

Machine Learning Exam

Basics

| | |
|---|---|
| Instructions | 1 |
| Files | 1 |
| Description of the credit_train dataset | 2 |

Instructions

1. **Explore** and **preprocess** the [credit_train.csv](#) dataset.
2. **Try multiple models** and **select the most promising one**
3. **Fine-tune its hyperparameters**.

You should:

- **Analyze and explain the features & the label** : *Is this feature important for the task ? What is its relationship with the other features or the label ?*
- **Explain your choices**: *Why did you choose this preprocessing step ? Why did you choose this metric ? Why did you choose this model ? ...*
- **Get the best performance possible**: *The performance of your model will be evaluated.*
- **Organize your Notebook and add as many comments as needed**: *The performance of your model will be evaluated.*

The [credit_test.csv](#) dataset is the test set and thus, it does not contain any label. **You should predict the labels and save them as a Numpy file** (see the save function [here](#)):

```
np.save("predictions.npy", predictions)
```

Files

You should return **two files**:

- Your Notebook
- Your predictions.npy file

Description of the credit_train dataset

Attribute description

1. 'checking_status': Status of existing checking account, in Deutsche Mark.
2. 'duration': Duration in months
3. 'credit_history': Credit history (credits taken, paid back duly, delays, critical accounts)
4. 'purpose': Purpose of the credit (car, television,...)
5. 'credit_amount': Credit amount
6. 'savings_status': Status of savings account/bonds, in Deutsche Mark.
7. 'employment': Present employment, in number of years.
8. 'installment_commitment': Installment rate in percentage of disposable income
9. 'personal_status': Personal status (married, single,...) and sex
10. 'other_parties': Other debtors / guarantors
11. 'residence_since': Present residence since X years
12. 'property_magnitude': Property (e.g. real estate)
13. 'age': Age in years
14. 'other_payment_plans': Other installment plans (banks, stores)
15. 'housing': Housing (rent, own,...)
16. 'existing_credits': Number of existing credits at this bank
17. 'job': Job
18. 'num_dependents': Number of people being liable to provide maintenance for
19. 'own_telephone': Telephone (yes,no)
20. 'foreign_worker': Foreign worker (yes,no)

Label :

1. 'class': good or bad credit risks