High Level Design (HLD)

Credit risk assessment using South German credit data

December 31, 2021

Abstract

Normally, most of the bank's wealth is obtained from providing credit loans so that a marketing bank must be able to reduce the risk of non-performing credit loans. The risk of providing loans can be minimized by studying patterns from existing lending data. One approach, that addresses this problem, is to use machine learning (ML) techniques. ML makes it possible to find hidden information from data sets by way of classification.

The goal of this project is to build a model which will predict whether a person, described by the attributes of the dataset, is a good or bad credit risk.

Revision number: 1.0

Last date of revision: December 31, 2021

Contents

Chapter 1

Introduction

1.1 Description of High Level Design (HLD) document

Return to table of contents

The purpose of this document is to add the necessary details to the current project description to represent a suitable model for coding. This document is also intended to help detect contradictions prior to coding, and can be used as a reference manual for how the modules interact at a high level. In general a HLD document:

1. presents design aspects and defines them in detail
2. describes user interface
3. describes hardware and software interfaces
4. describes performance requirements
5. lists and describes non-functional attributes like security, reliability, maintainability, portability, reusability, application compatibility, resource utilization, serviceability.

1.2 Scope

Return to table of contents

The document presents the structure of the system, such as the database architecture, application architecture, application flow, and technology architecture. The HLD uses non-technical and slightly-technical terms which should be understandable to administrators of the system.

1.3 Definitions

Return to table of contents

Term	Definition
ML	Machine Learning
Database	South German credit data hosted on Astra DB
DM	Deutsche Mark

Chapter 2

General Description

2.1 Product perspective

Return to table of contents

Credit risk assessment solution is an ensemble of ML models which provides a client's probability of default and generates suggestion to a credit officer regarding a loan approval.

2.2 Problem statement

Return to table of contents

To create a ML solution for banking system of South German which assists credit officers in loan providing process (refuse or proceed) and calculates probability of default for an applicant.

2.3 Proposed solution

Return to table of contents

The solution provided in this document is a credit assessment tool constructed on South German credit data. The tool accepts some characteristics of an applicant as an input. Then it processes the information provided and outputs default probability in line with a suggestion regarding approval.

2.4 Future improvements

Return to table of contents

It will be very helpful to augment the existing data as it contains only 1000 observations. In addition collection of some new characteristics about clients can turn handy too.

2.5 Technical requirements

Return to table of contents

A loan officer can access the credit assessment tool via computer, laptop, tablet, and smartphone.

2.6 Data requirements

Return to table of contents

The credit assessment tool needs the following information about a client to procceed:

1. status of the debtor's checking account with a bank (loan provider)
2. credit duration in months
3. history of compliance with previous or concurrent credit contracts
4. purpose for which the credit is needed
5. credit amount in DM
6. debtor's savings
7. duration of debtor's employment with current employer
8. credit installments as a percentage of debtor's disposable income
9. combined information on sex and marital status
10. presence of another debtor or a guarantor for the credit
11. length of time (in years) a debtor lives in the present residence
12. a debtor's most valuable property
13. age
14. installment plans from providers other than a credit-giving bank
15. type of housing a debtor lives in
16. number of credits including the current one the debtor has (or had) at a bank (loan provider)
17. quality of debtor's job
18. number of persons who financially depend on a debtor
19. presence of a telephone landline registered on a debtor's name
20. whether a debtor is a foreign worker or not.

2.7 Tools used

Return to table of contents

Google Colab is used as a notebook environment to test ML models and create visualizations.

Sublime text is used as a code editor.

Visualizations are done using matplotlib and seaborn.

Heroku is used for deployment.

Astra DB hosts the data.

CQL is used to retrieve the data.

Front end development is done using HTML and CSS.

Flask is used for backend development.

Github is used as a version control system.

2.7.1 Hardware requirements

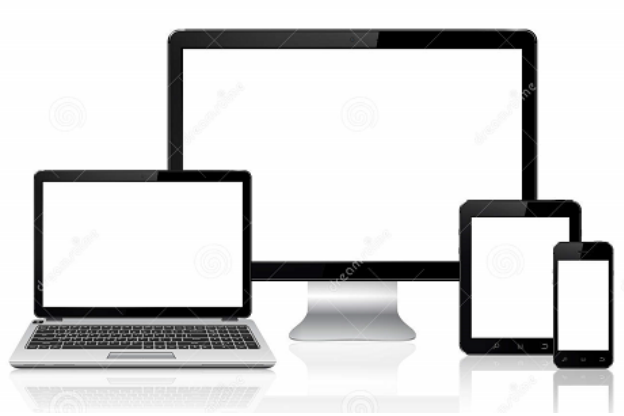
Return to table of contents

Computer

Laptop

Tablet

Smartphone



2.8 Constraints

Return to table of contents

The credit assessment solution must be user friendly and users do not need to know the inner workings to be able to exploit it.

2.9 Assumptions

Return to table of contents

The main objective of this project is to assess risk for new applicants and generate an advice regarding credit issuance. All of that will be possible due to a ML model applied to the database. Moreover all aspects of the project have the ability to work together in accordance with designer's expectation.

Chapter 3

Design details

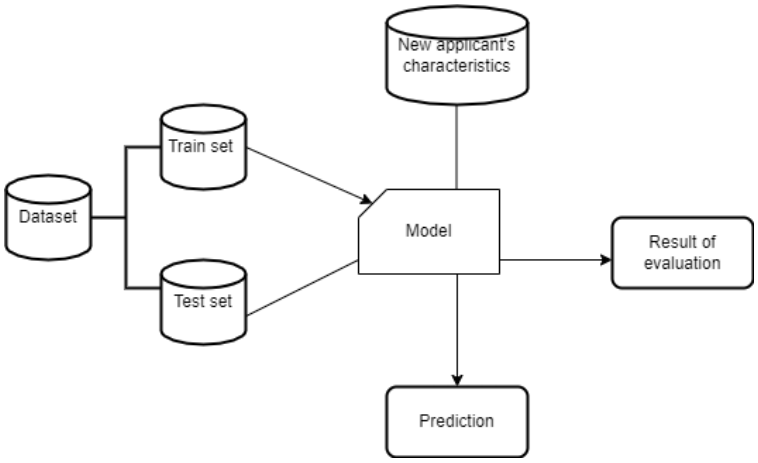
3.1 Process flow

Return to table of contents

For credit risk assessment we will use 3 ML models to ensemble one model. Process flow diagram is depicted below.

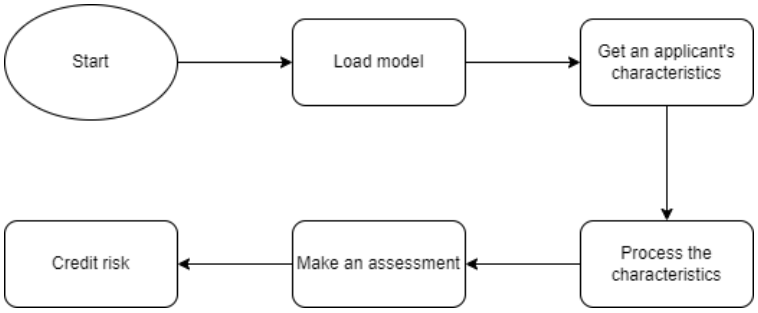
3.1.1 Model training and evaluation

Return to table of contents



3.1.2 Deployment process

Return to table of contents



3.2 Event log

Return to table of contents

The logging is applied in model building process almost after each step, but logging results will not be accessible for end users (credit officers). Moreover logging is made accessible via console and a file for developers. Logging is mandatory as it allows to find debug issues more easily.

3.3 Error handling

Return to table of contents

Anything falling outside the normal and intended usage should be considered as an error. The user interface is constructed in a way that, during inputting a new applicants characteristics, chances to make a mistake leading to an error are negligible. Therefore an end user does not see error messages during explotation of the tool.

Chapter 4

Performance

The credit risk assessment tool is used to identify applicants with high default probability to minimize a bank's risks. Therefore it should be as accurate as possible. It is very important to retrain the model for enhancing its performance as soon as new data will be available.

4.1 Reusability

Return to table of contents

The code and its components should be reused with no problems.

4.2 Application compatibility

Return to table of contents

This project has different components and python is used as an interface between them. Each component has its own task to perform, and python ensures proper transfer of information.

4.3 Resource utilization

Return to table of contents

When any task is performed, it will likely use all the processing power available until it is finished.

4.4 Deployment

Return to table of contents

Chapter 5

Conclusion

The credit risk assessment tool will estimate default probability of an applicant and assist a credit officer to decide whether to issue a credit or not.

Chapter 6

References

1. google.com

2. filesmerge.com

3. app.diagrams.net