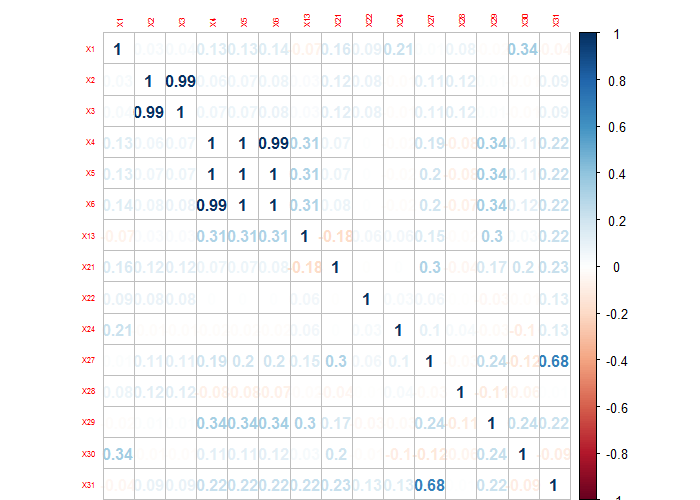
**AIM: -**

Predict the interest rate on the loan given pertaining parameters related to loan. Build machine learning/statistical models in R to predict the interest rate assigned to a loan.

**INSIGHTS: -**

* **Following is the correlation plot.**



* Data which is highly correlated provides same information to the model while inflating data accuracy. Hence it is very important to remove highly correlated data.
* We can see that variables “X4”,”X5”, “X6” are highly correlated with each other, also “X27” & “X31” and “X2”& “X3”are correlated to each other. So we remove variables “X2”, “X4” and “X6” and “X31”.

**MODEL BUILDING:-**

Models used for predicting the interest rate of load data is as follows:

* CART MODEL
* REGRESSION TREES (Rpart)

**CART Model:-**

* The accuracy of CART model is 93.74%.
* Variable importance for CART model is as follows:

|  |  |  |
| --- | --- | --- |
| Rank | variable | Overall |
| 1 | X9 | 3.49093 |
| 2 | X8 | 3.381349 |
| 3 | date\_iss | 0.665197 |
| 4 | X7 | 0.507598 |
| 5 | X30 | 0.355183 |
| 6 | X14 | 0.116701 |
| 7 | X24 | 0.06027 |
| 8 | month\_iss | 0.020132 |
| 9 | X6 | 0.012127 |
| 10 | X28 | 0.011219 |
| 11 | month\_earl\_cr | 0 |
| 11 | X11 | 0 |
| 11 | X12 | 0 |
| 11 | X13 | 0 |
| 11 | X17 | 0 |
| 11 | X20 | 0 |
| 11 | X21 | 0 |
| 11 | X22 | 0 |
| 11 | X27 | 0 |
| 11 | X29 | 0 |
| 11 | X31 | 0 |
| 11 | X32 | 0 |
| 11 | year\_earl\_cr | 0 |

* As from above figure we can extract that X9 is the factor that contributes most for the interest rate prediction and so on. But the data in X9 & X8 is same so we have removed variable X8.

**REGRESSION Tree Model:-**

* The accuracy using this model is 94.27%.
* Variable importance from this model is as follows:

|  |  |  |
| --- | --- | --- |
| Rank | variable | Overall |
| 1 | X9 | 3.597165 |
| 2 | X8 | 3.488138 |
| 3 | date\_iss | 0.679095 |
| 4 | X7 | 0.474328 |
| 5 | X30 | 0.370468 |
| 6 | X14 | 0.093834 |
| 7 | X24 | 0.076829 |
| 8 | month\_iss | 0.020958 |
| 9 | X6 | 0.018653 |
| 10 | X28 | 0.017819 |
| 11 | X11 | 0 |
| 11 | X12 | 0 |
| 11 | X13 | 0 |
| 11 | X17 | 0 |
| 11 | X20 | 0 |
| 11 | X21 | 0 |
| 11 | X22 | 0 |
| 11 | X27 | 0 |
| 11 | X29 | 0 |
| 11 | X31 | 0 |
| 11 | X32 | 0 |
| 11 | month\_earl\_cr | 0 |
| 11 | year\_earl\_cr | 0 |

**Analysis of text:**

From the variable X10, X16, X17:

Finally, we analysis peoples who are taking loans and reason behind the loans. After doing text analysis on variables X10-Employer job title and X16-Reason for loan and X17- Loan category, as provided by borrower we get following output:

X10:- From this wordcloud we can see that most people who applied for loan are managers, directors, teachers etc. it is beneficial to know people from which field and posts are getting loans so that we can focus on them and try and get focus on people from other categories as well.



X16:- From this wordcloud it shows reasons behind taking loan. We can see that most of people who took loan did it to return borrowed money or pay their debt and loan and bills.



X17: From this wordcloud it shows the loan category for which the people take loans and mostly people take foe debt consolidation which means: Debt consolidation involves taking out a new loan to pay off a number of other debts.

