Artificial Intelligence formulated this projection for compatibility purposes from the original article published at Global Journals. However, this technology is currently in beta. Therefore, kindly ignore odd layouts, missed formulae, text, tables, or figures.

1 2	A Machine Learning-based Model for Implementing the Cybersecurity for Organizations 'Assets
3	Ramze Sanea
4	Received: 1 January 1970 Accepted: 1 January 1970 Published: 1 January 1970
5	

Abstract 6

16

Abstractâ??"In the current era, cybersecurity problems are one of the critical problems that 7 threaten the organizations? assets, since they may cause a big financial and moral loss. and in 8

the parallel the with the advent of the Machine Learning and Artificial Intelligence techniques, 9

It is important and wise to use these technologies to help in achieving cybersecurity for 10

organizations? assets, due to accurate work of these systems and saving time, effort, and cost. 11

So, this research develops a model that uses machine learning technology to detect the 12

vulnerability in the information security of the organizations ?assets to avoid as possible the 13

lack of the information security in organizations ?assets and thus avoid the financial and moral 14 loss that such organizations may face.

15

Index terms— information security policy, ISP, machine learning, dataset, vulnerable, CART. Abstract-In the current era, cybersecurity problems are one of the critical problems that threaten the 17

18 organizations' assets, since they may cause a big financial and moral loss. and in the parallel the with the 19 advent of the Machine Learning and Artificial Intelligence techniques, It is important and wise to use these 20 technologies to help in achieving cybersecurity for organizations' assets, due to accurate work of these systems 21 and saving time, effort, and cost. So, this research develops a model that uses machine learning technology to 22 detect the vulnerability in the information security of the organizations' assets to avoid as possible the lack of the 23 information security in organizations' assets and thus avoid the financial and moral loss that such organizations 24 may face. 25

1 I. Introduction 26

achine learning aims to answer the question of how to create machines that make a decision on their own depending 27 on the learned dataset. It is a mix between computer science and statistics, as well as it is at the heart of artificial 28 intelligence and data science, and is one of today's fastest expanding technological topics. 29

The information security policy is one of the most critical information security measures. This important 30 directiongiving document, however, is not always easy to create, and its writers struggle with issues such as 31 what defines a policy. As a result, policymakers are forced to rely on existing sources for direction. The many 32 worldwide information security standards are one of these sources. 33

A dataset is a set of data, usually represented in the form of a table. Each column in the table represents 34 a specific feature, and each row returns to an element of the dataset that is used in the learning phase when 35 building a model. 36

CART stands for Classification and Regression Tree, can easily handle both numerical and categorical CART 37 stands for Classification and Regression Tree, can easily handle both numerical and categorical variables, but 38 there is tow common disadvantages the first one is CART split one variable at a time, the other one is CART 39

may create unstable decision tree. 40

2 II. Research Objectives 41

? The main goals of the research are to suggest best practices to cover the vulnerabilities that belong to user 42 assets. ? Assisting the user in identifying unknown vulnerabilities that could be destructive and recommending 43

appropriate solutions while keeping the budget in mind. 44

3 III. Research Methodology 45

A suitable dataset is required for any machine learning model, so before implementing any model, it is necessary 46 to provide a suitable dataset that is used to train the machine, and the next step is to split the dataset into 47 two subsets using an appropriate splitting algorithm, the first one being used to train the machine. After that, 48 the appropriate ML algorithm must be chosen and used for dataset exploration and pattern recognition with 49 minimum human intervention. Finally, after the algorithm chooses the second subset, it is used to test and 50 evaluate the model. Before input, the dataset should be converted to a numerical value so that the algorithm 51 can deal with it and add a new column that contains the unique numerical key points to the solution. Then the 52 numerical dataset will be entered as an input to train and test the model, in addition to the user asset and the 53 budget that he has. 54

4 b) Processing 55

After the dataset enters the model, it should be split into two subdatasets. The first one is to train and the other 56 is to test, and to do so, the machine uses a kfold algorithm to divide the origenal dataset into test dataset with 57 0.30 of the dataset at all and the remain data should be the train dataset. The random forest algorithm uses the 58 training dataset by building a number of decision trees. Then combine them to one tree. 59

c) Output 5 60

61

62

63

64

67

The final step is to predict the solution key for the user asset, but before that, the machine uses the testing subdataset to evaluate the usage algorithm and the model. So the input data from the user enters into the model and takes a suitable path on the tree that has already been built in the processing phase, and when the machine reaches the leaf that is the prediction that the user needs, the result is converted to the related solution that maps to the prediction key value. Styling one of the important factor to develop an efficient ISP so it should be write 65 in clear and consist manner in addition to fit with the organizational culture, the second factor is Development 66 so it should be updated from time to time to be suitable with the organization requirement. the other factor for develop in effective ISP is commitment because when the employee in the organization see the header and all the 68

manger commit to the ISP this means all the employee should Committed to it. 69

VI. Application Example 6 70

After splitting phase the machine calculate information gain for each feature in the splitting datasets to indicate 71 the impurity for each one of them using the following equation. 72

7 Fig. 4: Information gain 73

When the machine using random force algorithm it must calculate nodes importance for each created decision 74 tree using the following equation. After the machine learn the dataset that entered it well be ready to suggest 75 the key of the solution for the most important vulnerability that the machine see depending on the taken path 76 that the machine take so when the key predicted the check if this key is exist in the interred dataset so if it exist 77 the machine return the solution that related to this key, on other hand if the predicted key does not exist in the 78 dataset the machine try to approximate this key to the nearest value that exist in the dataset and return the 79

8 VII. Conclusion 81

solution that related with it.

Machine learning is a wondrous method by which to solve a critical problem. One of them is organization security, 82 so it can help to build an information security policy in a short time with high accuracy. 83

Assets are the most important thing that the organization has so ot will be secure as possible we can and we 84 can let it more secure by cover the vulnerability for all one of this assets by choose the best solution for all one 85 of it. 86

In this research the machine learning used to suggest a solution for one vulnerability that belong to the user 87

asset, so the machine depending on both mathematical equation and suitable path in the tree that built in 88 learning phase to choose what one of the vulnerability should choose to suggest the solution for protect against 89

it. 90

80



Figure 1: Fig. 1 :



 $\mathbf{2}$

Figure 2: Fig. 2 :



Figure 3: Fig. 3 :

Gain(T,X) = Entropy(T) - Entropy(T,X)

T = target variable

 $\mathbf{5}$

X = Feature to be split on

• Entropy(T,X) = The entropy calculated after the data is split on feature X

Figure 4: Fig. 5 :

$$ni_j = w_j C_j - w_{left(j)} C_{left(j)} - w_{right(j)} C_{right(j)}$$

- ni sub(j) = the importance of node j
- w sub(j) = weighted number of samples reaching node j
- C sub(j) = the impurity value of node j
- left(j) = child node from left split on node j
- right(j) = child node from right split on node j

Figure 5: Fig. 6:

$$fi_i = \frac{\sum_{j:node \ j \ splits \ on \ feature \ i} ni_j}{\sum_{k \in all \ nodes} ni_k}$$

- fi sub(i) = the importance of feature i
- ni sub(j) = the importance of node j

 $\mathbf{78}$

Figure 6: Fig. 7 : Fig. 8 :

Figure 7:

- [Ram et al. ()] 'A process model for analyzing and managing flexibility in information systems'. L Ram , Kumar
 , C Antonis , Stylianou . European Journal of Information Systems 2013.
- Briand and Labiche ()] 'A UML-Based Approach to System Testing'. Lionel Briand , Yvan Labiche . Journal of
 Carleton University 2002.
- ⁹⁵ [Lejk and Deeks ()] An Introduction to Systems Analysis Techniques, M Lejk , Deeks . 2002. Perason Education
 ⁹⁶ Limited. (2nd Ed)
- [Maciaszek ()] Requirments Analysis and System Design, Developing Information Systems with UML, A
 Maciaszek . 2001. Addison-Wesely.
- 99 [Riechert ()] 'Research Information Standardization as a Wicked Problem: Possible Consequences for the
- Standardization'. Mathias Riechert . Proceedings of The International Conference of Current Research
 Information Systems, (The International Conference of Current Research Information SystemsAt Rome)
- 102 2014. 12.
- 103 [Satzinger ()] J Satzinger . Systems Analysis and Design, 2002. Thomson Learning. (2nd ed.)
- [Vizulis and Diebelis ()] 'Self-Testing Approach and Testing Tools'. Valdis Vizulis , Edgars Diebelis . Journal of
 University of Latvia 2012. 787.
- [Dennis ()] Systems Analysis and Design: An-Object-Oriented Approach with UML, A Dennis . 2002. John Wiley
 Sons Inc.