

# Final Project

Lalida Kungval

05/02/2024

## Introduction

The burgeoning field of credit scoring is pivotal to financial institutions, where accurate credit risk assessments dictate lending decisions and influence economic access. “Credit Approval Data Analysis Using Classification and Regression Models,” a seminal paper published in the *International Journal of Research and Analytical Reviews*, employs advanced statistical techniques such as logistic regression and Classification and Regression Trees (CART) to enhance decision-making processes in credit approvals. This analysis is not only technical but also deeply intertwined with ethical considerations, underscoring the dual significance of accuracy and fairness in financial assessments.

Credit scoring systems leverage a variety of data points to predict the creditworthiness of applicants, deploying models that process sociodemographic and financial data to segregate applicants into ‘good’ or ‘bad’ credit risks. This classification significantly affects individuals’ financial opportunities and is typically quantified through accuracy metrics. However, the implications of these models extend far beyond mere numerical values, influencing societal norms and individual lives.

This paper delves into the statistical methodologies applied in credit scoring, critically examining their precision and the ethical dimensions they engender. The analysis is twofold: it evaluates the statistical methods for their robustness and predictive power, and it scrutinizes the moral frameworks that underpin these methodologies. By integrating statistical analysis with philosophical inquiry, the paper aims to foster a comprehensive understanding of how credit scoring models not only predict outcomes but also influence societal norms and individual lives.

The impending sections will provide a meticulous analysis of the methods used for credit scoring, including a personal replication of results and application to novel datasets. Simultaneously, the paper will address the normative concerns associated with these models—particularly issues of bias and fairness—using established philosophical principles discussed in class. The objective is to present a balanced critique that appreciates the statistical ingenuity of credit scoring while questioning its broader ethical implications.

As financial technologies evolve, so too must our critical engagement with their impacts. This paper invites a reflective inquiry into the intersection of technology, ethics, and finance, proposing that the true measure of a model’s success extends beyond its predictive accuracy to its conformity with ethical standards. Through this analysis, the paper strives to enlighten not only the statistical community but also policymakers and practitioners about the profound responsibilities entailed in crafting algorithms that shape financial destinies.

## Analysis of Methods

Data used to replicate the methods in the paper was obtained from UCI Machine Learning Repository. The data cleaning process ensued, aligning categorical variables such as *Male*, *PriorDefault*, *Employment*, and *Approval* into a binary format. This transformation aligns with the project’s thematic focus on applying advanced statistical techniques like logistic regression and Classification and Regression Trees (CART) to

refine decision-making processes in credit approvals. Following are the first few rows of the data table and summary statistics of the variables in the data.

```
## Loading required package: ggplot2
```

```
## Loading required package: lattice
```

```
## Loading required package: Matrix
```

```
## Loaded glmnet 4.1-8
```

```
##   Male   Age  Debt Married BankCustomer EducationLevel Ethnicity YearsEmployed
## 1    0 30.83 0.000      u          g          w          v          1.25
## 2    1 58.67 4.460      u          g          q          h          3.04
## 3    1 24.50 0.500      u          g          q          h          1.50
## 4    0 27.83 1.540      u          g          w          v          3.75
## 5    0 20.17 5.625      u          g          w          v          1.71
## 6    0 32.08 4.000      u          g          m          v          2.50
```

```
##   PriorDefault Employed CreditScore DriversLicense Citizen Zipcode Income
## 1             1         1           1             f      g      202      0
## 2             1         1           6             f      g       43    560
## 3             1         0           0             f      g      280    824
## 4             1         1           5             t      g      100      3
## 5             1         0           0             f      s      120      0
## 6             1         0           0             t      g      360      0
```

```
##   Approved
```

```
## 1         1
## 2         1
## 3         1
## 4         1
## 5         1
## 6         1
```

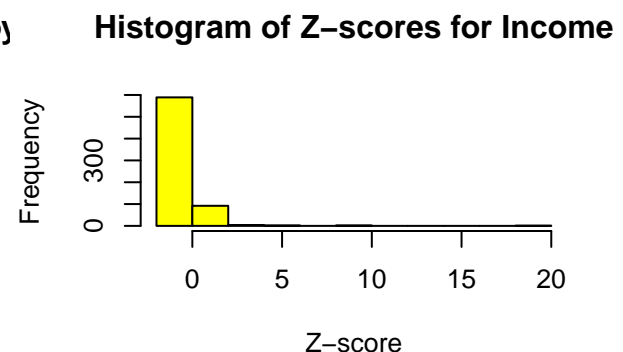
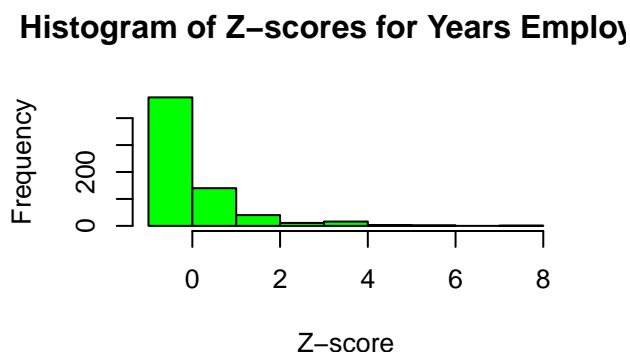
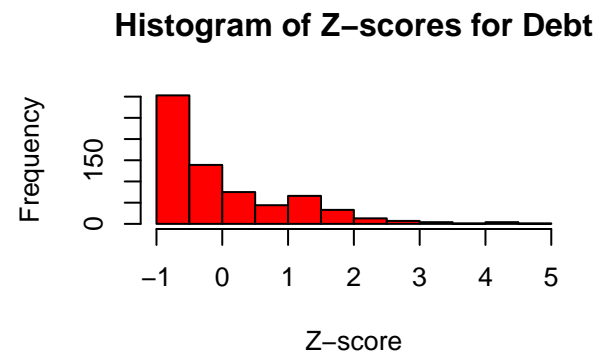
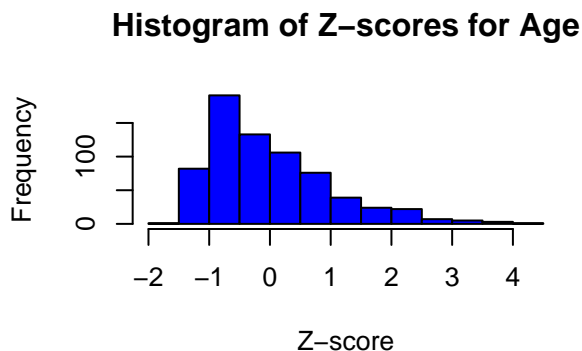
```
##           Male           Age           Debt           Married
## Min.      :0.0000   Min.      :13.75   Min.      : 0.000   Length:690
## 1st Qu.:0.0000   1st Qu.:22.60   1st Qu.: 1.000   Class :character
## Median :0.0000   Median :28.46   Median : 2.750   Mode  :character
## Mean      :0.3097   Mean      :31.57   Mean      : 4.759
## 3rd Qu.:1.0000   3rd Qu.:38.23   3rd Qu.: 7.207
## Max.      :1.0000   Max.      :80.25   Max.      :28.000
## NA's      :12     NA's      :12
```

```
## BankCustomer      EducationLevel      Ethnicity      YearsEmployed
## Length:690      Length:690      Length:690      Min.      : 0.000
## Class :character  Class :character  Class :character  1st Qu.: 0.165
## Mode  :character  Mode  :character  Mode  :character  Median : 1.000
##                                     Mean      : 2.223
##                                     3rd Qu.: 2.625
##                                     Max.      :28.500
```

```
##   PriorDefault      Employed      CreditScore      DriversLicense
## Min.      :0.0000   Min.      :0.0000   Min.      : 0.0   Length:690
## 1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.: 0.0   Class :character
```

```
## Median :1.0000   Median :0.0000   Median : 0.0   Mode  :character
## Mean   :0.5232   Mean    :0.4275   Mean    : 2.4
## 3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.: 3.0
## Max.   :1.0000   Max.    :1.0000   Max.    :67.0
##
## Citizen      Zipcode      Income      Approved
## Length:690   Min.      : 0   Min.      : 0.0   Min.      :0.0000
## Class :character 1st Qu.: 75   1st Qu.: 0.0   1st Qu.:0.0000
## Mode  :character Median : 160   Median : 5.0   Median :0.0000
##              Mean  : 184   Mean  : 1017.4   Mean  :0.4449
##              3rd Qu.: 276   3rd Qu.: 395.5   3rd Qu.:1.0000
##              Max.   :2000   Max.   :100000.0   Max.   :1.0000
##              NA's    :13
```

From the summary statistics, we can see that there are missing values. Following the methods used in the paper, missing numerical values are replaced with the mean and missing categorical values are replaced with the mode. Then the next step is to standardize the continuous variables by calculating z-scores. Below are the histograms that show the distributions of the continuous variables that have been standardized. All the variables are skewed to the right. This is supported by the skewness calculations which all result in positive values. These variables that show positive skewness are to be taken logarithm before being trained on the logistic model.



```
## [1] "Skewness for Age: 1.15092209834212"
```

```
## [1] "Skewness for Debt: 1.48234627832632"
```

```
## [1] "Skewness for Years Employed: 2.87877156820516"
```

```
## [1] "Skewness for Income: 13.0835769330132"
```

The data is then split into training and test sets using the probabilities of 75% and 25%, respectively. This means that 75% of the observations are randomly assigned to the training set, while the remaining 25% form the test set. To establish a benchmark for evaluating more complex models, a baseline model is constructed where the prediction for each application's approval is based on the mode of the training set's outcomes. This simplistic approach provides a minimal accuracy threshold against which to measure the efficacy of more sophisticated statistical models.

The baseline model is then evaluated on the test set to assess its generalization capability:

```
## [1] "Baseline Model Test Accuracy: 0.569767441860465"
```

After establishing the baseline model's accuracy, the logistic regression and CART models were developed to enhance the prediction accuracy and handle the complex nonlinear relationships potentially present in the data. The logistic regression model utilized stanardized *Age* along with transformed logarithmic variables for *Debt*, *YearsEmployed*, *CreditScore*, and *Income* to handle non-linearity and improve the model's performance. These variables are specifically used to replicate the model of the paper. The results from the logistic regression model are as follows:

```
##
## Call:
## glm(formula = Approved ~ z_Age + DebtLog + YearsEmployedLog +
##       CreditScoreLog + IncomeLog, family = binomial(link = "logit"),
##       data = trainingSet)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -2.15313    0.26642  -8.082 6.39e-16 ***
## z_Age          0.03857    0.11302   0.341  0.73291
## DebtLog        0.13668    0.13766   0.993  0.32079
## YearsEmployedLog 0.91182    0.15905   5.733 9.87e-09 ***
## CreditScoreLog  1.03908    0.14494   7.169 7.55e-13 ***
## IncomeLog       0.11582    0.03752   3.087  0.00202 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 712.87  on 517  degrees of freedom
## Residual deviance: 525.26  on 512  degrees of freedom
## AIC: 537.26
##
## Number of Fisher Scoring iterations: 5
```

The model coefficients suggest same significant relationships as the paper where *z\_Age* and *DebtLog* are not significant. All other variables have significant relationships with the the approval decision. Notably, the *CreditScoreLog* and *YearsEmployedLog* has a substantial positive effect, indicating that higher credit scores or higher years employed strongly correlate with approval chances. Following these relationships, the model is further simplified by removing *z\_Age* and *DebtLog*.

```
##
## Call:
## glm(formula = Approved ~ YearsEmployedLog + CreditScoreLog +
##      IncomeLog, family = binomial(link = "logit"), data = trainingSet)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -2.01195    0.20758  -9.692 < 2e-16 ***
## YearsEmployedLog  0.95320    0.15277   6.240 4.39e-10 ***
## CreditScoreLog   1.04709    0.14412   7.266 3.72e-13 ***
## IncomeLog       0.11793    0.03747   3.147 0.00165 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 712.87  on 517  degrees of freedom
## Residual deviance: 526.41  on 514  degrees of freedom
## AIC: 534.41
##
## Number of Fisher Scoring iterations: 4

##           Actual
## Predicted  0  1
##           0 87 24
##           1 11 50

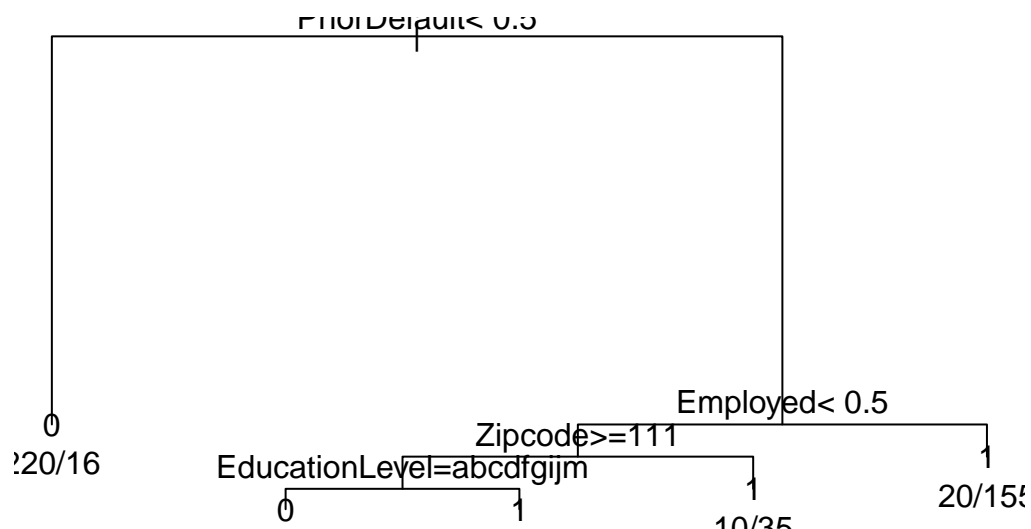
## [1] "Logistic Regression Model Accuracy: 0.796511627906977"
```

The simplified model is then tested on the test set to determine the model's accuracy. The confusion matrix for this logistic regression model reveals that while the model is fairly accurate, there are still a number of false positives and negatives that could impact the decision-making process in credit approvals. The accuracy of the logistic regression model on the test set is approximately 79.65%. This indicates a decent improvement over the baseline model and shows that the model can effectively distinguish between the classes.

The paper then fits a training set to a Classification and Regression Trees (CART) model by prioritizing the split at the variable *PriorDefault*. The CART model is then applied to a test set to evaluate its predictive performance. The model's effectiveness is initially indicated by a confusion matrix and an overall accuracy of 87.79%.

```
## NULL
```

```
## NULL
```



```
##          Actual
## Predicted  0  1
##           0 95 18
##           1  3 56
```

```
## [1] "CART Model Accuracy: 0.877906976744186"
```

```
## [1] "Confusion Matrix for Logistic Regression Model (Male):"
```

```
##          Actual
## Predicted  0  1
##           0 23  8
##           1  4 14
```

```
## [1] "Confusion Matrix for Logistic Regression Model (Non-Male):"
```

```
##          Actual
## Predicted  0  1
##           0 64 16
##           1  7 36
```

```
## [1] "Confusion Matrix for CART Model (Male):"
```

```
##           Actual
## Predicted  0  1
##           0 27  5
##           1  0 17
```

```
## [1] "Confusion Matrix for CART Model (Non-Male):"
```

```
##           Actual
## Predicted  0  1
##           0 68 13
##           1  3 39
```

To evaluate potential biases in the logistic regression and CART models used for credit scoring, I've segregated the test set data based on different variables or sociodemographic groups. The first variable is *Male* and I analyzed the predictive performance for each gender group separately. This allows us to see if the model behaves differently across genders, which could indicate bias in how credit approvals are predicted. For the logistic regression model, male applicants experienced 23 true negatives and 14 true positives, suggesting a balanced decision-making process. However, there were 8 false positives and 4 false negatives, indicating some errors in prediction. Non-male applicants had 64 true negatives and 36 true positives, with a slightly higher number of false positives (16) and false negatives (7) compared to male applicants. This might suggest that non-male applicants are more likely to be incorrectly classified as creditworthy, potentially reflecting a leniency in the model's judgment, which could be a sign of underlying bias.

The CART model displayed no false negatives for male applicants, showing a high level of certainty in approving creditworthy individuals within this group. There were 27 true negatives and 17 true positives, with only 5 false positives. In contrast, non-male applicants had 68 true negatives and 39 true positives, but also 3 false negatives and 13 false positives. The absence of false negatives for male applicants and their presence for non-male applicants might indicate that the model's criteria or decision rules are more favorable or attuned to characteristics typically associated with male applicants, leading to fewer missed opportunities for credit approval among this group.

```
## [1] "Confusion Matrices for Logistic Regression Model by Marital Status:"
```

```
## $u
##           Actual
## Predicted  0  1
##           0 58 18
##           1  9 45
```

```
## $y
##           Actual
## Predicted  0  1
##           0 29  5
##           1  2  5
```

```
## $l
##           Actual
## Predicted  1
##           0  1
```

```
## [1] "Confusion Matrices for CART Model by Marital Status:"
```

```
## $u
##           Actual
## Predicted  0  1
##           0 65 12
##           1  2 51
##
## $y
##           Actual
## Predicted  0  1
##           0 30  5
##           1  1  5
##
## $l
##           Actual
## Predicted  1
##           0  1
##           1  0
```

The confusion matrices for the logistic regression and CART models, segmented by marital status, reveal distinct patterns in how the models predict credit approval across different groups. In the logistic regression model, for individuals possibly identified as unmarried (u), there are 58 true negatives and 45 true positives, suggesting a relatively balanced approach to approvals and denials. However, there are 18 false positives and 9 false negatives, indicating some misclassifications, particularly in overestimating creditworthiness. For those possibly married (y), the model displays more conservative behavior with fewer approvals (5 true positives) and a smaller number of misclassifications (5 false positives and 2 false negatives). The data for individuals possibly legally separated (l) is very sparse, showing only a single observation, which limits any substantive analysis for this group.

Similarly, the CART model shows stronger performance in the \$u category with 65 true negatives and 51 true positives, along with fewer errors (12 false positives and 2 false negatives) compared to the logistic regression. The pattern in the y category is similar to the logistic regression with a conservative number of approvals and few errors, suggesting cautious credit evaluation for married individuals.

The differences in model performance by marital status suggest that there might be an inherent bias, particularly in favoring the unmarried group with more accurate predictions. The logistic model's higher false positive rate for the unmarried may point to an overestimation of creditworthiness, potentially leading to financial risks for lenders. This discrepancy highlights the need for recalibration of the models to ensure fair treatment across all marital statuses.

```
## [1] "Confusion Matrices for Logistic Regression Model by Age Group:"
```

```
## $Young
##           Actual
## Predicted  0  1
##           0 54 16
##           1  7 20
##
## $'Middle-aged'
##           Actual
## Predicted  0  1
##           0 33  7
##           1  4 28
##
## $Senior
```



```

##           Actual
## Predicted 1
##           0 1
##           1 2

## [1] "Confusion Matrices for CART Model by Age Group:"

## $Young
##           Actual
## Predicted 0 1
##           0 61 11
##           1 0 25
##
## $'Middle-aged'
##           Actual
## Predicted 0 1
##           0 34 7
##           1 3 28
##
## $Senior
##           Actual
## Predicted 1
##           0 0
##           1 3

```

The confusion matrices for the logistic regression and CART models, segmented by age groups, provide insights into how each model performs across different age demographics. In the logistic regression model, the young group shows a relatively balanced approach to credit approval with 54 true negatives and 20 true positives, although there are 16 false positives indicating a tendency to overestimate creditworthiness among younger individuals. The model performs well among the middle-aged with 33 true negatives and 28 true positives, demonstrating a robust ability to identify creditworthy individuals while maintaining a low rate of false negatives and positives. However, the senior group presents limited data with only a few observations, which includes 1 true negative and 2 false negatives, suggesting a potential underestimation of creditworthiness but the small sample size makes definitive conclusions difficult.

The CART model mirrors some of these patterns but with notable differences. For the young group, it shows high sensitivity with no false negatives and 25 true positives, albeit at the cost of 11 false positives, pointing to a similar overestimation of creditworthiness as seen in the logistic model. The middle-aged group again shows effective model performance with 34 true negatives and 28 true positives, and only a small number of false negatives and positives, indicating the model's efficacy in accurately assessing credit approval. Data for seniors in the CART model is also sparse, featuring only 3 true positives, which does not provide enough information for a comprehensive evaluation.

The performance across these age groups reveals strengths in both models for identifying creditworthy individuals, particularly in the middle-aged category. However, the variability in model performance, especially with the younger individuals experiencing more false positives and the senior group lacking sufficient data for evaluation, points to areas needing further attention. Both models could benefit from additional tuning and possibly a reassessment of how age-related features are weighted. Enhancing the fairness and accuracy of credit scoring algorithms through better-balanced data representation across all age groups and refined model configurations could lead to more equitable financial decision-making.

```

## [1] "Confusion Matrices for Logistic Regression Model by Ethnicity:"

## $v

```

```

##           Actual
## Predicted  0  1
##           0 60 11
##           1  7 26
##
## $h
##           Actual
## Predicted  0  1
##           0 10  9
##           1  2 16
##
## $bb
##           Actual
## Predicted 0  1
##           0 9 0
##           1 1 4
##
## $ff
##           Actual
## Predicted 0  1
##           0 5 1
##           1 0 2
##
## $z
##           Actual
## Predicted 1
##           0 1
##           1 1
##
## $o
##           Actual
## Predicted 1
##           0 1
##
## $j
##           Actual
## Predicted 0  1
##           0 1 1
##
## $dd
##           Actual
## Predicted 0
##           0 1
##           1 1
##
## $n
##           Actual
## Predicted 0  1
##           0 1 0
##           1 0 1

## [1] "Confusion Matrices for CART Model by Ethnicity:"

## $v

```

```

##           Actual
## Predicted  0  1
##           0 66  8
##           1  1 29
##
## $h
##           Actual
## Predicted  0  1
##           0 11  8
##           1  1 17
##
## $bb
##           Actual
## Predicted 0  1
##           0 9  0
##           1 1  4
##
## $ff
##           Actual
## Predicted 0  1
##           0 5  0
##           1 0  3
##
## $z
##           Actual
## Predicted 1
##           0 0
##           1 2
##
## $o
##           Actual
## Predicted 1
##           0 1
##           1 0
##
## $j
##           Actual
## Predicted 0  1
##           0 1  1
##           1 0  0
##
## $dd
##           Actual
## Predicted 0
##           0 2
##           1 0
##
## $n
##           Actual
## Predicted 0  1
##           0 1  0
##           1 0  1

```

The confusion matrices for both the logistic regression and CART models, categorized by ethnicity, reveal

variations in how each model handles credit approval decisions across different ethnic groups. This detailed analysis highlights the complexities and challenges in ensuring fair treatment across all demographics.

In the logistic regression model, ethnicity v demonstrates a balance with 60 true negatives and 26 true positives, although the presence of 11 false positives suggests a slight tendency towards overestimating creditworthiness. Ethnicity h shows balanced results with 10 true negatives and 16 true positives, but the relatively high number of 9 false positives might indicate a potential bias towards more favorable outcomes. For ethnicity bb, the model performs effectively with no false negatives and a good balance of true negatives and positives. Ethnicity ff displays a conservative approach with few approvals, though the small numbers suggest limited data. Other groups such as z, o, j, dd, and n suffer from sparse data, leading to variability in predictions ranging from perfect classifications to mixed results with both true and false positives.

The CART model reflects a similar pattern but with subtle differences. Ethnicity v shows robust performance with 66 true negatives and 29 true positives, and minimal errors, indicating high sensitivity and precision. Ethnicity h also shows improvement over the logistic model, with fewer false positives enhancing the prediction accuracy. Consistencies with the logistic model are observed in ethnicities bb and ff, where the CART model also demonstrates good accuracy. However, for the groups with sparse data, the performance varies significantly, from no errors in some groups to an even split in true and false classifications in others, which hampers a comprehensive evaluation.

The analysis indicates that while both models generally perform well for certain ethnic categories, demonstrating the capability to effectively distinguish between creditworthy and non-creditworthy individuals, there is noticeable variability, especially among groups with limited data. This inconsistency might be due to the representation of ethnic data or the sample size available for each group, raising concerns about the uniform application of decision-making processes across all ethnic groups.

## Analysis of Normative Consideration

The ethical dimension of automated credit scoring systems emerges prominently in the analysis of logistic regression and CART models used for determining creditworthiness. These systems are crucial for financial institutions, yet they carry significant ethical implications that necessitate rigorous scrutiny. The potential for biases and discrimination highlighted by the predictive performance discrepancies across different demographic groups underscores the pressing need for fairness and transparency in credit lending practices.

The variations in model performance across gender, marital status, age, and ethnicity raise ethical concerns regarding fairness. For instance, the observed tendency for models to favor certain demographic groups, such as unmarried individuals or specific ethnicities, suggests potential systemic biases. These biases can perpetuate inequality and restrict financial opportunities for unfairly disadvantaged groups. Ethical credit scoring must aim for impartiality, ensuring that all individuals are assessed based on their genuine creditworthiness without prejudice rooted in demographic characteristics.

The complexities and potential biases inherent in statistical models used for credit scoring also highlight the importance of transparency and accountability in their deployment. Financial institutions must ensure that these models are not only accurate but also understandable and fair. This involves clear communication about how credit decisions are made and allowing individuals the opportunity to contest decisions that may adversely affect them. Transparency fosters trust and ensures that applicants are aware of the factors influencing their credit evaluations.

Ethically, the use of credit scoring models touches on principles of justice and respect for persons. From a deontological perspective, ensuring that each individual is treated as an end in themselves requires that credit scoring systems do not reduce people merely to their demographic identifiers. Furthermore, from a utilitarian perspective, the broad societal impact of these systems, which can either support or hinder economic access, must be considered. Ensuring that these systems promote overall well-being involves rigorous testing for biases and continual adjustments to safeguard against unintended harmful consequences.

The critique of the methodologies and their implications must be grounded in ethical principles discussed in class, such as fairness, justice, and respect for persons. The normative analysis should critically examine

how well the credit scoring practices align with these principles and where they fall short. This examination should not only point out the deficiencies but also suggest ways in which credit scoring systems might be improved to better adhere to ethical norms.

This analysis should ultimately inform policy and practice, guiding the development of more equitable credit scoring models. Policymakers and practitioners should consider the ethical implications of deploying automated systems in credit decisions. Recommendations might include the implementation of more robust anti-discrimination measures, the development of fairness-enhancing technologies, and greater regulatory oversight to ensure that these technologies serve the public interest without exacerbating social inequalities.

## Conclusion

The analysis of logistic regression and CART models in credit scoring has highlighted significant disparities in performance across different demographic groups, revealing both statistical and ethical challenges. These findings underscore the necessity for ongoing adjustments to enhance fairness and reduce biases in automated credit decision-making systems. Ethical considerations, particularly concerning justice and equity, are paramount as they reflect the potential of these models to influence societal norms and individual opportunities significantly.

This paper advocates for an integrated approach that marries rigorous statistical methodologies with robust ethical scrutiny. It calls for financial institutions and policymakers to adopt more sophisticated techniques that not only improve predictive accuracy but also ensure that these advances in financial technology do not perpetuate social inequalities. Furthermore, the engagement of various stakeholders in a dialogue about the technical and moral dimensions of credit scoring is crucial for refining these models in line with evolving societal values.

Ultimately, the success of credit scoring models should be measured not only by their statistical accuracy but also by their adherence to ethical standards. This analysis provides a foundation for future enhancements, aiming to align technological advancements in credit scoring with broader commitments to fairness and justice.