IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

AN ANOMALY BASED INSIDER THREAT DETECTION FROM CYBER ACTIVITIES

[1]Kamatchi K, [2]Vinoth Kumar Y, [3]Dr. E. Uma

[1]PG Student, [2]PG Student, [3]Assistant Professor (SL. Gr)

[1]Department of Information Technology,

[1]College of Engineering Guindy, Chennai, India

Abstract: The threat of penetrating a company's end result in a collection of actions aimed at violently misusing systems, networks, data and resources. Preventing internal threats is not difficult, as the trusted partner of the concerned company, who allows access to those confidential services. Progressive analysis in identifying internal threat specifically focuses on retrieving methods of uncontrolled behavioral injury for detective task discomfort or abnormal changes in user behavior over time. However, the unpleasant task is that one does not earn honestly in a threatening environment. We propose an improvement logging strategy, a technology that detects internal threats from the user activity chain. Initially, a single-day selection area unit calculated from the user activity log. The array of feature attribute selectors is predefined over the data for each feature over time. Each label sets the data attribute and extracts from the truth below. The XG-boost classifier is used to classify whether the user is malicious or not, and its performance has been compared to various popular methods such as random forest. The CMU Insider Threat Dataset is only used for internal threat notification, providing approximately 14-GB browsing logs for the general public such as logon, application connection, file transfer and e-mail log files.

Index Terms **– Time Series Vector, XG-Boost, Malicious, Non-Malicious.**

I. INTRODUCTION

Internal users are trusted partners of the company, which grant them access to the company's resources, information and networks. In unsolicited cases, these people get out of anger and their greed, their work or their own benefit. This can cause them to misbehave and they can damage the company's resources and underestimate their name. In accordance with IBM 2016, IBM's internet cyber security index survey, which deals with threats, occurred more than an hour in isolation by intruders, in all 2015 reported security breaches or attacks worldwide. Therefore, preventing internal threats is an important issue. Internal threats are widely regarded as a challenge faced by large businesses. However, what we get from a wide variety of threats is the most important thing the company's record of major threats. Issues of detecting disorder in cyber security protection are constantly being taken seriously. This is an important policy that needs to be addressed in the style of cyber security systems.

The internal threat is extra sophisticated compared to the oldest threatening incidents. It's completely up to articulate and workout. User cannot break their managers and still harm the company. Internal threats occur in cases where the user is generally given certain rights to use the company's resources. The use of company computers and access to confidential information is generally seen as traditional computer activity. This may be an explanation that their inaccurate actions can only be known by existing cyber security measures. Solve internal security issues; introduce user security and business behavior. Influenced by these concepts, machine learning strategies are often used to evaluate user performance, resulting in an increase in the effectiveness of internal threat detection. As an improvement of existing strategies, an approach to identify internal threat by dividing the range of user activities is proposed. Initially, a set of one-day options are calculated from the user activity log. The feature vector of the series is built next to one-day feature figures over time. To isolate information of actual internal threat action that only includes certain types of bad things, use low-cost information retrieval in periodic non-threatening attack situations.

The proposed algorithmic rule compares the XG-boost classification and its performance with the alternative isolation random forest technique. The CMU promotes results by containing only threat information, strategy, and fourteen Internet browsing logs, as well as publicly threatened information that includes logon, device association, file transfer and email log files. In this paper, we tend to in brief explore the objectives and ways of time series analysis in machine learning. At identical time, we tend to get a way of however the big computing power of recent day's computers has helped statistical hugely in their ever-challenging tasks in statistic analysis. Finally, user activity deviates from the conventional usage activity. This means that the planned model classifies the user as a malicious user.

II. LITERATURE REVIEW

Gawai et al. (2017) machine learning algorithms have been developed for threats analyzing online business and labour activity, as well as login and logoff and email and internet operations. The author is paid several times as a substitute for each work. There are 42 developer features in 5 domains: email usability, email content, log-on/ log-off practicality, application time tasks and network functionality. They used 2 random recognition algorithms. It is fundamental to look for a cooperative disorder that utilizes divisional

forest. The second technique utilizes a random forest monitoring practice. The authors started the subsidiary foreign terrorist organization 0.77 for unproven employment practices with 73.4% accuracy.

Cinque et al. (2012) applied youth-oriented approach is one of the most widely used and developed strategies with progress and science history. With the creation of good depository records and therefore the involvement of many experts, these strategies can be very useful and effective. However, it is also dangerous. Investigation requires intensive and domain experience. In addition, responsive policies are slower than growing web threats.

Roy et al. (2015) developed by experts, the world health organization uses data processing strategies to extract aspects of operational problems. Dedicated integration strategies are used with the level of honesty of the domain of knowledge. The business executive catcher seeks to reduce the need for domain-specific information. Internally the catcher tried to promote the role of machine learning rather than human activity.

Aditham et al. (2017) developed a collaborative disorder diagnosis research system that supports LSTM. Instead of analyzing the log, it analyzes the memory behaviour of the knowledge centres in most large web networks. He used LSTM to see some variation in knowledge modification with each knowledge method. However, they should only perform basic testing and additional on-line analysis. Tiwari et al. (2017) developed LSTM and RNN systems to detect intrusions. Tragically, they supported the tested knowledge, suggesting that performance should exploit real knowledge in the future. Althubiti et al. (2018) LSTM wanted to work with the CNN system to handle cyber bullying communications radio to detect internal threats. The program generates a varied amount for each user's online session. However, during the entire processing, the system requires technicians to refer to the suspicious actions of the user associated with such internal threats. The log helps to detect internal threats at each session level instead of recording level. This can slow down their online activities after online analysis.

Moore et al. (2011) proposed anomaly detection methods by analyzing user roles and their historical access patterns from computer use and network activity logs. The list of models provided by the board can be used, as the installation of the re-establishment system for the base is for human hazards. It is an additional measure of measuring structural disorder by combining structural and operational knowledge from consumer activities. Features used in this work are: attachment count for sent emails, transfer counts for removable drives, multi-machine work, print job counts, blacklist delivery for websites, referral rates for all transfers and specific transfer tasks for transfer operations used for loading. Analysis of the impact of those strategies within the broader knowledge used for the logs of the 5500 population project has shown encouraging results. However, this set of data is not made public and provides an abstract framework as a whole and does not provide information on the technical implications needed to maintain such ambiguous knowledge.

Fan et al. (2003) projected necessary aspects of time series feature analysis is known as classification in machine learning. Time-series prognostication has been performed preponderantly exploitation statistical-based ways. These embody the well-known autoregressive moving average model are based on the evolution of the increments are used occasionally to reduce first-order non-stationary. However, differencing usually amplifies the high frequency noise within the statistic, and nice effort is so needed to see the order of an ARIMA model. Also, ARIMA models are mostly restricted to capturing the first-order non-stationary in an exceedingly statistic knowledge.

Weron et al. (2008) proposed a short statistic prognostication; it is extension of Fan. The ARIMA model assumes a linear relationship between the lagged variables and produces solely a coarse approximation to real-world advanced systems and usually fails to accurately predict the evolution of nonlinear and non-stationary processes. ARIMA model performance often degrades significantly whenever time trends features are gift within the extremely unsteady statistic knowledge.

Engle et al. (1982) explained the autoregressive conditional heterodasticity model to capture the second-order moment non-stationarity. This model represents the variance of the error term as operate of its regressive terms, thereby permitting a lot of stingy illustration of the time-series. Further, threshold nonlinear ARIMA models were developed by Tong (1990) that with success applied for statistic prognostication in social science and neuro science among different fields. Lineesh et al. (2010) proposed moving ridge bases to decompose statistic into orthogonal trend series, so used autoregressive model and threshold autoregressive models to forecast every series, respectively. Krishnamurthy et al. (2002) and Yadav et al. (1994) combined to develop a hidden Markov model and AR models beneath a Markov regime, wherever AR parameters switch in time in line with the realization of a finite-state Markov process, for non linear statistic prognostication. However, most of those ways tend to be restricted for nonlinear and stationary time series prognostication by the native one-dimensionality assumption implicit with an AR-type structure. Over the past few decades, artificial neural networks those exhibit superior performance on classification and regression issues in machine learning domain. The characteristics of the neural network approaches have non linear, knowledge driven, non-parametric, versatile and universal.

Lapedes et al. (1987) explained Feed-forward Neural Network models parameterized with a back-propagation algorithmic program have been utilized for non linear statistic prognostication. They're well-known to be at ancient applied math ways like regression and Box–Jenkins approaches in useful approximation, however they assume the dynamics underlying statistic are time-invariant. De Groot et al. (1991) proposed Feed-forward Neural Network with repeated feedback connections have additionally been tried for statistic prognostication. Such dynamic repeated neural network models permit prognostication of nonlinear time series occurring in varied fields. Grudnitski et al. (1993) designed a repeated network structure of nonlinear autoregressive models with exogenous input for multi-step prognostication of chaotic statistic. Various types of Radial Basis operate neural network models using dynamic regularization. Yee et al. (1999) orthogonal statistical procedure learning rule are investigated to capture totally different forms of trends and volatility within the statistic.

Barreto et al. (2010) reviewed time series prognostication approaches mistreatment self-organizing map neural network models. The local approximation property inherent in these models will improve the prognostication accuracy of nonlinear statistic compared to world models like feed forward neural network. They obviate the requirement to specify the amount of neurons before by permitting the network architecture to grow supported the information. Zhang et al. (2001) developed ensemble or hybrid neural network models such as wavelet neural network models have additionally been tried for non linear statistic prognostication. SVM-based prognostication ways use a category of generalized regression models, such as Support Vector Regression (SVR) and Least-Squares Support Vector Machines (LSSVMs).

Smola et al. (2004) that are parameterized using hogged quadratic programming ways. SVMs are categorised into linear, gaussian, polynomial, and multilayer perceptron classifiers. A linear regression is then created by minimizing the structural risk decrease the higher bound of the generalization error, resulting in higher prognostication performance than conventional techniques. Huang (2008) proposed extreme learning machine (ELM), anew sort of neural network for regression and classification issues. Though ANN is found to be a sure-fire prognostication tool in sizable amount of applications, it suffers from the constraints like recording machine technique, over fitting and gets cornered in native minima. A combination of moving ridge and Takagi Sugeno Kang (TSK) fuzzy rules based mostly

system proposed by Chang et al. (2007) is applied to predict monetary statistic knowledge of Taiwan stock market symbolic logic theory is most well-liked by several researchers as a result of it's an efficient tool to handle uncertainties. A fuzzy statistic technique supported a multiple period modified equation derived from adjustive expectation model is employed to forecast the Taiwan exchange. A fuzzy neural network is employed by Yu et al. (2005) to forecast monetary statistic wherever genetic algorithmic program and gradient descent learning algorithmic program are used as an alternative in an reiterative manner to adjust the parameters till the error is a smaller amount than the desired worth. A hybrid neuron fuzzy architecture supported kalman filter developed by Slim (2006) has been applied to predict financial statistic taking Mackey glass statistic as experimental knowledge.

A combination of improved particle swarm improvement (PSO) algorithmic program and fuzzy neural network has been adopted by Fu-yuan Huang (2008) to predict shanghai securities market indices. He has additionally applied genetic fuzzy neural network to forecast Shenzhen stock indices. A neural fuzzy model Kuo et al. (2009) has been applied to forecast sales knowledge of a renowned store Franchise Company in Taiwan where weights are generated by generic algorithm. Interval type-2 fuzzy neural networks are used to forecast monetary statistic knowledge. Each PSO and differential evolution (DE) algorithms are used for training the weights of the network. Most of the models reviewed on top of involve instruction execution, wherever the model is work and updated intermittently mistreatment batches of historic knowledge. However, the curse of dimensionality because of the preventative procedure effort, memory necessities and large knowledge sizes hampers their pertinence to several real-world issues, particularly for online method watching. A range of successive also called on-line or recursive forecasting models, like Hidden Markov Models introduced by HMMs Rabiner (1989) are investigated to surmount this limitation. AN HMM could be a special category of mixture models, where the discovered statistic y(t) is treated as a operate of the underlying, unobserved states vector. A state vector could also be reconstructed from autoregressive terms of y(t). Unscented Kalman Filter by Wan et al. (2000) are introduced to be at the constraints of Extended Kalman filters rather than native linearization and to avoid the Jacobian matrix calculation inherent in Kalman Filter, Extended Kalman filters choose low sample of points to realize a lot of correct estimate of native dynamics, and the evolution of those sample points is propagated at every estimation step. All the above ways is applied to classification of nonlinear and non-stationary time series databases that also area crucial side of knowledge mining.

III. RESEARCH METHODOLOGY

The main contribution of the proposed work is to extract the time series feature by setting the time window length then apply the Bi-LSTM model to predict deviation of the user activities and result is label as malicious or not malicious. Learning from the imbalanced dataset shows low accuracy so sampling the data using SMOTE sampling technique. As a classifier XG-Boost technique is used to classify whether the user is malicious or not malicious

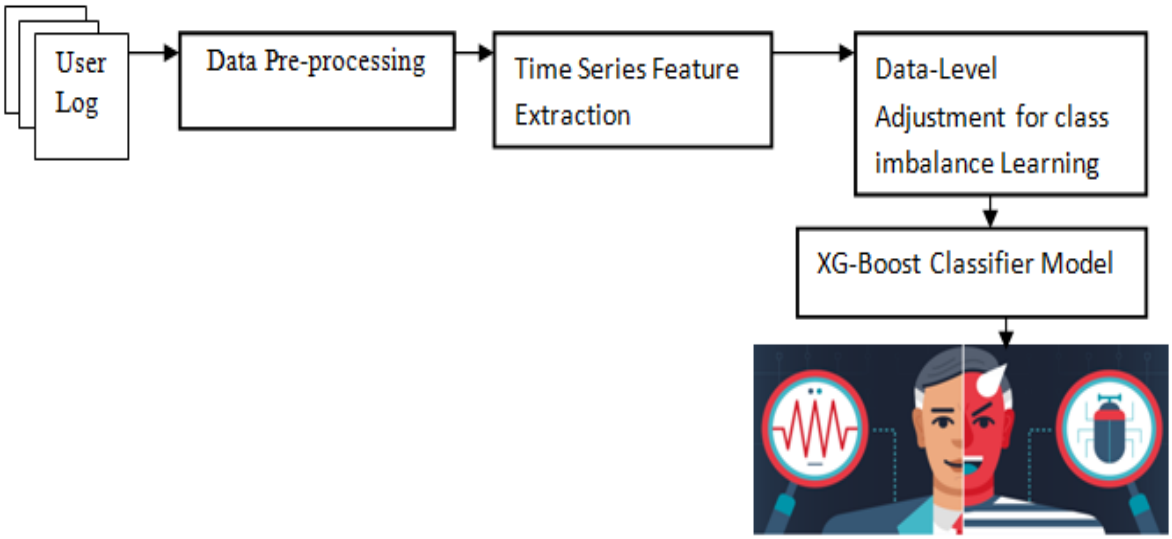

Fig. 1 Insider Threat Detection Framework

3.1 Source of Data

Dataset: The CMU Insider Threat Data set is the public insider threat dataset which contain log data of user activities. Version 4.2 of the data is used in the work, since it focuses only on behavioral characteristics. This data set consists of the user-activity logs of 1000 insiders over a period of 17 months from January 2017 to May 2018. The different user activity logs captured in the CMU Data are as follows. The logon activity details are stored in a file named logon.csv. The device.csv file contains the device connection information for each user, including the computer ids and the time stamps at which a device is connected/disconnected. The file transfer details for each user are stored in file.csv. The e-mail exchange details are contained in a file named email.csv. The http log is provided by the file http.csv.

3.2 Data Pre-Processing

In several Machine learning activities, the information set would possibly contain categorical data, primarily non-numeric information. The machine learning model can't settle for explicit knowledge thus initial rework the text knowledge into numeric knowledge using label coding technique.

1. Identify the categorical data.
2. Apply the label encoder algorithm for the text data.
3. Finally, Text data are encoded into numeric form.

3.3 Time Series Feature Extraction

The encoded information is next fed into the statistic feature extraction model wherever set the time window length is fifty as shown in below Figure 2 and extract the whole statistic feature of the user activities then apply the Bi-Directional Long Short Term Memory rule. Bi-LSTM model permits to specifying merge mode, that moves from the forward and backward path should be integrated before transferring to consecutive layer. The choices are total which ends are added at the same time, "Mull" Effects are increased along, "Concat" the results are classified along (by default), providing doubly the output worth within the next column, "avg" the middle of results is taken, the default mode is synchronization. The matter is outlined as a sequence of random values between zero and one. This sequence is considered the addition of a tangle for every given favorite in times based mostly order. The binary label (0 or 1) is related to every entry. Once the overall range of input values in an exceedingly sequence exceeds the limit, then the output worth overflows from zero to one. The result from the Bi-LSTM model is to see the deviation of the user activities if the user deviates from the regular activities suggests that then label user as malicious otherwise non-malicious.

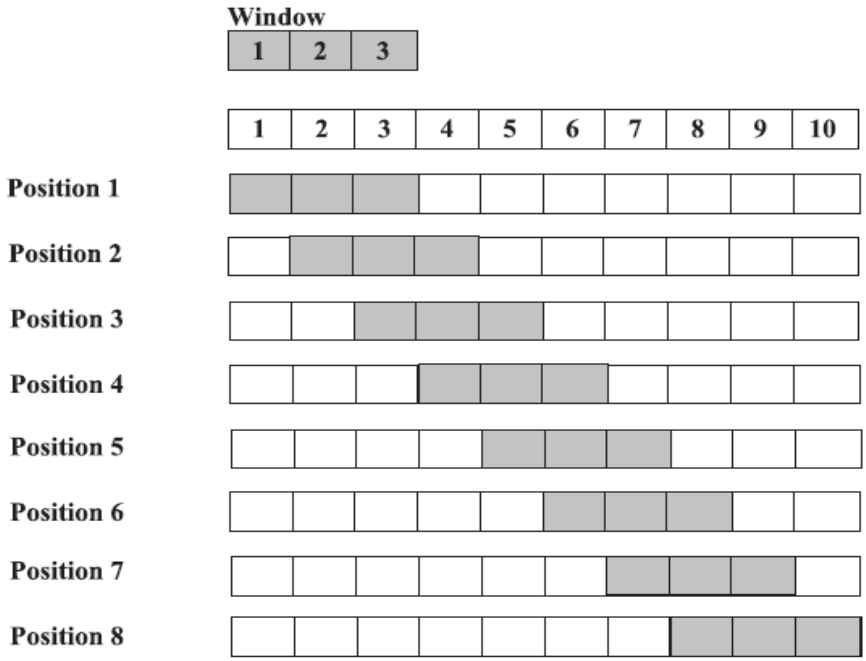

Fig. 2 Window Mechanism

3.4 Data Level Adjustment for Class Imbalance Learning

The category imbalance magnitude relation is calculated by the quantity of majority sample divided by the minority sample if the bulk same magnitude relation is high mean apply random beneath down-sampling technique to beneath sample the bulk class instances. Suppose the minority sample is low means that apply the artificial Minority Oversampling technique to up-sample the category instances. The new coaching sample is fed into the classifier; if the classifier happy with its performance means that proceed there with classifier otherwise alter the information and sophistication instances. XG-Boost classifier model accustomed to classify the user is malicious or not malicious user as shown in Figure 3. The classifier performance will be calculated by exactness, recall, f-score exactness is that the magnitude relation of properly foretold positive observations to the entire foretold positive observations. The recall is that the magnitude relation of properly foretold positive observations to all observations in actual category. F Score is the weighted average of exactness and Recall. Therefore, this score takes each false positives and false negatives formula as below.

$$Precision = \frac{TP}{TP+FP} \tag{3.1}$$

$$Recall = \frac{TP}{TP+FP} \tag{3.2}$$

$$fscore = \frac{2PR}{P+R} \tag{3.3}$$

Where,

1. TN - True Negative: case was malicious and predicted malicious.
2. TP - True Positive: case was non-malicious and predicted non-malicious.
3. FN - False Negative: case was non-malicious but predicted malicious.
4. FP - False Positive: case was malicious but predicted non-malicious.

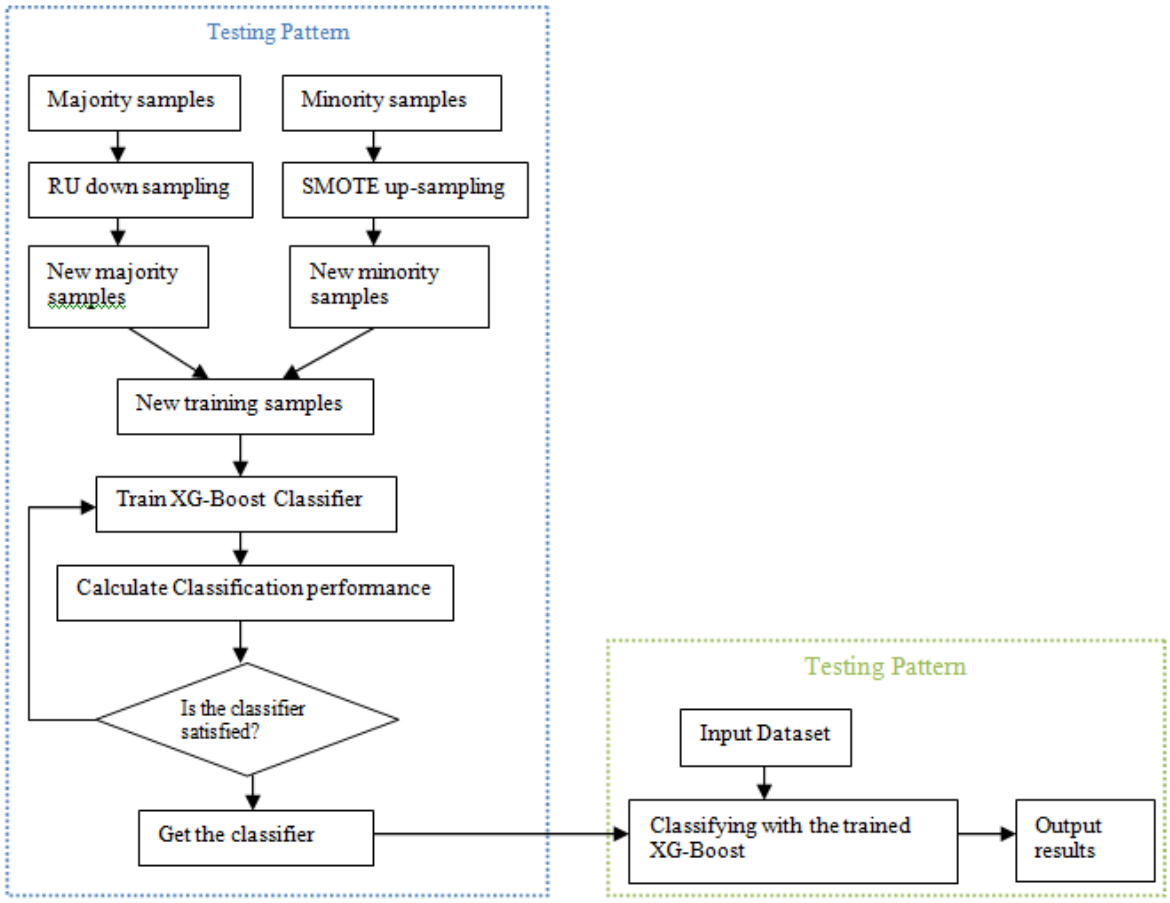

Fig. 3 Class Imbalance Learning

IV. RESULTS AND DISCUSSION

4.1 Data Pre-processing

	A	B	C	D	E
1	id	date	user	pc	activity
2	{X1D9-S0E	01-02-2010 06:49	NGF0157	PC-6056	Logon
3	{G2B3-L6E	01-02-2010 06:50	LRR0148	PC-4275	Logon
4	{U6Q3-U0'	01-02-2010 06:53	LRR0148	PC-4124	Logon
5	{I0N5-R7N	01-02-2010 07:00	IRM0931	PC-7188	Logon
6	{D1S0-N6F	01-02-2010 07:00	MOH0273	PC-6699	Logon
7	{S6P1-M4I	01-02-2010 07:07	LAP0338	PC-5758	Logon
8	{M6O6-F9	01-02-2010 07:08	MHH0180	PC-9822	Logon
9	{D0P1-U5I	01-02-2010 07:08	NOB0181	PC-3446	Logon
10	{F9K3-Q1F	01-02-2010 07:13	AHC0142	PC-8893	Logon

Fig. 4 Sample Dataset

In Figure 4 shows data concerning browsing information set and its information. The raw dataset may be fed into the model before its regenerate into a computer-readable type. The label encoder algorithmic program is employed to convert hierarchal information into numeric information that's economical for the process within the model.

	date	user	Feature
0	02-01-2010 06:49	NGF0157	10
1	02-01-2010 06:50	LRR0148	10
2	02-01-2010 06:53	LRR0148	10
3	02-01-2010 07:00	IRM0931	10
4	02-01-2010 07:00	MOH0273	10
5	02-01-2010 07:07	LAP0338	10
6	02-01-2010 07:08	MHH0180	10
7	02-01-2010 07:08	NOB0181	10
8	02-01-2010 07:13	AHC0142	10
9	02-01-2010 07:14	CTR0341	10
10	02-01-2010 07:24	LRR0148	13

Fig. 5 Encoded Dataset

In Figure 5 depicts the encoded knowledge as a model clear kind that's a numerical type. The model is capable of process only numeric information and not hierarchical information that's textual information. Once the feature coding is over, produce a fixed-size

feature vector for a convenient model process. The amount of options can depend upon the datasets the CMU insider threat, once encoding has twenty-seven features. Since the check size is a 0.5, these are going to be divided into equal components for training and testing. The model used for this method is the long short-term memory and support vector machine that could be a supervised learning technique. Model weights are saved once the training section. The planned system uses a neural network to come up with a user's web usage behavior patterns by analyzing daily online behavior. The system will endlessly monitor the standing of the user's web usage and realize abnormal event records. The primary half is the historical user's computing usage behavior analysis that is predicated on the historical log record of a selected user to coach for every user's behavior pattern. It records the foremost recent common log records and uses the model to predict this action of this user. If the extraordinary action is expected or the prophet is systematically separated from the specific actions, it should be investigated.

4.2 Time Series Feature Extraction

The time series feature will be extracted by exploitation window mechanism and also the result is initial time logon/off, whether or not USB connected or disconnect feature the subdivided device property options are weekend traditional affiliation, whether or not mail send to insider mail or outsider mail, file access either document, zip file, text, image file as shown in below Figure 6.

```
{'Login': 0, 'Logoff': 1, 'Device_Connected': 2, 'Device_Disconnected': 3, 'Email_From': 4, 'Email_To':
5, 'Files': 6, 'Sites': 7, 'Weekday_Normal_Login': 8, 'Weekday_After_Login': 9, 'Weekend_Login': 10, 'We
ekday_Normal_Logoff': 11, 'Weekday_After_Logoff': 12, 'Weekend_Logoff': 13, 'Weekday_Normal_Connect': 14
, 'Weekday_After_Connect': 15, 'Weekend_Connect': 16, 'Weekday_Normal_Disconnect': 17, 'Weekday_After_Di
sconnect': 18, 'Weekend_Disconnect': 19, 'Mail_Insider': 20, 'Mail_Outsider': 21, 'Bcc_Mail': 22, 'File_
exe': 23, 'File_jpg': 24, 'File_zip': 25, 'File_txt': 26, 'File_doc': 27}
```

Fig. 6 Time Series Feature

The training dataset is used for learning that is such as parameters e.g., weight of, e.g. several ways that hunt for dynamic relationship training information tend to quantify information that means they'll find and exploit the relationships that are evident in non-aggregate training information. Experimental information is independent information for training information, however, that follows a similar distribution of similarity as training information. If the model such as the training dataset is additionally precisely the same because of the check dataset, a sufficiently great amount of information has occurred. Higher preparation of training information, compared to check information, sometimes points to the foremost applicable. Finally, prediction information is the same because the actual information means that labels as non-malicious otherwise labelled as malicious.

1. Decide the information parameters.
2. Model those parameters.
3. Load and save information. Perform stop word elimination, and clean information, take away special characters.
4. Save parameters to the file.
5. Train information for prediction.
6. Use the Bi-LSTM model for prediction.
7. Repeat steps 5 to 7 until the end of the training data.

4.3Data Level Adjustment for Class Imbalance Learning

Learning from the difference dataset poses a crucial challenge thus classification and particularly a matrix of disparate prices like as well as within the training part. Usually, for categories C, the price matrix may be an area unit of size C × C, wherever the road item I and column j represent the price of not correcting the sample from category j as category i. Within the case of binary classification issues like insider threat detection, the price matrix is in size 2 × 2. Of specific importance are algorithmic rule level changes, modifications to the category distribution of real-world information and collaborative-based learning. The primary class of methods is wide utilized in classification exploitation neural networks or alternative classifications. The primary category-based learning algorithmic rule isn't appropriate for classifying information by unequal class distribution. The value-sensitive nature of the back-propagation learning algorithmic rule takes under consideration the fluctuations within the reading rate and therefore the output values of the somatic cell at the reading level of the little category was chosen to be more than that of the bulk category.

An increased error of back-propagation to possess a better impact on the minority class than the bulk section, and therefore the network is additional training little class samples more accurately rather than adjusting the reading rate, the negative value per subject is increased by the anticipated price within the corresponding output somatic cell. Similarly, any classifier may be born-again to its own version with a pointy value within the second stage of vital learning ways to decision, information-level adjustment is created to scale back uneven data distribution. These ways are sample training information in accordance with the negative value given by the price matrix. Each under-sampling and over sampling variation have their benefits and limitations whereas a sample size will usually lose necessary information, a little sample size will result in over-accuracy.

One of the algorithms for information correction is termed the artificial Minority Over-sampling Technique (SMOTE). It eliminates a minority class by adding new samples to the training set. The algorithmic rule selects one sample at a time, releases the vector into a little sample, finds the nearest neighbour of the component, compares the vector distinction between these, multiplies this distinction by a random range generated within the vary (0, 1), and at last, adds this weighted distinction within the feature example exploitation of this system, small sample sizes are generated domestically for the initial samples till the information set becomes linear. However, the SMOTE algorithmic rule isn't optional in nature, wherever any of the opposite sub-system samples may be won't to add additional.

The XG-Boost model that includes internal embedded action will leave only a little digital footprint on the analysis knowledge. The provision of large-scale information storage at reduced prices, and increased operational potency yield the gathering and analysis of an increasing quantity of experimental knowledge. However, a few ranges of some many risky tasks will simply be obscured by an outsized number of normal tasks, creating them harder to observe, resulting in the event of upper complexness once developing an

efficient analytics system. From the attitude of improper adoption, ask for to point out the final behaviour of the user as being like the fundamental model of that user or cluster of users. Any damaging behaviour may be cited as a deviation from the fundamental model. Here are two challenges within which a typical user's activity model includes a non-complex relationship to sense modality knowledge e.g. that they're compelled to undertake an unattended or supervised learning approach to deal with these challenges use XG-Boost thanks to its ability to represent non-linear relationships and therefore the vital understanding that a postgraduate student is ready to develop a general character thus any feature show that exemplifies traditional behaviour of users may be tricked with a nominal error. However, if the feature show is performed by users' malicious behaviour, the trained model should manufacture a mistake.

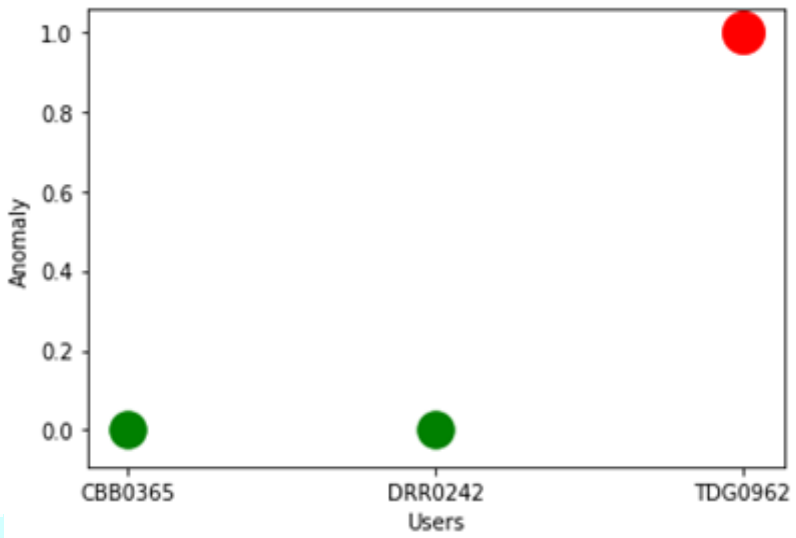

Fig. 7 Classification

The individual user is malicious or non-malicious with the assistance of XG-Boost classifiers wherever malicious denotes the abnormal user and non-malicious user denotes the conventional user as shown in Figure 7.

The macro average calculates a matrix for every malicious and non-malicious label and finds it's corresponding ways. A weighted scale calculates the matrix for every label and measures their ratio supported the quantity of relevant conditions for every label. This changes the macro to account for label imbalance, this might have an F1 result that's not between preciseness and recall. The subsequent Table one provides an outline of analysis metrics.

Table 1 Performance Evaluation

Classification Report				
	Precision	**Recall**	**F1-score**	**support**
0	0.95	0.94	0.94	67343
1	0.94	0.95	0.95	67343
Accuracy	0.94			134686
Macro avg	0.94	0.94	0.94	134686
weighted	0.94	0.94	0.94	134686

V CONCLUSION

Finally, the proposed system creates user models to represent unstable but internal threat computer usage behavior and supports historical vulnerability datasets that detect disorder. It can monitor user behavior endlessly by analyzing log sequence with numeric. This scalability detection system becomes a new technology for securing web security settings. Internal risk is widespread in organizations. However, thanks to management, it is still questionable in recent years. The result of its characteristic features and the behavior of insiders is not found by ancient methods alone.

The demand for its care will attract a lot of attention in the future Insider threat has two parts to manage the historical record with the threat detection system and to complete the online cyber security monitoring period with the Bi-LSTM with the time series analysis system. Through the investigation, it was found that most of the bullying incidents described were of the basic computing activity described. Although creating a training model with internal bullying records is very labor intensive, internal bullying often analyzes employees' computing usage behavior. Detection Insider provides security over cyber security by increasing the speed and accuracy of threat detection. The proposed system captures patterns that represent the user's general consumption behavior to distinguish normal behavior from harmful actions. Experiments show the improved performance of the proposed system on current log-based anomaly detection methods.

REFERENCES

[1] Aditham, S. Ranganathan, N. and Katkoori, S. 2017. LSTM-based memory profiling for predicting data attacks in distributed big data systems. In 2017 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW), 1259-1267.

[2] Althubiti, S. Nick, W. Mason, J. Yuan, X. and Esterline, A. 2018. Applying Long Short-Term Memory Recurrent Neural Network for Intrusion Detection. In Southeast Conference, 1-5.

[3] Barreto, G. Hammer, B. and Hitzler, P. 2007. Time Series prediction with the self-organizing map: a review perspectives of neural-symbolic integration, in Perspectives of Neural-Symbolic Integration, eds. B. Hammer and P. Hitzler, Springer, Berlin, 135– 158. 26

[4] Batuwita, R. and V. Palade, V. 2010. FSVM-CIL: Fuzzy support vector machines for class imbalance learning. IEEE Transaction Fuzzy System 18(3): 558–571.

[5] Chang, P. C. Fan, C. Y. Chen, S. H. 2007. Financial time series data forecasting by Wavelet and TSK fuzzy rule based system. Institute of Electrical and Electronics Engineers, Fourth International Conference on Fuzzy Systems and knowledge Discovery.

[6] Chawla, N. Lazarevic, Hall, O. and Bowyer, K. 2003. SMOTE Boost: Improving prediction of the minority class in boosting. In Proceedings of 7th European Conference Principle Practical Knowledge Discovery Databases, 107–119.

[7] Chawla, N. Bowyer, K. Hall, L. and Kegelmeyer, W. 2002. SMOTE: Synthetic minority over-sampling technique. Journal of Artificial Intelligence,16(1): 321–357.

[8] Cinque, M. Cotroneo, D. and Pecchia, A. 2012. Event logs for the analysis of software failures a rule-based approach. IEEE Transactions on Software Engineering, 39(6): 806-821.

[9] De Groot, C. and Wuertz, D. 1991. Analysis of univariate time series with connectionist nets: a case study of two classical examples. Neuro computing, 3(4): 177–192.

[10] Engle, R. F. 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica, 50(4): 987–1007.

[11] Fan, J. and Yao, Q. 2003. Nonlinear Time Series: Nonparametric and Parametric Methods, Springer-Verlag, New York.

[12] Fu-yuan, H. 2008. Forecasting Stock Price using a Genetic Fuzzy Neural Network. Institute of Electrical and Electronics Engineers, International Conference on Computer Science and Information technology. 978-0-7695-3308.

[13] Grudnitski, G. and Osburn, L. 1993. Forecasting S&P and gold futures prices: an application of neural networks. Journal of Futures Markets, 13(6): 631–643.

[14] Gavai, G. Sricharan, K. Gunning, D. Hanley, J. Singhal, M. and Rolleston, R. 2015. Supervised and Unsupervised methods to detect Insider Threat from Enterprise Social and Online Activity Data. JoWUA,6(4): 47-63.

[15] Huang F. 2008. Integration of an Improved Particle Swarm Optimization Algorithm and Fuzzy Neural Network for Shanghai Stock Market Prediction. IEEEWorkshop on Power Electronics and Intelligent Transportation System, 978-07695-3342.

[16] Hu, S. Liang, Y. Ma, L. and He, Y. 2009. MSMOTE Improving classification performance when training data is imbalanced. In Proceedings of 2nd International Workshop Computer Science Engineering, 13–17.

[17] Krishnamurthy, V. and Yin, G. 2002. Recursive algorithms for estimation of hidden Markov models and autoregressive models with Markov regime. IEEE Transactions on Information Theory, 48(2): 458–476.

[18] Kukar, M. and Kononenko, I. 1998. Cost-sensitive learning with neural networks. In Proceedings of 13th European Conference Artificial Intelligence.Hoboken, NJ, USA: Wiley, 445–449.

[19] Kuo, H. Horng, S. J. Chen, Y. H. Run, R. Kao, T. Chen, R. Lai, J. Lin, T. 2009. Forecasting TAIFEX based on fuzzy time series and particle swarm optimization, Expert System with Applications.

[20] Lapedes, A. and Farber, R. 1987. Nonlinear signal processing using neural networks: prediction and system modelling. Report, Los Alamos Nationary Laboratory, Los Alamos.

[21] Lineesh, M. C. and John, C. J. 2010. Analysis of non-stationary time series using wavelet decomposition. Nature and Science, 8(1): 53–59.

[22] Moore, A. P. Cappelli, D. M. Caron, T. C. Shaw, E. Spooner, D. and Trzeciak, R. F. 2011. A preliminary model of insider theft of intellectual property (No. MU/SEI-2011-TN-013). Carnegie-Mellon University Pittsburgh Pa Software Engineering.

[23] Rabiner, L. 1989. A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE, 77(2): 257–286.

[24] Roy, S. Konig, A.C. Dvorkin, I. and Kumar, M. Perfaugur. 2015. Robust diagnostics for performance anomalies in cloud services. In 2015 IEEE 31st International Conference on Data Engineering, 1167-1178.

[25] Seiffert, C. Khoshgoftaar, T. J. Van Hulse, J. and Napolitano, A. 2010. RUSBoost hybrid approach to alleviating class imbalance. IEEE Transaction System Cybern,40(1): 185–197.

[26] Slim, C. 2006. Neuro-Fuzzy Network based on Extended Kalman Filtering for financial time series. Proceeding of World Academy of Science, Engineering and Technology, 15, ISSN 1307-6884.

[27] Smola, A. and Schiolkopf, B. 2004. A tutorial on support vector regression. Statistics and Computing, 14(3): 199–222.

[28] Tiwari, T. Turk, A. Oprea, A. Olcoz, K. and Coskun, A. K. 2017. User-profile-based analytics for detecting cloud security breaches. In 2017 IEEE International Conference on Big Data, 4529-4535.

[29] Tong, H. 1990. Non-Linear Time Series: A Dynamical System Approach, Clarendon Press, Oxford, UK.

[30] Tulyakov, S. Jaeger, S. Govindaraju, V. and Doermann, D. 2008. Review of classifier combination methods in Machine Learning in Document Analysis and Recognition. Berlin, Germany: Springer, 361–386.

[31] Wan, E. A. and van, D. Merwe, R. 2000. The unscented Kalman filter for nonlinear estimation, in The IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium, IEEE, Lake Louise, Alberta, Canada, 153–158.

[32] Weron, R. and Misiorek, A. 2008. Forecasting spot electricity prices: a comparison of parametric and semi parametric time series models. International Journal of Forecasting, 24(4): 744–763.

[33] Yadav, P. K. Pope, P. F. and Paudyal, K. 1994. Threshold autoregressive modelling in finance: the price differences of equivalent assets. Mathematical Finance, 4(2): 205–221.

[34] Yee, P. and Haykin, S. 1999. A dynamic regularized radial basis function network for nonlinear, non-stationary time series prediction. IEEE Transactions on Signal Processing, 47(9): 2503–2521.

[35] Ye, L. and Keogh, E. 2011. Time series shapelets: A novel technique that allows accurate, interpretable and fast classification. Data Mining Knowledge Discovery, 22(1): 149–182.

[36] Yu, L. and Zhang, Y. Q. 2005. Evolutionary Fuzzy Neural Networks for Hybrid financial Prediction. Institute of Electrical and Electronics Engineers Transaction on Systems Man and Cybernetics, 35(2).

[37] Zhang, G. P. and Berardi, V. L. 2001. Time series forecasting with neural network ensembles: an application for exchange rate prediction. Journal of the Operational Research Society, 5(2): 652–664.