**Data Science Challenge**

The data given is of credit records of individuals with certain attributes. Please go through following to understand the variables involved:  
1. serial number : unique identification key

2. **account\_info :** Categorized details of existing accounts of the individuals. The balance of money in account provided is stated by this variable

3. purpose: This variable signifies why the loan was taken  
A40 signifies that the loan is taken to buy a new car  
A46 signifies that the loan is taken for education  
A47 signifies that the loan is taken for vacation  
A48 signifies that the loan is taken for re skilling  
A49 signifies that the loan is taken for business and establishment  
A410 signifies other purposes

4. savings\_account: This variable signifies details of the amount present in savings account of the individual:  
A61 signifies that less than 100 units (excluding 100) of currency is present  
A62 signifies that greater than 100 units (including 100) and less than 500 (excluding 500) units of currency is present  
A63 signifies that greater than 500 (including 500) and less than 1000 (excluding 1000) units of currency is present.  
A64 signifies that greater than 1000 (including 1000) units of currency is present.  
A65 signifies that no savings account details is present on record

5. \*\*employment\_st: Catergorical variable that signifies the employment status of everyone who has been alloted loans  
A71 signifies that the individual is unemployed  
A72 signifies that the individual has been employed for less than a year  
A73 signifies that the individual has been employed for more than a year but less than four years  
A74 signifies that the individual has been employed more than four years but less than seven years  
A75 signifies that the individual has been employed for more than seven years

6. gurantors: Categorical variable which signifies if any other individual is involved with an individual loan case  
A101 signifies that only a single individual is involved in the loan application  
A102 signifies that one or more co-applicant is present in the loan application  
A103 signifies that guarantor are present.

1. resident\_since: Numerical variable that signifies for how many years the applicant has been a resident

8. property\_type: This qualitative variable defines the property holding information of the individual  
A121 signifies that the individual holds real estate property  
A122 signifies that the individual holds a building society savings agreement or life insurance  
A123 signifies that the individual holds cars or other properties  
A124 signifies that property information is not available

9. age: Numerical variable that signifies age in number of years

10. installment\_type: This variable signifies other installment types taken  
A141 signifies installment to bank  
A142 signifies installment to outlets or stores  
A143 signifies that no information is present

11. housing\_type: This is a categorical variable that signifies which type of housing does a applicant have.  
A151 signifies that the housing is on rent  
A152 signifies that the housing is owned by the applicant  
A153 signifies that no loan amount is present on the housing and there is no expense for the housing)

12. credits\_no: Numerical variable for number of credits taken by the person

13. job\_type: Signifies the employment status of the person  
A171 signifies that the individual is unemployed or unskilled and is a non-resident  
A172 signifies that the individual is unskilled but is a resident  
A173 signifies that the individual is a skilled employee or official  
A174 signifies that the individual is involved in management or is self-employed or a  
highly qualified employee or officer

14. liables: Signifies number of persons dependent on the applicant

15. telephone: Signifies if the individual has a telephone or not  
A191 signifies that no telephonic records are present  
A192 signifies that a telephone is registered with the customer’s name

16.foreigner: Signifies if the individual is a foreigner or not (considering the country of residence of the bank)  
A201 signifies that the individual is a foreigner  
A202 signifies that the individual is a resident  
Objective of the problem: The objective of the problem is to predict the values of credit\_amount variable as per serial number variable. Please view the sample submissions file for better understanding. The solution must be presented in the form of a csv with predicted values of the response variable credit\_amount along with it’s corresponding serial number.  
*Evaluation Metric :* Normalized root mean squared error. The score is calculated by (1-rmse/normalization factor)\*100.  
Submission Limit: . Please note that individual submission limits 50

The objective of the problem is to predict the values of credit\_amount variable as per serial number variable. Please view the sample submissions file for better understanding. The solution must be presented in the form of a csv with predicted values of the response variable credit\_amount along with it’s corresponding serial number.

**Evaluation Algorithm**

Root Mean Square Error (RMSE)