ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 5, Issue 5 , May 2018

Big Data Security Modeling and Analysis Using Machine Learning Techniques

Feba Babu, Kishore Sebastian

P.G. Student, Department of Computer Science, St. Joseph College Of Engineering, Palai, Kerala, India
Assistant Professor, Department of Computer Science, St. Joseph College Of EngineeringPalai, Kerala, India

ABSTRACT: Presently, Associations and countries are getting the opportunity to be exposed against a wide grouping of security ruptures against their information establishment. Cyber threat is evident from the extending rate of digital assaults against PCs and essential establishment. Cyber security is a basic piece in development areas. The main challenges of large set of data include storage, capturing, Analysis, sharing, transfer, visualization, search, querying, updating and information privacy.In this paper focus on detect phishing webpages with keyword based features for grouping phishing URLs and classifications using machine learning techniques.Thenhost on to the Azure Machine Learning Studio process with three models thatare,Two-Class Neural Network, Two-Class Boosted Decision Tree and Two-Class Decision Jungle.

KEY WORDS: Machine Learning, Cyber attack, Phishing attack,URLs

I.INTRODUCTION

Big data systems are extremely basic piece of this cutting edge organisations, because presently a days we are living in advanced world so at regular intervals a huge number of information is getting generate. This plenteous information just past the technology's. As little as a long time back nearly individuals just reasoning how to store tens to hundred of gigabytes information in our PC? be that as it may, today individuals figuring how to store tens to hundred of terabytes information? IBM overview gave that consistently 2.5 quintillion bytes of learning square measure made most that 90% of the information inside the world these days has been made inside the most recent 2 years. Intel's Info graphic uncovers each sixty seconds, 639,800GB of overall data is transferred, One moment of net time, 204 million messages sent. on-line inhabitants read twenty million photographs on Flickr. Twitter forms 100,000 new tweets and 320 new Twitter accounts are made.

The number and many-sided quality of cyber attacks has been expanding relentlessly lately. The significant players in todays cyber conflicts are efficient and vigorously subsidized groups with particular objectives and destinations, some of which are working under a state umbrella. Foes are focusing on the correspondence and data frameworks of government, military and modern associations and are ready to utilize a lot of cash, time and aptitude to achieve their objectives.

The period of big data and digital security is here. What's more, that implies both happenstance also, hazard for generally organizations. On the off chance that you are in the digital security field you are likely exceptionally comfortable with huge information, which is the term used to depict an extremely extensive informational collection that is mined and broke down to discover designs and conduct patterns. It is for the most part characterized as being thick in variety, velocity and volume. From a digital security point of view enormous information has introduced new potential outcomes as far as examination also, security answers for ensure information and anticipate future digital assaults. However, similarly as enormous information has opened up new potential outcomes for digital security groups, it has likewise given digital offenders the chance to get to mass amounts of touchy and individual data using progressed technologies. When digital offenders target enormous informational collections, the reward is regularly certainly justified regardless of the exertion expected to infiltrate security layers, which is the reason huge information displays such an awesome open door for organizations as well as for digital culprits. They have significantly more to pick up when they follow such a vast informational collection. Thus, organizations have significantly more to lose should they confront a digital assault without the best possible safety efforts set up.

The main objectives of the project,

1**)** To overview and analyse the present procedures and solutions of anti-phishing, and increase further learning through the comprehension of these techniques.
2) To lead an examination of new phishing attacks and potential threats.
3) To collect the proposed system requirements.
4) To design the proposed systems architecture.
5)To evaluate the resulting system.

II. SIGNIFICANCE OF THE SYSTEM

In this paper focus on detect phishing webpages with keyword based features for grouping phishing URLs and classifications using machine learning techniques.Thenhost on to the Azure Machine Learning Studio process with three models thatare,Two-Class Neural Network, Two-Class Boosted Decision Tree and Two-Class Decision Jungle.
The study of literature survey is presented in section III, Methodology is explained in section IV, section V covers the experimental results of the study, and section VI discusses the future study and Conclusion.

III. LITERATURE SURVEY

BIG data systems in huge associations today have turned out to be always mind boggling, multi-layered, multi-merchant, physically or intelligently circulated. The unpredictability offers ascend to a multi-faceted system of gadgets and applications all conceivably speaking to an attack vector or passage point into the corporate basic system. In expansion, associations are continually looked with a tremendous volume of recently found programming vulnerabilities and exposures. With the underlying driver of major cyber security issues to a great extent fixated on programming vulnerabilities, associations must keep up successful defencelessness' administration programs including recognizable proof, evaluation, remediation and revealing.
Looked with vast overabundances of uncertain vulnerabilities, associations can wind up responsive and caught off guard for new inundations of vulnerabilities and moving danger scene, especially if there are back to back a very long time of high volume and coordinated assaults abusing various vulnerabilities. Actually, it is additionally exceptionally trying for huge associations to secure their basic computerized resources and digital framework because of the mind boggling setups and limitations. In this manner, it is basic to extend our comprehension on the multiplication of newfound vulnerabilities. A factual structure to find patterns and examples in helplessness divulgences empowers associations to becomemore proactive in dealing with these vulnerabilities.

MingJian Tang, Mamoun Alazab, et al. Study on not only handling persistent volatilities in the information as well as further disclosing multivariate reliance structure among various defenselessness dangers. In sharp difference to the current investigations on univariate time arrangement, we think about the more general multivariate case endeavoring to catch their captivating connections. Through our broad experimental examinations utilizing this present reality powerlessness information, A composite model can successfully catch and save long haul reliance between various defenselessness and endeavor disclosures.

S.Arunet al. Observing the client who buys things from Merchant. Phishing is an on the web trick that endeavors to swindle individuals of their own data, for example, charge card or ledger information. Then distinguish, find and evacuate the phishing E-mail. The client points of interest will be put away in web registry.Here show how the online business procedures can be executed with various situations that incorporate checking open administration approach commitments.Experimental comes about demonstrate that this approach can be a compelling method to expel phishing pages facilitated on servers around the globe. Moreover, there is degree to attempt improvement on more forceful procedures to address the issue of a non-responsive host Administrator that neglects to close down a phishing site.

Ram B. Basnetet al. Anatomy of phishing URLs that are made with the particular goal of imitating a put stock in outsider to trap clients into disclosing individual information. Dissimilar to past work in this area,use various freely accessible highlights on URL alone; in addition, compare execution of various machine learning methods and assess the viability of continuous utilization of this method.Applying it on realworld informational indexes, This proposed approach is very compelling in distinguishing phishing URLs with a blunder rate of 0.3% false positive rate of 0.2%

what's more, false negative rate of around 0.5% in this manner enhancing past outcomes on the vital issue of phishing discovery.

Gaurav patel et ala novel, algorithm to distinguish phishing sites, in light of the attributes of the hyperlinks utilized as a part of the phishing attacks and the substance in the site. way to deal with detect phishing sites by consolidating the approach in light of the attributes of hyperlinks utilized as a part of the phishing assaults and the substance examination approach utilizing TF-IDF algorithm and in this way expelling every one of the downsides from the current calculation known as Link Guard algorithm . Consequently utilizing the Phis Guard calculation the issue of false-positives and false-negatives in the Link Guard has evacuated and the precision of Link Guard has made strides.

IV. METHODOLOGY

The Internet is currently a prominent means for giving stimulation, conveying with companions, conveying internet business, and conveying educating materials. Be that as it may, a few people the world over are exploiting the obscurity given by the Internet to trick people with counterfeit offers, or by distorting themselves as real organizations. Phishing is the online trick that endeavour’s to dupe individuals of their own data, for example, charge card or ledger data, and username and watchword certifications. The online culprits are known as Phishers. Expectedly, mass E-mailing with a phishing join is the most famous approach to bait the casualties. Be that as it may, SMS messages, visit rooms, counterfeit include standards, counterfeit occupation offers, and phony program instruments have developed as another stage among Phishers. Analysts have proposed strategies to avert phishing assaults, Phishers are ending up progressively advanced in their methodologies. Phishing assaults frequently include thorough arranging and consolidate procedures to sidestep existing hostile to phishing instruments. The volume of phishing attack recommends that current hostile to phishing instruments are inadequate. This is basically because of actuality that they just adopt a responsive or inactive strategy to stemming the issue. That is, they just channel speculate messages, however don’t really successfully close down the issue at its source.

A) SYSTEM ARCHITECTURE

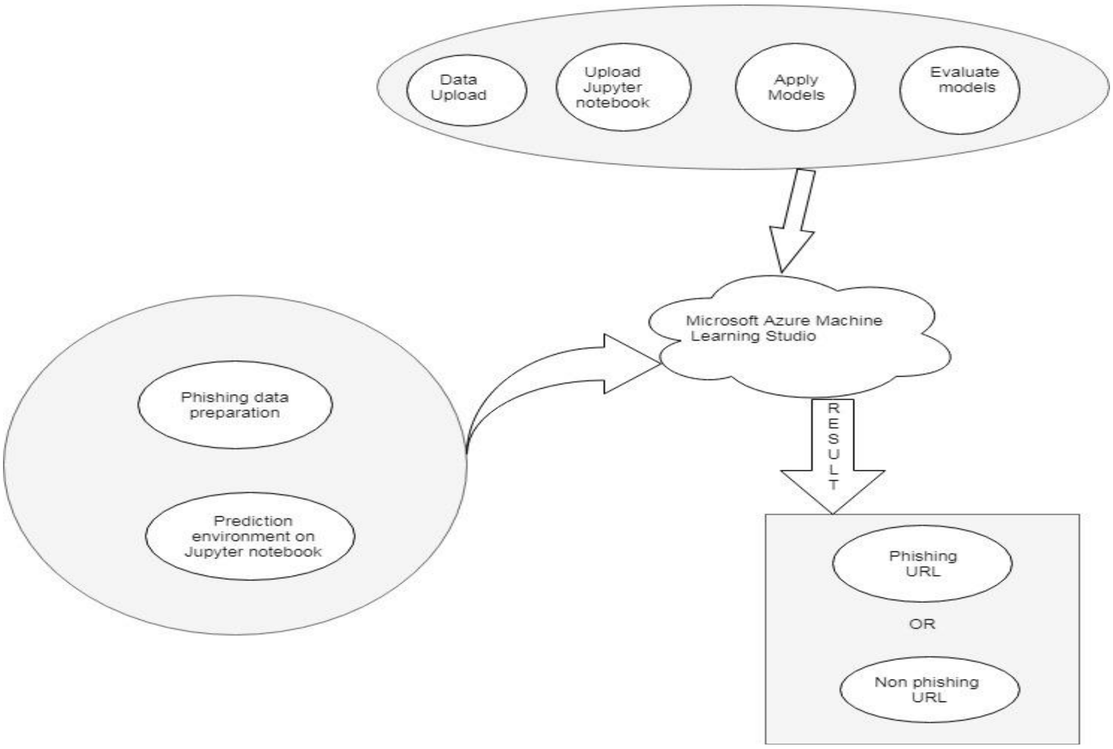

Data Upload
Upload Jupyter notebook
Apply Models
Evaluate models
Microsoft Azure Machine Learning Studio
Phishing data preparation
Prediction environment on Jupyter notebook
RESULT
Phishing URL
OR
Non phishing URL

Figure 1. System Architecture

The proposed system as shown in Fig3.1 General process of the system. This initial phase Business understanding focuses on understanding the project objectives and requirements from a business perspective, then converting this knowledge into a data mining problem definition and a preliminary plan designed to achieve the objectives.
The phishing data understanding phase starts with initial data collection and proceeds with activities that enable you to become familiar with the data, identify data quality problems, discover first insights into the data, and/or detect interesting subsets to form hypotheses regarding hidden information.
The phishing data preparation phase covers all activities needed to construct the final dataset from the initial raw data. Data preparation tasks are likely to be performed multiple times and not in any prescribed order. Tasks include table, record, and attribute selection, as well as transformation and cleaning of data for modelling tools.
Then create predictive environment on the jupyter notebook and open a accounting Machine learning studio workspace and upload dataset into the cloud.
Then various modelling techniques are selected and applied, and their parameters are calibrated to optimal values. Typically, there are several techniques for thesame data mining problem type. Some techniques have specific requirements on the form of data. Therefore, going back to the data preparation phase is often necessary. Here used three modules, one is Two-Class Neural Network module in AzureMachine Learning Studio, to create a neural network model that can be used to predicta target that has only two values. another is Two-Class Boosted Decision Treemodule in Azure Machine Learning Studio, to create a machine learning model that is based on the boosted decision trees algorithm and Two-Class Decision Junglemodule in Azure Machine Learning Studio, to create a machine learning model that is based on a supervised ensemble learning algorithm called decision jungles.
Evaluate the model stage in the project, you have built a model (or models)that appears to have high quality from a data analysis perspective. Before proceeding to final deployment of the model, it is important to thoroughly evaluate it and review the steps executed to create it, to be certain the model properly achieves the business objectives. A key objective is to determine if there is some important business issue that has not been sufficiently considered. At the end of this phase, adecision on the use of the data mining results should be reached.

B) WORK FLOW OF THE SYSTEM

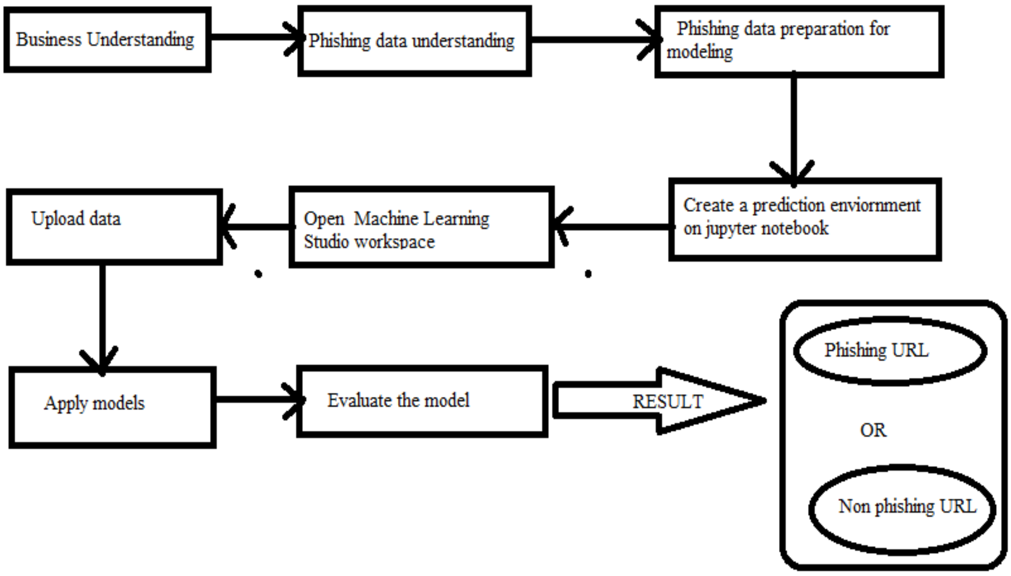

Business Understanding
Phishing data understanding
Phishing data preparation for modeling
Create a prediction enviornment on jupyter notebook
Open Machine Learning Studio workspace
Upload data
Apply models
Evaluate the model
RESULT
Phishing URL
OR
Non phishing URL

Figure 2. Work Flow of the system

C) MODULES

1. Business understanding
2. Phishing data understanding
3. Phishing data preparation for modeling
4. Create code in python
5. Create account on cloud
6. Upload data apply models
7. Evaluate the model

1. Business understanding: The principal objective is to altogether comprehend, from a business point of view, what the client truly needs to achieve. Frequently the client has numerous contending goals and limitations that must be legitimately adjusted. The examiners objective is to reveal imperative elements, toward the starting, that can impact the result of the task. A conceivable result of dismissing this progression is to use an awesome arrangement of exertion delivering the correct responses to the wrong questions. Record the data that is thought about the associations business circumstance toward the start of the project. List the phases to be executed in the undertaking, together with their span, assets required, sources of info, yields, and conditions. Where conceivable, make unequivocal the substantial scale cycles in the data mining process.

2. Phishing data understanding: List the dataset obtained, together with their areas, the strategies used to obtain them, and any issues experienced. Record issues experienced and any resolutions achieved. Describe the information that has been obtained, including Describe after effects of this assignment, including first discoveries or starting theory and their effect on the rest of the undertaking. In the event that suitable, incorporate diagrams and plots to demonstrate information attributes that recommend encourage examination of fascinating information subsets. List the after effects of the information quality check; if quality issues exist, list conceivable arrangements. Arrangements the organization of the information, the amount of information (for instance, the quantity of records and fields in each table), the personalities of the fields, and whatever other surface highlights which have been found. Assess whether the information obtained fulfils the applicable requirements. Describe consequences of this assignment, including to start with discoveries or beginning theory and their effect on the rest of the venture. On the off chance that proper, incorporate diagrams and plots to demonstrate information qualities that propose assist examination of fascinating information subsets.to information quality issues for the most part depend vigorously on the two information and business knowledge.
Data set contains Using the IP Address,Long URL to Hide the Suspicious Part, Using URL Shortening Services TinyURL, URLs having@ Symbol, Redirecting using //, Adding Prefix or Suffix Separated by (-) to the Domain, HTTPS(Hyper Text Transfer Protocol with Secure Sockets Layer), Domain Registration Length, Favicon, Using Non-Standard Port, The Existence of HTTPS Token in the Domain Part of the URL, Request URL, URL of Anchor, Links in <Meta >, < Script > and < Link >, Server Form Handler (SFH), Submitting Information to Email, Abnormal URL,Website Forwarding, Status Bar Customization, Disabling Right Click, Using Pop-up Window, IFrame Redirection, Age of Domain, DNS Record, Website Traffic, PageRank,Google Index, Number of Links Pointing to Page, Statistical-Reports Based Feature.

3. Phishing data preparation for modeling: Depict the dataset(s) that will be utilized for the demonstrating and the real investigation work of the project.Decide on the information to be utilized for investigation. Criteria incorporate significance to the information mining objectives, quality, and specialized limitations, for example, limits on information volume or information composes. Note that information choice spreads determination of segments and determination of records in a table. Describe what choices what's more. moves were made to address the information quality issues revealed amid the Confirm Data Quality undertaking of the Data Understanding stage. Changes of the information for cleaning purposes and the conceivable effect on the examination results ought to be considered. Describe the making of totally new records. The first field being a one of a kind identifier for each record or the last field being the result field the model is to predict.

4.Create code in python: Create a predictive environment on jupyter notebook using the random forestalgorithm.Random forest algorithm can use both for classification and the regression kind of problems. The Same algorithm both for classification and regression, the more trees in the forest the more robust the forest looks like. In the same way in the random forest classifier, the higher the number of trees in the forest gives the high accuracy results.

5.Create account on cloud: Azure Machine Learning is a cloud predictive analytics service that makes itpossible to quickly create and deploy predictive models as analytics solutions. The machine learning tools are mostly cloud-based services, Azure Machine Learning Studio is a collaborative visual development environment that helps you build, test, and deploy predictive analytics solutions in the cloud. So Sign in to your Machine Learning Studio workspace and get start.

6.Upload data apply models: Upload data or connect to data already in the cloud for the next step**.**

7.Evaluate the model: select the genuine displaying procedure that will be utilized. Despite the fact that you may have effectively chosen an apparatus amid the Business Understanding stage, this assignment alludes to the particular demonstrating strategy. Depict the planned arrangement for preparing, testing, and assessing the models. An essential part of the arrangement is deciding how to partition the accessible dataset into preparing, test, and approval datasets. With any demonstrating apparatus, there are frequently a substantial number of parameters that can be balanced. Rundown the parameters and their picked esteems, alongside the justification for the decision of parameter settings. Outline consequences of this undertaking, list characteristics of produced models , and rank their quality in connection to each other.

I. Two-Class Neural Network

A neural network is an arrangement of interconnected layers. The sources of info are the first layer, and are associated with a yield layer by a non-cyclic diagram contained weighted edges and hubs. Between the info and yield layers you can embed various covered up layers. Most prescient errands can be refined effectively with just a single or a hardly any concealed layers. In any case, late research has demonstrated that profound neural systems (DNN) with numerous layers can be exceptionally viable in complex errands, for example, picture or discourse acknowledgment. The progressive layers are utilized to demonstrate expanding levels of semantic profundity.

The connection amongst sources of inputs and outputs is found out from preparing the neural organize on the info information. The bearing of the chart continues from the information sources through the concealed layer and to the yield layer. All hubs in a layer are associated by the weighted edges to hubs in the following layer. To process the yield of the system for a specific info, an esteem is computed at every hub in the shrouded layers furthermore, in the yield layer. The esteem is set by figuring the weighted aggregate of the estimations of the hubs from the past layer. An actuation work is then connected to that weighted entirety.

Configuration of Two-Class Neural Network

1. Add the Two-Class Neural Network module experiment in Studio
2. Create trainer mode option.
3. Hidden layer specification.
4. Use Number of hidden nodes, and type the number of hidden nodes. The default is one hidden layer with 100 nodes.
5. For Learning rate, define the size of the step taken at each iteration, before correction.
6. A larger value for learning rate can cause the model to converge faster, but it can overshoot local minima.
7. For learning rate, define the size of the step taken at each iteration, before correction. A larger value for learning rate can cause the model to converge faster, but it can overshoot local minima.
8. For Number of learning iterations, specify the maximum number of times the algorithm should process the training cases. The initial learning weights diameter, specify the node weights at the start of the learning process.
9. For The momentum, specify a weight to apply during learning to nodes from previous iterations.
10. In The type of normalizer, select a method to use for feature normalization.
11. Select the Shuffle examples option to shuffle cases between iterations.
12. For Random number seed, type a value to use as the seed.
13. Select the Allow unknown categorical levels option to create a grouping for unknown values in the training and validation sets. The model might be less precise on known values but provide better predictions for new (unknown)values.
14. Add a tagged dataset to the experiment.
15. Run the experiment.

II. Two-Class Boosted Decision Tree

A boosted decision tree is a group learning technique in which the second tree amends for the blunders of the main tree, the third tree revises for the mistakes of the to begin with and second trees, et cetera. Expectations depend on the whole troupe of trees together that makes the prediction.

Configuration of Two-Class Boosted Decision Tree

1. Azure Machine Learning Studio, add the Boosted Decision Tree module to experiment.
2. Create trainer mode option.
3. Hidden layer specification.
4. For Maximum number of leaves per tree, indicate the maximum number of terminal nodes that can be created in any tree.
5. For Minimum number of samples per leaf node, indicate the number of cases required to create any terminal node in a tree.
6. For Learning rate, type a number between 0 and 1 that defines the step size while learning.
7. For Number of trees constructed, indicate the total number of decision trees tocreate in the ensemble. By creating more decision trees, you can potentiallyget better coverage, but training time will increase.
8. For Random number seed, optionally type a non-negative integer to use asthe random seed value. Specifying a seed ensures reproducibility across runsthat have the same data and parameters.
9. Select Allow unknown categorical levels option to create a group for unknownvalues in the training and validation sets.
10. Train the model

III. Two-Class Decision Jungle

By permitting tree limbs to consolidate, a choice DAG normally has a lower memory impression and preferable speculation execution over a choice tree, but at the cost of to some degree longer preparing time. Choice wildernesses are non-parametric models that can speak to non-straight choice limits. They perform coordinated feature choice and grouping and are versatile within the sight of uproarious features.

Configuration of Two-Class Decision Jungle

1. Azure Machine Learning Studio, add the Two-Class Decision Jungle moduleto experiment.
2. For Resampling method, pick the technique used to make the individualtrees. How you need the model to be prepared, by setting the Create trainer mode alternative.
3. For Number of decision DAGs, demonstrate the greatest number of graphs that can be made in the ensemble.
4. For Maximum depth of the decision DAGs, demonstrate the greatest number depth of graphs.
5. For Maximum width of the decision DAGs, demonstrate the greatest number width of graphs.
6. In Number of optimization steps per decision DAG layer, demonstrate how manyiterations over the data to perform when building each DAG.
7. Select the Allow unknown values for categorical features option to create agroup for unknown values in testing or validation data.
8. Add a tagged dataset to the experiment, and connect one of the training modules.

8. Evaluate the model

Evaluation brings about terms of business achievement criteria, including a last proclamation with respect to whether the task as of now meets the underlying business objectives. Summarize the procedure audit and feature exercises that have been missed and those that ought to be repeated. List the potential further activities, along with the explanations behind and against each option. Deploy the model and coordinate the workflow in applications by calling a web service.

V. EXPERIMENTAL RESULTS

It is implemented on adataset from UCI,Uses Two-class NN, Decision Jungle and Boosted trees to predict if a site is a phishing site or not. Figure 3. Added Conda and Python to the Environment Variables experiment in Machine Learning Studio. The output from the Evaluate Model module of Two-Class Neural Network and Two-Class Boosted Decision Tree shown in Figure 4 and Figure 5.The output from the Evaluate Model module of output from the Evaluate Model of Two-Class Boosted Decision Tree and Two-Class Decision Jungle shown in Figure 6 and Figure 7.Output from the Predictive experiment shown in Figure 8.Then output from the Predictive experiment web service shown in Figure 9.Figure 10 shows that Result of the prediction.

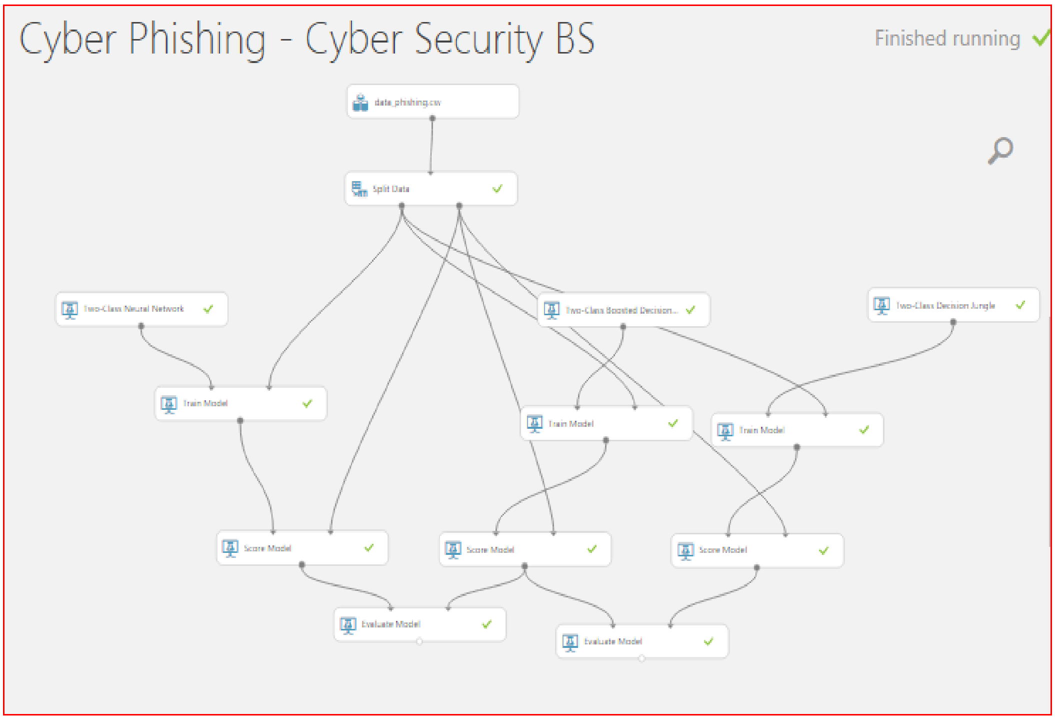
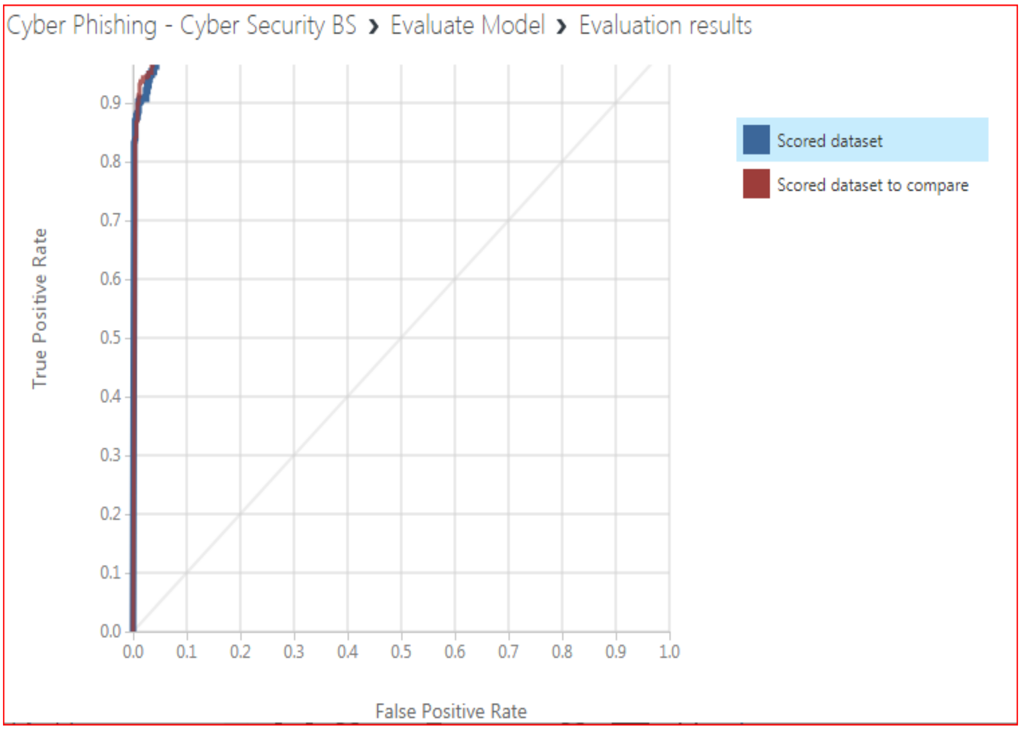
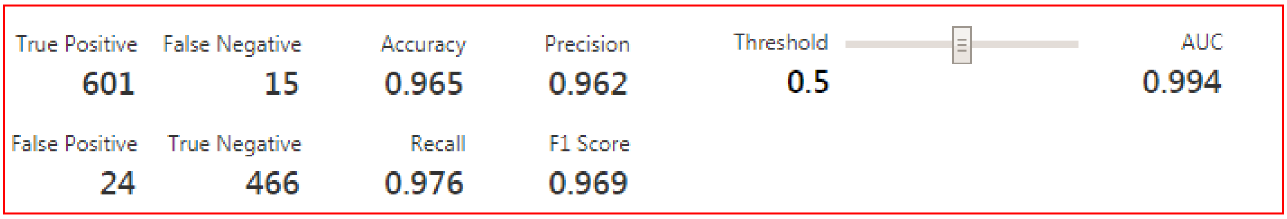


Figure 3.Added Conda and Python to the Environment Variables experiment



Cyber Phishing - Cyber Security BS › Evaluate Model › Evaluation results

Figure 4.output from the Evaluate Model of Two-Class Neural Network and Two-Class Boosted Decision Tree



True Positive	False Negative	Accuracy	Precision	Threshold	AUC
601	15	0.965	0.962	0.5	0.994
False Positive	True Negative	Recall	F1 Score		
24	466	0.976	0.969		

Figure 5.output from the Evaluate Model of Two-Class Neural Network and Two-Class Boosted Decision Tree

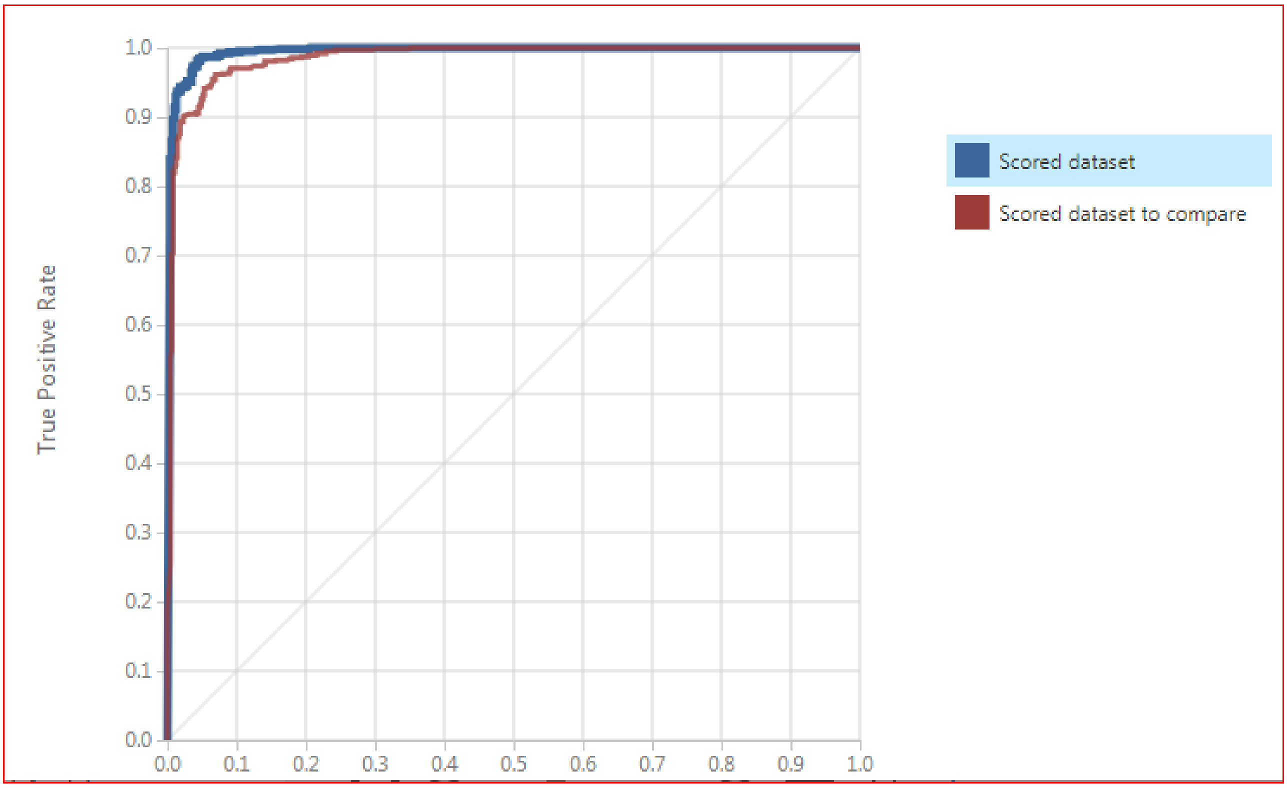
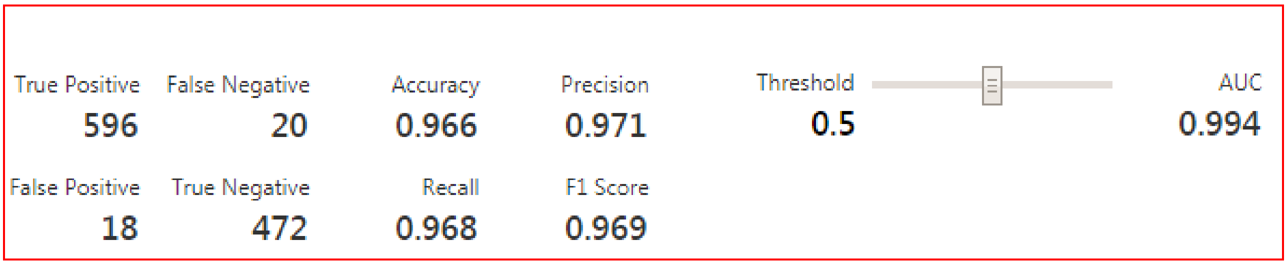


Figure 6.output from the Evaluate Model of Two-Class Boosted Decision Tree and Two-Class Decision Jungle



True Positive 596	False Negative 20	Accuracy 0.966	Precision 0.971	Threshold 0.5	AUC 0.994
False Positive 18	True Negative 472	Recall 0.968	F1 Score 0.969		

Figure 7 .output from the Evaluate Model of Two-Class Boosted Decision Tree and Two-Class Decision Jungle

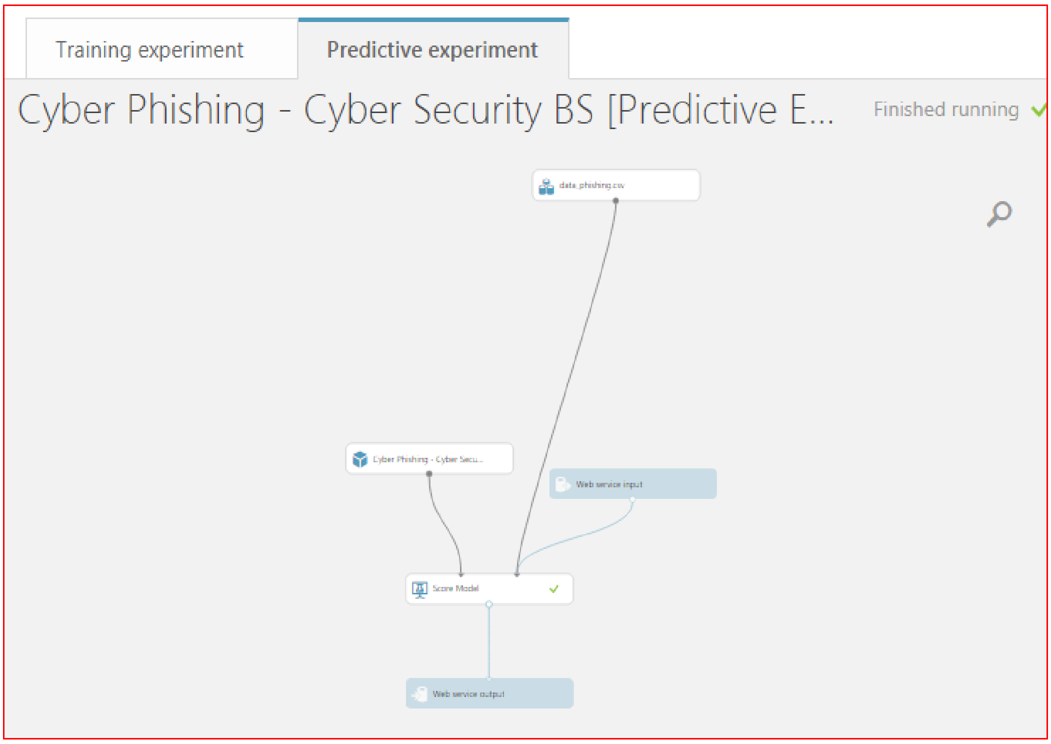
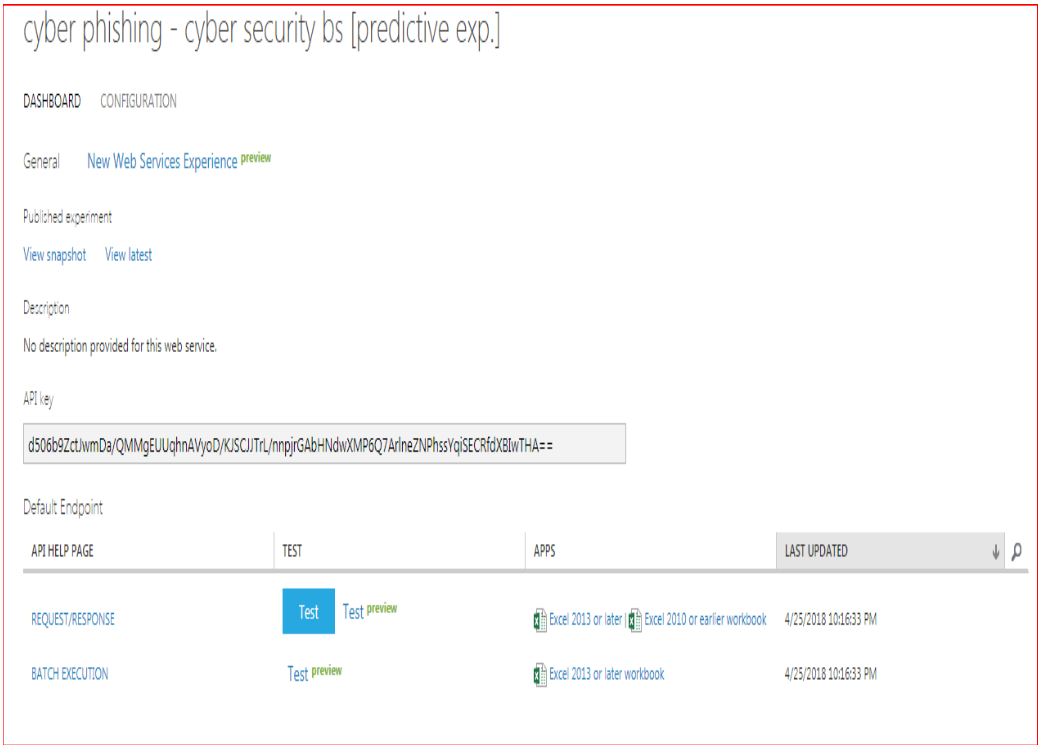


Figure 8.output from the Predictive experiment


cyber phishing - cyber security bs [predictive exp.]

DASHBOARD CONFIGURATION

General New Web Services Experience preview

Published experiment

View snapshot View latest

Description

No description provided for this web service.

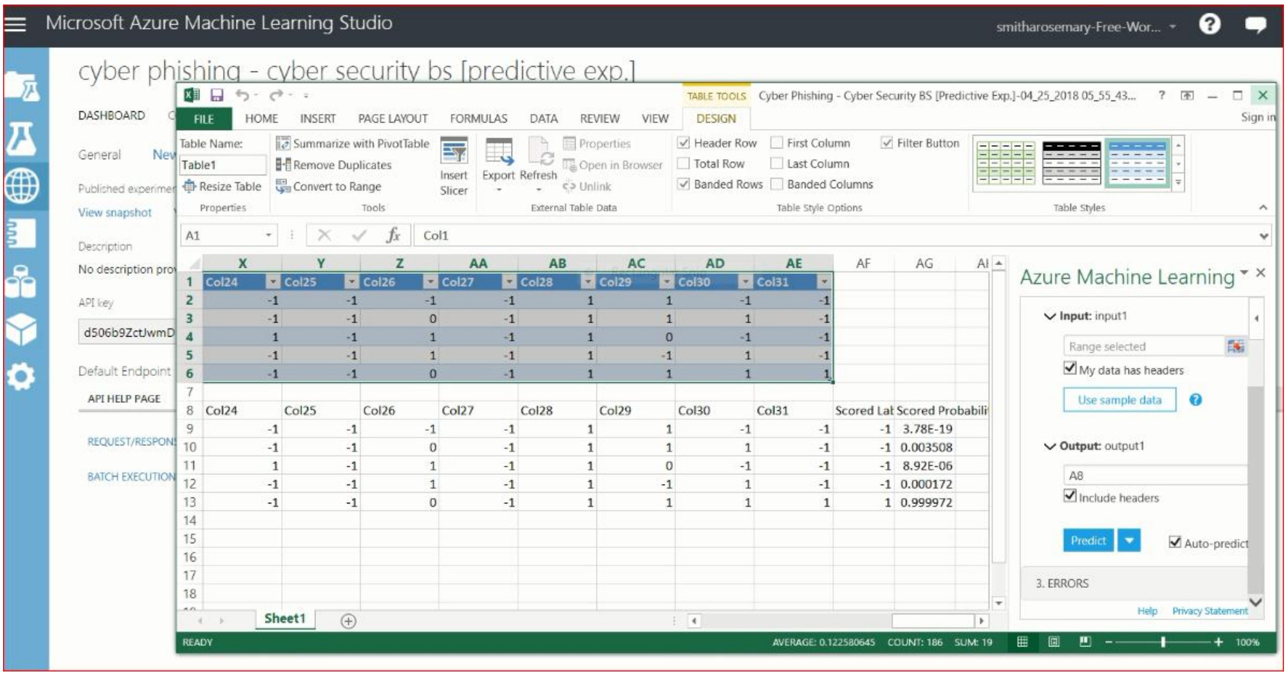
API key

d506b9ZctJwmDa/QMMgEUUqhnAVyoD/KJSCJJTrL/nnpjrGAbHNdwXMP6Q7ArlneZNPhssYqiSECRfdXBIwTHA==

Default Endpoint

API HELP PAGE	TEST	APPS	LAST UPDATED
REQUEST/RESPONSE	Test Test preview	Excel 2013 or later | Excel 2010 or earlier workbook	4/25/2018 10:16:33 PM
BATCH EXECUTION	Test preview	Excel 2013 or later workbook	4/25/2018 10:16:33 PM

Figure 9.output from the Predictive experiment web service



	X	Y	Z	AA	AB	AC	AD	AE	AF	AG
1	Col24	Col25	Col26	Col27	Col28	Col29	Col30	Col31		
2	-1	-1	-1	-1	1	1	-1	-1		
3	-1	-1	0	-1	1	1	1	-1		
4	1	-1	1	-1	1	0	-1	-1		
5	-1	-1	1	-1	1	-1	1	-1		
6	-1	-1	0	-1	1	1	1	1		
7										
8	Col24	Col25	Col26	Col27	Col28	Col29	Col30	Col31	Scored Lat	Scored Probabili
9	-1	-1	-1	-1	1	1	-1	-1	-1	3.78E-19
10	-1	-1	0	-1	1	1	1	-1	-1	0.003508
11	1	-1	1	-1	1	0	-1	-1	-1	8.92E-06
12	-1	-1	1	-1	1	-1	1	-1	-1	0.000172
13	-1	-1	0	-1	1	1	1	1	1	0.999972
14										
15										
16										
17										
18										

Figure 10.Result of the prediction

VI.CONCLUSION AND FUTURE WORK

So many techniques are existing to predict phishing attack on Url's but in this project take an extended look at the process of developing a predictive analytics solution in Machine Learning Studio.Here develop a model in Machine Learning Studio, and then deploy it as an Azure Machine Learning web service where the model can make predictions using new data.

In future work, plan to develop a efficient predictive environment using another approach and deploy it Machine Learning Studio for a large-scale real-world test and real world webpage testing for predict weather itis phishig attack or nonphishingattack.

REFERENCES

[1] MingJian Tang, Mamoun Alazab, Senior Member, IEEE, and Yuxiu LuoT (2016) ,Big Data for Cybersecurity:VulnerabilityDisclosure Trends and Dependencies.

[2] William Melicher,BlaseUr,SeanM.Segreti,SarangaKomanduri, LujoBauer,Nicolas Christin,Lorrie Faith Cranor,Fast,Lean,and Accurate: Modeling Password Guessability Using Neural Networks. 2016.

[3] GJ.Peters, Ed. New York: McGraw Hill, 1964, 15-64 S.Arun, D.Anandan,T.Selvaprabhu, B.Sivakumar, P.Revathi, H.Shine,Detecting phishing attacksin purchasing process through proactive approach,Advanced Computing:An International Journal (ACIJ), Vol.3, No.3, May 2012

[4] Ram B. Basnet, Andrew H. Sung, Quingzhong Liu,Learning to detectphishing urlsIJRET:eISSN: 2319-1163pISSN: 2321- 7308,

[5] GAURAV PATEL, An Approach to Detect Phishing Websites

[6] Token Based Security for Prevention of Phishing Attack at ClientSideInternational Association of Scientific Innovation and Research (IASIR)

[7] UCI Machine Learning Repository,https://archive.ics.uci.edu/ml/datasets/phishing+websites