

Credit Card Fraud Detection Using Adaboost

B.Nithin, Rohit Ravula, Shaik Gangina Sulthana

Dept of CSE

St.Peter's Engineering College Hyderabad, AP, India

Abstract – Fraud is cheating or wrongful or a culprit activity, its main aim is focus financial or personal sign. In this proposed system we uses two mechanisms namely, (i) fraud prevention and (ii) fraud detection, for avoiding loss from fraud, that detecting details from fraud. In the first fraud prevention mechanism. Is most defensive and proactive strategy, it prevents the misrepresentation from starting. At that point, the second mechanism fraud detection is guessing the fraudster. This component is required for a fake exchange, but it guesses the fraudster, in the time exchange endeavored by fraudster. Credit card fraud is a serious problem in financial services. Billions of dollars are lost due to credit card fraud every year. There is a lack of research studies on analyzing real-world credit card data owing to confidentiality issues. In this project, machine learning algorithm is used to detect credit card fraud. Standard models are first used. Then, hybrid methods which use AdaBoost . To evaluate the model efficiency, a publicly available credit card data set is used. Then, a real-world credit card data set from a financial institution is analyzed. In addition, noise is added to the data samples to further assess the robustness of the algorithms. The experimental results positively indicate that Boosting Algorithms achieves good accuracy rates in detecting fraud cases in credit cards.

Keywords– Creditcardfraud, Application Machine Learning.

I. INTRODUCTION

Credit card fraud is a serious problem in financial services. Billions of dollars are lost due to credit card fraud every year. There is a lack of research studies on analyzing real-world credit card data owing to confidentiality issues. In this paper, machine learning algorithms are used to detect credit card fraud. Standard models are firstly used. Then, hybrid methods which use AdaBoost applied. To evaluate the model efficacy, a publicly available credit card data set is used. Then, a real-world credit card data set from a financial institution is analyzed. In addition, noise is added to the data samples to further assess the robustness of the algorithms. The experimental results positively indicate that the majority voting method achieves good accuracy rates in detecting fraud cases in credit cards.

II. IMPLEMENTATION

Method of Implementation

1. Python

The python libraries used in this project are SCIKIT-LEARN, PANDAS, SEABORN, and MATPLOTLIB. We further discuss about the modules in detail.

2. Scikit-Learn:

If you are a Python programmer or you are looking for a robust library you can use to bring machine learning into a production system then a library that you will want to seriously consider is scikit-learn.

- Installation: pip install scikit-learn
- It is the main Library in our project where we import our Adaboost Classifier by following line `>>>from sklearn.ensembleimport AdaBoostClassifier`

3. Pandas:

- Pandas are an open-source, BSD-licensed Python library providing high-performance, easy-to-use data structures and data analysis tools for the Python programming language. Python with Pandas is used in a wide range of fields including academic and commercial domains including finance, economics, Statistics, analytics, etc.

- Installation: pip install pandas

- Importing the library : import pandas as pd

4. Matplotlib:

- Matplotlib.pyplot is a plotting library used for 2D graphics in python programming language. It can be used in python scripts, shell, web application servers and other graphical user interface toolkits.

- Installation: pip install matplotlib

- Import the library: import matplotlib.pyplot as p

5. Seaborn:

- Seaborn is a library for making statistical graphics in Python. It is built on top of matplotlib and closely integrated with pandas data structures.

- Installation : pip install seaborn

- Importing the library : import seaborn as sns

6. Problem Definition:

Fraud has been increasing drastically with the progression of state-of-art technology and worldwide communication. Fraud can be avoided in two main ways: prevention and

detection. Prevention avoids any attacks from fraudsters by acting as a layer of protection. Detection happens once the prevention has already failed. Therefore, detection helps in identifying and alerting as soon as a fraudulent transaction is being triggered. Recently, card not-present transactions in credit card operations have become popular among web payment gateways. According to the Nilson Report in October 2016, more than \$31 trillion were generated worldwide by online payment systems in 2015, increasing 7.3% than 2014.

Worldwide losses from credit card fraud have been rising to \$21 billion in 2015, and will possibly reach \$31 billion by 2020. However, there has been an extreme increase in fraudulent transactions that affect the economy dramatically. Credit card fraud can be classified into several categories. The two types of frauds that can be mainly identified in a set of transactions are Card-not present (CNP) frauds and Card-present (CP) frauds. Those two types can be described further by bankruptcy fraud, theft/counterfeit fraud, application fraud, and behavioral fraud. Our study aims at addressing four fraud natures that belong to the CNP fraud category described above and we propose a method to detect those frauds real time.

Machine learning is this generation's solution which replaces such methodologies and can work on large datasets which is not easily possible for human beings. Machine learning techniques fall into two main categories; supervised learning and unsupervised learning. Fraud detection can be done in either way and only can be decided when to use according to the dataset. Supervised learning requires prior classification to anomalies. During the last few years, several supervised algorithms have been used in detecting credit card fraud. The data which is being used in this study is analyzed in two main ways: as categorical data and as numerical data. The dataset originally comes with categorical data. The raw data can be prepared by data cleaning and other basic preprocessing techniques. First, categorical data can be transformed into numerical data and then appropriate techniques are applied to do the evaluation. Secondly, categorical data is used in the machine learning techniques to find the optimal algorithm.

III. PROPOSED SYSTEM

There are many machine learning models in this era which are used for credit card fraud detection. But, using Adaboost is something which increases the accuracy, other parameters and performance of the model. AdaBoost, short for Adaptive Boosting, is a machine learning meta-algorithm formulated by Yoav Freund and Robert Schapire, who won the 2003 Godel Prize for their work. It can be used in conjunction with many other types of learning algorithms to improve

performance. The output of the other learning algorithms ('weak learners') is combined into a weighted sum that represents the final output of the boosted classifier. AdaBoost is adaptive in the sense that subsequent weak learners are tweaked in favor of those instances misclassified by previous classifiers. AdaBoost is sensitive to noisy data and outliers. In some problems it can be less susceptible to the over fitting problem than other learning algorithms. The individual learners can be weak, but as long as the performance of each one is slightly better than random guessing, the final model can be proven to converge to a strong learner.

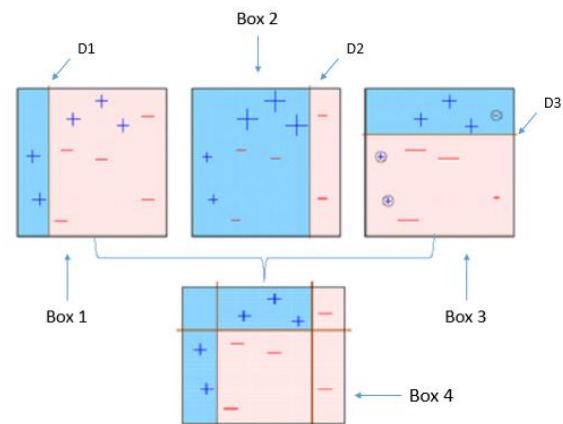


Fig. 1 AdaBoost is sensitive to noisy data and outliers

Advantages:

1. Adaboost is easy to implement.
2. It's not prone to over fitting.
3. Works better than other models as ensemble learning (Adaboost) is used.
4. More accurate and makes weak learners as strong learners.

Software requirements:

1. Operating system : Windows 7 / Linux /OSX
2. Coding Language : Python
3. Python Modules : Numpy, Pandas, Seaborn, Matplotlib, Scikit-Learn
4. Tool : Vim editor, Spyder
5. Database : excel, CSVfiles

IV. MODULE DESCRIPTION



Fig.2 Model Description.

1. **User Module:** In this module the User loads the dataset and take cares of the other input to feed the machine learning model.
2. **Machine learning module:** In this module the dataset is separated in 70% and 30% as training and testing data and Training data is fed to model to train the model.
3. **Fraud Risk Estimation Module:** This module is all about detection of fraud where the Model should be trained in such a way that every fraudulent transaction is detected.
4. **Result Module:** In this module the result is obtained as either fraudulent transaction or genuine transaction. Figure 1 below provides Activity Diagram

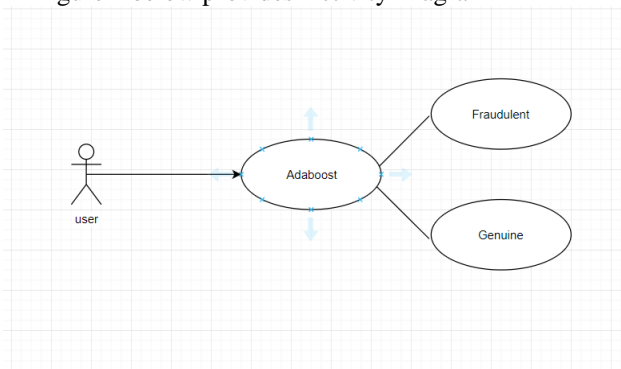


Fig. 3 Activity Diagram.

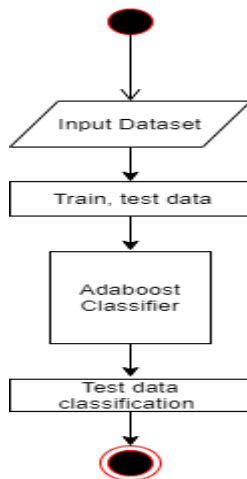


Fig.4 Below provides Credit Card fraud detection process.

Discussion of Results

```

C:\WINDOWS\system32\cmd.exe
Microsoft Windows [Version 10.0.17134.950]
(c) 2018 Microsoft Corporation. All rights reserved.

C:\Users\WITHTIN>cd Desktop
C:\Users\WITHTIN\Desktop>cd ccf
C:\Users\WITHTIN\Desktop\ccf>python ccf_0.py
    
```

Fig.5 Running the python code in command line

```

C:\WINDOWS\system32\cmd.exe
Microsoft Windows [Version 10.0.17134.950]
(c) 2018 Microsoft Corporation. All rights reserved.

C:\Users\WITHTIN\Desktop>cd ccf
C:\Users\WITHTIN\Desktop\ccf>python ccf_0.py
[*]dataset loaded
    
```

Fig.6 we get an output of dataset loaded one the dataset is loaded.

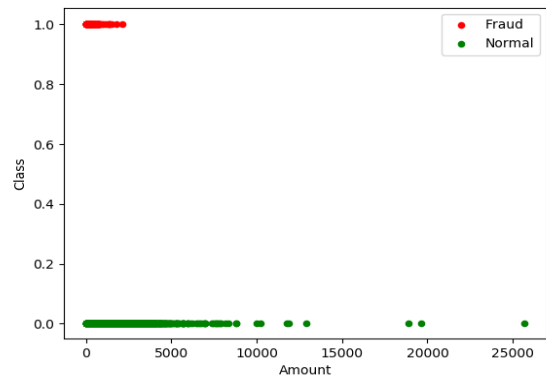


Fig.7 we visualize a graph with class(0|1) on y axis and amount on x axis determining the amount done in a fraud transaction.

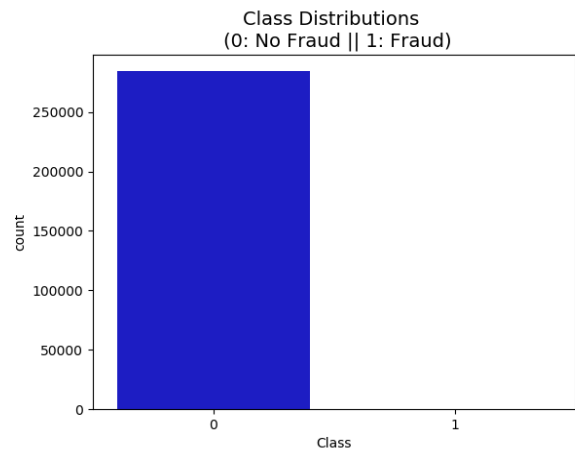


Fig.8 we visualize between genuine and fraudulent transactions in the dataset resulting in a great imbalance as we have transactions of 284,807 in which only 492 are fraudulent (0.17).

```

C:\WINDOWS\system32\cmd.exe
Microsoft Windows [Version 10.0.17134.950]
(c) 2018 Microsoft Corporation. All rights reserved.

C:\Users\WITHTIN\Desktop>cd ccf
C:\Users\WITHTIN\Desktop\ccf>python ccf_0.py
[*]dataset loaded
[*]data fabrication
    
```

Fig.9 so to overcome the Imbalance we took a fraction of dataset where the genuine and fraudulent are balanced resulting in data fabrication.

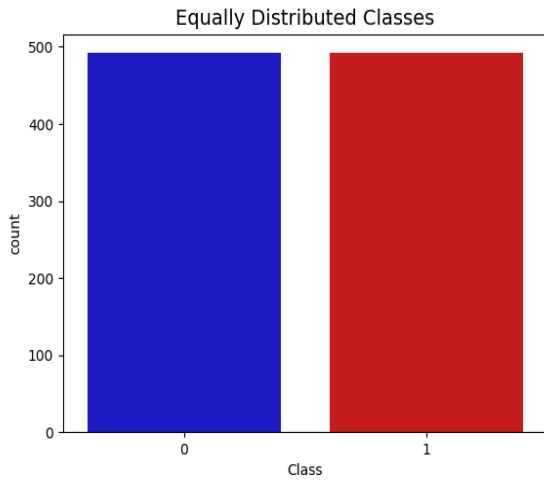


Fig.10 This is the visualization done after data fabrication resulting in data balance.

```

C:\WINDOWS\system32\cmd.exe
Microsoft Windows [Version 10.0.17134.950]
(c) 2018 Microsoft Corporation. All rights reserved.

C:\Users\WETHEN\ Desktop
C:\Users\WETHEN\Desktop\ ccf4
C:\Users\WETHEN\Desktop\ ccf4\python ccf4_0.py
[+]dataset Loaded
-----
[+]data fabrication
[+]dataset sample
      Time      V1      V2      V3      V4      ...      V26      V27      V28      Amount      Class
81794  59089.0 -0.256070  0.871836  1.408880 -0.272186 ...  0.554828 -0.851561  0.813347  49.95  0
251891  155548.0  1.878238  1.325638 -2.333469  4.231511 ... -0.002601 -0.826667  0.805238  7.06  1
6668   8246.0  1.836273 -0.148851  0.988459  1.682407 ... -0.314981  0.834172  0.822541  54.37  0
154720  182676.0 -5.552122  5.678134 -9.775528  8.416295 ...  0.420919 -0.297557 -0.946184  119.74  1
182782  68357.0  1.232684 -0.548931  1.887873  0.894082 ... -0.322687  0.888885  0.835427  19.59  1
[5 rows x 31 columns]

[+] Labels sample
81794  0
251891  1
6668   0
154720  1
182782  1
Name: Class, dtype: int64

[+] Features sample
      Time      V1      V2      V3      V4      ...      V25      V26      V27      V28      Amount
81794  59089.0 -0.256070  0.871836  1.408880 -0.272186 ...  0.144452  0.554828 -0.851561  0.813347  49.95
251891  155548.0  1.878238  1.325638 -2.333469  4.231511 ...  0.171089  0.805081 -0.826667  0.805238  7.06
6668   8246.0  1.836273 -0.148851  0.988459  1.682407 ...  0.555087 -0.314981  0.834172  0.822541  54.37
154720  182676.0 -5.552122  5.678134 -9.775528  8.416295 ... -0.184925  0.420919 -0.297557 -0.946184  119.74
182782  68357.0  1.232684 -0.548931  1.887873  0.894082 ...  0.527258 -0.322687  0.888885  0.835427  19.59
[5 rows x 30 columns]

```

Fig.13 Creating Features dataframe which contains all the columns of dataset except Class and printing first five rows of it.

```

C:\WINDOWS\system32\cmd.exe
Microsoft Windows [Version 10.0.17134.950]
(c) 2018 Microsoft Corporation. All rights reserved.

C:\Users\WETHEN\ Desktop
C:\Users\WETHEN\Desktop\ ccf4
C:\Users\WETHEN\Desktop\ ccf4\python ccf4_0.py
[+]dataset Loaded
-----
[+]data fabrication
[+]dataset sample
      Time      V1      V2      V3      V4      ...      V26      V27      V28      Amount      Class
81794  59089.0 -0.256070  0.871836  1.408880 -0.272186 ...  0.554828 -0.851561  0.813347  49.95  0
251891  155548.0  1.878238  1.325638 -2.333469  4.231511 ... -0.002601 -0.826667  0.805238  7.06  1
6668   8246.0  1.836273 -0.148851  0.988459  1.682407 ... -0.314981  0.834172  0.822541  54.37  0
154720  182676.0 -5.552122  5.678134 -9.775528  8.416295 ...  0.420919 -0.297557 -0.946184  119.74  1
182782  68357.0  1.232684 -0.548931  1.887873  0.894082 ... -0.322687  0.888885  0.835427  19.59  1
[5 rows x 31 columns]

```

Fig.11 Then we print the dataset sample which returns first five rows and dimensions (5 X 31).

```

C:\WINDOWS\system32\cmd.exe
Microsoft Windows [Version 10.0.17134.950]
(c) 2018 Microsoft Corporation. All rights reserved.

C:\Users\WETHEN\ Desktop
C:\Users\WETHEN\Desktop\ ccf4
C:\Users\WETHEN\Desktop\ ccf4\python ccf4_0.py
[+]dataset Loaded
-----
[+]data fabrication
[+]dataset sample
      Time      V1      V2      V3      V4      ...      V26      V27      V28      Amount      Class
81794  59089.0 -0.256070  0.871836  1.408880 -0.272186 ...  0.554828 -0.851561  0.813347  49.95  0
251891  155548.0  1.878238  1.325638 -2.333469  4.231511 ... -0.002601 -0.826667  0.805238  7.06  1
6668   8246.0  1.836273 -0.148851  0.988459  1.682407 ... -0.314981  0.834172  0.822541  54.37  0
154720  182676.0 -5.552122  5.678134 -9.775528  8.416295 ...  0.420919 -0.297557 -0.946184  119.74  1
182782  68357.0  1.232684 -0.548931  1.887873  0.894082 ... -0.322687  0.888885  0.835427  19.59  1
[5 rows x 31 columns]

[+] Labels sample
81794  0
251891  1
6668   0
154720  1
182782  1
Name: Class, dtype: int64

[+] Features sample
      Time      V1      V2      V3      V4      ...      V25      V26      V27      V28      Amount
81794  59089.0 -0.256070  0.871836  1.408880 -0.272186 ...  0.144452  0.554828 -0.851561  0.813347  49.95
251891  155548.0  1.878238  1.325638 -2.333469  4.231511 ...  0.171089  0.805081 -0.826667  0.805238  7.06
6668   8246.0  1.836273 -0.148851  0.988459  1.682407 ...  0.555087 -0.314981  0.834172  0.822541  54.37
154720  182676.0 -5.552122  5.678134 -9.775528  8.416295 ... -0.184925  0.420919 -0.297557 -0.946184  119.74
182782  68357.0  1.232684 -0.548931  1.887873  0.894082 ...  0.527258 -0.322687  0.888885  0.835427  19.59
[5 rows x 30 columns]

[+] operation of test, train data

```

Fig.14 Then we take these Label and Feature Dataframes and split them into Training and Testing Data.

```

C:\WINDOWS\system32\cmd.exe
Microsoft Windows [Version 10.0.17134.950]
(c) 2018 Microsoft Corporation. All rights reserved.

C:\Users\WETHEN\ Desktop
C:\Users\WETHEN\Desktop\ ccf4
C:\Users\WETHEN\Desktop\ ccf4\python ccf4_0.py
[+]dataset Loaded
-----
[+]data fabrication
[+]dataset sample
      Time      V1      V2      V3      V4      ...      V26      V27      V28      Amount      Class
81794  59089.0 -0.256070  0.871836  1.408880 -0.272186 ...  0.554828 -0.851561  0.813347  49.95  0
251891  155548.0  1.878238  1.325638 -2.333469  4.231511 ... -0.002601 -0.826667  0.805238  7.06  1
6668   8246.0  1.836273 -0.148851  0.988459  1.682407 ... -0.314981  0.834172  0.822541  54.37  0
154720  182676.0 -5.552122  5.678134 -9.775528  8.416295 ...  0.420919 -0.297557 -0.946184  119.74  1
182782  68357.0  1.232684 -0.548931  1.887873  0.894082 ... -0.322687  0.888885  0.835427  19.59  1
[5 rows x 31 columns]

[+] Labels sample
81794  0
251891  1
6668   0
154720  1
182782  1
Name: Class, dtype: int64

```

Fig.12 Creating Label dataframe which contains the class column of the dataset and printing the first five rows of it.

```

C:\WINDOWS\system32\cmd.exe
[5 rows x 30 columns]
-----
[+] operation of test, train data
-----
[+]initializing classifiers

```

Fig.15 Then we Initialize the classifier and fit the training and testing data into it and make predictions over test data.

```
C:\WINDOWS\system32\cmd.exe
[5 rows x 30 columns]
-----
[+] speration of test, train data
-----
[+]initializing classifiers
-----
[+]accuracy: 95.1219512195122
```

Fig.16 after predicting the test data we get the Accuracy of our model.

```
C:\WINDOWS\system32\cmd.exe
[5 rows x 30 columns]
-----
[+] speration of test, train data
-----
[+]initializing classifiers
-----
[+]accuracy: 95.1219512195122
confusion matrix
[[123  4]
 [ 8 111]]
```

Fig.17 we get a Confusion matrix where we get a estimation of the performance of our model.

V. CONCLUSION AND FUTURE WORK

A study on credit card fraud detection using machine learning algorithms. A number of standard models which include NB, SVM have been used in the empirical evaluation. A publicly available credit card data set has been used for evaluation using individual (standard) models and hybrid models using AdaBoost and combination methods. The MCC metric has been adopted as a performance measure, as it takes into account the true and false positive and negative predicted outcomes. The best MCC score is 0.823, achieved using Adaboost. A real credit card data set from a financial institution has also been used for evaluation. The same individual and hybrid models have been employed. To further evaluate the hybrid models, noise from 10% to 30% has been added into the data samples. This shows that the Adaboost method is stable in performance in the presence of noise. For future work, the methods studied in this paper will be extended to online learning models. In addition, other online learning models will be investigated. The use of online learning will enable rapid detection of fraud cases, potentially in realtime. This in turn will help detect and prevent fraudulent transactions before they take Place, which will reduce the number of losses incurred every day in the financial sector.

VI. FUTURE SCOPE

We believe that our study will have profound impact on both research and Industry communities. In the future, more advanced deep learning models such as Convolutional Neural Networks can be explored for feature learning. We will also consider improving the

current feature mapping method through ideas in transferring learning.

REFERENCES

- [1]. Nancy R Mead, Carol C Woody, "Cyber Security Engineering – A practical Approach for Systems and Software Assurance" , SEI Series in Software Engineering, Addison Wesley Pearson, ISBN No 978-93-325-8589-8, 2017
- [2]. Dr.G.Anil Kumar, Dr.M.Upendra Kumar, Dr.Sesham Anand, Dr.D.Shravani, "Novel Design of Machine Learning for Malicious Software Analysis – Malicious URL Case Study", Vol 6 Issue 4 October 2018 – December 2018 Pp 292-298 International Journal of Interdisciplinary Research and Innovations (IJIRI) ISSN 2348-1218 (print) ISSN 2348-1226 (online)
- [3]. <https://github.com/surajr/URL-Classification/find/master> (last accessed on 02.10.2019)
- [4]. S. Carolin Jeeva, Elijah Blessing Rajsingh, "Intelligent phishing url detection using association rule mining", Springer Open, Human-centric Computing and Information Sciences, Jeeva and Rajsingh Hum.Cent.Comput.Inf.Sci.(2016), DOI: 10.1186/s13673-016-0064-3
- [5]. Anjali B. Sayamber, Arati M. Dixit, "Malicious URL Detection and Identification", Internal Journal of Computer Applications(0975 – 8887) , volume 99 – No.17, (2014)
- [6]. Min-Sheng Lin, Chien-Yi Chiu, Yuh-Jye Lee and Hsing-Kuo Pao, "Malicious URL Filtering – A Big Data Application", IEEE International Conference on Big Data, Page No.589-596,(2013)
- [7]. Piyush AnastaRumao, "Using Two Dimensional Hybrid Feature Dataset to Detect Malicious Executables" , International Journal of Innovative Research in Computer and Communication Engineering, Vol.4, Issue 7,(2016), DOI: 10.15680/IJIRCCE.2016.0407158
- [8]. Matthew G. Schultz and Eleazar Eskin, Erez Zadok, Salvatore J. Stolfo, "Data Mining Methods for Detection of New Malicious Executables", Security and Privacy, S&P Proceedings. 2001 IEEE Symposium, SP'01, 38,(2001)
- [9]. Ankit Kumar Jain and B. B. Gupta, "A Novel approach to protect against phishing attacks at client side using auto-updated white-list", Springer Open, EURASIP Journal on Information Security, Jain and Gupta EURASIP Journal on Information Security,(2016),DOI. 10.1186/s1335-0160034-3
- [10]. Chia-Mei Chen, D.J. Guan, Qun-Kai Su, "Feature set identification for detecting suspicious URLs using Bayesian classification in social networks", ELSEVIER, Information Sciences 289 (2014),pg: 133-147.

- [11]. Asrian Stefan Popescu, Dragos Teodor Gavrilut, Dumitru Bogdan Prelipcean, "A Study on Techniques for Proactively Identifying Malicious URLs", International Symposium on Symbolic and Numeric Algorithms for Scientific Computing, IEEE (2016), DOI: 10.1109/SYNASC.2015.40
- [12]. Amruta Rajeev Nagaonkar, Umesh L. Kulkarni, "Finding the malicious URLs using Search Engines", International Conference on Computing for Sustainable Global Development, IEEE(2016)
- [13]. B. B. Gupta, Aakanksha Tewari, Ankit Kumar Jain, Dharma P. Agrawal, "Fighting against phishing attacks: state of the art and future challenges", Springer (2016).
- [14]. Adrian-Stefan Popescu, Dragos-Teodor Gavrilut, Daniel-Ionut Irimia, "A practical approach for clustering large data flows of malicious URLs", J Comput Virol Hack Tech , Springer (2016),pg:37-47, DOI. 10.1007/s11416-015-0239-x
- [15]. Cheng Cao and James Caverlee, "Detecting Spam URLs in Social Media via Behavioral Analysis", Springer International Publishing Switzerland (2015),pp. 703-714
- [16]. Mohammad Saiful Islam Mamun, Arash Habibi Lashkari, Natalia Stakhanova, Ali A. Ghorbani, "Detecting Malicious URLs using Lexical Analysis", Springer International Publishing(2016), pp. 467-482
- [17]. Aaron Blum, Brad Wardman, Thamar Solorio, Gary Warner, "Lexical Feature Based Phishing URL Detection using Online Learning", ACM (2010).
- [18]. Hyunsang Choi, Bin B. Zhu, Heejo Lee, "Detecting Malicious Web Links and Identifying Their Attack Types".
- [19]. Justin Ma, Lawrence K. Saul, Stefan Savage, Geoffrey M. Voelker, "Identifying Suspicious URLs : An Application of Large-Scale Online Learning", International Conference on Machine Learning (2009).
- [20]. Mehedy Masud, Latifur Khan, and Bhavani Thuraisingham, "Data Mining Tools for Malware Detection", International Standard Book Number-13: 978-1-4665-1648-9, (2011).
- [21]. Justin Ma, Lawrence K. Saul, Stefan Savage, Geoffrey M. Voelker, "Beyond Blacklists: Learning to Detect Malicious Web Sites from Suspicious URLs", KDD (2009).
- [22]. Seow Wooi Liew, Nor Fazlida Mohd Sani, Mohd. Taufik Abdullah, Razali Yaakob, Mohd Yunus Sharum, "Improvement Of Classification Features To Increase Phishing Tweets Detection Accuracy", Journal of Theoretical and Applied Information Technology(2018), E-ISSN: 1817-3195
- [23]. B. B. Gupta, Nalin A.G. Arachchilage, Konstantinos E. Psannis, "Defending against Phishing Attacks: Taxonomy of Methods, Current Issues and Future Directions"
- [24]. Khulood Al Messabi, Monther Aldwairi, Ayesha Al Yousif, Anoud Thoban, Fatna Belqasmi, "Malware Detection using DNS Records and Domain Name Features", ACM(2018), <https://doi.org/10.1145/3231053.3231082>
- [25]. Yoshitaka Nakamura, Shihori Kanazawa, Hiroshi Inamura, Osamu Takahashi, "Classification of unknown Web sites based on yearly changes of distribution information of malicious IP addresses", IEEE(2018)
- [26]. Grega Vrbancic, Iztok Fister Jr., Vili Podgorelec, "Swarm Intelligence Approaches for Parameter Setting of Deep Learning Neural Network: Case Study on Phishing Websites Classification", ACM(2018), <https://doi.org/10.1145/3227609.3227655>
- [27]. Nancy R. Mead, Carol C. Woody, "Cyber Security Engineering", Pearson , ISBN : 978-93-325-8589-8
- [28]. Mark Stamp, "Introduction to Machine Learning with Applications in Information Security", CRC Press, ISBN: 978-1-1386-26782
- [29]. Clarence Chio, David Freeman, "Machine Learning and Security", O'REILLY ISBN:978-93-5213-693-3
- [30]. Andrea Isoni, "Machine Learning for the Web", PACKT ISBN:978-1-78588-660-7
- [31]. Sumeet Dua, Xian Du, "Data Mining and Machine Learning in Cybersecurity", CRC Press ISBN:978-1-4398-3942-3
- [32]. Cylance Data Science Team, "Introduction to Artificial Intelligence for Security Professionals" CYLANCE Press ISBN: 978-0-9980169-0-0
- [33]. Shaik. Irfan Babu , Dr. M.V.P. Chandra Sekhar Rao , G. Nagi Reddy, "Research Methodology on Web Mining for Malware Detection" , IJCTT V12(4): 152-160, (2014), DOI:10.14445/22312803/IJCTT-V12P131.