

DDOS HUJUMLARINI ANIQLASHDA MASHINALI O'QITISH ALGORITMLARINING QIYOSIY TAHLILI: RANDOM FOREST, KNN, XGBOOST VA LIGHTGBM

Raxmatov Furqat Abdirazzoqovich¹, Toshtemirov Muxammadi Shokir o'g'li²

¹*Muhammad al-Xorazmiy nomidagi Toshkent axborot texnologiyalari universiteti “Kompyuter tizimlari” kafedrası dotsenti*

²*Muhammad al-Xorazmiy nomidagi Toshkent axborot texnologiyalari universiteti 2-kurs magistranti*

¹*e-mail: furkat.rahmatov@gmail.com, ²e-mail: muxammadi0799@gmail.com*

Annotatsiya: Ushbu maqolada uchta xalqaro benchmark dataset (CIC-DDoS2019, UNSW-NB15, Network Intrusion) asosida to'rtta mashinali o'qitish algoritmi — Random Forest (RF), K-Nearest Neighbors (KNN), XGBoost va LightGBM — ning DDoS hujumlarini aniqlashdagi samaradorligi qiyosiy tahlil qilinadi. Algoritmlar nafaqat aniqlik (accuracy, F1-score), balki hisoblash samaradorligi (o'qitish va bashorat vaqti) bo'yicha ham baholandi. Tajribalar shuni ko'rsatdiki, XGBoost CIC-DDoS2019 datasetida eng yuqori aniqlikka (93.47%) erishdi, LightGBM esa Network-Intrusion datasetida 99.92% natija ko'rsatdi. Hisoblash samaradorligi tahlili Random Forest algoritmining 88.4–687.9 ms bashorat vaqti bilan real vaqt DDoS aniqlash tizimi uchun eng maqbul variant ekanligini isbotladi. KNN ning 8,000–10,000 ms bashorat vaqti esa real vaqt tizimlarida qo'llash uchun yaroqsiz ekanligi aniqlandi.

Kalit so'zlar: DDoS hujumlari, Random Forest, KNN, XGBoost, LightGBM, mashinali o'qitish, tarmoq xavfsizligi, hisoblash samaradorligi, intrusion detection.

Abstract: This paper presents a comparative analysis of four machine learning algorithms — Random Forest (RF), K-Nearest Neighbors (KNN), XGBoost, and LightGBM — for detecting DDoS attacks based on three international benchmark datasets (CIC-DDoS2019, UNSW-NB15, and Network Intrusion). The algorithms were evaluated not only by accuracy (accuracy, F1-score), but also by computational efficiency (training and prediction time). Experiments demonstrated that XGBoost achieved the highest accuracy on the CIC-DDoS2019 dataset (93.47%), while LightGBM showed 99.92% on the Network-Intrusion dataset. Computational efficiency analysis confirmed that Random Forest, with a prediction time of 88.4–687.9 ms, is the most suitable algorithm for real-time DDoS detection systems. KNN's prediction time of 8,000–10,000 ms was found to be unsuitable for real-time applications.

Keywords: DDoS attacks, Random Forest, KNN, XGBoost, LightGBM, machine learning, network security, computational efficiency, intrusion detection.

1. KIRISH

Internet tarmoqlarida DDoS hujumlarining soni va murakkabligi yildan yilga ortib bormoqda. Bunday hujumlarni aniqlash uchun mashinali o'qitish algoritmlariga asoslangan yondashuvlar faol tadqiq qilinmoqda. Biroq adabiyotlarda algoritmlar odatda faqat aniqlik ko'rsatkichlari bo'yicha taqqoslanadi — real vaqt tizimlari uchun muhim bo'lgan hisoblash samaradorligi (bashorat tezligi) ko'pincha e'tibordan chetda qoladi [1].

Ushbu tadqiqot ikkita asosiy savolga javob beradi. Birinchi savol: to'rtta mashinali o'qitish algoritmi (RF, KNN, XGBoost, LightGBM) bir xil sharoitda qaysi biri yuqori aniqlik ko'rsatadi? Ikkinchi savol: real vaqt DDoS aniqlash tizimi uchun qaysi algoritmi aniqlik va tezlik balansini eng yaxshi ta'minlaydi?

2. METODOLOGIYA

Tadqiqotda uchta xalqaro benchmark dataset ishlatildi. CIC-DDoS2019 — 431,371 qator, 11 turdagi DDoS hujumi (18 ta klass 11 taga birlashtirildi). UNSW-NB15 — 257,673 qator, 7 turdagi tarmoq hujumi. Network Intrusion (CIC-IDS2017) — 844,000+ qator, DDoS va normal traffic. Har bir datasetda ma'lumotlar tozalanib, muvozanatlashtirilgan namuna olindi (CIC-DDoS2019 uchun 12,000, UNSW-NB15 uchun 15,000, Network Intrusion uchun 8,000 ta har bir klassdan). Ma'lumotlar 80% o'qitish va 20% sinov uchun ajratildi.

To'rtta algoritmi quyidagi parametrlar bilan o'qitildi. Random Forest: `n_estimators=200`, `class_weight='balanced'`. KNN: `n_neighbors=5`, euclidean metrika, StandardScaler bilan masshtablash. XGBoost: `n_estimators=200`, `max_depth=6`, `learning_rate=0.1`. LightGBM: `n_estimators=200`, `max_depth=6`, `learning_rate=0.1`.

Baholash uchun to'rtta metrika ishlatildi: Accuracy, F1-Score, Precision, Recall. Bundan tashqari, har bir algoritmi uchun o'qitish vaqti (soniya) va bashorat vaqti (millisekund) o'lchandi.

3. NATIJALAR

3.1 Algoritmning aniqligini taqqoslash

1-jadval.

To'rt algoritmi taqqoslash jadvali — Accuracy va F1-Score

Dataset	RF	KNN	XGBoost	LightGBM
CIC-DDoS2019 (Acc)	91.49%	91.72%	93.47%	92.95%
CIC-DDoS2019 (F1)	91.45%	91.50%	93.27%	93.04%
UNSW-NB15 (Acc)	74.98%	68.12%	77.08%	77.41%
UNSW-NB15 (F1)	75.60%	67.79%	77.39%	77.86%

Network-Intrusion (Acc)	99.80%	99.10%	99.90%	99.92%
Network-Intrusion (F1)	99.80%	99.10%	99.90%	99.92%

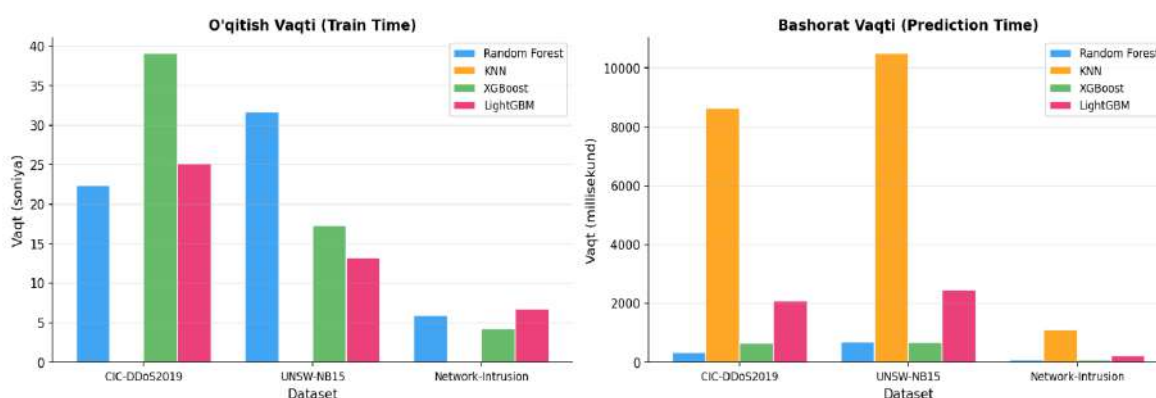
Jadvaldan ko'rinadiki, gradient boosting algoritmlari (XGBoost va LightGBM) aniqlik bo'yicha RF va KNN dan biroz ustun. XGBoost CIC-DDoS2019 da 93.47%, LightGBM Network-Intrusion da 99.92% accuracy ko'rsatdi. UNSW-NB15 datasetida LightGBM (77.41%) va XGBoost (77.08%) RF dan (74.98%) yuqori natija berdi. KNN faqat CIC-DDoS2019 da RF ga yaqin natija ko'rsatdi (91.72% vs 91.49%), qolgan datasetsda sezilarli darajada orqada qoldi.

3.2 Algoritmlarning hisoblash samaradorligi tahlili.

2-jadval.

Algoritmlar hisoblash samaradorligi

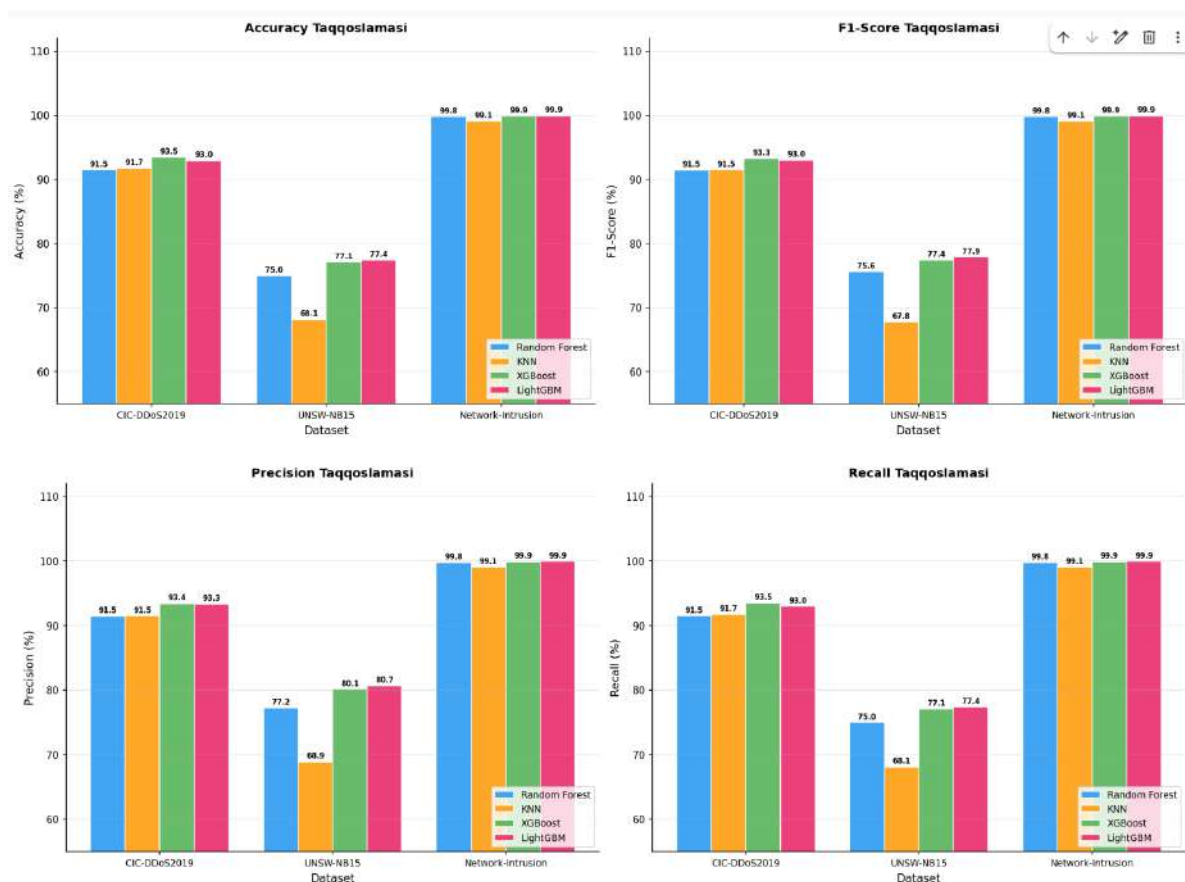
Algoritm	CIC-DDoS2019		UNSW-NB15		Network-Intrusion	
	Train (s)	Pred (ms)	Train (s)	Pred (ms)	Train (s)	Pred (ms)
Random Forest	22.4	333.4	31.7	687.9	5.9	88.4
KNN	0.03	8630.6	0.02	10502.9	0.01	1098.3
XGBoost	39.1	643.5	17.3	664.2	4.2	76.3
LightGBM	25.1	2086.3	13.3	2466.6	6.8	236.8



1-rasm. To'rt algoritmlarning hisoblash samaradorligi grafiklari

Hisoblash samaradorligi tahlili quyidagi muhim xulosalarga olib keldi. KNN o'qitish vaqti juda past (0.01–0.03 soniya), lekin bashorat vaqti 8,630–10,503 ms ga yetadi — bu real vaqt tizimlarida mutlaqo yaroqsiz. Random Forest barcha datasetsda eng past bashorat vaqtini ko'rsatdi: Network-Intrusion da 88.4 ms, CIC-DDoS2019 da 333.4 ms. XGBoost Network-Intrusion da eng tez bashorat qildi (76.3 ms), lekin CIC-DDoS2019 da 643.5 ms ko'rsatdi. LightGBM bashorat tezligi bo'yicha o'rtacha ko'rsatkich namoyish etdi (236.8–2,466 ms).

3.2 Umumiy taqqoslash natijalari.



2-rasm. To'rt algoritmlarning taqqoslash grafiklari

4. MUHOKAMA

Natijalar shuni ko'rsatadiki, aniqlik va bashorat tezligi bir-biriga qarama-qarshi ko'rsatkichlar bo'lib, real tizim uchun optimal algoritmni tanlashda ikkalasini birga hisobga olish kerak.

Gradient boosting algoritmlari (XGBoost, LightGBM) aniqlik bo'yicha ustunlik qilsa-da, bashorat vaqti RF ga nisbatan 2–7 marta yuqori. Masalan, CIC-DDoS2019 da XGBoost 643.5 ms, RF esa 333.4 ms bashorat vaqti ko'rsatdi. Yuqori trafik sharoitida (sekundiga minglab paketlar) bu farq kritik ahamiyat kasb etadi.

KNN algoritmining bashorat vaqti (8,000–10,500 ms) uni real vaqt tizimlaridan to'liq chiqaradi. Bu holat KNN ning o'qitish bosqichida hech qanday model qurilmasligi — bashorat paytida barcha namunalar bilan masofani hisoblash zaruratidan kelib chiqadi.

Random Forest aniqlik va tezlik balansini eng yaxshi ta'minladi — barcha datasetsda 88.4–687.9 ms bashorat vaqti bilan yuqori aniqlikni saqladi. Shuning uchun real vaqt DDoS aniqlash tizimi uchun Random Forest eng maqbul algoritm hisoblanadi.

XULOSA

Ushbu tadqiqotda to'rtta mashinali o'qitish algoritmi uchta xalqaro datasetda keng qamrovli taqqoslandi. Asosiy xulosalar:

1. Aniqlik bo'yicha: XGBoost (93.47%) va LightGBM (99.92%) eng yuqori natijalarni ko'rsatdi, Random Forest esa barqaror va yuqori aniqlikni ta'minladi (88.81% o'rtacha);
2. Bashorat tezligi bo'yicha: Random Forest eng tez bashorat qildi (88.4–687.9 ms), KNN esa real vaqt tizimlarida qo'llashga yaroqsiz (8,000–10,500 ms);
3. Real vaqt tizimi uchun: aniqlik va tezlik balansini hisobga olgan holda Random Forest eng maqbul algoritm hisoblanadi;
4. KNN cheklovi: katta hajmli tarmoq ma'lumotlarida KNN ning bashorat vaqti juda yuqori — bu algoritmni real vaqt DDoS aniqlashda ishlatish tavsiya etilmaydi.

FOYDALANILGAN ADABIYOTLAR

- [1]. M. Waqas et al., "A Systematic Review of DDoS Attack Detection Techniques," *IEEE Access*, vol. 10, pp. 45891–45924, 2022.
- [2]. N. Farnaaz and M. A. Jabbar, "Random Forest Modeling for Network Intrusion Detection System," *Procedia Computer Science*, vol. 89, pp. 213–217, 2016.
- [3]. Sharafaldin et al., "Toward Generating a New Intrusion Detection Dataset," in *Proc. ICISSP*, 2018, pp. 108–116.
- [4]. N. Moustafa and J. Slay, "UNSW-NB15: A Comprehensive Data Set for Network Intrusion Detection Systems," in *Proc. MilCIS*, 2015.
- [5]. J. Zhang et al., "Random-Forests-Based Network Intrusion Detection Systems," *IEEE Trans. Systems, Man, Cybernetics*, vol. 38, no. 5, pp. 649–659, 2008.
- [6]. T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proc. KDD*, 2016, pp. 785–794.