет негізінде құралған

Сондай-ақ, әлеуметтік желілерді қолдану көлемінің өсуіне байланысты онлайн платформаларда қорлау, кемсіту жағдайларының өсуін байқауға болады. Екінші жағынан, 2022 жылдың наурыз айында басталған *BalaQorgau* бастамасы балалар мен олардың ата-аналарынан 600-ден астам өтініш алды [31]. Осы өтініштердің жүзге жуығы оқушылардың әлеуметтік желілерде кибербуллингтің құрбаны болғанын көрсеткен. Бірқатар балалар, мысалы, әлеуметтік желідегі парақшаларына жүктеген бейнелер мен фотосуреттер кемсітетін немесе кемсіту, қорлау ниеті бар пікірлермен бірге келгенін хабарлады. Бұған қоса, басқа қолданбалар балалардың фотосуреттері Photoshop көмегімен теріс өңделіп, кейін оларды келеке ету мақсатында әлеуметтік желілерде бөлісілгенін хабарлады [1].

Кибербулинг жасалу 13-15 жас аралығындағы балалар арасында жиі кездеседі. Яғни, психологтордың айтуынша, 13-15 жас аралығында бала психикасы енді дамып, сырттан келетін психикалық әлімжеттіктерді көтеру деңгейі төмен болғандықтан, мектеп ішінде, ата – аналар тіпті, қоғам тарапынан баланы кибербулингтен өзін – өзі қорғай алуын қамтамасыз ету қажеттілігі туындап отыр дейді [32]. Мысалы, Интернет-тролльдардан қорғануға арналған Ұлыбритания мектептерінде пән бар [33]. Сабақ барысында білім алушыларға әлеуметтік медиа платформаларында орын алатын кемсітушілік пен қудалау жағдайларына қалай әрекет ету керектігі, сондай-ақ осындай жағдайлардан туындауы мүмкін күйзеліске қарсы тұру әдістері туралы нұсқау беріледі. Германияда 2023 жылдан бастап балаға порнографиялық фотосуреттерді таратқан немесе іздеген немесе дұрыс емес сөйлесуге әрекеттенген адамдарға жаза күшейтіледі [34]. Жаңа заңға енгізілген түзетулерге сәйкес, әлеуметтік желіде аккаунт ашқан адамдар баланың атынан орынсыз сөйлескендер де қамауға алынады [35]. Бұл баламен жыныстық қатынасқа түсуге әрекеттенген

тұлғаларды қылмыстық жауапкершілікке тартатын бұрынғы заңнан айырмашылығы ретінде заңның жыл сайын қаталдана түсуінен байқауға болады [36] .

Бірқатар елдер жасөспірімдердің әлеуметтік желілерді пайдалану кезінде олардың қорғалуына кепілдік беру үшін әртүрлі заңнамалар қабылдады. 2000 жылы қабылданған «Балалардың Интернеттегі құпиялылығын қорғау» актісі (COPPA) осы заңдардың бірі болып табылады [37]. Бұл жерде балалардың құдалауына, зорлық-зомбылыққа ұшырауына және жағымсыз ескертулер алуына жол бермеуге арналған нормалар нақты және қысқаша сипатталған.

Қазақстанда, жоғарырда атап өткеніміздей балаларда және олардың ата – аналырында пайда болған проблемаларды шешуге көмек беретін *BalaQorgau* жобасы 2022 жылдың наурыз айында қосылды. Яғни мектептерге QR кодтар ілінді, осы кодты сканерлеу аркылы сайтка кіріп, проблемасын айту арқылы мамандардан көмек ала алады [7]. Ал, 2022 жылы заңға біршама өзгертулер мен толықтырулар енгізілді. «Байланыс туралы» заңның 41-1-бабының 1, 1-1 және 1-2-тармақтарындағы нормаларда көрсетілген [38]. Ал 41-бапта «жеке адамның, қоғамның және мемлекеттің мүдделеріне нұқсан келтіретін қылмыстық мақсаттарда, сондай-ақ кәмелетке толмағандарды сексуалдық қанауды және балалар порнографиясын насихаттайтын ақпаратты таратқан байланыс желілерінің және (немесе) құралдарының жұмысын, байланыс қызметтерінің көрсетілуін уақытша тоқтата тұру бойынша шаралар қабылдау туралы» жазылған [39].

2024 жылдың қыркүйегінен бастап сынып сағаттары аясында мектептерде балалар қауіпсіздігі сабақтарын енгізілді. 2024-2029 мектептерде «КИВА» онлайн контенттен келетін жағымсыз пікірлерге қарсы тұру бағдарламасын жүзеге асыру бойынша Білім министрлігімен бірлесіп жұмысты ұйымдастырылды. 2024 жылдың қыркүйегінде – 110 мектепке енгізілді, кейінгі жылдарына жылына 1,5 мың мектепке енгізу көзделіп отыр [40].

Дегенмен, Ақпарат және қоғамдық даму министрі Асқар Омаров: «Заңдарда тиісті баптар болмағандықтан Meta, Telegram және т.б. цифрлы алпауыттар физикалық тұрғыдан Қазақстан территориясында тіркелмегендіктен біз олардан заңдарымызды сақтауды талап ете алмаймыз. Заңды өкілді тағайындау осы жағдайды түзетуге мүмкіндік береді»-, дейді [41].

Онлайн контенттен келетін жағымсыз пікірлер ретінен тек қана балалар мен жасөспірімдер ғана қорғансыз емес. Сонымен қатар, ересек адамдар да онлайн желі қолданушылар тарапынан кибербуллингке ұшырап отыр. Мысалы, Қазақстандық блогерлермен болған сұхбаттардан әлеуметтік желіге жүктелген видеолар мен фотоларға жазылған ғадауат тілді пікірлер эмоциональдық күйлеріне белгілі бір дәрежеде әсер еткен. Мысалы, өз – өзіне деген сенім төмендеп, өзінің соншалықты дарынсыз, ұсқынсыз екеніне сене бастаған [41]. Дегенмен, психологқа немесе жақындарының қолауымен ол апатиялық күйден шыға алғандарын алға тартуда [42].

Ал, онлайн желі қолданушылардың бейәдеп пікірін көтере алмай, өз- өзіне қол жұмсаған ересек адамдар қатары бар екені белгілі болып отыр. Мысалы, 2021

жылы шілде айында қазақ қызы жазушы Аягул Мантай киббербуллинг құрбаны болды. «Әлеуметтік желіде өзіне үздіксіз бағытталған лас пікірлерге шыдай алмай өз - өзіне қол жұмсады» , деп жазады адвокаты Ляззат Ахатова [43].

Сонымен қатар, Жанар Хамитова атты Қазақстандық әншіні де кибербуллингтің құрбаны болып, өз – өзіне қол жұмсады деп жазды қазақ журналистері. Яғни, оны әлеуметтік желіде «ажырасқан сайқал» деген сөздерді нысанға ала отырып, оны желі қолданушылары қоғамнан алшақтатуға тырысқан [44].

2024 жылы Самбодан спорт шеберіне үміткер Алтай тауының тумасы Ксения Чепонова өзі жаттығып жүрген Новосібірде өз-өзіне қол жұмсады. Ол Telegram желісінде тобындағы қыздар Sibir.Realii сайтна ұлтына қатысты жағымсыз пікірлер қалдырып, онлайн қорлық көрсеткені туралы айтқан [45].

Физикалық тұрғыда жасалған буллингтің әсері бірден байқалады, яғни қандай да бір дене жарақаты арықлы анықтауға болады. Ал, кибербуллинг кезінде сырт көздің байқау қиындығы туындайды.

Балаларды қорғау, балалардың өмірін құтқару мандатына ие БҰҰ Балалар қоры ЮНИСЕФ 10 түрлі интернеттегі қорлық көрсету мәселесін талқылайды [33]:

1. Интернет желідегі қалжың мен қорлауды қалай ажыратуға болады?

Барлық достар бір-бірін мазақ етеді, бірақ кейде әзіл немесе ренжіту әрекеті арасындағы шекараны анықтау қиын, әсіресе онлайн контент желісінде. Бұл кейде «біз әзілдеп жатырмыз» немесе «тым шынайы қабылдама, бұл жай ғана әзіл» деген мағынада айтылуы мүмкін.

Бірақ кемсіту мен ренжітуді сезінсе немесе басқалар әзілдеп қана қоймай, мазақ етіп жатыр деген ой келетін болса, онда бұл жағдайда әзіл шектен шығып кетті деп түсіну керек. Егер ондай «әзілді» тоқтатуды сұрағаннан кейін де жағдай жалғаса берсе және сол «әзілге» деген реніш сезімі басылмаса, оны қорлау деп санауға болады.

Егер адам өзін нашар сезінсе және ол тоқтамаса, онда көмекке жүгіну керек. Кибербуллингті тоқтату тек бұзақыларды ауыздықтау ғана емес, бұл әрбір адамның желіде де, шынайы өмірде де құрметке лайық екенін мойындау.

2. Кибербуллингтің салдары қандай болмақ?

Интернетте қорқыту орын алғанда, адам өзін барлық жерде, тіпті үйде де шабуылдап жатқандай сезінуі мүмкін. Одан құтылу мүмкін емес сияқты көрінуі мүмкін. Салдары ұзақ мерзімді болуы мүмкін және адамға көптеген жолдармен әсер етеді:

• Психологиялық – адам ренжиді, өзін ақымақ, ұялшақ кейде тіпті ашулы сезінуі мүмкін

• Эмоционалды түрде – ұялу сезімі ұялайды немесе өзіне ұнайтын нәрселерге деген қызығушылықты жоғалту

• Физикалық – шаршау сезімі (ұйқының жоғалуы) немесе тіпті асқазанның ауыруы және бас ауруы сияқты белгілер

Адамды келекелеп немесе қорлап жатқандай сезіну бұл туралы ашық айтуды немесе мәселені шешуге тырысуды қиындатады. Ең төтенше жағдайларда кибербуллинг адамның өз өмірін қиюына әкелуі мүмкін.

Кибербуллинг адамдарға әртүрлі жолдармен әсер етеді. Бірақ оны жеңуге болады және адам өз сенімін де, денсаулығыңызды да қалпына келтіре алады [46].

3. *Интернетте басқа желі қолданушылар қорлап жатса, кіммен сөйлесу керек? Неліктен бұл туралы хабарлау маңызды?*

Егер адамды әлеуметтік желіде қорлап жатса, агрессорды блоктауға тырысу және оның мінез-құлқы туралы ресми түрде онлайн платформада тікелей хабарлау маңызды. Әлеуметтік медиа платформалары өз пайдаланушыларының қауіпсіздігін қамтамасыз етуге міндетті [36].

Мысал ретінде - жақында ғана қайта іске қосылған Facebook-тің интернет желісінде қорқытудың, кемсітудің алдын алу орталығы пайда болған [47]. Желі қолданушылары осы ақпараттық ресурсты пайдалана отырып, өздерін және жақын адамдарын кибербуллинг құрбаны болудан сақтай алады. Сондай-ақ қол жетімді басқа да қызметтер бар, мысалы, пікірлерді модерациялау мүмкіндігі, сізге бейтаныс немесе сізді қорлау, кемсіту пиғылы бар пікірлерді жазған адамдарға тыйым салып бұғаттау сияқты қосымша функцияларды атауға болады. Жыл сайын осы әлеуметтік желіні жақсарту мақсатында жұмыс істейтін ізденушілер кемсіту, қорлау әрекеттерін болдырмауға арналған мүмкіндіктерді жақсартуға күш салады.

Оның үстіне, Instagram жақында шектеулер деп аталатын жаңа мүмкіндікті іске қосты. Бұл функцияны енгізу арқылы біз жіберілетін жағымсыз хабарлар мен түсініктемелердің санын азайтқымыз келеді. Белгілі бір адам Instagram-да киберқорқытуға ұшырап, кемсітетін немесе қорлайтын сипаттағы ескертулер таратылған жағдайда, әлеуметтік медиа платформасында қысқа мерзімге шектеу функциясын бірден тоқтату мүмкіндігі бар. Сонымен қатар, Instagram-да пікір жіберу мүмкіндігі платформада тіркелмеген пайдаланушылар үшін шектелген. Тамыз айының соңында тағы бір жаңа функция пайдаланушыларға қолжетімді болды. Пайдаланушылар қауіп-қатер немесе бопсалауды қамтитын жеке хабарламаларды жіберген жағдайда, пікірлер бірден көрінбейтін қалтаға тасымалданады [7].

Қорқытуды тоқтату үшін жасалатын ең маңызды нәрсе - оны анықтау және жағдайды хабарлау.

Facebook желісіндегі кемсіту кезінде қалай әрекет ету керектігін немесе басқа біреудің қорқытып жатқанын көрсеңіз не істеу керектігін айтатын нұсқаулық бар [8]. Сондай-ақ Instagram-да ата-аналарға, қамқоршыларға және сенімді ересектерге киберқорқытумен күресу бойынша нұсқаулар беретін ата-аналарға арналған нұсқаулық және қауіпсіздік құралдары туралы ақпарат алатын орталық платформа бар [48].

4. *Жақын адамдардың кибербуллингке ұшырағаны туралы хабарлауға қалай көмектесуге болады, егер ол оны қаламаса?*

Кез келген адам кибербуллингтің құрбаны болуы мүмкін. Егер көмектесу мақсатында ештеңе жасалмаса, онда ол адамда бәрі оған қарсы, немесе өзінің ешкімге қажет еместігі туралы сезім болуы мүмкін.

Киберқорқыту оқиғасы туралы хабарлау Instagram және Facebook-те әрқашан анонимді болып табылады және бұл туралы бізге хабарлағаныңызды ешкім ешқашан білмейді деп көрсетілген [47,48].

Желі қолданушы өзімен не болғанын хабарлай алады, бірақ тікелей қолданбада қол жетімді опцияларды пайдаланып басқа адам тарапына келген кибербуллинг туралы не болғанын хабарлауға болады. Бірдеңені хабарлау туралы қосымша ақпаратты Instagram анықтамалық орталығынан және Facebook анықтамалық орталығынан табуға болады [10-11].

5. Интернетті қолданудан бас тартпай кибербуллингті қалай тоқтатуға болады?

Кибербуллингті тоқтауға күш салынып жатыр және бұл кибербуллинг туралы есеп берудің маңызды себептерінің бірі болып табылады. Онлайн контентте бөлісетін немесе пікір алмасатын хабарламалар, суреттер, видеолар басқаларды ренжітуі мүмкін екенін әрқашан есте ұстаған жөн. Интернетте де, өмірде де бір-бірімізге мейірімді болу керектігін үнемі есте ұстаған жөн.

Интернет қолданушылар үшін Instagram мен Facebook қауіпсіз және позитивті платформа болып қалуы маңызды – адамдар өздерін қауіпсіз сезінген жағдайда ғана онлайн қарым-қатынас жасауда өздерін жайлы сезінеді. Кибербуллинг бұған кедергі келтіріп, теріс түсініктер тудыруы мүмкін екені анық. Сондықтан Instagram мен Facebook кибербуллингпен күресте көш бастауға дайын [47,48].

Ғадауат пікірлерден туатын интернет буллингтен қорғану үшін екі тиімді жолды қарастыруға болады. Біріншіден, пайдаланушылардың өздерін қорлауға немесе олардың көріністерін көруге жол бермейтін технологияларды қолдану. Мысалы, қорлау немесе ренжітетін пікірлерді автоматты түрде сүзу және жасыру үшін жасанды интеллект технологиясын пайдаланатын параметрді пайдалануға болады.

Екіншіден, Facebook және Instagram пайдаланушыларының қажеттіліктеріне бейімделген опцияларды ұсыну,шектеу опциясы аккаунтты жасырын түрде қорғауға және сонымен бірге агрессордың әрекетін бақылауға мүмкіндік беретін құралдардың бірі ретінде қолданысқа енгізу.

6. Жеке бастық ақпаратпен әлеуметтік желіде адамды манипуляциялау немесе қорлау үшін пайдаланылуын қалай болдырмауға болады? – деген сұраққа ЮНИСЕФ мамандарының пікірі:

Жеке бастың ақпараттарымен желіде бөліспеуге кеңес береді. Мысалы, телефон нөмер, мекен – жай, қандай да бір жеке ақпараттармен, құжаттармен.

Қолданатын әлеуметтік медиа қолданбаларыңызда құпиялылық параметрлерін пайдалану жолын үйрену қажеттілігі туындайды. Сонымен қатар, төмендегі әрекеттерді орындауға кеңес береді [33]:

Аккаунттың құпиялылық параметрлерін өзгерту арқылы профилді кім көре алатынын, тікелей хабарлар жіберетінін немесе жазбаларға пікір білдіретінін шешуге болады.

Сізге қорлау, кемсіту пиғылы бар пікірлер, хабарламалар мен фотосуреттер туралы хабарлап, оларды жоюды сұрауға болатындығын ескертеді.

«Достықтан шығару тізімі» опциясына қоса, басқа пайдаланушылардың профилді көруіне немесе қандай да бір байланысуына жол бермеу үшін оларды толығымен блока қоюға болады.

Сондай-ақ белгілі бір адамдардың пікірлерін толығымен блоктамастан тек олар үшін көрсетуді таңдауға болады.

Профильдегі жазбаларды жоюға немесе белгілі бір адамдардан жасыруға болады.

7. Киберқорқыту үшін қандай жаза түрі қолданылады?

Қорқытып, үркітуге қарсы заңдар, әсіресе кибербуллинг, басқа заңдармен салыстырмалы түрде жаңа және әлі тоылқ түрде қабылданған жоқ. Сондықтан да көптеген елдер кибербуллерлерге жазаны белгілеу кезінде осы саладағы басқа заңдардың ережелерін, мысалы, қудалауға қарсы заңдарды қолданады.

Кибербуллингке қарсы арнайы заңдары бар елдерде ауыр эмоционалды күйзелісті тудыратын қасақана онлайн әрекеттер заңсыз болып саналады. Осы елдердің кейбірінде киберқорқытудың құрбандары белгілі бір адамның киберқорқыту мақсатында белгілі бір кезеңге немесе өмір бойы негізде пайдаланатын электрондық құрылғыларды пайдалануын шектеуге, белгілі бір тұлғаның байланысына тыйым салуға және қорғауға өтініш бере алады.

Facebook-те қауымдастық стандарттарының тізімі бар [9], ал Instagram-да қауымдастық нұсқаулары бар [49], қорқыту немесе қудалау сияқты осы ережелерді бұзатын мазмұнды тапса, оны жоятындықтарын айтып өткен.

Twitter барлық пайдаланушылардың ашық әңгімеге еркін және қауіпсіз қатыса алуын қамтамасыз ету үшін саясаттарsy қатаң түрде орындаймыз деп сендіреді. Бұл ережелер бірқатар салаларды, соның ішінде тақырыптарды қамтиды. Яғни:

- Зорлық-зомбылық
- Балаларды зорлық – зомбылыққа пайдалану
- Қиянат/қорлау
- Жек көру насихаты
- Өзін-өзі өлтіру немесе өзіне зиян келтіру
- Заңсыз мазмұн, соның ішінде зорлық-зомбылық бейнелері

Осы ережелердің бір бөлігі бойынша ережелерді бұзатын жағдайларда шара қолдануға мәжбүр екенін алға тартуда [50].

8. Киберқорқытуға қарсы тұратын онлайн құрылғылар:

Facebook пен Instagram кибербуллингтен қорғау үшін құрылған онлайн құрылғылар:

Агрессордан келген барлық хабарларды елемеу опциясын немесе ол адамға ескертпестен аккаунтыңызды жасырын қорғау үшін шектеу қоя опциясын пайдалануға болатындығын айтып отыр [50,51].

Жазбаларыңыздағы пікірлерді модерациялай яғни, пікірлерге қандай да бір талаптар қою құқығын пайдалану мүмкіндігі .

Параметрлерді тек жазылушылар тікелей хабарлар жібере алатындай етіп өзгертуге болады.

Instagram-да ережелерді бұзуы мүмкін пост жариялағалы жатқаны туралы хабарландыру жіберетіндіктерін және оны қайта қарауды желі қолданушылардан сұрайтындығын алға тартуда [49].

Бұл тарауда «ғадауат» сөзінің кездесу аймақтары, мағынасы түрлі сөздіктердегі мағыналарын ашып көрсеткен. Сонымен, қатар ғадауат тілді сөздердің түрлері, психологиялық әсері және сондай пікірлермен бетпе – бет келгенде қорғану шаралары туралы айтылған. Қазақстандағы әлеуметтік желіні, онлайн контентті қолдану статистикасы мен әлеуметтік желілердің ғадауат тілді сөздерден тұратын пікірлерден қорғану функциялары туралы мәлімет берілген.

2 ҒАДАУАТ СӨЗДЕРДІ АНЫҚТАУ АЛГОРИТМДЕРІНДЕГІ ҒЫЛЫМИ ЗЕРТТЕУ ЖҰМЫСТАРЫНА ШОЛУ

Соңғы жылдары NLP- ні және машиналық және терең оқыту алгоритмдерін қолданып әртүрлі тілдер үшін әлеуметтік желілерде ,адауат тілді сөздерді анықтауда зерттеушілер қатары көбеюде. Онлайн контенттегі ғадауат тілді анықтау үшін сапалы және қолжетімді деректер қорынан бөлек, оларды нақты анықтай алатын өлшемділік көрсеткіші бойынша жақсы нәтижеге жеткізетін оқыту алгоритмдерін де дұрыс таңдау маңызды. Бұл бөлімде зерттеу жүргізу үшін деректер қорын қалыптастыруға, аннотаторлық нұсқауларға, үлестерге және шектеулерге назар аудара отырып, Twitter және Facebook желілерінен алынған деректер қорына және ресурсы аз тілді тілдерге арналған онлайн контенттегі ғадауат сөздерді анықтауда машиналық және терең оқыту алгоритмдерінің тиімділігін салыстырады.

Деректер жинағының ауқымы, тілдің әртүрлілігі және қолданылған машиналық және терең оқыту алгоритмдерінің өнімділік сапасы зерттелген ғылыми жұмыстар арасында арасындағы ерекшеліктерді көруге болады. Кейбір зерттеу жұмыстары исламофобия немесе кибербуллинг үшін жаңа деректер жиынтығын ұсынады, ал келесі ғылыми зерттеу жұмыстары қол жетімді деректер корпусын пайдаланған.

Бұл тарауда ғадауат тілді сөздерді зерттеуде деректер жинағын құру анағұрлым маңызды, және қажетсіз ақпараттардан таза болуы керек екенін көрсетеді. Ғадауат тілді сөздер әр елдің пайдалану жиілігі жоғары онлайн платформалары мен географиялық орналасуы бойынша өзеріп отыратындығы белгілі болған, сондықтан оны жинақтау және класстарға бөлу әдістері техникалық жағынан жақсы жетілген және үлкен жауапкершілікті талап ететіндігі туралы баяндалған. Онлайн контенттегі ғадауат тілді анықтауда аз ресурсты тілді деректер қорын пайдала отырып, машиналық және терең оқыту алгоритмдерінің өнімділік көрсеткіштері салыстырыла отырып, тиімді әдістерге зерттеу жүргізген.

2.1 Онлайн мазмұндағы ғадауат сөздерді анықтау үшін қолданылатын машиналық және терең оқыту әдістері

Ғылыми зерттеу жұмысы бойынша ғылыми әдебиеттерге шолу бойынша төмендегі міндеттер қатары қарастырылды:

1. Машиналық оқытудың ең тиімді алгоритмдерін ғадауат тілді сөздерді анықтау үшін шолу жасау;
2. Ғылыми зерттеулерде ғадауат тілді сөздерді анықтауда қандай деректер қорының түрлері пайдаланылды.

Ғадауат тілді сөздерді анықтауда машиналық оқытудың қай әдісі тиімдірек екенін білу мақсатында осы зерттеу сұрақтары таңдалды. Осылайша ең жақсы нәтиже алу үшін ғылыми зерттеу жұмыстарында тиімді деп табылған әдісті зерттеп, мәселені шешудегі ықтималдығының қаншалықты деңгейде екендігін анықтауға болады [52]. Сондай-ақ, екінші зерттеу міндеті: болашақтағы ғылыми

зерттеулер үшін деректер базасының деңгейін анықтап, қандай зерттеулер қажет ететіндігін көруге болады.

Ғылыми зерттеу жұмыстары мен ғылыми мақалаларды IEEE Xplore [53], Springer Link [54], Science Direct [55]және ACM Digital Library [56] сияқты онлайн дерекқор кітапханаларынан алынды. Барлық дерлік ғылыми зерттеу жұмыстары 2015 жылдан бастап қазіргі кезеңге дейінгі уақыт аралықтарын қамти отыра, қазақ, ағылшын, орыс тілінде жазылған және осы зерттеу жұмысының мақсатына сәйкес келетін біршама критерийлерді қамтыды. Ғылыми зерттеу жұмыстарын іздеу кезінде пайдаланылған кілттік сөздер немесе сөз тіркестері келесідей етіп таңдалып алынды:

• ғадауат тіл, адамның жеке басына тіл тигізу арқылы қорлау тілі немесе кемсіту тілі, кибербуллинг, хейтспич;

• нейрондық желі, машиналық оқыту алгоритмдері және терең оқыту алгоритмдері.

Жоғарыда ұсынылған алғашқы іздеу жолын қарастырғанда онлайн желі кітапханаларының кеңейтілген іздеу мүмкіндігіндегі «ұсынылған мәтіннен немесе кілттік сөз бойынша мәліметтер қорына қол жеткізу» болып табылды. Ол ғадауат тілді сөздерді анықтауға күш салған барлық ғылыми зерттеу жұмыстарын іздеп, сүзгіден өткізуге арналған. Екінші іздеу жолы онлайн іздеу мүмкіндігінен *«дәл келетін ақпаратты табу»* үшін пайдаланылады. Бұл іздеу жолын нейрондық желілерге, машиналық оқыту алогритмдеріне және терең оқыту алгоритмдерін пайдалана отырып онлайн контенттегі барлық ғадауат тілді пікірлерді анықтауға қатысы бар барлық ғылыми мақалалар, еңбектер мен интернет ақпараттарын сүзу үшін пайдаланылады.

Зерттеу жұмысында [57] сөздер ұқсастығын пайдалану концепциясындағы нейрондық тіл үлгісі сипатталған. Сөйлемде қатар тұрған сөздер, яғни сөздер тіркестері бір – бірімен байланыста бола алатындығын ескерген. Үздіксіз сөздер қапшығы (Continuous Bag Of Words (CBOW)) үлгісін пайдаланған, яғни кілттік сөзді пайдалану арқылы сөйлем құрастырлған сөзді нақты дәлдікпен табуға тырысады. Бұл модель осы типтегі сөздердің жанында орналасқан сөздерді табады және оның ғадауат тілді сөздер қатарына жата ма, жоқ па деген болжам жасау үшін кілттік сөзге сүйенеді. Бұл әдіс өте тиімді екенін байқауға болады, өйткені ол аз дайындықты қажет етеді және жоғары перспективалы нәтиже береді.

Келесі ғылыми зерттеу жұмысының мақсаты [52] Twitter желісіндегі ғадауат тілді сөздерінің болуын анықтау әдісі ретінде терең оқытуды пайдалану мүмкіндігін зерттеу болды. Қазіргі таңда әлеуметтік желілердегі белсенділіктің артуы нәтижесінде интернет қолданушылары белсенділіктің артуына байланысты теріс пиғылдағы және бейәдеп тілдегі сөз тіркестері жиі қолданыста. Зерттеулер нәтижесіне сүйенер болсақ, кездейсоқ орман (RF), логистикалық регрессия (LR), градиентті күшейтілген шешім ағаштары (GBDT), қолдау векторлық машиналары (SVM) және терең нейрондық желілер (DNN) осы зерттеу жұмыстарында талқыланып және бағалаушы метрикаларының біршамасы жақсы нәтижелік көрсеткішке жеткен.

Келесі ұсынылатын [58] зерттеу аясында терең оқытудың үш түрлі алгоритмдері қолданылған: конволюционды нейрондық желілер (CNN), ұзақ мерзімді жады (LSTM) және FastText. Зерттеу жұмысы авторлары бұл әдісті char n-gram, Word Term Frequency-Inverse Document Frequency (TF-IDF) және Сөздер қапшығы (BOW) сияқты басқа дәстүрлі тәсілдермен салыстырғанда әлдеқайда жоғары тиімділік деңгейін көрсеткенін байқаған.

Вектор (BoWV) келесі ғылыми мақалада [59] әлеуметтік желілерде жеті модельді қолдану арқылы ғадауат тілді сөздерді анықтау талқыланды. Әлеуметтік медиа адамдардың желідегі өзара әрекеттесу тәсілін өзгертті, сондықтан олар шабуылдаушыларға ғадауат тілді мазмұнды жариялау арқылы пікірлерге әсер етуге мүмкіндік береді. Деректер жинағында Twitter және Wikipedia-дағы әртүрлі көздерден жиналған шамамен 300 000 жазбадан тұратын мәліметтер қоры бар. Нәтижелерге сәйкес, логистикалық регрессияны қолданатын және лингвистикалық n-граммдарға негізделген қарапайым әдіс DNN, GRU немесе CNN сияқты күрделі үлгілерге қарағанда табыстырақ болып келеді деп көрсетеді.

Зерттеу нәтижелеріне сүйене отырып, болашақта модельдерді нақтылауға назар аудармай, деректер жиынтығына көбірек көңіл бөліп, зерттеу жүргізу ұсынылады [52].

Зерттеу жұмысында [60] ұзақ мәтіндік құжаттарда машиналық оқыту әдісін қолдану арқылы ғадауат сөздерді анықтау талқыланды. Бұл зерттеу жұмысының айырмашылығы - индонезия тіліндегі мәліметтер қоры жинағы болып табылады. Бұл сондай-ақ 2017 жылдан бастап, дүние жүзінде ғадауат сөздерді қолданатын саяси тәжірибелер санының артып келе жатқанын көрсетеді. Деректер жинағы саясат туралы жарияланатын Facebook жазбаларынан жиналды. Деректерді 20 мен 26 жас аралығындағы тұрғылықты мекен – жайлары жан- жақта орналасқан әйелдер мен ер адамдардан тұратын 3 аннонаторлар жинақтады. Нәтиже көрсеткендей, SVM-ді TF-IDF, ғадауат сөздер, позитивті сөздер, char quad-gram және word unigram-мен біріктіру 85% F1 ұпайымен жақсы нәтиже берген.

Келесі ғылыми зерттеу жұмысы [61] HateDefender деп аталатын жаңа жүйеге арналған. Сонымен қатар, LSTM терең нейрондық желіге негізделген. Ғадауат сөздерді анықтаудың орташа дәлдігі 90,82%, ал кемсіту тілін анықтау 89,10% құрайды. Бұл жұмыста [57] Ченг ғылымии зерттеу жұмысының авторы болып табылатын басқа зерттеудің деректер жинағын пайдаланады, олар ғадауат, қорлаушы сөздер немесе ондай мазмұндас сөздер жоқ деп тегтелген твиттерден тұрады.

Ұсынылып отырған ғылыми зерттеу жұмысы [62] әлеуметтік желілердегі ғадауат тілді сөздерді анықтауға арналған. Дініне, этникалық тегіне немесе әлеуметтік мәртебесіне негізделген белгілі бір пайдаланушы топтарына бағытталған шабуылдарға байланысты өшпенділік сөздердің өскендігін алға тартқан. Зерттеу жұмысы 3 түрлі тіл деректер жиынтығын пайдаланады; 32% өшпенділік ниетті білдіретін 16 000 ағылшын твиттері, 32% ғадауат тілді 4 000 итальяндық твиттер және 34% бейәдеп сөздер тіркесі кездесетін 5 000 неміс

твиттері. Бұл зерттеуде LSTM, GRU және екі бағытты LSTM (BiLSTM) FastText әдістері қолданылды. Нәтиже көрсеткендей, итальяндық деректер жинағы үшін ең жоғары балл орташа F1 ұпайы 0,801 болатын униграмманы қолдану арқылы алынған. Дегенмен, неміс және ағылшын тілдеріне арналған деректер жинағы ағылшын тілі үшін 0,785 және неміс тілі үшін 0,718 орташа F1 ұпайы бар униграмманы қажет етпей - ақ жоғары ұпайға жетеді деп келтірген.

Ғылыми зерттеу жұмысы [63] ғадауат тлді сөздерді анықтау мақсатында пікірлердің арасында кемсіту мен қорлау ниеті кездесетін, тілге тән таңбалар жүйелерін зерттеуді басшылыққа алған. Жинақталған деректер базасы кроудфлоуер – дегі деректер базасымен сонымен қатар, басқа ғылыми зерттеу мақалаларында қолмен сыныпқа бөлініп жинақталған деректер жиынымен бірге қолданған. Деректер базасын оқу жаттығуы негізінде 21 мың пікірлерді, ал, тестілеу үшін және тексеру жаттығу алгоритмдері үшін 2 010 пікірлер кездеседі деп жазған, әр сынып 670 пікірлерден тұрады деп қарастырған.

Зерттеу нәтижесі бойынша, бұл зерттеудің әдісі пікірлердің ғадауат тілді сөздерден тұратынын немесе ғадауат емес, бейтарап пікірлер қатарынан екендігін екілік классификацияға бөлу кезінде 86,9 пайыздық дәлдікке жеткен және үштік классификациялау кезінде пікірлерді ғадауат тілді сөздер кездеспейтін, бейтарап пікірлер немесе кемсіту, қорлау пиғылды сөздер кездеседі деп көрсеткен кезде 79,1 пайыздық дәлдікпен анықталған [63].

Келесі ұсынылып отырған ғылыми мақалада [64] Оңтүстік Африка тұрғындарының ішінде Twitter әлеуметтік желісін пайдаланушыларынан келген твиттердегі ғадауат тілді сөздерді машиналық оқыту әдісімен анықтау туралы баяндаған. Twitter әлеуметтік желісі - Оңтүстік Африкадағы ең танымал, пайдаланушысы көп онлайн контенттегі желі. Бұл жұмыста көрсетілгендей, 2019 жылдың 5 мамырында басталып, 13 маусымында аяқталатын кезеңде Twitter Achiver құралы жалпы саны 21 350 твиттерді жинау үшін пайдаланылды. Сесото, isiZulu және Африкаан тілдерінен алынған терминдерді қамтитын ағылшын тілінде жазылған аралас кодты твиттерді қоспағанда, ағылшын тілінде жазылмаған твиттер ескерілмеді.

Келесі жұмыста [65] Support Vector Machines (SVM), Random Forest, Gradient Boosting және Logistic Regression сияқты машиналық оқыту әдістері қолданылды. Өшпенділік сөздері үшін максималды шынайы оң көрсеткіш 0,894 болғанда, n-gram тәсілін пайдаланған оңтайландырылған SVM үлгісі 0,646 жалпы дәлдікке қол жеткізді. Екінші жағынан, ғадауат тілді сөздердің шынайы оң көрсеткіші өте төмен болды, ол 0,069 деңгейінде болды.

Зерттеу жұмысы [66] урду тілінің сөздік қорындағы ғадауат тілді сөздерді және урду тілдік қорының роман тілінде кездесетін сөздерді анықтауға бағытталған. Зерттеу жұмысында пайдаланылған деректер қоры римдік урду тілінің деректер қоры Github [67] сайтында 147 000 желі пайдаланушысының аккаунты мен пайдаланушы пікірлерімен кез – келген адам өз зерттеу жұмысына деректер базасы ретінде қолдана алады.

Дегенмен, ғадауат тілді сөздер тізбегін анықтау мақсатында қолданысқа ие стандартқа сай құрастырылған урду тілі бойынша деректер базасы

кездеспейді, оның орнына ғылыми зерттеу жұмысының авторлары Үнді елі мен Пәкістанда жасақталған саяси, ойын-сауықтық, діндік және спорттық жанрдағы ютуб желісіндегі желі пайдаланушылардың пікірлерін қолмен жасақтап, деректер базасы ретінде қалыптастырды [68].

Деректер қорын жасақтау үшін 3 магистрант пен жергілікті ерікті тұрғындар тағайындалған. Зерттеу жұмысында келесі алгоритмдер қолданылады: LogitBoost - бұрыннан үйренген деректерді пайдалана отырып, модельді жаттықтыратын әдіс. Ол AdaBoost тәсіліне негізделген, сонымен қатар, ғылыми зерттеуде SimpleLogistic деп аталатын қарапайым регрессия функциясы қолданылды.

Яғни, жоғарыдағы зерттеу жұмысында тірек векторлық машиналары (SVM) алгоритмі, аңғал Байес (NB), хедффинг ағаштары және K-ең жақын көршілері (KNN) оқыту алгоритмдері пайдаланылды. Символдық триграммалармен жұмыс жүргізу арқылы SimpleLogistic урду тілі бойынша деректер қорында 96,1 пайыздық және римдік урду деректер қорында 97,9 пайыздан тұратын F1 ұпайына қол жеткізді.

Ғылыми зерттеу жұмысында [61] Индонезия халқының Twitter әлеуметтік желісін қолданушыларының твиттері үшін ғадауат тілді сөздерді табуға негізделген. Индонезия әлеуметтік желілерді көптеген қосымша мақсаттарда пайдаланушы елдердің қатарына жатады. Ғылыми зерттеу жұмысындағы деректер қоры Twitter API негізінде қол арқылы жасақталып, сонымен қатар, оны қолмен сүзгіге салып, әр сыныпқа біөліп қарастырды және 20 адам ерікті түрде деректер базасын құруға келісім беріп, деректер қорын түзіп шыққан. Құрастырылған деректер базасы 100 пайыз аннотаторлар көмегімен, келісілген түрде жасақталып отырды, сол себепті өзгеше белгілі твиттер жинағы деректер базасынан алынып тасталынды. Жалпы көлемі 2017 твиттер жинақталды. Naive Bayes, Support Vector Machines және Decision Data Trees (RFDT) жоғарыдағы деректер базасын анықтау мақсатындағы алгоритмдер қатары түзілді. Сөздердің униграммалық ықтималдығына назар аударып, есеп жүргізу арқылы жеткен нәтиже 87,43% F1 ұпай берсе, Naive Bayes алгоритмінің нәтижесі SVM және RFDT-ге қарағанда әлдеқайда артықшылығы бар екендігі байқалады.

Келесі ұсынылып отырған ғылыми зерттеу жұмысы [69] хинди-ағылшын аралас код деректерінде ғадауат тілді сөздерді анықтауға арналған. Үндістан халқының шамамен 44%-ы хинди тілінде сөйлейді, сондықтан Twitter және Facebook-те хинди-ағылшын тілін қолдану өте жоғары. Қолданылған деректер жинағы 3 зерттеу жұмысының комбинациясы болып табылады, сондықтан барлығы 10 000 мәтін болды және олар ғадауат тілді сөздерден тұратын пікірлер және бейтарап деп 2 сыныпқа тең бөлінді. Келесі әдістер қолданылды: SVM, SVM-Radial Basis Function (SVM-RBF) және Random Forest. Нәтиже FastText-пен біріктірілген SVM-RBF 85,81% F1-балын беретінін көрсетеді, бұл 75,11% F1-балын беретін word2vec-пен біріктірілген SVM-RBF-ден жоғары.

Кезекті ғылыми зерттеу мақаласында [70] Twitter-де кибержеккөрініш сөздерін анықтау талқыланды. Ол мүгедектік, нәсіл, жыныстық бағдар немесе жалпылама ғадауат сөздері бар твиттерге түсініктеме беру үшін CrowdFlower

пайдаланылған. Жіктеу BoWV-мен бірге SVM және RFDT көмегімен жасалды. Нәтиже жалпы F1-балы 0,96 екенін көрсетті.

Ғылыми жұмыста [71] онлайн контентте кездесетін ғадауат тілді сөздерді анықтаудың әдістері ұсынылған. Желі пайдаланушылардың (яғни аккаунттардың) өзге әлеуметтік желі пайдаланушыларына (яғни, өзге аккаунттың қолданушыларына) жазатын, айтатын пікірлері (мысалы, твиттер) аккаунттары киберәлемнің негізігі бөлігі ретінде көрсеткен.

Деректер қоры базасы 9 588 аннотациялаудан өткен твиттер мен пікірлерден құралған. Анықтау әдісі ретінде RFDT пайдаланылған [71]. Бұл зерттеу жұмысында мәтіндік айнымалыларға қарағанда метадеректерге басымдық беріп отырған.

Ұсынылып отырған ғылми жұмыста [72] көптеген әлеуметтік медиа платформалары үшін ғадауат сөздердің онлайн классификаторын жасау туралы қарастырған. Ересектердің шамамен 22% -ы онлайн платформаларда кемсітуді бастан кешірді. Бұл зерттеуде пайдаланылған деректер жинағы басқа ғылыми зерттеу жұмыстарынан алынды. Youtube, Wikipedia, Twitter және Reddit сияқты 4 әлеуметтік медиа деректер жинағы қолданылған. Барлығы 197 566 пікір жиналды және пікірлердің 80% ғадауат тілді емес деп белгіленді, ал қалған 20% -да бейәдеп сөздер жиыны кездесетіні туралы айтылған.

Әлеуметтік желідегі ғадауат тілді хабарламалар бір-біріне сәйкес белгілі мақсаттар мен іс –әрекеттердің кең ауқымды күрделі құбылыс ретінде қарастыруға болады [73]. Кибербуллинг және ғадауат тілді сөздер біздің қоғамға кері әсерін тигізуіне байланысты соңғы бірнеше онжылдықта зерттеушілердің қызығушылығын арттырған кемсіту тілінің типтік мысалдары болып табылады. Бұл спам-хабарламаларды басқа әлеуметтік медиа хабарламаларының арасында автоматты түрде анықтау үшін зерттеулер жүргізіліп жатқанын алға тартуда [73].

Келесі зерттеу жұмысында [74] Машиналық оқыту алгоритмдері ғадауат сөздерді анықтауға және жалпы әлеуметтік желілер мазмұнын талдауға үлкен үлес қосқанын атап өткен. Ғадауат тілді сөздер және кибербуллинг сияқты қорлайтын пікірлер соңғы бірнеше онжылдықта NLP-тің ең көп зерттелген бағыттары болып табылады [75]. Машиналық оқыту алгоритмдері бейәдеп тілді пікірлерді анықтау және жіктеу үшін әлеуметтік желілер деректерін талдау мақсатында осы бағытта өте пайдалы болғандығы туралы баяндалған. ML алгоритмдері бойынша зерттеулердің жетістіктері қызметтің көптеген салаларына айтарлықтай әсер етті, нақты әлемдегі мәселелерде үлкен көлемдегі деректерді талдау үшін жұмыс өнімділігі құралдар мен модельдердің пайда болуына әкелгендігі туралы айтылады [8].

Бұл ғылыми зерттеуде [60] авторлар ғадауат тілді сөз тіркестерін анықтау үшін қолданылуы мүмкін сегіз түрлі стратегиялар мен тәсілдерге қысқаша шолу жасаған. TF-IDF, лексикондар, N-граммалар, сезімдік талдаулар, үлгіге негізделген әдістер, сөз құрылымдары, сөздер пакеті және ережеге негізделген тәсілдер осы тізімге енгізілген сегіз әдіс болып табылады. Олардың ғылыми зерттеуінде терең оқыту да, ансамбльдік әдістер назарға алынбаған.

Ансамбльдік тәсілдер – белгілі бір нәтижеге жету үшін көптеген модельдерді бағалайтын әдістер [76]. Келесі ұсынылып отырған ғылыми зерттеу жұмысы авторлары табиғи тілді өңдеуде ғадауат сөздерді автоматты түрде анықтау салаларына әдеби тілдегі өшпенділік сөздерін анықтау ерекшеліктерін талдады, олар: жәй сырт белгілері (simple surface features), сөзді жалпылау (word generalization), көңіл-күй бойынша талдаулар (sentiment analysis), лексика ресурсы (lexical resources), лингвистика ерекшелігі (linguistic features), білім жолындағы ерекшелік (knowledge-based features), мета-мәлімет (meta-information) және мультимодаль мәлімет (multimodal information) [77].

Кезекті ғылыми зерттеу жұмысы [78] авторлары Twitter платформасында жариялаған твиттер негізінде терроризмді қолдайтын желі қолданушының мінез-құлқын талдау тәсілін ұсынады. Зерттеу барысында деректер қоры Twitter API арқылы жиналғандығы анықталды. Твит бағасы sentiwordnet сөздігіне сілтеме жасау арқылы есептеледі, содан кейін твит аңғал Байес классификаторы арқылы жіктелетіндігін байқауға болады. Пайдаланушының мінез-құлқын талдау құралы белгілі бір пайдаланушының твиттер тарихын бақылау үшін snapbird құралын пайдаланған. Деректер Twitter деректерін талдау құралы болып табылатын followthehashtags.com арқылы алынғандығын көруге болады.

Ендігі кезекте классикалық машиналық оқыту алгоритмдеріне тоқталатын боламыз. Бұл әдіс қолдануға жеңіл әдіс деп те аталады. Бұл әдіс оқу мақсаттары үшін пайдаланылатын қол арқылы немесе автомат режиміндегі кодпен алмастырлыған деректер базасына сай жасақталған. Мұндай таңбалы деректер базасы пікірлерді ғадауат тілді сөздер немесе бейтарап сөз жағдайында қарастыру және оларды жіктеуге негізделген модель жасауда оқу алгоритмдерін үйрету мақсатында қолданылады (кесте 2.1. –де көрсетілгендей).

Кесте 2.1 - Ғылыми зерттеу жұмыстардағы машиналық оқыту алгоритмдерінің пайдаланылуы

Ғылыми зерттеу жұмыстары	Классификатор	Жаңалығы	Ерекшелігін анықтау	Бағалау көрсеткіші
1	2	3	4	5
[79]	NB, RF, LG, DT, SVM, DL	Исламифобия сөздерін анықтау	Сөздерді енгізу	Accuracy, precision, recall және F1
[80]	DL	Контекстегі ғадауат сөздер	Енгізу	Accuracy, precision, recall және F1 бағалау
[81]	Ансамбльдік әдістер	Мульти мета оқыту моделі	n – gram дағы характерлер n – gram дағы сөздер	Precision, Recall және F1 бағалау

2.1-кестенің жалғасы

1	2	3	4	5
[82]	GRU	Амхар тілі бойынша жаңа оқыту	Word2 Vec	accuracy, ROC, AUC
[74]	SVM, NB, DT, RF	Араб тілді контекстерден ғадауат сөздерді анықтау	BoW және TF – IDF	Accuracy, precision, recall
[83]	NB, LR, SVM, KNN,DY, RF	Алмастырушы кодтардың адресстері	TF – IDF	Шатастыру матрикасы
[84]	LR және LSTM	Ғадауат сөздердің мульти – лингвистік анализі	BoW	F1 бағалау
[85]	RF	Ғадауат тілді сөздер үшін жақсартылған RF	Векторлар санамасы	F1 бағалау, Precision, Recall
[58]	Lexicon, RNN	Араб тіліндегі мәліметтер қорын құру	N – gram, ендіру	F1 бағалау, Precision, Recall, AUROC
[83]	SVM, NB және RF	Эмоциялық анализ жүргізу	N – gram	Precision және Recall
[71]	RF, SVM	3 түрлі мәлімттер қорын салыстыру және	Unigrams	Precision, Recall және F1 бағалау
[86]	n – Gram сөздер	Кибербулингті анықтау	BoW	F1 бағалау, Precision және Recall

Кесте 2.1 берілген мысалдар көрсеткендей векторлық машиналары (SVM), Naive Bayes (NB), Логистикалық регрессия (LR), Шешім ағаштары (DT), K-Ең жақын көрші (KNN) және т.б. ғадауат тілді сөздерді анықтау үшін жиі қолданылатын ML алгоритмдері кестеде жинақталған. Кестеден байқағанымыздай SVM SM деректерін ғадауат тілді сөздерді немесе бейтарап деп жіктеу үшін зерттеушілердің ең көп пайдалану санына ие. Кездейсоқ орман - бұл рейтингте екінші, логистикалық регрессия және аңғал Бейс те айтарлықтай деңгейде жақсы қолданылған.

Келесі кезекте мәтіндік тексттерді өңдеу үшін көп пайдаланатын негізгі әдістердің мысалдары және олардың ерекшеліктеріне байланысты сипаттық ақпараты ұсынылған:

1) *Naive Bayes (аңғал Байес):* Бұл әдіс ықтималдық теориясы қарастыратын қарапайым және өнімділігі жоғары алгоритм. Ол мәтіндік деректерді жіктеуде кеңінен қолданылады және мәтінді жіктеудің негізгі үлгісі ретінде жиі қолданылады [19].

2) *Қолдау векторлық машиналары (SVMs):* Бұл алгоритмдер әртүрлі класстарды барынша бөлетін көпөлшемді кеңістікте (параметрлік кеңістік) гипержазықтықты табуға бағытталған [76]. Олар шағын және орта өлшемді деректер жиыны бар мәтінді жіктеу тапсырмалары үшін тиімді. Алгоритмдердің бұл түрі ML есептерін әрі қарай түсінуге және эксперименттерді бастауға жақсырақ.

3) *Шешім ағашы (Decision tree):* Бұл алгоритмдер мәтіндерді жіктеу үшін пайдалануға болатын ағашқа негізделген шешім үлгісін жасайды. Оларды түсіндіру және енгізу оңай, бірақ шамадан тыс орнатуға бейім болуы мүмкін, сондықтан бастапқыда осы модельге өлшемді азайту тапсырмасын қолдану мағынасы бар [21].

4) *Кездейсоқ орман әдістері (Random Forest):* Бұл алгоритмдер шешім ағаштарының ансамблін жасайтын және мәтіндерді жіктеу үшін пайдаланатын шешім ағаштарының жетілдірілген нұсқасы болып табылады. Олар жалғыз шешім ағаштарына қарағанда шамадан тыс орнатуға аз сезімтал және көбінесе жіктеу тапсырмаларын жақсы орындайды [87].

5) *К-ең жақын көршілер (KNN) әдісі.* Бұл қарапайым және тиімді жіктеу және регрессия әдісі. Екі жағдайда да кіріс деректері мүмкіндіктер кеңістігіндегі ең жақын k жаттығу мысалдарынан тұрады [88]. KNN классификациясында шығыс класс мүшелігі болып табылады. Нысан көршілерінің көпшілігі бойынша жіктеледі, ал нысанға жақын көршілерінің арасында жиі кездесетін класс тағайындалады.

6) *Логистикалық регрессия (Logistic regression).* Бұл таңбаланған оқу деректерінің жиынтығын алатын және жаңа, көрінбейтін деректердің екілік белгісін болжауды үйренетін бақыланатын оқыту алгоритмі [89]. Логистикалық регрессияның мақсаты берілген кіріс сигналының белгілі бір класқа жататын болу ықтималдығын болжайтын ең жақсы параметрлерді немесе салмақтарды табу болып табылады.

Мәтінді жіктеу тапсырмасы үшін сәйкес әдісті таңдау өте маңызды және ол толықтай деректер жиынтығының өлшемі мен күрделілігіне, сонымен қатар қолданыстағы ресурстарға байланысты. Осыны ескере отырып, алғашқы зерттеулер ағылшын тіліндегі бұрыннан бар деректер жиынтығын пайдаланады деп келісілді. Бастапқы нүкте болған Twitter деректер жинағын пайдалану арқылы машиналық оқытудың негізінде жатқан тұжырымдамалар туралы толық білімге қол жеткізілді [88].

2.2 Нейрондық желілер классификациясы және енгізу параметрінің баптаулары

Мәтінді талдау үшін әртүрлі классификаторлар бар болғанымен, мәтіндік деректер қорын ұсынудың күрделілігі айтарлықтай қиындық тудырады [38]. Мәтін ерекшелік ретінде қарастырылатын көптеген сөздерді қамтуы мүмкін; дегенмен, олардың көпшілігі мақсатты сыныпқа қатысты ұйымдастырылмаған, маңыздылығы аз, шулы (артық деректердің көп болуы) немесе артық болуы мүмкін. Бұл классификаторлар арасында шатасуларға әкелуі мүмкін, осылайша

олардың өнімділігін төмендетеді [30, 32]. Демек, шулы, аз ақпараттандыратын және артық құрамдастарды жою үшін мүмкіндікті таңдау әдістерін қолдану маңызды. Бұл мүмкіндік кеңістігін азайтады, басқару мүмкіндігін арттырады және жіктеуіштердің дәлдігі мен өнімділігін арттырады.

Ғадауат тілді сөздер кездесетін белгілерді таңдау тәсілі төрт негізгі қадамнан тұрады: белгілер жиынын жасау, ішкі жиынды бағалау, тоқтату (stop words) критерийлерін орнату және жіктеу нәтижелерін тексеру [39]. Бастапқыда ізденуші мүмкіндіктің ішкі жиынын анықтау үшін зерттеу әдісі пайдаланылады, ол кейін келесі қадамда белгіленген қабылдау критерийлері бойынша бағаланады. Үшінші кезеңде тоқтату критерийлеріне жеткеннен кейін жиынтықты құру және бағалау процестері аяқталады, нәтижесінде барлық ұсынылған мәліметтер қорынан оңтайлы белгілер жиыны таңдалады. Белгілер жиыны кейіннен соңғы кезеңде тексеру жиынын пайдаланып бағаланады.

Функцияларды таңдау тәсілдерін мүмкіндіктердің ішкі жиындарын құру әдісіне қарай төрт топқа бөлуге болады: *сүзгі үлгісі* [26,62,88], *сөздерді «орау»* үлгісі [34, 40], *ендіру үлгісі* [41] және *гибридті модель* [36]. Мәтінді жіктеуге арналған мүмкіндіктерді таңдаудың көптеген әдістері олардың жоғары дәлдігі мен өнімділігіне байланысты сүзгі тәсілдеріне сүйенеді. [31, 34] жалпы деректерге қолданылатын әртүрлі ғадауат тілді сөздер кездесетін белгілерді таңдау әдістерін жан-жақты талдау мен салыстыруды ұсынады.

Ағымдағы зерттеулер машиналық оқыту әдістерін қолданатын онлайн платформалар мен әлеуметтік желілерде кибербуллингті анықтау үшін арнайы әзірленген оқыту алгоритмдерінің тапшылығын көрсетеді. Зерттелген көптеген зерттеулер (мысалы, [34, 37, 30–32]) машиналық оқыту үлгілерін оқытудағы маңызды қасиеттерді анықтау үшін ғадауат тілді сөздер кездесетін белгілерді таңдауды пайдаланбады.

Кейбір мәтіндік деректер жинақтары өте үлкен және сызықты түрде бөлінбейді, сондықтан классикалық ML оны тиімді талдай алмайды. Сызықтық бөлінбейтін деректердегі мағыналы тенденцияларды болжау мәселесін шешу үшін DL алгоритмі ұсынылды. DL жай ғана жасанды нейрондық желі (ANN) деп аталатын ML алгоритмінің кеңейтімі болып табылады [90]. Тереңдігі әдетте берілген мәселенің шешу күрделілігіне байланысты. Мысалы, кескінді өңдеу тапсырмасы әдетте SM мәтінін болжау тапсырмаларына қарағанда тереңірек қабаттарды қажет етеді [91].

Зерттеушілердің назары қайталанатын нейрондық желі (RNN) және конволюционды нейрондық желіге (CNN) аударылды, өйткені олар сөйлем семантикасын жақсырақ бейнелей алуында болып табылады. CNN әсіресе сөздердің мазмұнын талдауда сөздердің семантикасы мен синтаксисін түсіруде тиімді екені дәлелденді [89].

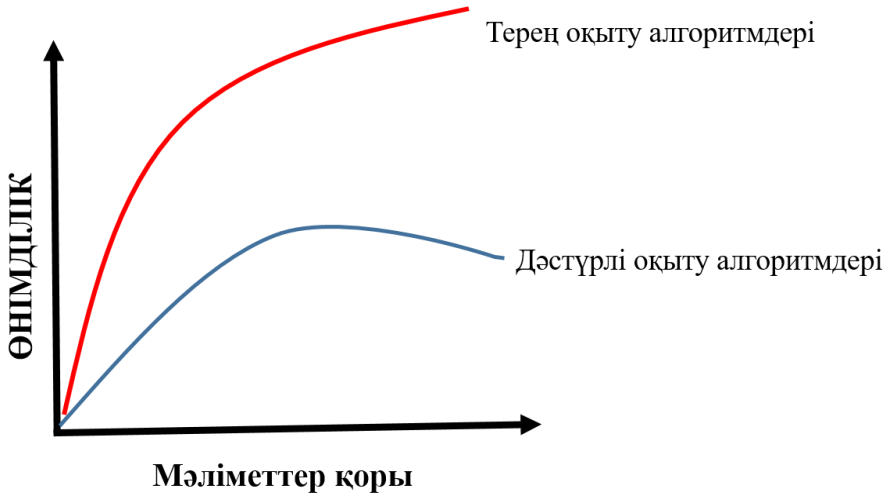
Әлеуметтік желілерде өшпенділік сөздерін анықтау үшін әртүрлі терең оқыту тәсілдері қолданылды. Ұзақ қысқа мерзімді жад (LSTM) және қақпалы қайталанатын бірліктер (GRU) зерттеуде пайдаланылған CNN және RNN екі ерекше түрі болды [88]. SemEval-2019 6-тапсырмасының шешімі табылды, ол қатысушылардан әлеуметтік медиа платформаларында жарияланған қарсы

мәтіндерді тану және санаттауды талап етті. Бұл стратегиялар мәселені шешу үшін қолданылды. Зерттеушілер сонымен қатар [92] ұсынған екі модельді, атап айтқанда LSTM-CNN және CNN-LSTM әдістерін зерттеді. Бұл модельдер сыналған. Зерттеуді аяқтағаннан кейін авторлары BiLSTM-CNN моделі жақсырақ F1 жалпы ұпайын ұсынды деген қорытындыға келді.

Бұдан басқа, келесі зерттеу ғадауат тілді сөздерін анықтау мақсатында үш түрлі терең нейрондық желі (DNN) алгоритмдерін сынады. LSTMs, CNNs және FastText Convolutional Neural Networks осы зерттеуде пайдаланылған DNN түрлері болды [93].

Терең оқыту (DL) мен машиналық оқыту (ML) арасындағы негізгі айырмашылық DL үлкен деректер жиынын тиімді өңдеуді қажет етеді, ал ML кішірек деректер жиынынан үйрену үшін графиктерді пайдаланады [94].

Терең оқыту алгоритмдерінің жаңа білімді меңгеру қарқыны қызыл сызықпен көрсетіледі. Ол тік ось деп те аталатын өнімділік осінің деректер жиынының көлемінің ұлғаюымен үнемі өсетінін көрсетеді, бұл алгоритмнің барған сайын тиімдірек болып келе жатқанының көрсеткіші болып табылады. Бұл дегеніміз, деректер неғұрлым көп болса,терең оқытудың өнімділігі соғұрлым жақсы болады. Ғадауат тілді сөздерді автоматты түрде анықтау бойынша жүргізілген алдыңғы зерттеулердің үлкен саны негізінен әлеуметтік желілердегі бейәдеп тілді сөздердің әртүрлі формаларын анықтау үшін дәстүрлі машиналық оқытуға бағытталған (2.1-суретте көрсетілгендей). Әлеуметтік медиада жасалған деректер күнделікті жоғары жылдамдықпен өсіп отырады, сол себепті көлемдері өте үлкен болып табылады [95]. Бұл мәселені шешу үшін терең оқыту алгоритмдерін қолдану тиімді деп саналады. Келесі ұсынылып отырған 2.2 – кестеде өшпенділік сөздерін анықтау үшін терең оқытудың кейбір салыстырмалы өнімділігін көрсетеді.



Сурет 2.1 - Мәліметтер қорының өнімділігі
(сурет автордың жеке өнімі)

Кесте 2.2 - Ғадауат сөздерді анықтауда салыстырмалы ғылыми зерттеу жұмыстары

Ғылыми зерттеу жұмыстары	Оқытудың мақсаты	Ерекшеліктерді тану әдістері	Терең оқыту алгоритмдері	Бағалау көрсеткіші
[95]	Сөзбен кемсіту мәселелерін шешу	Сөздер енгізу	CNN	std deviations =0.84
[96]	Араб тіліндегі пікірлерде ғадауат сөздерді анықтау	n – gram характерлері және CBOW	CNN RNN	Pr = 0.81 Rc = 0.78 A = 83 Auc = 0.89 F1 = 93.35
[93]	Ғадауат сөздерді анықтау	CBOW Skip – gram жалғастырушы	CNN, LSTM, CNN+GR U	Pr = 0.93 Rc = 0.93 F1 = 0.93
[97]	Расизм, сексизм бойынша пікірлерді классификациялау	n – gram TF IDF	CNN LTSM	Pr = 0.93 Rc = 0.93 F1 = 0.93
[43]	Ғадауат сөздерді анықтау және анықтама беру ӘЖ бойынша	NA	Deep LSTM	A = 90.82 Pr = 83.82 Rc = 84.23

2.3 Ғадауат тілді сөздерді анықтауда өнімділікті бағалау метрикасы

Тиімділікті бағалау – әдетте өнімділікті бағалау метрикасын қолдану арқылы жүзеге асырылатын барлық оқыту бойынша зерттеу мәселесі. Тиімділікті бағалау көрсеткіштері нақты мәндер мен болжамды мәндер арасындағы айырмашылық арқылы алынатын логикалық-математикалық құрылымдар [98].

Бейәдеп сөздерді анықтау үлгілерінің өнімділігін бағалау әдетте классикалық дәлдік, қайта шақыру және F1 балл көрсеткіштерін пайдалану арқылы жүзеге асырылады. Әдетте ғадауат тілді сөздер деректер жинағының теңгерімсіз сипатына байланысты пайдаланылады. Кез келген теңдестірілген деректер жиынтығы үшін дәлдікті анықтау ең жақсы нұсқа болып табылады [8].

Ұсынылатын модель пікірлерді ғадауат тілді сөздер және ғадауат сөйлеу тіліне жатпайтын бейтарапты сөйлеу тілі деп жіктеуге үйретілді делік. Мысалы, 5 пікір ғадауат тілді сөздер және 15-і бейтарап мағынадағы 20 пікірлер жиынтығы бар. Модель 6 пікірді бейәдеп сөздер ретінде анықтай алды. Анықталған 6 твиттің (пікірдің) 4-і іс жүзінде ғадауат тілді сөздер (шынайы позитивтер) және 2-еуі бейтарап сөздер (жалған позитив) болды. Модель ғадауат тілді сөздер болып табылатын 2 пікірді (жалған теріс) қате жіктеді және 13 пкір бейтарап пікір (шын негативтер) ретінде нақты алынып тасталды.

Дәлдік (Pr - precision)

Дәлдік (precision) – шынайы оң және жалпы болжамдардың арақатынасы. Үлгінің өнімділігін бағалау мақсатында дәлдікті пайдаланады.

Математикалық нұсқасын төмеендегідей етіп көрсетуге болады (2.1):

$$Pr = \frac{TP}{TP+FP}, \qquad (2.1)$$

Pr - зерттеу мақсатында қолднау үшін дәлдіктің қысқаша нұсқасы. Дәлдік жай ғана модельмен дұрыс анықталған оң классификациялардың бір бөлігін білдіреді [8]. Мысалы, жоғарыдағы мысалдан дұрыс анықталған нақты позитивтердің үлесі 4. Сонда үлгі дәлдігі 4/6 (шын позитивтер/барлық позитивтер) = 0,67.

TP – шын позитивтің қысқаша мағыналы нұсқасы. Жоғарыдағы көрсетілген мысал бойынша TP – 4. 5 ғадауат тілді пікірден үлгі 4-ін бейәдеп сөз ретінде дұрыс анықтай алды.

FP жалған позитивті білдіреді. Бұл ғадауат тілді сөздер ретінде жіктелген өшпенділік тудырмайтын пікірлерге қатысты. Жоғарыдағы мысал бойынша 2 пікір ғадауат тілді сөздерден тұратын пікір ретінде жіктеліп жіберілген және шындығында олар бейәдеп сөзді пікірлер қатарына жатпайтындығын байқауға болады [99].

Қайтару (Rc - recall)

Rc – таңдау кеңістігіндегі дұрыс болжамдар мен барлық дұрыс бақылаулар санының қатынасы. Математикалық тұрғыдан (2.2):

$$Pc = \frac{TP}{TP+FP}, \quad (2.2)$$

Rc бұл формулада Recall дегенді білдіреді. Бұл дұрыс анықталған нақты позитивтердің үлесін білдіреді. Мысалды еске түсіретін болсақ 4/5 (шын позитив/барлық оң) = 0,8. Бұл дегеніміз модель 80% ғадауат тілді сөздерден құралған пікірлерді дұрыс анықтай алды дегенді білдіреді.

FN жоғарыдағы көрсетілген мысал үшін жалған теріс мағынаны білдіреді. Бұл модель ғадауат тілді сөз ретінде анықталмаған ғадауат тілді пікірлерге қатысты. Модель оларды ғадауат тілді емес, яғни бейтарап пікірлер деп санады, ал олар шынайы мағынада ғадауат тілді сөздерден тұратын пікірлер болды. Жоғарыдағы мысалда тек бір пікір бейәдеп тілді пікір деп қате жіктелді.

F-MEASURE

Дәлдік пен F1-балы (F) жұмыс өнімділгінің өлшенген гармоникалық орташа мәнін (whm) білдіреді деп есептеледі. Бұл статистиканы пайдалану жиі кездеседі, әсіресе деректер жинағы теңгерілмеген жағдайларда. Ол [8,100] және [15] сияқты ғылыми басылымдарда өшпенділік сөздерін анықтау үлгілерінің тиімділігін бағалауда қолданылған.Математикалық тұрғыдан (2.3):

$$F = 2 * \frac{Pr*Rc}{Pr+Rc}, \quad (2.3)$$

F-өлшемі немесе F1-балы үшін қысқа және теңдестірілмеген класс үлестіруімен үлгі өнімділігін тексеру үшін пайдаланылады. Нақты өмірдегі

мәтінді жіктеу тапсырмаларының көпшілігінде теңдестірілмеген сыныпты бөлу орын алады, сондықтан F1 ұпайы модельді тексеру үшін ақылды метрика болып табылады [101]. Жоғарыдағы мысалдан F = 2(0,67∗0,8)/(0,67 + 0,8) = 1,072/1,47 = 0,72, бұл жай ғана модельдің F1-өлшемі 72% дегенді білдіреді.

Дәлдік (Accuracy A)

Дәлдік – дұрыс болжау мен жалпы бақылаулардың арақатынасы. Модельдің дәлдігі, егер бізде FP және FN мәні екі класты есеп үшін тең болатын симметриялы деректер жинағы болған жағдайда ғана жақсы деп саналады. Көптеген және теңгерімсіз деректер жиындарында дәлдік ең жақсы нұсқа емес, сондықтан F1-балы сияқты басқа бағалау параметрлерін қарастыруға болады. Келесі ғылыми зерттеулерде,[102], [103] және [80] дәлдік қолданылды. Математикалық түрде дәлдікті (A) мына түрде көрсетуге болады (2.4):

$$A = \frac{TP+TN}{TP+FP+FN+TN}, \quad (2.4)$$

Деректер жинағын жинауда және бейәдеп сөздерді анықтауда кездесетін қиындықтар:

Бірінші мәселе - ғадауат тілді сөздердің әлемнің әртүрлі аймақтарындағы сөйлеу ерекшеліктеріне байланысты болуы;

Екінші мәселе ретінде - деректер жиынтығының аздығы кездеседі. Twitter мәліметтері бойынша, мысалы, бір твит ең көбі 140 таңбадан тұруы мүмкін. Твиттен жиналған ақпарат мұндай жағдайларда белгілі бір перспективаны жалпылау үшін сәйкес келмеуі мүмкін. Әлеуметтік желілердегі қысқа хабарламалар мен пікірлерді өңдеу қиындық тудыратын мәселелер қатарын толықтырады.

Үшінші мәселе ретінде – ғадауат тілді сөздерді анықтауда классқа бөлуде кездесетін қиындықтар. Мысалы, пікірді бейәдеп сөздер классына да бейтарап пікірлер классна да жатқыза алмауында. Пікірлерді екі классқа бөлер кезде дисбаланстық күй туғызуында.

Мәдени бағыттағы өзгерістер ғадауат тілді сөздердің анықтамасына тікелей әсер етеді немесе бейәдеп тілді сөздер мәдениет пен дәстүрге байланысты өзгереді. АҚШ-та қалыпты сөйлеу деп саналатын нәрсені, мысалы, Нигерияда кемсіту мақсатында қолданылғансөз тіркесі ретінде қарастыруға болады. Адамдардың мәдениеті мен дәстүрі сөйлеуді кемсітіп, қорлайтын немесе ол классқа жатпайды деп жіктеуде үлкен рөл атқарады. Сарапшылар әлеуметтік желілер провайдерлеріне өздерінің платформаларында ғадауат тілде сөйлеу мәселесін кешенді түрде шешуге кеңес берді.

Пандемия немесе табиғи апат құрбандары стереотиптері болуы мүмкін. Мысал ретінде, COVID19 пандемиясы, онда қытайлар әлемнің көптеген жерлерінде стереотипке айналды, «Қытай вирусы», «қытай ауруы» деген сияқты сөздерді бейәдеп сөздер сыныбына немесе бейтарап сөздер сынбына жатқызуда біршама дау туғызатыны анық.

Төмендегі кестеде мәліметтер қорын жинау кезінде орындалған ғылыми зерттеу жұмыстарына анализ жасалған [91].

Кесте 2.3 - Мәліметтер қорын жасақтаудағы ғылыми жұмыстар салыстырмасы

Ғылыми зерттеу жұмысы	Мәліметтер жинау көзі	Қосылған үлес	Шектеулер
1	2	3	4
[90]	Twitter	Жаңа деректер қоры жасақталды. Аннотаторларға арналған нұсқаулық жасалды Исламиафобияға анықтама берілді Inter – coder бойынша келісушілік 89,9% жеткен	Жиналған дерекқорлар Ұлыбритания саясаткерлерін бақылаушылармен шектеледі. Деректер қорын жинау шектеулерге байланысты біркелкі болып шықпады. Тек исламифобияға назар аударылды. Сөздердің мазмұнына назар аударылмады. Мәліметтерді қайта өңдеу кезде сандық белгілермен қатар құнды мәліметтердің жойылғандығын байқауға болады.
[104]	Twitter	Геторгенді пікірлердің қамтылуы Көптеген ғадауат тілді сөздер қарастырылды	Бейәдеп сөздер бойынша жалпылама түрде жасақталған нұсқаулықты жетілдіру қажеттілігі туды. Денсаулығына, отбасылық жағдайына, ұстанатын жынысына (трансгендер) сияқты «сезімтал салалар» қарастырылды. Көңіл – күйді білдіретін таңбалар мен сандық белгілер алдын – ала өңдеу кезінде жойылып отырылды.
[90]	Twitter	Гетерогенді қоғамды қамту жақсы жүргізілді Аннотатор ретінде мәліметтер қорын санаттарға бөлуден бұрын оқуытудан өткен Оңтүстік Африка саясаткерлері алынды.	Оңтүстік Африка тілінде жазылмаған басқа мәліметтер қорына назар аударылмады. Сандық символдар қайта өңдеу процессі кезінде жойылып отырды Аз мөлшердегі аннотатор жұмыс істеді, яғни бір пікірді бірінші аннотатор ғадауат сөздер классына жатқызса, екнішісі бейтарап пікір қатарына жатқызып отырған.

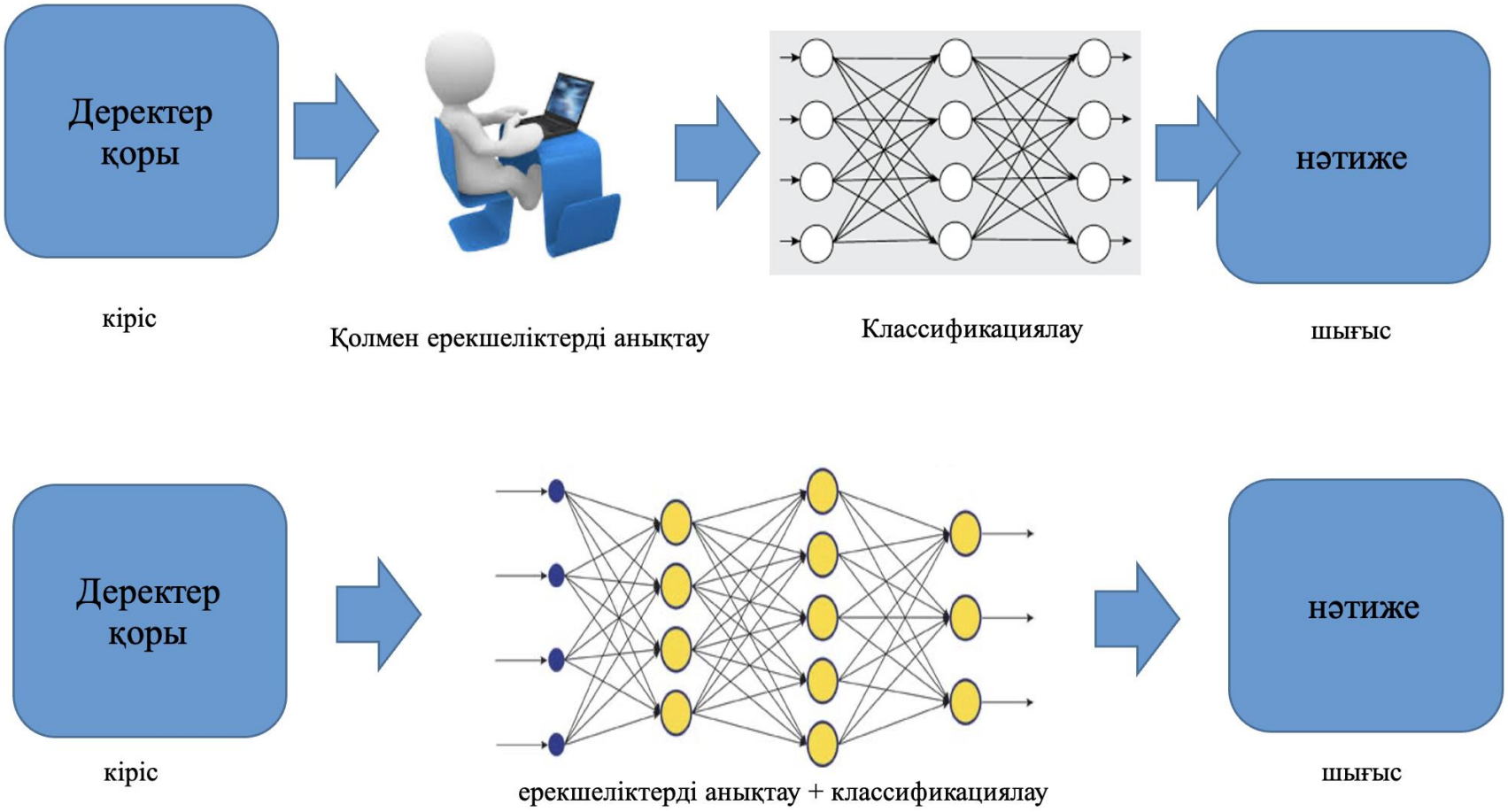
2.3-кестенің жалғасы

1	2	3	4
[105]	Facebook	Аннотаторға арналған нұсқаулықта жақсы мысалдар келтірілген Ғадауат пікірлер мәліметтер қорының жақсы қамтылуы	Дайын мәліметтер қоры пайдаланылған Араласқан тілде жазылған пікірлерді қарастырмаған Мағынасы бар сандық белгілерді ескермеген. Трансгендер және отбасылық жағдай бойынша мәселе ескерілмеген
[106]	Twitter	Ағылшын тілінен басқа кибер кемсітуді жақсы зерттеген Кросс валидация деңгейі 10- ға дейін жақсы көтерілген.	Тек испан тіліндегі пікірлер зерттелген Сандық мәтіндер жойылған Эмодзилер ескерілмеген Код аралас деректер жазбалары қарастырылмаған
[107]	Facebook	Аннотаторға арналған нұсқаулықтағы мысалдардың жақсы келтірілуі Ғадауат тілді пікірлердің қамтылу аймағының кең болуы	Мағына беретін сандық белгілердің алынып тасталуы Отбасылық жағдайға байланысты қолданылатын ғадауат сөздердің ескерілмеуі
[108]	Twitter	Текстің мазмұнына жіті назар аударылды Анықтамалар мен мысалдар аннотаорларға жақсы көрсетілген	Аннотаторлаға толықтай көмектесе алатын нұсқаулық жоқ.

Ғадауат тілді сөздер деректер қоырының критерийлерін белгілеу, арнайы деректер жинақтарын құру және әртүрлі мәселелерді зерттеу, соның ішінде исламофобия мен кибер-кемсітушілік Twitter негізіндегі зерттеулерге негізгі үлес болып табылады ([88], [102], [103], [104], [106]). Әртүрлі қауымдастықтарды және денсаулық, гендерлік сәйкестік және отбасы мәртебесі сияқты нәзік тақырыптарды зерттей отырып, бұл зерттеулер инклюзивтілікті де атап өтті. Дегенмен, зерттеушілер жиі аралас кодты мәтіндерді және эмодзилер сияқты эмоционалды белгілерді елемейді және зерттеулер тілге тән бейімділіктерге байланысты белгілі бір жерлерде (мысалы, Ұлыбритания, Оңтүстік Африка немесе испан тілінде сөйлейтін адамдар) шектеледі. Деректер жиынтығын әзірлеу және жіктеу сенімділігі сәйкес келмейтін аннотаторлар мен әлсіз оқыту стандарттары арқылы одан әрі кедергі болды.

Facebook негізіндегі зерттеулердегі аннотация стандарттары ([103], [105]) өшпенділік сөздерінің деректер жиынын кеңінен қамтуды көрсетті және нақты мысалдар келтірді. Осы артықшылықтарға қарамастан, тек бұрыннан бар деректер жиынын пайдалану, сандық белгілер сияқты күрделі идентификаторларды есепке алмау және гендерлік бейімділік пен отбасы мәртебесі сияқты әсер ететін субъектілерді болдырмау сияқты шектеулер болды.

Зерттеулер көрсеткендей, әртүрлі деректер жинақтары мен жақсырақ ережелерді пайдалану өшпенділік сөздерін анықтауды жақсартуы мүмкін, бірақ ол сонымен қатар көптілділік, аннотаторларды оқыту және сандық деректер мен эмодзилерді есте сақтау сияқты мәселелерді шешуге көбірек күш салу керек екенін көрсетеді. 2.2-суретте функцияларды шығарудың және мүмкіндіктерді іздеудің негізгі мақсаты - әртүрлі контексттерден мәтінді машиналық оқыту үлгісімен оңай талдауға болатын кілт сөздер жинағына түрлендіру. Құжатқа тән фразалар жиілігі сияқты қосымша мәтіндік ақпараттар да осы процедура арқылы жасалады. Дұрыс кілт сөздерді таңдау және оларды кодтау жолын анықтау - бақыланатын машиналық оқытудағы маңызды қадамдар. Классификация жүйесінің идеалды үлгілерді анықтау мүмкіндігі көп жағдайда осы іздеу фразалары мен түйінді сөздерді таңдауға байланысты [73].



Сурет 2.2 – Мәліметтер қорын жинау алгоритмі
(сурет ізденушінің авторлық өнімі)

Бұл тарауда ғадауат тілді сөздерді анықтаудағы IEE Xplore, Springer Link, Science direct сияқты онлайн дерекқорларынан алынған Индонезия, Оңтүстік Африка, Испан және Урду тілдеріндегі ғадауат сөздерді анықтау зерттеу жұмыстары талқыланған. Яғни, деректер қорының классификациясы, ғылыми жаңалығы, машиналық және терең оқыту алгоритмдері және бағалау көрсеткіштеріне салыстырмалы зерттеу жүргізген.

3 ТАБИҒИ ТІЛДЕРДІ ӨҢДЕУ ЖӘНЕ МАШИНАЛЫҚ ОҚЫТУ ӘДІСТЕРІН ОНЛАЙН КОНТЕНТТЕ БЕЙӘДЕП СӨЗДЕРДІ АНЫҚТАУДА ҚОЛДАНУ

Онлайн контенттегі ғадауат тілді автоматты түрде анықтау үшін Natural Language Processing (NLP) және Machine Learning (ML) әдістемелерін пайдалану зерттеудің негізгі саласына айналды.

Бұл тарауда NLP және ML-дің ауқымды, әр түрлі мәтіндік деректер жиынындағы ғадауат тілді сөздерді анықтаудағы теориялық және практикалық маңыздылығын зерттейді. Зерттеу мәтінді жіктеу әдістерінің дамуын зерттейді, мысалы, сөз қапшығы (BoW) және терминдік жиілікке кері құжат жиілігі (TF-IDF) сияқты дәстүрлі әдістерден бастап және Transformers екі жақты кодтаушы өкілдіктері (BERT) сияқты анағұрлым жетілдірілген және контекстке сезімтал терең оқыту үлгілерінің ерекшеліктерін зерттейді.

Бақыланатын оқыту әдістерін, соның ішінде тірек векторлық машиналары (SVM), логистикалық регрессия, кездейсоқ ормандар және таңбаланған деректер жиынында оқытылатын терең нейрондық желілерді қолдануға маңызды назар аударылады. Зерттеу деректер аннотациясындағы кедергілерді, соның ішінде субъективтілікті, сыныптық бөлу теңгерімсіздігін және доменге тән нормалардың қажеттілігін, әсіресе нәсіл, дін, гендерлік сәйкестік және саяси іс – әрекеттердегі сезімталдығы жоғары сөйлемдерін қатарынан ғадауат тілді сөздерді автоматты анықтауды зерттейді.

Сондай-ақ, ғадауат тілді сөздерді анықтауға арналған бірнеше машиналық оқыту әдістемелерін мұқият салыстыра отырып, дәстүрлі әдістерден BERT сияқты контекстік үлгілердің артықшылығына баса назар аударады. Сонымен қатар, токенизация, эмодзилерді басқару, сандық таңбаларды өңдеу және тоқтату сөзін жою сияқты алдын ала өңдеу әдістерінің салдары үлгі тиімділігіне әсер ету ықпалдылығы мұқият зерттеледі.

Онлайн контенттегі ғадауат тілді сөздерді жедел және дәл анықтау арқылы интернет қауіпсіздігін арттырудағы NLP және ML мен DL революциялық мүмкіндіктерін көрсетеді.

3.1 Табиғи тілдерді өңдеу және машиналық оқыту әдістерінің рөлі мен маңызы

Табиғи тілді өңдеу (NLP) – күрделі және шешмі қиын мәселелерді шеше алатын машиналық аудару, машинаның сұрақтарға адамша жауап беурін қамтамасыз ететін жасанды интеллектің бір саласы. NLP адам тілдерін түсінудегі практикалық мәселелерді шешу үшін модельдерді, жүйелерді және алгоритмдерді жобалауды және енгізуді қамтиды.

NLP-ді іргелі (немесе негізгі) және қолданбалы зерттеулер болып табылатын екі негізгі салаға бөліп қарастыруға болады. Бірінші санатта, адам тіліне негізделген күрделі жүйелерді құрудағы мәселелерді көруге болады. Бұл тапсырмалардың кейбірі тілдік модельдеу, морфологиялық талдау, синтаксистік өңдеу немесе талдау және семантикалық талдау болып табылады.

Сонымен қатар, NLP мәтіндерден тиісті ақпаратты автоматты түрде алу тіларалық мәтіндерді аудару, құжаттарды орталықтандыру, сұрақтарға автоматты түрде жауап беру, құжаттарды жіктеу және кластерлеу сияқты тақырыптарды қамтиды.

Мәтінді алдын ала өңдеу көптеген мәтіндермен жұмыс істеу алгоритмдерінің негізгі компоненттерінің бірі болып табылады. Ол әдетте *токенизация*, *лемматизация* және *стемминг* сияқты тапсырмалардан тұрады [109]. Таңбалар тізбегінен сөйлемді бөліп алғаннан кейін келесі қадам - тексеру процесі, оның нәтижесінде мәтін жеке сөздер немесе бірнеше сөзден тұратын тізбектер түріндегі бөліктерге (лексемалар деп аталады) бөлінеді. Токенизация процесі үшін n − gram моделі пайдаланылады, мұндағы n – тізбекті құрайтын сөздердің саны, ал n грамм – бір сөз [110]. Мәтін жазылған тілдің грамматикалық құрылымына қарай қалпына келтіру процесі орындалуы керек. NLP-де жиі қолданылатын қалыпқа келтіру әдістері - стемпинг және лемматизация. Бұл сөздердің әртүрлі формаларға қарай өзгертуді және бастапқы қалыптағы мағынасын өзгертпей отырып, мәтіндерді машинаның ыңғайлы формасына келтіру.

Стемминг (түбір сөзге келтіру) және лемматизация.

Стемминг дегеніміз – бір түбірлі сөздердің барлығын жалпы формаға ауыстыру, әдетте әр сөзден жалғау, шылау және жұрнақтарын алып тастау [111]. Түбір сөзге келтірудің негізгі мақсаты – сөздің зат есім, сын есім, етістік, үстеу т.б түрлі сөз формаларын оның түбір сөз түріне келтіру [20].

Дегенмен, автоматты морфологиялық талдауда сөздің түбірі оның грамматикалық құрылымына анықтама ретінде қолданылатын жұрнақтарға қарағанда аз қызығушылық тудыруы мүмкін [112]. Лемматизация сөздің негізгі түбіріне жетуді көздейді. Бұл үдеріс берілген сөйлемдегі сөз бөлігін және сөздің контекстін тану арқылы сөз формаларын олардың түбір формасына келтірумен айналысады. Шын мәнінде, сөзді лингвистикалық тұрғыдан дұрыс түбір сөзге келтіру алгоритмі лемматизатор деп аталады.

Лемматизация сөздерді сөздік және морфологиялық талдау арқылы дұрыс орындауды, көбінесе тек флексия жалғауларын алып тастауды және лемманы (әр түрлі тұжырымдарды дәледеу барысында қолданылатын көмекші сөйлем) қайтаруды білдіреді [113].

NLP-де мәтіндердегі сөздердің кездесуін санайтын статистикалық әдістер жиі қолданылады. Олар жолды өңдеуде қолданылатын қажетті айнымалыларды қамтамасыз етеді. Құжаттарды көрсетудің әртүрлі тәсілдері бар, олардың ішінде ең көп қолданылатыны график немесе терминдер векторы түрінде болады. Ғылыми зерттеу жұмыстарында n мүшенің жиыны құрылады [114], одан кейін циклдік график тұрғызылады және төбелер арасындағы ұқсастық есептеледі, авторлары өз мақалаларында терминдердің реті, жиілігі және контексті туралы ақпаратты қамтитын мәтіндік графикті ұсынған [115]. Екінші жағынан, [116] жұмыс векторлық тәсілді графиктік тәсілмен салыстырады. Термин векторы ретінде құжатты ұсыну терминдерден тұрады, ал вектордың мәндеріне сәйкес термин яғни сөз салмақтары болып табылады.

Сөздердің салмақтық өлшемдері

Термин жиілігі (TF) өлшемі қарапайым, бірақ тиімді терминді бағалау функциясы болып табылады. Ол әрбір терминнің барлық құжаттарда неше рет кездесетінін санауға негізделген [117] және төмендегі формуламен анықталады (3.1):

$$TF_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}}, \quad (3.1)$$

мұндағы: $n_{i,j}$ – d_j құжатындағы ti терминінің өңделмеген саны, ал бөлгіш – d_j құжатындағы барлық $n_{k,j}$ мүшелерінің бастапқы санының қосындысы. Сан неғұрлым көп болса, термин соғұрлым өзекті болады – бұл шара терминнің құжатқа қандай дәрежеде жататынын дәл анықтауға мүмкіндік береді.

Тағы бір өлшем екілік талдау болып табылатын кері құжат жиілігі (IDF). Ол өңделген құжаттардың санына және кемінде бір мерзімі $\{d : ti \in d\}$ болатын құжаттар санына негізделген. Бұл ақпаратты іздеуде ең пайдалы және кеңінен қолданылатын ұғымдардың бірі [118]. Ол мына формуламен өрнектеледі (3.2):

$$IDF_i = log \frac{n_d}{\{d:t_i \in d\}}, \quad (3.2)$$

Атрибут құжатта термин бар болса 1 мәнін, ал жоқ болса 0 мәнін қабылдайды. Үшінші өлшем – Термин жиілігі-кері құжат жиілігі (TF-IDF), ол қазіргі заманғы ақпаратты іздеу жүйелерінде ең көп қолданылатын терминдік салмақ схемаларының бірі болып табылады. Танымалдығына қарамастан, TF-IDF жиі эмпирикалық әдіс болып саналады, әсіресе ықтималдық тұрғысынан көптеген ықтимал вариациялары бар [119]. Анықтау бойынша, TFIDF - TF және IDF екі өлшемін көбейтетін метрика. TF-IDF сөздер, мәтіндік құжаттар және жеке санаттар арасындағы сәйкестікті анықтайды [120]. TF-IDF мына формуламен өрнектеледі (3.3):

$$(TF - IDF)_{i,j} = TF_{i,j} * IDF_i, \quad (3.3)$$

$TF_{i,j}$ мүшелерінің жиілігі IDF_i мәтініндегі жиілікке кері шамаға көбейтіледі. Бұл шама сонымен қатар жергілікті салмақ терминін анықтайды. Бұл бірнеше құжатта кездесетін сөздердің салмақтарымен салыстырғанда бір құжатта бірнеше рет пайда болатын сөздердің төменгі салмақтарын анықтау мақсатында қолданылады.

Кластерлеу – бақылаусыз оқыту әдісі, онда белгілер белгісіз және кейбір жағдайларда тіпті кездеспеуі де мүмкін [98]. Ол бір-біріне ұқсас және сонымен бірге басқа кластерлерге жататын объектілерге ұқсамайтын кластерлеу объектілерімен айналысады. Кластер әдетте кластердің орталығы болып табылатын центроидпен анықталады [121]. Деректерді кластерлеу бақыланбайтын үлгіні тануда қиын мәселе болып табылады, өйткені кластерлердің өлшемдері әртүрлі болуы мүмкін. Сонымен қатар, алгоритмдерді

іске қоспас бұрын кластерлердің санының жиілігін анықтау қажет. Кластерлеу – бұл қандай да бір ұқсастық өлшеміне негізделген табиғи топтар немесе кластерлерді анықтау процесі [122]. Сондықтан ұқсастық өлшемдері көптеген кластерлік алгоритмдердің негізгі элементтері болып табылады [18].

Секциялық кластерлеу алгоритмдері деректер жиынын кластерлеу сипатына тікелей әсер ететін жарамдылық өлшемі деп аталатын таңдалған критерийге сәйкес бірнеше кластерлерге бөледі [8,111]. Оларды оңтайландыру мәселелері ретінде қарастыруға болады, себебі олар таңдалған критерийлерді азайтуға тырысады. Бөлу алгоритмдерінің бірі K орташа әдісі болып табылады. Бұл әдіс бір кластердегі нүктелер арасындағы орташа квадраттық қашықтықты азайтуға бағытталған кең таралған кластерлеу әдісі [56]. Бұл мәселе NP қиын және оның шешімін жергілікті іздеуді қолдану арқылы Ллойд ұсынған [123]. K-орташаларының алгоритмі үлгілерді тең дисперсиялы n топқа бөлуге әрекет жасау арқылы деректерді кластерлейді, инерция немесе квадраттардың кластер ішілік қосындысы деп аталатын критерийді азайтады (3.4).

$$\sum_{i=0}^{n} min\ _{\mu \in C}\ (||x_i - \mu_i\ ||^2),\ (3.4)$$

мұндағы K-орталар алгоритмі n үлгі x жиынын дискретті C кластерлеріне бөледі, олардың әрқайсысы кластер ішіндегі үлгілердің орташа μ мәнімен анықталады. Бұл орташа мәндер әдетте кластер орталықтары немесе центроидтар деп аталады. К-ортасының жұмысы екі сатылы циклды қамтиды: бірінші қадамда әрбір үлгі өзіне жақын центроидқа тағайындалады, ал екінші қадамда алдыңғыға тағайындалған барлық үлгілердің орташа мәнін есептеу арқылы жаңа центроидтар анықталады. Содан кейін ескі және жаңа центроидтар арасындағы айырмашылық есептеледі және алгоритм бұл қадамдарды айырмашылық шекті мәннен төмен түскенше қайталайды. K-орталықтар әдісі центроидтарды бір-бірінен алшақ болатындай етіп инициализациялау арқылы жергілікті минимумға ықтимал жинақтау мәселесін шешеді.

K-орталар алгоритмі. Эксперименттік нәтижелер бойынша ұсынылған жүйе тақырыптары сәйкес келетін жұмыстарды баяндама тезистерінен алынған түйінді сөздерге сәйкес жіктей алатынын көрсетті.

Келесі ғылыми зерттеу жұмысында мәтіндік талдау негізінде ғылыми мақалаларды автоматты түрде кластерлеу бойынша жұмыс жүргізген. Осы мақсатта NLP әдістерін және K-means алгоритмін құжаттарды кластерлеуге арналған машиналық оқыту алгоритмі ретінде қолданған [124]. Ұсынылған тәсілдің бірінші қадамы PDF файлдары ретінде жиналған барлық ғылыми мақалаларды мәтіндік файлдарға түрлендіру болып табылады. Содан кейін мақалалардың мазмұны үш жинаққа бөлген. Бірінші жинақта әр мақаланың дерексіз бөлігі ғана қамтылса, екінші топтама мақалалардың кіріспе бөлімі болса, үшінші топтама мақалалардың толық мазмұны рефератсыз көрсетілген. Сонымен қатар, түйінді сөздер кластерлеу жұмысын бағалау үшін пайдаланылатын мақалалардан алынады.

3.2 Машиналық оқыту әдістерін ғадауат тілді сөздерді анықтауда қолдану (Decision Tree, Random Forest және Naïve Bayes, Logistic regression және K nearest neighbors)

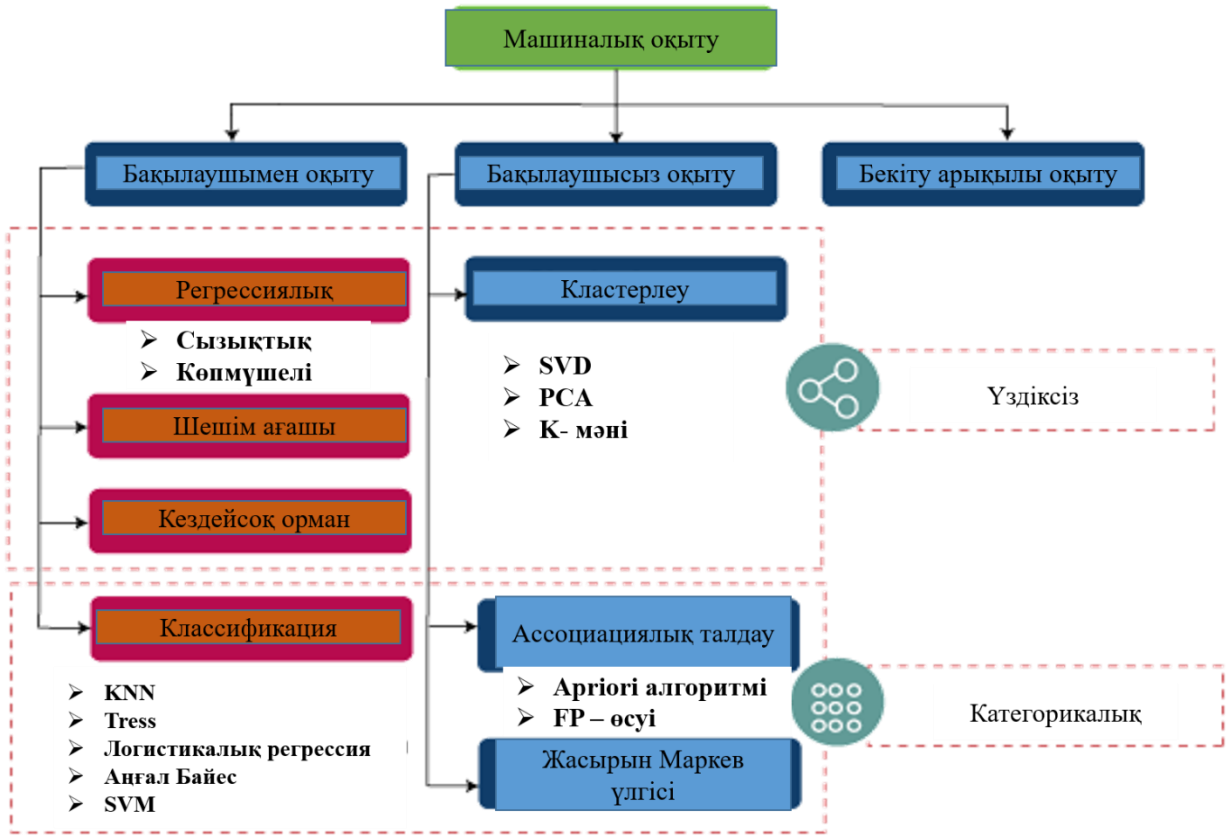
Machine Learning алгоритмдері - бұл деректер қорынан жасырын үлгілерді үйренетін, нәтижені болжайтын және қайта қайта тәжірибе жасау арқылы өнімділікті жақсартатын жасанды интеллект әдістерінің класы. Бұл әдістер математикалық статистика, ықтималдықтар теориясы, графикалық теория сонымен қатар сандық мәліметтер қорымен жұмыс істеуде қолданылады.

Мәліметтер қорымен жұмыс жасауға негізделген машиналық оқыту дойбы ойыны үшін алғаш рет 1950 жылы бағдарламаларды жасақтау кезінен бастап пайда бола бастады [125].

Компьютерлердің функционалдық қасиеттері артқан сайын есептеулерді құратын заңдылықтар мен болжамдар күрделене бастағандықтан, машиналық оқыту көмегімен іске астатын есеп пен мәселе қатары арта бастады.

Машиналық оқыту алгоритмдерінің 3.1-суретте көрсетілгендей үш түрі бар:

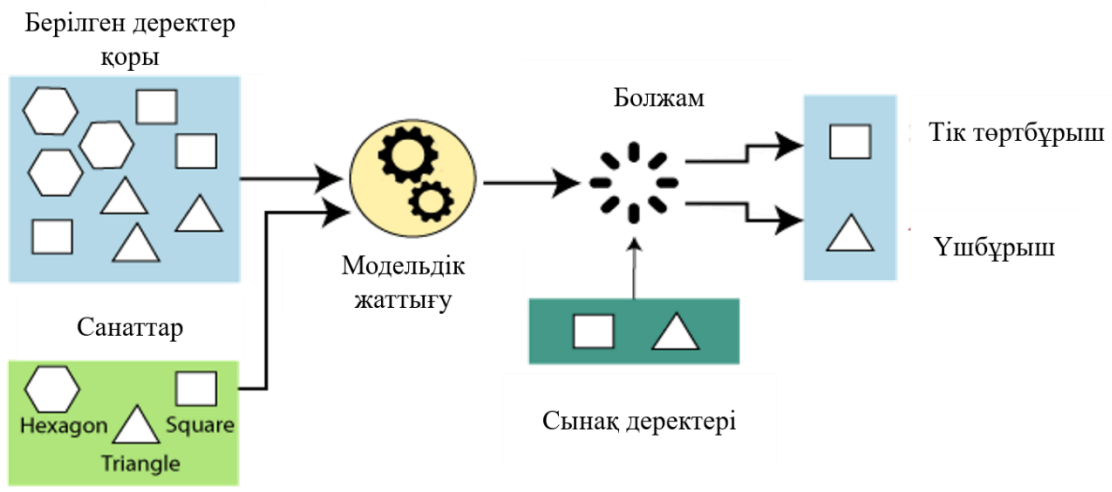
1. Бақылаушымен оқыту а
2. Бақылаушының көмегінсіз, бақылаушысыз оқыту
3. Бекіту арқылы оқыту



Сурет 3.1 – Машиналық оқыту түрлері
(Сурет машиналық оқыту алгоритмдері түрлерін ашып көрсету мақсатындағы ізденушінің авторлық өнімі болып табылады)

1. Бақылаушымен оқыту алгоритмі – машина үйрену үшін сыртқы бақылауды қажет ететін машиналық оқыту түрі. Бақылаушымен оқу үлгілері белгіленген деректер жиынтығы арқылы оқытылады. Оқыту және өңдеу аяқталғаннан кейін модель дұрыс нәтижені жете алатындығын тексеру үшін үлгі сынақ деректерін беру арқылы тексеріледі [48].

Бақылаушымен оқытудың мақсаты кіріс мәліметтерін шығыс деректерімен салыстыру болып табылады. Бақылаушымен оқыту бақылауға негізделген және ол мұғалімнің бақылауында оқушының бір нәрсені меңгеруі сияқты болып келеді (сурет 3.2).



Сурет 3.2 – Бақылаушымен оқыту
(Сурет бақылаушымен оқыту машиналық оқыту алгоритмін ашып көрсету мақсатындағы ізденушінің авторлық өнімі болып табылады)

Бақылаушымен оқытуды екі категорияға бөліп қарастыруға болады:

Классификациялық яғни жіктеу алгоритмі - оқыту деректері негізінде жаңа бақылаулар санатын анықтау үшін қолданылады. Классификацияда бағдарлама берілген деректер жиынтығынан немесе бақылаулардан үйренеді, содан кейін жаңа бақылауды бірнеше сыныптар немесе топтарға жіктейді. Мысалы, Иә немесе Жоқ, 0 немесе 1, Спам немесе Спам емес, мысық немесе ит, т.б.. Сыныптарды мақсат/белгі немесе санат деп атауға болады.

Регрессиялық талдау - бір немесе бірнеше тәуелсіз айнымалылары бар тәуелді (мақсатты) және тәуелсіз (болжаушы) айнымалылар арасындағы қатынасты модельдеуге арналған статистикалық әдіс яғни, регрессиялық талдау үздіксіз айнымалыны болжауға көмектеседі. Әлемде ауа-райы, сатуды болжау, маркетинг тенденциялары және т.б. сияқты болашақ болжамдарды қажет ететін әртүрлі жағдайлар бар, мұндай жағдайда адамзатқа болжамды дәлірек жасай алатын технология қажет. Мұндай жағдайда статистикалық әдіс болып табылатын және машиналық оқытуда және деректер қоырн өңдеуде қолданылатын регрессиялық талдау қажет.

Кейбір танымал бақылаушымен оқыту алгоритмдерінің түрлеріне тоқталар болсақ: Қарапайым сызықтық регрессия, шешім ағашы, логистикалық регрессия, KNN алгоритмін қарастыруға болады [8].

2. Бақылаушысыз оқыту - бұл машинаға деректерден үйрену үшін ешқандай сыртқы бақылауды қажет етпейтін машиналық оқыту түрі. Бақыланбайтын үлгілерді жіктелмеген немесе санатталмаған таңбаланбаған деректер жиынын пайдаланып оқытуға болады және алгоритм сол деректерге ешқандай бақылаушысыз әрекет етуі керек. Бақылаушысыз оқытуда модельде алдын ала анықталған нәтиже болмайды және ол деректердің үлкен көлемінен пайдалы түсініктерді табуға тырысады [94].

Бақыланбайтын оқытуды регрессия немесе жіктеу алгоритмдеріне тікелей қолдану мүмкін емес, өйткені бақылаудағы оқытудан айырмашылығы, бізде кіріс деректері бар, бірақ сәйкес шығыс деректері жоқ. Бақылаушысыз оқытудың мақсаты – деректер жиынының негізгі құрылымын табу, сол деректерді ұқсастықтары бойынша топтау және осы деректер жиынын дұрыс форматта көрсету. Оны класстерлеу және ассоциялық талдау деп екіге бөліп қарастыруға болады.

3. Бекіту арқылы оқыту өзіне іс - әрекеттер жасау арқылы қоршаған ортамен әрекеттеседі және кері байланыс арқылы үйренеді. Кері байланыс агентке сыйақы түрінде беріледі, мысалы, әрбір жақсы әрекеті үшін оң сый, ал әрбір жаман әрекеті үшін теріс сыйақы алады [118].

Жиі қолданыстағы машиналық оқыту алгоритмдер:

1. Сызықтық регрессиялық алгоритмдер.
2. Логистикалық регрессиялық алгоритмдер.
3. Шешім ағашы.
4. SVM.
5. Аңғал Байес.
6. KNN.
7. K – Means Clustering.
8. Кездойсақ орман.

Сызықтық регрессия - болжамды талдау үшін қолданылатын ең танымал және қарапайым машиналық оқыту алгоритмдерінің бірі. Мұнда болжамды талдау бір нәрсенің болжамын анықтайды, ал сызықтық регрессия жалақы, жас сияқты үздіксіз сандарға болжам жасайды.

$y= a_0+a*x+b$ мұнда, y= тәуелді айнымалы, x= тәуелсіз айнымалы

Сызықтық регрессияны екі негізгі категорияға бөлуге болады:

Қарапайым сызықтық регрессия деп аталатын бұл әдіс тәуелді айнымалының мәні туралы болжам жасау үшін жалғыз тәуелсіз айнымалыны пайдаланады [26].

Көп сызықты регрессия тәсілі тәуелді айнымалының мәні туралы болжам жасау үшін бірнеше тәуелсіз айнымалыларды пайдалануды қамтиды.

Логистикалық регрессия - бұл категориялық айнымалыларды немесе дискретті мәндерді болжау үшін қолданылатын бақыланатын оқыту алгоритмі. Оны машиналық оқытудағы жіктеу мәселелері үшін пайдалануға болады және

логистикалық регрессия алгоритмінің нәтижесі Иә немесе ЖОҚ, 0 немесе 1 деген санаттарды қолданады [99].

Шешім ағашының алгоритмі – бұл негізінен жіктеу мәселелерін шешу мақсатында қолданылатын бақыланатын оқыту алгоритмінің бір түрі. Дегенмен, ол регрессия қиындықтарын шешу үшін де пайдалануға қабілетті. Ол категориялық айнымалылармен де, үздіксіз айнымалылармен де жұмыс істей алады. Ол түйіндер мен бұтақтарды қамтитын ағаш тәрізді құрылымды көрсетеді және одан әрі бұтақтарда жапырақ түйініне дейін кеңейетін тамыр түйінінен басталады. Ішкі түйін деректер жиынының мүмкіндіктерін көрсету үшін пайдаланылады, тармақтар шешім ережелерін көрсетеді, ал жапырақ түйіндері мәселенің нәтижесін көрсетеді.

Векторлық машина алгоритмі немесе SVM бақыланатын оқыту алгоритмі болып табылады, оны жіктеу және регрессия мәселелері үшін де пайдалануға болады. Дегенмен, ол негізінен жіктеу мәселелері үшін қолданылады. SVM мақсаты деректер жиынын біршама сыныпқа бөлуде гипержазықтық немесе шешім шекарасын қалыптастыру.

Гипержазықтықты анықтауға көмектесетін деректер нүктелері тірек векторлары ретінде белгілі, сондықтан ол тірек векторлық машина алгоритмі деп аталады.

Аңғал Bayes классификаторы – объектінің ықтималдығының негізінде болжам жасау үшін қолданылатын бақыланатын оқыту алгоритмі. Алгоритм Bayes теоремасына негізделген және «айнымалылар бір-бірінен тәуелсіз» деген аңғал болжамға сүйенетіндіктен, аңғал Bayes деп аталады.

Bayes теоремасы шартты ықтималдыққа негізделген; бұл оқиғаның (B) орын алғанын ескерсек, (A) оқиғасының орын алу ықтималдығын білдіреді. Bayes теоремасының теңдеуі келесі түрде берілген (3.5):

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}, \quad (3.5)$$

K-жақын көршілер – бақылаушы арқылы оқыту алгоритмі, оны жіктеу және регрессия есептері үшін де қолдануға болады. Бұл алгоритм жаңа деректер нүктесі мен қол жетімді деректер нүктелері арасындағы ұқсастықтарды қабылдау арқылы жұмыс істейді. Осы ұқсастықтардың негізінде жаңа деректер нүктелері ең ұқсас санаттарға қойылады [126].

K - мәнді кластерлеу мәселелерін шешу үшін қолданылатын ең қарапайым бақылаушысыз оқыту алгоритмдерінің бірі болып табылады. Деректер жиындары ұқсастықтары мен айырмашылықтары негізінде K түрлі кластерлерге топтастырылған, яғни ортақ белгілерінің көпшілігі бар деректер жиындары басқа кластерлер арасында ортақтығы өте аз немесе мүлдем жоқ бір кластерде қалады [127].

Кездейсоқ орман алгоритмі (Random Forest) – машиналық оқытуда жіктеу және регрессия есептері үшін пайдаланылуы мүмкін бақылаушымен орындалатын оқыту алгоритмі. Бұл бірнеше классификаторларды біріктіру арқылы болжамды қамтамасыз ететін және модельдің өнімділігін жақсартатын

ансамбльді оқыту әдісі. Бұл дегеніміз шешім ағаштарының жиынтығы. Регрессиялық есепте олардың жауаптары орташаланады, классификациялық есепте шешім көпшілік дауыспен қабылданады. Барлық шешім ағаштары келесі схемаға сәйкес дербес орындалады:

Деректер қорын енгізу;

Ұсынылған деректер базасындағы үлгіні кездейсоқтыпен алу;

Содан соң, алгоритм үлгіге сүйене отырып, шешім ағашын жасап шығарады.

Ағаштың әрбір жапырағында n-нен көбірек деректер базасы кездеспейінше немесе ұсынылған межеге толғанша тоқтаусыз беріліп отырады. Содан соң әр шешім ағашы бойынша болжамдық нәтижесін алып отырады [128].

Бұл уақытта әр болжамдық нәтижелерге ұпай беру басталады: ең дұрыс нәтижелік ұсынымы таңдалып, сәйкесінше ағашты бөліп отырады және таңдаудың сандық көрсеткіші аяқталғанша осы алгоритм қайталанып отырады.

Соңында ең көп дауыс жинаған болжам нәтижесі таңдалады. Бұл болжамның соңғы нәтижесі болып табылады (3.7).

$$\text{a (x)} = \frac{1}{N}\sum_{i=1}^{N} b_i(x), \quad (3.7)$$

мұндағы: N – ағаштар саны;

i – бұтақтар есептеуіші;

b – шешуші ағаштар;

x – деректер негізінде жасалған үлгі.

Кездейсоқ орман (Random Forest)- бұл бірнеше пәндерде қолданылатын бақыланатын машиналық оқыту әдісі. Бұл тәсілді жұмыстардың кең ауқымында қолдануға болады; дегенмен, көп жағдайда ол классификация және регрессия мәселелерін шешу үшін пайдаланылады. Көлемі үлкен деректермен жұмыс жасауда өте дәл нәтижелерге қол жеткізу кезінде кездейсоқ орман алгоритмін қолдану тиімді.

3.3 Text classification Bag-of-Words to BERT (BagOfWords, Word2Vec, Continuous Bag-of-Words, Continuous Skip-gram)

Word2Vec (болжауға негізделген модель) Google-да Томас Миколов 2013 жылы [129] (машиналық оқыту саласында жұмыс істейтін чех компьютер ғылымдары ғалымы) енгізген. Бұл берілген корпустағы сөздер мен сөз тіркестерінің векторлық көріністерін үйрену үшін екі деңгейлі нейрондық желіні пайдаланатын семантикалық оқыту жүйесі [67]. Word2Vec үлгісі мәтінді кіріс ретінде қабылдайды және корпустың сөздік қорын сипаттайтын мүмкіндік векторларын шығарады. Бұл әдістің жиілікке негізделген үлгілерден айырмашылығы көршілес сөздерге негізделген сөздерді болжау үшін болжамды талдау тәсілін қолданады. Word2Vec үлгісінің екі нұсқасы бар:

Үздіксіз сөздер қапшығы (CBOW): Бұл модель көрші сөздерге негізделген негізгі сөздің ықтималдығын барынша арттыруды үйренеді (ағымдағы сөздің

айналасындағы сөздер терезесі). Осылайша, модель контекстке қарап, сөздерді болжауға үйренеді.

Сондықтан ол сөзді барлық нақты контексте қолдануға болатынын жалпылауды үйренеді. Оқытудың бұл түрі сирек кездесетін сөздерге аз көңіл бөледі, себебі мақсаты ұсынылған кілттік сөздерді табуға негізделеді.

Мысалы, «бүгін расында саяхаттауға ... күн болып тұр», бұл жағдайда модель сөз ықтималдығына байланысты «тамаша» немесе «ыңғайлы» деп болжауға тырысады. Бұл модель сирек кездесетін сөздерді үйрене алмайды.

Skip-Gram (SG): Бұл модель ағымдағы сөзге негізделген көрші сөздерді (контекст) болжауды үйренеді. Мәселен, мысалда модель «тамаша» сөзін қабылдайды және «бүгін расында саяхаттауға ... күн болып тұр» сөйлемі үшін оның контекстін болжайды. Бұл модель сирек және жиі кездесетін сөздерге бірдей назар аударады, мұнда, «әдемі», «ғажап» сияқты сөздерге және оның көршілес сөз тіркестерін жаңа үлгі ретінде қарастырады. Сондықтан бұл үлгі әрбір контекст үшін екі немесе одан көп үлгілерді қажет етеді.

Миколовтың айтуынша [130], SG шағын оқу корпусымен жақсы жұмыс істейді, тіпті сирек кездесетін сөздерді тиімді меңгереді және анықтауда жоғары тиімділік көрсетеді. Сонымен, сөзді ендіруді үйрету үшін осы Skip-Gram үлгісін таңдаймыз, өйткені бізде оқу деректері, яғни мәліметтер қоры шектеулі.

Алгоритм BERT

Табиғи тілді өңдеу мәселелерін шешу үшін әртүрлі нейрондық желілер модельдері жиі қолданылады, солардың ішінде перспективті алгоритмдердің бірі Google корпорациясының BERT нейрондық желісі болып табылады [131]. Ол таңбалардың арнайы тізбегі енгізілген сөздердің векторлық бейнелерін яғни таңбалауыштарды (токендеуді) пайдаланады. BERT Transformer архитектурасына негізделген. Деректердің үлкен көлеміне үйренген модель мәтіндегі сөз тіркестері мен сөйлемдер арасындағы байланысты түсінуге қабілетті болып келеді. Содан кейін оны қажетті тапсырмаға арналған таңбаланған деректер бойынша қосымша оқытуға болады. Жіктеу үшін қабаттарды қосу арқылы мәтіндерді автоматты түрде санаттаудың тиімді құралын алуға болады [132].

Ұсынылып отырған ғылыми жұмыста transformers деп аталатын кітапханасны Python бағдарламалау тілінде қолданған. Бұл жерде түрлі есептерге, яғни мәтінді, текстті жіктеп қарастыруға ыңғалы бұрыннан қолданылып келе жатқан үлгілер қатары кездеседі. Салыстырмалы тестілеуде түрлі өлшемді деректер базасы ретінде қарастырылған біршама модель таңдап алынған. Олардың негізгі сипаттамаларына тоқталып өтер болсақ:

– bert-base-multilingual-cased (bert – ке негізделген -көп тілді-регистр). Толық BERT нейрондық желісі Wikipedia -ғы мәтіндердің үлкен жинағында 104 тілді қамтып қолданылады.

Деректер базасын құру және BERT желісінде оқытуды жүргізу

Бірінші кезең деректер базасын құруда сыныптар белгісін өзгерістерге енгізу ретінде қарастырылады. Алғашқы деректер базасында сызықтық

мәтіндермен келтірілген, сол себепт сандық символдармен алмастыру қажеттіліг туындайды.

Екінші кезеңде деректер базасын оқыта отырып, оларды валидацияға бөліп тастау. Валидация деректер базасының өлшемі ретінде оқыту үшін деректер қорының 10 пайызын қарастырды. Модельдің бірқатар санаттарындағы деректер көлемінің аздығына байланысты валидация жиынының көлемінің ұлғаюы анықтау тиімділігінің төмендеуіне әкеледі.

Үшінші кезең – хабарламалардың векторлық кескіндерін құру және таңбалауыштарды қосу. Жоғарыда айтылғандай, BERT арнайы таңбалар тізбегі енгізілген мәтіннің векторлық көрінісін пайдаланады. Деректерді дайындау үшін Transformer кітапханасының BertTokenizer класы пайдаланылды.

Төртінші кезең – нейрондық желіні оқытуда қолданылатын оңтайландыру алгоритмін таңдау және сәйкес объектіні құру.

Бесінші қадам – нәтижеге қол жеткізу үшін random seed – ді орнату. Егер бұл кезең орындалмаса, әр жолы нәтиже кездейсоқ болады, өйткені терең оқыту алгоритмдері стохастикалық сипатта болады.

Алтыншы кезең - дайындалған үлгіні жүктеу және оны қосымша оқыту. Модель 4 кезеңде оқытылады: деректер жинағы шағын болғандықтан, көп кезеңнен тұратын модельдермен қайта дайындалуы мүмкін, ал аз болса оқыту динамикасы көрібей қалу қаупі бар. Әр кезеңнен кейін модель әртүрлі кезеңдерде алынған нәтижелерді талдауға мүмкіндік беретіндей сақталады [132].

Бұл тарауда табиғи тілдерді өңдеу, яғни стемминг арқылы сөздердің түбірін табу, сөздердің аймақтық өлшемдері, кластерлеу, к- орталар алгоритмдері баяндалған. Сонымен қатар, машиналық оқыту алгоритмдерін текстті танудағы қызметі зерттелген. Терең оқыту алгоритмдерін мәтіндерді анықтаудаға ерекшеліктерін ашып көрсеткен.

Ары қарай, біздің жұмысымыз қазақстандық медиа кеңістігіндегі онлайн контентте кездесетін ғадауат тілді пікірлерді зерттеуді мақсат етті, бірақ бұл этапта бізге кедергілер кездесе бастады, деректер қоры мен зерттеу әдістерінің жоқтығынан біз бәрін өзіміз жасауымыз керек деген қорытындыға келдік. Осыған байланысты жоспар құрылды, жоспар келесі міндеттерді қамтыды:

1) Қазақ тілі үшін ғадауат тілді сөздері бар пікірлерден тұратын деректер қорын жинақтау.

2) Деректер жиынын қолмен класстарға бөлу.

3) Қазақ тіліндегі ғадауат тілді сөздері бар пікірлерге арналған тілдік корпус жинақтау.

4) Жинақталған деректер жиынында машиналық және терең оқыту талдауын жүргізу.

4 ВЕБ – КОНТЕНТТЕ ҒАДАУАТ ТІЛДІ СӨЗДЕРДІ АНЫҚТАУҒА АРНАЛҒАН СЕМАНТИКАЛЫҚ МОДЕЛЬДІ ЗЕРТТЕУ ЖӘНЕ ҚҰРУ

Бұл тарауда Natural Language Processing (NLP) және Machine Learning (ML) әдістері арқылы онлайн контенттегі ғадауат тілді сөздерін анықтаудың жүйелі әдістемесі қолданылады. Қазақ тілді ғадауат сөздерден тұратын деректер көздерін анықтауға және жинауға бағытталған алғашқы кезеңнен басталды.

Интернет-ресурстардың әртүрлі жинағы талданды, әсіресе ғадауат тілді пікірлер жиі кездесетін қазақ тілді онлайн платформалар, соның ішінде әлеуметтік медиа, форумдар және жаңалықтар веб-сайттары басты назарға алынды. Бұл тараудың мақсаты ғадауат тілді сөздердің формаларын, айтылу, жазылу ерекшеліктерін ескере отырып және тілдік ерекшеліктерді басшылыққа ала отырып қазақ тілі үшін деректер корпусын құру болды.

Қажетті ғадауат тілді сөздерден тұратын деректер корпусы жинақталғаннан кейін мәліметтер қорын машиналық және терең оқыту алгоритмдеріне оқытуға әзірлеуде жүргізілген зерттеу жұмыстары мен нәтижелері талқыланылады.

Нақты және жақсы нәтижеге қол жеткізу үшін белгіленген стандарттар жиынтығына сәйкес деректер корпусының ерекшеліктері мен сапасына қарай қолмен класстарға жіктелу процессі баяндалады. Таңбалау процедурасы мәтін мазмұнының белгіленген жіктеулерге, соның ішінде ғадауат тілді сөздер, д бейтарап мәтіндік деректер деп жіктеліп отырды.

Модельдеу кезеңінде зейін механизмімен BERT (Трансформаторлардан екі бағытты кодтаушы өкілдіктер) артықшылықтарын біріктіретін гибридті терең оқыту архитектурасы пайдаланылды. BERT мәтіндегі терең контекстік көріністер мен семантикалық байланыстарды түсіну қабілеті үшін таңдалды, бұл әдеттегі NLP үлгілерінен әлдеқайда жоғары қабілеттілікке ие екендігі сипатталған. Зейін механизмі модельдің ғадауат тілді сөздерді немес сөз тіркесінің ішіндегі маңызды компоненттерге назар аудару қабілетін арттыру үшін қосылды, осылайша түсіндірмелілік пен дәлдікті арттырады.

Модельді оқытудан бұрын деректер жинағы кіші әріптерді, тоқтау сөздерді жоюды, тыныс белгілерін басқаруды, эмодзилерді өңдеуді және таңбалауды қамтитын кешенді алдын ала өңдеуден өтті. Семантикалық маңызды таңбалауыштарды, соның ішінде хэштегтерді, пайдаланушы ескертулерін және сандық белгілерді сақтауға ерекше назар аударылды.

Соңғы гибридті модель оқу жылдамдығы, топтама өлшемі және өнімділікті арттыру үшін реттелген гиперпараметрлері бар GPU жеделдетілген есептеу ресурстарын пайдалана отырып, аннотацияланған деректер жиынында оқытылды. Модельдің тиімділігі кәдімгі жіктеу критерийлері, соның ішінде recall, accuracy, precision және F1 ұпайы арқылы бағаланды. Модельдің жаңа деректерге жалпылануын қамтамасыз ету үшін кросс-валидация әдістері қолданылды.

Сонымен қатар, ұсынылған әдіс ғадауат тілді сөздерді айтарлықтай дәлдікпен анықтау үшін мұқият жинақталып, тазартылған деректер жиынтығын және жетілдірілген NLP архитектурасын пайдаланады. BERT және зейін механизмін біріктіру жүйеге тілдің синтаксистік және семантикалық атрибуттарын түсінуге мүмкіндік береді, бұл оны мазмұнды модерациялау жүйелерінде немесе цифрлық қауіпсіздік құрылымдарында нақты уақытта іске асыруға сәйкес етеді.

4.1 Мәліметтер қорын жинақтауға дайындық кезеңі

Онлайн контенттегі бейәдеп немесе ғадауат тілді мәтіндерді анықтауда семантикалық талдау модельдерін құру мақсатында ең бірінші көлемі үлкен мәліметтер қоры корпусы қажет болады. Бұл корпустағы деректер қоры машиналық оқыту алгоритмдерін оқыту және тестілеу мақсаттарында қолданылады. Алғашқы кезеңде қол жетімді ресурстардағы қазақ тілді ғадауат пікірлер іздестірілді. Яғни, қазақ тілді желі қолданушылары жиі қолданатын «VK», «Youtube», Facebook», «Instagram», «Twitter» әлеуметтік желілер пікірлері арасынан табылды [49,49,50,133,134].

- Tik –tok – жастар мен жасөспірімдер аккаунттарында бірін – бірі қаралау, кемсіту жиі кездеседі;
- Instagram – блогерлер, әнші, актер және актрисалардың парақшаларына жағымсыз пікір қалдыру бойынша көш бастап тұр;
- Facebook -саяси парақшаларда, журналистердің және қоғам қайраткелрелінің жеке парақшаларындағы халықтан келген пікірлерарасында халыққа, ұлтқа бөліп тіл тигізу жиі ұшырасады;
- Tengrinews, nur.kz, zakon.kz, 24kz, habar.kz сияқты ақпараттық танымдық сайттар мен телевизияның әлеуметтік парақшаларында жарияланған жаңалықтарға ызасын, келіспеушілігін ғадауат тілді сөздерді пайдалана отырып жеткізу;
- Youtube – әлеуметтік желісіндегі кәмлеттік жастан асқан пайдалнушыларға арналған парақшаларындағы жиі кездесетін пікірлерден менсінбеу, кекету және қорлау мақсатында жазылған пікірлерді жиі кездестіруге болады.

Ғылыми зерттеу жұмысында мәліметтер корпусына ғадауат тілді мәтіндермен бірге бейтарап мазмұнды деректерді жинақтау жоспарланды. Ғадауат тілді сөздерден тұратын мәліметтер қорына адамды жынысына, нәсіліне, жасына, әлеуметтік жағдайына, бет –бейнесіне, ақыл ойының дамуы бойынша және т.б. қамтитын мәтіндер тізбектері таңдалып алынды. Бейтарап санаттағы мәтіндерге қорлау, жек көрініш немесе кемсіту үлгілері кездеспейтін қазақ тілді аудиториядағы онлайн контенттегі пікірлер топтамасы таңдалды.

Бірінші кезекте қазақ тілінде жазылған жазбалар мен бейнежазаблар жазылған аккаунттар таңдалып алынды және пікірлерді жинақтау мақсатында парсинг жасау қажет болды.

Іздеу нәтижесінде қазақ тілді веб контенттегі 100 ден астам желі парақшалары табылды. «Facebook», «Instagram» сияқты әлеуметтік желілерге

парсинг жасау біршама қиындық тудырып отырды. Себебі әлеуметтік желілердің қорғаныс деңгейі жоғары болғандықтан, парсинг жасау барысында автоматты түрде бұғатталып отырды. Ал, кей желі парақшаларында ғадауат тілді пікірлер кездеспеді.

Мәліметтерді жинақтау.

Жоғары нәтижеге қол жеткізу үшін, мәліметтер көздерінен ақпаратқа қол жеткізуде кіріктірілген әдістерді қолднау қажеттілігі туындайды.

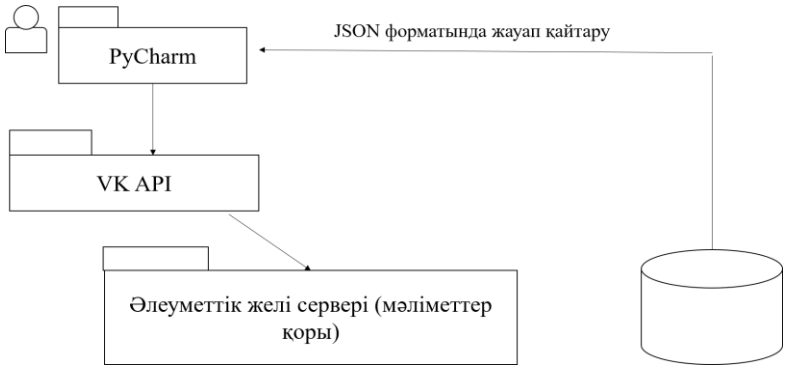
Бағдарлмалық жасақтаманы үш модульге бөліп көрестуге болады:

1) мәліметтерді жинақтау модулі – ашық көздер арқылы ақпаратқа қол жеткізуге және оны ары қарай өңдеуден өткізуге жауап береді; «VK» әлеуметтік желісіндегі деректерге қол жеткізуде Python бағдарламалау тілі қолданылды. Ресми түрде VK API қолданыла отырып, қазақ тіліндегі аккаунттар ішінара талданды, сондай – ақ деректер Solr деректерқорында, Youtube желіснен алынған деректер csv форматында, Facebook», «Instagram әлеуметтік желілерінен алынған деректер қоры xlsx форматында сақталды.

2) кілттік сөздерді іздеу модулі - үлкен көлемді деректер қоры ішінен кілттік сөздерді іздеуде қызмет атқарады.

3) құжаттарды сараптау модулі – ақпараттың тазалығына жауап береді. Құжаттарды фильтрден өткізу дәрежесі бойынша Bidirectional Long Short Term Memory терең оқыту алгоритмі қолданылады.

Деректер қорына қол жеткізу үшін ТМД кеңінен таралған VK әлеуметтік желісі пайдаланылды. Мәліметтер қорын жинау үшін Python 3.7 қолданылды. Суретте деректер жинау процессі берілген (Сурет 4.1).



Сурет 4.1 – Мәліметтер қорын жинау схемасы
(Сурет деректер қорына қол жеткізу схемасында беріле отырып, ізденушінің авторлық өнімі болып табылады)

Pycharm Community Edition бағдарламасы әлеуметтік желідегі API – мен байланыстыру арықылы сұрау кітапханасы жасалды [135]. Деректер қорына қол жеткізу үшін VK, Youtube әлеуметтік желілерінде серверлерге https сұранысын пайдаланып, қарастырып отырған әлеуметтік желілердің деректер қорынан қажет деп танылған ақпараттарды алу үшін интерфейс құрастырылды.

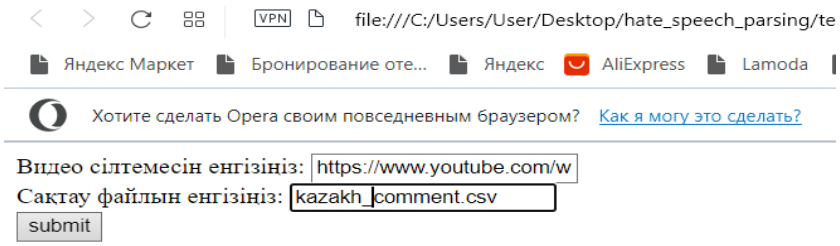
Әдістер – деректер қоры операцияларына сәйкес орналасқан шартты командалар тізбегі. Жүйеде кездесетін әдістер бөлімшелерге бөлініп

қарастырылған. GET параметрлері үшін кіріс дрекетерді http сұрауларынан кейін ұсыну қажет. Егерде, сұранысты беру іс – әркеті жақсы нәтиже көрсетсе, сұралған мәліметтер қоры JSON форматы негізінде қайтарып береді. Мәліметтерді талдау мақсатында «pandas», «spacy», «plotly», «bokeh», «numpy», «googletrans» пакеті бар Python бағдарламалық тіліндегі басты кітапхана мақсатында пайдаланылады. Керекті мәліметтерді іздеу үшін ғадауат тілді сөздер қатарына енетін сөздер қатары анықталды. Мәліметтер қоры жинақталғанымен оның дұрыс таңбаланғанына көз жеткізу мақсатында барлық пікірлер қолмен тексерілді. Мысалы, «сен шошқа, сүмелек» текті мәтіндер тізбегі -1 санымен таңбаланса, бейтарап мазмұндас пікірлер «шошқаның жем – шөбінің бағасы жыл сайын өсуде» сияқты сөйлемдер – 0 санымен таңбаланды [136].

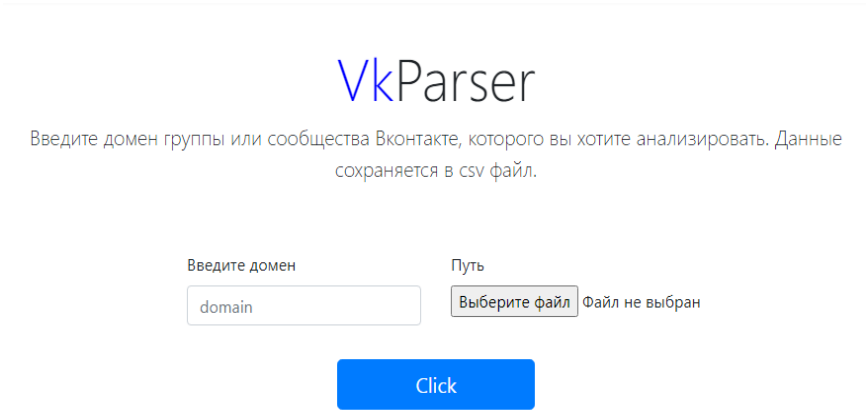
Ескере кететін жайт, қазақ тілді мәліметтер қоры болғандықтан үш ерікті аннотатор яғни, қазақ тілі маманы Н. Алимбеков, педагог – психолог Д. Әмір және ғылыми зерттеу жүргізуші А.Тоқтарова жинақталған мәліметтер қорын екі классқа бөліп отырды.

Деректер қорын жинау үшін парсер құрастыру.

Ұсынылып отырған зерттеу үшін үлкен көлемдегі деректер қоры қажет. Қазақ тілі үшін дайын ғадауат тілді сөздерден құралған мәліметтер корпусы болмағандықтан, VK, Youtube, Facebook, Telegram, Instagram, сонымен қатар елдегі жаңалықтар порталдарынан ашық ақпарат көздерін пайдалана отырып, мәліметтер қорын жинақтайтын парсер жазылды. Парсердің (шолғыш) нәтижесін көрсету үшін веб – қосымша құрылды. 4.2-4.3 суреттерде әртүрлі дереккөздерден мәліметтер жинақтауға арналған парсер бағдарлмасы ұсынылған.



Сурет 4.2 – Мәліметтер қорын жинақтауда құрастырылған Youtube үшін парсер



Сурет 4.3 – Мәліметтер қорын жинақтауда құрастырылған VK үшін парсер

Қорытындылай келе, жасақталған парсер көмегімен ғадауат тілді бағыттағы деректер қоры жинақталды. Модельдері одан ары оқыту мен тестілеу мақсатында деректер қоры корпусы жасақталды, «ғадауат тілді сөздер» және «бейтарап» бағыты бойынша екі класстан тұрады.

4.2 Веб – ресурстардағы ғадауат тілді деректерді анықтауда қажетті корпус құру арқылы семантикалық моделін құру

Деректер қоры корпусына енгізілген ғадауат тілді қолданушы пікірлерінің саны 20 000 – ға жуық, ал бейтарап пікірлер саны 20 000- ға жуық сөйлем қатарын құрайды.

Жинақталған корпус .csv форматына келтіріліп құжат ретінде сақталды. Ғадауат тілді корпусты құжатты 4 бағанға бөліп қарастырдық: пікірлердің рет бойынша орналасқан реттік нөмері, пікірдің өзі, пікірдің таза нұсқасы (стоп сөздерден, тыныс белгілерден, эмоджилардан және т.б тазалануы) және класстар «1/0» атрибуттары. Сәйкесінше «1» атрибуты – ғадауат тілді пікірлер кездесетін пікірлер тізімі, «0» атрибуты – бейтарап пікірлер жиыны (4.4-сурет).

1267	Бұлардың бәрін жинап псих болницаға апару керек	1
1268	Осы топастардың әке шешесі бар ма?	1
1269	Топастықтың да шегі болу керек. Не сандырақ? Жап-жаңв машина болса, гарантийный срогы болуы керек қой.	1
1270	осындай долбоебтарды кайдан тауып аласың	1
1271	Ей сақалды. Казёл. Тех. Ақау. Салғырттықтан болады. Сенің дұға оқығаныңнан болмайды.. Халықты тупойға санамам. Ия рас ҚҰДАЙҒА СЕНЕМІЗ! сенің ертегің жараспайды. Нақты дәлел видеомен жіберіңіз. До и после.	1
1272	Жалпы қазақтар рухани азып, адам болудан қалып барады екен. Масқара. Режимді өзгертпесе бәрі мал боп бітер.	1
1273	Ол жақтағы ӨЗБЕК ТУҒАНДАР ..МІҢКЕСІЗ МӘҢГҮРТ..САН МЛНДАҒАН ҚАЗАҚ МАЗАҚТАН АРТЫҚ СӨЙЛЕЙДІ...ҚАЗАҚ ТІЛІН.	1
1274	Ініне тығылған крысаға қарап отырады екен ғой, қалған жалбақайлар)))	1
1275	Сгиннің баласы, кобитті қайдан тауып алып жүр	1
1276	Ёмаё, мынау сразім қабыл боп кететін қарғыс қой 🤦	1
1277	қатқан б*қты жібітіп отырған отырыстарын😐	1
1278	Өмір бойы , осылай отырсын. Бүкіл коорупционерлерді осылай мыжылау керек. Атым жазасынан да күшті..	1
1279	қантарда баланды атса , басқаша сөйлер едін, соқыр сымақтар сендердейлердің кесірінен қазақ осындай халге жетті.	1
1280	Дәлел жоқ.бір қазақ қиналса бәрі мәз болатын недеген надандық тасжүректік.	1
1281	Жаза өтесе жақсығой,жазасын өтемей қашып жүргендер қаншама,өмір тілейік ағайын	1
1282	Осындай ақпаратты қайдан аласыңдар?	1
1283	Өтірік. Нұрсұлтан шал тұрғанда олай бола қоймас. Жай алдау мұның бәрі.Болат нашарбаевта Өтірік инвалид болып қалған жоқ па еді.	1
1284	Дұрыс емес өздеріңіздің сауатсыздығызды көрсетіп тұрсыздар ! Оттайбермеңдерші ? Сауатсыз мұғалімсіздер	1
1285	Қазір сауатсыз,жемқор басшылар көбейіп кетті.Мықты басшы болмаса мектеп оңалмайды.	1

Сурет 4.4 - Ғадауат тілді сипаттағы хабаралама мысалы (шикі деректер қоры)
(Сурет автордың мәліметтер қорының скрині)

Қазақ тіліндегі ғадауат тілді сөздерді қамтитын пікірлерді сыныптарға бөлуде сөз корпусына тән ерекшелігі бар екенін байқауға болады [36]:

1) қазақ тілі әріптерің кириллица әріптерімен ауыстыра отырып жазу (мысалы, «сұм_рай» - «сум_рай», «өл_п қ_л» - «олип к_л» және «оңб_ған» - «онб_ган»),

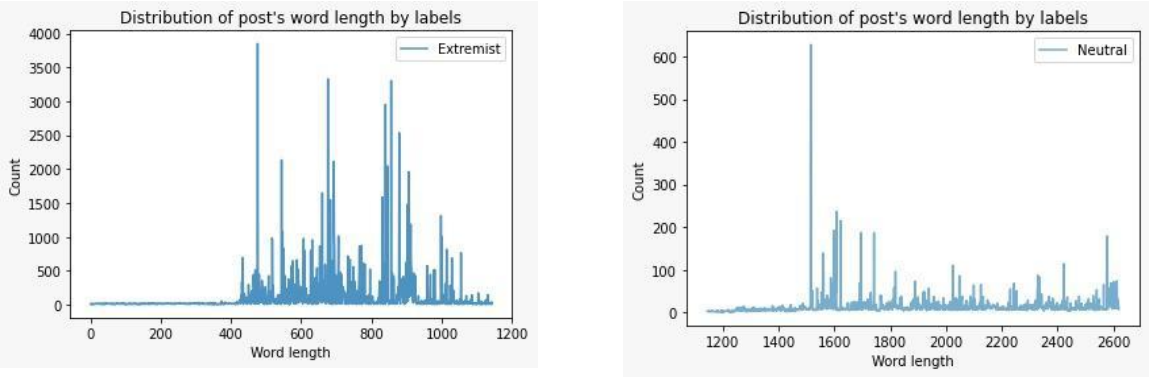
2) қазақша пікір жазып жатқанда әріптерді латын әріптерімен алмастыра қолдану (мысалы, «ш_шқа» - «w_wka», «қ_ншық» - «k_nwyk» және «х_йуан» - «x_iyan»),

3) теңеу сөздерді пайдалана отырып немесе метафраны қолданып кемсіту, жеке басының қадір – қасиетіне нұқсан келтіру (мысалы, «м_ктабас», «ор_скұл» және «од_қбас»),

4) тұрғылықты өлкесіне байланысты ғадауат тілді кемсіту сөздерін қолдана отырып пікір қалдыру (мысалы, «юж_ндар», «т_хас», «ор_стар», «к_ртопбас»),

5) орфография қателерін түзетпей жазу немесе әріпті символмен алмастыру (мысалы, «к.т __екенсің», «м@__л»),

6) орыс тілі сөздігінен келген сөзді қосымша жұрнақ жалғау арқылы «қазақ тілінің сөзі жасау» (мысалы, «д_лбандар», «с_касың», «ч_рттар». Сурет 4.5.-те мәліметтер корпусындағы ғадауат тілді сөздердің ұзындығы белгіленген.



а б

а - ғадауат тілді пікірлерді үлестірім графигі; б - бейтарап мәліметтердің үлестірім графигі

Сурет 4.5 – мәліметтер корпусындағы ғадауат тілді пікірлерді сөз ұзындығы белгісі бойынша бөлу

Сөздер бұлты. Мәліметтер қорын визуалды етіп көрсету мақсатында сурет 4-6 – да көрсетілгендей сөз бұлттары қолданылады.

а б

а - ғадауат тілді сөздер бұлты; б - бейтарап сөздер бұлты

Сурет 4.6 - ғадауат тілді және бейтарап пікірлер корпусын визуализациялау

Мәтіндерді талдау. Мәтіндік текстке талдау жүргізу 3 нақты форматты қамтиды[5]: *классификациялау*, *мазмұнды рефлексия* және *көңіл-күйлік талдау*.

Барша мәтіндік жіктеуші есептерін 2 – ге бөліп қарастыруға болады:

а) екілік классификация пайдаланушыға ұсынылаған мәтіннің негізгі өзегін табуға көмектеседі;

ә) екіден көп класты классификация мәтіндік тексттің тақырыбын анықтауға және бірнеше санаттық сыныптардың біріне тағайындалуына септігін тигізеді [8].

Мәтін мазмұнының рефлексиясы (мәтінді қорытындылау) келесідей жұмыс істейді: NLP жүйесі үлкен мәтінді енгізеді және үлкен мәтіннің мазмұнын көрсететін кішірек мәтінді шығарады [26].

Жазылу тілін тану және оны синтездеу

Тілді тану дегеніміз – сөйлеп жатқан сөзді мәтін түрінде цифрлық мәліметке ауыстыру процессі ретінде қарастырамыз. Сөйлеу синтезі басылған мәтіннен сөйлеу сигналын құра отырып, қарама-қарсы бағытта жұмыс істейді.

Сөйлеуді синтездеу және тану дауыстық көмекшілер, IVR жүйелері және смарт үйлер сияқты көптеген салаларда қолданылады.

Диалогтық жүйелерді дамыту

Диалогтық жүйелер ретінде болады:

ақылды көмекшілер (Yandex.Alice, Siri, Alexa);

чат боттары – диалог сценарийлеріне сәйкес келетін мәтіндік жүйелер (Messenger боты, Facebook Messenger);

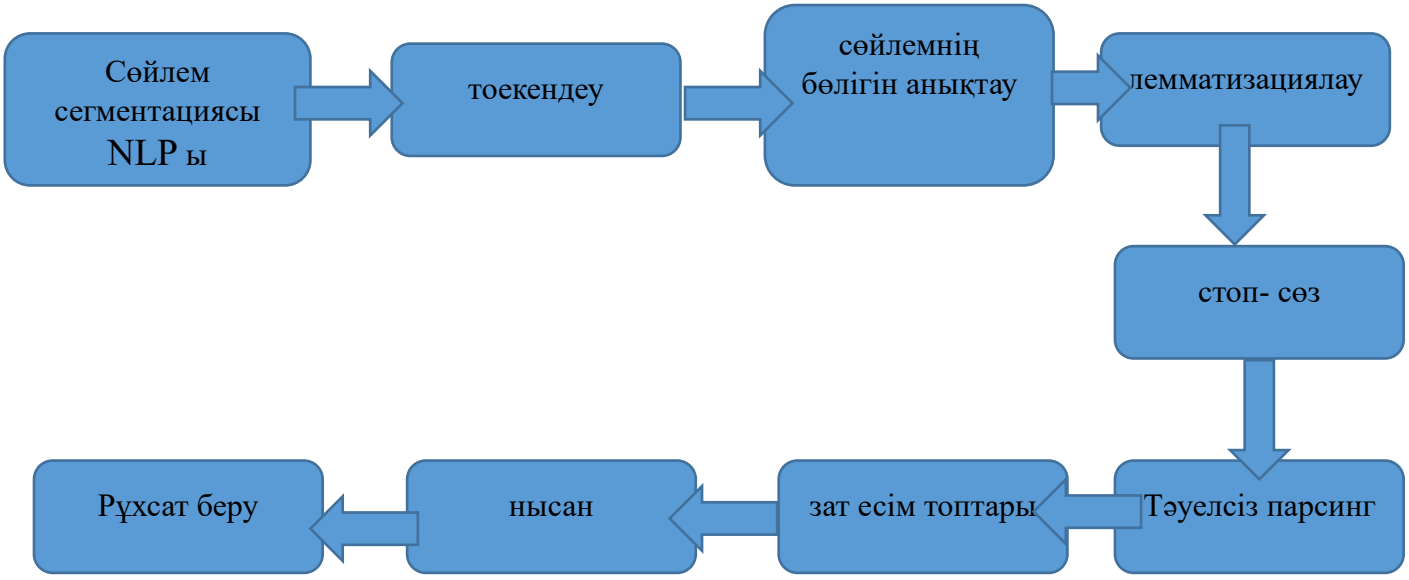
QA жүйелері.

Олардың барлығы NLP құралдарына сүйенеді: сөйлеуді тану, мағынаны, контекстті бөлектеу, ниетті анықтау, содан кейін жоғарыда айтылғандардың негізінде диалог құру (дұрысы сөйлеу синтезі арқылы).

Табиғи тілді өңдеу

Тапсырмалардың көпшілігі архитектураны мұқият таңдауды, сондай-ақ мүмкіндіктерді қолмен жинауды және өңдеуді талап етті. Алайда қазіргі таңда, нейрондық желілер классикалық модельдермен салыстырғанда дәлірек нәтижелер бере бастады және NLP мәселелерін шешудің жалпы тәсілін қалыптастырды.

NLP конвейері. Сурет 4.6.1 -де сөйлемдерді өңдеуге арналған NLP қадамдарын сипаттайтын блок-схема көрсетілген.



Сурет 4.6.1 - NLP конвейері
(Сурет NLP конвейерінің схемасын баяндау мақсатында сызылған автордың жеке өнімі)

Келесі қадамдарға тоқталар болсақ:

1. Алдымен мәтінді сөйлемдерге бөлу керек. Бұл табиғи тілді өңдеудегі (NLP) ең іргелі кезең және кез келген кейінгі кезеңдерді талдау үшін қажет.

2. Токенизация: Сөйлемдерді ажырату үшін сөздер немесе лексемалар қолданылады.

3. Бізде әр лексемаға грамматикалық қызмет (зат есім, етістік, сын есім, т.б.) тағайындайтын сөз бөлігін белгілеу бар.

4. Жүйелілікті өңдеу үшін сөздер лемматизацияланады немесе олардың негізгі немесе сөздік формасына қысқартылады.

5. Келесі қадамда– «және», «те» және т.б. сияқты мағыналы мазмұн қоспайтын жалпы терминдерді жою.

6. Сөз тіркесінің құрылымын және оның тәуелдіктерін түсіну үшін тәуелдіктерді талдау сөздер арасындағы байланыстарды зерттейді.

7. Зат есімдік тіркестерді анықтау: Бұл қадам мақсатты тексеру үшін зат есімдердің топтарын немесе сөз тіркестерін анықтауды қамтиды.

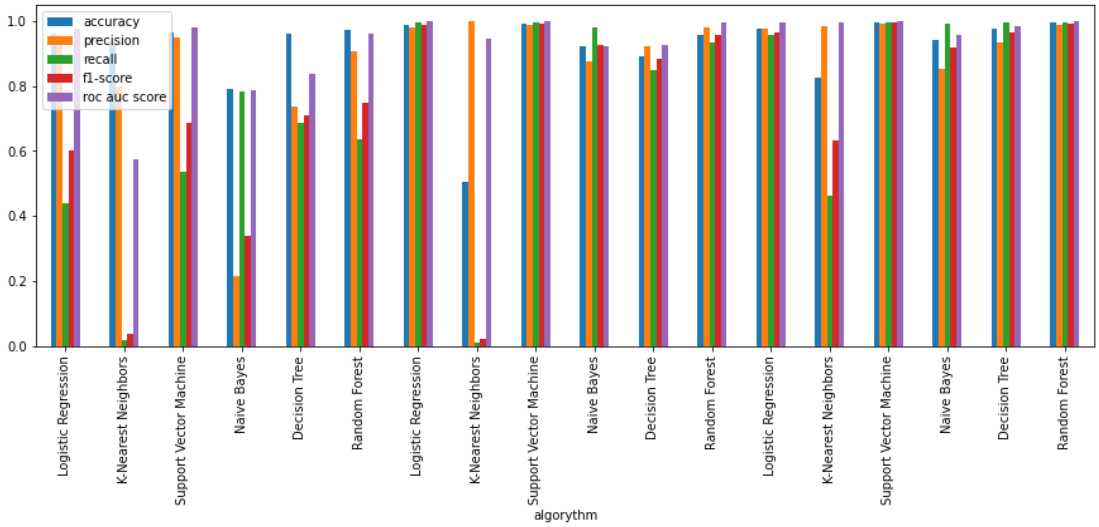
8. Объектіні тану: Мәтіннен белгілер мен нысандарды алу.

9. Авторизация: процестің соңында рұқсаттармен айналысатын кезең бар сияқты. Бұл талдау нәтижелері белгілі бір қолданбаға қолданылып жатқанын немесе рұқсат берілгенін білдіруі мүмкін [137].

4.3 Мәтіндегі ғадауат тілді анықтау үшін машиналық оқытуды қолдану

1. Функцияны өңдеу

Бұл бөлімде біз әртүрлі мүмкіндіктер тіркесімін пайдалана отырып, ғадауат сөздерді жіктеу үшін әртүрлі машиналық оқыту алгоритмдерін қолдану нәтижелерін салыстырамыз. Қазіргі заманғы зерттеулерде классификаторларды құру және оқытудың келесі кең таралған әдістері қарастырылады: шешім ағашы, кездейсоқ орман, тірек векторлық машина, k-ең жақын көршілер, логистикалық регрессия, аңғал Бейс (4.7-сурет).



Сурет - 4.7. Әдістердің өнімділік деңгейлерін салыстыру

4.1-кестеде әртүрлі мүмкіндіктер жиынтықтары бойынша SVM, Шешім ағаштары, Кездейсоқ ормандар (RF), K-Ең жақын көршілер (KNN), Naive Bayes алгоритмдері және Логистикалық регрессия (LR) сияқты әртүрлі машиналық оқыту алгоритмдерінің өнімділігі салыстырылады. Бағаланатын көрсеткіштер: дәлдік (ACC), дәлдік (прес.), еске түсіру, F1-балл (F1) және қисық астындағы аумақ (AUC).

Кесте 4.1 - Әдістер ауытқуының нәтижесі

Әдістер	**Ерекшеліктер**	**Acc.**	**Prec.**	**Recall**	**F1**	**AUC**
SVM	Statistical Features +TF-IDF	0.8204	0.2423	0.7593	0.3673	0.8622
	Statistical Features +TF-IDF +POS	0.8412	0.2512	0.6625	0.3643	0.8263
	Statistical Features +TF-IDF +POS + LIWC	0.1065	0.0641	0.8834	0.1196	0.5357
	Statistical Features +TF-IDF	0.9444	0.9529	0.201	0.332	0.6472
Шешім ағашы	Statistical Features + TF-IDF + POS	0.9444	0.8969	0.2159	0.348	0.6395
	Statistical Features +TF-IDF +POS + LIWC	0.9444	0.8812	0.2208	0.3532	0.6274
	Statistical Features	1234	1234	1234	1234	1234
	Statistical Features +TF-IDF	0.9368	1.0	0.0794	0.1471	0.9179
RF	Statistical Features +TF-IDF +POS	0.9369	1.0	0.0819	0.1514	0.9151
	Statistical Features +TF-IDF +POS + LIWC	0.9364	1.0	0.0744	0.1386	0.914
	Statistical Features +TF-IDF	0.9335	0.8421	0.0397	0.0758	0.5847
KNN	Statistical Features +TF-IDF +POS	0.9354	0.8158	0.0769	0.1406	0.6105
	Statistical Features +TF-IDF +POS + LIWC	0.9351	0.7037	0.0943	0.1663	0.701
	Statistical Features +TF-IDF	0.9681	0.8942	0.6079	0.7238	0.9739
Байес алгоритмдері	Statistical Features +TF-IDF +POS	0.9625	0.806	0.598	0.6866	0.9687
	Statistical Features +TF-IDF +POS + LIWC	0.9543	0.7304	0.531	0.6149	0.9599
	Statistical Features +TF-IDF	0.9601	0.9568	0.4392	0.602	0.9759
ЛР	Statistical Features +TF-IDF +POS	0.9598	0.9418	0.4417	0.6014	0.9759
	Statistical Features +TF-IDF +POS + LIWC	0.9409	0.6647	0.2804	0.3944	0.9336
Ескерту – Әдебиет негізінде құралған [4]						

Жоғарыдағы кестеде көрсетілгендей векторлық машиналарды қолдау (SVM): Дәлдік 0,1065 пен 0,9444 аралығында. POS және LIWC мүмкіндіктерін қосу әдетте Еске алуды жақсартады, бірақ дәлдік пен AUC азайтады.

Шешім ағаштары: Жоғары дәлдік (0,9368–0,9444) байқалады. Дегенмен, алгоритмдер жиі қайта шақырудың төмен деңгейін көрсетеді (0,3-тен төмен), бұл қалыпты F1 ұпайларына әкеледі.

Кездейсоқ ормандар (RF): дәйекті дәлдікке (шамамен 0,9335–0,9369) және жоғары дәлдікке қол жеткізеді, бірақ еске түсіру төмен болып қалады, бұл оның теңдестірілген жіктеулердегі тиімділігін шектейді.

KNN: 0,9681 ең жоғары дәлдікпен бәсекеге қабілетті өнімділікті ұсынады. TF-IDF мүмкіндіктері басқалармен біріктірілгенде дәлдік пен F1 ұпайы жақсарады.

Naive Bayes: F1 ұпайы мен жоғары AUC (0,9759 дейін) қол жеткізе отырып, дәлдік пен еске түсіруді жақсы теңестіреді.

Логистикалық регрессия (LR): тұрақты жоғары дәлдік (0,9409–0,9598), 0,9759-ға дейінгі күшті AUC. Статистикалық мүмкіндіктерді TF-IDF және POS көмегімен біріктіру метрикадағы ең жақсы жалпы теңгерімге қол жеткізеді.

LIWC мүмкіндіктерін қосу кейде еске түсіруді арттырады, бірақ AUC және Precision сияқты басқа көрсеткіштерді төмендетуі мүмкін.

4.3.1 Оқыту үшін мәліметтерді дайындау

Біздің деректер жиынтығы таңбаланған мәтіндерден тұрады. Әрбір мәтінде бейтарап материал үшін «0» және экстремистік мәтін үшін «1» деген белгі бар. Деректерді өңдеу мақсатында біз мәтіндермен айналысатындықтан, осы нақты сценарийде сүзу мен векторлауды ескердік. Мәтінді сүзу деп аталатын әдіс шу мен экстремалды мәндердің мөлшерін азайту мақсатында жүзеге асырылады.

Мәтіндерді сүзгіден өткізуге келгенде келесі алгоритм пайдаланылды:

Бірінші қадам- барлық кейіпкерлерді бір регистрге өзгерту және артық таңбалардан құтылу.

Екінші қадам- жиі кездесетін сөздерден құтылу (тоқтау сөздер «stop words»).

Үшінші қадам- инкубациялық және лемматизацияны орындау.

Төртінші қадам- мәтінді сөздерге немесе лексемаларға бөлуді көздейтін лексема үлгісін, сондай-ақ лексеманың ішіне кіруі мүмкін сөздердің мөлшерін білдіретін сөздердің n-граммалық үлгісін анықтаңыз [98].

Мәтінді нейрондық желіге қоймас бұрын біз әр сөзді сандық мәнге айналдырдық. Бұл нейрондық желілер тек сандық деректердегі үлгілерді тануды үйренетіндіктен жасалды, осылайша олар үйрене алады. Сөзді кодтау немесе токенизация - бұл процедураны сипаттау үшін қолданылатын термин [138].

Біз токенизация әдісі ретінде сөзді енгізуді қолдандық. Осы процесс арқылы сөздер сөзді кірістіру деп аталатын тығыз векторларға айналады. Басқаша айтқанда, ендірілген сөздер қосымша ақпаратты өлшемдердің кішірек жалпы санына жинай алады. Семантикалық мағынаны геометриялық кеңістікте бейнелеу осының мақсаты болып табылады. Енгізу кеңістігі деп аталатын бұл кеңістікте бір-біріне ұқсас мағыналары бар сөздер сандар немесе түстерді қамтитын ұяшық кеңістіктері сияқты бір-біріне ұқсас орындарға қойылады.

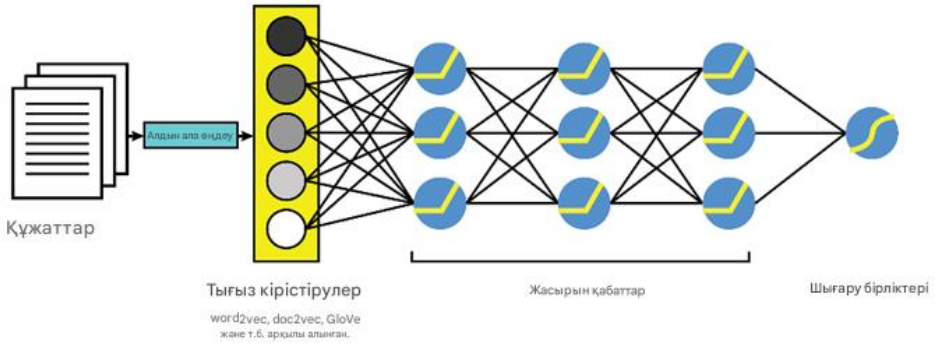
Деректерді токенизациялау мақсатында біз Keras бумасына кіретін Tokenizer функциясын қолдандық. Бұл утилита мәтіндік корпусты сандар кестесіне векторлауға мүмкіндік береді.

```
from keras.preprocessing.text import Tokenizer
from keras.preprocessing.sequence import pad_sequences
from keras.preprocessing import text, sequence

tokenizer = Tokenizer(num_words=20000)
tokenizer.fit_on_texts(list(X_train))

X_train = tokenizer.texts_to_sequences(X_train)
X_test  = tokenizer.texts_to_sequences(X_test)

X_train = sequence.pad_sequences(X_train, maxlen=200)
X_test  = sequence.pad_sequences(X_test,  maxlen=200)

print('X_train shape:', X_train.shape)
print('X_test shape: ', X_test.shape)
```

Токенизаторда екі негізгі параметр пайдаланылды. Біріншісі сөздік өлшемін анықтауға жауап беретін num_words параметрі болды. Екіншісі pad_sequence() функциясы болды, ол сөз тізбегін нөлдермен толтыру үшін пайдаланылды. Мәтіндердегі әртүрлі ұзындықтағы сөздер мәселесін шешу үшін pad_sequence() аргументі қолданылады. Сөздік өлшемін анықтауға жауап беретін сөз_саны опциясының мәні 20000 мәніне орнатылды. Біз бұл параметрді жаңа ғана қостық. Әрбір мәтіннің белгілі бір сөз ұзындығы бар екендігі бізді толғандыратын мәселелердің бірі болып табылады. Осыған байланысты maxlen опциясы пайдаланылуы тиіс тізбектердің ұзындығын көрсету үшін ұсынылды; осы өлшемнен ұзын тізбектер осы параметр арқылы қысқартылады [139].

Мәтінді категориялау үшін терең оқыту моделін құру

Кіріс деңгейі, жасырын қабат және шығыс қабаты нейрондық желілерді талқылағанда бізге таныс сөздер. Жасырын қабаттардың саны нейрондық желі топологиялары мен терең оқыту стратегиялары арасындағы негізгі айырмашылық болып табылады деген қорытынды жасауға болады. Бір немесе бірнеше жасырын қабаттары болуы мүмкін терең оқытумен салыстырғанда, қарапайым нейрондық желіде бір ғана жасырын қабат бар екендігі сурет 4.8- де келтірілген [126].



Сурет 4.8 - Терең оқыту моделі

Ескерту – Әдебиет негізінде құралған [137]

Біз кіріс нейрондарының қабатынан бастаймыз, онда біз функция векторларын енгіземіз, содан кейін мәндер жасырын қабатқа ауыстырылады. Әрбір қосылымда мәнді алға жібереміз, ал мән салмаққа көбейтіледі және мәнге ығысу қосылады. Бұл әрбір қосылымда орын алады және соңында біз шығыс қабатының мәнін аламыз. Шығыс деңгейі бір немесе бірнеше шығыс түйіндерінен тұрады. Біздің жағдайда, бір түйін, өйткені бізде екілік жіктеу тапсырмасы бар.

Нейрондық желінің мысалы келесідей: Әрбір кіріс түйіні p салмағы W салмағына көбейтіледі, содан кейін нәтижеге ығысу b қосылады. Бұл әрбір шығыс түйінінің мәнін анықтауға мүмкіндік береді. Осыдан кейін біз алған барлық мәндерді қосып, оларды f функциясына енгізуіміз керек. Функциялардың көптеген түрлері бар және белгілі бір функция белсендіру функциясы ретінде белгілі. Жасырын қабаттар үшін жиі ReLU деп аталатын тұрақты сызықтық бірлік пайдаланылады. Екілік классификация есептерінде шығыс қабаты үшін сигма функциясы пайдаланылады, ал көп класты жіктеу сұрақтарында шығыс деңгей үшін softmax функциясы қолданылады. Онда біз сигма функциясын қолданамыз.

Кері таралу – бұл алгоритм жаттығу процесін бастау үшін бастапқыда кездейсоқ мәндермен белгіленген салмақтарды жаттықтыру үшін қолданылатын процедура. Есептелген нәтиже мен болжанған нәтиже (сонымен қатар мақсатты нәтиже ретінде белгілі) арасындағы қатенің мөлшерін азайту үшін бұл процедура оңтайландырушы деп те аталатын градиенттің төмендеуі сияқты оңтайландыру әдістерін қолдану арқылы жүзеге асырылады. Қате жоғалту функциясы арқылы анықталатын функция болып табылады және оңтайландырушы функцияның жоғалуын азайту үшін пайдаланылады. Бұл контексте "Адам" деп аталатын оңтайландырушы және "кросс энтропия" деп аталатын жоғалту функциясы пайдаланылады. Мәтінді жіктеу үшін конволюционды нейрондық желіні әзірлеу

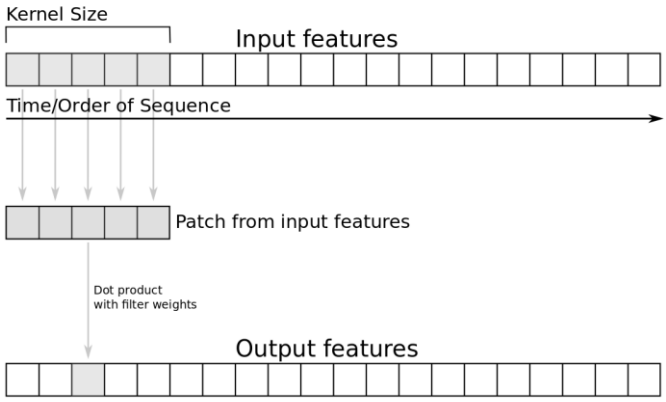
Конволюционды нейрондық желілер кескіндерден мүмкіндіктерді алу және оларды нейрондық желілерде пайдалану арқылы кескін классификациясы мен компьютерлік көруді өзгертті. Оларды кескінді өңдеуде пайдалы ететін

қасиеттер оларды реттілікпен өңдеуге де қолайлы етеді. CNN көмегімен мәтінді жіктеу жағдайында конволюционды ядро сөздерді ендірулер бойынша сырғытады, тек оның міндеті - бірден бірнеше сөздің ендірілуін қарау. Конвульсия ядросының өлшемдері де осы тапсырмаға сәйкес өзгеруі керек.

Сөздерді ендіру ретін көру үшін терезенің реттіліктегі бірнеше (әдетте 3 немесе 5) сөз ендірілгенін қарауын қалаймыз. Ядролар өлшемдері 3х300 немесе 5х300 (енгізу ұзындығы 300) сияқты кең төртбұрыш болады. Біздің жағдайда 4х200, өйткені біз таңбалау кезінде реттілік ұзындығын 200 етіп орнаттық. Әрбір ядро ұясының сәйкес салмағы бар. Ядро ендірілген сөздің үстінен сырғыған кезде, ядро салмағы сөзді ендіру мәніне көбейтіледі, содан кейін шығыс мәнін алу үшін барлық көбейтілген мәндер қосылады.

Конволюционды нейрондық желі осы ядролардың көпшілігін қамтиды және желі үйретілген сайын ядроның бұл салмақтары үйренеді. Әрбір өзек сөзді және оның айналасындағы сөздерді дәйекті терезеде қарауға арналған. Осылайша, конволюция операциясын терезеге негізделген мүмкіндікті шығару ретінде қарастыруға болады. Бұл конвульсия операциясының тағы бір жақсы қасиеті бар. Еске салайық, ұқсас сөздердің ұқсас кірістірулері болады, ал конволюция операциясы осы векторлардағы сызықтық операция ғана. Осылайша, конволюциялық ядро ұқсас сөздердің әртүрлі жиындарына қолданылғанда, ол бірдей шығыс мәнін шығарады.

Сөздердің толық тізбегін өңдеу үшін бұл ядролар сөз ендірілгендер тізімін дәйекті түрде жылжытады. Бұл 1D конволюциясы деп аталады, себебі ядро тек бір өлшемде қозғалады: уақыт. Бір ядро бірінші сөзді ендіруге, содан кейін келесі сөзді енгізуге, келесіге және т.б. қарай отырып, кіріс ендірілгендер тізімі бойынша бір-бірден жылжиды. Нәтижедегі шығыс мүмкіндік векторы болады. Төмендегі сурет 4.9- да мұндай конвульсияның қалай жұмыс істейтінін көруге болады.



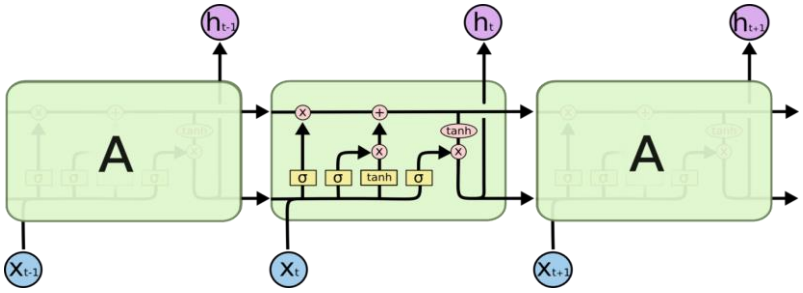
Сурет 4.9 - Тізбекті модельдеуге арналған уақытша 1D конволюциясы

Біздің конвультивтік мүмкіндік векторларының әрқайсысын өңдеуден алынған максималды мәндер біріктіріліп, соңғы қабатқа өтеді. Ол MaxPooling деп аталады.

Енді осы желіні Кераста қалай пайдалануға болатынын көрейік. Біріншіден, біз input_dim параметрлері бар ендіру қабатын қосуымыз керек - сөздік өлшемі, біз пайдаланғымыз келетін бірегей сөздердің саны; input_length – тізбектің ұзындығы; output_dim - енгізілген айнымалының өлшемі. Содан кейін біз түйіндердің 50% алып тастау үшін алып тастау деңгейін орнатамыз. Енді біз ядро өлшемі 4 болатын 100 сүзгіден тұратын конволюция қабатын қосамыз, осылайша әрбір конволюция 4 сөзді енгізу терезесін және қайта белсендіру функциясын ескереді. Max Pooling қабатын қоспас бұрын, нормалау қабатын қосамыз. Біріктіру қабатынан кейін 8 шығыс өлшемін алу үшін тығыз қабатты қосыңыз және relu белсендіру функциясын пайдаланыңыз. Соңында біз шығыс қабатын орнатамыз. Біз екілік жіктеуді орындағандықтан, біз сигма тәрізді белсендіру функциясын қолданамыз және шығыс қабатта 1 нәтиже аламыз.

Қайталанатын нейрондық желі – объектіні жіктеу және сөйлеуді анықтау сияқты көптеген күрделі компьютерлік мәселелерді шешуге арналған терең оқыту алгоритмі. RNNs алдыңғы оқиғалардан алынған ақпарат негізінде әрбір оқиғаны түсіну арқылы бірізді болып жатқан оқиғалар тізбегін өңдеуге арналған. RNN қайталанатын деп аталады, себебі ол әрбір келесі элемент үшін бірдей тапсырманы орындайды және нәтиже алдыңғы есептеуге байланысты. RNN жоғалып кететін градиент мәселесіне байланысты нақты өмір сценарийлерінде сирек қолданылады. Бұл RNN үшін ең үлкен өнімділік мәселелерінің бірі. Іс жүзінде RNN архитектурасы өзінің ұзақ мерзімді жады мүмкіндіктерін шектейді, олар бір уақытта тек бірнеше ретті есте сақтаумен шектеледі. Демек, RNN жады қысқарақ реттілік пен қысқа уақыт кезеңдері үшін ғана пайдалы. Жоғалып бара жатқан градиент мәселесі дәстүрлі RNN-нің жад мүмкіндіктерін шектейді - тым көп уақыт қадамдарын қосу кері таралуды пайдалану кезінде градиент мәселесі мен ақпараттың жоғалу мүмкіндігін арттырады.

LSTM (ұзақ қысқа мерзімді жад) жоғалып кететін градиент мәселесін шешуге арналған және оларға дәстүрлі RNN-мен салыстырғанда ақпаратты ұзақ уақыт бойы сақтауға мүмкіндік береді. Сондықтан біз дәстүрлі қайталанатын нейрондық желіні емес, LSTM пайдаланамыз. Суретке сілтеме



Сурет 4.9 - LSTM архитектурасы

Ескерту – Әдебиет негізінде құралған [139]

Суретте деректерді дәйекті өңдеу үшін кеңінен қолданылатын қайталанатын нейрондық желінің (RNN) түрі Ұзақ қысқа мерзімді жад (LSTM) бірлігінің архитектурасы көрсетілген. Ол үш уақыт қадамы бойынша LSTM өңделмеген құрылымын көрсетеді: (t-1), (t) және (t+1).

Әрбір жасыл блок LSTM ұяшығын білдіреді. Әрбір уақыт қадамында LSTM ұяшығы ағымдағы енгізуді (x_t), алдыңғы уақыт қадамынан жасырын күйді (h_{t-1}) өңдейді және жаңа жасырын күйді (h_t) жасайды. Сонымен қатар, ұяшық әрбір блок арқылы өтетін көлденең сызық ретінде бейнеленген ұяшық күйін сақтайды және жаңартады.

Орталық LSTM блогы оның ішкі операцияларын көрсету үшін кеңейтілген. Диаграмма негізгі құрамдастарды бөлектейді: кіріс қақпасы (σ), ұмыту қақпасы (σ), шығыс қақпасы (σ) және үміткер ұяшық күйі (tanh). Бұл қақпалар ақпарат ағынын реттейді, LSTM ұзақ мерзімді тәуелділіктерді сақтауға және жойылатын градиент мәселелерін басқаруға мүмкіндік береді.

- ұмыту қақпасы: ұяшық күйінен қандай ақпаратты алып тастау керектігін шешеді;

- кіріс қақпасы: қандай жаңа ақпаратты сақтау керектігін анықтайды;

- шығыс қақпасы: ұяшықтан шығатын шығысты басқарады [127].

LSTM желісіне мәтін енгізу үшін алдымен мәтіндік деректерді сандық тізбектерге түрлендіру қажет. Нейрондық желі моделін құру кезінде біз мәтінді сандық мәндерге түрлендірдік. Әрі қарай біз сол сандық мәндермен жұмыс істейміз.

Осы нейрондық желі үшін конволюциялық нейрондық желі сияқты дәл сол модельді жасақталды. Тек конволюциялық қабаттың орнына екі бағытты (Екі бағытты) LSTM қабаты қосылды.

Екі бағытты LSTM-лер дәйектілікті жіктеу мәселелерінде үлгі өнімділігін жақсарта алатын дәстүрлі LSTM кеңейтімі болып табылады. Енгізу тізбегінің барлық уақыт аралығы қол жетімді тапсырмаларда қос бағытты LSTMs енгізу тізбегіндегі біреудің орнына екі LSTM оқытады. Біріншісі сол күйінде енгізу ретін, ал екіншісі енгізу ретінің инверттелген көшірмесін білдіреді. Бұл желіні қосымша контекстпен қамтамасыз етіп, мәселені тезірек және одан да толық зерттеуге әкелуі мүмкін. Генеративті терең оқытудың бұл түрімен шығыс қабаты бір уақытта кері және тура күйлерден үйрене алады [139].

Эксперименттердің сипаттамасы

Алгоритм бірнеше кезеңнен тұрады:

1) Сандар мен арнайы таңбалар жойылады, тек сөздер таңдалады.

2) Сөз соңынан жұрнақтар мен бағыныңқы сөйлемдер алынып тасталады.

3) Дерекқорда әрбір жұрнақтың немесе қабылданған сөздің болуы немесе болмауы тексеріледі.

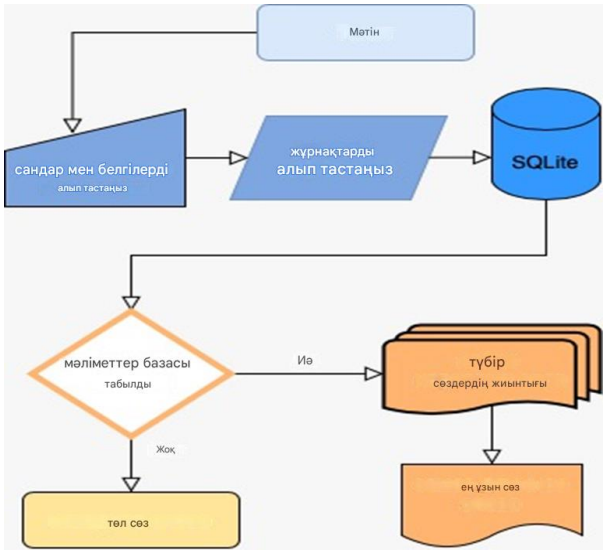
4) Деректер базасында бірнеше сөз табылса, ең ұзын сөз қайтарылады.

5) Егер дерекқордан жоғарыда аталған әдістермен сәйкестік болмаса, дерекқорда іздеу берілген сөздің басынан басталады.

6) Дерекқорда анықталмаған жағдайда енгізілген сөздің өзі қайтарылады [127].

Қазақ тілінде жұрнақтар мағынасы мен қызметіне қарай екі түрге бөлініп қарастырылады: 1) сөз тудырушы жұрнақ түбір сөзге жалғанған соң жаңа сөз шығады. Мысалы, «қой-шы», «оқу- шы», «жас-тық», «шаға-ла».

2) Сөзді түрлендіретін, жалғанған сөзге үстеме мағына ұсынатын, сөздің түрін өзгертетін жұрнақтар. Мысалы, «ақ-шыл», «ақ -шылтым», «қара-лау», «жасыл-рақ», «айт-ып», «айт-қалы» [4]. (Сурет 4.10)



Сурет 4.10 - Сөздің түбірін анықтау алгоритмі
(Сурет сөздердің түбірін анықтауды көрсету мақсатында автордың қолымен сызылды)

Жұрнақ – сөз бен сөзді байланыстыратын, сөздіктегі қатынасты көбейтетін және сөзге грамматикалық мағына қосатын жалғау.

Қазақ тілінде сөздің түбіріне жалғанатын жалғаулар сөздің бастапқы қалпын өзгертеді. Сөздің негізгі түбіріне жалғауларды жалғау кезінде «п» әрпі «б» әрпімен, «қ» әрпі «ғ» әрпімен, «к» әрпі «г» әрптерімен ауысып отырады.

Мысалы, «і» жалғауын алар болсақ, «мектеп» сөзіне жалғағанда «мектебі» сөзі пайда болады. Осындай ережелерді негізге ала отырып, түбір сөзді нақтылап алу мақсатында checkKazChar() әдісі пайдаға асты. Деректер базасында жалғауы бар сөздердің бар - жоғына көз жүгіртіп өтті [36].

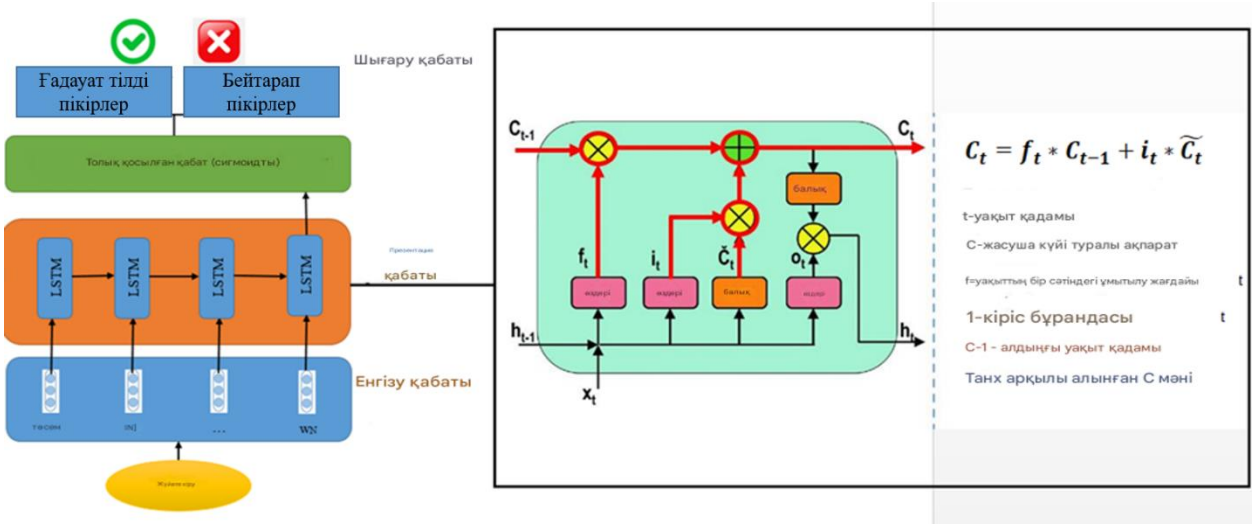
Деректер қорында мұндай сөз кездеспген жағдайда, сөздің ең соңғы әріпін алып checkKazChar() тәсілімен қаралды. «Мектебі» сөзі «мектеп» сөзімен ауыстырылды. [99].

Келесі кезеңде мәтіндегі әрбір сөздің жұрнағы жоғарыда сипатталған түбір алу алгоритміне сәйкес алынып тасталады, тек түбірлер ғана қалады. Енгізілген мәтіндегі сөздердің негізін анықтау үшін қазақ тіліндегі 24 мың және 76 мыңға жуық сөздік қорын қамтитын қазақ тіліндегі екі дерекқор пайдаланылды.

Кілтсөзді анықтау мәселесінің сипаттамасы

TF-IDF (ағылшын тілінен TF – термин жиілігі, IDF – кері құжат жиілігі) – құжат жинағының немесе корпустың бөлігі болып табылатын құжат контексіндегі сөздің маңыздылығын бағалау үшін қолданылатын статистикалық көрсеткіш. Сөздің салмағы бұл сөздің құжатта кездесу жиілігіне пропорционал және жинақтағы барлық құжаттарда сөздің кездесу жиілігіне кері пропорционал. TF-IDF өлшемі мәтінді талдау және ақпаратты іздеу тапсырмаларында жиі пайдаланылады, мысалы, кластерлеу кезінде құжаттардың жақындық өлшемін есептеу кезінде іздеу сұрауына құжаттың сәйкестігінің критерийлерінің бірі ретінде. TF-IDF-те жоғары салмақ белгілі бір құжатта жиілігі жоғары, ал басқа құжаттарда жиілігі төмен сөздерге беріледі.

Ұсынылған модель иерархиялық көп сатылы процесс және бірнеше қабаттардан тұрады. Бірінші қабат - алдын ала берілген енгізу арқылы векторлық кеңістікте сөздерді көрсететін сөз енгізу қабаты. Әрі қарай әр мәтіннің көрсету матрицасын алу үшін LSTM қолданатын көрсету қабаты келеді, содан кейін мәтіндердің оңтайлы көрінісін қамтамасыз ету үшін ағымдағы сөйлемнің мәтінмәні мен дисплейін біріктіретін әртүрлі әдістерді қолданады. Келесі қабат мәтінді жіктеу нәтижесін беретін шығыс қабаты болып табылатындығы Сурет 4.11- де көрсетілген.



Сурет 4.11 - LSTM қабаттарын жіктеу

Ескерту – Әдебиет негізінде құралған [121]

Сөзді енгізу деңгейі әрбір сөзді семантикалық және синтаксистік ақпаратты сақтай алатын үлкен векторлық кеңістікте салыстыруға жауап береді. Матрицаның әрбір бағанында сәйкес сөздің енгізілген сөзі сақталады.

Сөздерді ендіру – векторлық кескінді пайдаланып сөздер мен құжаттарды көрсету әдістерінің категориясы. Вектор — сөздің үздіксіз кеңістікке проекциясы. Сөздің векторлық кеңістіктегі көрінісі мәтіннен алынады және сөзді қолдану барысында оның жанында орналасқан сөздерге негізделеді. x1 деп алдық. Біріншіден, біз әрбір сөздің тұрақты сөзін енгізу үшін сөз векторларын пайдаланамыз. Бұл қабат арқылы сөйлем матрица түрінде

беріледі: X ∈ Rm × T, мұндағы m - вектор сөзінің өлшемі, ал T - сөйлемнің ұзындығы. Біріншіден, сөз енгізу үлгілері хабарламалардағы сөздерді векторларға түрлендіреді. Содан кейін сөздер арасындағы контекстік тәуелділік қашықтығын зерттеу үшін LSTM-ге сөйлемдегі сөздер тізбегі енгізіледі.

Сөз енгізу деңгейі кездейсоқ сандармен инициализацияланады және оқу жиынындағы барлық деректерде векторлау тапсырмасын орындайды. Бұл қабат желідегі бірінші жасырын қабат болып табылады және үш аргумент алады:

input_dim: мәтіндік деректер сөздігінің өлшемі.

output_dim: сөздер векторланған векторлық кеңістіктің өлшемі. Ол әрбір сөз үшін осы қабаттағы шығыс векторларының өлшемін анықтайды.

енгізу_ұзындығы: енгізу реттерінің ұзындығы.

LSTM негізіндегі өкілдік – визуализация деңгейі

LSTM желісіндегі негізгі элемент ұяшықтың күйі болып табылады, яғни диаграмманың жоғарғы жағынан өтетін көлденең сызық. LSTM ұяшық күйіндегі ақпаратты жоюға немесе оған жаңа ақпаратты қосуға қабілетті, бірақ бұл мүмкіндік қақпа деп аталатын құрылыммен басқарылады. Ол сигма тәрізді нейрондық желілерден және нүктелерді көбейту операцияларынан тұрады. LSTM-дегі бірінші қадам ұяшық күйінен қандай ақпаратты жою керектігін шешу болып табылады. Бұл шешімді «ұмыту механизмі» деп аталатын сигма тәрізді қабат қабылдайды. Оның кірісіне ht-1 және xt мәндері беріледі, ал шығысы 0 мен 1 арасындағы сан болып табылады.

Эксперименттік бөлігі

	precision	recall	f1-score	support
0	0.86	0.91	0.89	293
1	0.88	0.82	0.85	232
accuracy			0.87	525

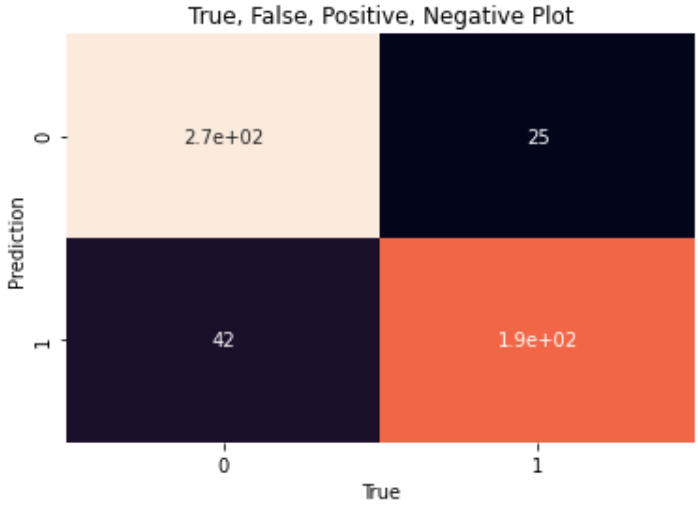
Шатасу матрицасы нақты мақсатты мәндерді машиналық оқыту үлгісімен болжанған мәндермен салыстырады. Бұл бізге жіктеу моделінің қаншалықты жақсы орындайтыны және қандай қателіктер жіберетіні туралы түсінік береді. Мақсатты айнымалының екі мәні бар: оң немесе теріс. Бағандар мақсатты айнымалының нақты мәндерін білдіреді.

Нағыз оң (TP) = 2,7 e + 02 (733); 560 үлгісіне сәйкес оң класс деректері дұрыс жіктеледі.

Шынайы теріс (TN) = 1,9 e + 02 (516); 516 үлгісі бойынша теріс кластың деректері дұрыс жіктеледі.

Жалған оң (FP) = 25; үлгі бойынша теріс класстың 25 деректері оң класқа жататындар ретінде қате жіктеледі.

Жалған теріс мәндер (FN) = 42; 42-модельге сәйкес оң класстың деректері теріс сыныпқа жататындар ретінде қате жіктеледі. (Сурет 4.12)



Сурет 4.12 - Үлгі қателерінің визуалды көрінісі
(Сурет автордың үлгі қателерін анықтау кезіндегі нәтижесі)

Біз AUC-ROC мәні 0,86 екенін көреміз, бұл классификатор оң класс мәндерін теріс класс мәндерінен жақсы ажырата алатындығын көрсетеді.

Келесі нәтижелер Word2vec және биграммалар тіркесімін терең оқыту алгоритмдерін енгізу арқылы алынады.

	precision	recall	f1-score	support
0	0.66	0.96	0.78	257
1	0.91	0.40	0.56	217
accuracy			0.71	474

Нағыз оң (TP) = 2,5 e + 02 (680); 680 үлгісіне сәйкес оң класс деректері дұрыс жіктеледі.

Жалған теріс мәндер (TN) = 87; үлгі бойынша теріс класстың 87 деректері дұрыс жіктелген.

Жалған оң (FP) = 9; 9 модельге сәйкес теріс кластың деректері оң классқа жататындар ретінде дұрыс емес жіктеледі.

Жалған теріс (FN) = 1,3 e + 02 (353); 353 үлгісіне сәйкес оң класс деректері теріс класс деректері ретінде қате жіктеледі.

Матрицадан көріп отырғанымыздай, Word2vec және биграммаларды пайдалану процесінде модельдегі жалған теріс нәтижелердің саны өте көп (353 деректер бойынша қате жіктеу жасалған). Жоғарыда келтірілген нәтижеге сүйене отырып, экстремистік мәтіндерді жіктеуде осы белгілер комбинациясын пайдаланудың тиімсіздігін байқауға болады.

AUC-ROC мәні 0,68, бұл классификатордың оң және теріс класс нүктелерін айыруда жақсы жұмыс істемейтінін білдіреді (ол 0,5-ке жақын болған сайын, классификатордың өнімділігі төмен). Жіктеуіш барлық деректер үшін кездейсоқ класты немесе тұрақты классты болжайды. Жіктеуіш үшін AUC-ROC мәні неғұрлым жоғары болса, оның оң және теріс класстарды ажырату қабілеті соғұрлым жоғары болады. Эксперименттің келесі бөлімінде мәтіндерді тереңдетіп оқыту алгоритмдеріне негізделген белгілер мен

биграммалар сөмкесінің көмегімен жіктеу жүргізілді, бұл әдістеме келесі нәтижелерді көрсетті.

AUC-ROC мәні 0,77 болып табылады, бұл жіктеуіш Word2vec әдісімен және биграммалармен салыстырғанда Word сөмкесі әдісі мен биграммалар негізінде жақсырақ жұмыс істейтінін көрсетеді.

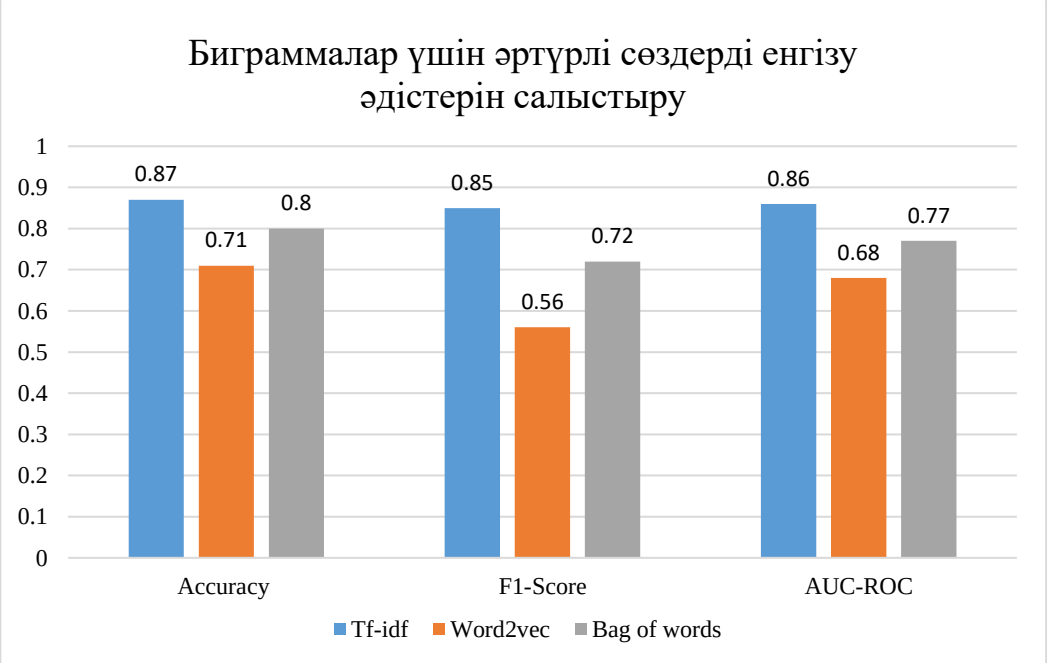
TF-IDF әдісі, Word2vec, Сөздер мен биграммалар қаптамасы негізіндегі мәтінді жіктеудің салыстырмалы нәтижелері кестеде және диаграмма түрінде берілген (4.3-кесте, 4.13 – сурет).

4.3-кестеден көріп отырғанымыздай, барлық бағалау параметрлері бойынша ең жақсы нәтиже TF-IDF+bigram әдісімен (үлгінің дәлдігі 0,87, F1-өлшемі 0,85-ке тең және ROC AUC мәні 0,86-ға тең) көрсетілген.

Тәжірибе нәтижелері бойынша келесі ең жоғары нәтиже Bag of words+bigram негізінде алынған, модель дәлдігі 0,8, F1-өлшемі 0,72, AUC-ROC мәні 0,77.

Кесте 4.3 - Дәлдік, F1-балл және AUC-ROC әдістерін салыстыру

Терең оқыту алгоритмінде қолданылатын мүмкіндіктер	Accuracy	F1-Score	AUC-ROC
TF-IDF+bigram	0.87	0.85	0.86
Word2vec+bigram	0.71	0.56	0.68
Bag of words+bigram	0.80	0.72	0.77



Сурет 4.13 - Биграммалар үшін әртүрлі сөздерді енгізу әдістерін салыстыру

Word2vec+bigram негізіндегі модель бұл тапсырма үшін әсіресе қолайлы болмайды, өйткені жіктеу нәтижелері төмен ұпайларды көрсетті, оның ішінде AUC-ROC мәні 0,68, бұл модельдің оң және теріс класс деректерін анық ажырату қабілетінің төмендігін көрсетеді.

Бұл алгоритмдердің дәлдігін арттыру мақсатында қазақ тіліне арналған стемпинг алгоритмі қолданылды. Түйінді алгоритмді қолдану нәтижесінде TF-IDF+bigram мүмкіндіктерінің комбинациясы келесі нәтижені көрсетті.

Нәтижелер

1) Алғаш рет қазақ тілінің ерекшеліктерін ескере отырып, TF-IDF әдісінің биграммаларға қолданылуымен ерекшеленетін семантикалық талдау моделі құрылды, бұрын стюминг алгоритмімен кірістірілген қабатқа қолданылған LSTM желісінің сөздері және экстремистік мәтіндерді анықтау дәлдігін арттыру.

2) N-граммалар мен сөз енгізу әдістерінің тіркесіміне негізделген және экстремистік мәтіндерді жіктеу сапасын жақсартатын мүмкіндіктер жиынтығын құру әдісі әзірленді.

Сәулет дизайны

Жүйенің архитектурасы оның негізгі құрамдас бөліктерін, олардың қарым-қатынастарын (құрылымдарын) және олардың бір-бірімен өзара әрекеттесуін сипаттайды. Бағдарламалық жасақтаманың архитектурасы мен дизайны бизнес стратегиясы, сапа атрибуттары, адам динамикасы, дизайн және АТ ортасы сияқты бірнеше факторларды қамтиды. Ғылыми зерттеу жұмысы бойынша бізге әлеуметтік желілерді талдайтын бағдарламалық құрал жасау керек. Веб-қосымшаны әзірлеу бірнеше кезеңнен тұрады:

− зерттеудің мақсаттары мен міндеттерін анықтау (техникалық тапсырманы құру);

− қолданбалы құрылымды әзірлеу;

− дизайн макеттерін құру;

− макет;

− бағдарламалау және қолданбалы логиканы дамыту;

− тестілеу және жөндеу.

Жобалау кезеңінде зерттеу іс-әрекетінің негізгі мақсаттары анықталады, маңызды талаптар жасалады және кең идея құрастырылады.

Жобаның мақсаттары мен міндеттерін анықтау

Жобаның мақсаты – ExWeb бағдарламалық қамтамасыз етуді әзірлеу, кешенді зерттеу жүргізу және тартылған пайдаланушыларды анықтау алгоритмдерін және сілтемелерді графикалық визуализациялау әдістерін әзірлеу. Қойылған мақсаттарға жету үшін келесі міндеттер анықталды:

- алдын ала дайындалған машиналық оқыту үлгісін пайдалана отырып, Vkontakte, Youtube, Twitter әлеуметтік желілерінің мазмұнын талдау;

- желіні пайдаланушылар арасындағы байланысты және олардың белгілі бір топтармен байланысын анықтау;

- қоғамдастық көшбасшысының анықтамасы.

UML диаграммаларында класстар арасындағы қатынастарды белгілеудің келесі негізгі түрлері жиі кездеседі:

- ассоциация бір нысанның (сыныптың) объектілерінің басқа бір объектінің объектілерімен бір сыныптың объектілерінен екіншісіне өтуге болатындай байланыстырылғанын көрсетеді;

- іске асыру - модельдің екі элементі арасындағы қатынас, онда бір элемент екіншісімен көрсетілген әрекетті жүзеге асырады;

- жинақтау - бұл бүтін мен оның бөліктері арасындағы қатынастың бір түрі;

- композицияның контейнер класы даналарының қызмет ету мерзіміне және қамтылған сыныптардың даналарына қатаң тәуелділігі бар. Егер контейнер жойылса, онда оның барлық мазмұны да жойылады.

Қолданбалы құрылымды әзірлеу кезінде оның мазмұнымен және ақпараттық стратегиясымен байланысты барлық факторларды ескеру қажет. Бұл стратегия қолданбаны пайдаланушыларға ақпаратты қысқа және қарапайым түрде табуға мүмкіндік беретін әдістер мен тәсілдерді белгілейді. Бұл кезеңде бірінші орындалатын нәрсе - типтік беттер мен олар көрсететін негізгі функциялар арасындағы байланыстарды суреттейтін қолданба картасын құру [140].

Қолданба картасы диаграмма түрінде берілген, блок-схема деп те аталады, онда әрбір бет бір тіктөртбұрыш түрінде көрсетіледі. Әрбір бет арасындағы байланыстар, сондай-ақ беттер арқылы өтуге мүмкіндік беретін шарлау құрылымы көрсетіледі.

Бұған қоса, беттегі мәтін мен көрнекі бейнелердің орналасуын, сондай-ақ пайдаланушылардың осы құрамдас бөліктермен өзара әрекеттесу тәсілін көрсету үшін негізгі үлгі беттер үшін сым жақтаулары әзірленеді. Болашақта бет жақтауында кеңейту үшін орын болуы керек.Wireframe - бұл дизайн және веб-әзірлеу процесінің алғашқы қадамы. Ол әркім жобаның ішінде не болып жатқанын нақты көре алатындай практикалық карта ретінде пайдалануға арналған. Рамка пайдаланушылардың біздің қолданбамызбен қалай әрекеттесетіні туралы түсінік береді.

Біздің веб-қосымшаның қаңқасын жасау кезінде бізді ақпарат пен навигация дизайны қызықтырады. Ақпараттық дизайн – ақпаратты оқуға және түсінуге оңай етіп беру жолы. Навигация дизайны - оны анықтау және тиімді пайдалану оңай болуы үшін веб-қосымшада навигация қалай көрсетіледі.

Басқаша айтқанда, дизайнер екі негізгі нәрсені анықтауы керек:

- қолданбадағы мазмұнды көрсетудің ең жақсы жолы;

− қолданбадағы мазмұнды табуды қалай жеңілдетуге болады.

Графикті визуализациялау алгоритмі

Графиктің өзі қосылған шыңдардың тығыз орналасуын қамтамасыз ететін қуат графигін көрсету алгоритмі деп аталатын алгоритм арқылы көрсетілді. Сонымен қатар, график шыңдарын жеке қажетті шыңдарды ескере отырып, сүйреп апаруға және жылжытуға болады. ВКонтакте әлеуметтік желісінде

көрсетілген пайдаланушы аты шыңдарда жазылған. Меңзерді жоғарғы жағына апарған кезде, осы пайдаланушының жеке бетіне сілтемені көре аласыз.

Визуализацияны жасау үшін веб-қосымша әзірленді. JavaScript, HTML және CSS графиктерді көрсету үшін веб-бағдарлама пайдаланған тілдер болды. Визуализация мақсатында d3.js бумасы пайдаланылды. Дисплей үшін SVG визуалды пішімін пайдалану туралы шешім қабылданды, себебі ол фотосуреттің анықтығын бұзбай суретті қалағаныңызша үлкейтуге және кішірейтуге мүмкіндік береді.

Деректер жинау. Үтірмен бөлінген топтар тізімін енгізген кезде және осы топтардан қажетті жазбаларды қамтитын уақыт кезеңін көрсеткенде, жазбалар графигін алуға болады. Мұнда деректерді жинау кезінде белгілі бір уақыт аралығында көрсетілген топтардың жазбалары, сондай-ақ пікірлер немесе ұнатулар қалдырған пайдаланушылар туралы ақпарат жиналды. Атап айтқанда, тізімдегі әрбір топ үшін келесі қадамдардан тұратын get_posts әдісі іске қосылды:

Топтағы барлық жазбаларды алу мақсатында get_posts әдісі - жарияланым күні бойынша кему ретімен сұрыпталған хабарламалар тізімінде қажетті хабарламалар қатарда пайда болатын жазбаны табу үшін екілік іздеу алгоритмін пайдалану. Осылайша, біз соңғысынан бастап барлық жазбаларды алудан гөрі жақсы өнімділікті қамтамасыз етеміз. Орташа алғанда, алгоритм қажетті жазбаны табуға уақыт жұмсайды, мұнда n - топтағы барлық жазбалардың саны.

Барлық жазбаларды алғаннан кейін, әрбір жазба үшін оның барлық пікірлері немесе ұнатулары get_post_weights әдісімен алынады.
Содан кейін осы деректер негізінде график құрастырылады. Графикті визуализация сұрауы қайта шақырылғанда көп уақытты қажет етпеуі үшін мәліметтер базасын пайдалану жоспарлануда.

Мәліметтер бойынша график құру. Мұндағы пост жоғарғы жағында, ал оның астындағы пікірлер немесе лайктардың жалпы саны оның салмағы болып табылады. Шыңдар арасындағы жиектер салмағы - бұл жазбаны ұнатқан немесе пікір қалдырған жазбалардың бірдей пайдаланушыларының саны.

Графикалық талдау. Графикті талдау үшін NetworkX кітапханасының көмегімен график құрастырылды. Графикті талдау кезінде графтың әртүрлі қасиеттері есептелді, мысалы, Дәрежелік Орталықтық, Жақындық Орталықтық, Аралық Орталықтық, төбелер саны, жиектер саны және графиктің оң жағында көруге болатын тығыздық. өзі. Бұл графикте Чи квадраты жоқ. Мұнда ең ықпалды пайдаланушыларды табудың қажеті жоқ, тек пайдаланушыға қосымша пайдалы ақпаратты көрсету қажет.

Графикалық визуализация. Графиктің өзі қосылған шыңдардың тығыз орналасуын қамтамасыз ететін қуат графигін көрсету алгоритмі деп аталатын алгоритм арқылы көрсетілді. Сонымен қатар, график шыңдарын жеке қажетті шыңдарды ескере отырып, сүйреп апаруға және жылжытуға болады. Төбенің идентификаторы шыңдарға жазылады. Меңзерді үстіңгі жаққа апарсаңыз, хабарлама мәтінінің бір бөлігін көре аласыз, басқан кезде хабарлама бетіне өтуге болады.

Веб-қосымша пішіміндес визуализация әзірленді. Веб-қосымшадағы графикті салу үшін html, css және javascript пайдаланады. Визуализация үшін d3.js кітапханасы пайдаланылды. svg графикалық дисплей пішімі ретінде таңдалды, себебі ол кескінді сапаны жоғалтпай, қалағаныңызша бірнеше рет үлкейтуге және кішірейтуге мүмкіндік береді.

Django Python құрылымы веб-қосымшаның сервері ретінде пайдаланылады. Html бетінде svg элементі орналастырылды, онда график d3.js көмегімен JavaScript сценарийі арқылы сызылған. Тіпті html бетін жасау сатысында Джанго деректерді алғаннан кейін біз жасаған қажетті деректерді жібереді. Әрі қарай түс, күш алгоритмінің тартылу күші, график шыңдарының бір-бірінен бастапқы қашықтығы, svg элементіндегі графиктің ортасы сияқты әртүрлі параметрлерді таңдаймыз.

Содан кейін svg элементінің ішінде кейбір <g> элементтерін жасау керек. Бұл элементтер графиктің болашақ шыңдары мен жиектерін белгілейді. Төбелер мен жиектерді бір-бірінен ажырату үшін біз оларды тиісінше класс түйіндері мен сілтемелерімен белгілеуіміз керек. Содан кейін әрбір элемент сәйкес деректермен толтырылады және элементтер үшін қосымша параметрлерді орнатуға болады.

Онлайн желі қолданушылар парақшасына кіріп, осы топтың қажетті жазбалары бар уақыт кезеңін көрсеткенде, пайдаланушы жазбаларының ағашын алуға болады. Мұнда деректерді жинау кезінде белгілі бір уақыт аралығында көрсетілген топтың жазбалары, сондай-ақ пікір қалдырған немесе ұнатқан пайдаланушылар туралы ақпарат жиналды. Атап айтқанда, келесі қадамдардан тұратын топ үшін get_posts әдісі іске қосылды:

Парақшадағы барлық жазбаларын алу үшін get_posts әдісі - жарияланым күні бойынша кему ретімен сұрыпталған хабарламалар тізімінде қажетті хабарламалар қатарда пайда болатын жазбаны табу үшін екілік іздеу алгоритмін пайдалану. Осылайша, біз соңғысынан бастап барлық жазбаларды алудан гөрі жақсы өнімділікті қамтамасыз етеміз. Орташа алғанда, алгоритм қажетті жазбаны табуға уақыт жұмсайды, мұнда n - топтағы барлық жазбалардың саны.

Содан кейін осы деректер негізінде ағаш салынады. Ағашты визуализация сұрауы қайта шақырылғанда көп уақытты қажет етпеуі үшін дерекқорды пайдалану жоспарлануда.

Қатынас ағашы түріндегі деректерді визуализациялау. ВКонтакте әлеуметтік желісіне өтпей-ақ, жазбалар мен пайдаланушылар арасындағы сілтемелерді ыңғайлы көру үшін топтың түбірі, ал жазбалар оның балалары болып табылатын ағаштың визуализациясы жүзеге асырылды. Бұл ретте ағаштың жапырақтары тиісті жазбалардың астына пікір қалдырған қолданушылар. Пайдаланушылар үшін аты- жөні, тегі және аватары бойынша ақпаратты алуға болады. Яғни, енді пайдаланушы түбір, қолданушының балалары - бұл пайдаланушы жазбаларының астына пікір қалдырған топтар, ал жапырақтар - пайдаланушы пікір қалдырған жазбалар.

Мысалыға алар болсақ, «Оған ертең келу керек», «Сен келдің». Екі сөйлемді алып қарайтын болсақ, «кел» етістігін байқаймыз, дегнмен

жалғауларының әртүрлі аяқталғанына назар аудару қажет. Егер текст ақпараттық құралдарда өңделуге жіберілсе, «кел» сөзін пайдалану барсыныда бірдей ұғымда айтып жатқанымызды түсіну үшін, ол әрбір сөздің шығу түбірін білуі қажет. Олай болмаған жағдайда, «келу» және «келдім» таңбалауыштары (токен) мүлдем басқаша қабылданады. NLP-де лемматизация процессі деп атайды [131].

1.*Тоқтау сөздер (стоп -слова)* сөздің мағынасын табу және сүзгіден өткізу. Қазақ тілінде көмекші сөздер тізбегі көптеп кездеседі, мысалы, «немесе», «сондай -ақ», «егер». Мәтінді статистикалық анықтау барысында бұл сөздер тізбегі басқаларға қарағанда жиі ұшырасатындықтан, өзіндік мәселе тудындатады. Сол себепті оларды тоқтау сөзі ретінде белгілеп, сөйлем немесе пікір «арамшөптерден» арылады [141].

2. *Есімдік сөздер*. Қазақ лексиконында «оған», «әлдекім», «мұның», «анабір жерде» деген сөз тіркестерінен тұратын есімдіктер бар екені анық. Мысалы, «Дүйсенбі күні мен *Оралда* қаласына бардым. *Бұл жерде* тұратын халықтың саны басқа қалалармен салыстырмалы түрде аз.» Яғни, назар аударатын болсақ «Орал» сөзін «бұл жер» деген есімдікпен ауыстырылып айтылды [5].

Адам контексттің мағынасына назар аудару арқылы бұл сөздердің басқа сөйлеммен байланасын оңай ажырата алады. Дегенмен, NLP моделі есімдіктің орналасу реті мен мағынасын ажырата алмайды, себебі ол бір уақыт аралығында бір сөйлемді ғана қарастыра алады.

3. Тәуелділіктерді талдау. Бұл алгоритмнің мақсаты ретінде – әр таңбалауышта бір ғана ата – анадан яғни түбірден тұратын таңбалауыш. Түбірдің негізі ретінде етістікті алуға болады. Сондай-ақ екі сөздің арасындағы байланыс түрін орнату керек [99]. (Сурет 4.14)

Сурет 4.14- сөздерден сөйлем векторын құрастыру

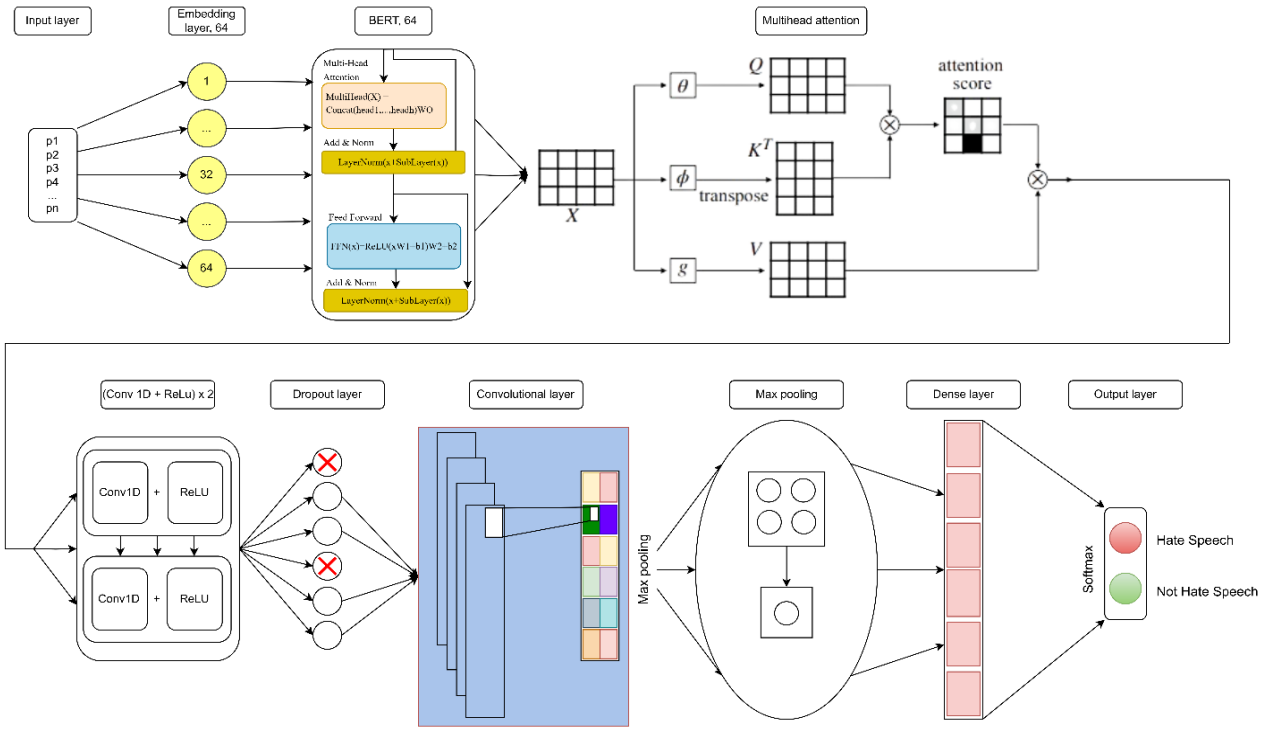
Мысалы ретінде, «Алматы» сөзін қарастырсақ, «Алматы» мен «қаласы» сөзін бірге алсақ, Алматының қала екенін білуге болады, ары қарай векторға сүйенсек оның астана екенін білуге болады, одан ары қарай сызықтармен жүретін болсақ, мәдени астана екеннін білуге болады. Сызық бойымен жүріп отырсақ, «Алматы қаласы Қазақстанның мәдени астанасы» екенін білуге болады.

4. *Өңделген мәтінді векторлық түрге аудару.* Бұл қадамда сөздің векторлық көрінісін құруғакүш салады. Осылайша, бір контексте қолданылатын өзара сөздің ұқсас векторлары қалыптасады [99].

Белгілі бір мәселені шешу үшін кейбір қадамдарды алып тастауға немесе жаңа қадам қосуға болады. Дегенмен, бұл конвейер NLP-тен практикалық мән алуға мүмкіндік беретін барлық типтік қадамдар мен тәсілдерді қамтиды[5].

4.4 Bert пен зейін механизімі бар гибридті нейрондық желіні ұсыну

Зерттеу нәтижелері талданғаннан кейін Bert (Bidirectional Encoder Representations from Transformers) моделі мен зейін (attention) механизмін пайдаланатын қысқа мерзімді және ұзақ мерзімді жадының сурет 4.15- те көрсетілген нейрондық желі моделі құрылды. Бұл модель ең басқа модельдермен салыстырмалы түрде жақсы нәтижелер берді. Мұнда өңделмеген мәтіндік деректерді енгізуден басталып, мәтіннің ғадауат тілді сөздер класына жататынын немесе жатпайтынын болжаумен аяқталатын бүкіл процестің жұмыс архитектурасы берілген.



Сурет 4.15 - Bert пен зейін механизімі бар гибридті нейрондық желі

Жоғарыда ұсынылған 4.15-суретте мәтіндік енгізулер мен конволюциялық нейрондық мүмкіндіктерді шығаруды біріктіру арқылы мәтін мазмұнындағы ғадауат тілді сөздерді нақты анықтауға арналған ұсынылған гибридті нейрондық желінің архитектурасы көрсетілген.

Ұсынылған модель әрбір pip_ipi таңбалауышы 64 өлшемді ендіру қабатын пайдаланып тығыз векторға түрленетін, таңбаланған мәтін реттілігін сипаттайтын кіріс қабатымен басталады. Енгізу қабаттары көп басты зейін (multi head attention) механизмін пайдаланатын BERT негізіндегі

трансформаторға беріледі. Бұл зейін механизмі модельге сұрау, кілт және мән матрицалары бойынша зейін ұпайларын есептеу арқылы сөздер арасындағы контекстік қатынастарды шығаруға мүмкіндік береді, осылайша бүкіл таңбалау тізбегі бойынша семантикалық көріністі жақсартады.

Зейін механизімі деңгейінің шығысы қалыпқа келтіріледі және кейіннен градиент ағынын сақтау және жаттығу тұрақтылығын жақсарту үшін одан әрі деңгейді қалыпқа келтіру арқылы алға жіберілетін позициялық желіге беріледі. Содан кейін алынған мәтіндік кіріс қабаттары ғадауат тілді сөздермен байланысты ерекше фразалар мен тілдік көрсеткіштерді анықтау үшін қажет реттілікпен жергілікті n-грамм үлгілерін түсіретін ReLU белсендіру функцияларын пайдаланып екі дәйекті бір өлшемді конвульстік қабаттар (Conv1D) арқылы өңделеді.

Шамадан тыс орнатуды жеңілдету үшін конвультациядан кейін түсіру қабаты, одан кейін мүмкіндіктер картасын жақсартатын қосымша конвульсиялық модуль қосылады. Максималды біріктіру ең маңызды мүмкіндіктерді сақтай отырып, көріністі кішірейту үшін пайдаланылады. Содан кейін мүмкіндіктер тураланады және толығымен қосылған қалың қабат арқылы жіберіледі. Соңғы санаттарға екілік санаттар үшін ықтималдықтарды жасайтын softmax қабаты арқылы қол жеткізіледі: ғадауат тілді сөздер және ғадауат тілді емес сөздер.

Ұсынылған тәсіл мәтіндегі ғаламдық семантикалық қатынастарды да, жергілікті кемсітушілік ерекшеліктерді де тиімді түсіре отырып, трансформаторға негізделген контекстуализация мен конволюциялық үлгіні танудың артықшылықтарын біріктіреді, осылайша ғадауат тілді сөзді анықтауға сенімді негіз береді. Бұл гибридті архитектура зейінге негізделген кодтауды конволюционды қабаттармен біріктіру модельдің әртүрлі және шулы пайдаланушы жасаған кірістерге жалпылау мүмкіндігін қалай жақсартатынын, осылайша жіктеу дәлдігін жақсартатынын көрсетеді.

1. *Кіріс қабаттары (Input layer).* Енгізу қабаты ғадауат тілді анықтаудың ұсынылған үлгісінің негізгі құрамдас бөлігі ретінде қызмет етеді, ол өңделмеген мәтіндік деректерді нейрондық архитектурада кейіннен өңдеуге жарамды сандық пішімге түрлендіруге жауапты. Атап айтқанда, әрбір кіріс сөйлем алдымен дискретті элементтер тізбегіне белгіленеді: $X = \{p_1, p_2, ..., p_n\}$, әрбір таңбалауыш pip_ipi бүтін санмен ұсынылған, ол бұрыннан анықталған сөздіктегі i-ші сөздің индексі болып табылады. Бұл сандық көріністің көмегімен модель енгізу тәртібін сақтай отырып, мәтіндік деректерді жылдам кодтай алады, яғни nnn ретінің ұзындығы оқыту және қорытынды жасау кезінде кірістің барлық топтамаларының бірдей өлшемге ие екендігіне көз жеткізу үшін орнатылады немесе динамикалық түрде толтырылады.

Бұл қабаттың негізгі жұмысы енгізу деректерін келесі қадамдарға дайындау болып табылады. Табиғи тілді өңдеуге (NLP) келетін болсақ, кірістің түпнұсқалық көрсетілімінің сапасы оқудың кейінірек қаншалықты жақсы өтетініне үлкен әсер етеді. Осылайша, кіріс қабаты өңделмеген мәтінді үйренген көріністермен байланыстырады. Бұл келесі ендіру және

трансформаторлық қабаттарға маңызды семантикалық және синтаксистік үлгілерді таңдауға мүмкіндік береді. Бұл қадам модель әртүрлі сөйлем ұзындықтары мен тіл үлгілерін өңдей алатынына көз жеткізеді. Бұл әсіресе әлеуметтік медиа деректері үшін өте маңызды, мұнда өшпенділік сөздері әртүрлі формалар мен сөз тіркестерінде болуы мүмкін. Кіріс деңгейі дискретті таңбалауыштарды құрылымдық реттілікке айналдыру арқылы контекстке негізделген ендірулерді және зейін механизмдерді пайдалануды жеңілдетеді.

Бұл үлгінің енгізу қабаты өңделмеген мәтінді әрі қарай үйрену және ендіру үшін пайдалануға болатын пішімге айналдыруға жауапты (4.1):

$$X = \{p_1, p_2, ..., p_n\}, \qquad (4.1)$$

мұндағы: X — бұл жалпы сөйлем немесе мәтін, яғни модельге кіріс болатын дерек; p_1, p_2, ..., p_n — сөйлемдегі сөздердің немесе токендердің векторлық көрсетілімдері (embedding).

2. *64 өлшемді ендіру қабаты (Embedding Layer)*. Енгізу қабаты әрбір дискретті енгізу таңбалауышын сөздердің семантикалық және синтаксистік қасиеттерін қамтитын үздіксіз, төмен өлшемді векторлық кеңістікке салыстыратын өшпенділік сөзін анықтау архитектурасының маңызды құрамдас бөлігі болып табылады. Енгізу функциясы әрбір таңбалауыш индексін $p_i \in Z$ кіріс тізбегінен тығыз векторға $e_i \in R^{64}$ түрлендіреді, 64 енгізу көлемі мен есептеу тиімділігін оңтайландыру үшін таңдалған ендіру өлшемін білдіреді. Трансформация $E \in R^{n \times 64}$ матрицаны береді, n реттілік ұзындығын білдіреді (4.2).

$$e_i = Embedding(p_i) \in R^{64}$$
$$E = [e_1, e_2, ..., e_n] \in R^{n \times 64}, \qquad (4.2)$$

Бұл қабат мағыналық жағынан ұқсас сөздердің кірістіру кеңістігінде тығыз орналасуын қамтамасыз ете отырып, сөздердің таратылған көрінісін жеткізуге бағытталған. Бұл реттеу модельдің лингвистикалық вариациялар мен өшпенділік сөздерінің қайталанатын формалары бойынша жалпылау мүмкіндігін жеңілдетеді.

Бұл тәсіл оқудың тиімділігін арттырады және сөздер арасындағы жасырын қатынастарды тиімдірек көрсетеді. Сонымен қатар, кірістіру салмақтарын кездейсоқ түрде инициализациялауға және жаттығу кезінде реттеуге немесе Word2Vec немесе GloVe сияқты сыртқы корпустармен алдын ала жаттығуға болады. Трансформаторға негізделген архитектураларда бұл ендірулер әдетте BERT сияқты контекстік қабаттармен бірге дәл реттелген. Бұл модельге қоршаған контекстке жауап ретінде сөз көріністерін реттеуге мүмкіндік береді, бұл ғадауат тілді сөздердің контекстке тәуелді формаларын дәл түсіндіру үшін маңызды. Енгізу қабаты кейінгі қабаттардағы зейін механизмдері мен конволюционды сүзгілер арқылы тиімді өңдеу үшін маңызды болып табылатын негізгі тілдік сипаттамалары бар өңделмеген таңбалауыштық

тізбегін жақсарта отырып, модельдің семантикалық интерфейсі ретінде қызмет етеді.

3. *BERT трансформері.* Ұсынылған үлгі архитектурасындағы BERT кодтау бөлігі зейін механизмдері және кейінгі алға қарай түрлендірулер арқылы қажетті контекстік ақпаратты алу үшін өте маңызды. Бұл компонент енгізу тізбегі ішіндегі қысқа және ұзақ ауқымды тәуелділіктерді модельдейді, бұл жүйеге әрбір сөзді қоршаған контекстке қатысты түсіндіруге мүмкіндік береді. Бұл мүмкіндік көбінесе нюанстық контекстік белгілерге тәуелді ғадауат тілді сөздерді дәл анықтау үшін өте маңызды. Енгізу матрицасы берілген $E \in R^{n \times d}$, көп бағытты ішкі зейін механизмі алдымен үш үйренген проекцияны есептейді: сұраулар $Q=EW^Q$, кілттер $K=EW^K$ және маңыздылығы $V=EW^V$, бұл жерде $W^Q, W^K, W^V \in R^{d \times d_k}$ – жатықтыруға бейім салмақ матрицалары, Содан кейін зейін механизімі нәтижесі келесі формулас арқылы шығарылады: $Attention(Q,K,V) = soft\max\left(\frac{QK^T}{\sqrt{d_k}}\right)V$, әр токен өзіне қатысты барлық басқа токендерге назар аудара алады, олардың маңыздылығына байланысты. Бірнеше "бас" (multi-head) арқылы алынған зейін сигналдары біріктіріліп, сызықтық түрде түрленеді, сосын қалдық байланыс (residual connection) және қабаттық қалыптандыру (layer normalization) арқылы өтеді. Содан кейін ReLU белсендіруімен екі толық қосылған қабаттан тұратын алдыңғы байланысты нейрондық желіге (feed-forward network) түседі, яғни $FFN(x) = \mathrm{Re}\,LU(xW_1 + b_1)W_2 + b_2$ анықталады.

Бұл екі деңгейлі механизм (BERT +зейін механизімі) модельге мәтіндерді күрделі және иерархиялық жолмен түсінуге көмектеседі. Бұл механизм бір сөздер қолданылса да,ғадауат тілді сөздер немесе ғадауат тілді сөз емес ажыратуға мүмкіндік береді. Негізгі артықшылығы - сөздердің контекстіне назар аудару.

BERT негізіндегі кодтауыш модельдің ішкі құрылымына енгізілгендіктен, ол алдын ала дайындалған контекстік деректердің және динамикалық белгілерін шығару мүмкіндігінің артықшылығын пайдаланады. Бұл әдіс модельдің беріктігін арттырады және оның ғадауат тілді сөздерді анықтауда мәтіндердің бейімделу қабілетін жақсартады.

Self-Attention (өзіндік зейін механизмі) модельге әрбір сөзді өңдегенде, сол сөзге қатысты маңызды бөліктерге ғана назар аударуға мүмкіндік береді.

Ал Multi-head Attention — әртүрлі типтегі мағыналық және грамматикалық байланыстарды бір уақытта байқауға көмектеседі.

Бұл екі негізгі компоненттен тұрады:

a. Multi-head Self-Attention

Әрбір токен үшін $x \in R^{64}$ есептейміз (4.3):

$$Q = XW^Q,\ K = XW^K,\ V = XW^V$$

$$Attention(Q,K,V) = soft\max\left(\frac{QK^T}{\sqrt{d_k}}\right)V\ , \qquad (4.3)$$

Содан кейін зейін механизмдері біріктіріледі (Multi-head барлық бастардың шығыстарын біріктіреді) (4.4):

$$MultiHead(X) = Concat(head_1,...,head_h)W^O, \qquad (4.4)$$

мұндағы: W^Q, W^K, W^V, W^O оқытылған салмақ матрицалары.

b. Residual Connection + Layer Norm (4.5):

$$X' = LayerNorm(X + LayerNorm(X + MultiHead(X))), \qquad (4.5)$$

Бұл қалдық қосылымды қосады және қабатты қалыпқа келтіруді қолданады.

c. Feed Forward Network (FFN) трансформатор архитектурасының негізгі компоненттерінің бірі болып табылады. Ол әрбір позициямен (әр таңбалауыштағы вектор) жеке жұмыс істейді және таңбалауыштың берілген векторлық көрінісін тереңірек өңдейді, пайдалы және жоғары деңгейлі функцияларды жасайды.

Алғашқы қабат кірістен жоғары өлшемді (мысалы, кеңейтілген) кеңістікке өтеді:

$$FFN(x) = \mathrm{Re}\,LU(xW_1 + b_1)W_2 + b_2, \qquad (4.6)$$

мұндағы: x — кіріс векторы;

W_1— бірінші қабаттың салмақ матрицасы;

b_1 — бірінші қабаттың ығысу (bias) векторы;

ReLU— белсендіру функциясы

Екінші қабат қайтадан өлшемді тарылтып, модельдің келесі қабаттарымен үйлесімді ету үшін (4.7):

$$X'' = LayerNorm(X' + FFN(X')), \qquad (4.7)$$

мұндағы: W_2 — екінші қабаттың салмақ матрицасы;

b_2— екінші қабаттың ығысу (bias) векторы.

Feed Forward Network (FFN) трансформатор архитектурасында әрбір токеннің векторлық көрінісі екі қабатты толық қосылған желі арқылы жеке өңделеді. Бұл процесс ReLU белсендіру функциясын пайдаланып күрделі иерархиялық мүмкіндіктерді үйренуге мүмкіндік береді, осылайша модельдің жалпы өнімділігін, бейімделгіштігін және жалпы үйрену мүмкіндігін жақсартады.

4. 1D конволюциялық қабаттар. Бір өлшемді конволюционды қабаттар BERT сияқты үлгіде алынған мәтіндік векторлық көріністерден (яғни әрбір сөздің контекстік көрінісі) бірегей жергілікті мүмкіндіктерді шығару үшін пайдаланылады. Олар мәтіндегі жергілікті n-граммаларды, яғни сөздер мен шағын сөз тіркестерінің арасындағы үлгілер мен тізбектерді тауып сипаттайды. Бұл үлгілер жек көрушілікті тануға көмектесетін маңызды синтаксистік және семантикалық анықтамаларға айналады. BERT шығыс матрицасы берілген $H \in R^{n \times d}$, әрбір конволюциялық сүзгі $f \in R^{k \times d}$ ені k қадамға негізделген

сырғымалы терезені пайдаланып, мүмкіндіктер картасын жасайтын реттілік бойынша қолданылады (4.8-4.9):

$$h_1 = \mathrm{Re}\,LU\left(Conv1D_1\left(X''\right)\right), \qquad (4.8)$$

$$h_2 = \mathrm{Re}\,LU\left(Conv1D_2\left(h_1\right)\right), \qquad (4.9)$$

мұндағы: b - қиғаш термин және ReLU - заттарды түзу сызықта жұмыс істемейтін белсендіру функциясы.

Модель сүзгі тереңдігі жоғары бірнеше Conv1D қабаттарын жинақтау арқылы кірістен жоғары деңгейлі лингвистикалық үлгілерді біртіндеп алып тастайды. Бұл иерархиялық мүмкіндікті үйрену әсіресе қайталанатын сөз құрылымдары немесе жай ғана зейін қою қиын болатын сөйлем құрылымдары арқылы көрсетілуі мүмкін өшпенділік сөйлеудің нәзік немесе болжамды түрлерін табу үшін өте пайдалы. Конволюционды қабаттар сонымен қатар бүкіл дәйектілік бойынша параллель жұмыс істейді, бұл ақпараттың ретін сақтай отырып, уақыт пен жадты үнемдейді, бұл табиғи тілді пайдаланатын жұмыстар үшін маңызды. Бұл ғадауат тілді сөздерді анықтауды дәлдірек етеді.

5. Dropout Layer. Dropout Layer – кейбір нейрондарды кездейсоқ «өшіру», олардың белсенділігін уақытша тоқтату арқылы нейрондық желіні шамадан тыс оқыту мәселесін азайтуға көмектесетін техникалық әдіс. 1D конволюционды қабаттардан кейін алдын ала анықталған ықтималдықпен нейрондық белсендірулердің бір бөлігін нөлге кездейсоқ орнатады (4.10).

$$h_{drop} = Dropout\left(h_2\right), \quad (4.10)$$

Бұл механизм белгілі бір нейрондарға немесе функция детекторларына тәуелділікті болдырмау арқылы желіні сенімдірек өкілдіктерді үйренуге мәжбүр етеді, осылайша оның көрінбейтін деректерге жалпылау қабілетін жақсартады. Мәтіндік мазмұн айтарлықтай тілдік вариативтілік пен сыныптық теңгерімсіздікті көрсетуі мүмкін өшпенділік сөздерін анықтау контекстінде сабақты тастау жиі үлгілерге немесе басым сыныптарға шамадан тыс бейімделу қаупін азайтуда маңызды рөл атқарады. Сонымен қатар, оқуды тастап кетуді пайдалану тығыз және конвульсиялық қабаттармен үйлескенде әсіресе тиімді, өйткені ол үйренілген мүмкіндіктердің желіде әртүрлі және артық болмайтынын қамтамасыз етеді. Қорытындылау кезінде оқуды тастап кету әдетте өшіріледі және оқу кезеңімен сәйкестікті сақтау үшін оқуды тастау көрсеткіші бойынша нәтиже масштабталады. Осы ықтималдық қабатын қосу арқылы ұсынылған архитектура оның шу мен шамадан тыс орнатуға төзімділігін арттырып қана қоймай, сонымен қатар әртүрлі мәтіндік домендерде өшпенділік сөздерін дәл анықтай алатын сенімдірек және түсіндірілетін үлгіге ықпал етеді.

6. Convolutional Feature Representation Layer (Visual Representation). Ұсынылған архитектураның визуалды көрсетілімінде бейнеленген Конволюциялық мүмкіндікті көрсету қабаты алдыңғы конволюциялық және түсіру қабаттарынан алынған жоғары деңгейлі мүмкіндіктерді одан әрі

нақтылауға арналған аралық абстракциялау кезеңі ретінде қызмет етеді. Бұл қабат кеңістіктік локализацияланған үлгілерді біріктіреді және кіріс мәтінінің семантикалық мәнін қамтитын дискриминативті және ықшам көрініске айналдырады.

Алдыңғы 1D конволюционды қабаттар негізгі n-gram мүмкіндіктерін және қысқа диапазондағы тәуелділіктерді түсіруге бағытталғанымен, бұл қосымша конволюциялық блок тереңірек деңгейде жұмыс істейді, өшпенділік сөзі мен өшпенділік білдірмейтін сөйлеу өрнектерін саралау үшін маңыздырақ абстрактілі, сыныпқа тән үлгілерді үйренеді. Құрылымдық жағынан бұл қабат сүзгі ені мен арна тереңдігі өзгеретін бір немесе бірнеше конвульстік операцияларды қамтуы мүмкін, бұл модельге реттілік бойынша көп масштабты мүмкіндіктерді біріктіруге және қысуға мүмкіндік береді. Көрнекі көрініс иерархиялық жинақталған конвульсиялық дизайнды көрсетеді, бұл түсіру қабатының шығысы бірізді Conv1D түрлендірулері арқылы өтеді, олардың әрқайсысынан кейін ReLU сияқты сызықты емес белсендіру функциялары бар. Бұл процесс мәтін тізбегіндегі ең көрнекті және сыныпқа қатысты аймақтарды бөлектейтін ықшамдалған мүмкіндіктер картасын жасауға әкеледі. Бұл нақтыланған мүмкіндік карталары пайдаланушы жасаған мазмұндағы жиі қорлайтын тілді сипаттайтын нәзік тілдік белгілерді, идиоматикалық өрнектерді немесе контекстік инверсияларды өңдеуде ерекше құнды. Сонымен қатар, конволюциялық жолды тереңдете отырып, модель шағын синтаксистік вариацияларға инвариантты сенімді иерархиялық көріністерді үйрену үшін жақсы жабдықталған. Осы мамандандырылған конволюциялық мүмкіндікті көрсету қабатын қосу осылайша, модельдің гетерогенді мәтін енгізулері бойынша жалпылау мүмкіндігін арттырады және төменгі ағынды біріктіру және тығыз қабаттар үшін бай, семантикалық ақпараттандырылған мүмкіндіктер жиынтығын қамтамасыз ету арқылы жалпы жіктеу дәлдігін күшейтеді.

Бұл қосымша жинақталған Conv қабаттарының көрнекі қысқаша мазмұны. Олар әрі қарай токендік векторлардан жоғары деңгейлі кеңістіктік мүмкіндіктерді шығарады.

7. Max Pooling. Ұсынылған ғадауат тілді сөздерді анықтау архитектурасындағы максималды біріктіру қабаты алдыңғы конвульсиялық мүмкіндік карталарындағы ең маңызды мүмкіндіктерді таңдамалы түрде сақтайтын өлшемді азайту механизмі ретінде жұмыс істейді. Уақыт өлшемі бойынша максимум біріктіру әрекетін қолдану арқылы модель айнымалы ұзындық ретін бекітілген өлшемді көрініске қысады, осылайша маңызды ақпаратты сақтай отырып, есептеу күрделілігін азайтады. Берілген енгізу мүмкіндіктері картасы үшін $h \in R^{n \times d}$ мұндағы nnn реттілік ұзындығын және d мүмкіндік арналарының санын білдіреді, максималды біріктіру әрбір мүмкіндік арнасы бойынша алдын ала анықталған терезе өлшемі ішіндегі ең үлкен мәнді есептейді, нәтижесінде біріктірілген көрініс пайда болады, $h_{pool} \in R^{m \times d}$ мұнда $m<n$ болып табылады. Кеңістік өлшемдерін азайту және ең көрнекті ерекшеліктерді сақтау үшін максималды жинақтау қабаты қолданылады (4.11):

$$h_{pool} = MaxPool(h_{conv}), \quad (4.11)$$

Бұл әрекет ең көрнекті белсендірулердің — ғадауат тілді класстарға тән үлгілердің көрсеткіші — келесі қабаттарға таратылуын қамтамасыз етеді. Табиғи тілді өңдеу контекстінде конвульсиялық нәтижелер бойынша максималды біріктіру маңызды емес немесе артық мүмкіндіктерді алып тастай отырып, негізгі лексикалық және фразалық белгілерді түсіруде тиімді екенін дәлелдеді. Бұл архитектурада максималды біріктіру қабаты кеңістікте бөлінген мүмкіндіктерден тығыз қабат жіктелуі үшін қолайлы неғұрлым ықшам, векторланған пішінге өтуді қамтамасыз етеді. Оған қоса, максимум біріктіру аудармалық инвариантты лық дәрежесін енгізеді, бұл модельді сөйлемдегі жек көретін сөз тіркестерінің позициялық өзгермелілігіне сенімдірек етеді. Күрделі, жоғары өлшемді мүмкіндік карталарын маңызды жиынтық статистикаға айналдыру арқылы максималды біріктіру қабаты әртүрлі және құрылымдалмаған мәтіндік кірістердегі өшпенділік сөздерін анықтаудағы модельдің тиімділігі мен тиімділігіне айтарлықтай үлес қосады.

8. Dense Layer. Ұсынылған ғадауат тілді анықтау архитектурасындағы тығыз қабат жоғары деңгейлі шығарылған мүмкіндіктерді жіктеуге қолайлы сызықты түрде бөлінетін кеңістікке түрлендіру үшін маңызды құрамдас ретінде қызмет етеді. Конволюциялық жолдың ең көрнекті ерекшеліктерін конденсациялайтын максималды біріктіру операциясынан кейін шығыс бір өлшемді векторға тегістеледі және толығымен қосылған нейрондық қабат арқылы өтеді. Математикалық түрде біріктірілген мүмкіндік векторы берілген $h \in R^d$ тығыз қабат сызықтық түрлендіруді, одан кейін сызықты емес белсендіру функциясын қолданады, әдетте ReLU, шығыс келесідей анықталады:

$$h_{dense} = \sigma(h_{pool} W + b), \quad (4.12)$$

мұндағы: $W \in R^{d \times k}$ және $b \in R^k$ үйренуге болатын салмақтар мен ауытқулар, ал k жасырын бірліктердің санын білдіреді.

Бұл деңгей модельге абстракцияланған мүмкіндіктер арасындағы күрделі өзара әрекеттесулерді үйренуге мүмкіндік береді, сыныптың бөліну мүмкіндігін жақсартатын пішінде көрсетілімдерді нақтылайды. Тілдік сигналдар контекстке тәуелді құрылымдардың ішінде нәзік немесе ендірілген болуы мүмкін өшпенділік сөздерін анықтау контекстінде тығыз қабат маңызды емес шуды басу кезінде кемсітушілік үлгілерді күшейте отырып, көптеген мүмкіндік өлшемдерінен ақпаратты біріктіру және синтездеу үшін қызмет етеді. Бұдан басқа, толық қосылған қабат ретінде ғадауат тілді сөздер және ғадауат тілді емес сөзге тән нюансты семантикалық вариацияларды модельдеуге қабілетті жоғары сыйымдылықты функция жуықтаушысы ретінде әрекет ететін, конволюциялық мүмкіндіктерді шығару кезеңдерін соңғы шығыс қабатымен байланыстырады. Модельдің күрделілігі мен жалпылауды теңестірудегі тығыз қабаттың рөлі ерекше маңызды, бұл жүйенің әртүрлі мәтіндік кірістер мен ғадауат тілді сөздердің көріністері бойынша дәл және сенімді болжамдар жасау қабілетіне тікелей ықпал етеді.

9. Output Layer (Softmax). Ұсынылған ғадауат тілді сөздерді анықтау архитектурасының шығыс қабаты тығыз қабаттан жоғары өлшемді мүмкіндік векторын мақсатты сыныптар бойынша ықтималдық үлестіріміне салыстыру арқылы соңғы жіктеуді орындау үшін softmax белсендіру функциясын пайдаланады. Түрлендірілген $h_{dense} \in R^k$, мүмкіндік векторын ескере отырып, мұндағы k – соңғы тығыз қабаттағы бірліктердің саны, шығыс қабаты сынып ықтималдығын шығару үшін softmax функциясынан кейін сызықтық түрлендіруді қолданады (4.13-4.15):

$$\hat{y} = Soft\max\left(h_{dense} W_o + b_o\right), \quad (4.13)$$

$$\hat{y}_0 = P\left(Not\ hate\ speech\right), \quad (4.14)$$

$$predicted\ Class = \arg\max\left(\hat{y}p\right), \quad (4.15)$$

Мұндағы $W_o \in R^{k \times C}$ және C=2 екілік классификация үшін (ғадауат тілді сөздер және ғадауат тіліді емес сөздер), softmax функциясы келесідей анықталады (4.16):

$$Soft\max\left(z_i\right) = \frac{e_{z_i}}{\sum_{j=1}^{C} e^{z_j}}, \quad (4.16)$$

Бұл деңгей әрбір сынып белгісіне сенімділік ұпайларын тағайындайды, бұл модельге болжамдардағы сенімділігін сандық бағалауға мүмкіндік береді. Болжалды белгі ретінде анықталатын ең жоғары ықтималдығы бар сынып таңдалады. Жек сөйлейтін сөздерді анықтау контекстінде мұндай ықтималдық нәтиже түсіндіру және шешім қабылдау үшін маңызды, әсіресе мазмұнды модерациялау немесе заңды есеп беру сияқты сезімтал қолданбаларда. Сонымен қатар, жаттығу кезінде softmax шығысы категориялық кросс-энтропия жоғалтуын пайдалануға мүмкіндік береді, бұл олардың сенімділік деңгейлеріне негізделген қате жіктеулерді тиімді жазалайды, модельді дәл және калибрленген болжамдар жасауға ынталандырады. Абстрактілі оқытылған мүмкіндіктерді түсіндірілетін сынып ықтималдығына түрлендіру арқылы softmax шығыс деңгейі қорлайтын және қорлайтын мәтіндік мазмұнды ажырату үшін бүкіл желінің үйренген дискриминациялық мүмкіндіктерін инкапсуляциялай отырып, модельдің соңғы шешім нүктесі ретінде қызмет етеді.

Сонымен, мәтіндегі ғадауат тілді сөздерді анықтауға арналған ұсынылған модель архитектурасы контекстік ендірулер мен конволюциялық нейрондық мүмкіндіктерді шығаруды біріктіреді. Модель 64 өлшемді кірістіру қабаты арқылы тығыз векторға түрленетін таңбаланған мәтін тізбегін қабылдайтын енгізу деңгейінен басталады. Бұл ендірмелер BERT негізіндегі трансформаторға беріледі, оның құрамына көп негізді зейін механизмі кіреді. Модель сөздер арасындағы контекстік тәуелділіктерді сұрау, кілт және мән матрицалары бойынша зейін өлшемдерін есептеу, токендер тізбегі бойынша семантикалық көрсетуді жақсарту арқылы үйренеді. Зейін механизімі деңгейінің шығысы қалыпқа келтіріліп, алға қарай позициялық желі арқылы беріледі. Алынған

мәтіндік ендірулер реттіліктегі n-граммдардың жергілікті үлгілерін түсіріп, ReLU белсендіру функциялары бар екі дәйекті бір өлшемді конвульстік қабаттар арқылы өңделеді. Соңғы классификацияға екілік классқа сәйкес ықтималдықтарды шығаратын softmax қабаты арқылы қол жеткізіледі: ғадауат тілді сөздер және ғадауат тілді емес сөздер қатары.

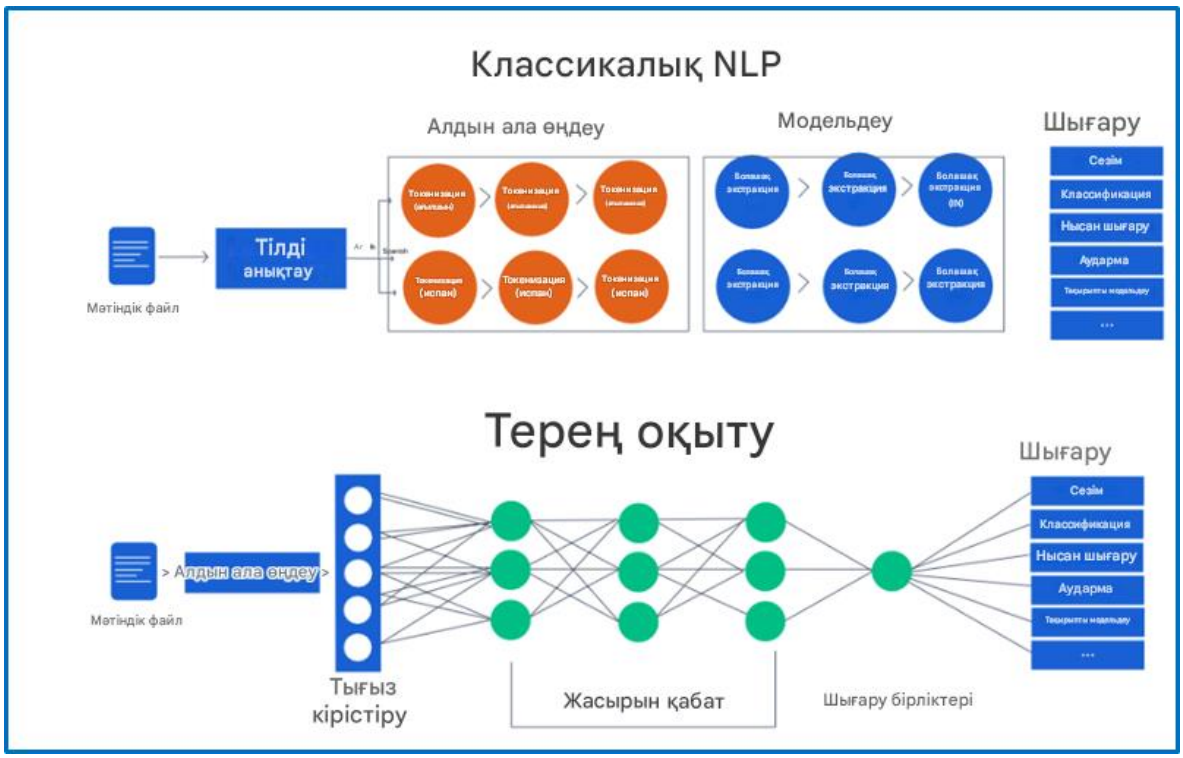
Ұсынылған үлгі архитектурасы контекстік ақпаратты түсіру және әрбір сөзді қоршаған контекст аясында түсіндіру үшін BERT кодтау бірлігін пайдаланады. Бұл жүйеге көбінесе контекстік белгілерге сүйенетін ғадауат тілді сөздерді дәл анықтауға мүмкіндік береді. Модель мәтінді енгізудің күрделі иерархиялық көріністерін үйрену үшін өзіне-өзі зейін механизмдерін, көп жақты зейінді және алға қарай түрлендірулерді пайдаланады. Архитектура алдын ала дайындалған контекстті модельдеудің және динамикалық мүмкіндіктерді шығарудың күшті жақтарын пайдаланады, оның беріктігін және әртүрлі және ғадауат тілді сөздерге жатпайтын мәтіндерді жалпылау мүмкіндігін арттырады.

Ұсынылған ғадауат тілді сөздерді анықтау архитектурасы конвульсиялық мүмкіндікті көрсету деңгейін, максималды біріктіру деңгейін, тығыз қабатты және шығыс деңгейін қамтиды. Конволюциялық мүмкіндікті көрсету деңгейі алдыңғы конволюциялық және түсіру қабаттарынан жоғары деңгейлі мүмкіндіктерді нақтылайды, кеңістіктік локализацияланған үлгілерді неғұрлым дискриминативті және ықшам көрініске айналдырады.

Максималды біріктіру қабаты алдыңғы конверсиялық мүмкіндік карталарындағы ең маңызды мүмкіндіктерді таңдамалы түрде сақтау арқылы есептеу күрделілігін азайтады. Тығыз қабат ерекше белгілерді жіктеуге қолайлы сызықты түрде бөлінетін кеңістікке айналдырады. Шығыс деңгейі жоғары өлшемді мүмкіндік векторын тығыз қабаттан мақсатты класстар бойынша ықтималдық үлестіріміне салыстыру арқылы соңғы жіктеуді орындау үшін softmax белсендіру функциясын пайдаланады. Softmax шығыс деңгейі модельдің ғадауат тілді сөздер және ғадауат тілді емес сөздер мазмұнын ажырату үшін барлық үйренген қабілетін қамтитын модельдің соңғы шешім нүктесі ретінде қызмет етеді.

4.5 Ғадауат сөздерді анықтау үлгісі

Бейәдеп тілді сөздерді автоматтандырылған анықтау әлеуметтік медиа платформаларында ғадауат тілді сөздердің таралуымен күресудің перспективалы стратегияларының бірі болып табылады. Кекесін мақсатта сөйлейтін сөздерді анықтау доменінде бірнеше деректер жиынтығы бар. Модельді жасау, жаттықтыру және сынау үшін Kaggle (Kaggle, 2020) ғадауат тілді деректер жинағы қолданылды. (Сурет 4.16)



Сурет 4.16 - Терең оқыту
(Сурет классикалық табиғи тілді өңдеу алгоритмдерін ұсыну мақсатында сызылған автордың жекелей өнімі)

Python (2022) NLP басталған кезде деректерді алдын ала өңдеуді ұсынады, өйткені талданатын деректер құрылымы жоқ болғандықтан жөндеу қажет. Табиғи тілді өңдеудегі бес негізгі тапсырма:

- ✓ классификация;
- ✓ сәйкестік;
- ✓ аударма;
- ✓ құрылымдық болжау;
- ✓ шешім қабылдаудың ретті процесі.

Таңдалған бағдарламалау тілі - Python бағдарламалау тілі. Бұл жерде Natural Language Toolkit (NLTK), numpy және Matplotlib сияқты тиісті кітапханаларды орнатудан басталды. Таңдау құралы - Anaconda жұмыс істейтін Jupyter ноутбугі. Жұмысты бастау үшін төменде көрсетілгендей қажетті кітапханалар сурет 4.17- да көрсетілгендей импортталды.

```
In [1]: import numpy as np
import pandas as pd
import re
import string
import warnings
import seaborn as sns
import matplotlib.pyplot as plt
from matplotlib import style
style.use('ggplot')
import nltk
from nltk.tokenize import word_tokenize
from nltk.stem import WordNetLemmatizer
from nltk.corpus import stopwords
stopwords=set(stopwords.words('english'))
from wordcloud import WordCloud
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix,ConfusionMatrixDisplay
```

Сурет 4.17 - Қажетті кітапханаларды импорттау

Деректер жинағы төмендегі сурет 4.18 көрсетілгендей панда кітапханасы арқылы жүктелді.

```
In [2]: tweet_df = pd.read_csv('hate_speech.csv')
tweet_df.head()
```

Out[2]:

	Unnamed: 0	count	hate_speech	offensive_language	neither	class	tweet
0	0	3	0	0	3	2	!!! RT @mayasolovely: As a woman you shouldn't...
1	1	3	0	3	0	1	!!!!! RT @mleew17: boy dats cold...tyga dwn ba...
2	2	3	0	3	0	1	!!!!!!! RT @UrKindOfBrand Dawg!!!! RT @80sbaby...
3	3	3	0	2	1	1	!!!!!!!! RT @C_G_Anderson: @viva_based she lo...
4	4	6	0	6	0	1	!!!!!!!!!!! RT @ShenikaRoberts: The shit you...

Сурет 4.18 - Деректер жиынын жүктеу

Алдын ала өңдеу.

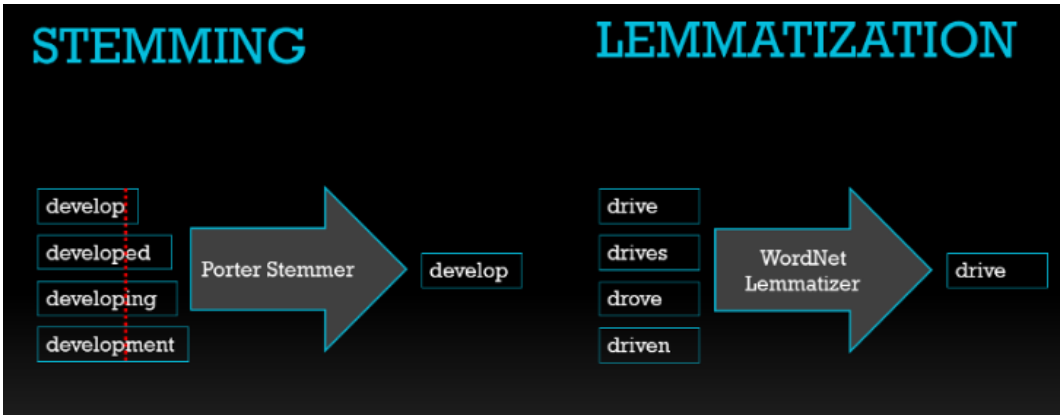
Деректерді модельдеуге дайындау үшін алдын ала өңдеу әрекеттерінің сериясы орындалды. Оларға біркелкі болу үшін барлық мәтіннің регистрін кішірейту, тыныс белгілері мен интервал сияқты жолдың қажетсіз бөліктерін алып тастау, қайталанатын және бос мәндерді жою кіреді [4]. (Сурет 4.19)

```
In [4]: # Lower case all the words of the tweet before any preprocessing
tweet_df['tweet'] = tweet_df['tweet'].str.lower()

# Removing punctuations present in the text
punctuations_list = string.punctuation
def remove_punctuations(tweet):
    temp = str.maketrans('', '', punctuations_list)
    return tweet.translate(temp)

In [5]: tweet_df['tweet']= tweet_df['tweet'].apply(lambda x: remove_punctuations(x))
tweet_df.head()

In [6]: tweet_df = tweet_df.drop_duplicates('tweet')
```

Сурет 4.19 - Алдын ала өңдеуден өткізу

Лемматизация алдын ала өңдеудің тағы бір маңызды қадамы болып табылады. Жоғарыдағы прагматикалық талдауда айтылғандай, лемматизация сөздер тобын олардың мағынасын бекіту үшін контексттендіреді. Бұл процесс қалыпқа келтіру деп аталады. Қалыпқа келтірудің тағы бір әдісі - стемминг.

Шектеу және лемматизация жағдайға байланысты бірге немесе дербес қолданылуы мүмкін. Айырмашылық олардың қалыпқа келу тәсілінде жатыр. Екеуі де түбір сөзді алуға ұмтылғанымен, стемпинг мұны танымал префикстер мен жұрнақтардың тізімін пайдаланып сөздердің басын немесе соңын кесу арқылы жасайды. Екінші жағынан, лемматизация мағынасы жақын сөздерді біріктіріп (сурет 4.20), түбір сөзді жасаудың бір мазмұнға біріктіреді.



Сурет 4.20 - Лемматизация және штрихтау

Ескерту – Әдебиет негізінде құралған [142]

NLTK кітапханасының WordNetLemmatizer сыныбын пайдалана отырып, төменде көрсетілгендей, деректер жиынындағы твиттерді лемматизациялауға (сурет 4.21) мүмкіндік бар.

```
In [7]: lemmatizer = WordNetLemmatizer()
        def lemmatizing(data):
            tweet = [lemmatizer.lemmatize(word) for word in data]
            return data

In [8]: tweet_df['tweet'] = tweet_df['tweet'].apply(lambda x: lemmatizing(x))
```

Сурет 4.21 - Лематизацияны қолдану

Деректер жиынын теңестіру

Жоғары және төмен үлгілеу деректер жиынын теңестіру үшін қолданылатын кейбір әдістерді теңестіру қажет болады. Теңгерімсіз деректер жинағы дегеніміз не және ол неге теңестірілуі керек?

Теңгерімсіз сыныптар деректер жиынының сыныптарында деректер нүктелерінің тең емес қатынасы болған кезде пайда болады. Бұл модельдің

дәлдігіне нұқсан келтіруі мүмкін. Кейде, егер оның деректер нүктелері айтарлықтай аз болса, азшылық класы толығымен еленбеуі мүмкін (GeeksforGeeks, 2021). Мұндай жағдайды болдырмау үшін жоғарыдан іріктеу жүзеге асырылады. Бұл азшылық класындағы үлгілерді кездейсоқ түрде көшіру төменде сурет 4.22 – де көрсетілген.

```
In [11]: class_2 = tweet_df[tweet_df['class'] == 2]
         class_1 = tweet_df[tweet_df['class'] == 1].sample(n=3500)
         class_0 = tweet_df[tweet_df['class'] == 0]

         balanced_df = pd.concat([class_0, class_0, class_0, class_1, class_2], axis=0)
```

Сурет 4.22 – Векторизаторлау

Деректер (Мәтін) талдауға дайындау үшін алдын ала өңдеу әрекеттерінің сериясынан өтуі керек. Осындай әрекеттердің бірі мәтінді сандық деректер ретінде көрсету болып табылады. Juputer жазу кітапшасынан алынған келесі кодты (сурет 4.23) қарастырамыз.

```
In [1]: import pandas as pd
        import numpy as np
        import matplotlib.pyplot as plt
        import seaborn as sns

In [2]: simple_train = ['call you tonight', 'Call me a cab', 'Please call me... PLEASE!'
```

Сурет 4.23 – Векторизациялау

Simple_train - бұл алгоритмге енгізуге болмайтын өңделмеген мәтіннің тізімі, сондықтан белгіленген өлшемі бар сандық мүмкіндік векторларына түрлендіру қажет.

Countvectorizer() — мәтін жолын таңбалауыш санау матрицасына түрлендіретін sklearn мүмкіндігін шығару функциясы (sklearn, n.d).
Функция жолды сәйкес сөздердің массивіне шашады. Одан кейін оқу деректерін төменде сурет 4.24- те көрсетілгендей құжаттық терминдік матрицаға түрлендіру орындалады.

```
In [3]: # import and instantiate CountVectorizer (with the default parameters)
        from sklearn.feature_extraction.text import CountVectorizer

        vect = CountVectorizer()

        # learn the 'vocabulary' of the training data (occurs in-place)
        vect.fit(simple_train)

        # examine the fitted vocabulary
        vect.get_feature_names_out()

Out[3]: array(['cab', 'call', 'me', 'please', 'tonight', 'you'], dtype=object)
```

Сурет 4.24 – Векторизациялау

Біз зерттеу жұмысындағы үлгіде CountVectorizer қолданбасын деректер жиынындағы қосымша мүмкіндіктер үшін қолдандым, содан кейін токенизаторды біздің оқу деректеріне орналастырдым. Бұл төменде көрсетілгендей оқу және валидация деректерін векторларға түрлендіру үшін өте маңызды. (Сурет 4.24.1)

```
In [15]: vect = TfidfVectorizer(ngram_range=(1,3)).fit(tweet_df['tweet'])

In [16]: feature_names = vect.get_feature_names()
         print("Number of features: {}\n".format(len(feature_names)))
         print("First 20 features: \n{}".format(feature_names[:20]))

Number of features: 455568

First 20 features:
['00', '007', '007 httptcoon5t60rmfb', '007beardownjedi', '007beardownjedi the', '007beardownjedi the afl', '007hertzrumble',
'007hertzrumble httptcoqyn1bc7mxs', '007hertzrumble httptcoqyn1bc7mxs via', '007mh', '007mh lilduval', '007mh lilduval damn',
'00jackie', '00jackie darknight420', '00jackie darknight420 allahthefairy', '00jackie no', '00jackie no wanna', '00sexilexi00',
'00sexilexi00 freeze', '00sexilexi00 freeze bitch']
```

Сурет 4.24.1 – Векторизациялау

Модельді құру және үйрету

TfidfVectorizer және Logistic Regression көмегімен үлгіні жасадым. Деректер жинағы алдымен 80% оқытуға және 20% тестілеуге бөлінген (сурет 4.25), оны оқыту үшін Логистикалық регрессия үлгісіне өткізген.

```
In [18]: x_train, x_test, y_train, y_test = train_test_split(X, Y, test_size=0.2, random_state=42)

In [19]: print("Size of x_train:", (x_train.shape))
         print("Size of y_train:", (y_train.shape))
         print("Size of x_test: ", (x_test.shape))
         print("Size of y_test: ", (y_test.shape))

Size of x_train: (22079, 455568)
Size of y_train: (22079,)
Size of x_test:  (5520, 455568)
Size of y_test:  (5520,)
```

Сурет 4.25 -Деректер жиынын бөлу

Модельді жетілдіру. Sklearn's GridSearchCV сыныбын пайдалана отырып, модельдің гиперпараметрлерін берілген модель үшін оңтайлы мәндерді анықтау үшін дәл реттеуге болады (Great Learning Team, 2022). Бұл модельдің өнімділігін жақсартуға мүмкіндік береді. Келесі кезекте GridSearchCV

қолданылды, оның өнімділігін төменде көрсетілгендей 93% дәлдікке дейін жақсартылды. (Сурет 4.26)

```
In [24]: param_grid = {'C':[100, 10, 1.0, 0.1, 0.01], 'solver' :['newton-cg', 'lbfgs','liblinear']}
         grid = GridSearchCV(LogisticRegression(), param_grid, cv = 5)
         grid.fit(x_train, y_train)
         print("Best Cross validation score: {:.2f}".format(grid.best_score_))
         print("Best parameters: ", grid.best_params_)

         Best Cross validation score: 0.92
         Best parameters:  {'C': 100, 'solver': 'lbfgs'}

In [25]: y_pred = grid.predict(x_test)

In [26]: logreg_acc = accuracy_score(y_pred, y_test)
         print("Test accuracy: {:.2f}%".format(logreg_acc*100))

         Test accuracy: 93.26%
```

Сурет 4.26 - үлгіні жетілдіру

Модельді сақтау. Ақырғы өнімді пайдалану керек, сондықтан модельді сақтау өте маңызды. Бұл жағдайда Python's Pickle кітапханасы пайдалы. (Сурет 4.27)

```
In [28]: import pickle
         # Saving model to pickle file
         with open("nlp_model.pkl", "wb") as file: # file is a variable for storing the newly created file, it can be anything.
             pickle.dump(logreg, file) # Dump function is used to write the object into the created file in byte format.
```

Сурет 4.27 - Үлгіні сақтау

Болжамдарды орындау. Келесі кезекте ғадауат тілді сөздерді анықтау алгоритмінің тиімділігін бағалау үшін екі сынақ кірісі зерттелді. «Мен сені өлтіремін» деген бастапқы кіріс агрессивті және ғадауат тілді мәлімдемені құрайды. Модель мұны өшпенділік және ғадауат тілді пікір деп дәл анықтады.

«Мен сені сүйемін» деген екінші кіріс жағымды және бейтарап мәлімдемені білдіреді. Модель бұл енгізуді ғадауат тіл ретінде дұрыс жіктеді. Бұл тұжырымдар модельдің ғадауат тілді және бейтарап тілід ажырату мүмкіндігін көрсетеді. Түсінікті болу үшін шығыс фразасында аздап түзету ұсынылады, себебі «ғадауат тілді сөз және ғадауат тілді сөз емес» белгісі сурет 4.28- де көрсетілген.

```
In [39]: inp="I will kill you"
         inp = vect.transform([inp]).toarray()

         if (logreg.predict(inp))==1:
             print("No hate speech and Offensive speech")

         else:
             print("Hate and offensive speech")

Hate and offensive speech
```

```
In [40]: inp="I love you"
         inp = vect.transform([inp]).toarray()

         if (logreg.predict(inp))==1:
             print("No hate speech and Offensive speech")

         else:
             print("Hate and offensive speech")

No hate speech and Offensive speech
```

Сурет 4.28 - Болжамдарды орындау

Орналастыру. Модельді оқыту және сынау аяқталғаннан кейін мен үлгіні қолдануға дайын етіп сақтадым. Пайдаланудың қарапайымдылығы үшін біз пайдаланушыларға болжау үшін фразаларды теруге мүмкіндік беретін пайдаланушы интерфейсін жасадық.

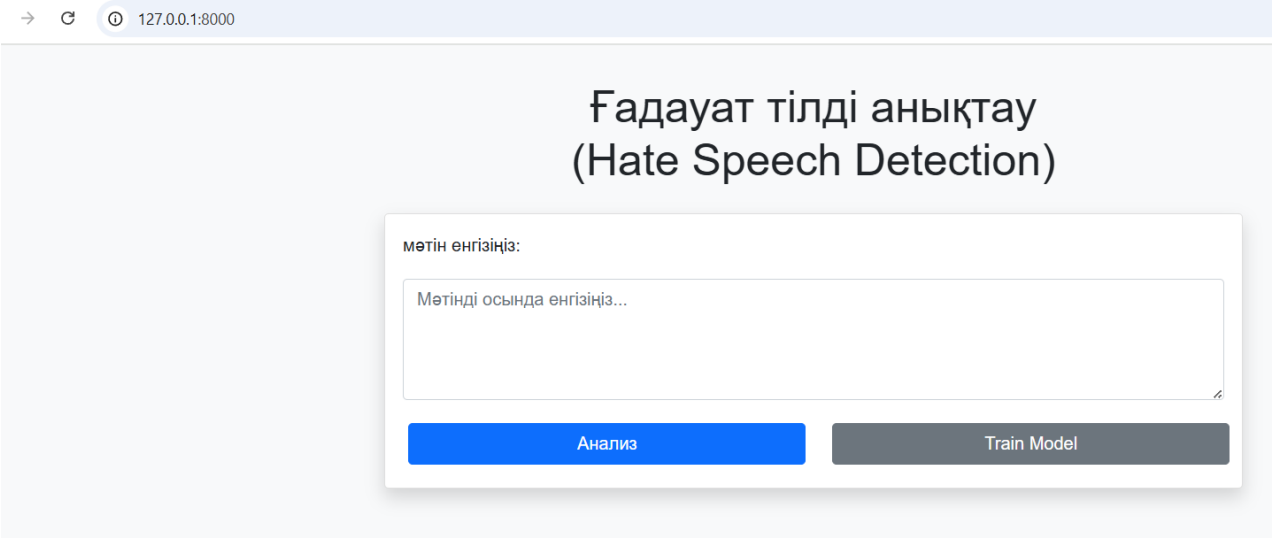
Бұл платформаның өзара әрекеттесуі үшін Python Flask құрылымы, HTML және Jinja коды арқылы жүзеге асырылды. App.py файлы POST әдісі арқылы HTML пішінінен фразаны алады және оны болжау үшін үлгіге жібермес бұрын өңдейді. Модель болжамды джинджа кодын пайдаланып index.html файлына қайтарады.

Жүктелген кезде app.py жергілікті хост мекенжайын жасайды: 127.0.0.1:8000, ол index.html файлын іске қосу үшін сурет 4.29- да көрсетілгендей веб -шолғышқа қойылады. Индекс бетінде болжау үшін сөз тіркесін теру мүмкіндігі бар.

Терілген сөз тіркесі болжау үшін үлгі арқылы іске асырылады және сәйкесінше жіктеледі.

```
 * Running on http://127.0.0.1:8000
Press CTRL+C to quit
 * Restarting with stat
2025-04-12 23:19:45.717758: I tensorflow/core/util/port.cc:153] oneDNN custom operations are on. You may see slightly
 different numerical results due to floating-point round-off errors from different computation orders. To turn them o
ff, set the environment variable `TF_ENABLE_ONEDNN_OPTS=0`.
2025-04-12 23:19:46.806078: I tensorflow/core/util/port.cc:153] oneDNN custom operations are on. You may see slightly
 different numerical results due to floating-point round-off errors from different computation orders. To turn them o
 x  set the environment variable `TF_ENABLE_ONEDNN_OPTS=0`.
```

Сурет 4.29 - Жоба қалтасына кіру және жобаны іске қосу



Сурет 4.30 – Интерфейс

Ұсынып отырған модель машиналық оқыту алгоритмдерімен салыстыра қарағанда барлық дерлік бағалау параметрлері бойынша жоғары нәтижеге жеткендігін байқауға болады және нәтижесінде ұсынылған модель әлеуметтік желілердегі қазақ тіліндегі ғадауат тілді сөздерден тұратын пікірлерді жоғары дәлдікпен анықтай алады деп қорытындылай аламыз.

Қорытындылай келе, қазақ тілі үшін декректер қорын жасақтауда әлеуметтік желілерге, онлайн контентке жүргізілген алдын – ала салыстырмалы зерттеулер жүргіізілген.

Сонымен қатар, деректерді талдаудан өткізу үшін Python кітапханалары және деректерді классификациялаудың қолмен жүргізілгендігі туралы баяндалған. Жасақталған деректер қорын пайдаланып ғадауат тілді сөздерді анықтауда машиналық оқыту, терең оқыту алгоритмдерін қолдану және мәтіндерді тануда жоғары дәлдікке жеткізетін модель ұсынған.

5 ЗЕРТТЕУДІҢ САЛЫСТЫРМАЛЫ НӘТИЖЕЛЕРІ МЕН ТАЛДАУЛАР

Бұл бөлімде ұсынылған гибридті модельдің әртүрлі веб-платформалар арасында ғадауат тілді сөздерді анықтаудағы тиімділігін бағалауды сипаттайды. Бұл модельдің ғадауат тілді сөздерді және бейтарап тілді нақты анықтаудағы өнімділігін, оның көптеген тілдік параметрлердегі әмбебаптығын және ғадауат тілді сөздердің ерекшеліктерін ескере отырып нақты нәтижеге жету көрсеткіштерін бағалайды.

BERT архитектурасын зейін архитектурасымен біріктіретін ұсынылған модель машиналық оқытудың дәстүрлі әдістеріне қарағанда семантикалық тілді түсінуде жақсы көрсеткіштер көрсетті.

Зерттеудің маңызды құрамдас бөлігі көптілді деректер жиынтығы бойынша модельдің тиімділігін бағалау болып табылады. Нәтижелер ұсынылған әдістің көптеген тілдерде тиімділігін сақтағанын көрсетті, бұл модель көптілді мазмұнды модерациялау орталарында қолдануға жарамды екенін көрсетеді. Бұл жаңалық трансформаторға негізделген дизайнның морфологиялық күрделіліктерді, сөз тіркестерін және мәдениет жағынан бір – бірінен ерекшеленетін тілдің әртүрлілігіне бейімделуін көрсетеді.

Нәтижелер ұсынылған парадигманың ғылыми жаңалығы мен практикалық маңыздылығын растайды. Олардың көптеген тілдік және құрылымдық параметрлердегі ғадауат тілді сөздерді дәл анықтаудағы тиімділігі олардың нақты уақыттағы интернет бақылау жүйелеріне қосылуға сәйкестігін көрсетеді. Нәтижелер тілді түсінуді жақсарту және әлеуметтік жауапты AI қолданбаларын ілгерілету үшін терең контекстік кірістірулерді зейін механизмдерімен біріктірудің маңыздылығын көрсетеді.

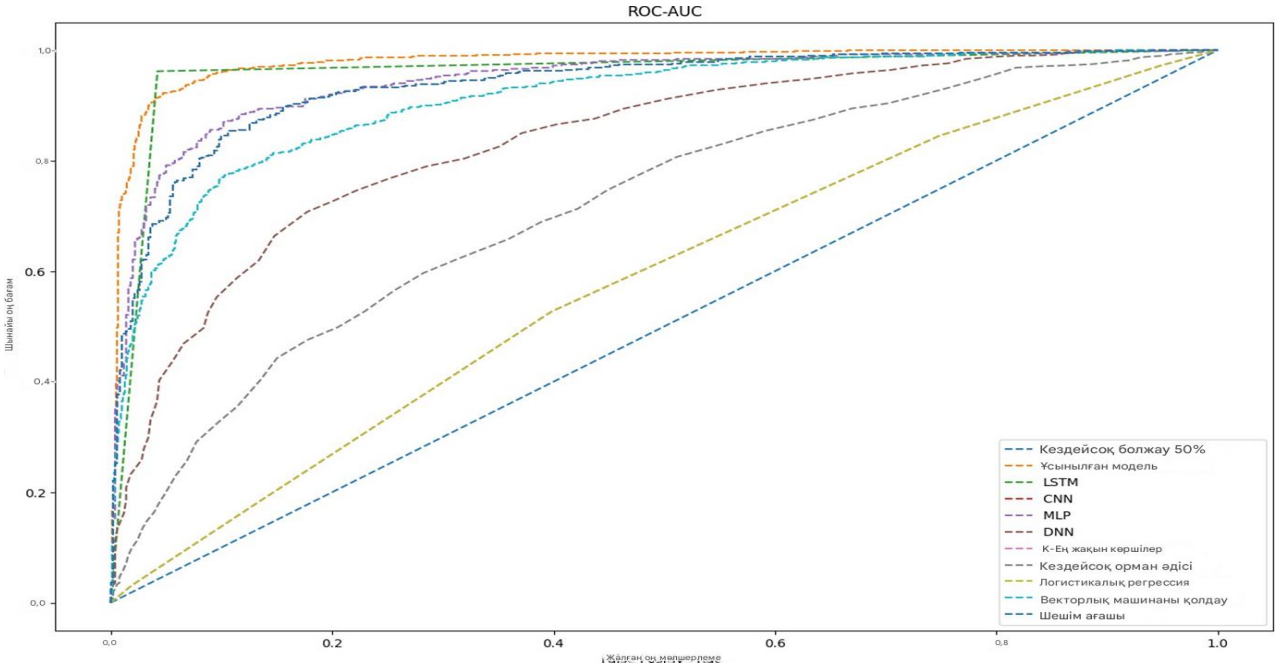
5.1 ROC қисығының салыстырмалы нәтижелері

Келесі кезекте классикалық статистикалық әдістер мен заманауи терең оқыту алгоритмдері мен ұсынылған модельдің нәтижелерін қамти отырып, жіктеу үлгілері үшін ROC-AUC (Қабылдағыштың жұмыс сипаттамасы – қисық астындағы аймақ) салыстырмалы талдауын ұсынады. ROC қисығы шынайы оң жылдамдықты (сезімталдық) жалған оң жылдамдықпен (1 – ерекшелік) көрсетеді, ал AUC әрбір модельдің жалпы өнімділігін сандық түрде көрсетеді.

Бұл зерттеу ROC-AUC талдауын қолдана отырып, бірнеше классикалық және заманауи жіктеу алгоритмдерінің салыстырмалы бағасын ұсынады. Модельдер логистикалық регрессия және шешім ағаштары сияқты статистикалық алгоритмдерден бастап қазіргі заманғы терең оқыту модельдеріне дейін, соның ішінде конволюционды нейрондық желілер (CNN), ұзақ қысқа мерзімді жады (LSTM) желілері және толық қосылған терең нейрондық желілер. Сонымен қатар, жаңадан ұсынылған модель оның жіктеу мүмкіндігін салыстыру үшін осы белгіленген әдістермен қатар бағаланады.

Бұл талдаудың мақсаты графикалық визуализациямен және сандық интерпретациямен қамтамасыз етілген екілік класстарды ажыратуда әрбір

модельдің қалай орындайтыны туралы нақты, сандық түсінік беру болып табылады. ROC қисықтары мен AUC мәндерін талдау арқылы зерттеу күрделі деректер жиындарында жоғары өнімділік классификациясы үшін ең тиімді модельдеу тәсілдерін анықтауға бағытталған. (Сурет 5.1)



Сурет 5.1- Ұсынылған модельдің нәтижелері: AUC ROC

Жоғарыдағы суретте қисық астындағы аумақ (AUC) метрикасымен бағаланатын бірнеше санаттау үлгілері үшін Қабылдағыштың жұмыс сипаттамасының (ROC) қисықтарының салыстыруын көрсетеді. Бұл емтихан әр модельдің әртүрлі санаттау деңгейлеріндегі сыныптарды саралау мүмкіндігі туралы түсінік береді.

Көлденең ось жалған оң жылдамдықты көрсетеді, ал тік ось шынайы оң жылдамдықты білдіреді. Оңтайлы жіктеуіш сол жақ шекараға және кейіннен ROC кеңістігінің жоғарғы шекарасына тығыз орналасқан қисық сызықты көрсетеді, нәтижесінде AUC 1,0-ге жақындайды. Толығымен ерікті классификатор (0,0)-ден (1,1) диагональды сызықпен тураланып, шамамен 0,5 AUC береді.

Ұсынылған модель 0,05 жалған оң жылдамдықпен 1,0 шынайы оң жылдамдықпен теңестіруден бұрын у осі бойынша төмен тік траекторияны сақтай отырып, жоғары өнімділікті көрсететінін көрсетеді. Қисық жоғарғы сол жақ шекарамен тығыз тураланады, бұл шамамен 1,0-ге тең AUC мәнін білдіреді. Бұл модельдің ғадауат тілді сөздерді анықтауда жоғары қабілетке ие екенін көрсетеді.

Терең оқыту алгоритмдерінің ішінде LSTM және CNN конфигурациялары жоғары өнімділікке ие, бұл олардың ROC қисықтарымен дәлелденеді, олар салыстырмалы түрде төмен жалған оң жылдамдықтарда жоғары шынайы оң көрсеткіштерді сақтай отырып, тік көтеріледі. Екі үлгі де х осінің бастапқы 10%

шегінде у осіндегі 0,9 шегінен асады. MLP және DNN қисықтары біршама төмен, бірақ тұрақты өнімділікті сақтайды, визуалды түрде 0,85 пен 0,95 аралығындағы AUC мәндерін көрсетеді.

Классикалық машиналық оқыту алгоритмдерінің тиімділігі, соның ішінде логистикалық регрессия, SVM және кездейсоқ орман сияқты алгоритмдері бір қалыпта орналасқан. Қисықтар тұрақты түрде көтеріліп, графиктің оң жағында 0,85 пен 0,9 арасындағы нақты оң жылдамдыққа жетеді. Олардың траекториялары төмен жалған позитивті жылдамдықтарда сезімталдықтың төмендегенін көрсетеді, бұл олардың терең оқу баламалары ретінде баламалы дәлдікке жету үшін жоғары шекті талап ететінін көрсетеді.

K-ең жақын көршілер мен шешім ағаштары кездейсоқ емес үлгілер арасында ең аз тиімділікті көрсетеді. Олардың ROC қисықтары басқаларынан айтарлықтай төмен, бір қисық 0,8 шынайы оң жылдамдықтан, тіпті 0,5 жалған оң жылдамдықтан аспайды. Бұл AUC мәндерін көрсетеді, мүмкін 0,6-дан 0,7-ге дейін ауытқиды, бұл олардың жалпылау мүмкіндігіндегі шектеулерді көрсетеді.

Кездейсоқ болжау үшін анықтамалық қисық қосылған, ол бастапқыдан жоғарғы оң жақ бұрышқа дейін диагональ бойынша созылады. Бұл сызық AUC 0,5 мәнін білдіреді және негізгі сызық ретінде қызмет етеді; барлық модельдер бұл шекті деңгейден асып түседі, бұл олардың негізгі жіктеу мүмкіндігін растайды.

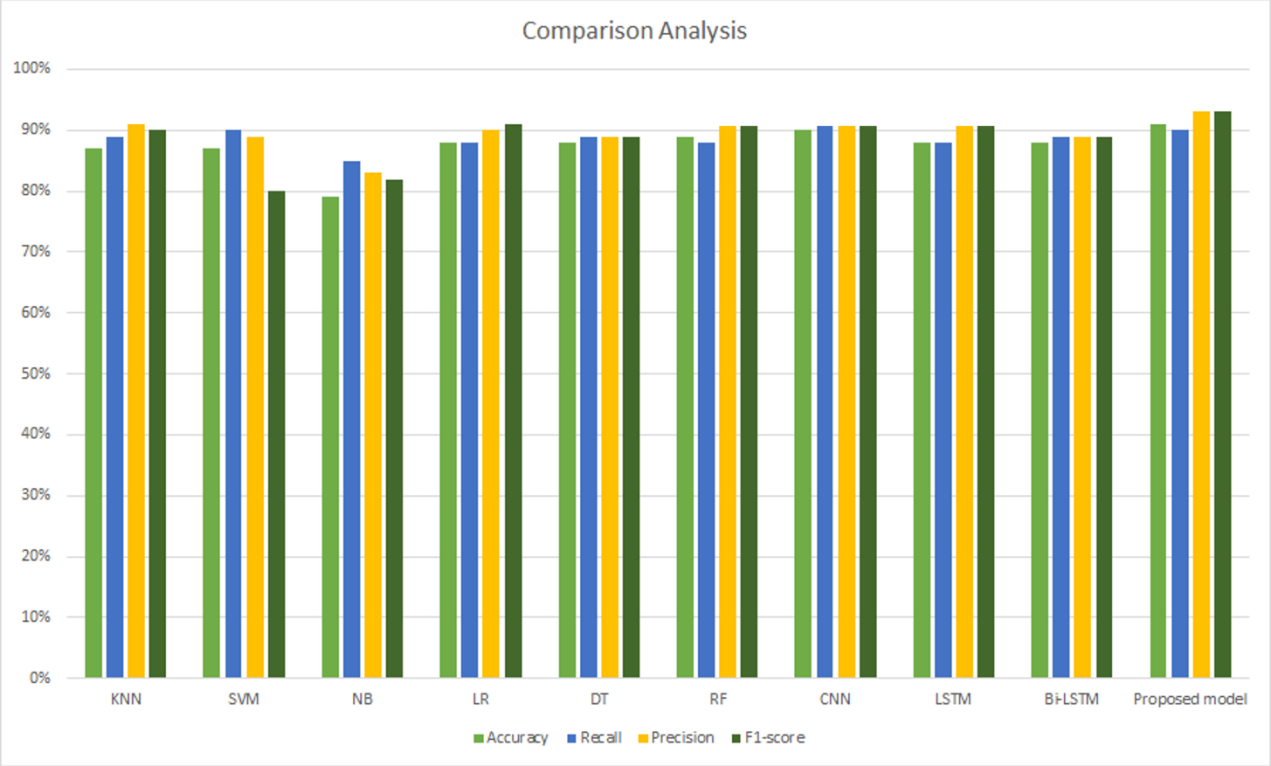
Салыстырмалы ROC-AUC талдауы үлгілер арасындағы нақты өнімділік градиентін көрсетеді. Ұсынылған модель LSTM және CNN архитектурасын қоса алғанда, барлық базалық үлгілерден айтарлықтай асып түседі. Логистикалық регрессия және қолдау векторлық машиналары сияқты кәдімгі модельдер аралық дәлдікті қамтамасыз етеді, бірақ шешім ағаштары және K-ең жақын көршілері сияқты қарапайым алгоритмдерде бұл жағдайда ғадауат тілді сөздерді анықтау өнімділігі әлдеқайда төмен екенін байқауға болады.

5.2 Өнімділік көрсеткіші бойынша салыстырмалы талдау нәтижелері

Сандық бағалау классификациялық жүйелерді құру мен бағалауда маңызды, себебі ол модельдің тиімділігін, жалпылау қабілетін және нақты әлемде қолдануға жарамдылығын анықтайды. Деректер жиынының күрделілігінің артып келе жатқанын және модельдеу әдістерінің әртүрлілігін ескере отырып, жалғыз өнімділік метрикасына сүйену жиі жеткіліксіз. Модель өнімділігінің бірнеше өлшемдерін бағалау үшін мультиметрикалық бағалау жүйесі пайдаланылады.

Бұл бөлім төрт жалпы қабылданған бағалау метрикасы арқылы әртүрлі жіктеу алгоритмдерінің өнімділігін мұқият тексеруді ұсынады: дәлдік, еске түсіру, дәлдік және F1 ұпай. Бұл көрсеткіштер жағдайларды дәл анықтау, қате болжауларды азайту және сыныптар арасындағы сәйкестікті қамтамасыз ету бойынша әрбір үлгінің дағдысын жан-жақты бағалауды қамтамасыз етеді. Дәлдік болжамдардың жалпы дұрыстығын білдіреді, еске түсіру сәйкес мысалдарға сезімталдықты бағалайды, дәлдік оң болжамдардың дұрыстығын сандық түрде анықтайды.

Бағалау кәдімгі машиналық оқыту алгоритмдері, ансамбльдік әдістер және терең оқыту шеңберлері сияқты әртүрлі үлгілерді қамтиды. Мақсаты стандартталған эксперименттік дизайнда олардың әртүрлі оқыту механизмдеріне жататын өнімділік ауытқуларын көрсете отырып, олардың санаттау қабілеттерін бағалау болып табылады. Зерттеудің мақсаты бірнеше бағалау өлшемдері бойынша осы көрсеткіштерді жүйелі талдау арқылы сенімділік пен сенімділікті қамтамасыз ете отырып, жоғары болжамды дәлдікке қол жеткізетін үлгілерді әзірлеу болып табылады. 5.2-суретте өнімділіктің төрт түрлі стандартты өлшемі негізінде бағаланған он түрлі алгоритмдер үлгісінің салыстырмалы салыстыруын қамтамасыз етеді, яғни accuracy, recall, precision және F1-score бағалау метрикалары пайызбен көрсетілген.



Сурет 5.2- Алгоритмдер арасындағы салыстырмалы анализ

Нәтижелер көрсеткендей, ұсынылған модель барлық төрт критерий бойынша ең жоғары өнімділікке ие болады, әр ұпай 95%-дан сәл жоғары. Бұл модельдің ең тиімді екенін көрсетеді. Бұл жалпылаудың жоғары екенін және өнімділіктің сезімталдық пен ерекшелік жағынан теңестірілгенін көрсетеді.

Екі бағытты ұзақ қысқа мерзімді жад (Bi-LSTM) моделі жақыннан кейін тізбекті деректерде алға және кері тәуелділіктерді түсіруде үлкен тұрақтылық пен жүйелілікті көрсетеді. Барлық көрсеткіштер 94–95%-ға жақын, бұл модельдің жақсы жұмыс істеп тұрғанын көрсетеді.

Жоғары өнімділікті сәйкесінше шамамен 92% және 94% аралығында болатын accuracy, recall, precision және F1-score мәндері бар LSTM үлгісі де көрсетеді. Демек, бұл уақыттық қатарларды немесе дәйекті сипаттамаларды модельдеу үшін оның сәйкестігін негіздейді.

Конволюциялық нейрондық желілер (CNNs) және ұзақ қысқа мерзімді жад (LSTM) шығаратын шамамен бірдей метрикалық мәндер шамамен 93% кластерленген, бұл нейрондық желілердің екі түрінің де кеңістіктік мүмкіндіктерді шығарудан пайда алатын жіктеу тапсырмаларында баламалы мүмкіндіктерге ие екенін көрсетеді.

Кездейсоқ орман алгоритмінің нәтижелері жақсы, барлық көрсеткіштер 93–94% жетеді. Бұл дисперсияны азайту және болжамдардың дәлдігін арттыру тұрғысынан модельдің ансамбльдік артықшылығын көрсетеді.

Шешім ағаштары үшін орташа балл барлық төрт параметр бойынша 89% және 91% арасында, бұл олардың ұпайлары сәл төмен екенін көрсетеді. Бір ағашты классификаторлардың белгілі артық сәйкестік тенденциялары осы тенденцияларға сәйкес келетін өнімділіктің төмендеуіне сәйкес келеді.

Логистикалық регрессияны пайдаланған кезде өнімділіктің аздап төмендеуі байқалады, өлшемдер аймақта 87%-дан 89%-ға дейін төмендеді. Ол әлі де сәтті болғанына қарамастан, деректер ішіндегі сызықтық емес қатынастарды түсіру кезінде оның кейбір кемшіліктері бар.

Naive Bayes алгоритмі басқа алгоритмдердің көпшілігіне қарағанда төмен көрсеткіштерге ие, ұпайлары шамамен 80%-дан 83%-ға дейін өзгереді. Бұл мүмкіндіктердің тәуелсіздігі туралы болжамдар қолданылған деректер жиынына сәйкес келмеуі мүмкін екенін көрсетеді.

Әдетте 88%-дан 90%-ға дейінгі диапазонға түсетін Қолдау векторлық машиналары (SVMs) осы жіктеу мәселесі бойынша қанағаттанарлық өнімділік деңгейін көрсететін логистикалық регрессия нәтижелерімен салыстырылатын нәтижелер береді.

K-Nearest Neighbours (KNN) 87 пайыздан 90 пайызға дейінгі ұпайларымен жеткілікті жоғары өнімділікке қол жеткізеді, дегенмен ол ең жоғары рейтингтерді алатын үлгілерге қарағанда біршама аз сәйкес келеді.

Демек, ұсынылған модель барлық талданған критерийлерге қатысты ең жақсы теңдестірілген және ең жоғары нәтижелі нәтижелерді береді. Терең оқыту алгоритмдерінің өнімділігі, атап айтқанда Bi-LSTM және LSTM стандартты үлгілерден үнемі асып түседі. Кездейсоқ ормандар сияқты ансамбльдік әдістер дәлдік пен түсіндіру арасындағы күшті тепе-теңдікті ұсынады.

Ұсынылған модель accuracy, recall, precision және F1-score бойынша ең жақсы нәтиже береді, әр санат үшін 95%-дан жоғары ұпайлар бар. Екі бағытты ұзақ қысқа мерзімді жад (Bi-LSTM) моделі 94-95%-ға жуық ұпаймен екінші орында келеді. LSTM моделінде жоғары дәлдік, еске түсіру, дәлдік және F1 ұпай сандары бар, бұл оны уақыттық қатарлар немесе дәйекті сипаттар үшін қолайлы етеді. Конволюциялық нейрондық желілер (CNNs) және ұзақ қысқа мерзімді жад (LSTM) метрикалық мәндерде ұқсас, бұл оларды жіктеу тапсырмаларында бірдей жақсы етеді. Кездейсоқ ормандар дисперсияны азайту және болжамдарды дәлірек ету арқасында жақсы өнімділікке ие. Шешім ағаштарының ұпайлары төмен, ал Logistic Regression және Naive Bayes жақсырақ жұмыс істейді. Қолдау векторлық машиналары (SVMs) жақсы нәтижелерге ие және K-En Nearest Neighbors (KNN) жақсы жұмыс істейді, бірақ тұрақты емес.

5.3 Көптілді деректер жиынындағы салыстырмалы өнімділікті талдау нәтижелері

Жіктеу үлгілерін жан-жақты бағалау ғадауат тілді сөздерді анықтау сияқты табиғи тілді өңдеу тапсырмаларының тиімді алгоритмдерін әзірлеу және таңдауда өте маңызды. Сандық талдау accuracy, recall, precision және F1-score және қабылдағыштың жұмыс сипаттамасы қисығының астындағы аумақты (AUC-ROC) қоса алғанда, бірнеше параметрлер бойынша әртүрлі модельдердің салыстырмалы өнімділігі туралы ақпаратты қамтамасыз етеді. Бұл көрсеткіштер бірге берілген үлгінің дәлдігі мен беріктігін теңдестірілген бағалауды қамтамасыз етеді, әсіресе теңгерімсіз немесе шулы мәтіндік деректер жиыны контекстінде қолданылады.

Бұл бөлімде қазақ, орыс және ағылшын тілдеріндегі ғадауат тілді мәтіндер корпусын қоса алғанда, көптілді деректер жиынын пайдалана отырып, машиналық және терең оқыту алгоритмдерінің жүйелі салыстырмалы талдауы қарастырылған. Мақсат - әртүрлі тілдік контексттерде ғадауат тілді сөздерді анықтаудың оңтайлы әдісін ұсыну. Кешенді бағалау логистикалық регрессия және кездейсоқ ормандар сияқты дәстүрлі жіктеуіштерден LSTM, BiLSTM, CNN және RNN сияқты жетілдірілген нейрондық архитектураға дейінгі үлгілердің кең ауқымын қамтиды. Барлық деректер жиынына біркелкі бағалау критерийлерін қолдану арқылы бұл талдау әрбір модельдің жалпылануы мен бейімделгіштігін көрсетуге бағытталған. Көптеген тілдерде жоғары өнімділікті қамтамасыз ететін модельдерді анықтауға ерекше назар аударылады, олардың көптілді және нақты әлемдегі уытты мазмұнды модерациялау жүйелерінде қолдану мүмкіндігін көрсетеді. (Кесте 5.1)

Кесте 5.1 - Үш тілді деректер жиынындағы өнімділіктің салыстырмалы талдауы

Мәлі метте р қоры	**Тәсіл**	**Алгоритм**	**Accuracy**	**Precision**	**Rec all**	**F1-score**	**AUC-ROC**
1	2	3	4	5	6	7	8
		Ұсынылған модель	**92.1%**	**93%**	**92.5%**	**92.9%**	**95%**
Қазақ тілінд егі мәлім еттер қоры	**Терең оқыту**	LSTM	91.0%	92.5%	92.8%	92.5%	92%
		BiLSTM	91.3%	92.6%	93%	92.7%	92%
		CNN	89.2%	89%	89%	88.7%	90%
		RNN	89.7%	89.5%	90%	89.9%	94%
	Машина лық оқыту	LR	87.3%	85.2%	86.2%	85.1%	78%
		RF	85.6%	83.9%	83.1%	83.7%	92%
		DT	87.4%	83.2%	86.3%	85.1%	80%

5.1-кестенің жалғасы

1	2	3	4	5	6	7	8
		NB	60.2%	52.4%	58.5%	64.2%	65%
		KNN	85.1%	85.4%	82.2%	85.6%	77%
		SVM	86.2%	85.3%	83.7%	85.8%	78%
Орыс тілінд егі мәлім еттер қоры (*Russi an Langu age Toxic Comm ents*)	**Терең оқыту**	**Ұсынылған модель**	**91.5%**	**93%**	**92%**	**92.4%**	**94%**
		BiLSTM	91%	92%	92%	92%	93%
		CNN	88.7%	89%	89%	88.9%	90%
		LSTM	90%	91%	91.5%	91.2%	93%
		RNN	89%	89%	91%	90%	92%
	Машина лық оқыту	LR	87.3%	85.6%	87.4%	60.2%	85.1%
		RF	85.2%	83.9%	83.2%	52.4%	85.4%
		DT	86.2%	83.1%	86.3%	58.5%	82.2%
		NB	85.1%	83.7%	85.1%	64.2%	85.6%
		KNN	75%	90%	76%	68%	77%
		SVM	87.3%	85.6%	87.4%	60.2%	85.1%
Ағыл шын тілінд егі мәлім еттер қоры (hate-speech-and-offensive-language)	**Терең оқыту**	**Ұсынылған модель**	**95%**	**96%**	**96.5%**	**96%**	**97%**
		LSTM	93%	92.5%	94%	95%	95%
		MLP	91.5%	92%	93%	92.5%	94%
		CNN	90%	92%	91%	91.5%	93%
		RNN	91%	93%	92%	92%	92%
	Машина лық оқыту	LR	89.3%	90%	89.5%	89.3%	89%
		RF	88.6%	89%	88%	88.5%	89%
		DT	87.4%	83.2%	86.3%	85.1%	80%
		NB	86.2%	82.4%	85.5%	84.5%	85%
		KNN	87.1%	85.4%	87.2%	86.4%	87%
		SVM	87.2%	87.3%	86.7%	85.8%	88%

Кеңейтілген бағалау әртүрлі тілдік орталарда модельдің жалпылану мүмкіндігін мұқият тексеруді қамтамасыз ету үшін ағылшын тіліндегі уытты пікірлердің үшінші деректер жинағын қамтиды. Модельдер бес стандартты көрсеткіш арқылы қайта бағаланды: дәлдік, дәлдік, еске түсіру, F1 ұпайы және AUC-ROC.

Ұсынылған үлгі барлық көрсеткіштер бойынша 95% accuracy, 96% recall, 96,5% precision 96% F1-score және 97% AUC-ROC көрсеткіштеріне қол жеткізе отырып, ағылшын тіліндегі деректер жиынында тамаша өнімділікке қол

жеткізді. Бұл нәтижелер болжау дәлдігінің жақсарғанын ғана емес, сонымен қатар дәлдік пен еске түсіру арасындағы күшті тепе-теңдікті көрсетеді, бұл оның жалған позитивті және жалған теріс мәндерді азайтудағы тиімділігін көрсетеді.

Терең оқыту әдістемелерінің ішінде LSTM 93% accuracy және 95% F1 score қол жеткізді, ал RNN 92% және 92% F1 ұпайларының салыстырмалы еске түсіру жылдамдығына ие болды. MLP және CNN үлгілері біршама нашар өнімділікті көрсетті, CNN 90% accuracy 91,5% F1 ұпайына қол жеткізді, ал MLP 91,5% дәлдік пен 92,5% F1 ұпайына қол жеткізді.

Кәдімгі машиналық оқыту әдістері терең оқыту үлгілерімен салыстырғанда өнімділіктің айтарлықтай төмендеуін көрсетті. Логистикалық регрессия және кездейсоқ орман алгоритмі сәйкесінше 89,3% және 88,6% дәлдікке қол жеткізе отырып, қолайлы өнімділікті көрсетті және екі үлгі үшін де AUC-ROC мәндері 89% құрады. Шешім ағашының жіктеуіштері 87,4% дәлдікке және 80% айтарлықтай төмен AUC-ROC деңгейіне қол жеткізді, бұл күрделі улы тіл үлгілерін анықтаудағы шектеулі тиімділікті көрсетеді. Naive Bayes тағы да ең нашар орындады, дәлдік 86,2% және F1 көрсеткіші 84,5%. K-ең жақын көршілер мен тірек векторлық машиналар сәл жоғары нәтиже берді, SVM 87,2% дәлдікке және 85,8% F1 ұпайына қол жеткізді, ал KNN 87,1% дәлдікке және 86,4% F1 ұпайына қол жеткізді.

Үш деректер жиыны (қазақ, орыс және ағылшын) бойынша өнімділікті бағалау кезінде ұсынылған үлгі әр конфигурациядағы барлық баламалардан тұрақты түрде асып түсті. Қазақ тіліндегі деректер жинағы 92,1% дәлдікке және 95% AUC-ROC деңгейіне қол жеткізді. Орыс тіліндегі деректер жинағы 91,5% дәлдікке және 94% AUC-ROC деңгейіне қол жеткізді. Сайып келгенде, ағылшын деректер жинағы максималды дәлдік мәндеріне 95% және AUC-ROC 97% жетті. Бұл тұрақты үстемдік модельдің әртүрлі тілдік құрылымдарға және уытты пікірлерде байқалатын контекстік белгілерге сенімді бейімделуін көрсетеді.

LSTM, BiLSTM және RNN қоса алғанда, терең оқыту әдістемелері барлық тіл домендері бойынша дәстүрлі алгоритмдерден, әсіресе еске түсіру және AUC-ROC метрикаларында жоғары көрсеткішке ие болды, олар ғадауат тілді сөздерді анықтау сияқты теңгерімсіз мәтінді санаттау мәселелерін шешу үшін өте маңызды.

Бұл тарау үш тілдік бір-бірінен ерекшеленетін деректер жиыны: қазақ, орыс және ағылшын тілдерінде ғадауат тілді пікірлерді анықтау үшін қолданылатын бірнеше жіктеу үлгілерінің жан-жақты салыстырмалы сараптамасын ұсынады. Бағалау кәдімгі машиналық оқыту алгоритмдерін, орнатылған терең оқыту алгоритмдерімен accuracy, recall, precision және F1-score және AUC-ROC сияқты бірнеше өнімділік көрсеткіштері арқылы бағаланатын ұсынылған модельді қамтиды.

Нәтижелер ұсынылған модельдің барлық деректер жинақтары мен бағалау көрсеткіштері бойынша жоғарылатылған тиімділігін дәйекті түрде көрсетеді. Ұсынылған архитектураның бағалау көрсеткіштері бойынша жоғары

нәтижеге жеткенін, AUC-ROC 1,0-ге жақындайды және басқа көрсеткіштер тұрақты түрде 95%-дан асқандығын көрсетеді. ROC қисығының талдауы бұл нәтижелерді диаграммада келтірілген салыстырмалы нәтижелер ретінде растайды, өйткені ұсынылған модельдің траекториясы жоғарғы сол жақ шекараға дерлік сәйкес келеді, бұл классификацияның оңтайлы өнімділігін білдіреді.

Терең оқытудың базалық алгоритмдерінің арасында BiLSTM және LSTM үлгілері сенімді және бәсекеге қабілетті өнімділігімен ерекшеленеді, әсіресе recall және F1 score бағалаулары, мәтіндік деректердегі дәйекті қарым-қатынастар мен контекстік нәзіктіктерді түсіру қабілетін көрсетеді. CNN үлгілері, әсіресе, кеңістіктік мүмкіндіктерді шығару тиімді болған жағдайда, күшті өнімділікті көрсетті, ал MLP және RNN дизайндары орташа, бірақ дәйекті нәтижелер берді.

Дәстүрлі машиналық оқыту алгоритмдері — логистикалық регрессия, кездейсоқ орман, қолдау векторлық машиналары, шешім ағаштары, ең жақын көршілер және қарапайым Бейс — әдетте олардың терең оқу баламаларынан артта қалды. Кездейсоқ ормандар және SVM сияқты бірнеше модельдер нақты контексттерде сенімді өнімділікті көрсеткенімен, басқалары Naive Bayes және Decision Trees сияқты күрделі тілдік сипаттамаларды, әсіресе қазақ және орыс деректер жиынтығын басқарудағы кемшіліктерді анықтады.

Салыстырмалы бағаналы диаграмма және ROC қисықтары терең оқыту мен классикалық әдістер арасындағы өнімділік алшақтығын көрсетеді, біріншісі дәлдікте де, еске түсіруде де ерекше артықшылықты көрсетеді. Бұдан басқа, кешенді кестелік қорытындылар ұсынылған модельдің жеке көрсеткіштер бойынша да, барлық бағаланған критерийлер бойынша тепе-теңдікті сақтау қабілеті бойынша тұрақты артықшылығын көрсетеді.

Бұл тұжырымдар көптілді табиғи тілді өңдеуде, әсіресе ғадауат тілді пікірлерді жіктеу сияқты күрделі тапсырмалар үшін терең оқытудың маңызды функциясын растайды. Ұсынылған модельдің жетістігі кеңейтілген нейрондық желілердің әртүрлі лингвистикалық орталарда мазмұнды модерациялау үшін масштабталатын, дәл және контекстке сезімтал шешімдерді қамтамасыз ету мүмкіндігін көрсетеді.

Қорытындылай келе, зерттеу классикалық классификаторлар интерпретация мен есептеу тиімділігін қамтамасыз еткенімен, қазіргі заманғы терең оқыту әдістерімен біртіндеп асып түсетінін растайды. Ұсынылған модель әртүрлі тілдер мен мәдени контексттерде жақсы жұмыс істеуге қабілетті анағұрлым интеллектуалды және инклюзивті автоматтандырылған модерация жүйелерін дамытуға көмектесетін өте тиімді және бейімделгіш шешімді ұсынады.

Бұл зерттеудің мақсаты қазақ, орыс және ағылшын тілдеріндегі онлайн контентте кездесетін ғадауат тілді сөздері бар пікірлерді анықтаудың тиімді үлгілерін әзірлеу болды. Осы мақсатқа жету үшін теңдестірілген және сапалы мәліметтер базасы құрылды. Атап айтқанда, үш тәуелсіз аннотатор 20 000 ғадауат тілді сөздерге жататын және 20 000 бейтарап пікірлерді жинап, оларды

екі сыныпқа бөлді. Дауыс беру әдісі әрбір пікірдің шынайы классификациясын анықтап, аннотацияның сапасын қамтамасыз етті.

Зерттеу жұмысы аясында әртүрлі машиналық және терең оқыту әдістерін сынақтан өтті. Дегенмен, зерттеу тобының арнайы әзірлеген гибридті моделі ең жақсы нәтиже көрсетті. Бұл модель BERT заманауи контекстік тіл үлгісінің архитектурасын оған біріктірілген қосымша зейін механизмін пайдалана отырып, мәтіннің маңызды бөліктеріне көбірек салмақ беру мүмкіндігін біріктіреді.

Гибридті модельдің тиімділігі көптілді деректермен жұмыс істегенде басқа алгоритмдермен салыстырмалы түрде қарағанда жақсы нәтижеге жете білді. Модель әртүрлі тілдердегі мәтіндердің мағыналық және құрылымдық ерекшеліктерін ескере отырып, жоғары дәлдікпен жіктеуді орындауға қабілеттілігін көрсетті. Бұл нәтижелер BERT үлгісінің алдын ала дайындалған тілдік білімінің тапсырмаға арнайы бейімделудегі рөлін және контекстік сәйкестікті анықтаудағы зейін механизмін көрсетеді.

Салыстырмалы талдау барысында ұсынылып отырған гибридті үлгі машиналық оқытудың дәстүрлі үлгілерінен қазақ тілі бойынша 17%, орыс тілі үшін 15% және ағылшын тілі бойынша 5% артық болды. Бұл айырмашылықтар модельдің тілдік мүмкіндіктерін және контекстті түсініп, ғадауат тілді пікірлерді анықтауда жоғары нәтиже көрсетті.

Тұтастай алғанда, BERT және зейін механизміне негізделген бұл гибридті архитектура көптілді мәтіндермен жұмыс істеуде жоғары нәтижелерге қол жеткізуге мүмкіндік беретін перспективалы бағыттардың бірі болып саналады. Ол онлайн контенттегі ғадауат тілді пікірлерді автоматты түрде анықтау жүйелерінде және мәтінді талдаудың басқа салаларында кең қолдану аясын таба алады.

ҚОРЫТЫНДЫ

Бұл ғылыми зерттеу жұмысында желі пайдаланушыларынан келген ғадауат тілді сөздерді анықтау жағдайларын жіктеу тапсырмасы үшін терең оқыту алгоритмдерін әзірлеуге арналды. Зерттеу әртүрлі машиналық оқыту және терең оқыту архитектураларын, соның ішінде қайталанатын нейрондық желілерді, конволюциялық нейрондық желілерді және табиғи тілді өңдеу құралдарын зерттеді және олардың ғадауат тілді сөздерді анықтаудағы тиімділігін зерттеді.

Зерттеу нәтижелері терең оқыту үлгілері дәстүрлі машиналық оқыту алгоритмдерімен салыстырғанда ғадауат тілді сөздерді анықтауда жоғары дәлдікті көрсетті. Мәтіндік ағындардағы күрделі үлгілер мен мүмкіндіктерді тиімді түсіру арқылы бұл модельдер мәтінді жіктеудің көп қырлы мәселелерін шешуге өте қолайлы екенін дәлелдеді. Зерттеу сонымен қатар әртүрлі терең оқыту архитектурасын біріктірудің ықтимал артықшылықтарын көрсетеді.

Түрлі деректер базасында жасалған іс – тәжірибелер, зерттеулер мен талқылаулар, жасақталған деректер қорын талқылаудан өткізу және алгоритмдерді тестілеуде пайдаланған нейрондық желілер арқылы қазіргі таңда күннен күнге онлайн контентте өршіп келе жатқан онлайн контенттегі ғадауат тілді сөздер кездесетін пікірлерді анықтауда көмегін тигізетінін және жасақталған модельді одан кейінгі ғылыми зерттеулерде және ғадауат тілді сөздерді автоматты түрде анықтау құралдарында пайдалануға болатынын көрсетті.

Ғылыми зерттеу жұмысында келесі нәтижелерге қол жеткізілді:

1) Қазақ тіліне арналған алдын ала өңделген және қолмен жіктелген деректер қоры машиналық және терең оқыту тапсырмалары үшін жиналды.

2) Зейінді механизмі бар терең нейрондық желі екілік және үш класты жіктеу тапсырмасы үшін әзірленді және оқытылды.

3) Анықтау тапсырмасында машиналық оқыту мен терең оқыту арасында салыстырмалы талдау жүргізіліп, өнімділігі жоғары модель таңдалып алынды.

ПАЙДАЛАНЫЛҒАН ӘДЕБИЕТТЕР ТІЗІМІ

1. Нұрыш Ж. Медиадағы ғадауат тілі дегеніміз не және гендерлік стереотиптерге не жатады? Оразай Қыдырбаев түсіндіреді [Электрондық ресурс] // Жаңа репортер. – 2021. – URL: https://inlnk.ru/1PYRmd (қол жеткізу күні: 26.01.2025).
2. The Devil of Hate Speech: How a Journalist Should Avoid Hate Speech in Their Materials - Cabar School [Электрондық ресурс]. – URL: https://school.cabar.asia/en/articles/the-devil-of-hate-speech-how-a-journalist-should-avoid-hate-speech-in-their-materials/?print=print (қол жеткізу күні: 26.01.2025).
3. ғадауат [Электрондық ресурс] // Sozdikqor.kz. – URL: http://sozdikqor.kz/search?q=%D2%93%D0%B0%D0%B4%D0%B0%D1%83%D0%B0%D1%82 (қол жеткізу күні: 26.01.2025).
4. Электронды каталог | Қазақстан Республикасының Ұлттық академиялық кітапханасы [Электрондық ресурс]. – URL: https://inlnk.ru/DB02g2 (қол жеткізу күні: 26.01.2025).
5. Toktarova, A., Azhibekova, Z., Kazbekova, G., & Temirbekova, F. (2023). Онлайн контенттегі ғадауат сөздер классификациясы, анықтау және қоғамға тигізер әсері. Bulletin of Abai KazNPU. Series of Physical and mathematical sciences, 82(2), 288-296.
6. Махаббат пен ғадауат майданы немесе Абайдағы рух түсінігі [Электрондық ресурс] // Қазақ әдебиеті. – URL: https://qazaqadebieti.kz/14773/mahabbat-pen-adauat-majdany-nemese-abajda-y-ruh-t-sinigi (қол жеткізу күні: 26.01.2025).
7. hate speech [Электрондық ресурс] // Cambridge Dictionary. – URL: https://dictionary.cambridge.org/dictionary/english/hate-speech (қол жеткізу күні: 26.01.2025).
8. Toktarova A., Azhibekova, Z., Kazbekova, G., & Temirbekova, F Automated Hate Speech Classification using Emotion Analysis in Social Media User Generated Texts // Journal of Theoretical and Applied Information Technology. – 2022. – Vol. 100. – P. 6621–6634.
9. Ugarte R. Classifying and Identifying the Intensity of Hate Speech [Электрондық ресурс] // Items. – URL: https://items.ssrc.org/disinformation-democracy-and-conflict-prevention/classifying-and-identifying-the-intensity-of-hate-speech/ (қол жеткізу күні: 26.01.2025).
10. Погонцева Д.В. К вопросу о дискриминации по внешнему облику // Северо-Кавказский психологический вестник. – 2011. – Т. 9, № 2. – С. 47–50.
11. Кибербуллинг [Электрондық ресурс] // Бикин – БезФормата. – URL: https://bikin.bezformata.com/listnews/kiberbulling/108336082/ (қол жеткізу күні: 26.01.2025).
12. hate speech noun – Definition, pictures, pronunciation and usage notes [Электрондық ресурс] // Oxford Advanced Learner’s Dictionary. – URL:

https://www.oxfordlearnersdictionaries.com/definition/english/hate-speech (қол жеткізу күні: 26.01.2025).

13. Hate speech Definition & Meaning [Электрондық ресурс] // Merriam-Webster Dictionary. – URL: https://www.merriam-webster.com/dictionary/hate%20speech (қол жеткізу күні: 26.01.2025).

14. HATE SPEECH definition and meaning [Электрондық ресурс] // Collins English Dictionary. – URL: https://www.collinsdictionary.com/dictionary/english/hate-speech (қол жеткізу күні: 26.01.2025).

15. Sultan, D., Suliman, A., Toktarova, A., Omarov, B., Mamikov, S., & Beissenova, G. (2021, January). Cyberbullying detection and prevention: Data mining in social media. In 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence) (pp. 338-342). IEEE.

16. Lake D.A. Rational extremism: Understanding terrorism in the twenty-first century // Dialogue IO. – Cambridge University Press, 2002. – Vol. 1, № 1. – P. 15–28.

17. The psychology of extremism and terrorism: A Middle-Eastern perspective [Электрондық ресурс] // ScienceDirect. – URL: https://www.sciencedirect.com/science/article/abs/pii/S1359178906000929 (қол жеткізу күні: 26.01.2025).

18. Is Extremism the 'New' Terrorism? The Convergence of 'Extremism' and 'Terrorism' in British Parliamentary Discourse [Электрондық ресурс] // Taylor & Francis Online. – URL: https://www.tandfonline.com/doi/abs/10.1080/09546553.2019.1598391 (қол жеткізу күні: 26.01.2025).

19. Galland O., Muxel A. Radicalism in Question // Radical Thought among the Young: A Survey of French Lycée Students. – Brill, 2020. – P. 1–23.

20. Bittner E. Radicalism and the Organization of Radical Movements // American Sociological Review. – 1963. – Vol. 28, № 6. – P. 928–940.

21. Инга Сикорская: ғадауат тілі дегеніміз не және журналистикада оған қалай жол бермеуге болады? [Электрондық ресурс] // Cabar School. – URL: https://school.cabar.asia/kz/video/inga-sikorskaja-adauat-tili-degenimiz-ne-zh-ne-zhurnalistikada-o-an-alaj-zhol-bermeuge-bolady/ (қол жеткізу күні: 26.01.2025).

22. Щербинин Ю.В. Hate speech [Электрондық ресурс]. – URL: https://surl.li/grdoti (қол жеткізу күні: 26.01.2025).

23. (20+) В эфире на Facebook [Электрондық ресурс] // Facebook. – URL: https://www.facebook.com/watch/live/?ref=watch_permalink&v=1528556700847322 (қол жеткізу күні: 26.01.2025).

24. язык ненависти [Электрондық ресурс]. – URL: https://magazine.mospsy.ru/phorum/read.php?f=2&i=3540&t=3540 (қол жеткізу күні: 26.01.2025).

25. Язык ненависти: юридические и психологические аспекты [Электрондық ресурс] // Defenders Belarus. – URL: https://www.defendersbelarus.org/hatespeech (қол жеткізу күні: 26.01.2025).

26. Khoja Akhmet Yassawi International Kazakh-Turkish University, et al. Automated offensive language classification through “emotional” comments from network users // Bulletin of the National Engineering Academy of the Republic of Kazakhstan. – 2023. – P. 92–102.

27. Team I.M. Over the past year, Instagram has introduced new features to make the platform more accessible to those with visual impairments [Электрондық ресурс] // Internet Matters. – 2019. – URL: https://www.internetmatters.org/hub/news-blogs/instagram-the-fight-against-online-bullying/ (қол жеткізу күні: 21.01.2025).

28. Digital 2024: Kazakhstan [Электрондық ресурс] // DataReportal. – 2024. – URL: https://datareportal.com/reports/digital-2024-kazakhstan (қол жеткізу күні: 26.01.2025).

29. THE TENGE. Количество интернет-пользователей в Казахстане [Электрондық ресурс] // THE TENGE. – URL: https://the-tenge.kz/articles/internet-i-kazakhstan (қол жеткізу күні: 26.01.2025).

30. Қазақстанда интернетті пайдаланушылар саны қарқынды өсім келеді [Электрондық ресурс] // Ranking.kz. – 2023. – URL: https://ranking.kz/kz/digest-kz/industries-digest-kz/kazakstanda-internetti-paydalanushylar-sany-karkyndy-osim-keledi.html (қол жеткізу күні: 22.01.2025).

31. Жоба туралы - Balaqorgau [Электрондық ресурс]. – URL: https://bala.gov.kz/kk/about.html (қол жеткізу күні: 21.01.2025).

32. Психолог Александра Бочавер: Что такое кибербуллинг и как правильно на него реагировать [Электрондық ресурс] // Вместе против буллинга. – URL: https://bullying.shkolamoskva.ru/theory/read/psiholog-aleksandra-bochaver-chto-takoe-kiberbulling-i-kak-pravilno-na-nego-reagirovat/ (қол жеткізу күні: 26.01.2025).

33. UNICEF Қазақстан [Электрондық ресурс]. – URL: https://www.unicef.org/kazakhstan/kk?yandex-source=desktop-maps (қол жеткізу күні: 21.01.2025).

34. Как травят друг друга немецкие школьники – DW – 02.12.2020 [Электрондық ресурс] // DW.com. – URL: https://www.dw.com/ru/kak-travjat-v-nemeckih-shkolah/a-55796370 (қол жеткізу күні: 26.01.2025).

35. Детский булинг в немецких школах [Электрондық ресурс] // Laru Helps Ukraine e.V. – 2023. – URL: https://laruhelpsukraine.com/info/detskij-buling-v-nemeczkih-shkolah/ (қол жеткізу күні: 26.01.2025).

36. Toktarova A., et al. Онлайн контенттегі қазақ тілді бейәдеп пікірлерді машиналық оқытуда жинақтау // Bulletin of Abai KazNPU. Series of Physical and Mathematical Sciences. – 2023. – Vol. 81, № 1. – P. 265–272.

37. Құпиялық саясаты – Әйелдер гольф күні [Электрондық ресурс]. – URL: https://womensgolfday.com/kk/privacy-policy/ (қол жеткізу күні: 26.01.2025).

38. Қазақстан Республикасының кейбір конституциялық заңдарына Мемлекет басшысының 2022 жылғы 16 наурыздағы Жолдауын іске асыру мәселелері бойынша өзгерістер мен толықтырулар енгізу туралы [Электрондық

ресурс] // Әділет. – URL: https://adilet.zan.kz/kaz/docs/Z2200000156 (қол жеткізу күні: 26.01.2025).

39. Әкімшілік құқық бұзушылық туралы 41-бап. Әкімшілік жаза түрлері [Электрондық ресурс]. – URL: https://kodeksy-kz.com/ob_administrativnyh_pravonarusheniyah/41.htm (қол жеткізу күні: 26.01.2025).

40. Кибербуллинг: Қатыгездік қайдан, қауіп неден? [Электрондық ресурс] // Egemen.kz. – URL: https://egemen.kz/article/307981-kiberbulling-qatygezdik-qaydan-qauip-neden (қол жеткізу күні: 26.01.2025).

41. Как избежать языка вражды в медиа? Советы коллегам от Комиссии по этике БАЖ [Электрондық ресурс] // БАЖ. – URL: https://baj.media/ru/nakirunki-i-kampanii/kak-izbezhat-yazyka-vrazhdy-v-media-sovety-kollegam-ot-komissii-po-etike-bazh/ (қол жеткізу күні: 26.01.2025).

42. Полиция Туркестанской области начала расследование по делу о доведении до самоубийства писательницы Аягуль Мантай [Электрондық ресурс] // Медиазона Центральная Азия. – URL: https://mediazona.ca/news/2021/07/30/mantai (қол жеткізу күні: 26.01.2025).

43. Дядя рассказал о последнем дне жизни Жанар Хамитовой [Электрондық ресурс] // Tengrinews.kz. – 2019. – URL: https://tengrinews.kz/show/dyadya-rasskazal-o-poslednem-dne-jizni-janar-hamitovoy-379380/ (қол жеткізу күні: 26.01.2025).

44. Хейтспич в сети: распознать, противостоять и предотвратить [Электрондық ресурс] // Ландшафт цифровых прав и свобод. – 2024. – URL: https://drfl.kz/ru/online-hate-speech/ (қол жеткізу күні: 26.01.2025).

45. Какие меры в отношении ложной информации принимает Facebook [Электрондық ресурс] // Facebook Help Center. – URL: https://www.facebook.com/help/1952307158131536/?locale=ru_RU (қол жеткізу күні: 26.01.2025).

46. Условия использования | Справочный центр Instagram [Электрондық ресурс]. – URL: https://help.instagram.com/581066165581870/?locale=ru_RU (қол жеткізу күні: 26.01.2025).

47. Управление конфиденциальностью в Instagram [Электрондық ресурс]. – URL: https://help.instagram.com/safety (қол жеткізу күні: 21.01.2025).

48. Schultz M. What Are the Twitter Rules? [Электрондық ресурс] // Twenvy. – 2023. – URL: https://www.twenvy.com/what-are-the-twitter-rules/ (қол жеткізу күні: 21.01.2025).

49. Burnap P., Williams M.L. Us and Them: Identifying Cyber Hate on Twitter Across Multiple Protected Characteristics // EPJ Data Science. – 2016. – Vol. 5, № 1. – P. 11.

50. A Deep Learning Approach for Automatic Hate Speech Detection in the Saudi Twittersphere [Электрондық ресурс] // MDPI. – URL: https://www.mdpi.com/2076-3417/10/23/8614 (қол жеткізу күні: 21.01.2025).

51. IEEE Xplore [Электрондық ресурс]. – URL: https://ieeexplore.ieee.org/Xplore/home.jsp (қол жеткізу күні: 26.01.2025).

52. SpringerLink [Электрондық ресурс]. – URL: https://link.springer.com/ (қол жеткізу күні: 26.01.2025).
53. ScienceDirect [Электрондық ресурс]. – URL: https://www.sciencedirect.com/ (қол жеткізу күні: 26.01.2025).
54. ACM Digital Library [Электрондық ресурс]. – URL: https://dl.acm.org/ (қол жеткізу күні: 26.01.2025).
55. Cheng J., Danescu-Niculescu-Mizil C., Leskovec J. Antisocial Behavior in Online Discussion Communities // Proceedings of the International AAAI Conference on Web and Social Media. – 2015. – Vol. 9, № 1. – P. 61–70.
56. Ayo F.E., et al. Machine Learning Techniques for Hate Speech Classification of Twitter Data: State-of-the-Art, Future Challenges and Research Directions // Computer Science Review. – 2020. – Vol. 38. – P. 100311.
57. Alkomah F., Ma X. A Literature Review of Textual Hate Speech Detection Methods and Datasets // Information. – 2022. – Vol. 13, № 6. – P. 273.
58. Gröndahl T., et al. All You Need is "Love": Evading Hate Speech Detection // Proceedings of the 11th ACM Workshop on Artificial Intelligence and Security. – 2018. – P. 2–12.
59. Rohmawati U.A.N., Sihwi S.W., Cahyani D.E. SEMAR: An Interface for Indonesian Hate Speech Detection Using Machine Learning // 2018 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI). – 2018. – P. 646–651.
60. Al-Saqqa S., Awajan A., Hammo B. A Survey of Hate Speech Detection for Arabic Social Media: Methods and Datasets // Procedia Computer Science. – 2024. – Vol. 251. – P. 224–231.
61. Nayel H.A., Shashirekha H.L. DEEP at HASOC2019: A Machine Learning Framework for Hate Speech and Offensive Language Detection // FIRE (working notes). – 2019. – Vol.2517. - P. 336–343.
62. Kumari K., et al. Bilingual Cyber-aggression Detection on Social Media Using LSTM Autoencoder // Soft Computing. – 2021. – Vol. 25, № 14. – P. 8999–9012.
63. Canhoto A.I., Clear F. Artificial Intelligence and Machine Learning as Business Tools: A Framework for Diagnosing Value Destruction Potential // Business Horizons. – 2020. – Vol. 63, № 2. – P. 183–193.
64. Jain A., Mandowara J. Text Classification by Combining Text Classifiers to Improve the Efficiency of Classification // International Journal of Computer Application. – 2016. – Vol. 6, № 2. – P. 181–192
65. Build Software Better, Together [Электрондық ресурс] // GitHub. – URL: https://github.com (қол жеткізу күні: 26.01.2025).
66. Khan M.M., Shahzad K., Malik M.K. Hate Speech Detection in Roman Urdu // ACM Transactions on Asian and Low-Resource Language Information Processing. – 2021. – Vol. 20, № 1. – P. 1–19.
67. Sharma D., Singh A., Singh V.K. THAR – Targeted Hate Speech Against Religion: A High-quality Hindi-English Code-mixed Dataset with the Application of

Deep Learning Models for Automatic Detection // ACM Transactions on Asian and Low-Resource Language Information Processing. – 2024. – Vol.43– P.1027-1033

68. Yu M., et al. Deep Learning for Real-time Social Media Text Classification for Situation Awareness: Using Hurricanes Sandy, Harvey, and Irma as Case Studies // Social Sensing and Big Data Computing for Disaster Management. – Routledge, 2020. – P. 33–50.

69. Singh J.P., et al. Attention-Based LSTM Network for Rumor Veracity Estimation of Tweets // Information Systems Frontiers. – 2022. – Vol. 24, № 2. – P. 459–474.

70. Forestiero A. Metaheuristic Algorithm for Anomaly Detection in Internet of Things Leveraging on a Neural-driven Multiagent System // Knowledge-Based Systems. – 2021. – Vol. 228. – P. 107241.

71. Website P.T. How to Help Someone That is Experiencing Online Abuse – Tips from Twitter [Электрондық ресурс] // Phambano Technology Development Centre NPC. – 2019. – URL: https://www.phambano.org.za/thembasouthernafrica/how-to-help-someone-that-is-experiencing-online-abuse-tips-from-twitter/ (қол жеткізу күні: 21.01.2025).

72. Comito C., Forestiero A., Pizzuti C. Word Embedding Based Clustering to Detect Topics in Social Media // IEEE/WIC/ACM International Conference on Web Intelligence. – Thessaloniki, Greece: ACM, 2019. – P. 192–199.

73. MacAvaney S., et al. Hate Speech Detection: Challenges and Solutions // PLOS ONE. – 2019. – Vol. 14, № 8. – P. e0221152.

74. Tranfield D., Denyer D., Smart P. Towards a Methodology for Developing Evidence-Informed Management Knowledge by Means of Systematic Review // British Journal of Management. – 2003. – Vol. 14, № 3. – P. 207–222.

75. Guest G., MacQueen K.M., Namey E.E. Applied Thematic Analysis. – Sage Publications, 2011. - Vol 7 – P.133-147

76. Zampieri M., et al. SemEval-2019 Task 6: Identifying and Categorizing Offensive Language in Social Media (OffensEval) // arXiv:1903.08983. – arXiv, 2019. – Vol.2. – P.45-56

77. Mussiraliyeva S., et al. Applying Machine Learning Techniques for Religious Extremism Detection on Online User Contents // Computers, Materials & Continua. – 2022. – Vol. 70, № 1. – P. 915–934.

78. Matamoros-Fernández A., Farkas J. Racism, Hate Speech, and Social Media: A Systematic Review and Critique // Television & New Media. – 2021. – Vol. 22, № 2. – P. 205–224.

79. Strossen N. Freedom of Speech and Equality: Do We Have to Choose // Journal of Law & Policy. – 2016. – Vol. 25. – P. 185.

80. Neupane D., Seok J. Bearing Fault Detection and Diagnosis Using Case Western Reserve University Dataset with Deep Learning Approaches: A Review // IEEE Access. – 2020. – Vol. 8. – P. 93155–93178.

81. Chetty N., Alathur S. Hate Speech Review in the Context of Online Social Networks // Aggression and Violent Behavior. – 2018. – Vol. 40. – P. 108–118.

82. Yuan S., Wu X. Deep Learning for Insider Threat Detection: Review, Challenges and Opportunities // Computers & Security. – 2021. – Vol. 104. – P. 102221.

83. Paz M.A., Montero-Díaz J., Moreno-Delgado A. Hate Speech: A Systematized Review // Sage Open. – 2020. – Vol. 10, № 4. – P. 2158244020973022.

84. Salloum S.A., et al. Machine Learning and Deep Learning Techniques for Cybersecurity: A Review // Proceedings of the International Conference on Artificial Intelligence and Computer Vision (AICV2020) / Eds. Hassanien A.-E. et al. – Cham: Springer International Publishing, 2020. – Vol. 1153. – P. 50–57.

85. Saha P., et al. Hateminers: Detecting Hate Speech Against Women // arXiv:1812.06700. – arXiv, 2018. – Vol 8. – P.115-128

86. de Andrade C.M.V., Gonçalves M.A. Profiling Hate Speech Spreaders on Twitter: Exploiting Textual Analysis of Tweets and Combinations of Multiple Textual Representations // CEUR Workshop Proceedings. – 2021. – Vol. 2936. – P. 2186–2192.

87. Cer D. Universal Sentence Encoder // arXiv preprint arXiv:1803.11175. – 2018. - Vol.1 – P.49-61.

88. Wadhwa P., Bhatia M.P.S. Classification of Radical Messages in Twitter Using Security Associations // Case Studies in Secure Computing: Achievements and Trends. – Boca Raton, FL: CRC Press, 2014. – P. 273–294.

89. Rangel F., et al. Profiling Hate Speech Spreaders on Twitter: Task at PAN 2021 // Proceedings of the Working Notes of CLEF 2021, Conference and Labs of the Evaluation Forum. – Bucharest, Romania, 2021. – P. 1772–1789.

90. Davidson T., et al. Automated Hate Speech Detection and the Problem of Offensive Language // Proceedings of the International AAAI Conference on Web and Social Media. – 2017. – Vol. 11, № 1. – P. 512–515.

91. Watanabe H., Bouazizi M., Ohtsuki T. Hate Speech on Twitter: A Pragmatic Approach to Collect Hateful and Offensive Expressions and Perform Hate Speech Detection // IEEE Access. – 2018. – Vol. 6. – P. 13825–13835.

92. Mullah N.S., Zainon W.M.N.W. Advances in Machine Learning Algorithms for Hate Speech Detection in Social Media: A Review // IEEE Access. – 2021. – Vol. 9. – P. 88364–88376.

93. Waseem Z. Are You a Racist or Am I Seeing Things? Annotator Influence on Hate Speech Detection on Twitter // Proceedings of the First Workshop on NLP and Computational Social Science. – 2016. – P. 138–142.

94. Aljarah I., Habib M., Hijazi N., Faris H., Qaddoura R., Hammo B., Abushariah M., Alfawareh M. Intelligent Detection of Hate Speech in Arabic Social Network: A Machine Learning Approach [Электрондық ресурс]. – URL: https://journals.sagepub.com/doi/abs/10.1177/0165551520917651 (қол жеткізу күні: 22.01.2025).

95. Dias D.S., Welikala M.D., Dias N.G. Identifying Racist Social Media Comments in Sinhala Language Using Text Analytics Models with Machine Learning // 2018 18th International Conference on Advances in ICT for Emerging Regions (ICTer). – IEEE, 2018. – P. 1–6.

96. Toktarova A., et al. Hate Speech Detection in Social Networks Using Machine Learning and Deep Learning Methods // International Journal of Advanced Computer Science and Applications. – 2023. – Vol. 14, № 5. – P.142-153.

97. Toktarova A., et al. Offensive Language Identification in Low Resource Languages Using Bidirectional Long-Short-Term Memory Network // International Journal of Advanced Computer Science and Applications. – 2023. – Vol. 14, № 6. – P.173-182.

98. Li Y., Chen Z. Performance Evaluation of Machine Learning Methods for Breast Cancer Prediction // Applied Computational Mathematics. – 2018. – Vol. 7, № 4. – P. 212–216.

99. Hand D.J., Christen P., Kirielle N. F*: An Interpretable Transformation of the F-measure // Machine Learning. – 2021. – Vol. 110, № 3. – P. 451–456.

100. Ydyrys A., Sultanova N. Identifying Spam Messages for Kazakh Language Using Hybrid Machine Learning Model // Natural and Technical Sciences. – 2023. – Vol. 62, № 1. – P. 142–150.

101. Bilakhanova A., Ydyrys A., Sultanova N. Kazakh Language, Question-Answering System, Natural Language Processing, Deep Learning Approach, Accuracy // Natural and Technical Sciences. – 2023. – Vol. 62, № 1. – P. 113–121.

102. Mühlhoff R. Human-Aided Artificial Intelligence: Or, How to Run Large Computations in Human Brains? Toward a Media Sociology of Machine Learning // New Media & Society. – 2020. – Vol. 22, № 10. – P. 1868–1884.

103. Del Vigna12, F., Cimino23, A., Dell'Orletta, F., Petrocchi, M., & Tesconi, M. (2017, January). Hate me, hate me not: Hate speech detection on facebook. In Proceedings of the first Italian conference on cybersecurity (ITASEC17) (pp. 86-95).

104. Watanabe H., Bouazizi M., Ohtsuki T. Hate Speech on Twitter: A Pragmatic Approach to Collect Hateful and Offensive Expressions and Perform Hate Speech Detection // IEEE Access. – 2018. – Vol. 6. – P. 13825–13835.

105. S Schieb C., Preuss M. Governing hate speech by means of counterspeech on Facebook //66th ica annual conference, at fukuoka, japan. – 2016. – P. 1-23.

106. Twitter Hate Speech Detection: A Systematic Review of Methods, Taxonomy Analysis, Challenges, and Opportunities [Электрондық ресурс]. – URL: https://ieeexplore.ieee.org/abstract/document/10025718 (қол жеткізу күні: 23.01.2025).

107. Alshalan R., Al-Khalifa H. A Deep Learning Approach for Automatic Hate Speech Detection in the Saudi Twittersphere // Applied Sciences. – 2020. – Vol. 10, № 23. – P. 8614.

108. Al-Hassan A., Al-Dossari H. Detection of Hate Speech in Social Networks: A Survey on Multilingual Corpus // Proceedings of the 6th International Conference on Computer Science and Information Technology. – ACM, 2019. – Vol. 10. – P. 10–5121.

109. Alrehili A. Automatic Hate Speech Detection on Social Media: A Brief Survey // 2019 IEEE/ACS 16th International Conference on Computer Systems and Applications (AICCSA). – 2019. – P. 1–6.

110. Schmidt A., Wiegand M. A Survey on Hate Speech Detection Using Natural Language Processing // Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media / Eds. Ku L.-W., Li C.-T. – Valencia, Spain: Association for Computational Linguistics, 2017. – P. 1–10.

111. Al-Garadi M.A., et al. Predicting Cyberbullying on Social Media in the Big Data Era Using Machine Learning Algorithms: Review of Literature and Open Challenges // IEEE Access. – 2019. – Vol. 7. – P. 70701–70718.

112. Novák V. Fuzzy Sets in Natural Language Processing // An Introduction to Fuzzy Logic Applications in Intelligent Systems / Eds. Yager R.R., Zadeh L.A. – Boston, MA: Springer US, 1992. – P. 185–200.

113. Kowsher M., et al. Lemmatization Algorithm Development for Bangla Natural Language Processing // 2020 Joint 9th International Conference on Informatics, Electronics & Vision (ICIEV) and 2020 4th International Conference on Imaging, Vision & Pattern Recognition (icIVPR). – IEEE, 2020. – P. 1–8.

114. An Overview of Stemming and Lemmatization Techniques [Электрондық ресурс] // Taylor & Francis. – URL: https://www.taylorfrancis.com/chapters/edit/10.1201/9781003430421-31/overview-stemming-lemmatization-techniques-vinay-kumar-pant-rupak-sharma-shakti-kundu (қол жеткізу күні: 27.01.2025).

115. Dai S., et al. AI-based NLP Section Discusses the Application and Effect of Bag-of-Words Models and TF-IDF in NLP Tasks // Journal of Artificial Intelligence General Science (JAIGS). – 2024. – Vol. 5, № 1. – P. 13–21.

116. Sultan D., et al. Cyberbullying-Related Hate Speech Detection Using Shallow-to-Deep Learning // Computers, Materials & Continua. – 2023. – Vol. 74, № 1. – P. 2115–2131.

117. Gebre B.G., et al. Improving Native Language Identification with TF-IDF Weighting // Proceedings of the Eighth Workshop on Innovative Use of NLP for Building Educational Applications / Eds. Tetreault J., Burstein J., Leacock C. – Atlanta, Georgia: Association for Computational Linguistics, 2013. – P. 216–223.

118. Soufyane A., Abdelhakim B.A., Ahmed M.B. An Intelligent Chatbot Using NLP and TF-IDF Algorithm for Text Understanding Applied to the Medical Field // Emerging Trends in ICT for Sustainable Development / Eds. Ben Ahmed M., Boudhir A.A., Alimi A.M. – Cham: Springer International Publishing, 2021. – P. 3–10.

119. Rodríguez A., Argueta C., Chen Y.-L. Automatic Detection of Hate Speech on Facebook Using Sentiment and Emotion Analysis // 2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC). – 2019. – P. 169–174.

120. Weir G., et al. Cloud-based Textual Analysis as a Basis for Document Classification // 2018 International Conference on High Performance Computing & Simulation (HPCS). – 2018. – P. 672–676.

121. Lloyd, S., Beckman, M., Pearl, D., Passonneau, R., Li, Z., & Wang, Z. (2022). Foundations for NLP-assisted formative assessment feedback for short-answer tasks in large-enrollment classes. arXiv preprint arXiv:2205.02829.

122. Al-Azzawy D.S., Al-Rufaye F.M.L. Arabic Words Clustering by Using K-means Algorithm // 2017 Annual Conference on New Trends in Information & Communications Technology Applications (NTICT). – IEEE, 2017. – P. 263–267.

123. Mendon-Plasek A. Irreducible Worlds of Inexhaustible Meaning: Early 1950s Machine Learning as Subjective Decision Making, Creative Imagining and Remedy for the Unforeseen // BJHS Themes. – 2023. – Vol. 8. – P. 65–80.

124. Toktarova A. et al. Automated Hate Speech Classification using Emotion Analysis in Social Media User Generated Texts //J. Theor. Appl. Inf. Technol. – 2022. – T. 100. – P. 6621-6634.

125. Sultan D., et al. Cyberbullying Detection and Prevention: Data Mining in Social Media // Proceedings of the 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence). – 2021. – P. 338–342.

126. Sultan D., et al. Cyberbullying-Related Hate Speech Detection Using Shallow-to-Deep Learning // Computers, Materials & Continua. – 2023. – Vol. 74, № 1. – P. 2115–2131.

127. Общее описание и реализация Word2Vec с помощью PyTorch [Электрондық ресурс] // Хабр. – URL: https://habr.com/ru/articles/801807/ (қол жеткізу күні: 23.01.2025).

128. Mikolov T., et al. Distributed Representations of Words and Phrases and Their Compositionality // Advances in Neural Information Processing Systems. – 2013. – Vol. 26. – P.1025-1036.

129. Tenney I., Das D., Pavlick E. BERT Rediscovers the Classical NLP Pipeline // arXiv:1905.05950. – arXiv, 2019. - Vol.42. – P.135-147.

130. González-Carvajal S., Garrido-Merchán E.C. Comparing BERT Against Traditional Machine Learning Text Classification // JCCE. – 2023. – Vol. 2, № 4. – P. 352–356.

131. Team I.M. За последний год Instagram представил новые функции, чтобы сделать платформу более доступной для людей с нарушениями зрения [Электрондық ресурс] // Internet Matters. – 2019. – URL: https://www.internetmatters.org/ru/hub/news-blogs/instagram-the-fight-against-online-bullying/ (қол жеткізу күні: 21.01.2025).

132. О нас | ВКонтакте [Электрондық ресурс] // VK. – URL: https://vk.com/about (қол жеткізу күні: 22.01.2025).

133. Nguyen Q. Hands-On Application Development with PyCharm: Accelerate Your Python Applications Using Practical Coding Techniques in PyCharm. –Packt Publishing Ltd, 2019. – 475 p.

134. Тоқтарова А.Б., et al. Онлайн контенттегі қазақ тілді бейәдеп пікірлерді машиналық оқытуда жинақтау // Вестник КазНПУ имени Абая. Серия «Физико-математические науки». – 2023. – Vol. 81, № 1. – P. 265–272.

135. Jahan M.S., Oussalah M. A Systematic Review of Hate Speech Automatic Detection Using Natural Language Processing // Neurocomputing. – 2023. – Vol. 546. – P. 126232.

136. Sharma D. et al. Hate Speech Detection Research in South Asian Languages: A Survey of Tasks, Datasets and Methods //ACM Transactions on Asian and Low-Resource Language Information Processing. – 2025. – Т. 24. – №. 3. – P. 1-44.

137. Toktarova A., et al. Automatic Offensive Language Detection in Online User Generated Contents // Journal of Theoretical and Applied Information Technology. – 2021. – Vol. 99, № 9. – P. 2054–2067.

138. Ensemble Based Hinglish Hate Speech Detection [Электрондық ресурс] // IEEE Conference Publication. – URL: https://ieeexplore.ieee.org/abstract/document/9432352 (қол жеткізу күні: 23.01.2025).

139. Sharma H. K. et al. Detecting hate speech and insults on social commentary using nlp and machine learning //Int J Eng Technol Sci Res. – 2017. – Т. 4. – №. 12. – С. 279-285.