

# Data Mining Probabilistic Classifiers for Extracting Knowledge from Maternal Health Datasets



Sourabh, Vibhakar Mansotra

**Abstract:** Data Mining is an important sub-process of Knowledge Discovery in Databases (KDD) or Knowledge Discovery Process (KDP) methodology that is mainly used for applying various data mining techniques and algorithms on the target data. In this research paper, the authors have made an attempt to discover knowledge by classifying the maternal healthcare data of Jammu and Kashmir State of India (now declared as Union Territory by the Government of India). The data for the present research work was collected from a web portal named as Health Management Information System (HMIS) facilitated by Ministry of Health and Family Welfare (MoHFW), Government of India. The data consists of diverse health parameters pertaining to the maternal health of women and for this study, the maternal healthcare data of all districts of Jammu and Kashmir State was considered. Two data mining classifiers viz. Bayesian TAN and Naïve Bayes were applied for classifying the districts of Jammu and Kashmir State into High MMR and Low MMR districts based on the available past data from 2014 to 2018. Additionally, evaluation measures viz. Accuracy, F-measure, Area under the Curve (AUC), and Gini have been used to evaluate the performance of the models developed by Bayesian TAN and Naïve Bayes.

**Index Terms:** Data Mining, KDD, Classification, Bayesian TAN, Naïve Bayes, Maternal Health, Accuracy, F-measure, AUC, Gini.

## I. INTRODUCTION

Maternal health is an important public healthcare challenge throughout the world and reflects the entire spectrum of social development. Poor health among mothers has remained a question of worry for a long time. The better provisioning of primary health services as regards maternal health can be measured by the level of health outcome such as Maternal Mortality Ratio (MMR). MMR is a health outcome that is directly related to health parameters such as antenatal care, postnatal care, and number of institutional deliveries which form a part of the primary health services. MMR is the number of deaths due to pregnancy related causes per 100,000 live births [1].

The information as regards maternal health is required not only to understand the status of health of the population but also to know the requisite need of the population residing in a defined geographical area. Various schemes related to maternal health have been initiated in India for providing facilities to the beneficiaries including Janani Suraksha Yojana (JSY) in 2005, Accredited Social Health Activist (ASHA) in 2007, Monthly Village Health and Nutrition Day (VHND) in 2007, Mother Child Tracking System/Reproductive Child Health (MCTS/RCH) in 2009, Janani Shishu Suraksha Karyakaram (JSSK) in 2011, Mother and Child Tracking Facilitation Centre (MCTFC) in 2014, Kilkari in 2014, ANM on Line (ANMOL) in 2016, and Pradhan Mantri Surakshit Matritva Abhiyan (PMSMA) in 2016 [2]. The aim of Janani Suraksha Yojana (JSY) was to promote institutional deliveries by providing cash incentives to the beneficiaries who opt to deliver in the public health facility [3]. Accredited Social Health Activist (ASHA) is a community level worker to focus on women and children. The main task for ASHA workers is to encourage the women to seek Ante Natal Checkups (ANC) and give birth in health centres [4, 5, 6]. Monthly Village Health and Nutrition Day (VHND) is an Anganwadi activity for maternal and child care that includes nutrition also [7]. Mother Child Tracking System/Reproductive Child Health (MCTS/RCH) is a web based application to track all pregnant women and children to use full services [8, 9]. Janani Shishu Suraksha Karyakaram (JSSK) provides number of free facilities to pregnant women and infants like free delivery including C-section, to-and-fro transport, free diet, diagnosis, drugs, and blood transfusion [10]. Mother and Child Tracking Facilitation Centre (MCTFC) is a helpdesk that provides relevant information and guidance directly to pregnant women and parents of children [2]. The scheme Kilkari sends 72 audio messages related to pregnancy and child care to the mothers or parents of the children [11]. ANM on Line (ANMOL) is a tablet based application to empower Auxiliary Nursing Midwifery (ANM) for their day to day activities [2]. Pradhan Mantri Surakshit Matritva Abhiyan (PMSMA) is a free of cost Ante Natal Checkup (ANC) of pregnant women on 9th of every month [12]. Despite these various schemes launched by Ministry of Health & Family Welfare (MoHFW), Government of India (GoI), the pace of reducing MMR is very slow. The National Health Policy (NHP) 2002 has set a target to reduce the MMR from 254 to 100 by 2010 but they have achieved the MMR of 178 till 2010 [13, 14].

Revised Manuscript Received on December 30, 2019.

\* Correspondence Author

**Sourabh\***, Ph.D. from Department of Computer Science & IT, University of Jammu, India.

**Vibhakar Mansotra**, Professor in Department of Computer Science and IT, University of Jammu, India.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Now the National Health Policy (NHP) 2017 has again set the target from MMR 130 to MMR 100 by 2020 [15]. It means there is not a big change in the reduction of MMR from 2002 to 2017. India is spending only 1.4% of GDP on healthcare that is very less in comparison to US, UK, China and many other countries [16]. Even the percentage of GDP on healthcare that is spending by the neighboring countries including Bhutan, Nepal, and Maldives is more in comparison to India [17]. Data Mining has started playing a very important role in healthcare and intelligent decision making. Numerous researchers have applied data mining to medical data for prediction of specific diseases. In the present research work, data mining has been used on maternal health data for extracting knowledge for the purpose of decision making. Bayesian Tree Augmented Network (TAN) and Naïve Bayes have been used as the classifiers for classifying the maternal health data of Jammu and Kashmir State (now declared as Union Territory by the Government of India) for the years 2014-18 for knowledge discovery. Additionally, the evaluation measures including Accuracy, F-measure, AUC, and Gini have been used for the interpretation/evaluation of the models constructed by Bayesian TAN and Naïve Bayes.

## II. REVIEW OF LITERATURE

Numerous researchers have applied the data mining techniques on maternal health data. Some of the important works which assisted in the present study are mentioned in this section. A. Kamat et al. [18] developed a graphical user interface predictor tool to predict mode of delivery viz. normal and not-normal. The Naïve Bayes approach was observed to show better performance as compared to ID3. B. N. Lakshmi et al. [19] predicted the risk levels during pregnancy with accuracy of 71.30% for the standardized data and 66.08% for non-standardized data. M. W. L. Moreira et al. [20] in their study found that the TAN classifier produced the best results in terms of accuracy for predicting four hypertensive disorders during pregnancy than Averaged One Dependence Estimator (AODE) and Naïve Bayes classifiers. In the research work of S. Gupta et al. [21], Random Decision Tree was used for classifying the data into priority and non-priority districts and they observed that the number of home deliveries was more and visit of new born in 24 hours was less in priority districts as compared to non-priority districts in Uttar Pradesh. P. Kour et al. [22] classified the maternal health data of the districts of Jammu and Kashmir State, India for the year 2016-17 using Naive Bayes classifier into priority and non-priority districts. S. Gupta et al. [23] applied three methods viz. Exponential Smoothing (ES), Multi Layer Perception (MLP), and Long Short Term Memory (LSTM) to forecast the maternal healthcare data of Uttar Pradesh and results revealed that the Exponential Smoothing method has performed better than the other two methods. S. Shastri and

V. Mansotra [24] designed a conceptual framework KDD-MHCI based on Knowledge Discovery Process (KDP) for discovering knowledge from maternal health and child immunization databases. F. S. Gharehchopogh et al. [25] classified the data pertaining to pregnant women using C4.5 with 86.25% accuracy for predicting the deliveries as normal and C-section. C. Sunder et al. [26] utilized Naïve Bayes probabilistic model based classifier for classifying fetal heart Rate (FHR) into three class labels viz. normal, suspicious and pathological. M. W. L. Moreira et al. [27] developed models for hypertensive diseases in risk pregnancies using Naïve Bayes and Averaged One-dependence Estimators Classifier (AODE). Naïve Bayes indicated better performance in terms of precision and F-measure. In the research work of G. Sahle [28], the algorithm J48 depicted performance than algorithm JRip in terms of accuracy for predicting post-natal care visits in Ethiopia. The factors that affect postnatal care were place of delivery, assistance of health delivery professional, prenatal care health professional, and age. The research study of R. Mehta et al. [29] was carried out to predict the chance of high risk maternal patients which was helpful for timely detection and providing quality care. For this work, the decision tree showed best results, Support Vector Machine (SVM) reflected poor results whereas other classifiers including Naïve Bayes, K-Nearest Neighbour (KNN), and Artificial Neural Network (ANN) indicated average results. S. Sharmilan et al. [30] used a novel hybrid approach by combining ANN and NB algorithms that gave more accuracy than the individual algorithms for pregnancy complications diagnosis. The research of O. B. Alaba et al. [31] presented a predictive model for maternal mortality in Nigeria using three classification techniques viz. Naïve Bayes, Decision Tree classifier, and Multilayer Perceptron classifier for predicting the chances of having a safe delivery. Perceptron model emerged the most efficient with an accuracy of 84%.

## III. METHODOLOGY

### A. Data Mining Classifiers

In the present research work, two probabilistic classifiers of data mining viz. Bayesian Tree Augmented Naïve Bayes (TAN) and Naïve Bayes have been used for the classification of the districts of Jammu and Kashmir State using the Maternal Healthcare (State-MH) data. A brief overview of the two Bayesian classifiers used in the present research paper is presented in the following subsections.

#### *Bayesian TAN*

The Bayesian Tree Augmented Naïve Bayes (TAN) approach of Bayesian network is an extension of Naïve Bayes where the input parameters have relationship with each other in addition to the class label and thus considered as more realistic than Naïve Bayes.

It works better than Naïve Bayes if there are correlations between the input predictors but if there are not enough correlations between the input parameters then both classifiers viz. Naïve Bayes and Bayesian TAN perform in almost the same manner.

#### Naïve Bayes

Naïve Bayes classifier of data mining is a Bayesian classifier that is based on the Bayes theorem. Bayes theorem can be defined as:

$$P(H|X) = P(X|H) \cdot P(H) / P(X) \quad [32]$$

Where  $P(H|X)$  is the posterior probability of  $H$  conditioned on  $X$  and  $P(H)$  is the prior probability of  $H$ . In the same way,  $P(X|H)$  is the posterior probability of  $X$  conditioned on  $H$  and  $P(X)$  is the prior probability of  $X$ . The word naïve in Naïve Bayes classifier means that all the input parameters of the model are independent of each other and only depend on the class label. As Naïve Bayes is a statistical classifier, it predicts the class label probabilities such as the probability that a given tuple belongs to a particular class.

#### B. Proposed Algorithm for State-MH Dataset

In the present research problem, a number of maternal health parameters have been used initially for the classification into High MMR and Low MMR. Let  $D_s = \{P_1, P_2, P_3, \dots, P_{n-1}, P_n\}$  be the maternal health dataset with  $n$  number of parameters. The proposed algorithm is shown in Algorithm 1 and the goal is to evaluate the two classifiers viz. Bayesian TAN and Naïve Bayes to find out the best between them for the purpose of knowledge discovery. The data preprocessing of the dataset has been performed using group mean substitution and nearest non-outlier group substitution. Three feature selection methods viz. One R feature evaluation, Correlation feature evaluation and Relief F feature evaluation have been used in the present study and the top twenty ranked aggregated parameters among these 3 feature selection methods have been used finally for the purpose of modeling. The performance of the two supra mentioned classifiers have been evaluated by using the data mining classification evaluation measures viz. Accuracy, F-measure, AUC, and Gini.

#### ALGORITHM 1

**INPUT:** MH Dataset,  $D_s = \{P_1, P_2, P_3, \dots, P_{n-1}, P_n\}$

**OUTPUT:** To find the Best\_Classifier

#### // Data Preprocessing

1. Apply Imputation by Group Mean Substitution.
2. Apply Nearest Non-outlier Group Substitution.

#### // Feature Selection

Initially,  $P_s = \{\}$

3. Apply Feature Selection Methods.
4. One R Feature Evaluation:  $\{D_s\}$
5. Correlation Feature Evaluation:  $\{D_s\}$
6. Relief F Feature Evaluation:  $\{D_s\}$

7. Find aggregated top 20 ranked features among 3 FS methods.
8.  $P_s = 1/n \sum (1R, COR, ReliefF)$
9. Output:  $P_s$  is new Dataset for Modeling.

#### // Modeling & Prediction

##### A) Apply Bayesian TAN:

10.  $O_1 = \text{Bayesian TAN } (P_s)$
11.  $PE_{BT} = O_1$  (Performance\_Evaluation)

##### B) Apply Naïve Bayes:

12.  $O_2 = \text{Naïve Bayes } (P_s)$
13.  $PE_{NB} = O_2$  (Performance\_Evaluation)

#### // Comparison for Best\_Classifier

14. Find the Best\_Classifier (Accuracy, F-measure, AUC, and Gini)
15. Compare ( $PE_{BT}$ ,  $PE_{NB}$ )

**Output:** Best\_Classifier for Knowledge Discovery

#### Algorithm 1: State-MH Algorithm

The best classifier between Bayesian TAN and Naïve Bayes has to be selected for the discovery of knowledge from the maternal healthcare dataset.

#### C. State-MH Flow Diagram

The working of State-MH algorithm is shown through the flow diagram in figure 1. The dataset used in the present research work is of maternal healthcare from Jammu and Kashmir State (State-MH) from 2014-2018. Bayesian TAN and Naïve Bayes have been used as classifiers for the classification of the maternal healthcare data whereas Accuracy, F-measure, AUC, and Gini have been used to evaluate the models produced by these aforementioned classifiers. The preliminary screening of the parameters has been carried out by using three feature selection methods and the aggregated top twenty ranked features among three feature selection methods have been used for modeling for finding the best subset of features for training.



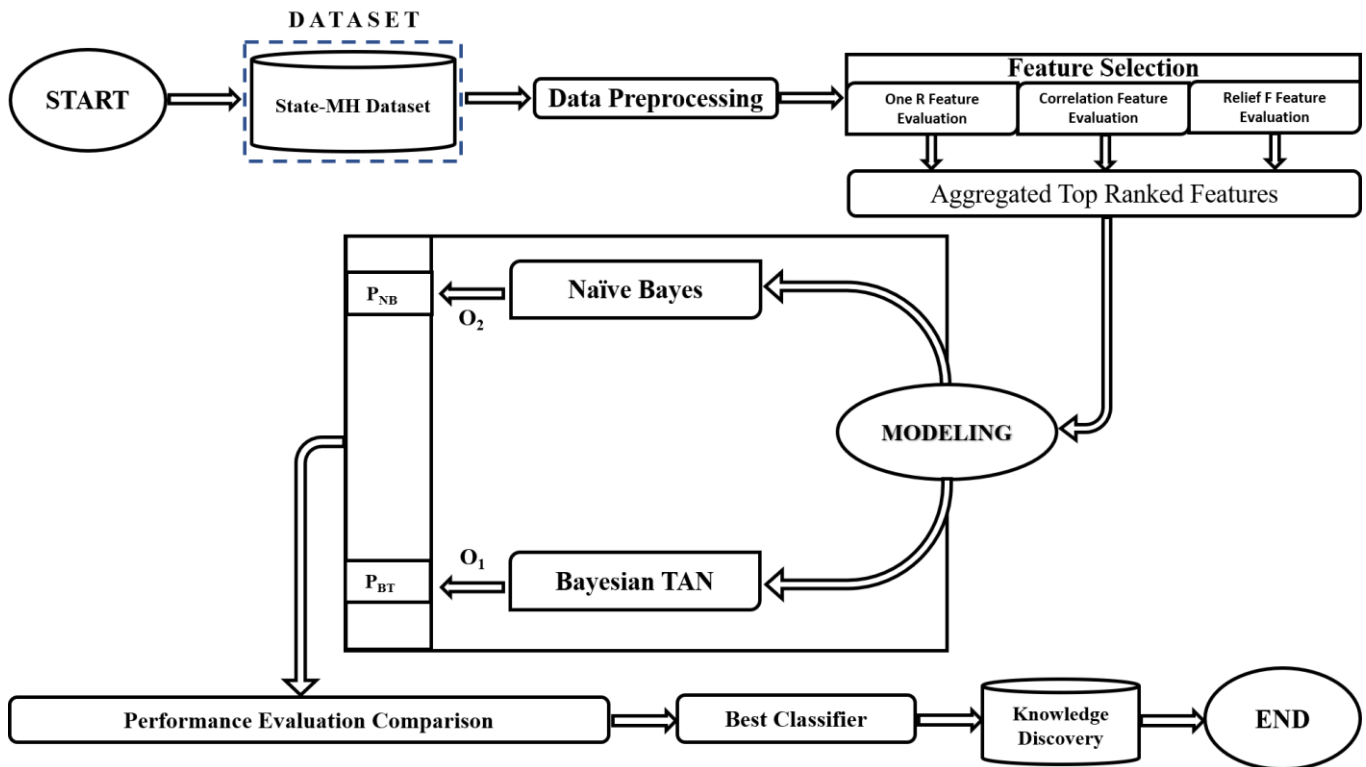


Fig. 1. State-MH Flow Diagram

## D. State-MH Dataset

The maternal healthcare dataset has been collected from the Health Management Information System (HMIS) portal of Ministry of Health & Family Welfare, Government of India [33]. The dataset initially included 33 parameters of all the districts of Jammu and Kashmir State and after the feature selection of the input parameters, only top 20 aggregated ranked parameters were used for the purpose of modeling as shown in figure 2.

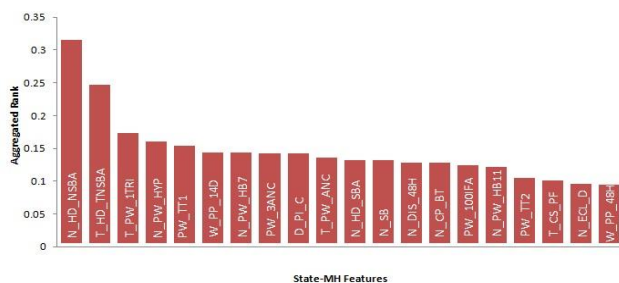


Fig. 2. Top 20 Aggregated Ranked Features

The class label used for the classification is MMR with two values viz. High and Low. The goal of the research work is to find out the best classifier between Bayesian TAN and Naïve Bayes for the discovery of knowledge from maternal healthcare data of Jammu and Kashmir State.

## IV. EXPERIMENTS AND RESULTS

### A. Experiments with Bayesian TAN

The Bayesian TAN structure is shown in figure 3 where the input parameters that were selected for the

modeling out of 20 aforementioned parameters are shown and depict a relationship with the class label MMR and correlation with one other head node parameter i.e. T\_PW\_ANC. The casual relationships between head node T\_PW\_ANC and other input parameters are shown in figure 3.

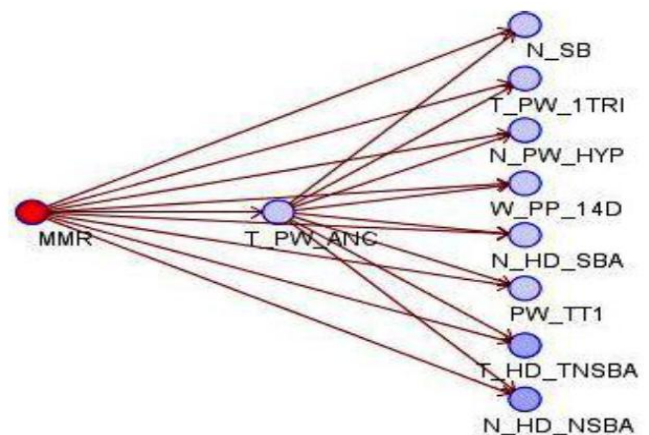


Fig. 3. Bayesian TAN Structure

Table 1 demonstrates the conditional probability of total number of pregnant women who have registered for ANC checkup (T\_PW\_ANC). In low MMR areas, 75% of the districts have T\_PW\_ANC value less than equal to 21,911 whereas in high MMR areas, all of the districts have T\_PW\_ANC value less than 21,911 and the rest of the values of T\_PW\_ANC with regard to low and high MMR districts are shown in table 1.

**Table 1: Conditional Probabilities of T\_PW\_ANC**  
T\_PW\_ANC

Parents		Probability			
MMR	T_PW_ANC	<= 21,910.6	21,910.6 ~ 50,615.5	50,615.5 ~ 79,320.4	> 79,320.4
LOW		0.75	0.12	0.06	0.06
HIGH		1.00	0.00	0.00	0.00

Tables 2-9 explains the conditional probabilities of the parameters N\_HD\_NSBA, T\_HD\_TNSBA, PW\_TT1, N\_HD\_SBA, W\_PP\_14D, N\_PW\_HYP, T\_PW\_1TRI, and N\_SB in relation to the head parameter T\_PW\_ANC.

**Table 2: Conditional Probabilities of N\_HD\_NSBA given**  
T\_PW\_ANC  
N\_HD\_NSBA

Parents		Probability					
MMR	T_PW_ANC	<= 475.6	475.6 ~ 949.2	949.2 ~ 1,422.8	1,422.8 ~ 1,896.4	> 1,896.4	
LOW	<= 21,910.6	0.75	0.00	0.00	0.17	0.08	
LOW	21,910.6 ~ 50,615.5	1.00	0.00	0.00	0.00	0.00	
LOW	50,615.5 ~ 79,320.4	1.00	0.00	0.00	0.00	0.00	
LOW	> 79,320.4	1.00	0.00	0.00	0.00	0.00	
HIGH	<= 21,910.6	0.17	0.17	0.33	0.17	0.17	

**Table 3: Conditional Probabilities of T\_HD\_TNSBA**  
given T\_PW\_ANC  
T\_HD\_TNSBA

Parents		Probability					
MMR	T_PW_ANC	<= 485.6	485.6 ~ 968.2	968.2 ~ 1,450.8	1,450.8 ~ 1,933.4	> 1,933.4	
LOW	<= 21,910.6	0.67	0.08	0.00	0.17	0.08	
LOW	21,910.6 ~ 50,615.5	1.00	0.00	0.00	0.00	0.00	
LOW	50,615.5 ~ 79,320.4	1.00	0.00	0.00	0.00	0.00	
LOW	> 79,320.4	1.00	0.00	0.00	0.00	0.00	
HIGH	<= 21,910.6	0.17	0.17	0.17	0.33	0.17	

**Table 4: Conditional Probabilities of PW\_TT1 given**  
T\_PW\_ANC  
PW\_TT1

Parents		Probability					
MMR	T_PW_ANC	<= 4,694	4,694 ~ 8,002	8,002 ~ 11,310	11,310 ~ 14,618	> 14,618	
LOW	<= 21,910.6	0.25	0.25	0.42	0.08	0.00	
LOW	21,910.6 ~ 50,615.5	0.00	0.00	0.00	0.00	1.00	
LOW	50,615.5 ~ 79,320.4	0.00	0.00	0.00	0.00	1.00	
LOW	> 79,320.4	0.00	0.00	0.00	0.00	1.00	
HIGH	<= 21,910.6	0.33	0.33	0.33	0.00	0.00	

**Table 5: Conditional Probabilities of N\_HD\_SBA given**  
T\_PW\_ANC

N\_HD\_SBA

Parents		Probability			
MMR	T_PW_ANC	<= 131.4	131.4 ~ 262.8	262.8 ~ 394.2	> 394.2
LOW	<= 21,910.6	0.92	0.00	0.00	0.08
LOW	21,910.6 ~ 50,615.5	1.00	0.00	0.00	0.00
LOW	50,615.5 ~ 79,320.4	1.00	0.00	0.00	0.00
LOW	> 79,320.4	1.00	0.00	0.00	0.00
HIGH	<= 21,910.6	0.67	0.17	0.17	0.00

**Table 6: Conditional Probabilities of W\_PP\_14D given**  
T\_PW\_ANC  
W\_PP\_14D

Parents		Probability					
MMR	T_PW_ANC	<= 4,695.4	4,695.4 ~ 8,189.8	8,189.8 ~ 11,684.2	11,684.2 ~ 15,178.6	> 15,178.6	
LOW	<= 21,910.6	0.83	0.17	0.00	0.00	0.00	
LOW	21,910.6 ~ 50,615.5	0.00	0.00	0.50	0.50	0.00	
LOW	50,615.5 ~ 79,320.4	0.00	0.00	0.00	0.00	1.00	
LOW	> 79,320.4	0.00	0.00	0.00	0.00	1.00	
HIGH	<= 21,910.6	0.83	0.17	0.00	0.00	0.00	

**Table 7: Conditional Probabilities of N\_PW\_HYP given**  
T\_PW\_ANC  
N\_PW\_HYP

Parents		Probability			
MMR	T_PW_ANC	<= 165.8	165.8 ~ 325.6	325.6 ~ 565.3	> 565.3
LOW	<= 21,910.6	0.75	0.25	0.00	0.00
LOW	21,910.6 ~ 50,615.5	0.00	0.50	0.50	0.00
LOW	50,615.5 ~ 79,320.4	0.00	0.00	0.00	1.00
LOW	> 79,320.4	0.00	0.00	0.00	1.00
HIGH	<= 21,910.6	0.83	0.00	0.00	0.17

**Table 8: Conditional Probabilities of T\_PW\_1TRI given**  
T\_PW\_ANC  
T\_PW\_1TRI

Parents		Probability			
MMR	T_PW_ANC	<= 11,578.8	11,578.8 ~ 21,939.6	21,939.6 ~ 37,480.8	> 37,480.8
LOW	<= 21,910.6	0.83	0.17	0.00	0.00
LOW	21,910.6 ~ 50,615.5	0.00	0.50	0.50	0.00
LOW	50,615.5 ~ 79,320.4	0.00	0.00	1.00	0.00
LOW	> 79,320.4	0.00	0.00	0.00	1.00
HIGH	<= 21,910.6	1.00	0.00	0.00	0.00

**Table 9: Conditional Probabilities of N\_SB given T\_PW\_ANC**

Parents		Probability			
MMR	T_PW_ANC	<= 177.8	177.8 ~ 351.6	351.6 ~ 612.3	> 612.3
LOW	<= 21,910.6	0.92	0.08	0.00	0.00
LOW	21,910.6 ~ 50,615.5	0.00	0.50	0.50	0.00
LOW	50,615.5 ~ 79,320.4	0.00	0.00	0.00	1.00
LOW	> 79,320.4	0.00	0.00	0.00	1.00
HIGH	<= 21,910.6	0.67	0.33	0.00	0.00

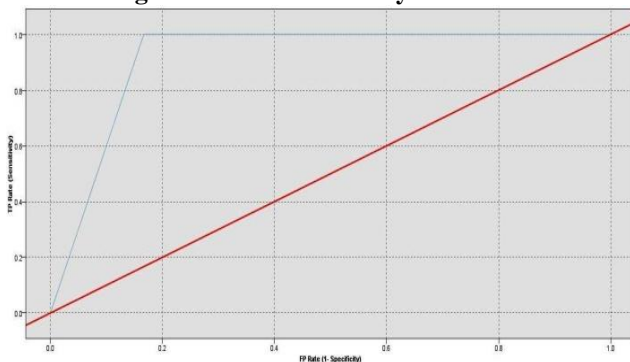
The confusion matrix for the Bayesian TAN model is shown in table 10 indicating that out of six high MMR districts, one district match with the criteria of low MMR districts whereas out of 16 low MMR districts, 01 district match with the criteria of high MMR districts.

**Table 10: Confusion Matrix for Bayesian TAN**

Partition	HIGH	LOW
HIGH	05 (TP)	01 (FN)
LOW	01 (FP)	15 (TN)

The Receiver Operating Characteristic (ROC) curve is shown in figure 4 and the Accuracy, F-measure, AUC, and Gini values of Bayesian TAN are shown in table 11.

**Fig. 4. ROC curve for Bayesian TAN**



**Table 11: Bayesian TAN Evaluation Measures**

State-MH Evaluation using Bayesian TAN	
Accuracy	90.91%
F-measure	0.833
AUC	0.953
Gini	0.906

#### B. Experiments with Naïve Bayes

The target field used for building the model was MMR and the input fields used were: N\_HD\_NSBA, T\_HD\_TNSBA, T\_PW\_1TRI, N\_PW\_HYP, PW\_TT1, W\_PP\_14D, N\_SB, and N\_HD\_SBA. The testing phase of Naïve Bayes is shown in figure 5.

**Fig. 5. Naïve Bayes Testing Phase**

Testing Phase

NB-MHCI

T\_PW\_1TRI: 4751

PW\_TT1: 4367

N\_PW\_HYP: 9118

N\_HD\_SBA: 1115

N\_HD\_NSBA: 94238

T\_HD\_TNSBA: 96

N\_SB: 7852

W\_PP\_14D: 9653

Calculate Probability

Reset Value

PREDICTED VALUE IS: LOW

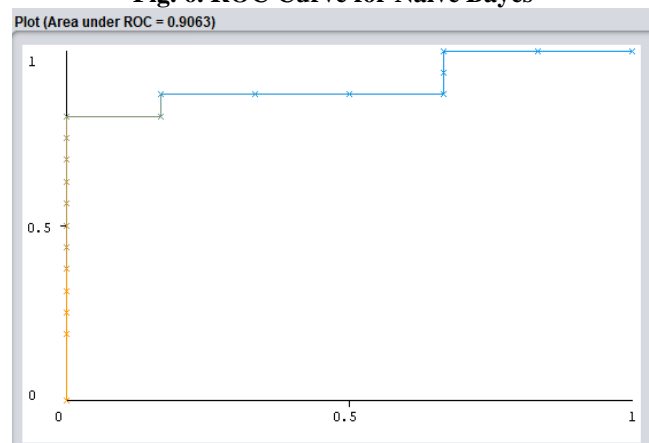
The confusion matrix for the Naïve Bayes model is shown in table 12 indicating that out of six high MMR districts, two districts match with the criteria of low MMR districts whereas out of 16 low MMR districts, no district match with the criteria of high MMR districts.

**Table 12: Confusion Matrix for Naïve Bayes**

Partition	HIGH	LOW
HIGH	04 (TP)	02 (FN)
LOW	00 (FP)	16 (TN)

The Receiver Operating Characteristic (ROC) curve is shown in figure 6 and the Accuracy, F-measure, AUC, and Gini values of Naïve Bayes are shown in table 13.

**Fig. 6. ROC Curve for Naïve Bayes**



**Table 13: Naïve Bayes Evaluation Measures**

State-MH Evaluation using Naïve Bayes	
Accuracy	90.91%
F-measure	0.800
AUC	0.906
Gini	0.812

#### C. Comparison and Discussion of the Results



The classifiers viz. Bayesian TAN and Naïve Bayes have achieved accuracy of 90.91% both but as far as the F-measure, AUC, and Gini are concerned the Bayesian TAN is performing better than Naïve Bayes. The Bayesian TAN has achieved the F-measure of 0.833 whereas F-measure of Naïve Bayes is only 0.800. Similarly, an AUC of 0.953 has achieved by Bayesian TAN whereas the Naïve Bayes has slightly less AUC of 0.906. Moreover, Gini of Naïve Bayes is slightly less than Bayesian TAN. The Gini of Bayesian TAN is 0.906 whereas Naïve Bayes has Gini of only 0.812 as shown in table 14. It is concluded that the performance of Bayesian TAN is slightly better than the performance of Naïve Bayes in terms of F-measure, AUC, and Gini. Therefore, Bayesian TAN classifier has been selected as the best classifier between Bayesian TAN and Naïve Bayes and thus considered for the discovery of knowledge.

**Table 14: Comparison between the performances of Bayesian TAN and Naïve Bayes**

Evaluation of the State-MH Classifiers		
State-MH Model Measures	Bayesian TAN	Naïve Bayes
Accuracy	90.91%	90.91%
F-measure	0.833	0.800
AUC	0.953	0.906
Gini	0.906	0.812

#### D. Knowledge Discovery from State-MH Data using Bayesian TAN

- In case of low MMR areas, when the total number of women registered for ANC was less than 21,911, then in 75% of the low MMR districts in this range have less than 475 home deliveries each that were attended by non-trained SBA whereas in high MMR areas, only 17% districts have less than 475 home deliveries attended by non-trained SBA and 83% districts have more than 475 home deliveries attended by non-trained SBA. Therefore, there is still need to encourage the people in high MMR areas of the State to prefer institutional deliveries rather than home deliveries.
- On the other hand, again in case of low MMR areas, when the total number of women registered for ANC was less than 21,911, then 67% of low MMR districts in this range have discerned to be less than 485 home deliveries attended by trained or non-trained SBA whereas in high MMR areas, only 17% districts were found to be less than 485 and rest districts of high MMR areas have greater than 485 home deliveries attended by trained or non-trained SBA.
- The intake of PW\_TT1 was also better in low MMR areas as compared to high MMR areas. Therefore, more awareness methods must be adopted by the Government in high MMR areas to awake the women regarding the benefits of Tetanus Toxoid during pregnancy.
- When the total number of women registered for ANC was less than 21,911, then 92% districts of low MMR areas in this range were noted with less than 131 home deliveries each that were attended by trained SBA whereas only 67% districts were found in case of high MMR areas.

- When the total number of pregnant women registered for ANC is less than 21,911, then in all districts of high MMR areas, the total number of pregnant women who registered within first trimester is less than 11,579 whereas in 17% districts of low MMR areas in this range have more than 11,579 pregnant women who registered within first trimester. Therefore, more awareness methods must be adopted by the Government to awake the women regarding the benefits of registering themselves in the first trimester.
- The number of pregnant women with hypertension (BP>140/90) were found more in high MMR areas as compared to low MMR areas.
- When the total number of pregnant women registered for ANC is less than 21,911, then in 67% districts of high MMR areas have less than 178 cases of still births and 33% districts have more than 178 cases. In contrast, in low MMR areas in this range, 92% districts have less than 178 cases of still births and only 8% districts have greater than 178 cases.

## V. CONCLUSION

The future of a country or state depends on the health of its citizens, particularly the mothers and children. In this paper, two data mining approaches of probabilistic classifiers viz. Bayesian TAN and Naïve Bayes were applied on the maternal healthcare dataset of Jammu and Kashmir State (State-MH) of India (declared now as a Union Territory by Government of India) from 2014-18 for knowledge discovery from the data and the final results indicated the Bayesian TAN as a better classifier than Naïve Bayes and thus the knowledge was discovered by using the TAN approach of Bayesian networks. The present study reveals that the situation in low MMR areas is better than high MMR areas and therefore efforts should be made for spreading awareness regarding important health indicators including atleast three antenatal checkups of pregnant women, registration within first trimester (within 12 weeks), institutional deliveries, and postnatal checkup within 48 hours after delivery to improve the maternal health. Further, it is pertinent to declare here that as institutional deliveries or facility based births are often promoted for reducing maternal and neonatal mortality, it is highly recommended that awareness programmes be held for institutional deliveries i.e. in a hospital and assisted by a skilled attendant as skilled birth attendants have the potential to handle normal deliveries, thereby, detecting complications at early stage and treat them appropriately and in all this, economic status can also be considered as a more crucial determinant than access. In the future, we will extend this work at national level with a variety of classification algorithms and ensemble learning.

## REFERENCES

1. S. Gupta, S. N. Singh, and D. Kumar, "An Empirical Analysis of Maternal Health Data: A Case Study of India," In Proc. International Conference on Next Generation Computing Technologies, 2016, pp. 490-493.
2. Ministry of Health and Family Welfare, "Annual Report 2018-19," Government of India, New Delhi, India, 2018. [Accessed: Oct. 10, 2019]. [Online]. Available: <https://mohfw.gov.in/documents/Publications/annual-report-2018-19>
3. P. Sinoj and S. Mishra, "Status of Maternal Health and its services in Gujarat: Some Observations," *International Journal of Social Science & Interdisciplinary Research*, vol. 1, no. 9, pp. 90-98, 2012.
4. F. N. Fathima et al., "Assessment of 'Accredited Social Health Activists'- A National Community Health Volunteer Scheme in Karnataka State, India," *J Health Popul Nutr*, vol. 33, no. 1, pp. 137-145, 2015.
5. M. Choudhary et al., "Evaluation of Knowledge of ASHA Workers regarding various Health Services under NRHM in Saurashtra Region of Gujarat," *National Journal of Community Medicine*, vol. 6, no. 2, pp. 193-197, 2015.
6. S. Shet, K. Sumit, and S. Phadnis, "A study on assessment of ASHA's work profile in the context of Udipi Taluk, Karnataka, India," *Clinical Epidemiology and Global Health*, vol. 6, no. 3, pp. 143-147, 2018.
7. V. Saxena et al., "Planning and preparation of VHND through convergence: Sharing experiences from Uttarakhand," *Clinical Epidemiology and Global Health*, vol. 3, no. 3, pp. 125-131, 2015.
8. R. Gera et al., "An in-depth assessment of India's Mother and Child Tracking System (MCTS) in Rajasthan and Uttar Pradesh," *BMC Health Services Research*, vol. 15, no. 315, pp. 1-14, 2015.
9. R. K. Rai, "Tracking women and children in a Continuum of Reproductive, Maternal Newborn, and Child Healthcare (RMNCH) in India," *Journal of Epidemiology and Global Health*, vol. 4, no. 3, pp. 239-243, 2014.
10. S. Chaudhary, R. Rohilla, V. Kumar, and S. Kumar, "Evaluation of Janani Shishu Suraksha Karyakram scheme and out of pocket expenditure in a rural area of Northern India," *Journal of Family Medicine and Primary Care*, vol. 6, no. 3, pp. 477-481, 2017.
11. A. LeFevre et al., "Are stage-based health information messages effective and good value for money in improving maternal newborn and child health outcomes in India? Protocol for an individually randomized controlled trial," *Trials*, vol. 20, no. 272, pp. 1-12, 2019.
12. R. Sharma, R. Manchanda, and N. Chandil, "Impact of 'ANC Education Session' on Rural Mothers of South Delhi: A Step to prevent Maternal Mortality and Morbidity," *International Journal of Current Advanced research*, vol. 7, no. 9(B), pp. 15319-15322, 2018.
13. Ministry of Health and Family Welfare, "National Health Policy 2002," Government of India, New Delhi, India, 2002. [Accessed: Oct. 12, 2019]. [Online]. Available: [https://www.nhp.gov.in/sites/default/files/pdf/National\\_Health\\_Policy.pdf](https://www.nhp.gov.in/sites/default/files/pdf/National_Health_Policy.pdf)
14. Niti Aayog, "Maternal Mortality Ratio (MMR) (Per 100000 Live Births)," Government of India, New Delhi, India. [Accessed: Oct. 22, 2019]. [Online]. Available: <https://niti.gov.in/content/maternalmortality-ratio-mmr-100000-live-births>
15. Ministry of Health and Family Welfare, "National Health Policy 2017," Government of India, New Delhi, India, 2017. [Accessed: Oct. 31, 2019]. [Online]. Available: [https://mohfw.gov.in/sites/default/files/91475629414897531\\_21.pdf](https://mohfw.gov.in/sites/default/files/91475629414897531_21.pdf)
16. Business Insider, "Health Budget 2019: India might increase spending on healthcare but that may not be enough," 2019. [Accessed: Nov. 02, 2019]. [Online]. Available: <https://www.businessinsider.in/healthbudget-2019-india-might-increase-spending-on-healthcare-but-that-may-not-beenough/articleshow/69956156.cms>
17. Hindustan Times, "India spends less on health than Bhutan, Nepal, and Maldives," 2019. [Accessed: Nov. 03, 2019]. [Online]. Available: <https://www.hindustantimes.com/editorials/indiaspends-less-on-health-than-bhutan-nepal-andmaldives/story-vtRAzI3gzng0sbocvy9V3L.html>
18. A. Kamat, V. Oswal, and M. Dater, "Implementation of Classification Algorithms to Predict Mode of Delivery," *International Journal of Computer Science and Information Technologies*, vol. 6, no. 5, pp. 4531-4534, 2015.
19. B. N. Lakshmi, T. S. Indumathi, and N. Ravi, "A Comparative Study of Classification Algorithms for Predicting Gestational Risks in Pregnant Women," In Proc. International Conference on Computers, Communications and Systems, 2015, pp. 42-46.
20. M. W. L. Moreira et al., "Performance Evaluation of the Tree Augmented Naïve Bayes Classifier for Knowledge Discovery in Healthcare Databases," In Proc. XVII Workshop de Informática Médica, 2017, pp. 1835-1844.
21. S. Gupta, S. N. Singh, and P. K. Jain, "Data Mining on Maternal Healthcare Data for Decision Support," In Proc. IEEE 5<sup>th</sup> International Conference on Computing for Sustainable Global Development, 2018, pp. 4989-4993.
22. P. Kour et al., "Classification of Maternal Healthcare Data using Naïve Bayes," *International Journal of Computer Sciences and Engineering*, vol. 7, no. 3, pp. 388-394, 2019.
23. S. Gupta, S. N. Singh, and P. K. Jain, "Time Series Forecasting to Improve Predictive Modelling in Public Maternal Healthcare Data," *Recent Patents on Engineering*, vol. 13, no. 1, 2019, <https://doi.org/10.2174/1872212113666190304163200>
24. S. Shastri and V. Mansotra, "KDD-Based Decision Making: A Conceptual Framework Model for Maternal Health and Child Immunization Databases," In Proc. Springer 3<sup>rd</sup> International Conference on Advances in Computer Communication and Computational Sciences, 2018, pp. 243-253.
25. F. S. Gharehchopogh, P. Mohammadi, and P. Hakimi, "Application of Decision Tree Algorithm for Data Mining in Healthcare Operations: A Case Study," *International Journal of Computer Applications*, vol. 52, no. 6, pp. 21-26, 2012.
26. C. Sundar, M. Chitradevi, and G. Geetharamani, "An Analysis on the Performance of Naïve Bayes Probabilistic Model Based Classifier for Cardiocogram Data Classification," *International Journal on Computational Sciences & Applications*, vol. 3, no. 1, pp. 17-26, 2013.
27. M. W. L. Moreira et al., "An Inference Mechanism using Bayes-based Classifiers in Pregnancy Care," In Proc. IEEE 18<sup>th</sup> International Conference on e-Health Networking, Applications and Services, 2016.
28. G. Sahle, "Ethiopic maternal care data mining: discovering the factors that affect postnatal care visit in Ethiopia," *Health Information Science and Systems*, vol. 4, no. 4, pp. 1-8, 2016.
29. R. Mehta, N. Bhatt, and A. Ganatra, "A Survey on Data Mining Technologies for Decision Support System of Maternal Care Domain," *International Journal of Computer Applications*, vol. 138, no. 10, pp. 20-24, 2016.
30. S. Sharmilan and H. T. Chaminda, "Predictive Data Mining System to diagnose Pregnancy Complications," In Proc. International Conference on Computational Modeling & Simulation, 2017.
31. O. B. Alaba, P. A. Idowu, O. A. Abass, and I. O. Faniyi, "Development of a Predictive Model for Maternal Mortality in Nigeria using Data Mining Technique," In Proc. 11<sup>th</sup> International Multi-Conference on ICT Applications, 2017, pp. 25-32.
32. J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. Waltham, MA: Morgan Kaufmann Publishers, 2012.
33. Health Management Information System, "HMIS Standard Reports," Government of India, New Delhi, India, 2018. [Accessed Nov. 04, 2019]. [Online]. Available: <https://nrhmmis.nic.in/SitePages/Home.aspx>

## AUTHORS PROFILE



**Sourabh** is pursuing Ph.D. from Department of Computer Science & IT, University of Jammu under the guidance of Professor Vibhakar Mansotra and currently working as Assistant Professor in Department of Computer Science and IT, Kathua Campus, University of Jammu. He is a life member of Computer Society of India and has published more than 20 research papers in International Journals. He has more than 10 years of teaching experience and his main research work focuses on Analysis and Design of Algorithms, Data Mining, and Machine Learning.



**Dr. Vibhakar Mansotra** is the senior most Professor in Department of Computer Science and IT, University of Jammu. He has completed his M.Tech. from IIT Delhi and Ph.D. from University of Jammu. He is also Dean Mathematical Sciences and Director IT, University of Jammu. Additionally, he is holding the charge of Coordinator IGNOU (S.C-1201), University of Jammu. He has more than 27 years of teaching experience at University of Jammu. He has been conferred with the Best Teacher Award in Information Technology by the Amar Ujala B-School Excellence Awards. He is a life member of Computer Society of India (CSI) and presently holding the charge of Chairperson Division-IV. He has published several research papers in National and International Journals and his research interests are Computer Graphics, Data Mining, Information Retrieval, and Machine Learning.

