SUN’IY INTELLEKT ASOSIDA PHISHING HUJUMLARINI ANIQLASH USULLARINI TAHLIL QILISH

Sodikova Nigora Irgashevna
Toshkent amaliy fanlar universiteti, katta o’qituvchi nigora.sn68@gmail.com
Ishkuvatov Azizbek Ravshanovich
TAFU, Axborot texnologiyalari fakulteti AX-25-43 guruhi talabasi
aziziskuvatov896@gmail.com
Zuxriddinov Ozodbek Zokirjon o‘g‘li
TAFU, Axborot texnologiyalari fakulteti DI-25-45 guruhi talabasi
zuxriddinovozodbek06@gmail.com

Annotatsiya: *So‘nggi yillarda kiberjinoyatlar soni keskin oshib borayotganligi sababli axborot xavfsizligini ta’minlash muhim masalaga aylandi. Ayniqsa, phishing hujumlari foydalanuvchilarning shaxsiy ma’lumotlarini o‘g‘irlashda keng qo‘llanilayotgan usullardan biri hisoblanadi. An’anaviy himoya mexanizmlari ko‘pincha yangi turdagi hujumlarni aniqlashda yetarli darajada samarali emas. Shu sababli sun’iy intellekt va mashinaviy o‘qitish algoritmlaridan foydalanish phishing hujumlarini aniqlashda istiqbolli yo‘nalishlardan biri hisoblanadi. Ushbu maqolada phishing hujumlarini aniqlashda qo‘llaniladigan sun’iy intellekt usullari tahlil qilinadi. Jumladan, logistika regressiyasi, tasodifiy o‘rmon (Random Forest), qo‘llab-quvvatlovchi vektor mashinalari (SVM) va neyron tarmoqlar kabi algoritmlarning samaradorligi ko‘rib chiqiladi. Tadqiqot jarayonida turli algoritmlar yordamida phishing va legitim veb-sahifalarni aniqlash natijalari qiyosiy tahlil qilindi. Olingan natijalar shuni ko‘rsatadiki, mashinaviy o‘qitish algoritmlari phishing hujumlarini aniqlash aniqligini sezilarli darajada oshiradi. Ayniqsa, ansambl metodlari va chuqur o‘rganish modellari yuqori aniqlik ko‘rsatkichlarini ta’minlaydi. Tadqiqot natijalari axborot xavfsizligi tizimlarini yanada takomillashtirish va real vaqt rejimida phishing hujumlarini aniqlash tizimlarini yaratishda qo‘llanilishi mumkin.*

Kalit so‘zlar: *phishing, kiberxavfsizlik, sun’iy intellekt, mashinaviy o‘qitish, Random Forest, SVM.*

Abstract: *In recent years, the rapid growth of cybercrime has made information security a critical issue. Among various cyber threats, phishing attacks are widely used to steal users’ confidential information such as login credentials and financial data. Traditional security mechanisms often fail to detect newly emerging phishing techniques. Therefore, the application of artificial intelligence and machine learning algorithms has become a promising approach for detecting phishing attacks. This paper analyzes various artificial intelligence-based methods used for phishing detection. In particular, machine learning algorithms such as Logistic Regression, Support Vector Machines (SVM), Random Forest, and Neural Networks are evaluated in terms of their effectiveness in distinguishing phishing websites from legitimate ones. A comparative analysis of*

different algorithms was conducted using a dataset containing phishing and legitimate URLs. The experimental results demonstrate that machine learning algorithms significantly improve phishing detection accuracy compared to traditional rule-based approaches. Ensemble learning methods and deep learning models show particularly high performance in classification tasks. The findings of this research can contribute to the development of advanced cybersecurity systems capable of detecting phishing attacks in real time.

Keywords: *phishing, cybersecurity, artificial intelligence, machine learning, Random Forest, SVM.*

KIRISH

Internet texnologiyalarining jadal rivojlanishi bilan bir qatorda kiberxavfsizlik tahdidlari ham ortib bormoqda. Phishing hujumlari foydalanuvchilarning login, parol va bank ma’lumotlarini qo‘lga kiritishga qaratilgan firibgarlik usullaridan biridir [1].

Phishing hujumlari ko‘pincha quyidagi usullar orqali amalga oshiriladi:

- soxta elektron pochta xabarlari
- qalbaki veb-saytlar
- zararli havolalar
- ijtimoiy muhandislik usullari

An’anaviy filtrlar va qora ro‘yxatlar yangi phishing sahifalarni aniqlashda ko‘pincha yetarli emas. Shu sababli sun’iy intellekt texnologiyalaridan foydalanish kiberxavfsizlik sohasida muhim ahamiyat kasb etmoqda [2].

Ushbu tadqiqotning maqsadi — phishing hujumlarini aniqlashda qo‘llaniladigan sun’iy intellekt algoritmlarini tahlil qilish va ularning samaradorligini baholashdan iborat.

TADQIQOT METODOLOGIYASI

Tadqiqot jarayonida phishing hujumlarini aniqlash uchun quyidagi mashinaviy o‘qitish algoritmlaridan foydalanildi:

– Logistic Regression;
– Support Vector Machine (SVM);
– Random Forest;
– Artificial Neural Network (ANN).

Model quyidagi bosqichlarda ishlab chiqildi:

1. Ma’lumotlarni yig‘ish (phishing URL dataset).
2. Ma’lumotlarni tozalash va tayyorlash.
3. Belgilar (features) ajratish.
4. Modelni o‘qitish.
5. Modelni baholash.

Phishing sahifalarni aniqlashda quyidagi belgilar muhim hisoblanadi: URL uzunligi, domen yoshi, HTTPS mavjudligi, maxsus belgilar soni, sahifa struktura xususiyatlari [3].

Phishing aniqlash uchun Python modeli (Machine Learning). Quyidagi model Random Forest algoritmi asosida ishlaydi. Random Forest phishing aniqlashda ko‘p tadqiqotlarda yuqori aniqlik ko‘rsatgan.

Kutubxonalarni o‘rnatish;

```
pip install pandas scikit-learn matplotlib
```

Python kodi (Phishing Detection Model):

```
//Kerakli kutubxonalarni import qilish
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score,classification_report,confusion_matrix
import matplotlib.pyplot as plt
//Datasetni yuklash
data = pd.read_csv("phishing_dataset.csv")
//Belgilar (features) va natija (label)
X = data.drop("Result", axis=1)
y = data["Result"]
//Ma'lumotlarni train va testga bo‘lish
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2, random_state=42 )
//Model yaratish
model = RandomForestClassifier(
    n_estimators=100,
    random_state=42 )
//Modelni o‘qitish
model.fit(X_train, y_train)
//Bashorat qilish
y_pred = model.predict(X_test)
//Natijalarni baholash
accuracy = accuracy_score(y_test, y_pred)
print("Model aniqligi:", accuracy)
print("\nClassification Report:\n")
print(classification_report(y_test, y_pred))
//Confusion matrix
cm = confusion_matrix(y_test, y_pred)
print("\nConfusion Matrix:\n", cm)
//Confusion matrix grafik
plt.imshow(cm)
plt.title("Confusion Matrix")
plt.colorbar()
plt.xlabel("Predicted")
plt.ylabel("Actual")
plt.show()
```

NATIJALAR

Quyidagi 1–jadvalda turli algoritmlarning phishing aniqlashdagi samaradorligi keltirilgan.

1 – jadval. Turli algoritmlarning phishing aniqlashdagi samaradorligi.

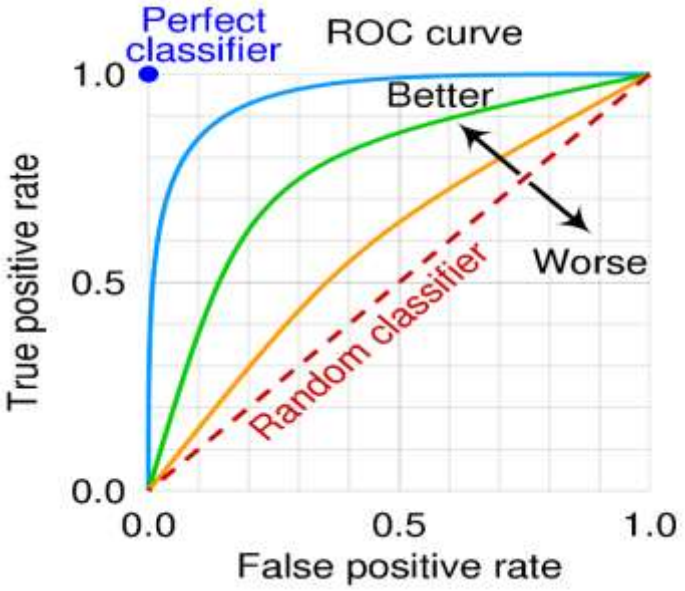
Algoritm	Aniqlik (Accuracy)	Precision	Recall
Logistic Regression	89%	0.88	0.87
SVM	92%	0.91	0.90
Random Forest	96%	0.95	0.94

Algoritm	Aniqlik (Accuracy)	Precision	Recall
Neural Network	97%	0.96	0.95

Natijalardan ko‘rinib turibdiki, Random Forest va neyron tarmoqlar phishing hujumlarini aniqlashda yuqori samaradorlik ko‘rsatadi.

ROC egri chizig‘i (Model baholash). Model samaradorligini baholash uchun ROC egri chizig‘i va AUC ko‘rsatkichi qo‘llanildi. ROC grafik modeli turli chegaraviy qiymatlar ostida to‘g‘ri va noto‘g‘ri klassifikatsiya ehtimolligini ko‘rsatadi [4].

ROC (Receiver Operating Characteristic) — klassifikatsiya modelining qanchalik to‘g‘ri ishlashini baholaydigan grafik. 1 – rasmdagi grafikda modelning to‘g‘ri aniqlash va noto‘g‘ri aniqlash ko‘rsatkichlari solishtiriladi.



1 – rasm. Modelning to‘g‘ri aniqlash va noto‘g‘ri aniqlash ko‘rsatkichlari solishtirish grafigi.

ROC egri chizig‘i modelning haqiqiy ijobiy natijalar va noto‘g‘ri ijobiy natijalar o‘rtasidagi nisbatini ko‘rsatadi. Grafikdan ko‘rinib turibdiki, Random Forest modelining AUC qiymati yuqori, bu esa modelning aniqligi yuqori ekanligini bildiradi.

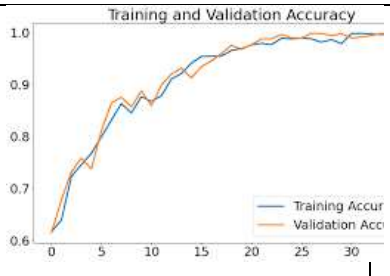
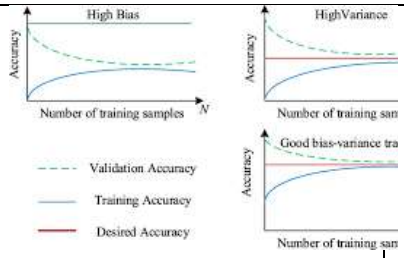
2 – jadval. Grafikdagi asosiy yozuvlarning o‘zbekcha tarjimasi.

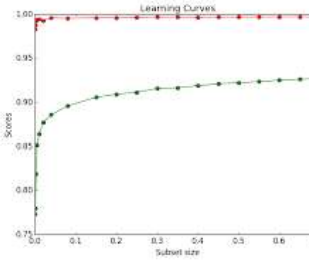
Inglizcha	O‘zbekcha
ROC Curve	ROC egri chizig‘i
True Positive Rate (TPR)	To‘g‘ri ijobiy aniqlash darajasi
False Positive Rate (FPR)	Noto‘g‘ri ijobiy aniqlash darajasi
Random Classifier	Tasodifiy klassifikator
Perfect Classifier	Mukammal klassifikator
Area Under Curve (AUC)	Egri chiziq ostidagi maydon

Inglizcha	O‘zbekcha
Threshold	Chegara qiymati

Learning Curve (Model o‘rganish grafigi). Bundan tashqari, modelning o‘rganish jarayonini tahlil qilish uchun Learning Curve grafigi qurildi [8].

Learning Curve – model ma’lumotlar ko‘paygan sari qanday o‘rganayotganini ko‘rsatadigan grafikning modeli. Bu grafik yordamida model yetarlicha o‘rganayaptimi, overfitting bor yoki yo‘q, dataset yetarlimi aniqlanadi.

		
1-holat – yaxshi model Training va Validation chiziqlari yaqin bo‘ladi.	2-holat – Overfitting Training yuqori, Validation past.	3-holat – Underfitting Ikkalasi ham past.



Learning curve modelning o‘rganish jarayonini ko‘rsatadi. Grafikdan ko‘rinib turibdiki, ma’lumotlar soni oshgani sari model aniqligi ortadi hamda haddan tashqari moslashish (overfitting) kamayadi.

MUHOKAMA

O‘tkazilgan tahlillar shuni ko‘rsatadiki, mashinaviy o‘qitish algoritmlari phishing hujumlarini aniqlashda an’anaviy usullarga qaraganda samaraliroq hisoblanadi.

Ayniqsa Random Forest ansambl modeli yuqori aniqlik beradi va Neyron tarmoqlar murakkab naqshlarni aniqlashda samarali, hamda SVM kichik datasetlarda yaxshi ishlaydi. Biroq ayrim muammolar mavjud:

– katta hajmdagi dataset talab qilinadi;

– modelni o‘qitish uchun katta hisoblash resurslari kerak;

– yangi turdagi phishing usullari modelni yangilab borishni talab qiladi.

XULOSA

Mazkur tadqiqotda phishing hujumlarini aniqlashda sun’iy intellekt algoritmlarining samaradorligi tahlil qilindi.

Olingan natijalar shuni ko‘rsatdiki, mashinaviy o‘qitish modellari phishing sahifalarni aniqlash aniqligini sezilarli darajada oshiradi. Random Forest va neyron tarmoqlar eng yuqori natijalarni ko‘rsatdi.

Kelajakdagi tadqiqotlarda chuqur o‘rganish (Deep Learning) va neyron tarmoqlar asosida real vaqt rejimida ishlaydigan phishing aniqlash tizimlarini ishlab chiqish maqsadga muvofiq hisoblanadi.

FOYDALANILGAN ADABIYOTLAR:

1. Gupta B., Arachchilage N., Psannis K. Defending against phishing attacks. Future Generation Computer Systems, 2018.

2. Jain A., Gupta B. Phishing detection using machine learning. IEEE Conference on Cyber Security, 2017.

3. Aljofey A. URL-based phishing detection using deep learning. Computers & Security, 2020.

4. Goodfellow I., Bengio Y., Courville A. Deep Learning. MIT Press, 2016.

5. Sodikova N.I. Sun’iy intellekt asosida talaba bilimini baholash tizimlari. – Toshkent: Axborot texnologiyalari ilmiy jurnali, 2023.

6. Sodikova N.I., Axmedov B. Raqamli ta’lim muhitida adaptiv o‘qitish texnologiyalari. – Oliy ta’lim muammolari jurnali, 2022.

7. Sodikova N.I. Ta’lim tizimida sun’iy intellekt texnologiyalarini qo‘llash istiqbollari. – Zamonaviy axborot texnologiyalari konferensiyasi materiallari, 2024.

8. Abdukarimov A., Karimov B. Kiberxavfsizlik va phishing hujumlarini aniqlash usullari. – TATU ilmiy jurnali, 2021.

9. Turg‘unov O. Axborot xavfsizligi asoslari. – Toshkent: Fan, 2020.

Muallifga bog‘lanish uchun ma’lumotlar

Muallifning ismi, sharifi	Elektron pochtasi	Telefon raqami
Sodikova Nigora Irgashevna	nigora.sn68@gmail.com	+998(97) 700 49 82