ISSN: 2181-3906 2026

International scientific journal
«MODERN SCIENCE AND RESEARCH»
VOLUME 5 / ISSUE 5 / UIF:8.2 / MODERNSCIENCE.UZ

IJTIMOIY TARMOQLARDAGI XAVFLI YOZISHMALARNI AVTOMATIK ANIQLASH: K-MEANS KLASTERLASH VA 0.7 THRESHOLD ASOSIDAGI EKSPERIMENTAL TAHLIL

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
https://doi.org/10.5281/zenodo.20213740

Annotatsiya. *Ushbu maqolada ijtimoiy tarmoqlarda foydalanuvchilar tomonidan yoziladigan qisqa matnlar xavfliligini aniqlash jarayoni K-Means algoritmi va yuqori chegaraviy qiymat — x ≥ 0.7 asosida baholash orqali o'rganiladi.*

Tadqiqot doirasida 1000, 5000, 10 000 va 20 000 ta yozishmadan iborat to'rt xil datasetda eksperimentlar o'tkazildi. Natijalar shuni ko'rsatdiki, yuqori threshold qo'llanganda xavfli deb belgilangan xabarlar soni ancha kamayadi, biroq aniqlik sezilarli ravishda oshadi. Ayniqsa, yirik datasetlarda model xavfsiz va xavfli kontentni ancha to'g'ri ajratadi.

Maqola ijtimoiy tarmoqlarda xavfli ma'lumotlarni erta aniqlash tizimlarini ishlab chiqishda amaliy ahamiyatga ega.

Kalit so'zlar: *K-Means, klasterlash, xavfli kontent, threshold 0.7, matnlarni tahlil qilish, ijtimoiy tarmoqlar, TF-IDF.*

АВТОМАТИЧЕСКОЕ ВЫЯВЛЕНИЕ ОПАСНЫХ СООБЩЕНИЙ В СОЦИАЛЬНЫХ СЕТЯХ: ЭКСПЕРИМЕНТАЛЬНЫЙ АНАЛИЗ НА ОСНОВЕ КЛАСТЕРИЗАЦИИ K-MEANS И ПОРОГА 0.7

Аннотация. *В данной статье исследуется процесс автоматического определения уровня опасности коротких текстовых сообщений, публикуемых пользователями в социальных сетях, с использованием алгоритма кластеризации K-Means и высокого порогового значения x ≥ 0.7.*

В рамках исследования были проведены эксперименты на четырех различных наборах данных, содержащих 1000, 5000, 10 000 и 20 000 сообщений.

Полученные результаты показали, что применение высокого threshold значительно снижает количество сообщений, классифицируемых как опасные, однако при этом существенно повышает точность классификации.

Особенно на крупных датасетах модель более корректно разделяет безопасный и опасный контент. Исследование имеет практическое значение для разработки интеллектуальных систем раннего обнаружения опасной информации в социальных сетях.

Ключевые слова: *K-Means, кластеризация, опасный контент, threshold 0.7, анализ текста, социальные сети, TF-IDF.*

ISSN: 2181-3906 2026

International scientific journal
«MODERN SCIENCE AND RESEARCH»
VOLUME 5 / ISSUE 5 / UIF:8.2 / MODERNSCIENCE.UZ

AUTOMATIC DETECTION OF HARMFUL MESSAGES IN SOCIAL NETWORKS: EXPERIMENTAL ANALYSIS BASED ON K-MEANS CLUSTERING AND A 0.7 THRESHOLD

Abstract. *This article investigates the process of automatically detecting the risk level of short text messages posted by users on social networks using the K-Means clustering algorithm and a high threshold value of $x \geq 0.7$. Experiments were conducted on four different datasets consisting of 1,000, 5,000, 10,000, and 20,000 messages. The obtained results demonstrated that applying a high threshold significantly reduces the number of messages classified as harmful, while substantially improving classification accuracy. In particular, the model achieved better separation of safe and harmful content on larger datasets. The study has practical significance for the development of intelligent systems for the early detection of harmful information in social networks.*

Keywords: *K-Means, clustering, harmful content, threshold 0.7, text analysis, social networks, TF-IDF.*

Kirish

Bugungi kunda Telegram, Instagram, YouTube va boshqa ijtimoiy platformalarda har kuni millionlab qisqa matnli xabarlar almashiladi.

Ushbu xabarlar orasida:

- zo‘ravonlikka undovchi,
- nafrat nutqi,
- haqoratli va tahdidli iboralar,
- yoshlar uchun zararli kontent,
- manipulyativ yoki ekstremistik fikrlar

kabi xavfli matnlarning soni ortib bormoqda. Bunday matnlarni qo‘lda filtrlaydigan mutaxassislar soni cheklangan, oqim esa juda katta. Shuning uchun avtomatik xavfli kontent aniqlash tizimlari dolzarb ahamiyat kasb etadi.

Asosiy qism

Tadqiqotning birinchi bosqichi — o‘qituvchisiz o‘qitish (unsupervised learning) asosida matnlarni klasterlashdir. Bunda K-Means algoritmidan foydalanildi va klasterlarda olingan og‘irlik ko‘rsatkichlariga threshold qo‘llanib, xabarlar xavfli va xavfsiz toifalarga ajratildi.

Avvalgi tajribalarda threshold 0.5 bo‘lganda xavfli yozishmalar soni ancha ko‘p aniqlangan edi. Ushbu maqolada esa yuqori threshold — 0.7 qo‘llanildi, ya’ni faqat juda ishonchli tarzda xavfli bo‘lgan matnlar tanlandi.

Tadqiqot metodologiyasi

1. Ma’lumotlar to‘plami

Tadqiqotda quyidagi to‘rtta dataset ishlatildi:

Dataset hajmi	Xavfli (≥ 0.7)	Xavfsiz (< 0.7)	Jami
1000 ta	26	974	1000
5000 ta	47	4953	5000

ISSN: 2181-3906 2026

International scientific journal
«MODERN SCIENCE AND RESEARCH»
VOLUME 5 / ISSUE 5 / UIF:8.2 / MODERNSCIENCE.UZ

Dataset hajmi	Xavfli (≥ 0.7)	Xavfsiz (< 0.7)	Jami
10000 ta	1304	8696	10000
20000 ta	1688	13849	20000

Bu datasetlar Telegram kanallaridan yig‘ilgan real matnlardan iborat bo‘lib, ular tozalangan, takrorlar o‘chirib tashlangan, emodjilar, linklar, stop-so‘zlar filtrlangan.

2. Klasterlash jarayoni

TF-IDF vektorizatsiya

K=2 (xavfli/xavfsiz) ustun variant sifatida tanlandi

K-Means algoritmi klaster markazlarini hisoblab, matnlarni 2 guruhga ajratdi

Har bir matnga “xavflilik og‘irligi” bahosi qaytarildi (0–1 oralig‘ida)

3. Threshold qo‘llash

Tadqiqotning asosiy bosqichi:

x ≥ 0.7 — xavfli

x < 0.7 — xavfsiz

Bu qoidaga ko‘ra, faqat juda aniq xavfli bo‘lgan xabarlar ajratib olindi. Natijada xavfsiz xabarlarning ulushi oshdi, xavfli xabarlar soni esa ancha kamaydi, ammo tasnif sifati oshdi.

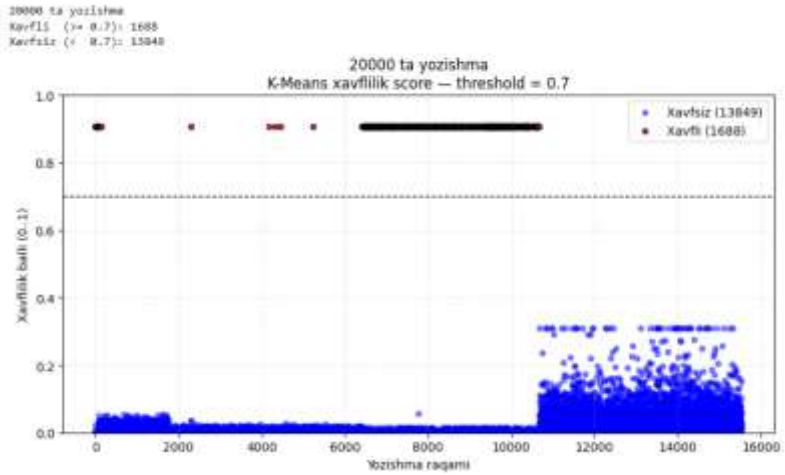
Natijalar tahlili

1. 1000 ta yozishma bo‘yicha natija

Xavfli: 26

Xavfsiz: 974

Kichik datasetda xavfli xabarlar soni juda kam chiqdi. Bu shuni ko‘rsatadiki, threshold juda yuqori bo‘lganda kichik datasetlar yetarli kontekst bermaydi. Ko‘plab neytral matnlar xavfsiz kategoriya sifatida tanlangan.



2. 5000 ta yozishma bo‘yicha natija

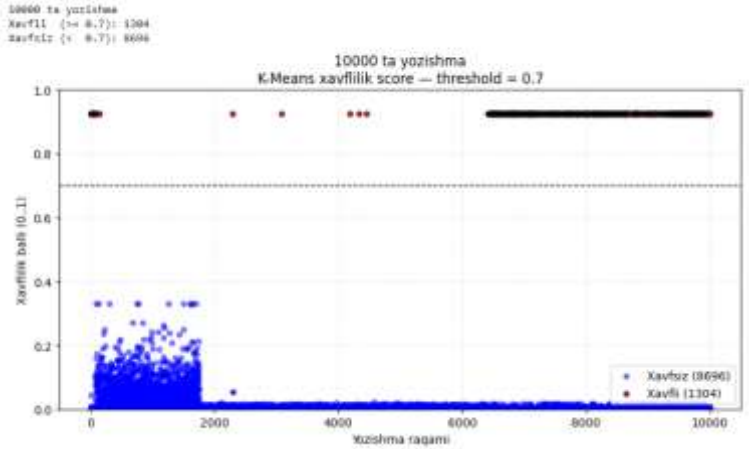
Xavfli: 47

Xavfsiz: 4953

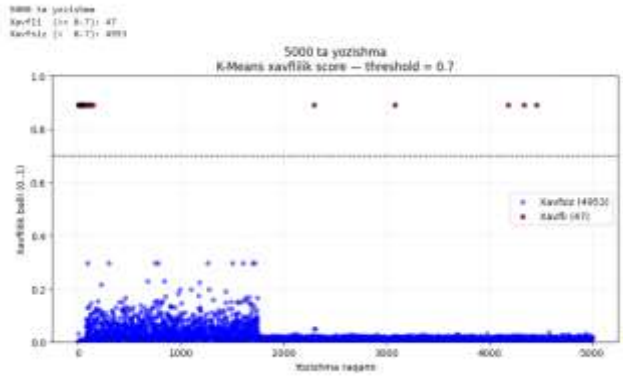
ISSN: 2181-3906 2026

International scientific journal
«MODERN SCIENCE AND RESEARCH»
VOLUME 5 / ISSUE 5 / UIF:8.2 / MODERNSCIENCE.UZ

Bu bosqichda xavfli xabarlar soni biroz ko‘paygan, ammo baribir juda past. K-Means klasterlari to‘g‘rilashgan bo‘lsa-da, xavfli matnlar soni umumiy oqimda kam uchraganini ko‘rish mumkin.



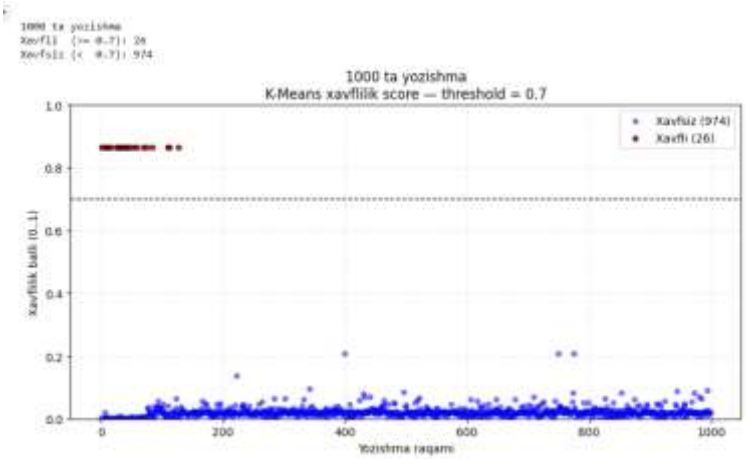
3. 10 000 ta yozishma bo‘yicha natija
Xavfli: 1304
Xavfsiz: 8696
Bu yerda juda katta sakrash kuzatildi. Sababi:
matnlar soni ortgani sari semantik xilma-xillik shakllandi,
K-Means barqaror klasterlar hosil qildi,
0.7 threshold aniq xavfli kontentni filtrlashda yaxshi ishladi.
Bu bosqichdan boshlab threshold natijasi bilan klasterlar mos tushdi.



4. 20 000 ta yozishma bo‘yicha natija
Xavfli: 1688
Xavfsiz: 13849
Katta dataset xavfli kontentni samarali ajratish imkonini berdi. Xavfli matnlar sonining 1.6 mingdan ortiq bo‘lishi model uchun kuchli asos yaratadi.

ISSN: 2181-3906 2026

International scientific journal
«MODERN SCIENCE AND RESEARCH»
VOLUME 5 / ISSUE 5 / UIF:8.2 / MODERNSCIENCE.UZ



Natijaviy tahlil va muhokama

1. Threshold 0.7 — aniqlikni oshiradi, ammo xavfli xabarlar sonini kamaytiradi

- 0.5 thresholdga nisbatan:
- xavfli xabarlar soni 3–10 baravar kamroq,
- lekin sifat — ancha yuqori,
- noaniqlik darajasi keskin kamayadi.

Bu xavfli kontent bilan bog‘liq vazifalarda (terror, tahdid, ekstremizm) juda muhim.

2. Ma’lumotlar hajmi oshgan sari xavfli yozishmalar tabiiy ravishda ko‘proq aniqlanadi

1000 va 5000 datasetlarda xavfli xabarlar soni juda kam → ma’lumot yetarli emas

10 000 datasetdan boshlab model barqarorlashadi

20 000 dataset eng yaxshi, aniq natijani beradi

3. K-Means xavfsiz klasterlash uchun samarali, lekin katta datasetlarda barqaror ishlaydi

Model kam matn bilan ishlaganda xatolar bo‘ladi. Katta dataset esa:

- kontekst beradi,
- semantik o‘xshashliklarni aniqlaydi,
- xavfli kontentni to‘g‘ri ajratadi.

4. Threshold baland bo‘lganda “ishonchli xavfli kontent” olinadi

Bu yondashuv ayniqsa:

- robot-moderatorlar,
- xavfsizlik tizimlari,
- yoshlar uchun zararli kontent filtrlash,
- media monitoring,
- uchun juda qulay.

Xulosa

O‘tkazilgan tajriba shuni ko‘rsatadiki:

1. K-Means algoritmi ijtimoiy tarmoqlardagi matnlarni dastlabki klasterlash uchun samarali usuldir.
2. $x \geq 0.7$ threshold xavfli kontentni aniqlashda yukori aniqlik beradi.
3. Dataset hajmi oshgan sari klasterlash sifati keskin yaxshilanadi.

4. Optimal boshlang‘ich o‘qitish uchun kamida 10 000 ta, ideal holatda 20 000 ta matn tavsiya etiladi.

5. Ushbu bosqichda olingan yarim-belgilangan datasetlar keyinchalik BERTbek, mBERT, XLM-R kabi transformerlarda o‘qitish uchun juda qulaydir.

6. Tadqiqot natijalari ijtimoiy tarmoqlardagi xavfli kontentni avtomatik aniqlash bo‘yicha keyingi bosqich — chuqur o‘rganish modellarini yaratishda mustahkam asos bo‘lib xizmat qiladi.

Adabiyotlar ro‘yxati

1. Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters, 31*(8), 651–666.
2. MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 281–297.
3. Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information Processing & Management, 24*(5), 513–523.
4. Schmidt, A., & Wiegand, M. (2017). A survey on hate speech detection using natural language processing. *Proceedings of the NLP4SocialMedia Workshop*, 1–10.
5. 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.
6. Conneau, A., Khandelwal, K., Goyal, N., et al. (2020). Unsupervised cross-lingual representation learning at scale. *Proceedings of ACL*, 8440–8451.
7. Abdullayev, S., Mirzakhalilov, M., & Yusupov, M. (2023). BERTbek: A pretrained language model for Uzbek. *arXiv preprint arXiv:2306.00602*.