International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE)
ISSN: 0976-1353 Volume 28 Issue 5 – MAY 2021.

A HYBRID APPROACH FOR REDUCING CARBON EMISSION IN INTERCLOUD ENVIRONMENT

Karthik Vishwanathan Iyer [#1], *S. Sivaraja* [#2], *A. Subramaniyan* [#3], *Dr.R.Kanniga Devi* [*4]

[#]*Student , Department of Computer Science and Engineering, Kalasalingam Academy of Research and Education, Anand Nagar, Krishnankoil, India, karthikviyer72@gmail.com*

[#]*Student , Department of Computer Science and Engineering, Kalasalingam Academy of Research and Education, Anand Nagar, Krishnankoil, India, rockeyraj001@gmail.com*

[#]*Student , Department of Computer Science and Engineering, Kalasalingam Academy of Research and Education, Anand Nagar, Krishnankoil, India, gunaasubramaniyan55@gmail.com*

[*]*Associate Professor, Department of Computer Science and Engineering, Kalasalingam Academy of Research and Education, Anand Nagar, Krishnankoil, India, rkannigadevi@gmail.com*

***Abstract*— In Cloud systems, Virtual Machines (VMs) are scheduled to hosts according to their instant resource usage (e.g. to hosts with most available RAM) without considering their overall and long-term utilization. Also, in many cases, the scheduling and placement processes are computational expensive and affect performance of deployed VMs. In this work, a Cloud VM scheduling algorithm that takes into account already running VM resource usage over time by analyzing past VM utilization levels in order to schedule VMs by optimizing performance by using knn and Naive Bayes classification technique. The Euclidean distance of knn is measured and then virtual machine is scheduled on the physical machine. The Cloud management processes, like VM placement, affect already deployed systems so the aim is to minimize such performance degradation. Moreover, overloaded VMs tend to steal resources from neighboring VMs, so the work maximizes VMs real CPU utilization. The results show that our solution refines traditional Instant-based physical machine selection as it learns the system behavior as well as it adapts over time. The concept of VM scheduling according to resource monitoring data extracted from past resource utilizations (including PMs and VMs).The count of the physical machine gets reduced by four using K-NN & NB classifier than Support Vector Machine (ACO) classifier. The task performed by 28 physical machine when using ACO is reduced by 24 physical machine by using knn&nb classifier algorithm also the error rates gets decreased by 0.025%.**

Index Terms- ***Virtual Machines, Machine Learning, Classification Algorithms, Cloud.***

I. INTRODUCTION

The Cloud Computing is the next generation computational paradigm. It is rapidly consolidating itself as the future of distributed on-demand computing .By using the concept of virtualization, Cloud Computing is emerging as vital backbone for the varieties of internet businesses. On the other hand, Internet enabled business (e-Business) is becoming one of best business model in present era. To fulfill the need of internet enabled business, computing is being transformed to a model consisting of services that are commoditized and delivered in a manner similar to traditional utilities such as water. Users can access services based on their requirements without regard to where the services are hosted or how they are delivered. Several computing paradigms have promised to deliver this utility computing. Cloud computing is one such reliable computing paradigm. Cloud computing architecture consists of a front end and a back end. These two ends are connected by Internet or Intranet. The front end comprises of client devices like thin client, fat client or mobile devices etc. The clients need some interface and applications for accessing the cloud computing system. The back end consists of the various servers and data storage systems. There is also a server called "Central Server". A central server is used for administering the cloud system. It also monitors the overall traffic and fulfilling the client demands in real time. The paper is structured as follows. The related works is discussed in Section 2, Section 3 elaborates scope and project plan, System requirements and Project design are provided in Section 4, Section 5 consists of Project testings, Section 6 Conclusion and future works.

II. RELATED WORKS

We [1] Youngyuet al has proposed this paper Remote data integrity checking (RDIC) enables a data storage server, say a cloud server, to prove to a verifier that it is actually storing a data owner's data honestly In this paper, we investigated a new primitive called identity-based remote data integrity checking for secure cloud storage. We formalized the security model of two important properties of this primitive, namely, soundness and perfect data privacy. We provided a new construction of of this primitive and showed that it achieves soundness and perfect data privacy. Both the numerical analysis and the implementation demonstrated that the proposed protocol is efficient and practical. In PDP, the data owner generates some metadata for a file, and then sends his data file together with the metadata to a remote server and deletes the file from its local storage. To generate a proof that the server stores the original file correctly, the server computes a response to a challenge from the verifier. The verifier can verify if the file keeps unchanged via checking the correctness of the response. PDP is a practical approach to checking the integrity of cloud data since it adopts a spot-checking technique. Specifically, a file is divided into blocks and a verifier only challenges a small set of randomly chosen clocks for integrity checking [2] UsmanWaziret al has proposed this paper Cloud computing provides distributed resources to the users globally. Cloud computing contains a scalable architecture which provides on-demand services to the organizations in different domains. However, there are multiple challenges exists in the cloud services. Different techniques has been proposed for different kind of challenges exists in the cloud services. This paper reviews the different models proposed for SLA in cloud computing, to overcome on the challenges exists in SLA. Challenges related to

Performance, Customer Level Satisfaction, Security, Profit and SLA Violation. We discuss SLA architecture in cloud computing. Then we discuss existing models proposed for SLA in different cloud service models like SaaS, PaaS and IaaS. In next section, we discuss the advantages and limitations of current models with the help of tables. In the last section, we summarize and provide conclusion In this survey, we discuss some of SLA parameters for consumers that must consider these parameters before signing SLA in cloud platform.

[3] PritiNarwalet al has proposed this paper Cloud Computing is a new evolutionary and dynamic platform that makes use of virtualization technology. In Cloud computing environment, virtualization abstracts the hardware system resources in software so that each application can be run in an isolated environment called the virtual machine and hypervisor does the allocation of virtual machines to different users that are hosted on same server. Although it provides many benefits like resource-sharing, cost-efficiency, high-performance computability and decrease in hardware cost but it also imposes a number of security threats. The threats can be directly on Virtual Machines (VMs) or indirectly on Hyper-visor through virtual machines that are hosted on it. This paper presents a review of all possible security threats and also their countermeasures by using GameTheoretic approaches. Game Theory can be used as a defensive measure because of independent and strategic rational decision making nature of cloud users where each player would compete for best possible solution in a secure manner is dealt with. For future work, games with complete information have been discussed so far but still there is a lot of work to explore in those games where player is unaware of other player's estimated loss. In that case, it would become difficult for a player to make a decision to opt or not to opt for a secured hypervisor.

[4] Nitin Kumar Sharmaet.al has proposed this paper Attribute Based Access Control (ABAC) models are designed with the intention to overcome the shortcomings of classical access control models (DAC, MAC and RBAC) and unifying their advantages. In ABAC, the access control is provided based on generic attributes of entities. Many organizational security policies condition access decisions on attributes. OWL can be used to formally define and process security policies that can be captured using ABAC models. We have defined models, domains, data and security policies in OWL and used a reasoner to decide what is permitted. In this paper we present a way to represent the ABACα model using Web Ontology Language (OWL).In ongoing work we are modeling more complex policies by capturing the ABACβ model in OWL.

[5]Ziad Ismail.al has proposed this paper The new developments in cloud computing have introduced significant security challenges to guarantee the confidentiality, integrity, and availability of outsourced data. A Service Level Agreement (SLA) is usually signed between the cloud provider and the customer. For redundancy purposes, it is important to verify the cloud provider's compliance with data backup requirements in the SLA. There exists a number of security mechanisms to check the integrity and availability of outsourced data. This task can be performed by the customer or be delegated to an independent entity that we will refer to as the verifier. However, checking the availability of data introduces extra costs, which can discourage the customer of performing data verification too often. The interaction between the verifier and the cloud provider can be captured using game theory in order to find an optimal data verification strategy. Interestingly, our results show that a NE of the game exists and when it is achieved, the CP cannot improve his utility by acting dishonestly. At the NE, it is as if the trust of the TPA in the CP's actions outweigh any belief of a potential misconduct.

[6] Jin Liet.al has proposed this paper Identity-Based Encryption (IBE) which simplifies the public key and certificate management at Public Key Infrastructure (PKI) is an important alternative to public key encryption. However, one of the main efficiency drawbacks of IBE is the overhead computation at Private Key Generator (PKG) during user revocation. Efficient revocation has been well studied in traditional PKI setting, but the cumbersome management of certificates is precisely the burden that IBE strives to alleviate. In this paper, aiming at tackling the critical issue of identity revocation, we introduce outsourcing computation into IBE for the first time and propose a revocable IBE scheme in the server-aided setting.Therefore, even if a revoked user and either of the KU-CSPs collude, it is unable to help such user re-obtain his/her decryptability. Finally, we provide extensive experimental results to demonstrate the efficiency of our proposed construction.

[7] Xun Yiet.al has proposed this paper Recently, the paradigm of data mining-as-a-service in cloud computing environment has been attracting interests. In this paradigm, a company (data owner), lacking data storage, computational resources and expertise, stores its data in the cloud and outsources its mining tasks to the cloud service provider (server). In order to protect the privacy of the outsourced database and the association rules mined, k-anonymity, k-support, and k-privacy techniques have been proposed to perturb the data before it is uploaded to the server. However, additional modifications to our protocols are needed in order to yield the correct results. We plan to consider this as our future work and will investigate its performance when deploying our solutions in the cloud environment provided by the cloud service providers.

III. SCOPE AND PROJECT PLAN

Scope And Objective

- To enhance the vm scheduling .
- To improve the efficiency of classification algorithm
- To implement k-nearest neighbor's method
- To implement naive bayes (nb) method
- To define the weight of the PM according to the resource usage of the VMs optimization scheme

Project Plan

Module Description

- Vm Scheduling
- Classification Algorithm
  - ✓ k-nearest neighbor's method
  - ✓ Naive Bayes (NB) method
- Optimization Scheme

Vm Scheduling

The algorithm enhances the VM selection phase based on real time monitoring data collections and analysis of physical and virtual resources. Our aim is to strengthen VM scheduling .In order to incorporate criteria related to the actual VM utilization levels, so VMs can be placed by minimizing the penalization of overall performance levels. The optimization schemes involve analytics to the already deployed VMs to include (a) maximization of utilization levels and (b) minimization of the performance drops.A monitoring engine that allows online resource usage monitoring data collection from VMs. The engine is capable of collecting system data based on interval and stores it to an online cloud service that makes it available for data processing. Data is collected every a tiny time interval (e.g. 1 second) and is stored in a temporary local file.

Classification Algorithm

When supervised machine learning algorithms are considered for classification purpose, the input dataset is desired to be a labeled one.

K-Nearest Neighbor's Method

K-nearest neighbors is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure (e.g., distance functions). K-NN has been used in statistical estimation and pattern recognition. The k-nearest neighbor's algorithm (k-NN) is a non-parametric method used for classifications and regression.

Naive Bayes (NB) method

The Naive Bayes Classifier technique is based on Bayesian theorem and is particularly used when the dimensionality of the inputs is high. The Bayesian Classifier is capable of calculating the most possible output based on the input. It is also possible to add new raw data at runtime and have a better probabilistic classifier. A naive Bayes classifier considers that the presence (or absence) of a particular feature (attribute) of a class is unrelated to the presence (or absence) of any other feature when the class variable is given. For example, a fruit may be considered to be an apple if it is red, round. Even if these features depend on each other or upon the existence of other features of a class, a naive Bayes classifier considers all of these properties to independently contribute to the probability that this fruit is an apple. Algorithm works as follows,

P (label |features) = P (label) * P (features |label)
P (features)
P(C|X) = P(X|C)P(C)
P(X)
P(C|X)=P(X1|C)* P(X2|C)* ……P(Xn|C)*P(C)

In equation P(c|x) is the posterior probability of class (target) given predictor (attribute) of class. P(c) is called the prior probability of class. P(x|c) is the likelihood which is the probability of predictor of given class. P(x) is the prior probability of predictor of class. Bayes theorem provides a way of calculating the posterior probability, P(c|x), from P(c), P(x), and P(x|c). Naive Bayes classifier considers that the effect of the value of a predictor (x) on a given class (c) is independent of the values of other predictors

Optimization Scheme

The aim of this optimization schemes is to define the weight of the PM according to the resource usage of the VMs. This will reveal information about the already deployed VMs status, like indications that a workload is running or not. To achieve this we provide two optimization schemes. Here classification of the VM status about its current resource usage is classified using the knn and nb shown in fig 4.1.Initially the virtual machine resource usage dataset is collected and monitored and then the collected data is classified using the machine learning methods like K-NN and NB.

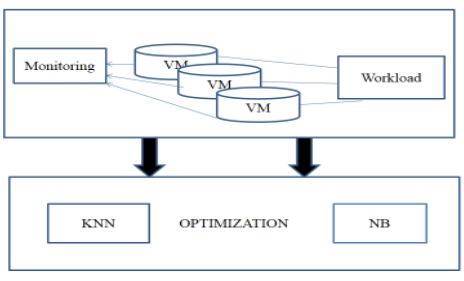


Figure 1: VM resource monitoring process

IV. SYSTEM REQUIREMENTS AND DESIGN

Hardware Requirements:

- Processor Type ` : Pentium i3
- Speed : 3.40GHZ
- RAM : 4GB DD2 RAM
- Hard disk : 500 GB
- Keyboard : 102 Standard Keys
- Mouse : Optical Mouse

Software Requirements:

- Operating System : Windows 10
- Front end : Netbeans • IDE 8.2
- Coding Language : Java

V. SYSTEM ANALYSIS

Existing System

The concept of VM scheduling according to resource monitoring data extracted from past resource utilizations (including PMs and VMs) and the resource data are classified using the optimization methods K-NN and NB, thus performing the scheduling.A classification model attempts to draw some conclusion from observed values. Given one or more inputs a classification model will try to predict the value of one or more outcomes. Outcomes are labels that can be applied to a dataset. There are two approaches to machine learning: supervised and unsupervised. In a supervised model, a training dataset is fed into the classification algorithm.The k-nearest neighbor's algorithm (k-NN) is a non-parametric method used for classification and regression. In both cases, the input consists of the k closest training examples in the feature space. The output depends on whether k-NN is used for classification or regression.

Disadvantages:

Virtual Machines are scheduled to hosts according to their instant resource usage (e.g. to hosts with most available RAM) without considering their overall and long-term utilization. Also, in many cases, the scheduling and placement processes are computational expensive and affect performance of deployed VMs. Thus the traditional VM placement algorithm does not consider past VM resource utilization levels.

Advantages:

Simple to implement Flexible to feature / distance choices Naturally handles multi-class cases Can make probabilistic predictions.Handles continuous and discrete data Not sensitive to irrelevant features.

VI. PROJECT DELIVERABLES

The multi-domain dataset includes the various resource utilization from of cloud resources such as bandwidth, memory ,cpu .the entire domain consist of 1000 labeled instances which are assumed here as the past resources utilization history record.

Table 1: characteristic features of dataset

Simulation Setup

No.of .virtual machine: 16
No.of.physical machine: 20
No.of classifiers: 02

Performance Evaluation

The absolute error is defined as the absolute value of the difference between the measured value and the true value. Thus, let:

e_a = the absolute error

x_m = the measured value

x_t = the true value

The formula for computing absolute error is:

$$e_a = | x_m - x_t |$$

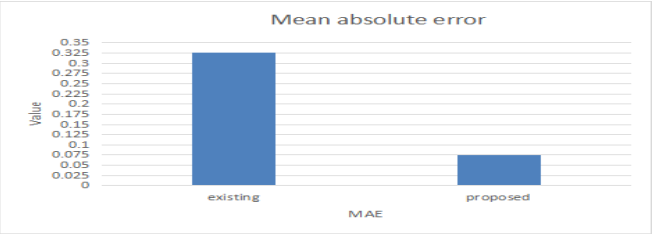


Figure 2:Mean Absolute Error

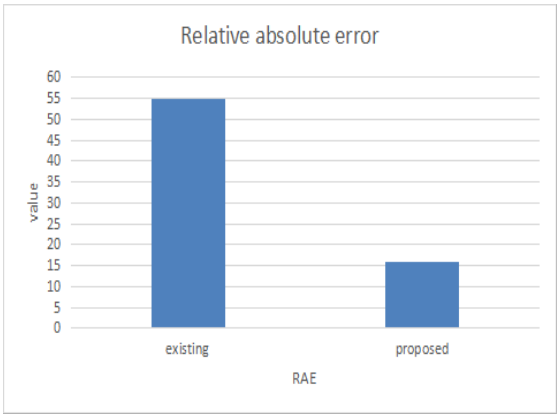


Figure 3: Relative Absolute Error

Fundamental Design Knowledge

Input Design:
The input design is the link between the information system and the user. It comprises the developing specification and procedures for data preparation and those steps are necessary to put transaction data in to a usable form for processing can be achieved by inspecting the computer to read data from a written or printed document or it can occur by having people keying the data directly into the system.
The input design consist two forms

- Pre-processing Form
- Attribute Selection Form

Pre-Processing Form:
The amount of data decreased due to the existence of missing values, gross errors and dead band errors, which have been removed from the dataset. However, replacing the removed and missing values is beyond the scope of this research. Furthermore, the dataset does not include seasonal effects for accurate prediction of upcoming environmental variable.In order to identify temporal outliers, are utilized to obtain temporal patterns and differences of the real-life data set, named Wisconsin Lung Cancer is used. The data set is publicly available on UCI machine learning repository

Attribute Selection Form:
Concept-independent preprocessing and concept-specific sampling. The first step is concept-independent, thus allowing for learning di®erent concepts at a later stage. The second step is concerned with the actual selection of a relevant subset of the instances once the concept of interest has been identified. Following the selection of the subset, the relevant learning algorithm (EigFusion) is applied to obtain the classifier.The DNNs as the classifier algorithm, multiple hyperplanes would need to be learned, one for each class. Essentially, the second step dictates the reuse of the training dataset with varying labels. Such a scenario is common in data repositories where the data is available for pre-processing with periodic updates adding/removing instances.

Output Design:
The output Design which meets the requirements of the end user and presents the information clearly. In any system results of processing are communicated to the users and to other system through outputs. In output design it is determined how the information is to be displaced for immediate need and also the hard copy output. It is the most important and direct source information to the user. Efficient and intelligent output design improves the system's relationship to help user decision-making.

- The output design consist Result Extraction Form
- Comparison And Graph Analysis Form

Resu lt Extraction that presented a novel clustering algorithm EigFusion is designed to perform clustering on rich structured multivariate datasets. it have shown that the applicability of deep neural network are not limited to classification problems. Even in the absence of labelled training instances for DNN, generating labels on-the-fly effectively increases clustering performance. The result on the integration of authorship analysis with topical clustering of documents show significant improvements over traditional Eig-Fusion and confirms that there is great benefit in incorporating additional dimensions of similarity into a unified clustering solution. Since spatiote mporal outlier detectio n might

CLOUD RESOURCE S	3	1000

turn out to be useful in many different research fields, we hope that this work will spark further interest in such problems that are challenging and relatively unexplored.

Design Constraints And Standards

The application considers two constraints, namely, economic, and sustainability. Economic constraint is satisfied by providing plausible success rate and survey communications. The main operation of the application pertains to recommendation algorithms, providing personalized suggestions to organization with respect to their success rate and tender statistics. The entire application is built on top of open-source frameworks leveraging java programming. Hence the application is sustainable. This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSC.2016.2638900, IEEE Transactions on Services Computing

VII. PROJECT TESTING

Types Of Testing:

There are three testing phases

- Unit testing
- Validation testing
- Output Testing

Unit Testing:

Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program input produces valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application. It is done after the completion of an individual unit before integration.We implemented the two algorithms implementation using the selecting text file generation optimization. Additionally, we implemented the Sequential Pattern, Hybrid and Inverse Term Frequency algorithms as presented in the previous section. All these algorithms were implemented in Java using several of the data structures provided by the JAVA Standard Template Library. All experiments reported in this thesis were performed on a 1.40GHz Intel Processor with 1GB main memory, running Windows 8.

Validation Testing:

The procedure of validation testing is to keep the system safe from errors. The basic purpose is to conform that the system satisfies the necessary conditions. This testing is planned when the incorrect data is given, the user receives the error message. A test plan output lines the classes of test to be conducted and test a procedure device specific test case that will be used to demonstrate conformity with the requirements. Being different from normal system behaviour, intrusion detection is a perfect candidate for applying outlier detection techniques. The key challenges for outlier detection are :-

- Huge Data Volume: This calls for computationally efficient techniques.
- Streaming Data: This requires on-line analysis.
- False alarm rate: Smallest percentage of false alarms among millions of data objects can make be overwhelming for an analyst.

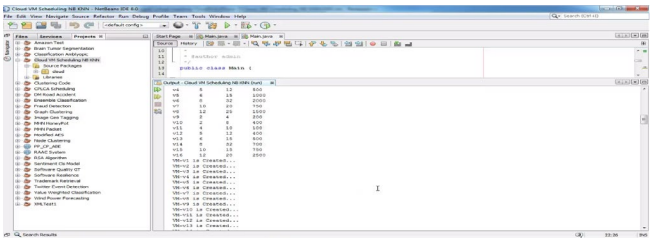
Output Testing:

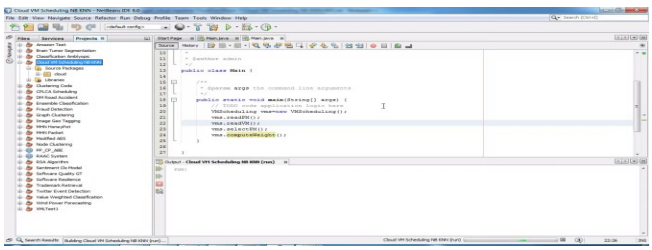
After performing the validation testing the next step is the output testing of the proposed system. Since no system could be useful if it does not produce the required output format. Asking the users about format required by them test the output generated or displayed by the system under consideration. Here the output form is considered on screen and in printed format.

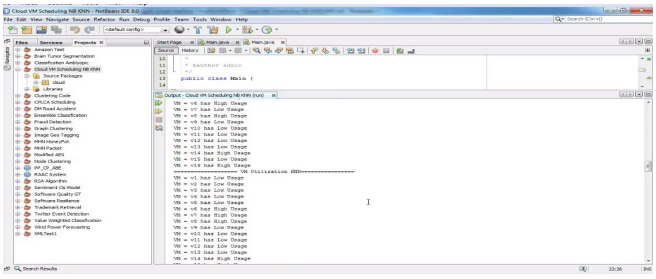
An important aspect for any outlier detection technique is the manner in which the outliers are reported. Typically, the outputs produced by outlier detection techniques are one of the following two types:

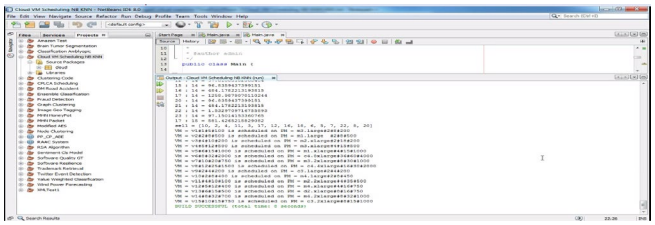
- Scores: Scoring techniques assign an outlier score to each instance in the test data depending on the degree to which that instance is considered an outlier. Thus the output of such techniques is a ranked list of outliers. An analyst may choose to either analyze top few outliers or use a cut-off threshold to select the outliers.
- Labels: Techniques in this category assign a label (normal or anomalous) to each test instances.

Sample Pictures:









VIII. CONCLUSION

Different virtual machine placement algorithms were used for scheduling by choosing physical machines according to the system data (i.e. usage of CPU, memory, bandwidth) in cloud system. The present VM placement doesn't take into account of realtime VM resource utilization levels. Here we a new VM placement algorithm based on past VM usage experiences is proposed then the VM usage is monitored and the data gets trained using machine learning models (K-NN&NB) to calculate the prediction of the VM resource usage, to place VMs accordingly. An algorithm that allows VM placement according to PM and VM usage levels and computational learning method based on the concept of analyzing past VM resource usage according to historical records to optimize the PM selection phase was introduced. Also, a VM placement algorithm based on real time virtual resource monitoring was introduced where machine learning models is used to train and learn from previous virtual machine resources usage. Thus, a monitoring engine is assumed with resource usage data. The count of the physical machine gets reduced by 4 by using knn&nb classifier than Support Vector Machine (SVM) classifier. The task performed by 28 physical machine when using SVM is reduced by 24 physical machine by using knn&nb classifier algorithm also the error rates gets reduced by 0.025%.

Future Work:

The proposed work allows data processing based on a timeframe window to define the PMs or VMs actual behavior. In case of VM placement method, result highlights the major improvements. The future research work may be carried out with further experimentation relevant to various machine learning models like random forest, decision trees to improve the performance.

REFERENCES

[1] Y. Yong, M. H. Au, and G. Ateniese, Identity-basedremote data integrity checking with perfect data privacypreserving for cloud storage, IEEE Transactions onInformation Forensics and Security, vol. 12, no. 4, pp.767–778, 2017.

[2] U. Wazir, F. G. Khan, S. Shah, Service level agreementin cloud computing: A survey, International Journal of Computer Science and Information Security, vol. 14, no.6, p. 324, 2016.

[3] P. Narwal, D. Kumar, and M. Sharma, A review ofgame-theoretic approaches for secure virtual machineresource allocation in cloud, in Proceedings of theSecond International Conference on Information andCommunication Technology for Competitive Strategies,2016.

[4] N. K. Sharma and A. Joshi, Representing attribute basedaccess control policies in owl, in 2016 IEEE TenthInternational Conference on Semantic Computing(ICSC),2016, pp. 333–336.

[5] Z. Ismail, C. Kiennert, J. Leneutre, and L. Chen,Auditing a cloud provider's compliance with databackup requirements: A game theoretical analysis, IEEETransactions on Information Forensics and Security, vol.11, no. 8, pp. 1685–1699, 2016.

[6] E. Furuncu and I. Sogukpinar, Scalable risk assessmentmethod for cloud computing using game theory(CCRAM), Computer Standards & Interfaces, vol. 38, pp.44–50, 2015.

[7] J. Li, J.W. Li, and X. F. Chen, Identity-based encryptionwith outsourced revocation in cloud computing, IEEETransactions on Computers, vol. 64, no. 2, pp. 425–437,2015.

[8] X. Yi, F. Y. Rao, and E. Bertino, Privacy-preservingassociation rule mining in cloud computing, inProceedings of the 10th ACM Symposium on Information,Computer and Communications Security, 2015, pp.439–450.

[9] J. Lou and Y. Vorobeychik, Equilibrium analysis ofmulti-defender security games, in Proceedings of theTwenty-Fourth International Joint Conference on ArtificalIntelligence (IJCAI), 2015, pp. 596–602.