Identifying Best Classification Algorithm

Prabhat Ale
22

Dataset Generation

employment_status:
1. Employed
2. Self Employed
3. Unemployed

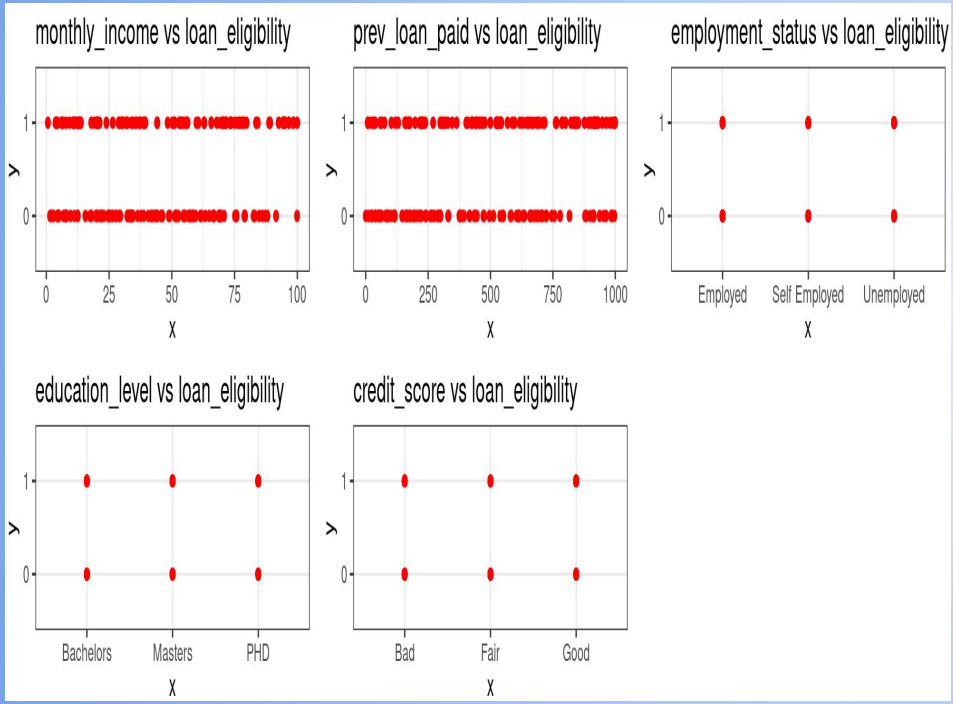
education_level:
1. Bachelors
2. Masters
3. PhD

loan_eligibility:
0 -> not eligible
1 -> eligible

credit_score:
1. Bad
2. Fair
3. Good

```r
set.seed(22)
## Generating one binary/dichotomous dependent variable
loan_eligibility <- as.factor(sample(x = 0:1, size = 200, replace = TRUE))
## Generating two continuous independent variables
monthly_income <- runif(n = 200,min = 0, max = 100)
prev_loan_paid <- runif(n = 200, min = 0, max = 1000)
## Generate three categorical independent variables
employment_status <- as.factor(sample(x = c("Employed", "Self Employed", "Unemployed"), size = 200, replace = TRUE))
education_level <- as.factor(sample(x = c("Bachelors", "Masters", "PHD"), size = 200, replace = TRUE))
credit_score <- as.factor(sample(x = c("Bad", "Fair", "Good"), size = 200, replace = TRUE))
```

Bivariate Logistic Regression Analysis



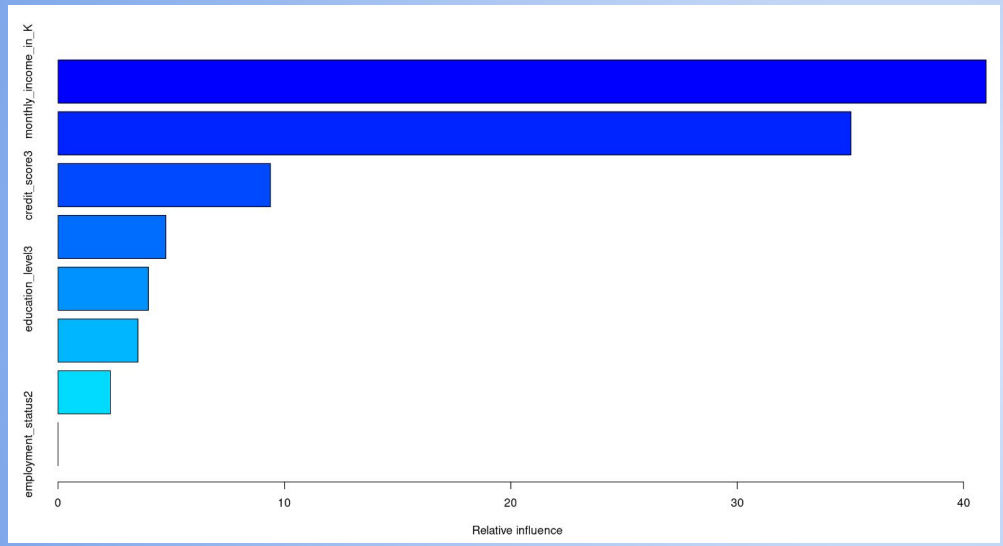
Independent Variable	Accuracy	Sensitivity	Specificity	Confusion Matrix
monthly_income_in_K	0.5676	0.6500	0.4706	0 1 0 13 9 1 7 8
prev_loan_paid_in_K	0.6216	0.7000	0.5294	0 1 0 14 8 1 6 9
employment_status	0.3784	0.5500	0.1765	0 1 0 11 14 1 9 3
education_level	0.5405	1.0000	0.000	0 1 0 20 17 1 0 0
credit_score	0.5946	0.4500	0.7647	0 1 0 9 4 1 11 13

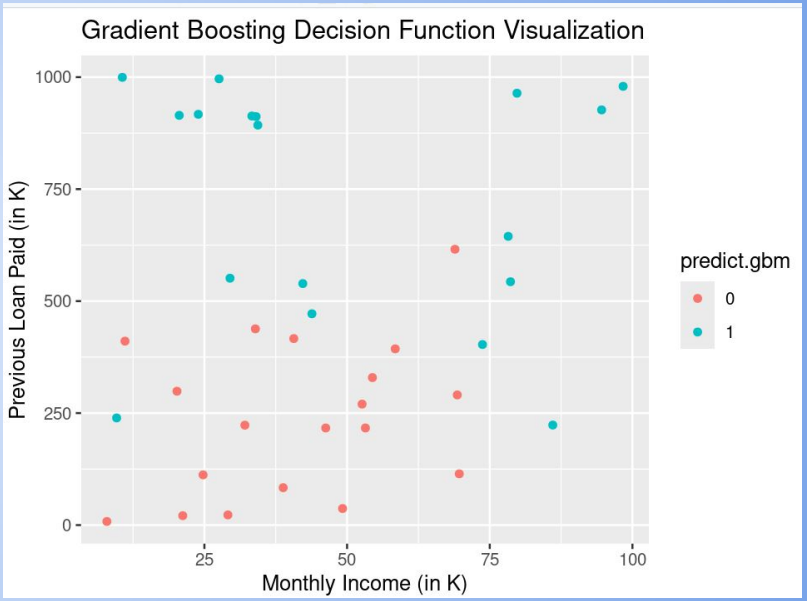
Model Experimentation: Best ML Algorithm For Loan Eligibility Prediction

Algorithms	Accuracy	Sensitivity	Specificity	Confusion Matrix
Multivariate Logistic Regression	0.5676	0.6500	0.4706	0 1 0 13 9 1 7 8
Naive Baye's Algorithm	0.5405	0.6000	0.4706	0 1 0 12 9 1 8 8
Linear SVM	0.5676	0.6000	0.5294	0 1 0 12 8 1 8 9
Decision Tree	0.5405	1.0000	0.000	0 1 0 20 17 1 0 0
Bagging	0.3514	0.4000	0.2941	0 1 0 8 12 1 12 5
Random Forest	0.3784	0.4500	0.2941	0 1 0 9 12 1 11 5
Tuned Random Forest	0.4054	0.4000	0.4118	0 1 0 8 10 1 12 7
Gradient Boosting Algorithm	0.5946	0.6000	0.5882	0 1 0 12 7 1 8 10

	var	rel.inf
monthly_income_in_K		40.994400
prev_loan_paid_in_K		35.024061
credit_score3		9.380856
credit_score2		4.763797
education_level3		3.989496
education_level2		3.532163
employment_status3		2.315227
employment_status2		0.000000

Visualizing Key Features Impacting Loan Approval





Conclusion

- **Bivariate analysis indicates that previous loan payment history is a significant predictor of loan eligibility.**
- **Multivariate model evaluation reveals Gradient Boosting outperforms other models with an accuracy of approximately 0.59.**
- **Monthly income and previous loan payments are pivotal factors in determining loan eligibility.**
- **A higher monthly income coupled with higher previous loan payments correlates with a greater likelihood of loan approval.**
- **Despite other features, previous loan payments alone can effectively model loan eligibility, yielding an accuracy of around 0.62.**

Any Questions?

Thank You