

PREDICTION OF WOMEN SURVIVAL RATE DURING PREGNANCY USING CLASSIFICATION AND CLUSTERING

D. Catherine Remy, Research Scholar VIT, University, Vellore, India.
S. Rita Samikannu, Associate Professor, VIT University, Vellore, India.
M.K.Nallakaruppan, Assistant Professor VIT University, Vellore, India.

Abstract

A statistical survey shows that In India a mother dies in every eight minutes in India at the same amount of time 20 women suffer from lifelong health impairments that result from complications during their pregnancies. The majority of these deaths are among women in the 15-29 age group, at the prime of their reproductive lives. The most tragic aspect of these deaths is that about 90 percent of them are avoidable in nature, in case of women are able to receive the right kind of intervention. This paper deals with the health record of pregnant women obtained from a primary health center, Vellore. This is a district in Tamilnadu and India. This paper analyses various compound reasons regarding the mother mortality rate and the Impact on these factors that affect the survival rate of the women. This paper performs Decision table based classification and Filtered Clustering to prove the correctness of the results obtained from the survey regarding which of the fourteen factors involved in the analysis has the highest impact on the survival status of the women during the period of pregnancy. This paper classifies these fourteen parameters and the impact of the every single parameter on the survival status of the women and how much percentage is the survival rate obtained from this survey and out of which is the primary key attribute that has the highest impact on the survival rate is predicted by the rigorous classification and the clustering algorithms

Keywords: Decision Table based Classification, Filtered Clustering, Attribute Selection and Survival rate prediction.

I.INTRODUCTION

There are three stages of the delay that could impact a pregnant women and increase the risk of the mortality .First delay is lack of education in the family members to read the pregnancy related warning signs and report the same to the doctor and more so there are several prevailing health conditions for women which also would impact on the survival rate .For example in India Most

prosperous states like Maharashtra and Gujrat where 90% women are anemic.23.1% women India

receive iron based supplements and on a whole 56.2% of the pregnant women in India are anemic.Lack of knowledge in conveying these things to doctor. This is termed as the "First Delay".

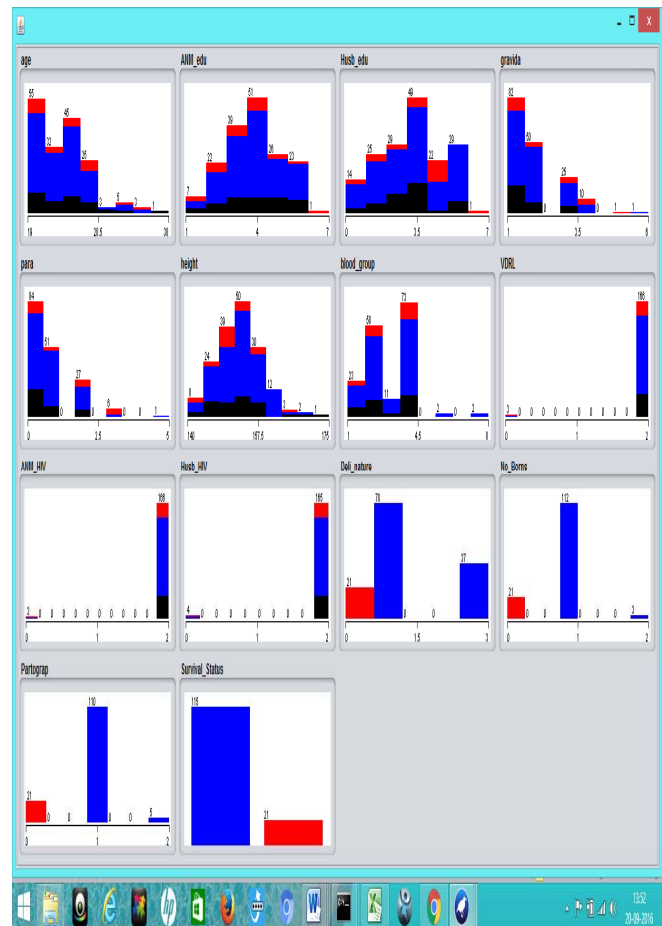
The second delay occurs during the transportation time in India it is reported that the 70% of people are not prepared to take the pregnant women to hospital that is there is lack of transportation facility and availability of vehicles. Apart from this 40% of death occurs during excessive bleeding and sepal infections at home.

The Third delay occurs in hospitals lack of obstetricians to attend the critical pregnant women in hospitals is recorded with an average of 30% and only 10% anaesthetists are available. From this survey we clearly understand that the reason for the first two delays are because of lack of literacy and awareness among the pregnant women and her husband and the family members. This paper addresses not only the education but also the other factors that could impact the survival rate.

II.DATA GATHERING

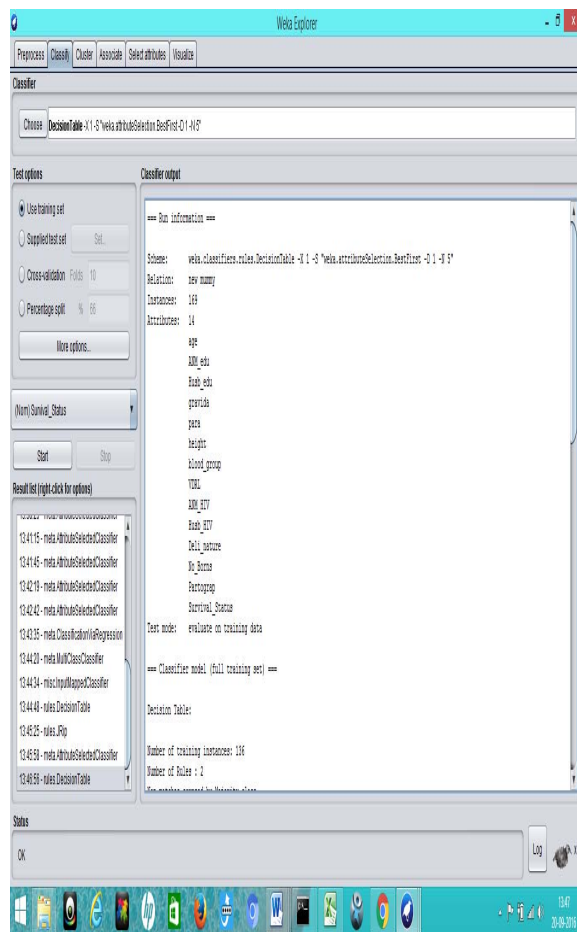
In this survey there are 14 parameters that includes Age, Education of the wife, Education of the husband, Gravida, No of pregnancies, Last menstrual Period, Estimated Date of the Delivery ,Height, Blood Group ,Venereal Disease Research Laboratory Test ,Husband HIV details, Wife HIV Details, Nature of Delivery, No Of children born,Partograp of children born and the Survival Status of the children. There are 169 data items collected in all the 169 women belong to rural areas in Vellore which is a district in Tamilnadu, a state of the Indian country. This datum is collected over a period of cohort study of one year during the period of Jun 2014-July 2015.

The primary classifier is selected as Survival status which is the classifier against all other 13 attributes which are taken for comparison regarding how the survival rate is impacted by all the other classifiers and which of the 13 classifier has a uniform normal distribution of values and the maximum impact factor when compared with survival status and which provides the normalized curve of the highest impact factor is chosen for the further analysis of data through the justified classification and the clustering process. Once the parameter has the highest impact factor is arrived later the classification of data relevant to that parameter and clustering is done to whether that data is fitting within the range of the prediction



(a) Comparison of the Survival rate and other parameters under measurement for the prediction.

III.DECISION TABLE BASED CLASSIFICATION

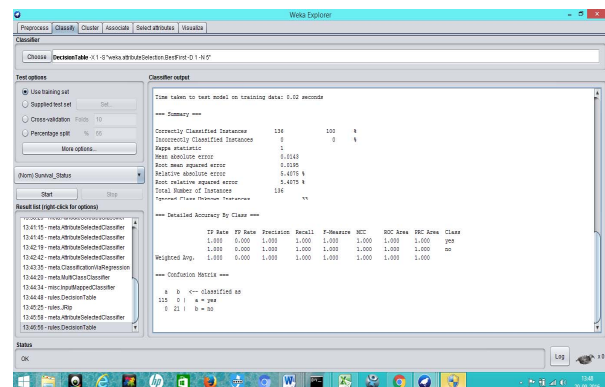


(b) Classification using Decision Table

We are need to perform the classification process as the next level that is ensuring the correctness of data prediction through the rigorous classification algorithm. Out of the other available algorithms (as we are taking a single attribute to match the other attribute) it is most appropriate that we need to use a predictive classification process. In this regard we make use of the **Attribute Selection Best Fit based algorithms** and we have chosen **Decision Table based classification algorithm**. According to this algorithm each and every decision corresponds to a variable ,relation or a predicate whose possible values are listed among the condition alternative. Each action is a procedure or an operation to perform, and the entries specify whether the action is set to be performed for the condition alternatives the entry corresponds to. A condition may be

considered to be a primary influencing factor may be proven insignificant once the relevant action is performed when no condition has actual influence on the resultant action This algorithm ran around all the 169 data items and found that 136 out of 169 has complete details and 33 out of 169 are the missing data elements and based on the classification we have estimated several parameters like **TP-Rate, FP-Rate,Precision,Recall,F Measure, MCC, Region of the convergence area, PRC Area** and the **class**

There are 136 classified instances that predicts the survival rate comparing all the other attributes in two major categories such as two nominal values “yes” or “no”. After running the classification process we observe that 115 data items are classified under “Yes” and 21 data items are classified as “No”. This classification clearly shows that the **survival status** is higher when comparing all the data attributes with survival status as classifier which are taken from the training set of data

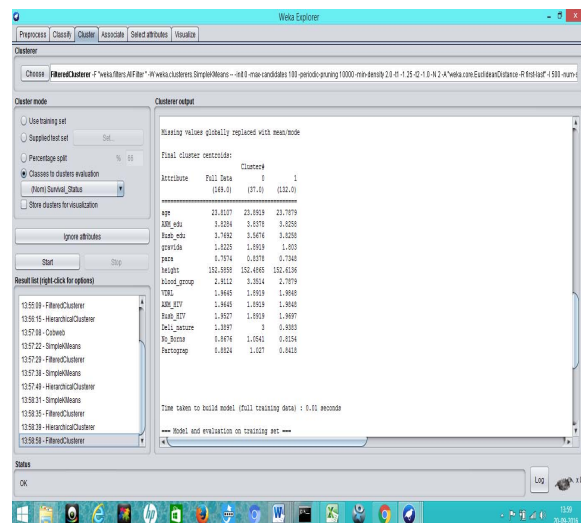


(c) Result Analysis of the Clustering Process

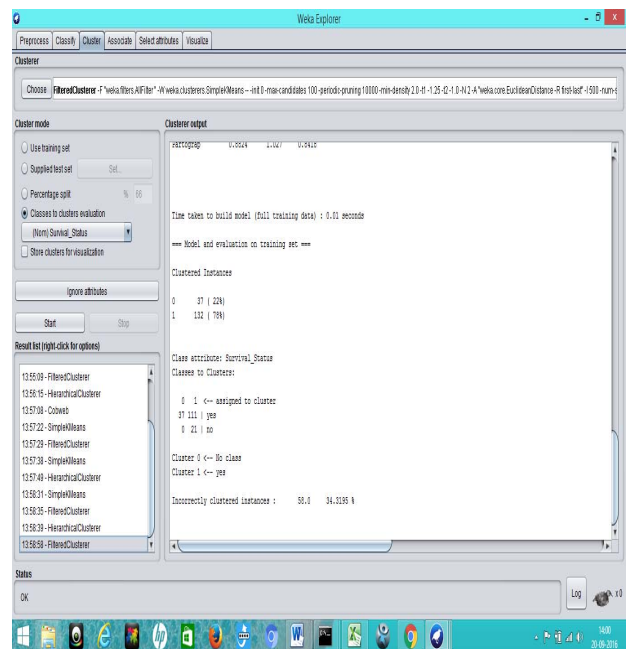
V.FILTERED CLUSTERING OF DATA

Filtered clustering is chosen to see how many data items are classified into survival status attribute to see how many data items are fitting to the Decision “Yes” and how many data items are fitting decision “No”. Filtered clustering is done to check the survival rate prediction. Out of 169 data items 132

(78%) are found to be in cluster 1 and 32 data items (28%) are found to be in cluster 0. Cluster 1 corresponds to the decision “Yes” and Cluster 0 corresponds to the decision “No”. The clustering also shows the survival status is higher than the mortality rate by attribute based filtered clustering process. There are 136 data items which are complete which are only taken for classification. But in case of clustering it produces results even with missing data items included and still produces the prediction efficiency of 78% for the survival rate status. This assures that the research shows that majority of the factors used in the study has the positive impact on the survival status



(d) Filtered Based Clusters obtained from the Predicates



(e) Results obtained from the Clustering Process

VI.CONCLUSION

The results which are obtained from the process of filtering, classification and the clustering shows that the three factors have the prime impact on the survival rate of a child .They are Husband’s Literacy rate, Wife’s Literacy rate and the Height. These three parameters influence the survival rate and all the other parameters also have some significant impact on the survival rate. The survival rate is found to be positive in the survey these parameters could be observed in the upcoming years to see the consistency in the impact of these parameters on the survival.

REFERENCES

1. TulaSalud: An m-health system for maternal and infant mortality reduction in Guatemala AndreMartinez-Fernandez, Isabel Lobos-Medina, César AugustoDíaz-Molina, Moisés Faraon Chen-Cruz and Ignacio Prieto-Egido, TulaSalud, Journal of Telemedicine and Telecare 2015, Vol. 21(5) 283–291.
2. <http://www.armman.org/motherslife>
3. Leveraging human capital to reduce maternal mortality in India: enhanced public health system or public-private partnership? Karl Krupp and Purnima Madhivana Human Resources for Health 2009 7:18 **DOI:** 10.1186/1478-4491-7-18
4. Maternal Mortality in Colombia in 2011: A Two Level Ecological Study-Luz Mery Cárdenas-Cárdenas, Karol Cotes-Cantillo, Pablo Enrique Chaparro-Narváez, Julián Alfredo Fernández-Niño, Angel Paternina-Cañedo, Carlos Castañeda-Orjuela, Fernando De la Hoz-Restrepo Published: March 18, 2015 <http://dx.doi.org/10.1371/journal.pone.0118944>