HW 4: Fairness and Classification

Details:

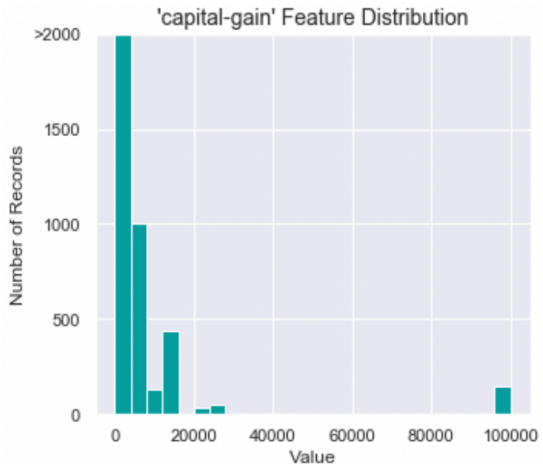
Name : Siddhanth Kalyanpur
Miner username : mb13
Miner Score : 0.86
Miner Rank : 33

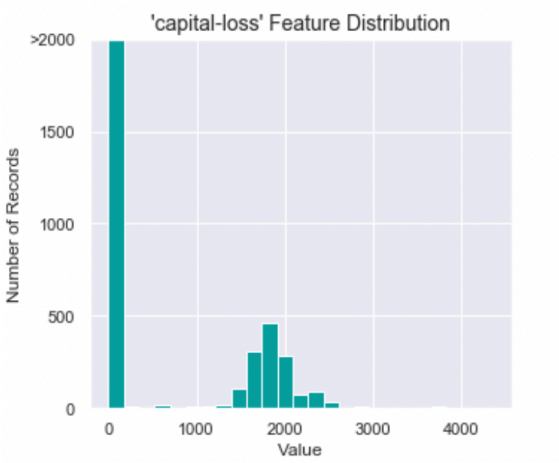
Part 1 : Prediction

1. Preprocessing

- Adult Data set doesn't have any Nan as those are replaced by "?". I checked for the later and replaced them with NaN and dropped the row with Nan.

- The data is skewed for "Capital-gain" and "Capital-loss" hence applying Log transformation to these columns so it doesn't affect our prediction.





- As a part of Feature Engineering I have used LabelEncoder class to normalise labels as they were categorical and thus transformed non-numerical labels to numerical labels.

- Dropped the last column (Income) as our target feature and user list traversal to convert labels above 50k as "1" else "0".

- Used StandardScaler for scaling all the features to a common scale.

- Maximum Accuracy achieved was 86% with the above feature engineering.

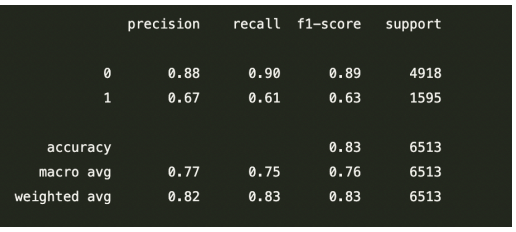
- Please find below in the table the performance of each classifier and it's hyper parameter tuning. The best classifiers are highlighted.

2. Classification

Classifier	Accuracy	Hyper Parameter Tuning
KNN	0.84	Default values used.
Logistic Regression	0.85	No Hyper Parameter Tuning. Tried Feature reduction(PCA) with this but reduced the accuracy.
Decision Tree	0.85	Since the first 10 features contribute to 90% of the variance tuned in depth feature between 7-10.
Random Forest	0.86	Since the Target variables are imbalanced used class weights of 1.5:1 for 1:0. Used an array of n_estimators and max_features and identified best parameters with trial and error.
SVM	0.82	Used the iteration variable from a list of iterators . Found best result for 1000.
Ada Boost	0.83	Used underlying Decision Tree as underlying estimator and 100 estimators.
Cat Boost	0.86	Use learning rate to reduce gradient step. learning rate = 0/04

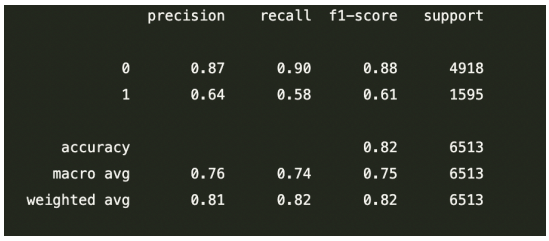
- The classification report for individual classifiers can be found in below. F-1 score, Accuracy and Recall can be found in the below snap shots of the classification report of each classifier.

- KNN



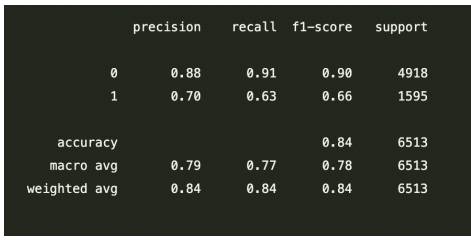
	precision	recall	f1-score	support
0	0.88	0.90	0.89	4918
1	0.67	0.61	0.63	1595
accuracy			0.83	6513
macro avg	0.77	0.75	0.76	6513
weighted avg	0.82	0.83	0.83	6513

Ada Boost



	precision	recall	f1-score	support
0	0.87	0.90	0.88	4918
1	0.64	0.58	0.61	1595
accuracy			0.82	6513
macro avg	0.76	0.74	0.75	6513
weighted avg	0.81	0.82	0.82	6513

Random Forest



	precision	recall	f1-score	support
0	0.88	0.91	0.90	4918
1	0.70	0.63	0.66	1595
accuracy			0.84	6513
macro avg	0.79	0.77	0.78	6513
weighted avg	0.84	0.84	0.84	6513

- Cat Boost

Decision Tree

accuracy			0.84	6513
macro avg	0.79	0.77	0.78	6513
weighted avg	0.84	0.84	0.84	6513

	precision	recall	f1-score	support
0	0.86	0.94	0.90	4918
1	0.76	0.54	0.63	1595
accuracy			0.85	6513
macro avg	0.81	0.74	0.77	6513
weighted avg	0.84	0.85	0.84	6513

- **Conclusion: Random Forest and Cat Boost gave me the best accuracy of 86%.**

Part 2 : Fairness Diagnosis

- We use the same pre processed data from part 1 and calculate the Demographic Disparity, inequality of odds and Equal opportunity for the sensitive features (gender ,race).
- You can find the Fairness diagnosis for each classifier in table as seen below :

Sensitive Feature	Fairness Diagnosis	Logistic Regression	Cat Boost	Ada Boost	KNN	Random Forest	Decision Tree
Race	Demographic Disparity	0.096	0.096	0.093	0.117	0.096	0.056
	Equal Opportunity	0.098	0.098	0.015	0.191	0.098	0.015
	Inequality of Odds	0.091	0.098	0.015	0.191	0.098	0.015
Sex	Demographic Disparity	0.175	0.175	0.165	0.185	0.175	0.098
	Equal Opportunity	0.073	0.073	0.066	0.111	0.073	0.066
	Inequality of Odds	0.071	0.073	0.066	0.111	0.073	0.066

- The high Disparity metrics on the prediction shows that there is some gender bias. The male population have a higher chance of getting an income of >50000 than the female population.
- The opportunity results show that the model correctly predicts the male population having income greater than 50000 showing a bias. Logistic regression has the worst prediction. Logistic regression has the maximum difference in opportunity showing heavy bias in favour of the majority gender(male).
- Decision tree and Ada Boost have low bias than random forest but poor accuracy and f1-score.
- Race also has some bias in favour of the majority class but, it is less compared to gender.

Part 3 : Fairness Mitigation

- Part 1:

- We remove the sensitive attributes (Sex, Race) and and find the Fairness measures on the best classifier(Random Forest)
- We find the below results for Disparity, odds and equality and we find that the bias has reduced slightly more so for sex than Race. The accuracy, F-1 and recall are similar to the prediction before removing the sensitive features.

```
Calculating disparity for  sex
count of classes [0,1]:  [2164, 4349]
count of 1's in classes [0,1]:  [224, 1330]
probability for class 0  is:  0.10351201478743069
probability for class 1  is:  0.3058174292940906

Disparity for attribute sex :  0.20230541450665993
None
Calculating disparity for  race
count of classes [0,1]:  [56, 216]
count of 1's in classes [0,1]:  [7, 53]
probability for class 0  is:  0.125
probability for class 1  is:  0.24537037037037038

Disparity for attribute race :  0.12037037037037038
None
```

```
equal opportunity for attribute sex:  0.10489429628626867
equal opportunity for attribute race:  0.025000000000000022
```

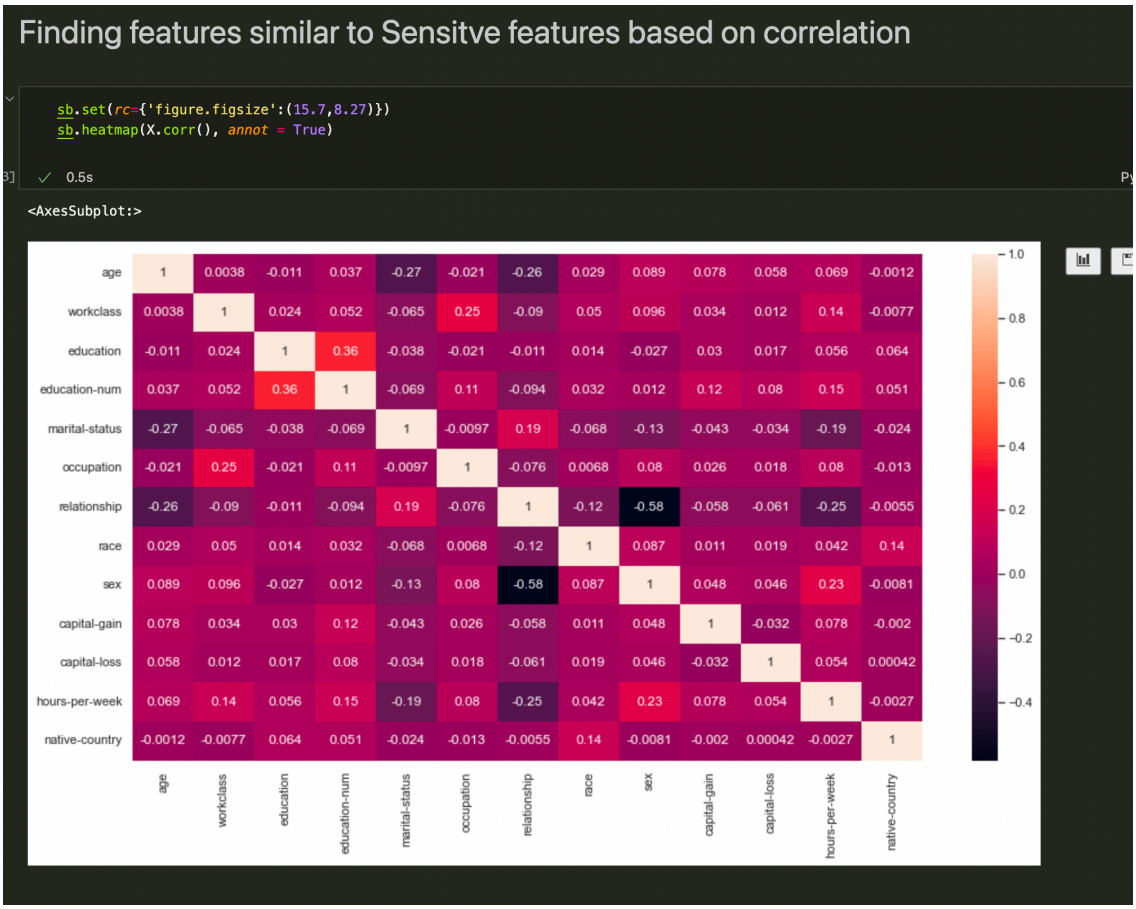
```
  equality of odds (true positive) for sex:  0.10489429628626867
  equality of odds (false positive) for sex:  0.09250219470559012

  equality of odds (true positive) for race:  0.025000000000000022
  equality of odds (false positive) for race:  0.11516563146997931
```

	precision	recall	f1-score	support
0	0.90	0.91	0.91	4918
1	0.72	0.70	0.71	1595
accuracy			0.86	6513
macro avg	0.81	0.81	0.81	6513
weighted avg	0.86	0.86	0.86	6513

- Part 2 :

- I find the correlation heat map for all the features to identify the attributes that correlate the most with the sensitive attribute.



- Based on the heat map we remove **Race, Sex, Marital-Status, Occupation, Relationship** as the correlated sensitive features.

- We find that the Demographic Disparity and Equal Opportunity reduced significantly.

  Demographic parity (Sex) : 0.09
  Demographic parity (Race) :0.07
  Race Average Equality of Opportunity : 0.11
  Race Average Equality of Opportunity **:**

- **The Bias has reduced significantly after removing the correlated features . Accuracy of the classifier reduced to 84% from 86% hence displaying Fairness Accuracy tradeoff.**