EDUCATIONAL SENSITIVE INFORMATION RETRIEVAL SYSTEM

Prashant Deshmukh[1], Prof. Shyam Gupta[2]
[1]*Student, Department of computer Engineering, Siddhant College of Engineering Sudumbre Pune, Savitribai Phule Pune University, Pune, Maharashtra, India*[1]
[2]*Professor, Department of computer Engineering, Siddhant College of Engineering Sudumbre Pune, Savitribai Phule Pune University, Pune, Maharashtra, India*[2]

ABSTRACT: Now a day's social networking and cloud services will collect a large amount of high-dimensional and complex data for third party statistical analysis and data mining. Although data analysis is beneficial to users and external parties, they constitute a serious privacy risk disclosure of user sensitive information. Proposed system used Support Vector Machines algorithm with discriminant component model to analyse the dataset, and make an accurate prediction of the impact on bad habits of students' comprehensive performance. Secondly, though various kinds of data transformation for privacy protection have been achieved and applied, there is little research designing two preference classification tasks simultaneously. In order to prevent the privacy leakage of sensitive information in the data analysis, cryptographic methods such as encryption and decryption are used. The goal is to preserve the statistical properties of one preference classification as far as possible, and to realize the data security of the other preference classification.

KEYWORDS: Data mining, SVM, Student performance analysis, sensitive task.

I. INTRODUCTION

The rapid growth of data generated in our society presents important challenges to database system design. In addition to the need for efficient information systems to manage high dimensional and complex data generated by multiple devices, a large amount of data is creating new privacy and security and social concerns. In fact, as the amount of data acquisitions are greatly increased in the past few decades, many users of personal devices (such as smart mobile phone, smart watch, wristbands) created, is a large part of the user specific data. Sharing user data can cause serious privacy problems because it contains a behavioural pattern that may reveal sensitive personal information when a third party mines data. For a recent target event, the store can view the history of personal purchases to predict whether a customer has changed purchase behaviour, it may show that a customer experiences a major life event such as pregnancy. Therefore, the design of privacy protection is crucial while protecting the user's sensitive information by enabling the application of the technology to obtain meaningful results. Pre-processing of a data is one of the first and critical step to data mining .The results of data pre-processing is directly inputted to mining model and obtained the final results. A clean data set can not only increase the accuracy of mining, but also raise the efficiency of algorithm dramatically. In general case, data pre-processing refers to as data cleaning, data integrate, data transition, data reduction, et al., processed before the implementation of data mining algorithm. Whereas, the technique of data mining concern many comprehensive area such as mathematic, computer, statistic, artificial intelligent, computer visual, et al. many application domain requires various function of data pre-processing.

Proposed system used Support Vector Machines algorithm with discriminant component model to analyse the dataset, and make an accurate prediction of the impact on bad habits of students' comprehensive performance. Secondly, though various kinds of data transformation for privacy protection have been achieved and applied, there is little research designing two preference classification tasks simultaneously. In order to prevent the privacy leakage of sensitive information in the data analysis, cryptographic methods such as encryption and decryption are used.

II. RELATED WORK

[1] "**Geometric Data Perturbation Approach for Privacy Preserving in Data Stream Mining**," T.Ankleshwaria,J.S.Dhobi, [1], Proposed system includes an effective approach of geometric transformation based data perturbation of data stream for mining has been proposed.

[2] "Approximate algorithms with generalizing at tribute values for k-anonymity" H. Park and K. Shim, [2], This paper, we give a presentation and diagram of some measurable exposure impediment issues that are of exceptional significance to general wellbeing considers and surveys.

[3] "A Cryptographic Privacy Preserving Approach over Classification", G. Nageswara Rao,M.Sweta Harini,Ch.Ravi Kishore,[3], We introduce a cryptographic based ap- proach that will ensure the protection of data sets.

[4] "A New Privacy Preserving Measure: p-Sensitive t-Closeness",C. N. Sowmyarani G. N.Srinivasan K.Sukanya [4], Preserving a sensitive information has turned into an incredible test in the zone of research under data privacy.

[5] "Towards a Methodology for statistical disclosure control" T. Dalenius, [5], Proposed framework give a presentation and diagram of some factual exposure impediment issues that are of uncommon pertinence to general wellbeing thinks about and overviews. Usama Fayyad,Gregory Piatetsky Shapiro, and Padhraic Smyth.

[6] "From Data Mining to Knowledge Discovery in Databases" Usama Fayyad,Gregory Piatetsky Shapiro, and Padhraic Smyth [6],The article makes reference to specific realworld applications, explicit datamining strategies, challenges engaged with realworld utilizations of learning revelation, and ebb and flow and future research bearings in the field.

[7] "A Study on Privacy Preserving Data Mining:Techniques, Challenges and Future Prospects" Ronica Raj,Veena Kulkarni [7], This study describes about various techniques of privacy preserving data mining.It also analyzes their advantages and limitations and comes up with a conclusion that a single technique does not exceed all the parameters such as performance, data utility, level of uncertainty, resistance to data mining algorithms and complexity

III. PROPOSED SYSTEM

Pre-processing: Pre-processing of an information is one of the first and basic advance to information mining. The results of information pre-processing is specifically inputted to mining model and acquired the last outcomes. Pre-processing of information incorporates extraction of valuable information from vast measure of information extraction is vital piece of the proposed framework since it increment the precision of mining, yet additionally raise the proficiency of calculation drastically.

SVM Classifier: A Support Vector Machine (SVM) is a regulated machine learning calculation that can be utilized for both characterization and relapse purposes. SVMs are all the more ordinarily utilized in order issues. In proposed system SVM classifier is used to classify the data into two main categories as sensitive data and insensitive data.

Data Privacy maintenance: Cryptography provides a suitable privacy model and includes complete proofs and exact quantization methods. The purpose of encryption is to make the original information after encryption becomes unreadable, only the user who holds the decryption key to recover the encrypted text.

Discriminant component Model: In the cooperative learning environment, we must put forward a feasible method to reduce the privacy leakage and ensure the security of the data. By reducing data dimension and removing some components, it prevents illegal data reconstruction. In proposed system we used a new approach where sensitive information hiding can be proposed as a pair of classification transformation tasks which base on idea of Discriminant Component Analysis (DCA) for creating reduced dimensional subspaces in collaborative learning environment.

Student performance analysis Education is an important part of social development. We obtain the main factors restricting the students' performance through the depth of excavation of student information. In order to analyse which attributes play a more important role in knowledge extraction, we only list the top five attributes that are most relevant to classification performance.
Attributes like alcohol consumption Alcohol has a lot of bad effects on the way we live. Proposed system analysed student performance through academic records, family background, friend circle... Etc., and suggest improvement regarding academics of the student.

IV. ALGORITHM AND PSEUDO CODE

1 SVM: A Support Vector Machine (SVM) is a supervised machine learning algorithm that can be used for both classification as well as regression purposes. SVMs are more mostly used in tasks of classification. In proposed system SVM classifier is used to classify the data into two main categories as sensitive data and insensitive data.
SVM works as follows
1 Assumed training examples plotted in space. These information focuses are relied upon to be isolated by a clear hole.
2 It predicts a straight hyper plane partitioning 2 classes.
3 The essential concentration while drawing the hyper plane is on amplifying the separation from hyper plane to the closest information purpose of either class. 4 The drawn hyper plane called as a most maximum-margin hyper plane.

2. AES Algorithm For Encryption.

Input:
128_bit /192 bit/256 bit input(0,1)
secret key(128_bit)+plain text(128_bit).
Process:
10/12/14-rounds for-128_bit /192 bit/256 bit input
Xor state block (i/p)
Final round:10,12,14
Each round consists: sub byte, shift byte, mix columns, add round key.
Output:
cipher text(128 bit)

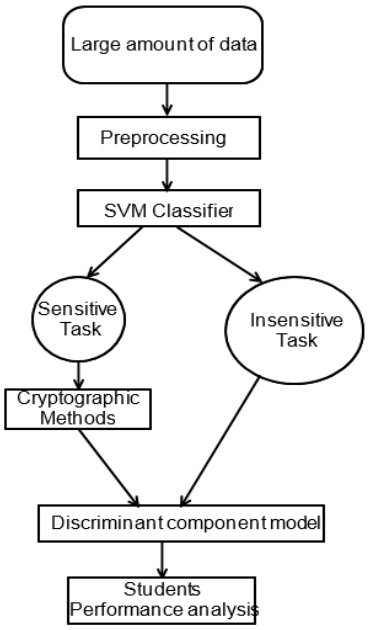
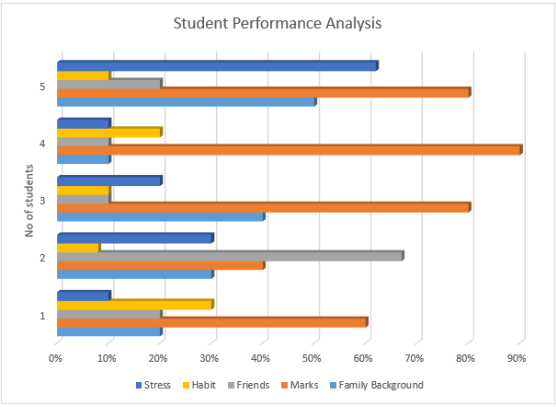
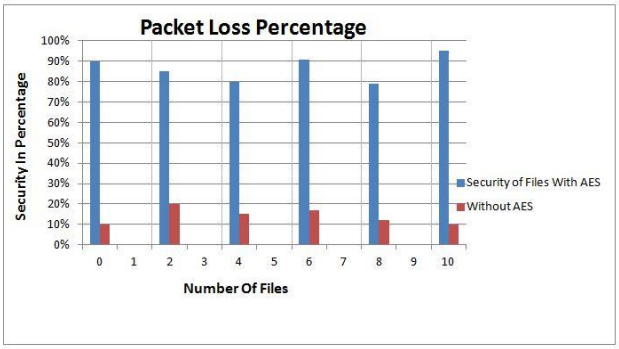


Fig. 1. System Architecture

V. RESULTS AND SYSTEM SCREENSHOTS

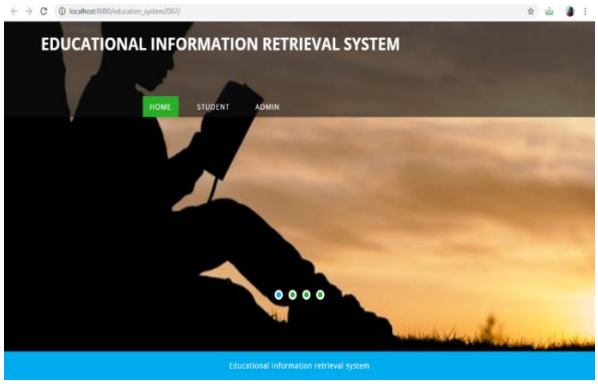


Graph 1: Student performance analysis graph

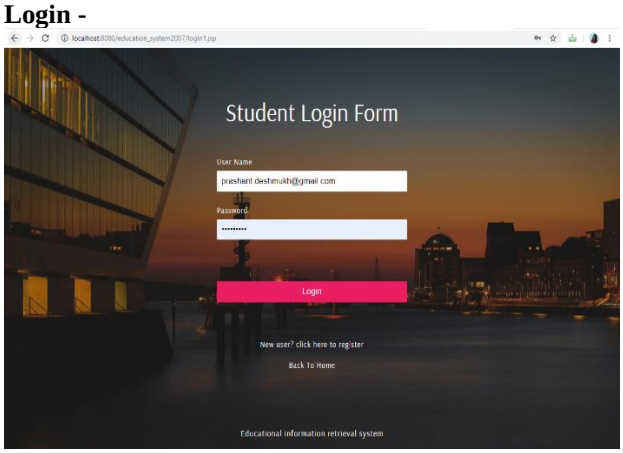


Graph 2: Security with AES

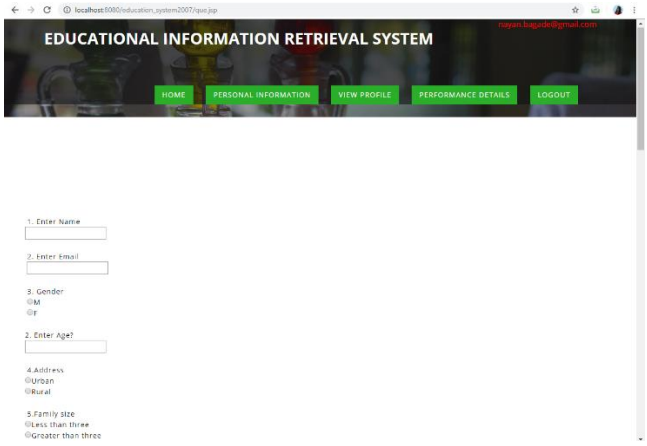
Home -



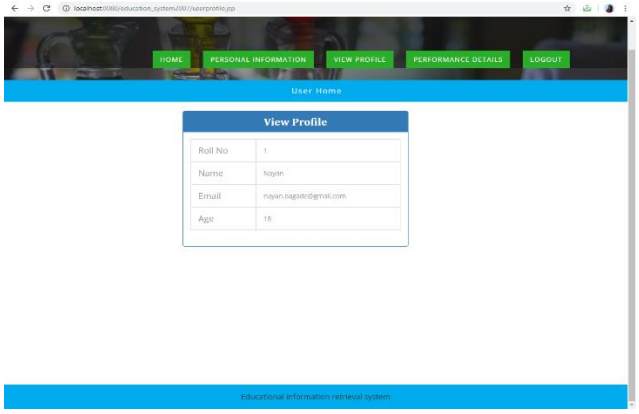
Login -



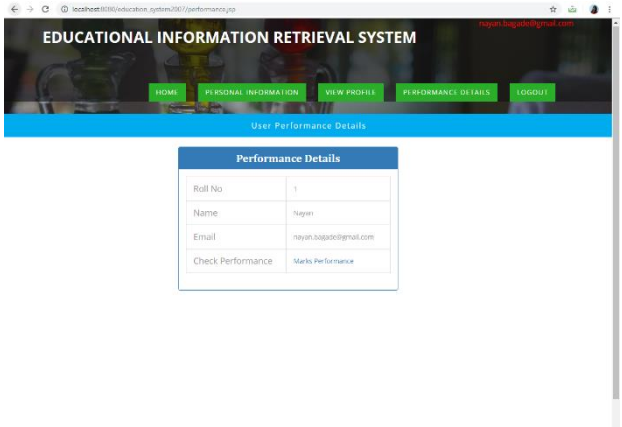
Student Info -



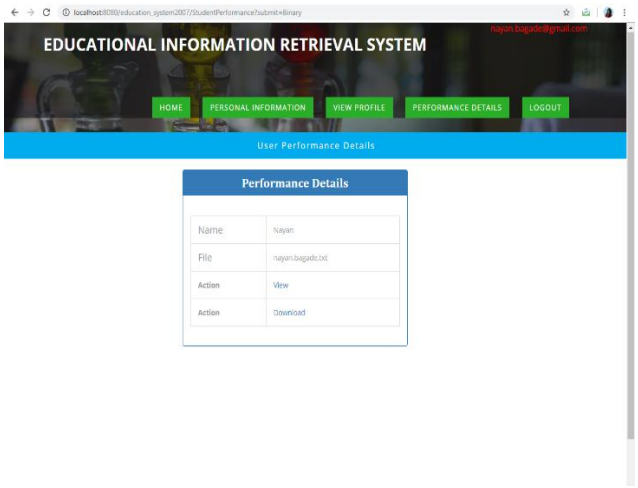
Student Profile -



Student Performance Details -



Student result –



VI. CONCLUSION

Proposed system represent algorithm of discriminant component analysis to predict the influence of students performance, and to analyse the main factors for non-dominant components. In order to protect the sensitive information in the data analysis process, we propose the classification of sensitive tasks and insensitive task. The crypto-graphic method is used, which can suppress the leakage of sensitive information while preserving the characteristics of insensitive information

VII. REFERENCES

[1] T. Ankleshwaria, J. S. Dhobi, Geometric Data Perturbation Approach for Privacy Preserving in Data Stream Mining, Engineering Universe for Scientific Research and Management, vol. 6, no. 4, pp. 1-6,2014

[2] H. Park and K. Shim, Approximate algorithms with generalizing attribute values for k-anonymity, Information System, vol. 35, no. 8, pp. 933-955, Dec. 2010

[3] G. Nageswara Rao, M. Sweta Harini, Ch. Ravi Kishore, A Cryptographic Privacy Preserving Approach over Classification, Springer, 2014.

[4] C. N. Sowmyarani G. N. Srinivasan K. Sukanya, A New Privacy Preserving Measure: p-Sensitive t-Closeness, Proceedings of International Conference on Ad- vances in Computing Advances in Intelligent Systems and Computing, vol. 174, pp. 57-62, 2013.

[5] T. Dalenius, Towards a methodology for statistical disclosure control, Statistik Tidskrift, vol. 15, pp. 429-444, 1977

[6] Majid, M.Asger, Rashid Ali, Privacy preserving Data Mining Techniques: Cur- rent Scenario and Future Prospects, IEEE 2012.

[7] L.Golab and M.T.Ozsu, Data Stream Management issu es-A Survey Technical Report, 2003.