Emerging Trends in Data Science, Artificial Intelligence and Machine learning

Dr.P.Rizwan Ahmed

Dr.T.Aasif Ahmed

S. Abirami

ISBN: 978-81-955856-9-4

AN PUBLICATIONS

CONTENTS

Manipulate Machine Learning to Project the Next Buying Day for a Personal Retail Client

M.Prithi
Asst.Professor, Department of Computer Applications
Marudhar Kesari Jain College for Women, Vaniyambadi

Dr.K.Tamizharasi
Guest Lecturer, Department of Computer Science
Govt Arts and Science College Idappadi.

ABSTRACT

Target marketing has become more popular in recent years and know when a customer will need a product that can be of great value for a business. However, predicting this is a difficult task. This paper report on a study to predict when customers will buy fast-moving retail products using machine learning techniques. This is done by analysing the purchase history of customers when participating Retailers. These predictions will be used to personalize discount offers for customers when they are about to make a purchase. Such proposals will be sent to the participating customer's mobile device and, ultimately, physical, paper-based marketing will generally be reduced.

Keywords: Deep Learning, Machine Learning, Sales Prediction, Baseline model, E-Commerce.

INTRODUCTION:

We live in a rapidly changing world when it comes to technology. The use of paper becomes redundant; for example, more forms are filled out online, newspapers can be read online and assignments can be submitted electronically. Signing up for a new mobile app is the new norm, including apps to track sleep patterns or fitness levels or even just an app to play game. It becomes easier to collect information from customers when companies have loyalty programs track their buying behavior. We live in an age where search engines suggest your next word, online Shopping is no longer scary and people book cars using apps. The fact that technology is Easier to develop and collect information from customers. With these changes, the questions, however is: "How do companies use this information to gain a competitive advantage?" Do they use this information for the benefit of the customer? "; "How can a company use information about its customers to make people available different experiences? '.Marketing has moved from product-oriented to customer-oriented and in this information-rich field customer behavior at that time can help marketers choose the most effective marketing strategy for their customers [1]. If a company knows exactly what customers want to buy at a particular

time, the company can market to customer needs and potentially gain a competitive advantage. Predicting customer buying behavior opens the door for companies to market to a specific individual at the right time. This will have a different impact on customers than traditional brochures in newspaper.

A study was conducted to determine whether machine learning can be used to predict when individuals will purchase fast-moving retail products, based on their past buying behavior. Figure 1 illustrates how a retailer can use predictive tools to market to customers. Customers buy items from a retail store. Retail stores capture customer purchase history and provide sales data for recommended next purchase date (NPD) prediction. NPD predictor using machine learning technique that relies on certain sales data to determine when a customer will repurchase a certain product. This information is then provided to the marketing team, who can use it to advertise to individuals at the appropriate time.

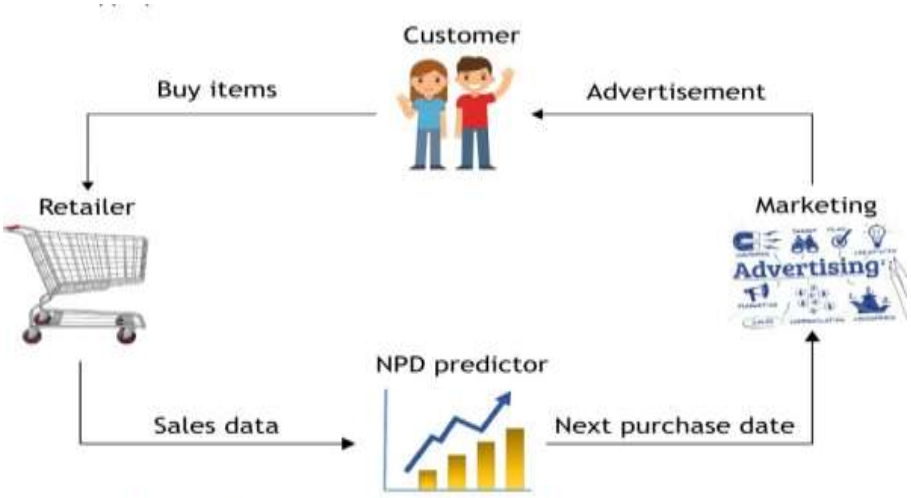


Figure 1: Illustrating how retailers can use the NPD predictor

2.1 Machine learning techniques used

This article focuses on four techniques for developing an NPD forecasting tool. Two of these techniques – repetition neural networks (RNNs) and linear regression - will only consider the sequence of the user product pair, while two other techniques - artificial neural networks (NNs) and extreme gradient enhancement - will look to the overview using all data from all product-user pairs. A brief explanation of how these

Technical work will be launched later.

2.1.1 Artificial Neural Network (NN)

An artificial neural network, referred to as a "neural network" (NN) in this article, is a collection of Nerve cells are organized into layers. As shown in Figure 2, there are three types oflayers:

Input layer - this layer feeds data into the system for further processing by subsequent layers. Hidden layer(s) - this layer(s) is located between the input and output layers, where the artificial neurons have gathers weighted inputs and produces outputs through an activationfunction.

Output layer — this layer is the last layer of neurons (it could be one or more neurons), which produces a given output for the system

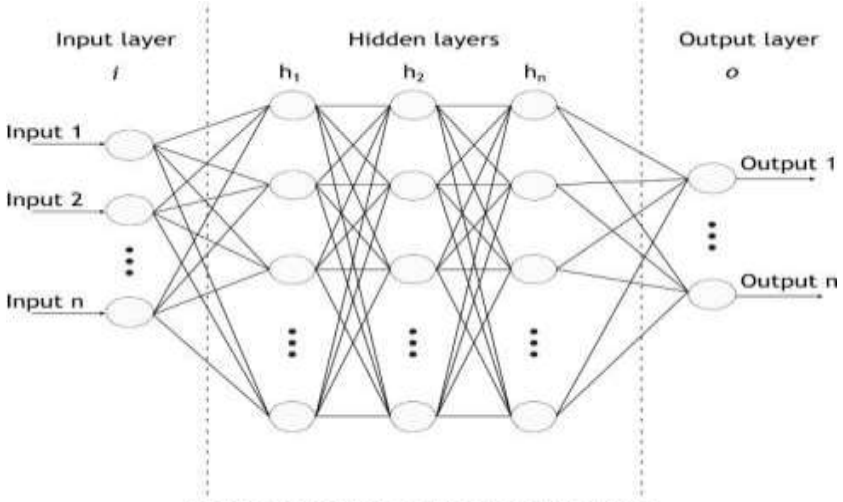


Figure 2: Neural network architecture

2.1.2 Recurrent Neural Networks (RNNs)

RNN is a neural network capable of modeling sequential data [7]. This technique is used for Sequential signals such as speech recognition, stock prediction, and language translation. Like with a flow Direct neural network, RNN also has input layer, hidden layer and output layer. However, one An additional loop is added to transmit information. RNN uses the output of the previous step as entered for the current step. Figure 3 shows the architecture of an RNN, with 1 as the input for primes step, 0 is the hidden value of the first step (must be initialized) and 1 is the output variable for the first class. It is said to repeat because 1 is then

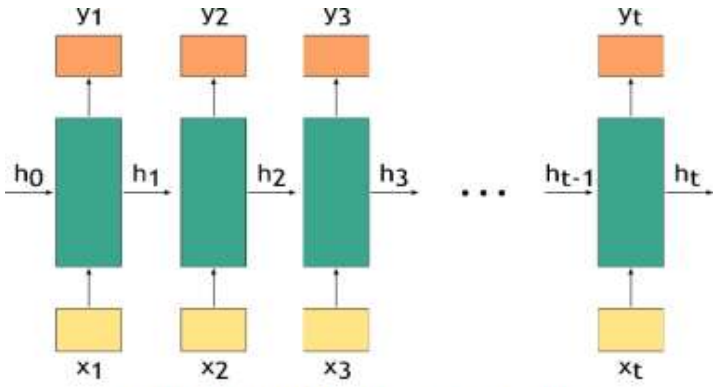


Figure 3: Unfolded RNN architecture

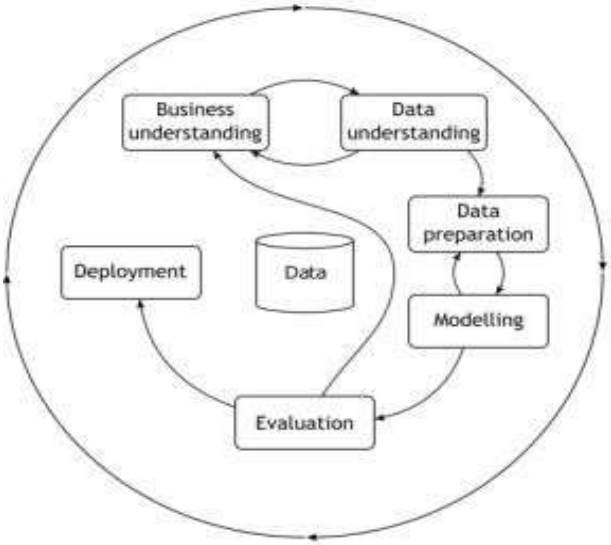
transmitted and contains information about first iteration. These are representations of RNNs, but they can also have multiple entries with a single output or an input with multiple outputs.

2.1.3 Linear regression

Linear regression is a statistical approach to model the relationship between a dependent variable and one or more independent variables. For linear regression, a straight line is adjusted through the data in terms of specific mathematical criteria. For example, this line can be adjusted to minimize the sum of squares space between data and line. This allows estimation of dependent variables [8]. Here widely used technique, mainly for prediction or forecasting. Linear regression is usually modeled as = (Xi,) +, where represent the dependent variable,Xirepresents the independent variable, represents an unknown parameter and represent error period. The goal then is to estimate the function (Xi, B) best fit the data. Parameter is estimated using various tools provided by regression analysis, such as least squares method, find out the value of B minimizes the squared error between the line and the data. After estimating this value, The data can then be adjusted for the prediction.

METHODOLOGY

To develop and evaluate the proposed NPD model, the cross-industry standard process for data mining (CRISP-DM) process [10], depicted in Figure 6, was used. This process has six stages: business understanding, data understanding, data preparation, modelling, control, and deployment. These will be discussed briefly.



3.1 Business savvy

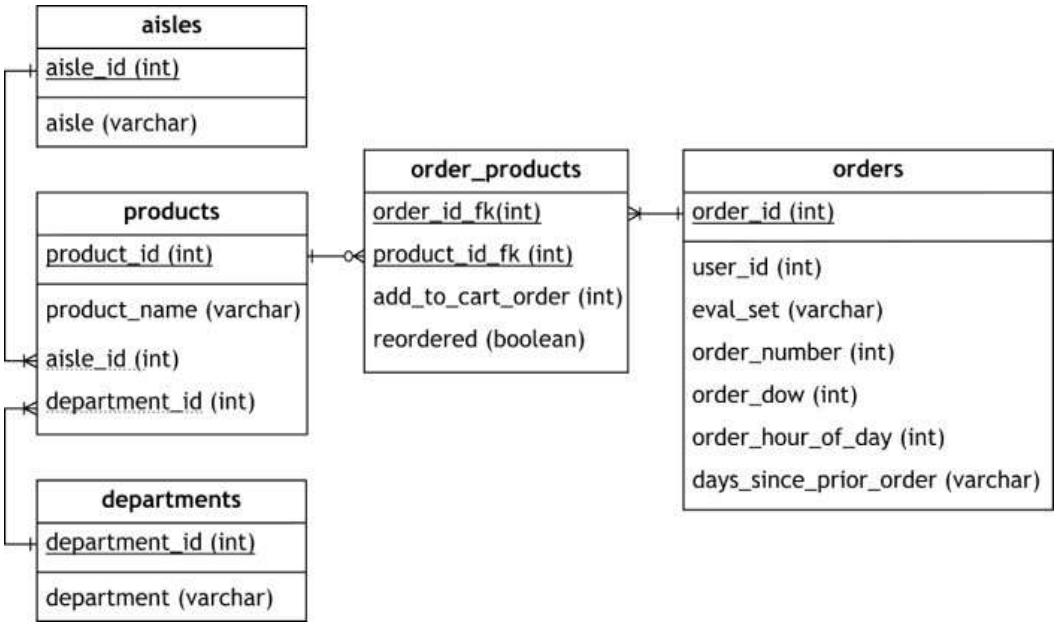
The concept of business insights is covered in the first part of this article. It's finally being used to predict the next purchase date of the customer-product pair, so that the company can profitfrom using personalized marketing strategies with these customers.

3.2 Understanding data

The dataset used to develop the NPD predictor is the Instacart online grocery dataset from 2017 [11]. It includes over three million orders from over 200,000 users. Relational data set The structure consists of five tables. These tables are: orders, departments, products, aisles and orders some products. The relational structure can be seen in Figure 7. Each user has at least four commands, with order of products purchased in each order. The command table contains the function 'days_since_prior_order', provides a measure of the relative time between two orders from the user. Here will be manipulated to predict the NPD for a product-user pair. 3.3 Prepare data DND for each product-user pair is the target feature, and since this feature is not explicitly available in dataset, it must be derived. Feature available describes the time between commands to the user is in the orders table 'days_since_prior_order'. However, this feature does not record

the days between for which a particular product has been ordered. The first entity created is the capture entity target variable.

The next step in the data preparation phase is to create features that describe the desired output feature. In the following subsection, the functions that have been created are described.



3.3.1 Sequence-based features

Feature 1:

Number of days between orders by product The dataset only identifies the dates between user orders. This is a useful feature, but we want to predict the NPD for a user-product pair, we have to convert the 'days_since_prior_order' function in the 'days_between_orders_per_product' feature, as this will get the desired value at guess. This is done by building a table for each user that includes the sequence number for the user has a matching 0 or 1, indicating that the user purchased a product in that particular domain purchase case. Then the 'days_between_orders' feature is used to calculate 'days_between_orders_per_product' considers cases where the product is purchased.

Feature 2: Days since prior order per product

The ‘days_since_prior_order_per_product’ feature was created to capture some of the user’s behaviour, along with the user-product behaviour. An example of how this feature is created, again for ‘original beef jerky’, can be seen in Figure 9. Assuming that the customer purchased the product, this feature evaluates how many days ago the user made a purchase (not necessarily

the product that is being evaluated, but any product). As seen in the example, the product was purchased with order 2; thus the customer previously ordered 15 days ago, and the first entry for the sequence will be 15; but in order 3 this product was not purchased, so no record for this instance will be taken, as the product was not ordered. The product was again ordered with order 4; therefore an entry will be made, and the previous order was made 29 days ago.

3.4 Modelling

After the features and datasets were created, the models were developed and tested. The first one The method performed is to predict NPD using only sequence-based features. This is formed using RNN, used for time series prediction.

3.4.1 RNN . Deployment

RNNs are trained only using sequence-based features, as shown in Figure 10.

First implementation took the entire sequence of feature 1 except the last entry in the sequence to form model and the last entry in the series as a test variable with which the prediction can be compared.

The second implementation of an RNN model takes the sequence of trait 1 and trait 2. The target variable is preserved (last entry of property string 1) and last entry of both sequences were removed from training. Therefore, the size of the training set is equal to the product-user pair the size of the string has no final variable. Since enough data is required to train the model, the user product pair was filtered to have at least 20 entries, leaving a dataset with 112,796 product-user pairs predicted by the NPD. RNNs are implemented in Python using PyTorch, an open source machine learning library. Thesequences have been scaled before training, parameters can be set as optimizer,learning rate, criteria, number of hidden layers and activation function. Adam Optimizer used with a learning rate of 0.01 and 10 hidden classes. An adjustable linear unit trigger function (relu) was used with the square root squared error loss criterion.

3.4.2 Regression implementation

The second sequence-based approach (Section 2.1.3) is implemented using linear regression. how sequence data is represented for linear regression, with the above purchase case x-axis and sequence value (number of days between purchases) on the y-axis. For this implementation, next purchase version value, in this case 47, has been predicted. Regression model implemented in Python using Scikit-learning's machine learning librarySGDRegressor (SGD stands for'stochastic gradient descent'). Data has been scaled and default value parameters were used, with a squared loss function and a learning rate of 0.001.

3.5 Evaluation (results)

To evaluate the models, the absolute error in days was calculated for each user-product pair for each technique. The absolute errors were sorted from small to large and plotted, as seen in Figure 15. The graph on the left-hand side shows all the user-product pairs' absolute prediction errors. The NN implementation performs best, as it has the smallest absolute error over number of user-product pair instances. In the graph on the right-hand side, only the first 40 000 user-product pair instances are plotted. This shows that the NN has a much smaller gradient than the other techniques. This also shows that the NN can predict at least 40 000 user-product pairs with an absolute error of less than one-and-a-half days.

4. CONCLUSION

A user-product pair's next purchase date can be predicted using machine learning. You can see it in. This paper uses a technique that looks at all data, including data from other product-user pairs, which performs better than the NPD prediction approach using only data from product pairs - other users, a product-user pair. NN outperforms other algorithms with XGBoost being the second best algorithm, that is better than any sequential approach.

Data is the new oil, and therefore a resource that must be used and managed with care. It is natural, for industrial engineers to develop system building blocks to support business operations and decision making using machine learning. This study shows how machine learning can be used to predict purchase date for a particular customer. A retailer can use the results to develop a system offer individual customers individual discounts on specific products on specific days. This design is different typical loyalty programs, as offers are personalized, based on the needs of each customer purchase history. Forecasting NPD is a system for the future and when implemented reduces waste insert paper (often found in newspapers) and can help retailersbetter plan their inventory, which can Save on logistics costs and reduce pollution caused by distribution.

This research identifies opportunities for future work, including:

1.. Extend analysis to more datasets.

2. Implement the NPD predictor in the retail chain using the NN algorithm.

3. Explore ways to combine sequence-based attempts with non-sequence-based attempts.

4. Define a marketing strategy to target specific users, using the NPD predictor.

REFERENCES

[1] Hosseini, M. & Shabani, M. 2015. New approach to customer segmentation based on changes in customer value. Journal of Market Analysis, 3(3), pp. 110-121.

[2] Linoff, G.S. & Berry, M.J. 2011. Data mining techniques: For marketing, sales, and customer relationship management. Indianapolis: John Wiley & Sons.

[3] Raorane A.A., Kulkarni, R.V. & Jitkar, B.D. 2012. Association rule — Extracting knowledge using market basket analysis. Research Journal of Recent Sciences, 1(2), pp. 19-27.

[4] Cumby, C., Fano, A., Ghani, R. & Krema, M. 2005. Building intelligent shopping assistants using individual consumer models. IUI 05 — 2005 International Conference on Intelligent User Interfaces. San Diego, California, USA.

[5] Cumby, C., Fano, A., Ghani, R. & Krema, M. 2004. Predicting customer shopping lists from point-of-sale purchase data. KDD ’04 Seattle, Washington, USA.

[6] Els, Z. 2019. Development of a data analytics-driven information system for instant, temporary personalised discount offers. M.Eng thesis, Stellenbosch University.

AN PUBLICATIONS

(Reg.Under MSME,Govt.of India)

CERTIFICATE OF PUBLICATION

This is to certify that M.PRITHI, Asst.Professor, Department of Computer Applications Marudhar Kesari Jain College for Women, Vaniyambadi , has published his /her chapter Manipulate Machine Learning to Project the Next Buying Day for a Personal Retail Client in the edited book on "Emerging Trends in Data Science, Artificial Intelligence and Machine learning" bearing ISBN: 978-81-955856-9-4

".

Date : 05-11-2022

Editor –Iin-Chief
(Dr.M.Charles Arockiaraj)

Cyber Law, Policies, Procedures in Banking Sector

Volume - 1

Compiled by

Dr. M. Inbavalli

Dr. G. Deepalakshmi

B. Sakthimala

பாரத் பதிப்பகம்

A STUDY OF CYBER SECURITY USING MACHINE LEARNING TECHNIQUES

Ms.M.Prithi [1] , Dr.G.Tamizharasi [2]
[1]Assistant Professor, Marudhar Kesari Jain College For Women & Research Scholar,

Computer Science, Periyar University, Salem, Tirupathur, Tamilnadu, India.

[2] Research Supervisor, Computer Science, Periyar University Constituent College of Arts & Science, Idappadi, Salem, Tamilnadu, India

ABSTRACT

Within the ever-growing and quickly increasing field of cyber security, it is nearly impossible to quantify or justify the explanations why cyber security has such an outsized impact. Permitting malicious threats to run any place, at any time or in any context is a long way from being acceptable, and may cause forceful injury. It particularly applies to the Byzantine web of consumers and using the net and company information that cyber security groups are finding it hard to shield and contain. Cyber security may be a necessary thought for people and families alike, also for businesses, governments, and academic establishments that operate inside the compass of world network or net. With the facility of Machine Learning, we will advance cyber security landscape. Businesses these days are gathering immense amounts of user information. Information is at the heartbeat of any business-critical system you'll be able to think about. This co-jointly includes infrastructure systems that are being implemented these days. Today's high-tech infrastructure, that has network and cyber security systems, is gathering tremendous amounts of data and analytics on almost all the key aspects of mission-critical systems. Whereas people still give the key operational oversight and intelligent insights into today's infrastructure, machine learning and AI are gaining pace and gathering immense momentum in most of the areas of today's systems, whether or not it's positioned on premise or within the cyber security house.

Keywords:Machine Learning, cyber security, k-means, Random Forest, SVM etc.

INTRODUCTION

Cyber-attacks are increasing within the cyber world. There ought to be some advanced security measures taken to scale back or avoid the amount of cyber-attacks. There are various attacks like D-Dos attacks, Man within the middle, information escape, PROBE, User-To-Root, Remote-To-Local.

These attacks are utilized by the hackers or intruders to realize the unauthorized access to any non-public network, websites, information or perhaps in our personal computers. Therefore, outside or internal hackers use using advanced techniques or finding ways to tickle or break any defense systems to shield the sensitive information, information, money data. Sensible intrusion munitions ought to stop or try and manage varied innovative attacks created or programmed by the hackers.

Cyber security refers to the science of technologies, processes, and practices designed to shield networks, devices, programs, and information from attacks, damage, or unauthorized access. Cyber security can also be stated because it's security, within the year 2016, witnessed several advancements in machine learning techniques like self-driven cars, linguistic communication process, health sector, and sensible virtual assistant. They need to be used for locating helpful data from varied audit datasets, which are applied to the matter of intrusion detection.

With the assistance of Machine learning technology, we will deploy these ideas in cyber security to boost the protection measures within the intrusion detection system. Initially, we've got to feed the information into the machine- learning model. The model gets trained by the dataset sample and makes it a trained model. Once we feed the dataset sample, future step is to use and apply the machine- learning formula.

Machine learning formula plays an important role in rising the protection measures in this intrusion detection system. ML algorithms are classified into 2 types: supervised learning and unsupervised learning. They're differentiated by the information (i.e., input) that they settle for. Supervised learning refers to algorithms that are given a group of labelled training information, with the task of understanding what differentiates the labels. Unsupervised learning refers to algorithms that are given unlabeled training information, with the task of inferring the classes all by itself. Typically, the labelled information is incredibly rare, or even the task of labelled data is itself terribly exhausting and we may not be able or ready to sight if labels actually exist.

RELATED WORK

In the budding stages of building intrusion detection systems, cyber analytics support was studied as a mixture of ML/DM (Machine learning/Data mining). Anomaly-based techniques model the system behavior and traditional network, and helps to spot anomalies as deviations from traditional behavior.

Its advantage is that the profiles of traditional activity are custom-made for each system, application or network, thereby creating it tough for attackers to grasp those activities that they can perform undiscovered and in a very clandestine manner. They give the impression of being appealing, owing to their distinctive ability to sight zero-day attacks.

Hybrid techniques mix anomaly detection and misuse. They're used to boost detection rates of acknowledged intrusions and reduce the rates of unknown attacks. Again, intelligent intrusion detection systems will solely be engineered if there's accessibility of a good information set. An information set with a large quantity of quality data and the one that mimics real time which will solely facilitate to train the associated check of an intrusion detection system.

A Comprehensive Cybersecurity Audit Model

Nowadays, non-public firms and public establishments are dealing with constant and complicated cyberthreats and cyberattacks. As a general warning, organizations should build and develop a cybersecurity culture and awareness so as to defend against cyber criminals. Data Technology like IT and data Security like InfoSec audits that were economical within the past, try to converge into cybersecurity audits to handle cyber threats, cyber risks and cyberattacks that evolve in an aggressive cyber landscape[1]. However, the rise in variety and quality of cyberattacks and therefore the convoluted cyber threat landscape is challenging the running cybersecurity audit models and fitting proof for a brand new cybersecurity audit model. This text reviews the simplest practices and methodologies of world leaders within the cybersecurity assurance and audit arena. By means of the analysis of this approaches and theoretical background, their real scope, strengths and weaknesses are highlighted looking forward at a good and cohesive synthesis. As a result, this text presents an inspired and comprehensive cybersecurity audit model as a proposal to be utilized for conducting cybersecurity audits in organizations and Nation States. The CyberSecurity Audit Model (CSAM) evaluates and validates audit, preventive, rhetorical and detective controls for all structure useful areas [2]. CSAM has been tested, enforced and validated at the side of the Cybersecurity. A research case study is being conducted to validate each model and therefore the findings are revealed consequently.

Feature selection to detect botnets using machine learning algorithms

A completely unique technique to try to feature choices to sight botnets at their section of Command and control (C&C) is conferred. A significant downside is that researchers have proposed options supported in their experience, however there's no technique to judge these options since a number of these options might get a lower detection rate than alternative.

To the current aim, we discover the feature set supported connections of botnets at their section of C&C, that maximizes the detection rate of those botnets. A Genetic formula (GA) was accustomed and chosen as the set of options that offers the best detection rate. We tend to use the machine learning formula C4.5, this formula did the classification between connections belonging or not to a botnet. The datasets employed in this paper were extracted from the repositories ISOT and ISCX [3].

Some tests were done to induce the simplest parameters in a GA and the formula C4.5. We tend to co-jointly perform experiments so as to get the simplest set of options for every botnet analyzed

(specific), and for every kind of botnet (general) too. The results are shown at the tip of the paper, within which a substantial reduction of features and the higher detection rate than the related work conferred were obtained.

Intrusion Detection using Deep Belief Network

The problems existing in intrusion detection using neural network, including redundant data, large amount of data, long-time training, are easy to fall into the local optimum. An intrusion detection technique using deep belief network

(DBN) and probabilistic neural network (PNN) is proposed. First, the raw information is regenerated to low-dimensional data whereas holding the essential attributes of the data by using the nonlinear wit of DBN. Second, to get the simplest learning performance, particle swarm optimization formula is employed to optimize the amount of hidden-layer nodes per layer. Next, PNN is deployed to classify the low dimensional information [4]. Finally, the KDD CUP 1999 dataset is utilized to check the performance of the tactics mentioned on the top. The experiment result shows that the tactic performs higher than the standard PNN, PCA-PNN and non-optimized DBN-PNN.

Over the previous couple of years' machine learning has migrated from the laboratory to the forefront of operational systems. Amazon, Google and Facebook use machine learning on a daily basis to boost client experiences, instructed purchases or connect individuals socially with new applications and facilitate personal connections. Machine learning's powerful capability is additionally there for cybersecurity. Cybersecurity is positioned to leverage machine learning to boost malware detection, sorting events, acknowledge breaches and alert organizations to security problems.

Machine learning may be accustomed to determine advanced targeting and threats like organization identification, infrastructure vulnerabilities and potential mutually beneficial vulnerabilities and exploits. Machine learning will considerably modify the cybersecurity landscape [5].

Malware by itself will represent as many as three million new samples in a hour. Ancient malware detection and malware analysis is unable to pace with new attacks and variants. New attacks and complicated malware are ready to bypass network and endpoint detection to deliver cyber-attacks at alarming rates. New techniques like machine learning should be leveraged to handle the growing malware downside. This proposition describes the machine learning, that can be accustomed to detect and highlight advanced malware for cyber defense analysts. The results of our initial analysis and a discussion of future analysis to increase machine learning is presented.

A Comparative Survey on the Influence of Machine Learning Techniques on Intrusion Detection System (IDS)

With the large growth of laptop networks and content usage of users, there's a necessity for secure and reliable networks. Because it is determined that the various form of network attacks is raised

over a amount of your time, it's necessary to create the supply of effective automatic tools so as to spot the attack detection situations. Intrusion Detection System is one among the attack systems that detect intrusions returning from the net. Many approaches were determined within the literature for intrusion detection over the network. Within the recent past, mining techniques were prevailing so as to visualize the intrusion detection [6]. The characteristics of incoming intrusions were known by using the well mined data over the information given within the network. Whenever an identical object is found within the characteristics of the well-mined information then it's declared as an intrusion. Supporting this criterion varied intrusion detection models were developed within the recent analysis and therefore the accuracy is improved. A quick review is carried out over the sooner approaches. The entire approach is divided into information preprocessing approaches and detection approaches. Further, the information preprocessing approaches are divided into Feature extraction and have transformation models that support operating methodology over the options. Similarly, the detection approaches are classified as machine learning and organic process approaches.

Improving Cybersecurity Assurance Model

Every time a gaggle of auditors are taking part in an IT, data Security or compliance audit, there'll be consistent phases like designing, shaping objectives and scope, elucidating terms of engagements, conducting the audit, corroboratory proof, evaluating risks, news the audit findings and schedule follow up tasks. Designing a cybersecurity audit isn't totally different than any kind of audit. This however will take a great deal of effort thanks to the quality of the many cybersecurity domains.

However, most cyber capabilities aren't reviewed by the inner audits' scope. This specific framework includes risk/compliance management, development life cycle, security program, third-party management, information/asset management, access management, threat/vulnerability management, of implementing cybersecurity controls as a part of an overall framework and strategy, the necessity for assurance which will be achieved by management reviews, cyber risk assessments, information management and protection, risk analytics, crisis management and resiliency, security operation and security awareness and training. Moreover, Deloitte's framework is aligned with trade frameworks just like the National Institute of Standards and Technology (NIST), data Technology Infrastructure Library (ITIL), Committee of Sponsoring Organizations of the Treadway Commission (COSO) and world organization for Standardization (ISO).

In addition, there aren't any metrics to live cybersecurity audits and therefore the cybersecurity audit topic is poorly understood because it transforms extremely quickly. Khan considers that to hide a significant scope for designing a cybersecurity audit, the auditors should embrace all relevant areas of any organization; these areas are client operations, finance, human resources, IT systems and applications, legal, purchasing, regulatory affairs, physical security and every one of the applicable third parties that have relationships with the business.

Database Intrusion Detection System Using Octraplet and Machine Learning.

Several Intrusion detection systems are developed for host systems and networks. However, comparatively few notable works are there for information intrusion detection. One among the latest works revealed was a technique by Chung et al. proposes a misuse detection approach for information intrusion detection. Here frequent information patterns are well-mined and hold on as traditional profiles.

The main disadvantage is that it doesn't produce role profiles. The users perform totally different actions supported by their roles. User profiles can't be used as the only criteria. Users can perform actions supported roles and that they will be detected malicious. Lee et al. a proposed true time intrusion detection system supported time signatures. Real time information systems use temporal information objects and their values change with time [7]. Thus whenever time it is updated a device dealing is generated. The temporal information is updated over a certain amount of your time. If a dealing tries to vary the temporal information that has been updated already over that amount, an alarm is raised. But the disadvantage of this technique is that it focuses on updates solely and not role profiles. Hu Panda. uses log files to come up with user profiles. Frequently accessed information and tables and hold on for comparison. The problem with this approach is that, maintenance of knowledge is incredibly tough once the dimensions of information is simply too giant and variety of users conjointly increase dynamically.

Today's cyber security threats are too varied and arrive too quick for strictly manual defense. Machine learning in addition provides power and increased speed to tackle enormous volumes of attacks with myriad variations. Nevertheless, the $64000 key to investing AI for cyber protection is to use it with human intelligence, combination of power, speed, skills and judgement. Artificial Intelligence and Machine Learning are often very nice and helpful in detective work in cyber security attacks. The work that human has to do are often through with the assistance of machine learning at a lot of quicker pace and with high accuracy. Implementing numerous Machine-Learning Techniques can facilitate US to discover the cyber security attacks.

DISCUSSION AND FINDINGS

The machine learning security corporations typically deploy train methods on giant information sets to "learn" what to look out for and the way to react to totally different things on networks. Machine learning is much too powerful in its title, though, and approach could be a natural suited antivirus defense and malware scanning:

- Machine learning approach can facilitate US to stop sensitive information leaks, corporate executive intrusion detection system and malware detection.

- Machine learning is more and more permitting businesses these days to work with higher potency, accuracy, agility, and intelligence.
- This approach could facilitate to unravel the safety problems.
- Machine learning rule will discover and determine any new uncommon pattern or behavior that may arise from intrusions from outside.
- Security holes caused by allowances created to users or programmers or directors ought to be infatuated a lot of care.

CONCLUSION

This paper represents a review of Machine Learning and DL unit methods for network security domain. The literature paper, that has largely targeted on the last four year, introduces the most recent applications of ML and DL unit within the field of intrusion detection systems. Sadly, the foremost effective methodology of intrusion detection has not nevertheless been established, and therefore the analysis remains occurring. Every approach for implementing an intrusion detection system has its own pros and cons, a degree apparent from the discussion conducted for comparisons among the varied ways. Thus, it's tough to settle on a specific methodology to implement an intrusion detection system over the others. Datasets for network intrusion detection are important resources for coaching and testing systems. The Machine Learning and DL methods don't work while not the representative information, and getting such a dataset is tough and long. However, there are several issues with the accessible existing public dataset, like uneven information, or out-of-date content and therefore the likes are similar. These issues have mostly restricted the event of analysis during this explicit space. Network info updates in no time, that brings to the ML and DL model coaching with larger problem. The Model has to be retrained semi-permanent/long-termed and quickly. Thus progressive learning and long learning are the long run focus within the study of this field within the future.

REFERENCES

1. J. Cano, "Cyberattacks-The Instability of Security and Control Knowledge", *ISACA Journal*, vol.

   5, pp. 1-5, 2016.
2. C. Hollingsworth, "Auditing from FISMA and HIPAA: Lessons Learned Performing an In-House Cybersecurity Audit", *ISACA Journal*, vol. 5, pp. 1-6, 2016.
3. Li X, Wang J, Zhang X, "Botnet Detection Technology Based on DNS", J. Future Internet, 2017.
4. Y J Hu, Z H Ling, "DBN-based Spectral Feature Representation for Statistical Parametric Speech Synthesis", *IEEE Signal Processing Letters*, vol. 23, no. 3, pp. 21-325, 2016.

5. Dinil Mon Divakaran et al., “Evidence gathering for network security and forensics”, *Digital Investigation*, pp. 56-65, 2017.

6. S Fong, R Wong, A V Vasilakos, “Accelerated PSO Swarm Search Feature Selection for Data Stream Mining Big Data”, *IEEE Transactions on Services Computing*, vol. 9, no. 1, pp. 33-45, 2016.

7. M. Khan, “Managing Data Protection and Cybersecurity- Audit’s Role”, *ISACA Journal*, vol. 1, pp. 13, 2016