iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** V **Month of publication:** May 2022

DOI: https://doi.org/10.22214/ijraset.2022.43255

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue V May 2022- Available at www.ijraset.com

Credit Card Defaulters Using Machine Learning

Akshay A. Bardiya
Sipna College of Engineering and Technology, Amravati

Abstract: ***As an increasing number of purchasers depend upon the credit score card to pay their regular purchases in on line and bodily retail store, the quantity of issued credit score playing cards and the overpowering quantity of credit score card debt via way of means of the cardholders have hastily accelerated. Therefore, maximum monetary establishments need to address the problems of credit score card default similarly to the credit score card fraud together with credit score card dump. Both the credit score card verification implemented to the cardholders and the default threat control after card issued are essential to the destiny achievement of maximum monetary establishments According to the Federal Reserve financial statistics, the default charge on credit score loans throughout all industrial banks is at an all-time excessive for the beyond sixty-six months, and it's miles possibly to hold to climb in the course of 2022. The delinquency charge suggests the share of beyond-due loans in the borrower's whole mortgage portfolio. The mountain climbing delinquencies will bring about a good sized amount of cash lose from the lending establishments, together with industrial banks. Therefore, banks ought to have a threat prediction version and be capable of classify the maximum relative traits which can be indicative of humans who've a better possibility of default on credit score. In 2013, patron spending encompassed about 69% of USA gross home product. Of the $3.098 trillion of tremendous patron credit score within the United States within the closing region of 2013, they have been revolving credit score card for over 25% of it ($857.6 billion). A small boom within the accuracy of figuring out excessive-threat loans may want to save you losses of over $eight billion. Because of the dangers inherent in one of these huge part of the economy, constructing fashions for patron spending behaviors to restrict threat exposures on this region is turning into extra critical. For this to be a feasible option, the predictions want to be moderately correct. A strong version isn't most effective a beneficial device for the lending establishments to determine on credit score applications, however it could additionally assist the customers to be privy to the behaviors that could harm their credit score scores. The number one motivation at the back of threat prediction is to make use of monetary statistics, for instance, enterprise transactional statistics, change statistics and patron transactions, and so on to foresee the patron's enterprise overall performance or character credit score card statistics and to lower bathrooms and vulnerability. Several threat prediction fashions are primarily based totally on statistical methods, inclusive of nearest neighbor, discriminant evaluation, and logistic regression. The aim of credit score default prediction is to assist monetary establishments determine whether or not or now no longer to lend to a patron. The ensuing check is mostly a threshold price that lets in the choice-makers to make the lending choice. The well-known version relies upon at the monetary ratios, earnings account, and statistics at the stability sheet.***

I. INTRODUCTION

As increasingly more purchasers depend upon the credit score card to pay their ordinary purchases in on line and bodily retail store, the quantity of issued credit score playing cards and the overpowering quantity of credit score card debt via way of means of the cardholders have swiftly increased. Therefore, maximum economic establishments need to address the troubles of credit score card default further to the credit score card fraud which include credit score card dump. Both the credit score card verification carried out to the cardholders and the default hazard control after card issued are critical to the destiny achievement of maximum economic establishments

According to the Federal Reserve financial statistics, the default charge on credit score loans throughout all industrial banks is at an all-time excessive for the beyond sixty-six months, and its miles in all likelihood to preserve to climb at some stage in 2022. The delinquency charge suggests the share of beyond-due loans within the borrower's complete mortgage portfolio. The mountaineering delinquencies will bring about a huge amount of cash lose from the lending establishments, which include industrial banks. Therefore, banks should have a hazard prediction version and be capable of classify the maximum relative traits which are indicative of humans who've a better chance of default on credit score.

In 2013, purchaser spending encompassed about 69% of USA gross home product. Of the $3.098 trillion of wonderful purchaser credit score within the United States within the remaining area of 2013, they have been revolving credit score card for over 25% of it ($857.6 billion).

©IJRASET: All Rights are Reserved

A small growth within the accuracy of figuring out excessive-hazard loans may want to save you losses of over $eight billion. Because of the dangers inherent in this kind of massive part of the economy, constructing fashions for purchaser spending behaviors to restrict hazard exposures on this area is turning into extra critical. For this to be a possible option, the predictions want to be moderately accurate. A strong version isn't simplest a beneficial device for the lending establishments to determine on credit score applications, however it may additionally assist the customers to be privy to the behaviors which can harm their credit score scores. The number one motivation in the back of hazard prediction is to make use of economic statistics, for example, enterprise transactional statistics, alternate statistics and patron transactions, and so on to foresee the patron's enterprise overall performance or person credit score card statistics and to lower bogs and vulnerability. Several hazard prediction fashions are primarily based totally on statistical methods, which includes nearest neighbor, discriminant analysis, and logistic regression. The intention of credit default prediction is to assist economic establishments determine whether or not or now no longer to lend to a patron. The ensuing check is often a threshold fee that lets in the decision-makers to make the lending decision. The popular version relies upon at the economic ratios, profits account, and statistics at the stability sheet.

II. LITERATURE SURVEY

Prediction of credit card default requires the use of various machine learning techniques. Some of the work done by various researchers is summarized below:

E-commerce industry is growing rapidly and this leads to the increased usage of credit card payments for online purchases. In this paper investigation of the performance of logistic regression, random forest and decision tree for credit card fraud detection is carried out. The dataset for credit card fraud detection is gathered from kaggle and this dataset consists of over 2, 84,808 credit card transaction data of a European bank. Fraud transactions are considered as positive class and the genuine transactions are considered as negative class. Dataset consists of imbalanced 0.172% of fraud transitions and the reaming transactions are genuine. Performance is evaluated based on accuracy, sensitivity, error rate and specificity. [1]

Real-time Credit Card Fraud/scam Detection Using Machine Learning. [2] This paper centers around four principle fraud events in certifiable transactions. Every fraud is tended to strategy is chosen through an assessment. Significant key territory which we discourse in our venture is constant credit card scam identification.

Ref. [3] proposes model for providing the measures for loss probabilities as well as the evaluation of credit risk. The data used for this purpose consists of account level data from six different banks. Three different machine learning techniques including the decision trees, random forests and logistic regression are evaluated in the proposed model. A credit card amount that is not recovered for a period of more than 90 days is considered as non-recoverable or a default.

[4] In this paper the author has used a case sensitive method which is based on Bays maximum risk and then it is presented using proposed cost measure. The dataset is based on the real life transaction data obtained from a European company and maintaining the confidentiality of the personal data. The accuracy of the algorithm used is 50%. The main significance of this paper is to reduce the cost. The result obtained was 23%.

Credit Card scam identification- Machine Learning methods [5] Credit Card Scam identification database was utilized in an analysis. Since the database was profoundly non balanced, destroyed strategy was utilized in over sampling. Later on, highlight determination was done and database was part in two sections, preparing data and testing data. The techniques utilized for the investigation were Logistic Regression, Random Forest, Naive Bayes with Multilayer Perception.

Credit Card Fraud/scam Detection using Deep Learning [6] this paper is tied in with developing a credit card scam identification framework utilizing Deep Learning Neural Networks. Regardless of whether or not the Neural Network is prepared above an extensive variety of emphases, that isn't always effectively accurate to categorize the data as fraud or valid because of skewness of the database. We make use of two sampling systems: Under-Sampling, from lessening number of legitimate perceptions and Over-Sampling, in which the fraud class perception is copied. Detection of Credit Card Fraud/scam Transactions Using Machine Learning Algorithms and Neural Networks [7] Credit card fraud coming about because of abuse for the framework is characterized like burglary or abuse of someone's credit card data that is utilized for individual increases unescorted by the consent of the owner of card. For identifying these scams, this is essential for checking the use examples for a client by the previous transaction. Contrasting the utilization example and present day transaction, we could categorize this like one or the other scam or a real transaction. In this research, the procedures utilized are KNN, Naïve Bayes, Logistic Regression, Chebyshev Functional Link Artificial Neural Network (CFLANN), Multi-Layer Perceptron and Decision Trees.

This paper checks and investigates the performance of Random Forest, SVM, logistic regression and Decision tree on a highly skewed credit card fraud data. The dataset was gathered by a European cardholders consisting of about 2, 84,786 transactions. The result obtained was 97.7% accuracy by Logistic regression, 97.5% by SVM and 98.6% precise accuracy obtained by Random Forest. [8]

In this paper one of the best data mining algorithm called machine learning algorithm was introduced, which was used to recognize the credit card fraud. A half bread grouping framework with exception recognition was utilized in order to differentiate between misrepresentations of internet recreations. The framework obtained online calculations with factual data in order to distinguish various extraction types. This framework attained extreme location rate at 98% along with 0.1% fault rate. [9]

This paper discusses about supervisor based classification using Bayesian network classifiers such as Naïve Bayes, K2, Tree Augmented Naïve Bayes (TAN, logistics and J48 classifiers. The datasets are pre-processed by using normalization and principal component analysis. Two datasets were used dummy dataset which represented the characteristics of the credit card data and newly generated dataset using data normalization and principal component analysis technique. All these classifiers achieved over 95% accuracy. [10]

III. PROBLEM STATEMENT

Credit card default are increasing heavily because of fraud financial loss increasing drastically. Every year due to fraud, Billions of amounts loss. To predict the default in credit card, there is lack of research. Taking this point in mind we are developing a system which can predict and find the defaulter person, so that financial sector may secure from this fraud.

IV. PROPOSED SYSTEM

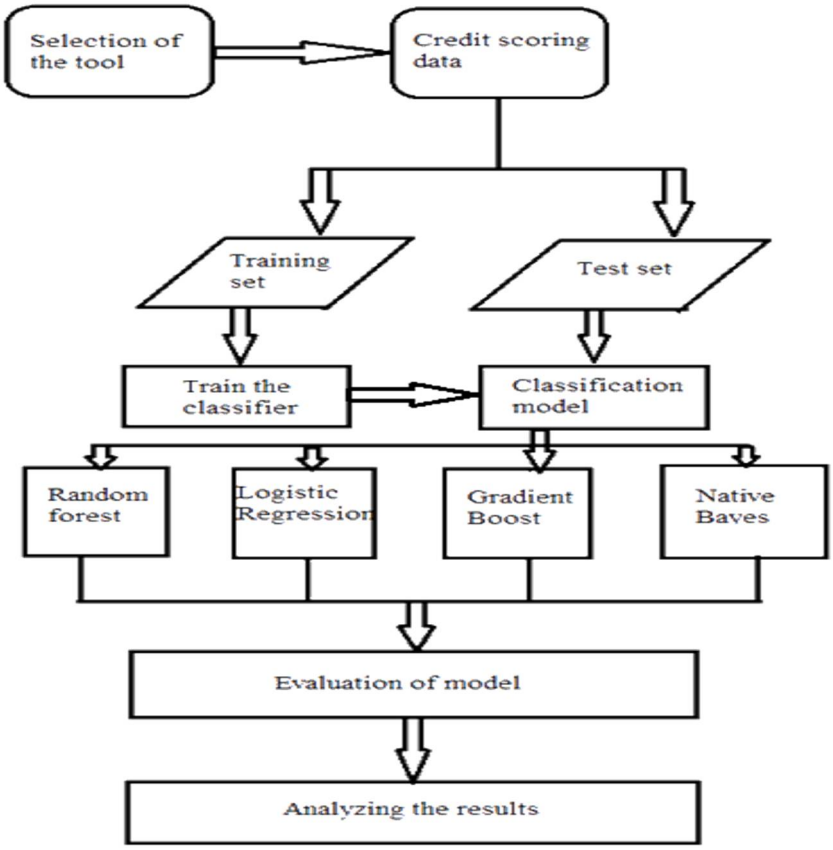


Fig 1: System Flow Diagram

In propose system, we propose four different algorithms to find the best accuracy in predicting the default candidate for the next month. The credit card data set is divided into two parts- the training set as well as the test set. The classification model is trained using the training set and the remaining observations passes to next level to perform the prediction task using different techniques.

A brief introduction of these techniques is as follows [16]:

1) *Random Forest:* Random forest is a type of supervised learning algorithm getting to know set of rules this is used broadly in Classification and Regression problems. It builds decision trees on distinct samples and takes their majority vote for classification and average in case of regression. In Random Forest model, Random means each tree is only trained on a random subset of samples drawn from the training set (with repetition) and possibly a random subset of features and Forest because there are several trees.
2) *Logistic Regression:* Logistic regression model attempts to pick out the correlation between the dependent and the independent variables. It is basically used for binary classification where the target variable is binary and one or more independent variables can be continuous or binary. This model uses the logistic function to identify or to track the probability of the output with respect to the input. The classification is applied such that a threshold is provided, and all the probability values greater than a certain threshold are assigned one class and the values less than the threshold are assigned the other class.
3) *Gradient Boost:* The main purpose of Gradient Boost model is to help weak prediction models becomes stronger. It works by building one tree at a time, and correct errors made by previously tree. It can be used for predicting not only continuous target variable (as a Regressor) but also categorical target variable (as a Classifier).
4) *Naive Bayes:* Naive Bayes classifier is one of the supervised learning algorithms that is primarily based totally on Bayes theorem and makes use of the probabilistic features. It assumes the independence among all the features for a particular model. Naive Bayes classifier is a simplest method to integrate as it assumes conditional independence. The probabilities of conditional method are identified for the attributes and classification is performed such that the class with the most probable hypothesis or maximum a posteriori (MAP) is assigned to the given element.

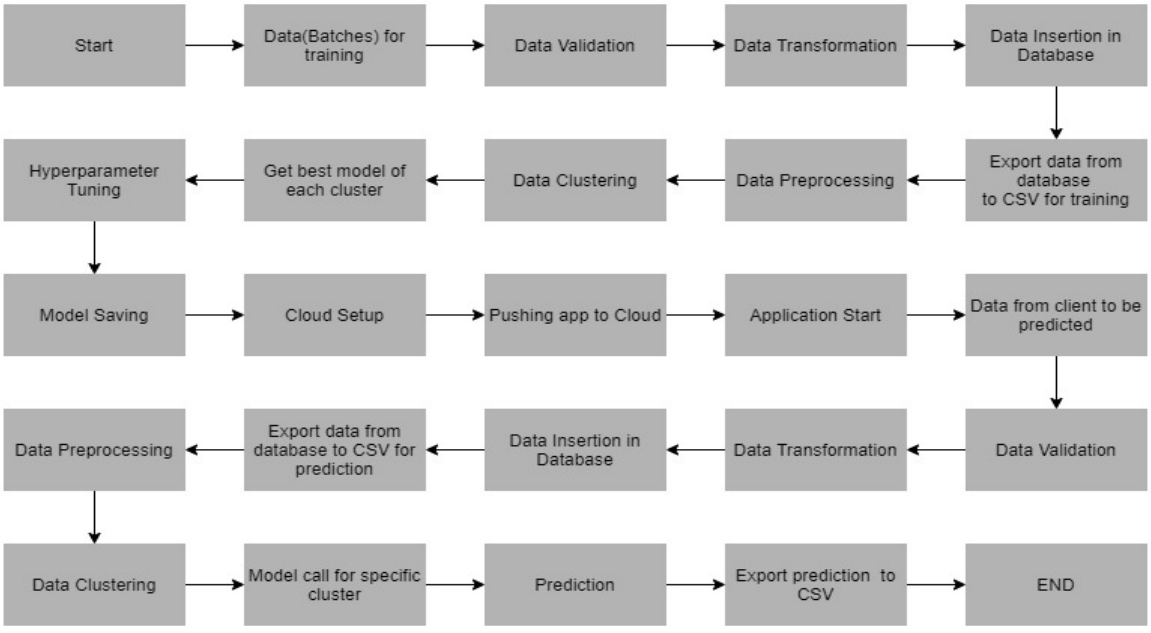


Fig 2: Architecture Flow Diagram

V. CONCLUSION

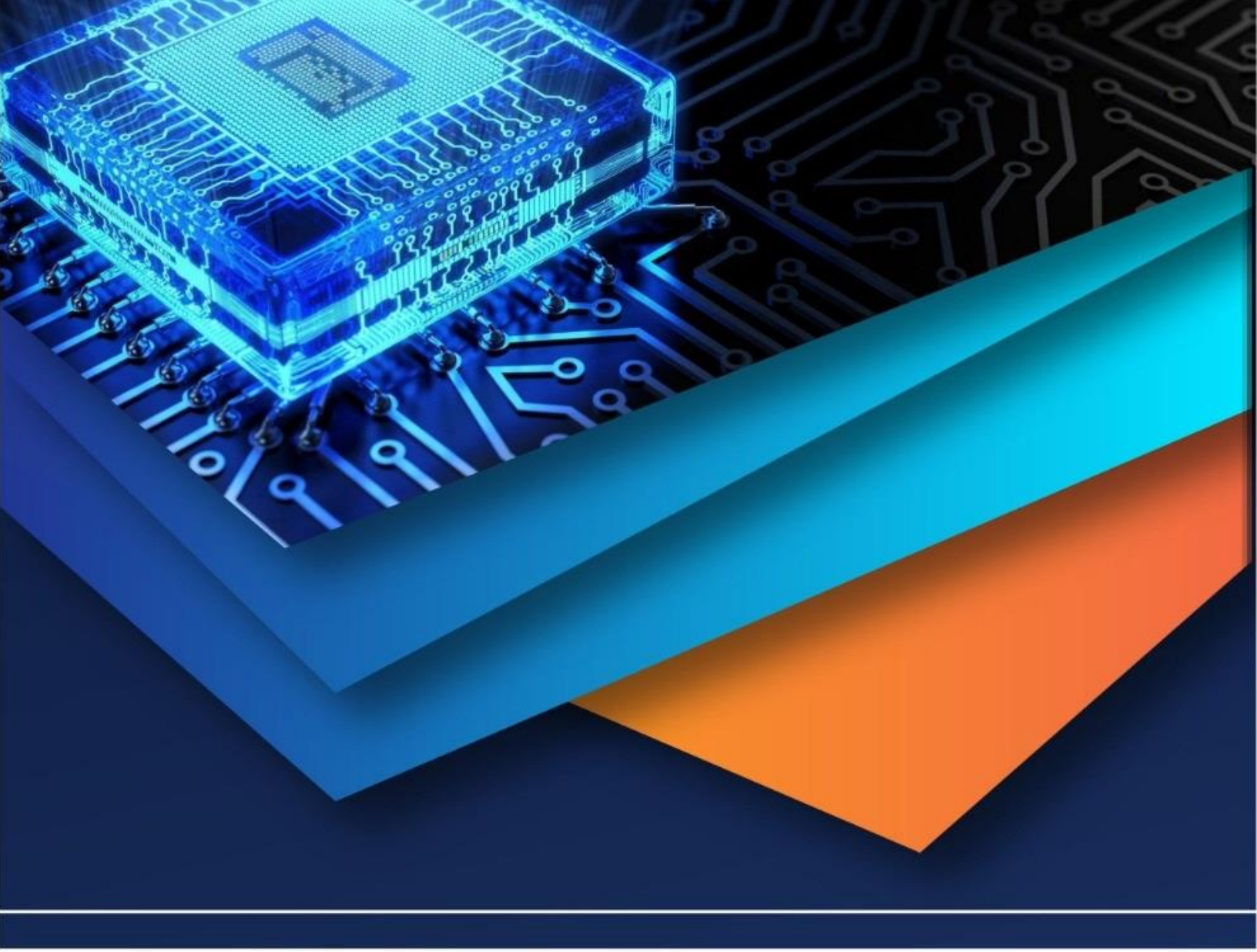
Credit card fraud cases are increasing day by day and it is one of the major concerns in financial service sectors. This occurs when no proper security measures are taken into consideration are. In this paper an attempt is made to identify the number of fraudulent transactions in a particular dataset by using various machine learning algorithms such as local outlier factor and isolation forest method. Only a part of dataset was used in order to speed up the computational process. Future scope of improvement includes:

1) Large number of datasets can be stored using cloud storage and then fetch the datasets from the cloud storage repository.
2) User interface is not provided in this project.

REFERENCES

[1] M. BM and H. Mohapatra, "Human centric software engineering," International Journal of Innovations & Advancement in Computer Science (IJIACS), vol. 4, no. 7, pp. 86-95, 2015.
[2] Anuruddha Thennakoon; Chee Bhagyani; Sasitha Premadasa; Shalitha Mihiranga; Nuwan Kuruwitaarachchi 2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence);10-11 Jan. 2019
[3] Zhou H, Lan Y, Soh Y, Huang G and Zhang R (2012), "Credit risk evaluation with extreme learning machine", IEEE International Conference on Systems, Man, and Cybernetic(SMC), Seoul, 2012, pp. 1064-1069.
[4] H. Mohapatra, C Programming: Practice, Vols. ISBN: 1726820874, 9781726820875, Kindle, 2018.
[5] Dejan Varmedja; Mirjana Karanovic; Srdjan Sladojevic; Marko Arsenovic; Andras Anderla; 2019 18th International Symposium INFOTEH-JAHORINA (INFOTEH) 20-22 March 2019
[6] Pranali Shenvi; Neel Samant; Shubham Kumar; Vaishali Kulkarni; 2019 IEEE 5th International Conference for Convergence in Technology (I2CT) 29-31 March 2019
[7] Deepti Dighe; Sneha Patil; Shrikant Kokate; 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)16-18 Aug. 2018
[8] H. Mohapatra and A. Rath, Advancing generation Z employability through new forms of learning: quality assurance and recognition of alternative credentials, ResearchGate, 2020.
[9] H. Mohapatra and A. Rath, Fundamentals of software engineering: Designed to provide an insight into the software engineering concepts, BPB, 2020.
[10] V. Ande and H. Mohapatra, "SSO mechanism in distributed environment," International Journal of Innovations & Advancement in Computer Science, vol. 4, no. 6, pp. 133-136, 2015.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)