

PREDICTING THE DELIVERY TYPE OF PREGNANCY: A COMPARATIVE STUDY

E.G.K.M. Gamlath¹, S. Abeysundara² & S.S.K.N. Rajaguru³

Correspondence: seu.gamlath@gmail.com

ABSTRACT

One of the most beautiful things in the world is the birth of a child. The advancement of technology gives a mother several options to deliver a child to the world. However, unexpected incidences at the time of delivery may cause her to change the delivery type of the baby. Therefore, it would be essential and advantageous if there is a method that can predict the best way to make this happen based on the information collected from the mother and the baby to be born. This study intends to examine the best model to predict the delivery type. This study was conducted based on the medical records available from January 2015 to December 2015 in the General Hospital at Ampara, Sri Lanka. Births of 1400 babies from the mentioned metropolis are analysed in this study. There are two popular delivery types of birth, namely, Vaginal and Cesarean. The maternal factors such as ethnicity, age, the number of pregnancies, the number of babies along with the infant's gender, weight, head circumference, height and shoulder-length were recorded. Several classification models, namely, logistic regression, decision trees, support vector machines and a naive-Bayes classifier, were used to find the best-suited model to predict the delivery type. Moreover, the most significant factors that affect the type of birth were identified. Results indicated that the maternal age, delivery time, infant's weight and infant shoulder-length have a statistically significant association with the type of delivery. The logistic regression model was obtained by getting delivery type as the dependent variable and the model was in a better fit with 68.21%. In the decision tree, the accuracy of the model was 71.43 %. Also, the most significant factor was the delivery time in the decision tree. In Naïve Bayes Classifier, the accuracy was 67.68% and Support Vector Machines outcome suggested that the accuracy was 79.11%. Study outcomes suggest that Support Vector Machines can be used to predict the delivery type in higher accuracy.

Keywords: Logistic regression model Decision Tree, Support Vector Machine and Naiver Bayes Classifier

CHAPTER 1 - INTRODUCTION

1.1 Introduction

The life of a baby begins long before he or she is born. A new individual human being begins at fertilization, when the sperm and ovum meet to form a single cell. Universe is an enigma waiting to be unfolded and deciphered by man because the human sentient ability and the consciousness is essential to decipher the mysteries of this Universe. So, every human being is so important to world and if a baby dies at birth then it is non-compensable as each person has different capabilities, everybody is important to the world and being a human being is something rare to be happen. So, there is a vital necessity to think deliberately about childcare.

There are methods of child deliveries at present. Namely, vaginal delivery, cesarean section (C-Section), vaginal birth after cesarean, vacuum extraction,

¹ Department of Mathematical Sciences, South Eastern University of Sri Lanka.

² Department of Statistics and Computer Science, University of Peradeniya, Sri Lanka.

³ Department of Statistics and Computer Science, University of Peradeniya, Sri Lanka.

forceps delivery. In a vaginal birth, the baby is born through the birth canal though it is difficult to predict when exactly you will go into labor, but most women give birth at around 38-41 weeks of pregnancy. A cesarean section or C-section is the delivery of a baby through a surgical incision in the mother's abdomen and uterus. In the past, a C-section ended any hope of future vaginal deliveries. But today, largely to changes in surgical technique, vaginal birth after cesarean (VBAC) is possible in many cases. VBAC isn't right for everyone, though. Sometimes a pregnancy complication or underlying condition prevents the possibility of a successful VBAC. A vacuum extraction is a procedure sometimes done during the course of vaginal childbirth. During vacuum extraction, a health care provider applies the vacuum (a soft or rigid cup with a handle and a vacuum pump) to the baby's head to help guide the baby out of the birth canal. A forceps delivery is a type of operative vaginal delivery. It's sometimes needed in the course of vaginal childbirth. In a forceps delivery, a health care provider applies forceps (an instrument shaped like a pair of large spoons or salad tongs) to the baby's head to help guide the baby out of the birth canal.

Pros and Cons are in all types of deliveries. Vaginal Delivery is the more natural way to give birth. Mother's body is naturally equipped to give birth vaginally without medical intervention. Labor starts with your cervix dilating, and it ends with a newborn baby. Women have a sense of empowerment and accomplishment after vaginal birth. They are active participants in the childbirth experience. They must push to help move their baby through the birth canal and into the world. Shorter hospital stay is possible when the birth is given as a vaginal delivery. As cons though most vaginal births are uncomplicated, unforeseen complications can occur during labor and delivery, including maternal hemorrhaging (bleeding) and mother is at risk for perineum tearing from a vaginal delivery. This can range from mild tears to fourth-degree lacerations that tear into your rectum. This can add to your healing time.

In a planned cesarean section (c-section) may be more convenient for women. Because the baby's delivery date is usually scheduled ahead of time, mom may have less stress and anxiety about labor. Women may feel more in control, because they know when their baby will be born and they can better plan for work leave, their baby's nursery, etc. Mother can avoid post term pregnancies with a planned c-section. Most c-sections are typically scheduled between 39 or 40 weeks of gestation. When compared to a vaginal delivery and an unplanned c-section, scheduled cesareans have a reduced risk of postpartum hemorrhage. As cons, s c-section is a major abdominal surgery that comes with surgical risks and complications from anesthesia. Anesthesia side effects may include severe headache, nausea, and vomiting. Anesthesia may also affect the baby, causing him or her to be sluggish or inactive when

born. Women with planned cesarean sections have longer hospital stays and a longer postpartum recovery period than women with vaginal deliveries.

To date, limited research work has been done to determine the factors of mother and infant that affect the type of delivery. Maternal factors of Ethnicity, Age, Number of Pregnancies and number of babies and the infant factors of Gender, Weight, Head Circumference, Height and Shoulder Length may affect for the delivery type.

1.2 Motivation

The purpose of this study is to figure out which factors have significant effect for the delivery type and to find the best statistical model from Logistic regression, Decision Tree, Support Vector Machine and Naiver Bayes Classifier to predict the delivery type.

1.3 Research Objectives

In this research the main objective is to figure out the best statistical method to predict the delivery type using the statistical methods of Logistic regression model, Decision Tree, Support Vector Machine and Naiver Bayes Classifier. Examining the characteristics and factors that highly associated with the delivery type and to bring out the awareness about the types of deliveries among the pregnant mothers as well as the whole population are the sub objectives.

CHAPTER 2 - LITERATURE REVIEW

To date, very little research has been done to determine the factors of mother and infant that affect the type of delivery. No single strategy for lowering cesarean rates has proven successful over time. Healthy People 2020 calls for reducing cesarean births by 10% for low-risk prim gravid women.

(Ronsmans, 2012) conducted a research to identify factors driving the rapid increase in caesarean section in China between 1988 and 2008. The strong regional variation in the caesarean section rate have been observed.

A study on vaginal birth after caesarean (VBAC) rate and evaluate the importance of ethnicity, body mass index (BMI), parity, previous vaginal delivery conducted by (AM, 2013). In this study considered labor as a factor that determine the success of a VBAC, and serious maternal and perinatal outcomes were determined. (Landon MB, 2004) carried out a research to find the Maternal and Perinatal Outcomes Associated with a Trial of Labor after Prior Cesarean Delivery and to see the influence of Maternal age at delivery, ethnic group, married, smoking during the pregnancy, body mass index at delivery.

A mixed-methods study on factors that influence the decision of women to have a cesarean delivery has been carried out by (M, 2013). By (Faisal-Cury

A, 2007) carried out a research to estimate the prevalence and risk factors for antenatal anxiety and antenatal depression. Antenatal anxiety and antenatal depression were associated with similar socio-demographic and socio-economic risk factors, (Gamlath E.G.K.M., 2018) carried out a logistic regression approach to study the maternal and infant factors that affect for the cesarean delivery and the normal delivery. (Rasoli, 2019) carried out a research to assess the effects of reviewing written childbirth scenarios on the selection of delivery method. This study has been showed that a combination of childbirth scenarios and training based on theory of planned behaviour could be used to reduce the incidence of unnecessary CS deliveries.

A research named an evidence-based approach to determining route of delivery for twin gestations by (Robinson B.K, 2011) and there they have been figured out that around 50% of twin pregnancies deliver preterm, and major factors with prematurity include respiratory distress syndrome, necrotizing enterocolitis, intraventricular hemorrhage, and sepsis. (González MS, 2015) have done a research to analyse the comparative risks of this anal sphincter injury in relation to the type of intervention in vaginal delivery. They have analyzed the incidence of obstetric anal sphincter injury for each mode of vaginal delivery: spontaneous delivery, vacuum, Thierry spatulas, and forceps. They founded that the type of intervention in a vaginal delivery was a modifiable intrapartum risk factor for obstetric anal sphincter injury and tearing can be occur in any type of delivery, but proportions vary significantly.

CHAPTER 3 - THEORY AND METHODOLOGY

3.1 Background

This study was conducted based on the medical records available over a period from January 2015 to December 2015 by analyzing 1400 babies from the in Ampara hospital. Mandatory data analysis was done using Minitab 18, SPSS 16, Excel and R software. In 2017 E.G.K.M.Gamlath did an logistic regression approach to find the factors affecting the delivery type using SPSS software. However, the accuracy of the model has only been checked using the Hosmer and Lemeshow Test. In this study we do all the analysis with four methods of Logistic Regression, Decision Tree, Naïve Bayes Classifier and Support Vector Machines by dividing the data set into train and test data with the ration 80:20 and finally classify the variables greatly affect to the delivery type and the best model that can be used for the prediction of the delivery type.

3.2 Methodology

The data were collected from the General Hospital Ampara. Maternal Factors: Age, Ethnicity, No_Pregnancies (Number of Pregnancies) and No_Babies (Number of Babies), Infant Factors: Gender, Weight, Height, Head Circumference, Shoulder Length and the type of delivery were taken for 1400 babies from 2015 January to 2015 December.

Sstatistics was obtained using SPSS 16. Association with variables have been found using Minitab 18 at 5% significance level. Logistic Regression, Support Vector Machines, Naiver Bayes Classifier and Decision Tree have been carried out to using R software to find out the best method of to figure out the delivery type and the factors that mostly affect to the type of delivery.

CHAPTER 4 – RESULTS AND DISCUSSION

4.1 Decision Tree

The statistical method decision tree has been used to check the accuracy of the method to find the delivery type and to figure out the most significant factor that affect to the type of delivery. The accuracy of the model for train data was 75.71% and Accuracy for test data it was 71.43 % with P value $6.724e^{-14}$ and 0.02427 respectively. That is misclassification for train data was 0.2429 and for test data it was 0.2857. Biggest split was delivery time. If the delivery time is greater than 270 days then there is a normal delivery, if it is less than 270 days then if the age of the mother is less than 25 then high chance to have a normal delivery. If it is ≥ 25 then it depends on the infant height, if height is ≥ 53 then depend on the number of pregnancies. If number of pregnancies is ≥ 3 then they have normal delivery, if less than 3 then cesarean delivery. Also, if the height is less than 53 then if number of pregnancies less than 4 then cesarean delivery otherwise it if height is less than 50 then cesarean delivery. If height is ≥ 50 then again depend on the weight of the infant. If weight is ≥ 3.1 then have a cesarean delivery and < 3.1 then a normal delivery.

Table 1: Summary of Train and Test Data

| Parameter | Train Data | Test Data |
|------------------------|----------------|------------------|
| Accuracy | 0.7571 | 0.7143 |
| 95% CI | 0.7309, 0.782) | (0.6575, 0.7665) |
| No Information Rate | 0.6545 | 0.6571 |
| P-Value [Acc > NIR] | $6.724e^{-14}$ | 0.02427 |
| Kappa | 0.4311 | 0.3327 |
| Mcnemar's Test P-Value | $3.435e^{-08}$ | 0.03365 |
| Sensitivity | 0.8772 | 0.8370 |
| Specificity | 0.5297 | 0.4792 |
| Pos Pred Value | 0.7794 | 0.7549 |
| Neg Pred Value | 0.6949 | 0.6053 |
| Prevalence | 0.6545 | 0.6571 |
| Detection Rate | 0.5741 | 0.5500 |
| Detection Prevalence | 0.7366 | 0.7286 |
| Balanced Accuracy | 0.7035 | 0.6581 |
| 'Positive' Class | 0 | 0 |

4.2 Logistic regression

The logistic regression method is a statistical method which is used when the data set is categorical. Here, the misclassification for the train data was 0.31875 and for test data it was 0.3178571. Age, delivery time, Number of babies, head circumference and height were statistically significant in 0.05 significance level. Each one-unit change in number of pregnancies will increase the log odds of having a normal delivery by 0.016, and its p-value indicates that it is somewhat significant in determining the delivery type. Similarly, head circumference, shoulder length positively related to have a cesarean delivery.

Delivery time is negatively related to the log odds of having a cesarean delivery (i.e., controlling for the other variables, mothers who has longer delivery time are less likely to have a cesarean delivery). Similarly, number of babies, weight and height also negatively related to the log odds of having a cesarean delivery.

The interpretation of age is different from others, going to Age-1 (20 ≤ age < 25) from age-0 (age < 20) will increase the log odds of having a cesarean delivery by 0.670309. Going from Age-1 (20 ≤ age < 25) to Age-2 (25 ≤ age < 30) will increase it by 1.725605. Going from age-2 (25 ≤ age < 30) to Age-3 (30 ≤ age < 35) will increase it by 1.799419. Going from Age-3 (30 ≤ age < 35) to Age-4 (age ≥ 35) will increase it by 2.568833. Similarly, in ethnicity from Sinhalese to Tamil log odds of having a cesarean delivery increases by 0.145853. From Tamil to Muslim log odds of having a cesarean delivery decreases by 0.290084.

The difference between Null deviance and Residual deviance tells us that the model is a good fit. Greater the difference better the model. Null deviance is the value when only the intercept has in the equation with no variables and Residual deviance is the value when taking all the variables into account. It makes sense to consider the model good if that difference is big enough. In our results the difference is 171.4 which is big, so the model has a good fit.

Table 2: Summary of Logistic Regression

| Coefficients | Estimate | Std. Error | z value | Pr(> z) |
|---------------|-----------|------------|---------|--------------|
| Intercept | 12.050880 | 2.558736 | 4.710 | 2.48e-06 *** |
| Ethicity2 | 0.145853 | 0.218002 | 0.669 | 0.503467 |
| Ethicity3 | -0.290084 | 0.430355 | -0.674 | 0.500274 |
| Age1 | 0.670309 | 0.475259 | 1.410 | 0.158420 |
| Age2 | 1.725605 | 0.471978 | 3.656 | 0.000256 *** |
| Age3 | 1.799419 | 0.481265 | 3.739 | 0.000185 *** |
| Age4 | 2.568833 | 0.511427 | 5.023 | 5.09e-07 *** |
| Delivery_Time | -0.041629 | 0.006845 | -6.082 | 1.19e-09 *** |
| No_Pregnancy | 0.016439 | 0.124727 | 0.132 | 0.895141 |
| No_babies | -0.435142 | 0.146638 | -2.967 | 0.003003 ** |
| Gender2 | -0.095748 | 0.137753 | -0.695 | 0.487011 |
| Weight | -0.184592 | 0.263680 | -0.700 | 0.483889 |
| Head | 0.141923 | 0.051639 | 2.748 | 0.005989 ** |
| Height | -0.139630 | 0.036297 | -3.847 | 0.000120 *** |
| Shoulder | 0.026203 | 0.031295 | 0.837 | 0.402436 |

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
 (Dispersion parameter for binomial family taken to be 1)
 Null deviance: 1444.0 on 1119 degrees of freedom
 Residual deviance: 1272.6 on 1105 degrees of freedom
 AIC: 1302.6
 Number of Fisher Scoring iterations: 4

4.3 Naive Bayes Classifier

Naive bayes classifier has been used to check the efficiency of the method for identifying the delivery type. In this case, for the train dataset, the priori probability of a normal delivery was 0.6561489 and a cesarean delivery it was 0.3438511. That is, the frequency of having a normal delivery is 65.6% and cesarean delivery is 34.4%. Conditional probabilities are the likelihood of each variable. The likelihood of Normal birth given Sinhalese that is $P(\text{Normal}/\text{Sinhalese})$ was 0.85696671. Similarly, the other conditional probabilities were,

$$P(\text{Normal}/\text{Tamli}) = 0.11344020, P(\text{Normal}/\text{Muslim}) = 0.02959309,$$

$$P(\text{Cesarean}/\text{Sinhalese}) = 0.85411765, (P(\text{Cesarean}/\text{Tamil}) = 0.12000000,$$

$$(P(\text{Cesarean}/\text{Muslim}) = 0.02588235.$$

Then, the likelihood of Normal and Cesarean deliveries with respect to age is found. When age is less than 20 the probability of having a normal delivery, that is $P(\text{Normal}/\text{age} < 20)$ was 0.05548705. Similarly, the other probabilities have been found for the different age groups. Those are mentioned below. $P(\text{Normal}/20 \leq \text{age} < 25) = 0.31442663$, $P(\text{Normal}/25 \leq \text{age} < 30) = 0.30579531$, $P(\text{Normal}/30 \leq \text{age} < 35) = 0.24290999$, $P(\text{Normal}/\text{age} \geq 35) = 0.08138101$, $P(\text{Cesarean}/\text{age} < 20) = 0.0141176$, $P(\text{Cesarean}/20 \leq \text{age} < 25) = 0.17882353$, $P(\text{Cesarean}/25 \leq \text{age} < 30) = 0.32705882$, $P(\text{Cesarean}/30 \leq \text{age} < 35) = 0.28235294$, $P(\text{Cesarean}/\text{age} \geq 35) = 0.19764706$.

Then, the likelihood of Normal and Cesarean deliveries with respect to gender of the infant is found and mentioned as follows. $P(\text{Normal}/\text{Female}) = 0.4944513$, $P(\text{Normal}/\text{Male}) = 0.5055487$, $P(\text{Cesarean}/\text{Female}) = 0.5223529$, $P(\text{Cesarean}/\text{Male}) = 0.4776471$.

Other variables of Number of Pregnancies, Number of Babies, weight, head circumference, height and shoulder length of the infant are continuous and found mean and standard deviations for each variable as shown in the table.

Finally, the confusion matrix of train data represents, 716 deliveries predicted as normal and 166 deliveries predicted as cesarean and it was true. Also 259 Cesarean deliveries predicted as normal and 95 normal deliveries wrongly predicted as cesarean deliveries. Misclassification in train data is 0.2864078.

In the confusion matrix of test data, 88 deliveries predicted as normal and 23 deliveries predicted as cesarean and it was true. Also 35 Cesarean deliveries predicted as normal and 18 normal deliveries wrongly predicted as cesarean deliveries. Misclassification in test data is 0.3231707.

4.4 Support vector Machines

In support Vector Machines we did the analysis using the kernels of radial, linear, polynomial and sigmoid using train and test data sets. Number of support vectors were 740, 755, 793 and 674 respectively. Misclassification rate for train data were 0.2455357, 0.2991071, 0.3178571 and 0.3705357 respectively. Also, misclassification rate for train data were 0.2455357, 0.2991071, 0.3178571 and 0.3705357 accordingly. Then by using the train data the model was tuned. Sampling method used was 10-fold cross validation and best parameters were epsilon 0 and cost was 4. The better performance was 0.2803571. In the best model number of support vectors were 713 and the classification was radial. Misclassification for train data was 0.2089286 and misclassification for test data was 0.2089286.

Table 1: Summary of Accuracy of the Four Models

| Model Type | Misclassification of train data | Misclassification of test data |
|-------------------------|---------------------------------|--------------------------------|
| Logistic Regression | 0.31875 | 0.3178571 |
| Decision Tree | 0.2429 | 0.2857 |
| Naïve Bayes Classifier | 0.2864078 | 0.3231707 |
| Support Vector Machines | 0.2089286 | 0.2089286 |

Table 4: Support Vector Machines Summary

| | Radial | Linear | Polynomial | Sigmoid | Best Model | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|--------------------------------|---|------------------|------------------|------------------|------------------|-----|---|---|-----|----|---|-----|-----|---|---|----------|---|---|------|----|---|---|----|---|-----|-----|---|---------|---|--|---|---|-----|-----|-----|---|---|-----|----|--|---------|----------|--|--|---|---|---|-----|-----|---|-----|-----|---|---------|----------|--|--|---|---|---|-----|----|---|-----|-----|
| SVM - Type | C-classification | C-classification | C-classification | C-classification | C-classification | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| SVM - Kernel | radial | linear | polynomial | sigmoid | radial | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Cost or degree | 1 | 1 | 3 | 1 | 4 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| gamma | 0.06666667 | 0.06666667 | 0.06666667 | 0.06666667 | 0.06666667 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| # of Support Vectors | 740 | 755 | 793 | 674 | 713 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| # of classes | 2 | 2 | 2 | 2 | 2 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Levels | 0 1 | 0 1 | 0 1 | 0 1 | 0 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Summary for Levels | (358 382) | (375 380) | (380 413) | (335 339) | (336 377) | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Confusion Matrix – Train Data | <table border="0"> <tr><td>Act_Val</td><td>Pred_Val</td><td></td></tr> <tr><td></td><td>0</td><td>1</td></tr> <tr><td>0</td><td>677</td><td>56</td></tr> <tr><td>1</td><td>219</td><td>168</td></tr> </table> | Act_Val | Pred_Val | | | 0 | 1 | 0 | 677 | 56 | 1 | 219 | 168 | <table border="0"> <tr><td>Act_Val</td><td>Pred_Val</td><td></td></tr> <tr><td></td><td>0</td><td>1</td></tr> <tr><td>0</td><td>679</td><td>54</td></tr> <tr><td>1</td><td>281</td><td>106</td></tr> </table> | Act_Val | Pred_Val | | | 0 | 1 | 0 | 679 | 54 | 1 | 281 | 106 | <table border="0"> <tr><td>Act_Val</td><td>Pred_Val</td><td></td></tr> <tr><td></td><td>0</td><td>1</td></tr> <tr><td>0</td><td>727</td><td>6</td></tr> <tr><td>1</td><td>350</td><td>37</td></tr> </table> | Act_Val | Pred_Val | | | 0 | 1 | 0 | 727 | 6 | 1 | 350 | 37 | <table border="0"> <tr><td>Act_Val</td><td>Pred_Val</td><td></td></tr> <tr><td></td><td>0</td><td>1</td></tr> <tr><td>0</td><td>589</td><td>144</td></tr> <tr><td>1</td><td>271</td><td>116</td></tr> </table> | Act_Val | Pred_Val | | | 0 | 1 | 0 | 589 | 144 | 1 | 271 | 116 | <table border="0"> <tr><td>Act_Val</td><td>Pred_Val</td><td></td></tr> <tr><td></td><td>0</td><td>1</td></tr> <tr><td>0</td><td>676</td><td>57</td></tr> <tr><td>1</td><td>177</td><td>210</td></tr> </table> | Act_Val | Pred_Val | | | 0 | 1 | 0 | 676 | 57 | 1 | 177 | 210 |
| Act_Val | Pred_Val | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | 0 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0 | 677 | 56 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | 219 | 168 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Act_Val | Pred_Val | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | 0 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0 | 679 | 54 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | 281 | 106 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Act_Val | Pred_Val | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | 0 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0 | 727 | 6 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | 350 | 37 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Act_Val | Pred_Val | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | 0 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0 | 589 | 144 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | 271 | 116 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Act_Val | Pred_Val | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | 0 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0 | 676 | 57 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | 177 | 210 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Misclassification - Train Data | 0.2455357 | 0.2991071 | 0.3178571 | 0.3705357 | 0.2089286 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Prediction Summary – Test Data | <table border="0"> <tr><td></td><td>0</td><td>1</td></tr> <tr><td>896</td><td>224</td><td></td></tr> </table> | | 0 | 1 | 896 | 224 | | <table border="0"> <tr><td></td><td>0</td><td>1</td></tr> <tr><td>960</td><td>160</td><td></td></tr> </table> | | 0 | 1 | 960 | 160 | | <table border="0"> <tr><td></td><td>0</td><td>1</td></tr> <tr><td>1077</td><td>43</td><td></td></tr> </table> | | 0 | 1 | 1077 | 43 | | <table border="0"> <tr><td></td><td>0</td><td>1</td></tr> <tr><td>860</td><td>260</td><td></td></tr> </table> | | 0 | 1 | 860 | 260 | | <table border="0"> <tr><td></td><td>0</td><td>1</td></tr> <tr><td>853</td><td>267</td><td></td></tr> </table> | | 0 | 1 | 853 | 267 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | 0 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 896 | 224 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | 0 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 960 | 160 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | 0 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1077 | 43 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | 0 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 860 | 260 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | 0 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 853 | 267 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Confusion Matrix – Test Data | <table border="0"> <tr><td>Act_Val</td><td>Pred_Val</td><td></td></tr> <tr><td></td><td>0</td><td>1</td></tr> <tr><td>0</td><td>677</td><td>56</td></tr> <tr><td>1</td><td>219</td><td>168</td></tr> </table> | Act_Val | Pred_Val | | | 0 | 1 | 0 | 677 | 56 | 1 | 219 | 168 | <table border="0"> <tr><td>Act_Val</td><td>Pred_Val</td><td></td></tr> <tr><td></td><td>0</td><td>1</td></tr> <tr><td>0</td><td>679</td><td>54</td></tr> <tr><td>1</td><td>281</td><td>106</td></tr> </table> | Act_Val | Pred_Val | | | 0 | 1 | 0 | 679 | 54 | 1 | 281 | 106 | <table border="0"> <tr><td>Act_Val</td><td>Pred_Val</td><td></td></tr> <tr><td></td><td>0</td><td>1</td></tr> <tr><td>0</td><td>727</td><td>6</td></tr> <tr><td>1</td><td>350</td><td>37</td></tr> </table> | Act_Val | Pred_Val | | | 0 | 1 | 0 | 727 | 6 | 1 | 350 | 37 | <table border="0"> <tr><td>Act_Val</td><td>Pred_Val</td><td></td></tr> <tr><td></td><td>0</td><td>1</td></tr> <tr><td>0</td><td>589</td><td>144</td></tr> <tr><td>1</td><td>271</td><td>116</td></tr> </table> | Act_Val | Pred_Val | | | 0 | 1 | 0 | 589 | 144 | 1 | 271 | 116 | <table border="0"> <tr><td>Act_Val</td><td>Pred_Val</td><td></td></tr> <tr><td></td><td>0</td><td>1</td></tr> <tr><td>0</td><td>676</td><td>57</td></tr> <tr><td>1</td><td>177</td><td>210</td></tr> </table> | Act_Val | Pred_Val | | | 0 | 1 | 0 | 676 | 57 | 1 | 177 | 210 |
| Act_Val | Pred_Val | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | 0 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0 | 677 | 56 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | 219 | 168 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Act_Val | Pred_Val | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | 0 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0 | 679 | 54 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | 281 | 106 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Act_Val | Pred_Val | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | 0 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0 | 727 | 6 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | 350 | 37 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Act_Val | Pred_Val | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | 0 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0 | 589 | 144 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | 271 | 116 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Act_Val | Pred_Val | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | 0 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0 | 676 | 57 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | 177 | 210 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Misclassification – Test Data | 0.2455357 | 0.2991071 | 0.3178571 | 0.3705357 | 0.2089286 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

CHAPTER 5 – CONCLUSION

All the four methods of Logistic regression, Decision Tree, Naïve Bayes Classifier and Support Vector Machines the lowest misclassification error in train and test data from support vector machines and it was 0.2089286. So, the best model to predict the delivery type using the considered factors is the support vector machines (Table 3).

Study started to find the factors affecting for the Delivery Type and the best model to predict the delivery type. Four factors of mother (ethnicity, age, number of pregnancies and number of babies) and five factors of the infant (gender, weight, head circumference, height and shoulder length) have been considered at the beginning for 1400 babies which were born in Ampara General Hospital from January 2015 to December 2015. Descriptive statistics have been calculated using SPSS 16 and it showed that there are 34% of cesarean deliveries.

Then the association of the considered variables with the type of delivery have been found using Minitab 2018 and the maternal age, delivery time, infant's weight and infant's shoulder length were statistically significant at 5% significance level. Then the Logistic regression was carried out to find the logistic model, getting the dependent variable as delivery type. Age, delivery time, number of babies, head circumference, height of the infant was statistically significant in 95% confidence interval and the misclassification error was 0.31875 for the test data and 0.3178571 for the train data. Each one-unit change in number of pregnancies will increase the log odds of having a normal delivery by 0.016, and its p-value indicates that it is somewhat significant in determining the delivery type. Similarly, head circumference, shoulder length positively related to have a cesarean delivery. Delivery time is negatively related to the log odds of having a cesarean delivery (i.e., controlling for the other variables, mothers who has longer delivery time are less likely to have a cesarean delivery). Similarly, number of babies, weight and height also negatively related to the log odds of having a cesarean delivery.

In decision tree the accuracy of the model for train data was 75.71% and accuracy for test data it was 71.43 % with P value $6.724e^{-14}$ and 0.02427 respectively. That is misclassification for train data was 0.2429 and for test data it was 0.2857. Biggest split was delivery time.

In Naïve Bayes Classifier, the priori probability of a normal delivery in train data set was 0.6561489 and a cesarean delivery it was 0.3438511. That is the frequency of having a normal delivery is 65.6% and cesarean delivery is 34.4%. In the confusion matrix of train data represents, 716 deliveries predicted as normal and 166 deliveries predicted as cesarean and it was true. Also 259 Cesarean deliveries predicted as normal and 95 normal deliveries wrongly predicted as cesarean deliveries. Misclassification in train data is 0.2864078. Similarly, in the confusion matrix of the test data, 88 deliveries predicted as normal and 23 deliveries predicted as cesarean and it was true. Also 35 Cesarean deliveries predicted as normal and 18 normal deliveries wrongly predicted as cesarean deliveries. Misclassification error in test data is 0.3231707.

In support Vector Machines we did the analysis using the kernels of radial, linear, polynomial and sigmoid using train and test data sets. Number of support vectors were 740, 755, 793 and 674 respectively. Misclassification rate for train data were 0.2455

357, 0.2991071, 0.3178571 and 0.3705357 respectively. Also, misclassification rate for train data were 0.2455357, 0.2991071, 0.3178571 and 0.3705357 accordingly. Then by using the train data the model was tuned. Sampling method that was used is 10-fold cross validation and best parameters were epsilon 0 and cost was 4. The better performance was 0.2803571. In the best model number of support vectors were 7 13 and the classification was radial. Misclassification for train data was 0.2089286 and misclassification for the test data was 0.2089286.

According to the results in the analysis, all the four methods of Logistic Regression, Decision Tree, Naïve Bayes Classifier and Support Vector Machines the lowest misclassification error in train and test data gave by the Support Vector Machines and it was 0.2089286. So, the best model to predict the delivery type using the considered factors is the support vector machines. So, the type of delivery can be predicted with 0.791071 accuracy using the support vector machines. Similarly, by comparing all the models, delivery time, age of the mother, height of the infant, number of pregnancies had a significant effect of selecting the delivery type.

REFERENCES

(n.d.).

AM, v. d. (2013). Factors affecting vaginal birth after caesarean section at Middlemore Hospital, Auckland, New Zealand. *PubMed Central*, 126(1383):49-57.

Faisal-Cury A. (2007). Prevalence of anxiety and depression during pregnancy in a private setting sample. *Springer Link*, 10(1):25-32.

Gamlath E.G.K.M. (2018). Determinants of Type of Delivery in Sri Lanka: A Logistic Regression Approach. *Seventh Annual Science Research Sessions - 2018* (p. 11). Sammanthurai, Sri Lanka: South Eastern University of Sri Lanka.

González MS. (2015). Mode of Vaginal Delivery: A Modifiable Intrapartum Risk Factor for Obstetric Anal Sphincter Injury. *Obstetrics and Gynecology International*, doi: 10.1155/2015/679470.

Landon MB. (2004). Maternal and perinatal outcomes associated with a trial of labor after prior cesarean delivery. *The New England Journal of Medicine*, 351:2581-2589.

M, R. (2013). Factors That Influence Women's Decision Making for Childbirth. *The Journal of Perinatal Education*, 22(3): 171-180.

Promotion, O. o. (2010). U.S. Department of Health and Human Services.

Rasoli, M. (2019). Effects of reviewing childbirth scenarios on choice of delivery type: a randomized controlled trial. *Turkish Journal of Obstetrics and Gynecology*, 16(1): 15-22.

Robinson B.K. (2011). An Evidence-Based Approach to Determining Route of Delivery for Twin Gestations. *PubMed Central*, 4(3-4): 109-116.

Ronsmans, C. (2012). Factors influencing rising caesarean section rates in China between 1988 and 2008. *PubMed Central*, 90(1): 30–39A.