# Credit Risk Prediction Using machine Learning

**Waruni Wijayasinghe**

Springboard Data Science Career Track

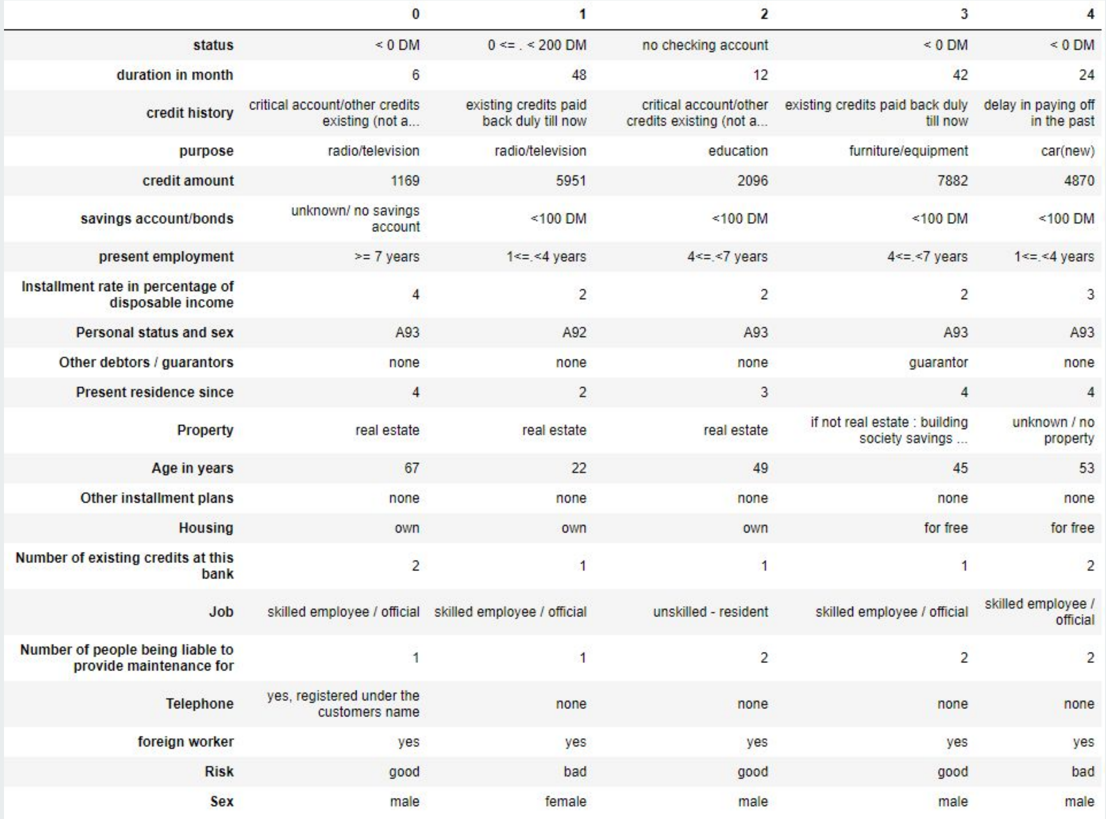
March 2022

# Business Problem

# Data

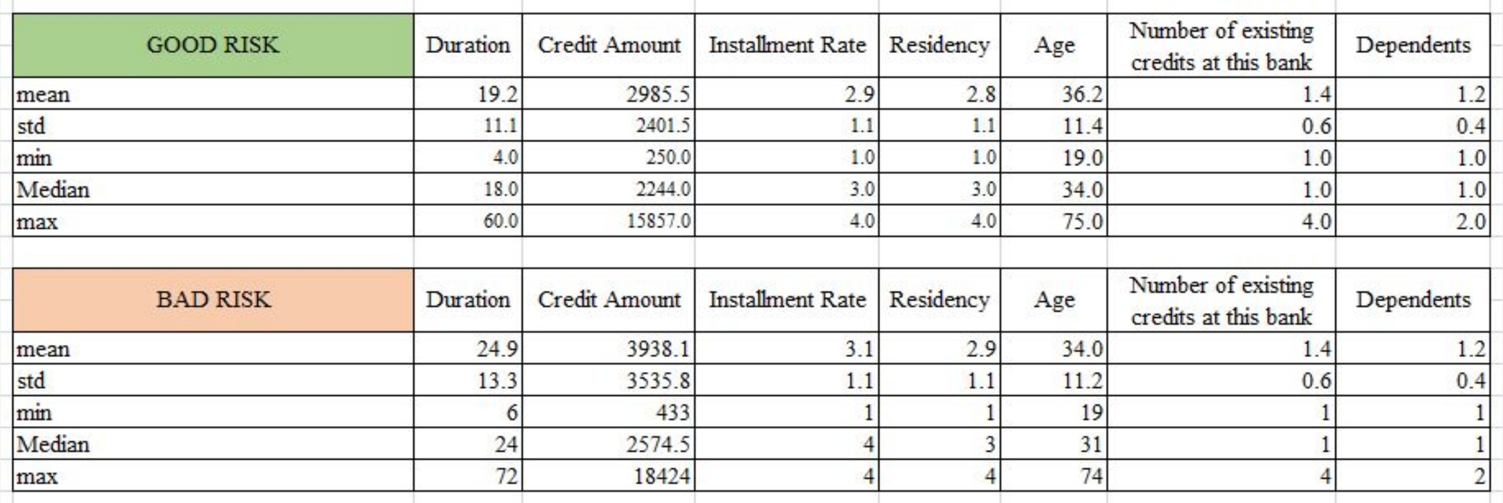
German Credit Risk Data

UCI Machine Learning Repository.



| | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| **status** | < 0 DM | 0 <= . < 200 DM | no checking account | < 0 DM | < 0 DM |
| **duration in month** | 6 | 48 | 12 | 42 | 24 |
| **credit history** | critical account/other credits existing (not a... | existing credits paid back duly till now | critical account/other credits existing (not a... | existing credits paid back duly till now | delay in paying off in the past |
| **purpose** | radio/television | radio/television | education | furniture/equipment | car(new) |
| **credit amount** | 1169 | 5951 | 2096 | 7882 | 4870 |
| **savings account/bonds** | unknown/ no savings account | <100 DM | <100 DM | <100 DM | <100 DM |
| **present employment** | >= 7 years | 1<=.<4 years | 4<=.<7 years | 4<=.<7 years | 1<=.<4 years |
| **Installment rate in percentage of disposable income** | 4 | 2 | 2 | 2 | 3 |
| **Personal status and sex** | A93 | A92 | A93 | A93 | A93 |
| **Other debtors / guarantors** | none | none | none | guarantor | none |
| **Present residence since** | 4 | 2 | 3 | 4 | 4 |
| **Property** | real estate | real estate | real estate | if not real estate : building society savings ... | unknown / no property |
| **Age in years** | 67 | 22 | 49 | 45 | 53 |
| **Other installment plans** | none | none | none | none | none |
| **Housing** | own | own | own | for free | for free |
| **Number of existing credits at this bank** | 2 | 1 | 1 | 1 | 2 |
| **Job** | skilled employee / official | skilled employee / official | unskilled - resident | skilled employee / official | skilled employee / official |
| **Number of people being liable to provide maintenance for** | 1 | 1 | 2 | 2 | 2 |
| **Telephone** | yes, registered under the customers name | none | none | none | none |
| **foreign worker** | yes | yes | yes | yes | yes |
| **Risk** | good | bad | good | good | bad |
| **Sex** | male | female | male | male | male |

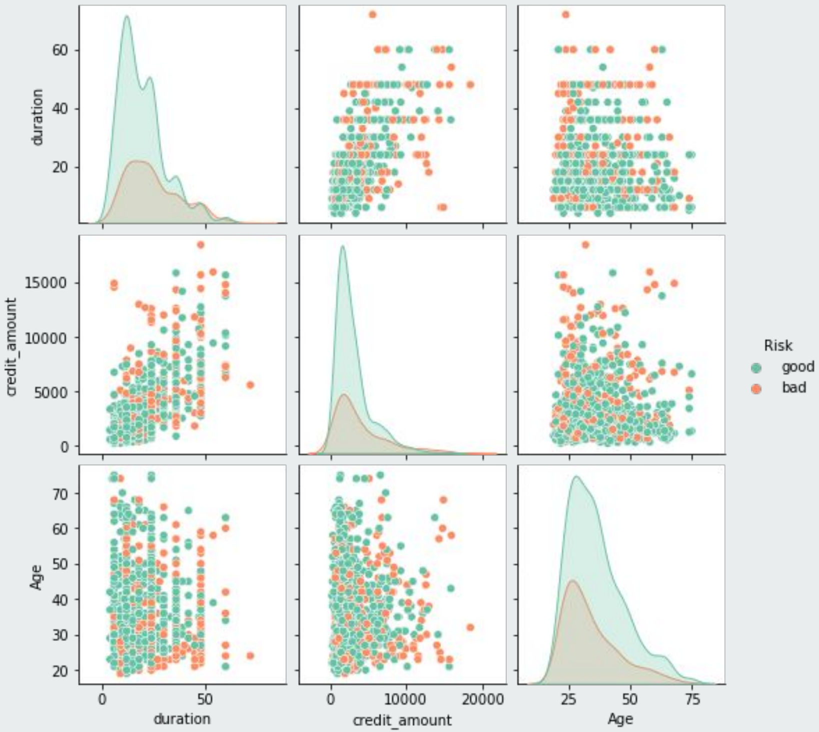
# Data Exploration



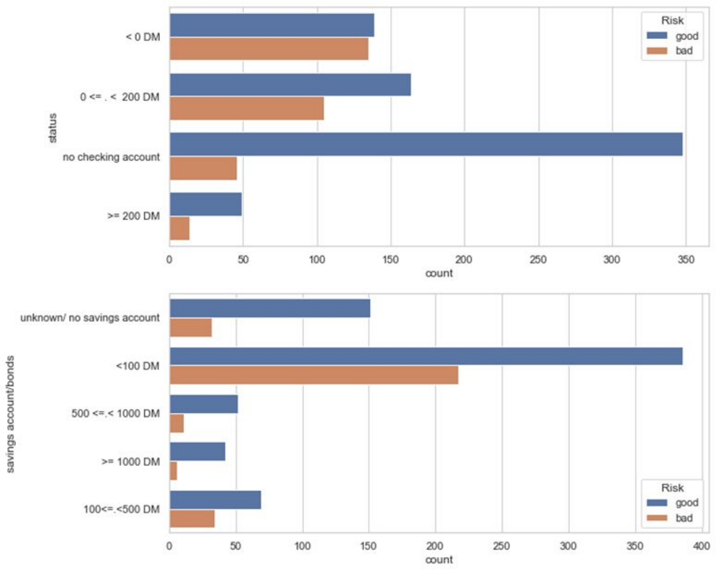
| GOOD RISK | Duration | Credit Amount | Installment Rate | Residency | Age | Number of existing credits at this bank | Dependents |
|---|---|---|---|---|---|---|---|
| mean | 19.2 | 2985.5 | 2.9 | 2.8 | 36.2 | 1.4 | 1.2 |
| std | 11.1 | 2401.5 | 1.1 | 1.1 | 11.4 | 0.6 | 0.4 |
| min | 4.0 | 250.0 | 1.0 | 1.0 | 19.0 | 1.0 | 1.0 |
| Median | 18.0 | 2244.0 | 3.0 | 3.0 | 34.0 | 1.0 | 1.0 |
| max | 60.0 | 15857.0 | 4.0 | 4.0 | 75.0 | 4.0 | 2.0 |

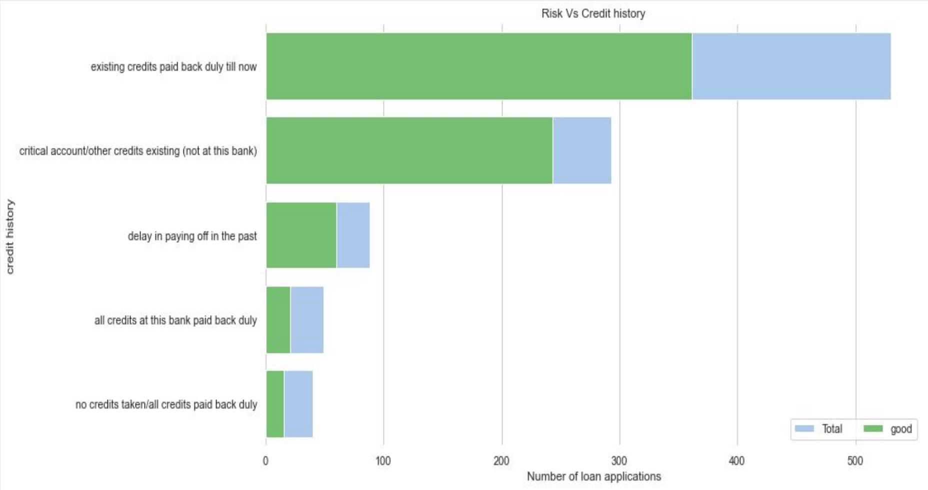
| BAD RISK | Duration | Credit Amount | Installment Rate | Residency | Age | Number of existing credits at this bank | Dependents |
|---|---|---|---|---|---|---|---|
| mean | 24.9 | 3938.1 | 3.1 | 2.9 | 34.0 | 1.4 | 1.2 |
| std | 13.3 | 3535.8 | 1.1 | 1.1 | 11.2 | 0.6 | 0.4 |
| min | 6 | 433 | 1 | 1 | 19 | 1 | 1 |
| Median | 24 | 2574.5 | 4 | 3 | 31 | 1 | 1 |
| max | 72 | 18424 | 4 | 4 | 74 | 4 | 2 |

# Data Exploration



# Data Exploration



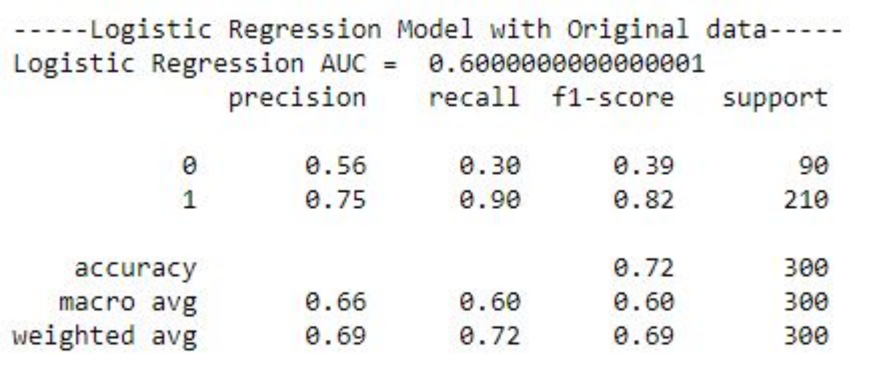


# Imbalance Data

- Random Undersampling
- Random Oversampling
- Synthetic Minority Oversampling (SMOTE)

| Resampling Techniques | F1 Score |
|---|---|
| **Original** | 0.82 |
| **Undersample** | 0.69 |
| **Oversample** | 0.68 |
| **SMOTE** | 0.79 |

# Baseline Model : Logistic Regression



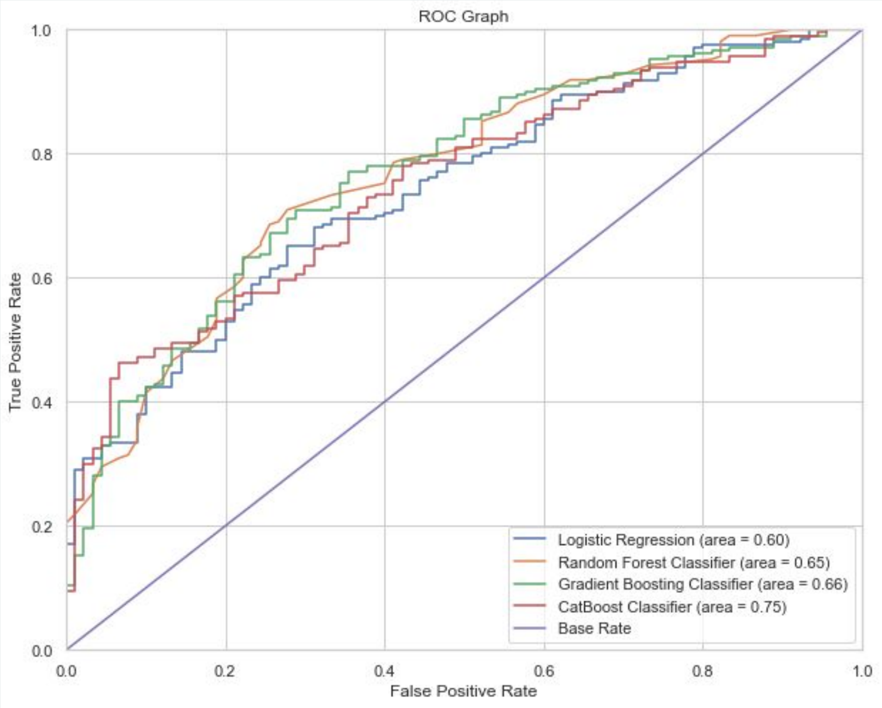
```
-----Logistic Regression Model with Original data-----
Logistic Regression AUC =  0.6000000000000001
              precision    recall  f1-score   support

           0       0.56      0.30      0.39        90
           1       0.75      0.90      0.82       210

    accuracy                           0.72       300
   macro avg       0.66      0.60      0.60       300
weighted avg       0.69      0.72      0.69       300
```

# Other Models

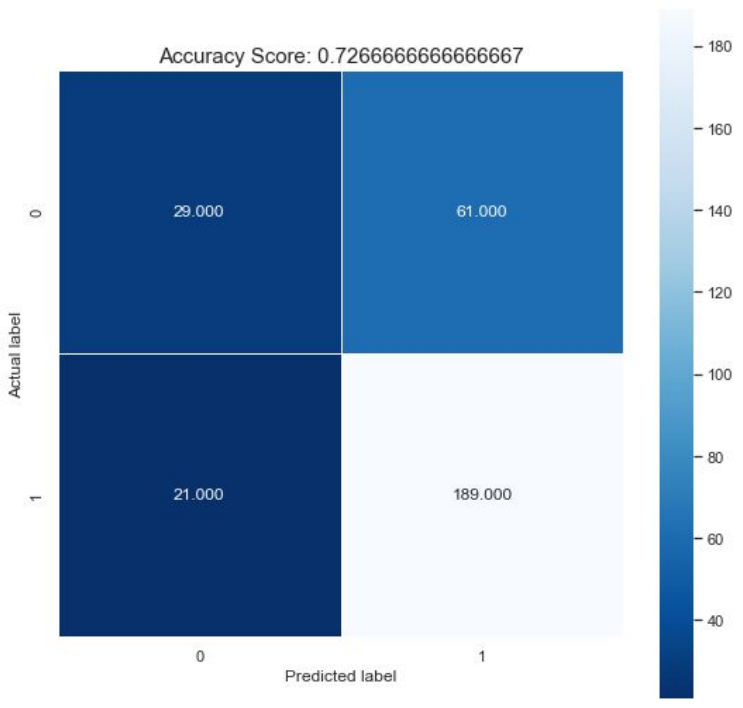
1. Random Forest Classifier
2. Gradient Boosting Classifier
3. CatBoost Algorithm

# Model Evaluation

**CATBOOST!**



# Model Prediction With CatBoost



# Questions or Comments?

# Thank You!