

Data Mining on Some Factors Related to the Stillbirth of Mothers Referred to Health Centers in Some Provinces of Iran

Nazanin Rahmati

Islamic Azad University, Shahrood

Hadiseh Kamalgharibi

Ferdowsi University of Mashhad

Mohammad Mehdi hosseini (✉ hosseini_mm@yahoo.com)

Islamic Azad University, Shahrood

Alireza Jalali

Islamic Azad University, Shahrood

Research Article

Keywords: relationship, risk, health, mothers

Posted Date: January 6th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-134695/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Data mining on some factors related to the stillbirth of mothers referred to health centers in some provinces of Iran

Nazanin Rahmati¹, Hadiseh Kamalgharibi², Mohammad Mehdi Hosseini^{3*}, Alireza Jalai⁴

1. Department of Computer Engineering, shahrood branch, Islamic Azad university, shahrood, Iran.
Naz.rahmati@gmail.com
2. Department of Computer Engineering, Ferdowsi University of Mashhad, Mashhad, Iran.
hadis.kamalgharibi@gmail.com
3. Department of Computer Engineering, shahrood branch, Islamic Azad university, shahrood, Iran.
Hosseini_mm@yahoo.com
4. Department of Computer Engineering, shahrood branch, Islamic Azad university, shahrood, Iran.
Drjalali16@gmail.com

Abstract:

Purpose: This study aimed to determine the relationship between maternal risk factors and stillbirth in some provinces. In the field of health, databases contain a wide range of public health variables. The purpose of this article is to explore the concepts related to data mining and its application in the field of health.

Methods: This study is a case-control. The study population consisted of 3340 mothers who referred to health centers of some provinces of the country in 2016. The 1459 mothers had a history of stillbirth and 1881 of them had a live birth at the last pregnancy. Then, data mining algorithms were used to investigate the desired data.

Results: Number Abortion , Mo Blood Group, Number Live Birth, Parity pregnancy, tanesh life, Mo Age ,Neonate weight, Smoking during pregnancy, Mo Educate, Number Death Neonate, Job Number, City Number, and Type Delivery Maternal risk factors are stillbirths. The accuracy of Naïve Bayes 82.71%, Logistic Regression 87.42%, Deep Learning 86.06%, Decision Tree 87%, Gradient Boosted 84.49%, Svm 86.79%.

Conclusion: This study supports the hypothesis that Factors affecting stillbirth in some algorithms are somewhat identical and in others different and that different preventive and treatment strategies might be required. The overall conclusion of this study shows that Logistic Regression models and Decision Tree are the most suitable models for prediction (classification) Stillbirth mothers (among the models investigated in this study). On the most important predictors, variables NumberAboration, Paritypregnancy, taneshlife, NumberDeathNeonate, City Number are presented as the most critical predictors in this study.

1. Introduction

Stillbirth is one of the most common and concerning outcomes of pregnancy which causes physiological and physiological consequences for mother and is costly for family and society. Today, prevention of stillbirth is one of the challenges of maternal care. Stillbirth is related different conditions and its main reason is not clearly understood and is a multifactorial phenomenon. Therefore, stillbirth can be considered as one of the most common consequences of pregnancy. One of the most challenging issues in health care is the transformation of raw clinical data into meaningful information, followed by continuous production of large amounts of data, as many health care organizations have been faced with a lack of knowledge despite the abundance of data. In other words, confronting with very large data sets and the development of databases compared to the past decades have created new needs Such as automated data abstraction, extraction of stored information, and discovery of patterns from raw data, of which data mining is an example. The purpose of predictive data mining in the medical and clinical field is to get a model that can predict outcomes using patient-specific information and make decisions based on it.

Determining the reliable accuracy of predictive models, especially in the field of medical problem prediction, is generally complicated. One of the reasons for the complexity is that vital data is not collected at an important time and place. Much of the past research in the field of survival prediction has been applied first to various statistical techniques and then to artificial neural networks (which are subsets of data mining techniques). Of course, other techniques such as decision tree, decision machine and business networks (Bayesian belief networks) have also been studied in a few studies, which are discussed here.

2. Related Work

In the following, the results of surveys and searches showed that there are no study comparing two predictive algorithms, decision tree method and support vector machine method, so far. Therefore, the purpose of this study was to determine and compare the performance of these algorithms with each other in predicting maternal stillbirth. According to the [1-3],the results of these studies showed that the prediction of stillbirth in mothers over 35 years old were higher than those of mothers 35 years old and younger . which is consistent with the results of other studies. Hajian et al. suggested that stillbirth risk may be related to increased rates of miscarriage due to aging, a history of previous stillbirths, multiparity, and placental abruption, which together increase the chance of stillbirth in older mothers [4]. In [3, 5] showed that the chances of stillbirth in mothers With primary education were higher than

those with college educations, The reason for such a relationship can be the high awareness of educated mothers of the consequences of pregnancy, more observance of pregnancy-related issues and more attention to referral to health centers for prenatal care. Akolekar et al [1] reported that women with a history of miscarriage had an increased risk of stillbirth. which were consistent with the results of other studies.

in [6] stated the risk of stillbirth will increase if women have a history of abortion or stillbirth . The most important variables that were examined in the model of relationship between lifestyle and stillbirth are [7, 8].

tobacco use status in mother and her spouse, maternal exposure to cigarette smoke, parental kin ratio, maternal physical activity. Among these variables, family ratios, mother's education, mother's place of residence, pregnancy-related illness and exposure to cigarette smoke remained in the model after adjustment and increased the chances of stillbirth. One of the influential factors in the lifestyle model is the relative parental ratio of the fetus, which is 77% higher among couples who have a family relative than those who did not have a family ratio [7].

common genetic factor in particular relatives and races increases the chance of malformations and death in the fetus and neonate[9]. A study by Hyland at the Roswell Oral Cancer Institute in Buffalo, New York, showed that pregnant women who are exposed to high levels of smoke had a higher rate of abortion, stillbirth and fetal mortality [10]. In a meta-analysis study of Amina's 2014 study of factors and factors associated with stillbirth in middle and low-income countries. The study found that significant factors including demographic and lifestyle variables, maternal age, low family income, and low education were significantly associated with stillbirth. On the other hand, smoking during the first trimester of pregnancy increases the risk of stillbirth by 2.4 times [2]. In a cohort study conducted by Jason Gardasil et al., Maternal and Fetal Risk Factors for stillbirth. There was a significant relationship between maternal pregnancy rating (first pregnancy > 3), smoking, diabetes, mental health problems, bleeding, and maternal mortality [9]. A Study on the Frequency of Intrauterine Death and Some Related Factors conducted by Zarei et al, It was concluded that factors such as maternal age, ethnicity, history of hypertension, diabetes, pregnancy complications (preeclampsia), history of intrauterine death, fetal weight and fetal malformations increase intrauterine death [11]. In a study conducted by Ulla Waldenström et al, they concluded that with increasing age in mothers, with the first pregnancy, the risk of stillbirth increases regardless of education. The risk of stillbirth increases with age in low- and middle-aged mothers in the second pregnancy [12]. In a study conducted by Sharma et al., the study found that a 37% incidence of stillbirth in mothers with a history of There was more stillbirth. There was also a significant relationship between late and premature births in mothers with a history of stillbirth [13].

The results of the [14] show that Mothers aged 20-24 years had a lower risk of stillbirth compared to younger mothers and also over 35 years old. A history of preeclampsia, a history of preterm birth, and a history of stillbirth and abortion, the first pregnancy was associated with an increased risk of stillbirth. In the second pregnancy, the risk of stillbirth was 10 times higher. In general, the chances of stillbirth in mothers over 35 were higher than mothers under 35 years of age. The chances of stillbirths in mothers with less education were also much higher than in the rest of the study groups. Occupation also has a significant association with stillbirth, so mothers who are exposed to stressful occupations have a higher chance of stillbirth. The chances of stillbirth in mothers with a history of stillbirth and abortion are also higher than those of mothers who did not have a miscarriage or stillbirth.

3. Materials and methods

3.1. Study population

Data collection method at the stage of selecting and using algorithms is to study texts and articles, and to examine the most famous predictive classification algorithms. Data collection method at the stage of evaluating the software is to ask the user using a predetermined checklist (table 6) based on the software usability assessment criteria. The present study is a case-control one. The study population consisted of 3340 mothers who referred to public health centers of Fars, Hormozgan, Kermanshah, Hamedan, Kohgiluyeh and Boyerahmad, Yazd, South Khorasan, Golestan, Mashhad and Zahedan. The 1459 mothers had a history of stillbirth and 1881 group of them had a live birth at the last pregnancy. The data collection tool in this study is a checklist that was collected by epidemiology researchers and given to me as a format of SPSS file. This study examines the possibility of maternal stillbirth with regard to previous maternal pregnancies with 2 possible occurrences (live birth, death birth). In the study, the features discussed were separated from a huge amount of information obtained by epidemiological researchers. Inefficient data as well as zero data were separated and then the requested information was added to the SQL server 2014 software as an Excel file. At first, created the tables and imported the data into the corresponding tables and then, using XAMPP software and building the corresponding Local Host, transferred the data from SQL to Local Host using Full Convertor software. Then to import the data into MATLAB 2013, selected Database Explore from the App section and after that chose Microsoft Ansi Driver. Then, added the Local Host specifications to MATLAB, then linked MATLAB to the Database. Also, used Rapid Miner studio for data mining. Finally, it should be noted that due to the wide range of variables used in the questionnaire and due to differences in variables in terms of risk mechanisms for mothers and related consultations with professors, it was decided that modeling should be based on similar variables and on specific goals separately.

The variables used in this paper are as follows: maternal age, maternal occupation, maternal education, maternal blood group, city name, type of delivery, number of live births, number of abortions, maternal ethnicity, whether the marriage was family, infant weight, environmental stress and mother's stress and those around her, smoker's mother and those around her and whether the mother has contraceptive method was examined.

This population-based case-control study was conducted in 10 provinces of the country and it included the following algorithms: Naive Bayes, Logistic Regression, Deep Learning, Decision Tree, Gradient Boosted Trees, Support Vector Machine method were used to predict neonatal birth. Data analysis was performed by data mining using Rapid Miner software and SVM algorithm in MATLAB. In this research, Rapid Miner data mining software was used for modeling to categorize mothers into their future birth. These data include general variables of maternal age, occupation, education, etc., as well as variables such as maternal life stress and maternal smoking. Given the desired output after data mining, various data mining techniques can be used to produce the desired output. Data mining tasks can be divided into two groups: descriptors and predictors. Descriptive explorers have the task of identifying data properties in databases. These types of explorers extract new and valuable information from the data. Predictive explorers have the task of deducing data for the purpose of predicting action, which creates a system model for us [15]. Due to the differences in data mining tasks, data mining methods can be categorized as categorization, clustering, regression, associative rules, sequence discovery, prediction, etc. In fact, by Descriptive or predictive explorers can be created with the help of data mining techniques. Data mining methods and applications are illustrated in Figure 1.

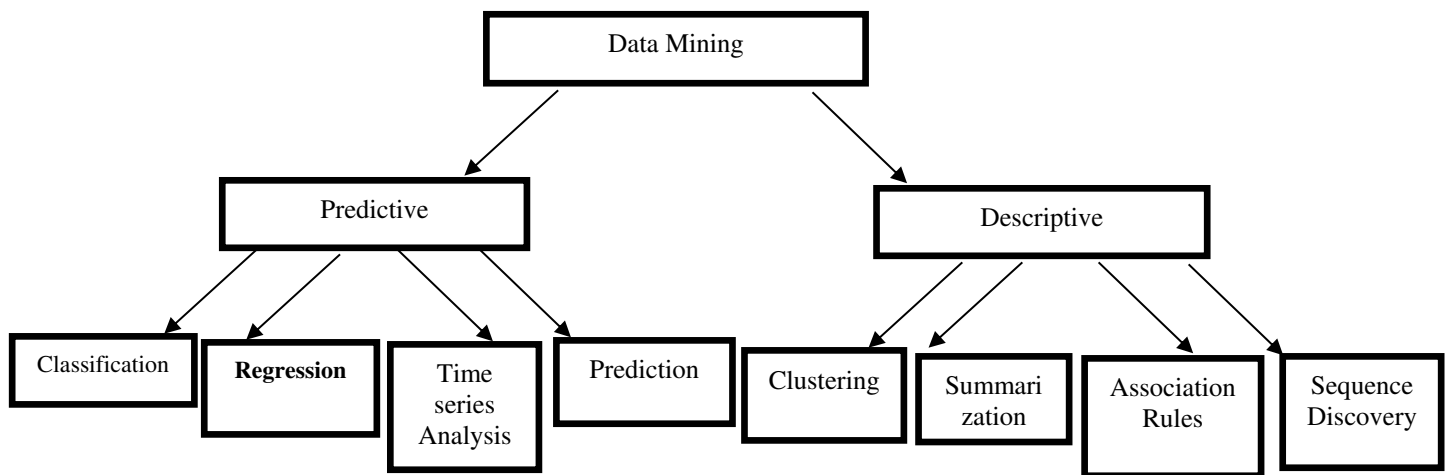


Figure 1. Data Mining Methods and Applications

An explorer is a predictor that maps data according to predefined classes [15]. Since classes are defined before data mining, the class is a type of supervised training. Classification algorithms must define classes based on data attribute values. Classification techniques include decision tree techniques, rule-based methods, memory-based arguments, artificial neural networks, Bayesian networks, belief networks. Bayesian and vector machines are the decision [16]. There are many variables that have a significant impact on the calculation of live birth mothers but have not been considered in previous studies due to the limitations of the methods used - so in this study, Data processing was attempted (clearing-integration-data conversion and identifying).

Using an integrated data mining method to show more effective relationships on survival and predictive factors. The dependent variable of the study, which is called survival, was considered to be "0" as non-survival and "1" as survival. Already, the relationship between this dependent variable and the independent / predictor variables of the study (all of these independent variables were used as inputs for all models and the reason for choosing them is to review previous studies, consult with respected clinical professors, and evaluate the efficacy of each model). In the next step, then used binary dependent variable predictor data mining methods to predict the survival of mothers with live births. These classification models show the relationships between dependent variable and predictor variables. Applies as well as predictive variables based on their importance in forecasting next, confusion matrix, lift chart, etc. can be used to evaluate the models. Each of the data mining models used in this study provides us with a different set of important predictor variables. An overview of integrated model production with data mining techniques is shown in Figure 2.

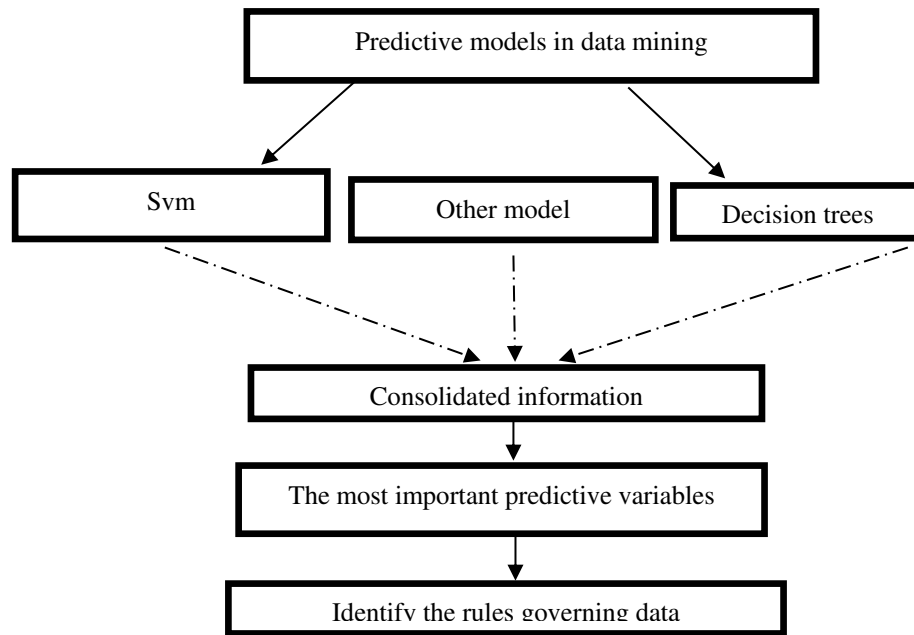


Figure 2. Overview of integrated model generation with data mining techniques

3.1.1. Decision Tree Algorithm

It is a decision support tool that uses trees to model. Decision tree is commonly used in various research and operations. It is specifically used in decision analysis to identify the strategy most likely to be targeted. Another use of decision trees is to describe conditional probability calculations. Decision trees place data recursively in branches according to a predetermined criterion to maximize the accuracy of the prediction results using a tree structure [17, 18]. The most popular decision tree algorithms are ID3, C4.5, and C5. Compared to other machine learning methods, the advantage of decision trees is that they are not black box models and hence the rules can easily be derived. The advantage, he explained, is that these algorithms are widely used in medicine [17].

3.1.2. Support Vector Machine

It is one of the classification methods that aims to create a hyper plane with the largest margin of samples available at the boundary between two classification classes that are linearly separated from each other. If the samples are not linearly separable, the data will be mapped to a larger dimension so that they can be linearly separated in this new space. This method is generally very accurate in classification. Decision machine is one of the most popular and powerful techniques in machine learning. Using this technique has led to some good modeling successes and in some cases has covered neural network deficiencies. This technique has been used to design data points in multidimensional space and to draw a separate metaphor for differentiation. The results of the operation are different. Of course, this space is not linear. The vector machine formulation has a unique optimum state and training always tends to this optimum state. This technique works well with noise data. However, this technique, such as neural network, is an efficient model because it has the properties of a black box. It often reveals hidden relationships, but no complete information is available on how to do this.

3.1.3. Logistic regression

Generalized logistic regression is a linear regression used to predict binary or multilevel dependent variables, since the response variable is a discrete variable so it cannot be directly modeled with linear regression so instead of predicting point estimates An event is designed to predict the probability of that event occurring. In a two-class problem, the probability of more than 50% means that the class assigned to class one will be zero, otherwise this technique will be used to smooth the data [19]. Relation 1 represents the formula. Calculate logistic regression.

$$\begin{aligned} \text{logit}(p) &= \ln \frac{p}{1-p} = \alpha + \beta_1 x_{1,i} + \dots + \beta_k x_{k,i} \\ i &= 1, \dots, n \quad p = \Pr(Y_i = 1) \\ p &= \Pr(Y_i = 1|X) = \frac{e^{\alpha + \beta_1 x_{1,i} + \dots + \beta_k x_{k,i}}}{1 + e^{\alpha + \beta_1 x_{1,i} + \dots + \beta_k x_{k,i}}} \end{aligned} \quad (1)$$

3.1.4. Neural Networks

Neural networks are used to investigate the complex relationships between predictor variables and dependent variables such as nonlinear and multiple functions. According to formal definition, neural networks are

highly sophisticated analytical techniques capable of predicting new observations from other observations are following the implementation of a so-called "learning" process of existing data. Neural networks are the most popular artificial intelligence-based data modeling algorithms that are used with good predictive performance in clinical medicine. In a study by Hornick et al [20], they experimentally demonstrated that the Due to their structure and size, multilayer perceptron networks are capable of learning complex arbitrary nonlinearity with a level of arbitrary precision. Multilayer neural networks are basically organized from a set of nonlinear neurons (called perceptron's) and are interconnected in a forward multilayer structure.

3.1.5. Bayesian Navigation techniques and Bayesian belief networks

Bayesian reasoning is probability-based. In fact it gives us a set of traits with prior conditional probabilities, Bayesian theory is used to assign particular probabilities to different hypotheses. There are many different techniques for converting continuous variables to discrete variables, although the choice of a discretization technique can have a significant impact on model performance. This method is also used to model the probability of missing values. Bayesian models and Bayesian belief networks are two general forms of these models [21]. In the Bayes Navigation classification it is assumed that the different results are unique and the effects of all conditional predictions are independent of all other predictions, while the assumption seems to be relatively accurate and the Bayesian predictor has shown that it works well in many situations. The bourbizine networks represent a statistical relationship structure between variables and provide an easy interpretation by creating a distant graph as well as a conditional probability table for each node of the population under study. In Bayesian networks, data is assumed to have a stable distribution, indicating that the Structure of these networks are able to describe the joint distribution of data [22] The relationship 2 shows the calculation formula on Navi techniques and Bayesian Belief Networks 3 represents a formula for calculating techniques which are listed below.

$$P(C|F_1, \dots, F_n) = \frac{p(C)P(F_1, \dots, F_n|C)}{P(F_1, \dots, F_n)} \quad (2)$$

$$P(R = T | G = T) = \frac{P(G=T | R=T)}{P(G=T)} = \frac{\sum_{S \in \{T, F\}} P(G=T, S, R=T)}{\sum_{S \in \{T, F\}} P(G=T, S, R)} \quad (3)$$

3.2. Performance criteria for model evaluation

To compare the classification models, the three criteria of accuracy, sensitivity, and specificity performance were used, which how to calculate them will be told. First, the elements in the confusion matrix (The matrix displaying classification results is called the confusion matrix or confusion matrix, Table 1 represents this matrix) was identified and then the relationships based on these elements, which will be discussed. The values in the confusion table are TP (true positive) and TN (true negative) and FP (false positive) and FN (false negative).

Table 1 Displays the Confusion Matrix for the Class Classification Problem

Real Classification		Classification by mode	
		Positive	Negative
	Positive	TP	FN
	Negative	FP	TN

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad (5)$$

$$\text{Specificity} = \frac{TN}{TN+FP} \quad (6)$$

3.2.1. Lift chart

A lift chart is a parametric-defined graph resembling a rock diagram used in predictive analysis based on statistical techniques, each binary classification corresponding to a point being placed in the parametric space of the graph. However, in some applications, the representation may be downward. At each point in the curve is a combination of a classification that is closer to the curve indicating that the sample or point is ahead. His real nose is closer. Its relation is as relation 2.

3.2.3 Roc

Another important criterion used to determine the performance of a cluster is the Area under Curve (AUC). The AUC represents the Receiver Operating Characteristic (ROC) level where the greater the value of a cluster, the greater the final performance. ROC diagrams are a way of checking the performance of clusters, in fact

ROC curves are two-dimensional curves in which DR is the True Positive Rate (TPR) on the Y axis and similarly to the FAR or The same False Positive Rate (FPR) is plotted on the X-axis, in other words a ROC curve of relative reconciliation between Shows profits and costs.

4. Experimental Results

The 3340 mothers were studied in this study, according to Table 2, the 1061 of whom were in the age range of 14–24. The 1958 of them were in the age range of 25–35 years and 320 of them were in the age range of 36–62. The 1234 of the mothers had high school education, two of them had secondary education, three of them had primary education, three of them were illiterate and two of them had a college degree. The 3065 were housewives, 191 were employees and the rest were freelancers.

Table 2 Frequency and frequency percentage of variables used in predicted models

Variable name	Description	Variable type	Values	Frequency	(%)
MoAge	Mother's age	Categorical	(1) 14-24	1061	31.78
			(2) 25-35	1958	58.65
			(3) 36-46	313	9.38
			(4) 47-57	6	0.18
			(5) 58-62	1	0.03
					0.03
MoEducate	Mother's education	Categorical	(1) illiterate	199	5.96
			(2) elementary	769	23.04
			(3) middle school	803	24.05
			(4) High school	1234	36.96
			(5) academic	334	10.01
JobNumber	Mother's job	Categorical	(1) housewife	3065	91.8
			(2) employee	191	5.73
			(3) farmer	6	0.18
			(4) rancher	3	0.09
			(5) carpet waver	42	1.26
			(6) others	32	0.96
	Mother's blood type	Categorical		670	20.07
					8.33

MoBloodGroup			(1) A+	278	20.01
			(2) A-	668	6.11
			(3) B+	204	10.58
			(4) B-	353	29.44
			(5) AB+	983	3
			(6) O+	100	2.46
			(7) O-	82	
			(8) AB-		
CityNumber	Name of city	Categorical	(1)Fars	123	3.69
			(2)Golestan	73	2.19
			(3)Kermanshah	982	29.42
			(4)Hormozgan	17	0.51
			(5)Lorestan	173	5.19
			(6)Kohgiluyeh and Boyerahmad	233	6.98
			(7)Mashhad	403	12.07
			(8)Birjand	213	6.38
			(9)Gorgan	78	2.34
			(10) Yazd	775	23.22
			(11) Zahedan	32	0.96
			(12) Qeshm	238	7.13
TypeDelivery	Type of delivery	Categorical	(1)Vaginal birth	2264	67.81
			(2)Cesarean	1075	32.2
			(3)section		
			Water birth	1	0.03
NumberLiveBirth	The number of live birth	Categorical	0	249	7.46
			1	1307	39.15
			2	1058	31.69
			3	456	13.66
			4	158	4.74
			5	70	2.1
			6	41	1.23
NumberDeathNeonate	The number of stillbirth	Binary	0		
			1	2908	87.1
NumberAoration	The number of abortions	Categorical		432	12.94
			0	2916	87.34
			1	313	9.38
			2	80	2.4
			3	30	0.9

EthnicNumber	Mother's ethnicity	Categorical	(1)Fars	2031	60.83
			(2)Lor	234	7.01
			(3)Turkish	592	17.73
			(4)Kurdish	70	2.1
			(5)Arab	41	1.23
			(6)Others	17	0.51
			(7)Lak	208	6.23
			(8)Afghan	12	0.36
			(9)Turkmen	113	3.39
			(10) Bandari	21	0.63
familymarriage	cousin marriage	Binary	Yes	1011	30.28
			No	2328	69.73
Neonateweight	Baby weight	Categorical	(1)Less than 2500	1125	33.7
			(2)Between 2500 and 4000	1790	53.61
			(3)More than 4000	52	1.56
			(4)Unknown	372	11.15
taneshlife	Is there stress in the workplace if the mother is working?	Categorical	(1)Always calm	725	21.72
			(2)Relatively calm	1735	51.97
			(3)Sometimes stressful	755	22.62
			(4)Always stressful	124	3.72
Contraceptionbeforepregnancy	Whether or not mother has contraceptive methods	Binary	Yes	1521	45.56
			No	1818	54.45
	Whether or not mother use any tobacco during pregnancy?	Binary	Yes	2834	84.88
			No	506	15.16
Smokingparents	Did anyone smoke at work or people around her?	Binary	Yes	900	26.96
			No	2440	73.08

Smokingduringpregnancy	What was the result of the mother's previous Pregnancy?	Binary	(1)Live birth (2)Death birth	1655 1685	49.57 50.47
------------------------	---	--------	---------------------------------	--------------	----------------

Predictive performance of six models for the validation cohort is shown by the ROC curves in Figure 3. SVM model achieved the highest AUC of 0.94.03The AUC of Naive Bayes, Logistic Regression, Deep Learning, Decision Tree and the Gradient Boosted Trees model was 88.26% 92.44%, 93.46%, 78.12%, 90.40%, respectively (Figure 3).

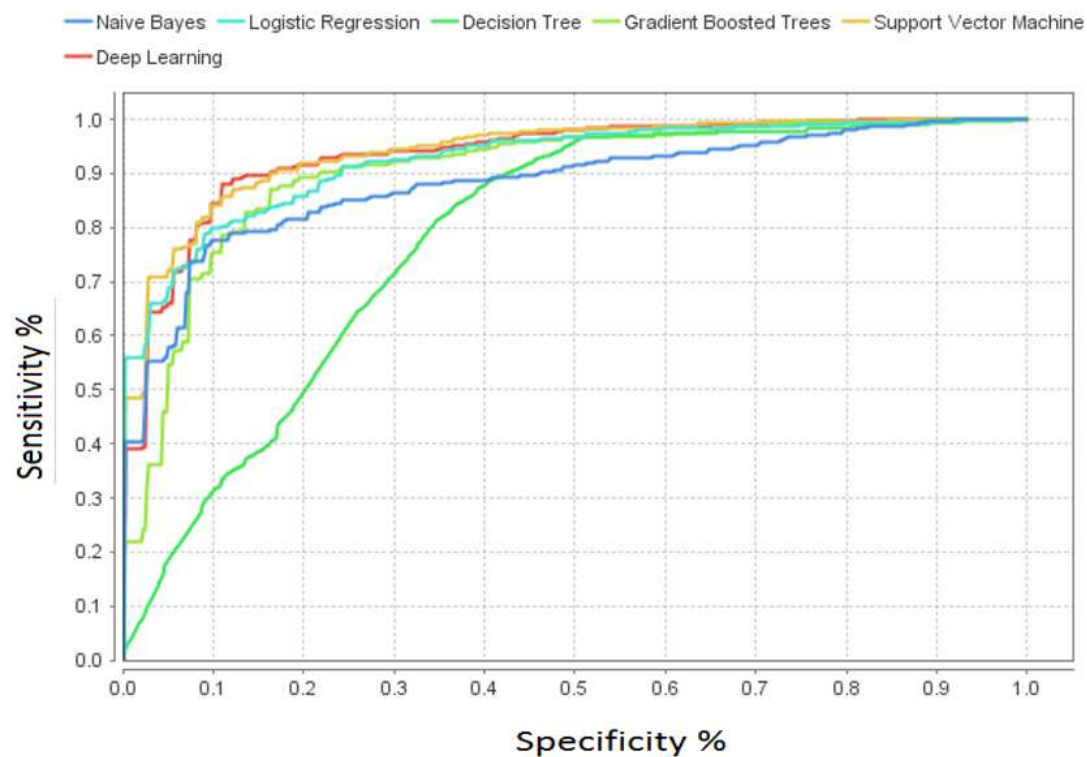


Figure 3. Receiver operating characteristic (ROC) curve of predictive models. logistic regression; SVM, Naive Bayes; Logistic Regression; Deep Learning; Decision Tree; Gradient Boosted Trees.

SVM had highest sensitivity in all the models. svm had the best Time, Deep Learning had the best recall and sensitivity, Decision Tree had the best specificity and Logistic Regression had the best accuracy. The ten most important variables in the models included Number abortion, Number live birth, Number Death Birth, Tanesh Life, Mo Age, Smoking during Pregnancy, Mo Educate, Parity Pregnancy, and Neonate Weight (Figure 4). At this stage, considering that the logistic regression model has the highest accuracy (87.42%) and the decision tree which performs better in terms of prediction than the other models, it was considered that, they are the most important model in Table 3.

Table 3 Comparison of each model for the validation cohort

Model	accuracy	AUC	precision	recall	f_measure	sensitivity	specificity	Time (ms)
Naive Bayes	82.71%	88.26%	92.28%	85.34%	88.67%	85.34%	72.66%	114950.0
Logistic Regression	87.42%	92.44%	89.15%	95.92%	92.39%	95.92%	53.79%	74768.0
Deep Learning	86.06%	93.46%	85.54%	99.08%	91.81%	99.08%	37.33%	170658.0
Decision Tree	87.00%	78.12%	88.20%	96.24%	92.03%	96.24%	54.03%	72816.0
Gradient Boosted Trees	84.49%	90.40%	84.95%	97.61%	90.83%	97.61%	35.49%	704442.0
Support Vector Machine	86.79%	94.03%	86.74%	98.28%	92.15%	98.28%	43.67%	1282276.0

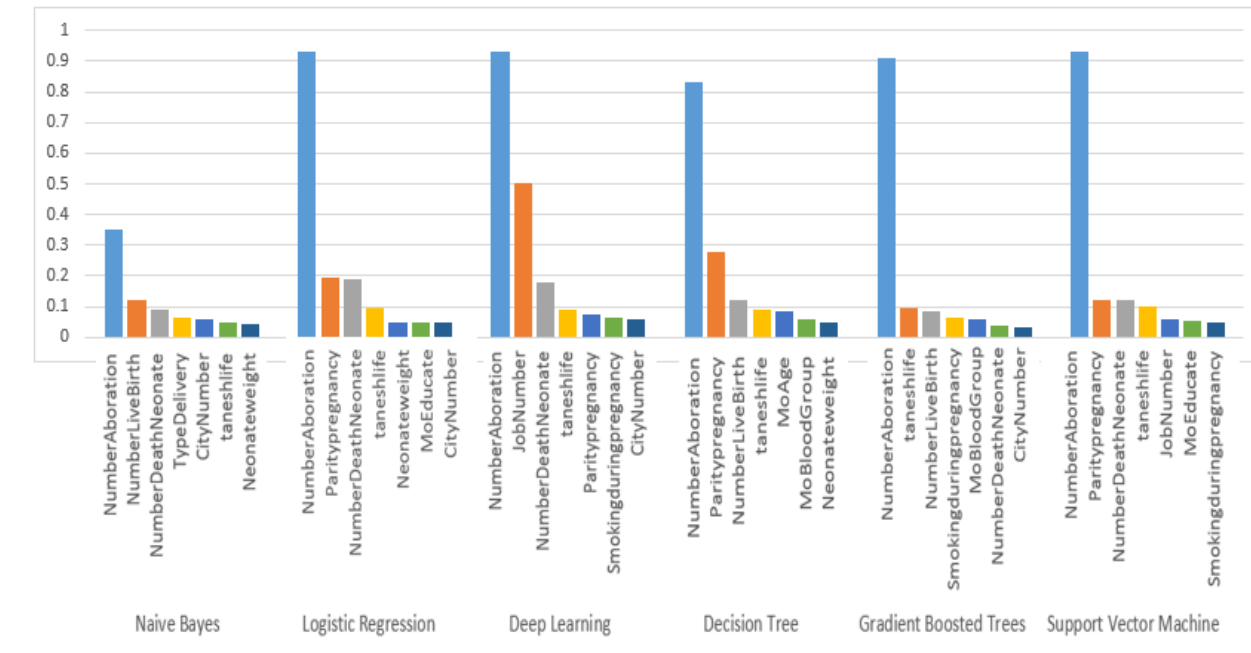


Figure 4. The most important factors in all of the models.

Predictive variables selected for integration are selected from Table 4 which include (Number abortion, Mo Blood Group , Number Live Birth, Parity pregnancy, tanesh life, Mo Age ,Neonate weight, Smoking during pregnancy, Mo Educate, Number Death Neonate, Job Number, City Number, and Type Delivery. These predictor variables are used as inputs for neural network models and decision vector machines. Bottom to the accuracy and predictive modelling and evaluation of the final report.

Table 4 an overview of the predictions and the number of duplicates in the models used in the study

Model	Naive Bayes	Logistic Regression	Deep Learning	Decision Tree	Gradient Boosted Trees	Support Vector Machine	number of repetitions
NumberAboration	*	*	*	*	*	*	6
MoBloodGroup				*	*		2
NumberLiveBirth	*			*	*		3
Paritypregnancy		*	*	*		*	4
taneshlife	*	*	*	*	*	*	6
MoAge				*			1

Neonateweight	*	*		*				3
Smokingduringpregnancy				*		*	*	3
MoEducate		*					*	2
NumberDeathNeonate	*	*	*			*	*	5
Contraceptionbeforepregnancy								0
EthnicNumber								0
JobNumber			*				*	2
CityNumber	*	*	*			*		4
familymarriage								0
Smokingparents								0
TypeDelivery	*							1

Table 5 shows the accuracy of the predictor models in the information integration process. As it can be seen, the decision vector is the most precise inverse of the pre-integration information, although the accuracy is about 5% lower than before. The findings and the neural network are reduced by about 7% in terms of accuracy than the vector machine model.

Table 5 Accuracy of the models used in research after information integration

Model	accuracy	AUC	precision	recall	f_measure	sensitivity	specificity
Logistic Regression	80.71%	91.36%	80.53%	99.60%	89.05%	99.60%	10.08%
Decision Tree	82.81 %	70.77%	83.87%	96.81%	89.86%	96.81%	30.71%

5. Discussion

This study is a case-control one. The data collection tool in this study is a checklist that was collected by epidemiology researchers. This study examines the possibility of maternal stillbirth with regard to previous maternal pregnancies with 2 possible occurrences (live birth, death birth). This population-based case-control study was conducted in 10 provinces of the country and it included the following algorithms: Naive Bayes, Logistic Regression, Deep Learning, Decision Tree, Gradient Boosted Trees, Support Vector Machine method were used to predict neonatal birth. Data analysis was performed by data mining using Rapid Miner software and SVM algorithm in MATLAB. These data include general variables of maternal age, occupation, education, etc., as well as variables such as maternal life stress and maternal smoking. The present study supports the hypothesis that Factors affecting stillbirth in some algorithms are somewhat identical and in others different and that different preventive and treatment strategies might be required. The overall conclusion of the present study shows that Logistic Regression models and Decision Tree are the most suitable models for prediction (classification) Stillbirth mothers (among the models investigated in this study).

6. Conclusion

This study supports the hypothesis that Factors affecting stillbirth in some algorithms are somewhat identical and in others different and that different preventive and treatment strategies might be required. The overall conclusion of the present study shows that Logistic Regression models and Decision Tree are the most suitable models for prediction (classification) Stillbirth mothers (among the models investigated in this study). On the most important predictors, variables Number Abortion, Parity pregnancy, gestational age, Number Death Neonate, City Number are presented as the most important predictors in this study.

Declarations:

- Funding:

The author(s) received no financial support for the research, authorship, and/or publication of this article.

- Conflicts of interest/Competing interests:

I hereby disclose all of my conflicts of interest and other potentially conflicting interests.

- ✓ *All authors have participated in (a) conception and design, or analysis and interpretation of the data; (b) drafting the article or revising it critically for important intellectual content; and (c) approval of the final version.*
- ✓ *This manuscript has not been submitted to, nor is under review at, another journal or other publishing venue.*
- ✓ *The authors have no affiliation with any organization with a direct or indirect financial interest in the subject matter discussed in the manuscript.*
- ✓ *The following authors have affiliations with organizations with direct or indirect financial interest in the subject matter discussed in the manuscript.*

-Authors' contributions:

Nazanin Rahmati, Hadiseh Kamalgharibi, Mohammad Mehdi Hosseini, Alireza Jalai contributed to the design and implementation of the research, to the analysis of the results, and to the writing of the manuscript. Rahmati and Kamalgharibi gathering of the data and preprocessing them, Rahmati and Dr.Hosseini implemented an algorithm, analyzed the result, and writing the basic manuscript. Also, Dr.Jalali has participated in this project for writing, guidance, and feedback throughout this project. Finally, all authors approve of the final version.

- Availability of data and material:

Due to the nature of this research, participants of this study did not agree for their data to be shared publicly, so the availability of data and material is not available. but The datasets used and/or analyzed during the current study are available from the first author(Nazanin Rahmani) on reasonable request.

- Code availability

Due to the nature of this research, participants of this study did not agree for their data to be shared publicly, so code of data is not available.

Ethical Approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and Informed consent was obtained from all individual participants involved in the study. Due to the nature of this research which did not have a clinical trial and most of the work was done in the form of a questionnaire and summary of statistics and analysis of results, the ethics committee approval is not required for such studies. so this study has been granted an exemption from requiring ethics approval by Shahrood Branch, Islamic Azad University ethical committee. (Committee ID: IR.IAU.SHAHROOD.REC).

Written informed consent was obtained from all subjects before the study. for subjects are under 18, from a parent and/or legal guardian Written informed consent was obtained from legally authorized representatives before the study.

Acknowledgement

We would like to express my special thanks of gratitude to all pregnant mothers, which helped us in doing a lot of Research and to led to know us about so many new things. we are really thankful to them. Secondly, we would also like to thank our friends in Bahar Hospital who helped us a lot in finishing this project within the limited time.

References

- [1] Akolekar R, Bower S, Flack N, Bilardo CM, Nicolaides KH (2011) Prediction of miscarriage and stillbirth at 11–13 weeks and the contribution of chorionic villus sampling Prenatal diagnosis 31:38-45
- [2] Aminu M, Unkels R, Mdegela M, Utz B, Adaji S, Van Den Broek N (2014) Causes of and factors associated with stillbirth in low-and middle-income countries: a systematic literature review BJOG: An International Journal of Obstetrics & Gynaecology 121:141-153
- [3] Rasmussen S, Irgens LM, Skjærven R, Melve KK (2009) Prior adverse pregnancy outcome and the risk of stillbirth Obstetrics & Gynecology 114:1259-1270
- [4] Hajian-Tilaki K, Esmailzadeh S, Sadeghian G (2014) Trend of stillbirth rates and the associated risk factors in Babol, Northern Iran Oman medical journal 29:18
- [5] Ashish K, Nelin V, Wrammert J, Ewald U, Vitrakoti R, Baral GN, Målqvist M (2015) Risk factors for antepartum stillbirth: a case-control study in Nepal BMC pregnancy and childbirth 15:146
- [6] Conde-Agudelo A, Belizán JM (2000) Maternal morbidity and mortality associated with interpregnancy interval: cross sectional study Bmj 321:1255-1259
- [7] Bittles AH, Black ML (2010) The impact of consanguinity on neonatal and infant health Early human development 86:737-741
- [8] Mohammadi MM, Hooman HA, Afrooz GA, Daramadi PS (2012) The relationship between consanguineous marriage and death in fetus and infants Journal of research in medical sciences: the official journal of Isfahan University of Medical Sciences 17:448
- [9] Gardosi J, Madurasinghe V, Williams M, Malik A, Francis A (2013) Maternal and fetal risk factors for stillbirth: population based study Bmj 346:f108
- [10] Hyland A, Piazza KM, Hovey KM, Ockene JK, Andrews CA, Rivard C, Wactawski-Wende J (2015) Associations of lifetime active and passive smoking with spontaneous abortion, stillbirth and tubal ectopic pregnancy: a cross-sectional analysis of historical data from the Women's Health Initiative Tobacco control 24:328-335.
- [11] Zarei R, Aghaie N (2010) Assessing of the frequency of interauterine fetal death and related factors in Ahvaz Imam Khomeini hospital SCIENTIFIC MEDICAL JOURNAL 8:43-437.
- [12] Waldenström U, Cnattingius S, Norman M, Schytt E (2015) Advanced maternal age and stillbirth risk in nulliparous and parous women Obstetrics & Gynecology 126:355-362
- [13] Sharma PP, Salihu HM, Kirby RS (2007) Stillbirth recurrence in a population of relatively low-risk mothers Paediatric and perinatal epidemiology 21:24-30
- [14] Maghsoudlou S et al. (2015) Consanguineous marriage, prepregnancy maternal characteristics and stillbirth risk: a population-based case–control study Acta obstetrica et gynecologica Scandinavica 94:1095-1101
- [15] Arabameri, A., Pradhan, B., Pourghasemi, H. R., Rezaei, K., & Kerle, N. (2018). Spatial modelling of gully erosion using GIS and R programing: A comparison among three data mining algorithms. Applied sciences, 8(8), 1369.
- [16] Hussain, S., Dahan, N. A., Ba-Alwib, F. M., & Ribata, N. (2018). Educational data mining and analysis of students' academic performance using WEKA. Indonesian Journal of Electrical Engineering and Computer Science, 9(2), 447-459.
- [17] Dreiseitl S, Ohno-Machado L (2002) Logistic regression and artificial neural network classification models: a methodology review Journal of biomedical informatics 35:352-359.

- [18] Khalil A, Syngelaki A, Maiz N, Zinevich Y, Nicolaides KH (2013) Maternal age and adverse pregnancy outcome: a cohort study *Ultrasound in Obstetrics & Gynecology* 42:634-643
- [19] Pham, B. T., Prakash, I., & Bui, D. T. (2018). Spatial prediction of landslides using a hybrid machine learning approach based on random subspace and classification and regression trees. *Geomorphology*, 303, 256-270.
- [20] Hornik K, Stinchcombe M, White H (1990) Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks *Neural networks* 3:551-560
- [21] Michie D, Spiegelhalter DJ, Taylor C (1994) Machine learning *Neural and Statistical Classification* 13:1-298
- [22] Lucas PJ, Van der Gaag LC, Abu-Hanna A (2004) Bayesian networks in biomedicine and health-care *Artificial intelligence in medicine* 30:201-214.

Table 6 the checklist that given to the mothers

Dear mother

The present questionnaire was designed to identify the factors associated with stillbirth. Your cooperation in providing accurate information can be an effective step in better understanding the risk factors for neonatal mortality and ways to prevent it. Be sure that your information will remain confidential to the group. Thank you for helping us with this research.

1. name of city 2.name of city/village 3.name of health center

Questionnaire type: 1.Stillbirth 2. Neonatal mortality 3. Live birth	
1.How old was the mother at birth?	2. How old was the father when the baby was born?
3. accommodations: 1) city 2) village	4. Race: 1) Fars 2) Lurs 3) Turkish 4) Kurdish 5) Arab 6) Other (mentioned)
5. Mother's level of education: 1) Illiteracy 2) Elementary 3) Middle School 4) High school 5) Academic	
6. Father's level of education: 1) Illiteracy 2) Elementary 3) Middle School 4) High school 5) Academic	
7. Mother's job: 1) Housewife 2) Clerk 3) Farmer 4) Rancher 5) Carpet waver 6) Other:.....	
8. Do you have a night shift work experience? 1) Yes 2) No if yes, how many years?..... year(s)	
9. Father's job: 1) Self-employment 2) Clerk 3) Farmer 4) Rancher 5) Other:.....	
10. Do you have a close family relationship (cousin, etc.) with your husband? 1) Yes 2) No	
11. How did the mother give birth? 1) Natural childbirth 2) Caesarean 3) Water birth	
12. If the previous question is 2, what is the cause of cesarean section? 1) Selective cesarean section 2) History of cesarean section 3) Failure to progress delivery 4) Fetal problems or abnormal fetal display 5) Other (mentioned)	
13. How many pregnancies has this mother had?	
14. The number of deliveries should be mentioned separately: -The number of live births..... -The number of stillbirths (over 20 weeks pregnant) -The number of abortions (under 20 weeks of pregnancy)	
15. Has this pregnancy been twin or more? 1) Yes 2) No	16) What was the sex of the newborn? 1) Female 2)Male 3) Unknown
17. What was the weight of the newborn? 1) Less than 2500 2) Between 2500 and 4000g 3) More than 4000g 4) Unknown	
18. What was the exact age of the newborn? weeks of pregnancy Don't know	
19. How far has this pregnancy been from her previous pregnancy? 1) Less than a year 2) 1-3 years 3) More than 3 years 4) First pregnancy	
20. What was the result of the mother's previous pregnancy? 1) Live birth 2) Stillbirth 3) Abortion 4) This was the first pregnancy of mother 5) Other (ectopic pregnancy, etc.) should be mentioned	
21. Was the pregnancy wanted by the couple or unintentionally? 1) Both wanted 2) Only the man 3) only the woman 4) Both unwanted	
22. If the above question is 1, was the pregnancy concurrent with the use of contraceptives? 1) Yes (mentioned) 2) No	
23. Have you used contraceptive methods before this pregnancy? 1) Yes 2) No	

24. If yes, which methods..... 1) Pill 2) IUD 3) Injectable ampoules 4) Condom 5) Other	
25. Have your period usually been regular? 1) Yes 2) No	
26. What is the average interval between the start of your two periods? days	27. How long does your period usually last?days
28. Has the mother permanently taken any supplements during this pregnancy? 1) Yes, regularly 2) Yes, irregularly 3) No	
29. Has the mother used anti-nausea medicine during the first 3 months of pregnancy? 1) Yes 2) No If yes, what is the medicine?	
30. During this pregnancy, have the mother had a specific illness that led to the use of medication or surgery? 1) Yes (name of surgery or severe illness) 2) No	
31. During a week, how many meals do you usually include meat (beef or veal) chicken, other birds and fish:	
32. Do you usually remove tallow before cooking the chicken? 1) Yes 2) No	
33. How many times do you consume dairy products? 1) Daily 2) 1-2 times a week 3) once every two weeks 4) 1-2 times a month 5) I do not consume at all	
34. Do you eat more special foods (for example, rice, pasta, kebab), please name? How many times do you eat? 1) Daily 2) 1-2 times a week 3) once every two weeks 4) 1-2 times a month 5) I do not eat at all	
35. How many times do you eat cake or sweetmeat? 1) Daily 2) 1-2 times a week 3) once every two weeks 4) 1-2 times a month 5) I do not eat at all	
36. How many times do you eat pickles (pickles, lemon juice or verjuice)? 1) Daily 2) 1-2 times a week 3) once every two weeks 4) 1-2 times a month 5) I do not eat at all	
37. How many times do you eat fruits, vegetables (dried and fresh)? 1) Daily 2) 1-2 times a week 3) once every two weeks 4) 1-2 times a month 5) I do not eat at all	
38. How many minutes do you normally do physical activity per day besides housework? (Walking, public sports, etc.)? minutes	
39. How is your sleeping place? 1) Dark 2) Semi-dark 3) Bright	
40. How did you sleep during pregnancy? 1) Lying on your side 2) Lying on your back	
41. How do you think your sleep was during pregnancy in general? 1) Very good 2) Good 3) Bad 4) Very bad	
42. Did you go to bed at night regularly during pregnancy? 1) Always 2) Often 3) Sometimes 4) Seldom 5) Never	
43. How many hours did you usually sleep during a day (including the afternoon hours) during pregnancy? hours	
44. In general, my past family life was 1) Always calm 2) Relatively calm 3) Sometimes intense 4) Often intense 5) Always intense	
45. The worst event in your life (mention 1): 1) Related to the loss of relatives 2) Related to economic or family factors 3) Related to job factors 4) Related to the socio-political factor 5) Related to your health or family health	
46. If you have a job, how is your work environment always? 1) Always calm 2) Relatively calm 3) Sometimes intense 4) Often intense 5) Always intense	
47. If you experience any of the following during your pregnancy, mark: Has your husband ever hurled or kicked you? 1) Yes 2) No Has your husband ever hit you by a fist, a knife or other device? 1) Yes 2) No Has your husband ever hurt you so you don't need medical attention? 1) Yes 2) No Has your husband ever hurt you in such a way that you don't go to school or work? 1) Yes 2) No Has your husband ever deprived you of sleep? 1) Yes 2) No	

48. Where did you live during the war? (mention the name of city)					
49. If you were born after 1367, which city did your parents live in?					
50. Has the mother used any tobacco (cigarettes, hookah, etc.) during pregnancy? 1) Yes 2) No					
51. If yes, how long? 1) Less than a year 2) Between 1 and 5 years 3) More than 5 years					
52. Have your parents smoked during your childhood? 1) Yes 2) No					
53. Has anyone smoked at your place of residence or work? 1) Yes 2) No If yes, how many years? Year(s)					
The esteemed questioner should refer to the place where the pregnancy care was done (more) and get the appropriate answer to the following questions by reading the file.					
54. Weight before pregnancy: kg		55. Mother's weight at the last prenatal visit: kg		56. Mother's height: m	
57. Which of the following diseases did the mother have before giving birth to the newborn? 1) Hepatitis B 2) Hepatitis C 3) Epilepsy 4) Heart disease 5) Bile 6) Type 2 diabetes 7) kidney Diseases 8) Digestive diseases 9) Blood pressure 10) Cancer 11) Hypothyroidism 12) Others					
58. Have oral and dental examinations been performed for the mother during pregnancy care if necessary? 1) Yes 2) No					
59. Does the mother have any of the following complications during pregnancy? 1) Yes 2) No					
	Complication	Reference		Complication	Reference
	1-Bleeding - spotting			2- Dysuria	
	3- Water breaking			4- Pain in the abdomen or side	
	5- Unilateral leg and thigh pain			6- Toothache	
	7- Severe or blood vomiting			8- purulent, smelly vaginal discharge	
	9- Decreased or no fetal movement			10- skin itching	
	11- psychological symptoms			12-Domestic violence	
	13- Hit			14- A common complaint	
	15- Headaches – blurred vision			16- Increased blood pressure	
	17- Ague			18- Skin rashes	
	19- Swelling of the hands and face				
60. Mother's blood type: 1) A+ 2) A- 3) B+ 4) B- 5) AB+ 6) O+ 7) O- 8) AB-					
61. What was the amount of hemoglobin in the mother's blood during the first test? 1) less than 7g/dl 2) between 7 and 11g/dl 3) between 11 and 15g/dl 4) more than 15g/dl 5) It's not done					
62. What was the amount of hemoglobin in the mother's blood during the second test? 1) less than 7g/dl 2) between 7 and 11g/dl 3) between 11 and 15g/dl 4) more than 15g/dl 5) Not done					

63. If there has been a positive urine culture on the mother's test results, was the mother been referred? 1) Yes with reference 2) Yes without reference 3) The test was negative 4) Not done
64. If the GCT and GTT mother's tests results were abnormal, was the mother been referred? 1) Yes with reference 2) Yes without reference 3) The test was negative 4) Not done
65. If the FBS mother's test results were abnormal, was the mother referred? 1) Yes with reference 2) Yes without reference 3) The test was negative 4) Not done
66. Has the mother had type 2 diabetes during pregnancy? 1) Yes 2) No
67. If the above answer is 1, has the mother taken insulin? 1) Yes 2) No If she has taken insulin, how many units has been taken? unit
68. Does the mother have thalassemia minor? 1) Yes 2) No

Figures

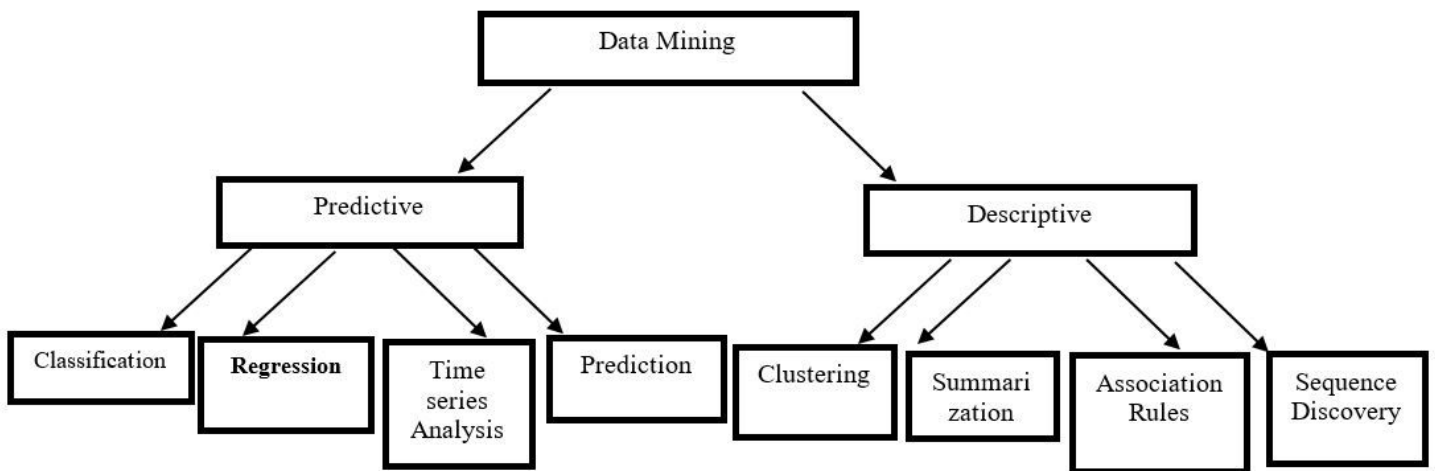


Figure 1

Data Mining Methods and Applications

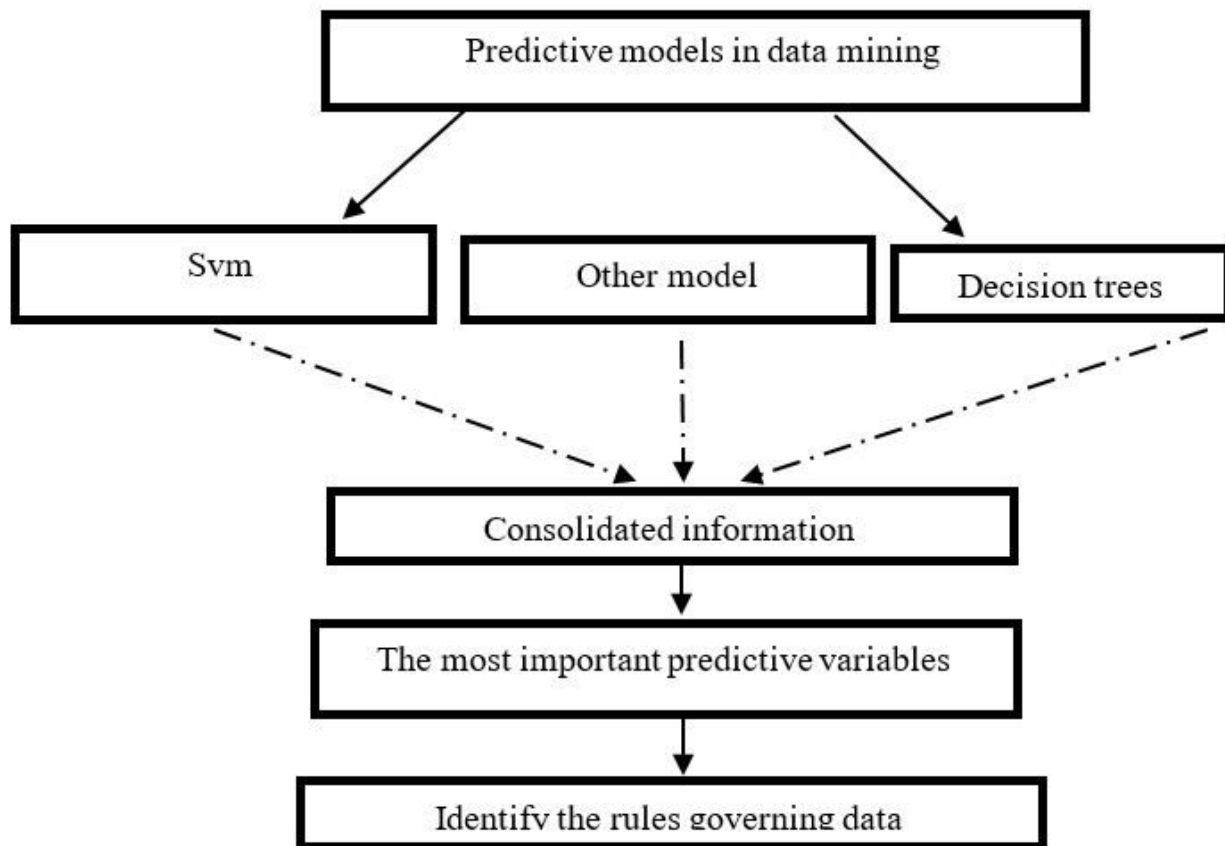


Figure 2

Overview of integrated model generation with data mining techniques

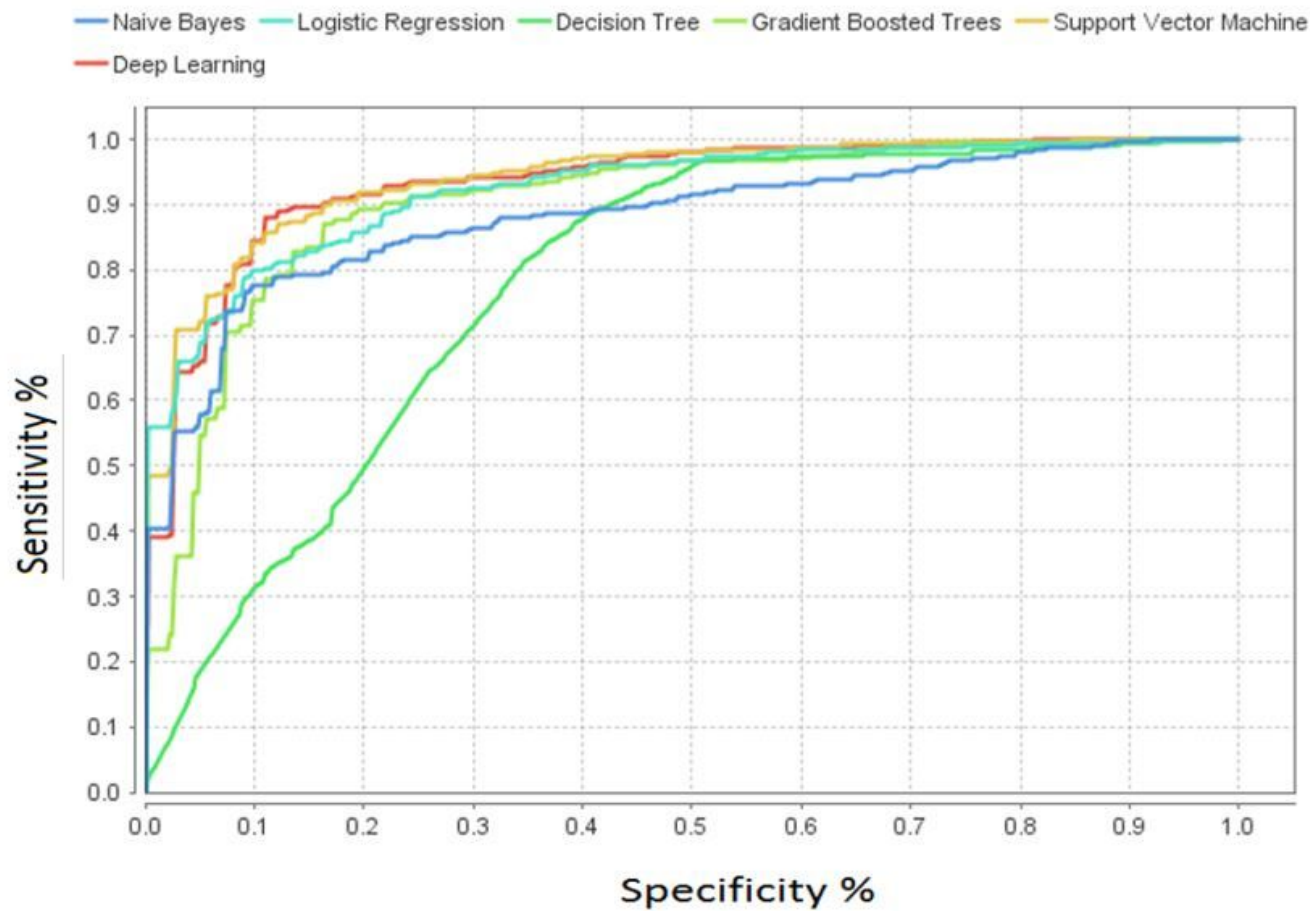


Figure 3

Receiver operating characteristic (ROC) curve of predictive models. logistic regression; SVM, Naive Bayes; Logistic Regression; Deep Learning; Decision Tree; Gradient Boosted Trees.

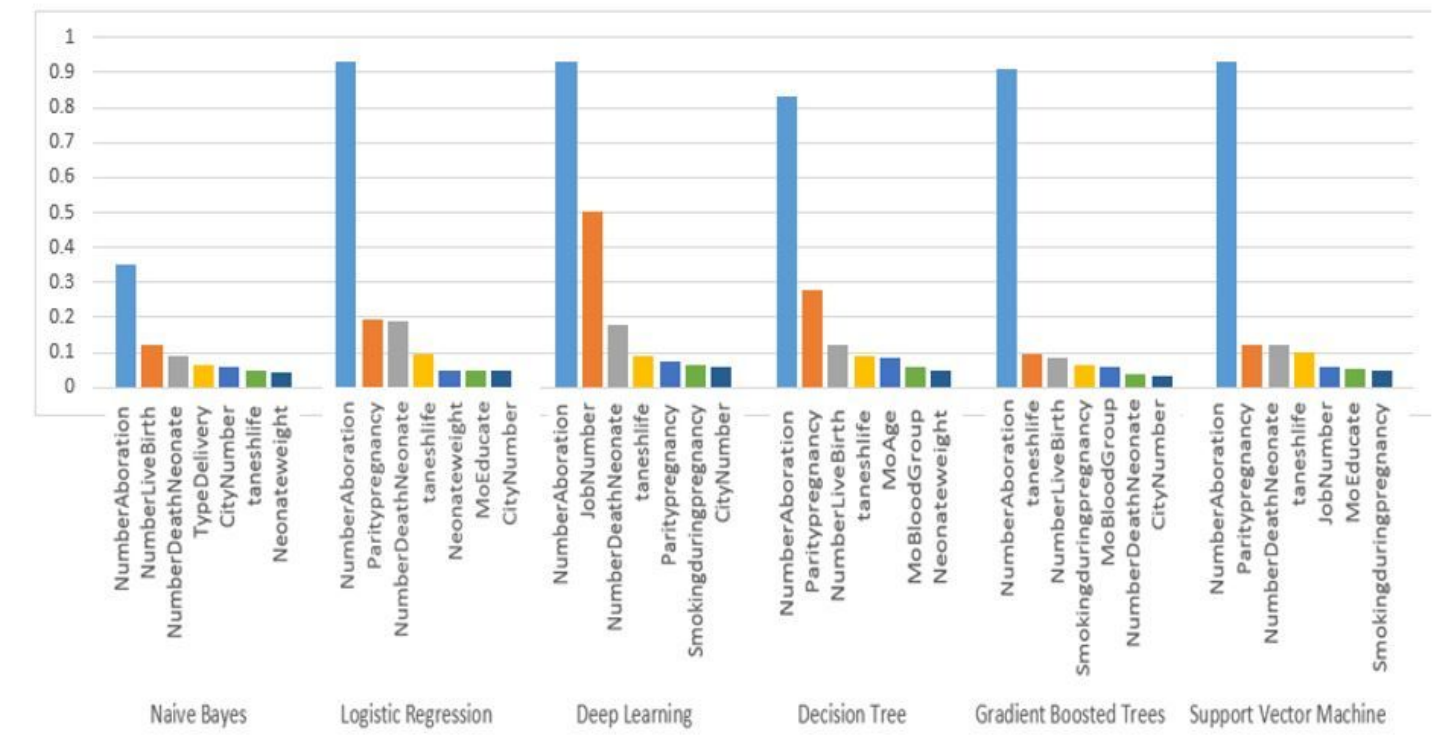


Figure 4

The most important factors in all of the models.