predict the customer lifetime value (Regression Problem)

**Problem Statement:**

Vahanbima is one of the leading insurance companies in India. It provides motor vehicle insurances at best prices with 24/7 claim settlement.  It offers different types of policies for  both personal and commercial vehicles. It has established its brand across different region I India.   
  
Around 90% of the businesses today use personalized services. The company wants to launch different personalized experience programs for customers of VahanBima. The personalized experience can be dedicated resources for claim settlement, different kinds of services at doorstep, etc. Inorder to do so, they would like to segment the customers into different tiers based on their customer lifetime value (CLTV).

APPROACH:

As the problem is to predict the customer lifetime value,the target “CLTV” is given as numerical value which is continuous.For the numerical target attribute we should consider it as a Regression problem.so Regression modules can work on this problem

Firstly I read all the test and the train data sets and I have done preprocessing paralelly on test and the train data

Steps done in preprocessing:

1)checking the Missing Values

2)changing the datatypes to appropriate datatypes.

3)Doing One-Hot encoding on categorical attributes and standardization on Numerical attributes

4)splitting the train data into train and validation data

Error Metrics for the Regression are:

a)RMSE(Root Mean Squared Error)

b)MSE(Mean Squared Error)

c)R2\_score

but in this problem we have to see the evaluate metric is R2\_score.

5)Built the model like Decision Tree Regressor with Hyperparameter tuning.

6)Built the model Random Forest Regressor with Hyperparameter Tuning

7)At last done the XG boosting technique.

From the above models the Random Forest Regressor is giving the best 0.083 of R2\_score