A Comparison of Logistic Regression and Linear Discriminant Analysis in Predicting of Female Students Attrition from School in Bangladesh

Mohammad Nayeem Hasan
Department of Statistics
Shahjalal University of Science & Technology Sylhet, Bangladesh
nayeem5847@gmail.com

IEEE
Advancing Technology for Humanity

INTRODUCTION

- Classifying or predicting the binary outcome variable, some methods are available such as linear discriminant (LD) analysis, logistic regression (LR) analysis, different data mining methods like a decision tree, Support vector machine, and artificial neural network models

- Several studies have been conducted to compare different classified models on different fields

- From that, many studies specified that linear discriminant analysis is more powerful for classifying a variable into several groups, where another confirmed logistic regression is useful and some concluded that both models had a similar performance

IEEE

OBJECTIVE

To compare the predictive accuracy and also the classification accuracy of the logistic regression model and linear discriminant model, using real data of school attrition of female students in Bangladesh

IEEE

METHODOLOGY (Study Design)

- Data were used from the Multiple Indicator Cluster Survey (MICS)
- It is based on a sample of 51895 households (43474 rural, 8421 urban) interviewed with a response rate of 98.5%
- It provides a comprehensive picture of children and women in the seven divisions (Dhaka, Chittagong, Sylhet, Rajshahi, Rangpur, Barisal, Khulna) in Bangladesh
- Women were aged between 15-49 years
- A total of 4800 women generated as a subsample age between 15-17 years

IEEE

METHODOLOGY (Data Mining Approach)

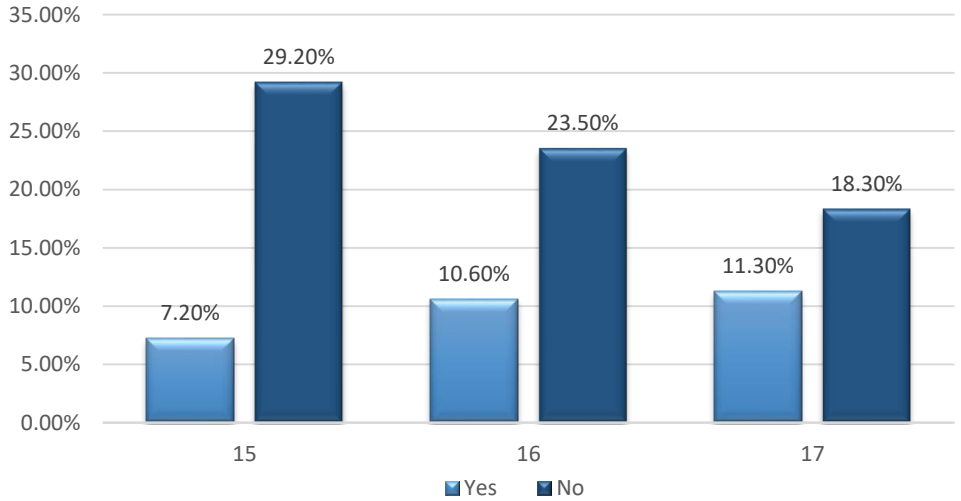
- The predictive accuracy for both models was estimated by the area under the receiver operating characteristics (AUROC) curves
- Higher AUROC indicated a better-predicted accuracy of the models
- The overall classification accuracy for both models was determined by comparing the predicted values with the actual events,

$$Accuracy = \frac{True\ Positive + True\ Negative}{Total}$$

IEEE

RESULTS

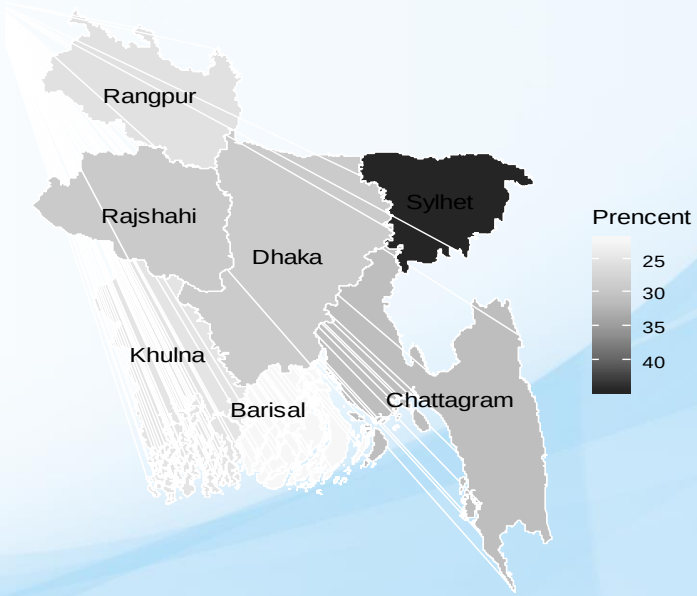
- Among 15-year-old women, 7.20% were out of school and 29.20% of women were attending school. Similarly, the percentage of 16 and 17-year-old women were 10.60% and 11.30%, respectively



Percentage of women by age and attrition

RESULTS

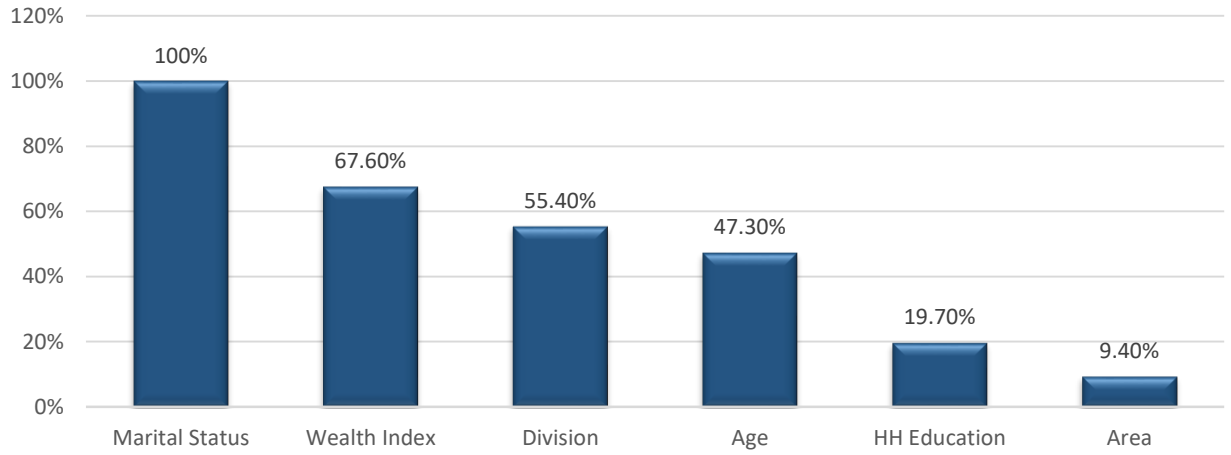
The Sylhet division had the highest attrition (44.47%) among girls age between 15-17 and the lowest percentage was recorded in Barisal division (22.18%)



Division wise School Attrition

RESULTS

- Variable importance measure marital status, wealth index, division, the age of women and household education were the first five effective factors on school attrition, respectively.



Factors affecting on attrition in the order of their importance

RESULTS

TEST OF AUROC FOR LR AND LD

Model	Sensitivity (%)	1-Specificity (%)	AUROC
LR	45.81%	91.60%	80.63%
LD	46.81%	91.01%	80.57%

The AUROC curve for logistic regression was 80.63%, while it was 80.57% for the linear discriminant ($p<0.001$).

IEEE

RESULTS

TEST OF CLASSIFICATION ACCURACY

Model		Yes	No	Accuracy (%)
LR	Yes	638	283	78.33%
	No	757	3122	
LD	Yes	661	304	78.38%
	No	734	3101	

The logistic regression and linear discriminant models classified 78.33% and 78.38% respectively, in-school and out-of-school correctly

IEEE

KEY FINDINGS

- In this study, overall classification rates were good for both and could be helpful in either classification of the class of attrition
- LR slightly exceeds LD in the correct classification rate and LR performed better than LD
- When taking into account sensitivity, specificity and AUROC, the differences in the AUROC were negligible
- From the result we can conclude that, discriminant analysis is preferred for classification and logistics regression should be used for prediction

IEEE

CONCLUSIONS

- In this paper we have compared linear discriminant analysis and logistic regression analysis using the overall classification accuracy

- By comparing the power of their prediction in terms of sensitivity and specificity and check the AUROC to compare the overall accuracy

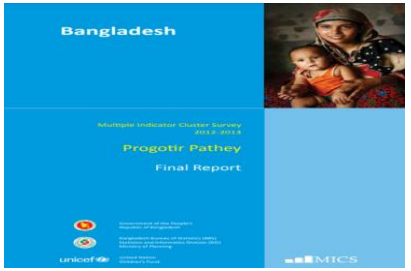
- From the predicting performance and classification accuracy, we decided that the overall classified rates of both models were good and it could be helpful in predicting the school attrition

IEEE

Acknowledgement

unicef

MICS



IEEE


OBRIGADO
gracias
どうも
ARIGATO
hvala
DANK U
謝謝
takk
MERSI
merci
obrigado
danke schön
köszi
PALDIES
muchas gracias
ありがとう
TEŞEKKÜR EDERIM
MOLTE GRAZIE
GO RAIBH MAITH AGAT
danke
grazas
GRAZZI
THANKS
THANK YOU
благодаря
TAK
どうも
quan
TAK
asante
muchas gracias
vielen dank
PALDIES
OBRIGADO
mesi
DANK U
danke
감사합니다
köszi
благодаря
grazie
DZIEKI
TACK
Gràcies
MULTUMESC
多謝
TEŞEKKÜR EDERIM
muchas gracias
obrigado
спасибо
NA GODE