5/1/2024

Gender Fairness in Predictive Modelling: A Case Study in Car Insurance

Introduction

Predictive models are critical for evaluating risk and managing claims for both policyholders and insurance firms. While these models provide valuable insights, their fairness, particularly concerning gender equity, has increasingly come under scrutiny. Lindholm et al. (2022) investigated discrimination and fairness in insurance pricing, emphasising the importance of transparency to promote equity. Similarly, Xin and Huang (2023) advocate for the adoption of fairness standards and governmental actions to mitigate bias in these pricing models. Gender bias in car insurance claims prediction models may stem from systemic discrimination, cultural stereotypes, and historical data imbalances. Such biases can manifest in several ways, including the overestimation of risk for one gender while underestimating it for another, leading to gender-based disparities in pricing and claim determinations. This could impact the fairness of financial charges as well as influence the trust and treatment of policyholders. Hence, this study aims to rigorously analyse the presence of gender bias in the car insurance claims prediction model.

The necessity for unbiased models in the insurance industry cannot be overstated. They ensure equitable treatment of all policyholders, irrespective of gender, enhancing consumer trust and promoting fairness. Moreover, unbiased models help insurance companies maintain their reputations and adhere to increasingly stringent legal standards regarding the fairness and transparency of insurance practices. Addressing gender bias in car insurance claim prediction algorithms is crucial for maintaining policyholder trust and preventing discriminatory practices. Failure to do so risks damaging the reputations of insurance firms, leading to potential business losses and legal repercussions. For the long-term sustainability and ethical integrity of the insurance sector, eliminating gender bias is imperative.

Model Development

Dataset Description

The dataset used in this study was sourced from Kaggle providing comprehensive insights into various attributes related to individuals and their car insurance details. It includes 10,000 observations across 19 features. Table 1 in the appendix presents a comprehensive description to the dataset. Gender was designated as a protected variable due to its implications in fairness considerations, which plays a crucial role in this experiment. Gender discrimination can skew the outcomes in the processing of insurance claims and premium settings, making the guarantee of gender equity essential when assessing and modelling insurance claim predictions.

Data Exploration

The initial step involved checking for missing values within the dataset. This inspection revealed the presence of missing values as seen in Figure 1. This could potentially affect the ML algorithm.

```
In [9]: #Checking number of null in each column
        Car.isnull().sum()

Out[9]: ID                      0
        AGE                     0
        GENDER                  0
        RACE                    0
        DRIVING_EXPERIENCE      0
        EDUCATION               0
        INCOME                  0
        CREDIT_SCORE          982
        VEHICLE_OWNERSHIP       0
        VEHICLE_YEAR            0
        MARRIED                 0
        CHILDREN                0
        POSTAL_CODE             0
        ANNUAL_MILEAGE        957
        VEHICLE_TYPE            0
        SPEEDING_VIOLATIONS     0
        DUIS                    0
        PAST_ACCIDENTS          0
        OUTCOME                 0
        dtype: int64
```

Fig 1: Checking for missing values in the dataset.

Figure 2 shows the distributions of features were not significantly skewed. This indicated that there was no need to address outliers in the dataset. This uniformity ensures that the assessment of the

model will be based on representative and unbiased data.

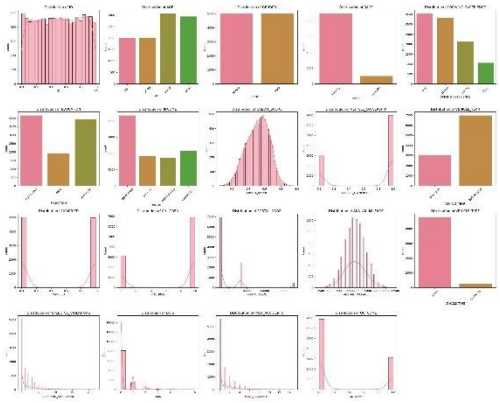

Fig 2: Distribution of Features in the dataset.

The plot of gender distribution showed a balanced ratio of male to female as seen in Figure 3. This indicates that all individuals were fairly represented, which is crucial for promoting fairness in model predictions. The absence of significant biases or imbalances, particularly concerning gender, is critical to prevent unfair treatment of groups and skewed model results. The equitable representation of genders supports the fairness of the insurance claim prediction modeling process, thereby reducing the risk of biased outcomes.

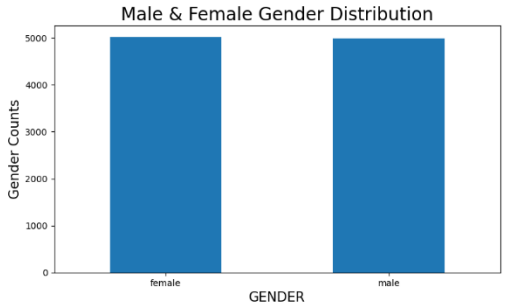

Fig 3: Distribution of Protected Column

Data Cleaning

Columns such as ID and Postal Code were removed from the dataset. These attributes were deemed unnecessary for predictive modelling and their exclusion helps increase the efficiency of the model efficiency by reducing the input dimensionality. Also, categorical features within the dataset were encoded into numerical representations. This conversion is essential as most ML algorithms require numerical input. The encoding ensures that categorical data are effectively incorporated into the model. Additionally, missing values in numerical columns were addressed through median imputation, which provides a comprehensive estimate less sensitive to outliers than the mean. This choice enhances the reliability of the dataset for modelling. Subsequent checks for duplicate entries led to the identification and removal of twenty-four duplicate records as illustrated in Figure 4. Removing these duplicates is crucial for maintaining data integrity and ensuring that each observation contributes uniquely to the algorithm.

```
In [49]: # Checking duplicates from dataset
         Car.duplicated().sum()

Out[49]: 24

In [50]: #Removing Duplicates from dataset
         Car.drop_duplicates(inplace=True)

In [51]: # Checking duplicates from dataset after removing duplicates
         Car.duplicated().sum()

Out[51]: 0
```

Fig 4: Removal of duplicates after replacing missing values.

A review of the initial records of the dataset was conducted to ensure the data cleaning processes were correctly applied as shown in Figure 5.

```
Car.head()
```

	AGE	GENDER	RACE	DRIVING_EXPERIENCE	EDUCATION	INCOME	CREDIT_SCORE	VEHICLE_OWNERSHIP	VEHICLE_YEAR	MARRIED	CHILDREN	ANNU
0	3	0	0	0	0	2	0.629027	1	0	0	1	
1	0	1	0	0	1	1	0.357757	0	1	0	0	
2	0	0	0	0	0	3	0.493146	1	1	0	0	
3	0	1	0	0	2	3	0.206013	1	1	0	1	
4	1	1	0	1	1	3	0.388366	1	1	0	0	

Fig 5: Head of the dataset after cleaning

Multi-Collinearity Analysis

To assess the multicollinearity among features in the dataset, a thorough correlation analysis was conducted. The results indicated the absence of highly correlated variables as shown in Figure 6, suggesting minimal redundancy within the feature set. This is crucial as it enhances the performance of the predictive modelling by ensuring that each variable contributes unique information to the prediction of insurance claim outcomes. The lack of significant correlations among the variables allows the modelling process

to proceed with confidence, as it minimises the risk of collinearity affecting the accuracy and interpretability of the algorithm.

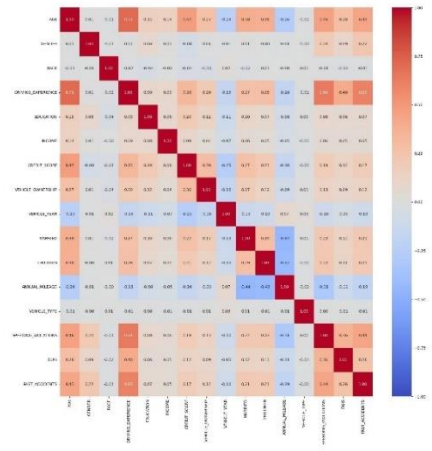

Fig 6: Correlation Analysis

Data Splitting

The dataset was divided into training and testing sets using an 80-20 ratio as seen in Figure 7. This allocation ensures that 80% of the data is utilised for training the model, which allows the model to learn from a diverse and comprehensive set of examples. The remaining 20% is set aside for testing, enabling assessment of the performance of the model on new, untested data. This standard splitting ratio is pivotal for validating the ability of the model to generalise, which is crucial for its application in real-world scenarios. The experiment enhances the reliability and dependability of the model in predictive tasks by training on a broad base of data and testing on separate data.

Data Splitting

```
# Splitting the data into training and testing sets
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.2, stratify=Y, random_state=2)


# Checking the shape of the resulting sets
print("Shape of X_train:", X_train.shape)
print("Shape of X_test:", X_test.shape)
print("Shape of Y_train:", Y_train.shape)
print("Shape of Y_test:", Y_test.shape)
```

```
Shape of X_train: (7980, 16)
Shape of X_test: (1996, 16)
Shape of Y_train: (7980,)
Shape of Y_test: (1996,)
```

Fig 7: Data Splitting

Normalisation

Normalisation was done to the input features in both sets, identified as X_train and X_test, using the MinMaxScaler. This method scales the feature values to a consistent range, typically between 0 and 1 as illustrated in Figure 8, which is crucial for preventing any one feature with larger numerical ranges from dominating the learning process of the model. Normalisation ensures that all features contribute equally to the predictions of the model, thereby enhancing the fairness and effectiveness of the training.

```
X_train.head()
```

	AGE	GENDER	RACE	DRIVING_EXPERIENCE	EDUCATION	INCOME	CREDIT_SCORE	VEHICLE_OWNERSHIP	VEHICLE_YEAR	MARRIED	CHILDREN	A
0	1.000000	1.0	0.0	1.000000	1.0	0.666667	0.430315	1.0	0.0	0.0	1.0	
1	0.000000	1.0	0.0	0.000000	0.0	0.333333	0.262278	1.0	1.0	0.0	1.0	
2	0.333333	1.0	0.0	0.333333	0.5	0.333333	0.366406	0.0	1.0	0.0	1.0	
3	1.000000	1.0	0.0	1.000000	0.0	0.666667	0.737907	1.0	1.0	1.0	1.0	
4	0.000000	1.0	0.0	0.000000	0.0	0.000000	0.597594	0.0	1.0	1.0	0.0	

Fig 8: Head of the dataset after normalisation.

Machine Learning Algorithm

The Random Forest algorithm was chosen for this experiment. The Random Forest algorithm is particularly well-suited for predicting insurance claims due to its capability to handle mixed data types, its resistance to overfitting, and its ability to model non-linear relationships. This algorithm can effectively process both numerical and categorical features, making it highly versatile across various data structures. Unlike single decision trees, which are prone to overfitting the training data, Random Forest reduces this risk by averaging the results of multiple trees, thereby enhancing the generalisability of the results. Moreover, it is adept at capturing complex, nonlinear interactions between features, which are commonly present in real-world datasets such as those involved in insurance claim predictions.

Performance Evaluation

The performance of the Random Forest model, utilised for predicting car insurance claims, is assessed using a comprehensive set of evaluation metrics.

Accuracy and Classification Report

Accuracy provides a general indication of the overall correctness of a model in making

predictions. The classification report as presented in Figure 10 offers an extensive assessment, detailing precision, recall, and other crucial metrics for each class. This helps in understanding the performance of the model across various categories. After applying the trained model to the test data, it achieved an accuracy of 81% as seen in Figure 9.

Random Forest Classification Report:

	precision	recall	f1-score	support
0	0.84	0.88	0.86	1373
1	0.71	0.64	0.67	623
accuracy			0.81	1996
macro avg	0.78	0.76	0.77	1996
weighted avg	0.80	0.81	0.80	1996

Fig 9: Classification Report.

Confusion Matrix

The confusion matrix provides valuable insights into the performance of the model by displaying the counts of true positives, true negatives, false positives, and false negatives as shown in Figure 10. Specifically, the model correctly identified 1,233 cases as negative (True Negative) and 413 cases as positive (True Positive), demonstrating its effectiveness in recognising both negative and positive outcomes. However, there were 140 instances of false positives and 210 false negatives, highlighting areas where the model may require further tuning.

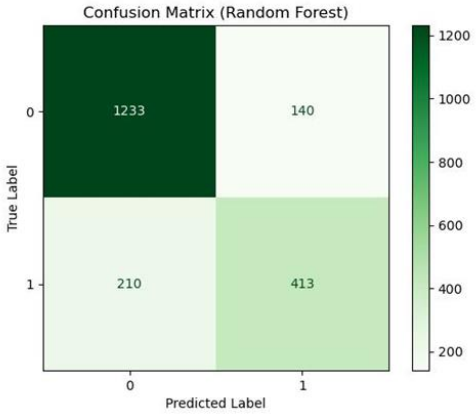

Fig 10: Confusion Matrix.

Receiver Operating Characteristic (ROC) Curve

The ROC curve and the Area Under the Curve (AUC) score are used to evaluate the discriminating ability of a model between positive and negative outcomes. An AUC of 0.78 as seen in Figure 11 indicates a reasonably good performance, confirming the capability of the model to distinguish between the classes effectively.

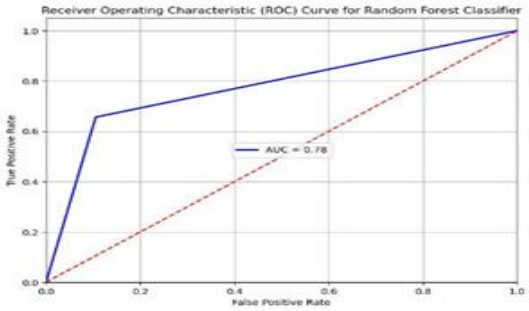

Fig 11: ROC-AUC Curve.

Fairness Criteria

To examine bias and promote fairness, the data was segmented by gender (male and female), and the Random Forest model was independently run on each subgroup. This experiment utilised three fairness criteria—equal opportunity, demographic parity (group fairness), and equal accuracy—to analyse potential biases between these groups.

Model Performance on Males

The performance evaluation of the Random Forest model on the male subgroup showed an accuracy of 83%, with recall at 74% and precision at 75% as shown in Figure 12. These metrics indicate the effectiveness of the model in correctly predicting outcomes for the male group.

```
Calculated Accuracy =  0.829465186680121
Calculated Recall =  0.7388724035608308
Calculated Precision =  0.7545454545454545
```

Fig 12: Calculated Metrics for Male Groups.

The classification report in Figure 13 reveals a positive rate of 33%, suggesting how frequently the model predicts positive outcomes within this subgroup.

```
Random Forest Classification Report:
              precision    recall  f1-score   support

           0       0.87      0.88      0.87       654
           1       0.76      0.74      0.75       337

    accuracy                           0.83       991
   macro avg       0.81      0.81      0.81       991
weighted avg       0.83      0.83      0.83       991

Positive Rate 0.32795156407669024
```

Fig 13: Classification Report for Male Groups.

The confusion matrix details the following counts: 247 True Positives, 581 True Negatives, 73 False Positives, and 90 False Negatives as presented in Figure 14. These numbers provide a deeper insight into the ability of the model to correctly classify both positive and negative outcomes.

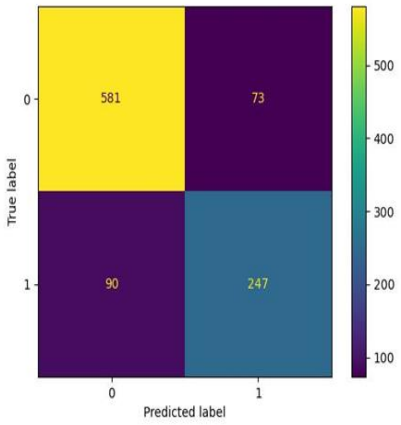

Fig 14: Confusion Metrics for Male Groups.

Model Performance on Females

Figure 15 revealed that the female subgroup yielded an accuracy of 81%, with recall at 58% and precision at 71%. This demonstrates a slightly lower performance compared to the male subgroup.

```
Calculated Accuracy =  0.8149253731343283
Calculated Recall =  0.5769230769230769
Calculated Precision =  0.717391304347826
```

Fig 15: Calculated Metrics for Female Groups.

The classification report for the female group shows a positive rate of 24% as seen in Figure 16. This indicates a lower likelihood of predicting positive outcomes compared to males.

```
Random Forest Classification Report:
              precision    recall  f1-score   support

           0       0.85      0.90      0.87       719
           1       0.71      0.59      0.65       286

    accuracy                           0.81      1005
   macro avg       0.78      0.75      0.76      1005
weighted avg       0.81      0.81      0.81      1005

Positive Rate 0.23681592039800994
```

Fig 16: Classification Report for Female Groups.

Figure 17 reveals that the female group records 166 True Positives, 652 True Negatives, 67 False Positives, and 120 False Negatives. These figures highlight the specific challenges of the model in predicting outcomes accurately for females, particularly in terms of higher false negative rates.

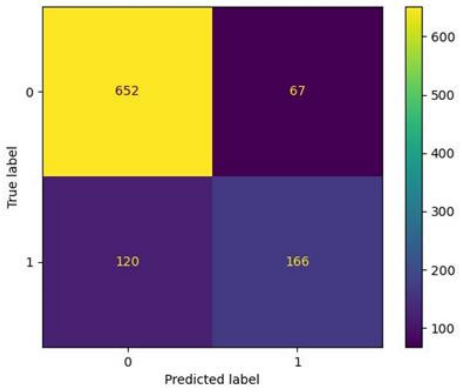

Fig 17: Confusion Metrics for Female Groups.

Fairness Criteria Application (Findings)

Equal Accuracy

Equal accuracy ensures the model performs equally well regardless of gender.

Evidence

Equal Accuracy for Males; 0.829
Equal Accuracy for Females; 0.814

In Figure 18, Equal Accuracy scores indicate a slight but noticeable difference in model accuracy, with 81% of females and 82% of males scoring higher. Despite the small difference, it suggests a tiny preference for male candidates, which may account for variations in acceptance rates or insurance premium computations based on gender.

Equal Accuracy accross the Protected Group

```
# Computing values for males
TN_m=572
FP_m=82
FN_m=87
TP_m=250

# Computing values for females
TN_f=650
FP_f=69
FN_f=118
TP_f=168

# Calculating Equal Accuracy for Males
total_predictions_m = TN_m + FP_m + FN_m + TP_m
equal_accuracy_m = (TP_m + TN_m)/ total_predictions_m

# Calculating Equal Accuracy for Females
total_predictions_f = TN_f + FP_f + FN_f + TP_f
equal_accuracy_f = (TP_f + TN_f)/ total_predictions_f

# Printing the result
print("Equal Accuracy for Males;", equal_accuracy_m)
print("Equal Accuracy for Females;", equal_accuracy_f)
```

```
Equal Accuracy for Males; 0.829465186680121
Equal Accuracy for Females; 0.8139303482587065
```

Fig 18: Equal Accuracy across groups.

Demographic Parity

Demographic Parity ensures similar positive outcome rates for all gender groupings.

Evidence

Demographic Parity for Males; 0.335
Demographic Parity for Females; 0.236

Demographic Parity (Group Fairness) across the Protected Group

```
# Computing values for males
TN_m=572
FP_m=82
FN_m=87
TP_m=250

# Computing values for females
TN_f=650
FP_f=69
FN_f=118
TP_f=168

# Calculating Demographic Parity for Males
total_predictions_m = TN_m + FP_m + FN_m + TP_m
Demograpghic_Parity_m = (TP_m + FP_m)/ total_predictions_m

# Calculating Demographic Parity for Females
total_predictions_f = TN_f + FP_f + FN_f + TP_f
Demograpghic_Parity_f =(TP_f + FP_f)/ total_predictions_f


# Printing the result
print("Demographic Parity for Males;", Demograpghic_Parity_m)
print("Demographic Parity for Females;", Demograpghic_Parity_f)
```

```
Demograpghic Parity for Males; 0.3350151362260343
Demograpghic Parity for Females; 0.23582089552238805
```

Fig 19: Demographic Parity across groups.

Concerning Group Fairness scores as shown in figure 19, where males scored 33% and females scored 24%, the differences are even more noticeable. This 9%-point difference shows a substantial bias and suggests that the model is more likely to produce positive results for men. Unfair treatment in policy offerings and pricing are two essential elements of ethical insurance practices that could result from such bias.

Equal Opportunities

Equal opportunity guarantees that male and female subgroups will recall favourable outcomes at comparable rates.

Evidence

Equal Opportunity for Males; 0.742
Equal Opportunity for Females; 0.587

Equal Opportunity (Recall) across the Protected Group

```
# Computing values for males
FN_m=87
TP_m=250

# Computing values for females
FN_f=118
TP_f=168

# Calculating Equal Opportunity for Males
equal_opportunity_m = (TP_m)/(FN_m+TP_m)

# Calculating Equal Opportunity for Females
equal_opportunity_f = (TP_f)/(FN_f+TP_f)


# Printing the result
print("Equal Opportunity for Males;", equal_opportunity_m)
print("Equal Opportunity for Females;", equal_opportunity_f)
```

```
Equal Opportunity for Males; 0.7418397626112759
Equal Opportunity for Females; 0.5874125874125874
```

Fig 20: Equal Opportunity across groups.

The inequality in equal opportunity as shown in Figure 20 is even more alarming, with men accurately predicting good outcomes at a rate of 74% compared to women's lower rate of 58%. The substantial difference of 15 percentage points suggests that the model's ability to identify eligible female applicants is inferior to that of their male counterparts.

Conclusion

These differences, particularly in Group Fairness and Equal Opportunity, show that the model's predictions hold bias and highlight the need for corrective actions to guarantee fair treatment for all protected groups. Regarding future studies and practical applications in the insurance sector, these findings bear important significance. Firstly, to guarantee justice and equity, bias in predictive models must be addressed. Future work should focus on creating algorithms that are impartial and fair by nature, by adding more fairness restrictions to the model's training process or investigating innovative methods for bias reduction. Secondly, continuous observation and assessment of prediction models are necessary for real-life situations. Insurance

companies have a responsibility to perform periodic fairness and bias assessments of their models and to implement strong fairness measures to avoid discriminatory outcomes and guarantee adherence to legal and ethical requirements. More importantly, putting fairness first in studies and practical applications can result in more open, responsible, and fair insurance policies that are helpful to both people and society at large.

References

Xin, X. and Huang, F., 2023. Antidiscrimination insurance pricing: Regulations, fairness criteria, and models. North American Actuarial Journal, pp.1-35.

Lindholm, M., Richman, R., Tsanakas, A. and Wüthrich, M.V., 2022. A discussion of discrimination and fairness in insurance pricing. arXiv preprint arXiv:2209.00858.

Appendix

Features	Description
ID	Unique number for each customer
AGE	Age group of customers
GENDER	Gender of customer
RACE	Is the customer from the majority or minority race
DRIVING EXPERIENCE	Customer's driving experience years
EDUCATION	Customer's education level
INCOME	What income class does the customer belong to
CREDIT SCORE	Customer's credit score
VEHICLE OWNERSHIP	Is the driver the owner of the car?
VEHICLE YEAR	When was the vehicle released?
MARITAL STATUS	Is the customer married or not?
CHILDREN	Does the customer have kids?
POSTAL CODE	Post code of Customer
ANNUAL MILEAGE	Customer's annual mileage
VEHICLE TYPE	What type of vehicle does the customer drive?
SPEEDING VIOLATIONS	How many speeding violations does the customer have
DUIS (Driving Under the Influence)	How many times has the customer been charged for driving under the influence?
PAST ACCIDENTS	How many accidents has the customer had in the past?
OUTCOME	What is the outcome of the insurance claim?