IJTIMOIY TARMOQ MATNLARIDA XAVF DARAJALARINI DETALLASHTIRISHNING AVTOMATIK KLASSIFIKATSIYA SIFATIGA TA'SIRINI BAHOLASH

Babomurodov Ozod Jurayevich DSc, professor. Jizzax shaxar sun'iy intellekt yo'nalishi bo'yicha hokim o'rinbosari.
babomurodovozod@gmail.com
Qo'yliyeva Feruzaxon Alisher qizi Toshkent davlat agrar universiteti asissenti.
qoyliyevaferuza@gmail.com

Annotatsiya: *Ushbu tadqiqotda ijtimoiy tarmoq matnlarida xavf darajalarini detallashtirishning avtomatik klassifikatsiya sifatiga ta'siri o'rganildi. Matnlarning semantik xususiyatlari XLM-R modeli yordamida shakllantirilib, olingan embeddinglar asosida MiniBatch K-Means, Logistic Regression va Linear SVM algoritmlari K=2, K=3 va K=4 holatlarida sinovdan o'tkazildi. Natijalar Accuracy, Precision, Recall va F1-score ko'rsatkichlari asosida baholandi. Tajribalar xavf darajalari sonining oshishi sinflar orasidagi semantik yaqinlikni kuchaytirib, klassifikatsiya aniqligining pasayishiga olib kelishini ko'rsatdi. Natijalar xavf darajasini detallashtirish va avtomatik aniqlash sifati o'rtasida maqbul muvozanatni tanlash zarurligini ko'rsatadi.*

Kalit so'zlar: *ijtimoiy tarmoq, xavfli kontent, xavf darajasi, XLM-R, Logistic Regression, Linear SVM, MiniBatch K-Means.*

Аннотация: *В данном исследовании изучено влияние детализации уровней риска в текстах социальных сетей на качество автоматической классификации. Семантические характеристики текстов были сформированы с использованием модели XLM-R, а полученные эмбеддинги использованы для тестирования алгоритмов MiniBatch K-Means, Logistic Regression и Linear SVM при K=2, K=3 и K=4. Результаты оценивались на основе показателей Accuracy, Precision, Recall и F1-score. Эксперименты показали, что увеличение количества уровней риска повышает семантическую близость между классами, что приводит к снижению точности классификации. Полученные результаты свидетельствуют о необходимости выбора оптимального баланса между степенью детализации уровней риска и качеством их автоматического определения.*

Ключевые слова: *социальная сеть, опасный контент, уровень риска, XLM-R, Logistic Regression, Linear SVM, MiniBatch K-Means.*

Abstract: *This study investigates the impact of risk-level granularity in social media texts on automatic classification performance. The semantic features of the texts were generated using the XLM-R model, and the resulting embeddings were used to evaluate MiniBatch K-Means, Logistic Regression, and Linear SVM for K=2, K=3, and K=4. The results were evaluated using Accuracy, Precision, Recall, and F1-score. The experiments showed that increasing the number of risk levels increases semantic similarity between classes, leading to a decrease in classification accuracy. The results indicate the need to achieve an appropriate balance between risk-level granularity and the quality of automatic risk detection.*

Keywords: *social media, harmful content, risk level, XLM-R, Logistic Regression, Linear SVM, MiniBatch K-Means.*

Kirish. Ijtimoiy tarmoqlarda har kuni katta hajmdagi matnli ma'lumotlar shakllanib, ular orasida haqorat, tahdid, zo'ravonlikka undash va nafrat mazmunidagi xavfli xabarlar ham uchraydi. Bunday kontentni qo'lda aniqlash ko'p vaqt va inson resursini talab qilishi sababli uni avtomatik aniqlash tabiiy tilni qayta ishlashning muhim vazifalaridan biri hisoblanadi.

So'nggi yillarda transformer arxitekturasiga asoslangan modellar matnning semantik mazmunini ifodalashda samarali qo'llanilmoqda. Xususan, XLM-R modeli ko'p tilli matnlarni yagona vektor fazosida ifodalash imkonini berib, turli til va yozuv shakllari uchraydigan ijtimoiy tarmoq ma'lumotlarini tahlil qilish uchun mos keladi.

Xavfli kontentni aniqlashda matnlarni faqat xavfsiz va xavfli sinflarga ajratish xavf darajasi haqida yetarli ma’lumot bermaydi. Chunki xavfli xabarlarning mazmuni va xavflilik darajasi turlicha bo‘lishi mumkin.

Ushbu tadqiqotda ijtimoiy tarmoq matnlari XLM-R yordamida semantik embeddinglarga aylantirilib, xavf darajalari K=2, K=3 va K=4 ko‘rinishida shakllantiriladi. Har bir holatda MiniBatch K-Means, Logistic Regression va Linear SVM algoritmlari qo‘llanilib, natijalar Accuracy, Precision, Recall va F1-score ko‘rsatkichlari asosida taqqoslanadi.

Tadqiqotning asosiy maqsadi xavf darajalarini detallashtirish va ularni avtomatik aniqlash sifati o‘rtasidagi bog‘liqlikni aniqlash hamda xavf haqidagi metama’lumotning axborotliligi va klassifikatsiya sifati o‘rtasidagi muvozanatni baholashdan iborat.

Adabiyotlar tahlili. Ijtimoiy tarmoqlarda xavfli va zararli matnlarni avtomatik aniqlash tabiiy tilni qayta ishlash sohasida keng o‘rganilmoqda. Dastlab statistik belgilar va an’anaviy mashinali o‘qitish algoritmlaridan foydalanilgan bo‘lsa, transformer modellarining rivojlanishi matnning kontekstual va semantik xususiyatlarini chuqurroq ifodalash imkonini berdi.

Devlin va boshqalar tomonidan taklif qilingan BERT modeli matnni ikki tomonlama kontekst asosida ifodalash imkonini yaratdi [1]. Ko‘p tilli matnlar uchun ishlab chiqilgan XLM-R modeli esa turli tillardagi matnlarning semantik xususiyatlarini yagona vektor fazosida shakllantirish imkonini beradi [2]. Bu xususiyat turli til va yozuv shakllari uchraydigan ijtimoiy tarmoq ma’lumotlarini tahlil qilishda muhim ahamiyatga ega.

Zampieri va boshqalar ijtimoiy tarmoq xabarlarida haqoratli kontentni bir nechta darajada tasniflash imkoniyatini o‘rgangan [3]. Davidson va boshqalar esa nafrat va haqorat mazmunidagi matnlarni ajratishda semantik jihatdan yaqin sinflar orasidagi chegarani aniqlash murakkabligini ko‘rsatgan [4]. Ushbu tadqiqotlar xavfli kontentni detallashtirishda sinflar soni ortishi klassifikatsiya vazifasini murakkablashtirishi mumkinligini ko‘rsatadi.

Schmidt va Wiegand zararli kontentni aniqlashda leksik, sintaktik va semantik xususiyatlarning ahamiyatini tahlil qilib, yashirin mazmun va kontekst avtomatik aniqlash jarayonini murakkablashtirishini ta’kidlagan [5]. Fortuna va Nunes esa zararli kontent sinflarining yagona ta’rifi mavjud emasligini hamda turli tadqiqotlarda ikki yoki undan ortiq sinflardan foydalanilishini ko‘rsatgan [6].

Transformer yordamida olingan yuqori o‘lchamli embeddinglarni an’anaviy klassifikatsiya algoritmlari bilan birgalikda qo‘llash ham amaliy ahamiyatga ega. Joachims yuqori o‘lchamli matn belgilarini klassifikatsiya qilishda Support Vector Machine algoritmining samaradorligini ko‘rsatgan [7]. Katta hajmdagi ma’lumotlarni klasterlash uchun esa Sculley tomonidan MiniBatch K-Means yondashuvi taklif qilingan bo‘lib, u ma’lumotlarning kichik qismlaridan foydalanish orqali hisoblash xarajatlarini kamaytiradi [8].

Ko‘rib chiqilgan tadqiqotlarda xavfli kontentni aniqlash va klassifikatsiya sifatini oshirishga katta e’tibor qaratilgan. Biroq bir xil semantik embeddinglar va ma’lumotlar asosida xavf darajalari sonini K=2, K=3 va K=4 ko‘rinishida bosqichma-bosqich oshirishning klassifikatsiya sifatiga ta’siri kamroq o‘rganilgan.

Shu sababli ushbu tadqiqotda XLM-R embeddinglari asosida MiniBatch K-Means, Logistic Regression va Linear SVM algoritmlarining K=2, K=3 va K=4 holatlaridagi natijalari taqqoslanib, xavf darajalarini detallashtirish bilan avtomatik klassifikatsiya sifati o‘rtasidagi bog‘liqlik baholanadi.

Tadqiqot metodologiyasi. Tadqiqotda ijtimoiy tarmoq matnlaridagi xavf darajalarini avtomatik aniqlash va xavf sinflari sonining klassifikatsiya sifatiga ta’sirini baholash uchun bir nechta ketma-ket bosqichlardan iborat tajriba tashkil etildi. Tajribaning umumiy jarayoni matnli ma’lumotlarni yig‘ish, dastlabki qayta ishlash, semantik vektorlarni shakllantirish, xavf darajalarini belgilash, klassifikatsiya va natijalarni baholash bosqichlaridan tashkil topdi.

Tajriba uchun Telegram ijtimoiy tarmog‘idan yig‘ilgan 10 000 ta matnli xabardan foydalanildi. Ijtimoiy tarmoq matnlari odatda norasmiy shaklda yozilishi, ortiqcha belgilar, havolalar, foydalanuvchi nomlari va boshqa elementlarni o‘z ichiga olishi sababli dastlab ma’lumotlarga oldindan ishlov berildi. Ushbu bosqichda matnlar kichik harflarga o‘tkazildi, URL

havolalar, @foydalanuvchi nomlari, hashtaglar, ortiqcha belgilar va takroriy xabarlar tozalandi. Shuningdek, lotin va kirill yozuvida uchraydigan matnlarning xususiyatlari hisobga olindi.

Dastlabki qayta ishlashdan o‘tgan matnlarning semantik xususiyatlarini shakllantirish uchun XLM-R modeli qo‘llanildi. Har bir matn modelga uzatilib, uning mazmunini ifodalovchi ko‘p o‘lchamli embedding vektori olindi. Natijada matnlar bevosita so‘zlar ketma-ketligi sifatida emas, balki ularning semantik mazmunini ifodalovchi sonli vektorlar ko‘rinishida keyingi algoritmlarga uzatildi.

Tadqiqotning asosiy farqli jihati xavf darajalarini turli detallashtirish holatlarida tekshirishdan iborat. Buning uchun uchta alohida tajriba tashkil qilindi: K=2, K=3 va K=4. K=2 holatida ma’lumotlar ikkita asosiy xavf guruhiga ajratildi. K=3 holatida xavf haqidagi ma’lumot uchta darajada, K=4 holatida esa to‘rtta darajada shakllantirildi. Shu orqali xavf darajalari sonining oshishi modellar tomonidan ushbu sinflarni ajratish sifatiga qanday ta’sir qilishi tekshirildi.

Hosil qilingan XLM-R embeddinglari asosida uchta algoritm qo‘llanildi: MiniBatch K-Means, Logistic Regression va Linear SVM. Ushbu algoritmlarning tanlanishi bir xil semantik belgilar asosida klasterlash va klassifikatsiya yondashuvlarining natijalarini taqqoslash imkonini beradi.

MiniBatch K-Means katta hajmdagi ma’lumotlarni nisbatan tez guruhlash imkoniyatiga ega bo‘lgan K-Means algoritmining optimallashtirilgan ko‘rinishi sifatida qo‘llanildi. Algoritm har bir iteratsiyada ma’lumotlarning to‘liq to‘plamidan emas, balki kichik qismlaridan foydalanadi. Tajribada klasterlar soni mos ravishda K=2, K=3 va K=4 qilib belgilandi.

Logistic Regression nazorat ostida o‘qitiladigan klassifikatsiya algoritmi sifatida qo‘llanildi. XLM-R orqali olingan embedding vektorlari ushbu algoritm uchun kiruvchi belgilar sifatida ishlatildi. Algoritmning vazifasi har bir matnning mavjud xavf sinflaridan biriga tegishliligini aniqlashdan iborat bo‘ldi.

Linear SVM ham XLM-R embeddinglari asosida o‘qitildi. Ushbu algoritm yuqori o‘lchamli belgilar fazosida sinflarni ajratuvchi optimal chegarani aniqlashga asoslanadi. Xavf darajalari soni oshirilganda sinflar o‘rtasidagi chegaralarning murakkablashishi sababli Linear SVM natijalarining o‘zgarishi ham alohida kuzatildi.

Har bir K qiymati uchun algoritmlar bir xil ma’lumotlar va bir xil semantik embeddinglar asosida sinovdan o‘tkazildi. Bu turli tajriba sharoitlarining natijaga ta’sirini kamaytirib, asosiy o‘zgaruvchi sifatida xavf darajalari sonini tahlil qilish imkonini berdi.

Modellarning ishlash samaradorligini baholash uchun Accuracy, Precision, Recall va F1-score ko‘rsatkichlaridan foydalanildi. Accuracy barcha namunalar ichida to‘g‘ri aniqlangan namunalar ulushini ifodalasa, Precision model tomonidan muayyan sinfga tegishli deb topilgan namunalar orasida to‘g‘ri aniqlanganlar ulushini ko‘rsatadi. Recall esa mavjud sinf namunalarining qanchasi model tomonidan to‘g‘ri topilganini ifodalaydi. Precision va Recall o‘rtasidagi muvozanatni baholash uchun F1-score ko‘rsatkichidan foydalanildi.

Tajribaning umumiy ketma-ketligi quyidagicha shakllantirildi:

Telegram matnlari → dastlabki qayta ishlash → XLM-R → embeddinglarni shakllantirish → K=2, K=3, K=4 xavf darajalari → MiniBatch K-Means, Logistic Regression, Linear SVM → Accuracy, Precision, Recall, F1-score → natijalarni taqqoslash.

Ushbu tajriba sxemasi bir xil ma’lumotlar to‘plamida xavf darajalari sonini o‘zgartirish orqali klassifikatsiya sifatidagi o‘zgarishlarni bevosita kuzatish imkonini beradi. Natijada xavf darajalarini yanada batafsil shakllantirish natijasida olinadigan qo‘shimcha axborot bilan avtomatik aniqlash sifatidagi o‘zgarish o‘rtasidagi bog‘liqlik baholanadi.

Tajriba natijalari va ularning tahlili. Tajribaning asosiy maqsadi xavf darajalari sonining oshishi avtomatik klassifikatsiya sifatiga qanday ta’sir qilishini aniqlashdan iborat. Shu maqsadda XLM-R modeli yordamida shakllantirilgan embeddinglar asosida MiniBatch K-Means, Logistic Regression va Linear SVM algoritmlari K=2, K=3 va K=4 holatlarida alohida sinovdan o‘tkazildi.

Har bir tajribada bir xil dastlabki ma’lumotlar va bir xil XLM-R embeddinglaridan foydalanildi. Faqat xavf darajalarining soni o‘zgartirildi. Bunday yondashuv natijalardagi

o‘zgarishlarni aynan xavf darajasining detallashtirilishi bilan bog‘liq holda tahlil qilish imkonini berdi.

1-jadval. Xavf darajalari soniga bog‘liq holda algoritmlarning Accuracy ko‘rsatkichlari

Xavf darajalari	MiniBatch K-Means	Logistic Regression	Linear SVM
K=2	85,60 %	84,00 %	85,00 %
K=3	49,20 %	77,00 %	82,00 %
K=4	39,60 %	67,00 %	67,00 %

1-jadval natijalaridan ko‘rinadiki, xavf darajalari ikkita sinfga ajratilgan K=2 holatida barcha algoritmlar nisbatan yuqori natija ko‘rsatdi. MiniBatch K-Means uchun Accuracy 85,60 %, Logistic Regression uchun 84,00 % va Linear SVM uchun 85,00 % ni tashkil etdi. Ushbu holatda uchta algoritm natijalari orasidagi farq katta emas.

K=3 holatiga o‘tilganda natijalarda sezilarli o‘zgarish kuzatildi. MiniBatch K-Means aniqligi 49,20 % gacha pasaydi. Logistic Regression 77,00 %, Linear SVM esa 82,00 % natija ko‘rsatdi. Ushbu holatda eng yuqori natija Linear SVM algoritmida qayd etildi. K=2 bilan taqqoslaganda Linear SVM aniqligi nisbatan kam pasayganligi ushbu algoritmning xavf darajalari soni oshganda sinflarni ajratish imkoniyati yuqoriroq ekanligini ko‘rsatdi.

K=4 holatida klassifikatsiya vazifasi yanada murakkablashdi. MiniBatch K-Means natijasi 39,60 % gacha kamaydi. Logistic Regression va Linear SVM algoritmlarida esa Accuracy 67,00 % ni tashkil etdi. Demak, xavf darajalari to‘rtta guruhga ajratilganda barcha algoritmlarda aniqlikning pasayishi kuzatildi.

Natijalarning o‘zgarishini yanada aniq ko‘rish uchun K=2 holatidagi natijalar K=4 bilan taqqoslandi. MiniBatch K-Means aniqligi 85,60 % dan 39,60 % gacha, ya’ni 46 foiz punktga kamaydi. Logistic Regression natijasi 84,00 % dan 67,00 % gacha, ya’ni 17 foiz punktga pasaydi. Linear SVM natijasi esa 85,00 % dan 67,00 % gacha kamayib, farq 18 foiz punktni tashkil etdi.

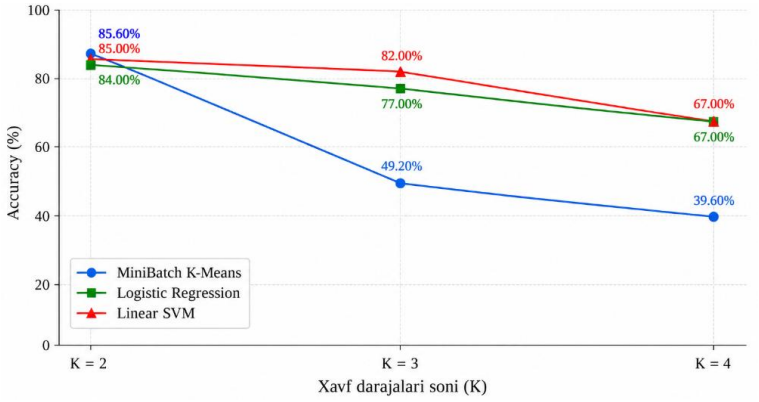
MiniBatch K-Means natijasining keskin pasayishi klasterlash algoritmining ma’lumotlarni oldindan belgilangan sinf chegaralariga emas, embedding fazosidagi o‘xshashlik va masofalarga asoslanib guruhlashi bilan izohlanishi mumkin. Xavf darajalari soni oshirilganda semantik jihatdan bir-biriga yaqin bo‘lgan matnlarni alohida klasterlarga ajratish qiyinlashadi. Natijada hosil qilingan klasterlar belgilangan xavf darajalari bilan to‘liq mos kelmasligi mumkin.

Logistic Regression va Linear SVM nazorat ostida o‘qitiladigan algoritmlar bo‘lganligi sababli K=3 va K=4 holatlarida MiniBatch K-Meansga nisbatan yuqoriroq natijalarni saqlab qoldi. Bunda modelga oldindan ma’lum sinf belgilari asosida o‘qitish imkoniyati berilishi sinflar orasidagi chegaralarni yaxshiroq shakllantirishga yordam berdi.

Linear SVM ayniqsa K=3 holatida 82,00 % Accuracy bilan boshqa algoritmlardan yuqori natija ko‘rsatdi. Bu natija uchta xavf darajasidan foydalanish xavf haqidagi ma’lumotni ikki sinfga nisbatan batafsilroq ifodalash bilan birga, klassifikatsiya sifatini ham nisbatan yuqori darajada saqlash imkonini berishini ko‘rsatadi.

K=4 holatida Logistic Regression va Linear SVM natijalarining bir xil 67,00 % darajaga tushishi to‘rtta xavf sinfi orasidagi farqlarni aniqlash murakkablashganligini ko‘rsatadi. Xavf darajalari soni ko‘paygan sari qo‘shni sinflardagi matnlar semantik jihatdan bir-biriga yaqinlashadi. Masalan, o‘rta xavf va yuqori xavf chegarasida joylashgan ayrim matnlarda har ikkala sinfga xos belgilar mavjud bo‘lishi mumkin. Bunday holatlarda model uchun aniq sinfni tanlash qiyinlashadi.

Olingan natijalar xavf darajasini detallashtirish ikki tomonlama ta’sirga ega ekanligini ko‘rsatdi. Bir tomondan, K qiymatining oshirilishi har bir matn uchun xavf haqida batafsilroq metama’lumot shakllantirish imkonini beradi. Ikkinchi tomondan esa sinflar sonining ko‘payishi avtomatik klassifikatsiya vazifasini murakkablashtirib, aniqlik ko‘rsatkichining pasayishiga olib keladi.



1-rasm. Xavf darajalari soniga bog'liq holda algoritmlarning klassifikatsiya aniqligi (Accuracy)

Tajriba natijalariga koʻra, K=2 holati yuqori aniqlik talab qilinadigan dastlabki xavfli kontentni aniqlash vazifalari uchun samarali hisoblanadi. K=3 holati esa xavfni batafsilroq tavsiflash zarur boʻlgan vazifalarda muvozanatli variant sifatida qaralishi mumkin. Ayniqsa, Linear SVM algoritmining K=3 holatida 82,00 % natija koʻrsatishi ushbu yondashuvning amaliy qoʻllanilishi mumkinligini koʻrsatadi. K=4 holatida xavf haqida yanada batafsil ma'lumot hosil qilinsa-da, klassifikatsiya aniqligining sezilarli pasayishi kuzatiladi.

Shunday qilib, tajriba xavf darajalari sonini shunchaki oshirish har doim ham avtomatik tahlil sifatini yaxshilamasligini koʻrsatdi. Xavf darajasini detallashtirishda hosil qilinadigan metama'lumotning axborotliligi va uni avtomatik aniqlash sifati oʻrtasidagi muvozanatni hisobga olish zarur.

Muhokamalar. Oʻtkazilgan tajribalar xavf darajalarini detallashtirish avtomatik klassifikatsiya sifatiga bevosita ta'sir qilishini koʻrsatdi. K=2 holatida uchala algoritmda ham Accuracy koʻrsatkichi 84–85,6 % oralig'ida boʻlib, xavfli kontentni ikkita asosiy guruhga ajratish nisbatan barqaror natija berdi. Xavf darajalari soni K=3 va K=4 gacha oshirilganda esa algoritmlar orasidagi farq aniqroq namoyon boʻldi.

K=2 holatida MiniBatch K-Meansning 85,60 % natija koʻrsatishi embedding fazosida ikkita asosiy guruhni ajratish imkoniyati mavjudligini koʻrsatadi. Biroq K=3 va ayniqsa K=4 holatiga oʻtilganda ushbu algoritm natijasi keskin pasaydi. Bu xavf darajalarining koʻpayishi bilan matnlar orasidagi semantik chegaralar klasterlash uchun yetarlicha aniq shakllanmasligini koʻrsatadi.

Nazorat ostida oʻqitiladigan Logistic Regression va Linear SVM algoritmlarida pasayish nisbatan kichik boʻldi. Linear SVM K=3 holatida 82,00 % natija bilan eng barqaror algoritm sifatida namoyon boʻldi. Ushbu holatda K=2 ga nisbatan xavf haqida batafsilroq ma'lumot shakllantiriladi, shu bilan birga klassifikatsiya aniqligi ham yuqori darajada saqlanadi.

K=4 holatida Logistic Regression va Linear SVM algoritmlarining aniqligi 67,00 % gacha pasaydi. Bu natija xavf darajalarini haddan tashqari detallashtirishda qoʻshni sinflar orasidagi semantik farq kamayishini koʻrsatadi. Bunda model umumiy xavf belgisini aniqlashi mumkin, ammo ushbu matnning aynan qaysi xavf darajasiga tegishli ekanligini aniqlash murakkablashadi.

Shu sababli xavf darajalari sonini tanlashda faqat klassifikatsiya aniqligiga asoslanish yetarli emas. K=2 yuqoriroq aniqlikni ta'minlaydi, ammo xavf haqida shakllantiriladigan ma'lumot nisbatan umumiy boʻladi. K=4 koʻproq metama'lumot hosil qilish imkonini beradi, biroq aniqlik sezilarli darajada kamayadi. Tajriba doirasida K=3 holati ushbu ikki talab oʻrtasidagi muvozanatli yechim sifatida namoyon boʻldi.

Olingan natijalar xavf darajasini detallashtirish va uni avtomatik aniqlash sifati oʻrtasida bogʻliqlik mavjudligini koʻrsatadi. Demak, ijtimoiy tarmoq matnlarini avtomatik tahlil qilish tizimini ishlab chiqishda xavf sinflari soni tizimdan olinishi kerak boʻlgan metama'lumot hajmi va klassifikatsiya sifatiga qoʻyiladigan talablar asosida tanlanishi kerak.

Xulosalar. Ushbu tadqiqotda ijtimoiy tarmoq matnlarida xavf darajalarini detallashtirishning avtomatik klassifikatsiya sifatiga ta’siri eksperimental ravishda baholandi. Matnlarning semantik xususiyatlarini shakllantirish uchun XLM-R modeli, keyingi tahlil bosqichida esa MiniBatch K-Means, Logistic Regression va Linear SVM algoritmlaridan foydalanildi. Tajribalar xavf darajalari K=2, K=3 va K=4 holatlarida amalga oshirildi.

Olingan natijalarga ko‘ra, K=2 holatida barcha algoritmlar yuqoriroq aniqlik ko‘rsatdi. MiniBatch K-Means 85,60 %, Logistic Regression 84,00 % va Linear SVM 85,00 % natijaga erishdi. Xavf darajalari soni oshirilganda klassifikatsiya sifati pasayishi kuzatildi. K=3 holatida Linear SVM 82,00 % bilan eng yuqori natijani saqlab qoldi. K=4 holatida esa Logistic Regression va Linear SVM algoritmlarining aniqligi 67,00 % ni tashkil etdi.

Tajriba natijalari xavf darajalarini ko‘paytirish natijasida xavf haqidagi metama’lumotning axborotliligi ortishi bilan bir vaqtda uni avtomatik aniqlash murakkabligi ham oshishini ko‘rsatdi. Ayniqsa, semantik jihatdan bir-biriga yaqin xavf darajalarini alohida sinflarga ajratishda klassifikatsiya aniqligining pasayishi kuzatildi.

K=2 holati xavfli kontentni dastlabki aniqlash uchun yuqori samaradorlikni ko‘rsatsa, K=3 holati xavf haqida batafsilroq ma’lumot olish va klassifikatsiya sifatini saqlash o‘rtasidagi muvozanatli variant sifatida namoyon bo‘ldi. K=4 holatida xavf darajasi yanada batafsil shakllantiriladi, biroq klassifikatsiya aniqligining sezilarli pasayishi ushbu yondashuvni qo‘llashda qo‘shimcha takomillashtirish zarurligini ko‘rsatadi.

Shunday qilib, ijtimoiy tarmoq matnlarida xavf darajasini shakllantirishda sinflar sonini maksimal darajada oshirish emas, balki xavf haqidagi metama’lumotning axborotliligi va uni avtomatik aniqlash sifati o‘rtasidagi optimal muvozanatni tanlash muhim hisoblanadi. Keyingi tadqiqotlarda xavf darajalari orasidagi semantik chegaralarni aniqroq shakllantirish va ko‘p sinfli klassifikatsiya sifatini oshirish usullarini ishlab chiqish rejalashtirilgan.

Foydalanilgan adabiyotlar.

1. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of NAACL-HLT, 4171–4186.
2. Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., & Stoyanov, V. (2020). Unsupervised cross-lingual representation learning at scale. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 8440–8451. https://doi.org/10.18653/v1/2020.acl-main.747
3. Zampieri, M., Malmasi, S., Nakov, P., Rosenthal, S., Farra, N., & Kumar, R. (2019). Predicting the type and target of offensive posts in social media. Proceedings of NAACL-HLT, 1415–1420.
4. Davidson, T., Warmsley, D., Macy, M., & Weber, I. (2017). Automated hate speech detection and the problem of offensive language. Proceedings of the International AAAI Conference on Web and Social Media, 11(1), 512–515.
5. Schmidt, A., & Wiegand, M. (2017). A survey on hate speech detection using natural language processing. Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media, 1–10. https://doi.org/10.18653/v1/W17-1101
6. Fortuna, P., & Nunes, S. (2018). A survey on automatic detection of hate speech in text. ACM Computing Surveys, 51(4). https://doi.org/10.1145/3232676
7. Joachims, T. (1998). Text categorization with support vector machines: Learning with many relevant features. Proceedings of the European Conference on Machine Learning, 137–142.
8. Sculley, D. (2010). Web-scale K-means clustering. Proceedings of the 19th International Conference on World Wide Web, 1177–1178.
9. Waseem, Z., & Hovy, D. (2016). Hateful symbols or hateful people? Predictive features for hate speech detection on Twitter. Proceedings of the NAACL Student Research Workshop, 88–93. https://doi.org/10.18653/v1/N16-2013

10. MacAvaney, S., Yao, H.-R., Yang, E., Russell, K., Goharian, N., & Frieder, O. (2019). Hate speech detection: Challenges and solutions. PLOS ONE, 14(8), e0221152. https://doi.org/10.1371/journal.pone.0221152
11. Vidgen, B., & Derczynski, L. (2020). Directions in abusive language training data, a systematic review: Garbage in, garbage out. PLOS ONE, 15(12), e0243300. https://doi.org/10.1371/journal.pone.0243300
12. Mathew, B., Saha, P., Yimam, S. M., Biemann, C., Goyal, P., & Mukherjee, A. (2021). HateXplain: A benchmark dataset for explainable hate speech detection. Proceedings of the AAAI Conference on Artificial Intelligence, 35(17), 14867–14875. https://doi.org/10.1609/aaai.v35i17.17745
13. Barbieri, F., Espinosa Anke, L., & Camacho-Collados, J. (2022). XLM-T: Multilingual language models in Twitter for sentiment analysis and beyond. Proceedings of the Thirteenth Language Resources and Evaluation Conference (LREC), 258–266.
14. Mozafari, M., Farahbakhsh, R., & Crespi, N. (2020). A BERT-based transfer learning approach for hate speech detection in online social media. Complex Networks and Their Applications VIII, 928–940.
15. Caselli, T., Basile, V., Mitrović, J., & Granitzer, M. (2021). HateBERT: Retraining BERT for abusive language detection in English. Proceedings of the 5th Workshop on Online Abuse and Harms (WOAH), 17–25.