Online Appendix to Posterior Rebalance as a Method to Reduce Classifier Bias on Imbalanced Problems

Krasanakis Emmanouil[1,*], Eleftherios Spyromitros-Xioufis[1], Symeon Papadopoulos[1], Yiannis Kompatsiaris[1]

Abstract

In this appendix, we present classifier performance across datasets and rebalance schemes proposed for the paper *Posterior Rebalance as a Method to Reduce Classifier Bias on Imbalanced Problems*. Measurements are obtained from ten-fold cross-validiation and are used to derive performance rankings. All values are rounded to the nearest second decimal point.

1. Base Classifier

Table 1: Base Classifier

Dataset	Logistic Regression				Smooth-kNN				LibSVM			
	wAUC	wTPr	Fair	wTPR·Fair	wAUC	wTPr	Fair	wTPR·Fair	wAUC	wTPr	Fair	wTPR·Fair
Sexuality #1	94%	86%	95%	81%	79%	72%	91%	66%	92%	83%	94%	78%
Religion #1	87%	94%	90%	85%	86%	98%	87%	85%	93%	98%	85%	83%
Politics #1	82%	74%	100%	74%	82%	74%	100%	74%	89%	81%	100%	81%
Sexuality #2	49%	67%	60%	40%	51%	89%	8%	7%	41%	89%	0%	0%
Religion #2	57%	71%	58%	41%	55%	85%	0%	0%	38%	87%	0%	0%
Politics #2	60%	64%	93%	59%	71%	76%	67%	51%	46%	70%	0%	0%
Cannabis User	74%	84%	82%	68%	58%	92%	10%	9%	68%	92%	0%	0%
Spam Collection	97%	95%	100%	95%	96%	97%	98%	95%	99%	98%	100%	97%
Cell Review Sentiment	69%	66%	98%	65%	69%	63%	95%	60%	56%	53%	65%	35%
Dermatology	100%	96%	100%	96%	99%	93%	99%	93%	98%	87%	96%	83%
Diabetes	83%	77%	91%	71%	69%	68%	85%	58%	51%	65%	0%	0%
E. Coli	95%	86%	92%	79%	95%	87%	93%	81%	97%	84%	92%	77%
Hepatitis	80%	83%	84%	70%	64%	70%	7%	5%	47%	79%	0%	0%
Hypothyroid	97%	97%	90%	87%	94%	96%	82%	78%	94%	93%	52%	49%
Liver Disorders	72%	68%	92%	63%	66%	66%	88%	58%	57%	60%	11%	7%
Lung Cancer	67%	69%	96%	66%	58%	81%	60%	49%	48%	69%	24%	16%
Lymphography	81%	73%	86%	63%	89%	82%	91%	74%	92%	82%	89%	73%

2. Rebalancing Experiments

Table 2: 1*lin* on Base Classifier

Dataset	Logistic Regression				Smooth-kNN				LibSVM			
	wAUC	wTPr	Fair	wTPR·Fair	wAUC	wTPr	Fair	wTPR·Fair	wAUC	wTPr	Fair	wTPR·Fair
Sexuality #1	94%	85%	96%	81%	79%	71%	93%	66%	92%	83%	95%	78%
Religion #1	87%	93%	92%	86%	86%	92%	98%	90%	93%	97%	97%	94%
Politics #1	82%	74%	100%	74%	82%	74%	100%	74%	89%	81%	100%	81%
Sexuality #2	50%	66%	68%	45%	52%	54%	98%	53%	42%	45%	100%	45%

*Corresponding author

Email addresses: maniospas@iti.gr (Krasanakis Emmanouil), espyromi@iti.gr (Eleftherios Spyromitros-Xioufis), papadop@iti.gr (Symeon Papadopoulos), ikom@iti.gr (Yiannis Kompatsiaris)

[1]Centre of Research & Technology - Hellas (CERTH), 6th km Xarilaou - Thermi, 57001, Thessaloniki, Greece

Dataset	wAUC	wTPr	Fair	wTPR·Fair	wAUC	wTPr	Fair	wTPR·Fair	wAUC	wTPr	Fair	wTPR·Fair
Religion #2	57%	70%	59%	41%	57%	51%	97%	49%	39%	48%	99%	48%
Politics #2	60%	64%	93%	59%	72%	69%	93%	64%	47%	48%	79%	38%
Cannabis User	74%	81%	83%	67%	58%	70%	84%	59%	68%	68%	100%	68%
Spam Collection	97%	95%	100%	95%	96%	96%	100%	95%	99%	96%	100%	96%
Cell Review Sentiment	69%	66%	98%	65%	69%	63%	95%	60%	56%	53%	63%	33%
Dermatology	100%	96%	100%	96%	99%	92%	99%	92%	98%	87%	97%	84%
Diabetes	83%	76%	100%	76%	69%	63%	98%	62%	51%	42%	56%	24%
E. Coli	95%	84%	92%	77%	95%	86%	93%	80%	97%	83%	93%	77%
Hepatitis	80%	81%	97%	78%	64%	67%	94%	63%	49%	63%	61%	39%
Hypothyroid	97%	94%	98%	92%	94%	95%	98%	93%	94%	86%	94%	81%
Liver Disorders	72%	65%	100%	65%	66%	66%	88%	58%	57%	60%	43%	26%
Lung Cancer	67%	66%	98%	64%	60%	63%	91%	57%	49%	53%	100%	53%
Lymphography	81%	73%	86%	63%	89%	82%	92%	76%	93%	82%	90%	74%

Table 3: 1*inv* on Base Classifier

Dataset	Logistic Regression				Smooth-kNN				LibSVM			
	wAUC	wTPr	Fair	wTPR·Fair	wAUC	wTPr	Fair	wTPR·Fair	wAUC	wTPr	Fair	wTPR·Fair
Sexuality #1	92%	77%	98%	76%	76%	65%	96%	62%	90%	45%	96%	44%
Religion #1	87%	93%	92%	86%	86%	92%	98%	90%	93%	97%	97%	94%
Politics #1	82%	74%	100%	74%	82%	74%	100%	74%	89%	81%	100%	81%
Sexuality #2	50%	66%	68%	45%	52%	54%	98%	53%	42%	45%	100%	45%
Religion #2	57%	70%	59%	41%	57%	51%	97%	49%	39%	48%	99%	48%
Politics #2	60%	64%	93%	59%	72%	69%	93%	64%	47%	48%	79%	38%
Cannabis User	74%	81%	83%	67%	58%	70%	84%	59%	68%	68%	100%	68%
Spam Collection	97%	95%	100%	95%	96%	96%	100%	95%	99%	96%	100%	96%
Cell Review Sentiment	69%	66%	98%	65%	69%	63%	95%	60%	56%	53%	63%	33%
Dermatology	100%	96%	100%	96%	99%	93%	99%	92%	97%	85%	97%	82%
Diabetes	83%	76%	100%	76%	69%	63%	98%	62%	51%	42%	56%	24%
E. Coli	95%	82%	92%	76%	94%	82%	93%	76%	96%	78%	92%	71%
Hepatitis	80%	81%	97%	78%	64%	67%	94%	63%	49%	63%	61%	39%
Hypothyroid	97%	91%	99%	90%	94%	95%	97%	92%	94%	78%	95%	74%
Liver Disorders	72%	65%	100%	65%	66%	66%	88%	58%	57%	60%	43%	26%
Lung Cancer	67%	66%	98%	64%	60%	63%	91%	57%	49%	53%	100%	53%
Lymphography	81%	73%	86%	63%	89%	82%	100%	82%	92%	79%	99%	78%

Table 4: 1*exp* on Base Classifier

Dataset	Logistic Regression				Smooth-kNN				LibSVM			
	wAUC	wTPr	Fair	wTPR·Fair	wAUC	wTPr	Fair	wTPR·Fair	wAUC	wTPr	Fair	wTPR·Fair
Sexuality #1	94%	71%	99%	71%	78%	65%	96%	62%	91%	43%	92%	39%
Religion #1	89%	94%	91%	86%	86%	92%	98%	90%	93%	98%	93%	91%
Politics #1	82%	74%	100%	74%	82%	74%	100%	74%	89%	81%	100%	81%
Sexuality #2	49%	66%	68%	45%	52%	54%	98%	53%	43%	89%	0%	0%
Religion #2	55%	70%	59%	41%	57%	77%	46%	36%	39%	87%	0%	0%
Politics #2	60%	64%	93%	59%	71%	69%	93%	64%	46%	70%	25%	17%
Cannabis User	73%	81%	83%	67%	58%	70%	84%	59%	68%	86%	67%	58%
Spam Collection	97%	95%	100%	95%	96%	98%	99%	97%	99%	97%	100%	97%
Cell Review Sentiment	69%	66%	98%	65%	69%	63%	95%	60%	56%	53%	64%	34%
Dermatology	100%	96%	100%	96%	98%	93%	99%	92%	97%	85%	97%	82%
Diabetes	83%	76%	98%	75%	69%	63%	98%	62%	51%	65%	1%	0%
E. Coli	95%	68%	79%	54%	94%	82%	93%	76%	96%	40%	50%	20%
Hepatitis	80%	81%	93%	76%	64%	67%	88%	59%	49%	77%	0%	0%
Hypothyroid	93%	93%	99%	92%	94%	95%	97%	92%	94%	3%	49%	2%
Liver Disorders	72%	67%	100%	67%	66%	66%	88%	58%	57%	61%	15%	9%
Lung Cancer	67%	66%	98%	64%	60%	63%	91%	57%	49%	66%	50%	33%
Lymphography	81%	73%	86%	63%	89%	82%	100%	82%	92%	44%	88%	39%

3. Sampling Experiments

Table 5: Sampling on Base Classifier

Dataset	Logistic Regression				Smooth-kNN				LibSVM			
	wAUC	wTPr	Fair	wTPR·Fair	wAUC	wTPr	Fair	wTPR·Fair	wAUC	wTPr	Fair	wTPR·Fair
Sexuality #1	88%	75%	98%	73%	74%	61%	97%	59%	87%	64%	96%	62%
Religion #1	84%	92%	94%	86%	85%	90%	98%	89%	81%	4%	0%	0%
Politics #1	71%	67%	100%	67%	79%	71%	100%	71%	89%	80%	100%	80%
Sexuality #2	51%	67%	68%	45%	55%	53%	100%	52%	40%	56%	60%	33%
Religion #2	45%	71%	66%	47%	62%	57%	99%	57%	51%	61%	87%	53%
Politics #2	61%	65%	95%	62%	67%	61%	95%	58%	65%	36%	26%	9%
Cannabis User	74%	84%	86%	73%	57%	69%	85%	59%	70%	72%	99%	72%
Spam Collection	96%	98%	99%	97%	96%	95%	100%	95%	98%	96%	97%	94%
Cell Review Sentiment	71%	66%	96%	64%	67%	62%	96%	59%	62%	59%	99%	59%
Dermatology	99%	95%	100%	95%	97%	86%	98%	85%	97%	86%	98%	85%
Diabetes	82%	75%	100%	75%	65%	59%	100%	59%	53%	65%	1%	1%
E. Coli	93%	77%	92%	71%	92%	76%	92%	70%	96%	80%	90%	72%
Hepatitis	78%	75%	96%	72%	61%	60%	96%	57%	54%	79%	0%	0%
Hypothyroid	97%	96%	99%	95%	96%	93%	98%	91%	90%	93%	35%	33%

Dataset	Logistic Regression				Smooth-kNN				LibSVM			
	wAUC	wTPr	Fair	wTPR·Fair	wAUC	wTPr	Fair	wTPR·Fair	wAUC	wTPr	Fair	wTPR·Fair
Liver Disorders	71%	66%	100%	66%	67%	64%	97%	62%	57%	60%	10%	6%
Lung Cancer	57%	59%	93%	55%	62%	56%	97%	55%	59%	75%	93%	69%
Lymphography	88%	82%	100%	82%	87%	71%	90%	64%	91%	81%	99%	80%

Table 6: 1*lin*+Sampling on Base Classifier

Dataset	Logistic Regression				Smooth-kNN				LibSVM			
	wAUC	wTPr	Fair	wTPR·Fair	wAUC	wTPr	Fair	wTPR·Fair	wAUC	wTPr	Fair	wTPR·Fair
Sexuality #1	88%	74%	98%	72%	73%	58%	97%	57%	86%	27%	21%	6%
Religion #1	84%	91%	94%	86%	85%	87%	99%	86%	79%	4%	0%	0%
Politics #1	71%	67%	100%	67%	79%	71%	100%	71%	89%	79%	98%	77%
Sexuality #2	51%	63%	93%	58%	56%	38%	71%	27%	40%	11%	1%	0%
Religion #2	45%	68%	68%	47%	62%	42%	73%	31%	51%	13%	0%	0%
Politics #2	61%	65%	95%	62%	67%	48%	73%	35%	65%	33%	15%	5%
Cannabis User	74%	84%	87%	72%	57%	50%	98%	49%	70%	16%	20%	3%
Spam Collection	96%	98%	99%	97%	96%	92%	100%	92%	98%	95%	100%	95%
Cell Review Sentiment	71%	66%	96%	64%	67%	62%	96%	59%	62%	59%	100%	59%
Dermatology	99%	95%	100%	95%	97%	86%	98%	84%	97%	86%	98%	84%
Diabetes	82%	68%	92%	63%	65%	55%	82%	45%	53%	65%	5%	3%
E. Coli	93%	76%	92%	70%	92%	76%	92%	69%	96%	75%	86%	65%
Hepatitis	78%	72%	99%	72%	61%	55%	94%	52%	55%	79%	0%	0%
Hypothyroid	97%	88%	99%	87%	96%	89%	99%	88%	90%	92%	64%	59%
Liver Disorders	71%	62%	94%	58%	67%	64%	97%	62%	57%	61%	18%	11%
Lung Cancer	57%	59%	93%	55%	64%	44%	82%	36%	60%	53%	100%	53%
Lymphography	88%	82%	100%	82%	86%	71%	90%	64%	91%	80%	99%	79%

Table 7: 1*inv*+Sampling on Base Classifier

Dataset	Logistic Regression				Smooth-kNN				LibSVM			
	wAUC	wTPr	Fair	wTPR·Fair	wAUC	wTPr	Fair	wTPR·Fair	wAUC	wTPr	Fair	wTPR·Fair
Sexuality #1	87%	70%	98%	69%	72%	56%	98%	55%	79%	2%	0%	0%
Religion #1	84%	91%	94%	86%	85%	87%	99%	86%	79%	4%	0%	0%
Politics #1	71%	67%	100%	67%	79%	71%	100%	71%	89%	79%	98%	77%
Sexuality #2	51%	63%	93%	58%	56%	38%	71%	27%	40%	11%	1%	0%
Religion #2	45%	68%	68%	47%	62%	42%	73%	31%	51%	13%	0%	0%
Politics #2	61%	65%	95%	62%	67%	48%	73%	35%	65%	33%	15%	5%
Cannabis User	74%	84%	87%	72%	57%	50%	98%	49%	70%	16%	20%	3%
Spam Collection	96%	98%	99%	97%	96%	92%	100%	92%	98%	95%	100%	95%
Cell Review Sentiment	71%	66%	96%	64%	67%	62%	96%	59%	62%	59%	100%	59%
Dermatology	99%	95%	100%	94%	97%	83%	98%	82%	97%	83%	98%	81%
Diabetes	82%	68%	92%	63%	65%	55%	82%	45%	53%	65%	5%	3%
E. Coli	92%	59%	76%	44%	91%	68%	92%	63%	91%	32%	32%	10%
Hepatitis	78%	72%	99%	72%	61%	55%	94%	52%	55%	79%	0%	0%
Hypothyroid	97%	85%	98%	83%	96%	89%	99%	88%	86%	74%	78%	57%
Liver Disorders	71%	62%	94%	58%	67%	64%	97%	62%	57%	61%	18%	11%
Lung Cancer	57%	59%	93%	55%	64%	44%	82%	36%	60%	53%	100%	53%
Lymphography	88%	82%	100%	82%	86%	70%	91%	64%	89%	63%	96%	60%

Table 8: 1*exp*+Sampling on Base Classifier

Dataset	Logistic Regression				Smooth-kNN				LibSVM			
	wAUC	wTPr	Fair	wTPR·Fair	wAUC	wTPr	Fair	wTPR·Fair	wAUC	wTPr	Fair	wTPR·Fair
Sexuality #1	88%	70%	99%	69%	73%	56%	98%	55%	88%	2%	0%	0%
Religion #1	88%	91%	94%	86%	85%	87%	99%	86%	82%	4%	0%	0%
Politics #1	72%	67%	100%	67%	79%	71%	100%	71%	89%	79%	99%	78%
Sexuality #2	52%	65%	91%	60%	56%	38%	71%	27%	40%	13%	6%	1%
Religion #2	45%	68%	68%	47%	63%	50%	92%	46%	51%	13%	0%	0%
Politics #2	62%	65%	95%	62%	67%	48%	73%	35%	65%	33%	15%	5%
Cannabis User	74%	84%	86%	73%	57%	50%	98%	49%	70%	28%	47%	13%
Spam Collection	99%	98%	99%	97%	96%	93%	100%	93%	98%	96%	99%	95%
Cell Review Sentiment	71%	66%	96%	64%	67%	62%	96%	59%	62%	59%	100%	59%
Dermatology	100%	95%	100%	94%	97%	83%	98%	82%	97%	82%	98%	80%
Diabetes	82%	71%	96%	68%	65%	55%	82%	45%	53%	66%	4%	2%
E. Coli	92%	57%	82%	47%	92%	69%	92%	63%	96%	18%	19%	3%
Hepatitis	78%	74%	99%	73%	61%	55%	95%	52%	55%	79%	0%	0%
Hypothyroid	98%	85%	98%	83%	96%	89%	99%	88%	90%	35%	41%	14%
Liver Disorders	71%	65%	98%	63%	67%	64%	97%	62%	57%	61%	16%	10%
Lung Cancer	56%	59%	93%	55%	64%	44%	82%	36%	60%	63%	99%	62%
Lymphography	87%	82%	100%	82%	86%	70%	91%	64%	91%	30%	73%	22%

4. SMOTE Experiments

SMOTE cannot run on datasets that contain classes with a single sample, as it requires at least two samples for each class.

Table 9: SMOTE on Base Classifier

Dataset	Logistic Regression				Smooth-kNN				LibSVM			
	wAUC	wTPr	Fair	wTPR·Fair	wAUC	wTPr	Fair	wTPR·Fair	wAUC	wTPr	Fair	wTPR·Fair
Sexuality #1	94%	85%	96%	82%	79%	71%	92%	66%	92%	83%	95%	79%
Religion #1	87%	94%	92%	87%	87%	97%	91%	88%	94%	98%	91%	89%
Politics #1	83%	74%	100%	74%	80%	71%	94%	67%	88%	80%	100%	80%
Sexuality #2	50%	86%	8%	7%	52%	83%	26%	21%	50%	88%	0%	0%
Religion #2	38%	80%	9%	7%	50%	72%	30%	22%	48%	84%	0%	0%
Politics #2	70%	65%	95%	62%	69%	69%	85%	58%	54%	71%	49%	35%
Cannabis User	64%	88%	21%	18%	60%	89%	40%	36%	68%	90%	20%	18%
Spam Collection	99%	99%	100%	99%	97%	97%	99%	96%	98%	96%	100%	96%
Cell Review Sentiment	72%	66%	99%	66%	68%	62%	86%	53%	62%	50%	10%	5%
Dermatology	100%	96%	100%	95%	99%	93%	100%	93%	98%	86%	96%	83%
Diabetes	83%	76%	100%	75%	68%	65%	100%	65%	54%	66%	13%	9%
E. Coli	-	-	-	-	-	-	-	-	-	-	-	-
Hepatitis	81%	81%	86%	69%	68%	66%	76%	50%	58%	79%	13%	10%
Hypothyroid	-	-	-	-	-	-	-	-	-	-	-	-
Liver Disorders	71%	60%	88%	53%	64%	61%	99%	60%	59%	56%	94%	53%
Lung Cancer	62%	72%	94%	68%	75%	75%	99%	74%	73%	72%	67%	48%
Lymphography	-	-	-	-	-	-	-	-	-	-	-	-

Table 10: 1*lin*+SMOTE on Base Classifier

Dataset	Logistic Regression				Smooth-kNN				LibSVM			
	wAUC	wTPr	Fair	wTPR·Fair	wAUC	wTPr	Fair	wTPR·Fair	wAUC	wTPr	Fair	wTPR·Fair
Sexuality #1	94%	85%	96%	81%	79%	70%	94%	66%	92%	82%	95%	78%
Religion #1	87%	93%	94%	87%	87%	90%	99%	89%	94%	93%	100%	92%
Politics #1	83%	74%	100%	73%	80%	71%	94%	66%	88%	80%	100%	79%
Sexuality #2	50%	85%	8%	7%	53%	47%	98%	47%	50%	59%	87%	51%
Religion #2	38%	79%	9%	7%	51%	43%	88%	38%	48%	31%	59%	18%
Politics #2	70%	67%	99%	67%	70%	65%	100%	65%	55%	28%	8%	2%
Cannabis User	63%	87%	21%	18%	59%	65%	92%	60%	68%	71%	99%	71%
Spam Collection	99%	99%	100%	99%	97%	93%	100%	93%	98%	92%	100%	92%
Cell Review Sentiment	72%	66%	99%	66%	68%	62%	86%	53%	62%	50%	10%	5%
Dermatology	100%	96%	100%	95%	99%	93%	99%	92%	98%	86%	96%	83%
Diabetes	83%	69%	93%	65%	69%	57%	79%	45%	54%	37%	6%	2%
E. Coli	-	-	-	-	-	-	-	-	-	-	-	-
Hepatitis	81%	81%	98%	80%	68%	63%	96%	61%	58%	26%	13%	3%
Hypothyroid	-	-	-	-	-	-	-	-	-	-	-	-
Liver Disorders	71%	54%	66%	36%	64%	61%	99%	60%	59%	45%	11%	5%
Lung Cancer	62%	72%	94%	68%	77%	59%	83%	50%	74%	63%	95%	59%
Lymphography	-	-	-	-	-	-	-	-	-	-	-	-

Table 11: *Tlin*+SMOTE on Base Classifier

Dataset	Logistic Regression				Smooth-kNN				LibSVM			
	wAUC	wTPr	Fair	wTPR·Fair	wAUC	wTPr	Fair	wTPR·Fair	wAUC	wTPr	Fair	wTPR·Fair
Sexuality #1	94%	86%	95%	82%	79%	71%	92%	66%	92%	84%	92%	77%
Religion #1	87%	94%	93%	87%	87%	98%	87%	85%	94%	98%	84%	82%
Politics #1	83%	74%	100%	74%	80%	71%	95%	67%	88%	80%	100%	80%
Sexuality #2	50%	85%	8%	7%	52%	81%	26%	21%	49%	84%	17%	14%
Religion #2	38%	79%	9%	7%	49%	68%	52%	35%	47%	63%	73%	46%
Politics #2	70%	67%	99%	67%	69%	69%	85%	58%	51%	61%	65%	40%
Cannabis User	63%	87%	21%	18%	59%	87%	59%	51%	68%	90%	20%	18%
Spam Collection	99%	99%	100%	99%	97%	97%	100%	97%	98%	86%	99%	96%
Cell Review Sentiment	72%	66%	99%	66%	68%	62%	86%	53%	62%	50%	10%	5%
Dermatology	100%	96%	100%	95%	99%	93%	100%	93%	98%	86%	96%	83%
Diabetes	83%	77%	93%	72%	69%	65%	100%	65%	54%	37%	6%	2%
E. Coli	-	-	-	-	-	-	-	-	-	-	-	-
Hepatitis	81%	81%	82%	66%	67%	66%	76%	50%	58%	26%	13%	3%
Hypothyroid	-	-	-	-	-	-	-	-	-	-	-	-
Liver Disorders	72%	64%	99%	64%	64%	61%	99%	60%	59%	45%	11%	5%
Lung Cancer	62%	72%	94%	68%	72%	66%	100%	66%	72%	75%	93%	69%
Lymphography	-	-	-	-	-	-	-	-	-	-	-	-

Table 12: *Tinv*+SMOTE on Base Classifier

Dataset	Logistic Regression				Smooth-kNN				LibSVM			
	wAUC	wTPr	Fair	wTPR·Fair	wAUC	wTPr	Fair	wTPR·Fair	wAUC	wTPr	Fair	wTPR·Fair
Sexuality #1	94%	86%	95%	82%	79%	71%	92%	66%	92%	84%	92%	77%
Religion #1	87%	94%	93%	87%	87%	98%	87%	85%	94%	98%	84%	82%
Politics #1	83%	74%	100%	74%	80%	71%	95%	67%	88%	80%	100%	80%
Sexuality #2	50%	85%	8%	7%	52%	83%	26%	21%	49%	80%	27%	21%
Religion #2	38%	79%	9%	7%	50%	72%	30%	22%	50%	71%	40%	29%
Politics #2	70%	67%	99%	67%	69%	69%	85%	58%	52%	51%	98%	50%
Cannabis User	63%	87%	21%	18%	60%	89%	40%	36%	67%	89%	31%	27%
Spam Collection	98%	99%	100%	99%	96%	97%	99%	97%	98%	97%	99%	96%
Cell Review Sentiment	72%	66%	99%	66%	68%	62%	86%	53%	62%	50%	10%	5%

Dataset												
Dermatology	100%	96%	100%	95%	99%	93%	100%	93%	97%	87%	97%	84%
Diabetes	83%	77%	93%	72%	69%	65%	100%	65%	54%	37%	6%	2%
E. Coli	-	-	-	-	-	-	-	-	-	-	-	-
Hepatitis	81%	81%	82%	66%	66%	70%	49%	35%	58%	26%	13%	3%
Hypothyroid	-	-	-	-	-	-	-	-	-	-	-	-
Liver Disorders	72%	64%	99%	64%	64%	61%	99%	60%	59%	45%	11%	5%
Lung Cancer	61%	72%	94%	68%	73%	66%	100%	66%	72%	72%	94%	68%
Lymphography	-	-	-	-	-	-	-	-	-	-	-	-

Table 13: $Texp$+SMOTE on Base Classifier

Dataset	Logistic Regression				Smooth-kNN				LibSVM			
	wAUC	wTPr	Fair	wTPR·Fair	wAUC	wTPr	Fair	wTPR·Fair	wAUC	wTPr	Fair	wTPR·Fair
Sexuality #1	94%	86%	95%	82%	79%	71%	92%	66%	92%	84%	95%	79%
Religion #1	89%	94%	93%	87%	87%	98%	85%	83%	94%	98%	86%	84%
Politics #1	83%	73%	100%	73%	80%	71%	94%	67%	88%	80%	100%	80%
Sexuality #2	50%	85%	8%	7%	52%	81%	26%	21%	49%	86%	8%	7%
Religion #2	38%	79%	9%	7%	49%	66%	53%	35%	45%	76%	47%	35%
Politics #2	70%	65%	95%	62%	69%	69%	79%	55%	51%	64%	63%	40%
Cannabis User	63%	88%	21%	18%	59%	87%	59%	51%	72%	90%	20%	18%
Spam Collection	99%	99%	100%	99%	97%	97%	100%	97%	98%	97%	99%	96%
Cell Review Sentiment	72%	66%	99%	66%	68%	62%	86%	53%	62%	50%	10%	5%
Dermatology	100%	96%	100%	95%	99%	93%	100%	93%	96%	86%	96%	82%
Diabetes	83%	76%	96%	74%	69%	65%	100%	65%	54%	37%	8%	3%
E. Coli	-	-	-	-	-	-	-	-	-	-	-	-
Hepatitis	79%	81%	82%	66%	67%	66%	76%	50%	58%	26%	17%	4%
Hypothyroid	-	-	-	-	-	-	-	-	-	-	-	-
Liver Disorders	71%	64%	98%	62%	64%	61%	99%	60%	59%	45%	12%	5%
Lung Cancer	61%	72%	94%	68%	72%	66%	100%	66%	72%	75%	93%	69%
Lymphography	-	-	-	-	-	-	-	-	-	-	-	-