**UCI Portuguese Bank Marketing Data Set**

We first checked for nulls in the data as a part of our data cleaning. Now, as given, the data has categorical variables. Hence, we use Label Encoding to convert the categorical data into columns suitable for regression. We followed this up with three approaches to perform classification:

1. Naïve-Bayes Classifier
2. Logistic Regression
3. K-Nearest Neighbours

Following are the results for the same:

1. **Naïve – Bayes Classifier**:

Confusion Matrix for train data:

[[33214 3334]   
 [ 2248 2392]]

Overall accuracy for train data: 0.864475   
Accuracy for Class 1 (No): 0.908777   
Accuracy for Class 2 (Yes): 0.515517

Confusion Matrix for test data:

[[3327 341]

[ 214 237]]

Overall accuracy for test data: 0.865259

Accuracy for Class 1 (No): 0.907034

Accuracy for Class 2 (Yes): 0.525499







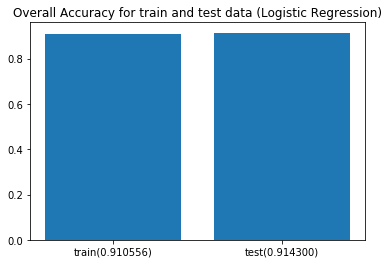
**2. Logistic Regression**:

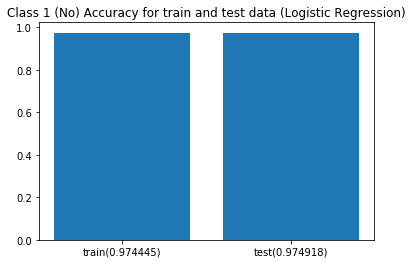
Confusion Matrix for LR on train data:   
[[35614 934]   
[ 2750 1890]]

Overall accuracy of LR or train data: 0.910556   
LR Accuracy for Class 1 (No): 0.974445   
LR Accuracy for Class 2 (Yes): 0.407328

Confusion Matrix for LR on test data:   
[[3576 92]   
 [ 261 190]]

Overall accuracy of LR on test data: 0.914300   
LR Accuracy for Class 1 (No): 0.974918   
LR Accuracy for Class 2 (Yes): 0.421286







**3. K-Nearest Neighbours**:

For train data:

Overall accuracy for neighbours n = 1 is 1.000000

Overall accuracy for neighbours n = 2 is 0.943794

Overall accuracy for neighbours n = 3 is 0.941973

Overall accuracy for neighbours n = 4 is 0.930732

Overall accuracy for neighbours n = 5 is 0.931388

For test data:

Overall accuracy for neighbours n = 1 is 1.000000

Overall accuracy for neighbours n = 2 is 0.943433

Overall accuracy for neighbours n = 3 is 0.939306

Overall accuracy for neighbours n = 4 is 0.932751

Overall accuracy for neighbours n = 5 is 0.931051

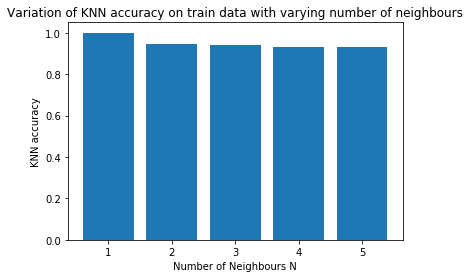
#Validation Test

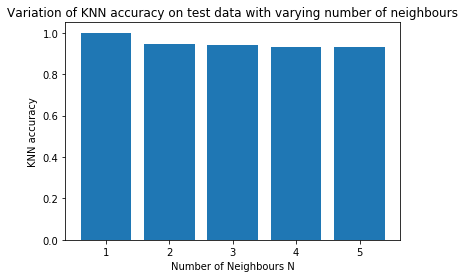
KNN accuracy for Class 1(No) on train data is 1.000000

KNN accuracy for Class 1(No) on test data is 1.000000

KNN accuracy for Class 2(Yes) on train data is 1.000000

KNN accuracy for Class 2(Yes) on test data is 1.000000









**Observation/Conclusion**

Based on the above data, we can easily observe that out of the approaches taken above, **K-nearest neighbours Classifier with K = 2** should be used for predicting customers who will/ will not be making a term deposit in the bank. This will bring down the Marketing Costs and help in increasing the revenue of the bank.