Statistical Modeling for Anaemia of Reproductive Aged Women in Baramati

Abstract : Anaemia is a global health problem and women in reproductive age (WRA) are amongst the most affected population. Its consequences include low birth weight and maternal mortality. The aim of this study is to develop a machine learning model to predict the anaemia by using complete blood count report. For this purpose 363 women with age group 15-49 were selected. According to WHO anaemia was categorized into mild, moderate, severe, and non-anaemia. Decision tree model under the CART was developed for prediction purpose. After fitting the model it was observed that the parameters PCV and MCHC are mostly affects the anaemia status. We can also diagnose anaemia at an early stage with this approach at a low cost. We must consume a meal that is enriched with vitamin B12, iron, and other nutrients to decrease anaemia in reproductive-aged women.

IndexTerms - Component, for matting, style, styling, insert.

II. RESEARCH METHODOLOGY

2.1 Population and Sample

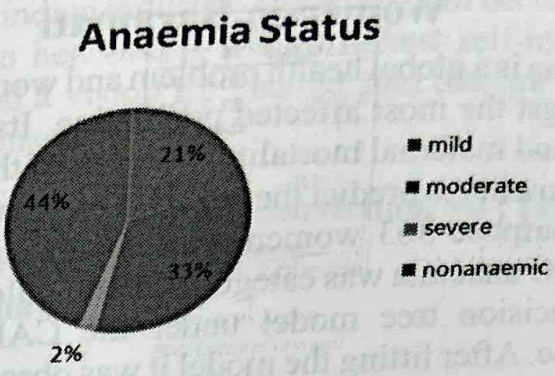
The study area is located near Baramati Town, District Pune. 363 CBC reports of women at child bearing age were included in the study as a sample. According to WHO Anaemia was categorized as mild, moderate, severe and no-anaemia. The Supervised machine learning model decision Tree was developed for prediction of anaemia status under the CART algorithm.

2.2. Theoretical framework

For classification and regression, Decision Trees (DTs) are one of the best non-parametric supervised learning method. The goal is to develop a model that predicts the value of a target variable by learning simple decision rules from data properties. Without any parameter settings, a decision tree classifier can be created, making it ideal for exploratory knowledge discovery.

CART is one of the most widely used methods for building decision trees in the machine learning. CART builds a binary decision tree by separating records at each node according to the function of a single attribute. The GINI Index is used by CART to determine the optimal split. The initial split produces two nodes, which we attempt to split in the same way, resulting in a root node. We go over all of the input fields again to find the possible splitters. If no split is found that significantly diminishes the diversity of a node, it is labelled as a leaf node. After a while, all that's left is a leaf node, and we've formed a whole decision tree. Due to over fitting, the complete tree that will be performs poorly when classifying a new batch of records. Each record in the training data is assigned to a leaf of the whole decision tree at the end of the tree-growing technique. It is now possible to assign a class to each leaf. High-dimensional data is no match for decision trees.

III. STATISTICALANALYSIS



From the above pie chart it was observed that nearly 66% women at reproductive age were found to be anaemic.

Decision tree:

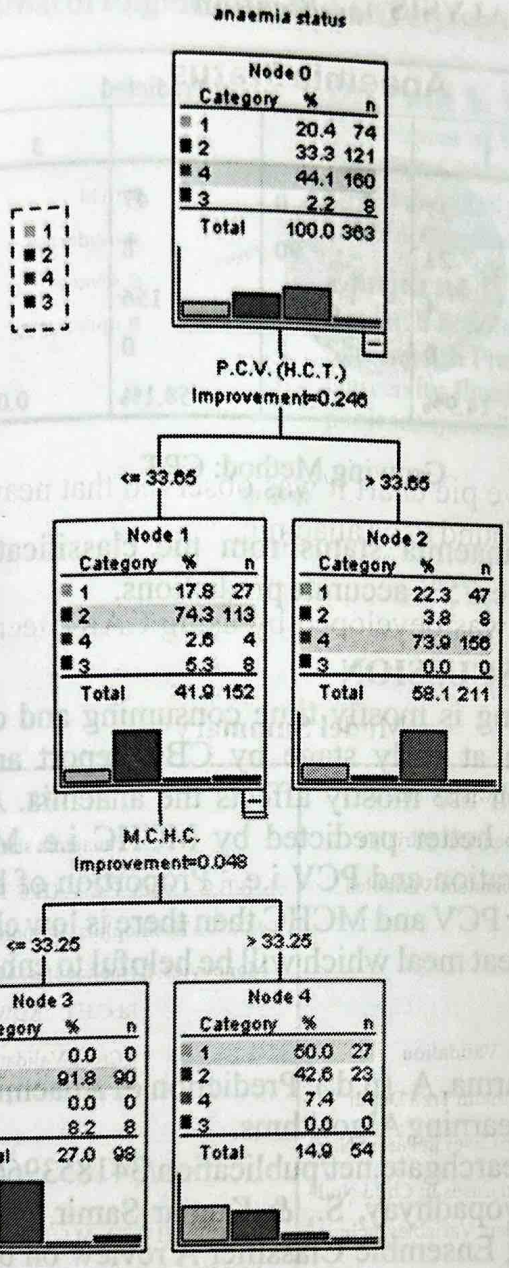
Decision tree was developed by using CART technique. Following output was found.

Model Summary

		CRT
Specifications	Growing Method	anaemia status
	Dependent Variable	R.B.C. Count, P.C.V. (H.C.T.), W.B.C. Count, M.C.V., Eosinophils, Basophils, Platelet Count, Monocytes, Lymphocytes, M.C.H., Neutrophils, M.C.H.C., RDW-CV
	Independent Variables	
	Validation	Cross Validation
	Maximum Tree Depth	5
	Minimum Cases in Parent Node	100
	Minimum Cases in Child Node	50
Results	Independent Variables Included	P.C.V. (H.C.T.), M.C.H., M.C.V., RDW-CV, M.C.H.C., R.B.C. Count, Platelet Count, W.B.C. Count, Neutrophils, Lymphocytes, Monocytes
	Number of Nodes	5
	Number of Terminal Nodes	3
	Depth	2

Above table shows model summary i.e. information about dependent variable and independent variables.

Principal
Tuljaram Chaturchand College
Baramati



Interpretation: From the above tree we found that the factors like PCV (HCT), MCHC, are more import than other. If the PCV is greater than 33.65 then there is 73.9% chance of person has no anaemia. If the PCV is less than 33.65 then there is 74.3 % chance of person having moderate anaemia and 17.8% chance that of mild anaemia. If the PCV is less than 33.65 and MCHC is less than 33.25 then there is 91.8 % chance of moderate anaemia and 8.2 % chance of severe anaemia that means if the MCHC is less than 33.25 then there is 100% chance of anaemia but if MCHC is greater than 33.25 then there is 50% chance of mild anaemia and 42.6% chance of severe anaemia but o% chance of severe ana[illegible]

Classification

Observed	Predicted				
	1	2	4	3	Percent Correct
1	27	0	47	0	36.5%
2	23	90	8	0	74.4%
4	4	0	156	0	97.5%
3	0	8	0	0	0.0%
Overall Percentage	14.9%	27.0%	58.1%	0.0%	75.2%

Growing Method: CRT

Dependent Variable: anaemia statusFrom the classification table it was observed that model gives75% accurate predictions.

I. RESULTS AND DISCUSSION

Anaemia testing is mostly time consuming and costly. This study diagnoses the anaemia at early stage by CBC report and also finds the important factors which are mostly affects the anaemia. According to this study Anaemia can be better predicted by MCHC i.e. Mean Corpuscular Haemoglobin Concentration and PCV i.e. Proportion of blood made up of cells. If we improve our PCV and MCHC then there is low chance of anaemia. So this study tells us to eat meal which will be helpful to enhance the PCV and MCHC values.

REFERENCES

1. Anand, P., & Sharma, A. (n.d.). Prediction of Anaemia among children using Machine Learning Algorithms. https://www.researchgate.net/publication/341853966
2. Dutta, S., Bandyopadhyay, S., & Kumar Samir, B. (2020). Diabetes Prediction Using Ensemble Classifier A review on bio medical image processing View project Medical Imaging View project Diabetes Prediction Using Ensemble Classifier. In International Journal of Medical and Health Sciences Journal Home (Vol. 9, Issue 2). https://www.researchgate.net/publication/341775512
3. Dithy, M. D., & Krishnapriya, V. (2019). Anemia selection in pregnant women by using random prediction (Rp) classification algorithm. International Journal of Recent Technology and Engineering, 8(2), 2623–2630. https://doi.org/10.35940/ijrte.B3016.078219
4. Chavan, R. H., Naik, D. N., Halbati, R. A., Patel, S. J., &Prajapati, M. A. (2018). Disease Prediction using Machine Learning. International

Research Journal of Engineering and Technology. www.irjet.net

Dr. Patil V. V.
Department of Statistics,
Tuljaram Chaturchand College, Baramati,
Dist. Pune (MS), India 413102
vaishutcc@gmail.com

Zanjurne P. S.
Research Scholar,
Shri Jagdish Prasad Jhabarmal Tibrewala
University Jhunjhunu, Rajasthan-333001
poojaszanjurne21@gmail.com

***

The lamp of education that you lighted so many years earlier
is enlightening millions of your daughters today.
Savitribai Phule