ii. Cloud Storage

Kamara & Lauter et al. [32] recommended providing a digital personal garage that may satisfy a variety of needs (privacy, integrity, authentication, etc.). The majority of requirements are met by encrypting files stored in the cloud. However, with collaboration technology, such encryption ends in rigour at each search approach and real-time altering through files. Figure 3 depicts the cryptographic garage carrier's structure, which can be utilised to address back-up, archival, fitness document structures, static statistics trading, and e-discovery security issues" [9]. It is made up of three main components: the Data Processor (DP), which reads data before transferring it to the cloud, the Data Verifier (DV), which verifies the data's integrity, and the Token Generator (TG), which allows the bearer issuer to access the files. enables recovery Before sending statistics to the cloud, Alice encrypts and encodes files with metadata (tags, timing, length, etc.) using a statistics processor, then transmits them to the cloud.

B. Proposed Model Main Building Block

This version includes cypher cloud version and cloud statistics encryption, both of which are mostly based on quantum cryptography, so that: I key technology and key control are primarily based on DKD to improve the supply and reliability of cloud computing encryption. Deploy decryption and strategy mechanisms.

(ii) Manipulate heavy computing techniques that aren't compatible with non-public computers.

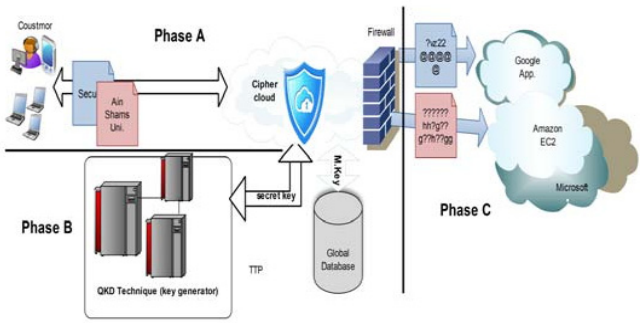

Fig. 4 Framework of Proposed Architecture

The suggested version performs a number of computations before generating data on the fly in a cloud environment; those calculations may be summarised in three easy steps, as illustrated in Fig. 4: Enterprise, DKD, and Open Cloud parts. - EC (Enterprise Control): EC (Enterprise Control) is a term that refer Customers execute various pre-processing activities on the entered data in this area before transferring it to the cloud environment via the following steps:

1. Aspects of the Customer: Stop using Mobile Scope for Users, Enterprises, and Remote Locations.
2. Cipher Cloud: Encryption and decryption issues for data and attached documents. This is further reinforced by the use of encryption algorithms such as AES, DES, and RSA.

DKD: DKD is a practical stationary technique in which all responsibilities are determined using quantum physics and computing factors. Although it is a blend of traditional cryptography, the concept of facts, and quantum physics [26], [27], it is not a natural mathematical progression. Within the proposed version, the DKD is the most significant portion; it has been understood as being dependent on the 1/3 section (TTP), which is responsible for key technology, key control, and key distribution. These keys are used to encrypt user-uploaded files or documents, which are mostly based on a set of fully symmetric encryption principles (AES). Furthermore, it has been taken into consideration thus far because it lies in the midst of the planned version, it is extremely difficult to notice or exploit. It is, however, simple to use and maintain, and it eliminates the computational layout complexity that classical cryptography entails.
Open Cloud Phase: This is the ideal portion for absorbing and calculating the percentage of files, packages, or attached documents on the Internet, such as Google Apps and Amazon EC2.

IV. SECURITY CHALLENGES IN CLOUD

When it relates to privacy and security, the cloud poses significant risks. People should certify, just like merchants, that the cloud of harassment is free of issues such as knowledge loss or data theft.

There is a possibility that a malicious user or hacker would enter the cloud under the guise of a normal user, affecting many of us who are infected or using the afflicted cloud. Cloud computing can be used in a variety of scenarios:

i. Data theft
ii. Data integrity
iii. Security issue
iv. Loss of sensitive data
v. Corrupt code
vi. Data abbreviation
vii. Data security at the business level
viii. Data security at the user level

The present generation of cloud computing features give no privacy against untrustworthy cloud operators, therefore vital data such as medical records, financial records, or high-impact corporate data are not asked to be stored.

To deal with this, we have a tendency to employ a variety of approaches that vary from theory to practise. The most

typical application of coding is to provide confidentiality by abstracting all useful information from plaintext. Coding transforms useless knowledge into something that can't be accessed.

To solve this challenge, we're developing cryptosystem algorithms that make it easier to do a range of calculations on encrypted raw data, ranging from traditional computations to special-purpose computations. Fully homomorphic encryption, searchable encryption, organised encryption, and practical encryption are all examples of homomorphic cryptography research.

a. Storage proofs. A buyer must check whether the cloud operator's Data Victimization Proof of Storage has been tampered with. It is frequently avoided by shoppers who save a clone of the information associate in nursing, despite the fact that it does not require any data to be stored back. In reality, the work is excellent.

b. ***Secure Storage system.*** We get a tendency to strive to design cloud storage systems that protect client data from a hostile cloud provider in terms of confidentiality, security, and integrity. The systems will need to be high-functioning and use the newest cryptanalytic coding techniques such as homomorphic encryption, searchable encryption, verifiable computation, and proof of storage, among others, to ensure confidentiality without sacrificing security.

V. CONCLUSION

Cloud computing is a rapidly evolving technology that has become the new trend, with many companies and large corporations migrating to the cloud. However, due to security concerns, the insulation is lagging behind. Cloud security is the ultimate edifice that can smash the flaws in huge multinationals', corporations', and organizations' cloud acceptance.

In the cloud, there are numerous security algorithms that can be used. DES, Triple-DES, AES, and Blowfish are examples of fundamentally symmetric algorithms. Symmetric algorithms are used in DES and AES because they are relatively secure. AES is more difficult to implement than DES. The algorithms used by RSA and Diffie-Hellman Key Exchange are very different.

Cryptography can be used for cloud security in a variety of ways. Cryptography will be utilized for cloud data access control, cloud data trust management, verifiable computing, cloud knowledge authorization and authentication, and secure data storage, to name a few applications.

Aside from these, full lattice based mainly cryptography and ID based cryptography are two major areas in the gifting world that ensure cloud data security. In this area, there is still a lot of research to be done.

REFERENCES

[1] Peter Mell, and Tim Grance, "The NIST Definition of Cloud Computing,"2009,

[2] http://www.wheresmyserver.co.nz/storage/media/faq-files/clouddef- v15.pdf, Accessed April 2011.

[3] Frank Gens, Robert P Mahowald and Richard L Villars. (2009, IDC Cloud Computing 2010).

[4] IDC, "IDC Ranking of issues of Cloud Computing model," ed, 2009, http://blogs.idc.com/ie/?p=210, Accessed on July 2011.

[5] Cloud Computing Use Case Discussion Group, "Cloud Computing Use Cases Version 3.0," 2010.

[6] Cloud Security Alliance (CSA). (2010). Available: http://www.cloudsecurityalliance.org/

[7] Meiko Jensen, Jörg Schwenk, Nils Gruschka and Luigi Lo Iacono, "On Technical Security Issues in Cloud Computing," in IEEE ICCC, Bangalore 2009, pp. 109-116.

[8] Bernd Grobauer, Tobias Walloschek and Elmar Stöcker, "Understanding Cloud-Computing Vulnerabilities," IEEE Security and Privacy, vol. 99, 2010.

[9] Balachandra Reddy Kandukuri, Ramakrishna Paturi and Atanu Rakshit, "Cloud Security Issues," in Proceedings of the 2009 IEEE International Conference on Services Computing, 2009, pp. 517-520.

[10] Kresimir Popovic , Zeljko Hocenski, "Cloud computing security issues and challenges," in The Third International Conference on Advances in Human-oriented and Personalized Mechanisms, Technologies, and Services, 2010, pp. 344-349.

[11] S. Subashini, Kavitha, V., "A survey on security issues in service delivery models of cloud computing," Journal of Network and Computer Applications, vol. In Press, Corrected Proof.

[12] Lohr, Steve. "Cloud Computing and EMC Deal." New York Times. Feb. 25, 2009. pg. C 6.

[13] McAllister, Neil. "Server virtualization." InfoWorld. Feb. 12, 2008. Retrieved March 12, 2008.

[14] http://www.infoworld.com/article/07/02/12/07FEvirtualserv_1.html

[15] Markoff, John. "An Internet Critic Who Is Not Shy About Ruffling the Big Names in High Technology." New York Times. Apr. 9, 2001. pg.C6

[16] G. Clarke, Microsoft's Azure Cloud Suffers First Crash, The Register, March 16, 2009, http://www.theregister.co.uk/ 2009/03/16/azure_cloud_crash/

[17] P. Ferrie, Attacks on Virtual Machine Emulators, White Paper, Symantec Corporation, January 2007, http://www.symantec.com/avcenter/reference/Virtual_Machine_Threats. pdf

[18] G. Fowler, B. Worthen, The Internet Industry is on a Cloud – Whatever That May Mean, The Wall Street Journal,March 26, 2010

[19] L. Youseff, M. Butrico, D. D. Silva, Toward a Unified Ontology of Cloud Computing, Grid Computing Environments Workshop, held with SC08, November 2008.

[20] http://www.cs.ucsb.edu/~lyouseff/CCOntology/CloudOntology.pdf
[21] G. Ateniese, R. Di Pietro, L. V. Mancini, and G. Tsudik. Scalable and e_cient provable datapossession. In Proceedings of the 4th international conference on Security and privacy in communication netowrks (SecureComm '08), pages 1{10, New York, NY, USA, 2008. ACM.
[22] Hodges, A. (2005), 'Can quantum computing solve classically unsolvable problems"
[23] H.K. Lo, H.F. Chau, Unconditional security of quantum key distribution over arbitrary long distances. Science 1999; 283(5410): 2050-2056.
[24] http://www.doc.ic.ac.uk/~nd/surprise_97/journal/vol4/spb3/
[25] L. Lydersen, Wiechers, C., Wittman, C., Elser, D., Skaar, J. and Makarov, V. Hacking commercial quantum cryptography systems by tailored bright illumination. Nat. Photonics 4, 686, 2010.
[26] K. Inoue, Quantum Key Distribution Technologies. IEEE Journal of Selected Topics in Quantum Electronics, vol. 12, no.4, July/August 2006.
[27] http://ewh.ieee.org/r10/bombay/news4/Quantum_Computers.htm
[28] http://www.bbc.co.uk/news/scienceenvironment-16636580
[29] http://science.howstuffworks.com/science-vs-myth/everyday-myths/quantum-cryptology.htm
[30] G. Brassard, T. Mor and B. C. Sanders, "Quantum cryptography via parametric downconversion", in Quantum Communication, Computing, and Measurement ,P. Kumar, G. Mauro D'Ariano and O. Hirota (editors), Kluwer Academic/Plenum Publishers, New York, 2000, pp. 381.
[31] P. D. Townsend, \Experimental investigation of the performance limits for first telecommunications-window quantum cryptography systems", IEEE Photonics Technology Letters, Vol. 10, 1998, pp. 1048.
[32] J. Brodkin. Gartner: Seven cloud-computing security risks. http://www.infoworld.com/d/security-central/gartnerseven-cloud- computing-security-risks-853, 2008.
[33] C.C.A: CipherCloud Gateway Architecture, www.ciphercloud.net.
[34] M. v. Dijk and A. Juels. On the impossibility of cryptography alone for privacy-preserving cloud computing. In Hot topics in Security (HotSec'10), pages 1{8. USENIX Association, 2010.
[35] Kamara and Lauter . CS2: A Searchable Cryptographic Cloud Storage System,IJSIR,2012.
[36] G. Ateniese, S. Kamara, and J. Katz. Proofs of storage from homomorphic identi_cation protocols. In Advances in Cryptology - ASIACRYPT '09, volume 5912 of Lecture Notes in Computer Science, pages 319{333. Springer, 2012.

ABOUT THE AUTHOR

Jaishree Jain is an Assistant Professor Computer Science & Engineering Department, Ajay Kumar Garg Engineering College, Ghaziabad, affiliated to AKTU, Lucknow, Uttar Pradesh, INDIA. She has 11 years teaching experience in CSE & IT Departments that include Chandigarh University, Punjab and college of AKTU, University, Lucknow since July 2010. She had done her B.Tech. in Information Technology from Uttar Pradesh Technical University, Lucknow. She had qualified GATE two times in 2007 & 2008 with 94.86 percentile. She had done her M.Tech. in Software Engineering from MNNIT, Allahabad. She had published about 27 research papers in SCI/SCOPUS/ UGC-Care/ UGC/National and International Journals Conferences and Chapter published in International Book. She had also been published one patent in August 2018. Her research interests include Image Processing, Cloud Security, and Steganography. She is also the reviewer of Journal of Supercomputing which is listed in Springer Journal and reviewed many papers till now. She is the member of professional bodies' i.e. life time member of ISTE, lifetime member of ICSES and member of Engineering Council of India (ECI) and position entitled as a "Professional Engineer" by ECI. INDIA.

THE IMPLEMENTATION OF GABASS IN TO BANK DATA

[1]Pradeep Gupta, [2]Jay Kant Pratap Singh
[1]Assistant Professor, Ajay Kumar Garg Engineering College,Ghaziabad, UP,India
[2]Assistant Professor, Ajay Kumar Garg Engineering College,Ghaziabad, UP,India
[1]guptapradeep@akgec.ac.in [2]yadavjaykant@akgec.ac.in

Abstract: **Subgroup identification assesses the acceptability of a subset of the best characteristics as a group. Subset determination is characterized in many approaches, here we are discussing the implementation tool with what suitable parts are required and how they are applied. The method proposed is GABASS and it is applied over bank data. The implementation of the GABASS algorithm is discussed in this paper.**

Keywords: **Genetic Algorithm, GUI, JAVA, TANGARA.**

I. INTRODUCTION

In this paper, we are discussing complete details of the implemented tool along with the descriptions of the result obtained in the form of snapshots. Java tool developed to automatically select a subset of GABAS-based features. A GUI is also included with the tool, which is further illustrated in-depth.

This paper aims to improve the classification accuracy of current work tests on banking data by using 11 existing features and 601 examples. The data is divided into two parts one is training and the other is testing,the verification process is repeated several times, so that at least once during the procedure, each occurrence in the dataset can be utilized as training data. K-fold cross-validation, 2-fold cross-validation, and one-off cross-validation are the most prevalent approaches in cross-validation.K data is separated into related portions in k-cycle cross-validation. The test set is chosen from among the k components, and the remainder is combined to create the classifier [2].The classification algorithm is trained and tested ten times in the proposed approach.The cross-validation data is separated into ten subgroups, with each subgroup subdivided based on the classification rule established into the remaining nine subgroups.For each train test setting, ten separate test outcomes are obtained, and the common result algorithm determines the test's correctness.

The data set is collected from the source bank ARFF file as input in this proposed approach. The ARFF attribute is an ASCII text file that describes a set of circumstances in which a collection of characteristics can be shared. After that, all of the database's features are encrypted. Some characteristics are chosen at random. The classification accuracy is calculated using the features that have been chosen. GABASs were used to increase classification accuracy and refresh the attribute set. The procedure is repeated until the completion criteria are satisfied. We'll get a subset of the top attributes and classification accuracy after that's done.The accuracy of the system classification given above was calculated using the TANGARA programme SIPINA Toll.

II. JAVA

The appropriate language for implementing the method is JAVA, which is platform neutral thanks to the JVM. In JAVA JDK 1.8, our feature selection mechanism is implemented as GABAS. It may run on any JVM-enabled operating system, including Windows, Linux, and UNIX. The (org.um. feri. ears. algorithm. GABAS) package is used to develop the GABAS algorithm. Discretization of feature value could aid in calculating each feature's fitness and information gain value. We use the discretization package [3], which is supported by Weka 3.7.0, for each feature, to discretize a group of values. The result is a feature subset that may be used to build a Nave Bayes classifier to categorize the accuracy.

A. SIPINA Tool of TANGARA

TANGARA is SIPINA's successor, and it uses supervised learning methods, association rules, feature selection, and the creation of custom algorithms. TANGARA is open-source software because it is written in Java. TANGARA's major goal is to make data mining software that is simple to use, especially in terms of the user interface and how to use it. The second purpose of TANGARA is to build an architecture that would allow users to simply add their own data mining algorithms and compare their results [4].

II. WORKING OF GABASS

a) Using java code, generate random subsets of attributes from bank data. There are 2n subsets of data D with n attributes. In general, the cost of computing data set D is O(n x |D| x log(|D|), where n is the number of characteristics and D is the number of instances. For m characteristics and n instances, the number of comparisons necessary is m*n2 [1].

A. GUI

As mentioned above that the Weka 3.7.0 is supported. Here we have to deal with datasets provided by the bank. The human-computer interface (GUI) employs windows, icons, pull-down menus, and a pointer, all of which can be controlled with a mouse. The utility has a graphical user interface (GUI) as depicted in Figure 1, with three command buttons labeled, (1) File, (2) Preprocess (3) Classify.

An ARFF format dataset file is viewed and taken as input by clicking the file button. An ARFF (Attribute-Relation File Format) report is an ASCII file that delineates a once-over of properties and shares a game plan.Figure 1 depicts one such list of characteristics. When you click on a specific attribute, you'll see its value and the number of times it's been used.

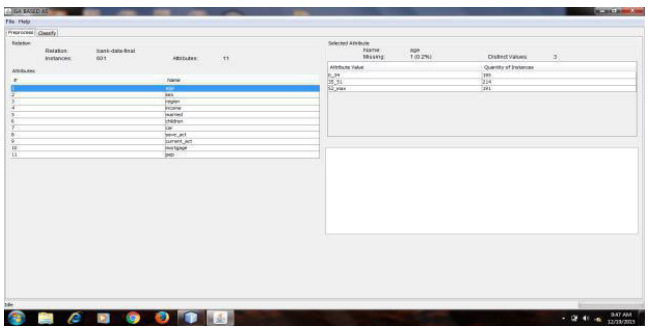

Figure 1 Attribute list

b) In figure 1, the initial population 11 is created by using the subsets generated as chromosomes.

c) The Naive Bayes classifier method is applied on individual subsets to compute the accuracy taken asa fitness value. This classifier has a minimum error rate. It performs consistently before and after reductions of numbers of attributes [5]. The fitness function is F= a-cnr+nr/2, where **"a"** is **"accuracy"**, **"cnr"** is **"cases** not covered**"** and **"nr"** is **"number** ofrules**"**.

A. Classification

After pressing the Classify button, Figure 2 presents a number of input boxes, checkboxes, and two buttons. User-defined values for various parameters are captured via input boxes. All checkboxes are optional and used to let the user make decisions depending on their preferences.

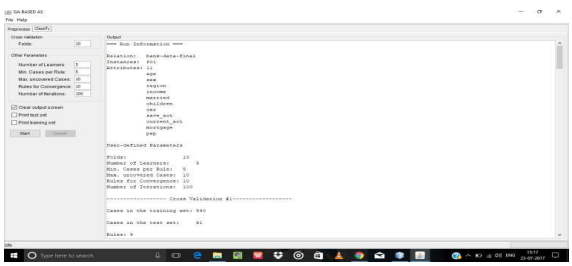

Figure 2 Lists of Parameters

B. Parameters

In figure 2, the first parameter is the number of learners. In this tool, 5 numbers of learners are taken and their accuracy compared to show best among them. "Forward Selection Multicross Validation, Bootstrap Backward Elimination, Relief, MIFS, and GABASS" are some of the techniques used in this study.

The four other parameters are "Minimum Cases per Rule," "Maximum Uncovered Cases," "Number of Rules for Convergence," and "Number of Iterations" for genetic algorithm optimization. Chromosomes that aregenerated using a genetic algorithm consist of 14 bits, out ofwhich the first two bits represent numbers of iterations, the nextthree bits represent minimum cases per rule, the next three bits represent maximum uncovered cases and the last four bits represents numbers rule convergence [5].

For example, Chromosomes samples: 010010110100

represents-

Number of Iterations- 200

Minimum Cases per Rule- 3

Maximum Uncovered Cases- 11

Number of Rule for convergence- 6

C. Genetic Algorithm

The genetic search begins with a population with zero characteristics and randomly created rules. The notion of survival of the fittest is used to generate a new population that will follow the laws of the fittest in the current population, as well as their children. The genetic operators cross over and mutations are used to create offspring. The generation process continues until a population P emerges, with every rule satisfying the fitness criterion [1]. By clicking on the start button below the parameters in figure 2, the aforesaid conditions are satisfied by the implemented tool. The output is shown in figure 3 which contains the number of rules as follow-

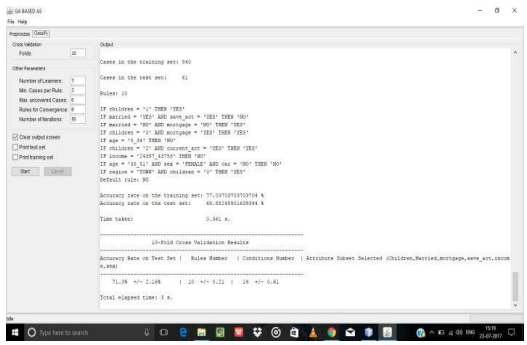

Figure 3 User Input Parameters and its Resulting Rules with Accuracy

The roulette wheel selection is used to establish a new population over two chromosomes with a cross over the

probability of 0.8 and a mutation probability of 0.001. Using this in a 10-fold cross validation until the requirements are met with the highest level of accuracy.

IV. RESULT

The best result is found by taking the value of k fold cross validation 10. When the parameters entered by the user are –

Number of Learners= 5

Minimum Cases per Rules= 5 Maximum Uncovered Cases=10 Rules for Convergence= 10 Number of Iterations= 100

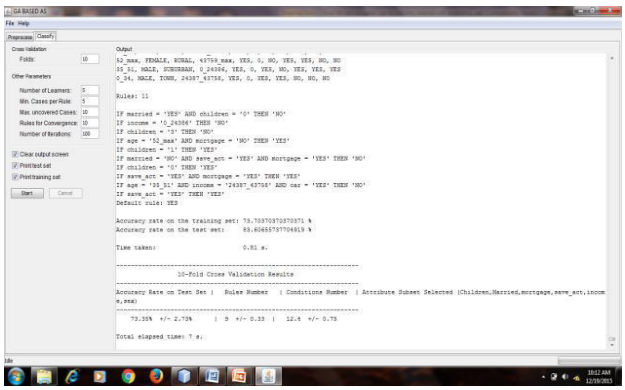

Figure 4 Selected subset of Attribute

Where the number of instances for the training set is 540 and for the testing set is 61. The accuracy observed at the bottom in figure 4 is 73.35% +/- 2.73% i.e., 76% approx. a subset of selected attributes is shown in the last of the above figure i.e., (Children, married, mortgage, save_acc, income, sex).

The graph in Fig 5 depicts the comparison between Random and GABASS. This comparison was made based on classification accuracy and GABASS was discovered to be superior to other methods.

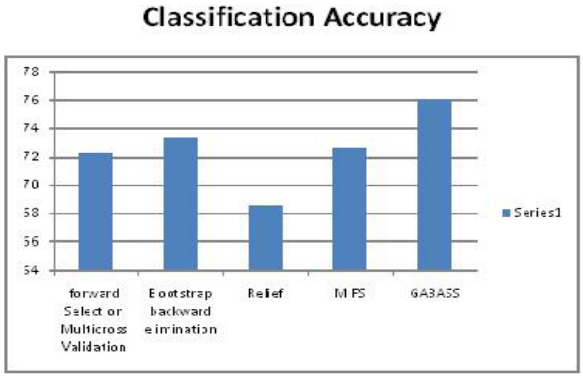

Figure 5 The Graph Shows the Comparison between Random and GABASS

V. CONCLUSION

Feature selection is a problem in development, hence a genetic calculation-based property subset determination using a Naive Bayes classifier is used. When there is a large population, GABASS has been found to be the best technique for determining the reason. The GABASS produces excellent results, and their strength resides in their ability to quickly adapt to a variety of scenarios.

VI. REFERENCES

[1] M.Anbarsi, E. Anupriya, N.CH.S.N. Iyengar "Enhanced prediction of heart disease with feature subset selection using genetic algorithm**"** vol.2(10),2010

[2] Asha Gowda Karegowda, M.A. Jayaram, A. S. Manjunath**"**Feature Subset Selection using Cascaded GA & CFS: A Filter Approach in Supervised Learning**"** International Journal of Computer Applications (0975 – 8887) Volume 23– No.2, 2011

[3] P. Xuan, M.Z. Guo, J. Wang, C.Y. Wang, X.Y. Liu and Y. Liu **"**Genetic algorithm-based efficient feature selection forclassification of pre-miRNAs**"** Genetics and Molecular Research 10 (2): 588-603 (2011)

[4] Ricco Rakotomalala, "TANAGRA: UN logicielgratuit pour l'enseignement et la recherche", in Actes de EGC'2005, RNTI-E-3, vol. 2, pp.697-702, 2005.

[5] Data Mining: Concepts, Methodologies, Tools, and Applications, Volume 1 Management Association, Information Resources IGI Global, 30-Nov-2012.

ABOUT THE AUTHORS

Pradeep Gupta received his B.E. (CSE) in 2006, MTech (CSE) in 2011.He has 13 years of experience in teaching.He is currently employed as an Assistant Professor at Ghaziabad's Ajay Kumar Garg Engineering College. Artificial Intelligence, Machine Learning, Deep Learning, and Cyber Security are some of his research interests.

Jay Kant Pratap Singh Yadav completed his B. Tech. and M. Tech. (NIT, Surat) and currently working as an Assistant Professor at Ajay Kumar Garg Engineering College in Ghaziabad, Uttar Pradesh, in the Department of Computer Science and Engineering.He is a lifetime member of IAENG and other academic societies and published several research papers in reputed international journals and conferences. His area of interest is Machine Learning, Soft Computing, Digital Image Processing, Computer Vision.