Evaluating the fairness in the performance of machine learning methods

Ming Yuan

A thesis
submitted in partial fulfillment of the
requirements for the degree of

Master of Science

University of Washington

2018

Committee:

Ankur Teredesai, Chair

Vikas Kumar

Program Authorized to Offer Degree:
Computer Sciences and Systems

University of Washington

Abstract

Evaluating the fairness in the performance
of machine learning methods

Ming Yuan

Chair of the Supervisory Committee:
Professor Ankur Teredesai
Department of Computer Science and Systems

Machine learning plays an increasingly important role in our lives, tackling both prevalent and specialized but high-risk problems. Motivated by legislation, responsibility to ensure transparency and accountability of machine learning methods and needs to maintain public's trust on the algorithms used in our lives, researchers have paid much attention to the fairness issue in machine learning. There are many methods developed to measure, reduce and even eliminate the fairness issue for both general and specific settings or algorithms. In this project, we focus on fairness in classification machine learning problems in healthcare which is one critical field of the application of machine learning. We found a general way to detect the fairness issue in the performance of machine learning methods and found the general solutions to address the issue in all the dimensions of data, method and metrics. We also introduced fairness threshold to help reduce the fairness issue without retraining the model and performance boundary to help analyze the effect of the methods we tried.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I am grateful for all the help UWT and KenSci has provided.

First I would like to express my deepest appreciation to Prof. Ankur Teredasai, who is my committee chair and advisor. He has helped and guided me a lot to find the direction through the journey, and always been encouraging and supporting me. Furthermore, I would also like to say thank you to my advisor, Vikas Kumar, without whose help and contributions this work would not have been completed.

I would also express my gratitude to my parents and friends, who have continuously supported me with their love and helped me by their best. They are the reason why I become the person who I am today.

Chapter 1
INTRODUCTION

Machine Learning has been deeply involved in our life in many ways in the last ten years. There are widespread algorithms [1] which are parts of everyone's everyday life. For instance, recommendation systems, which could suggest items to people by predicting their preference for that item, play an important role in many e-commerce sites such as Amazon, Netflix, or the like, and social media platforms [2]. There are also higher-stakes specialized algorithms [1] which are less prevalent but demand higher accuracy, like the ones used for criminal justice[3] and financial decision [4].

1.1 Motivations

Considering the high level of impact, fairness in machine learning becomes a critical issue. Fairness is motivated in many fields by national and international legislation. The Universal Declaration of Human Rights [5], which is a milestone document in the history of human rights, outlined equality and freedom from discrimination as basic human rights [6]. The Declaration, although is not binding, has been elaborated and ratified in subsequent international treaties, national constitutions, and other laws [7, 8, 9].

Furthermore, the concern about the fairness issue is also about ensuring transparency and accountability of the machine learning methods being used [10]. We should be able to explain the performance of the algorithm and guarantee what would happen when the algorithm are involved in making decisions in our real lives. To find the ways to detect and address the fairness issue could also help maintain public trust and protect the social contract [11] because the technology plays such an important role in everyone's life that it will influence their interests a lot.

1.2 Objectives

Fairness is critical in many places, but healthcare is important. This is because the algorithms used in the field of health care can be both widespread and specialized. Health issue is one of the most important issues in everyone's daily life, and it could be involved in decisions with life-or-death consequences which makes the accuracy of the methods being used critical. For instance, the FutureMatch, a framework proposed by Dickerson and Sandholm[12], is used to help make decision in kidney exchanges, which is to determine which patients receive which kidneys.

In this thesis, our focus is to identity in healthcare the protected features such as age, gender, race and measure and address the fairness issue in the performance of Machine Learning algorithms over the healthcare datasets which is used for classification problem. To be more specific, we have studied the following problems:

- Identify fairness in performance of machine learning methods;
- Determine if protected features and sampling affect the performance and fairness of machine learning methods;
- Find the fairness threshold, the performance corresponding to which is comparatively fair and also comparable.

Chapter 2

RELATED WORK

There has been plenty of evidence showing that when machine learning algorithms learn from data, intentional or unintentional discrimination can happen [13]. In particular, the researchers showed concerns for and discussed about the fairness issue in machine learning systems, including the fields of natural language processing [14], image classification [15], etc, and the problems of target advertising [16], judicial sentencing [17], etc. And all these discussion have proven that the machine learning methods could inherit bias from data and create disparate treatment across categories [18]. Regarding to the fairness issue showed in these findings, researchers are paying more and more attention to fairness in machine learning. For instance, many companies, like Google [19] and Amazon [20], have announced the AI ethics board emphasizing the importance of fairness in AI.

2.1 Fairness in Healthcare

The fairness issue in healthcare has began to draw the researchers' and public's attention [21] as well. The awareness and study of unfairness in this field depend not only on the development of detecting and addressing methods, but also on discovery in medicine. Herman et al. [22] have found the difference in the level of HbA1c, which is a widely used index to predict onset of diabetes, across different race, which means the performances of the machine learning model for predicting the diabetes using HbA1c as an index [23, 24] may also be different across race categories.

2.2 Detect and Measure the fairness issue in Machine Learning

Researchers have developed many methods to detect and measure the fairness issue based on different definition of fairness in ML. Zliobaite has listed [25] several statistical methods and comparison functions to detect or measure the fairness issue in ML based on the prediction or classification results, like Normalized Difference [26] which is normalized mean difference for binary classification used to quantify the difference between groups of people. Tramer et al. [27] has introduced unwarranted associations (UA) framework to detect the fairness issue in data-driven applications by investigating associations between application outcomes and sensitive user attributes. There are also some detection methods developed for specific problems or algorithms, like the detection methods for ranking algorithm [28, 29].

2.3 Address the Fairness issue in Machine Learning

The solutions to address the fairness issue are mainly developed for specific algorithms or specific settings, and they have been designed under the following three strategies [30]:

- Pre-processing: preprocessing the datasets to eliminate sources of the fairness issue in data. Feldman et al. [31] proposed a method to remove unfairness in dataset by changing the attributes to repaired version.

- In-processing: sanitizing algorithms or adjusting machine learning process, such as adding the regularizer to the model. Calders et al. [32] developed constrained linear regression models imposing constraints into the process of minimizing squared error to control the fairness issue. Karmiran et al. [33] designed a regularization approach for prediction algorithm with probabilistic discriminative models to remove prejudice which is one source of the fairness issue.

- Post-processing: adjusting trained model or performance to make it more fair. Karmiran et al. [34] proposed the construction of a decision tree classifier without the fairness issue, which made adjustment on both training process and trained model.

Chapter 3
DESIGN OF THE EXPERIMENT

3.1 Definition

3.1.1 Fairness in Machine Learning

In the context of machine learning fairness can be defined as: Each category of the protected features should have similar performance.

For instance, the algorithms used to assess a criminal defendant?s probability of becoming a recidivist who refer to the re-offending criminal have been increasingly used across the nation, probation, judges and parole officers [17]. However, the study [17] has shown that these algorithms may perform differently across different race. Far more black defendants were incorrectly predicted to be at a higher risk of recidivism than white defendants by Northpointe's tool, called COMPAS (which is short for Correctional Offender Management Profiling for Alternative Sanctions [35]).

Table 3.1: Example: Performance of Northpointe's assessment tool across different races

	White	African American
Labeled Higher Risk, But Didn't Re-Offend	23.5%	44.9%
Labeled Labeled Lower Risk, Yet Did Re-Offend	47.7%	28.0%

We could see the performance of 'White' category and 'African American' category are significantly different, and this is what we called "unfairness".

Fairness in machine learning could be studied in three different dimensions corresponding to the three strategies to address the fairness issue:

- Fairness in dataset: Studies in dataset are mainly about the attributes of dataset, which

could be the source of the fairness issue. For instance, when one or several categories are under-represented, when the dataset is outdated or incomplete [36], or when the dataset inherits unintentional perpetuation and promotion of historical biases [37], the machine learning methods may learn these disparateness and unfairness and give out unfair outcomes. To detect this kind of attributes and to make adjustment to reduce or eliminate the fairness issue from the source are critical.

- Fairness in model/algorithm: Studies in model/algorithm are about the design of model/algorithm. A poorly designed matching systems [38], the inappropriately chosen features and assumptions like correlation necessarily implies causation could all give rise to the fairness issue [37]. Thus, we need to examine the design of the model/algorithm and could improve the fairness in outcomes by regulating the model/algorithm.

- Fairness in metrics/results: Metrics/Results dimension focus on the fairness issue in performance of the model/algorithm. We could develop the detection methods like statistical test of the performance and post-processing solutions to coordinate the performance.

3.1.2 Protected Features

Protected Features are the features whose categories have different performance. We have two types of protected features:

- Known protected features: The features protected by the law, such as age, gender, disability, race, etc [Equality Act 2010].

- Unknown protected features: The potential protected features need to be examined. They could be found based on experience and related knowledge, like race in the example of predicting onset of diabetes based on HbA1c. They could also be found by experiments, which means to check all the features one by one to see the performance of whom is different across categories.

3.2 Dataset

We have chosen three public dataset, which are dealing with the common and prevalent problem in healthcare and whose target variables are nominal.

Table 3.2: Information of the Datasets

Name	Description	Size (nRows x nFeatures)	Target Variable	Protected Feature(s)
Length of Stay (LOS) [39]	Predict hospital length of stay	46630 x 212	Binary (shorter stay or longer stay)	age, gender, race
Thyroid Disease Dataset [40]	Predict the state of thyroxine-binding proteins, which is relative to thyroid disease	3772 x 29	Multi-class (negative, increased binding protein, decreased binding protein)	age, gender
Pima Indians Diabetes Dataset (PIMA) [41]	Predict the onset of diabetes based on diagnostic measures (medical predictor variables)	768 x 9	Binary (onset or not)	age

3.3 The Methodology

A standard machine learning process should include the steps of collecting and preprocessing data, selecting features, training the models and getting the metrics. This project has focused on a general way to detect and address the fairness issue in metrics step and introduced optimal threshold to help analyze the effect of proposed solutions.

To be more specific, we measured the fairness by comparing the performance of each category and comparing the optimal threshold chosen based on the performance of each category, and found the fairness threshold, which allows us to have more fair outcomes without much drop in performance. And performance in this project includes AUC score, precision, recall and f1 score.

The optimal threshold was found based on Youden's index [42], which is

$$J = sensitivity + specificity - 1.$$

Youden's index has been used as the measure of diagnostic effectiveness. It could also be used for selecting the optimal cut-point on ROC-curve [43], which is the optimal threshold we want.

Besides, we also considered addressing methods implemented on data and features, including training without the protected features, training without the important features related to the protected feature to help reduce the fairness issue and sampling to balance the size of each category on training dataset.

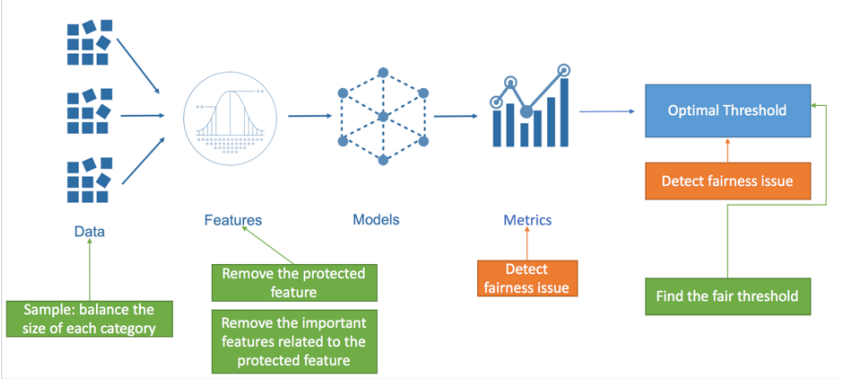


Figure 3.1: An overview of the contributions

3.3.1 Machine Learning Models

The following three popular and well-studied and well-applied in healthcare domain machine learning models are the ones involved in this project:

- Logistic Regression [44]:

  A binomial classification method used when the outcome variable is binary variable;

  A method to predict the probability that the outcome is '1', which means the item is classified to the group represented by 1, with given input data.

- Random Forest [45]:

  An ensemble learning method for classification and regression problems;

  A combination of decision tree and bagging.

- XGBoost (eXtreme Gradient Boosting) [46]:

  An implementation of gradient boosted decision trees with high efficient and good performance for classification and regression problems.

Chapter 4

EXPERIMENTS AND RESULTS

4.1 Detect the Fairness Issue in Performance

To measure the fairness in performance, we first compared the performance of each category. As the Figure 4.1 shows, we also computed the variance of the performance of all the categories, which are 'male' and 'female' in this example, to quantify the difference. The less the variance is, the less the difference is, which means the performance is more fair. In this example, the model trained without the protected feature, gender, has more fair performance than the one trained with all the features.

	Features	entire dataset	Male	Female	var
AUC score	with 'gender'	0.6831	0.6908	0.6724	0.0001693
	without 'gender'	0.6837	0.6887	0.6769	0.0000696
Precision	with 'gender'	0.7156	0.7225	0.7046	0.0001602
	without 'gender'	0.7274	0.7306	0.723	0.0000289
Recall	with 'gender'	0.755	0.7582	0.7506	0.0000289
	without 'gender'	0.7563	0.7590	0.7526	0.0000205
F1 score	with 'gender'	0.6625	0.6682	0.6549	0.0000884
	without 'gender'	0.6641	0.6683	0.6585	0.0000480

Figure 4.1: Performance of random forest models trained with and without the protected feature (gender) on LOS dataset

What's more, we also found the optimal threshold for entire dataset and the ones for all the categories. And used the variance of the optimal thresholds as another criteria to measure the fairness.

Features	entire dataset	Male	Female	var
with 'gender'	0.25	0.244	0.239	3.03333E-05
without 'gender'	0.279	0.279	0.279	0

Figure 4.2: Optimal thresholds of random forest models trained with and without the protected feature (gender) on LOS dataset

4.2 Threshold for Fairness and Performance Boundary

To find the threshold for fairness, we set the optimal threshold for entire dataset, optimal thresholds for each category, and the average, median, maximum and minimum value of the thresholds of all the categories as the candidates. And then, we computed the difference of performance corresponding to each threshold, including precision, recall and f1 score, across each category, and chose the one with minimum difference as the fairness threshold.

entire dataset		Male		Female		Average		Median		Maximum		Minimum		var
0.2208	0.003844	**0.2343**	0.004178	**0.2228**	0.003667	**0.2286**	0.003244	**0.2286**	0.003244	**0.2343**	0.004178	**0.2228**	0.003667	5.31E-05

Figure 4.3: Thresholds from the experiment choosing gender as the protected feature and trained with all the features on LOS dataset before sampling, where the fairness thresholds are marked as blue

To help analyze effect on performance of choosing different threshold, we plotted performance boundary, including precision boundary, recall boundary and f1 score boundary, for each fairness threshold candidate.

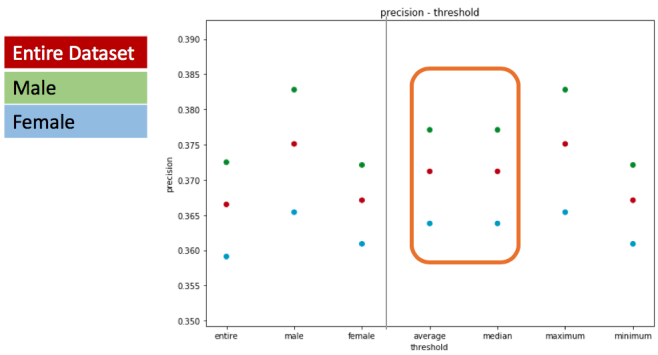


Figure 4.4: Precision boundary of the experiment in Figure 4.3, where the boundaries of fairness thresholds are marked by orange rectangular

The performance boundary shows that best performance does not mean fair. For instance, the threshold based on male which is also the maximum threshold in Figure 4.4 has the highest precision, but the it also has maximum difference, which means the performance corresponding to this threshold is the least fair one. Furthermore, model with fair outcomes can have comparable performance. For instance, the fairness threshold in Figure 4.4 performs even a bit better than the performance corresponding to optimal threshold for entire dataset, which we normally care about.

4.3 Effect of Removal of Protected Features

To make the performance of the protected feature's each category similar to each other, removing the protected features so that they could not influence the performance directly is one apparent choice. Figure 4.5 is the performance of methods training with and without protected features on LOS dataset, where the increase in performance are marked as red and the decrease in variance of the performance of all the categories are marked as blue.

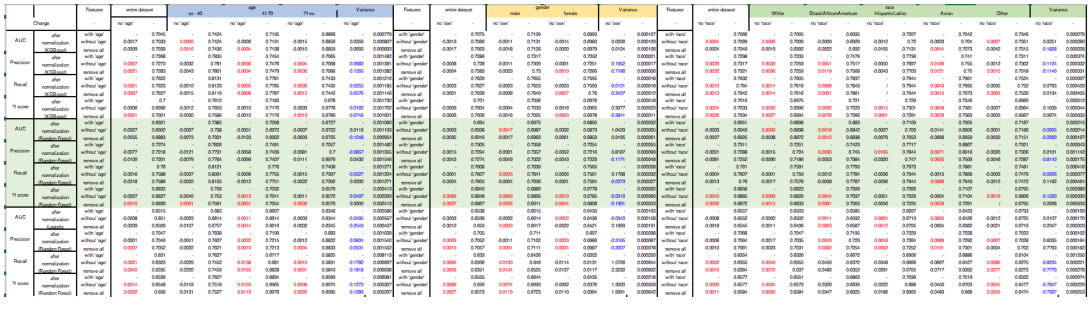


Figure 4.5: Performance of methods trained with and without protected features on LOS dataset

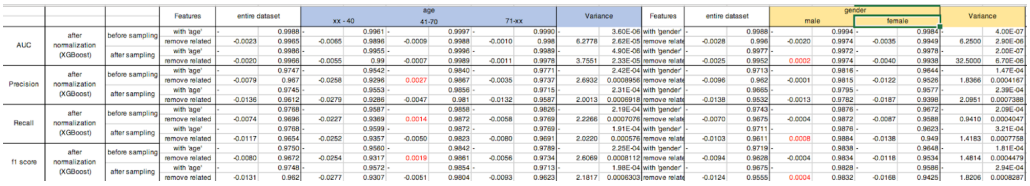
The results show that:

- The variance of the performance is more likely to decrease 10% - 30%, sometimes over 60%, which means removal of protected features could help the performance become more fair;

- Although it seems that the performance is more likely to increase, it mostly increases under 5%, which could be interpreted as the performance does not show significant change.

4.3.1 Effect of Removal of Important Features Related to Protected Features

A further idea is whether we need to remove the features, especially the ones with high importance scores, related to protected features to eliminate the indirect influence of the protected features on the outcomes.



			Features	entire dataset	age					Variance	Features	entire dataset	gender			Variance
					xx - 40		41-70		71-xx				male		female	
AUC	after normalization (XGBoost)	before sampling	with 'age'	-	0.9988	-	0.9961	-	0.9997	-	0.9990	-	3.60E-06	with 'gender'	-	0.9988	-	0.9994	-	0.9984	-	4.00E-07
			remove related	-0.0023	0.9965	-0.0065	0.9896	-0.0009	0.9988	-0.0010	0.998	6.277E	2.62E-05	remove relate	-0.0028	0.996	-0.0020	0.9974	-0.0035	0.9949	6.2500	2.90E-06
		after sampling	with 'age'	-	0.9986	-	0.9955	-	0.9996	-	0.9989	-	4.90E-06	with 'gender'	-	0.9977	-	0.9972	-	0.9978	-	2.00E-07
			remove related	-0.0020	0.9966	-0.0055	0.99	-0.0007	0.9989	-0.0011	0.9978	3.755E	2.33E-05	remove relate	-0.0025	0.9952	0.0002	0.9974	-0.0040	0.9938	32.5000	6.70E-06
Precision	after normalization (XGBoost)	before sampling	with 'age'	-	0.9747	-	0.9542	-	0.9840	-	0.9771	-	2.42E-04	with 'gender'	-	0.9713	-	0.9816	-	0.9644	-	1.47E-04
			remove related	-0.0079	0.967	-0.0258	0.9296	0.0027	0.9867	-0.0035	0.9737	2.693E	0.0008956	remove relate	-0.0096	0.962	-0.0001	0.9815	-0.0122	0.9526	1.8366	0.0004167
		after sampling	with 'age'	-	0.9745	-	0.9553	-	0.9856	-	0.9715	-	2.31E-04	with 'gender'	-	0.9665	-	0.9795	-	0.9577	-	2.39E-04
			remove related	-0.0136	0.9612	-0.0279	0.9286	-0.0047	0.981	-0.0132	0.9587	2.001E	0.0006918	remove relate	-0.0138	0.9532	-0.0013	0.9782	-0.0187	0.9398	2.0951	0.0007388
Recall	after normalization (XGBoost)	before sampling	with 'age'	-	0.9768	-	0.9587	-	0.9856	-	0.9826	-	2.19E-04	with 'gender'	-	0.9743	-	0.9876	-	0.9672	-	2.09E-04
			remove related	-0.0074	0.9698	-0.0227	0.9369	0.0014	0.9872	-0.0058	0.9769	2.226E	0.0007076	remove relate	-0.0070	0.9675	-0.0004	0.9872	-0.0087	0.9588	0.9410	0.0004047
		after sampling	with 'age'	-	0.9768	-	0.9599	-	0.9872	-	0.9769	-	1.91E-04	with 'gender'	-	0.9711	-	0.9876	-	0.9623	-	3.21E-04
			remove related	-0.0117	0.9654	-0.0252	0.9357	-0.0050	0.9823	-0.0080	0.9691	2.022E	0.000576	remove relate	-0.0103	0.9611	0.0008	0.9884	-0.0138	0.949	1.4183	0.0007758
f1 score	after normalization (XGBoost)	before sampling	with 'age'	-	0.9750	-	0.9560	-	0.9842	-	0.9789	-	2.25E-04	with 'gender'	-	0.9719	-	0.9838	-	0.9648	-	1.81E-04
			remove related	-0.0080	0.9672	-0.0254	0.9317	0.0019	0.9861	-0.0056	0.9734	2.606E	0.0008112	remove relate	-0.0094	0.9626	-0.0004	0.9834	-0.0118	0.9534	1.4814	0.0004479
		after sampling	with 'age'	-	0.9748	-	0.9572	-	0.9854	-	0.9713	-	1.98E-04	with 'gender'	-	0.9675	-	0.9828	-	0.9586	-	2.94E-04
			remove related	-0.0131	0.962	-0.0277	0.9307	-0.0051	0.9804	-0.0093	0.9623	2.181E	0.0006303	remove relate	-0.0124	0.9555	0.0004	0.9832	-0.0168	0.9425	1.8206	0.0008287

Figure 4.6: Performance of methods trained with and without the important features related to the protected features on Thyroid dataset, where the increase in performance are marked as red

Figure 4.6 is the performance of methods trained with and without the important features related to the protected features on Thyroid dataset. The important features refer to the top 10 features sorted by their important scores, which is calculated by how much the performance measure would improve on each attribute split and weighted by the number of observations the node is responsible for while training with Gradient Boosting method [47]. And it shows that:

- The variance of the performance does not decrease;
- The performance is more likely to decrease.

, which means removal of related important features is not helpful for both performance and the fairness in performance. It is not necessary to remove these related features.

4.4 Effect of Sampling

Taking the distribution of the protected feature 'age' and the one of the protected feature 'race' in LOS dataset as examples, there could be several categories under-represented, which may introduce the fairness issue to the performance after training on them.

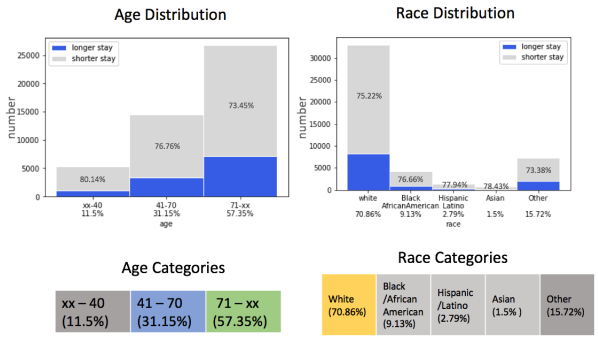


Figure 4.7: Distribution of 'age' and 'race' in LOS dataset

To reduce the influence of underrepresentation of the several categories, we tried to sample on training dataset to balance the size of each category. And as Figure 4.8 and Figure 4.9 shows, the results turn out to be more fair:

Table 4.1: Effect of Sampling

Size of Dataset	small	big
Performance: AUC score, precision, recall, f1 score	The performance of logistic regression and random forest may improve slightly (under 10%, mostly around 1%) after sampling , which may also happen when the size of the category is small	Although the performance is more likely to decrease a bit, it mostly decreases under 5%, which means the performance does not show significant change
Variance: variance of the performance of all the categories	the variance of the performance is more likely to decrease10% - 30%, sometimes over 50%, especially the performance of logistic regression	
Accuracy Boundary	the differences in performance across categories reduce	

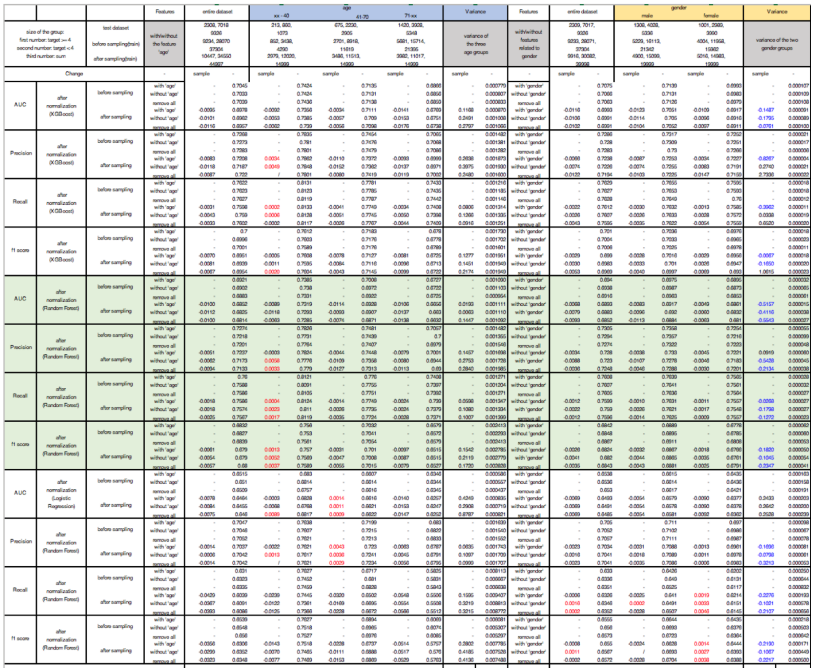


Figure 4.8: Performance of methods training on LOS dataset before and after sampling , where the increase in performance are marked as red and the decrease in variance of the performance of all the categories are marked as blue

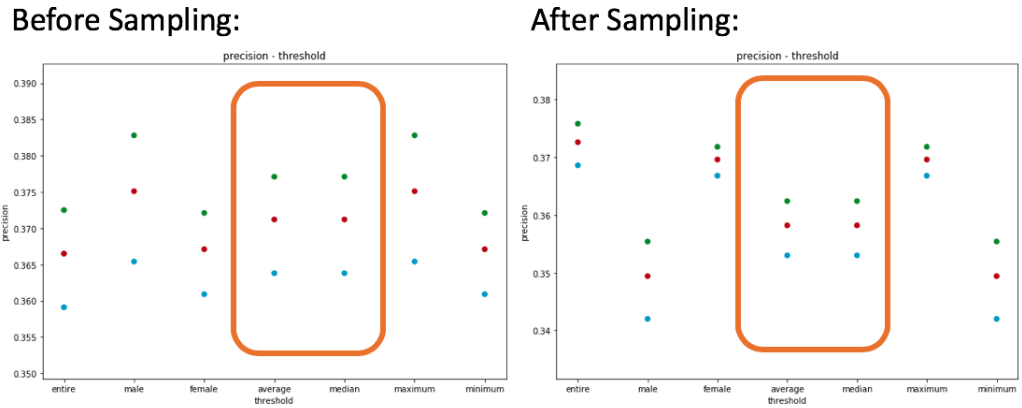


Figure 4.9: Performance Boundaries of the experiments choosing gender as the protected feature and trained with all the features on LOS dataset before (left) and after (right) sampling

4.5 Fairness in Methods

Table 4.2: AUC scores and variance of AUC scores of all the categories of the experiments choosing race as the protected feature (RF is short for Random Forest and LR for Logistic Regression)

	AUC score			Variance of AUC scores		
	XGBoost	RF	LR	xgboost	RF	LR
with 'race', before sampling	0.7066	0.6951	0.6557	0.000279	0.000215	0.000169
without 'race', before sampling	0.7069	0.6949	0.6552	0.000286	0.000214	0.000170
with 'race', after sampling	0.6944	0.6887	0.6436	0.000441	0.000318	0.000331
without 'race', after sampling	0.6918	0.6869	0.6431	0.000391	0.000305	0.000299

By comparing the AUC scores and the variance of AUC scores of the three methods, we could see the rank of the three methods based on AUC score is:

$$Logisticregression < RandomForest < XGBoost$$

, and the rank based on the variance in AUC score is:

$$Logisticregression < RandomForest < XGBoost(Race)$$

This could lead us to the conclusion that an accurate model does not necessarily imply fair outcomes.

Chapter 5

CONCLUSION

Identifying unfairness is challenging but doable. In this work, we proposed a general way to detect the fairness issue in performance of machine learning methods. We could first compare the performance across each category and use the variance of the performance as a criteria to measure the fairness. Second, we could compare the optimal thresholds chosen based on each category, the variance of which could also be used to help detect the fairness issue.

We also found several ways to address the fairness issue. First, we could train the model without the protected feature, which will help reduce the fairness issue but will not cause a significant drop in performance. The second method focuses on the data dimension, which is critical in a machine learning process, because the characteristics like underrepresentation could be inherited or even exemplified in machine learning process. When the size of dataset is small, or the one or several category is under-represented, we could sample the training dataset first to balance the size of each category. Plus, sampling will help a lot in reducing the fairness issue especially when we are using logistic regression. Finally, when we have detected the fairness issue and preferred to address it without training the model again, we could find the fairness threshold, which could make the performance more fair but also comparable.

Furthermore, the comparison among AUC scores of different models and the one among variance of AUC scores of different models lead us to the conclusion that an accurate model does not necessarily imply fairness. The model with higher accuracy may have less fair outcomes.

BIBLIOGRAPHY

[1] David Madras. Fairness in machine learning: An overview, 2017.

[2] James Bennett, Stan Lanning, et al. The netflix prize. In *Proceedings of KDD cup and workshop*, volume 2007, page 35. New York, NY, USA, 2007.

[3] Richard Berk. *Criminal justice forecasts of risk: A machine learning approach*. Springer Science & Business Media, 2012.

[4] David West, Scott Dellana, and Jingxia Qian. Neural network ensemble strategies for financial decision applications. *Computers & operations research*, 32(10):2543–2559, 2005.

[5] UN General Assembly. Universal declaration of human rights. *UN General Assembly*, 1948.

[6] Wikipedia. Anti-discrimination law, 2018.

[7] Wikipedia. Universal declaration of human rights, 2018.

[8] United Nations. Human rights law.

[9] Bernd Rechel. *Minority Rights in Central and Eastern Europe*. Routledge, 2008.

[10] Indre Zliobaite. Fairness-aware machine learning: a perspective. *arXiv preprint arXiv:1708.00754*, 2017.

[11] Global Future Council on Human Rights 2016-2018. How to prevent discriminatory outcomes in machine learning. Technical report, World Economic Forum, 03 2018.

[12] John P Dickerson and Tuomas Sandholm. Futurematch: Combining human value judgments and machine learning to match in dynamic environments. In *AAAI*, pages 622–628, 2015.

[13] Solon Barocas and Andrew D Selbst. Big data's disparate impact. *Cal. L. Rev.*, 104:671, 2016.

[14] Tolga Bolukbasi, Kai-Wei Chang, James Y Zou, Venkatesh Saligrama, and Adam T Kalai. Man is to computer programmer as woman is to homemaker? debiasing word embeddings. In *Advances in Neural Information Processing Systems*, pages 4349–4357, 2016.

[15] BBC NEWS. Google apologises for photos app's racist blunder, 2015.

[16] Amit Datta, Michael Carl Tschantz, and Anupam Datta. Automated experiments on ad privacy settings. *Proceedings on Privacy Enhancing Technologies*, 2015(1):92–112, 2015.

[17] Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner. Machine bias: There?s software used across the country to predict future criminals. *And it?s biased against blacks. ProPublica*, 2016.

[18] Michael Skirpan and Micha Gorelick. The authority of" fair" in machine learning. *arXiv preprint arXiv:1706.09976*, 2017.

[19] Google AI. Artificial intelligence at google our principles.

[20] Axon. Axon ai and policing technology ethics board.

[21] R Hart. If you?re not a white male, artificial intelligence?s use in healthcare could be dangerous. *Quartz. com*, 2017.

[22] William H Herman and Robert M Cohen. Racial and ethnic differences in the relationship between hba1c and blood glucose: implications for the diagnosis of diabetes. *The Journal of Clinical Endocrinology & Metabolism*, 97(4):1067–1072, 2012.

[23] Mark Thomas Hayes, Lynette Anne Hunt, Jonathan Foo, Yulia Tychinskaya, and Richard Strawson Stubbs. A model for predicting the resolution of type 2 diabetes in severely obese subjects following roux-en y gastric bypass surgery. *Obesity surgery*, 21(7):910–916, 2011.

[24] Ioannis Kavakiotis, Olga Tsave, Athanasios Salifoglou, Nicos Maglaveras, Ioannis Vlahavas, and Ioanna Chouvarda. Machine learning and data mining methods in diabetes research. *Computational and structural biotechnology journal*, 15:104–116, 2017.

[25] Indre Zliobaite. A survey on measuring indirect discrimination in machine learning. *arXiv preprint arXiv:1511.00148*, 2015.

[26] Indre Zliobaite. On the relation between accuracy and fairness in binary classification. *arXiv preprint arXiv:1505.05723*, 2015.

[27] Florian Tramer, Vaggelis Atlidakis, Roxana Geambasu, Daniel Hsu, Jean-Pierre Hubaux, Mathias Humbert, Ari Juels, and Huang Lin. Fairtest: Discovering unwarranted associations in data-driven applications. In *2017 IEEE European Symposium on Security and Privacy (EuroS&P)*, pages 401–416. IEEE, 2017.

[28] Meike Zehlike, Francesco Bonchi, Carlos Castillo, Sara Hajian, Mohamed Megahed, and Ricardo Baeza-Yates. Fa* ir: A fair top-k ranking algorithm. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pages 1569–1578. ACM, 2017.

[29] Ashudeep Singh and Thorsten Joachims. Fairness of exposure in rankings. *arXiv preprint arXiv:1802.07281*, 2018.

[30] Abhishek Tiwari. Bias and fairness in machine lea, 2017.

[31] Michael Feldman, Sorelle A Friedler, John Moeller, Carlos Scheidegger, and Suresh Venkatasubramanian. Certifying and removing disparate impact. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 259–268. ACM, 2015.

[32] Toon Calders, Asim Karim, Faisal Kamiran, Wasif Ali, and Xiangliang Zhang. Controlling attribute effect in linear regression. In *Data Mining (ICDM), 2013 IEEE 13th International Conference on*, pages 71–80. IEEE, 2013.

[33] Toshihiro Kamishima, Shotaro Akaho, Hideki Asoh, and Jun Sakuma. Fairness-aware classifier with prejudice remover regularizer. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 35–50. Springer, 2012.

[34] Faisal Kamiran, Toon Calders, and Mykola Pechenizkiy. Discrimination aware decision tree learning. In *Data Mining (ICDM), 2010 IEEE 10th International Conference on*, pages 869–874. IEEE, 2010.

[35] Equivant. Case management for supervision.

[36] Kate Crawford. The hidden biases in big data. *Harvard Business Review*, 1, 2013.

[37] Executive Office of the President, Cecilia Munoz, Domestic Policy Council Director, Megan (US Chief Technology Officer Smith (Office of Science, Technology Policy)),

DJ (Deputy Chief Technology Officer for Data Policy, Chief Data Scientist Patil (Office of Science, and Technology Policy)). *Big data: A report on algorithmic systems, opportunity, and civil rights*. Executive Office of the President, 2016.

[38] Latanya Sweeney. Discrimination in online ad delivery. *Queue*, 11(3):10, 2013.

[39] Alistair EW Johnson, Tom J Pollard, Lu Shen, H Lehman Li-wei, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G Mark. Mimic-iii, a freely accessible critical care database. *Scientific data*, 3:160035, 2016.

[40] Dua Dheeru and Efi Karra Taniskidou. UCI machine learning repository, 2017.

[41] R S Johannes. Using the adap learning algorithm to forecast the onset of diabetes mellitus. *Johns Hopkins APL Technical Digest*, 10:262–266, 1988.

[42] William J Youden. Index for rating diagnostic tests. *Cancer*, 3(1):32–35, 1950.

[43] Enrique F Schisterman, Neil J Perkins, Aiyi Liu, and Howard Bondell. Optimal cut-point and its corresponding youden index to discriminate individuals using pooled blood samples. *Epidemiology*, pages 73–81, 2005.

[44] David R Cox. The regression analysis of binary sequences. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 215–242, 1958.

[45] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[46] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794. ACM, 2016.

[47] Jason Brownlee. Feature importance and feature selection with xgboost in python, 2016.

Appendix A

RESULTS ON LOS DATASET

A.1 Introduction

The results on each dataset contain two performance sheets and one threshold sheets.

One performance sheet is for the performance of 10-fold cross validation on train dataset and the other one is for the performance on test dataset. The performance sheet contains the performance (which includes AUC score, precision, recall and f1 score) of entire dataset and each category obtained from each experiment with each machine learning method. It also contains the information about the variance of the performance of all the categories and how much the performance and the variance would change after implementing different different addressing solutions, include sampling, removal of protected features and removal of important features related to the protected feature (which is only contained in the performance sheet of 10-ford cross validation on train dataset of thyroid dataset). The increase in performance has been marked as red and the decrease in variance has been marked as blue. The short line, '-', represents that the corresponding performance is the baseline, and the slash, '/', represents that there is no significant different between the corresponding performance and the baseline.

The threshold sheet contains the candidate thresholds of fairness threshold and corresponding performance difference of each experiment. The chosen fairness threshold is marked as blue.

A.2 Performance sheet for 10-fold cross validation on train dataset

A.2.1 Performance while choosing 'age' as the protected feature

			Features	entire dataset			age: xx - 40			age: 41-70			age: 71-xx			Variance		
size of the group: first number: target = 1 second number: target = 0 third number: sum		test dataset	with/without the feature 'age'	2308, 7018 9326			213, 860, 1073			675, 2230, 2905			1420, 3928, 5348			variance of the three age groups		
		before sampling(train)		9234, 28070 37304			852, 3438, 4290			2701, 8918, 11619			5681, 15714, 21395					
		after sampling(train)		10447, 34550 44997			2979, 12020, 14999			3486, 11513, 14999			3982, 11017, 14999					
Change			-	sample	no 'age'	-	sample	no 'age'	-	sample	no 'age'	-	sample	no 'age'	-	sample	no 'age'	-
AUC	after normalization (XGBoost)	before sampling	with 'age'	-	-	0.7045	-	-	0.7424	-	-	0.7135	-	-	0.6866	-	-	0.000779
			without 'age'	-	-0.0017	0.7033	-	/	0.7424	-	-0.0006	0.7131	-	-0.0015	0.6856	-	0.0359	0.000807
			remove all	-	-0.0009	0.7039	-	0.0016	0.7436	-	0.0004	0.7138	-	-0.0010	0.6859	-	0.0693	0.000833
		after sampling	with 'age'	-0.0095	-	0.6978	-0.0092	-	0.7356	-0.0034	-	0.7111	-0.0141	-	0.6769	0.1168	-	0.000870
			without 'age'	-0.0101	-0.0023	0.6962	-0.0053	0.0039	0.7385	-0.0057	-0.0030	0.709	-0.0153	-0.0027	0.6751	0.2491	0.1586	0.001008
			remove all	-0.0116	-0.0030	0.6957	-0.0062	0.0046	0.739	-0.0056	-0.0018	0.7098	-0.0176	-0.0046	0.6738	0.2797	0.2253	0.001066
Precision	after normalization (XGBoost)	before sampling	with 'age'	-	-	0.7268	-	-	0.7835	-	-	0.7454	-	-	0.7065	-	-	0.001482
			without 'age'	-	0.0007	0.7273	-	-0.0032	0.781	-	0.0030	0.7476	-	0.0004	0.7068	-	-0.0682	0.001381
			remove all	-	0.0021	0.7283	-	-0.0043	0.7801	-	0.0034	0.7479	-	0.0030	0.7086	-	-0.1350	0.001282
		after sampling	with 'age'	-0.0083	-	0.7208	0.0034	-	0.7862	-0.0110	-	0.7372	-0.0093	-	0.6999	0.2638	-	0.001873
			without 'age'	-0.0118	-0.0029	0.7187	0.0049	-0.0018	0.7848	-0.0152	-0.0014	0.7362	-0.0137	-0.0040	0.6971	0.3975	0.0304	0.001930
			remove all	-0.0087	0.0017	0.722	/	-0.0078	0.7801	-0.0080	0.0064	0.7419	-0.0119	0.0004	0.7002	0.2480	-0.1458	0.001600
Recall	after normalization (XGBoost)	before sampling	with 'age'	-	-	0.7622	-	-	0.8131	-	-	0.7781	-	-	0.7433	-	-	0.001216
			without 'age'	-	0.0001	0.7623	-	-0.0010	0.8123	-	0.0005	0.7785	-	0.0003	0.7435	-	-0.0255	0.001185
			remove all	-	0.0007	0.7627	-	-0.0015	0.8119	-	0.0008	0.7787	-	0.0012	0.7442	-	-0.0576	0.001146
		after sampling	with 'age'	-0.0031	-	0.7598	0.0002	-	0.8133	-0.0041	-	0.7749	-0.0034	-	0.7408	0.0806	-	0.001314
			without 'age'	-0.0043	-0.0011	0.759	0.0006	-0.0006	0.8128	-0.0051	-0.0005	0.7745	-0.0050	-0.0013	0.7398	0.1266	0.0160	0.001335
			remove all	-0.0033	0.0005	0.7602	-0.0002	-0.0020	0.8117	-0.0026	0.0023	0.7767	-0.0044	0.0001	0.7409	0.0916	-0.0479	0.001251
f1 score	after normalization (XGBoost)	before sampling	with 'age'	-	-	0.7	-	-	0.7612	-	-	0.7183	-	-	0.678	-	-	0.001730
			without 'age'	-	-0.0006	0.6996	-	-0.0012	0.7603	-	-0.0010	0.7176	-	-0.0003	0.6778	-	-0.0162	0.001702
			remove all	-	0.0001	0.7001	-	-0.0030	0.7589	-	-0.0010	0.7176	-	0.0013	0.6789	-	-0.0746	0.001601
		after sampling	with 'age'	-0.0070	-	0.6951	-0.0005	-	0.7608	-0.0078	-	0.7127	-0.0081	-	0.6725	0.1277	-	0.001951
			without 'age'	-0.0081	-0.0017	0.6939	-0.0011	-0.0017	0.7595	-0.0084	-0.0015	0.7116	-0.0096	-0.0018	0.6713	0.1451	-0.0010	0.001949
			remove all	-0.0067	0.0004	0.6954	0.0020	-0.0005	0.7604	-0.0043	0.0025	0.7145	-0.0099	-0.0004	0.6722	0.2174	-0.0010	0.001949
AUC	after normalization (Random Forest)	before sampling	with 'age'	-	-	0.6921	-	-	0.7385	-	-	0.7008	-	-	0.6727	-	-	0.001090
			without 'age'	-	-0.0027	0.6902	-	-0.0007	0.738	-	-0.0051	0.6972	-	-0.0007	0.6722	-	0.0119	0.001103
			remove all	-	-0.0055	0.6883	-	-0.0073	0.7331	-	-0.0123	0.6922	-	-0.0003	0.6725	-	-0.1248	0.000954
		after sampling	with 'age'	-0.0100	-	0.6852	-0.0089	-	0.7319	-0.0114	-	0.6928	-0.0106	-	0.6656	0.0193	-	0.001111
			without 'age'	-0.0112	-0.0039	0.6825	-0.0118	-0.0036	0.7293	-0.0093	-0.0030	0.6907	-0.0137	-0.0039	0.663	0.0063	-0.0009	0.001110
			remove all	-0.0100	-0.0055	0.6814	-0.0063	-0.0046	0.7285	-0.0074	-0.0082	0.6871	-0.0138	-0.0036	0.6632	0.1447	-0.0171	0.001092
Precision	after normalization (Random Forest)	before sampling	with 'age'	-	-	0.7274	-	-	0.7826	-	-	0.7481	-	-	0.7057	-	-	0.001482
			without 'age'	-	-0.0077	0.7218	-	-0.0121	0.7731	-	-0.0056	0.7439	-	-0.0081	0.7	-	-0.0857	0.001355
			remove all	-	-0.0100	0.7201	-	-0.0079	0.7764	-	-0.0099	0.7407	-	-0.0111	0.6979	-	0.0432	0.001546
		after sampling	with 'age'	-0.0051	-	0.7237	-0.0003	-	0.7824	-0.0044	-	0.7448	-0.0079	-	0.7001	0.1457	-	0.001698
			without 'age'	-0.0062	-0.0088	0.7173	0.0056	-0.0061	0.7776	-0.0109	-0.0121	0.7358	-0.0080	-0.0081	0.6944	0.2753	0.0177	0.001728
			remove all	-0.0094	-0.0144	0.7133	0.0033	-0.0043	0.779	-0.0127	-0.0181	0.7313	-0.0113	-0.0144	0.69	0.2840	0.1690	0.001985
Recall	after normalization (Random Forest)	before sampling	with 'age'	-	-	0.76	-	-	0.8121	-	-	0.776	-	-	0.7408	-	-	0.001271
			without 'age'	-	-0.0016	0.7588	-	-0.0037	0.8091	-	-0.0006	0.7755	-	-0.0015	0.7397	-	-0.0527	0.001204
			remove all	-	-0.0018	0.7586	-	-0.0020	0.8105	-	-0.0012	0.7751	-	-0.0022	0.7392	-	0.0000	0.001271
		after sampling	with 'age'	-0.0018	-	0.7586	0.0004	-	0.8124	-0.0014	-	0.7749	-0.0024	-	0.739	0.0598	-	0.001347
			without 'age'	-0.0018	-0.0016	0.7574	0.0023	-0.0017	0.811	-0.0026	-0.0018	0.7735	-0.0024	-0.0015	0.7379	0.1080	-0.0097	0.001334
			remove all	-0.0025	-0.0025	0.7567	0.0017	-0.0006	0.8119	-0.0035	-0.0032	0.7724	-0.0028	-0.0026	0.7371	0.1007	0.0386	0.001399
f1 score	after normalization (Random Forest)	before sampling	with 'age'	-	-	0.6832	-	-	0.756	-	-	0.7032	-	-	0.6579	-	-	0.002413
			without 'age'	-	-0.0007	0.6827	-	-0.0040	0.753	-	0.0013	0.7041	-	-0.0011	0.6572	-	-0.0497	0.002293
			remove all	-	0.0010	0.6839	-	0.0001	0.7561	-	0.0031	0.7054	-	0.0000	0.6579	-	0.0000	0.002413
		after sampling	with 'age'	-0.0051	-	0.679	0.0013	-	0.757	-0.0031	-	0.701	-0.0097	-	0.6515	0.1542	-	0.002785
			without 'age'	-0.0054	0.0000	0.679	0.0052	-0.0001	0.7569	-0.0047	-0.0003	0.7008	-0.0087	0.0000	0.6515	0.2119	-0.0022	0.002779
			remove all	-0.0057	0.0015	0.68	0.0037	0.0025	0.7589	-0.0055	0.0007	0.7015	-0.0079	0.0018	0.6527	0.1720	0.0154	0.002828
AUC	after normalization (Logistic Regression)	before sampling	with 'age'	-	-	0.6515	-	-	0.683	-	-	0.6607	-	-	0.6346	-	-	0.000586
			without 'age'	-	-0.0008	0.651	-	-0.0023	0.6814	-	0.0011	0.6614	-	-0.0003	0.6344	-	-0.0495	0.000557
			remove all	-	-0.0009	0.6509	-	-0.0107	0.6757	-	0.0014	0.6616	-	-0.0002	0.6345	-	-0.2543	0.000437
		after sampling	with 'age'	-0.0078	-	0.6464	-0.0003	-	0.6828	0.0014	-	0.6616	-0.0140	-	0.6257	0.4249	-	0.000835
			without 'age'	-0.0084	-0.0014	0.6455	-0.0068	-0.0088	0.6768	0.0011	0.0008	0.6621	-0.0153	-0.0016	0.6247	0.2908	-0.1389	0.000719
			remove all	-0.0075	-0.0006	0.646	0.0089	-0.0016	0.6817	0.0009	0.0009	0.6622	-0.0147	-0.0008	0.6252	0.8787	-0.0168	0.000821
Precision	after normalization (Random Forest)	before sampling	with 'age'	-	-	0.7047	-	-	0.7638	-	-	0.7199	-	-	0.683	-	-	0.001639
			without 'age'	-	-0.0001	0.7046	-	-0.0041	0.7607	-	0.0022	0.7215	-	-0.0012	0.6822	-	-0.0604	0.001540
			remove all	-	0.0007	0.7052	-	-0.0022	0.7621	-	0.0019	0.7213	-	0.0004	0.6833	-	-0.0531	0.001552
		after sampling	with 'age'	-0.0014	-	0.7037	-0.0022	-	0.7621	0.0043	-	0.723	-0.0063	-	0.6787	0.0535	-	0.001743
			without 'age'	-0.0006	0.0007	0.7042	0.0013	-0.0005	0.7617	0.0035	0.0015	0.7241	-0.0045	0.0006	0.6791	0.1097	-0.0195	0.001709
			remove all	-0.0014	0.0007	0.7042	/	0.0000	0.7621	0.0029	0.0006	0.7234	-0.0056	0.0012	0.6795	0.0999	-0.0207	0.001707
Recall	after normalization (Random Forest)	before sampling	with 'age'	-	-	0.631	-	-	0.7627	-	-	0.6717	-	-	0.5825	-	-	0.008113
			without 'age'	-	0.0021	0.6323	-	-0.0229	0.7452	-	0.0138	0.681	-	0.0010	0.5831	-	-0.1782	0.006667
			remove all	-	0.0040	0.6335	-	-0.0220	0.7459	-	0.0165	0.6828	-	0.0031	0.5843	-	-0.1818	0.006638
		after sampling	with 'age'	-0.0429	-	0.6039	-0.0239	-	0.7445	-0.0320	-	0.6502	-0.0548	-	0.5506	0.1595	-	0.009407
			without 'age'	-0.0367	0.0086	0.6091	-0.0122	-0.0113	0.7351	-0.0169	0.0297	0.6695	-0.0554	0.0004	0.5508	0.3219	-0.0631	0.008813
			remove all	-0.0393	0.0078	0.6086	-0.0125	-0.0106	0.7366	-0.0228	0.0261	0.6672	-0.0566	0.0011	0.5512	0.3215	-0.0675	0.008772
f1 score	after normalization (Random Forest)	before sampling	with 'age'	-	-	0.6539	-	-	0.7627	-	-	0.6894	-	-	0.6069	-	-	0.006081
			without 'age'	-	0.0014	0.6548	-	-0.0143	0.7518	-	0.0103	0.6965	-	0.0008	0.6074	-	-0.1273	0.005307
			remove all	-	0.0032	0.656	-	-0.0131	0.7527	-	0.0119	0.6976	-	0.0026	0.6085	-	-0.1289	0.005297
		after sampling	with 'age'	-0.0356	-	0.6306	-0.0143	-	0.7518	-0.0228	-	0.6737	-0.0514	-	0.5757	0.2802	-	0.007785
			without 'age'	-0.0299	0.0073	0.6352	-0.0070	-0.0070	0.7465	-0.0111	0.0224	0.6888	-0.0517	0.0005	0.576	0.4185	-0.0330	0.007528
			remove all	-0.0323	0.0067	0.6348	-0.0077	-0.0065	0.7469	-0.0153	0.0196	0.6869	-0.0529	0.0010	0.5763	0.4136	-0.0382	0.007488

Figure A.1: Performance of 10-fold cross validation while choosing 'age' as the protected feature (LOS dataset)

A.2.2 Performance while choosing 'gender' as the protected feature

			Features	entire dataset			gender						Variance		
							male			female					
size of the group: first number: target = 1 second number: target = 0 third number: sum		test dataset	with/without features related to gender	2309, 7017, 9326			1308, 4028, 5336			1001, 2989, 3990			variance of the two gender groups		
		before sampling(train)		9233, 28071, 37304			5229, 16113, 21342			4004, 11958, 15962					
		after sampling(train)		9916, 30082, 39998			4900, 15099, 19999			5016, 14983, 19999					
Change			-	sample	no 'sex'	-	sample	no 'sex'	-	sample	no 'sex'	-	sample	no 'sex'	-
AUC	after normalization (XGBoost)	before sampling	with 'gender'	-	-	0.7075	-	-	0.7139	-	-	0.6993	-	-	0.000107
			without 'gender'	-	-0.0013	0.7066	-	-0.0011	0.7131	-	-0.0014	0.6983	-	0.0206	0.000109
			remove all	-	-0.0017	0.7063	-	-0.0018	0.7126	-	-0.0020	0.6979	-	0.0124	0.000108
		after sampling	with 'gender'	-0.0116	-	0.6993	-0.0123	-	0.7051	-0.0109	-	0.6917	-0.1487	-	0.000091
			without 'gender'	-0.0106	-0.0003	0.6991	-0.0114	-0.0001	0.705	-0.0096	-0.0001	0.6916	-0.1795	-0.0163	0.000089
			remove all	-0.0102	-0.0003	0.6991	-0.0104	0.0001	0.7052	-0.0097	-0.0009	0.6911	-0.0761	0.0987	0.000100
Precision	after normalization (XGBoost)	before sampling	with 'gender'	-	-	0.7286	-	-	0.7317	-	-	0.7252	-	-	0.000021
			without 'gender'	-	-0.0008	0.728	-	-0.0011	0.7309	-	-0.0001	0.7251	-	-0.1952	0.000017
			remove all	-	-0.0004	0.7283	-	-0.0023	0.73	-	0.0019	0.7266	-	-0.7186	0.000006
		after sampling	with 'gender'	-0.0066	-	0.7238	-0.0087	-	0.7253	-0.0034	-	0.7227	-0.8267	-	0.000004
			without 'gender'	-0.0074	-0.0017	0.7226	-0.0074	0.0003	0.7255	-0.0083	-0.0050	0.7191	0.2740	4.9157	0.000021
			remove all	-0.0122	-0.0061	0.7194	-0.0103	-0.0039	0.7225	-0.0147	-0.0094	0.7159	2.7336	5.0618	0.000022
Recall	after normalization (XGBoost)	before sampling	with 'gender'	-	-	0.7629	-	-	0.7655	-	-	0.7595	-	-	0.000018
			without 'gender'	-	-0.0003	0.7627	-	-0.0003	0.7653	-	-0.0003	0.7593	-	-0.0131	0.000018
			remove all	-	-0.0001	0.7628	-	-0.0008	0.7649	-	0.0007	0.76	-	-0.3437	0.000012
		after sampling	with 'gender'	-0.0022	-	0.7612	-0.0030	-	0.7632	-0.0013	-	0.7585	-0.3962	-	0.000011
			without 'gender'	-0.0026	-0.0007	0.7607	-0.0029	0.0001	0.7633	-0.0028	-0.0017	0.7572	0.0338	0.6896	0.000019
			remove all	-0.0043	-0.0022	0.7595	-0.0035	-0.0013	0.7622	-0.0054	-0.0034	0.7559	0.6520	0.7955	0.000020
f1 score	after normalization (XGBoost)	before sampling	with 'gender'	-	-	0.701	-	-	0.7036	-	-	0.6976	-	-	0.000018
			without 'gender'	-	-0.0009	0.7004	-	-0.0004	0.7033	-	-0.0016	0.6965	-	0.3077	0.000023
			remove all	-	-0.0006	0.7006	-	-0.0016	0.7025	-	0.0003	0.6978	-	-0.3841	0.000011
		after sampling	with 'gender'	-0.0029	-	0.699	-0.0028	-	0.7016	-0.0029	-	0.6956	-0.0067	-	0.000018
			without 'gender'	-0.0030	-0.0010	0.6983	-0.0033	-0.0009	0.701	-0.0026	-0.0013	0.6947	-0.1650	0.0993	0.000020
			remove all	-0.0053	-0.0030	0.6969	-0.0040	-0.0027	0.6997	-0.0069	-0.0037	0.693	1.0615	0.2783	0.000023
AUC	after normalization (Random Forest)	before sampling	with 'gender'	-	-	0.694	-	-	0.6975	-	-	0.6895	-	-	0.000032
			without 'gender'	-	-0.0003	0.6938	-	0.0017	0.6987	-	-0.0032	0.6873	-	1.0429	0.000065
			remove all	-	-0.0035	0.6916	-	-0.0017	0.6963	-	-0.0061	0.6853	-	0.9195	0.000061
		after sampling	with 'gender'	-0.0068	-	0.6893	-0.0083	-	0.6917	-0.0049	-	0.6861	-0.5157	-	0.000015
			without 'gender'	-0.0079	-0.0015	0.6883	-0.0096	0.0004	0.692	-0.0060	-0.0042	0.6832	-0.4116	1.4819	0.000038
			remove all	-0.0093	-0.0059	0.6852	-0.0113	-0.0048	0.6884	-0.0063	-0.0074	0.681	-0.5543	0.7665	0.000027
Precision	after normalization (Random Forest)	before sampling	with 'gender'	-	-	0.7305	-	-	0.7358	-	-	0.7254	-	-	0.000055
			without 'gender'	-	-0.0015	0.7294	-	-0.0001	0.7357	-	-0.0052	0.7216	-	0.8187	0.000099
			remove all	-	-0.0042	0.7274	-	-0.0049	0.7322	-	-0.0043	0.7223	-	-0.1171	0.000048
		after sampling	with 'gender'	-0.0034	-	0.728	-0.0038	-	0.733	-0.0045	-	0.7221	0.0919	-	0.000060
			without 'gender'	-0.0088	-0.0069	0.723	-0.0107	-0.0071	0.7278	-0.0046	-0.0053	0.7183	-0.5428	-0.2384	0.000045
			remove all	-0.0036	-0.0044	0.7248	-0.0046	-0.0057	0.7288	-0.0030	-0.0028	0.7201	-0.2134	-0.3640	0.000038
Recall	after normalization (Random Forest)	before sampling	with 'gender'	-	-	0.7608	-	-	0.7639	-	-	0.7565	-	-	0.000028
			without 'gender'	-	-0.0001	0.7607	-	0.0003	0.7641	-	-0.0005	0.7561	-	0.1768	0.000032
			remove all	-	-0.0004	0.7605	-	-0.0004	0.7636	-	-0.0001	0.7564	-	-0.0373	0.000027
		after sampling	with 'gender'	-0.0012	-	0.7599	-0.0010	-	0.7631	-0.0011	-	0.7557	-0.0268	-	0.000027
			without 'gender'	-0.0022	-0.0012	0.759	-0.0026	-0.0013	0.7621	-0.0017	-0.0012	0.7548	-0.1798	-0.0082	0.000027
			remove all	-0.0012	-0.0004	0.7596	-0.0014	-0.0008	0.7625	-0.0009	0.0000	0.7557	-0.1272	-0.1366	0.000023
f1 score	after normalization (Random Forest)	before sampling	with 'gender'	-	-	0.6842	-	-	0.6889	-	-	0.6778	-	-	0.000062
			without 'gender'	-	0.0009	0.6848	-	0.0009	0.6895	-	0.0010	0.6785	-	-0.0216	0.000060
			remove all	-	0.0037	0.6867	-	0.0032	0.6911	-	0.0044	0.6808	-	-0.1391	0.000053
		after sampling	with 'gender'	-0.0026	-	0.6824	-0.0032	-	0.6867	-0.0018	-	0.6766	-0.1820	-	0.000050
			without 'gender'	-0.0041	-0.0006	0.682	-0.0044	-0.0003	0.6865	-0.0035	-0.0007	0.6761	-0.1045	0.0710	0.000054
			remove all	-0.0035	0.0028	0.6843	-0.0043	0.0020	0.6881	-0.0025	0.0037	0.6791	-0.2347	-0.1946	0.000041
AUC	after normalization (Logistic Regression)	before sampling	with 'gender'	-	-	0.6538	-	-	0.6615	-	-	0.6435	-	-	0.000163
			without 'gender'	-	-0.0003	0.6536	-	-0.0002	0.6614	-	0.0002	0.6436	-	-0.0340	0.000158
			remove all	-	-0.0012	0.653	-	0.0003	0.6617	-	-0.0022	0.6421	-	0.1669	0.000191
		after sampling	with 'gender'	-0.0069	-	0.6493	-0.0054	-	0.6579	-0.0090	-	0.6377	0.2433	-	0.000203
			without 'gender'	-0.0069	-0.0003	0.6491	-0.0054	-0.0002	0.6578	-0.0090	0.0002	0.6378	0.2642	-0.0178	0.000200
			remove all	-0.0069	-0.0012	0.6485	-0.0054	0.0003	0.6581	-0.0092	-0.0024	0.6362	0.2528	0.1758	0.000239
Precision	after normalization (Random Forest)	before sampling	with 'gender'	-	-	0.705	-	-	0.711	-	-	0.697	-	-	0.000098
			without 'gender'	-	0.0003	0.7052	-	-0.0011	0.7102	-	0.0023	0.6986	-	-0.3165	0.000067
			remove all	-	0.0010	0.7057	-	0.0001	0.7111	-	0.0024	0.6987	-	-0.2007	0.000078
		after sampling	with 'gender'	-0.0023	-	0.7034	-0.0031	-	0.7088	-0.0013	-	0.6961	-0.1696	-	0.000081
			without 'gender'	-0.0016	0.0010	0.7041	-0.0018	0.0001	0.7089	-0.0011	0.0024	0.6978	-0.0798	-0.2425	0.000061
			remove all	-0.0023	0.0010	0.7041	-0.0035	-0.0003	0.7086	-0.0006	0.0032	0.6983	-0.3213	-0.3467	0.000053
Recall	after normalization (Random Forest)	before sampling	with 'gender'	-	-	0.633	-	-	0.6426	-	-	0.6202	-	-	0.000250
			without 'gender'	-	0.0009	0.6336	-	0.0100	0.649	-	-0.0114	0.6131	-	1.5726	0.000644
			remove all	-	0.0033	0.6351	-	0.0154	0.6525	-	-0.0137	0.6117	-	2.3232	0.000832
		after sampling	with 'gender'	-0.0006	-	0.6326	-0.0025	-	0.641	0.0019	-	0.6214	-0.2276	-	0.000193
			without 'gender'	0.0016	0.0032	0.6346	0.0002	0.0126	0.6491	0.0033	-0.0101	0.6151	-0.1021	1.9908	0.000578
			remove all	0.0002	0.0041	0.6352	-0.0028	0.0151	0.6507	0.0046	-0.0111	0.6145	-0.2107	2.3960	0.000656
f1 score	after normalization (Random Forest)	before sampling	with 'gender'	-	-	0.6555	-	-	0.6644	-	-	0.6435	-	-	0.000218
			without 'gender'	-	0.0008	0.656	-	0.0074	0.6693	-	-0.0092	0.6376	-	1.3020	0.000503
			remove all	-	0.0027	0.6573	-	0.0119	0.6723	-	-0.0110	0.6364	-	1.9391	0.000642
		after sampling	with 'gender'	-0.0008	-	0.655	-0.0024	-	0.6628	0.0014	-	0.6444	-0.2190	-	0.000171
			without 'gender'	0.0011	0.0026	0.6567	/	0.0098	0.6693	0.0027	-0.0079	0.6393	-0.1067	1.6329	0.000449
			remove all	-0.0002	0.0034	0.6572	-0.0028	0.0115	0.6704	0.0038	-0.0087	0.6388	-0.2217	1.9290	0.000500

Figure A.2: Performance of 10-fold cross validation while choosing 'gender' as the protected feature (LOS dataset)

A.2.3 Performance while choosing 'race' as the protected feature

			Features	entire dataset			race: White			race: Black/AfricanAmerican			race: Hispanic/Latino			race: Asian			race: Other			Variance		
size of the group: first number: target = 1 second number: target = 0 third number: sum		test dataset	with/without features related to race	2308, 7018, 9326			1638, 4971, 6609			193, 659, 852			57, 203, 260			30, 110, 140			390, 1075, 1465			variance of the five race groups		
		before sampling(train)		9234, 28070, 37304			6551, 19882, 26433			771, 2636, 3407			230, 811, 1041			121, 439, 560			1561, 4302, 5863					
		after sampling(train)		23544, 76451, 99995			4956, 15043, 19999			4525, 15474, 19999			4418, 15581, 19999			4321, 15678, 19999			5324, 14675, 19999					
Change			-	sample	no 'race'	-	sample	no 'race'	-	sample	no 'race'	-	sample	no 'race'	-	sample	no 'race'	-	sample	no 'race'	-	sample	no 'race'	-
AUC	after normalization (XGBoost)	before sampling	with 'race'	-	-	0.7066	-	-	0.7005	-	-	0.8935	-	-	0.7207	-	-	0.7042	-	-	0.7346	-	-	0.000279
			without 'race'	-	0.0004	0.7069	-	0.0006	0.7009	-	-0.0009	0.8929	-	-0.0010	0.72	-	-0.0003	0.704	-	0.0007	0.7351	-	0.0251	0.000286
			remove all	-	-0.0024	0.7049	-	-0.0019	0.6992	-	-0.0022	0.892	-	-0.0105	0.7131	-	0.0044	0.7073	-	-0.0042	0.7315	-	-0.1828	0.000228
		after sampling	with 'race'	-0.0173	-	0.6944	-0.0177	-	0.6881	-0.0208	-	0.8791	-0.0010	-	0.72	-0.0278	-	0.6846	-0.0151	-	0.7235	0.5820	-	0.000441
			without 'race'	-0.0214	-0.0037	0.6918	-0.0237	-0.0055	0.6843	-0.0208	-0.0009	0.8785	-0.0042	-0.0042	0.717	-0.0166	0.0112	0.6923	-0.0170	-0.0012	0.7226	0.3685	-0.1133	0.000391
			remove all	-0.0183	-0.0035	0.692	-0.0215	-0.0057	0.6842	-0.0130	0.0057	0.883	0.0024	-0.0072	0.7148	-0.0048	0.0282	0.7039	-0.0130	-0.0021	0.722	0.3642	-0.2953	0.000311
Precision	after normalization (XGBoost)	before sampling	with 'race'	-	-	0.7298	-	-	0.7233	-	-	0.7479	-	-	0.7736	-	-	0.741	-	-	0.7311	-	-	0.000374
			without 'race'	-	0.0026	0.7317	-	0.0036	0.7259	-	0.0051	0.7517	-	-0.0050	0.7697	-	0.0189	0.755	-	-0.0012	0.7302	-	-0.1124	0.000332
			remove all	-	0.0032	0.7321	-	0.0035	0.7259	-	0.0119	0.7568	-	-0.0043	0.7703	-	0.0121	0.75	-	0.0010	0.7318	-	-0.1140	0.000331
		after sampling	with 'race'	-0.0075	-	0.7243	-0.0062	-	0.7188	-0.0033	-	0.7454	-0.0292	-	0.751	-0.0061	-	0.7365	-0.0034	-	0.7286	-0.5562	-	0.000166
			without 'race'	-0.0160	-0.0059	0.72	-0.0164	-0.0067	0.714	-0.0218	-0.0135	0.7353	-0.0309	-0.0068	0.7459	-0.0127	0.0121	0.7454	-0.0085	-0.0063	0.724	-0.4198	0.1604	0.000192
			remove all	-0.0164	-0.0058	0.7201	-0.0146	-0.0049	0.7153	-0.0235	-0.0086	0.739	0.0030	0.0288	0.7726	-0.0009	0.0174	0.7493	-0.0179	-0.0136	0.7187	0.6683	2.3305	0.000552
Recall	after normalization (XGBoost)	before sampling	with 'race'	-	-	0.7532	-	-	0.7505	-	-	0.7837	-	-	0.7944	-	-	0.7991	-	-	0.7524	-	-	0.000397
			without 'race'	-	0.0010	0.764	-	0.0014	0.7616	-	0.0008	0.7843	-	0.0000	0.7944	-	0.0043	0.7995	-	-0.0005	0.752	-	0.0793	0.000429
			remove all	-	0.0013	0.7642	-	0.0014	0.7616	-	0.0031	0.7861	-	0.0000	0.7944	-	0.0015	0.7973	-	0.0003	0.7526	-	0.0194	0.000405
		after sampling	with 'race'	-0.0025	-	0.7613	-0.0020	-	0.759	-0.0033	-	0.7811	-0.0097	-	0.7867	-0.0143	-	0.7847	-0.0008	-	0.7518	-0.3503	-	0.000258
			without 'race'	-0.0058	-0.0022	0.7596	-0.0059	-0.0023	0.7571	-0.0075	-0.0035	0.7784	-0.0060	0.0037	0.7896	-0.0099	0.0088	0.7916	-0.0039	-0.0036	0.7491	-0.1422	0.4249	0.000368
			remove all	-0.0060	-0.0022	0.7596	-0.0053	-0.0018	0.7576	-0.0098	-0.0035	0.7784	-0.0037	0.0061	0.7915	-0.0100	0.0059	0.7893	-0.0077	-0.0067	0.7468	-0.0370	0.5108	0.000390
f1 score	after normalization (XGBoost)	before sampling	with 'race'	-	-	0.7016	-	-	0.5975	-	-	0.721	-	-	0.729	-	-	0.7346	-	-	0.6969	-	-	0.000312
			without 'race'	-	0.0024	0.7033	-	0.0030	0.6996	-	0.0026	0.7229	-	0.0015	0.7301	-	0.0048	0.7381	-	-0.0007	0.6964	-	0.1039	0.000344
			remove all	-	0.0026	0.7034	-	0.0027	0.6994	-	0.0078	0.7266	-	0.0001	0.7291	-	0.0023	0.7363	-	-0.0003	0.6967	-	0.0674	0.000333
		after sampling	with 'race'	-0.0011	-	0.7008	-0.0013	-	0.6966	-0.0021	-	0.7195	-0.0058	-	0.7248	-0.0159	-	0.7229	0.0036	-	0.6994	-0.4139	-	0.000183
			without 'race'	-0.0054	-0.0019	0.6995	-0.0061	-0.0019	0.6953	-0.0072	-0.0025	0.7177	-0.0056	0.0017	0.726	-0.0035	0.0174	0.7355	-0.0007	-0.0050	0.6959	-0.0578	0.7745	0.000324
			remove all	-0.0063	-0.0026	0.699	-0.0061	-0.0022	0.6951	-0.0120	-0.0022	0.7179	-0.0001	0.0058	0.729	-0.0008	0.0177	0.7357	-0.0056	-0.0094	0.6928	0.1425	1.0805	0.000380
AUC	after normalization (Random Forest)	before sampling	with 'race'	-	-	0.6931	-	-	0.6898	-	-	0.683	-	-	0.7109	-	-	0.7005	-	-	0.7187	-	-	0.000215
			without 'race'	-	-0.0003	0.6949	-	0.0000	0.6898	-	0.0018	0.6842	-	-0.0027	0.709	-	-0.0141	0.6908	-	-0.0001	0.7186	-	-0.0005	0.000214
			remove all	-	-0.0037	0.6925	-	-0.0038	0.6872	-	0.0012	0.6838	-	-0.0079	0.7053	-	-0.0066	0.6959	-	-0.0050	0.7151	-	-0.2200	0.000167
		after sampling	with 'race'	-0.0092	-	0.6887	-0.0104	-	0.6826	-0.0102	-	0.676	0.0018	-	0.7122	-0.0200	-	0.6865	-0.0056	-	0.7147	0.4811	-	0.000318
			without 'race'	-0.0115	-0.0026	0.6869	-0.0136	-0.0032	0.6804	-0.0089	0.0031	0.6781	0.0062	0.0017	0.7134	-0.0071	-0.0012	0.6857	-0.0086	-0.0032	0.7124	0.4240	-0.0390	0.000305
			remove all	-0.0121	-0.0067	0.6841	-0.0137	-0.0070	0.6778	-0.0151	-0.0037	0.6735	-0.0096	-0.0192	0.6985	-0.0105	0.0031	0.6886	-0.0067	-0.0062	0.7103	0.3521	-0.2880	0.000226
Precision	after normalization (Random Forest)	before sampling	with 'race'	-	-	0.7311	-	-	0.7251	-	-	0.7423	-	-	0.7717	-	-	0.6867	-	-	0.7321	-	-	0.000943
			without 'race'	-	-0.0031	0.7288	-	-0.0015	0.724	-	0.0036	0.745	-	0.0165	0.7844	-	0.0071	0.6916	-	-0.0020	0.7306	-	0.2101	0.001142
			remove all	-	-0.0081	0.7252	-	-0.0090	0.7186	-	-0.0053	0.7384	-	-0.0320	0.747	-	0.0935	0.7509	-	-0.0046	0.7287	-	-0.8140	0.000175
		after sampling	with 'race'	-0.0115	-	0.7227	-0.0025	-	0.7233	-0.0314	-	0.719	-0.0438	-	0.7379	0.0810	-	0.7423	-0.0219	-	0.7161	-0.8552	-	0.000137
			without 'race'	-0.0073	0.0011	0.7235	-0.0061	-0.0051	0.7196	-0.0318	0.0032	0.7213	-0.0403	0.0202	0.7528	0.0510	-0.0207	0.7269	-0.0048	0.0154	0.7271	-0.8423	0.3177	0.000180
			remove all	-0.0099	-0.0065	0.718	-0.0085	-0.0149	0.7125	-0.0318	-0.0057	0.7149	0.0229	0.0355	0.7641	-0.0045	0.0070	0.7475	-0.0100	0.0074	0.7214	1.9236	2.7562	0.000513
Recall	after normalization (Random Forest)	before sampling	with 'race'	-	-	0.761	-	-	0.7591	-	-	0.7793	-	-	0.7973	-	-	0.7881	-	-	0.7481	-	-	0.000416
			without 'race'	-	-0.0004	0.7607	-	-0.0001	0.759	-	-0.0012	0.7784	-	-0.0036	0.7944	-	-0.0015	0.7859	-	-0.0003	0.7479	-	-0.0926	0.000377
			remove all	-	-0.0013	0.76	-	-0.0017	0.7578	-	-0.0008	0.7787	-	-0.0036	0.7944	-	0.0086	0.7949	-	-0.0012	0.7472	-	0.1162	0.000464
		after sampling	with 'race'	-0.0033	-	0.7585	-0.0017	-	0.7578	-0.0076	-	0.7734	-0.0133	-	0.7867	0.0015	-	0.7893	-0.0072	-	0.7427	-0.0625	-	0.000390
			without 'race'	-0.0022	0.0007	0.759	-0.0020	-0.0004	0.7575	-0.0053	0.0012	0.7743	-0.0060	0.0037	0.7896	0.0060	0.0029	0.7916	-0.0032	0.0038	0.7455	0.0660	0.0318	0.000402
			remove all	-0.0026	-0.0007	0.758	-0.0021	-0.0021	0.7562	-0.0072	-0.0004	0.7731	-0.0013	0.0085	0.7934	0.0001	0.0072	0.795	-0.0027	0.0034	0.7452	0.0541	0.2550	0.000489
f1 score	after normalization (Random Forest)	before sampling	with 'race'	-	-	0.6858	-	-	0.6823	-	-	0.7006	-	-	0.7266	-	-	0.7107	-	-	0.6795	-	-	0.000389
			without 'race'	-	0.0010	0.6865	-	0.0009	0.6829	-	0.0044	0.7037	-	-0.0051	0.7229	-	-0.0004	0.7104	-	0.0016	0.6805	-	-0.1520	0.000330
			remove all	-	0.0020	0.6872	-	0.0015	0.6833	-	0.0081	0.7063	-	-0.0056	0.7225	-	0.0259	0.7291	-	0.0000	0.6795	-	0.2928	0.000503
		after sampling	with 'race'	-0.0090	-	0.6798	-0.0092	-	0.676	-0.0041	-	0.6977	-0.0164	-	0.7147	0.0042	-	0.7137	-0.0112	-	0.6719	0.0543	-	0.000410
			without 'race'	-0.0070	0.0031	0.6817	-0.0072	0.0030	0.678	-0.0077	0.0009	0.6983	-0.0102	0.0011	0.7155	0.0155	0.0108	0.7214	-0.0082	0.0046	0.675	0.3519	0.0874	0.000446
			remove all	-0.0055	0.0056	0.6834	-0.0061	0.0046	0.6791	-0.0089	0.0033	0.7	0.0022	0.0132	0.7241	0.0030	0.0247	0.7313	-0.0034	0.0079	0.6772	0.2396	0.5200	0.000623
AUC	after normalization (Logistic Regression)	before sampling	with 'race'	-	-	0.6557	-	-	0.6503	-	-	0.6585	-	-	0.6697	-	-	0.6423	-	-	0.6733	-	-	0.000169
			without 'race'	-	-0.0008	0.6552	-	-0.0002	0.6502	-	0.0011	0.6592	-	0.0024	0.6713	-	0.0005	0.6426	-	-0.0012	0.6725	-	0.0107	0.000170
			remove all	-	-0.0018	0.6545	-	-0.0011	0.6496	-	0.0003	0.6587	-	0.0012	0.6705	-	-0.0064	0.6382	-	-0.0021	0.6719	-	0.2047	0.000203
		after sampling	with 'race'	-0.0185	-	0.6436	-0.0211	-	0.6366	-0.0173	-	0.6471	/	-	0.6697	-0.0218	-	0.6283	-0.0100	-	0.6666	0.9614	-	0.000331
			without 'race'	-0.0185	-0.0008	0.6431	-0.0214	-0.0005	0.6363	-0.0176	0.0008	0.6476	-0.0037	-0.0013	0.6688	-0.0202	0.0021	0.6296	-0.0109	-0.0021	0.6652	0.7528	-0.0968	0.000299
			remove all	-0.0180	-0.0014	0.6427	-0.0212	-0.0013	0.6358	-0.0162	0.0014	0.648	0.0007	0.0019	0.671	-0.0136	0.0019	0.6295	-0.0097	-0.0018	0.6654	0.6074	-0.0127	0.000326
Precision	after normalization (Random Forest)	before sampling	with 'race'	-	-	0.7068	-	-	0.7047	-	-	0.7195	-	-	0.7229	-	-	0.7228	-	-	0.7023	-	-	0.000102
			without 'race'	-	-0.0006	0.7064	-	-0.0017	0.7035	-	0.0049	0.723	-	0.0048	0.7264	-	0.0089	0.7292	-	0.0007	0.7028	-	0.6035	0.000164
			remove all	-	-0.0010	0.7061	-	-0.0023	0.7031	-	0.0082	0.7254	-	0.0032	0.7252	-	0.0101	0.7301	-	-0.0004	0.702	-	0.7793	0.000182
		after sampling	with 'race'	-0.0041	-	0.7039	-0.0077	-	0.6993	0.0068	-	0.7244	0.0090	-	0.7294	0.0001	-	0.7229	0.0026	-	0.7041	0.7500	-	0.000179
			without 'race'	-0.0048	-0.0013	0.703	-0.0068	-0.0009	0.6987	0.0024	0.0004	0.7247	-0.0001	-0.0043	0.7263	0.0053	0.0141	0.7331	-0.0023	-0.0041	0.7012	0.5104	0.3839	0.000248
			remove all	-0.0057	-0.0026	0.7021	-0.0074	-0.0020	0.6979	-0.0008	0.0006	0.7248	-0.0007	-0.0064	0.7247	0.0029	0.0129	0.7322	-0.0044	-0.0074	0.6989	0.4199	0.4436	0.000259
Recall	after normalization (Random Forest)	before sampling	with 'race'	-	-	0.6354	-	-	0.6326	-	-	0.6672	-	-	0.6906	-	-	0.6886	-	-	0.6104	-	-	0.001250
			without 'race'	-	0.0002	0.6355	-	0.0046	0.6355	-	-0.0450	0.6372	-	-0.0348	0.6666	-	-0.0667	0.6427	-	0.0280	0.6275	-	-0.8234	0.000221
			remove all	-	0.0016	0.6364	-	0.0070	0.637	-	-0.0480	0.6352	-	-0.0291	0.6705	-	-0.0717	0.6392	-	0.0277	0.6273	-	-0.7779	0.000278
		after sampling	with 'race'	-0.0466	-	0.6058	-0.0536	-	0.5987	-0.0300	-	0.6472	-0.0207	-	0.6763	-0.0202	-	0.6747	-0.0336	-	0.5899	0.3512	-	0.001689
			without 'race'	-0.0173	0.0309	0.6245	-0.0192	0.0411	0.6233	0.0046	-0.0110	0.6401	0.0030	-0.0114	0.6686	0.0178	-0.0307	0.654	-0.0319	0.0298	0.6075	1.6502	-0.8536	0.000585
			remove all	-0.0195	0.0300	0.624	-0.0218	0.0408	0.6231	0.0091	-0.0096	0.641	-0.0013	-0.0099	0.6696	0.0196	-0.0341	0.6517	-0.0362	0.0249	0.6046	1.2752	-0.8260	0.000632
f1 score	after normalization (Random Forest)	before sampling	with 'race'	-	-	0.6577	-	-	0.6551	-	-	0.6863	-	-	0.7036	-	-	0.7019	-	-	0.6322	-	-	0.000974
			without 'race'	-	0.0000	0.6577	-	0.0034	0.6573	-	-0.0326	0.6639	-	-0.0222	0.688	-	-0.0450	0.6703	-	0.0245	0.6477	-	-0.7647	0.000229
			remove all	-	0.0011	0.6584	-	0.0050	0.6584	-	-0.0347	0.6625	-	-0.0186	0.6905	-	-0.0483	0.668	-	0.0240	0.6474	-	-0.7397	0.000253
		after sampling	with 'race'	-0.0386	-	0.6323	-0.0449	-	0.6257	-0.0210	-	0.6719	-0.0117	-	0.6954	-0.0138	-	0.6922	-0.0312	-	0.6125	0.5051	-	0.001465
			without 'race'	-0.0144	0.0251	0.6482	-0.0161	0.0336	0.6457	0.0036	-0.0083	0.6663	0.0015	-0.0092	0.689	0.0140	-0.0181	0.6797	-0.0281	0.0278	0.6295	1.5696	-0.5982	0.000589
			remove all	-0.0163	0.0244	0.6477	-0.0182	0.0331	0.6464	0.0066	-0.0073	0.667	-0.0016	-0.0086	0.6894	0.0145	-0.0209	0.6777	-0.0318	0.0233	0.6266	1.4759	-0.5718	0.000627

Figure A.3: Performance of 10-fold cross validation while choosing 'race' as the protected feature (LOS dataset)

A.3 Performance sheet for test dataset

A.3.1 Performance while choosing 'age' as the protected feature

			Features	entire dataset			age									Variance		
							xx - 40			41-70			71-xx					
size of the group: first number: target>=4 second number: target<4 third number: sum		test dataset before sampling(train)	with/without the feature 'age'	2308,7018			213,860,			675,2230,			1420,3928,			variance of the three age groups		
		after sampling(train)		9234,28070			852,3438,			2701,8918,			5681,15714,					
				10447,34550			2979,12020,			3486,11513,			3982,11017,					
				44997			14999			14999			14999					
Change			-	sample	no 'age'	-	sample	no 'age'	-	sample	no 'age'	-	sample	no 'age'	-	sample	no 'age'	-
AUC	after normalization (XGBoost)	before sampling	with 'age'	-	-	0.7137	-	-	0.7217	-	-	0.7267	-	-	0.6825	-	-	0.000586
			without 'age'	-	-0.0014	0.7127	-	-0.0112	0.7136	-	-0.0017	0.7255	-	0.0012	0.6833	-	-0.1920	0.000473
			remove all	-	-0.0013	0.7128	-	-0.0144	0.7113	-	-0.0011	0.7259	-	0.0016	0.6836	-	-0.2122	0.000462
		after sampling	with 'age'	-0.0241	-	0.6965	-0.0441	-	0.6899	-0.0259	-	0.7079	-0.0179	-	0.6703	-0.3963	-	0.000354
			without 'age'	-0.0223	0.0004	0.6968	-0.0346	-0.0014	0.6889	-0.0225	0.0018	0.7092	-0.0195	-0.0004	0.67	-0.1882	0.0865	0.000364
			remove all	-0.0255	-0.0027	0.6946	-0.0373	-0.0074	0.6848	-0.0242	0.0006	0.7083	-0.0238	-0.0045	0.6673	-0.0830	0.1968	0.000423
Precision	after normalization (XGBoost)	before sampling	with 'age'	-	-	0.7375	-	-	0.7825	-	-	0.7558	-	-	0.6973	-	-	0.001899
			without 'age'	-	-0.0008	0.7369	-	0.0006	0.783	-	-0.0033	0.7533	-	0.0026	0.6991	-	-0.0470	0.00181
			remove all	-	-0.0069	0.7324	-	-0.0022	0.7808	-	-0.0087	0.7492	-	-0.0054	0.6935	-	0.0288	0.001954
		after sampling	with 'age'	-0.0190	-	0.7235	0.0098	-	0.7902	-0.0239	-	0.7377	-0.0205	-	0.683	0.5131	-	0.002873
			without 'age'	-0.0145	0.0037	0.7262	0.0075	-0.0016	0.7889	-0.0081	0.0129	0.7472	-0.0266	-0.0037	0.6805	0.6520	0.0405	0.00299
			remove all	-0.0210	-0.0090	0.717	-0.0328	-0.0443	0.7552	-0.0162	-0.0008	0.7371	-0.0248	-0.0098	0.6763	-0.1257	-0.4055	0.001708
Recall	after normalization (XGBoost)	before sampling	with 'age'	-	-	0.7666	-	-	0.8127	-	-	0.7769	-	-	0.7361	-	-	0.001469
			without 'age'	-	-0.0008	0.766	-	/	0.8127	-	-0.0018	0.7755	-	0.0008	0.7367	-	-0.0169	0.001444
			remove all	-	-0.0029	0.7644	-	-0.0012	0.8117	-	-0.0036	0.7741	-	-0.0022	0.7345	-	0.0145	0.00149
		after sampling	with 'age'	-0.0076	-	0.7608	0.0011	-	0.8135	-0.0097	-	0.7694	-0.0075	-	0.7306	0.1741	-	0.001725
			without 'age'	-0.0056	0.0012	0.7617	0.0011	/	0.8135	-0.0044	0.0035	0.7721	-0.0099	-0.0016	0.7294	0.2273	0.0277	0.001773
			remove all	-0.0077	-0.0030	0.7585	-0.0115	-0.0138	0.8024	-0.0061	/	0.7694	-0.0090	-0.0037	0.7279	-0.0649	-0.1920	0.001394
f1 score	after normalization (XGBoost)	before sampling	with 'age'	-	-	0.7069	-	-	0.7617	-	-	0.719	-	-	0.6714	-	-	0.002041
			without 'age'	-	-0.0030	0.7048	-	-0.0014	0.7606	-	-0.0042	0.716	-	-0.0015	0.6704	-	-0.0031	0.002034
			remove all	-	-0.0052	0.7032	-	-0.0038	0.7588	-	-0.0063	0.7145	-	-0.0042	0.6686	-	-0.0031	0.002034
		after sampling	with 'age'	-0.0153	-	0.6961	-0.0067	-	0.7566	-0.0168	-	0.7069	-0.0161	-	0.6606	0.1296	-	0.002305
			without 'age'	-0.0108	0.0016	0.6972	-0.0037	0.0016	0.7578	-0.0116	0.0011	0.7077	-0.0121	0.0026	0.6623	0.1218	-0.0100	0.002282
			remove all	-0.0114	-0.0013	0.6952	-0.0142	-0.0114	0.748	-0.0090	0.0017	0.7081	-0.0141	-0.0021	0.6592	-0.0276	-0.1418	0.001978
AUC	after normalization (Random Forest)	before sampling	with 'age'	-	-	0.6871	-	-	0.6909	-	-	0.7064	-	-	0.6433	-	-	0.001081
			without 'age'	-	-0.0033	0.6848	-	-0.0103	0.6838	-	-0.0011	0.7056	-	-0.0014	0.6424	-	-0.0469	0.001031
			remove all	-	-0.0055	0.6833	-	0.0041	0.6937	-	-0.0061	0.7021	-	-0.0042	0.6406	-	0.0285	0.001112
		after sampling	with 'age'	-0.0233	-	0.6711	-0.0061	-	0.6867	-0.0262	-	0.6879	-0.0270	-	0.6259	0.1625	-	0.001257
			without 'age'	-0.0237	-0.0037	0.6686	-0.0015	-0.0057	0.6828	-0.0283	-0.0033	0.6856	-0.0271	-0.0014	0.625	0.1355	-0.0691	0.00117
			remove all	-0.0190	-0.0012	0.6703	-0.0007	0.0095	0.6932	-0.0236	-0.0035	0.6855	-0.0211	0.0019	0.6271	0.1748	0.0394	0.001307
Precision	after normalization (Random Forest)	before sampling	with 'age'	-	-	0.7358	-	-	0.7895	-	-	0.7488	-	-	0.6956	-	-	0.002222
			without 'age'	-	0.0029	0.7379	-	0.0128	0.7997	-	0.0016	0.75	-	-0.0023	0.694	-	0.2586	0.002796
			remove all	-	-0.0019	0.7344	-	0.0149	0.8014	-	-0.0123	0.7396	-	0.0085	0.7015	-	0.1440	0.002542
		after sampling	with 'age'	0.0004	-	0.7361	-0.0261	-	0.769	0.0017	-	0.7501	0.0083	-	0.7014	-0.4525	-	0.001216
			without 'age'	-0.0103	-0.0079	0.7303	-0.0440	-0.0059	0.7645	0.0088	0.0087	0.7566	-0.0272	-0.0375	0.6751	-0.1241	1.0137	0.00245
			remove all	-0.0264	-0.0287	0.715	-0.0578	-0.0181	0.7551	-0.0050	-0.0189	0.7359	-0.0489	-0.0488	0.6672	-0.1597	0.7558	0.002136
Recall	after normalization (Random Forest)	before sampling	with 'age'	-	-	0.7578	-	-	0.8108	-	-	0.7642	-	-	0.7309	-	-	0.001611
			without 'age'	-	-0.0001	0.7577	-	/	0.8108	-	/	0.7642	-	-0.0004	0.7306	-	0.0071	0.001622
			remove all	-	0.0004	0.7581	-	0.0023	0.8127	-	-0.0008	0.7636	-	0.0016	0.7321	-	0.0243	0.00165
		after sampling	with 'age'	-0.0040	-	0.7548	-0.0091	-	0.8034	-0.0039	-	0.7612	-0.0021	-	0.7294	-0.1445	-	0.001378
			without 'age'	-0.0044	-0.0005	0.7544	-0.0091	/	0.8034	-0.0039	/	0.7612	-0.0029	-0.0012	0.7285	-0.1308	0.0232	0.00141
			remove all	-0.0057	-0.0013	0.7538	-0.0127	-0.0012	0.8024	-0.0043	-0.0012	0.7603	-0.0053	-0.0016	0.7282	-0.1607	0.0049	0.001385
f1 score	after normalization (Random Forest)	before sampling	with 'age'	-	-	0.6675	-	-	0.7461	-	-	0.6757	-	-	0.63	-	-	0.003421
			without 'age'	-	-0.0016	0.6664	-	-0.0055	0.742	-	-0.0006	0.6753	-	-0.0019	0.6288	-	-0.0535	0.003238
			remove all	-	0.0028	0.6694	-	0.0016	0.7473	-	0.0009	0.6763	-	0.0065	0.6341	-	-0.0432	0.003273
		after sampling	with 'age'	-0.0184	-	0.6552	-0.0315	-	0.7226	-0.0166	-	0.6645	-0.0162	-	0.6198	-0.2233	-	0.002657
			without 'age'	-0.0176	-0.0008	0.6547	-0.0240	0.0022	0.7242	-0.0170	-0.0011	0.6638	-0.0159	-0.0016	0.6188	-0.1361	0.0528	0.002797
			remove all	-0.0226	-0.0014	0.6543	-0.0337	-0.0007	0.7221	-0.0191	-0.0017	0.6634	-0.0235	-0.0010	0.6192	-0.1858	0.0029	0.002665
AUC	after normalization (Logistic Regression)	before sampling	with 'age'	-	-	0.651	-	-	0.6515	-	-	0.6641	-	-	0.6192	-	-	0.000536
			without 'age'	-	-0.0015	0.65	-	-0.0101	0.6449	-	-0.0003	0.6639	-	0.0002	0.6193	-	-0.0660	0.000501
			remove all	-	-0.0031	0.649	-	-0.0163	0.6409	-	-0.0023	0.6626	-	-0.0006	0.6188	-	-0.1057	0.00048
		after sampling	with 'age'	-0.0296	-	0.6317	-0.0232	-	0.6364	-0.0315	-	0.6432	-0.0318	-	0.5995	0.0310	-	0.000553
			without 'age'	-0.0297	-0.0016	0.6307	-0.0274	-0.0145	0.6272	-0.0298	0.0014	0.6441	-0.0313	0.0007	0.5999	-0.0070	-0.1004	0.000497
			remove all	-0.0288	-0.0022	0.6303	-0.0251	-0.0182	0.6248	-0.0288	0.0005	0.6435	-0.0310	0.0002	0.5996	0.0119	-0.1223	0.000485
Precision	after normalization (Random Forest)	before sampling	with 'age'	-	-	0.7066	-	-	0.7495	-	-	0.7156	-	-	0.6765	-	-	0.001335
			without 'age'	-	-0.0016	0.7055	-	-0.0035	0.7469	-	-0.0007	0.7151	-	-0.0037	0.674	-	0.0010	0.001336
			remove all	-	-0.0014	0.7056	-	-0.0147	0.7385	-	0.0013	0.7165	-	-0.0022	0.675	-	-0.2208	0.00104
		after sampling	with 'age'	-0.0116	-	0.6984	-0.0036	-	0.7468	-0.0099	-	0.7085	-0.0201	-	0.6629	0.3220	-	0.001764
			without 'age'	-0.0130	-0.0030	0.6963	-0.0064	-0.0063	0.7421	-0.0117	-0.0025	0.7067	-0.0191	-0.0027	0.6611	0.2344	-0.0654	0.001649
			remove all	-0.0218	-0.0117	0.6902	-0.0088	-0.0198	0.732	-0.0227	-0.0117	0.7002	-0.0256	-0.0078	0.6577	0.3365	-0.2123	0.00139
Recall	after normalization (Random Forest)	before sampling	with 'age'	-	-	0.6275	-	-	0.7614	-	-	0.6499	-	-	0.5503	-	-	0.011153
			without 'age'	-	-0.0027	0.6258	-	-0.0380	0.7325	-	0.0072	0.6546	-	-0.0044	0.5479	-	-0.2299	0.008588
			remove all	-	-0.0006	0.6271	-	-0.0514	0.7223	-	0.0149	0.6596	-	-0.0055	0.5473	-	-0.2951	0.007861
		after sampling	with 'age'	-0.0194	-	0.6153	-0.0281	-	0.74	-0.0146	-	0.6404	-0.0242	-	0.537	-0.0761	-	0.010304
			without 'age'	-0.0205	-0.0037	0.613	-0.0522	-0.0618	0.6943	-0.0087	0.0133	0.6489	-0.0277	-0.0080	0.5327	-0.1912	-0.3258	0.006946
			remove all	-0.0053	0.0138	0.6238	-0.0206	-0.0441	0.7074	-0.0024	0.0275	0.658	-0.0033	0.0158	0.5455	-0.1242	-0.3318	0.006885
f1 score	after normalization (Random Forest)	before sampling	with 'age'	-	-	0.6512	-	-	0.7549	-	-	0.6713	-	-	0.5738	-	-	0.008215
			without 'age'	-	-0.0023	0.6497	-	-0.0209	0.7391	-	0.0054	0.6749	-	-0.0040	0.5715	-	-0.1296	0.007151
			remove all	-	-0.0008	0.6507	-	-0.0334	0.7297	-	0.0116	0.6791	-	-0.0054	0.5707	-	-0.1968	0.006599
		after sampling	with 'age'	-0.0170	-	0.6401	-0.0154	-	0.7433	-0.0128	-	0.6627	-0.0221	-	0.5611	0.0147	-	0.008336
			without 'age'	-0.0180	-0.0033	0.638	-0.0348	-0.0402	0.7134	-0.0087	0.0095	0.669	-0.0257	-0.0077	0.5568	-0.0890	-0.2186	0.006514
			remove all	-0.0072	0.0092	0.646	-0.0155	-0.0335	0.7184	-0.0072	0.0174	0.6742	-0.0011	0.0160	0.5701	-0.1215	-0.3046	0.005797

Figure A.4: Performance on test dataset while choosing 'age' as the protected feature (LOS dataset)

A.3.2 Performance while choosing 'gender' as the protected feature

size of the group: first number: target>= 4 second number: target< 4 third number: sum		test dataset before sampling(train) / after sampling(train)	Features with/without features related to gender	entire dataset 2309, 7017, 9233, 28071, 9916, 30082, 39998			gender: male 1308, 4028, 5229, 16113, 4900, 15099, 19999			gender: female 1001, 2989, 4004, 11958, 5016, 14983, 19999			Variance: variance of the two gender groups		
Change			-	sample	no 'sex'	-	sample	no 'sex'	-	sample	no 'sex'	-	sample	no 'sex'	-
AUC	after normalization (XGBoost)	before sampling	with 'gender'	-	-	0.7056	-	-	0.7142	-	-	0.6934	-	-	0.000216
			without 'gender'	-	0.0006	0.706	-	0.0010	0.7149	-	/	0.6934	-	0.0684	0.000231
			remove all	-	-0.0010	0.7049	-	-0.0003	0.714	-	-0.0017	0.6922	-	0.0985	0.000238
		after sampling	with 'gender'	-0.0103	-	0.6983	-0.0126	-	0.7052	-0.0069	-	0.6886	-0.3629	-	0.000138
			without 'gender'	-0.0093	0.0016	0.6994	-0.0119	0.0017	0.7064	-0.0058	0.0012	0.6894	-0.3747	0.0486	0.000145
			remove all	-0.0094	/	0.6983	-0.0120	0.0003	0.7054	-0.0055	-0.0003	0.6884	-0.3918	0.0486	0.000145
Precision	after normalization (XGBoost)	before sampling	with 'gender'	-	-	0.7244	-	-	0.7365	-	-	0.7084	-	-	0.000395
			without 'gender'	-	-0.0039	0.7216	-	-0.0087	0.7301	-	0.0027	0.7103	-	-0.5035	0.000196
			remove all	-	-0.0015	0.7233	-	-0.0073	0.7311	-	0.0064	0.7129	-	-0.5805	0.000166
		after sampling	with 'gender'	0.0007	-	0.7249	-0.0103	-	0.7289	0.0162	-	0.7199	-0.8974	-	4.05E-05
			without 'gender'	0.0019	-0.0026	0.723	-0.0052	-0.0036	0.7263	0.0118	-0.0017	0.7187	-0.8526	-0.2864	2.89E-05
			remove all	0.0010	-0.0012	0.724	-0.0107	-0.0077	0.7233	0.0184	0.0065	0.7246	-0.9952	-0.9802	8.00E-07
Recall	after normalization (XGBoost)	before sampling	with 'gender'	-	-	0.7613	-	-	0.7674	-	-	0.7531	-	-	0.000102
			without 'gender'	-	-0.0014	0.7602	-	-0.0031	0.765	-	0.0011	0.7539	-	-0.3973	6.16E-05
			remove all	-	-0.0005	0.7609	-	-0.0026	0.7654	-	0.0024	0.7549	-	-0.4609	5.51E-05
		after sampling	with 'gender'	0.0001	-	0.7614	-0.0042	-	0.7642	0.0060	-	0.7576	-0.7867	-	2.18E-05
			without 'gender'	0.0007	-0.0009	0.7607	-0.0022	-0.0012	0.7633	0.0042	-0.0007	0.7571	-0.6883	-0.1193	1.92E-05
			remove all	0.0001	-0.0005	0.761	-0.0042	-0.0026	0.7622	0.0060	0.0024	0.7594	-0.9292	-0.8211	3.90E-06
f1 score	after normalization (XGBoost)	before sampling	with 'gender'	-	-	0.6996	-	-	0.7069	-	-	0.6899	-	-	0.000145
			without 'gender'	-	-0.0017	0.6984	-	-0.0041	0.704	-	0.0016	0.691	-	-0.4152	8.45E-05
			remove all	-	-0.0010	0.6989	-	-0.0035	0.7044	-	0.0022	0.6914	-	-0.4152	8.45E-05
		after sampling	with 'gender'	-0.0013	-	0.6987	-0.0089	-	0.7006	0.0091	-	0.6962	-0.9329	-	9.70E-06
			without 'gender'	-0.0016	-0.0020	0.6973	-0.0067	-0.0019	0.6993	0.0054	-0.0022	0.6947	-0.8746	0.0928	1.06E-05
			remove all	-0.0023	-0.0020	0.6973	-0.0102	-0.0049	0.6972	0.0087	0.0017	0.6974	-1.0000	-1.0000	0
AUC	after normalization (Random Forest)	before sampling	with 'gender'	-	-	0.6831	-	-	0.6908	-	-	0.6724	-	-	0.000169
			without 'gender'	-	0.0009	0.6837	-	-0.0030	0.6887	-	0.0067	0.6769	-	-0.5889	6.95E-05
			remove all	-	-0.0038	0.6805	-	-0.0032	0.6886	-	-0.0042	0.6698	-	0.0662	0.000181
		after sampling	with 'gender'	-0.0107	-	0.6758	-0.0145	-	0.6808	-0.0052	-	0.6689	-0.5818	-	7.08E-05
			without 'gender'	-0.0079	0.0037	0.6783	-0.0106	0.0009	0.6814	-0.0041	0.0078	0.6741	-0.6178	-0.6243	2.66E-05
			remove all	-0.0041	0.0028	0.6777	-0.0102	0.0012	0.6816	0.0043	0.0054	0.6725	-0.7706	-0.4153	4.14E-05
Precision	after normalization (Random Forest)	before sampling	with 'gender'	-	-	0.7156	-	-	0.7225	-	-	0.7046	-	-	0.00016
			without 'gender'	-	0.0165	0.7274	-	0.0112	0.7306	-	0.0261	0.723	-	-0.8196	2.89E-05
			remove all	-	-0.0074	0.7103	-	-0.0208	0.7075	-	0.0153	0.7154	-	-0.8052	3.12E-05
		after sampling	with 'gender'	0.0024	-	0.7173	-0.0093	-	0.7158	0.0230	-	0.7208	-0.9220	-	1.25E-05
			without 'gender'	-0.0087	0.0053	0.7211	-0.0130	0.0074	0.7211	-0.0008	0.0022	0.7224	-0.9723	-0.9360	8.00E-07
			remove all	0.0028	-0.0070	0.7123	0.0112	-0.0006	0.7154	-0.0105	-0.0179	0.7079	-0.0994	1.2480	2.81E-05
Recall	after normalization (Random Forest)	before sampling	with 'gender'	-	-	0.755	-	-	0.7582	-	-	0.7506	-	-	2.89E-05
			without 'gender'	-	0.0017	0.7563	-	0.0011	0.759	-	0.0027	0.7526	-	-0.2907	2.05E-05
			remove all	-	-0.0008	0.7544	-	-0.0024	0.7564	-	0.0017	0.7519	-	-0.6505	1.01E-05
		after sampling	with 'gender'	-0.0004	-	0.7547	-0.0017	-	0.7569	0.0013	-	0.7516	-0.5156	-	1.40E-05
			without 'gender'	-0.0017	0.0004	0.755	-0.0020	0.0008	0.7575	-0.0013	/	0.7516	-0.1512	0.2429	1.74E-05
			remove all	-0.0001	-0.0005	0.7543	0.0009	0.0003	0.7571	-0.0017	-0.0013	0.7506	1.0891	0.5071	2.11E-05
f1 score	after normalization (Random Forest)	before sampling	with 'gender'	-	-	0.6625	-	-	0.6682	-	-	0.6549	-	-	8.84E-05
			without 'gender'	-	0.0024	0.6641	-	0.0001	0.6683	-	0.0055	0.6585	-	-0.4570	4.80E-05
			remove all	-	-0.0005	0.6622	-	-0.0039	0.6656	-	0.0043	0.6577	-	-0.6471	3.12E-05
		after sampling	with 'gender'	-0.0045	-	0.6595	-0.0070	-	0.6635	-0.0011	-	0.6542	-0.5113	-	4.32E-05
			without 'gender'	-0.0063	0.0006	0.6599	-0.0058	0.0014	0.6644	-0.0071	-0.0006	0.6538	0.1708	0.3009	5.62E-05
			remove all	-0.0035	0.0006	0.6599	-0.0008	0.0024	0.6651	-0.0075	-0.0021	0.6528	1.4231	0.7500	7.56E-05
AUC	after normalization (Logistic Regression)	before sampling	with 'gender'	-	-	0.6506	-	-	0.6553	-	-	0.644	-	-	6.38E-05
			without 'gender'	-	-0.0003	0.6504	-	-0.0002	0.6552	-	0.0005	0.6443	-	-0.0690	5.94E-05
			remove all	-	-0.0018	0.6494	-	-0.0031	0.6533	-	0.0003	0.6442	-	-0.3511	4.14E-05
		after sampling	with 'gender'	-0.0094	-	0.6445	-0.0119	-	0.6475	-0.0057	-	0.6403	-0.5940	-	2.59E-05
			without 'gender'	-0.0091	/	0.6445	-0.0118	/	0.6475	-0.0059	0.0003	0.6405	-0.5875	-0.0541	2.45E-05
			remove all	-0.0097	-0.0022	0.6431	-0.0116	-0.0028	0.6457	-0.0071	-0.0011	0.6396	-0.5507	-0.2819	1.86E-05
Precision	after normalization (Random Forest)	before sampling	with 'gender'	-	-	0.7064	-	-	0.7106	-	-	0.7008	-	-	4.80E-05
			without 'gender'	-	-0.0011	0.7056	-	-0.0018	0.7093	-	/	0.7008	-	-0.2479	3.61E-05
			remove all	-	0.0030	0.7085	-	0.0014	0.7116	-	0.0057	0.7048	-	-0.5187	2.31E-05
		after sampling	with 'gender'	-0.0079	-	0.7008	-0.0077	-	0.7051	-0.0081	-	0.6951	0.0417	-	5.00E-05
			without 'gender'	-0.0041	0.0027	0.7027	-0.0044	0.0016	0.7062	-0.0039	0.0043	0.6981	-0.0914	-0.3440	3.28E-05
			remove all	-0.0066	0.0043	0.7038	-0.0083	0.0009	0.7057	-0.0047	0.0092	0.7015	-0.6190	-0.8240	8.80E-06
Recall	after normalization (Random Forest)	before sampling	with 'gender'	-	-	0.6477	-	-	0.6563	-	-	0.6361	-	-	0.000204
			without 'gender'	-	-0.0009	0.6471	-	0.0098	0.6627	-	-0.0154	0.6263	-	2.2475	0.000663
			remove all	-	0.0039	0.6502	-	0.0148	0.666	-	-0.0110	0.6291	-	2.3373	0.000681
		after sampling	with 'gender'	-0.0245	-	0.6318	-0.0280	-	0.6379	-0.0197	-	0.6236	-0.4990	-	0.000102
			without 'gender'	-0.0173	0.0065	0.6359	-0.0190	0.0191	0.6501	-0.0152	-0.0109	0.6168	-0.1632	4.4247	0.000554
			remove all	-0.0208	0.0078	0.6367	-0.0219	0.0212	0.6514	-0.0192	-0.0106	0.617	-0.1309	4.7896	0.000592
f1 score	after normalization (Random Forest)	before sampling	with 'gender'	-	-	0.6674	-	-	0.6751	-	-	0.6571	-	-	0.000162
			without 'gender'	-	-0.0007	0.6669	-	0.0067	0.6796	-	-0.0122	0.6491	-	1.8710	0.000465
			remove all	-	0.0036	0.6698	-	0.0111	0.6826	-	-0.0081	0.6518	-	1.9278	0.000474
		after sampling	with 'gender'	-0.0201	-	0.654	-0.0227	-	0.6598	-0.0167	-	0.6461	-0.4210	-	9.38E-05
			without 'gender'	-0.0141	0.0054	0.6575	-0.0149	0.0147	0.6695	-0.0128	-0.0082	0.6408	-0.1146	3.3902	0.000412
			remove all	-0.0170	0.0067	0.6584	-0.0179	0.0161	0.6704	-0.0160	-0.0073	0.6414	-0.1134	3.4829	0.000421

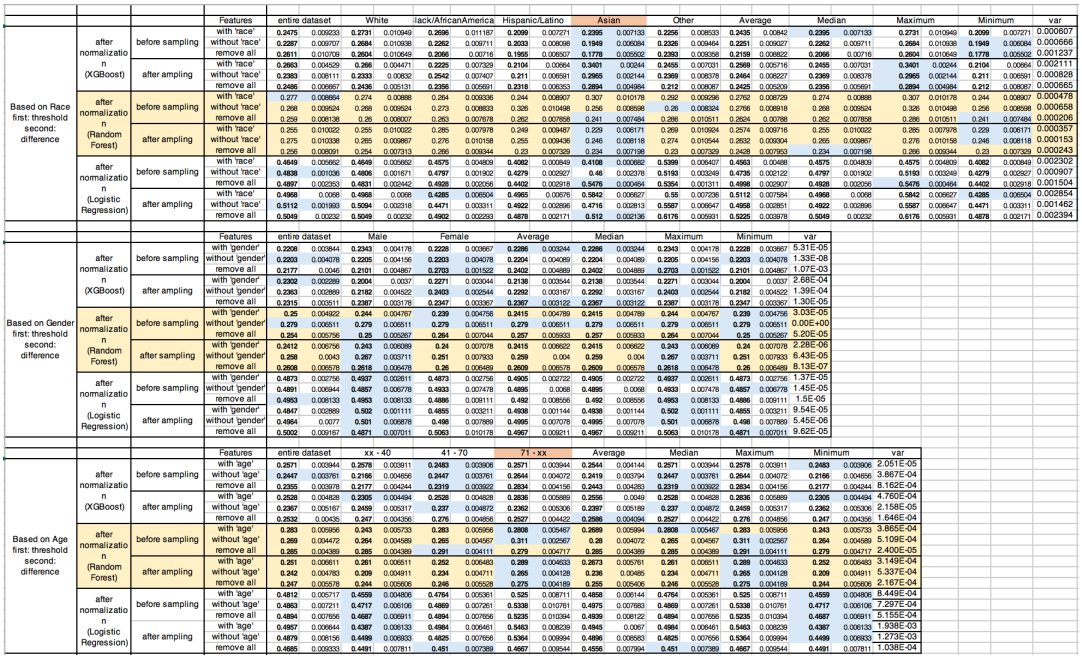
Figure A.5: Performance on test dataset while choosing 'gender' as the protected feature (LOS dataset)

A.3.3 Performance while choosing 'race' as the protected feature

size of the group: first number: target>=4 second number: target<4 third number: sum		test dataset	Features	entire dataset			White			Black/AfricanAmerican			race Hispanic/Latino			Asian			Other			Variance		
		before sampling(train)	with/without features related to race	2308,7018,			1638,4971			193,659,			57,203,			30,110,			390,1075,			variance of the five race groups		
		after sampling(train)		9234,28070,			6551,19882,			771,2636,			230,811,			121,439,			1561,4302,					
				23544,76451,			4956,15043,			4525,15474,			4418,15581,			4321,15678,			5324,14675,					
				99995			19999			19999			19999			19999			19999					
Change			-	sample	no 'race'	-	sample	no 'race'	-	sample	no 'race'	-	sample	no 'race'	-	sample	no 'race'	-	sample	no 'race'	-	sample	no 'race'	-
AUC	after normalization (XGBoost)	before sampling	with 'race'	-	-	0.711	-	-	0.7051	-	-	0.739	-	-	0.7198	-	-	0.6572	-	-	0.7223	-	-	0.000973
			without 'race'	-	-0.0003	0.7108	-	-0.0004	0.7048	-	-0.0011	0.7382	-	0.0013	0.7207	-	-0.0002	0.6571	-	0.0007	0.7228	-	-0.0002	0.000973
			remove all	-	0.0003	0.7112	-	-0.0001	0.705	-	-0.0060	0.7346	-	-0.0007	0.7193	-	0.0070	0.6618	-	0.0053	0.7261	-	-0.1527	0.000824
		after sampling	with 'race'	-0.0211	-	0.696	-0.0190	-	0.6917	-0.0150	-	0.7279	-0.0303	-	0.698	-0.0128	-	0.6488	-0.0350	-	0.697	-0.1741	-	0.000804
			without 'race'	-0.0276	-0.0069	0.6912	-0.0282	-0.0098	0.6849	-0.0187	-0.0048	0.7244	-0.0196	0.0123	0.7066	-0.0035	0.0092	0.6548	-0.0374	-0.0017	0.6958	-0.3052	-0.1589	0.000676
			remove all	-0.0308	-0.0096	0.6893	-0.0315	-0.0129	0.6828	-0.0215	-0.0125	0.7188	-0.0200	0.0099	0.7049	-0.0260	-0.0065	0.6446	-0.0388	0.0013	0.6979	-0.0214	0.0040	0.000807
Precision	after normalization (XGBoost)	before sampling	with 'race'	-	-	0.7102	-	-	0.7079	-	-	0.7708	-	-	0.7683	-	-	0.7305	-	-	0.6804	-	-	0.001517
			without 'race'	-	0.0031	0.7124	-	0.0028	0.7099	-	-0.0141	0.7599	-	0.0128	0.7781	-	/	0.7305	-	0.0096	0.6869	-	-0.1063	0.001356
			remove all	-	-0.0004	0.7099	-	0.0006	0.7083	-	-0.0304	0.7474	-	/	0.7683	-	/	0.7305	-	0.0079	0.6858	-	-0.3119	0.001044
		after sampling	with 'race'	0.0030	-	0.7123	0.0034	-	0.7103	-0.0243	-	0.7521	0.0057	-	0.7727	-0.0497	-	0.6942	0.0262	-	0.6982	-0.1939	-	0.001223
			without 'race'	0.0008	0.0010	0.713	0.0061	0.0055	0.7142	-0.0366	-0.0266	0.7321	0.0321	0.0393	0.8031	-0.0402	0.0099	0.7011	-0.0009	-0.0170	0.6863	0.5320	0.6983	0.002077
			remove all	0.0041	0.0007	0.7128	0.0079	0.0051	0.7139	-0.0120	-0.0182	0.7384	0.0112	0.0054	0.7769	-0.0402	0.0099	0.7011	0.0029	-0.0149	0.6878	0.1835	0.0103	0.001236
Recall	after normalization (XGBoost)	before sampling	with 'race'	-	-	0.7558	-	-	0.7548	-	-	0.7887	-	-	0.7923	-	-	0.7854	-	-	0.7289	-	-	0.000751
			without 'race'	-	0.0012	0.7567	-	0.0011	0.7556	-	-0.0044	0.7852	-	0.0049	0.7962	-	/	0.7854	-	0.0049	0.7325	-	-0.0801	0.000691
			remove all	-	-0.0001	0.7557	-	0.0003	0.755	-	-0.0089	0.7817	-	/	0.7923	-	/	0.7854	-	0.0040	0.7318	-	-0.1503	0.000638
		after sampling	with 'race'	0.0011	-	0.7566	0.0013	-	0.7558	-0.0060	-	0.784	0.0049	-	0.7962	-0.0465	-	0.7489	0.0119	-	0.7376	-0.1919	-	0.000607
			without 'race'	/	0.0001	0.7567	0.0020	0.0017	0.7571	-0.0104	-0.0089	0.777	0.0095	0.0095	0.8038	-0.0174	0.0304	0.7717	-0.0019	-0.0088	0.7311	0.0337	0.1766	0.000714
			remove all	0.0012	/	0.7566	0.0026	0.0016	0.757	-0.0031	-0.0060	0.7793	/	-0.0049	0.7923	-0.0174	0.0304	0.7717	/	-0.0079	0.7318	-0.1575	-0.1142	0.000538
f1 score	after normalization (XGBoost)	before sampling	with 'race'	-	-	0.6927	-	-	0.6906	-	-	0.7254	-	-	0.7278	-	-	0.7138	-	-	0.6721	-	-	0.000575
			without 'race'	-	0.0014	0.6937	-	0.0020	0.692	-	-0.0073	0.7201	-	0.0102	0.7352	-	/	0.7138	-	0.0027	0.6739	-	0.0151	0.000583
			remove all	-	-0.0001	0.6926	-	0.0009	0.6912	-	-0.0127	0.7162	-	/	0.7278	-	/	0.7138	-	0.0030	0.6741	-	-0.1878	0.000467
		after sampling	with 'race'	0.0043	-	0.6957	0.0022	-	0.6921	-0.0023	-	0.7237	0.0166	-	0.7399	-0.0027	-	0.7119	0.0177	-	0.684	-0.0947	-	0.00052
			without 'race'	0.0072	0.0043	0.6987	0.0088	0.0087	0.6981	-0.0036	-0.0086	0.7175	0.0136	0.0072	0.7452	-0.0031	-0.0004	0.7116	0.0071	-0.0077	0.6787	0.0365	0.1622	0.000605
			remove all	0.0087	0.0042	0.6986	0.0103	0.0090	0.6983	0.0040	-0.0064	0.7191	-0.0071	-0.0234	0.7226	-0.0031	-0.0004	0.7116	0.0086	-0.0060	0.6799	-0.3475	-0.4146	0.000305
AUC	after normalization (Random Forest)	before sampling	with 'race'	-	-	0.6848	-	-	0.682	-	-	0.6881	-	-	0.6667	-	-	0.6272	-	-	0.7013	-	-	0.000812
			without 'race'	-	0.0001	0.6849	-	0.0007	0.6825	-	-0.0173	0.6762	-	0.0081	0.6721	-	-0.0169	0.6166	-	0.0054	0.7051	-	0.3186	0.00107
			remove all	-	-0.0003	0.6846	-	-0.0013	0.6811	-	0.0013	0.689	-	0.0069	0.6713	-	0.0185	0.6388	-	-0.0004	0.701	-	-0.3146	0.000556
		after sampling	with 'race'	-0.0200	-	0.6711	-0.0195	-	0.6687	-0.0397	-	0.6608	-0.0493	-	0.6338	-0.0043	-	0.6245	-0.0088	-	0.6951	-0.0156	-	0.000799
			without 'race'	-0.0244	-0.0043	0.6682	-0.0258	-0.0057	0.6649	-0.0209	0.0020	0.6621	-0.0342	0.0241	0.6491	-0.0302	-0.0424	0.598	-0.0156	-0.0014	0.6941	0.1565	0.5491	0.001238
			remove all	-0.0222	-0.0025	0.6694	-0.0217	-0.0036	0.6663	-0.0389	0.0021	0.6622	-0.0349	0.0222	0.6479	-0.0279	-0.0056	0.621	-0.0111	-0.0027	0.6932	0.2565	-0.1251	0.000699
Precision	after normalization (Random Forest)	before sampling	with 'race'	-	-	0.7242	-	-	0.7264	-	-	0.7608	-	-	0.8336	-	-	0.8371	-	-	0.6766	-	-	0.004793
			without 'race'	-	-0.0057	0.7201	-	-0.0058	0.7222	-	-0.0055	0.7566	-	-0.0029	0.8312	-	/	0.8371	-	-0.0064	0.6725	-	0.0449	0.005008
			remove all	-	0.0003	0.7244	-	-0.0061	0.722	-	-0.0038	0.7579	-	-0.0029	0.8312	-	/	0.8371	-	0.0296	0.6968	-	-0.1628	0.004013
		after sampling	with 'race'	-0.0062	-	0.7197	-0.0014	-	0.7254	-0.0825	-	0.698	/	-	0.8336	-0.0864	-	0.7648	0.0324	-	0.6987	-0.3233	-	0.003244
			without 'race'	-0.0046	-0.0040	0.7168	-0.0026	-0.0070	0.7203	-0.0444	0.0358	0.723	/	-0.0029	0.8312	-0.0864	/	0.7648	0.0256	-0.0129	0.6897	-0.4024	-0.0773	0.002993
			remove all	-0.0039	0.0026	0.7216	0.0094	0.0047	0.7288	-0.0649	0.0153	0.7087	/	-0.0029	0.8312	-0.0864	/	0.7648	0.0067	0.0040	0.7015	-0.2976	-0.1310	0.002819
Recall	after normalization (Random Forest)	before sampling	with 'race'	-	-	0.7567	-	-	0.7565	-	-	0.7805	-	-	0.7885	-	-	0.7945	-	-	0.7311	-	-	0.000687
			without 'race'	-	-0.0008	0.7561	-	-0.0008	0.7559	-	/	0.7805	-	-0.0049	0.7846	-	/	0.7945	-	-0.0010	0.7304	-	-0.0221	0.000672
			remove all	-	0.0008	0.7573	-	0.0003	0.7567	-	0.0015	0.7817	-	-0.0049	0.7846	-	/	0.7945	-	0.0049	0.7347	-	-0.1367	0.000593
		after sampling	with 'race'	-0.0024	-	0.7549	-0.0022	-	0.7548	-0.0120	-	0.7711	/	-	0.7885	-0.0057	-	0.79	0.0029	-	0.7332	-0.1638	-	0.000575
			without 'race'	-0.0024	-0.0008	0.7543	-0.0024	-0.0009	0.7541	-0.0076	0.0045	0.7746	/	-0.0049	0.7846	-0.0057	/	0.79	0.0019	-0.0019	0.7318	-0.1433	0.0017	0.000576
			remove all	-0.0030	0.0001	0.755	-0.0022	0.0003	0.755	-0.0120	0.0016	0.7723	/	-0.0049	0.7846	-0.0057	/	0.79	-0.0020	/	0.7332	-0.0924	-0.0630	0.000538
f1 score	after normalization (Random Forest)	before sampling	with 'race'	-	-	0.6578	-	-	0.6667	-	-	0.7006	-	-	0.7025	-	-	0.7121	-	-	0.64	-	-	0.000911
			without 'race'	-	-0.0018	0.6666	-	-0.0016	0.6656	-	0.0026	0.7024	-	-0.0125	0.6937	-	/	0.7121	-	-0.0041	0.6374	-	0.0205	0.000929
			remove all	-	0.0051	0.6712	-	0.0051	0.6701	-	0.0086	0.7066	-	-0.0125	0.6937	-	/	0.7121	-	0.0067	0.6443	-	-0.1339	0.000789
		after sampling	with 'race'	-0.0120	-	0.6598	-0.0118	-	0.6588	-0.0210	-	0.6859	/	-	0.7025	-0.0038	-	0.7094	-0.0106	-	0.6332	0.1050	-	0.001006
			without 'race'	-0.0134	-0.0032	0.6577	-0.0140	-0.0038	0.6563	-0.0127	0.0111	0.6935	/	-0.0125	0.6937	-0.0038	/	0.7094	-0.0154	-0.0088	0.6276	0.2015	0.1096	0.001116
			remove all	-0.0171	-0.0002	0.6597	-0.0172	-0.0003	0.6586	-0.0231	0.0064	0.6903	/	-0.0125	0.6937	-0.0038	/	0.7094	-0.0191	-0.0019	0.632	0.2265	-0.0387	0.000967
AUC	after normalization (Logistic Regression)	before sampling	with 'race'	-	-	0.644	-	-	0.637	-	-	0.6525	-	-	0.5988	-	-	0.713	-	-	0.6646	-	-	0.001732
			without 'race'	-	-0.0006	0.6436	-	0.0002	0.6371	-	0.0014	0.6534	-	-0.0012	0.5981	-	0.0013	0.7139	-	-0.0011	0.6639	-	0.0240	0.001774
			remove all	-	-0.0022	0.6426	-	-0.0008	0.6365	-	-0.0012	0.6517	-	0.0008	0.5993	-	-0.0252	0.695	-	0.0002	0.6647	-	-0.2790	0.001249
		after sampling	with 'race'	-0.0154	-	0.6341	-0.0151	-	0.6274	0.0083	-	0.6579	-0.0048	-	0.5959	-0.0654	-	0.6664	-0.0087	-	0.6588	-0.4997	-	0.000866
			without 'race'	-0.0157	-0.0009	0.6335	-0.0187	-0.0035	0.6252	0.0078	0.0009	0.6585	-0.0060	-0.0023	0.5945	-0.0674	-0.0009	0.6658	-0.0119	-0.0043	0.656	-0.4992	0.0252	0.000888
			remove all	-0.0159	-0.0027	0.6324	-0.0192	-0.0049	0.6243	0.0083	-0.0012	0.6571	-0.0058	-0.0002	0.5958	-0.0637	-0.0236	0.6507	-0.0135	-0.0047	0.6557	-0.4395	-0.1923	0.0007
Precision	after normalization (Random Forest)	before sampling	with 'race'	-	-	0.7005	-	-	0.6966	-	-	0.7244	-	-	0.6948	-	-	0.7579	-	-	0.6945	-	-	0.000771
			without 'race'	-	/	0.7005	-	0.0003	0.6968	-	-0.0072	0.7192	-	0.0134	0.7041	-	-0.0033	0.7554	-	0.0035	0.697	-	-0.2140	0.000606
			remove all	-	0.0010	0.7012	-	0.0011	0.6974	-	-0.0084	0.7183	-	0.0101	0.7018	-	-0.0127	0.7483	-	0.0099	0.7015	-	-0.4245	0.000443
		after sampling	with 'race'	0.0006	-	0.7009	0.0011	-	0.6974	-0.0043	-	0.7213	0.0246	-	0.7119	-0.0110	-	0.7496	0.0081	-	0.7002	-0.4259	-	0.000442
			without 'race'	0.0021	0.0016	0.702	0.0027	0.0019	0.6987	0.0099	0.0059	0.7263	0.0229	0.0117	0.7202	0.0042	0.0120	0.7586	-0.0098	-0.0143	0.6902	0.1823	0.6187	0.000716
			remove all	/	0.0004	0.7012	-0.0003	-0.0003	0.6972	0.0187	0.0125	0.7303	0.0123	-0.0021	0.7104	-0.0029	-0.0047	0.7461	-0.0107	-0.0089	0.694	0.1152	0.1180	0.000495
Recall	after normalization (Random Forest)	before sampling	with 'race'	-	-	0.6374	-	-	0.633	-	-	0.6854	-	-	0.6731	-	-	0.7397	-	-	0.6056	-	-	0.002647
			without 'race'	-	-0.0020	0.6361	-	0.0041	0.6356	-	-0.0067	0.6397	-	-0.0343	0.65	-	-0.0616	0.6941	-	0.0310	0.6244	-	-0.7255	0.000726
			remove all	-	0.0005	0.6377	-	0.0087	0.6385	-	-0.0702	0.6373	-	-0.0287	0.6538	-	-0.0864	0.6758	-	0.0322	0.6251	-	-0.8566	0.000379
		after sampling	with 'race'	-0.0693	-	0.5932	-0.0610	-	0.5944	-0.0547	-	0.6479	-0.0343	-	0.65	-0.4198	-	0.4292	-0.0606	-	0.5689	2.0759	-	0.008141
			without 'race'	-0.0887	-0.0228	0.5797	-0.0884	-0.0252	0.5794	-0.0441	-0.0562	0.6115	-0.0474	-0.0474	0.6192	-0.1712	0.3404	0.5753	-0.1121	-0.0255	0.5544	-0.0055	-0.9112	0.000722
			remove all	-0.0925	-0.0244	0.5787	-0.0937	-0.0264	0.5787	-0.0405	-0.0562	0.6115	-0.0587	-0.0532	0.6154	-0.1756	0.2980	0.5571	-0.1118	-0.0241	0.5552	1.1892	-0.8980	0.000831
f1 score	after normalization (Random Forest)	before sampling	with 'race'	-	-	0.6584	-	-	0.6543	-	-	0.7008	-	-	0.6829	-	-	0.7476	-	-	0.6277	-	-	0.002093
			without 'race'	-	-0.0015	0.6574	-	0.0031	0.6563	-	-0.0502	0.6656	-	-0.0174	0.671	-	-0.0433	0.7152	-	0.0268	0.6445	-	-0.6536	0.000725
			remove all	-	0.0005	0.6587	-	0.0067	0.6587	-	-0.0531	0.6636	-	-0.0144	0.6731	-	-0.0638	0.6999	-	0.0284	0.6455	-	-0.8022	0.000414
		after sampling	with 'race'	-0.0570	-	0.6209	-0.0495	-	0.6219	-0.0405	-	0.6724	-0.0146	-	0.6729	-0.4010	-	0.4478	-0.0575	-	0.5916	3.0929	-	0.008566
			without 'race'	-0.0745	-0.0201	0.6084	-0.0734	-0.0222	0.6081	-0.0344	-0.0442	0.6427	-0.0308	-0.0336	0.6503	-0.1444	0.3665	0.6119	-0.1040	-0.0238	0.5775	0.1823	-0.8999	0.000857
			remove all	-0.0777	-0.0216	0.6075	-0.0779	-0.0233	0.6074	-0.0312	-0.0439	0.6429	-0.0401	-0.0398	0.6461	-0.1502	0.3283	0.5948	-0.1047	-0.0232	0.5779	1.1628	-0.8955	0.000895

Figure A.6: Performance on test dataset while choosing 'race' as the protected feature (LOS dataset)

A.4 Threshold sheet



			Features	entire dataset		White		Black/AfricanAmerica		Hispanic/Latino		Asian		Other		Average		Median		Maximum		Minimum		var
Based on Race first: threshold second: difference	after normalization (XGBoost)	before sampling	with 'race'	0.2475	0.009233	0.2731	0.010949	0.2695	0.011187	0.2099	0.007271	0.2395	0.007133	0.2256	0.008533	0.2435	0.00842	0.2395	0.007133	0.2731	0.010949	0.2099	0.007271	0.000607
			without 'race'	0.2397	0.009707	0.2694	0.010936	0.2262	0.009711	0.2033	0.008098	0.1949	0.009284	0.2326	0.009464	0.2251	0.009227	0.2262	0.009711	0.2694	0.010936	0.1949	0.009284	0.000666
			remove all	0.2611	0.010709	0.2604	0.010649	0.2066	0.00716	0.1955	0.006507	0.1778	0.005502	0.2393	0.009358	0.2159	0.008822	0.2066	0.00716	0.2604	0.010649	0.1778	0.005502	0.001237
		after sampling	with 'race'	0.2663	0.004529	0.266	0.004471	0.2225	0.007329	0.2104	0.00664	0.3401	0.00244	0.2455	0.007031	0.2569	0.005716	0.2455	0.007031	0.3401	0.00244	0.2104	0.00664	0.002111
			without 'race'	0.2383	0.008111	0.2333	0.00832	0.2542	0.007407	0.211	0.006591	0.2965	0.002144	0.2369	0.008378	0.2464	0.008227	0.2369	0.008378	0.2965	0.002144	0.211	0.006591	0.000828
			remove all	0.2486	0.006957	0.2436	0.005131	0.2356	0.006951	0.2318	0.006353	0.2894	0.004984	0.212	0.008357	0.2425	0.005209	0.2356	0.006951	0.2894	0.004984	0.212	0.006357	0.000665
	after normalization (Random Forest)	before sampling	with 'race'	0.277	0.008954	0.274	0.00888	0.264	0.008336	0.244	0.008607	0.307	0.010178	0.292	0.008296	0.2762	0.008729	0.274	0.00888	0.307	0.010178	0.244	0.008607	0.000478
			without 'race'	0.268	0.009524	0.268	0.009524	0.273	0.008833	0.326	0.010498	0.256	0.008598	0.26	0.008324	0.2766	0.008918	0.268	0.009524	0.326	0.010498	0.256	0.008598	0.000658
			remove all	0.259	0.008136	0.26	0.008007	0.263	0.007678	0.262	0.007858	0.241	0.007484	0.286	0.010511	0.2624	0.00798	0.262	0.007858	0.286	0.010511	0.241	0.007484	0.000206
		after sampling	with 'race'	0.256	0.010022	0.255	0.010022	0.285	0.007978	0.249	0.009487	0.229	0.006171	0.269	0.010924	0.2574	0.009716	0.255	0.010022	0.285	0.007978	0.229	0.006171	0.000357
			without 'race'	0.275	0.010338	0.265	0.009867	0.276	0.010158	0.255	0.009436	0.246	0.008118	0.274	0.010544	0.2632	0.009304	0.265	0.009867	0.276	0.010158	0.246	0.008118	0.000153
			remove all	0.256	0.008091	0.254	0.007313	0.266	0.009344	0.23	0.007329	0.234	0.007198	0.23	0.007329	0.2428	0.007953	0.234	0.007198	0.266	0.009344	0.23	0.007329	0.000243
	after normalization (Logistic Regression)	before sampling	with 'race'	0.4649	0.005662	0.4649	0.005662	0.4575	0.004809	0.4082	0.000849	0.4108	0.000682	0.5399	0.006407	0.4563	0.00488	0.4575	0.004809	0.4575	0.004809	0.4082	0.000849	0.002302
			without 'race'	0.4838	0.001036	0.4806	0.001671	0.4797	0.001902	0.4279	0.002927	0.46	0.002378	0.5193	0.003249	0.4735	0.002122	0.4797	0.001902	0.5193	0.003249	0.4279	0.002927	0.000907
			remove all	0.4897	0.002353	0.4831	0.002442	0.4928	0.002256	0.4402	0.002918	0.5476	0.000484	0.5354	0.001311	0.4998	0.002907	0.4928	0.002256	0.5476	0.000484	0.4402	0.002918	0.001504
		after sampling	with 'race'	0.4968	0.0068	0.4968	0.0068	0.4285	0.006504	0.4965	0.00676	0.5842	0.006627	0.55	0.007236	0.5112	0.007584	0.4968	0.0068	0.5842	0.006627	0.4285	0.006504	0.002854
			without 'race'	0.5112	0.001993	0.5094	0.002318	0.4471	0.003311	0.4922	0.002896	0.4716	0.002813	0.5587	0.006547	0.4958	0.002851	0.4922	0.002896	0.5587	0.006547	0.4471	0.003311	0.001462
			remove all	0.5049	0.00232	0.5049	0.00232	0.4902	0.002253	0.4878	0.002171	0.512	0.002136	0.6176	0.005931	0.5225	0.003978	0.5049	0.00232	0.6176	0.005931	0.4878	0.002171	0.002394

			Features	entire dataset		Male		Female		Average		Median		Maximum		Minimum		var
Based on Gender first: threshold second: difference	after normalization (XGBoost)	before sampling	with 'gender'	0.2208	0.003844	0.2343	0.004178	0.2228	0.003667	0.2286	0.003244	0.2286	0.003244	0.2343	0.004178	0.2228	0.003667	5.31E-05
			without 'gender'	0.2203	0.004078	0.2205	0.004156	0.2203	0.004078	0.2204	0.004089	0.2204	0.004089	0.2205	0.004156	0.2203	0.004078	1.33E-08
			remove all	0.2177	0.0046	0.2101	0.004857	0.2703	0.001522	0.2402	0.004889	0.2402	0.004889	0.2703	0.001522	0.2101	0.004857	1.07E-03
		after sampling	with 'gender'	0.2302	0.002589	0.2004	0.0037	0.2271	0.003044	0.2138	0.003544	0.2138	0.003544	0.2271	0.003044	0.2004	0.0037	2.68E-04
			without 'gender'	0.2363	0.002889	0.2182	0.004522	0.2403	0.002544	0.2292	0.003167	0.2292	0.003167	0.2403	0.002544	0.2182	0.004522	1.39E-04
			remove all	0.2315	0.003511	0.2387	0.003178	0.2347	0.003367	0.2367	0.003122	0.2367	0.003122	0.2387	0.003178	0.2347	0.003367	1.30E-05
	after normalization (Random Forest)	before sampling	with 'gender'	0.25	0.004922	0.244	0.004767	0.239	0.004758	0.2415	0.004789	0.2415	0.004789	0.244	0.004767	0.239	0.004758	3.03E-05
			without 'gender'	0.279	0.006511	0.279	0.006511	0.279	0.006511	0.279	0.006511	0.279	0.006511	0.279	0.006511	0.279	0.006511	0.00E+00
			remove all	0.254	0.005756	0.25	0.005267	0.264	0.007044	0.257	0.005933	0.257	0.005933	0.264	0.007044	0.25	0.005267	5.20E-05
		after sampling	with 'gender'	0.2412	0.006756	0.243	0.006089	0.24	0.007078	0.2415	0.006922	0.2415	0.006922	0.243	0.006089	0.24	0.007078	2.28E-06
			without 'gender'	0.258	0.0043	0.267	0.003711	0.251	0.007933	0.259	0.004	0.259	0.004	0.267	0.003711	0.251	0.007933	6.43E-05
			remove all	0.2608	0.006578	0.2618	0.006478	0.26	0.006489	0.2609	0.006578	0.2609	0.006578	0.2618	0.006478	0.26	0.006489	8.13E-07
	after normalization (Logistic Regression)	before sampling	with 'gender'	0.4873	0.002756	0.4937	0.002811	0.4873	0.002756	0.4905	0.002722	0.4905	0.002722	0.4937	0.002811	0.4873	0.002756	1.37E-05
			without 'gender'	0.4891	0.006944	0.4857	0.006778	0.4933	0.007478	0.4895	0.0068	0.4895	0.0068	0.4933	0.007478	0.4857	0.006778	1.45E-05
			remove all	0.4953	0.008133	0.4953	0.008133	0.4886	0.009111	0.492	0.008556	0.492	0.008556	0.4953	0.008133	0.4886	0.009111	1.5E-05
		after sampling	with 'gender'	0.4947	0.002899	0.502	0.001111	0.4855	0.003211	0.4938	0.001144	0.4938	0.001144	0.502	0.001111	0.4855	0.003211	9.54E-05
			without 'gender'	0.4964	0.0077	0.501	0.008878	0.498	0.007889	0.4995	0.007078	0.4995	0.007078	0.501	0.008878	0.498	0.007889	5.45E-06
			remove all	0.5002	0.009167	0.4871	0.007011	0.5063	0.010178	0.4967	0.009211	0.4967	0.009211	0.5063	0.010178	0.4871	0.007011	9.62E-05

			Features	entire dataset		xx - 40		41 - 70		71 - xx		Average		Median		Maximum		Minimum		var
Based on Age first: threshold second: difference	after normalization (XGBoost)	before sampling	with 'age'	0.2571	0.003944	0.2578	0.003911	0.2483	0.003906	0.2571	0.003944	0.2544	0.004144	0.2571	0.003944	0.2578	0.003911	0.2483	0.003906	2.051E-05
			without 'age'	0.2447	0.003761	0.2166	0.004656	0.2447	0.003761	0.2544	0.004072	0.2419	0.003794	0.2447	0.003761	0.2544	0.004072	0.2166	0.004656	3.867E-04
			remove all	0.2355	0.003978	0.2177	0.004244	0.2319	0.003922	0.2834	0.004156	0.2443	0.004283	0.2319	0.003922	0.2834	0.004156	0.2177	0.004244	8.162E-04
		after sampling	with 'age'	0.2528	0.004828	0.2305	0.004494	0.2528	0.004828	0.2836	0.005889	0.2556	0.0049	0.2528	0.004828	0.2836	0.005889	0.2305	0.004494	4.760E-04
			without 'age'	0.2367	0.005167	0.2459	0.005317	0.237	0.004872	0.2362	0.005306	0.2397	0.005189	0.237	0.004872	0.2459	0.005317	0.2362	0.005306	2.158E-05
			remove all	0.2532	0.00435	0.247	0.004356	0.276	0.004856	0.2527	0.004422	0.2586	0.004094	0.2527	0.004422	0.276	0.004856	0.247	0.004356	1.646E-04
	after normalization (Random Forest)	before sampling	with 'age'	0.283	0.005956	0.243	0.005733	0.283	0.005956	0.2808	0.005467	0.2689	0.005994	0.2808	0.005467	0.283	0.005956	0.243	0.005733	3.865E-04
			without 'age'	0.269	0.004472	0.264	0.004589	0.265	0.004567	0.311	0.002567	0.28	0.004072	0.265	0.004567	0.311	0.002567	0.264	0.004589	5.109E-04
			remove all	0.285	0.004389	0.285	0.004389	0.291	0.004111	0.279	0.004717	0.285	0.004389	0.285	0.004389	0.291	0.004111	0.279	0.004717	2.400E-05
		after sampling	with 'age'	0.261	0.006611	0.261	0.006511	0.262	0.006483	0.269	0.004633	0.2673	0.005761	0.261	0.006511	0.269	0.004633	0.262	0.006483	3.149E-04
			without 'age'	0.242	0.004783	0.209	0.004911	0.234	0.004711	0.265	0.004128	0.236	0.00485	0.234	0.004711	0.265	0.004128	0.209	0.004911	5.337E-04
			remove all	0.247	0.005578	0.244	0.005606	0.246	0.005528	0.275	0.004189	0.255	0.005406	0.246	0.005528	0.275	0.004189	0.244	0.005606	2.167E-04
	after normalization (Logistic Regression)	before sampling	with 'age'	0.4812	0.005717	0.4559	0.004806	0.4764	0.005361	0.525	0.006711	0.4858	0.006144	0.4764	0.005361	0.525	0.006711	0.4559	0.004806	8.449E-04
			without 'age'	0.4863	0.007211	0.4717	0.006106	0.4869	0.007261	0.5338	0.010761	0.4975	0.007683	0.4869	0.007261	0.5338	0.010761	0.4717	0.006106	7.297E-04
			remove all	0.4894	0.007656	0.4687	0.006911	0.4894	0.007656	0.5235	0.010394	0.4939	0.008122	0.4894	0.007656	0.5235	0.010394	0.4687	0.006911	5.155E-04
		after sampling	with 'age'	0.4957	0.006644	0.4387	0.006133	0.4984	0.006461	0.5463	0.008239	0.4945	0.0057	0.4984	0.006461	0.5463	0.008239	0.4387	0.006133	1.938E-03
			without 'age'	0.4879	0.008156	0.4499	0.006833	0.4825	0.007656	0.5364	0.009994	0.4896	0.008583	0.4825	0.007656	0.5364	0.009994	0.4499	0.006833	1.273E-03
			remove all	0.4685	0.009333	0.4491	0.007811	0.451	0.007389	0.4667	0.009544	0.4556	0.007994	0.451	0.007389	0.4667	0.009544	0.4491	0.007811	1.038E-04

Figure A.7: Threshold sheet (LOS dataset)

Appendix B

RESULTS ON THYROID DATASET

B.1 Performance sheet for 10-fold cross validation on train dataset

B.1.1 Performance while choosing 'age' as the protected feature

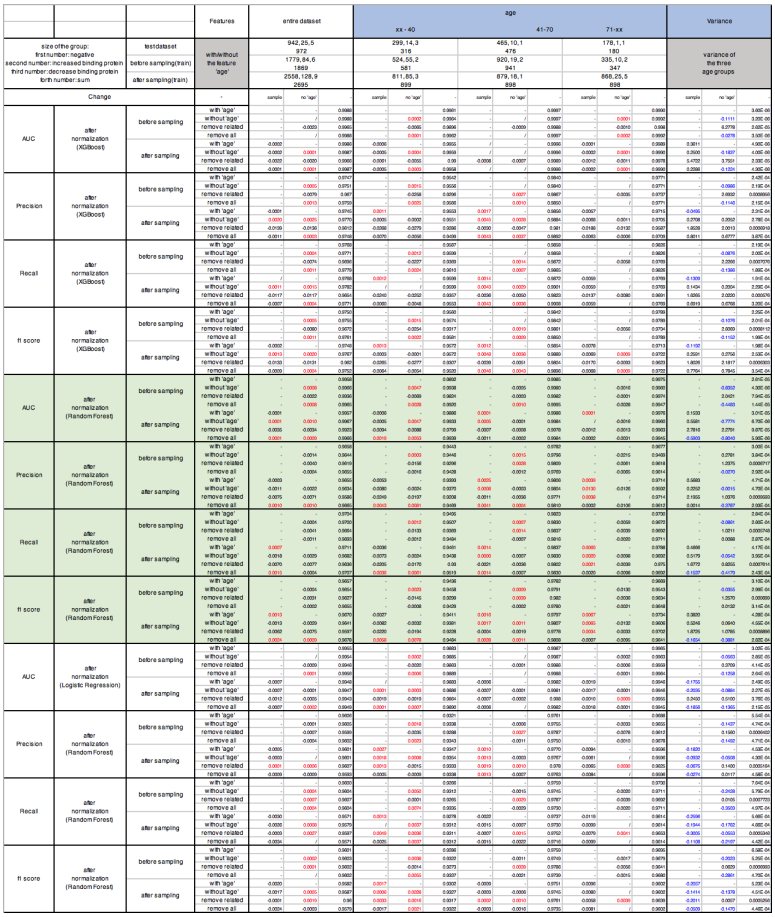


Figure B.1: Performance of 10-fold cross validation while choosing 'age' as the protected feature (Thyroid dataset)

B.1.2 Performance while choosing 'gender' as the protected feature

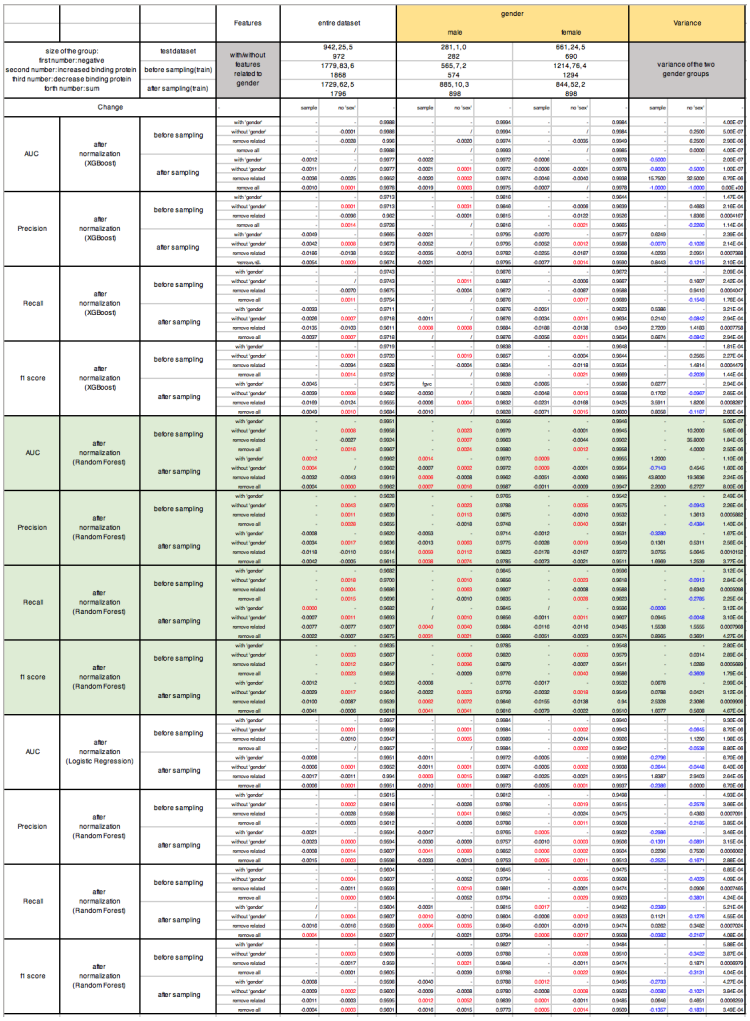


Figure B.2: Performance of 10-fold cross validation while choosing 'gender' as the protected feature (Thyroid dataset)

B.2 Performance sheet for test dataset

B.2.1 Performance while choosing 'age' as the protected feature

			Features	entire dataset			age									Variance		
							xx - 40			41-70			71-xx					
size of the group: first number: negative second number: increased binding protein third number: decrease binding protein forth number: sum		test dataset	with/without the feature 'age'	942,25,5 972			299,14,3 316			465,10,1 476			178,1,1 180			variance of the three age groups		
		before sampling(train)		1779,84,6 1869			524,55,2 581			920,19,2 941			335,10,2 347					
		after sampling(train)		2558,128,9 2695			811,85,3 899			879,18,1 898			868,25,5 898					
Change			-	sample	no 'age'	-	sample	no 'age'	-	sample	no 'age'	-	sample	no 'age'	-	sample	no 'age'	-
AUC	after normalization (XGBoost)	before sampling	with 'age'	-	-	0.998	-	-	0.9934	-	-	0.9989	-	-	0.9999	-	-	1.22E-05
			without 'age'	-	-0.0006	0.9974	-	-0.0024	0.991	-	-0.0001	0.9988	-	/	0.9999	-	0.9180	2.34E-05
			remove all	-	-0.0006	0.9974	-	-0.0022	0.9912	-	/	0.9989	-	-0.0001	0.9998	-	0.8443	2.25E-05
		after sampling	with 'age'	-0.0002	-	0.9978	-0.0003	-	0.9931	/	-	0.9989	-0.0001	-	0.9998	0.0820	-	1.32E-05
			without 'age'	0.0002	-0.0002	0.9976	0.0013	-0.0008	0.9923	/	-0.0001	0.9988	-0.0001	/	0.9998	-0.2821	0.2727	1.68E-05
			remove all	0.0001	-0.0003	0.9975	0.0005	-0.0014	0.9917	/	/	0.9989	/	/	0.9998	-0.1333	0.4773	1.95E-05
Precision	after normalization (XGBoost)	before sampling	with 'age'	-	-	0.9725	-	-	0.9532	-	-	0.9849	-	-	0.9779	-	-	0.0002771
			without 'age'	-	-0.0044	0.9682	-	-0.0066	0.9469	-	-0.0052	0.9798	-	/	0.9779	-	0.2320	0.0003414
			remove all	-	-0.0033	0.9693	-	-0.0038	0.9496	-	-0.0052	0.9798	-	/	0.9779	-	0.0339	0.0002865
		after sampling	with 'age'	-0.0092	-	0.9636	-0.0200	-	0.9341	-0.0075	-	0.9775	/	-	0.9779	1.2920	-	0.0006351
			without 'age'	-0.0029	0.0019	0.9654	-0.0078	0.0058	0.9395	-0.0019	0.0004	0.9779	-0.0001	-0.0001	0.9778	0.4332	-0.2296	0.0004893
			remove all	-0.0013	0.0046	0.968	-0.0071	0.0094	0.9429	0.0024	0.0048	0.9822	/	/	0.9779	0.6258	-0.2666	0.0004658
Recall	after normalization (XGBoost)	before sampling	with 'age'	-	-	0.9722	-	-	0.943	-	-	0.9853	-	-	0.9889	-	-	0.00065
			without 'age'	-	-0.0127	0.9599	-	-0.0268	0.9177	-	-0.0085	0.9769	-	/	0.9889	-	1.2334	0.0014517
			remove all	-	-0.0084	0.964	-	-0.0134	0.9304	-	-0.0085	0.9769	-	/	0.9889	-	0.4694	0.0009551
		after sampling	with 'age'	-0.0032	-	0.9691	-0.0033	-	0.9399	-0.0043	-	0.9811	/	-	0.9889	0.0674	-	0.0006938
			without 'age'	0.0065	-0.0031	0.9661	0.0207	-0.0034	0.9367	0.0021	-0.0021	0.979	-0.0057	-0.0057	0.9833	-0.5431	-0.0440	0.0006633
			remove all	0.0064	0.0011	0.9702	0.0068	-0.0034	0.9367	0.0086	0.0043	0.9853	/	/	0.9889	-0.1108	0.2241	0.0008493
f1 score	after normalization (XGBoost)	before sampling	with 'age'	-	-	0.972	-	-	0.9457	-	-	0.985	-	-	0.9834	-	-	0.0004939
			without 'age'	-	-0.0094	0.9629	-	-0.0187	0.928	-	-0.0070	0.9781	-	/	0.9834	-	0.8931	0.000935
			remove all	-	-0.0064	0.9658	-	-0.0095	0.9367	-	-0.0070	0.9781	-	/	0.9834	-	0.3252	0.0006545
		after sampling	with 'age'	-0.0058	-	0.9664	-0.0093	-	0.9369	-0.0060	-	0.9791	/	-	0.9834	0.3389	-	0.0006613
			without 'age'	0.0026	-0.0010	0.9654	0.0098	0.0002	0.9371	0.0003	-0.0007	0.9784	-0.0028	-0.0028	0.9806	-0.3591	-0.0939	0.0005992
			remove all	0.0032	0.0026	0.9689	0.0016	0.0014	0.9382	0.0056	0.0046	0.9836	/	/	0.9834	0.0445	0.0337	0.0006836
AUC	after normalization (Random Forest)	before sampling	with 'age'	-	-	0.9952	-	-	0.9922	-	-	0.9961	-	-	0.9955	-	-	4.40E-06
			without 'age'	-	0.0013	0.9965	-	-0.0010	0.9912	-	0.0013	0.9974	-	0.0043	0.9998	-	3.4773	1.97E-05
			remove all	-	-0.0014	0.9938	-	-0.0039	0.9883	-	-0.0001	0.996	-	/	0.9955	-	3.1818	1.84E-05
		after sampling	with 'age'	-0.0002	-	0.995	-0.0055	-	0.9867	0.0009	-	0.997	0.0041	-	0.9996	9.5455	-	4.64E-05
			without 'age'	-0.0015	/	0.995	-0.0036	0.0009	0.9876	-0.0005	-0.0001	0.9969	-0.0002	/	0.9996	0.9898	-0.1552	3.92E-05
			remove all	0.0015	0.0003	0.9953	0.0012	0.0028	0.9895	-0.0003	-0.0013	0.9957	0.0043	0.0002	0.9998	0.4511	-0.4246	2.67E-05
Precision	after normalization (Random Forest)	before sampling	with 'age'	-	-	0.964	-	-	0.9459	-	-	0.9773	-	-	0.9779	-	-	0.0003356
			without 'age'	-	-0.0015	0.9626	-	-0.0161	0.9307	-	0.0021	0.9794	-	/	0.9779	-	1.2881	0.0007679
			remove all	-	0.0049	0.9687	-	0.0087	0.9541	-	-0.0026	0.9748	-	/	0.9779	-	-0.5009	0.0001675
		after sampling	with 'age'	-0.0096	-	0.9547	-0.0200	-	0.927	-0.0144	-	0.9632	/	-	0.9779	1.0429	-	0.0006856
			without 'age'	-0.0009	0.0073	0.9617	0.0059	0.0099	0.9362	-0.0072	0.0094	0.9723	-0.0001	-0.0001	0.9778	-0.3334	-0.2534	0.0005119
			remove all	-0.0118	0.0027	0.9573	-0.0305	-0.0022	0.925	-0.0033	0.0087	0.9716	/	/	0.9779	3.9928	0.2198	0.0008363
Recall	after normalization (Random Forest)	before sampling	with 'age'	-	-	0.9722	-	-	0.9525	-	-	0.979	-	-	0.9889	-	-	0.0003534
			without 'age'	-	-0.0052	0.9671	-	-0.0232	0.9304	-	0.0043	0.9832	-	/	0.9889	-	1.9451	0.0010408
			remove all	-	-0.0021	0.9702	/		0.9525	-	-0.0043	0.9748	-	/	0.9889	-	-0.0490	0.0003361
		after sampling	with 'age'	-0.0095	-	0.963	-0.0232	-	0.9304	-0.0043	-	0.9748	/	-	0.9889	1.6384	-	0.0009324
			without 'age'	-0.0054	-0.0011	0.9619	0.0068	0.0068	0.9367	-0.0128	-0.0043	0.9706	-0.0057	-0.0057	0.9833	-0.4422	-0.3773	0.0005806
			remove all	-0.0160	-0.0086	0.9547	-0.0498	-0.0272	0.9051	/	/	0.9748	/	/	0.9889	4.9946	1.1609	0.0020148
f1 score	after normalization (Random Forest)	before sampling	with 'age'	-	-	0.9676	-	-	0.9491	-	-	0.9703	-	-	0.9834	-	-	0.0002992
			without 'age'	-	-0.0029	0.9648	-	-0.0200	0.9301	-	0.0109	0.9809	-	/	0.9834	-	2.0234	0.0009046
			remove all	-	0.0016	0.9691	-	0.0027	0.9517	-	0.0046	0.9748	-	/	0.9834	-	-0.1036	0.0002682
		after sampling	with 'age'	-0.0092	-	0.9587	-0.0217	-	0.9285	-0.0028	-	0.9676	/	-	0.9834	1.6631	-	0.0007968
			without 'age'	-0.0033	0.0030	0.9616	0.0062	0.0080	0.9359	-0.0098	0.0038	0.9713	-0.0028	-0.0028	0.9806	-0.3865	-0.3035	0.000555
			remove all	-0.0138	-0.0031	0.9557	-0.0406	-0.0166	0.9131	-0.0016	0.0058	0.9732	/	/	0.9834	4.3803	0.8110	0.001443
AUC	after normalization (Logistic Regression)	before sampling	with 'age'	-	-	0.9967	-	-	0.9901	-	-	0.9979	-	-	0.9997	-	-	2.60E-05
			without 'age'	-	-0.0014	0.9953	-	-0.0024	0.9877	-	-0.0011	0.9968	-	0.0001	0.9998	-	0.5231	3.96E-05
			remove all	-	-0.0013	0.9954	-	-0.0021	0.988	-	-0.0010	0.9969	-	0.0001	0.9998	-	0.4577	3.79E-05
		after sampling	with 'age'	-0.0008	-	0.9959	-0.0012	-	0.9889	-0.0006	-	0.9973	-0.0002	-	0.9995	0.2154	-	3.16E-05
			without 'age'	-0.0004	-0.0010	0.9949	/	-0.0012	0.9877	-0.0004	-0.0009	0.9964	-0.0002	0.0001	0.9996	-0.0556	0.1835	3.74E-05
			remove all	-0.0004	-0.0009	0.995	-0.0004	-0.0013	0.9876	-0.0003	-0.0007	0.9966	-0.0002	0.0001	0.9995	0.0185	0.2215	3.86E-05
Precision	after normalization (Random Forest)	before sampling	with 'age'	-	-	0.9679	-	-	0.9511	-	-	0.976	-	-	0.9779	-	-	0.0002234
			without 'age'	-	0.0009	0.9688	-	-0.0022	0.949	-	0.0005	0.9765	-	0.0084	0.9861	-	0.6616	0.0003712
			remove all	-	-0.0004	0.9675	-	-0.0022	0.949	-	0.0005	0.9765	-	-0.0001	0.9778	-	0.1871	0.0002652
		after sampling	with 'age'	0.0005	-	0.9684	-0.0012	-	0.95	0.0023	-	0.9782	-0.0001	-	0.9778	0.1714	-	0.0002617
			without 'age'	-0.0013	-0.0009	0.9675	/	-0.0011	0.949	/	-0.0017	0.9765	-0.0084	/	0.9778	-0.2856	0.0134	0.0002652
			remove all	-0.0002	-0.0011	0.9673	-0.0005	-0.0016	0.9485	/	-0.0017	0.9765	/	/	0.9778	0.0336	0.0474	0.0002741
Recall	after normalization (Random Forest)	before sampling	with 'age'	-	-	0.9589	-	-	0.9209	-	-	0.9727	-	-	0.9889	-	-	0.0012616
			without 'age'	-	-0.0075	0.9517	-	-0.0138	0.9082	-	-0.0065	0.9664	-	/	0.9889	-	0.3733	0.0017325
			remove all	-	-0.0087	0.9506	-	-0.0138	0.9082	-	-0.0065	0.9664	-	-0.0057	0.9833	-	0.2299	0.0015516
		after sampling	with 'age'	-0.0032	-	0.9558	-0.0068	-	0.9146	/	-	0.9727	-0.0057	-	0.9833	0.0862	-	0.0013703
			without 'age'	-0.0012	-0.0054	0.9506	/	-0.0070	0.9082	/	-0.0065	0.9664	-0.0057	/	0.9833	-0.1044	0.1323	0.0015516
			remove all	-0.0011	-0.0065	0.9496	-0.0034	-0.0104	0.9051	/	-0.0065	0.9664	/	/	0.9833	0.0929	0.2375	0.0016958
f1 score	after normalization (Random Forest)	before sampling	with 'age'	-	-	0.9622	-	-	0.9311	-	-	0.9742	-	-	0.9834	-	-	0.0007776
			without 'age'	-	-0.0044	0.958	-	-0.0091	0.9226	-	-0.0039	0.9704	-	0.0037	0.987	-	0.4411	0.0011206
			remove all	-	-0.0054	0.957	-	-0.0091	0.9226	-	-0.0039	0.9704	-	-0.0028	0.9806	-	0.2344	0.0009599
		after sampling	with 'age'	-0.0019	-	0.9604	-0.0046	-	0.9268	0.0007	-	0.9749	-0.0028	-	0.9806	0.1224	-	0.0008728
			without 'age'	-0.0010	-0.0035	0.957	/	-0.0045	0.9226	/	-0.0046	0.9704	-0.0065	/	0.9806	-0.1434	0.0998	0.0009599
			remove all	-0.0007	-0.0043	0.9563	-0.0023	-0.0068	0.9205	/	-0.0046	0.9704	/	/	0.9806	0.0788	0.1864	0.0010355

Figure B.3: Performance on test dataset while choosing 'age' as the protected feature (Thyroid dataset)

B.2.2 Performance while choosing 'gender' as the protected feature

size of the group: first number: negative, second number: increased binding protein, third number: decrease binding protein, forth number: sum			Features	entire dataset			gender: male			gender: female			Variance		
		test dataset	with/without features related to gender	942,25,5 972			281,1,0 282			661,24,5 690			variance of the two gender groups		
		before sampling(train)		1779,83,6 1868			565,7,2 574			1214,76,4 1294					
		after sampling(train)		1729,62,5 1796			885,10,3 898			844,52,2 898					
Change			-	sample	no 'sex'	-	sample	no 'sex'	-	sample	no 'sex'	-	sample	no 'sex'	-
AUC	after normalization (XGBoost)	before sampling	with 'gender'	-	-	0.998	-	-	1	-	-	0.9956	-	-	5.90E-06
			without 'gender'	-	-0.0002	0.9978	-	/	1	-	-0.0001	0.9965	-	0.0508	6.20E-06
			remove all	-	-0.0006	0.9974	-	/	1	-	-0.0008	0.9958	-	0.4746	8.70E-06
		after sampling	with 'gender'	-0.0004	-	0.9976	/	-	1	-0.0007	-	0.9959	0.4088	-	8.30E-06
			without 'gender'	-0.0004	-0.0002	0.9974	/	/	1	-0.0008	-0.0002	0.9957	0.4839	0.1084	9.20E-06
			remove all	-0.0002	-0.0004	0.9972	/	/	1	-0.0004	-0.0005	0.9954	0.1954	0.2530	1.04E-05
Precision	after normalization (XGBoost)	before sampling	with 'gender'	-	-	0.9725	-	-	1	-	-	0.9614	-	-	0.000747
			without 'gender'	-	-0.0005	0.972	-	/	1	-	-0.0007	0.9607	-	0.0350	0.000773
			remove all	-	-0.0033	0.9693	-	/	1	-	-0.0046	0.957	-	0.2390	0.000925
		after sampling	with 'gender'	-0.0051	-	0.9675	-0.0071	-	0.9929	-0.0052	-	0.9564	-0.1042	-	0.000669
			without 'gender'	-0.0032	0.0014	0.9689	/	0.0072	1	-0.0045	/	0.9564	0.2331	0.4247	0.000953
			remove all	-0.0053	-0.0034	0.9642	/	0.0072	1	-0.0077	-0.0071	0.9496	0.3759	0.9031	0.001273
Recall	after normalization (XGBoost)	before sampling	with 'gender'	-	-	0.9722	-	-	1	-	-	0.9609	-	-	0.000766
			without 'gender'	-	-0.0010	0.9712	-	/	1	-	-0.0016	0.9594	-	0.0755	0.000823
			remove all	-	-0.0084	0.964	-	/	1	-	-0.0121	0.9493	-	0.6801	0.001286
		after sampling	with 'gender'	-0.0010	-	0.9712	-0.0035	-	0.9965	/	-	0.9609	-0.1732	-	0.000633
			without 'gender'	0.0010	0.0010	0.9722	/	0.0035	1	0.0015	/	0.9609	-0.0702	0.2095	0.000766
			remove all	0.0064	-0.0010	0.9702	/	0.0035	1	0.0092	-0.0030	0.958	-0.3133	0.3954	0.000883
f1 score	after normalization (XGBoost)	before sampling	with 'gender'	-	-	0.972	-	-	1	-	-	0.9605	-	-	0.000779
			without 'gender'	-	-0.0008	0.9712	-	/	1	-	-0.0011	0.9594	-	0.0576	0.000824
			remove all	-	-0.0064	0.9658	-	/	1	-	-0.0090	0.9519	-	0.4852	0.001157
		after sampling	with 'gender'	-0.0028	-	0.9693	-0.0053	-	0.9947	-0.0021	-	0.9585	-0.1592	-	0.000655
			without 'gender'	-0.0007	0.0012	0.9705	/	0.0053	1	-0.0009	/	0.9585	0.0458	0.3155	0.000862
			remove all	0.0013	-0.0023	0.9671	/	0.0053	1	0.0019	-0.0050	0.9537	-0.0735	0.6366	0.001072
AUC	after normalization (Random Forest)	before sampling	with 'gender'	-	-	0.9952	-	-	1	-	-	0.9929	-	-	2.49E-05
			without 'gender'	-	0.0016	0.9968	-	-0.0001	0.9999	-	0.0023	0.9952	-	-0.5622	1.09E-05
			remove all	-	-0.0014	0.9938	-	/	1	-	-0.0020	0.9909	-	0.6586	4.13E-05
		after sampling	with 'gender'	-0.0019	-	0.9933	-0.0001	-	0.9999	-0.0027	-	0.9902	0.8996	-	4.73E-05
			without 'gender'	-0.0033	0.0002	0.9935	-0.0001	-0.0001	0.9998	-0.0047	0.0003	0.9905	3.0092	-0.0761	4.37E-05
			remove all	-0.0005	/	0.9933	/	0.0001	1	-0.0006	0.0001	0.9903	0.1308	-0.0127	4.67E-05
Precision	after normalization (Random Forest)	before sampling	with 'gender'	-	-	0.964	-	-	0.9929	-	-	0.9514	-	-	0.000862
			without 'gender'	-	-0.0009	0.9631	-	/	0.9929	-	-0.0013	0.9502	-	0.0591	0.000913
			remove all	-	0.0049	0.9687	-	/	0.9929	-	0.0069	0.958	-	-0.2934	0.000609
		after sampling	with 'gender'	-0.0027	-	0.9614	/	-	0.9929	-0.0037	-	0.9479	0.1740	-	0.001013
			without 'gender'	0.0001	0.0019	0.9632	/	/	0.9929	0.0008	0.0033	0.951	-0.0371	-0.1314	0.00088
			remove all	-0.0002	0.0074	0.9685	/	/	0.9929	-0.0004	0.0102	0.9576	0.0236	-0.3839	0.000624
Recall	after normalization (Random Forest)	before sampling	with 'gender'	-	-	0.9722	-	-	0.9965	-	-	0.9623	-	-	0.000582
			without 'gender'	-	-0.0042	0.9681	-	/	0.9965	-	-0.0060	0.9565	-	0.3688	0.000797
			remove all	-	-0.0021	0.9702	-	/	0.9965	-	-0.0030	0.9594	-	0.1772	0.000686
		after sampling	with 'gender'	-0.0021	-	0.9702	/	-	0.9965	-0.0030	-	0.9594	0.1772	-	0.000686
			without 'gender'	-0.0021	-0.0042	0.9661	-0.0036	-0.0036	0.9929	-0.0015	-0.0045	0.9551	-0.1020	0.0442	0.000716
			remove all	0.0042	0.0042	0.9743	/	/	0.9965	0.0060	0.0060	0.9652	-0.2887	-0.2887	0.000488
f1 score	after normalization (Random Forest)	before sampling	with 'gender'	-	-	0.9676	-	-	0.9947	-	-	0.9563	-	-	0.000738
			without 'gender'	-	-0.0021	0.9656	-	/	0.9947	-	-0.0031	0.9533	-	0.1589	0.000855
			remove all	-	0.0016	0.9691	-	/	0.9947	-	0.0020	0.9582	-	-0.1000	0.000664
		after sampling	with 'gender'	-0.0024	-	0.9653	/	-	0.9947	-0.0033	-	0.9531	0.1701	-	0.000863
			without 'gender'	-0.0010	-0.0007	0.9646	-0.0018	-0.0018	0.9929	-0.0004	-0.0002	0.9529	-0.0647	-0.0737	0.0008
			remove all	0.0023	0.0062	0.9713	/	/	0.9947	0.0033	0.0087	0.9614	-0.1649	-0.3577	0.000554
AUC	after normalization (Logistic Regression)	before sampling	with 'gender'	-	-	0.9967	-	-	0.9984	-	-	0.9958	-	-	3.50E-06
			without 'gender'	-	0.0001	0.9958	-	-0.0001	0.9983	-	0.0001	0.9959	-	-0.1429	3.00E-06
			remove all	-	-0.0013	0.9954	-	/	0.9984	-	-0.0020	0.9938	-	1.9714	1.04E-05
		after sampling	with 'gender'	-0.0004	-	0.9963	0.0009	-	0.9993	-0.0016	-	0.9942	2.7143	-	1.30E-05
			without 'gender'	-0.0005	/	0.9963	0.0009	-0.0001	0.9992	-0.0013	0.0004	0.9946	2.5667	-0.1769	1.07E-05
			remove all	-0.0014	-0.0023	0.994	0.0010	0.0001	0.9994	-0.0027	-0.0031	0.9911	2.2885	1.6308	3.42E-05
Precision	after normalization (Random Forest)	before sampling	with 'gender'	-	-	0.9679	-	-	0.9929	-	-	0.9611	-	-	0.000507
			without 'gender'	-	-0.0017	0.9663	-	/	0.9929	-	-0.0021	0.9591	-	0.1270	0.000572
			remove all	-	-0.0004	0.9675	-	/	0.9929	-	-0.0005	0.9606	-	0.0304	0.000523
		after sampling	with 'gender'	-0.0056	-	0.9625	/	-	0.9929	-0.0119	-	0.9497	0.8407	-	0.000934
			without 'gender'	-0.0037	0.0002	0.9627	/	/	0.9929	-0.0091	0.0007	0.9504	0.5815	-0.0317	0.000904
			remove all	-0.0041	0.0010	0.9635	/	/	0.9929	-0.0097	0.0017	0.9513	0.6540	-0.0741	0.000864
Recall	after normalization (Random Forest)	before sampling	with 'gender'	-	-	0.9589	-	-	0.9894	-	-	0.9464	-	-	0.000924
			without 'gender'	-	-0.0022	0.9568	-	-0.0036	0.9858	-	-0.0016	0.9449	-	-0.0948	0.000836
			remove all	-	-0.0087	0.9506	-	-0.0036	0.9858	-	-0.0108	0.9362	-	0.3313	0.00123
		after sampling	with 'gender'	-0.0032	-	0.9558	0.0035	-	0.9929	-0.0061	-	0.9406	0.4825	-	0.001369
			without 'gender'	/	0.0010	0.9568	0.0037	-0.0035	0.9894	-0.0015	0.0031	0.9435	0.2590	-0.2313	0.001053
			remove all	-0.0021	-0.0075	0.9486	0.0037	-0.0035	0.9894	-0.0046	-0.0092	0.9319	0.3435	0.2065	0.001652
f1 score	after normalization (Random Forest)	before sampling	with 'gender'	-	-	0.9622	-	-	0.9911	-	-	0.9504	-	-	0.000828
			without 'gender'	-	-0.0019	0.9604	-	-0.0018	0.9893	-	-0.0018	0.9489	-	-0.0127	0.000818
			remove all	-	-0.0054	0.957	-	-0.0018	0.9893	-	-0.0068	0.9439	-	0.2444	0.001031
		after sampling	with 'gender'	-0.0042	-	0.9582	0.0018	-	0.9929	-0.0069	-	0.9438	0.4573	-	0.001207
			without 'gender'	-0.0016	0.0007	0.9589	0.0018	-0.0018	0.9911	-0.0033	0.0021	0.9458	0.2576	-0.1480	0.001028
			remove all	-0.0029	-0.0042	0.9542	0.0018	-0.0018	0.9911	-0.0053	-0.0052	0.9389	0.3251	0.1315	0.001366

Figure B.4: Performance on test dataset while choosing 'gender' as the protected feature (Thyroid dataset)

B.3 Threshold sheet

			Features	entire dataset		Male		Female		Average		Median		Maximum		Minimum		var		
Based on Gender first: threshold second: error	after normalization (XGBoost)	before sampling	with 'gender'	**0.37157**	0.01594	0.55927	0.01304	0.37157	0.01594	0.46542	0.01449	0.46542	0.01449	0.55927	0.01304	0.37157	0.01594	1.17E-02		
			without 'gender'	**0.40312**	0.01358	0.48913	0.01401	0.40312	0.01358	0.44612	0.01261	0.44612	0.01261	0.48913	0.01401	0.40312	0.01358	2.47E-03		
			remove all	**0.36164**	0.01868	0.52365	0.01643	0.28614	0.01873	0.4049	0.01723	0.4049	0.01723	0.52365	0.01643	0.28614	0.01873	1.47E-02		
		after ampling	with 'gender'	**0.28099**	0.01594	0.46666	0.01234	0.33673	0.01449	0.4017	0.01401	0.4017	0.01401	0.46666	0.01234	0.33673	0.01449	9.08E-03		
			without 'gender'	**0.4458**	0.01353	0.51649	0.01138	0.4458	0.01353	0.48114	0.01304	0.48114	0.01304	0.51649	0.01138	0.4458	0.01353	1.67E-03		
			remove all	**0.29511**	0.01787	0.55253	0.01401	0.29511	0.01787	0.42382	0.01498	0.42382	0.01498	0.55253	0.01401	0.29511	0.01787	2.21E-02		
	after normalization (Random Forest)	before sampling	with 'gender'	**0.4**	0.01186	0.4	0.01186	0.4	0.01186	0.4	0.01186	0.4	0.01186	0.4	0.01186	0.4	0.01186	4.62E-33		
			without 'gender'	**0.5**	0.01331	0.5	0.01331	0.4	0.01454	0.45	0.01454	0.45	0.01454	0.5	0.01331	0.4	0.01454	3.33E-03		
			remove all	**0.5**	0.01234	0.5	0.01234	0.4	0.01739	0.45	0.01739	0.45	0.01739	0.5	0.01234	0.4	0.01739	3.33E-03		
		after sampling	with 'gender'	**0.5**	0.01234	0.6	0.01283	0.5	0.01234	0.55	0.01234	0.55	0.01234	0.6	0.01283	0.5	0.01234	3.33E-03		
			without 'gender'	**0.4**	0.00998	0.5	0.01261	0.4	0.00998	0.45	0.00998	0.45	0.00998	0.5	0.01261	0.4	0.00998	3.33E-03		
			remove all	**0.5**	0.01041	0.6	0.01041	0.5	0.01041	0.55	0.01041	0.55	0.01041	0.6	0.01041	0.5	0.01041	3.33E-03		
	after normalization (Logistic Regression)	before sampling	with 'gender'	**0.48994**	0.01648	0.49985	0.01599	0.2211	0.02066	0.36048	0.01916	0.36048	0.01916	0.49985	0.01599	0.2211	0.02066	0.025012		
			without 'gender'	**0.45206**	0.01626	0.49451	0.01529	0.22167	0.01685	0.35809	0.0175	0.35809	0.0175	0.49451	0.01529	0.22167	0.01685	0.021554		
			remove all	**0.44473**	0.01798	0.48271	0.01868	0.41476	0.02039	0.44874	0.01916	0.44874	0.01916	0.48271	0.01868	0.41476	0.02039	0.00116		
		after ampling	with 'gender'	**0.36271**	0.01819	0.48114	0.01744	0.36271	0.01819	0.42192	0.01868	0.42192	0.01868	0.48114	0.01744	0.36271	0.01819	0.004675		
			without 'gender'	**0.34941**	0.0175	0.44969	0.01723	0.34941	0.0175	0.39955	0.01868	0.39955	0.01868	0.44969	0.01723	0.34941	0.0175	0.003352		
			remove all	**0.40943**	0.02474	0.41974	0.02378	0.37616	0.02125	0.39795	0.02452	0.39795	0.02452	0.41974	0.02378	0.37616	0.02125	0.000519		

			Features	entire dataset		xx - 40		41 - 70		71 - xx		Average		Median		Maximum		Minimum		var
Based on Age first: threshold second: error	after normalization (XGBoost)	before sampling	with 'age'	**0.37157**	0.0006	**0.29029**	0.00235	**0.4237**	0.0006	**0.39163**	0.0006	**0.36854**	0.0006	**0.39163**	0.0006	**0.4237**	0.0006	**0.29029**	0.00235	3.235E-03
			without 'age'	**0.34252**	0.00212	**0.17023**	0.00505	**0.34252**	0.00212	**0.39649**	0.0027	**0.30308**	0.00155	**0.34252**	0.00212	**0.39649**	0.0027	**0.17023**	0.00505	9.699E-03
			remove all	**0.36164**	0.00247	**0.18041**	0.00505	**0.28614**	0.00155	**0.36164**	0.00247	**0.27606**	0.0019	**0.28614**	0.00155	**0.36164**	0.00247	**0.18041**	0.00505	7.356E-03
		after ampling	with 'age'	**0.4796**	0.00095	**0.38611**	0.00165	**0.4796**	0.00095	**0.14698**	0.00282	**0.33756**	0.002	**0.38611**	0.00165	**0.4796**	0.00095	**0.14698**	0.00282	2.466E-02
			without 'age'	**0.38759**	0.00177	**0.43114**	0.00107	**0.38759**	0.00177	**0.08327**	0.00773	**0.30067**	0.00247	**0.38759**	0.00177	**0.43114**	0.00107	**0.08327**	0.00773	2.584E-02
			remove all	**0.42178**	0.002	**0.42178**	0.002	**0.50264**	0.0006	**0.05224**	0.00913	**0.32555**	0.0027	**0.42178**	0.002	**0.50264**	0.0006	**0.05224**	0.00913	4.075E-02
	after normalization (Random Forest)	before sampling	with 'age'	**0.4**	0.0013	**0.5**	0.00165	**0.4**	0.0013	**0.7**	0.002	**0.53333**	0.00165	**0.5**	0.00165	**0.7**	0.002	**0.4**	0.0013	2.000E-02
			without 'age'	**0.3**	0.00282	**0.4**	0.00107	**0.5**	0.00095	**0.3**	0.00282	**0.4**	0.00107	**0.4**	0.00107	**0.5**	0.00095	**0.3**	0.00282	9.167E-03
			remove all	**0.5**	0.002	**0.5**	0.002	**0.4**	0.002	**0.6**	0.002	**0.5**	0.002	**0.5**	0.002	**0.6**	0.002	**0.4**	0.002	6.667E-03
		after ampling	with 'age'	**0.4**	0.00235	**0.4**	0.00235	**0.5**	0.00235	**0.6**	0.002	**0.5**	0.00235	**0.5**	0.00235	**0.6**	0.002	**0.4**	0.00235	9.167E-03
			without 'age'	**0.4**	0.00247	**0.4**	0.00247	**0.4**	0.00247	**0.4**	0.00247	**0.4**	0.00247	**0.4**	0.00247	**0.4**	0.00247	**0.4**	0.00247	0.000E+00
			remove all	**0.3**	0.00515	**0.3**	0.00515	**0.5**	0.00235	**0.7**	0.002	**0.5**	0.00235	**0.5**	0.00235	**0.7**	0.002	**0.3**	0.00515	3.667E-02
	after normalization (Logistic Regression)	before sampling	with 'age'	**0.48994**	0.00235	**0.2211**	0.00212	**0.49013**	0.00235	**0.48994**	0.00235	**0.40039**	0.00235	**0.48994**	0.00235	**0.49013**	0.00235	**0.2211**	0.00212	1.808E-02
			without 'age'	**0.43416**	0.00515	0.411	0.00585	0.48554	0.0034	0.48991	0.0034	0.46215	0.00375	0.48554	0.0034	0.48991	0.0034	0.411	0.00585	1.507E-03
			remove all	**0.44473**	0.00375	0.40687	0.0055	0.4494	0.00375	0.4883	0.00247	0.44819	0.00375	0.4494	0.00375	0.4883	0.00247	0.40687	0.0055	1.109E-03
		after ampling	with 'age'	**0.46964**	0.00177	0.18933	0.0061	0.4764	0.00177	0.49108	0.00142	0.3856	0.00247	0.4764	0.00177	0.49108	0.00142	0.18933	0.0061	2.106E-02
			without 'age'	**0.42591**	0.00422	0.25188	0.0089	0.43067	0.00387	0.49123	0.00247	0.39126	0.004	0.43067	0.00387	0.49123	0.00247	0.25188	0.0089	1.062E-02
			remove all	**0.39994**	0.00422	0.39994	0.00422	0.45183	0.00317	0.49147	0.00247	0.44775	0.00317	0.45183	0.00317	0.49147	0.00247	0.39994	0.00422	1.976E-03

Figure B.5: Threshold sheet (Thyroid dataset)

Appendix C

RESULTS ON PIMA DATASET

C.1 Performance sheet for 10-fold cross validation on train dataset

C.1.1 Performance while choosing 'age' as the protected feature

Diabetes_pima		Features		entire dataset			xx - 35			age 36-50			51-xx			Variance		
size of the group: first number: target = 0 second number: target = 1 third number: sum	test dataset	withWithout the feature 'age'		102, 52 154			73, 26, 99			20, 18 38			9, 8 17			variance of the three age groups		
	before sampling(train)			407, 207 614			294, 105 399			79, 72 151			34, 30 64					
	after sampling(train)			439, 309 748			182, 67 249			125, 125 250			132, 117 249					
	Change		-	sample	no 'age'	-	sample	no 'age'	-	sample	no 'age'	-	sample	no 'age'	-	sample	no 'age'	-
AUC	after normalization (XGBoost)	before sampling	with 'age'	-	-	0.83270	-	-	0.83600	-	-	0.77210	-	-	0.85280	-	-	0.001809
			without 'age'	-	-0.0014	0.83150	-	-0.0005	0.83560	-	-0.0058	0.76760	-	-0.0326	0.82500	-	-0.2604	0.001338
		after sampling	with 'age'	-0.0166	-	0.81890	-0.0050	-	0.83180	-0.0319	-	0.74750	-0.1433	-	0.73060	0.625200	-	0.002940
			without 'age'	-0.0207	-0.0056	0.81430	-0.0077	-0.0031	0.82920	-0.0259	0.0003	0.74770	-0.1212	-0.0077	0.72500	1.244400	0.0214	0.003003
Precision	after normalization (XGBoost)	before sampling	with 'age'	-	-	0.77890	-	-	0.79570	-	-	0.72610	-	-	0.81600	-	-	0.002221
			without 'age'	-	-0.0261	0.75860	-	-0.0113	0.78670	-	-0.0353	0.70050	-	-0.0533	0.77250	-	-0.0374	0.002138
		after sampling	with 'age'	-0.0403	-	0.74750	-0.0371	-	0.76620	-0.0416	-	0.69590	-0.1248	-	0.71420	-0.401200	-	0.001330
			without 'age'	-0.0236	-0.0091	0.74070	-0.0194	0.0068	0.77140	-0.0094	-0.0029	0.69390	-0.1313	-0.0603	0.67110	0.293300	1.0789	0.002765
Recall	after normalization (XGBoost)	before sampling	with 'age'	-	-	0.78010	-	-	0.80460	-	-	0.71470	-	-	0.78090	-	-	0.002173
			without 'age'	-	-0.0232	0.76200	-	-0.0127	0.79440	-	-0.0374	0.68800	-	-0.0579	0.73570	-	0.3056	0.002837
		after sampling	with 'age'	-0.0399	-	0.74900	-0.0252	-	0.78430	-0.0462	-	0.68170	-0.1158	-	0.69050	0.486900	-	0.003231
			without 'age'	-0.0239	-0.0069	0.74380	-0.0096	0.0032	0.78680	-0.0304	-0.0214	0.66710	-0.1068	-0.0484	0.65710	0.834700	0.6110	0.005205
f1 score	after normalization (XGBoost)	before sampling	with 'age'	-	-	0.77560	-	-	0.79290	-	-	0.71050	-	-	0.77330	-	-	0.001852
			without 'age'	-	-0.0231	0.75770	-	-0.0087	0.78600	-	-0.0359	0.68500	-	-0.0623	0.72510	-	0.3963	0.002586
		after sampling	with 'age'	-0.0467	-	0.73940	-0.0386	-	0.76230	-0.0442	-	0.67910	-0.1182	-	0.68190	0.204600	-	0.002231
			without 'age'	-0.0317	-0.0077	0.73370	-0.0202	0.0102	0.77010	-0.0399	-0.0315	0.65770	-0.1011	-0.0441	0.65180	0.718100	0.9915	0.004443
AUC	after normalization (Random Forest)	before sampling	with 'age'	-	-	0.82620	-	-	0.84250	-	-	0.72640	-	-	0.78890	-	-	0.003372
			without 'age'	-	-0.0001	0.82610	-	-0.0033	0.83970	-	0.0094	0.73320	-	0.0176	0.80280	-	-0.1323	0.002926
		after sampling	with 'age'	-0.0230	-	0.80720	-0.0185	-	0.82690	-0.0372	-	0.69940	-0.1127	-	0.70000	0.600200	-	0.005396
			without 'age'	-0.0314	-0.0087	0.80020	-0.0324	-0.0174	0.81250	-0.0221	0.0252	0.71700	-0.0831	0.0516	0.73610	-0.127500	-0.5269	0.002553
Precision	after normalization (Random Forest)	before sampling	with 'age'	-	-	0.75650	-	-	0.78020	-	-	0.68700	-	-	0.76940	-	-	0.002598
			without 'age'	-	-0.0096	0.74920	-	-0.0091	0.77310	-	-0.0246	0.67010	-	0.0333	0.79500	-	0.7125	0.004449
		after sampling	with 'age'	-0.0325	-	0.73190	-0.0155	-	0.76810	-0.0846	-	0.62890	0.0343	-	0.79580	2.080800	-	0.008004
			without 'age'	-0.0076	0.0158	0.74350	0.0066	0.0131	0.77820	0.0236	0.0906	0.68590	-0.1114	-0.1123	0.70640	-0.472900	-0.7070	0.002345
Recall	after normalization (Random Forest)	before sampling	with 'age'	-	-	0.76020	-	-	0.79180	-	-	0.68080	-	-	0.75240	-	-	0.003163
			without 'age'	-	-0.0086	0.75370	-	-0.0091	0.78460	-	-0.0306	0.66000	-	0.0348	0.77860	-	0.5624	0.004942
		after sampling	with 'age'	-0.0300	-	0.73740	-0.0129	-	0.78160	-0.0967	-	0.61500	0.0032	-	0.75480	1.529200	-	0.008000
			without 'age'	-0.0090	0.0129	0.74690	-0.0008	0.0031	0.78400	0.0121	0.0862	0.66800	-0.0918	-0.0632	0.70710	-0.296400	-0.5654	0.003477
f1 score	after normalization (Random Forest)	before sampling	with 'age'	-	-	0.75430	-	-	0.77400	-	-	0.67720	-	-	0.74680	-	-	0.002491
			without 'age'	-	-0.0086	0.74780	-	-0.0019	0.77250	-	-0.0285	0.65790	-	0.0376	0.77490	-	0.7965	0.004475
		after sampling	with 'age'	-0.0317	-	0.73040	-0.0163	-	0.76140	-0.1012	-	0.60870	-0.0017	-	0.74550	1.827800	-	0.007044
			without 'age'	-0.0115	0.0120	0.73920	0.0010	0.0156	0.77330	0.0105	0.0922	0.66480	-0.1023	-0.0669	0.69560	-0.301700	-0.5564	0.003125
AUC	after normalization (Logistic Regression)	before sampling	with 'age'	-	-	0.83050	-	-	0.84750	-	-	0.76990	-	-	0.76670	-	-	0.002095
			without 'age'	-	0.0045	0.83420	-	0.0020	0.84920	-	0.0003	0.77010	-	-0.0037	0.76390	-	0.0797	0.002262
		after sampling	with 'age'	-0.0122	-	0.82040	-0.0136	-	0.83600	-0.0221	-	0.75290	0.0181	-	0.78060	-0.144600	-	0.001792
			without 'age'	-0.0150	0.0016	0.82170	-0.0194	-0.0039	0.83270	-0.0106	0.0120	0.76190	0.0073	-0.0142	0.76950	-0.331600	-0.1563	0.001512
Precision	after normalization (Random Forest)	before sampling	with 'age'	-	-	0.76890	-	-	0.81510	-	-	0.70340	-	-	0.62680	-	-	0.008962
			without 'age'	-	0.0013	0.76990	-	-0.0177	0.80070	-	0.0073	0.70850	-	0.1827	0.74130	-	-0.7565	0.002182
		after sampling	with 'age'	-0.0023	-	0.76710	-0.0110	-	0.80610	0.0165	-	0.71500	0.1469	-	0.71890	-0.704200	-	0.002651
			without 'age'	-0.0018	0.0018	0.76850	/	-0.0067	0.80070	0.0092	0.0000	0.71500	0.0074	0.0388	0.74680	-0.140200	-0.2923	0.001876
Recall	after normalization (Random Forest)	before sampling	with 'age'	-	-	0.75850	-	-	0.81210	-	-	0.69290	-	-	0.58100	-	-	0.013360
			without 'age'	-	-0.0003	0.75830	-	-0.0248	0.79200	-	0.0095	0.69950	-	0.1843	0.68810	-	-0.7572	0.003244
		after sampling	with 'age'	0.0071	-	0.76390	-0.0121	-	0.80230	0.0123	-	0.70140	0.1680	-	0.67860	-0.675200	-	0.004339
			without 'age'	0.0092	0.0018	0.76530	0.0067	-0.0062	0.79730	0.0027	0.0000	0.70140	0.0484	0.0631	0.72140	-0.211200	-0.4102	0.002559
f1 score	after normalization (Random Forest)	before sampling	with 'age'	-	-	0.76120	-	-	0.81050	-	-	0.68470	-	-	0.54370	-	-	0.017812
			without 'age'	-	0.0001	0.76130	-	-0.0212	0.79330	-	0.0104	0.69180	-	0.2448	0.67680	-	-0.7745	0.004017
		after sampling	with 'age'	0.0033	-	0.76370	-0.0127	-	0.80020	0.0185	-	0.69740	0.2211	-	0.66390	-0.716900	-	0.005042
			without 'age'	0.0050	0.0018	0.76510	0.0032	-0.0055	0.79580	0.0081	0.0000	0.69740	0.0519	0.0723	0.71190	-0.298000	-0.4407	0.002820

Figure C.1: Performance of 10-fold cross validation while choosing 'age' as the protected feature (PIMA dataset)

C.2 Performance sheet for test dataset

C.2.1 Performance while choosing 'age' as the protected feature

Diabetes_pima		Features		entire dataset			xx - 35			age 36-50			51-xx			Variance		
size of the group: first number: target = 0 second number: target = 1 third number: sum	test dataset	with/without the feature 'age'		102, 52 154			73, 26, 99			20, 18 38			9, 8 17			variance of the three age groups		
	before sampling(train)			407, 207 614			294, 105 399			79, 32 144			34, 30 64					
	after sampling(train)			439, 309 748			182, 67 249			125, 125 250			132, 117 249					
Change			-	sample	no 'age'	-	sample	no 'age'	-	sample	no 'age'	-	sample	no 'age'	-	sample	no 'age'	-
AUC	after normalization (XGBoost)	before sampling	with 'age'	-	-	0.8296	-	-	0.8222	-	-	0.8441	-	-	0.8442	-	-	0.0001606
			without 'age'	-	-0.0160	0.8163	-	-0.0152	0.8097	-	-0.0174	0.8294	-	/	0.8442	-	0.8655	0.0002996
		after sampling	with 'age'	-0.0212	-	0.812	0.0041	-	0.8256	-0.0906	-	0.7676	-0.0308	-	0.8182	5.2049	-	0.0009965
			without 'age'	-0.0622	-0.0419	0.778	-0.0399	-0.0438	0.7894	-0.1045	-0.0152	0.7559	-0.0770	-0.0477	0.7792	0.8362	-0.7041	0.0002949
Precision	after normalization (XGBoost)	before sampling	with 'age'	-	-	0.7537	-	-	0.7908	-	-	0.7501	-	-	0.6019	-	-	0.0098838
			without 'age'	-	-0.0265	0.7337	-	-0.0101	0.7828	-	-0.0956	0.6784	-	/	0.6019	-	-0.1657	0.0082461
		after sampling	with 'age'	-0.0283	-	0.7324	-0.0250	-	0.771	-0.1304	-	0.6523	0.2098	-	0.7282	-0.6344	-	0.0036137
			without 'age'	-0.0569	-0.0295	0.7108	-0.0787	-0.0550	0.7286	-0.0505	0.0918	0.7122	0.1077	-0.0845	0.6667	-0.8959	-0.7154	0.0010285
Recall	after normalization (XGBoost)	before sampling	with 'age'	-	-	0.7597	-	-	0.798	-	-	0.7297	-	-	0.6111	-	-	0.0089437
			without 'age'	-	-0.0255	0.7403	-	-0.0127	0.7879	-	-0.0740	0.6757	-	/	0.6111	-	-0.1051	0.0080034
		after sampling	with 'age'	-0.0255	-	0.7403	-0.0253	-	0.7778	-0.1111	-	0.6486	0.1818	-	0.7222	-0.5304	-	0.0042002
			without 'age'	-0.0512	-0.0263	0.7208	-0.0759	-0.0519	0.7374	-0.0370	0.0834	0.7027	0.0910	-0.0768	0.6667	-0.8603	-0.7024	0.0012498
f1 score	after normalization (XGBoost)	before sampling	with 'age'	-	-	0.7506	-	-	0.791	-	-	0.7184	-	-	0.6046	-	-	0.0088277
			without 'age'	-	-0.0209	0.7349	-	-0.0081	0.7846	-	-0.0690	0.6688	-	/	0.6046	-	-0.0573	0.0083219
		after sampling	with 'age'	-0.0256	-	0.7314	-0.0228	-	0.773	-0.1121	-	0.6379	0.1594	-	0.701	-0.4824	-	0.0045696
			without 'age'	-0.0538	-0.0290	0.7102	-0.0750	-0.0534	0.7317	-0.0345	0.0873	0.6936	0.0347	-0.1076	0.6256	-0.6728	-0.3678	0.0028888
AUC	after normalization (Random Forest)	before sampling	with 'age'	-	-	0.8229	-	-	0.8213	-	-	0.8265	-	-	0.8571	-	-	0.0003742
			without 'age'	-	-0.0333	0.7955	-	-0.0421	0.7867	-	-0.0249	0.8059	-	-0.1061	0.7662	-	0.0534	0.0003942
		after sampling	with 'age'	-0.0137	-	0.8116	-0.0174	-	0.807	-0.0108	-	0.8176	-0.0151	-	0.8442	-0.0184	-	0.0003673
			without 'age'	-0.0701	-0.0572	0.7852	-0.0780	-0.0617	0.7572	-0.0676	-0.0575	0.7706	-0.1363	-0.1231	0.7403	-0.3840	-0.3724	0.0002305
Precision	after normalization (Random Forest)	before sampling	with 'age'	-	-	0.7396	-	-	0.7498	-	-	0.7363	-	-	0.7282	-	-	0.0001191
			without 'age'	-	-0.0264	0.7201	-	0.0161	0.7619	-	-0.0786	0.6784	-	-0.1734	0.6019	-	52.7708	0.0064041
		after sampling	with 'age'	0.0120	-	0.7485	-0.0037	-	0.747	0.0528	-	0.7752	0.1494	-	0.837	16.7926	-	0.0021191
			without 'age'	-0.0379	-0.0493	0.7116	-0.0396	-0.0360	0.7201	-0.0327	-0.0813	0.7122	/	-0.1300	0.7282	-0.4626	-0.9698	6.40E-05
Recall	after normalization (Random Forest)	before sampling	with 'age'	-	-	0.7468	-	-	0.7576	-	-	0.7297	-	-	0.7222	-	-	0.000348
			without 'age'	-	-0.0261	0.7273	-	0.0133	0.7677	-	-0.0740	0.6757	-	-0.1538	0.6111	-	16.7974	0.0061935
		after sampling	with 'age'	0.0086	-	0.7532	/	-	0.7576	/	-	0.7297	0.0770	-	0.7778	0.6761	-	0.0005833
			without 'age'	-0.0348	-0.0430	0.7208	-0.0400	-0.0400	0.7273	-0.0370	-0.0370	0.7027	/	-0.0715	0.7222	-0.5155	-0.7110	0.0001686
f1 score	after normalization (Random Forest)	before sampling	with 'age'	-	-	0.739	-	-	0.7523	-	-	0.724	-	-	0.701	-	-	0.0006603
			without 'age'	-	-0.0235	0.7216	-	0.0156	0.764	-	-0.0762	0.6688	-	-0.1375	0.6046	-	8.7413	0.0064322
		after sampling	with 'age'	-0.0003	-	0.7388	-0.0041	-	0.7492	-0.0184	-	0.7107	0.0705	-	0.7504	-0.2276	-	0.00051
			without 'age'	-0.0363	-0.0360	0.7122	-0.0389	-0.0350	0.723	-0.0420	-0.0241	0.6936	/	-0.0658	0.701	-0.6458	-0.5414	0.0002339
AUC	after normalization (Logistic Regression)	before sampling	with 'age'	-	-	0.8235	-	-	0.8155	-	-	0.8382	-	-	0.7792	-	-	0.0008857
			without 'age'	-	-0.0019	0.8219	-	0.0017	0.8169	-	-0.0035	0.8353	-	/	0.7792	-	-0.0767	0.0008178
		after sampling	with 'age'	-0.0159	-	0.8104	-0.0142	-	0.8039	-0.0351	-	0.8088	0.0334	-	0.8052	-0.9928	-	6.40E-06
			without 'age'	-0.0159	/	0.8104	-0.0142	/	0.8039	-0.0351	/	0.8088	0.0334	/	0.8052	-0.9928	/	6.40E-06
Precision	after normalization (Random Forest)	before sampling	with 'age'	-	-	0.7555	-	-	0.8269	-	-	0.6475	-	-	0.6019	-	-	0.0141481
			without 'age'	-	-0.0130	0.7457	-	-0.0204	0.81	-	/	0.6475	-	/	0.6019	-	-0.1543	0.0119652
		after sampling	with 'age'	-0.0549	-	0.714	-0.0890	-	0.7533	0.0477	-	0.6784	/	-	0.6019	-0.5949	-	0.0057307
			without 'age'	-0.0549	/	0.714	-0.0890	/	0.7533	0.0477	/	0.6784	/	/	0.6019	-0.5949	/	0.0057307
Recall	after normalization (Random Forest)	before sampling	with 'age'	-	-	0.7532	-	-	0.8182	-	-	0.6486	-	-	0.6111	-	-	0.0121768
			without 'age'	-	-0.0085	0.7468	-	-0.0123	0.8081	-	/	0.6486	-	/	0.6111	-	-0.1014	0.0109426
		after sampling	with 'age'	-0.0430	-	0.7208	-0.0741	-	0.7576	0.0418	-	0.6757	/	-	0.6111	-0.5573	-	0.0053905
			without 'age'	-0.0430	/	0.7208	-0.0741	/	0.7576	0.0418	/	0.6757	/	/	0.6111	-0.5573	/	0.0053905
f1 score	after normalization (Random Forest)	before sampling	with 'age'	-	-	0.7542	-	-	0.8211	-	-	0.6476	-	-	0.6046	-	-	0.0131373
			without 'age'	-	-0.0106	0.7462	-	-0.0147	0.809	-	/	0.6476	-	/	0.6046	-	-0.1160	0.0116131
		after sampling	with 'age'	-0.0509	-	0.7158	-0.0804	-	0.7551	0.0327	-	0.6688	/	-	0.6046	-0.5659	-	0.0057033
			without 'age'	-0.0509	/	0.7158	-0.0804	/	0.7551	0.0327	/	0.6688	/	/	0.6046	-0.5659	/	0.0057033

Figure C.2: Performance on test dataset while choosing 'age' as the protected feature(PIMA dataset)

C.3 Threshold sheet

			Features	entire dataset		xx - 35		36 - 50		51 - xx		Average		Median		Maximum		Minimum		var
Based on Age first: threshold second: error	after normalization (XGBoost)	before sampling	with 'age'	**0.6314**	0.0085	**0.6314**	0.0085	**0.6198**	0.0085	**0.6653**	0.0197	**0.6388**	0.0085	**0.6314**	0.0085	**0.6653**	0.0197	**0.6198**	0.0085	3.865E-04
			without 'age'	**0.6885**	0.0161	**0.7008**	0.0182	**0.6889**	0.0143	**0.6885**	0.0161	**0.6927**	0.0143	**0.6889**	0.0143	**0.7008**	0.0182	**0.6885**	0.0161	3.704E-05
		after ampling	with 'age'	**0.7617**	0.0195	**0.5695**	0.0064	**0.7791**	0.0253	**0.8004**	0.0253	**0.7163**	0.0082	**0.7791**	0.0253	**0.8004**	0.0253	**0.5695**	0.0064	1.137E-02
			without 'age'	**0.6363**	0.007	**0.6363**	0.007	**0.7196**	0.0082	**0.7681**	0.0314	**0.708**	0.0054	**0.7196**	0.0082	**0.7681**	0.0314	**0.6363**	0.007	4.248E-03
	after normalization (Random Forest)	before sampling	with 'age'	**0.645**	0.0235	**0.674**	0.0235	**0.682**	0.0195	**0.588**	0.0138	**0.648**	0.0235	**0.674**	0.0235	**0.682**	0.0195	**0.588**	0.0138	1.813E-03
			without 'age'	**0.692**	0.0181	**0.693**	0.0181	**0.805**	0.0162	**0.686**	0.0219	**0.728**	0.0154	**0.693**	0.0181	**0.805**	0.0162	**0.686**	0.0219	3.297E-03
		after ampling	with 'age'	**0.745**	0.0171	**0.61**	0.0191	**0.612**	0.0153	**0.514**	0.0126	**0.5787**	0.0198	**0.61**	0.0191	**0.612**	0.0153	**0.514**	0.0126	9.008E-03
			without 'age'	**0.649**	0.0266	**0.59**	0.0234	**0.651**	0.0229	**0.89**	0.0164	**0.7103**	0.0124	**0.651**	0.0229	**0.89**	0.0164	**0.59**	0.0234	1.770E-02
	after normalization (Logistic Regression)	before sampling	with 'age'	**0.6583**	0.0231	**0.5274**	0.0066	**0.7125**	0.0085	**0.7479**	0.0067	**0.6626**	0.0152	**0.7125**	0.0085	**0.7479**	0.0067	**0.5274**	0.0066	9.353E-03
			without 'age'	**0.6557**	0.0152	**0.5357**	0.0045	**0.7002**	0.0085	**0.7543**	0.0072	**0.6634**	0.0152	**0.7002**	0.0085	**0.7543**	0.0072	**0.5357**	0.0045	8.656E-03
		after ampling	with 'age'	**0.6261**	0.0092	**0.6064**	0.0049	**0.7177**	0.018	**0.7154**	0.0085	**0.6798**	0.0047	**0.7154**	0.0085	**0.7177**	0.018	**0.6064**	0.0049	3.419E-03
			without 'age'	**0.6261**	0.0092	**0.6064**	0.0049	**0.7177**	0.018	**0.7154**	0.0085	**0.6798**	0.0047	**0.7154**	0.0085	**0.7177**	0.018	**0.6064**	0.0049	3.419E-03

Figure C.3: Threshold sheet (PIMA dataset)