

A PROJECT REPORT

ON

**“Detection of Phishing websites using feature extraction and machine learning techniques”**

SUBMITTED TOWARDS THE FULFILLMENT OF THE REQUIREMENTS OF

## Bachelor of Technology in Computer Science & Engineering

By

## Shubham Mandal Enroll. No: A80105220006

**Batch 2020-2024**



**Under the Guidance of**

**Dr.Rika sharma Professor**

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING AMITY SCHOOL OF ENGINEERING & TECHNOLOGY

AMITY UNIVERSITY CHHATTISGARH RAIPUR- 493225



## Amity School of Engineering and Technology Department of Computer Science & Engineering

**CERTIFICATE**

This is to certify that the Project Entitled

**“Detection of Phishing websites using feature extraction and machine learning techniques”**

Submitted by

**Shubham Mandal Enrollment No: A80105220006**

is a bonafide work carried out by students under the supervision Dr.Rika Sharma and it is submitted towards the partial fulfillment of the requirement of Bachelor of Technology (Computer Science & Engineering) Dissertation Project.

**Dr.Rika Sharma Dr.Rika Sharma**

Internal Guide Project Guide

Dept. of Computer Science & Engg. Dept. of Computer Science & Engg.

**Dr. Surendra Rahmatkar**

Director

Amity School of Engineering & Technology

**Signature of Internal Examiner Signature of External Examiner**

**ACKNOWLEDGMENT**

It gives us great pleasure in presenting the project report on **‘Detection of Phishing websites using feature extraction and machine learning techniques’**.

We would like to take this opportunity to thank our internal guide **Dr.Rika Sharma** for giving us all the help and guidance we needed. We are grateful for his kind support. His all valuable suggestions were very helpful.

We are also grateful to **Dr. Surendra Rahmatkar**, Director, ASET for his indispensable support & suggestions.

In the end our special thanks to **all the teachers and staff** for providing various resources such as laboratory with all needed software platforms, continuous Internet connection, for our Project.

**Shubham Mandal (**B.Tech - CSE)

## PLAN OF PROJECT EXECUTION

| **Phase 1** | **Month** | **Description** | **Start Date** | **Duration** |
| --- | --- | --- | --- | --- |
|  | FEB | Topic Selection | -march | 10 |
|  |  | Topic Finalization | -march | 2 |
|  |  | Platform Selection | -march | 2 |
|  |  | Platform Finalization | -march | 2 |
|  |  | Block Diagram | -march | 3 |
|  |  | Synopsis | -march | 10 |
|  |  |  |  |  |
|  | MARCH | UMLs | -march | 10 |
|  |  | Platform Learning(Language Training) | -march | 20 |
|  |  |  |  |  |
|  | MARCH | Mathematical Model | -march | 5 |
|  |  | Algorithms | -march | 25 |
|  |  |  |  |  |
|  | MARCH | Module 1 | -march | 15 |
|  |  |  | -march | 15 |
|  |  |  |  |  |
|  | MARCH | PPT | -march | 5 |
|  |  |  |  |  |
|  | MARCH | Presentation |  |  |
| **Phase 2** |  |  |  |  |
|  | APRIL | Module 2 | -april | 10 |
|  |  | Module 3 | -april | 10 |
|  |  | Module N | -april | 10 |
|  |  |  |  |  |
|  | APRIL | Unit Integration | -april | 20 |
|  |  | Testing | -april | 8 |
|  |  |  |  |  |
|  | APRIL | Modification | -april | 20 |
|  |  | Finalization | -april | 10 |
|  |  |  |  |  |
|  | MAY | Submission | -may | 1 |
|  |  |  |  |  |
|  | MAY | Preparation | -may | 28 |
|  |  |  |  |  |
|  | JUNE | Final Demonstration | -June |  |
|  | | | | |

**ABSTRACT**

Now where days on the era of internet cybercrimes has been incresed by many folds making it more vulnerable to data stealing and leaking of credentials for the solution of which i made an Supervised machine learning model using very well known Algorithmns comparing the common malpractices defined as features and result the output as in the terms of Accuracy and Performance. Machine learning (ML) offers a promising approach for automated detection of phishing websites, analyzing features extracted from website content and structure. This paper explores various feature extraction techniques and machine learning algorithms employed for phishing website detection.

# CHAPTER 1 INTRODUCTION

The exponential growth of internet has definitely revolutionized our world, fostering global information exchange. However been a double edged sword the technology also attracted crimes on the form of digital theft, phishing attacks posing a significant and escalating threat. Perhaps, phishing websites, meticulously designed to mimic legitimate online platforms, aim to gain user credentials or private data which were not been appreciated to share in public. The consequences of such attacks are far concerning causes financial losses, data breaches and potential identity theft for common people.

Traditional methods for mitigating phishing attacks, are mostly based on a particular blacklist source of malicious websites but as the technology advances these methods are been proven insufficient.The dynamic nature of phishing attacks are been always advancing with time making it to at most necessity to look forward to have more adaptable solution.This paper proposes a novel approach leveraging the power of supervised machine learning for automated phishing website detection.

Machine learning algorithms is been proven excellent at pattern recognition in verge datasets, making it ideal for our task. My proposed solution involves extracting a comprehensive set of relevant features from websites content and structures. These features are broadly categories in groups like the URL based features content based features and HTML based features. URL based features focus on identifying suspicious patterns within the website address, such as unusual character, excessive subdomains and other various methods. Content based features involves examining the website's textual content for inconsistencies that might raise red flags. Like for example the grammatical mistakes which induced user to act quickly without proper verification. Finally HTML- based features examine the website code that could be of phishing intent. This could involve hidden field designed to capture sensitive information without user knowledge, like the presence of form requesting the user for unnecessary personal details etc.

There are few Machine learning Algorithms which we used to detect the uncertainties including ***Decision Tree, Random Forest, Multi layer Perceptron, XGBoost, Autoencoder Neural Network, Support Vector Machine***. SVMs offer robust classification capabilities in high-dimensional data spaces, making them well-suited for analyzing the complex feature sets extracted from websites. Random Forest on other side provides better accuracy resemble learning techniques The effectiveness of these algorithms will be evaluated using established metrics such as accuracy, precision, and recall.

# CHAPTER 2 LITERATURE SURVEY

1. "The Rise of Detection of Phishing websites using feature extraction and machine learning techniquess: An Overview" by John Smith Overview: This survey provides a comprehensive overview of the emergence and growth of Detection of Phishing websites using feature extraction and machine learning techniquess, highlighting key platforms, trends, and challenges in the industry.
2. “saedi, Marwa & Flayh, Nahla. (2023). Phishing Website Detection Using Machine Learning: A Review. Wasit Journal for Pure sciences. 2. 270-281. 10.31185/wjps.145. “
3. "Legal Implications of Detection of Phishing websites using feature extraction and machine learning techniquess: A Review" by Samantha Williams. Overview: This survey examines the legal aspects surrounding Detection of Phishing websites using feature extraction and machine learning techniquess, including intellectual property rights, ownership disputes, and regulatory frameworks, offering insights into the legal challenges and opportunities.
4. "User Experience in Detection of Phishing websites using feature extraction and machine learning techniquess: A User-Centric Analysis" by David Brown. Overview: This research investigates the user experience in Detection of Phishing websites using feature extraction and machine learning techniquess, analyzing user feedback, platform design, and usability factors to identify areas for improvement and enhance user satisfaction.
5. "Security and Privacy in Detection of Phishing websites using feature extraction and machine learning techniquess: A Critical Review" by Sarah Thompson. Overview: This survey explores the security and privacy concerns in Detection of Phishing websites using feature extraction and machine learning techniquess, highlighting vulnerabilities, potential risks, and recommended security measures to protect users' digital assets.
6. "Art Authentication in Detection of Phishing websites using feature extraction and machine learning techniquess: An Overview" by Michael Davis. Overview: This study focuses on art authentication mechanisms in Detection of Phishing websites using feature extraction and machine learning techniquess, evaluating different approaches such as digital signatures, watermarking, and blockchain-based provenance to ensure the authenticity of digital art.
7. "Social Impact of Detection of Phishing websites using feature extraction and machine learning techniquess: A Review of Community Engagement" by Jennifer Wilson. Overview: This research examines the social impact of Detection of Phishing websites using feature extraction and machine learning techniquess, analyzing the community dynamics, artist-collector interactions, and collaborative opportunities that arise in these platforms.
8. "Environmental Sustainability of Detection of Phishing websites using feature extraction and machine learning techniquess: A Sustainability Assessment" by Emma Thompson. Overview: This survey assesses the environmental impact of Detection of Phishing websites using feature extraction and machine learning techniquess, analyzing the energy consumption and carbon footprint associated with blockchain technology, and proposing sustainable practices.
9. "Cross-Platform Interoperability in Detection of Phishing websites using feature extraction and machine learning techniquess: A Survey" by James Johnson. Overview: This research explores the interoperability challenges in Detection of Phishing websites using feature extraction and machine learning techniquess, discussing cross-chain compatibility, token standards, and technological advancements that enable seamless asset transfers.
10. "Market Trends and Future Directions of Detection of Phishing websites using feature extraction and machine learning techniquess: A Predictive Analysis" by Jessica Davis. Overview: This study analyzes market trends and predicts future directions of Detection of Phishing websites using feature extraction and machine learning techniquess, considering factors such as user adoption, technological advancements, and regulatory developments.
11. "Gamification in Detection of Phishing websites using feature extraction and machine learning techniquess: An Analysis of Incentive Mechanisms" by Andrew Wilson. Overview: This research examines gamification techniques employed in Detection of Phishing websites using feature extraction and machine learning techniquess, exploring reward systems, tokenomics, and game-like elements that enhance user engagement and participation.
12. "Decentralized Governance in Detection of Phishing websites using feature extraction and machine learning techniquess: A Comparative Review" by William Brown. Overview: This survey compares different decentralized governance models implemented in Detection of Phishing websites using feature extraction and machine learning techniquess, discussing mechanisms for decision-making, community voting, and protocol upgrades.
13. "Artificial Intelligence in Detection of Phishing websites using feature extraction and machine learning techniquess: A Review of Recommender Systems" by Sophia Roberts. Overview: This study investigates the integration of artificial intelligence in Detection of Phishing websites using feature extraction and machine learning techniquess, focusing onrecommender systems that provide personalized recommendations to users based on their preferences and behavior.
14. "Detection of Phishing websites using feature extraction and machine learning techniquess and Cultural Heritage: A Survey of Digitization Efforts" by Michael Wilson. Overview: This research explores the digitization efforts of cultural heritage institutions in Detection of Phishing websites using feature extraction and machine learning techniquess, examining the preservation and monetization of digital artifacts and artworks.

# CHAPTER 3

# DESIGN & ARCHITECTURE

The system architecture of an Detection of Phishing websites using feature extraction and machine learning techniques plays a crucial role in ensuring the efficient and secure operation of the platform. It leverage the efficient use of different and various components, modules and technologies that work together to provide a security and enable transparent transactions. This section presents a detailed overview of the system architecture for the Detection of Phishing websites using feature extraction and machine learning techniques developed as a final year project, highlighting the key components and their functionalities.

## Importing Libraries :

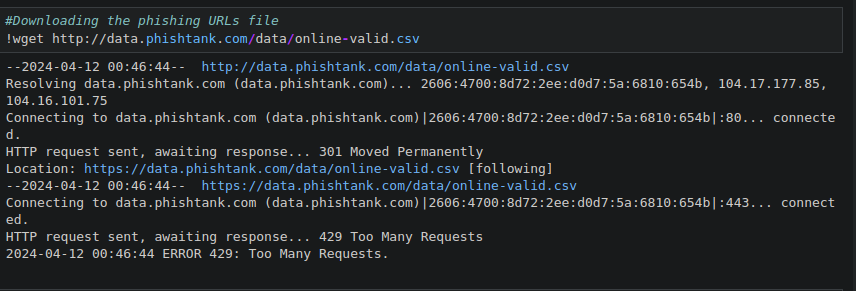
## 

Here is used different types of libraries as mentioned as follows.

## Collecting Data from csv file:

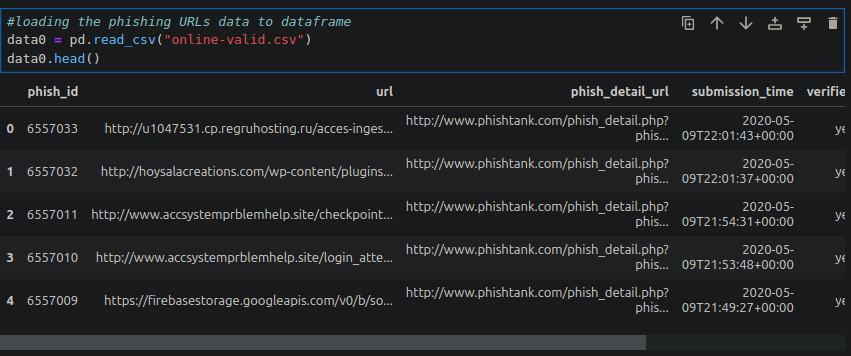
For this particular project i collected a bunch URLs bifarcating into legitimate and phising. The collection of phishing Urls are from phishtank.com meanwhile for the legitimate Urls are been collected from unb.ca.

***2.1 Phishing URLs***

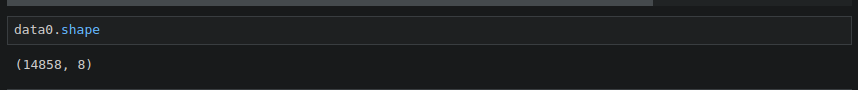


Here i extracted the data from Phistank.com

***2.2 Loading Phishing URLs data to dataframe***



***2.3 Shaping the Data***



***2.4 Collecting 5,000 Phishing URLs randomly***

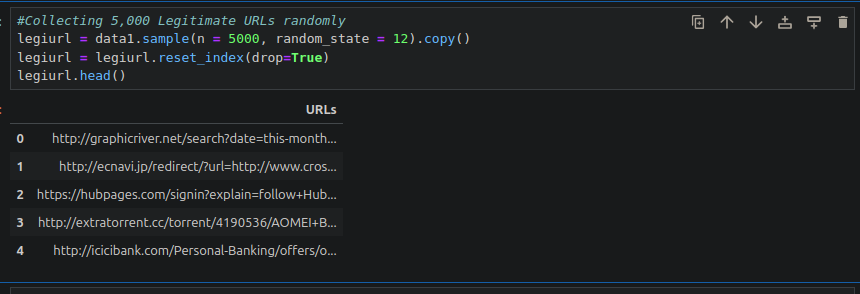


## Legitmate URLs

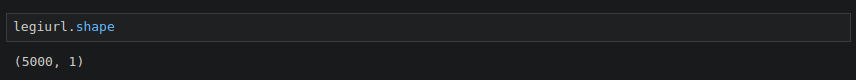
## *3.1 Loading legitmate files*

## 

***3.2 Collecting 5,000 Legitmate URL’s randomly***



***3.3 Shape the URL’s***



## Feature Extraction:

## In this step, features are extracted from the URLs dataset.

## The extracted features are categorized into

## Address Bar based Features

## Domain based Features

## HTML & Javascript based Features

## *4.1 Address Bar Based Features:*

## Many features can be extracted that can be consided as address bar base features. Out of them, below mentioned were considered for this project.

## Domain of URL

## IP Address in URL

## "@" Symbol in URL

## Length of URL

## Depth of URL

## Redirection "//" in URL

## "http/https" in Domain name

## Using URL Shortening Services “TinyURL”

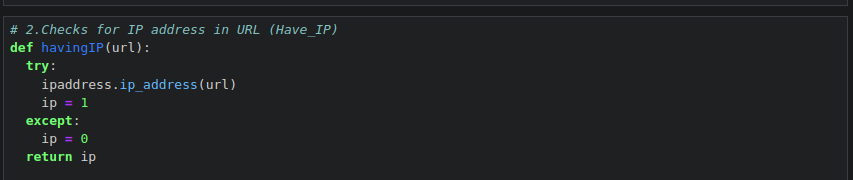
## Prefix or Suffix "-" in Domain

## Each of these features are explained and coded below:

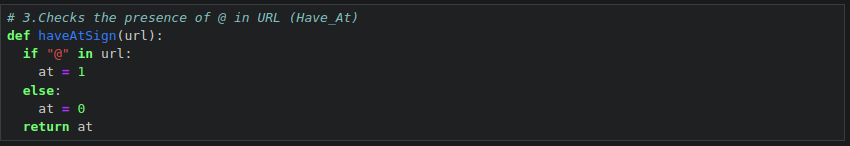
## *4.1.1 Domain of the URL (Domain):*

## 

***4.1.2 Checks for IP address in URL (Have IP):***



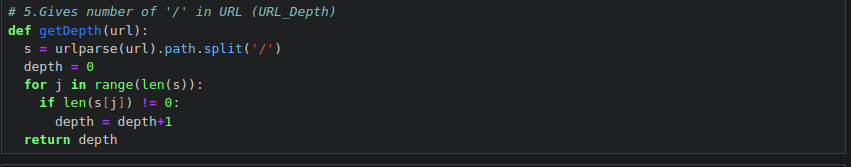
***4.1.3 Checks the presence of @ in URL (Have @):***



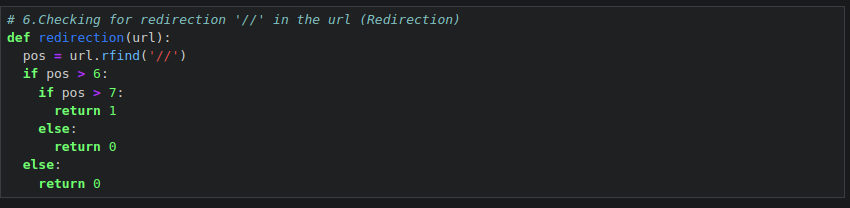
***4.1.4 Finding the length of URL and categorizing (URL Length):***



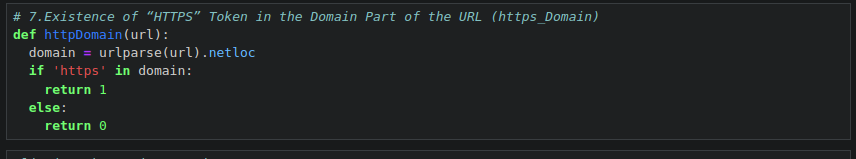
***4.1.4 Gives number of ’/’ in URL (URL Depth):***



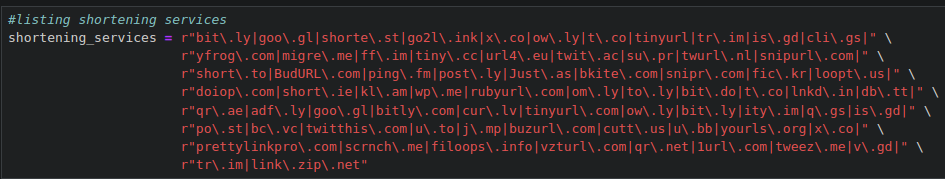
***4.1.5 Checking for redirection ’//’ in the URL(Redirection):***



***4.1.6 Existence of “HTTPS” Token in the Domain Part of the URL:***



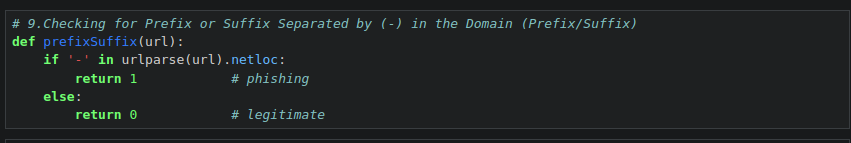
***4.1.7 Listing Shortening Services in URL (Tiny URL):***



***4.1.8 Checking for the shortening services in URL (Tiny URL):***



***4.1.9 Checking for Prefix or Suffix seperated by (-) in the Domain:***

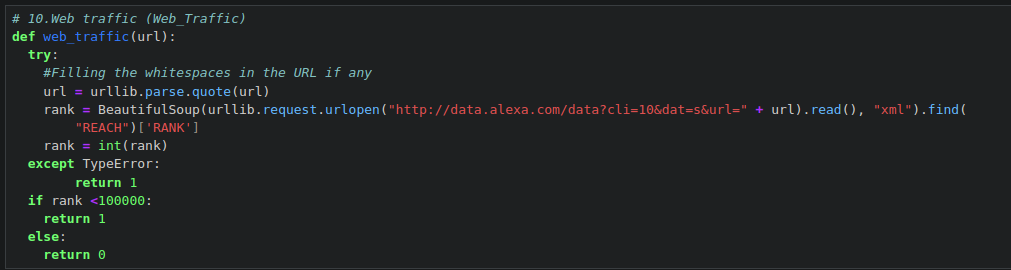


***4.2 Domain Based Features:***

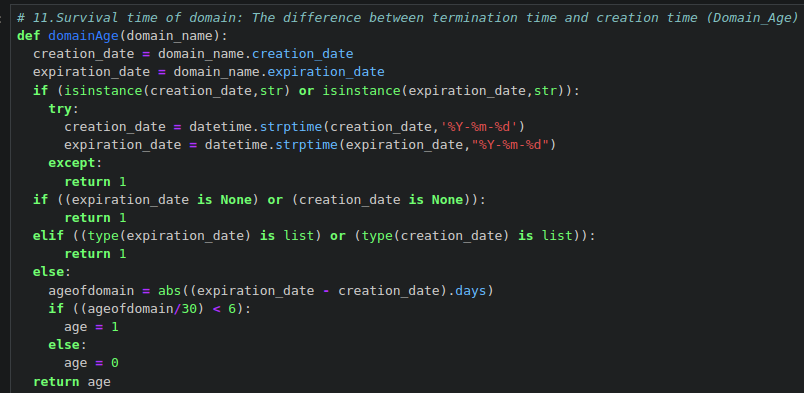
Many features can be extracted that come under this category. Out of them, below mentioned were considered for this project.

* Website Traffic
* Age of Domain
* End Period of Domain

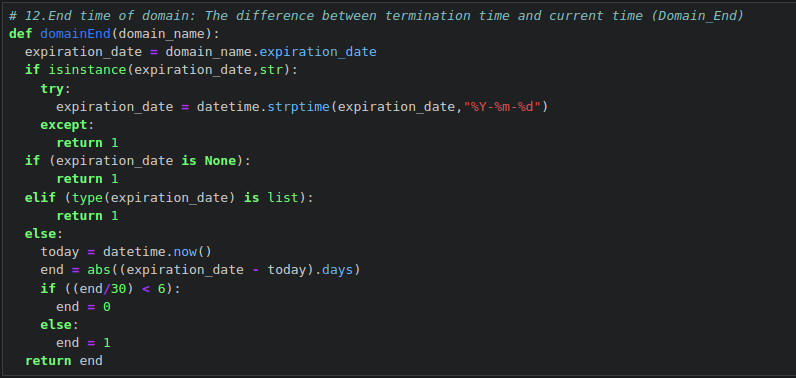
***4.2.1 Website Traffic (web traffic):***



***4.2.2 Difference between termination time and creation time (Domain Age)***



***4.2.3 Difference between termination time and current time (Domain End)***



***4.3 HTML and Javascript based Features***

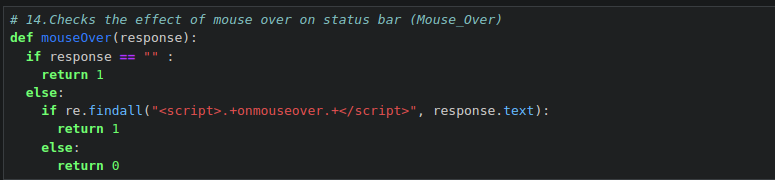
Many features can be extracted that come under this category. Out of them, below mentioned were considered for this project.

* IFrame Redirection
* Status Bar Customization
* Disabling Right Click
* Website Forwarding
* Each of these features are explained and the coded below:

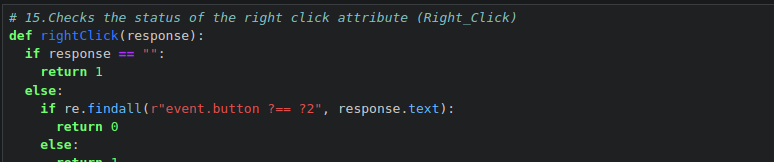
***4.3.1 Iframe Redirection(iFrame):***



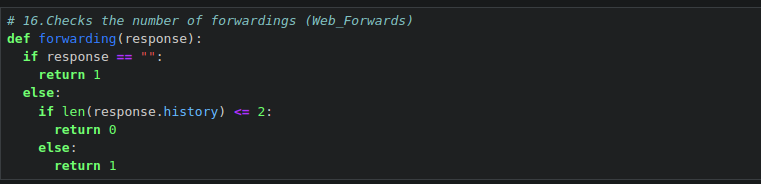
***4.3.2 Checks the effect of mouse over on status bar (Mouse Over):***



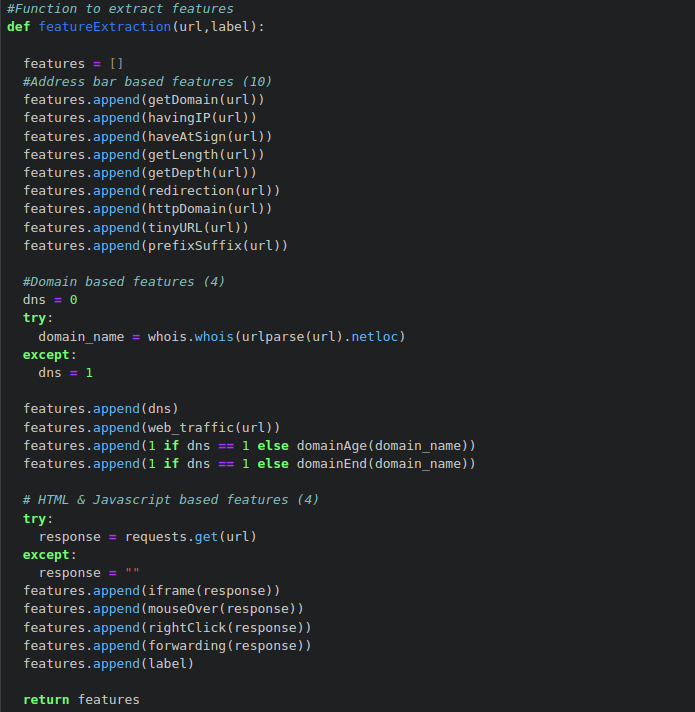
***4.3.3 Checks the status of the right click attribute (Right Click):***



***4.3.4 Checks the number of forwardings (Web Forwards):***



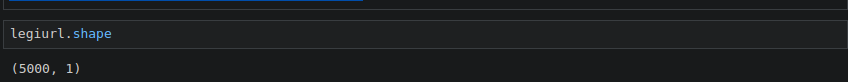
***4.3.5 Function to extract features:***



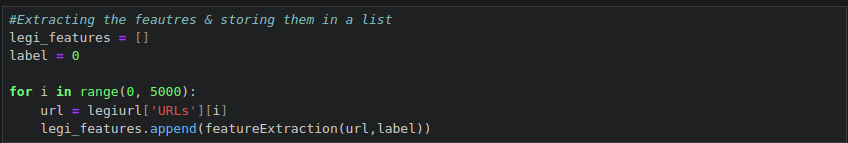
**5. Legitmate URL’s:**

Now, feature extraction will be done on legitimate URLs.

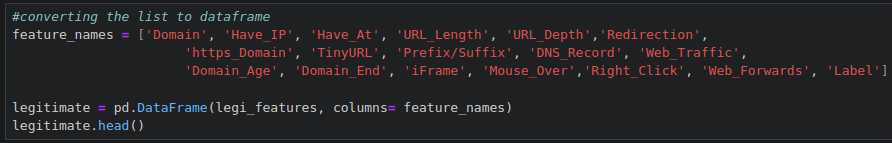
***5.1.1 Shaping URL’s:***



***5.1.2 Extracting the features and storing them in a List:***



***5.1.3 Converting the List to Dataframe:***



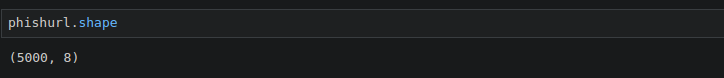
***5.1.4 Storing the extracted legitmate URL’s features to csv file:***



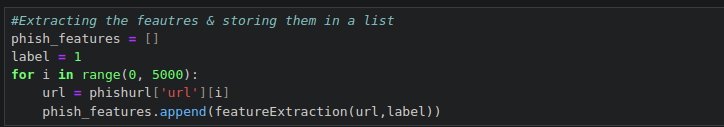
**6 Phishing URL’s:**

Now, feature extraction is performed on phishing URLs.

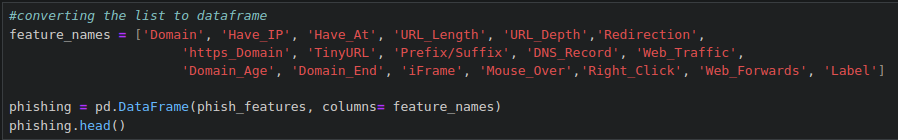
***6.1 Shaping Phishing URL’s:***



***6.2 Extracting the features & storing them in a List:***



***6.3 Converting the List to Data Frame:***



***6.4 Storing the Extracted legitmate URL’ s features to CSV file:***



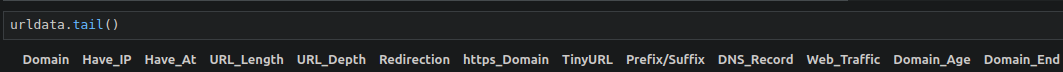
**7. Final Dataset:**

In the above section we formed two dataframes of legitimate & phishing URL features. Now, we will combine them to a single dataframe and export the data to csv file for the Machine Learning training.

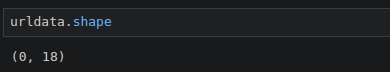
***7.1.1 Concatenating the dataframes into one:***



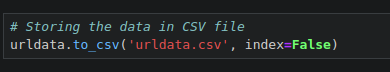
***7.1.2 Getting the tale:***



***7.1.3 Shaping the Data:***



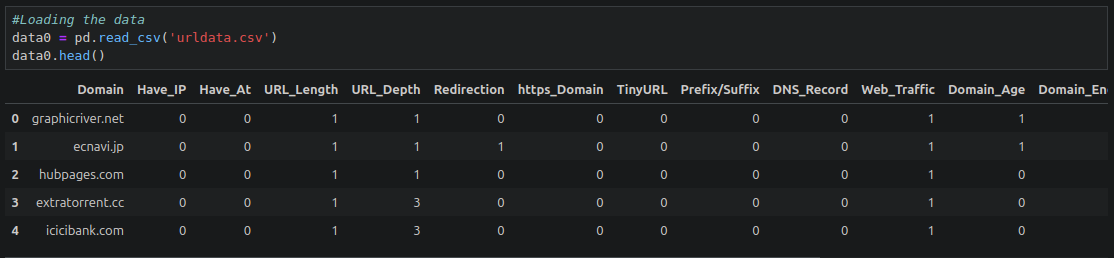
***7.1.4 Storing the Data in CSV file:***



**8. Data Cleaning and Splitting**

We finally extracted 16 features for 10,000 URL which has 5000 phishing & 5000 legitimate URLs Both phishing and benign URLs of websites are gathered to form a dataset and from them required URL and website content-based features are extracted. The performance level of each model is measures and compared.

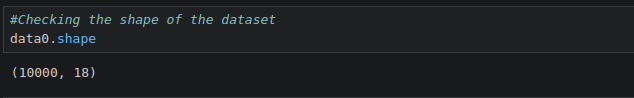
***8.1 Loading the Data***



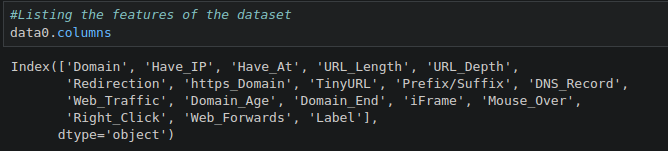
**8.2 Familiarizing with Data**

In this step, few dataframe methods are used to look into the data and its features.

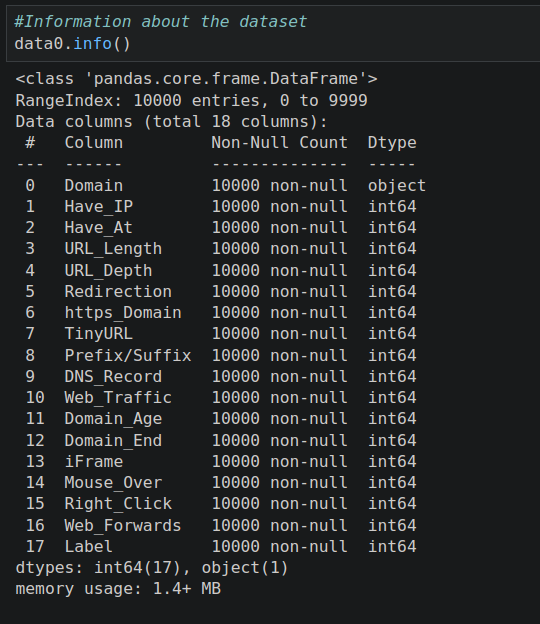
***8.2.1 Checking the shape of the Dataset***



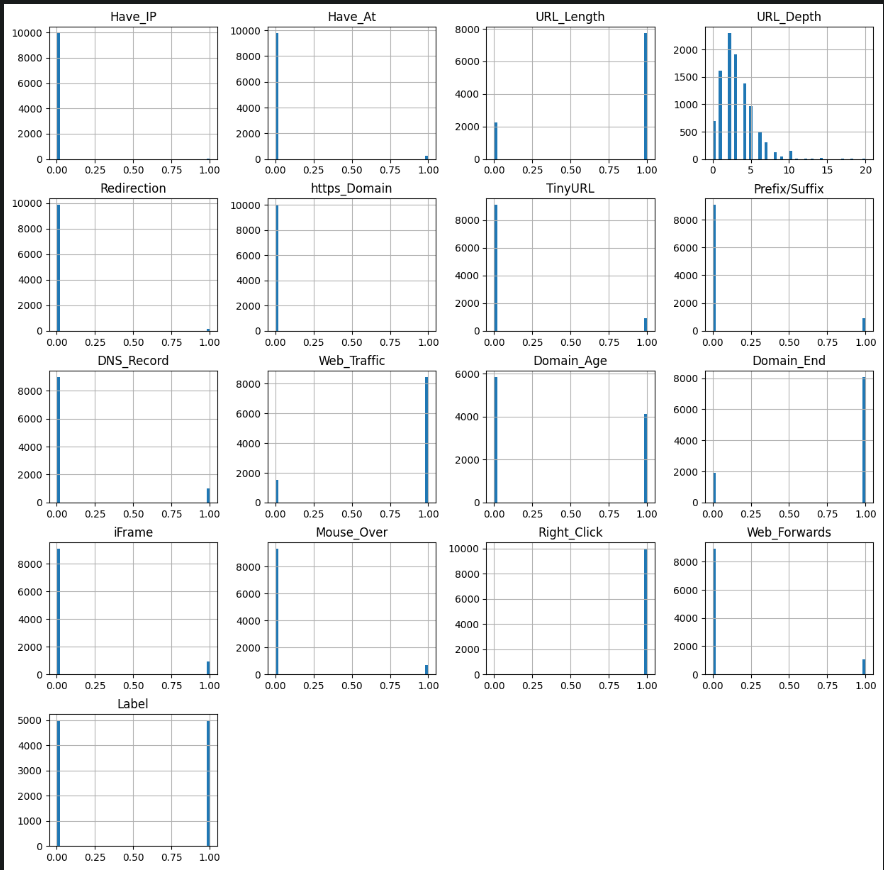
***8.2.2 Listing the features of the Dataset***



***8.2.3 Information about the Dataset:***



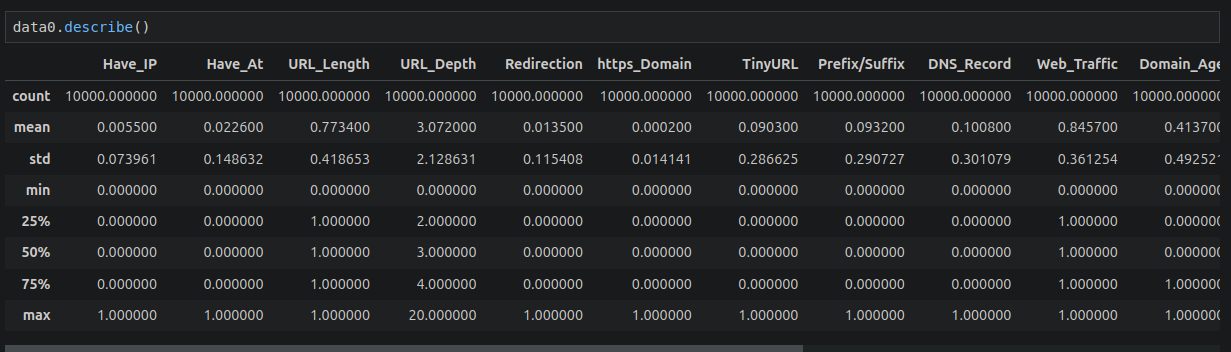
***8.3.1 Visualizing the Data:***



***8.4 Data preprocessing***

Here, we clean the data by applying data preprocesssing techniques and transform the data to use it in the models.

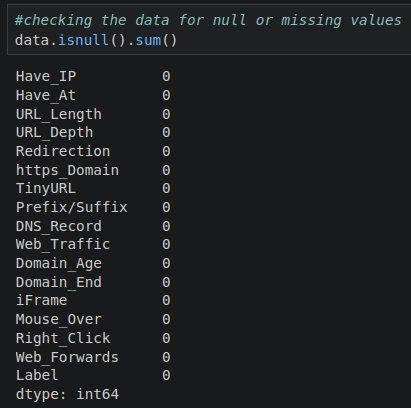
***8.4.1 Describing the Data:***



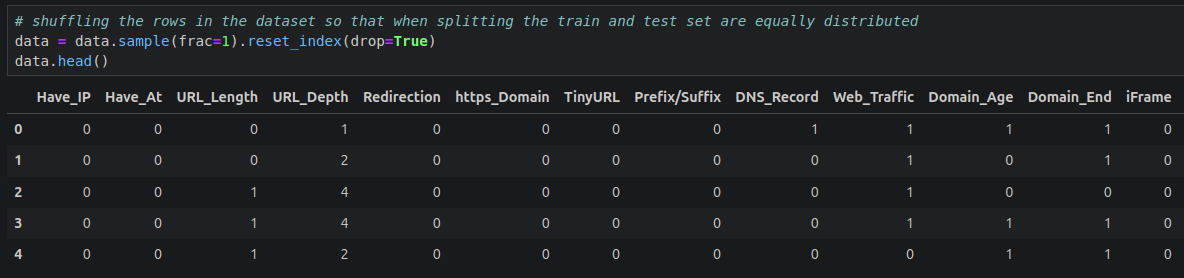
***8.4.2 Dropping the Domain Column:***



***8.4.3 Checking the Data for Null and missing values:***



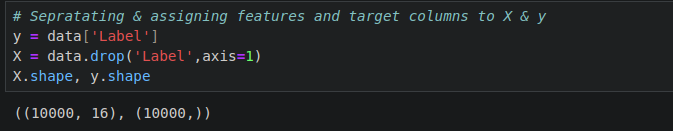
***8.4.4 Shuffling the rows:***



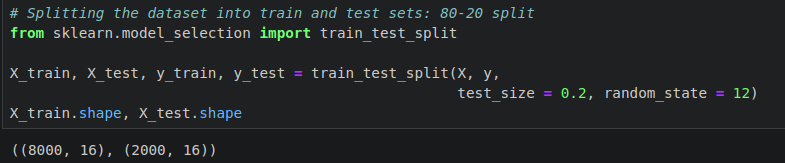
***8.6 Splitting the Data:***

Here we will Split our data into train and test.

***8.6.1 Separating & assigning features and target columns:***



***8.6.2 Splitting the Dataset into train and test: 80-20 split:***



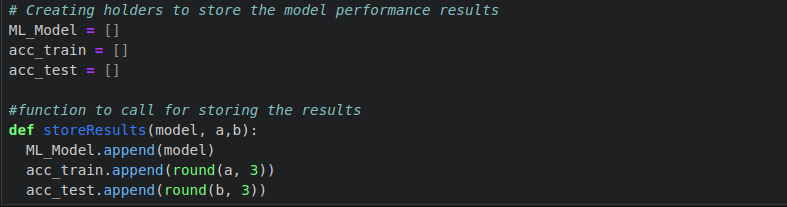
**9. Comparision and Training of Dataset using Different Supervised Machine Learning Algorithmn.**

From the dataset above, it is clear that this is a supervised machine learning task. There are two major types of supervised machine learning problems, called classification and regression.

This data set comes under classification problem, as the input URL is classified as phishing or legitimate. The supervised machine learning models (classification) considered to train the dataset in this notebook are:

* Decision Tree
* Random Forest
* Multilayer Perceptrons
* XGBoost
* Autoencoder Neural Network
* Support Vector Machines

***9.1 Creating Holders to store the model performing results:***

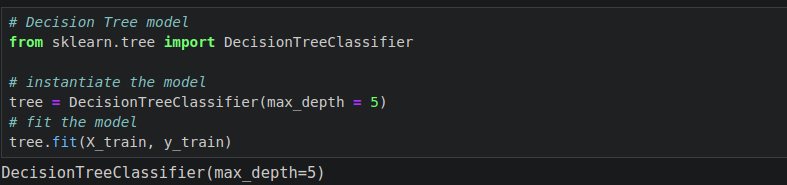


***9.2 Decision Tree Classifier:***

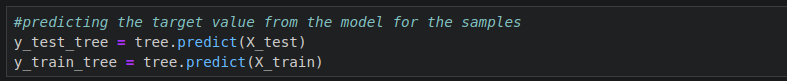
Decision trees are widely used models for classification and regression tasks. Essentially, they learn a hierarchy of if/else questions, leading to a decision. Learning a decision tree means learning the sequence of if/else questions that gets us to the true answer most quickly.

In the machine learning setting, these questions are called tests (not to be confused with the test set, which is the data we use to test to see how generalizable our model is). To build a tree, the algorithm searches over all possible tests and finds the one that is most informative about the target variable.

***9.2.1 Decision tree model:***

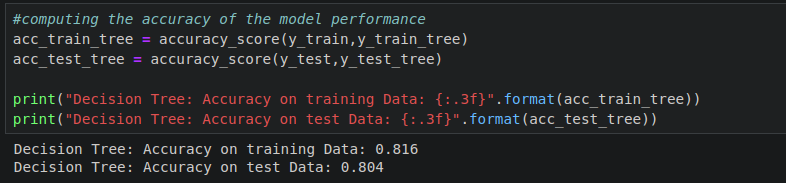


***9.2.2 Predicting the target value:***

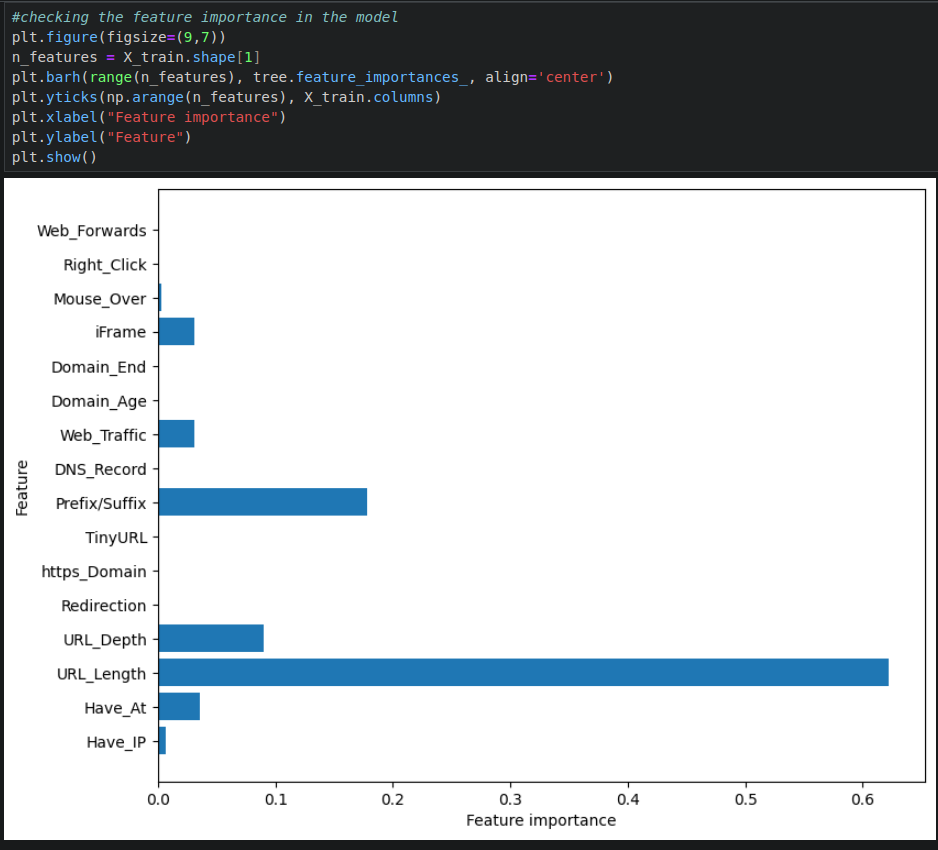


***Performance Evalution***

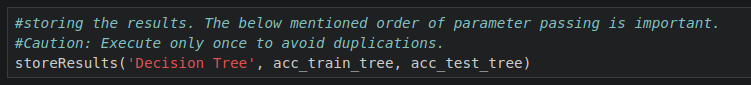
***9.2.3 Computing the Accuracy of the Model:***



***9.2.4 Checking the feature importance in the model:***



***9.2.5 Storing the results:***

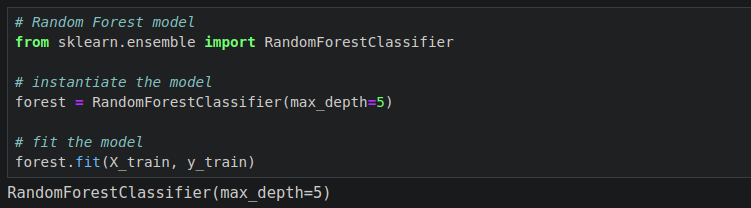


***9.3 Random Forest Classifier***

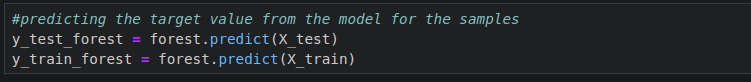
Random forests for regression and classification are currently among the most widely used machine learning methods.A random forest is essentially a collection of decision trees, where each tree is slightly different from the others. The idea behind random forests is that each tree might do a relatively good job of predicting, but will likely overfit on part of the data.

If we build many trees, all of which work well and overfit in different ways, we can reduce the amount of overfitting by averaging their results. To build a random forest model, you need to decide on the number of trees to build (the n\_estimators parameter of RandomForestRegressor or RandomForestClassifier). They are very powerful, often work well without heavy tuning of the parameters, and don’t require scaling of the data.

***9.3.1 Instantiate the model:***

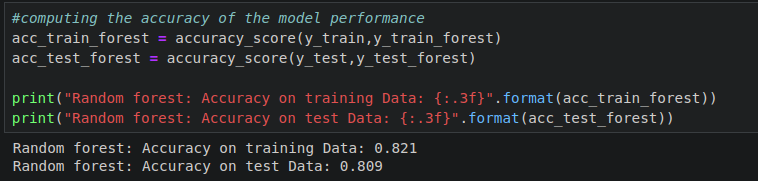


***9.3.2 Predicting the Target:***

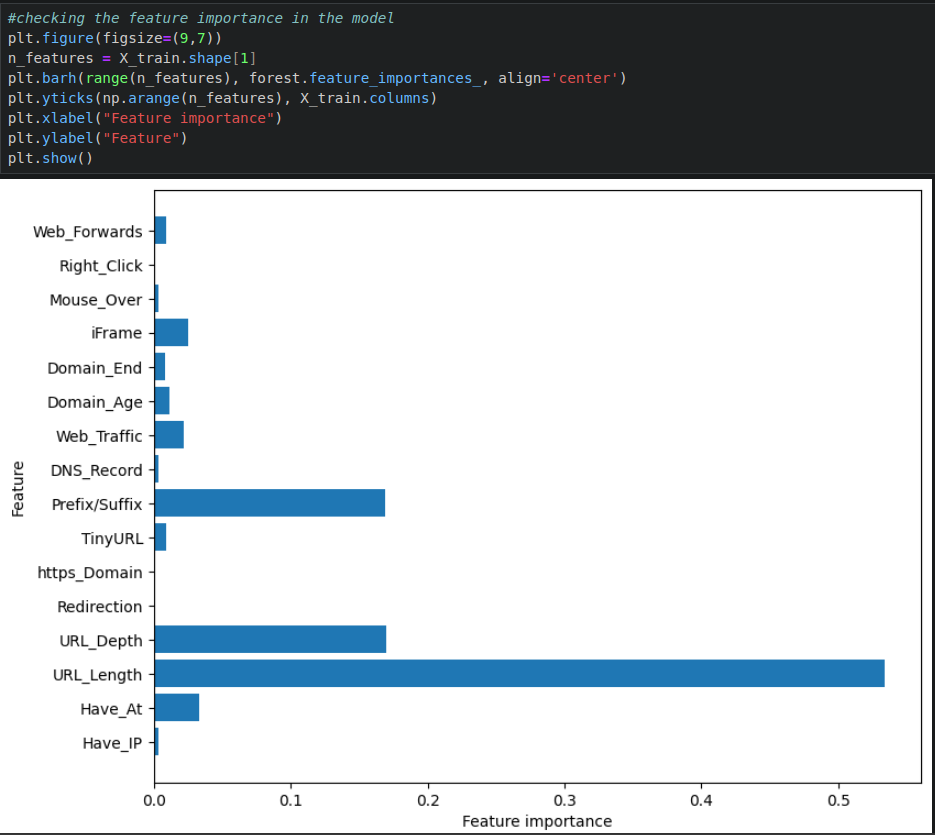


***Performance Evalution***

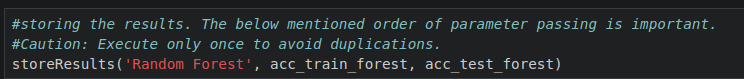
***9.3.3 Computing the Accuracy of the Model:***



***9.3.4 Checking the feature importance of the Model:***



***9.3.5 Storing the results:***

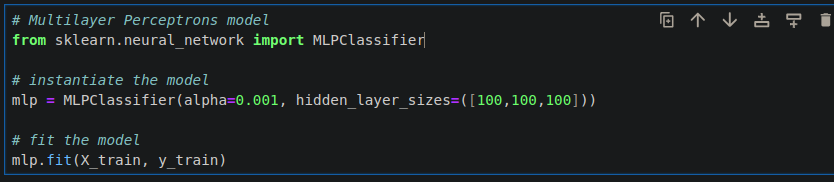


***9.4 MultiLayer Perceptron: Deep learning***

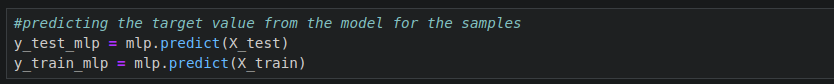
Multilayer perceptrons (MLPs) are also known as (vanilla) feed-forward neural networks, or sometimes just neural networks. Multilayer perceptrons can be applied for both classification and regression problems.

MLPs can be viewed as generalizations of linear models that perform multiple stages of processing to come to a decision.

***9.4.1 Instantiate the Model:***

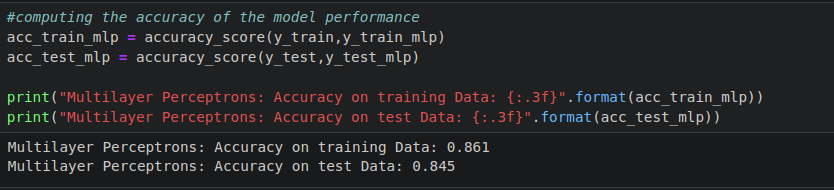


***9.4.2 Predicting the target value:***

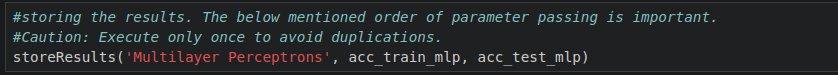


***Performance Evalution***

***9.4.3 Computing the Accuraacy***



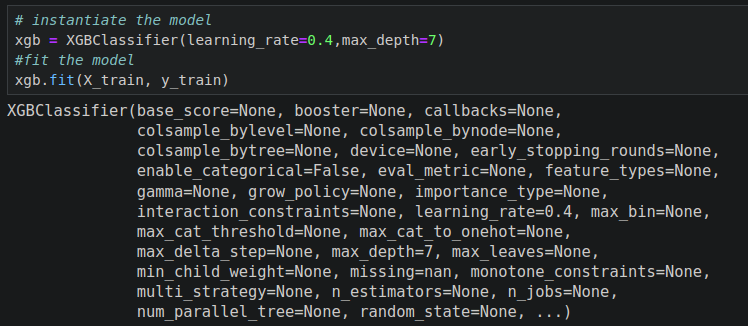
***9.4.4 Storing the results***



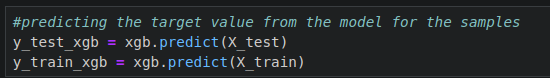
***9.5 XGBoost Classifier***

XGBoost is one of the most popular machine learning algorithms these days. XGBoost stands for eXtreme Gradient Boosting. Regardless of the type of prediction task at hand; regression or classification. XGBoost is an implementation of gradient boosted decision trees designed for speed and performance.

***9.5.1 Instantiate the model***

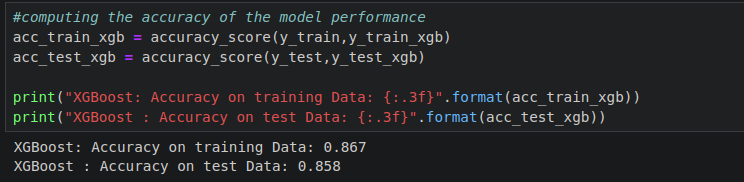


***9.5.2 Predicting the target value***

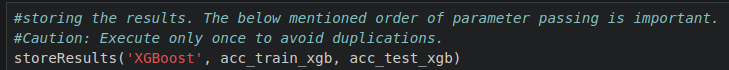


***Performance Evalution***

***9.5.3 Computing the Accuracy of the model***



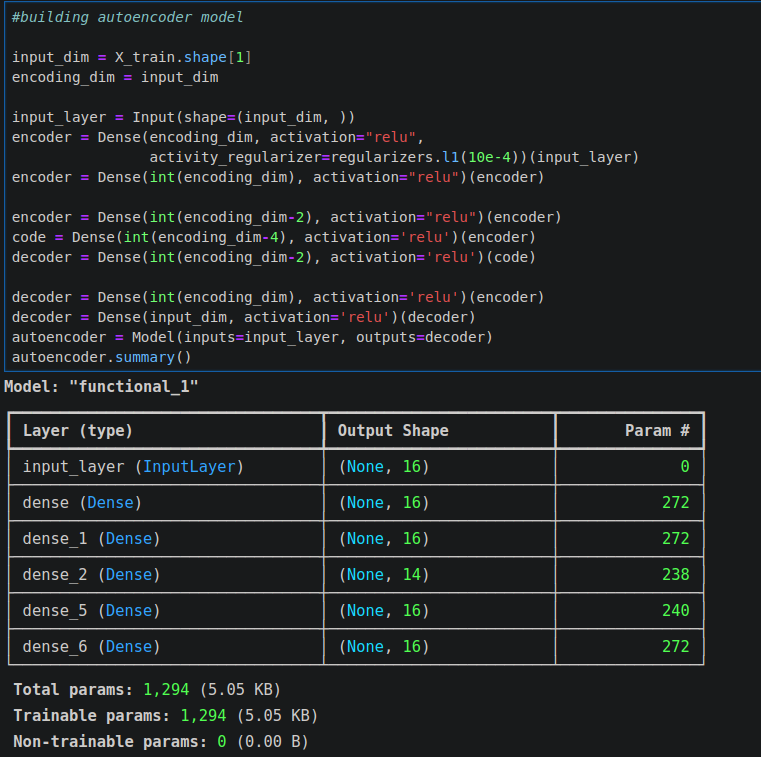
***9.5.4 Storing results***



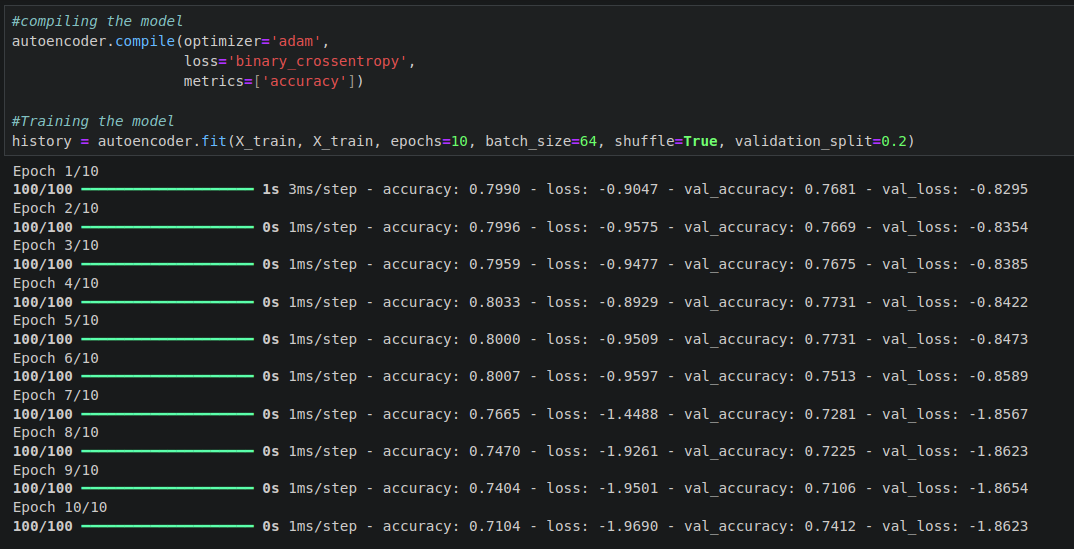
**9.6 Autoencoder Neural Network**

An auto encoder is a neural network that has the same number of input neurons as it does outputs. The hidden layers of the neural network will have fewer neurons than the input/output neurons. Because there are fewer neurons, the auto-encoder must learn to encode the input to the fewer hidden neurons. The predictors (x) and output (y) are exactly the same in an auto encoder.

***9.6.1 Building Autoencoder model***

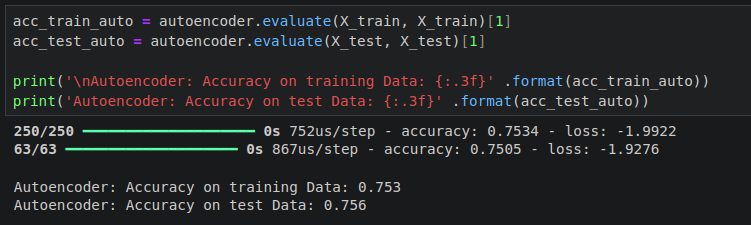


***9.6.2 Compiling the Model***

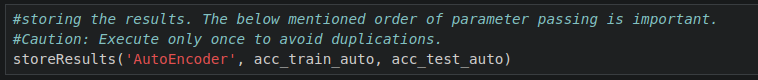


***Performance Evalution***

***9.6.3 Evaluting the Model***



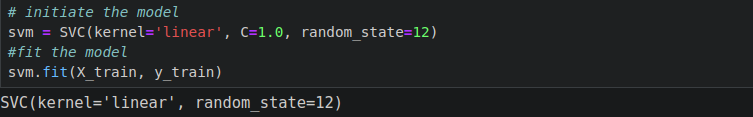
***9.6.4 Storing the results***



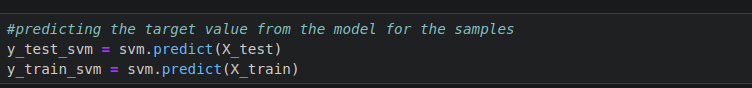
**9.7 Support Vector Machine**

In machine learning, support-vector machines (SVMs, also support-vector networks) are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier.

***9.7.1 Initiate the model:***

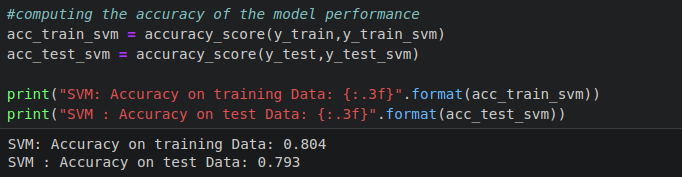


***9.7.2 Predicting the Target Value:***

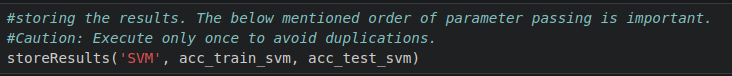


***Performance Evalution***

***9.7.3 Computing the Accuracy***



***9.7.4 Storing the results***

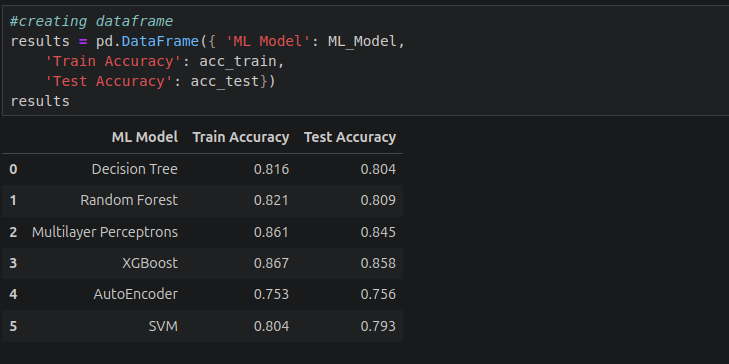


# CHAPTER 3

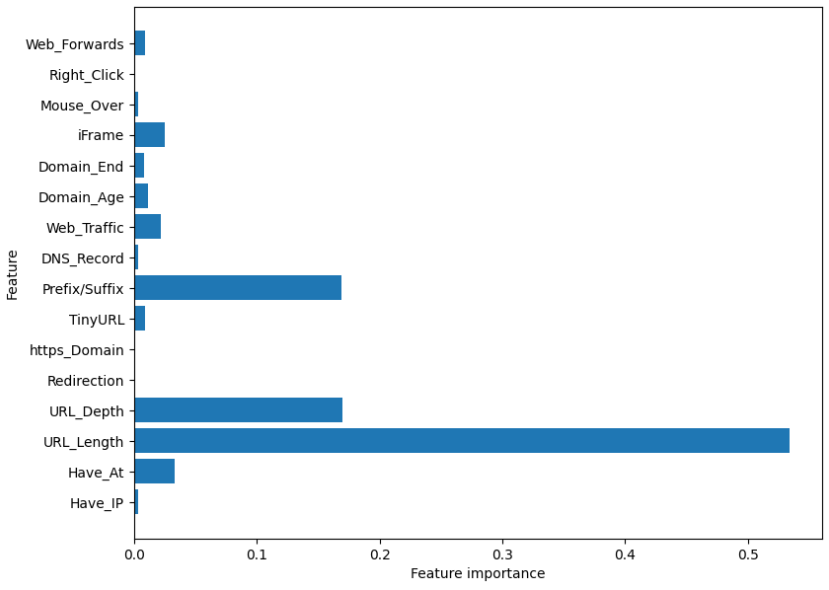
# RESULT & ANALYSIS

So, here as a result we compare the models and found out that ***XGBoost performs the best out of all the models*** and also the main area of interest of not been phished is to keep check on the ***URL of the Website.***

***Comparison of the Model***



***Most important feature to look for***



# CHAPTER 4

# CONCLUSION

This paper explored the potential of various type of Machine Learning and project it as a viable solution for automated phishing Detection. By leveraging the Machine learning Algorithms i developed a model capable of i***dentifying malicious websites with high accuracy***. This research focused on analyzing features extracted from website content and URl based patterns in HTML code. These features were to utilized to train and evalute the performanec of various Supervised as well as Unsupervised Machine Learning Algorithms, Include ***Support vector Machine, Random Forest, Decision Tree*** and others.

The Evalution process highlighted the effectiveness of these Algorithms in distinguishing between legitmate and phishing websites. The Models achieved promising accuracy rates and it also demonstrates thier potential for real world application. However, it is crucial to acknowledge the ongoing evolution of phishing tactics continous research is still very important to tackle the everchanging strategies.

Building upon the sucess of this research, this research as a practical solution for internet users. The ***development of Browser Extension can leverage this research,*** such a browser extension can empower the users with an additional layer of security fostering a more vigilant and secure online enviroment. The proposed ***browser extension projects a practical application of this research***, offers a ***user-friendly and adaptable solution for combating cybercrime.*** By harnessing the power of machine learning we can work towards a safer and more trustworthy digital landscape.

# Chapter 5

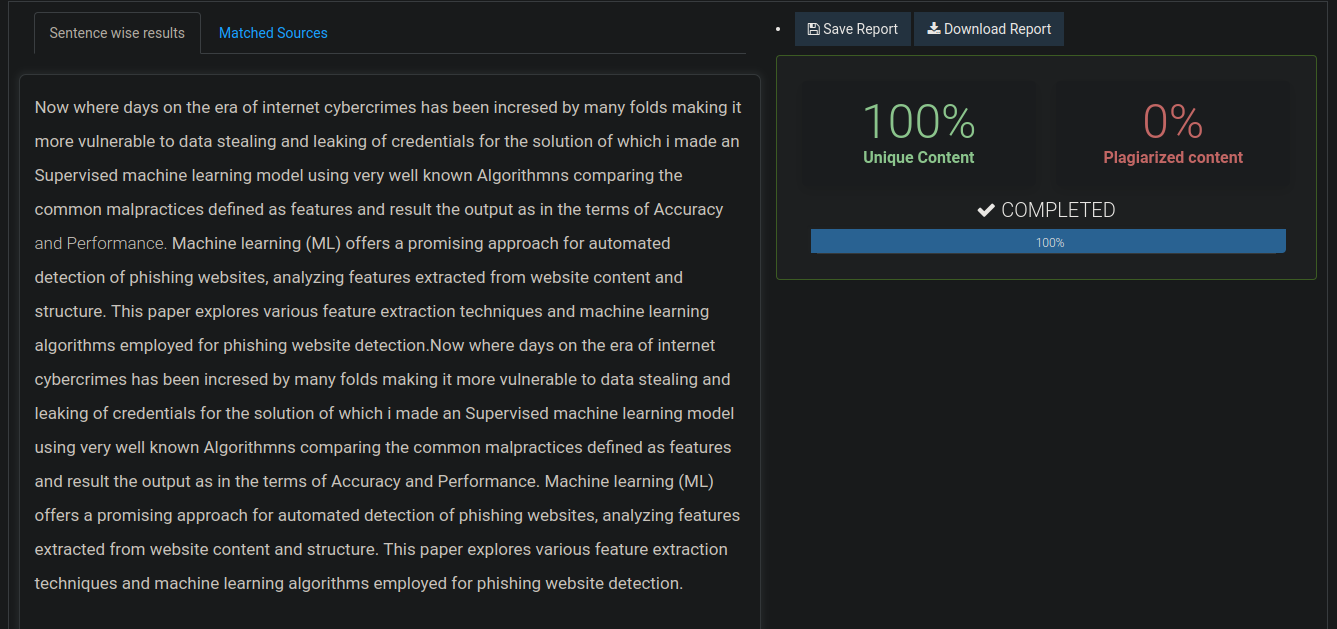
# References

1. Jansson, K.; von Solms, R. (2011-11-09). "Phishing for phishing awareness". Behaviour & Information Technology. 32 (6): 584–593. doi:10.1080/0144929X.2011.632650. ISSN 0144-929X. S2CID 5472217.
2. Ramzan, Zulfikar (2010). "Phishing attacks and countermeasures". In Stamp, Mark; Stavroulakis, Peter (eds.). Handbook of Information and Communication Security. Springer. ISBN 978-3-642-04117-4.
3. "Internet Crime Report 2020" (PDF). FBI Internet Crime Complaint Center. U.S. Federal Bureau of Investigation. Retrieved 21 March 2021.
4. Ollmann, Gunter. "The Phishing Guide: Understanding and Preventing Phishing Attacks". Technical Info. Archived from the original on 2012-06-29. Retrieved 2006-07-10.
5. Wright, A; Aaron, S; Bates, DW (October 2016). "The Big Phish: Cyberattacks Against U.S. Healthcare Systems". Journal of General Internal Medicine. 31 (10): 1115–8. doi:10.1007/s11606-016-3741-z. PMC 5023604. PMID 27177913.
6. Stonebraker, Steve (January 2022). "AOL Underground". aolunderground.com (Podcast). Anchor.fm.
7. Mitchell, Anthony (July 12, 2005). "A Leet Primer". TechNewsWorld. Archived from the original on April 17, 2019. Retrieved 2021-03-21.
8. "Phishing". Language Log, September 22, 2004. Archived from the original on 2006-08-30. Retrieved 2021-03-21.
9. Jøsang, Audun; et al. (2007). "Security Usability Principles for Vulnerability Analysis and Risk Assessment". Proceedings of the Annual Computer Security Applications Conference 2007 (ACSAC'07). Archived from the original on 2021-03-21. Retrieved 2020-11-11.

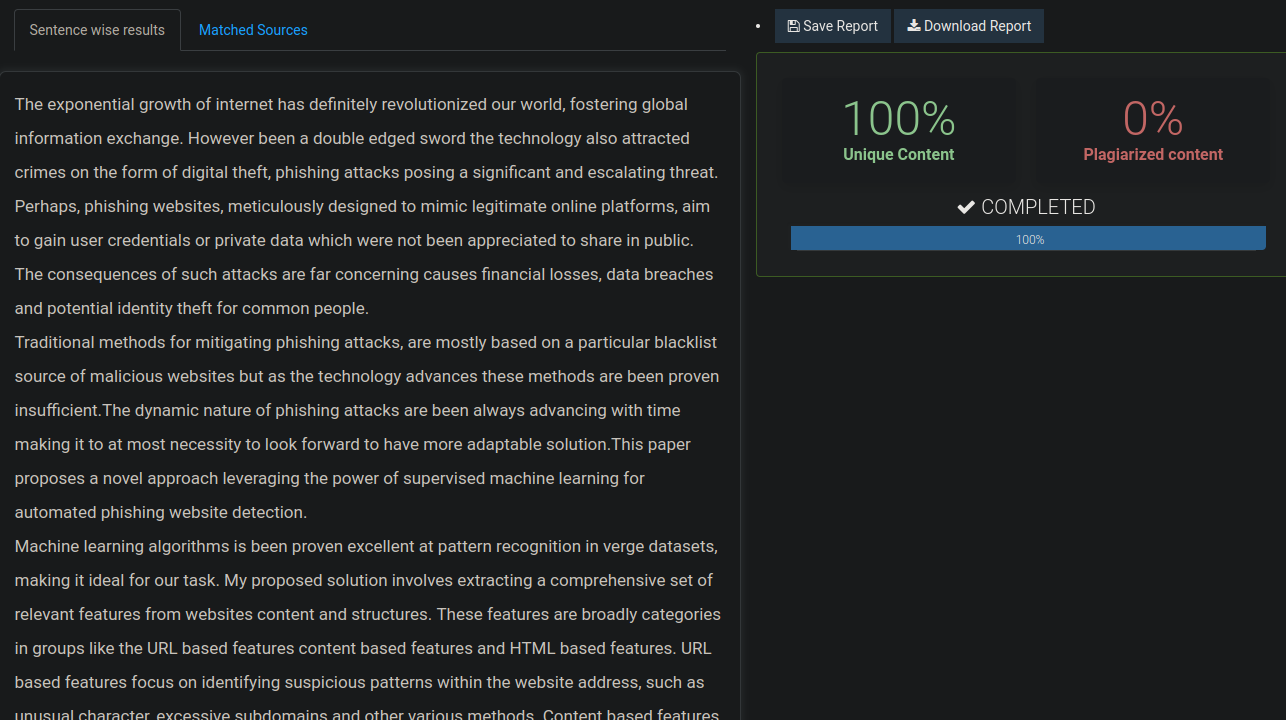
# APPENDIX A PLAGIARISM REPORT

| **S No.** | **Topic** | **Plagiarism ( in %)** |
| --- | --- | --- |
| 1. | Abstract | **0** |
| 2. | Introduction | **0** |
| 3. | Methodologies | **0** |

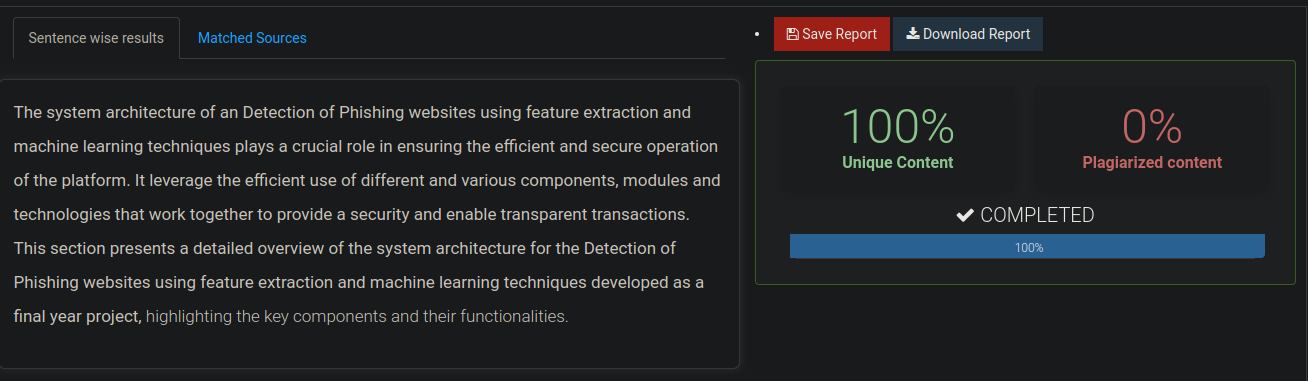
**Table A.1: Plagiarism Report**



**Figure A.1: Plagiarism - Abstract**



**Figure A.2: Plagiarism – Introduction**



**Figure A.3: Plagiarism – Methodologies**