Bank Credit Risk Using German Data

AGENDA

- ❖Introduction
- ❖Objective
- ❖Data Description
- ❖Architecture
- ❖Model Training and Evaluation Workflow
- ❖Deployment
- ❖Question

Introduction

★ Significance of Credit Risk: Credit risk is a crucial factor in the banking industry, impacting a wide range of services, including loans, credit cards, investments, and mortgages.

★ Challenges with Credit Cards: The increasing popularity of credit cards has led to a rising rate of defaults, posing significant challenges for banks.

★ Data-Driven Solutions: This project explores the use of data analytics to address these challenges by implementing a model that classifies customers into "Good" or "Bad" credit risk categories based on demographic and behavioral factors.

Objective

- **Development of a Model:** Create a model to categorize customer profiles into risk levels, identifying them as either "Good" or "Bad" credit risks.

- **Benefits:**
  - **Risk Prediction:** Helps financial institutions predict the risk level of customers requesting credit, enabling more informed decision-making.
  - **Customer Insights:** Offers deeper insights into the customer base, enhancing understanding and management of customer segments.
  - **Proactive Measures:** Allows financial institutions to take proactive steps to minimize potential losses.

Data Description

1. laufkont (status) : Status of the debtor's checking account with the bank.
   1: No checking account
   2: ... < 0 DM
   3: 0 <= ... < 200 DM
   4: ... >= 200 DM / salary for at least 1 year

2. laufzeit (duration) : Duration of the credit in months. Numerical value representing the credit duration.

3. moral (credit_history) : History of the debtor's credit at the bank.
   0: Delay in paying off in the past
   1: Critical account/other credits elsewhere
   2: No credits taken/all credits paid back duly
   3: Existing credits paid back duly till now
   4: All credits at this bank paid back duly

4. verw (purpose) : Purpose for which the credit is being requested.
   0: Others
   1: Car (new)
   2: Car (used)
   3: Furniture/equipment
   4: Radio/television
   5: Domestic appliances
   6: Repairs
   7: Education
   8: Vacation
   9: Retraining
   10: Business

5. hoehe (amount) : Amount of the credit requested. Numerical value representing the credit amount in DM.

6. sparkont (savings) : Savings account/bonds of the debtor.
   1: Unknown/no savings account
   2: ... < 100 DM
   3: 100 <= ... < 500 DM
   4: 500 <= ... < 1000 DM
   5: ... >= 1000 DM

7. beszeit (employment_duration) : Length of time the debtor has been employed at their current job.
   1: Unemployed
   2: < 1 year
   3: 1 <= ... < 4 years
   4: 4 <= ... < 7 years
   5: >= 7 years

8. rate (installment_rate) : Installment rate as a percentage of disposable income.
   1: >= 35%
   2: 25 <= ... < 35%
   3: 20 <= ... < 25%
   4: < 20%

9. famges (personal_status_sex) : Personal status and sex of the debtor.
   1: Male: divorced/separated
   2: Female: non-single or Male: single
   3: Male: married/widowed
   4: Female: single

10. buerge (other_debtors) : Other debtors or guarantors for the credit.
    1: None
    2: Co-applicant
    3: Guarantor

11. wohnzeit (present_residence) : Length of time the debtor has lived at their current residence.

    1: < 1 year
    2: 1 <= ... < 4 years
    3: 4 <= ... < 7 years
    4: >= 7 years

12. verm (property) : Type of property owned by the debtor.

    1: Unknown/no property
    2: Car or other property
    3: Building society savings agreement/life insurance
    4: Real estate

13. alter (age) : Age of the debtor in years.

    Numerical value representing the age of the debtor.

14. weitkred (other_installment_plans) : Other installment plans held by the debtor.
   1: Bank
   2: Stores
   3: None

15. wohn (housing) : Type of housing the debtor lives in.
   1: For free
   2: Rent
   3: Own

16. bishkred (number_credits) : Number of credits the debtor has in this bank.
   1: 1
   2: 2-3
   3: 4-5
   4: >= 6

17. beruf (job) : Job type of the debtor.
    1: Unemployed/unskilled -non-resident
    2: Unskilled -resident
    3: Skilled employee/official
    4: Manager/self-employed/highly qualified employee

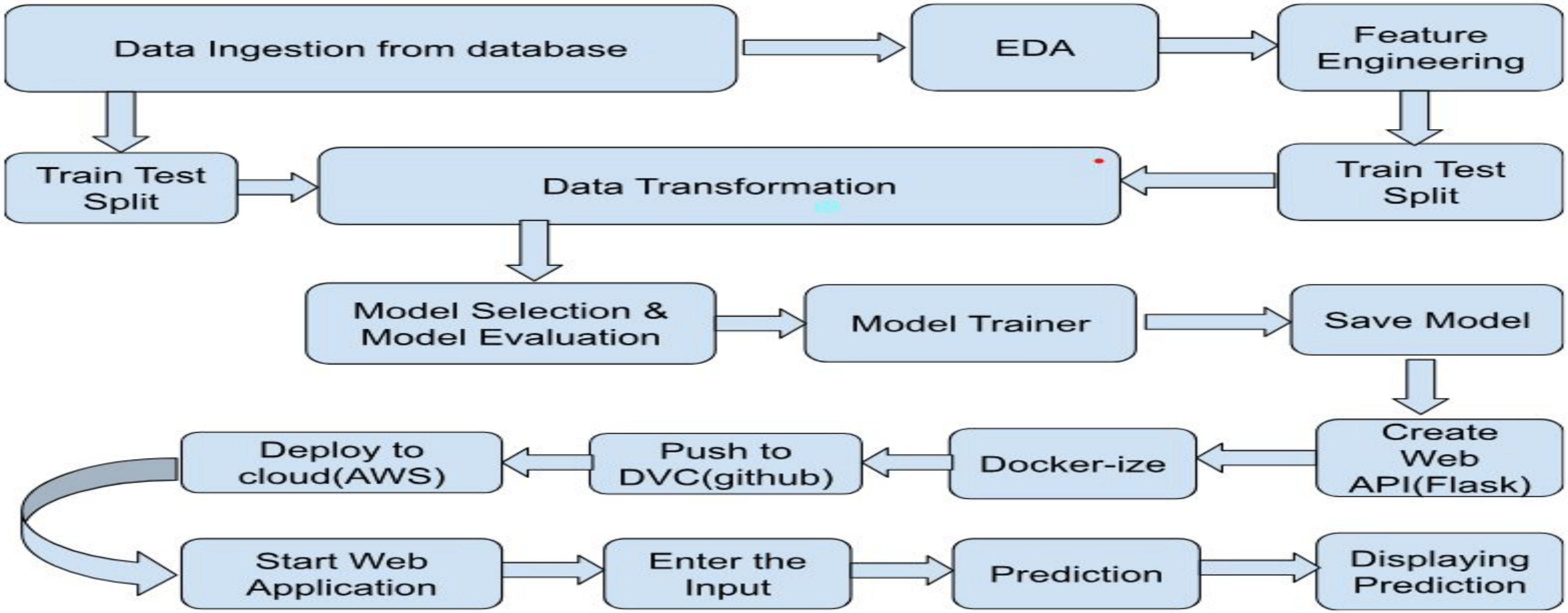
18. pers (people_liable) : Number of people who are financially dependent on the debtor.
    1: 3 or more
    2: 0 to 2

19. telef (telephone) : Whether the debtor has a telephone registered under their name.
    1: No
    2: Yes (under customer name)

20. gastarb (foreign_worker) : Whether the debtor is a foreign worker.
  1: Yes
  2: No

21. kredit (credit_risk) : The credit risk assigned to the debtor.
  0: Bad
  1: Good

Architecture



Data Ingestion from database

Exporting the data form database that is present in the database that we have to pull out from the database for model building. It can be directly given for data to data transformation for feature scaling

Data Exploration(EDA)

We divide the data into two types: numerical and categorical. We explore each type one by one. Within each type, we explore, visualize and analyze each variable one by one and note down our observations. Statistical tests are performed for each independent variable to see if the variable has any significance in determining the output for the target variable.

Feature Engineering

Encoded categorical variables

Train/Test Split

Split the data into 80% train set and 20% test set.

Data Transformation

Feature scaling is done here on train and test data

Model Selection & Model Evaluation

Models and trained and tested the data. Compared the performance of each model and selected the best one. Feature importance and/or hyper-parameter tuning performed to improve the performance of the selected model.

Model Trainer

Once we get best model we select that and used for model training for making prediction

Save the model

Saved the model by converting into a pickle file.

Create Web Application

Flask is used for creating Web Application so that user can put information from webpage

Dockerised

Dockerised the application for getting ready for deployment

Push to DVC(Github)

Pushed to Github and from there, deployed the application files to AWS, all using Github Actions as ci/cd pipeline.

Deploy to the Cloud

Selected AWS for deployment. Used the model to develop a flask application which can predict risk class for unseen data.

Application Web Application & Enter the input

User Start the application and enter the inputs.

Prediction & Display Prediction

After the inputs are submitted the application runs the model and makes predictions. The out is displayed as a message indicating whether the customer is classified as Safe or Not Safe.

FAQs

1) What is the data source?

The data is obtained from UCI Machine Learning Repository. Link : https://archive.ics.uci.edu/ml/datasets/South+German+Credit

2) What was the type of data?

The data contained both numerical and continuous type data.

3) What was the complete flow that you followed in this project?

Please refer to slide 13 .

4) How logs are managed?

We have a separate log files for each stage of the project

5) What techniques were you using for data pre-processing?

- ➢ Removing unwanted attributes
- ➢ Visualizing relation of independent variables with each other and output variables
- ➢ Cleaning data and imputing if null values are present.
- ➢ Encoding categorical variables

6) How training was done or what models were used?

- ➢ After loading the dataset, data pre-processing was done.
- ➢ For this project, we opted to train the data using the CatBoost Classifier.
- ➢ Hyper-parameter tuning, feature selection were performed during the various versions of modeling.
- ➢ The best model was selected.

7) How Prediction was done?

➢ The test files were provided.

➢ The test data also underwent preprocessing.

➢ Then the data was passed through the model and output was predicted.

8) What are the different stages of deployment?

➢ After training the model, we prepared all the necessary files required for deployment and uploaded in a document version control system called Github.

➢ We then connected to and deployed the model in AWS.

THANK YOU