Maulana Azad National Institute of Technology

Bhopal

Department

Master of Computer Applications

Session: 2021-2024

Semester: 4nd

Assignments(CA - 402)

Loan Prediction Using Support Vector Machine

Submitted by

Indrajeet Yadav

Scholar ID: **212120125**

Submitted to

Prof. G.S.Thakur

Loan Prediction Using Support Vector Machine

Dr. G.S.Thakur
Department of Computer Applications
Applications Maulana Azad National Institute
of Technology Bhopal,India
ghanshyamthakur@gmail.com , ghanshyamthakur@manit.ac.in

Indrajeet yadav
Department of Computer
Maulana Azad National Institute
of Technology Bhopal,India
212120125@stu.manit.ac.in

Abstract:-

Loan prediction is a significant challenge in the financial industry that involves assessing the risk associated with lending money to borrowers. Support vector machines (SVM) are a popular machine learning technique that can be used to predict whether a loan application should be approved or rejected based on various factors such as credit history, income, employment status, etc. This paper presents an SVM-based approach for loan prediction using a dataset of historical loan applications. The proposed method involves data preprocessing, feature engineering, model training, and evaluation. The results show that the SVM-based approach is effective in predicting loan approval with high accuracy and can be used as a reliable tool by financial institutions for loan approval decision-making.

Introduction:-

The loan prediction problem is a critical issue in the financial sector that has a significant impact on the stability and growth of financial institutions. In recent years, machine learning techniques such as Support Vector Machines (SVM) have been applied to this problem to improve the accuracy of loan prediction models. SVM is a powerful machine learning algorithm that can handle both linear and non-linear data and has been used extensively in classification and regression tasks.The objective of loan prediction is to identify the risk associated with lending money to potential borrowers. The risk assessment is based on various factors such as credit history, income, employment status, loan amount, and others. The traditional approach to

loan prediction involves manual analysis of the applicant's financial records, which can be time-consuming and error-prone.

The use of machine learning techniques such as SVM in loan prediction can significantly improve the accuracy and efficiency of the process. SVM can analyze large amounts of data and identify complex patterns in the data to predict loan approval or rejection. The SVM-based loan prediction model can assist financial institutions in decision-making and reduce the risk associated with lending money.

In this paper, we present an SVM-based approach to loan prediction using a dataset of historical loan applications. The proposed method involves data preprocessing, feature engineering, model training, and evaluation. The results demonstrate that the SVM-based approach is effective in predicting loan approval with high accuracy and can be used as a reliable tool by financial institutions for loan approval decision-making.

Methodology:-

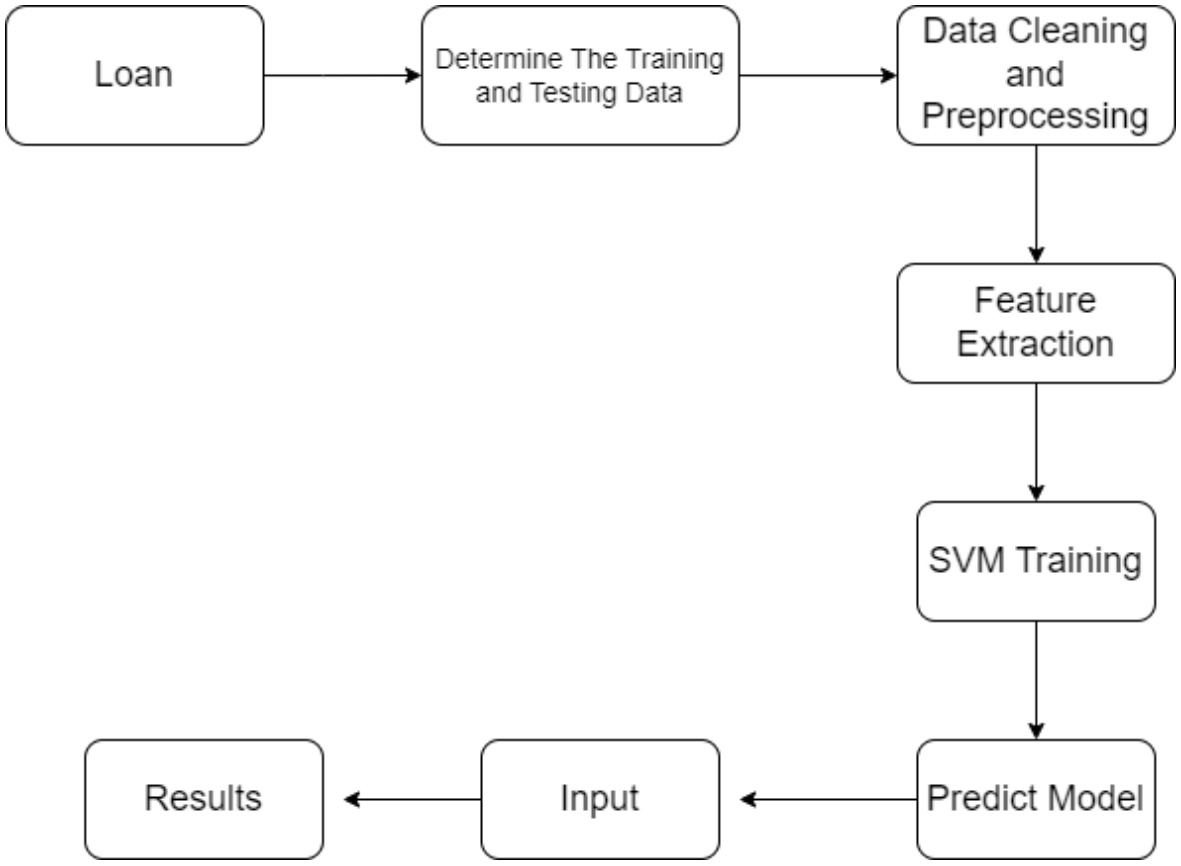
Support Vector Machine (SVM)-

A support vector machine (SVM) can be used when our information is completely in two classes. An SVM classifies information by detecting an ideal hyperplane that separates all information purposes of a class from an alternative class. The hyperplane for SVM means the largest difference between the two classes. The margin shows that the section parallel to the hyperplane has a maximum width with no internal information points. SVM has verified to be one of the most important economic kernel strategies. The success of SVM is mainly due to its high generalization capability. Not like many learning algorithms, SVM results in sensible demonstrations whereas previous data need not be included. In addition, the employment of positive fixed kernels within SVM can be taken as the associate degree embedding of the input field to the higher dimensional feature region where

the classification is met, while the exploitation does not explicitly use this feature area has been done. Therefore, the case for selecting a design for neural network

application is replaced by the case of selecting an acceptable kernel for a support vector machine. The support vector machine has shown power in binary classification. It's Wise Theoretical Foundation and Well Perfect Learning Algorithm Rules. This leads to stable information classification. The only disadvantage is this is it's time and memory once the size of the information is large. In the following section we will discuss the proposed methodology for email spam detection technique

Work Flow:-



Result Discussion:-

Preprocessing:-

The pre-processing step is used to remove the noises from the email which are irrelevant and need not be present. The preprocessing step includes

- Removal of Numbers
- Removal of Special Symbol
- Removal of URLS
- Stripping HTML
- Word Stemming

Feature Extraction:-

It involves selecting the most relevant features from the loan application data that can provide insights into the borrower's creditworthiness and repayment capacity. The following are some of the commonly used features in loan prediction:

- Credit history: The credit history of the borrower is one of the most important features in loan prediction. It includes factors such as the borrower's credit score, credit utilization ratio, and payment history.
- Income: The borrower's income is another important feature that can provide insights into their repayment capacity. It includes factors such as monthly income, source of income, and stability of income.
- Employment status: The borrower's employment status can provide insights into their ability to repay the loan. It includes factors such as employment type, job stability, and current employment status.

- Loan amount: The loan amount requested by the borrower is an important factor that can affect their ability to repay the loan.
- Loan purpose: The purpose of the loan can provide insights into the borrower's financial goals and their ability to repay the loan.
- Loan term: The loan term is the period over which the loan will be repaid. It can affect the borrower's ability to repay the loan and the lender's risk.

SVM Training:-

The preprocessed loan application data needs to be split into training and testing sets. The training set is used to train the SVM model, while the testing set is used to evaluate the model's performance.

Test Classifier:-

The classifier is tested with numerous training data to test the accuracy of the classifier. The proposed solution achieves up to 98 % accuracy in classifying emails.

Model Evaluation:-

Here, find the accuracy score on the Training data and the Testing data.

Dataset Description:-

The dataset used for loan prediction using SVM typically consists of loan application data, which includes information about the borrower and their loan application. The dataset can be obtained from various sources such as financial institutions, credit bureaus, or publicly available data sources. The following are some of the common attributes that can be included in the loan prediction dataset:

- Loan_ID: A unique identifier for each loan application.
- Gender: The gender of the borrower.
- Married: The marital status of the borrower.
- Education: The educational qualifications of the borrower.
- Self_Employed: The current employment status of the borrower.
- Applicant Income: The income of the borrower.
- Co-applicant Income: The income of the co-applicant (if applicable).
- Loan Amount: The amount of loan requested by the borrower.
- Loan Amount Term: The period over which the loan will be repaid.
- Credit History: The credit history of the borrower, including factors such as credit score, credit utilization ratio, and payment history.
- Property Area: The area of the property for which the loan is being requested.
- Loan Status: The status of the loan application, which can be approved or rejected.

Number of Rows and Columns

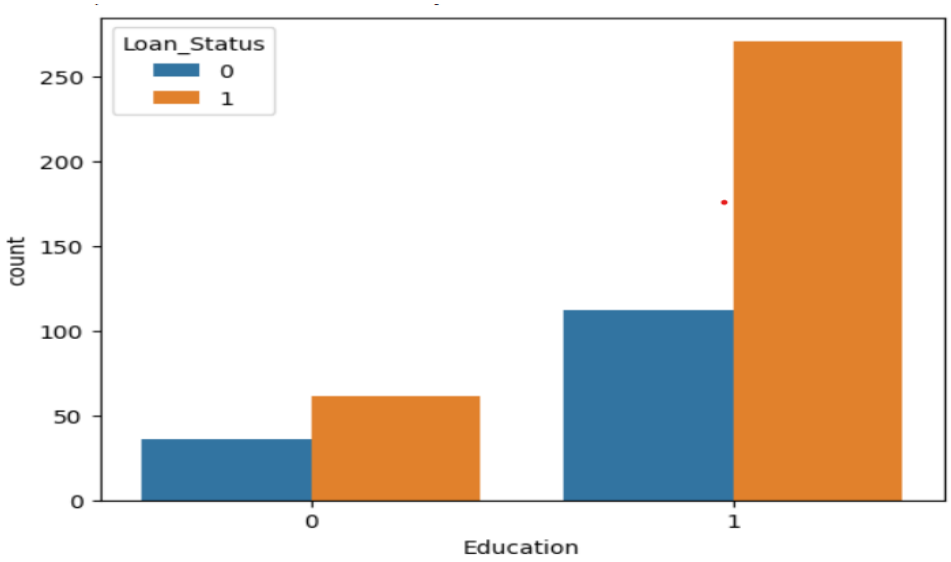
(614,13)

Tabular Form:-

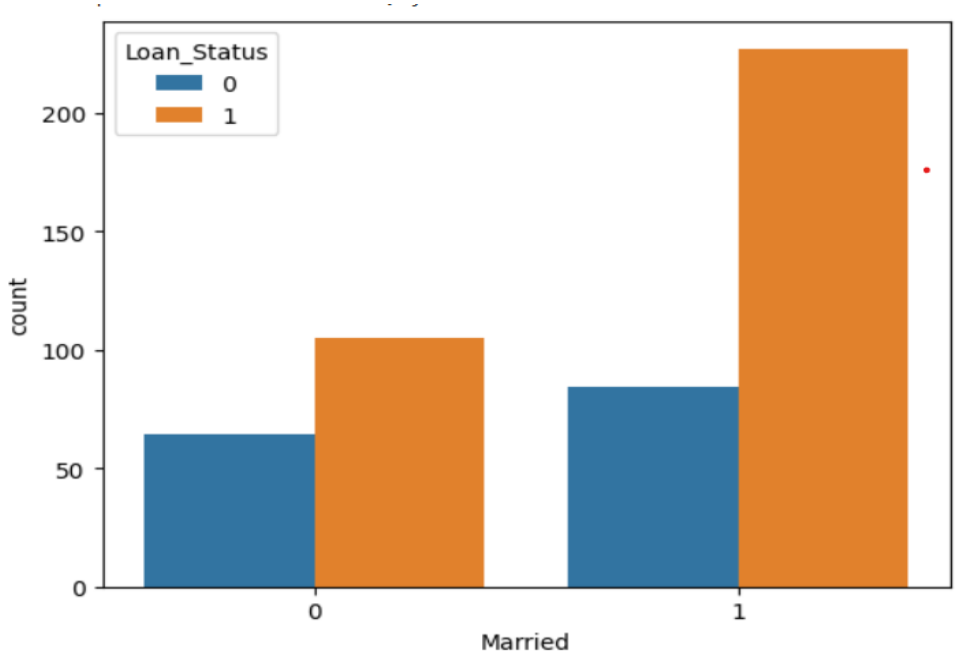
```
Data columns (total 13 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   Loan_ID            614 non-null    object
 1   Gender             601 non-null    object
 2   Married            611 non-null    object
 3   Dependents         599 non-null    object
 4   Education          614 non-null    object
 5   Self_Employed      582 non-null    object
 6   ApplicantIncome    614 non-null    int64
 7   CoapplicantIncome  614 non-null    float64
 8   LoanAmount         592 non-null    float64
 9   Loan_Amount_Term   600 non-null    float64
 10  Credit_History     564 non-null    float64
 11  Property_Area      614 non-null    object
 12  Loan_Status        614 non-null    object
```

Data Visualization; univariate analysis and to see correlations between variables:-

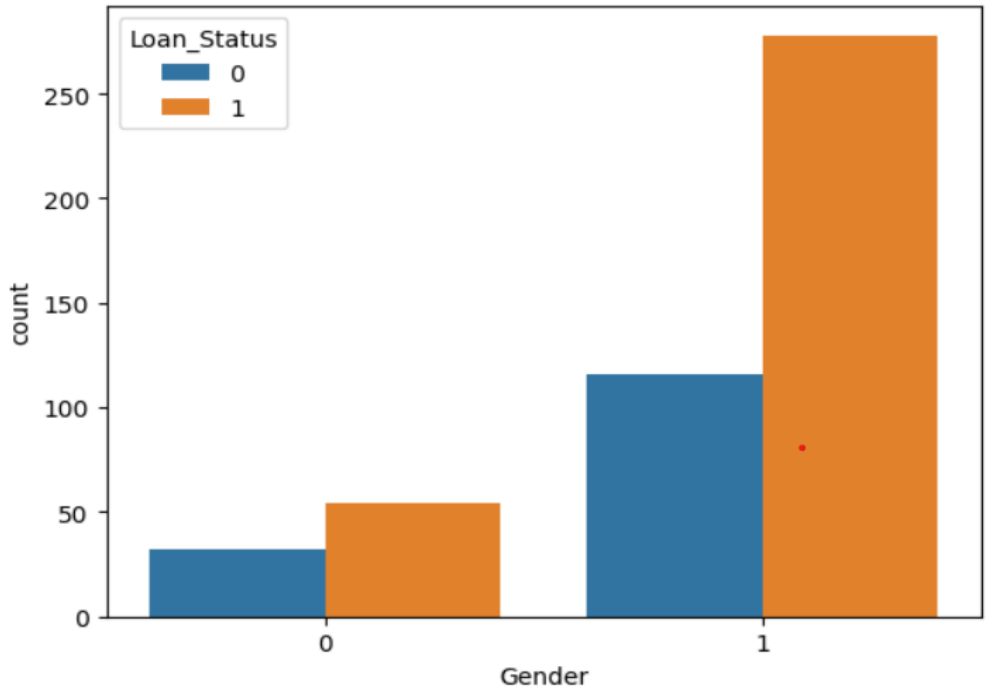
The correlation between Married and Loan Status:-



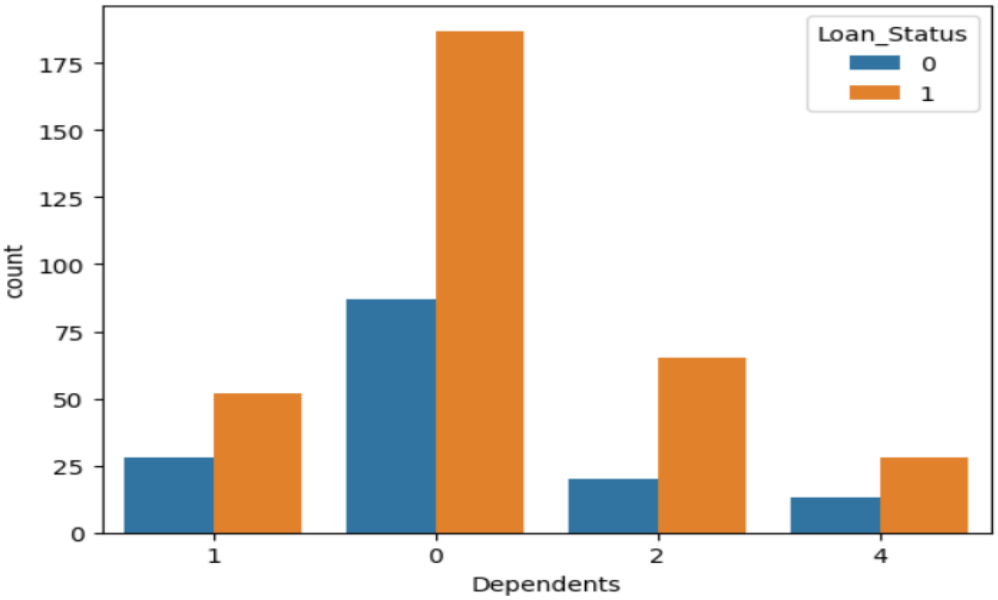
The correlation Education and Loan Status:-



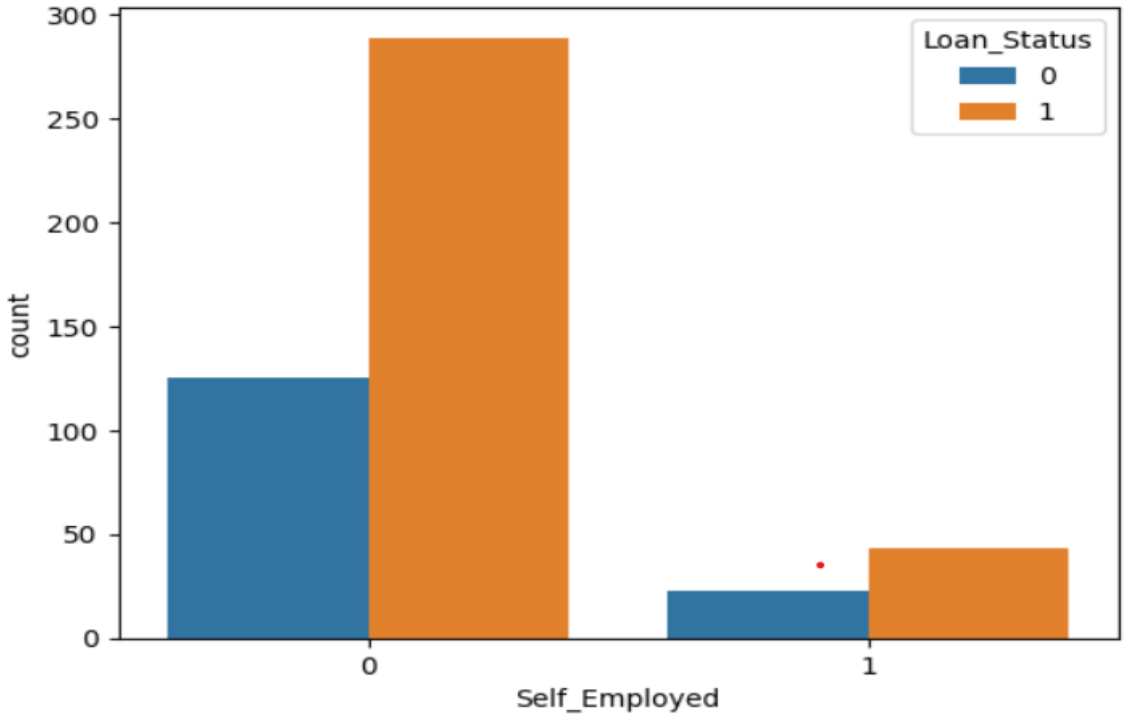
The correlation Gender and Loan Status:-



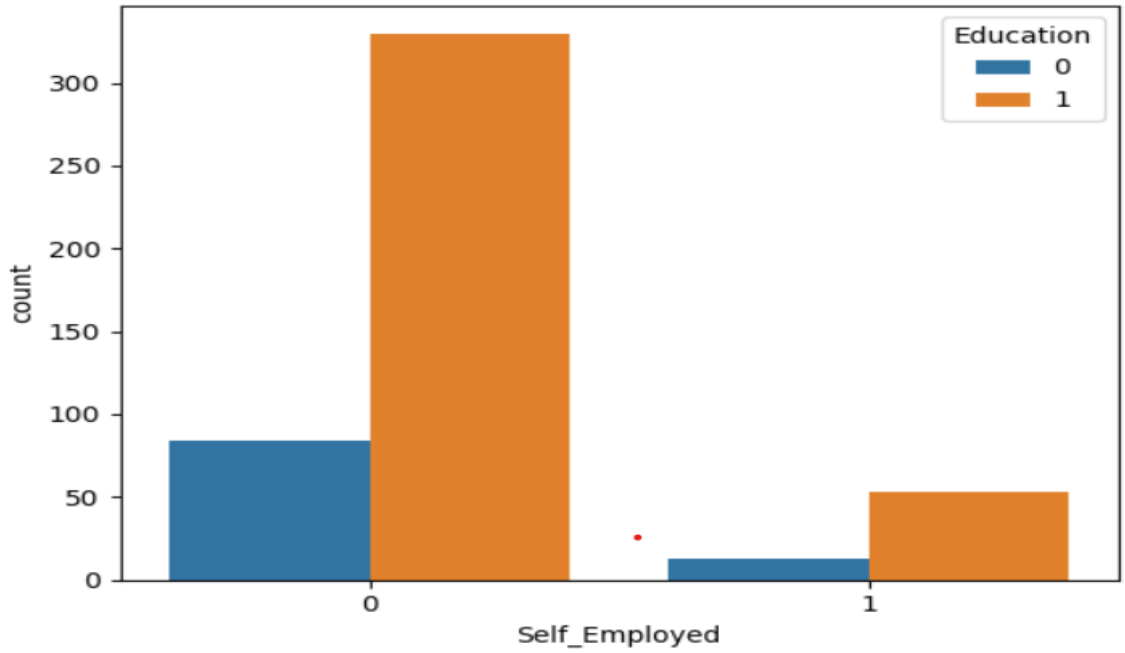
The Dependents Married and Loan Status:-



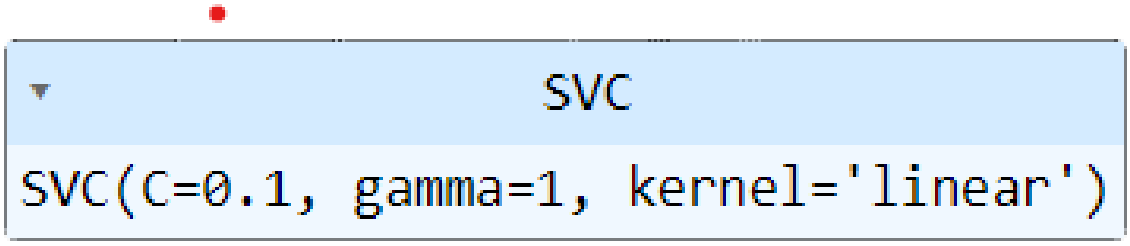
The correlation Self_Employed and Loan Status:-



The correlation Self_Employed and Education:-



Training using SVM model:-



Result Description:-

Model Evaluation to see how accurate our model is

***Accuracy Score**:-*

Training Accuracy	*Testing Accuracy*
0.7962962962962963	0.8333333333333334

Conclusion:-

The loan prediction problem can be tackled using a variety of machine learning algorithms, including SVM (Support Vector Machines). After training an SVM model on a labeled dataset, we can use it to predict whether a loan application will be approved or rejected based on certain features, such as the applicant's credit score, income, and loan amount.

SVM is a powerful algorithm that can provide accurate loan predictions when properly trained and tuned. However, it is important to note that no model is perfect and there is always room for improvement. It is important to continue to refine and improve the model over time by incorporating new data and refining the feature selection process. Additionally, it is important to monitor the model's performance regularly to ensure that it continues to perform well and to identify any potential issues that may arise.

References:-

[1]. Brijesh Bakariya and G.S. Thakur, "An Efficient Algorithm for Extracting Infrequent Itemsets from Weblog", The International Arab Journal of Information Technology (IAJIT),2019 (2)

[2]. Brijesh Bakariya and G.S. Thakur, "Mining Rare Itemsets from Weblog Data",National Academy Science Letters, 2015 page 1-5

[3]. Brijesh Bakariya and G.S. Thakur, "Pattern Mining Approach for Social Network Service", National Academy Science Letters, June 2017, Volume 40, Issue 3, pp 183–187

[4]. Brijesh Bakariya, G.S. Thakur,"An Efficient Algorithm for Extracting High Utility Itemsets from Web Log Data",published in The Institution of Electronics and Telecommunication Engineers (IETE) Technical Review SCI DOI: 10.1080/025646 02.2014.1000396, Volume-32, Issue-02, March- 2015, Pages 151-160.

ISSN 0256-4602 impact factor 1.304

[5]. C. Liu, K. Li, K. Li, and R. Buyya, “A new service mechanism for profit optimizations of a cloud provider and its users,” IEEE Trans. Cloud Comput., vol. 9, no. 1, pp. 14–26, Jan.–Mar. 2021

[6]. D. S. Rajput, R. S. Thakur, G. S. Thakur, “A Computational Model for Knowledge Extraction in Uncertain Textual Data using Karnaugh Map Technique”, International Journal of Computing Science and Mathematics (InderScience) ISSN: 1752-5055, Jan 2016, Vol. 7, Issue 2, pp. 166-176

[7]. D.S Rajput, R.S. Thakur and G.S. Thakur, “Fuzzy Association Rule Mining based Frequent Pattern Extraction from Uncertain Data” presented in IEEE 2nd World Congress on Information and Communication

Technologies (WICT- 2012) October 30-November 02, 2012 in IIITM Trivandrum, ISBN: 978-1-4673-4804-1

pp 709-714

[8]. D.S Rajput, R.S. Thakur, G.S. Thakur, “An Integrated Approach and framework for Document Clustering using Graph based Association Rule Mining” published in Advances in Intelligent Systems and Computing, Springer, ISBN 978-81-322-1602-5, Vol. 236 pp. 1421- 1438.

[9]. G. S. Thakur,Fuzzy Soft Traffic Accident Alert Model,National Academy of Science Letter(May–June 2014),Springer 37(3):26 1–268 DOI 10.1007/s40009-014-0235-6

[10]. G. Vishwakarma and G. S. Thakur, “Hybrid System for MPAA Ratings of Movie Clips Using Support Vector Machine,” in Advances in Intelligent Systems and Computing book series (AISC, volume 817), 2019, pp. 563–575

[11]. G.S. Thakur,Neeraj Sahu,Swatranta Sahu,“Hesitant Fuzzy Linguistic Term Set Based Document

Classification” ,In IEEE Third International Conference on Communication Systems and Network Technologies April 6-8,2013,Gwalior,India

[12]. G.S. Thakur,R.S. Thakur, “Design of 2-Level Clustering Framework for Time Series DataSets”, In an International Conference on Soft Computing for Problem Solving (SocPros11) ,IIT Roorkee India, December 16-18, 2011, Published Springer Proceeding.

[13]. G.S.Thakur,Ravi Singh, "New Hesitant Fuzzy Operators", in Fuzzy Information and Engineering Volume 6, Issue 3, September 2014, Pages 379–392

[14]. Ghanshyam Singh Thakur and Rekha Singh Thakur,” BSclassifier for Balance Scale Weight and Distance Database”, International Journal of Soft Computing Year: 2010 | Volume: 5 | Issue: 6 | Page No.: 211- 213.

[15]. Harshita Patel, G.S. Thakur, “Classification of Imbalanced Data using a Modified Fuzzy-Neighbor Weighted Approach”, published in IJIES in Vol. 10, Issue 1, 2017, pp. 56-64

[16]. Harshita Patel, G.S. Thakur, “Improved Fuzzy-Optimally Weighted Nearest Neighbor Strategy to Classify Imbalanced Data”, published in IJIES in Vol. 10, Issue 2, 2017, pp. 156-162

[17]. Harshita Patel, G.S. Thakur, “An Improved Fuzzy K-NN Algorithm for Imbalanced Data using Adaptive Approach”, published in IETE Journal of Research in Vol. 65, Issue 6, pp. 780-789, ISSN: 0974-780X

[18]. I. Naseem, R. Togneri, and M. Bennamoun, “Linear regression for face recognition,” IEEE Trans. Patt. Anal. Mach. Intel., vol. 32, no. 11, pp. 2106–2112, Jul. 2010.

[19]. K. K. Mohbey, S. Kumar, and V. Koolwal, “Advertisement prediction in social media environment using big data framework,” in Multimedia Big Data Comput. for IoT Appl. Berlin, Germany: Springer, 2020, pp. 323–341.

[20]. Krishna K. Mohbey & G. S. Thakur,Interesting User "Behaviour Prediction in Mobile E-commerce Environment using Constraints",published by Taylor & Francis in IETE Technical Review,18 Nov 2014