

Homework 5

German Credit Dataset

Logistic Regression

Vidit Kalani

MSCS 6520 Business Analytics

Spring 2018

Prof. RJ Nowling

These are my of the features for study in German Credit dataset

- Duration,
- Amount,
- InstallmentRatePercentage,
- Age,
- NumberExistingCredits,
- NumberPeopleMaintenance,
- Telephone,
- ForeignWorker,
- Class,
- CheckingAccountStatus.lt.0,
- CheckingAccountStatus.0.to.200,
- CheckingAccountStatus.gt.200,
- CheckingAccountStatus.none,
- CreditHistory.NoCredit.AllPaid,
- CreditHistory.ThisBank.AllPaid,
- CreditHistory.PaidDuly,
- CreditHistory.Delay,
- CreditHistory.Critical,
- SavingsAccountBonds.lt.100,
- SavingsAccountBonds.100.to.500,
- SavingsAccountBonds.500.to.1000,
- SavingsAccountBonds.gt.1000,
- SavingsAccountBonds.Unknown,
- EmploymentDuration.lt.1,
- EmploymentDuration.1.to.4,
- EmploymentDuration.4.to.7,
- EmploymentDuration.gt.7,
- EmploymentDuration.Unemployed,
- Personal.Male.Divorced.Seperated,
- Personal.Female.NotSingle,
- Personal.Male.Single,
- Personal.Male.Married.Widowed,
- Personal.Female.Single,
- Property.RealEstate,
- Property.Insurance,
- Property.CarOther,
- Property.Unknown,
- Housing.Rent,
- Housing.Own,
- Housing.ForFree

Round: 1

Features	LR Accuracies	kNN Accuracies
For All features	0.73	None
Duration + Amount	0.715	0.70
Duration + InstallmentRatePercentage	0.71	0.715
Duration + Age	0.72	0.715
Duration + NumberExistingCredits	0.71	0.725
Duration + NumberPeopleMaintenance	0.715	0.735
Duration + Telephone	0.725	0.725
Duration + ForeignWorker	0.715	0.72
Duration + CheckingAccountStatus	0.705	0.71
Duration + CreditHistory	0.725	0.71
Duration + SavingsAccountBonds	0.75	0.72
Duration + EmploymentDuratio	0.72	0.73
Duration + Personal	0.715	0.73
Duration + Property	0.73	0.70
Duration + Housing	0.72	0.7

Round: 2

Features	LR Accuracies	kNN Accuracies
Duration + SavingsAccountBonds + Amount	0.75	0.71
Duration + SavingsAccountBonds + InstallmentRatePercentage	0.765	0.715
Duration + SavingsAccountBonds + Age	0.755	0.73
Duration + SavingsAccountBonds + NumberExistingCredits	0.725	0.74
Duration + SavingsAccountBonds + NumberPeopleMaintenance	0.75	0.725
Duration + SavingsAccountBonds + Telephone	0.73	0.74
Duration + SavingsAccountBonds + ForeignWorker	0.75	0.71
Duration + SavingsAccountBonds + CheckingAccountStatus	0.695	0.735
Duration + SavingsAccountBonds + CreditHistory	0.72	0.73
Duration + SavingsAccountBonds + EmploymentDuratio	0.76	0.715
Duration + SavingsAccountBonds + Personal	0.75	0.715
Duration + SavingsAccountBonds + Property	0.725	0.71
Duration + SavingsAccountBonds + Housing	0.75	0.765

Round: 3

Features	LR Accuracies	Knn\Accuracies
Duration + SavingsAccountBonds + InstallmentRatePercentage + Amount	0.765	0.72
Duration + SavingsAccountBonds + InstallmentRatePercentage + Age	0.765	0.75
Duration + SavingsAccountBonds + InstallmentRatePercentage + NumberExistingCredits	0.755	0.735
Duration + SavingsAccountBonds + InstallmentRatePercentage + NumberPeopleMaintenance	0.765	0.73
Duration + SavingsAccountBonds + InstallmentRatePercentage + Telephone	0.75	0.775
Duration + SavingsAccountBonds + InstallmentRatePercentage + ForeignWorker	0.765	0.76
Duration + SavingsAccountBonds + InstallmentRatePercentage + CheckingAccountStatus	0.715	0.71
Duration + SavingsAccountBonds + InstallmentRatePercentage + CreditHistory	0.745	0.735
Duration + SavingsAccountBonds + InstallmentRatePercentage + EmploymentDuratio	0.765	0.735
Duration + SavingsAccountBonds + InstallmentRatePercentage + Personal	0.755	0.72
Duration + SavingsAccountBonds + InstallmentRatePercentage + Property	0.755	0.72
Duration + SavingsAccountBonds + InstallmentRatePercentage + Housing	0.765	0.73

Round: 4

Features	LR Accuracies	kNN Accuracies
Duration + SavingsAccountBonds + InstallmentRatePercentage + Amount + Age	0.765	0.735
Duration + SavingsAccountBonds + InstallmentRatePercentage + Amount + NumberExistingCredits	0.765	0.71
Duration + SavingsAccountBonds + InstallmentRatePercentage + Amount + NumberPeopleMaintenance	0.765	0.775
Duration + SavingsAccountBonds + InstallmentRatePercentage + Amount + Telephone	0.76	0.73
Duration + SavingsAccountBonds + InstallmentRatePercentage + Amount + ForeignWorker	0.765	0.74
Duration + SavingsAccountBonds + InstallmentRatePercentage + Amount + CheckingAccountStatus	0.715	0.665
Duration + SavingsAccountBonds + InstallmentRatePercentage + Amount + CreditHistory	0.745	0.705
Duration + SavingsAccountBonds + InstallmentRatePercentage + Amount + EmploymentDuratio	0.775	0.735
Duration + SavingsAccountBonds + InstallmentRatePercentage + Amount + Personal	0.765	0.705
Duration + SavingsAccountBonds + InstallmentRatePercentage + Amount + Property	0.77	0.74
Duration + SavingsAccountBonds + InstallmentRatePercentage + Amount + Housing	0.76	0.695

Logistic Regression performing

- Round 1:- Duration + SavingsAccountBonds feature are best accuracy.
- Round 2:- Duration + SavingsAccountBonds + InstallmentRatePercentage features are best accuracy.
- Round 3:- Duration + SavingsAccountBonds + InstallmentRatePercentage + Amount features are best accuracy.
- Round 4:- Duration + SavingsAccountBonds + InstallmentRatePercentage + Amount + EmploymentDuratio features are best accuracy.
- But I noticed that in Round 4 the most of features accuracy are same and all features in Round 4, the accuracy going up.
- So, I decided to stop at Round 4 because in Round 4, 4 out of 11 features accuracy are same(0.765).

Compare to LR Model and kNN Model

- In LR Model, I can start from selecting two features. And in kNN Model I can start with single feature.
- Round 1 in LR model and Round 2 in kNN model, the most of the all feature accuracy are nearly equal to same. In Round 1 in LR model, the high accuracy is Duration + SavingsAccountBonds. And In kNN model, the high accuracy is Duration + NumberPeopleMaintenance.
- In Round 2 in LR model and Round 3 in kNN model, the LR's all feature accuracy are going up. In Round 2 in LR model, the high accuracy is Duration + SavingsAccountBonds + InstallmentRatePercentage. And In kNN model, the high accuracy is Duration + NumberPeopleMaintenance + Housing.
- In Round 2 in LR model and Round 3 in kNN model, the LR's all feature accuracy are going up. In Round 2 in LR model, the high accuracy is Duration + SavingsAccountBonds + InstallmentRatePercentage + Amount. And In kNN model, the high accuracy is Duration + NumberPeopleMaintenance + Housing + Telephone.
- In Round 2 in LR model and Round 3 in kNN model, the LR's all feature accuracy are going up. In Round 2 in LR model, the high accuracy is Duration + SavingsAccountBonds + InstallmentRatePercentage + Amount + EmploymentDuratio. And In kNN model, the high accuracy is Duration + NumberPeopleMaintenance + Housing + Telephone + Age.
- I noticed that, in LR Model every round overall all accuracy is going up and after Round 3 some of the accuracy are stable, going up or going down. But in kNN Model every Rounds the accuracy frequency change and not stable.