Educational Administration: Theory and Practice
2024, 30(5), 12738-12741
ISSN: 2148-2403
https://kuey.net/

Research Article

Prediction Of Neonatal Jaundice(Hyperbilirubinemia) By Using Logistic Regression

Ragini Patil[1*], Dr.R.R. Kumbhar[2], Dr. S.V. Kakade[3]

[1*]Assistant Professor, Yashwantrao Chavan College of Science Karad, Email: patilrd1996@gmail.com
[2]Principal, Vivekananda college Kolhapur(Empowered Autonomous), Email: rrkumbhar@yahoo.co.in
[4]Professor, Krishna Institute of Medical Science Demmed to be an University karad, Email:satishvkakade@yahoo.co.in

Citation: Ragini Patil.et al (2024), Prediction Of Neonatal Jaundice (Hyperbilirubinemia) By Using Logistic Regression, *Educational Administration: Theory and Practice*, *30*(5), 12738-12741, Doi: 10.53555/kuey.v30i5.3360

ARTICLE INFO

ABSTRACT

The present study is undertaken with the aim to evaluate the utility of Logistic Regression classification in pre- identifying Neonatal Jaundice.In this research by using Logistic Regression classification we have predicted Neonatal Jaundice by using various parameters of neonates. Here our aim is to provide a benchmark and improve earlier ones in the field of Neonatal Jaundice diagnostics with the help of Logistic Regression classification technique. We have collected data of neonates by observing them from birth up to 72-96 hours of postnatal life. We have predicted Neonatal Jaundice by using Logistic regression. Then we have obtained accuracy, precision, recall, F1 score as evaluative measures. Also by using Receiver Operating Characteristic curve (ROC) area under curve (AUC) is 89.99%. This indicates that our model has higher discriminating power in prediction of Neonatal Jaundice.

Keywords: Hyperbilirubinemia(Jaundice), Logistic regression(LR), AUC, prediction, accuracy.

Introduction:

Neonatal Hyperbilirubinemia is the common abnormal physical findings during the first week of life. Neonatal Hyperbilirubinemia affects nearly 60% of term neonates during the first week of life according to Rennie J, Burman-Roy S, Murphy S.Neonatal Hyperbilirubinemia is a cause of concern for the parents as well as for the pediatricians. The Neonatal Hyperbilirubinemia recognition, follow-up, early treatment and prevention of bilirubin induced encephalopathy have become more difficult as a result of early discharges from the hospital. The concept of prediction offers an attractive option to pick up babies at risk for Neonatal Hyperbilirubinemia. Physical examination is not a reliable measure to check serum bilirubin levels. In this research we had studied various parameters of neonates and perform Logistic Regression to predict neonatal jaundice. By predicting newborns who at risk for significant Neonatal Hyperbilirubinemia (Jaundice) early at birth, pediatricians can design and implement the follow-up program in these high-risk groups, costeffectively.

Research Methodology:

In this study we included168 Neonates .Neonates will be followed from birth up to 72-96 hours of postnatal life. We have collected datawhichincludes following attributes of neonates as follows:

- Gender of baby
- Cord serum albumin (CSA) level
- Gestational age
- Blood group of mother
- Blood group of Baby
- Maternal weight of mother
- Birth weight of baby
- Type of Delivery (Normal, C section)
- Parity
- Bilirubin level
- Phototherapy required (Yes/ No)
- Transfusion (Yes/No)

• Jaundice (Yes/No)

Cord blood will be collected at birth. All the babies will be followed up daily for the development of jaundice. Peripheral venous blood will be collected for estimation of total serum bilirubin between 72-96 hours of life. Note that preterm babies, Neonatal sepsis, Instrumental delivery, Birth asphyxia, Respiratory distress, Meconium stained amniotic fluid, Neonatal jaundice observations within 24 hours of life were excluded. Our target variable is Jaundice (Yes/No).
Here data is divided into two parts for training and testing as 80% data is used to build or train and remaining 20% for test the model. As our target variable or dependent variable is Jaundice (Yes/No)is dichotomous. We have used Logistic Regression to predict Jaundice (Hyperbilirubinemia) of Neonates by using 12 parameters.

Logistic Regression Model:

Logistic Regression is a Supervised Machine Learning Classification Algorithm used to predict the target variable. As the target variable is binary, which means there are two possible classessay as positive and negative . Linear regression technique cannot be used so Logistic Regression is preferred as it deals with dichotomous target variable. Our target variable assumes only two values then by using logistic function we have;

$P(Y=1) = \frac{1}{1+e^{-(X\beta+\varepsilon)}}$

$P(Y=0) = \frac{e^{-(X\beta+\varepsilon)}}{1+e^{-(X\beta+\varepsilon)}}$

Where X : represents vector of K independent variables(regressors)
β : represents vector of unknown but constant regression coefficients or weights.
ε : represents error due to chance.
Y: Target variable or dependent variable
In our case Y=1 means that neonate have Jaundice(Positive) and Y=0 means that neonate does not have Jaundice(Negative)

Also in this study we have used cross- validation technique.

Cross- Validation technique:

In K- fold cross-validation complete data set is divided in to k- folds(subsets) say as D_1, D_2.....D_k. Training and testing is performed K- times. At each i^{th} iteration Subset Di is reserved for testing and remaining (k-1) subsets of dataset are used to train the model. Hence each sample is used same number of times(k-1) for training and once for testing.
In this research we used 10- fold cross-validation technique. That is number of folds K= 10.

Performance Evaluation of Model:

Performance of model can be evaluated by using confusion matrix and hence accuracy, precision, recall, f1-score, ROC curve.

Confusion Matrix

While evaluating performance of classificatory model confusion matrix is essential.

		Predicted	
		Yes	**No**
Actual	**Yes**	True Positive (TP)	False Negative (FN)
	No	False Positive (FP)	True Negative (TN)

Accuracy

Accuracy is the ratio of total number of correctly classified observations to the total number of observations. Accuracy is used to measure the performance of the model.

$Accuracy = \frac{TP+TN}{FP+FN+TP+TN}$

Precision

It is a measure to calculate how accurate a model's positive predictions are.

Precision=$\frac{TP}{FP+TP}$

Recall

It is used to measures the effectiveness of a classification model in identifying all relevant observations from a dataset.

Recall=$\frac{TP}{FN+TP}$

F1-Score

F1 score is a weighted average of the precision and recall. It is used to evaluate the overall performance of a classification model.

F1-Score: = $\frac{2*Precision*Recall}{Precision+Recall}$

ROCCurve

Receiver operating characteristic curve(ROC)is the graphical plot that plots true positive rate verses the false positive rate .If the Area under the ROC curve (AUC) is large then the model has high discriminating power. It tells that how good the fitted model can separatethe observations being tested into those with and without the Jaundice in question.

Results& Discusssion:

Table 1

Accuracy Train Data	Accuracy Test Data	Precision	Recall	F1-score
77.61%	74%	79 %	75%	77%

Cross Validation:

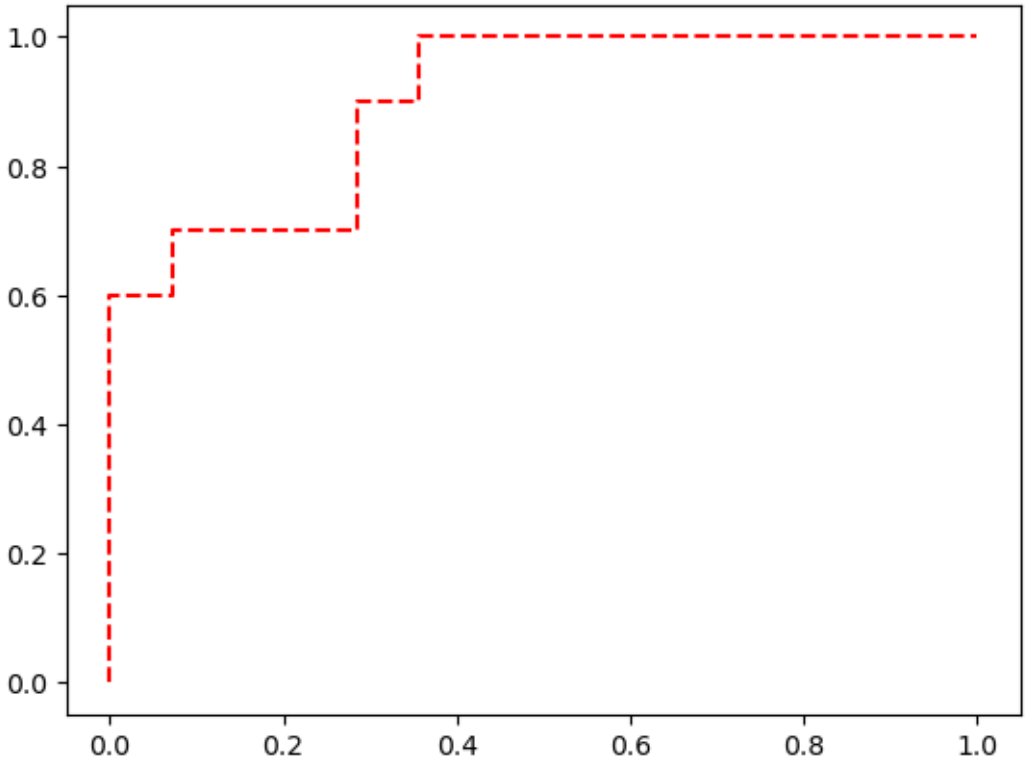
We have used 10- fold cross validation technique in Logistic regression model. This can reduces bias by giving each observation equal chance for training and once for testing. By using 10- fold cross validation at each iteration the accuracy of model obtained as follows.

Table 2

Iteration	Accuracy
1	0.78571429
2	0.78571429
3	0.78571429
4	0.71428571
5	0.69230769
6	0.53846154
7	0.69230769
8	0.53846154
9	0.76923077
10	0.84615385
Mean Accuracy	0.714835

ROC curve:

By using ROC curve the we found that area under curve (AUC) for our model is AUC=0.8999



By using Logistic Regression Model for prediction of Neonatal jaundice we observe that accuracy of our model on train data is 77.61% and for test data set is 74%. That is by using our model is built using our train data if we test it on remaining portion of data that is test data set then the 74% of observations are correctly classified. By usingcross validation technique every observation has equal chance of training and once goes for testing this reduces bias. And we calculate accuracy based on 10 fold cross validation where mean accuracy is 0.714835 that is 71.48%.While Precision is 79% that is by using logistic regression neonates who have actually jaundice of which79% are correctly predicted as these neonates had jaundice , Recall 75% and F1 Score is 77%. By using ROC curve we observe that for Logistic regression model in prediction of neonatal jaundice area under curve(AUC) is 0.8999 that is 89.99%. That is model has higher discriminating power in prediction of neonatal jaundice. This information might be used in new sample to identify high risk of individuals for whom special interventions might be necessary.

Bibliography

1. M. M. Eldibany, K. F. Totonchi, N. J. Joseph, and D. Rhone(1999): "Usefulness of certain red blood cell indices in diagnosing and differentiating thalassemia trait from iron- deficiency anemia"; *American Journal of Clinical Pathology. Vol.111(5) pp676–682.*
2. Rennie J, Burman-Roy S, Murphy S. Neonatal jaundice: summary of NICE guidance. *BMJ Br Med J. (*2010);**340**:c23409. doi: 10.1136/bmj.c2409. [PubMed] [CrossRef] [Google Scholar]
3. Sharareh R. Niakan Kalhori, Mahshid Nasehi, Xiao-Jun Zeng(2010): "A Logistic Regression Model to Predict High Risk Patients to Fail in Tuberculosis Treatment Course Completion"; *IAENG International Journal of Applied Mathematics, Vol.40(2), pp.1-8*
4. Joseph H Chou (2020): "Predictive Models for Neonatal Follow-Up Serum Bilirubin: Model Development and Validation" ; JMIR publications. doi: 10.2196/21222[Pubmed] central
5. Fouad H. Awad, Murtadha M. Hamad,Laith Alzubaidi (2023): "Robust Classification and Detection of Big Medical Data Using Advanced Parallel *K*-Means Clustering, YOLOv4, and Logistic Regression" doi: 10.3390/life13030691
6. R. D. Pati,l Dr.R.R. Kumbhar, Dr. S.V. Kakade(2023): "Study Of Classification Techniques (LogisticRegression, Support Vector Machine And Linear Discriminant Analysis) In Prediction Of Prevalence Of Heart Disease"Prarup Publication, Kolhapur; Statistics and Data Science;ISBN - 978-81-956739-9-5
7. Divyansh Khanna, Rohan Sahu, Veeky Baths, and Bharat Deshpande(2015):Comparative Study of Classification Techniques (SVM, Logistic Regression and Neural Networks) to Predict the Prevalence of Heart Disease International Journal of Machine Learning and Computing, Vol. 5