JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 1383 - 1390
https://publishoa.com
ISSN: 1309-3452

Secure Cloud Framework Based on Machine learning Approach

Prasenjit kumar Das

Research Scholar, NIT Silchar, Silchar, Assam-788010 Indiaprasenjitdas139@gmail.com

Nidul Sinha

Professor, NIT Silchar, Silchar, Assam-788010 IndiaNidul.sinha@gmail.com

Annappa B
[3]*NITKSurathkal,Karnataka,575025INDIA*
annappa@gmail.com

ABSTRACT
In the present business scenario, Cloud Computing has taken a centerstageduetoitscost-effectiveness,efficiency,andscalability.Therehasbeenbroaduseof cloud-based systems and its services by most of the organizations in current times.And in order to safeguard the different transactions of information over the cloudenvironment,itisverymuchessential toprovideasecureplatformfortheusers.Therefore cloud security plays a significant role in ensuring confidentiality, integrity,andavailabilityofinformation.Thispapermainlyfocusedontheuseof Machinelearning(ML)algorithms as atool tosecuredatastoredin thecloud.It is worthmentioning that Machine learning has been widely used in analyzing data anomalies,predicting threats, classify data's, etc. in cloud-based system. The main objective of thispaperistoproposeasecurecloudframeworkwithtwodistinguishedparti.e.classification and encryption. Here we mainly focused on the classification of data usingone of the Machine Learning Algorithm i.e. Hybrid Naïve Bayes Algorithm where dataareclassified into threelevels viz. basic, sensitive, and highly sensitive. The proposedML algorithm is experimented using cloudsim simulating tool, results are analyzed andcompared with existing other ML classification algorithms namely K-Nearest Neighbor(KNN)andSupportVectorMachine(SVM).

Keywords—Cloud Security, Machine Learning, Classification, Cloud threats

1. INTRODUCTION

Data is a set of essential information. It is an important asset for any organization thatcouldbeinanyforms,i.e.numbers,words,imagesetc.Andthere hasbeenanexponential increase of such data's everyday which we nowadays termed as 'Big data'.According to an author in [1], Big Data refers to set of data's whose size is beyond theability of conventional database applications to store, manage and analyze. Hence, "BigData" has become very complex in every discipline of any business, educational, researchorganizations. Themostpublicresourcedatathatareavailableover internetincludes

various forms starting from text to multimedia data, social media data both in the form ofstructured and unstructured manner. Therefore to store and operate such huge sets ofdata's, Cloud Computing Technology is widely used. According to recent surveys, morethan 90 percent business establishment using some or the other cloud services. But withits growing adoption there's always a constant threat or vulnerabilities in cloud. Fear ofdata loss, different security threats, availability of information etc. are major concern incloud. And to combat such vulnerabilities and threats, several organizations are turningtowards use of Artificial Intelligence (AI) and Machine Learning (ML). ML algorithmsare widely used to analyze different types of anomalies, detection of threats through dataprocessing. In this paper we use Machine learning techniques specifically theclassificationalgorithmsastooltosecureclouddata'sbasedoncertainparameters.Asweknow that, Machine Learning used for delivering decision-making easy identification ofthreats as well as the patterns without use of any human intervention is the primaryadvantage of using ML [2]. Moreover it provides scope for continuous improvement in arobust environment. In recent times Machine learning has been used across wide range ofapplicationthatdeals allkinds ofvariantdata.

2. RELATEDWORK

CloudSecurityisanemergingareainthefieldofInformationSecurityandwiththeexponentialgrowthofdigitalinformationit'splayingasignificantrole.SeveralOrganizations have come up with effective measures to ensure the privacy of user's data fromboth internal and external attacks. Time and again researchers also have highly contributed inhandling the issues related to cloud security. Several authors in [3],[4],[5] mentioned the needof effective security measuresin cloud based system.As an instancein [3] the authorsaddressed the actual security and privacy issues on real-time cloud environment. Here theyhave analysed the vulnerabilities of Amazon Machine Images (AMIs) using different tools tomitigate any attacks which might lead to loss of information. Similarly in [4] and [5] authorssuggestedthatsecurityshouldbeprovidedasaserviceandproposedamodeloranarchitectureforsecurityasservice.ItsignifiesthatCloudConsumersorthecloudvendorscanprovidethesecurityapplicationsandservicesasperrequirementoftheorganizations.In
[6] authors proposed an architecture that secure data stored in cloud environment using somecryptographic algorithms Advances EncryptionStandard (AES) and Deffie Hellmen keyexchangemethodstoensureconfidentiality.Butmaindrawbackoftheproposedmethodology is that the computational overhead of the system, that leads to slow down of thesystem.HencetoovercomesuchtimecomplexityissuesMachineLearningalgorithmsbecome more useful.In the recent time there has been much interest in Machine Learning(ML)techniquesfornetworkandcloudsecurity [6][20].Ithas been widely usedtodetectanomalies, classify huge data's into level of importance etc.The surveys conducted by theauthors of [7],[8] shows that there has been significant development in the use of Machinelearning algorithms to reduce security threats. Similarly, authors in [9], [10] have shownseveral usage of machine learning approaches to improve the cloud environment and reducesecurityissuesrelated.Theauthorsin[11]comeupwithadataclassificationapproachthat

used several parameters based on certain dimensions like access control, storage and content.And each dimension are further categorised accordingly. Here the authors have analysedlimited data elements and classified them on the basis of the given parameters. Workingmodule has not been implemented as well as simulation has not been carried out in thisparticular work. Some other frameworks proposed for data classification and recognitionmentionedin [12], [13]. Forinstance,in [12] theauthors putforward aframeworkforextractingfeaturesfromimagedataandbuiltaclassifiermodelusingSupportVectorMachine (SVM). Similarly in [13][18] authors proposed a model for traffic detection by usingNaïveBayesClassifierincloudenvironment.Theauthorsin[14]havepresentedthefeasibilityandpossibilitiesof usingsupervisedmachinelearningalgorithmstoenhancesecurity in cloud based systems tree based machine learning models to classify the anomaliesfound in cloud based environment. Similarly, in [16][19] the authors have used unsupervisedlearning methods and proposed an automated model for cloud network analysis and auto-tuning. Hence it can be observed that with changing demands there has been significantdevelopmentintheuseofmachinelearning algorithmsforpromoting cloudsecurity[17].

3. PROPOSEDFRAMEWORK

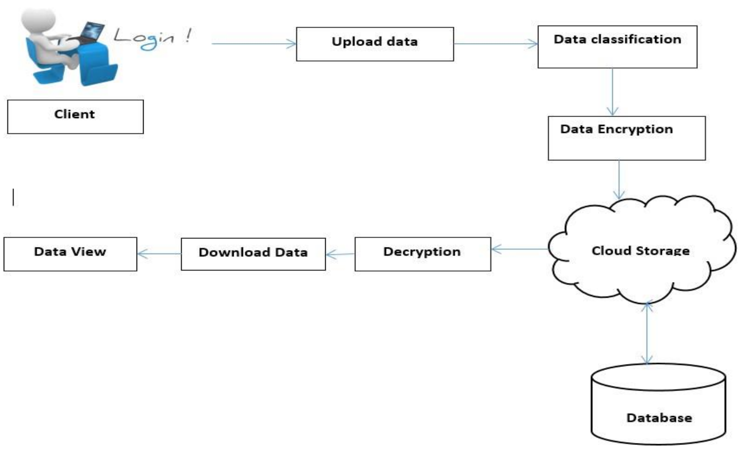

Figure1:CloudBasedFramework

In the above given figure 1, a simple overview of cloudbasedframework is designed withtwo very features. Firstly, classification of data's before storing in the cloud environmentusingclassifieralgorithmsandsecondlyencryptionofclassifieddata'sbasedonitsrequirement i.e. highly sensitive, sensitive and normal. In this paper we mainly focus on theclassification part where we classify cloud data's using different machine learning algorithmslike SVM,KNNandanImprovedNaïve Bayes algorithm.

Naïve Bayes' is a supervised classification algorithm based on probabilities and the mainessence of the classifier is based on Bayes Theorem. Here, in our proposed work we useddecisiontreealongwithNaïveBaye'salgorithmasaMetaClassifier,i.e.itisthe combinationofNaïveBayeswithDecisiontableofaDecisionTreealgorithm.MetaClassifier are generally defined as a proxy to the actual classifier, which is used to provideadditional data pre-processing. In this approach, we use Meta Learner scheme where theoutput of the Naïve Bayes is combined with Decision tablei.e. the Base Learner. BaseLearner are the algorithm used for building base classifiers which is Decision Tree. Here weuse level-0 model for thebase learner andlevel-1 model for the meta learner respectively.The predictions of level-0 are used as inputs of level-1 model to get the final prediction andthisprocessisalsoknownasensemblelearningasshowninfigure 2.

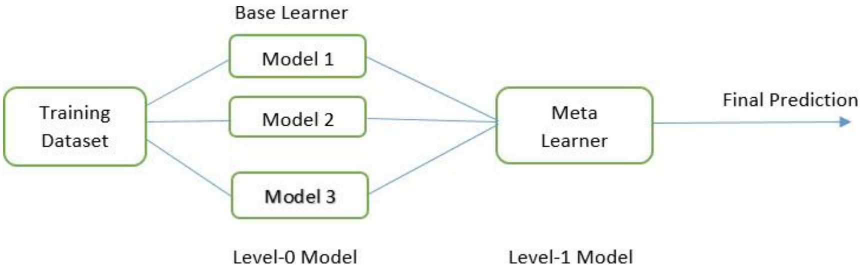

Figure2:EnsembleLearning

The hybrid Naïve Bayes approach would expected to work better than conventional classifierandtheparametersconsideredfortheevaluationofthegivenalgorithmsaredonebycalculatingClassificationtime,True positive RateandAccuracyrate.

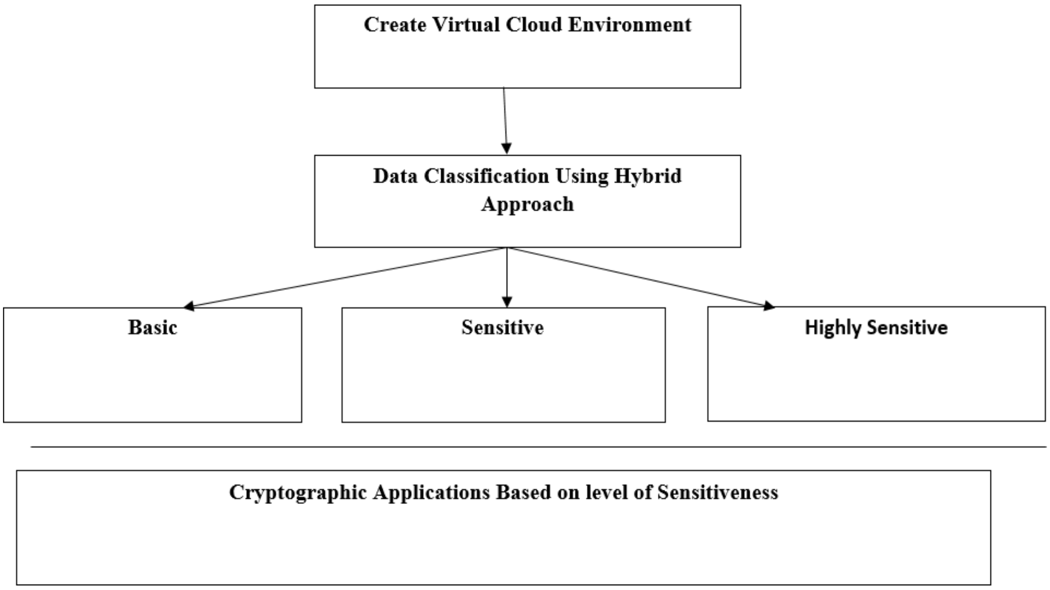
Create Virtual Cloud Environment
Data Classification Using Hybrid Approach
Basic
Sensitive
Highly Sensitive
Cryptographic Applications Based on level of Sensitiveness

Figure3:ProposedMethodology

As showninfigure3.thedataareclassifiedintothreecategoriesi.e.basic,sensitiveand highly sensitive respectively. After the successfully classifying the information, thosecan be used for further encryption based on its level of sensitiveness. In this work, wemainly focused on the first part of the methodology i.e. classification of data using hybridNaïve Bayes algorithm and compare its performance on given set of data's with other twoconventionalsupervisedclassifieri.e.SVM andKNN.

4. RESULTSANDDISCUSSION

This section displayed some experimental results and its details made in our performanceevaluation. In our experimental setup we use Cloudsim for simulation and Net Ide 8.2 inwindows Operating system. Cloudsim is a simulating tool that is used for simulating andmodelling of large scale data centres, describing the virtual machines, users and applications[16]. The experimental results here depicts the classification time, rate of accuracy and truepositiveratewhichhavebeenillustratedinthefollowingfigures.Beforestartingthesimulationitisnecessarytosettheproperti esofSaaS,PaaSandIaaSincloudsimenvironmentwhicharedescribedinthe giventables below:

Table1:ThepropertyofSaaSModel

IdNo.	Size ofcloudlet	Input Files(bytes)	Output Files(bytes)
0	4000	160	160
1	3000	135	135
0	4000	160	160

In the above Table 1, a SaaS model has been deployed with VMs in the cloud simulationenvironment. Here, Id No. represents the identification number of specific cloudlet, Lengthdescribesthecloudletsizeandthe Input/outputsize offilesmeasuredin bytes.

Table 2:The propertyofPaaSmodel

VMs IdNo.	MIPS	InputImage	Bandwidth(BW)	Processor Number	Virtual MachineManager
0	100	1500	1024	1	Xen
1	100	1500	1024	1	Xen

In the above Table 2, the property of PaaS model is described where Properties of virtualMachine created on the application deploymentlayers of the simulator. Here, the MIPSstands for Machine Instruction per second describes the CPU load and total capacity of VMsandhost.Bandwidth1000megabytesandprocessornumbersthatisusedinVMarementioned.

Table3:ThepropertyofIaaSmodel.

DataCentre ID	RAM in Mb	StorageLimit	ArchitectureofData	OperatingSystem	BW
2	2048	100000	X86	Windows	1000
3	2048	100000	X86	Windows	1000

Here in the table 3, data centers’ are assigned to VM storage limits, operating system andbandwidthis specified.

Table4:Comparingdifferent Classifiers

Classifier	**ClassificationTime(inms)**	**TPRateComparison**	**Accuracy(in %)**
KNN	1313	40.8	**50.222**
SVM	1022	53.5	**69**
HybridNaïve Bayes'	880	68.4	**78.41**

The above table shows the comparative analysis of different classifiers w.r.t. to classificationtime, True positive and the rate of accuracy. It is observed that the classifier Hybrid NaïveBayes gives much better results compared to other two classifiers. The performance analysisgraphsareshownbelow:

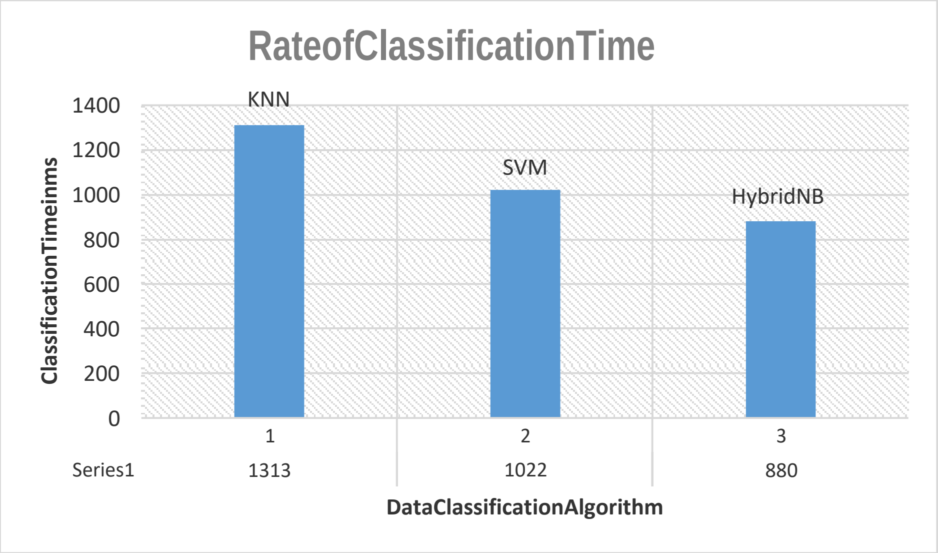

Figure4: Performance analysis of Classification Time

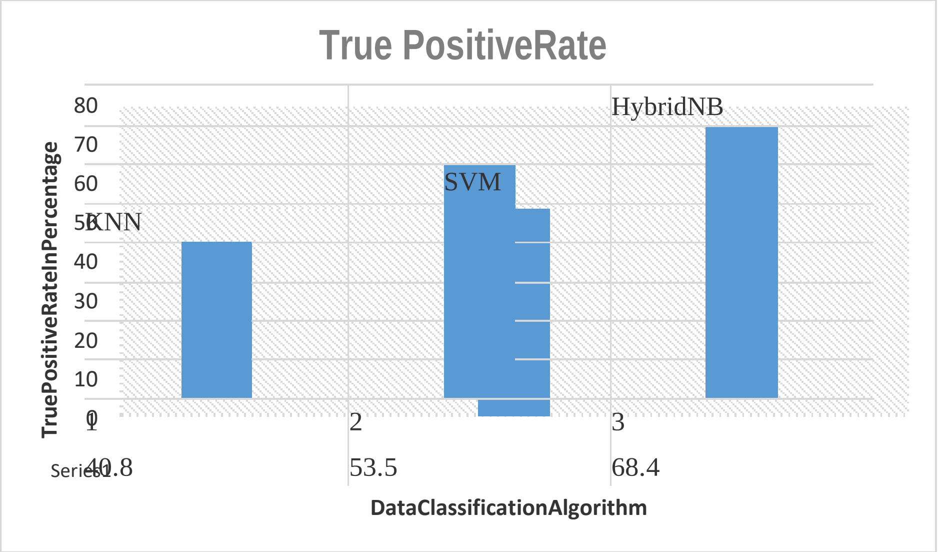

Figure5: Performance analysisofTPrate

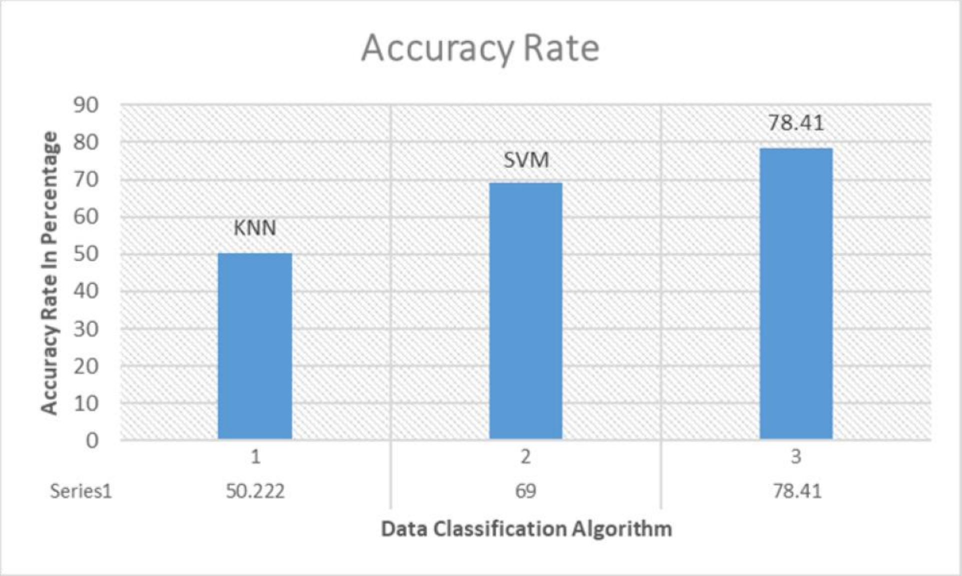

Figure6: Performance analysis of Accuracy Rate

As observed in given performance analysis graphs, it is to be mentioned that the classificationtime in figure 3 for the given set of information's is least in case of hybrid Naïve Bayescompared to other two classifiers i.e. SVM and KNN respectively. Moreover the true positiverate and the accuracy rate is comparatively better i.e. 68.4% and 78.41% respectively in caseof Hybrid Naïve Bayes than other two classifiers as shown in figure 4 and 5.From the aboveanalysis it can be seen that the proposed methodology results to better classification with useofHybridNaïveBayesalgorithm.

5. CONCLUSION

Data privacy preservation is one of the major issues while dealing with the storage in a cloudenvironment. And in recent times machine learning algorithms are widely used to ensureprivacy and security of information in cloud. In this paper we presented an efficientmethodof data classification using Hybrid Naïve Bayes algorithm. Naïve Bayes algorithm with theuse of decision tree as Meta classifier provided a better results compared to other supervisedalgorithmsKNNandSVM.Theproposedsystemhasbeensimulatedinadesignedsimulation environmentusing cloudsim simulator. And the experimental results depictthatthe given methodology took less computational time, better accuracy and true positive ratecompared to other two classifiers. The main objective of this work is to provide a betterclassification of data, based on levels of basic, sensitive and highly sensitive data so thatfurther it can be encrypted as required. In future, various available deep learning models canbe usedtoprovide cloudsecurityandthe performancecanbe compared.

REFERENCES

[1] Peter Mall, Timothy Grance.," The NIST Definition of cloud computing", NIST Special publication 800-145, 2011.

[2] Agarwal;S.Siddharth;P.Bansal"EvolutionofCloudComputingandRelatedSecurityConcerns.",SymposiumonColossalDataAnalysisandNetworking(CDAN),2016.

[3] S. Bugiel, S. Nurnberger, T. Poppelmann, A.R. Sadeghi, T. Schneider, AmazonIA: when elasticity snapsback, in: Proc of the 18th ACM Conference on Computer and Communications Security (CCS11), ACM,2011, pp. 389–400

[4] M. Hussain, H. Abdulsalam, SECaaS: security as a service for cloud-based applications, in: Proc. of theSecond KuwaitConferenceone-Services ande-Systems (KCESS11),ACM,2011,pp.1–4.

[5] Balaji, K., 2021. Load balancing in Cloud Computing: Issues and Challenges. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, *12*(2), pp.3077-3084.

[6] R. Laborde, F. Barre`re, A. Benzekri, Toward authorizationas a service: a study of the XACMLstandard, in: Proceedings of the 16th Communications & Networking Symposium, Society for ComputerSimulationInternational, 2013, pp. 1–7.

[7] Rewagad P, Pawar Y. "Use of digital signature with Diffie Hellman key exchange and AES encryptionalgorithmtoenhancedatasecurity incloudcomputing"inCommunicationSystemsandNetworkTechnologies (CSNT),2013InternationalConferenceon2013Apr6(pp.437-439).IEEE.

[8] TsaiC,HsuY,LinC,LinW(2009)Intrusiondetectionbymachinelearning:areview.ExpertSystAppl36(10):11994–12000.

[9] Balaji, K., P. Sai Kiran, and M. Sunil Kumar. "Power aware virtual machine placement in IaaS cloud using discrete firefly algorithm." *Applied Nanoscience* (2022): 1-9.

[10] Fernandes D, Soares L, Gomes J, Freire M, Inácio P (2014) Security issues in cloud environments: asurvey.IntJInfSecur13(2):113–170

[11] Balaji, K., P. Sai Kiran, and M. S Kumar. "An energy efficient load balancing on cloud computing using adaptive cat swarm optimization." *Materials Today: Proceedings* (2021).

[12] Palivela H, Chawande N, Wani A (2011) Development of server in cloud computing to solve issuesrelatedtosecurityandbackup. In:IEEECCIS, pp158–163

[13] Balaji, K., Kiran, P. S., & Kumar, M. S. (2020). Resource Aware Virtual Machine Placement in IaaS Cloud using Bio-Inspired Firefly Algorithm. *Journal of Green Engineering*, *10*, 9315-9327.

[14] Roshke S, Cheng F, Meinel C (2009) Intrusion detection in the cloud. In: Eighth IEEE internationalconferenceondependable, autonomicandsecurecomputing, pp729–734.

[15] Shaikh, Rizwana, and M. Sasikumar. "Data Classification for achieving Security in cloud computing."ProcediaComputerScience45(Elsevier):493-498, 2015.

[16] R.PakdelandJ.Herbert,"Acloud-baseddataanalysisframeworkforobjectrecognition,"inProceedings of the 5th International Conference on Cloud Computing and Services Science, 2015, pp.79–86

[17] T.T.N.SweSweAung,"NaveBayesclassifierbasedtrafficdetectionsystemoncloudinfrastructure," 6th InternationalConferenceonIntelligentSystems,ModellingandSimulation,2015.

[18] Bhamare D, Salman T, Samaka M, Erbad A, Jain R (2016) Feasibility of supervised machine learningforcloudsecurity.In:2016internationalconferenceoninformationscienceandsecurity (ICISS),Pattaya,pp1–5.

[19] Karn RR, Kudva P, Elfadel IA (2019) Dynamic auto selection and auto tuning of machine learningmodels forcloudnetworkanalytics.IEEETrans ParallelDistributeSyst30(5):1052–1064

[20] Rodrigo N. Calheiros1, Rajiv Ranjan2, Anton Beloglazov1, Cesar A. F. De Rose ´ 3 and RajkumarBuyya,"CloudSim:atoolkitformodelingandsimulationofcloudcomputingenvironmentsandevaluationofresource provisioningalgorithms"SOFTWARE–PRACTICEANDEXPERIENCESoftw. Pract. Exper. 2011; 41:23–50, 24 August 2010 in Wiley Online Library (wileyonlinelibrary.com).DOI:10.1002/spe.995.