WWW.IJRPR.COM

**International Journal of Research Publication and Reviews**

Journal homepage: www.ijrpr.com ISSN 2582-7421

# PHISHING WEBSITE DETECTION USING MACHINE LEARNING TECHNIQUES

***JAYAKUMAR R***

B.Sc., Department of Computer Science, Sri Krishna Adithya College of Arts and Science Coimbatore, Tamil Nadu, India

ABSTRACT :

Phishing URL detection is a web application that allows user to detecting phishing URL information for the company with efficient manner. Computer technology is widely been implemented. Hence the inception of computers had a great role in reducing large tasks to simpler one. So, that the proposed project is developed to automate the complete operations of the phishing URL detection. Initially admin has to login this application using their username and password. After successful login admin can add company URL information and company information in this application. This information will be maintained by a separate table. This project the user has to register to the website. After the registration user will get login id and password. The login id and password help to user login the website. After successful login user can upload URL information in this application. After that this application detect phishing URL information and display company information effectively. This system has been developed with an intention to make the system user-friendly thus reducing the manual work. The system has been developed with advanced features. The objective of our project is to establish computerization and maintain all the details throughout this website with efficient manner. Admin can also view all these information and he can give update in this website.

Keywords :Phishing attack, Machine learning

## INTRODUCTION :

Nowadays Phishing becomes a main area of concern for security researchers because it is not difficult to create the fake website which looks so close to legitimate website. Experts can identify fake websites but not all the users can identify the fake website and such users become the victim of phishing attack. The rise of bogus URLs (Uniform Resource Locators) poses a serious danger to online security, privacy, and trust. Phishing attempts, malware distribution, and other types of cybercrime sometimes use phishing URLs to trick people into providing sensitive information or downloading hazardous software. Detecting and neutralizing false URLs is therefore critical to protect users and against cyber-attacks. The fraudulent URL Detection Project attempts to create an intelligent system that can recognize and flag fraudulent URLs in real time. Using machine learning algorithms, natural language processing techniques, and data analytics, the project aims to improve the accuracy and effectiveness of URL detection methods, hence increasing cybersecurity posture and lowering the dangers associated with bogus URLs. Phishing has become the most serious problem, harming individuals, corporations, and even entire countries. The availability of multiple services such as online banking, entertainment, education, software downloading, and social networking has accelerated the Web's evolution in recent years. As a result, a massive amount of data is constantly downloaded and transferred to the Internet. Spoofed emails pretending to be from reputable businesses and agencies are used in social engineering techniques to direct consumers to phishing websites that deceive users into giving financial information such as usernames and passwords. Technical tricks involve the installation of malicious software on computers to steal credentials directly, with systems frequently used to intercept users online account usernames and passwords.

Machine learning technology consists of a many algorithms which requires past data to make a decision or prediction on future data. Using this technique, algorithm will analyze various blacklisted and legitimate URLs and their features to accurately detect the phishing websites including zero-hour phishing websites.

## SYSTEM SYSTEM

A System Requirements Specification (SRS) (also known as a Software Requirements Specification) is a document or set of documentation that describes the features and behaviour of a system or software application

### *2.1 FRONT-END*

ASP.NET:

ASP.NET is a key part of the wider Microsoft .Net initiative, Microsoft's new application development platform. The .Net by Microsoft Company is to overcome the difficulties in the ASP. Microsoft ensured the asp scripts without modification on the machine with the .Net Framework. The project contains Dynamic web pages. .Net is platform independent and is usedto develop dynamic web pages.

**NET Framework:**

The Microsoft .Net Framework is a new platform for building integrated, service- oriented applications to meet the needs of today's Internet businesses. Applications that gather information from, and interact with, a wide variety of sources, regardless of the platforms or languages in use. Microsoft Intermediate Language and JIT compiler, which make this reuse possible, are described as well as managed components, assemblies, and the Common Type System (CTS). At the heart of the .NET platform is a common language runtime engine and a base framework. The .NET base framework will allow developers to access the features of thecommon language runtime and also will offer many high-level services so that developers don't have to code the same services repeatedly. But more importantly, the .NET common languageruntime engine sitting under this library will provide the technologies to support rapid softwaredevelopment.

Visual Basic .NET (VB.NET) is an object-oriented computer programming language implemented on the .NET Framework. Everything in VB.NET is an object, including all of the primitive types (Short, Integer, Long, String, Boolean, etc.) and user-defined types, events, and even assemblies. All objects inherit from the base class Object. It's also possible to run VB.NETprograms on Mono, the open-source alternative to .NET, not only under Windows, but even Linux or Mac OSX.

### *2.2 BACKEND*

MICROSOFT SQL SERVER

Microsoft SQL Server is a Relational Database Management System (RDBMS) designed to run on platforms ranging from laptops to large multiprocessor servers. SQL Serveris commonly used as the backend system for websites and corporate CRMs and can support thousands of concurrent users. SQL Server comes with a number of tools to help you with yourdatabase administration and programming tasks. SQL Server is much more robust and scalablethan a desktop database management system such as Microsoft Access. Anyone who has ever tried using Access as a backend to a website will probably be familiar with the errors that weregenerated when too many users tried to access the database. Although SQL Server can also berun as a desktop database system, it is most commonly used as a server database system.

## MACHINE LEARNING ALGORITHM

Three machine learning classification model Decision Tree, Random forest and Support vector machine has been selected to detect phishing websites.

### *Decision Tree Algorithm*

One of the most widely used algorithm in machine learning technology. Decision tree algorithm is easy to understand and also easy to implement. Decision tree begins its work by choosing best splitter from the available attributes for classification which is considered as a root of the tree. Algorithm continues to build tree until it finds the leaf node. Decision tree creates training model which is used to predict target value or class in tree representation each internal node of the tree belongs to attribute and each leaf node of the tree belongs to class label. In decision tree algorithm, gini index and information gain methods are used to calculate these nodes.

### *Random Forest Algorithm*

Random forest algorithm is one of the most powerful algorithms in machine learning technology and it is based on concept of decision tree algorithm. Random forest algorithm creates the forest with number of decision trees. High number of tree gives high detection accuracy.

Creation of trees are based on bootstrap method. In bootstrap method features and samples of dataset are randomly selected with replacement to construct single tree. Among randomly selected features, random forest algorithm will choose best splitter for the classification and like decision tree algorithm; Random forest algorithm also uses gini index and information gain methods to find the best splitter. This process will get continue until random forest creates n number of trees.

Each tree in forest predicts the target value and then algorithm will calculate the votes for each predicted target. Finally random forest algorithm considers high voted predicted target as a final prediction.

### *Support Vector Machine Algorithm*

Support vector machine is another powerful algorithm in machine learning technology. In support vector machine algorithm each data item is plotted as a point in n-dimensional space and support vector machine algorithm constructs separating line for classification of two classes, this separating line is well known as hyperplane.

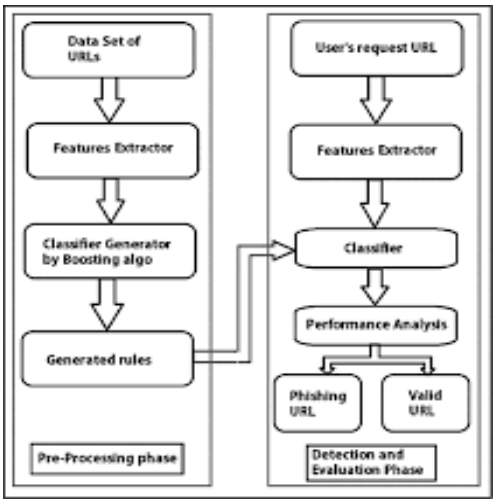
Support vector machine seeks for the closest points called as support vectors and once it finds the closest point it draws a line connecting to them. Support vector machine then construct separating line which bisects and perpendicular to the connecting line. In order to classify data perfectly the margin should be maximum. Here the margin is a distance between hyperplane and support vectors. In real scenario it is not possible to separate

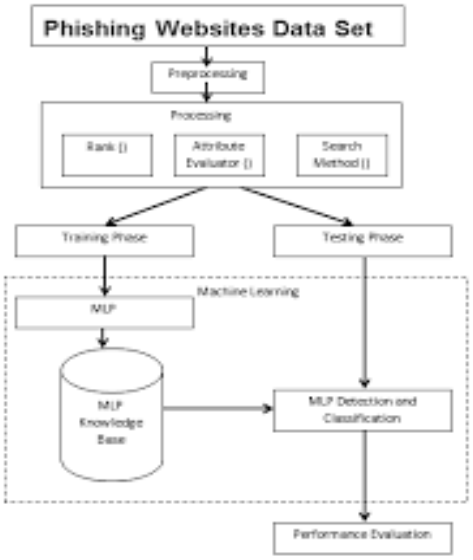
complex and non linear data, to solve this problem support vector machine uses kernel trick which transforms lower dimensional space to higher dimensional space.

## SYSTEM ARCHITECTURE AND IMPLEMENTATION

### *4.1DATA FLOW DIAGRAM*

DFD depict hoe data interact with the system. DFD are extremely useful in modelling many aspects of a business function because they systematically subdivide a task into basic parts, helping the analyst understand the system that they trying to model data flow diagram models a system by using external entities from which data flow to a process which transmission the data and creates output data which goes to other processes on external entities of files. Data may also flow to process as inputs.

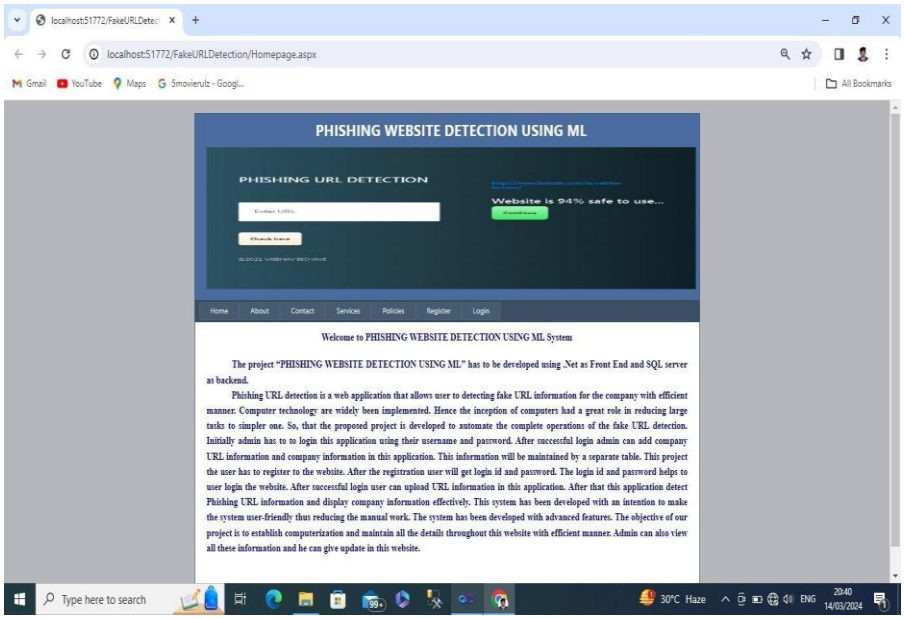


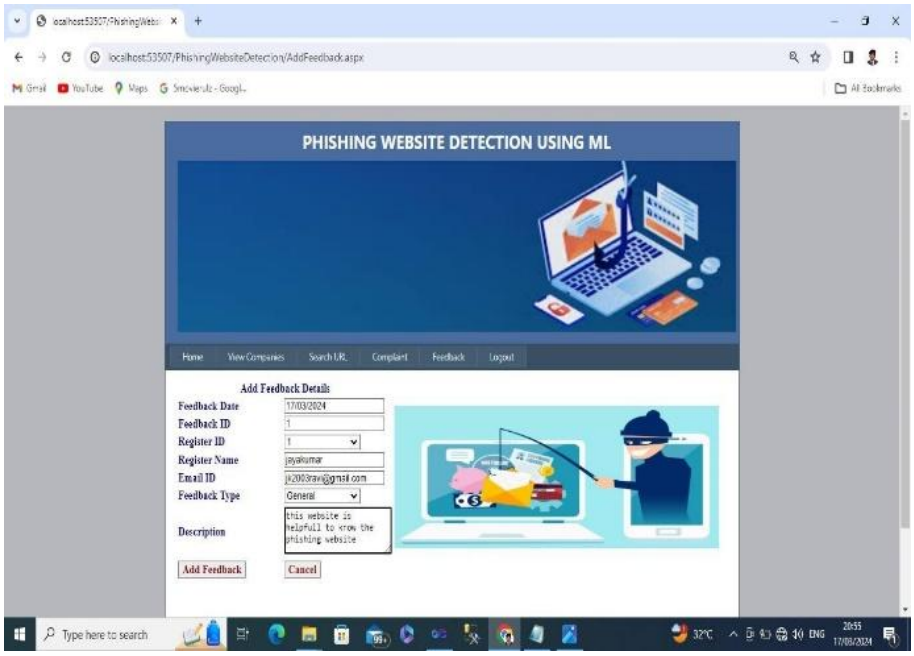



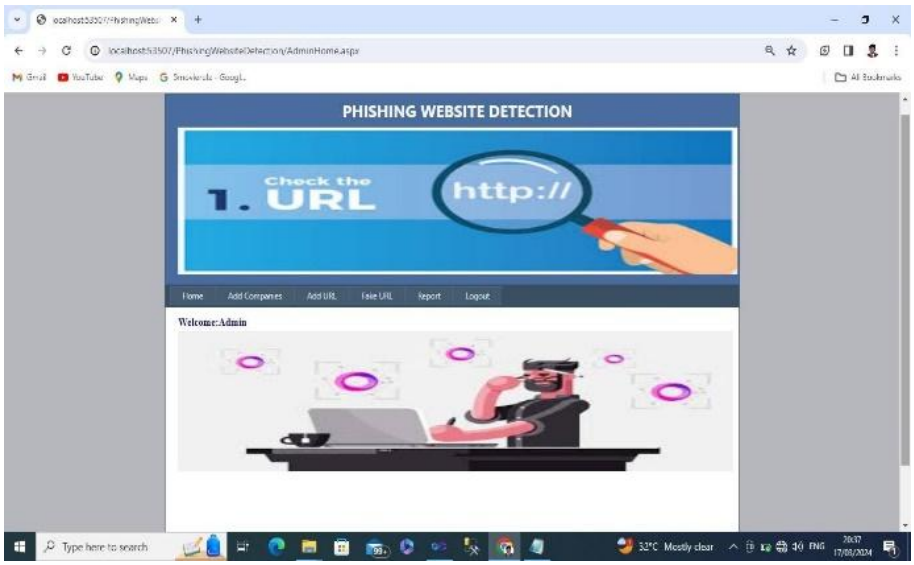

## CONCLUSION

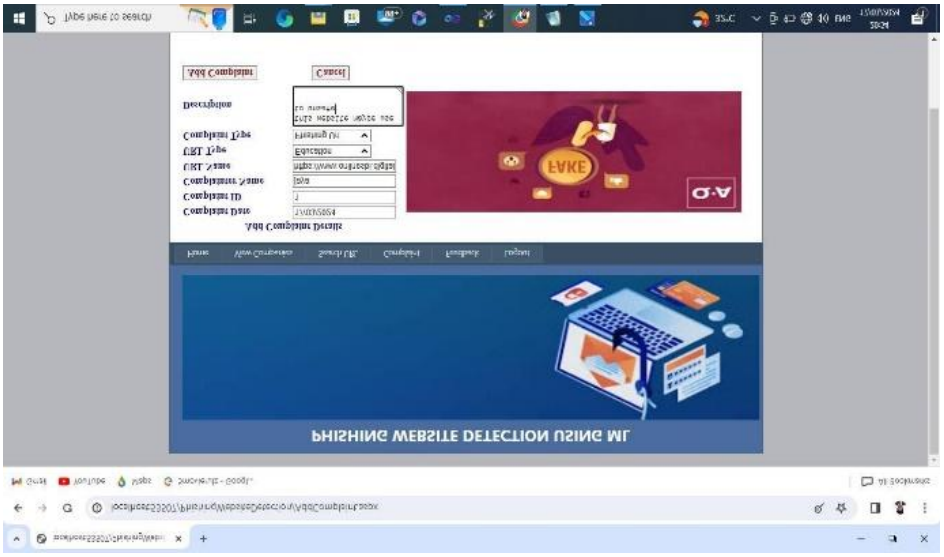
Implementation is the stage of the project when the theoretical design is turned out into a working system. Thus, it can be considered to be the most critical stage in achieving a successful new system and in giving the user, confidence that the new system will work and be effective. Proposed system successfully to automate the complete operations of the fake URL detection. Initially admin has to add company URL information and company information in this application. This project successfully helps user has upload URL information in this application. After that this application detects fake URL information and display company information effectively. This system has been developed with an intention to make the system user-friendly thus reducing the manual work. This proposed mobile application providing a robust, user-friendly solution for the user and company

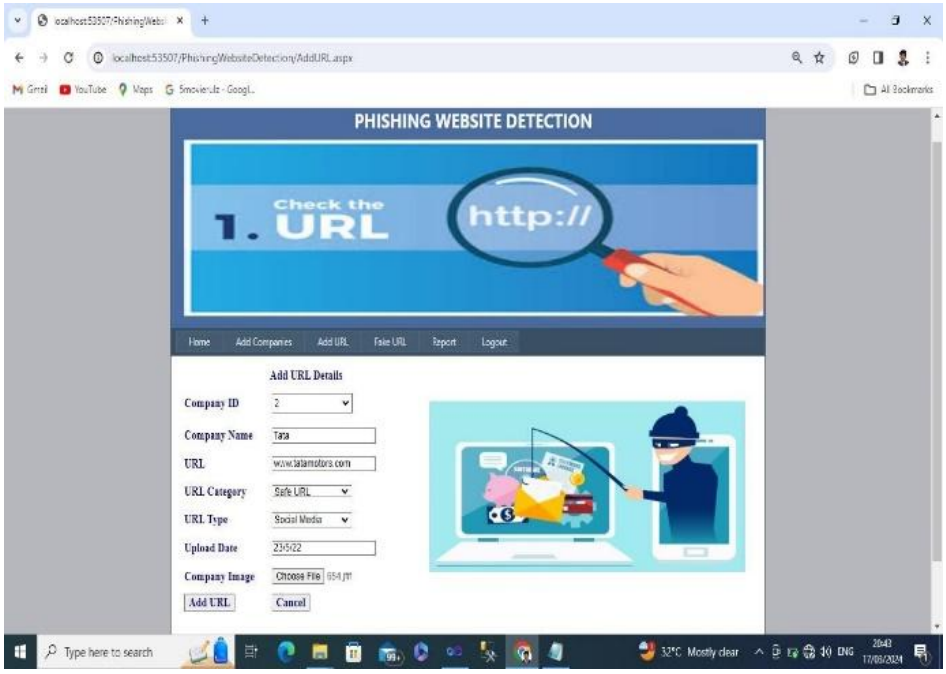
## SCREENSHOTS

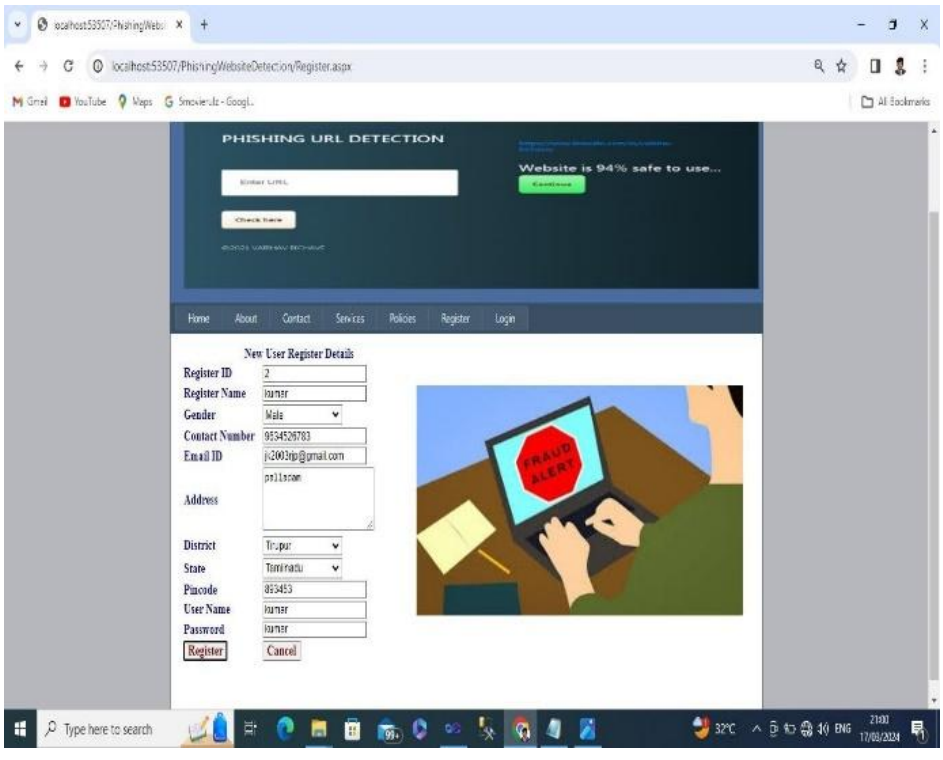



PHISHING WEBSITE DETECTION USING ML


PHISHING WEBSITE DETECTION
Check the
1. URL
http://
Home
Add Companies
Add URL
Fake URL
Report
Logout
Add URL Details
Company ID
Company Name
Tata
URL
www.tatamotors.com
URL Category
Safe URL
URL Type
Social Media
Upload Date
Company Image
Choose File
Add URL
Cancel


PHISHING URL DETECTION
Website is 94% safe to use...
Home
About
Contact
Services
Policies
Register
Login
New User Register Details
Register ID
Register Name
Gender
Male
Contact Number
Email ID
Address
District
State
Pincode
User Name
Password
Register
Cancel
FRAUD ALERT!

7.REFERENCES :


1. **"ASP IN A NUTSHELL", "**A. Keyton Weissinger**"**, Shroff Publishers anddistributors Pvt.Ltd, February 1999.
2. "ASP.NET Complete Reference", "A. Russel Jones", Sybex Publications, February 18,2002.
3. "C#.NET Black Book", "Steven Holzner", Dreamtech Publications, 2003 Edition.
4. "DATABASE SYSTEM CONCEPTS", "Henry F. Korth", Megraw-Hill, Third Edition 1997.
5. "Learning Visual Basic .Net Through Applications", Clayton crooks II.
6. "Professional VB.NET 1.1", "Alex Homer", Worx Publications, 2004 Edition.
7. "SOFTWARE ENGINEERING CONCEPT", "Richard Fairly Tata", McGraw-Hill Publications, Third Edition 1997.
8. "Software Engineering", "Roger S Pressman", Dreamtech Publications.
9. "SQL SERVER HIGH AVAILABILITY", "Paul Bertucci", Sams publishing, First Edition 2004.
10. "Visual Basic.NET Black Book", "Steven Holzner", Dreamtech Publications, 2000 Edition.