Evaluating the unfairness of the Default in Credit Card Clients in Taiwan dataset

By group 7

1.Dataset information

- 30000 instances
- Binary target variable representing whether a credit card holder defaults on their payment.
  - 1 represents "Yes," indicating that the individual defaults on the credit card payment.(bad)
- 23 Explanatory Variables (Features)
  - X1: **Amount of the given credit**
  - X2: **Gender**
    - (1 = male; 2 = female).
  - X3: **Education**
    - (1 = graduate school; 2 = university; 3 = high school; 4 = others).
  - X4: **Marital status**
    - (1 = married; 2 = single; 3 = others).
  - X5: **Age** (year).

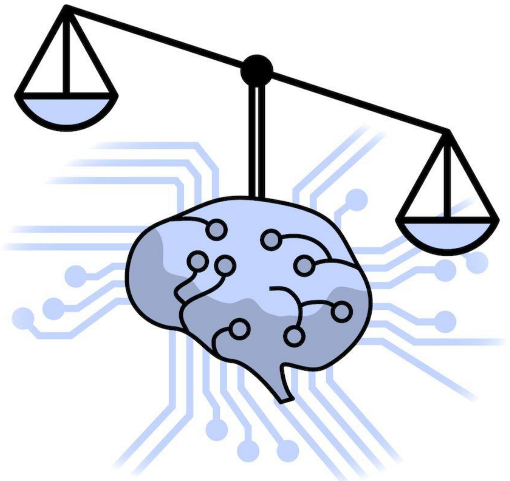
- X6 - X11: **Repayment status** for different months (from April to September, 2005)
    - X6 = the repayment status in September, 2005;
    - ; . . .;
    - X11 = the repayment status in April, 2005.
        - The measurement scale for the repayment status is: -1 = paid duly; 1 = payment delay for one month ; . . .; 9 = payment delay for nine months and above.
- X12-X17: **Amount of bill** statement for different months (NT dollar).
    - X12 = amount of bill statement in September, 2005;
    - ; . . .;
    - X17 = amount of bill statement in April, 2005.
- X18-X23: **Amount paid** for different months (NT dollar).
    - X18 = amount paid in September, 2005;
    - ; . . .;
    - X23 = amount paid in April, 2005.

Responsible AI Challenges:Fairness

- AI fairness is the aspect of responsible AI that checks that all individuals or groups free from any bias or discrimination.
- This is the main issue we have decided to tackle in this dataset.
- The reason we decided to evaluate this dataset's fairness is that having an unfair models in this situation will greatly affect people's lives due it being more likely to default the credit of certain groups.
- For our dataset we considered what type of harm would the model inflict by being unfair and we believe this to be a case of **Allocation of harms.**

Mitigation Strategies - Fairness

- The fairness of our binary classification model is measured using Demographic Parity and Equalized Odds
- **Pre-processing**:
    - Resample the training dataset
    - Identify and reduce biases in the data
- **In-processing**:
    - Data reweighting
    - Exponentiated model
- **Post-processing:**
    - Threshold adjustment

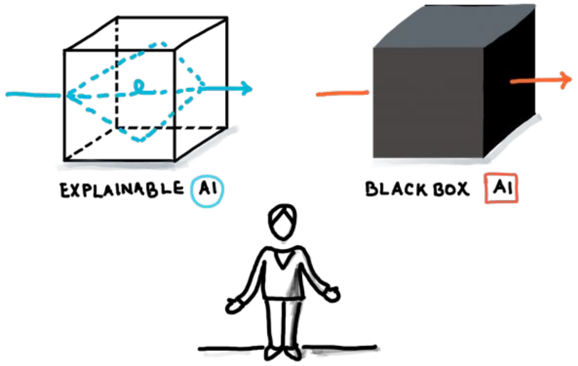


3.Responsible AI Challenges:Explainability

- Some AI models are often considered *black-box* hence the lack of transparency can lead to difficulties in being able to explain the output of the model(yes=1 or no=0).
- Having better explainability will also help us ensure that the model has achieved AI fairness.
- By being able to understand why the model has made a certain decisions and explain the output of the model we can make sure that there are not any biases being reinforced from the dataset.
- A lot of times using simpler models reduces accuracy but allows it to be explained better(also called Model accuracy and explainability trade-off)
- Understanding the dataset variables is important to understand the decision making process.

Mitigation Strategies - Explainability

- Use easily explainable models, such as Linear Regression
- Feature Importance in Linear Regression by using bagging
- Global explanation
- Local Explanation
- Documentation



References

Dataset: https://www.openml.org/search?type=data&status=active&id=42477&sort=runs

Paper: Yeh,I-Cheng. (2016). default of credit card clients. UCI Machine Learning Repository.

https://doi.org/10.24432/C55S3H.