Final Present

Adult Income Classification

ADULT INCOME CLASSIFICATION

Dataset

Details of information

ADULT INCOME CLASSIFICATION

Number of datasets

The dataset contains 32,561 cases from the 1994 Census Bureau database.

Feature

This dataset is tabular data with 15 features.

Predict

The prediction task is to determine whether a person makes over $50K a year.

Milestone 1

Task 1: Study the characteristics of the data and identify the quality issues in the selected.

ADULT INCOME CLASSIFICATION

Focus

This dataset focuses on the classification of who can earn money more than 50K per year.

Imbalance

This dataset is an imbalanced class, class 0 contains 24720 values and class 1 contains 7841 values.

Missing value & Type

Some categorical features have missing values and have many types in the categorical column.

Milestone 1

Task 1: Study the characteristics of the data and identify the quality issues in the selected.

ADULT INCOME CLASSIFICATION

Feature

Many features can be used in feature engineering to make a new feature.

Quality issues in the dataset

Bias: Imbalance class effective to the model when training.

Fairness and reliability: Because the model is bias, and unfairness makes the model unconvincing.

Performance: Bad quality of the dataset effect to the model performance.

ADULT INCOME CLASSIFICATION

Milestone 1

Task 2: Define the goals and a suitable measure for the quality issues.

ADULT INCOME CLASSIFICATION

Measures

Improve overall and specify metric performance Measures.

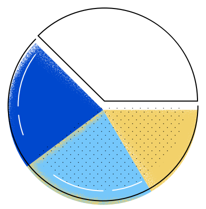
- F1-Score and Precision , because this model needs to classify who can earn more than 50K per year then it will focus on class 1.

Progress

- More EDA in the dataset .
- Apply sample techniques to handle imbalanced class.
- Create models using new preparation data and compare them with baseline models.

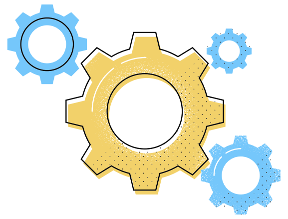
Milestone 1

Task 3: Explore different kinds of machine learning models developed with different modeling techniques. Then, choose the machine learning techniques, implement the models using scikit-learn, and train the models.



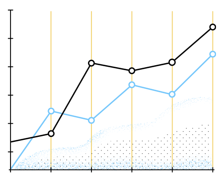
Step 1

EDA THE DATASET



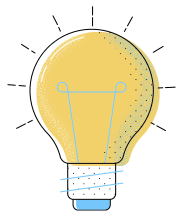
Step 2

PREPARE DATA



Step 3

TRAIN AND TEST DATA



Step 4

COMPUTE MODEL PERFORMANCE

For this part, we have studied different 6 machine learning model s for training and testing the dataset and each model prepares data the same method. The performance of baseline models and confusion matrix . TABLE 1: BASELINE MODEL

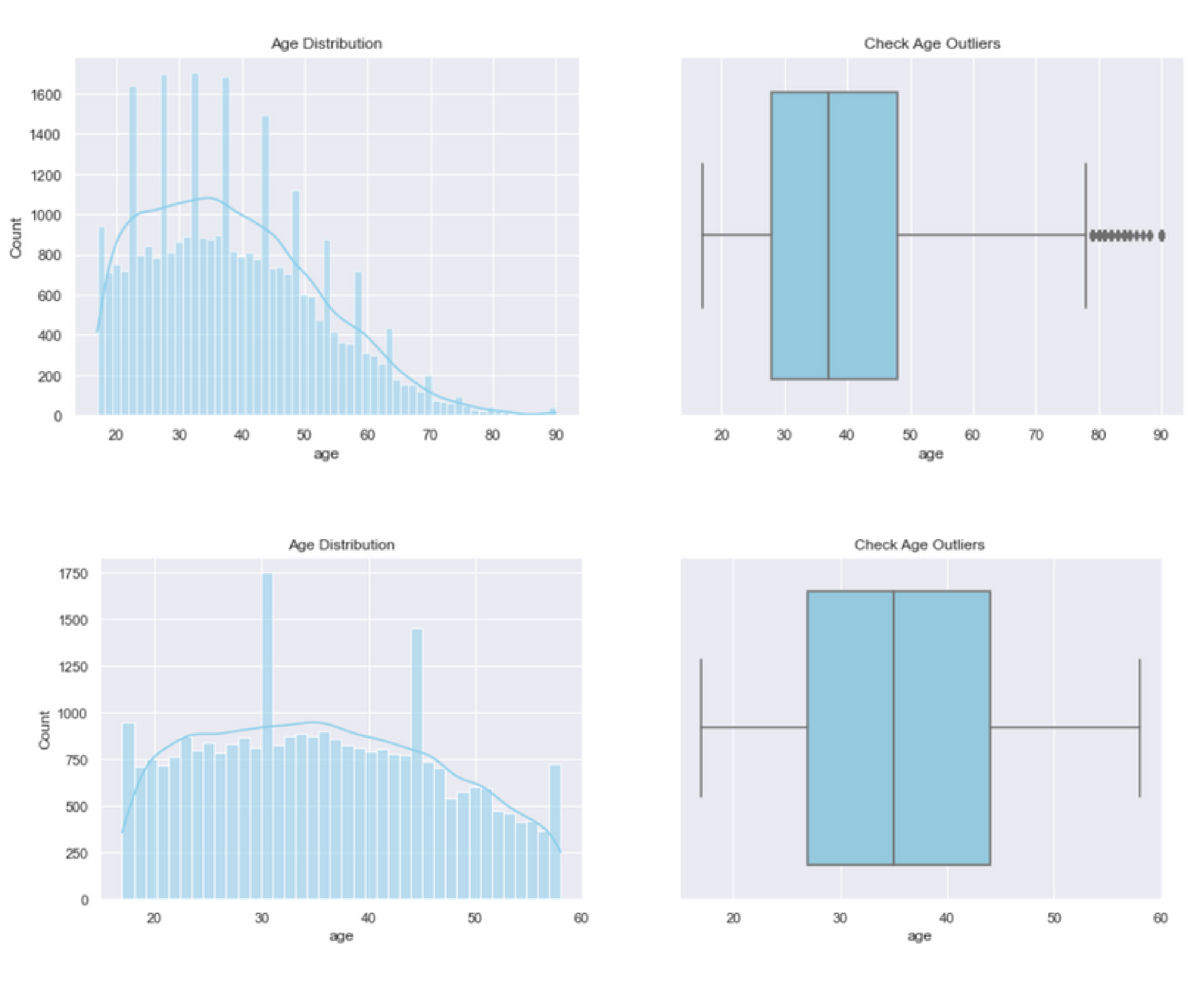
Model	Accuracy (%)	Recall (%)	Precision (%)	F1-Score (%)	Implementer
Random Forest	85.1674	85.1674	84.6325	84.7962	Rew
LightGBM	87.4910	87.4910	87.0745	87.1607	Rew
SVM	84.6760	84.6760	83.8554	83.7191	B
MLP	84.3792	84.3792	84.0164	84.1656	B
Logistic Regression	82.3012	82.3012	81.0242	80.7996	Joey
K-NN	83.1201	83.1201	82.6049	82.8066	Joey

Performance metrics

1 F1-Score: The model can classification between two classes well.

2 Precision: This score for model to focus predict in class 1.

3 Based on the performance of the baseline model, we chose to look over the dataset and discovered that age feature contains a high number of outliers. As a result, we remove it and retrain the model to know the difference between remove and do not remove.



Milestone 2

ADULT INCOME CLASSIFICATION

Milestone 2

Task 4: Model Comparison

TABLE 2: MODEL COMPARISON RESULTS

ADULT INCOME CLASSIFICATION

Model	Accuracy (%)	Precision (%)	Training time (s)	Training Memory used (MB)	Testing time (S)	Testing Memory used (MB)	Model size (MB in. pkl format)	Implementer
Random Forest	85.1674	84.8882	7.50	170.73	0.84	170.73	149.63	Rew
LightGBM	**87.4910**	**87.0745**	0.59	1.98	0.08	0.43	1.006	Rew
SVM	84.6760	83.8554	16.85	15.98	12.21	0.30	0.924	B
MLP	84.3792	83.6049	40.95	**1.34**	0.01	**0.00**	0.049	B

Milestone 2

Task 4: Model Comparison

TABLE 2: MODEL COMPARISON RESULTS

Model	Accuracy (%)	Precision (%)	Training time (s)	Training Memory used (MB)	Testing time (S)	Testing Memory used (MB)	Model size (MB in. pkl format)	Implementer
Logistic Regression	82.3012	81.0242	0.05	2.51	0.00	0.00	0.002	Joey
K-NN	83.1201	82.6049	0.11	2.31	4.20	0.48	4.767	Joey

ADULT INCOME CLASSIFICATION

Summary comparison

Model performance : LightGBM achieved the high accuracy and precision as 87.4910 and 87.0745, followed by Randon forest, MLP, SVM, K-NN and Logistic Regression

Training cost (Training r untime and m emory usage) : Logistic Regression is the best model in this section, with a training time (0.05 s) less than the MLP model (40 . 95 s), despite the MLP requiring less memory (1. 34 MB) than Logistic Regression (2.5 1 MB).

Inference cost (Testing runtime and memory usage) : Logistic Regression is the top model in this section , with testing time (0 s) and memory usage (0 MB) allowing it to outperform other models.

Model size : Logistic Regression is the lightest-weighted model and the worst model in this section is the Random Forest model. which has a size of 149.63 MB while other models have a size of less than 5 MB.

The best model is LightGBM because when determining the best model, we choose to prioritize performance. The LightGBM model outperforms the others, but it comes at a cost in terms of training and inference. Logistic Regression and MLP have potential in this area, however, their performance is roughly 3-5 % lower than LightGBM.

Milestone 2

TABLE 3: PERFORMANCE AFTER RMO

Model		Accuracy (%)	Recall (%)	Precision (%)	F1-core (%)
Random Forest	Baseline	85.2493	85.2493	84.7267	84.8882
	RMO	85.2971	85.2971	84.8507	85.0130
LightGBM	Baseline	87.4910	87.4910	87.0745	87.1607
	RMO	87.0786	87.0789	86.7286	86.8494
SVM	Baseline	84.6760	84.6760	83.8554	83.7191
	RMO	85.4888	85.4888	84.7465	84.6960

Milestone 2

TABLE 3: PERFORMANCE AFTER RMO

Model		Accuracy (%)	Recall (%)	Precision (%)	F1-core (%)
MLP	Baseline	84.4406	84.4406	83.6049	83.6365
	RMO	85.2295	85.2295	84.7802	84.9441
Logistic Regression	Baseline	82.3012	82.3012	81.0242	80.7996
	RMO	83.5607	83.5607	82.4881	82.3643
K-NN	Baseline	83.1201	83.1201	82.6049	82.8066
	RMO	83.7975	83.7975	83.2737	83.4731

Summary

As a result, deleting outliners from age feature helps enhance model performance except LightGBM model is not improved and is worse than baseline. However, when compared to other models, it was a good trade-off. The following stage will be to utilize sampling methods to try to improve the model based on datasets that already eliminate outliners.

Milestone 2

Applied sampling methods to improve models

TABLE 4: OVERSAMPLING RESULTS

Model	Method	Accuracy (%)	Recall (%)	Precision (%)	F1-core (%)
Random Forest	RMO	85.2971	85.2971	84.8507	85.0130
	SMOTE	84.1132	84.1132	84.5416	84.3007
	RandomOver-sampler	84.2823	84.2823	84.8754	84.5298
	ADASYN	83.4592	83.4595	84.3139	83.8005
	BorderlineSMOTE	83.6509	83.6509	84.3363	83.9343

Milestone 2

Applied sampling methods to improve models

TABLE 4: OVERSAMPLING RESULTS

Model	Method	Accuracy (%)	Recall (%)	Precision (%)	F1-core (%)
LightGBM	RMO	87.0786	87.0789	86.7286	86.8494
	SMOTE	85.4437	85.4437	86.3867	85.7913
	RandomOver-sampler	83.7975	83.7975	83.8186	84.5760
	ADASYN	84.6882	84.6882	86.8112	85.1946
	BorderlineSMOTE	84.7221	84.7221	86.2944	85.2390

Milestone 2

Applied sampling methods to improve models

TABLE 4: OVERSAMPLING RESULTS

Model	Method	Accuracy (%)	Recall (%)	Precision (%)	F1-core (%)
SVM	RMO	85.4888	85.4888	84.7465	84.6960
	SMOTE	79.0281	79.0281	84.7309	80.3985
	RandomOver-sampler	79.2536	79.2536	85.2842	80.6417
	ADASYN	76.2431	76.2431	84.4784	77.9881
	BorderlineSMOTE	76.5362	76.5362	84.4996	78.2449

ADULT INCOME CLASSIFICATION

Milestone 2

Applied sampling methods to improve models

TABLE 4: OVERSAMPLING RESULTS

Model	Method	Accuracy (%)	Recall (%)	Precision (%)	F1-core (%)
MLP	RMO	85.2295	85.2295	84.7802	84.9441
	SMOTE	79.8173	79.8173	84.6577	81.0504
	RandomOver-sampler	79.6933	79.6933	84.9099	80.9758
	ADASYN	79.4227	79.4227	84.5013	80..7044
	BorderlineSMOTE	78.2275	78.2275	84.1751	79.6688

ADULT INCOME CLASSIFICATION

Milestone 2

Applied sampling methods to improve models

TABLE 4: OVERSAMPLING RESULTS

Model	Method	Accuracy (%)	Recall (%)	Precision (%)	F1-core (%)
Logistic Regression	RMO	83.5607	83.5607	82.4881	82.3643
	SMOTE	77.1338	77.1138	82.3065	78.5472
	RandomOver-sampler	77.3368	77.3368	82.1936	78.6972
	ADASYN	72.0713	72.0713	81.7758	74..1889
	BorderlineSMOTE	72.1051	72.1051	81.7533	74.2187

ADULT INCOME CLASSIFICATION

Milestone 2

Applied sampling methods to improve models

TABLE 4: OVERSAMPLING RESULTS

Model	Method	Accuracy (%)	Recall (%)	Precision (%)	F1-core (%)
K-NN	RMO	83.7975	83.7975	83.2737	83.4731
	SMOTE	80.5164	80.5164	83.3863	81.4171
	RandomOver-sampler	79.1634	79.1634	83.4449	80.3550
	ADASYN	78.8364	78.8364	83.3960	80.0826
	BorderlineSMOTE	78.8702	78.8702	83.0788	80.0644

ADULT INCOME CLASSIFICATION

Milestone 2

Applied sampling methods to improve models

TABLE 5: UNDERSAMPLING RESULTS

Model	Method	Accuracy (%)	Recall (%)	Precision (%)	F1-core (%)
Random Forest	RMO	85.2971	85.2971	84.8507	85.0130
	RandomUnder-sampler	81.3846	81.3846	85.3104	82.4246
	Tomek Links	84.9588	84.9588	84.8448	84.8991
	Edited Nearest Neighbors	81.8356	81.8356	85.5715	82.8292
	AllKNN	80.9900	80.9900	85.3120	82.1006

Milestone 2

Applied sampling methods to improve models

TABLE 5: UNDERSAMPLING RESULTS

Model	Method	Accuracy (%)	Recall (%)	Precision (%)	F1-core (%)
LightGBM	RMO	87.0786	87.0789	86.7286	86.8494
	RandomUnder-sampler	82.9068	82.9068	86.8252	83.8754
	Tomek Links	87.3154	87.3154	87.1951	87.2504
	Edited Nearest Neighbors	82.9519	82.9519	86.6487	83.8904
	AllKNN	82.2866	82.2866	86.4597	83.3158

Milestone 2

Applied sampling methods to improve models

TABLE 5: UNDERSAMPLING RESULTS

Model	Method	Accuracy (%)	Recall (%)	Precision (%)	F1-core (%)
SVM	RMO	85.4888	85.4888	84.7465	84.6960
	RandomUnder-sampler	78.9830	78.9830	85.2522	80.4084
	Tomek Links	**85.5226**	**85.5226**	**84.8591**	**84.9754**
	Edited Nearest Neighbors	81.6214	81.6214	85.1041	82.5872
	AllKNN	81.1704	81.1704	84.8390	82.1831

ADULT INCOME CLASSIFICATION

Milestone 2

Applied sampling methods to improve models

TABLE 5: UNDERSAMPLING RESULTS

Model	Method	Accuracy (%)	Recall (%)	Precision (%)	F1-core (%)
MLP	RMO	85.2295	85.2295	84.7802	84.9441
	RandomUnder-sampler	78.9604	78.9604	85.0857	80.3746
	Tomek Links	83.6284	83.6284	84.7324	84.0440
	Edited Nearest Neighbors	80.5953	80.5953	84.5504	81.6750
	AllKNN	80.1105	80.1105	84.6308	81.2893

ADULT INCOME CLASSIFICATION

Milestone 2

Applied sampling methods to improve models

TABLE 5: UNDERSAMPLING RESULTS

Model	Method	Accuracy (%)	Recall (%)	Precision (%)	F1-core (%)
Logistic Regression	RMO	83.5607	83.5607	82.4881	82.3643
	RandomUnder-sampler	77.2128	77.2128	82.1841	78.5952
	Tomek Links	83.2788	83.2788	82.2446	82.3821
	Edited Nearest Neighbors	80.1556	80.1556	81.9302	80.8119
	AllKNN	79.7381	79.7381	81.9002	80.5041

Milestone 2

Applied sampling methods to improve models

TABLE 5: UNDERSAMPLING RESULTS

ADULT INCOME CLASSIFICATION

Model	Method	Accuracy (%)	Recall (%)	Precision (%)	F1-core (%)
K-NN	RMO	83.7975	83.7975	83.2737	83.4731
	RandomUnder-sampler	79.0394	79.0394	84.2840	80.3606
	Tomek Links	83.8426	83.8426	83.6403	83.7338
	Edited Nearest Neighbors	81.1929	81.1929	84.3122	82.1162
	AllKNN	81.0689	81.0689	84.1923	81.9974

Summary

According to the table, most sampling approaches are ineffective on models, particularly oversampling strategies, which do not help to improve model performance and make it worse than the original. On the other hand, undersampling helps to improve the model slightly.

Only some models have been upgraded, other models have some parts that are better than the original when using the Tomek Links method mostly.

ADULT INCOME CLASSIFICATION

Milestone 2

Applied sampling methods to improve models TABLE 6: OVERSAMPLING RESULTS TP AND FN BIAS

Model	Method	TP	FN	Precision (%)	F1-core (%)
Random Forest	RMO	1303	753	84.8507	85.0130
	SMOTE	1425	631	84.5416	84.3007
	RandomOver-sampler	1459	597	84.8754	84.5298
	ADASYN	1457	699	84.3139	83.8005
	BorderlineSMOTE	1442	614	84.3363	83.9343

Milestone 2

Applied sampling methods to improve models TABLE 6: OVERSAMPLING RESULTS TP AND FN BIAS

Model	Method	TP	FN	Precision (%)	F1-core (%)
LightGBM	RMO	1386	670	86.7286	86.8494
	SMOTE	1569	487	86.3867	85.7913
	RandomOver-sampler	1713	343	83.8186	84.5760
	ADASYN	1609	447	86.2112	85.1946
	BorderlineSMOTE	1617	439	86.2944	85.2390

Milestone 2

Applied sampling methods to improve models

TABLE 6: OVERSAMPLING RESULTS TP AND FN BIAS

Model	Method	TP	FN	Precision (%)	F1-core (%)
SVM	RMO	1150	906	84.7465	84.6960
	SMOTE	1738	318	84.7309	80.3985
	RandomOver-sampler	1777	279	85.2842	80.6417
	ADASYN	1800	256	84.4784	77.9881
	BorderlineSMOTE	1794	262	84.4996	78.2449

Milestone 2

Applied sampling methods to improve models TABLE 6: OVERSAMPLING RESULTS TP AND FN BIAS

Model	Method	TP	FN	Precision (%)	F1-core (%)
MLP	RMO	1274	782	84.7802	84.9441
	SMOTE	1703	353	84.6577	81.0504
	RandomOver-sampler	1730	326	84.9099	80.9758
	ADASYN	1704	352	84.5013	80.7044
	BorderlineSMOTE	1718	338	84.1751	79.6688

Milestone 2

Applied sampling methods to improve models

TABLE 6: OVERSAMPLING RESULTS TP AND FN BIAS

Model	Method	TP	FN	Precision (%)	F1-core (%)
Logistic Regression	RMO	996	1060	82.4881	82.3643
	SMOTE	1595	461	82.3065	78.5472
	RandomOver-sampler	1577	479	82.1936	78.6972
	ADASYN	1703	353	81.7758	74.1889
	BorderlineSMOTE	1702	354	81.7533	74.2187

Milestone 2

Applied sampling methods to improve models

TABLE 6: OVERSAMPLING RESULTS TP AND FN BIAS

Model	Method	TP	FN	Precision (%)	F1-core (%)
K-NN	RMO	1229	827	83.2737	83.4731
	SMOTE	1548	508	83.3863	81.4171
	RandomOver-sampler	1618	438	83.4449	80.3550
	ADASYN	1627	429	83.3960	80.0826
	BorderlineSMOTE	1596	460	83.0788	80.0644

Milestone 2

Applied sampling methods to improve models

TABLE 7: UNDERSAMPLING RESULTS TP AND FN BIAS

Model	Method	TP	FN	Precision (%)	F1-core (%)
Random Forest	RMO	1303	753	84.8507	85.0130
	RandomUnder-sampler	1699	357	85.3104	82.4246
	Tomek Links	1366	690	84.8448	84.8991
	Edited Nearest Neighbors	1704	352	85.5715	82.8292
	AllKNN	1716	340	85.3120	82.1006

Milestone 2

Applied sampling methods to improve models

TABLE 7: UNDERSAMPLING RESULTS TP AND FN BIAS

Model	Method	TP	FN	Precision (%)	F1-core (%)
LightGBM	RMO	1386	670	86.7286	86.8494
	RandomUnder-sampler	1777	279	86.8252	83.8754
	Tomek Links	1464	592	87.1951	87.2504
	Edited Nearest Neighbors	1758	298	86.6487	83.8904
	AllKNN	1769	287	86.4597	83.3158

ADULT INCOME CLASSIFICATION

Milestone 2

Applied sampling methods to improve models

TABLE 7: UNDERSAMPLING RESULTS TP AND FN BIAS

Model	Method	TP	FN	Precision (%)	F1-core (%)
SVM	RMO	1150	906	84.7465	84.6960
	RandomUnder-sampler	1783	273	85.2522	80.4084
	Tomek Links	1222	834	84.8591	84.9754
	Edited Nearest Neighbors	1668	388	85.1041	82.5872
	AllKNN	1663	393	84.8393	82.1831

ADULT INCOME CLASSIFICATION

Milestone 2

Applied sampling methods to improve models

TABLE 7: UNDERSAMPLING RESULTS TP AND FN BIAS

Model	Method	TP	FN	Precision (%)	F1-core (%)
MLP	RMO	1274	782	84.7802	84.9441
	RandomUnder-sampler	1770	286	85.0857	80.3746
	Tomek Links	1500	566	84.7324	84.0440
	Edited Nearest Neighbors	1661	395	84.5504	81.6750
	AllKNN	1689	367	84.6308	81.2893

ADULT INCOME CLASSIFICATION

Milestone 2

Applied sampling methods to improve models

TABLE 7: UNDERSAMPLING RESULTS TP AND FN BIAS

Model	Method	TP	FN	Precision (%)	F1-core (%)
Logistic Regression	RMO	996	1060	82.4881	82.3643
	RandomUnder-sampler	1581	475	82.1841	78.5952
	Tomek Links	1056	1000	82.2446	82.3821
	Edited Nearest Neighbors	1408	648	81.9302	80.8119
	AllKNN	1431	625	81.9002	80.5041

Milestone 2

Applied sampling methods to improve models

TABLE 7: UNDERSAMPLING RESULTS TP AND FN BIAS

Model	Method	TP	FN	Precision (%)	F1-core (%)
K-NN	RMO	1229	827	83.2737	83.4731
	RandomUnder-sampler	1699	357	84.2840	80.3606
	Tomek Links	1301	755	83.6403	83.7338
	Edited Nearest Neighbors	1609	447	84.3122	82.1162
	AllKNN	1603	453	84.1923	81.9974

ADULT INCOME CLASSIFICATION

Summary

The technique of sampling doesn't help much with model prediction bias. The oversampling method improves the model's ability to predict TP more than the baseline method, but it increases the FP of the model. Undersampling improves TP more than oversampling, but only for some models. Only some models with Tomek Links improve model prediction bias. Additionally, other methods of undersampling cause the model to predict TP greater than the baseline. It’s like oversampling methods.

ADULT INCOME CLASSIFICATION

Fairness metrics Milestone 3

We decided to use matrices such as Mean Difference, Disparate Impact , Average Odds Difference and Equal Opportunity Difference to measure model fairness. We used the Reweighing method the make model fairness, this method will be in the part of pre- processing .

ADULT INCOME CLASSIFICATION

Mean Difference : Examines if the model's predictions differ significantly between groups. Lower values indicate that predictions are more similar between groups, which is a sign of fairness.

Disparate Impact **:** Checks if a model treats different groups fairly. It looks at whether everyone has an equal chance of a favorable outcome. A value of 1 means equal treatment, while values far from 1 suggest potential unfairness.

Average Odds Difference : *Measures if a model makes errors unfairly. It looks at the difference in error rates between groups. Smaller values mean the model is more fair because it makes similar errors for everyone.*

Equal Opportunity Difference **:** Checks if the model gives everyone an equal chance to be correctly identified as positive. Smaller values indicate that both groups have similar opportunities to be correctly recognized.

Milestone 3

TABLE 8:COMPARISON FAIRNESS OF THE DATASET

Dataset	Reweight method	
	Mean Difference	Disparate Impact
Original	-0.2005	0.3531
Transformed	0.0000	1.000

Milestone 3

ADULT INCOME CLASSIFICATION

TABLE 9: COMPARISON MODEL PERFORMANCE AFTER APPLIED REWEIGHT

Model	Dataset	Performance			
		Accuracy (%)	Recall(%)	Precision (%)	F1-core (%)
Random Forest	Orginal	85.3372	85.3372	84.7942	84.9421
	Transformed	85.3687	85.3687	84.833	84.9813
LightGBM	Orginal	87.0339	87.0339	86.5608	86.6201
	Transformed	87.1806	87.1806	86.7286	86.7962

Milestone 3

TABLE 9: COMPARISON MODEL PERFORMANCE AFTER APPLIED REWEIGHT

Model	Dataset	Performance			
		Accuracy (%)	Recall(%)	Precision (%)	F1-core (%)
SVM	Orginal	85.0545	85.0545	84.3385	84.0741
	Transformed	80.1948	80.1948	79.0319	76.5241
MLP	Orginal	84.8240	84.8240	84.3486	84.5138
	Transformed	83.0017	83.0017	82.2199	81.0201

Milestone 3

ADULT INCOME CLASSIFICATION

TABLE 9: COMPARISON MODEL PERFORMANCE AFTER APPLIED REWEIGHT

Model	Dataset	Performance			
		Accuracy (%)	Recall(%)	Precision (%)	F1-core (%)
Logistic Regression	Orginal	82.7189	82.7189	81.6092	81.1802
	Transformed	80.9384	80.9384	79.3777	78.8361
K-NN	Orginal	83.7767	83.7767	83.2237	83.4188
	Transformed	84.6669	84.6669	84.1156	84.2880

Summary

According to the table above, the LightGBM Model has the best performance, and only a few models have improved their performance, but not significantly in term of fairness if we comparison only these metrics

Milestone 3

TABLE 10: COMPARISON MODEL PERFORMANCE ON FAIRNESS MATRIX

Model	Dataset	Fairness	
		Average Odds Difference	Equal Opportunity Difference
Random Forest	Orginal	-0.3042	0.3157
	Transformed	-0.1504	0.6247
LightGBM	Orginal	-0.3750	0.2082
	Transformed	-0.1427	0.6545

ADULT INCOME CLASSIFICATION

Milestone 3

ADULT INCOME CLASSIFICATION

TABLE 10: COMPARISON MODEL PERFORMANCE ON FAIRNESS MATRIX

Model	Dataset	Fairness	
		Average Odds Difference	Equal Opportunity Difference
SVM	Orginal	-0.5000	0.0000
	Transformed	-0.3494	0.2701
MLP	Orginal	0.4645	0.9706
	Transformed	-0.2740	0.4130

Milestone 3

TABLE 10: COMPARISON MODEL PERFORMANCE ON FAIRNESS MATRIX

Model	Dataset	Fairness	
		Average Odds Difference	Equal Opportunity Difference
Logistic Regression	Orginal	0.5000	1.0000
	Transformed	-0.2829	0.3788
K-NN	Orginal	-0.3733	0.2096
	Transformed	-0.1527	0.6142

ADULT INCOME CLASSIFICATION

Summary

Following the comparison of model performance, the fairness score is shown in this table, which compares the fairness score in original and transformed data using the reweight method.

In terms of AOD and EOD values, LightGBM in the Transformed dataset has the best fairness scores with an AOD of -0.1427 and an EOD of 0.6545. It exhibits the least disparity in error rates and provides more equitable opportunities for the unprivileged group to be correctly identified .

ADULT INCOME CLASSIFICATION

Result analysis

In terms of the performance that we evaluated on various models, if the model is based on bagging or boosting, it will outperform other models. The models' performance averages about 80%, but there is still a risk of false positives because most models fail in this area. Sampling methods are less important in improving model performance. Removing outliers in the age feature helps to enhance all models except LightGBM.

For fairness, we apply the reweight method, which requires bias reduction in the pre-processing section. The model's fairness was then determined using the Average Odds Difference and Equal Opportunity Difference. LightGBM has the highest fairness ratings in the Transformed dataset, with an AOD of -0.1427 and an EOD of 0.6545. It has the lowest error rate discrepancy and gives a more equitable opportunity for the underprivileged group to be appropriately identified.

ADULT INCOME CLASSIFICATION

Probability improvement :

- Reduce the bias from the data collection, find more useful information
- Features engineering and selection
- Try other bias mitigation methods such as in and post processing
- Hypermeter tuning based on bagging or boosting model

Solution or best practices to mitigate the quality issues throughout the system's life cycle
Experienced perspectives :

- Careful to check bias and outlier of the dataset.
- Pay attention to data privacy and protected attributes
- Using explainable model that helps to follow an important feature and compare in real situation.
- A clear prediction aim assists in the development of the model and the achievement of the goal.
-

Theory perspectives :

- Use the right performance measurement for the model.
- Make the AI enabled systems to be generalized for the task.
- Collecting and utilizing the appropriate quantity of data