## “ELECTRICITY ENERGY THEFT DETECTION IN SMART GRIDS USING DEEP

## NEURAL NETWOKS”

**A PROJECT REPORT**

Submitted to

## JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY ANANTAPURAMU

*In partial fulfillment of the requirements for the award of the degree of*

### Bachelor of Technology

In

### Computer Science and Engineering

##### By

**202K1A0504 - ARCOT NAVEEN**

**202K1A0535 - LAMBADE VENKATESH**

**202K1A0527 - JUJARE RANJITH KUMAR**

**202K1A0515 - D K KRISHNA MOHAN**

**202K1A0536 - M VINOD KUMAR**

**20HU1A0566 - MACHAM RAVI KIRAN**

Under the esteemed guidance of:

**B HUSSAIN,**

Assistant Professor.

DEPARTMENTOFCOMPUTERSCIENCEAND ENGINEERING

**CERTIFICATE**

This is to certify that the Project Report entitled “**ELECTRICITY ENERGY THEFT DETECTION IN SMART GRIDS USING DEEP NEURAL NETWORKS ”**is bonafide work of **ARCOT NAVEEN (202K1AO504),LAMBADE VENKATESH (202K1A0535), JUJARE RANJITH KUMAR (202K1A0527), D K KRISHNA MOHAN (202K1A0515), M VINOD KUMAR (202K1A0536), MACHAM RAVI KIRAN (20HU1A0566)** submitted to the Department of Computer Science & Engineering, in partial fulfillment of the requirements for the award of degree of Bachelor of Technology in **COMPUTER SCIENCE AND ENGINEERING** from Jawaharlal Nehru Technological University, Anantapuramu.

|  |  |
| --- | --- |
| **Signature of the Supervisor.** | **Signature of the Head of the Dept.** |
| **B HUSSAIN M.Tech.** | **Dr. WILLIAM ALBERT** |
| **Assistant Professor.** | **H.O.D.** |
| **Department of CSE** | **Department of CSE** |
| **Bheema Institute of Technology & Science** | **Bheema Institute of Technology & Science** |

#### DEPARTMENTOFCOMPUTERSCIENCEANDENGINEERING

**DECLARATION**

We hereby declare that the project Report entitled **“ELECTRICITY ENERGY THEFT DETECTION IN SMART GRIDDS USING DEEP NEURAL NETWORKS”** submitted by the Department of Computer Science and Engineering, **Bheema Institute of Technology & Science**, **Alur Road, Adoni,** in partial fulfillment of the requirements for the award of the degree of **Bachelor of Technology in Computer Science and Engineering** is are cord of bonafide work carried out by me under the supervision of Assistant Professor Mr. B HUSSAIN**,** We further declare that the work reported in this project has not been submitted and will not be submitted, either in part or in full, for the award of any other degree or diploma of this institute or any other institute or university.

### Signature

**ARCOT NAVEEN 202K1A0504 LAMABADE VENKATESH 202K1A0535 JUJARE RANJITH KUMAR 202K1A0527**

**D K KRISHNA MOHAN 202K1A0515**

**M VINOD KUMAR 202K1A0536**

**MACHAM RAVI KIRAN 20HU1A0566**

#### DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

The project report entitled **“ELECTRICITY ENERGY THEFT DETECTION IN SMART GRIDS USIND DEEP NEURAL NETWORKS”** is prepared and submitted by **ARCOT NAVEEN (202K1A0504), LAMBADE VENKATESH (202K1A0535), JUJARE RANJITH KUMAR (202K1A0527), D K KRISHNA MOHAN (202K1A0515), M VINOD KUMAR (202K1A0536), MACHAM RAVI KIRAN (20HU1A0566).** It has been found satisfactory in terms of scope, quality and presentation as partial fulfillment of the requirements for the award of the degree of **Bachelor of Technology in Computer Science and Engineering** in Bheema Institute of Technology & Science, Alur Road, Adoni, Kurnool, A.P.

Guide Head of the Department

**Mr. B.HUSSIAN Dr.**

Assistant Professor Professor

Department of CSE Department of CSE

Internal Examiner External Examiner

**ACKNOWLEDGEMENTS**

The satisfaction and euphoria that accompany the successful completion of any task would be incomplete without the mention of people who made it possible, whose constant guidance and encouragement crowned our efforts with success. It is a pleasant aspect that we have now the opportunity to express our guidance for all of them.

The first and foremost, We would like to express our deep gratitude to our Project Guide, **Mr. B. HUSSAIN,** Assistant Professor of Computer Science and Engineering Department, who has extended his support for the success of this project. His wide knowledge and logical way of thinking have made a deep impression on us. His understanding, encouragement and personal guidance have provided the basis for this thesis. His source of inspiration for innovative ideas and his kind support is well to all his students and colleagues.

We express our thanks to Project Coordinator, **Mrs. XXXXXXXXX .Tech (PhD)**, for her Continuous support and encouragement.

We wish to thank **Dr. XXXXXXXX,** Head of Computer Science and Engineering Department. His wide support, knowledge and enthusiastic encouragement have impressed us to better involvement into our project thesis and technical design also his ethical morals helped us to develop our personal and technical skills to deploy our project in success.

We wish to thank **Dr. XXXXXXXX**, Principal of Bheema Institute of Technology & Science. who has extended his support for the success of this project.

We express our sincere thanks to the project committee members, faculty and staff of Computer Science and Engineering Department, Bheema Institute of Technology & Science, for their valuable guidance and technical support.

Last but far from least, we also thank our family members and our friends for their moral support and constant encouragement, We are very much thankful to one and all who helped us for the successful completion of the project.

With gratitude

**ARCOT NAVEEN 202K1A0504**

**LAMBADE VENKATESH 202K1A0535**

**JUJARE RANJITH KUMAR 202K1A0527**

**D K KRISHNA MOHAN 202K1A0515**

**M VINOD KUMAR 202K1A0536**

**MACHAM RAVI KIRAN 20HU1A0566**

**INDEX**

|  |  |  |
| --- | --- | --- |
| **CHAPTER** | **CONTENT** | **PAGE.NO** |
|  | **ACKNOWLEDGEMENT** |  |
| **LIST OF FIGURES** |  |
| **ABSTRACT** | **1** |
| 1 | **INTRODUCTION**   * 1. BASICS OF THE PROJECT   2. MATIVATION   3. PROBLEM STATEMENT AND EXPLANATION   4. PURPOSE   5. INNOVATION OF THE PROJECT | **2-4** |
| **2** | **LITERATURESURVEY**   * 1. INTRODUCTION   2. EXISTING SYSTEM   3. PROPOSED SYSTEM | **5-9** |
| **3** | **KEYTECHNOLOGIES dnn**   * 1. CLIENT SERVER   2. USER INTERFACE DESIGN   3. COMMUNICATION   4. FEATURES OF THE LANGUAGE USED   5. IMPORTANCE OF JAVA TO THE INTERNET   6. FEATURES OF JAVA   7. ADVANTAGES   8. HYPER TEXT MARKUP LANGUAGE   9. ADVANTAGES   10. JAVA DATABASE CONNECTIVITY   11. TOMCAT6.0 WEBSERVER   12. BIBLIOGRAPHY | **10-26** |
| **4** | **IMPLEMENTATIONOFTHEPROJECT**   * 1. DATASETREQUIREMENTS   2. ARCHITECTUREDIAGRAM   3. INNOVATIONOFTHE PROJECT | **27-31** |

|  |  |  |
| --- | --- | --- |
|  | 4.4PLANOFACTIONTOCOMPLETEPROJECT |  |
| **5** | **SYSTEMSTUDY**   * 1. FEASIBIILITYSTUDY | **32-34** |
| **6** | **SYSTEMDESIGN**   * 1. UMLDIAGRAMS   2. GOALS   3. USECASE DIAGRAM   4. CLASSDIAGRAM   5. SEQUENCEDIAGRAM   6. DATAFLOWDIAGRAM   7. ALGORTHIMS | **35-40** |
| **7** | **SYSTEMREQUIREMENTS**   * 1. HARDWAREREQUIREMENTS   2. SOFTWAREREQUIREMENTS | **41-42** |
| **8** | **SYSTEMTESTING**   * 1. TYPESOFTESTING   2. TESTSTRATEGYANDAPPROACH   3. TESTINGMETHODOLOGIES   4. OTHERTESTINGMETHODOLOGIES   5. TESTINGSTRATEGY | **43-50** |
| **9** | **EXPERIMENTANALYSISANDRESULTS**   * 1. SCREENSHOTS | **51-55** |
| **10** | **CONCLUSION** | **56-57** |
| **11** | **REFERENCES** | **58-60** |

**LIST OF FIGURES**

Figure1:CompilingandinterpretingJavasourcecode 15

Figure2:Two-tierandThree-tire Models 20

Figure3:Tomcat6.0 webserver 24

Figure4:System Architecture 29

Figure5:UseCaseDiagram 37

Figure6:ClassDiagram 38

Figure 7:Sequence Diagram 39

Figure8:DataFlow Diagram 40

Figure9: Screen ShotsofData… 52-55

**ABSTRACT**

In our investigation of social semantic attack detection models, we focused on leveraging character-aware language models to detect various types of attacks such as phishing, spamming, defacement, and malware attacks. We developed three distinct models: a long short-term memory (LSTM)-based model, a convolutional neural network (CNN)- based model, and a CharacterBERT-based model. These models were designed to detect deceptive URLs by analyzing their semantic features.

Throughrigorousevaluation,wefoundthattheCharacterBERT-baseddetectionmodel exhibited exceptional performance. Across all evaluated metrics, it achieved an impressive overall detection accuracy of 99.65% when subjected to a 5-fold cross-validation process. Whenexamining the performance ofeach modelper attackclass, the CharacterBERT model outperformed the LSTM and CNN models, particularly excelling in detecting defacement attacks with a remarkable accuracy of 99.90%.

**CHAPTER-1: INTRODUCTION**

## CHAPTER -1 INTRODUCTION

#### BASICSOFTHEPROJECT

Numerous techniques are tested by researchers in the area for DR classification with encouraging results. Recent work for addressing blood vessel segmentation includes the applicationofCNN(LeNet-5architecture)asfeatureextractor[6].Threeheadsareused inthis modelatdifferentlayersoftheconvnentwhicharethenfedintothreerandomforests.Thefinal classifier achieved an accuracy of 0.97 and 0.98 on the DRIVE [7] and STARE dataset. An automatic segmentation of blood vessels in color fund us images is implemented by M.Melinscaketal[8]usingdeepmax-poolingconvnettoseparatethebloodvessels.Themodel contains a deep max-pooling convolutional neural networks to segment blood vessels. It deployed 10- layer architecture for achieving a maximum accuracy of around 0.94. It was carried around 4-convolutional and 4-max pooling layer with 2 additional fully connected layersforvesselsegmentation.AutomatedanalysisofDRusing imagesprocessingtechniques areintroducedbyAdarshetal[9].Inthisapproach,extractionofretinalbloodvessels,exudate, micro-aneurysms, haemorrhages and texture features takes place, followed byconstruction of MulticlassSVMusingareaoflesionsandtexturefeatures.Impressiveresultsarereportedusing thepubliclyavailabledatasetsDIARETDB0andDIARETDB1withaccuracyof0.96and0.946 respectively.

#### MOTIVATION:

The primary aim of this project is to develop an automated diabetic retinopathyscreening system, a detectionofMicro aneurysms in digitalfundus photographs is needed. Micro aneurysms are the first clinical sign of diabetic retinopathy. The number of microaneurysms is used to indicate the severity of the disease. Early micro- aneurysmdetectioncanhelp reduce the incidence of blindness.

#### PROBLEMSTATEMENTANDEXPLANATION:

The method proposed in this paper aims at detecting the various stages of Diabetic Retinopathy by using U-Net segmentation with region merging & Convolutional Neural Network.Theretinalsegmentationistheprocessofautomaticdetectionofboundariesofblood vessels within the retina. This allows classifier to learn important features such as retinal proliferation&retinaldetachment.Thedatalostduringretinalsegmentationisretracted

through region merging. Due to the inclusion of retinal segmentation this method completely outperforms previous models, with accuracy up to 93.33.

#### PURPOSE:-

We demonstrated that our approach enables successful segmentation of intra-retinal layers—evenwith low-quality imagescontaining speckle noise, low contrast,and different intensity ranges throughout—with the assistance of the ANIS feature.

#### INNOVATIONOFTHEPROJECT:

Millions ofpeople all around the world are suffering fromdiabetes which is a prominent cause of Diabetic Retinopathy. It is one of the leading causes of blindness and its effect can be reduced with proper treatment. So, it is necessary to build a system which can detect diabetic retinopathy without any expert guidance.

The inclusion ofretinal segmentation with region merging during preprocessing provides significant improvement overexisting model.Theproposedmodelprovidesresult asgoodasany supervised algorithm without explicitly determining features required for classification.

# CHAPTER-2: LITERATURESURVEY

## CHAPTER - 2 LITERATURESURVEY

#### INTRODUCTION

Social engineering is threatening the security of all computer networks despite the robustnessoftheir cryptographic methods, intrusiondetectionsystems, firewalls, or anti-virus software systems. Social engineering attacks refer to a specific category of cyber security attacksthatreliesonhumanerrorsandbehavioralattributesratherthansoftwarevulnerabilities. Social engineers manipulate victims of attacks with the intent of gaining confidential information and valuable data. Users are tricked into clicking on obfuscated URLs, downloading a harmful attachment, or sharing sensitive and private information.

The semantic attack is a subcategoryof socialengineering attacksthat are challenging to identify because the attacks utilize behavioral or cosmetic deception vectors (e.g., attacker create amaliciouswebsite that lookslike and behaves like the legitimate one), whichare very difficult to be identified or traced by computer programs.

The most common types of social semantic attacks include phishing, spamming, defacement,andmalwareattacks.InTable1,wediscussthesetypesofsemanticattacks.Inthe relatedworksection,wewillexplorethetechniquesusedtodetectthesesemanticattacks.Many tools, defense mechanisms, and machine learning and deep learning algorithms have been designed and developed to detect cyber attacks.

Semi-supervised machine learning models are used to generate text embeddings. BidirectionalEncoderRepresentationsfromTransformers(BERT)isalanguagerepresentation model that has recently become the most popular choice for building natural language processing(NLP)systems.ThetransformershavebeenusedefficientlyinNLPtasks,especially for analyzing time series data. In this research, we want to study the URL and its sequential patternto capturemaliciousactions. Wewant toinvestigatethe likelihoodofobtainingagood model performance for URL-specialized vocabulary if the standard word embedding is used.

Webuiltasimplerword-levelmodelasinsteadofBERT’sword-piecevocabulary.This method createsembedding fromthe characterofeachtokento build representations similarto the BERT models. Figure 1 shows a comparison between BERT and Character BERT.

The*contributions*ofthis researchare asfollows:

* We proposed novel character-aware language models for detecting URL-based social semantic attacks.
* We experimented three different algorithms (long short term memory (LSTM)-based detectionmodel,convolutionalneuralnetwork(CNN)-baseddetectionmodel,andCharacter BERT-baseddetectionmodel)fordetectingfourdifferentsocialsemanticattacks(phishing, spamming, defacement, and malware attacks).
* Webenchmarkedperformancesofdifferentmodelsfordifferentattacks.

#### EXISTINGSYSTEM

Over the past few years, researchers have designed and developed a varietyofdetection tools and techniques to detect cyberattacks. Many studies in the literature examined the efficiencyoftheexistingwebbrowsersecuritymeasurementandtheend-usertoolsindetecting the security vulnerability that leads to cyberattacks. In the field of cybersecurity, researchers are increasingly showing interest in deep learning and machine learning due to their growing popularity.

Many machine learning and advanced deep learning models have been proposed for social engineering attack detection. These models use various algorithms including support vector machine, decision trees, recurrent neural networks, transformers, convolutional neural networks, long short-term memory, multi-layer feed-forward networks to detect social engineeringattacks. Sahooet al. surveyedthe machine learning methodsusedinthedetection of malicious URL. They examined different aspects of malicious URL detection, such as features and models.

Many AI systems developers realize that training a system by showing it examples of desired input-output behavior is easier than programming it . By providing researchers with automaticfeatureextractionabilities,deeplearningalgorithmshaveentirelychangedthefield. The researchers can avoid the time-consuming task ofchoosing the utmost substantial feature extractionstrategies pertinent to a particular problemdue to their automatic feature extraction capabilities, which also guarantee higher detection rates than traditional machine learning classification algorithms.

The phishing attacks can be detected by employing machine learning techniques, text miningandNLPalgorithms,wherethealgorithmscanlearnandidentifythepatternsofthe

suspiciousactivities. In[49],theauthorsproposednoveldeep learning models for thepurpose ofwebsitephishingdetection.Usingonly10featuresfromtheirpriorwork,theypresentednew phishing URL detection methods by employing LSTM, CNN, and DNN. They achieved an accuracyof99.43% using the CNN model, 99.52% using the DNN model, and 99.57% using the LSTM model. A studyby Wang et al. It aims to detect phishing byutilizing Bidirectional LSTM(BLSTM)andrandomforestmodels.ThetestingresultsindicatethattheBLSTM-based phishingdetectionmodelwas significant inconfirming networksecurity, withadetectionrate of 95.47%, while only achieving 87.53% using random forest model.

Yi et al. demonstrated an approach to detect website phishing using a deep learning framework. They utilized two types of features for website phishing detection: the original feature(i.e.,thedirect featureofURL)andtheinteractionfeature(i.e.,theinteractionbetween websites). In their experiments using real IP flows fromInternet Service Providers (ISPs), the detectionmodelbasedonDeepBeliefNetworks(DBN)algorithmachievedatruepositiverate of 90%. In [6], researchers conducted a systematic studyof the effectiveness of deep learning algorithm architectures for phishing website detection. They utilized three types of website features (URLbased, content-based, and hybrid (URL and content together). Three deep learning algorithmarchitectures were applied in their study: Long Short-TermMemory, Fully ConnectedDeepNeuralNetwork,andConvolutionalNeuralNetwork.Theybuiltandevaluated themodelsusingfourpubliclyavailablephishingwebsitedatasets,achievingthebestaccuracy of 97.37%.

In [54], Tida and Hsu used pre-trained Google’s Bidirectional Encoder Representations fromTransformersbaseuncasedmodelstoclassifyhamorspamemailsinreal-timesituations

- 97%totalaccuracywasattained withan F1 score of0.96.Thisstudyfocused ontraining and testing using spam and ham emails/messages datasets. In the same vein, other researchers developed various different machine learning/deep learning models to detect spam emails. In recent study[40],theauthorsusedBERTforthedetectionofspamtext messages. TheBERT- based model, withanaccuracyof98.63%, outperformed other models based onNaive Bayes, Support Vector Machine, Logistic Regression, and Decision Tree.

To monitorand detect website defacement attacks,researchershave proposed avariety of tools and models. Some solutions onlyworkondynamic webpages, while others workonstatic webpages. The study in [20] presents a hybrid defacement detection model that achieved an accuracyof99.26%andafalsepositiverateof0.27%.In[9],authorsusedcomputervision

techniques to detect website defacement. Their detection system achieved true positive rates between97.42%and98.81%,andfalsepositiveratesbetween0.54%and1.52%.Davanzoetal.

[12] developed a system that employs the website’s HTMLsource code-related feature to monitor website defacement. Their dataset contains 300 legitimate websites and 320 defacements. Their experiments results show that the detection methods, i.e,Domain Knowledge, SVM, PulseParzen, and Hotelling, all show very low false negative rate and false positive rate values. Theyreportedthat the average value of false positive rate is less than 1%.

#### PROPOSEDSYSTEM

* + - We proposed novel character-aware language models for detecting URL-based social semantic attacks.
    - We experimented three different algorithms (long shortterm memory (LSTM)-based detection model, convolutional neural network (CNN)-based detection model, and CharacterBERT-baseddetectionmodel)fordetectingfourdifferentsocialsemanticattacks (phishing, spamming, defacement, and malware attacks).
    - Webenchmarkedperformancesofdifferentmodelsfordifferentattacks.
    - Statisticalparametersofthedatasetwereanalyzedtodeterminetheappropriateembedding dimension, utilizing the Hashing method.

# CHAPTER-3: KEYTECHNOLOGIES

* 1. **ClientServer**
     1. **Overview:**

## CHAPTER-3

**KEYTECHNOLOGIES**

Withthe variedtopic inexistence inthe fieldsofcomputers, Client Server isone, which has generated more heat than light, and also more hype than reality. This technology has acquiredacertaincriticalmassattentionwithitsdedicationconferencesand magazines.Major computervendorssuchasIBMandDEC,havedeclaredthatClientServersistheirmainfuture market. AsurveyofDBMS magazine reveled that 76% of its readers were actively looking at theclient serversolution.Thegrowthintheclientserverdevelopmenttoolsfrom$200million in 1992 to more than $1.2 billion in 1996.

Client server implementations are complex but the underlying concept is simple and powerful.Aclientisanapplicationrunningwithlocalresourcesbutabletorequestthedatabase and relate the services from separate remote server. The software mediating this client server interaction is often referred to as MIDDLEWARE.

The typical client either a PC or a Work Station connected through a network to a more powerful PC, Workstation, Midrange or Main Frames server usually capable of handling request from more than one client. However, with some configuration server may also act as client. Aserver may need to access other server inorder to process theoriginalclient request.

The keyclient server idea is that client as user is essentially insulated fromthe physical locationandformatsofthedataneedsfortheirapplication.Withthepropermiddleware,aclient input from or report can transparently access and manipulate both local database on the client machine and remote databases on one or more servers. An added bonus is the client server opens the door to multi-vendor database access indulging heterogeneous table joins.

WhatisaClientServer

Twoprominentsystemsinexistenceareclientserverandfileserversystems.Itisessential to distinguish between client servers and file server systems. Both provide shared network access to data but the comparison dens there! The file server simply provides a remote disk drive that canbe accessed byLAN applications on a file byfile basis. The client server offers full relational database services such as SQL-Access, Record modifying, Insert, Delete with full relational integrity backup/ restore performance for high volume of transactions, etc. the clientservermiddlewareprovidesaflexibleinterfacebetweenclientandserver,whodoeswhat, when and to whom.

WhyClientServer

Client server hasevolvedto solve aproblemthat has beenaroundsincetheearliest days ofcomputing:howbesttodistributeyourcomputing,datagenerationanddatastorageresources in order to obtain efficient, cost effective departmental an enterprise wide data processing. Duringmainframeerachoiceswerequitelimited. AcentralmachinehousedboththeCPUand DATA(cards,tapes,drumsandlaterdisks).Accesstotheseresourceswasinitiallyconfinedto batched runs that produced departmental reports at the appropriate intervals. A strong central information service department ruled the corporation. The role of the rest of the corporation limited to requesting new or more frequent reports and to provide hand written forms from which the central data banks were created and updated. The earliest client server solutions therefore could best be characterized as “SLAVE-MASTER”.

Time-sharing changed the picture. Remote terminal could view and even change the central data, subject to access permissions. And, as the central data banks evolved in to sophisticated relational database with non-programmer query languages, online users could formulate adhoc queries and produce local reports with out adding to the MIS applications softwarebacklog.Howeverremoteaccesswasthroughdumbterminals,andtheclient server remained subordinate to the Slave\Master.

* 1. FrontendorUserInterfaceDesign

Theentireuserinterfaceisplannedtobedevelopedinbrowserspecificenvironmentwith a touch of Intranet-Based Architecture for achieving the Distributed Concept. The browser specific components are designed by using the HTML standards, and the dynamism of the designed by concentrating on the constructs of the Java Server Pages.

* 1. CommunicationorDatabaseConnectivityTier

The Communication architecture is designed by concentrating on the Standards of Servlets and Enterprise Java Beans. The database connectivity is established by using the Java Data Base Connectivity. The standards of three-tire architecture are given major concentration to keepthestandardsofhigher cohesionand limitedcoupling for effectivenessoftheoperations.

* 1. FeaturesofTheLanguageUsed

Inmyproject,IhavechosenJavalanguagefor developingthecode

Initially the language was called as “oak” but it was renamed as “Java” in 1995. The primary motivationof thislanguagewastheneedforaplatform-independent(i.e.,architectureneutral)

language that could be used to create software to be embedded in various consumer electronic devices.

* Java isaprogrammer’s language.
* Javaiscohesiveand consistent.
* ExceptforthoseconstraintsimposedbytheInternetenvironment,Javagivesthe programmer, full control.

Finally,JavaistoInternetprogrammingwhereCwastosystemprogramming.

* 1. ImportanceofJavatotheInternet

JavahashadaprofoundeffectontheInternet.Thisisbecause;JavaexpandstheUniverse ofobjectsthatcanmoveaboutfreelyinCyberspace.Inanetwork,twocategoriesofobjectsare transmitted between the Server and the Personalcomputer. Theyare: Passive information and Dynamic active programs. The Dynamic, Self-executing programs cause serious problems in the areas ofSecurityand probability. But, Java addresses those concerns and bydoing so, has opened the door to an exciting new form of program called the Applet.

Javacanbeusedtocreatetwotypesofprograms

Applications and Applets: Anapplication is a programthat runs onour Computer under theoperatingsystemofthatcomputer.ItismoreorlesslikeonecreatingusingCorC++.Java’s ability to create Applets makes it important. An Applet is an application designed to be transmitted over the Internet and executed by a Java –compatible web browser. An applet is actuallya tinyJava program, dynamicallydownloaded across the network, just like an image. But the difference is, it is an intelligent program, not just a media file. It can react to the user input and dynamically change.

##### FeaturesOfJava

Security

Everytime youthatyoudownloada“normal”program, youareriskingaviralinfection. Priorto Java, most usersdid notdownloadexecutableprogramsfrequently,andthosewho did scanned them for viruses prior to execution. Most users still worried about the possibility of infecting their systems with a virus. In addition, another type of malicious programexists that must be guarded against. This type of program can gather private information, such as credit cardnumbers,bankaccountbalances,andpasswords.Javaanswersboththeseconcernsby

providinga“firewall”betweenanetworkapplicationandyourcomputer.WhenyouuseaJava- compatibleWebbrowser,youcansafelydownloadJavaappletswithoutfearofvirusinfection or malicious intent.

Portability

For programs to be dynamically downloaded to all the various types of platforms connected to the Internet, some means of generating portable executable code is needed .As you will see, the same mechanism that helps ensure security also helps create portability. Indeed, Java’s solution to these two problems is both elegant and efficient.

TheBytecode

The key that allows the Java to solve the security and portability problems is that the output of Java compiler is Byte code. Byte code is a highly optimized set of instructions designedtobeexecutedbytheJavarun-timesystem,whichiscalledtheJavaVirtualMachine (JVM). That is, in its standard form, the JVM is an interpreter for byte code.

Translating a Java program into byte code helps makes it much easier to run a program in a wide varietyofenvironments. The reasonis, once the run-time package exists for a given system, any Java program can run on it.

Although Java was designed for interpretation, there is technically nothing about Java that prevents on-the-flycompilation of byte code into native code. Sun has just completed its JustInTime(JIT)compiler forbytecode.WhentheJITcompiler isapartofJVM, itcompiles bytecodeintoexecutablecodeinrealtime,onapiece-by-piece,demandbasis.Itisnotpossible to compile an entire Java program into executable code all at once, because Java performs various run-time checks that can be done only at run time. The JIT compiles code, as it is needed, during execution.

JavaVirtualMachine(JVM)

Beyond the language, there is the Java virtual machine. The Java virtual machine is an important element oftheJavatechnology. The virtualmachinecanbeembeddedwithinaweb browser or an operating system. Once a piece of Java code is loaded onto a machine, it is verified.Aspartoftheloadingprocess,aclassloaderisinvokedanddoesbytecodeverification makessurethatthecodethat’shasbeengeneratedbythecompilerwillnotcorruptthemachine that it’s loaded on.

Java programming uses to produce byte codes and executes them. The first box indicates that theJavasourcecodeislocatedina.JavafilethatisprocessedwithaJavacompilercalledjavac. TheJavacompilerproducesafilecalleda.class file,whichcontainsthe bytecode.The.Class file is then loaded across the network or loaded locally on your machine into the execution environment is the Java virtual machine, which interprets and executes the byte code.

JavaArchitecture

Java architecture provides a portable, robust, high performing environment for development. Java provides portability by compiling the byte codes for the Java Virtual Machine, which is then interpreted on each platform by the run-time environment. Java is a dynamicsystem,abletoloadcodewhenneededfromamachineinthesameroomoracrossthe planet**.**

Compilationof code

Whenyoucompilethecode,theJavacompilercreatesmachinecode(calledbytecode) forahypotheticalmachinecalledJavaVirtualMachine(JVM).TheJVMissupposedtoexecute the byte code.The JVM is createdforovercoming the issue ofportability. The code is written and compiled for one machine and interpreted on all machines. This machine is called Java Virtual Machine.

CompilingandinterpretingJavaSourceCode

During run-time the Java interpreter tricks the byte code file into thinking that it is runningonaJavaVirtualMachine.Inrealitythiscould beaIntelPentiumWindows95orSun SARC station running Solaris or Apple Macintosh running systemand all could receive code from any computer through Internet and run the Applets.

|  |
| --- |
| **Java**  **Java Interpreter**  **Source (PC)**  **Code Bytecode**  **……….. Macintosh Java**  **……….. Interpreter**  **(Platform (Macintosh**  **SPARC indepen** **JavaInterpreter (Sparc)** |
|  |
|  |

Fig:3.6

Simple

Java was designed to be easy for the Professional programmer to learn and to use effectively. If you are an experienced C++ programmer, learning Java will be even easier. BecauseJavainheritstheC/C++syntaxandmanyoftheobjectorientedfeaturesofC++.Most of the confusing concepts from C++ are either left out of Java or implemented in a cleaner, more approachable manner. In Java there are a small number of clearly defined ways to accomplish a given task.

Object-Oriented

Javawasnotdesignedtobesource-codecompatiblewithanyotherlanguage.Thisallowed theJavateamthefreedomtodesignwithablankslate.Oneoutcomeofthiswasacleanusable, pragmatic approach to objects. The object model in Java is simple and easy to extend, while simple types, such as integers, are kept as high-performance non-objects.

Robust

Themulti-platformenvironment ofthe Web placesextraordinarydemandsonaprogram, because the programmust execute reliablyina varietyofsystems. The abilityto createrobust programs was given a high priority in the design of Java. Java is strictly typed language; it checks your code at compile time and run time. Java virtually eliminates the problems of memory management and deallocation, which is completelyautomatic. In a well-written Java program, all run time errors can –and should –be managed by your program.

Java Script

JavaScript is a script-based programming language that was developed by Netscape Communication Corporation. JavaScript was originally called Live Script and renamed as JavaScript to indicate its relationship with Java. JavaScript supports the development of both client and server components of Web-based applications. On the client side, it can be used to write programs that are executed bya Web browser within the context ofa Webpage. Onthe serverside,itcanbeusedtowriteWebserverprogramsthatcanprocessinformationsubmitted by a Web browser and then updates the browser’s display accordingly

Even though JavaScript supports both client and server Web programming, we prefer JavaScriptatClient sideprogrammingsincemostofthebrowserssupportsit.JavaScript isalmost as easy to learn as HTML, and JavaScript statements can be included in HTML documents by enclosing the statements between a pair of scripting tags <SCRIPTS>..</SCRIPT>.

<SCRIPTLANGUAGE=“JavaScript”>

JavaScriptstatements

</SCRIPT>

HereareafewthingswecandowithJavaScript:

Validatethecontentsofaformandmake calculations.

* Addscrolling orchangingmessagesto theBrowser’sstatusline.
* Animateimagesorrotateimagesthatchangewhenwemovethemouseover them.
* Detectthebrowserinuseanddisplaydifferentcontentfordifferentbrowsers.
* Detectinstalledplug-insandnotifytheuserifaplug-inisrequired.
* WecandomuchmorewithJavaScript,includingcreatingentireapplication.

**JavaScriptVsJava**

JavaScriptandJavaareentirelydifferentlanguages.Afewofthe mostglaringdifferences are:

* Java applets are generally displayed in a box within the web document; JavaScript can affect any part of the Web document itself.
* While JavaScript is best suited to simple applications and adding interactive features to Web pages, Java can be used for incredibly complex applications.

TherearemanyotherdifferencesbuttheimportantthingtorememberisthatJavaScriptandJava are different languages. They are both useful for different purposes, in fact they can be used together to combine their advantages.

* 1. **ADVANTAGES**
* JavaScript canbeusedforSever-sideandClient-sidescripting.
* Itismore flexiblethanVBScript.
* JavaScriptisthedefaultscriptinglanguagesatClient-sidesinceallthebrowserssupports it.
  1. HyperTextMarkupLanguage

Hypertext Markup Language (HTML),the languagesofthe World Wide Web(WWW), allows users to produces Web pages that include text, graphics and pointer to other Web pages (Hyperlinks).

HTML is not a programming language but it is an application of ISO Standard 8879, SGML(StandardGeneralizedMarkupLanguage),butspecializedtohypertextandadaptedto

the Web. The idea behind Hypertext is that instead of reading text in rigid linear structure, we caneasilyjumpfromonepointtoanotherpoint.Wecannavigatethroughtheinformationbased onourinterestandpreference.Amarkuplanguageissimplyaseriesofelements,eachdelimited with special characters that define how text or other items enclosed within the elements should be displayed. Hyperlinks are underlined or emphasized works that load to other documents or some portions of the same document.

HTML can be used to display any type of document on the host computer, which can be geographicallyat adifferent location. It isa versatile language andcanbeusedonanyplatform or desktop.

HTMLprovidestags(specialcodes)tomakethedocumentlookattractive.HTMLtagsare notcase-sensitive.Usinggraphics,fonts,differentsizes,color,etc.,canenhancethepresentation of the document. Anything that is not a tag is part of the document itself.

Basic HTMLTags :

<!-- --> Specifiescomments

<A>……….</A> Createshypertextlinks

<B>……….</B> Formatstextas bold

<BIG>……….</BIG> Formatstextinlargefont. Contains all tags and text in the HTML document

<CENTER>...</CENTER> Creates text

<DD>…</DD> Definitionofaterm

<DL>...</DL>

Createsdefinitionlist

<FONT>…</FONT> Formatstextwithaparticularfont

<FORM>...</FORM> Enclosesafill-outform Defines a particular frame in a set of frames

<H#>…</H#> Createsheadingsofdifferentlevels Contains tags that specify information about a document

<HR>...</HR> Createsa horizontalrule

<HTML>…</HTML> ContainsallotherHTMLtags Provides meta-information about a document

<SCRIPT>…</SCRIPT> Containsclient-sideorserver-sidescript

<TABLE>…</TABLE> Creates atable

<TD>…</TD> Indicatestabledatainatable

<TR>…</TR> Designatesatable row

<TH>…</TH> Createsaheadingina table

* 1. **ADVANTAGES**
* AHTMLdocument issmalland henceeasyto sendoverthenet. It issmallbecause it does not include formatted information.
* HTMLisplatformindependent.
* HTMLtagsare notcase-sensitive.

##### JavaDatabaseConnectivity

WhatIsJDBC?

JDBC is a Java API for executing SQL statements. (As a point of interest, JDBC is a trademarkednameandisnotanacronym;nevertheless,JDBCisoftenthoughtofasstandingfor Java Database Connectivity. It consists of a set of classes and interfaces written in the Java programming language. JDBCprovidesa standard API fortool/database developersand makes it possible to write database applications using a pure Java API.

Using JDBC, it is easyto send SQL statements to virtuallyanyrelationaldatabase. One can write a single program using the JDBC API, and the program will be able to send SQL statementsto the appropriate database. The combinations ofJava and JDBC lets a programmer write it once and run it anywhere.

WhatDoes JDBCDo?

JDBCmakesitpossibletodothreethings:

* Establishaconnectionwitha database
* SendSQLstatements
* Processtheresults.

JDBCversusODBCandother APIs

At this point, Microsoft's ODBC (Open Database Connectivity) API is that probably the most widelyused programming interface for accessing relationaldatabases. It offersthe ability to connect to almost all databases on almost all platforms.

So whynot just use ODBC fromJava?The answer is that you canuse ODBC fromJava, but this is best done with the help of JDBC in the formof the JDBC-ODBC Bridge.

Two-tierand Three-tierModels

TheJDBCAPIsupportsbothtwo-tierandthree-tiermodels fordatabaseaccess.

In the two-tier model, a Java applet or application talks directly to the database. This requires a JDBC driver that can communicate with the particular database management system being accessed. A user's SQL statements are delivered to the database, and the results of those statements are sent back tothe user. The database may be located onanother machine to which the user is connected via a network. This is referredto as a client/server configuration, withthe user's machine as the client, and the machine housing the database as the server. The network canbeanIntranet,which,forexample,connectsemployeeswithinacorporation,oritcanbethe Internet.

**Clientmachine**

**JDBC**

**JAVA**

**Application**

**DBMS**

**DBMS-proprietaryprotocol**

**Databaseserver**

**Clientmachine(GUI)**

**Application Server(Java) JDBC**

**Javaappletor Htmlbrowser**

**HTTP,RMI,orCORBAcalls**

**Servermachine(businessLogic)**

**Databaseserver**

DBMS

Fig:3.10

Inthethree-tier model, commands are sent to a"middle tier" ofservices, whichthensend SQL statements to the database. The database processes the SQL statements and sendsthe results back to the middle tier, which then sends them to the user. MIS directors find the three- tier model very attractive because the middle tier makes it possible to maintain control over access and the kinds of updates that can be made to corporate data. Another advantage is that when there is a middle tier, the user can employ an easy-to-use higher-level API which is translatedbythemiddletierintotheappropriatelow-levelcalls.Finally,inmanycasesthethree- tier architecture can provide performance advantages.

Untilnowthe middletierhastypicallybeenwritteninlanguagessuchasCorC++, which offer fast performance. However, with the introduction of optimizing compilers that translate Java byte code into efficient machine-specific code, it is becoming practical to implement the middletier inJava. This is a big plus, making it possibletotakeadvantageofJava'srobustness, multithreading, and security features. JDBC is important to allow database access from a Java middle tier.

JDBCDriver Types

TheJDBCdriversthatweareawareofat thistimefitintooneoffourcategories:

* + JDBC-ODBCbridgeplusODBCdriver
  + Native-APIpartly-Javadriver
  + JDBC-Net pureJavadriver
  + Native-protocolpureJavadriver

JDBC-ODBCBridge

Ifpossible, use a Pure Java JDBC driver instead ofthe Bridge and an ODBC driver. This completelyeliminatestheclientconfigurationrequiredbyODBC.Italsoeliminatesthepotential thattheJavaVMcouldbecorruptedbyanerrorinthenativecodebroughtinbytheBridge(that is,theBridgenativelibrary,theODBCdrivermanagerlibrary,theODBCdriverlibrary,andthe database client library).

WhatIstheJDBC-ODBCBridge?

The JDBC-ODBC Bridge is a JDBC driver, which implements JDBC operations by translating them into ODBC operations. To ODBC it appears as a normal application program. The Bridge implements JDBC for any database for which an ODBC driver is available. The Bridge is implemented as the sun.jdbc.odbc Java package and contains a native library used to access ODBC. The Bridge is a joint development of Intersolv and JavaSoft.

Java ServerPages(JSP)

Java server Pages is a simple, yet powerful technology for creating and maintaining dynamic-contentwebpages.BasedontheJavaprogramminglanguage,JavaServerPagesoffers proven portability, open standards, and a mature re-usable component model .The Java Server Pages architecture enables the separation ofcontent generation fromcontent presentation. This separationnoteasesmaintenanceheadaches,italsoallowswebteammemberstofocusontheir

areasofexpertise.Now,webpagedesignercanconcentrateonlayout,andwebapplication designers on programming, with minimal concern about impacting each other’s work.

##### FeaturesofJSP

Portability:

Java Server Pages files can be run on anyweb server or web-enabled application server that provides support for them. Dubbed the JSP engine, this support involves recognition, translation, and management of the Java Server Page lifecycle and its interaction components.

Components

ItwasmentionedearlierthattheJavaServerPagesarchitecturecanincludereusableJava components.Thearchitecturealsoallowsfortheembeddingofascriptinglanguagedirectlyinto theJavaServer Pages file. ThecomponentscurrentsupportedincludeJavaBeans, andServlets.

Processing

A Java Server Pages file is essentially an HTML document with JSP scripting or tags. The Java Server Pages file has a JSP extensionto the server as a Java Server Pages file. Before the page is served, the Java Server Pages syntax is parsed and processed into a Servlet on the server side. The Servlet that is generated outputs realcontent in straight HTML for responding to the client.

Access Models:

AJavaServerPagesfile maybeaccessed inat leasttwo different ways.Aclient’srequest comesdirectlyintoaJavaServerPage.Inthisscenario,supposethepageaccessesreusableJava Bean components that perform particular well-defined computations like accessing a database. The result of the Beans computations, called result sets is stored within the Bean as properties. The page uses such Beans to generate dynamic content and present it back to the client.

Inbothofthe above cases, the page could also containany valid Java code. Java Server Pages architecture encourages separation of content from presentation.

Stepsinthe executionofa JSPApplication:

1. The client sends a request to the web server for a JSP file by giving the name of the JSP file within the form tag of a HTML page.
2. This request is transferred to the JavaWebServer. At the server side JavaWebServer receives the request and if it is a request for a jsp file server gives this request to the JSP engine.
3. JSPengine isprogramwhichcanunderstandsthetagsofthejspandthenit convertsthose tags into aServlet programand it is storedat the server side. This Servlet is loaded inthe memoryandthenitisexecutedandtheresult isgivenbacktotheJavaWebServerandthen it is transferred back to the result is given back to the JavaWebServer and then it is transferred back to the client.

##### JDBCconnectivity

The JDBC provides database-independent connectivity between the J2EE platformand a widerangeoftabulardatasources.JDBCtechnologyallowsanApplicationComponentProvider to:

* + Performconnectionandauthenticationtoadatabaseserver
  + Managertransactions
  + MoveSQLstatementstoadatabaseengineforpreprocessingand execution
  + Executestoredprocedures
  + Inspectandmodifytheresults fromSelectstatements.

##### Tomcat6.0webserver

Tomcat is anopensource webserver developed byApache Group. Apache Tomcat is the servlet container that is used in the official Reference Implementation for the Java Servlet and Java Server Pages technologies. The Java Servlet and Java Server Pages specifications are developedbySunundertheJavaCommunityProcess.WebServerslikeApacheTomcatsupport onlywebcomponents while anapplicationserver supports webcomponents as wellas business components (BEAs Weblogic, is one of the popular application server).To develop a web applicationwithjsp/servlet installanywebserver likeJRun,Tomcat etctorunyourapplication.

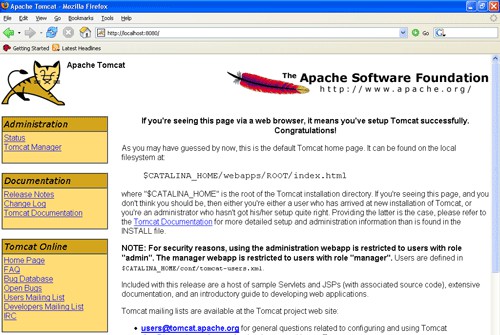


Fig:3.11

##### Bibliography:

ReferencesfortheProjectDevelopmentweretakenfromthefollowing BooksandWebSites.

Oracle

PL/SQLProgrammingbyScottUrman SQL complete reference by Livion JAVA Technologies

JAVACompleteReference

JavaScriptProgrammingbyYehudaShiran Mastering JAVA Security

JAVA2NetworkingbyPistoria JAVA Security by Scotl oaks Head First EJB Sierra Bates

J2EE ProfessionalbyShadabsiddiqui

JAVAserverpagesbyLarnePekowsley JAVA Server pages by Nick Todd HTML

HTMLBlackBookbyHolzner JDBC

JavaDatabaseProgrammingwithJDBCbyPatelmoss. Software Engineering by Roger Pressman

ALOGORITHMS:

##### ConvolutionalNeuralNetworks(CNN):

AConvolutionalNeuralNetwork(CNN) isatypeofdeeplearningalgorithmprimarily used for processing structured array data such as images. CNNs are highly effective for image recognition, classification, and analysis, thanks to their architecture, which mimics the connectivity pattern of neurons in the human brain. A typical CNN architecture consists of multiple layers: convolutional layers, pooling layers, and fully connected layers.

The convolutional layers apply a series of learnable filters to the input. Each filter activates certain features from the input. Pooling layers then reduce the spatial size of the convolved features, making the detectionoffeatures invariant to scale and orientationchanges. Finally, fully connected layers compile the data extracted by previous layers to form the final output predictions.

##### LongtermShortMemory(LSTM):

Long Short-Term Memory (LSTM) is a type of recurrent neural network (RNN) architecture used in the field of deep learning. LSTM networks are particularly well-suited for classifying,processing,andmakingpredictionsbasedontimeseriesdata,sincetherecanbelags ofunknowndurationbetweenimportant events inatimeseries. LSTMsweredevelopedtodeal with the vanishing gradient problem that can be encountered when training traditional RNNs. This issue arises as the gradient of the network's weights, used during training, becomes very small, effectively preventing the network from learning further.

An LSTM unit typically consists of a cell, an input gate, an output gate, and a forget gate. The cell remembers values over arbitrary time intervals and the three gates regulate the flow of information into and out of the cell.

**TECHNIQUES:**

##### CharacterBERT:

The CharacterBERT model's character-level analysis allows for a more nuanced understanding of language structure and syntax, which is particularly beneficial for languages with rich morphology or in scenarios where precise spelling may vary, such as user-generated contentorhistoricaltexts.ThiscapabilityalsoaidsinbettergeneralizationovertraditionalBERT models when faced with text from specific sectors that may not be well-represented in the training data of general language models.

Furthermore, CharacterBERT's architecture is designed to preserve the strengths of theoriginalBERT framework,maintainingdeepcontextuallearningandhighperformanceona range of NLP tasks, including sentiment analysis, named entity recognition, and question answering. This makes it a versatile and powerful tool for both research and practical applications in the field of natural language processing, where the ability to handle diverse linguistic phenomena at a granular level can significantly impact the effectiveness and adaptability of language models.

##### Hashing:

Hashingisaversatilecomputingtechniquethatinvolvesconvertingvariable-lengthinput data into a fixed-length numeric value, known as a hash code, through a hash function. This processisfundamentalinvariousapplicationssuchasdataretrieval,whereitpowershashtables to enable fast and efficient data access. Hash functions aim to minimize collisions—instances where different inputs produce the same output—and ensure that hash codes are distributed uniformly across the hash table.

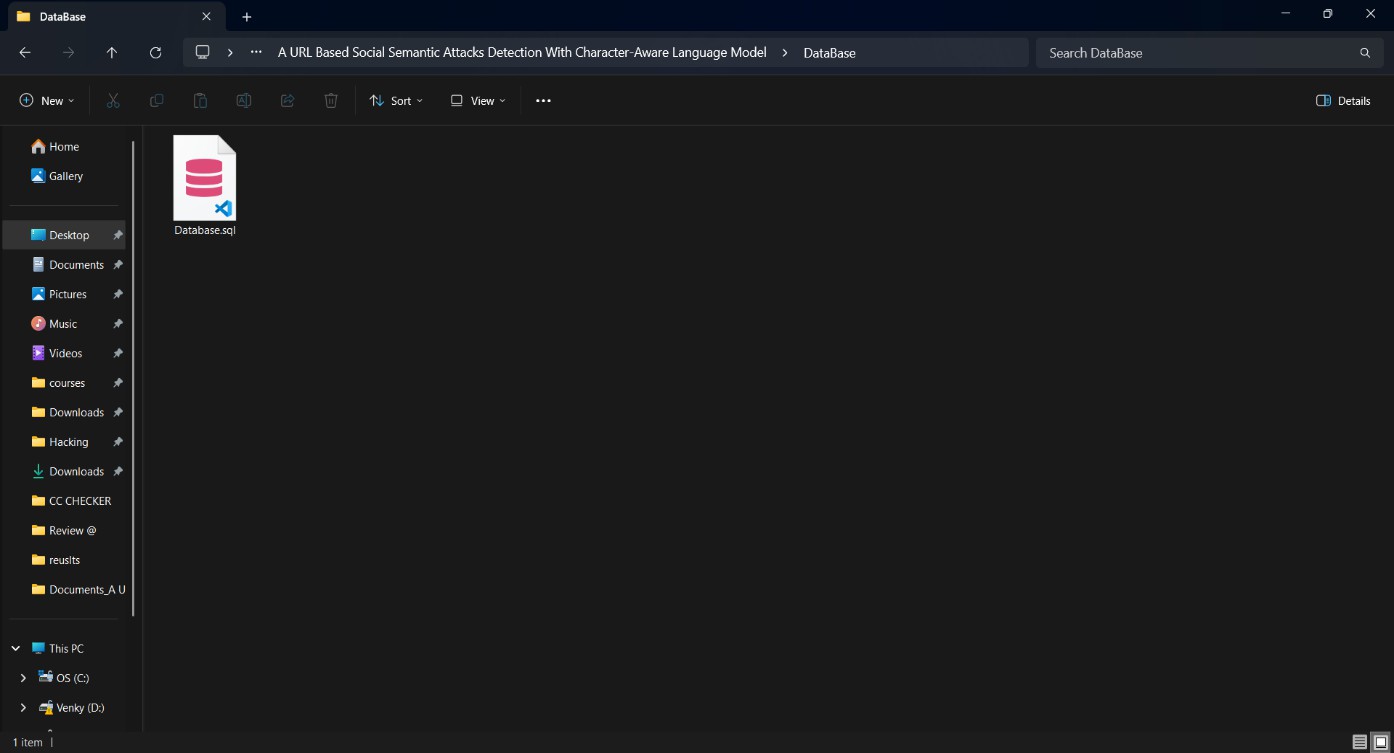
Intherealmofcryptography,hashingsecuresdataintegrity,aidsinthecreationofdigital signatures, and protects sensitive information. Cryptographic hash functions are designed to be bothcollision-resistantandirreversible,makingitchallengingtorecreatetheoriginalinputfrom its hashcode. Well-knownexamples includeSHA-256, whichiswidelyused for itsrobustness. Hashing's utility extends beyond security, facilitating tasks like load balancing and data fingerprinting by providing a quick way to compare large data sets.

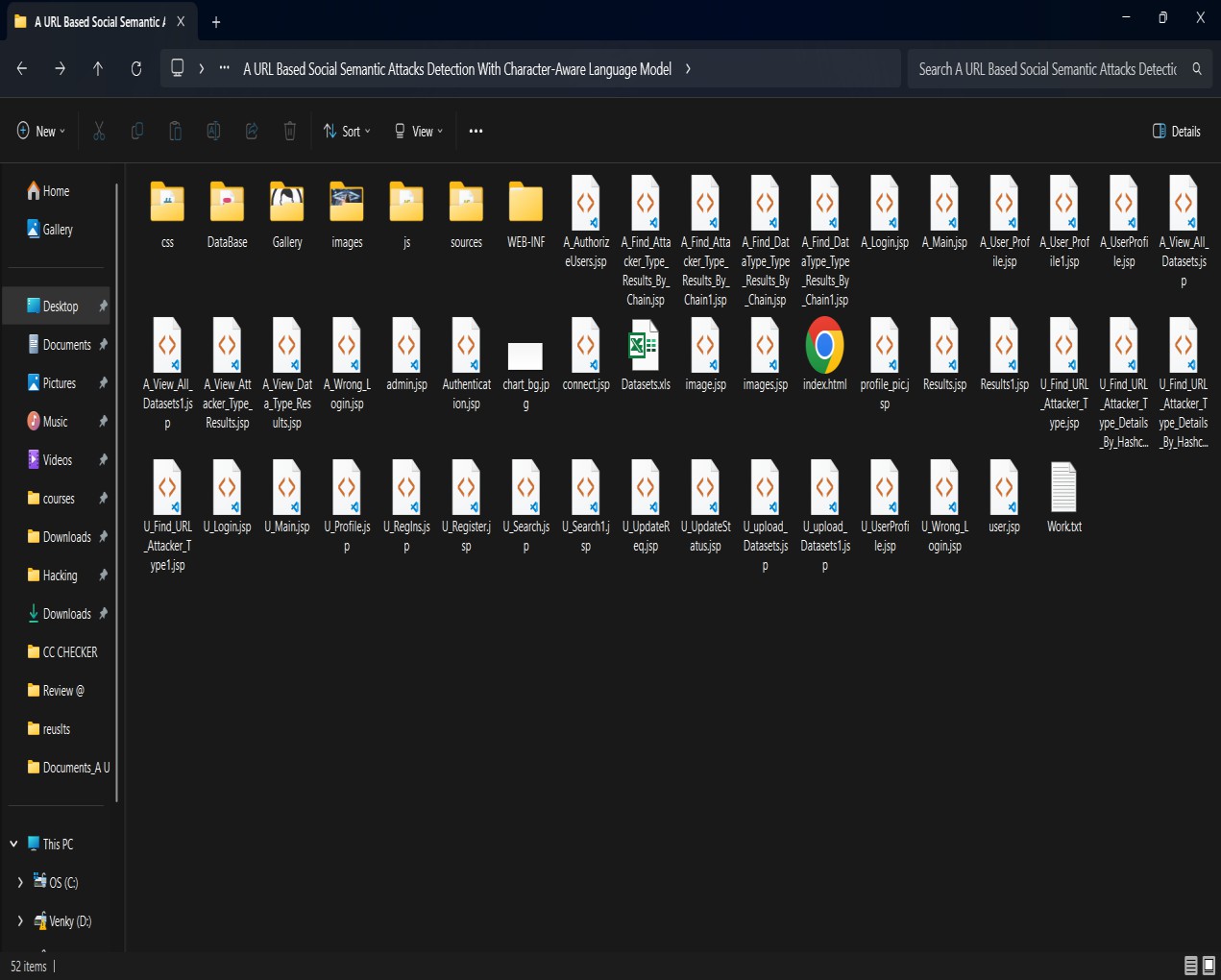
# CHAPTER-4: IMPLEMENTATIONOFTHEPROJECT

## CHAPTER-4

* 1. **IMPLEMENTATIONOFTHEPROJECT:**

### DataSetRequirements





##### ArchitectureDiagram:



Admin

AcceptingallInformation

Login,

DatasetsResultsStorage

ViewAll Users AndAuthorize,

Accessing

Data

View AllDatasets,

Find URLAttacker TypeResultsBy Chain,

Processall

userqueries

Find URLData TypeResults ByChain,

ViewURL Attacker Type Results,

**Storeandretrievals**

View URL Data Type Results.

**WEB**

**Database**

User

**WebServer**

RegisterandLogin,MyProfile,

UploadDatasets,

FindURLAttackerType,SearchData,

FindURLAttackerTypeDetailsByHashcode.

Fig:4.2

* 1. **INNOVATIONOFTHEPROJECT**

Millions of people all around the world are suffering from diabetes which is a prominent cause ofDiabetic Retinopathy. It is one ofthe leading causes of blindness and its effect can be reduced with proper treatment. So, it is necessary to build a system which can detect diabetic retinopathy without any expert guidance.

The inclusion of retinalsegmentation with region merging during preprocessing provides significantimprovementoverexistingmodel.Theproposedmodelprovidesresultasgoodasany supervised algorithm without explicitly determining features required for classification.

* 1. **PLANOFACTIONTOCOMPLETEPROJECT:**

###### WhatisSDLC?

SDLC stands for Software Development Life Cycle. A Software Development Life Cycle is essentially a series of steps, or phases, that provide a model for the development and lifecycle management of an application or piece of software.

SDLCistheprocessconsistingofaseriesofplannedactivitiestodevelopor alter thesoftwareproducts.

###### BenefitsoftheSDLCProcess

The intent ofa SDLC process it to help produce aproductthat is cost-efficient, effective, and of high quality. Once an application is created, the SDLC maps the proper deployment and decommissioning of the software once it becomes a legacy. The SDLC methodology usually contains the following stages: Analysis (requirements and design), construction, testing, release, and maintenance (response). Veracode makes it possible to integrate automated security testing into the SDLC process through use of its cloud based platform.

1. RequirementsGathering:

Inthisphasewegatheralltherequirementsfromtheclient,i.e.whataretheclientexpected input, output……

1. Analysis:

Inthisphasebasedupontheclient requirementsweprepareonedocumentationiscalled “High Level Design Document”. It contains Abstract, Functional Requirements, Non Functional Requirements, Existing System, Proposed System, SRS,………

1. Design:

It is difficult to understand the HighLevelDesign Document for allthe members, so to understand easilywe use “Low Level Design Document”. To design this document we use UML (Unified Modeling Language). In this we have Use case, Sequence, Collaboration……..

1. Coding:

Inthisphasewedevelopthecodingmodulebymodule.Afterdevelopingallthemodules we integrate them.

1. Testing:

Afterdevelopingwehavetocheckweatherclientrequirementsaresatisfiedornot.Ifnot we are again going to develop.

1. Implementation:

In testing phase if client requirements are satisfied, we go for implementation. i.e. weneed to deploy the application in some server.

1. Maintenance:

Afterdeployment,ifatallanyproblemscomefromtheclientside;weareproviding maintenance for that application.

# CHAPTER-5: SYSTEM STUDY

**CHAPTER-5**

## SYSTEMSTUDY

#### FEASIBILITYSTUDY

Thefeasibilityoftheproject isanalyzed inthisphaseandbusinessproposalisputforth with a very general plan for the project and some cost estimates. During system analysis the feasibilitystudyoftheproposedsystemistobecarriedout. This isto ensurethattheproposed system is not a burden to the company.For feasibility analysis, some understanding of the major requirements for the system is essential.

Threekeyconsiderationsinvolved inthefeasibilityanalysisare

* ECONOMICALFEASIBILITY
* TECHNICALFEASIBILITY
* SOCIALFEASIBILITY

#### ECONOMICALFEASIBILITY

This study is carried out to check the economic impact that the system will have on the organization. The amount of fund that the company can pour into the research and development ofthe system is limited. The expenditures must be justified. Thus the developed systemas well within the budget and this was achieved because most of the technologies used are freely available. Only the customized products had to be purchased.

#### TECHNICALFEASIBILITY

Thisstudyiscarriedouttocheckthetechnicalfeasibility,thatis,thetechnicalrequirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands onthe available technicalresources. This will lead to highdemandsbeingplacedontheclient.Thedevelopedsystemmusthaveamodestrequirement, as only minimal or null changes are required for implementing this system.

#### SOCIALFEASIBILITY

The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity. The level of acceptance by the users solelydepends onthe methods that are employed to educatethe user about the systemand tomakehimfamiliarwithit.Hislevelofconfidencemustberaisedsothatheisalsoabletomake some constructive criticism, which is welcomed, as he is the final user of the system.

# CHAPTER-6: SYSTEMDESIGN

## CHAPTER - 6 SYSTEMDESIGN

#### UMLDIAGRAMS

UML stands for Unified Modeling Language. UML is a standardized general-purpose modelinglanguageinthefieldofobject-orientedsoftwareengineering.Thestandardismanaged, and was created by, the Object Management Group.

ThegoalisforUMLtobecomeacommonlanguageforcreatingmodelsofobjectoriented computer software. In its current form UML is comprised of two major components: a Meta- model and a notation. In the future, some form of method or process may also be added to; or associated with, UML.

The Unified Modeling Language is a standard language for specifying, Visualization, Constructinganddocumentingtheartifactsofsoftwaresystem, aswellas forbusiness modeling and other non-software systems.

TheUMLrepresentsacollectionofbestengineeringpracticesthathaveprovensuccessful in the modeling of large and complex systems.

The UML is a very important part of developing objects oriented software and the software development process. The UML uses mostly graphical notations to express the design of software projects.

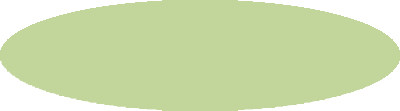
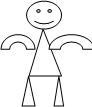
#### GOALS:

ThePrimarygoalsinthedesignoftheUMLareasfollows:

* + 1. Provideusersaready-to-use, expressivevisualmodelingLanguageso thattheycandevelop and exchange meaningful models.
    2. Provideextendibilityandspecializationmechanismstoextendthecore concepts.
    3. Beindependentofparticularprogramming languagesanddevelopmentprocess.
    4. Provideaformalbasisforunderstandingthemodelinglanguage.
    5. EncouragethegrowthofOO toolsmarket.
    6. Supporthigherleveldevelopmentconceptssuchascollaborations,frameworks,patternsand components.
    7. Integratebestpractices.

#### USECASEDIAGRAM:

A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overviewofthe functionalityprovided byasystemintermsofactors,their goals(represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.



Login,ViewAllUsersAndAuthorize,

View AllDatasets,

FindURLAttackerType Results

By Chain,

REGISTERANDLOGIN, My

Profile,

Admin

UploadDatasets,FindURL

AttackerType,

**User**

SearchData,FindURLAttacker

TypeDetailsByHashcode.

FindURLDataTypeResultsBy

Chain,

ViewURLAttackerTypeResults,

ViewURLDataType

Results

Fig:6.3

#### CLASSDIAGRAM:

In software engineering, a class diagram in the Unified Modeling Language (UML) is a type ofstatic structurediagramthat describes the structureofa systembyshowing the system's classes, their attributes, operations (or methods), and the relationships among the classes. It explains which class contains information.

Admin

Methods



rs

**Login**

Methods

**Register**

Methods Register(),Reset ()

UserName,Password,E-

Members

Members

mail,Mobile,Address,DOB,

Gender,Pincode,Image

RemoteUser

RegisterandLogin,MyProfile,UploadDatasets,FindURLAttacker Type, Search Data, Find URL Attacker Type Details By Hash code.

Methods

Members

Uid,Url,ranking,activeDuration,urlLen,attacktype,data type, A Hash code, D Hash code.

UserName,Password.

Login(),Reset(), Register ().

Uid,Url,ranking,activeDuration,urlLen,attacktype,datatype,A Hash code, D Hash code.

Login,ViewAll UsersAndAuthorize,ViewAllDatasets, FindURLAttackerType Results By Chain, Find URL Data Type Results By Chain,View URLAttacker Type Results, View URL Data Type Results.

Membe

Login,Register

UserName,Password

Fig:6.4

#### SEQUENCEDIAGRAM:

A sequence diagram in Unified Modeling Language (UML) is a kind of interaction diagramthat shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart. Sequence diagrams are sometimes called event diagrams, event scenarios, and timing diagrams.



Admin



RegisterandLogin,

Login,

MyProfile,

ViewAll UsersAndAuthorize,

UploadDatasets,

ViewAllDatasets,

Find URLAttackerType,

FindURLAttackerTypeResultsByChain,

SearchData,

FindURLData TypeResults ByChain,

ViewURLAttackerTypeResults,

Find URLAttackerTypeDetailsByHash

code.

ViewURLData TypeResults.

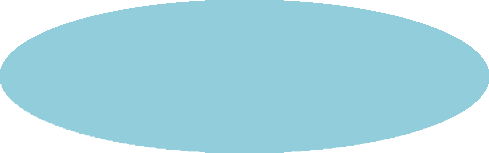
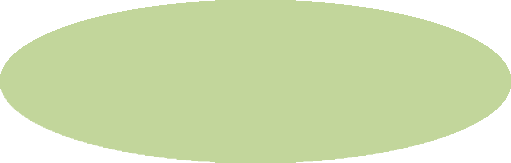
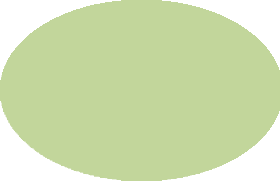
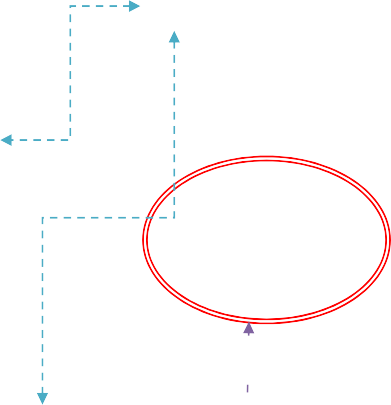
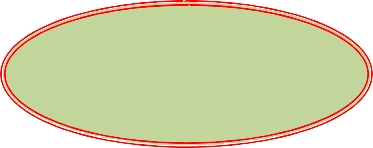
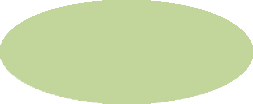
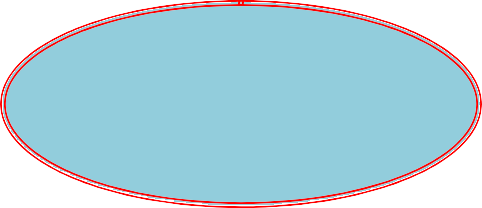
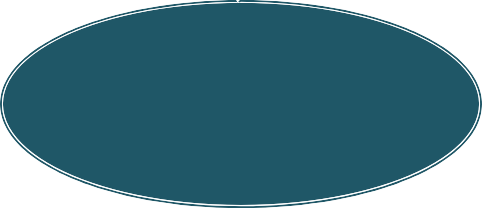
WebServer

User

Fig:6.5

#### DATAFLOWDIAGRAM:

Adataflowdiagram(DFD) illustratestheflowofinformationwithinasystem,depicting processes that manipulate data, data stores, and data flows between them. It provides a visual representation of how data moves through a system, showing inputs, outputs, and transformations. DFDs employ symbols like circles (representing processes), arrows (representing data flows), and rectangles (representing data stores) to convey the flow of data. Theyaid inunderstandingsystemfunctionality, identifyingpotentialbottlenecks,anddesigning efficient information architectures. DFDs are instrumental in system analysis and design, facilitatingcommunicationamongstakeholdersandensuringclarityinsystemdevelopment and maintenance processes.



Login,ViewAllUsersAnd

Authorize,ViewAllDatasets,

Admin

Login

System

RegisterandLogin,My

Profile,UploadDatasets,Find URL Attacker Type,

FindURLAttackerTypeResults

ByChain,FindURLDataType Results By Chain,

Response

Registerand

Loginwiththesystem

Request

SearchData,Find

URLAttackerType

DetailsByHashcode.

ViewURLAttackerType Results,

ViewURLDataTypeResults.

User

**Fig:6.6**

#### ALGORITHMS:

##### CNNWorkingProcedure

To demonstrate how to build a convolutional neural network based image classifier, we shall build a 7 layer neural network that will identifyand separate one image fromother. This networkthatweshallbuildisaverysmallnetworkthatwecanrunonaCPUaswell.Traditional neural networks that are very good at doing image classification have many more parameters andtakea lot oftimeiftrained onnormalCPU.However,ourobjectiveistoshowhowto build a real-world convolutional neural network using TENSORFLOW.

# CHAPTER-7: SYSTEMREQUIREMENTS

## CHAPTER – 7 SYSTEMREQUIREMENTS

#### HARDWAREREQUIREMENTS:

Minimumhardwarerequirementsareverydependentontheparticularsoftwarebeingdeveloped by a given Enthought Python / Canopy / VS Code user. Applications that need to store large arrays/objects in memory will require more RAM, whereas applications that need to perform numerous calculations or tasks more quickly will require a faster processor.

* + - Processor - Pentium–IV
    - RAM - 4GB(min)
    - HardDisk - 20GB

#### SOFTWAREREQUIREMENTS:

The appropriation of requirements and implementation constraints gives the general overviewoftheproject inregardstowhattheareasofstrengthanddeficit areandhowtotackle them.

* + - OperatingSystem - WindowsXP
    - CodingLanguage - Java/J2EE(JSP,Servlet)
    - FrontEnd - J2EE
    - BackEnd - MySQL

# CHAPTER-8: SYSTEMTESTING

## CHAPTER–8 SYSTEMTESTING

The purpose of testing is to discover errors. Testing is the process of trying to discover everyconceivable fault or weakness ina workproduct. It provides a wayto check the functionality of components, sub assemblies, assemblies and/or a finished product It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement.

#### TYPESOFTESTS

##### Unittesting

Unittestinginvolvesthedesignoftestcasesthatvalidatethattheinternalprogramlogic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application .it is done after the completion of an individual unit before integration. This is a structuraltesting,that reliesonknowledgeofitsconstructionand isinvasive.Unit testsperform basic tests at component level and test a specific business process, application, and/or system configuration. Unit testsensurethat eachuniquepathofabusinessprocessperformsaccurately to the documented specifications and contains clearly defined inputs and expected results.

##### Integrationtesting

Integration tests are designed to test integrated software components to determine if they actually run as one program.Testing is event driven and is more concerned with the basic outcome of screens or fields. Integration tests demonstrate that although the components were individuallysatisfaction, as shown bysuccessfully unit testing, the combinationofcomponents iscorrectandconsistent.Integrationtestingisspecificallyaimedatexposingtheproblemsthat arise from the combination of components.

##### Functionaltest

Functionaltestsprovidesystematicdemonstrationsthatfunctionstestedareavailableasspecified by the business and technical requirements, system documentation, and user manuals.

Functionaltestingiscenteredonthefollowingitems:

Valid Input :identified classes ofvalid input must be accepted. Invalid Input :identified classesofinvalid input must be rejected. Functions :identified functions must be exercised.

Output :identifiedclassesofapplicationoutputsmust be exercised. Systems/Procedures :interfacing systems or procedures must be invoked.

Organization and preparation of functional tests is focused on requirements, key functions, or specialtest cases. Inaddition, systematic coverage pertaining to identifyBusiness process flows; data fields, predefined processes, and successive processes must be considered fortesting.Beforefunctionaltesting iscomplete,additionaltestsareidentifiedandtheeffective value of current tests is determined.

##### SystemTest

Systemtesting ensures that the entire integrated software system meets requirements. It tests a configuration to ensure known and predictable results. An example of system testing is the configuration oriented system integration test. System testing is based on process descriptions and flows, emphasizing pre-driven process links and integration points.

##### WhiteBoxTesting

White Box Testing is a testing in which in which the software tester has knowledge of the inner workings, structure and language ofthe software,orat least its purpose. It is purpose. It is used to test areas that cannot be reached from a black box level.

##### BlackBoxTesting

BlackBoxTestingistestingthesoftwarewithoutanyknowledgeoftheinnerworkings, structure or language of the module being tested. Black box tests, as most other kinds of tests, must be written from a definitive source document, such as specification or requirements document, suchas specificationorrequirements document. It is a testing inwhichthe software under test is treated, as a black box .you cannot “see” into it. The test provides inputs and responds to outputs without considering how the software works.

##### UnitTesting

Unit testing is usually conducted as part of a combined code and unit test phase of the software lifecycle, although it is not uncommon for coding and unit testing to be conducted as two distinct phases.

##### Teststrategyandapproach

Fieldtestingwillbeperformedmanuallyandfunctionaltestswillbewrittenindetail.

Test objectives

* Allfieldentries mustworkproperly.
* Pagesmustbeactivatedfromtheidentifiedlink.
* Theentryscreen,messagesandresponsesmustnotbedelayed.

Featurestobe tested

* Verifythatthe entriesare ofthe correctformat
* No duplicateentriesshould beallowed
* Alllinksshouldtake theusertothe correctpage.

#### TESTINGMETHODOLOGIES

ThefollowingaretheTestingMethodologies:

* **UnitTesting.**
* **IntegrationTesting.**
* **UserAcceptanceTesting.**
* **OutputTesting.**
* **ValidationTesting.**

##### UnitTesting

UnittestingfocusesverificationeffortonthesmallestunitofSoftwaredesignthatisthemodule. Unit testingexercisesspecificpathsinamodule’scontrolstructuretoensurecompletecoverage and maximum error detection. This test focuses on each module individually, ensuring that it functions properly as a unit. Hence, the naming is Unit Testing.

Duringthistesting,eachmodule istestedindividuallyandthemodule interfacesareverified for the consistency with design specification. All important processing path are tested for the expected results. All error handling paths are also tested.

### IntegrationTesting

Integration testing addresses the issues associated with the dual problems of verification and program construction. After the software has been integrated a set of high order tests are conducted. The mainobjective inthis testing process is totakeunit tested modules and builds a program structure that has been dictated by design.

**ThefollowingarethetypesofIntegrationTesting:**

1)Top Down Integration

This methodisanincrementalapproachtotheconstructionofprogramstructure. Modulesare integrated bymoving downward throughthe controlhierarchy, beginning withthe main program module. The module subordinates to the main program module are incorporated

intothestructureineitheradepthfirst orbreadthfirstmanner.

Inthismethod,thesoftwareistestedfrommainmoduleandindividualstubsarereplaced when the test proceeds downwards.

1. **Bottom-up Integration**

This method begins the construction and testing with the modules at the lowest level in the program structure. Since the modules are integrated from the bottom up, processing required for modules subordinate to a given level is always available and the need for stubs is eliminated. The bottom up integration strategy may be implemented with the following steps:

* + The low-level modules are combined into clusters into clusters that perform a specificSoftware sub-function.
  + Adriver (i.e.)the controlprogramfortesting is writtento coordinatetest case input and output.
  + Theclusteristested.
  + Driversareremovedandclustersarecombined movingupward intheprogramstructure The bottom up approaches tests each module individually and then each module is module is integrated with a main module and tested for functionality.

## OTHERTESTINGMETHODOLOGIES

##### UserAcceptanceTesting

User Acceptance of a system is the key factor for the success of any system. The system under consideration is tested for user acceptance by constantly keeping in touch with the prospective system users at the time of developing and making changes wherever required. The system developedprovidesa friendlyuser interfacethat caneasilybeunderstoodevenbyapersonwho is new to the system.

##### OutputTesting

Afterperformingthevalidationtesting,thenextstepisoutputtestingoftheproposedsystem, sincenosystemcouldbeusefulifitdoesnotproducetherequiredoutputinthespecifiedformat. Asking the users about the format required by themtests the outputs generated or displayed by the system under consideration.Hence the output format is considered in 2 ways – one is on screen and another in printed format.

##### ValidationChecking

Validationchecksareperformedonthefollowingfields.

TextField:

The text field can contain only the number ofcharacters lesser than or equalto its size. The text fields are alphanumeric in some tables and alphabetic in other tables.Incorrect entry always flashes and error message.

NumericField:

The numeric field can contain only numbers from 0 to 9. An entry of any character flashes an error messages. Theindividual modules arecheckedfor accuracyandwhat it hastoperform.Eachmoduleis subjected to testrun along with sample data.The individually testedmodulesare integrated into a singlesystem.Testinginvolves executingthereal data informationis usedintheprogramthe existence of any program defect is inferred from the output.The testing should be planned sothat all the requirements are individually tested.

A successful test is one thatgives out the defects for the inappropriate data and produces and output revealing the errors in the system.

PreparationofTestData

Takingvariouskindsoftestdatadoestheabovetesting.Preparationoftestdataplaysavitalrole inthesystemtesting. Afterpreparingthetest datathesystemunderstudyistestedusingthattest data. While testing the system by using test data errors are again uncovered and corrected by using above testing steps and corrections are also noted for future use.

UsingLiveTest Data:

Live test data are those that are actually extracted from organization files. After a system is partiallyconstructed, programmers or analysts often ask users to key in a set of data fromtheir normalactivities.Then,thesystemspersonusesthisdataasawayto partiallytestthesystem.In other instances, programmers or analysts extract a set of live data fromthe files and have them entered themselves.

It is difficult to obtain live data in sufficient amounts to conduct extensive testing. And, although it is realistic data that will show how the system will perform for the typical processingrequirement,assumingthat thelivedata enteredareinfacttypical,such data

generally will not test all combinations or formats that can enter the system. This bias toward typical values then does not provide a true systems test and in fact ignores the cases most likely to cause system failure.

UsingArtificialTest Data:

Artificialtest dataarecreatedsolelyfortestpurposes,sincetheycanbegeneratedtotestall combinations of formats and values. In other words, the artificial data, which can quickly be prepared by a data generating utility program in the information systems department, make possible the testing of all login and control paths through the program.

The most effective test programs use artificial test data generated by persons other than those who wrotethe programs. Often, an independent teamoftesters formulates a testing plan, using the systems specifications.

The package “Virtual Private Network” has satisfied all the requirements specified as per software requirement specification and was accepted.

USERTRAINING

Whenever a new system is developed, user training is required to educate them about the working of the system so that it can be put to efficient use by those for whom the system has been primarily designed. For this purpose the normal working of the project was demonstrated to the prospective users. Its working is easily understandable and since the expected users are people who have good knowledge of computers, the use of this system is very easy.

MAINTAINENCE

Thiscoversawiderangeofactivitiesincludingcorrectingcodeanddesignerrors.Toreduce the need for maintenance in the long run, we have more accurately defined the user’s requirements during the process of system development. Depending on the requirements, this systemhas beendevelopedtosatisfythe needstothe largest possibleextent. Withdevelopment intechnology,itmaybepossibletoaddmanymorefeaturesbasedontherequirementsinfuture. Thecodinganddesigningissimpleandeasytounderstandwhichwillmakemaintenanceeasier.

#### TESTINGSTRATEGY:

A strategy for system testing integrates system test cases and design techniques into a well plannedseriesofstepsthatresultsinthesuccessfulconstructionofsoftware.Thetestingstrategy must co-operatetest planning, test case design, test execution, and theresultant data collection and evaluation .A strategy for software testingmustaccommodatelow-levelteststhat are necessaryto verifythat asmallsourcecodesegment hasbeencorrectlyimplementedaswell

as high level tests that validatemajorsystemfunctions against user requirements. Software testing is a critical element of software quality assurance and represents the ultimate review of specificationdesignandcoding.Testingrepresentsaninterestinganomalyforthesoftware.Thus, a seriesoftestingare performed for the proposedsystembefore the systemisready for user acceptance testing.

SYSTEMTESTING:

Software once validated must be combined with other systemelements (e.g. Hardware, people,database).Systemtestingverifiesthatalltheelementsareproperandthatoverallsystem functionperformance isachieved. It also teststo find discrepancies betweenthe systemand its original objective, currentspecifications and systemdocumentation.

UNITTESTING:

Inunittestingdifferent aremodulesaretestedagainstthespecificationsproducedduringthe designforthe modules. Unit testing isessentialfor verificationofthe codeproducedduring the coding phase, andhence the goals to testthe internallogicofthe modules. Using the detailed design descriptionas a guide, important Conrailpathsare tested to uncover errors within the boundaryofthe modules. Thistesting iscarriedout duringtheprogrammingstage itself. Inthis typeoftestingstep,eachmodulewasfoundtobeworkingsatisfactorilyasregardstotheexpected output from the module.

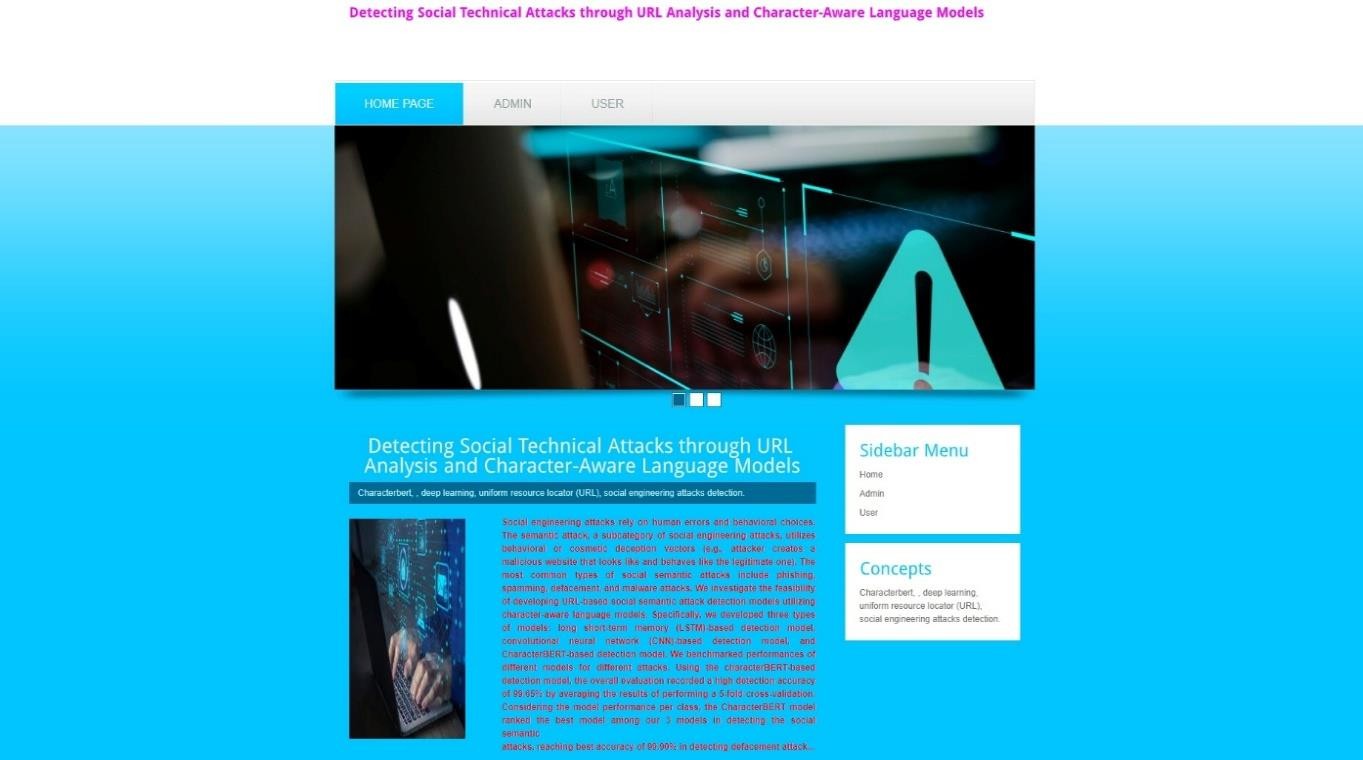
In Due Course, latest technologyadvancements will be taken into consideration. As part of technicalbuild-up many components ofthe networking systemwillbe generic in nature so that future projects can either use or interact with this.The future holds a lot to offer to the development and refinement of this project.

# CHAPTER-9: EXPERIMENT ANALYSIS AND RESULTS

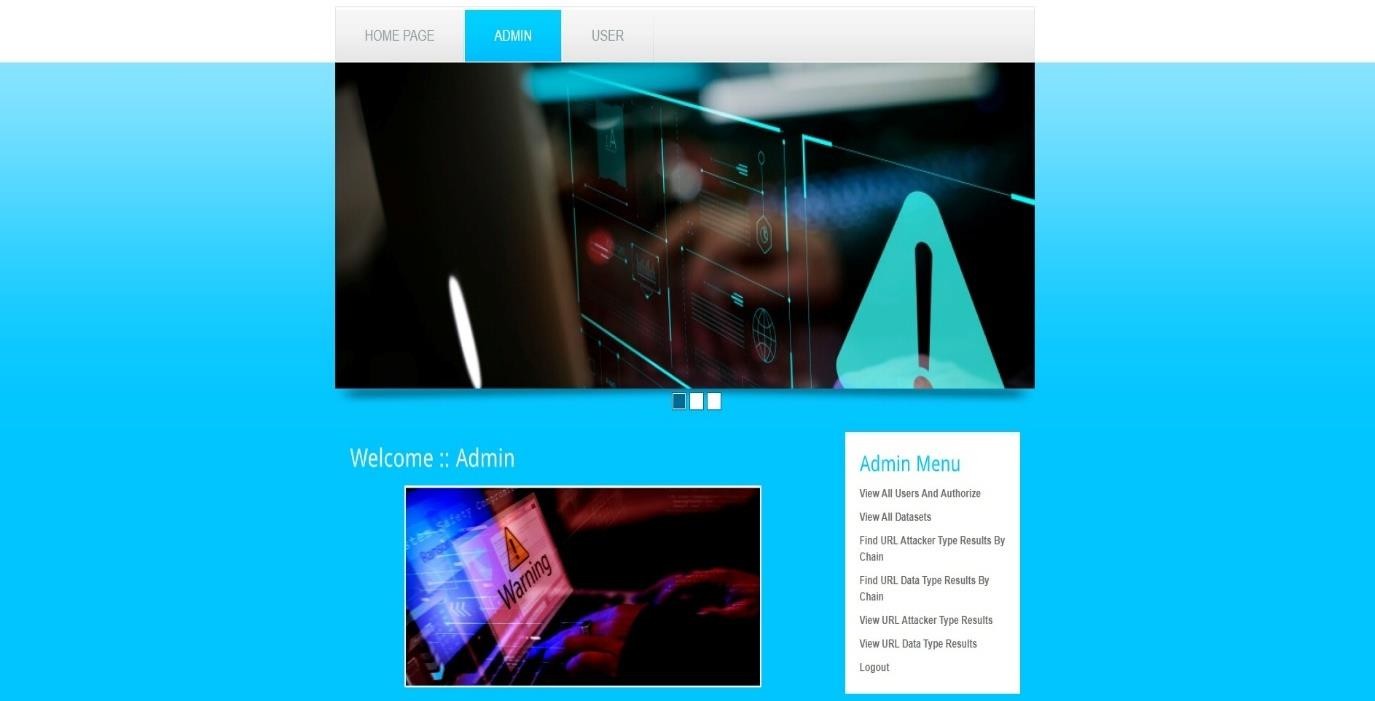
## CHAPTER - 9 EXPERIMENT ANALYSIS AND RESULT

* 1. SCREENSHOTS

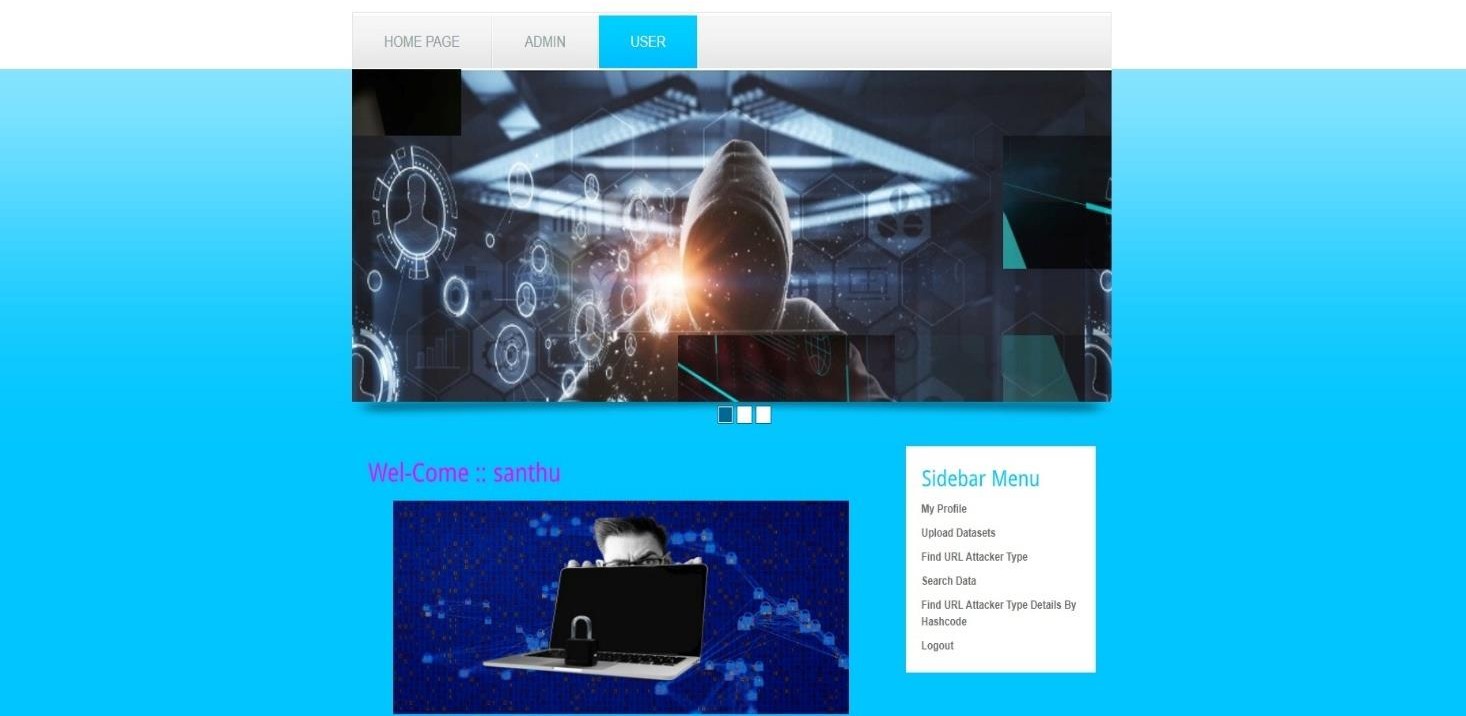
Home



Admin



User



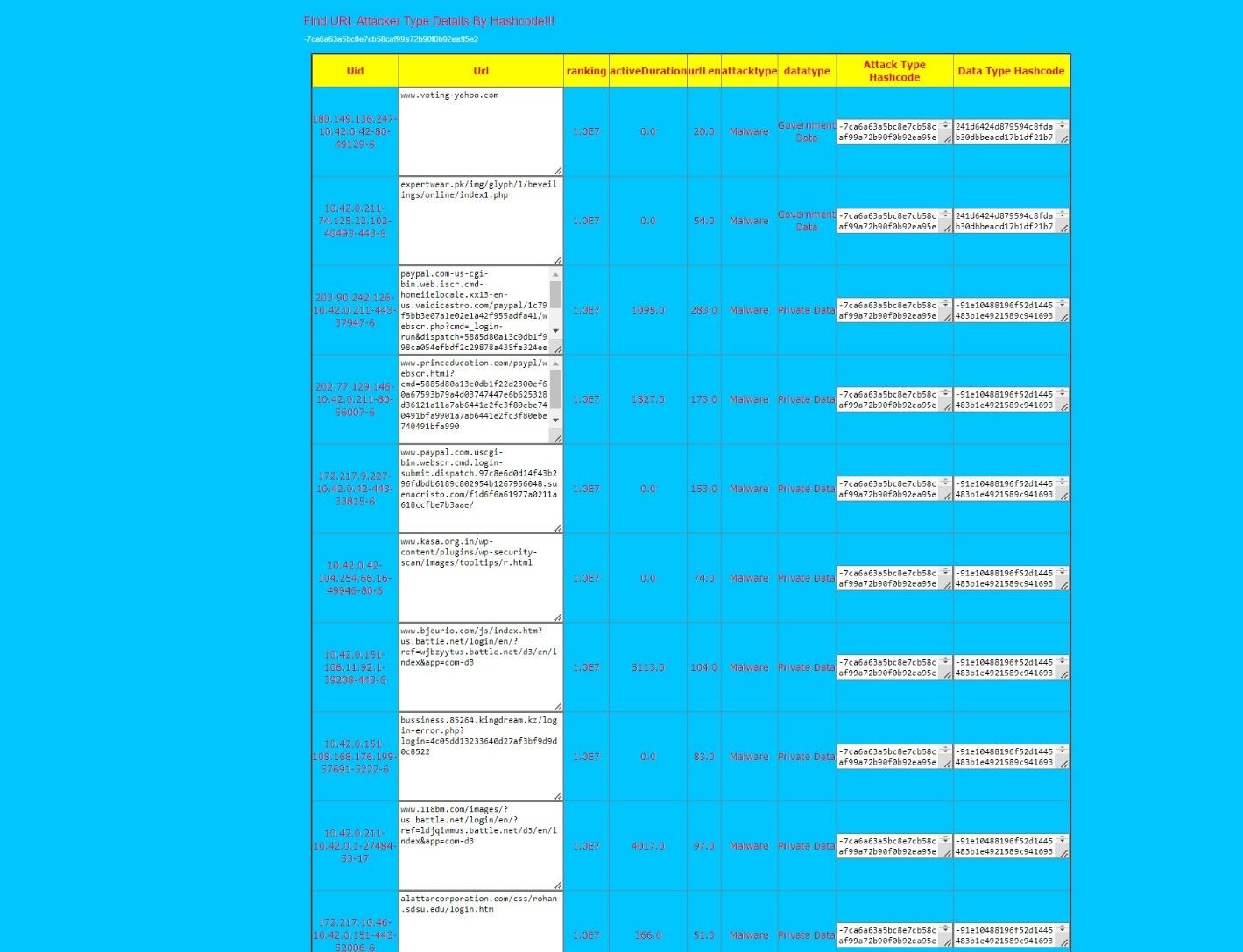
SearchData



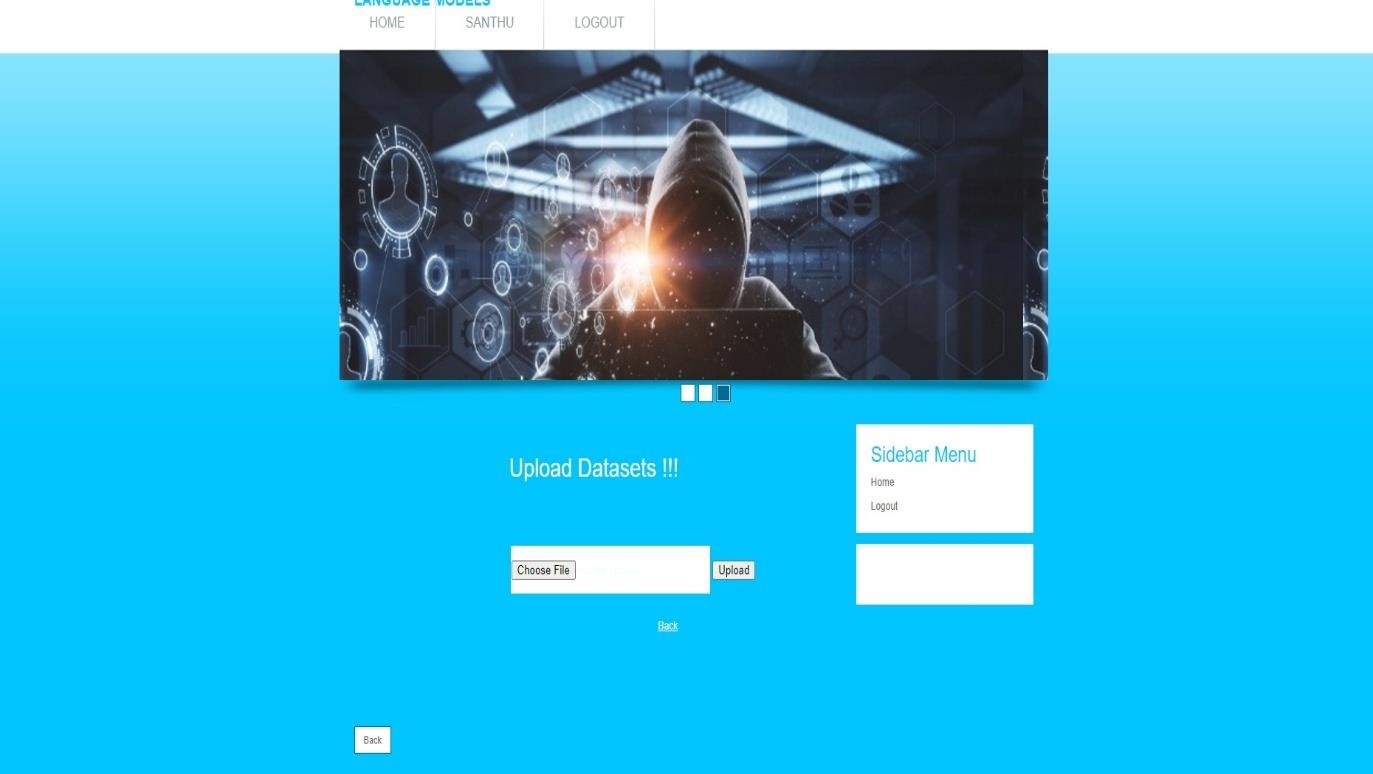
AllURLDataTyperesults



FindURLAttackertypeDetailsbyHashcode



**UploadDataset**



**AttackerTypeResults**



# CHAPTER-10: CONCLUSION

## CHAPTER-10

## CONCLUSION

Cyber security is an ever-changing field requiring vigilance and constant innovation to protect computer systems, networks, and data against cyber attacks and unauthorized access by unwantedparties.Thisresearchexaminesthemostcommontypesofsemanticattacks:phishing, spam, defacement, and malware, and details our findings after employing the Character BERT modelforsemanticattackdetectionbasedonURL. Thisstudycomparestheeffectivenessofthe Character BERT model to the LSTM model and the CNN model when detecting potential semantic attacks. The robustness of the Character BERT model enabled it to have a higher success rate in identifying social semantic attacks based on URLs, and for that reason, it outperformed the other models indetectingdifferent types ofsocialengineering attacks. Inthe future, we plan to consider moretypes ofsocialsemantic attacks. Moreover, we want to tryour detection models on other publicly-available social semantic attacks datasets to ensure their reproducibility.

# CHAPTER-11: REFERENCES

## CHAPTER–11 REFERENCES

1. S.AbrahamandI.Chengalur-Smith,‘‘Anoverviewofsocialengineering malware:Trends, tactics, and implications,’’ *Technol. Soc.*, vol. 32, no. 3, pp. 183–196, 2010.
2. N.Ahmed,R.Amin,H.Aldabbas,D.Koundal,B.Alouffi,andT.Shah,‘‘Machinelearning techniques for spamdetection inemailand IoTplatforms: Analysisand researchchallenges,’’ *Secur. Commun. Netw.*, vol. 2022, pp. 1–19, Feb. 2022.
3. A. Aleroud and L. Zhou, ‘‘Phishing environments, techniques, and countermeasures: A survey,’’ *Comput. Secur.*, vol. 68, pp. 160–196, Jul. 2017.
4. A. Aljofey, Q. Jiang, A. Rasool, H. Chen, W. Liu, Q. Qu, and Y. Wang, ‘‘An effective detectionapproach for phishing websites using URL and HTML features,’’ *Sci. Rep.*, vol. 12, no. 1, pp. 1–19, May 2022.
5. M. Almousa and M. Anwar, ‘‘Detecting exploit websites using browser-based predictive analytics,’’ in *Proc. 17th Int. Conf. Privacy, Secur. Trust (PST)*, Aug. 2019, pp. 1–3.
6. M.Almousa,T.Zhang,A.Sarrafzadeh,andM.Anwar,‘‘Phishingwebsite detection:How effective are deep learning-based models and hyperparameter optimization?’’ *Secur. Privacy*, vol. 5, no. 6, p. e256, Nov. 2022.
7. A.Altaher,‘‘PhishingwebsitesclassificationusinghybridSVMandKNNapproach,’’*Int.*

*J. Adv.Comput.Sci.Appl.*,vol.8,no.6,pp.1–6, 2017.

1. A. Barth, J.Caballero,and D.Song, ‘‘Secure content sniffing forweb browsers,orhowto stoppapersfromreviewingthemselves,’’in*Proc. 30th IEEE Symp. Secur.Privacy*, May2009,

pp.360–371.

1. K.Borgolte,C.Kruegel,andG.Vigna,‘‘Meerkat:Detectingwebsite defacementsthrough image-basedobjectrecognition,’’ in*Proc.24thUSENIXSecur.Symp.(USENIXSecurity)*,2015,

pp.595–610.

1. H. El Boukkouri, O. Ferret, T. Lavergne, H. Noji, P. Zweigenbaum, and J. Tsujii, ‘‘CharacterBERT: Reconciling ELMo and BERT for word-level open-vocabulary representations from characters,’’ 2020, *arXiv:2010.10392*.
2. K. Clark, M.-T. Luong, Q. V. Le, and C. D. Manning, ‘‘ELECTRA: Pretraining text encoders as discriminators rather than generators,’’ 2020, *arXiv:2003.10555*.
3. G. Davanzo, E. Medvet, and A. Bartoli, ‘‘Anomaly detection techniques for a web defacement monitoring service,’’ *Expert Syst. Appl.*, vol. 38, no. 10, pp. 12521–12530, Sep. 2011.
4. J.Devlin,M.-W.Chang,K.Lee,andK.Toutanova,‘‘BERT:Pre-trainingofdeep

bidirectionaltransformersforlanguageunderstanding,’’2018,*arXiv:1810.04805*.

1. R. Dhamija, J. D.Tygar, and M. Hearst,‘‘Whyphishing works,’’ in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, Apr. 2006, pp. 581–590.
2. M.Egele,P.Wurzinger,C.Kruegel,andE.Kirda,‘‘Defendingbrowsers againstdrive-by downloads: Mitigatingheapsprayingcode injectionattacks,’’ in*Proc. 6thInt.Conf.Detection IntrusionsMalware, Vulnerability Assessment(DIMVA)*.Como,Italy:Springer, Jul. 2009,pp. 88–106.
3. R.K.Gurjwar,D.R.Sahu,andD.S.Tomar,‘‘Anapproachtorevealwebsitedefacement,’’

*Int.J. Comput.Sci.Inf. Secur.*, vol. 11,no. 6, p. 73,2013.

1. T. S. Guzella and W. M. Caminhas, ‘‘A review of machine learning approaches to spam filtering,’’ *Expert Syst. Appl.*, vol. 36, no. 7, pp. 10206–10222, Sep. 2009.
2. K. Haynes, H. Shirazi, and I. Ray, ‘‘Lightweight URL-based phishing detection using naturallanguageprocessingtransformersformobiledevices,’’*Proc.Comput.Sci.*,vol.191,

pp.127–134, Jan. 2021.

1. R. Heartfield and G. Loukas, ‘‘Detecting semantic social engineering attacks with the weakest link: Implementation and empirical evaluation of a human-as—A-security-sensor framework,’’ *Comput. Secur.*, vol. 76, pp. 101–127, Jul. 2018.
2. X. D. Hoang and N. T. Nguyen, ‘‘Detecting website defacements based on machine learning techniques and attack signatures,’’ *Computers*, vol. 8, no. 2, p. 35, May2019.
3. C. Jackson, D.R. Simon, D.S.Tan, andA. Barth, ‘‘Anevaluationof extended validation and picture-in-picture phishing attacks,’’ in *11th Int. Conf. Financial Cryptogr. Data Secur., 1stInt.WorkshopUsableSecur. (FC/USEC)*.Scarborough,ON,Canada:Springer, Feb. 2007,

pp.281–293.

1. K. Jayaraman, W. Du, B. Rajagopalan, and S. J. Chapin, ‘‘ESCUDO: A fine-grained protection model for web browsers,’’ in *Proc.*
2. M.I.JordanandT.M.Mitchell,‘‘Machinelearning:Trends,perspectives,andprospects,’’

*Science*,vol.349,no.6245,pp.255–260,Jul.2015.

1. R. Jozefowicz, O.Vinyals, M. Schuster, N. Shazeer, andY.Wu, ‘‘Exploring the limits of language modeling,’’ 2016, *arXiv:1602.02410*.
2. C. Karlof, U. Shankar, J. D. Tygar, and D. Wagner, ‘‘Dynamic pharming attacks and locked same-origin policies for web browsers,’’ in *Proc. 14th ACM Conf. Comput. Commun. Secur.*, Oct. 2007, pp. 58–71.