Towards Fair Credit Risk Assessment: Developing Mathematical Models with Equity and Accuracy

Siyuan Wu[1,a], **Tingting Yang**[2,b,*]

[1]*Pomfret School*
[2]*Department of Mathematics, University of Wisconsin-Madison*
a. name-eman@163.com, b. Mobphyspde100@outlook.com
[*]*corresponding author*

Abstract: Credit default risk is an important factor in determining whether a person can apply for a credit card and continue to use it. Hothis studyver, the assessment of credit default risk should be fair and unbiased, especially as credit cards become an increasingly popular payment method. In this paper, this study analyze how to assess a user's credit default risk while emphasizing the importance of machine learning fairness. this study propose using principal component analysis (PCA) to extract key factors for judging credit default risk and a BP neural network to evaluate and analyze these factors. this study also propose adversarial representation learning which is aim to address discrimination against different minority group people. Our main purpose is to train the main network to generate features that the discriminator cannot use to accurately predict the sensitive attribute. By doing so, the learned features become fairer and do not contain discriminatory information. Therefore, adversarial representation learning is aimed at reducing discrimination against minority groups in machine learning models. Our approach serves as a natural method for ensuring that these parties act fairly and various adversarial objectives. this study demonstrate that selecting the appropriate objective is essential for achieving fair prediction. Through our approach, this study aim to ensure that our credit default risk assessment is fair and equitable for all users of different gender, race, and education.

Keywords: credit cards, credit default risk, principal component analysis

1. Introduction

A credit card is a form of non-cash payment where users don't need to pay with cash upfront, but instead keep a unified account and settle payments on a repayment date. Credit cards may also offer overdraft functionality, which is becoming increasingly popular [1–3]. However, credit cards have drawbacks, such as users being prone to overspending and accumulating debt beyond their repayment ability [4], resulting in credit default risk [4], which will result in failure to repay in time when the repayment date comes, and it will produce credit default risk.

Given the growing concern about credit default risk, it serves as a crucial factor in determining whether a user can continue to use a credit card. In this paper, this study use three aspects to evaluates and analyzes the credit default risk of the user: 1). whether the loan is paid off, 2). whether the installment is paid off, and 3). the credit card balance. These assessments determine the user's credit default risk level and their eligibility for holding a credit card again [4, 5].

Addressing fairness issues in credit risk assessment is crucial because credit decisions can significantly impact people's lives, and biased decisions can lead to systemic discrimination against certain groups. If credit risk assessment algorithms are biased, they can result in unjustified denial of credit or higher interest rates for individuals from certain demographic groups, such as people of color, women, or members of other historically disadvantaged communities [6].

This bias can perpetuate social and economic inequality by limiting access to financial resources and opportunities for these individuals, making it harder for them to achieve financial stability and independence. Furthermore, biased credit risk assessment algorithms may also negatively impact the credit card provider by hindering access to credit for potentially profitable ventures [7]. Therefore, it is imperative to address fairness issues in credit risk assessment to ensure that decisions are based on objective and non-discriminatory factors. By promoting fairness in credit risk assessment, this study can create a more equitable financial system that benefits individuals and society as a whole [8].

One study by [9] proposes a new hybrid deep learning model that combines a convolutional neural network (CNN) and long short-term memory (LSTM) neural network for credit card fraud detection. The results show that this model outperforms traditional machine learning algorithms such as logistic regression and decision trees. Another study by [10] proposes a novel framework that uses machine learning algorithms such as random forest and gradient boosting for credit card risk assessment. The study also employs an ensemble model that combines the predictions of multiple machine learning algorithms to improve the accuracy of credit card risk assessment. In addition to machine learning, researchers have also explored the use of big data analytics for credit card risk analysis. For instance, a study by [11] uses a big data analytics framework that combines various data sources such as credit reports, transaction records, and social media data to identify high-risk credit card users. The results show that this approach is more effective than traditional credit scoring models.

Furthermore, some studies have focused on developing new credit scoring models that incorporate alternative data sources such as mobile phone usage and online behavior. For example, a study by [11] proposes a new credit scoring model that uses mobile phone usage data to assess creditworthiness. The study shows that this model is more accurate than traditional credit scoring models that only rely on financial data. However, their framework cannot handle high dimensional data input.

The proposed framework introduces adversarial representation learning as a natural method for ensuring fair decision-making by third parties with unknown objectives, particularly in the context of group fairness in credit risk assessment. The framework establishes a connection between group fairness and various adversarial objectives, and through worst-case theoretical guarantees and experimental validation, demonstrates that the selection of the appropriate objective is critical to achieving fair prediction.

In summary, the main contribution of the proposed framework is the introduction of an approach that promotes fairness in credit risk assessment by using adversarial representation learning to mitigate bias and ensure that decisions are objective and non-discriminatory.

2. Problem formulation

Suppose this study have a dataset X with n credit card applications and p features, including information such as income, credit score, debt-to-income ratio, and other factors, such as race, gender, education level [4]. Each credit card application is labeled as either "fraudulent" or "non-fraudulent" based on whether the applicant defaulted on their credit card payments.

Our goal is to build a model that can not only accurately but also fairly predict whether a new credit card application will be “fraudulent” or “non-fraudulent”, based on the available features. This is a binary classification problem, and this study can use a logistic regression model to make the predictions.

Once this study have trained the regression model, this study can use it to predict the probability that a new credit card application will be “fraudulent” or “non-fraudulent”. Specifically, this study calculate the dot product of the model weights and the features of the new credit card application, and apply the sigmoid function to obtain a probability value between 0 and 1. this study can then set a threshold value (e.g., 0.5) to make a binary prediction.

To evaluate the performance of the regression model, this study can use metrics such as accuracy, precision, recall, and F1 score. this study can also use techniques such as cross-validation to estimate the generalization performance of the model on unseen data.

Overall, the Fair Credit Card Risk Assessment Problem involves building a regression model to predict the risk of credit card default based on a set of features, and evaluating the performance of the model using appropriate metrics.

this study can model the probability p_i that the ith credit card application is “non-fraudulent” using a regression model:

$$p_i = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip})}}$$

where β_0 is the intercept term, $\beta_1, \beta_2, \ldots, \beta_p$ are the weights associated with each feature x_{i1}, x_{i2}, . . ., x_{ip}, and e is the base of the natural logarithm.

this study can estimate the model parameters $\beta_0, \beta_1, \ldots, \beta_p$ using maximum likelihood estimation, which involves maximizing the likelihood function:

$$L(\beta_0, \beta_1, \ldots, \beta_p) = \prod_{i=1}^{n} p_i^{y_i} (1 - p_i)^{1-y_i}$$

To evaluate the performance of the regression model, this study can use metrics such as accuracy, precision, recall, and F1 score. this study can calculate these metrics using the predicted probabilities and the true labels of the credit.

3. Methods

3.1. PCA for feature selection

In this section, this study introduce a statistical method called principal component analysis to analyze the obtained data. This method can transform a set of potentially correlated variables into a set of linearly uncorrelated variables through an orthogonal transformation. For example, this study choose whether the loan is paid off and whether the instalment is paid off as a set of variables, both of which have a certain correlation, they all relate to the balance of the credit card.

this study then use projection tracking to assign weights to each of the principal component metrics from the main analysis. this study should try the best to achieve that only one projection is searched for each projection pursuit to ensure that the extracted projections are non-Gaussian distribution and reduce errors. Projection pursuit is a commonly used method for processing and analyzing high-dimensional data.

3.2. Selection of indicators based on PCA

There are many aspects to judge the credit default risk level of a user, such as whether the loans of different credit types are repaid. In order to identify the reliable key factors in assessing credit default risk, this study analyze it though three aspects: whether the loan of different credit type is paid off, whether the installment applied by credit card is paid off, and the balance of the credit card.

As mentioned in the previous section, the preprocessed data has some kind of correlations, but it is difficult to find it only from the data, so it needs to be converted into a principal component, and the information of the correlation between the data can be obtained through principal component analysis.

Following are the steps regarding the principal component analysis implementation process:

First, this study collect and preprocess the data that this study want to analyze. This includes identifying the variables or features that this study want to consider in our analysis, and transforming the data into a suitable format (e.g., standardized or normalized).

$$X' = \frac{X - \bar{X}}{\sigma} \tag{1}$$

Next, this study calculate the covariance matrix of the data. The covariance matrix is a square matrix that shows the relationships between all pairs of variables in our data. The diagonal elements of the matrix represent the variances of each variable, and the off-diagonal elements represent the covariances between pairs of variables.

$$S = \frac{1}{n-1}(X' - \bar{X})^T(X' - \bar{X}) \tag{2}$$

this study then calculate the eigenvectors and eigenvalues of the covariance matrix. Eigenvectors are vectors that are scaled by a scalar factor (the eigenvalue) when multiplied by the original matrix. In PCA, the eigenvectors represent the directions in which the data has the most variation. The eigenvalues represent the amount of variation explained by each eigenvector.

this study can then sort the eigenvectors in descending order based on their eigenvalues. The eigenvectors with the highest eigenvalues represent the directions with the most variation in the data, and therefore the most important features.

Finally, this study can project our data onto the new set of eigenvectors to obtain a lower-dimensional representation of the data that captures most of its variation. This allows us to identify the most important features for risk assessment:

$$X_{proj} = X'V_k \tag{3}$$

where V_k is a matrix containing the selected principal components as columns, and X_{proj} is the projected dataset.

3.3. Fairness learning

The proposed fairness loss for credit card risk assessment is based on the concept of demographic parity [12], which aims to ensure that credit decisions are made without regard to a person's demographic group membership, such as race ,gender and education [12].

this study propose a fairness loss that can be formulated as::

$$L_{Fair} = -\sum_{y\in Y}\sum_{a\in A} P(Y=y|A=a)P(Y=y) \times \log\left(\frac{1}{P(Y=y|A=a)}\right) \tag{4}$$

where Y is the set of possible credit decisions (e.g., approve or deny), A is the set of demographic groups, and $P(Y=y|A=a)$ is the probability of a credit decision y given a demographic group a. $P(Y=y)$ is the overall probability of the credit decision y in the entire population.

This loss function measures the negative average surprise among the different demographic groups a for all credit decisions (i.e. fraudulent and non-fraudulent). A lower value of the loss function indicates higher group fairness, i.e., a credit decision should be independent of demographic group membership.

By including this fairness loss in the credit card risk assessment model, the model is incentivized to make credit decisions that are unbiased and fair to all demographic groups.

Specifically, this study can use a fairness loss term that penalizes the model for making unfair predictions.

Based on the figure, this study can formulate the adversarial training framework:

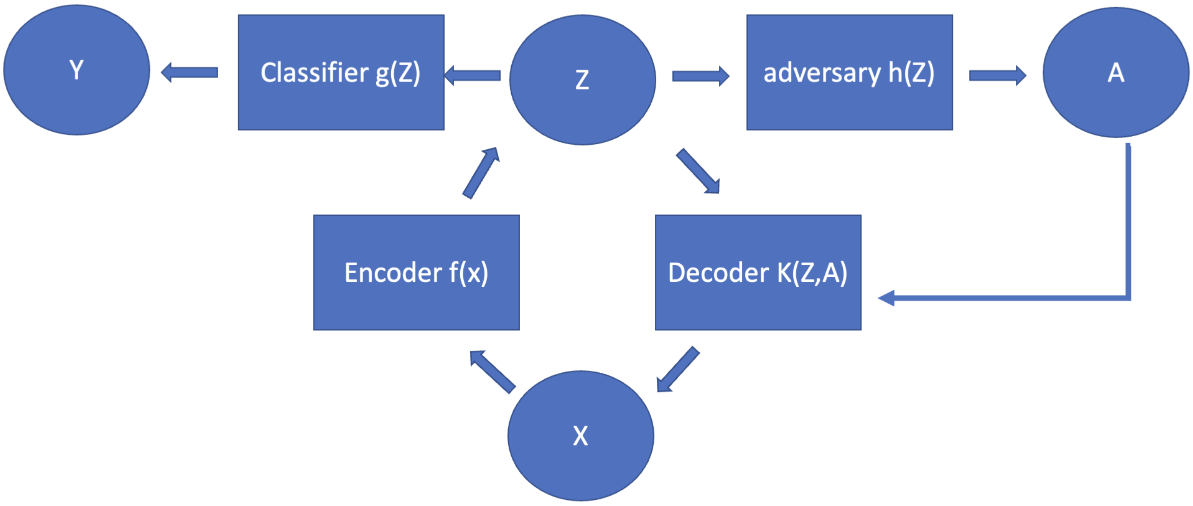


Figure 1: Adversarial Training Framework.The encoder f maps($f(x)$) X to Z, the decoder $K(Z, A)$ k reconstructs X from (Z, A), the classifier g predicts Y from Z, and the adversary h predicts A from Z.

Our approach, illustrated in Figure 1 above, involves a generalized model that learns a data representation Z, capable of reconstructing inputs X, classifying target labels Y, and safeguarding sensitive attribute A from an adversary. The hyperparameters α, β, and γ determine the desired balance between utility, reconstruction of inputs, and fairness. By setting any of these hyperparameters to zero, this study can omit the corresponding requirement and use the model in strictly supervised or unsupervised settings, as needed.

To train the model with fairness constraints, this study use an objective function inspired by [13], as shown below:

$$\begin{aligned} L(f,g,h,k) =&\alpha L_C(g(f(X,A)),Y) \\ &+\beta L_{Dec}(k(f(X,A),A),X) \\ &+\gamma L_{Fair}(h(f(X,A)),A) \end{aligned} \tag{5}$$

Here, L_C represents the classification loss. The encoder, decoder, and classifier work together to minimize L_C while maximizing the adversarial loss L_{fair} with respect to the adversary h. The expectation is taken over the tuples (X,Y,A) in the dataset.

The adversarial examples are generated by perturbing the input data x in such a way that it minimize the fairness loss. The resulting model is then tested on the original data to ensure that it is both accurate and fair. The resulting loss combination can find a good trade-off between accuracy and fairness.

By integrating the fairness objective into adversarial learning, this study can train models that not only achieve high accuracy but also avoid making unfair predictions based on sensitive attributes such as race, gender, or age.

3.4. Confidence evaluation of prediction

To make sure our model's prediction is reliable. this study define the confidence evaluation for our model. Assuming that this study have the predictions for each of the 30 experiments, this study can calculate the mean and standard deviation of the predictions. The confidence interval can then be calculated using the following formula:

$$\text{Confidence interval} = \text{Mean} \pm (z\text{-score}) \times \frac{\text{Standard deviation}}{\sqrt{\text{Sample size}}} \tag{6}$$

The z-score depends on the desired confidence level and can be looked up in a standard normal distribution table. For example, if this study want a 95% confidence interval, the z-score would be 1.96.

4. Data Exploration

this study integrate some data with different attributes and the main features are listed below:

From the Table 1, this study can see that this dataset can potentially be imbalanced for different races and genders because of the large difference in the number of observations between certain categories. For example, there are significantly more observations for male borrowers than female borrowers, which may lead to biased predictions or assessments if gender is included as a variable in the analysis. Similarly, there are more observations for white borrowers than black or Asian borrowers, which could also lead to biased assessments if race is included as a variable in the analysis.

An imbalanced dataset can result in several issues, such as overfitting, underfitting, and biased model predictions. In machine learning, it is generally recommended to have a balanced dataset where each class (in this case, each race and gender) has an equal number of observations. This is because many machine learning algorithms assume that the classes are balanced, and may not perform as well when faced with an imbalanced dataset. this study implement near miss algorithm to handle the data imbalance problem [14].

Table 1: Dataset description

Attribute types	**Attribute counts**
Credit card	18333667
Car loan	5019228
Mortgage	446901
Loan for business development	48856
Real estate loan	47181
Mobile operator loan	35
Microloan	9183
Unknown type of loan	313
Female	1912313
Male	16421354
High school	191391
University	123919
Grad	132913
white	6121999
Black	3919319
Asian	4919391

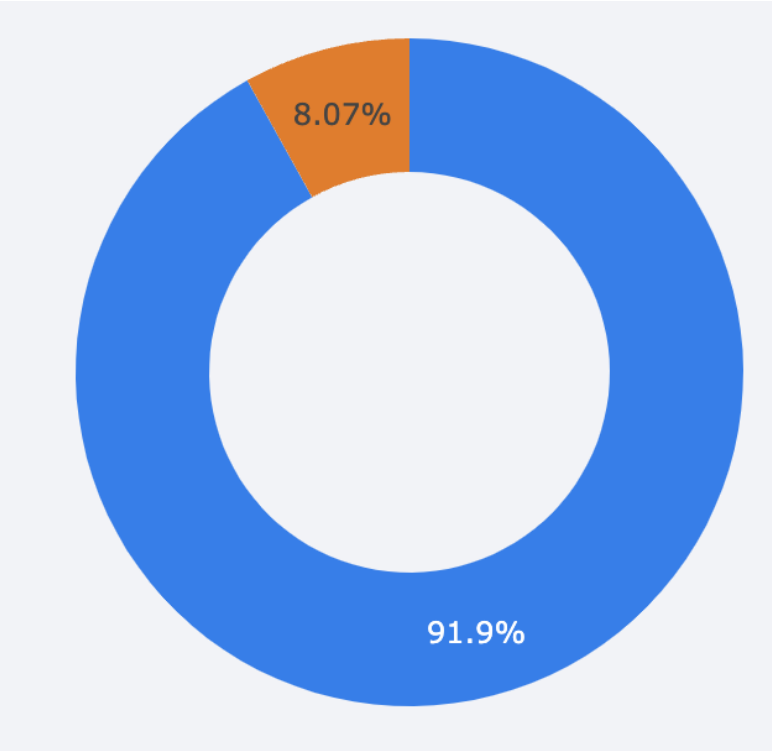


Figure 2: Analysis of different types of transactions

From the Figure 2, this study can observe one class (in this case, non-fraudulent transactions) is significantly more prevalent than the other class (fraudulent transactions). In this case, the non-fraudulent transactions account for over 90% of the dataset, while the fraudulent transactions only account for around 8%. this study also visualize the transaction in Figure 3.

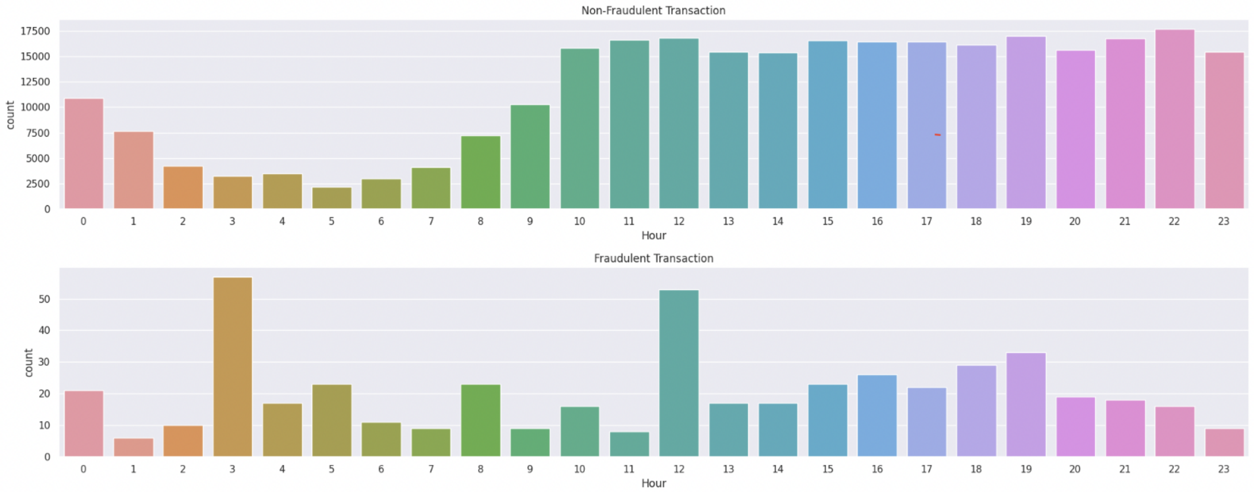


Figure 3: Transaction by different hours

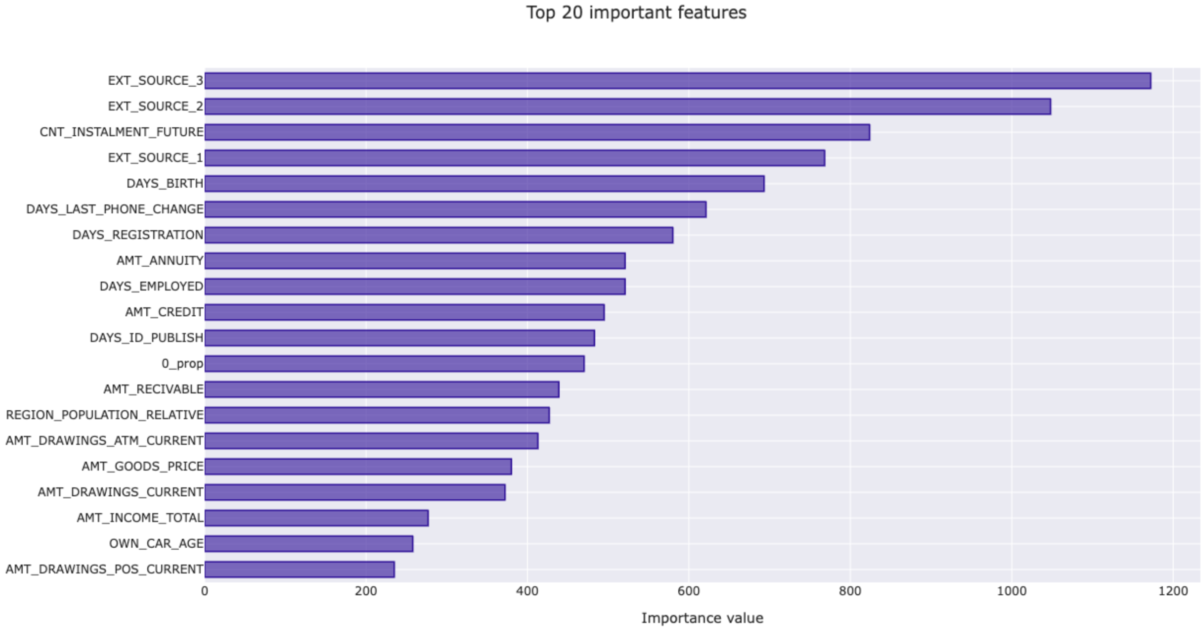


Figure 4: Analysis of important feature for prediction using PCA

This study furthermore analyze the most important top 20 features using PCA method in Figure 4. Overall, these features provide a good starting point for credit risk assessment. However, it's important to note that other factors, such as the borrower's age, education, and financial stability, should also be considered.

5. Experiments

5.1. Experiment procedure

Step 1: Import the dataset Use the read_csv method in pandas to read the csv file and this study can obtain six sets of data: application_train, bureau, bureau_balance, POS_CASH_balance, credit_card_balance and installments_payments. After reading the data in the csv file, then output the shape of each dataset.

Step 2: Transform and clean the acquired data Some special data whose values are relatively deviated from other values in the data set, if they are still retained, will affect subsequent data processing and other operations, resulting in large deviations in the results of data processing. Therefore, such data needs to be converted and processed. Preprocessing operations such as cleaning.

First is to remove constant values from the dataset and data that are not related to or cannot be explained by credit default risk. After removing the relevant values, start processing the missing data in the dataset. After processing all the data, calculate the proportion of each data to its total data dichotomy, convert the categorical variables to continuous variables, and finally create a new variable.

Step 3: Calculate the sample correlation index matrix. The sample correlation index matrix provides a measure of the strength and direction of the linear relationship between pairs of variables in the dataset. Here this study consider top 20 features.

For example, in PCA, the sample correlation index matrix is used to identify which variables are most strongly correlated with each other, and to determine the principal components of the data that capture the most important patterns of variation. In risk assessment models, the sample correlation index matrix can be used to identify which variables are most strongly associated with the outcome of interest (e.g., credit default), and to assess the overall level of multicollinearity in the dataset (i.e., the degree to which the independent variables are correlated with each other).

Step 4: Calculate the eigenvalues of the relevant index matrix and its eigenvalue vector

Step 5: Obtain the final principal components through principal component analysis.

Step 6: Calculate correlation score by PCA using Python and get the processed data

Step 7: Input the principal data into our model. In our model,this study use 10 layers neural network with learning rate as 0.001, and this study integrate the loss function in Equation 5 together. In our model, this study use grid search and find $\lambda = 0.2$ can find the best result.

6. Model Evaluation

6.1. Definition of evaluation criteria matrix

Confusion matrix is one of the standard formats for evaluating the accuracy of classification models. As shown in the figure below, each column of the confusion matrix represents the predicted class, and each row represents the true attribution class of the data.

True Positive: The sample category is a positive class, and the model recognizes it as a positive class.

False Positive: The sample category is a positive class, and the model recognizes it as a negative class.

True Negative: The sample category is a negative class, and the model recognizes it as a positive class.

False Negative: The sample category is a negative class, and the model recognizes it as a negative class.

Equalized Odds is a measure of fairness that evaluates whether a model satisfies the same true positive rate and false positive rate for different subgroups in a protected attribute. It can be mathematically formulated as follows:

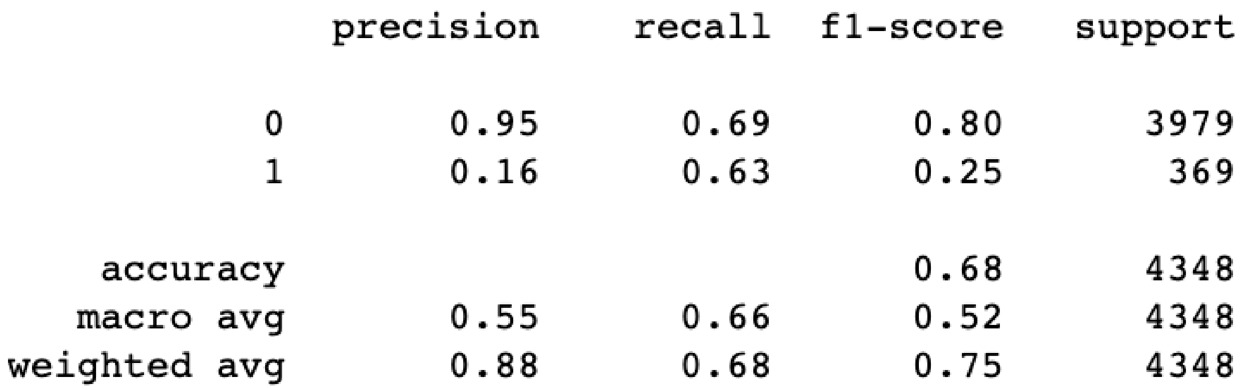
$$\begin{aligned}\Delta_{EO}(g) \triangleq & \left|\mathbb{E}_{\mathcal{Z}_0^0}[g] - \mathbb{E}_{\mathcal{Z}_1^0}[g]\right| \\ & + \left|\mathbb{E}_{\mathcal{Z}_0^1}[1-g] - \mathbb{E}_{\mathcal{Z}_1^1}[1-g]\right|,\end{aligned} \tag{7}$$

where $\Delta_{EO}(g)$: represents the equalized odds distance of the classifier g, which comprises the absolute difference in false positive rates plus the absolute difference in false negative rates.

AUC score, short for Area Under the Curve score, is a common performance metric used in binary classification tasks. It measures the overall performance of a model in distinguishing between positive and negative classes by calculating the area under the Receiver Operating Characteristic (ROC) curve.

6.2. Results Analysis of Performing Confusion Matrix

1. In the Precision Matrix by our model, this study can see from below Figure 5 that nearly 90% of the data is Re_paid, so the conclusion will be that the two cases of False are higher than the two cases of True.
2. When running the confusion matrix this time, as shown in Figure 5. There are two sets of data that, by definition, deviate from actual life situations. The definition of False Positive is that the customer cannot repay the loan, but the model still believes that it can be paid; the definition of False Negative is: the customer can actually pay, but the model thinks that it cannot pay.



```
              precision    recall  f1-score   support

           0       0.95      0.69      0.80      3979
           1       0.16      0.63      0.25       369

    accuracy                           0.68      4348
   macro avg       0.55      0.66      0.52      4348
weighted avg       0.88      0.68      0.75      4348
```

Figure 5: Precision matrix output from our model

Table 2: Performance comparison. The higher the AUC score represents better prediction performance while the lower EO score represents lower unfairness. And lower confidence interval represents higher confidence of our prediction

Methods	AUC score	EO	CI_{AUC}	CI_{EO}
RandomForest	0.931	16.88	0.021	2.131
ExtraTree	0.921	15.84	0.031	3.131
XGBoost	0.951	8.98	0.044	3.141
KNeightbors	0.919	10.51	0.031	4.455
LogisticRegression	0.912	7.55	0.029	5.513
SGDClassifier	0.913	8.11	0.031	6.511
Ours	**0.979**	**7.51**	0.028	2.321

6.3. Computing performance indicators

Using the test dataset, evaluate the performance of the models by calculating the AUC score [15], EO [16], and confidence interval level [17, 18].

6.4. Compared baselines

6.4.1. Random Forest

Random Forest is a machine learning algorithm commonly used for classification and regression tasks. It is an ensemble method that combines multiple decision trees to make more accurate and stable predictions. In Random Forest, each tree is built on a subset of the training data and a random subset of the features. During prediction, the output of each tree is aggregated to form the final prediction. Random Forest has several advantages, such as reducing overfitting, handling missing values and outliers, and providing feature importance measures. It has been applied to many fields, including credit risk assessment, where it can learn from historical data to predict the risk of default for new credit applicants [19].

6.4.2. Extra Tree

Extra Trees (or Extremely Randomized Trees) is an extension of Random Forest algorithm, where the decision trees are constructed in a slightly different way. In Random Forest, each tree is trained on a random subset of features and uses a bootstrap sample of the original data. In Extra Trees, instead of selecting the best split point based on information gain or Gini index, it selects the split point randomly from a subset of cut-points. This means that Extra Trees tends to have more random splits and can help to avoid overfitting.

Similar to Random Forest, Extra Trees can be used for classification and regression tasks, and it works well on high-dimensional datasets with a large number of features. In credit card risk assessment, Extra Trees can be used to predict the probability of default based on the customer's credit history and other relevant features. The model can be trained on a large dataset of past transactions and used to score new applicants for creditworthiness [20].

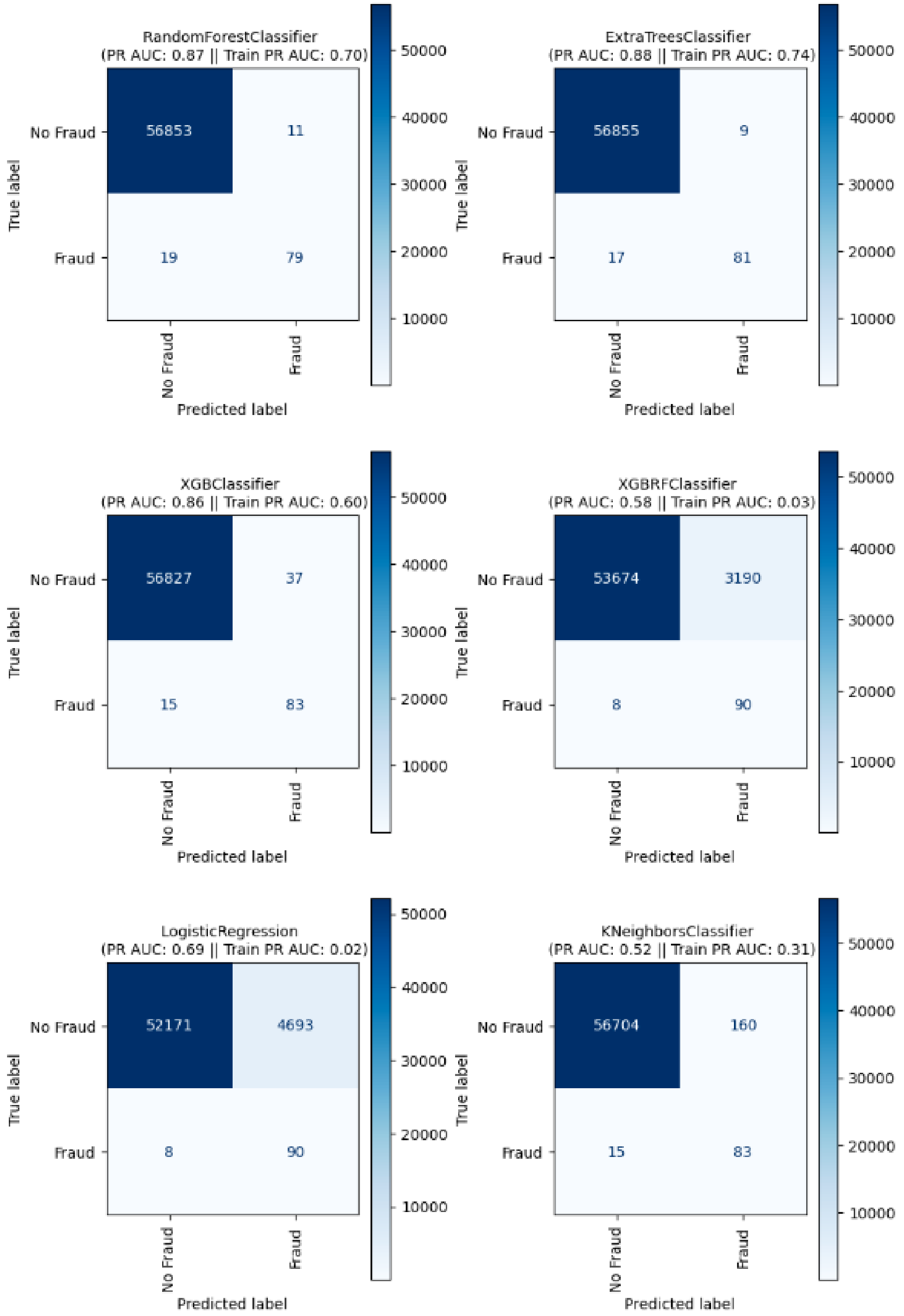


Figure 6: Confusion matrix evaluation of different methods

6.4.3. XGBoost

XGBoost is a powerful machine learning algorithm that can be used for both classification and regression problems. It builds an ensemble of decision trees sequentially, where each new tree corrects the errors of the previous ones. It employs a gradient descent optimization method to minimize the loss function by iteratively adding decision trees. XGBoost can handle missing data and also provides a way to handle imbalanced datasets.

The algorithm uses regularization techniques such as L1 and L2 regularization to prevent overfitting and reduce the impact of noisy data. Additionally, it supports parallel processing, making it faster than traditional gradient boosting algorithms [21].

6.4.4. KNeighbors

K-Nearest Neighbors (KNN) is a simple machine learning algorithm used for both classification and regression problems. In the context of credit risk assessment, KNN can be used to predict the creditworthiness of a borrower based on the similarity of their attributes (such as income, credit score, etc.) to those of previously labeled borrowers.

KNN can be useful for credit risk assessment because it is a non-parametric algorithm that does not make any assumptions about the underlying distribution of the data. However, it can be sensitive to the choice of distance metric and the value of k, and may not perform well with high-dimensional data [21].

6.4.5. Logistic Regression

Logistic Regression is a statistical model used for binary classification tasks where the outcome variable is binary (0 or 1). In credit risk assessment, the outcome variable is often whether a borrower is likely to default (1) or not (0). The logistic regression model estimates the probability of the outcome variable being 1 based on a set of predictor variables. The model uses a logistic function (sigmoid function) to transform the linear combination of predictor variables into a probability between 0 and 1. The model is trained using a labeled dataset and estimates the coefficients for each predictor variable. Once trained, the model can be used to predict the probability of default for new borrowers. If the probability exceeds a predefined threshold, the borrower is classified as a high-risk borrower. Logistic regression is a popular model in credit risk assessment due to its interpretability and ability to handle both continuous and categorical predictor variables [22].

7. Conclusion

With the development of science and technology, machine learning and deep learning models have been widely used in different industries, and credit default risk analysis is no exception. Machine learning models play an important role in analyzing the credit default risk of users [1, 23].

Based on the results in Table 2 and Figure 6, our model outperforms all the other methods in terms of AUC score, with a score of 0.979, which is significantly higher than the second-best performing method, XGBoost, with a score of 0.951. This indicates that our model has a higher ability to distinguish between positive and negative credit risk assessments.

Additionally, our model also achieved a lower EO score, with a value of 7.51, indicating a lower degree of unfairness compared to other methods. This suggests that our model is able to maintain a fair assessment of credit risk for all individuals, regardless of their sensitive attributes.

Finally, the lower confidence interval of our prediction also indicates a higher level of confidence in our model's performance compared to other methods. This suggests that our model's predictions are more reliable and less subject to chance.

Overall, our model performs better in both accuracy and fairness compared to other commonly used methods for credit risk assessment. Compared with the previous method of assessing risks by paper questionnaires, the risk level assessed by the machine learning model is more objective and more reliable. Therefore, the assessment of users' credit default risk based on the machine learning model can be implemented in the market.

This can have a positive impact on society by reducing the number of people who are unfairly denied access to credit and financial opportunities. It can also help to reduce poverty and increase economic growth by allowing more people to access credit and invest in their future.

Additionally, a fair credit risk assessment framework can promote trust in financial institutions and increase transparency in lending practices. This can help to build stronger relationships between financial institutions and their customers, which can lead to greater financial stability and prosperity for everyone.

In the future, this study will investigate the fairness of large language model. this study will also collect more various types of data to improve the performance of our model and evaluate on different datasets.

References

[1] *Amir E Khandani, Adlar J Kim, and Andrew W Lo. Consumer credit-risk models via machine-learning algorithms.* Journal of Banking & Finance*, 34(11):2767–2787, 2010.*

[2] *Khyati Chaudhary, Jyoti Yadav, and Bhawna Mallick. A review of fraud detection techniques: Credit card.* International Journal of Computer Applications*, 45(1):39–44, 2012.*

[3] *Aihua Shen, Rencheng Tong, and Yaochen Deng. Application of classification models on credit card fraud detection. In* 2007 International conference on service systems and service management*, pages 1–4. IEEE, 2007.*

[4] *Sumit Agarwal, Paige Marta Skiba, and Jeremy Tobacman. Payday loans and credit cards: New liquidity and credit scoring puzzles?* American Economic Review*, 99(2):412–417, 2009.*

[5] *Dejan Varmedja, Mirjana Karanovic, Srdjan Sladojevic, Marko Arsenovic, and Andras Anderla. Credit card fraud detection-machine learning methods. In* 2019 18th International Symposium INFOTEH-JAHORINA (INFOTEH)*, pages 1–5. IEEE, 2019.*

[6] *Tal Zarsky. The trouble with algorithmic decisions: An analytic road map to examine efficiency and fairness in automated and opaque decision making.* Science, Technology, & Human Values*, 41(1):118–132, 2016.*

[7] *Zhenpeng Chen, Jie M Zhang, Federica Sarro, and Mark Harman. Maat: a novel ensemble approach to addressing fairness and performance bugs for machine learning software. In* Proceedings of the 30th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering*, pages 1122–1134, 2022.*

[8] *Maria De-Arteaga, Stefan Feuerriegel, and Maytal Saar-Tsechansky. Algorithmic fairness in business analytics: Directions for research and practice.* Production and Operations Management*, 31(10):3749–3770, 2022.*

[9] *Mahmoud Abdallah, Nhien An Le Khac, Hamed Jahromi, and Anca Delia Jurcut. A hybrid cnn-lstm based approach for anomaly detection systems in sdns. In* Proceedings of the 16th International Conference on Availability, Reliability and Security*, pages 1–7, 2021.*

[10] *Tuong Le, Bay Vo, Hamido Fujita, Ngoc-Thanh Nguyen, and Sung Wook Baik. A fast and accurate approach for bankruptcy forecasting using squared logistics loss with gpu-based extreme gradient boosting.* Information Sciences*, 494:294–310, 2019.*

[11] *María Óskarsdóttir, Cristián Bravo, Carlos Sarraute, Jan Vanthienen, and Bart Baesens. The value of big data for credit scoring: Enhancing financial inclusion using mobile phone data and social network analytics.* Applied Soft Computing*, 74:26–39, 2019.*

[12] *Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. A survey on bias and fairness in machine learning.* ACM Computing Surveys (CSUR)*, 54(6):1–35, 2021.*

[13] *David Madras, Elliot Creager, Toniann Pitassi, and Richard Zemel. Learning adversarially fair and transferable representations. In* International Conference on Machine Learning*, pages 3384–3393. PMLR, 2018.*

[14] *Nhlakanipho Michael Mqadi, Nalindren Naicker, and Timothy Adeliyi. Solving misclassification of the credit card imbalance problem using near miss.* Mathematical Problems in Engineering*, 2021:1–16, 2021.*

[15] *Saharon Rosset. Model selection via the auc. In* Proceedings of the twenty-first international conference on Machine learning*, page 89, 2004.*

[16] *Yulu Jin and Lifeng Lai. Fairness-aware regression robust to adversarial attacks.* arXiv preprint arXiv:2211.04449*, 2022.*

[17] Clemens Kreutz, Andreas Raue, and Jens Timmer. Likelihood based observability analysis and confidence intervals for predictions of dynamic models. BMC Systems Biology*, 6(1):1–9, 2012.*

[18] JD Balakrishnan and Roger Ratcliff. Testing models of decision making using confidence ratings in classification. Journal of Experimental Psychology: Human Perception and Performance*, 22(3):615, 1996.*

[19] Nazeeh Ghatasheh. Business analytics using random forest trees for credit risk prediction: a comparison study. International Journal of Advanced Science and Technology*, 72(2014):19–30, 2014.*

[20] Trishita Saha, Saroj Kumar Biswas, Saptarsi Sanyal, Souvik Kumar Parui, and Biswajit Purkayastha. Credit risk prediction using extra trees ensemble method. In 2023 11th International Conference on Internet of Everything, Microwave Engineering, Communication and Networks (IEMECON)*, pages 1–8. IEEE, 2023.*

[21] Wenyu Qiu. Credit risk prediction in an imbalanced social lending environment based on xgboost. In 2019 5th International Conference on Big Data and Information Analytics (BigDIA)*, pages 150–156. IEEE, 2019.*

[22] Maria Aparecida Gouvêa and Eric Bacconi Gonçalves. Credit risk analysis applying logistic regression, neural networks and genetic algorithms models. In POMS 18th annual conference*, 2007.*

[23] Siddharth Bhatore, Lalit Mohan, and Y Raghu Reddy. Machine learning techniques for credit risk evaluation: a systematic literature review. Journal of Banking and Financial Technology*, 4:111–138, 2020.*