

A Report On
Analysis of Low Birth Weight in Babies and Its Causes.

by

Ravi Teja

Karishma Garewal

Keshavanand Siddharth

Executive summary

This report provides an analysis and evaluation of various causes for low birth weight in babies. Understandably, a lot of the causes are directly related to the parents.

The data set taken for this analysis has 17097 instances. Each instance has 38 variables out of which 28 variables were selected as input variables, 9 variables were rejected and one target variable was established. The basis on which these variables were selected is given in the report below. Methodology used for the analysis is decision tree. 3 different trees: Maximal, Optimal and Probability tree were made. Maximal and optimal tree are generated with the criterion of decision assessment measure and probability tree is generated with the criterion of average square error assessment measure.

Each of these models was compared based on their average square error, complexity and misclassification rate in which the data was partitioned into training and validation as 50 % and 50% respectively. A comparison model was taken with 70-30 partition of training and validation data was also considered in which we could conclude that the 50-50 dataset explains the model better.

Results of data analyzed show that probability tree is the best decision tree which is built with average square error assessment measure. It has a misclassification rate of 0.354 and an average square error of 0.219 for validation data and 15 leaves.

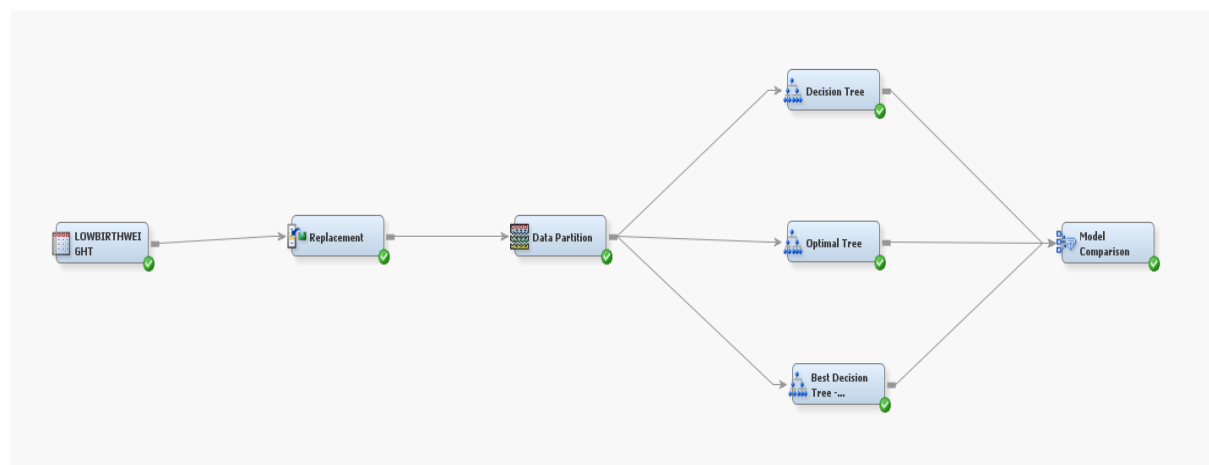
From the resultant decision tree we can conclude that the major factors affects the weight of babies are ethnicity of father, Hypertension during pregnancy in mother, Smoking habit of mother and the month of pregnancy where prenatal care began.

- Shown below is the table which enlists the reason for accepting or rejecting a corresponding variable.

Variables	Role	Level	Note
LBWT	Target	Binary	This is our Target variable as we need to understand the cause of low birth weight. And they are only two categories either 0 or 1. so It will be Binary
FAGE	Input	Nominal	Effect of Father's age on baby will be there.
MAGE	Input	Nominal	Effect of Mother's age on baby will be there because after certain age the level of fertility goes down
FEDUC	Input	Nominal	Parent's education plays an important role because they have to take care about the nutrition at the time of pregnancy
MEDUC	Input	Nominal	Parent's education plays an important role because they have to take care about the nutrition at the time of pregnancy
TOTALP	Input	Nominal	This will help parents in taking precautions for the current pregnancy with the knowledge of previous ones. Level of measurement will be nominal
PRENATAL	Input	Nominal	This is an important factor because parents will get recommendations from the doctor regarding Mother's diet and health care to take
BDEAD	Input	Nominal	Impact of perinatal loss will be there on mother's brain and will affect the subsequent pregnancy
TERMS	Rejected		Termination won't make any impact. It is usually the couple's decision to have a baby or not. They can have a healthy baby after certain time when they are ready
LOUTCOME	Input	Nominal	Studies suggest that last outcome of an adult might affect the subsequent baby
YrsLastFetalDeath	Rejected		As we are having LOUTCOME variable, these both variables almost mention same. So Rejecting the variable
YrsLastLiveBirth	Rejected		As we are having LOUTCOME variable, these both variables almost mention same. So Rejecting the variable
MARITAL	Input	Binary	Marital status of mother will affect the born baby weight. And as it has only two levels it will be binary
CHILDREN	Input	Nominal	Impact of previous children living will have an effect on current baby weight, because the parents need to take care of the previous baby and current baby.
ETHNICITYMOM	Input	Nominal	Ethnicity of parents will affect the born babies weight
ETHNICITYDAD	Input	Nominal	Ethnicity of parents will affect the born babies weight
CIGNUM	Rejected		Smoker variable also gives the same information. There is no requirement to find out the number of cigarettes.
DRINKNUM	Input	Nominal	Drinking habit of parents will affect born baby weight. And the measure will be nominal
DRINKER	Rejected		DRINKNUM variable also gives the same information. There is no requirement to find out the number of drinks
SMOKER	Input	Binary	Smoking habit of parents will affect born baby weight. And the measure will be binary
ANEMIA	Input	Binary	Anemia can cause a baby to born with low birth weight
CARDIAC	Input	Binary	Cardiac disease can cause low birth weight in babies
ACLUNG	Input	Binary	Lung diseases in Adults can cause serious birth defects to babies
DIABETES	Input	Binary	Studies states that diabetes can cause weight disorders in babies
HERPES	Input	Binary	Herpes is a STD which can affect the baby health condition after labour
HYDRAM	Input	Binary	hydramnios or polyhydramnios is an independent factor for low birth weight in babies
HEMOGLOB	Input	Binary	Hemoglobinopathy affects the new born baby weight as per recent studies and the level of measurement is binary
HYPERCH	Rejected		Previous hypertension in adult won't affect the baby weight
HYPERPR	Input	Binary	hypertension is the major factor that affects the born baby weight
ECLAMP	Input	Binary	Eclampsia can cause pre term birth or low birth weight

CERVIX	Input	Binary	Incompetent cervix can lead to low birth weight, deafness, blindness
PINFANT	Rejected		There is no particular reasoning that pre infant weight can affect the current baby weight
PRETERM	Input	Binary	previous preterm births can affect the current baby weight
RENAL	Input	Binary	Renal disease in adults can cause low birth weight in baby
RHSEN	Input	Binary	Rh sensitization can cause low birth weight
UTERINE	Input	Binary	Uterine bleeding can cause low birth weight
AMNIO	Rejected		Amniocentesis is a diagnostic test for genetic disorders. It won't affect the baby weight
ULTRA	Rejected		Ultra sound test is also one of the diagnostic tests. It won't affect the baby weight

- After the variable selection a sequence of 3 decision trees were build to predict the model with the above selected variables, decision trees were built using these variables. The Workflow as shown below

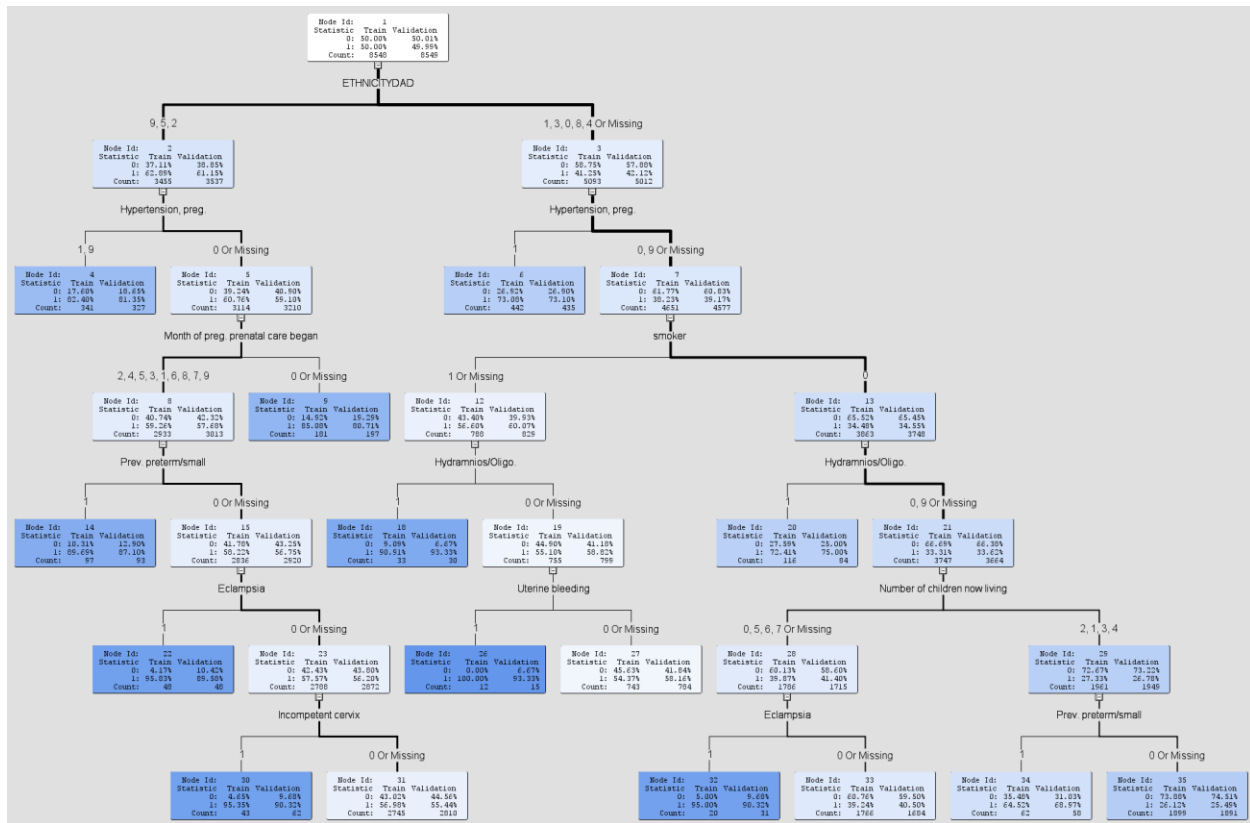


- After the decision trees have been made, choosing the best decision tree is the next objective. For this judgement the following parameters are taken into account.

The best decision tree selected as Probability Tree.

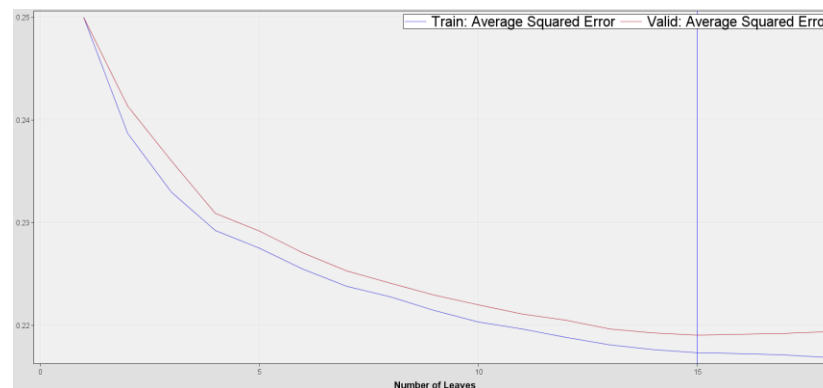
Fit Statistics						
Model Selection based on Valid: Misclassification Rate (_VMISC_)						
Selected Model	Model Node	Model Description	Valid: Misclassification Rate	Train: Average Squared Error	Train: Misclassification Rate	Valid: Average Squared Error
Y	Tree3	Probability Tree	0.35419	0.21735	0.34944	0.21909
	Tree2	Optimal Tree	0.35419	0.22376	0.34944	0.22495
	Tree	Decision Tree	0.35443	0.21691	0.34909	0.21942

- Below is the overview of the best decision tree.

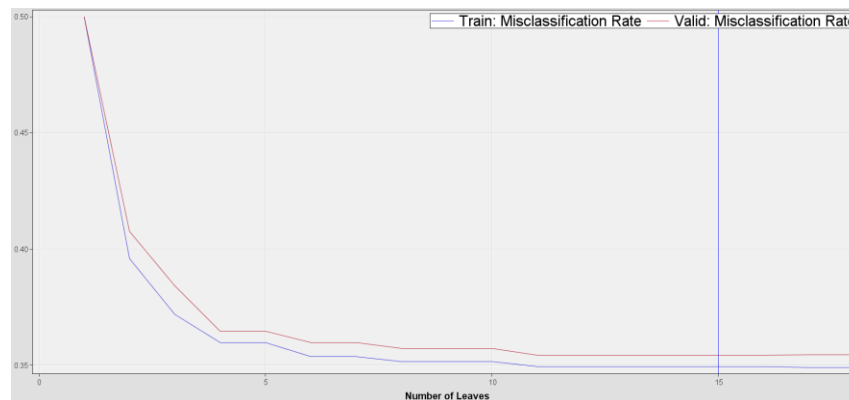


Below are the considered values for validating the model with validation data set.

- Average Square Error for Best Decision Tree



- Misclassification rate for the best decision tree:



- Variable Importance for Decision tree having 15 leaves:

Variable Name	Label	Number of Splitting Rules	Importance	Validation Importance
ETHNICITYDAD		1	1.0000	
HYPERPR	Hypertension, preg.	2	0.8110	
smoker		1	0.5763	
HYDRAM	Hydramnios/Oligo.	2	0.4696	
PRETERM	Prev. preterm/small	2	0.4338	
CHILDREN	Number of children now living	1	0.3902	
ECLAMP	Eclampsia	2	0.3680	
PRENATAL	Month of preg. prenatal care began	1	0.3434	
CERVIX	Incompetent cervix	1	0.2543	
UTERINE	Uterine bleeding	1	0.1597	
CARDIAC	Cardiac disease	0	0.0000	
FEDUC	Education of father (years)	0	0.0000	
BDEAD	Number previous live births now dead	0	0.0000	
MAGE	Age of mother	0	0.0000	
DRINKNUM	Average # of alcoholic drinks per week	0	0.0000	
FAGE	Age of father	0	0.0000	
ACLUNG	Ac/Ch Lung disease	0	0.0000	
ANEMIA	Anemia -- mother	0	0.0000	
MEDUC	Education of mother (years)	0	0.0000	
HEMOGLOB	Hemoglobinopathy	0	0.0000	
DIABETES	Diabetes	0	0.0000	
ETHNICITYMOM		0	0.0000	
RENAL	Renal disease	0	0.0000	
MARITAL	Marital status	0	0.0000	
LOUTCOME	Outcome of last delivery	0	0.0000	
TOTALP	Total pregnancies (including this one)	0	0.0000	
RHSEN	Rh sensitization	0	0.0000	
HERPES	Genital herpes	0	0.0000	

From the above Decision tree we conclude that the major factors that affect the baby weight are as below,

1. Father's Ethnicity
2. Hyper Tension during Pregnancy
3. Smoking habit of Mother
4. Month of pregnancy prenatal care began
5. Previous preterm/small
6. Hydramnios
7. Eclampsia
8. Uterine Bleeding

The first split being is EthnicityDAD and the competing splits are EthnicityMOM and Marital Status.