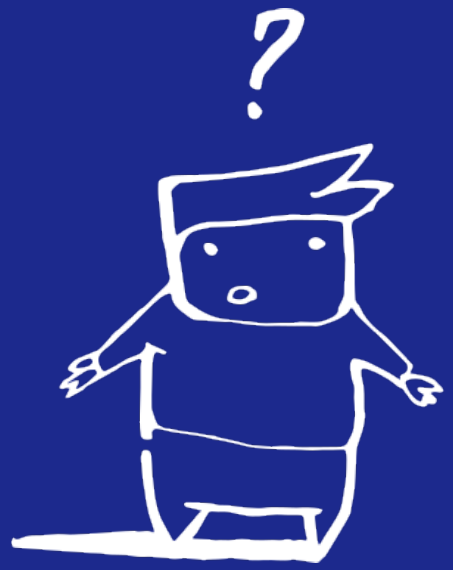
Cristian Valeria

Data Analytics Bootcamp | Final Project



KREDIT FATE

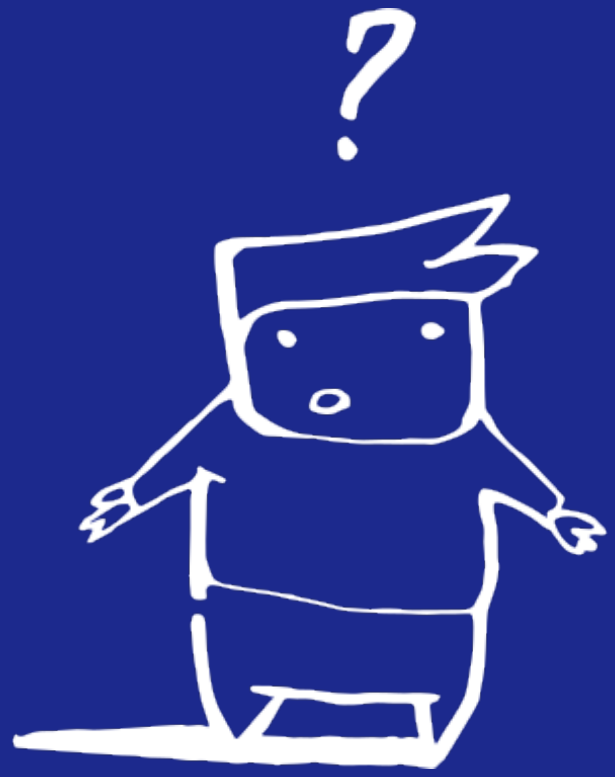
Can I get a credit on Germany?

The Problem

Get access to a credit is never something easy, and Germany is not an exception.

Thats why I want to know, given certain parameters, what is the **probability** to get a credit on Germany and what can I do to **improve** my chances.

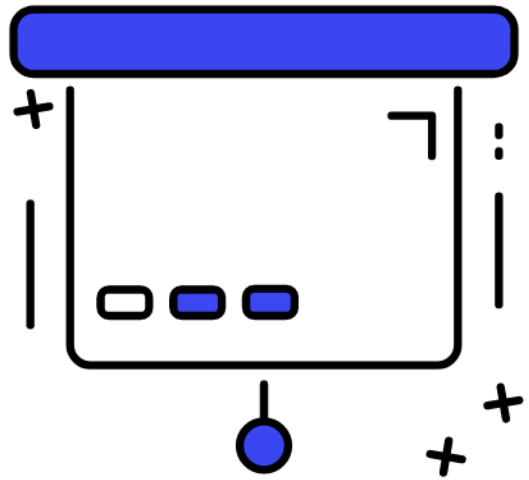
The idea is to build an interactive **app** that can predict if you will probably get a credit or not.



The DataSet

The Data Set was build by attribute from Professor Dr. Hans Hofmann's, and provided by UCI Machine Learning Repository

It contains **1000** applications. In witch the person obtained or not the credit.

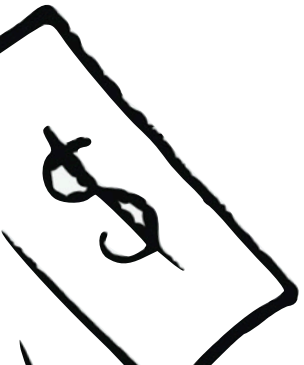


The DataSet

Variables

The Dataset contains **29** columns with the detail of every application, such as:

Age
Sex
Job Status
Savings
Credit amount
Credit History
Guarantors
Duration
Purpose



Some data processing

- **Drop:** 2 columns with more than 17% of missing values. not relevant columns
- **Dummify** categorical variables
- **y** = Creditability (obtain or not the credit)
- **y** imbalanced (70% Yes) —> RandomOverSampler

Modeling

LazyClassifier

Iteration of different models.

... What are the models that can apply better to this Classification problem ?

```
from lazypredict.Supervised import LazyClassifier

multiple_models = LazyClassifier(verbose=0, ignore_warnings=True, predictions=True)

models, predictions = multiple_models.fit(X_train, X_test, y_train, y_test)

100%|██████████| 29/29 [00:03<00:00,  8.14it/s]
```

```
models.sort_values(by='Accuracy', ascending=False)
```

Model	Accuracy	Balanced Accuracy	ROC AUC	F1 Score	Time Taken
ExtraTreesClassifier	0.87	0.87	0.87	0.87	0.32
RandomForestClassifier	0.86	0.86	0.86	0.86	0.45
LGBMClassifier	0.86	0.86	0.86	0.86	0.09
LabelSpreading	0.84	0.84	0.84	0.84	0.19
LabelPropagation	0.84	0.84	0.84	0.84	0.15
XGBClassifier	0.83	0.82	0.82	0.82	0.65
BaggingClassifier	0.82	0.82	0.82	0.82	0.07
DecisionTreeClassifier	0.81	0.81	0.81	0.81	0.03
ExtraTreeClassifier	0.78	0.78	0.78	0.78	0.02
NuSVC	0.77	0.77	0.77	0.77	0.16
SVC	0.76	0.76	0.76	0.76	0.14
AdaBoostClassifier	0.74	0.74	0.74	0.74	0.15
BernoulliNB	0.70	0.70	0.70	0.70	0.02
KNeighborsClassifier	0.69	0.69	0.69	0.69	0.03
LogisticRegression	0.69	0.69	0.69	0.69	0.05
CalibratedClassifierCV	0.69	0.69	0.69	0.69	0.38

Modeling

Applying models

Model	Version	Parameters
Logistic Regression	LR_1s	X_Scaled
K-NN Classifier	KNN_1	Standard Parameters K = 10
K-NN Classifier	KNN_2	RandomizedSearchCV K = 5
Random Forest	RF_1	Standar Parameters
Random Forest	RF_2	GridSearchCV: Best Parameters
Extra Trees Classifier	EXT_1	Standar Parameters
Extra Trees Classifier	EXT_2	GridSearchCV: Best Parameters

Modeling

Results

Best model that is not overfitted in the Train Set and have the best result in the Test set is:

Random Forest

Grid Search CV

- bootstrap= False,
- max_depth= 12,
- max_features= 'auto',
- min_samples_leaf= 1,
- min_samples_split= 2,
- random_state=25,
- n_estimators= 95)

	ACC	F1	PRECISION	RECALL	KAPPA
RF_1 Train	1.000000	1.000000	1.000000	1.000000	1.000000
EXT_1 Train	1.000000	1.000000	1.000000	1.000000	1.000000
RF_2 Train	0.997140	0.997135	1.000000	0.994286	0.994280
EXT_2 Train	0.989514	0.989453	0.996139	0.982857	0.979028
EXT_1 Test	0.878327	0.878788	0.872180	0.885496	0.756665
RF_2 Test	0.859316	0.854902	0.879032	0.832061	0.718570
RF_1 Test	0.855513	0.849206	0.884298	0.816794	0.710939
EXT_2 Test	0.847909	0.846154	0.852713	0.839695	0.695796
KNN_2 Train	0.776930	0.758264	0.828442	0.699048	0.553927
LR_1s Train	0.737846	0.738841	0.736742	0.740952	0.475688
LR_1s Test	0.699620	0.690196	0.709677	0.671756	0.399109
KNN_1 Train	0.693994	0.641341	0.775676	0.546667	0.388160
KNN_2 Test	0.615970	0.562771	0.650000	0.496183	0.231239
KNN_1 Test	0.539924	0.415459	0.565789	0.328244	0.078369

Overfitted

Best Model

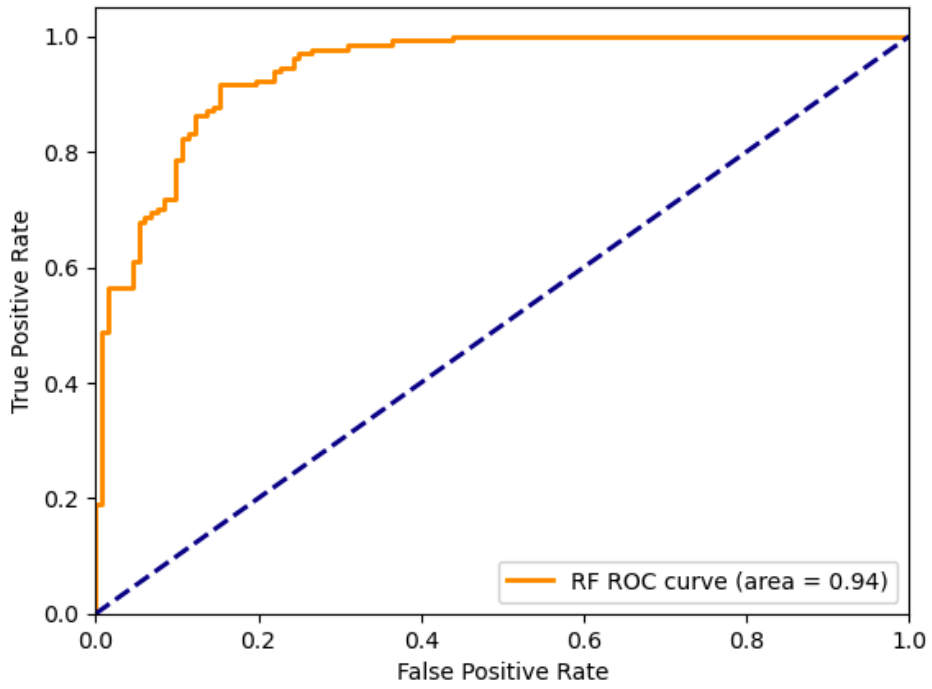
Modeling

Model Performance

Model have good performance on predict **True Positive** (Sensitivity) and **True Negative** (Specificity)

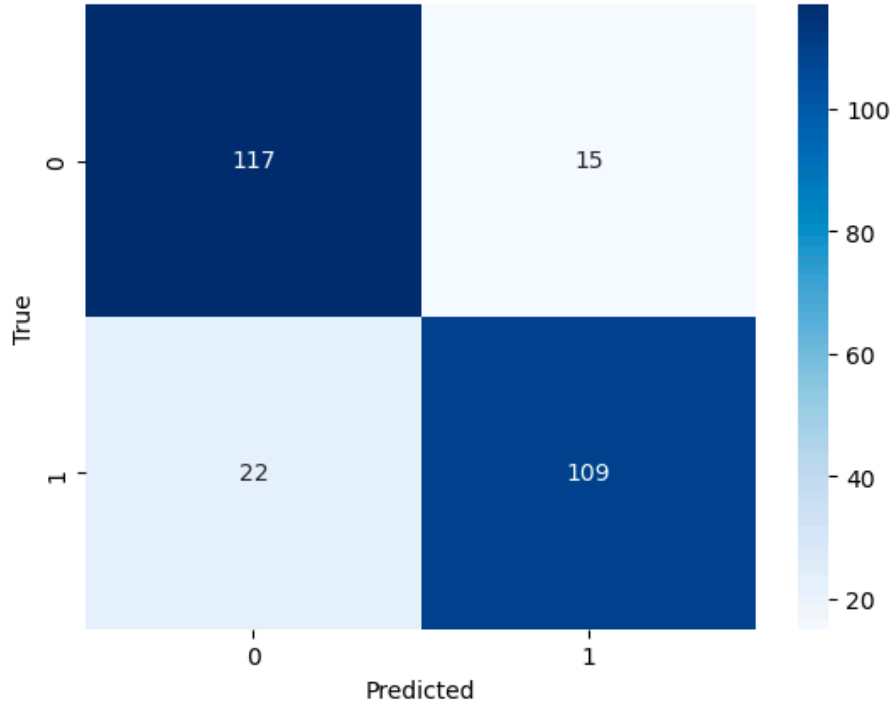
ROC Curve

AUC: 0,94



Confusion Matrix:

ACC: 0,86



Modeling

Best Features

One of relevant thing is to understand what **features** are the more important, that have more impact in my model, so I can understand what I can improve in my application to have bigger probability to get the credit.

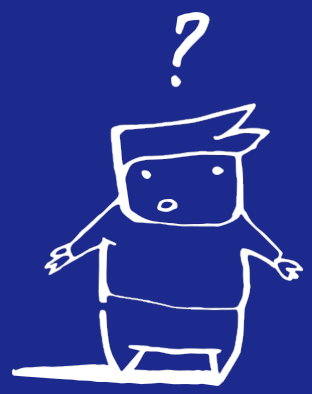
I select the ~**75**% (n =10) more important features for the app.

```
imp = pd.DataFrame(model_rf.feature_importances_, index = data.columns[:28])
imp.rename(columns={0:'importance'},inplace = True)
imp = imp.sort_values(by='importance',ascending=False)
imp['acc_importance'] = imp['importance'].cumsum()
imp
```

	importance	acc_importance
account_balance	0.153628	0.153628
credit_amount	0.124035	0.277663
age	0.105493	0.383155
duration	0.091974	0.475129
credit_history	0.071424	0.546553
length_of_current_employment	0.055424	0.601977
savings	0.055424	0.657401
property	0.046730	0.704131
job	0.030015	0.734146
guarantors	0.028417	0.762563
marital_status	0.026057	0.788619
concurrent_credits	0.023445	0.812065
purpose_radio_tv	0.020547	0.832611
telephone	0.019589	0.852201
no_of_credits_at_this_bank	0.019545	0.871745
purpose_car	0.014675	0.886420
housing_own	0.014599	0.901019
no_of_dependents	0.014262	0.915281
sex	0.014080	0.929362

*... ok, this is really nice, but im here for the **APP**, will I get the credit or not?*

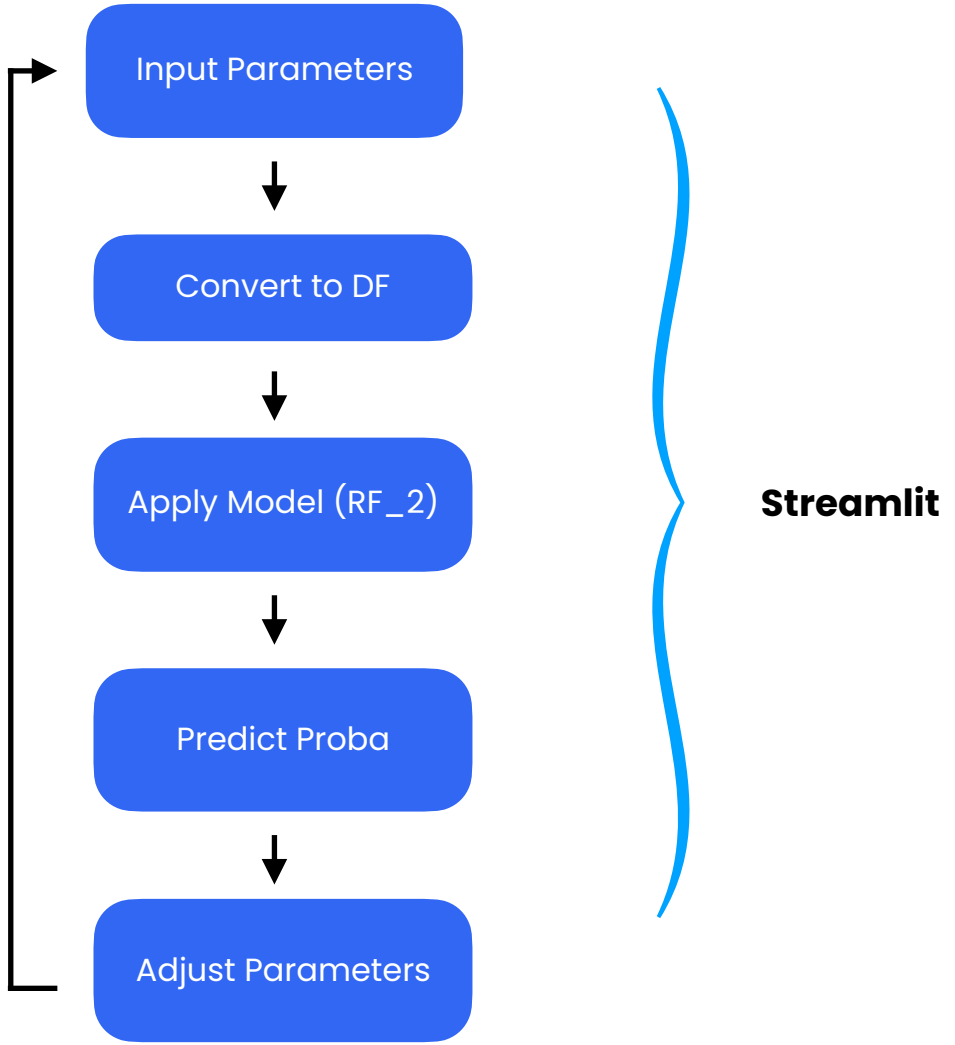
The App



KREDIT FATE

interactive **app** that can predict if you will get a credit or not, and with what probability.

How it works?



The App How it looks like

Hi, welcome to:

Kredit Fate is an App that tells you the chances of getting a credit on Germany.

Fill the questions below to find your Fate. If your chances are not so good dont give up and try to improve some parameters

What is your **Name**?:

Cris

What is your **Age**?:

35

18 100

Choose **Sex**: (1) Male, (2) Female

1

Credit **Amount**:

1400

In how many **month** you want to pay?:

12

Status of existing checking account: (1): < 0 Eu, (2): 0 <= < 400 Eu, (3): >= 400 Eu , (4): no checking account

1

1 4

Total **savings**: (1): < 200 Eu, (2) between 200 and 1000 Eu, (3) between 1000 and 2000 Eu, (4) >= 2000 Eu, (5) unknown/ no savings account

2

1 5

Do you have any **property**? (1) real state, (2), if not the previous, building society savings agreement/ life insurance, (3) if not the previous, car, (4) no property

3

1 4

Guarantors: (1): no guarantor, (2), (3)

1

1 3

Job Status: (0)/ unskilled - non-resident, (1) - resident, (2)skilled employee / official, (3)/ self-employed/highly qualified employee/ officer

1

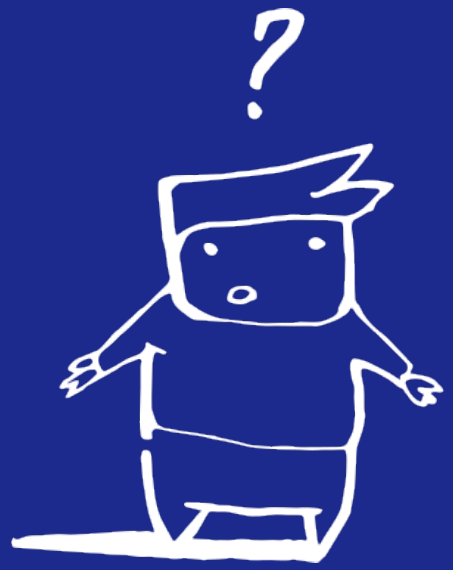
0 3

Predict

Cris , your chances to get the credit are low. Try to change some parameter, like the credit amount. 👎 probability of getting the credit: 0.59

Cristian Valeria

Data Analytics Bootcamp | Final Project



KREDIT FATE

Can I get a credit on Germany?