JOURNAL OF CRITICAL REVIEWS

ISSN- 2394-5125 VOL 10, ISSUE 07, 2023

MACHINE LEARNING TECHNIQUES FOR CYBER ATTACKS DETECTIONS

L.C.Usha Maheswari[1], P.Sruthi[2], Ch.Hamsini[3], G.Keerthika[4], M.Sreeja[5]

[1]Assistant Professor, Department of CSE(CS), MallaReddy Engineering College for Women, Hyderabad, TS, India.

Email: maheswari07usha@gmail.com

[2,3,4,5]UG Students, Department of CSE(CS), MallaReddy Engineering College for Women, Hyderabad, TS, India.

ABSTRACT:

Improvements in PCs and correspondence have brought about broad changes in comparison to the past. The use of new technologies gives individuals, companies and governments unbelievable benefits, even when they are messing against them. For eg, security of major data, safety of data stadiums, accessibility of information etc. Digital fear-based oppression is, depending on these questions, one of the biggest problems of our day. Digital fear, which has brought many problems to citizens and organizations, has come to an extent that could threaten the openness and protection of the nation through numerous events, such as criminal groups, professional individuals and digital activists. In this sense, IDS systems were built to maintain a strategic distance from digital attacks. Intrusion detection systems (IDS) At the moment, learning the computation of the SVM (Bolster Support Vector Machine) was used to identify port sweep efforts that depend on a new data set of CICIDS 2017, which achieved 69.79% individually. We should implement other algorithms, including CNN, ANN and Random Forest, rather than SVM, so that accuracy such as SVM – 93.29, CNN - 63.52, Random Forest – 99.93, ANN – 99.11 can be acquired.

Keywords: IDS, CNN, ANN, SVM, Digital terror.

1. INTRODUCTION

Improved PCs and correspondence developments have in comparison to the past, contributed to extensive and propelling improvements. The use of novel technologies gives individuals, companies and governments incredible advantages, whether or not they are destroyed. For eg, the safety of essential data, security of revealed data phases, information accessibility, etc. Digital fear-based oppression, depending on these questions, is nowadays one of the most critical issues. Digital terror, which has posed many problems for people and organizations, has reached a degree that could threaten accessible and security for country, through numerous meetings, such as criminal alliances,

professional people and digital activists. In this respect, IDS has been built to maintain a strategic distance from digital assaults. Right now the measurements of the SVM have been used to classify port sweeping activities based on the latest CICIDS 2017 data set with 97.80%, 69.79% of precise data is carried out individually. Rather than using the SVM, it is possible to join other forest algorithms, such as RF, CNN and ASN, which can be accurate, such as SVM – 93.29, CNN – 63.52, RF – 99.93, ANN - 99.11. Instead of SVM.

MOTIVATION

The use of new technologies gives individuals, companies and governments unbelievable benefits, even when they are messing against them. For eg, security of major data, safety of data stadiums, accessibility of information etc. Digital fear-based oppression is, depending on these questions, one of the biggest problems of our day. Digital fear, which has brought many problems to citizens and organisations, has come to an extent that could threaten the openness and protection of the nation through numerous events, such as criminal groups, professional individuals and digital activists. In this sense, IDS systems were built to maintain a strategic distance from digital attacks. Intrusion detection systems (IDS)

2. EXISTING SYSTEM :

The Almansob and Lomte KDD99 dataset was used in Blamless Bayes and Principal Component Analysis (PCA). Similarly, Chithik and Rabbani have also used PCA, SVM and KDD99 for IDS[10]. In Aljavarskaya et al. The papers, their evaluations and examinations have been transmitted based on the NSL-KDD data set for their IDS model [11] Composite inspectorates indicate that the KDD99 dataset is used continuously for IDS [6]–[10]. KDD99 is therefore older and provides little knowledge about cutting-edge new forms of attack, for example, multi-day misuse etc. In our investigation, we used an innovative and new dataset from CICIDS2017.

PROPOSED SYSTEM :

This paper tries to find which algorithm can predict best accuracy rates that help to better predict outcomes to identify cyber attacks or not. These predictions are made by four algorithms such as SVM, ANN, RF and CNN. This technique was used to detect cyber attack using machine learning technologies on the network.

Literature survey:

2.1 R. 2.1 2.1 "Harbor scanning and defense against them," Christopher, SANS Institute, 2001. 2001. - 2001.

Port scanning is one of the most common techniques used in network resource finding for attackers. All systems connected via modem to the LAN or the Internet run on common or unknown ports services. The intruder will see the following information on the target devices by way of port scanning, which services are executed by users, support

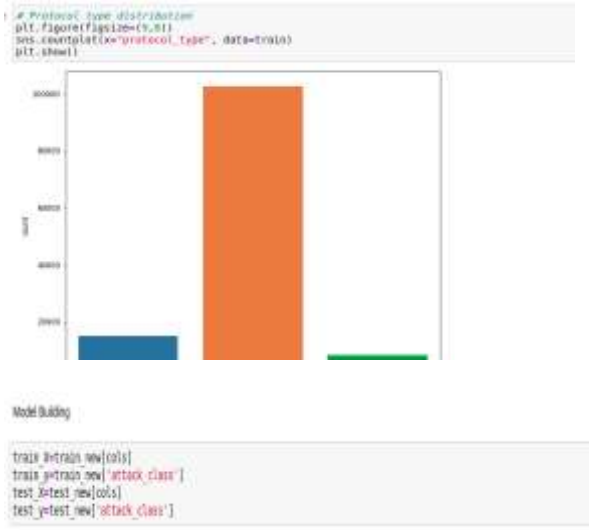
for anonymous logins and authentication of certain network services. The scan of the port is accomplished by sending a message to each port one by one. The reaction form indicates if the port is used and additional vulnerabilities can be tested. Port scanners are important for network security technicians because the target system may detect possible security vulnerabilities. Since port scans can be performed against your programs, port scans can be detected and open services information limited by the right tools. Each publicly available computer has and can operate open ports. The goal is to limit the visibility and refusal of approved users of open ports to closed ports.

2.2 pp. 2.2 p. 2.2 p. J. A. Hoagland, J. A. Staniford. - Staniford. M. McAlerney, Computer Security Journal, Vol. "Practical automated port detection," 10, 1-2, 118-24, 2002. 2002. 2002

Port scanning is a common method that is very important. Software attackers often use it to classify hosts or networks they are opposed to. This makes it preliminary to classify port scans for more serious attacks useful for system administrators and other network advocates. Network defenders also use their own networks to take into account and find vulnerabilities. Accordingly, attackers must determine whether or not network advocates scan the network regularly. But Defenders don't normally want to mask their ports' scanning, even if attackers. We'll certainly speak in the rest of this article about those attackers who search the network and supporters who attempt to check. Online mailing lists and newsgroups are ongoing the legal/ethical debate of port scanning. It is necessary to scan remote network port itself, without the owners' consent, as a legal and ethical activity. This is actually a grey field in most jurisdictions. But our experience in monitoring unwanted remote scanning is that virtually all of them come from endangered, hostile host systems.

Data preprocessing

Data EDA



Model Building

```
train X=train_new[cols]
train_y=train_new['attack_class']
test_X=test_new[cols]
test_y=test_new['attack_class']
```

ML Deploy

Logistic Regression

```
# Building Models
from sklearn.linear_model import LogisticRegression
logreg = LogisticRegression(random_state=0,solver='lbfgs',multi_class='multinomial')
logreg.fit( train_X, train_y)
logreg.predict(train_X)   #By default, it use cut-off as 0.5
```

```
list( zip( cols, logreg.coef_[0] ) )
```

```
logreg.intercept_
```

```
logreg.score(train_X,train_y)
```

Decision Trees

```
train_X.shape
```

```
param_grid = {'max_depth': np.arange(2, 12),
              'max_features': np.arange(10,23)}
```

```
train_y.shape
```

```
from sklearn.model_selection import GridSearchCV
from sklearn.tree import DecisionTreeClassifier, export_graphviz, export
tree = GridSearchCV(DecisionTreeClassifier(), param_grid, cv = 10,verbose=1,n_jobs=-1)
tree.fit( train_X, train_y )
```

```
tree.best_score_
```

```
tree.best_estimator_
tree.best_params_
```

```
train_pred = tree.predict(train_X)
```

```
print(metrics.classification_report(train_y, train_pred))
```

```
test_pred = tree.predict(test_X)
```

Random Forest

```
from sklearn.ensemble import RandomForestClassifier
paramgrid_rf = {'n_estimators': [50,60,70,80,90,100],
                'max_features': [2,3,4,5,6,7]}
```

```
from sklearn.model_selection import GridSearchCV
gscv_rf = GridSearchCV(estimator=RandomForestClassifier(),
                       param_grid=paramgrid_rf,
                       cv=10,
                       verbose=True, n_jobs=-1)

gscv_results = gscv_rf.fit(train_X, train_y)
```

```
gscv_results.best_params_
```

```
gscv_rf.best_score
```

```
rade_clf = RandomForestClassifier(oob_score=True,n_estimators=80, max_features=5, n_jobs=-1)
rade_clf.fit( train_X, train_y )
```

```
rade_test_pred = pd.DataFrame( {'actual':  test_y,
                                'predicted': rade_clf.predict( test_X ) } )
```

Support Vector Machine (SVM)

```
from sklearn.svm import LinearSVC
svm_clf = LinearSVC(random_state=0, tol=1e-5)
svm_clf.fit(train_X,train_y)
```

```
print(svm_clf.coef_)
print(svm_clf.intercept_)
print(svm_clf.predict(train_X))
```

```
from sklearn.svm import SVC
from sklearn.pipeline import make_pipeline
```

```
model = SVC(kernel='rbf', class_weight='balanced',gamma='scale')
```

```
model.fit(train_X,train_y)
```

```
from sklearn.model_selection import GridSearchCV
param_grid = {'C': [1, 10],
              'gamma': [0.0001, 0.001]}
grid = GridSearchCV(model, param_grid)
```

```
grid.fit(train_X,train_y)
```

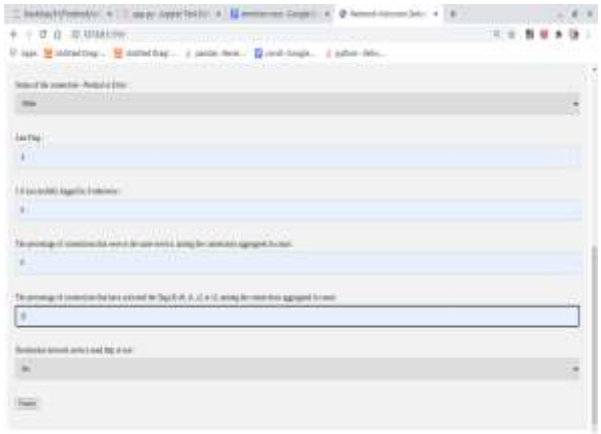
```
print(grid.best_params_)
```

From the score accuracy we concluding the DT & RF give better accuracy and building pickle file for predicting the user input

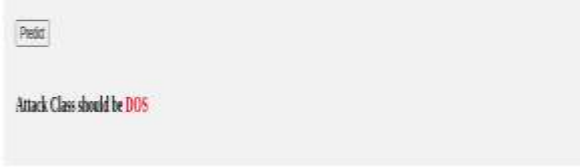
Application

Localhost - in cmd python app.py

Enter the input



Predict attack -



3. CONCLUSION

Right now, vector aid estimates, ANN, CNN, Random Forest, and deep learnings based on modern CICIDS2017 dataset have been relatively added. The findings show that the in-depth estimation of learning has obtained essentially better results than SVM, ANN, RF and CNN. With AI and in-depth learning calculations, apache Hadoop and sparkling inventions, we will make common use of port sweep efforts as well as other assault forms that rely on that dataset. All of this allows us to identify the network cyber threat. It occurs in a way that when we consider the many attacks that occurred over a long period of time, the features of these attacks are preserved in those datasets if they are remembered. We will also predict whether cyber attack is conducted or not using these datasets. This document aims to assess the best prediction algorithms to avoid the best outcomes of cyber attacks. This article can be found in four algorithms including SVM, ANN, RF, CNN.

FUTURE SCOPE

We add some ML algorithms for improvement to improve the precision.

REFERENCES

[1] K. Graves, Ceh: Official certified ethical hacker review guide: Exam 312-50. John Wiley & Sons, 2007.

[2] R. Christopher, “Port scanning techniques and the defense against them,” SANS Institute, 2001.

[3] M. Baykara, R. Das¸, and I. Karado ˘gan, “Bilgi g ¨uvenli ˘gi sistemlerinde kullanilan arac¸larin incelenmesi,” in 1st International Symposium on Digital Forensics and Security (ISDFS13), 2013, pp. 231–239.

[4] S. Staniford, J. A. Hoagland, and J. M. McAlerney, “Practical automated detection of stealthy portscans,” Journal of Computer Security, vol. 10, no. 1-2, pp. 105–136, 2002.

[5] S. Robertson, E. V. Siegel, M. Miller, and S. J. Stolfo, “Surveillance detection in high bandwidth environments,” in DARPA Information Survivability Conference and Exposition, 2003. Proceedings, vol. 1. IEEE, 2003, pp. 130–138.

[6] K. Ibrahimi and M. Ouaddane, “Management of intrusion detection systems based-kdd99: Analysis with lda and pca,” in Wireless Networks and Mobile Communications (WINCOM), 2017 International Conference on. IEEE, 2017, pp. 1–6.

[7] N. Moustafa and J. Slay, “The significant features of the unsw-nb15 and the kdd99 data sets for network intrusion detection systems,” in Building Analysis Datasets and Gathering

Experience Returns for Security (BADGERS), 2015 4th International Workshop on. IEEE, 2015, pp. 25–31.

[8] L. Sun, T. Anthony, H. Z. Xia, J. Chen, X. Huang, and Y. Zhang, “Detection and classification of malicious patterns in network traffic using benford’s law,” in Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), 2017. IEEE, 2017, pp. 864–872.

[9] S. M. Almansob and S. S. Lomte, "Addressing challenges for intrusion detection system using naive bayes and pca algorithm," in Convergence in Technology (I2CT), 2017 2nd International Conference for. IEEE, 2017, pp. 565–568.

[10] M. C. Raja and M. M. A. Rabbani, "Combined analysis of support vector machine and principle component analysis for ids," in IEEE International Conference on Communication and Electronics Systems, 2016, pp. 1–5.