ISSN:0975-9646

Intrusion Detection Using Data Mining Technique (Classification)

Dr.D.Aruna Kumari Phd[1] N.Tejeswani[2] G.Sravani[3] R.Phani Krishna[4]

[1]*Associative professor, K L University,Guntur(dt) ,*
[2]*B.Tech(1V/1V),ECM, K L University,Guntur(dt),*
[3]*B.Tech(1V/1V), ECM, K L University,Guntur(dt),*
[4]*B.Tech(1V/1V),ECM, K L University,Guntur(dt) ,*

Abstract: **This paper analysis and criticizes the way of using, functioning the intrusion detection system in data mining. Understanding the techniques and characteristics in IDS can improve the idealogy . We need to understand the methods and types involved in IDS and in which way they operate when an attack occurred. Intrusion Detection System (IDS) which is the increasingly key element of systems security is used to identify the malicious activities in a computer system and network. There are different approaches in IDS systems. An intrusion detection system (IDS) checks all the results in network activity and identifies suspicious patterns that may indicate network attack or system attacks from someone attempting to make threats or compromise a system. There IDS is being classified by .Misuse detection ,Anomaly detection.The methods and the knowledge used in techniques of intrusion detection is used to detect the threats occurred in many areas like banks, etc.**

Keywords*: *Data mining,intusion detection,preprocessing,classification.

1.INTRODUCTION:

1.1. DATAMINING: Data mining is an annalistic process designed to explore large amounts of meaningful data bases or to generate new patterns from the existing old raw data, in search of computing systematic relationships between variables .The unique processing goal in data mining process is automatic and semi automatic report of large amount of data is to extract the new information and meaningful patterns from the data set and transforms it into the understandable structure for further use. Data Mining is also known as analysis of Knowledge Discovery of Databases(KDD PROCESS), Misnomer, Buzzword. KDD is the process in which data is selected, preprocessing, transforming, data mining, Interpretation/extraction, cleaned, and finally knowledge is obtained . It is mostly used to obtain more absolute accurate results. Early methods used to identify the patterns in data is "BAYES THEOREM & REGRESSION ANALYSIS".

However ,Data Mining endure, in many variations in this theme, such as the cross industry standard process for data mining the six phases involved in it are:

1.Business Understanding
2.Data Understanding
3.Data Preparation
4.Modeling
5.Evaluation
6.Deployment

Data Mining also has six common classes. They are :

1.Anomaly Detection
2.Association Rule learning
3.classification
4.Regression
5.Clustering
6.Summarization

1.2 INTRUSION DETECTION: Intrusion Detection system is an essential task done for securing management systems for computers and networks that tried to detect the different attempts held by users and who are attempting to break into or misusing our systems. There can be many types of intruders they may be from outside the network or illegal users of network.

1.2.1. There are many approaches to solutions. They are:

1. Signature based
2.Anomaly based
3. Host based
4. Stack based
5. Network based

1.Signature Based: This possess an attack description that can be matched to sense activities. Most signature based analysis system are simple pattern matching system. In this system there are many advantages and disadvantages.

Some of the drawbacks in this system is:

1. They can not detect the novel attacks.
2. Suffer due to false alarms
3. Have to be programmed for every new attacks.

Advantages:

1.Simple to implement
2.light weight
3.low false positive rate

2.Anomaly Based: It observes the normal use of network as noise characterization which is distinct from noise is assumed as intrusion. There are also advantages and disadvantages in this system.

Some disadvantages are: Intrusions are accompanied by manifestations that are sufficiently unusual and raises the false alarm and compromise the effectiveness of the intrusion detection system.

3.Host based: Host operating system also known as application logs in audit information. This includes event like identification and authentication, file opens and program and then analyzed to detect trails.

4. Stack Based: These are integrated with TCP / IP stack which allows packets to be watched and traversed. Tcp/ip allows the IDS to pull packets from the stack before the OS or the application have the chance to preprocess the data in packets.

5.Network Based: Network based system looks for the attacks in the analysis data and signatures checks in network traffic. A filter is used to identify which system is used to discard or paused. It helps to filter out un-malicious activities.

1.3. FUNCTIONS OF INTRUSION DETECTION:

- It Monitors and analyzes both user and system activities
- Analyzes the system configurations and vulnerabilities
- Assesses the system and file integrity
- Has the Ability to recognize the patterns typical of attacks
- It Analysis the abnormal activity patterns
- Tracks the user policy violations

2.EXISTING WORK:

2.1. Data Preprocessing:

Data Preprocessing in Data mining is defined as lacking of the attribute values, lacking in certain attributes of interest, or contains only aggregate data.

Data Preprocessing is defined as the data is Incomplete, noisy, inconsistent.

2.1.1. Incomplete data comes from:

The Data is "Not applicable" valued data when the value is being collected. There are different considerations between the time and the data that had been collected and it is being analyzed.

2.1.2. Noisy data (incorrect values) may come from:

Noisy data comes from the Faulty data collection instruments There are many Human or computer errors at data entry this are Due to Errors in data transmission.

2.2. Classification: Classification predicts categorical class labels.

Classification classifies data (constructs a model) based on the training set and the values (class labels) in a classifying attribute and uses it in classifying new data.

2.2.1. Major Issues regarding classification:

- Data cleaning: Data cleaning is the Preprocess in which data is held in order to reduce noise and handle missing values
- Relevance analysis :This analysis Removes the irrelevant or redundant attributes contained in the data.
- Data transformation: This Generalizes and/or normalizes data.
- Speed and scalability: It checks The time to construct the model , and the time taken to use the model is called speed
- Robustness: Robustness handles the noise and the missing values
- Scalability: The scalability is that the Efficiency in disk-resident databases
- Interpretability: To understand and to insight the model

2.3. CLASSIFICATION TECHNIQUE (BAYESIAN CLASSIFICATION):

The classifier performs chance of prediction, and it predicts the class membership probabilities in the Bayesian Classification:

It consists of:

- Foundation: It is Based on the Bayes Theorem.
- Performance: Bayesian classifier, *naïve Bayesian classifier*, it compares the performance with decision tree and selects the neural network classifiers.
- Incremental: Training example can incrementally increase/decrease the probability that the hypothesis is correct -prior knowledge can be combined with the observed data.
- Standard: Even when the Bayesian methods are computationally intractable, they provide the standard of optimal decision making against which other methods can also be measured.

3. DESCRIPTION / METHODOLGY:

- In Our project it mainly describes about how the users are going to register and can login into our bank website and then by using the login details of the user "we can identify who are the Intruders Entering into our bank for misusing the Details of the bank
- In this project we Mainly provide how to analyze the data and classify the data in data mining using weka tool.
- We had created the html pages for our bank website and it named as "SPS BANK".
- We created the database for storing the users details and maintain their accounts and also we can update their databases.
- By using "WEKA" tool which is the most powerful data mining tool for analyzing the data (preprocessing) the data and further classification of data is done.
- In the Preprocessing of data "The attributes in database appear in the weka tool and analysis of data is shown in the graphical format" and their data sets are displayed in that format.
- In the classification step, the data is being classified using " bayesian algorithm" by declaring the folders and instances of the data present in the database and now classification is done and the data is represented in tree format .
- By using the net beans we had connected the client and server(databases with html pages) front end as Jsp and Back end as Mysql and also we created databases to update the users registration details.
- In this project we mainly classify whether the loan should be given to the correct user if the user is in "safe or risky" based on the details provided by the user.

3.1. ABOUT WEKA:

"WEKA" tool contains the collection of visualization tools and inbuilt algorithms for data analysis and predicts the modeling, together with graphical user interfaces for easy access to the functionality.

3.1.1. Advantages of Weka include:

This WEKA Tool has its Free availability under the GNU General Public License policy.

This WEKA supports Portability, because it is fully implemented in the Java programming language and thus maintained in almost any modern computing platform.

This is one of the Comprehensive data collection in data preprocessing and modeling techniques.

Our Poject Execution Using Weka:

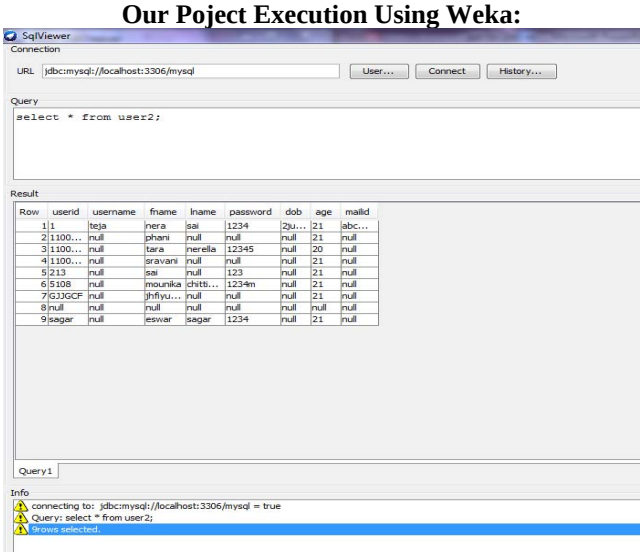


Fig 1.1 Connection of Weka To Data Base

This above fig explains about the connection of the most powerful data mining WEKA tool to database which consists the tables about the user and login details.

Data preprocessing using weka:

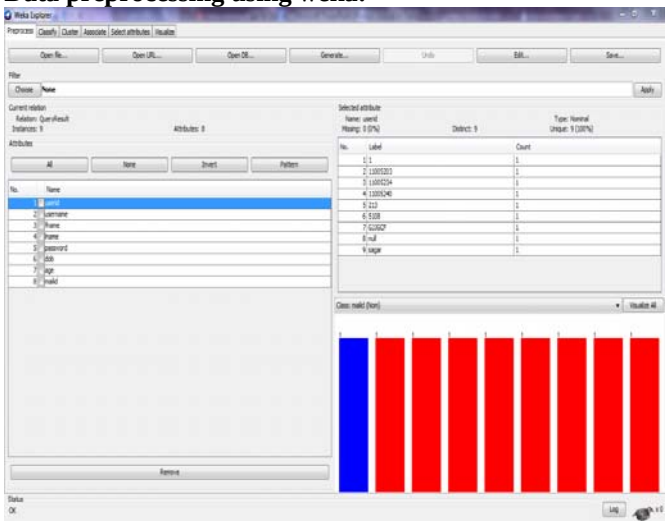


Fig 1.2: Data Preproccesing Using Weka

This above figure describes about the data preprocessing in WEKA tool, in data preprocessing the data is being analyzed.

Data classification using weka tool:

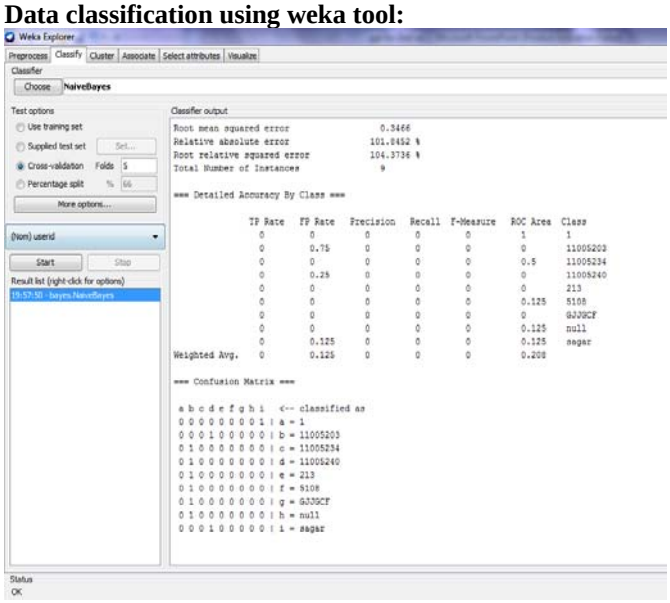


Fig 1.3 : Data Classification Using Weka

This above figure describes about the classification done in WEKA tool, in classification the data is being classified and the data set is displayed.

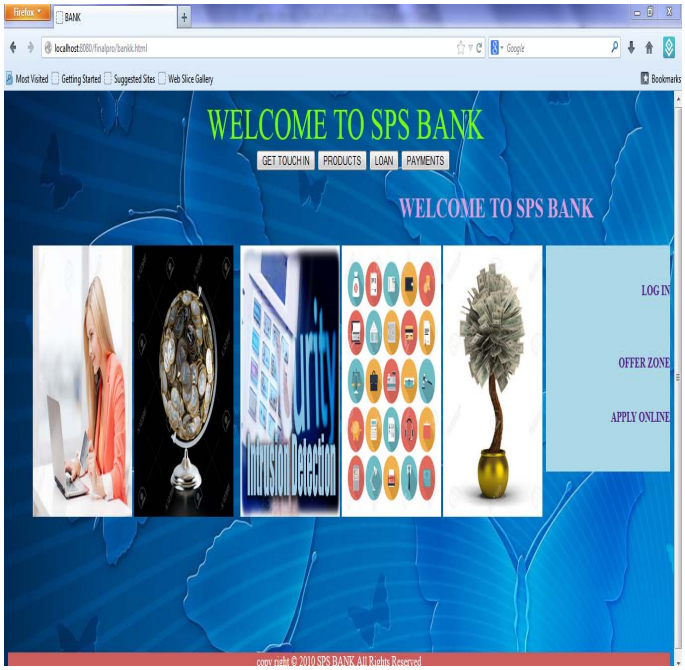
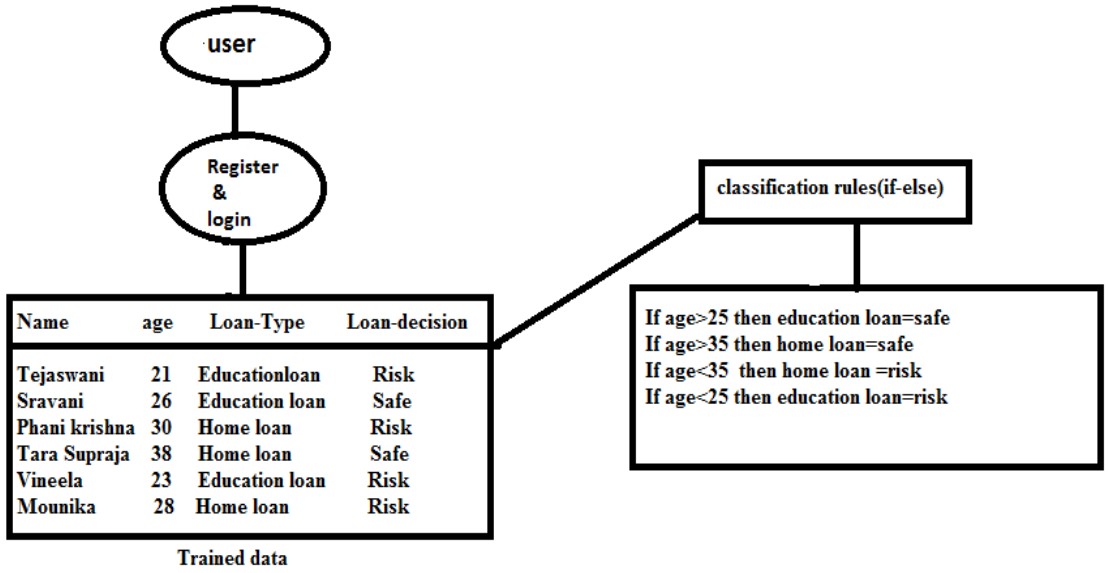


Fig 1.4: Snapshot of our SPS Bank Website

This figure describes about the front end of our SPS bank in which it describes and displays about our bank pages.

3.2. Historical If-Else Rules:

Block Diagram:



Name	age	Loan-Type	Loan-decision
Tejaswani	21	Educationloan	Risk
Sravani	26	Education loan	Safe
Phani krishna	30	Home loan	Risk
Tara Supraja	38	Home loan	Safe
Vineela	23	Education loan	Risk
Mounika	28	Home loan	Risk

Fig: Based on if-else rules

This figure describes about how the data is being classified by using historical If-else rules. By using this if rules accordingly.

If the age < 25 then the education loan decision is in risk ,else if the age>25 then the education loan decision is in safe.

If the age>35 then the home loan decision is in safe, else if the age<35 then home loan is in risk

4.CONCLUSIONS:

I hope this paper convinced about the data mining and intrusion detection system, IDS is a necessary tool in any environment. IDS requires a lot of planning and research. Once the research is done correctly there would be a lots of benefit. Though the need for computer security in both public and private is obvious Intrusion detection system provides the crucial component of the effective computer system. The work done in this paper is to understand the types in intrusion detection and how to provide security for the system. We use any one of the powerful tool in data mining and then provide security to highest level. In this paper we used one of the powerful tool named WEKA and classified the data. Although we have concentrated more on intrusion detection types it provides the security for computer network system. A total network security or computer security is a paradigm it can be viewed as kind of asymptote. And finally the security for bank whether the loan can be given or not is provided in this paper .

REFERENCES:

1. "**An overview of database centered intrusion detection system**" published in "International journal of engineering and advanced technology (IJEAT)" by "Ajayi adebowale,idowu S.A,otusile oluwabukola"
2. "**Network intrusion detection system using data mining and network behaviour analysis** " published in "International journal of computer science and information technology (IJCSIT)" by "Ahmed Youssef and ahmed emam"
3. **"Novel method for intrusion detection in datamining"** published in "International journal of advanced research in computer science and engineering" **by "**sherish johri"
4. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.46.4991&rep=rep1&type=pdf
5. http://dl.acm.org/citation.cfm?id=382923.
6. http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=1702202&abstractAccess=no&userType=inst.
7. http://www.techrepublic.com/resource-library/whitepapers/a-java-based-network-intrusion-detection-system-ids/.
8. https://www.usenix.org/legacy/publications/library/proceedings/sec98/full_papers/full_papers/lee/lee_html/lee.html.
9. http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=766909&abstractAccess=no&userType=inst.
10. http://neuro.bstu.by/ai/To-dom/My_research/Paper-0-again/For-research/D-mining/Anomaly-D/Fuzzy-by-GA/005slide.pdf.
11. http://en.wikipedia.org/wiki/Intrusion_detection_system.
12. http://www.sciencedirect.com/science/article/pii/S1110866513000418.
13. https://www.cs.sfu.ca/~jpei/publications/idmining-icde04.pdf.
14. https://www.google.co.in/url?sa=t&rct=j&q=&esrc=s&source=web&cd=7&cad=rja&uact=8&ved=0CFAQFjAG&url=http%3A%2F%2Fwww.it.iitb.ac.in%2F~praj%2Facads%2Fdm%2FIntrusion_Detection_Using_Data_Mining.ppt&ei=Rrv-VPXWGNecugSHvoLIDg&usg=AFQjCNHZ8yxc-G5ysMNMN0k5m7nUOV5kMg&sig2=ii5iNOXn-ff6tS5PCjP-hA&bvm=bv.87920726,d.c2E.

AUTHORS

I Dr..D.Aruna kumari published many papers of atmost of 50 papers in different journals and conferences related to web technologies stream ,completed the Phd and presuming as the associative professor in Kluniversity situated in vaddeswaram in Guntur dt

I N.Tejeswani published the 2 papers in web technology stream and presuming as B.Tech final year student in K L University

I G.Sravani published the 2 papers in webtechnology streams and continuing as student of B.Tech final year in K L University

I R.Phani Krishna is the student of B.Tech final year in K L University