Appendix D. Comparison of classification algorithms using data preprocessing techniques in binary class models from the perspective of evaluation metrics for the Nemenyi test.

In general, the XBGBoost (eXtreme Gradient-Boosting) algorithm was the one that presented the best results in all analyzed metrics. XBGBoost showed a statistically significant difference from the perspective of all metrics. XBGBoost maintained the first and second most significant distances compared to other algorithms. The LightGBM (Light Gradient-Boosting Model) obtained the third most considerable distance, ranking second as the best algorithm from the perspective of all metrics. RF (Random Forest) and kNN (k-Nearest Neighbors) were always within the margin of the critical distance limit and did not show significant statistical differences. RF had a slight advantage over kNN when compared. Next, the classification algorithms' detailed results are presented from each analyzed metric's perspective.

Precision (Table D.1): the XGBoost algorithm presented a significant statistical difference compared to the kNN and RF algorithms, with the first and second largest distances, with 54% and 29% above the critical margin limit, respectively. LightGBM, with the third most significant distance, also showed a difference compared to kNN, with 17% above the critical distance value. The comparison between LightGBM and RF showed a distance within the critical limit margin, with 92% of the critical distance value. In the other comparisons, all values were within the critical limit margin, varying between 17% and 37% of the critical distance value.

Critical Distance: 1.35						
Group 1	**Ranking**	**Group 2**	**Ranking**	**Distance**	**(%)**	**Hypothesis**
XGB	1.42	LGBM	1.92	0.50	37%	Equal
XGB	1.42	RF	3.17	1.75	129%	Different
XGB	1.42	KNN	3.50	2.08	154%	Different
LGBM	1.92	RF	3.17	1.25	92%	Equal
LGBM	1.92	KNN	3.50	1.58	117%	Different
RF	3.17	KNN	3.50	0.33	24%	Equal

Table D.1: Comparison of classification algorithms using data preprocessing techniques in binary class models from the perspective of **Precision metric** for the Nemenyi test.

Recall (Table D.2): the XGBoost algorithm showed a significant statistical difference compared to the kNN and RF algorithms, with the first and second most significant distances, with 57% and 38% above the critical limit margin, respectively. LightGBM, with the third largest distance, also showed a difference compared to kNN, with 20% above the critical distance value. The comparison between LightGBM and RF showed a slightly significant difference, with 1% above the critical distance value. In the other comparisons, all values were within the critical limit margin, varying between 18% and 37% of the critical distance value.

Critical Distance: 1.35						
Group 1	**Ranking**	**Group 2**	**Ranking**	**Distance**	**(%)**	**Hypothesis**
XGB	1.38	LGBM	1.88	0.50	37%	Equal
XGB	1.38	RF	3.25	1.87	138%	Different
XGB	1.38	KNN	3.50	2.12	157%	Different
LGBM	1.88	RF	3.25	1.37	101%	Different
LGBM	1.88	KNN	3.50	1.62	120%	Different
RF	3.25	KNN	3.50	0.25	18%	Equal

Table D.2: Comparison of classification algorithms using data preprocessing techniques in binary class models from the perspective of **Recall metric** for the Nemenyi test.

F1-Score (Table D.3): the XGBoost algorithm showed a significant statistical difference compared to the RF and kNN algorithms, with the first and second highest distances, with 72% and 29% above the critical limit margin, respectively. The LightGBM, with the third most significant distance, showed a difference when compared to the RF, with 54% above the critical distance value. The comparison between LightGBM and kNN showed a considerable difference, with 11% above the critical distance value. In the other comparisons, all values were within the critical limit margin, varying between 18% and 43% of the critical distance value.

Critical Distance: 1.35						
Group 1	**Ranking**	**Group 2**	**Ranking**	**Distance**	**(%)**	**Hypothesis**
XGB	1.42	LGBM	1.67	0.25	18%	Equal
XGB	1.42	KNN	3.17	1.75	129%	Different
XGB	1.42	RF	3.75	2.33	172%	Different
LGBM	1.67	KNN	3.17	1.50	111%	Different
LGBM	1.67	RF	3.75	2.08	154%	Different
KNN	3.17	RF	3.75	0.58	43%	Equal

Table D.3: Comparison of classification algorithms using data preprocessing techniques in binary class models from the perspective of **F1-Score metric** for the Nemenyi test.

FAR (Table D.4): the XGBoost algorithm showed a significant statistical difference compared to the kNN and RF algorithms, with the first and second largest distances, with 42% and 17% above the critical limit margin, respectively. LightGBM, with the third most significant distance, showed a difference when compared to kNN, with only 5% above the critical distance value. The comparison between LightGBM and RF showed a distance within the critical limit margin, with 80% of the critical distance value. In the other comparisons, all values were within the critical limit margin, varying between 25% and 37% of the critical distance value.

Critical Distance: 1.35						
Group 1	**Ranking**	**Group 2**	**Ranking**	**Distance**	**(%)**	**Hypothesis**
XGB	1.50	LGBM	2.00	0.50	37%	Equal
XGB	1.50	RF	3.08	1.58	117%	Different
XGB	1.50	KNN	3.42	1.92	142%	Different
LGBM	2.00	RF	3.08	1.08	80%	Equal
LGBM	2.00	KNN	3.42	1.42	105%	Different
RF	3.08	KNN	3.42	0.34	25%	Equal

Table D.4: Comparison of classification algorithms using data preprocessing techniques in binary class models from the perspective of **FAR metric** for the Nemenyi test.

Appendix E. Comparison of classification algorithms using data preprocessing techniques in multiclass models of specific Group 1 attacks from the perspective of evaluation metrics for the Nemenyi test.

In general, XGBoost (eXtreme Gradient-Boosting) was the algorithm that presented the best overall results in all metrics. In the Precision and FAR metrics, RF (Random Forest) achieved similar performances to XGBoost, getting the second best overall performance. The kNN (k-Nearest Neighbors) showed better results than the LightGBM (Light Gradient-Boosting Model) in detecting the specific attacks of Reconnaissance (UNSW-NB15) and Portscan (CIC-IDS2017). LightGBM was the algorithm that presented the worst results. Next, the classification algorithms' detailed results are presented from each analyzed metric's perspective.

Precision (Table E.1): the XGBoost and RF algorithms presented significant statistical differences when compared to the LightGBM and kNN algorithms, presenting the same distance values concerning the critical limit margin, with 33% and 26% above the value of the critical distance, for the first and second largest distances, respectively. In the other comparisons, all values were within the critical limit margin, varying between 0% and 7% of the critical distance value.

Critical Distance: 1.35						
Group 1	**Ranking**	**Group 2**	**Ranking**	**Distance**	**(%)**	**Hypothesis**
RF	1.62	XGB	1.62	0.00	0%	Equal
RF	1.62	KNN	3.33	1.71	126%	Different
RF	1.62	LGBM	3.42	1.80	133%	Different
XGB	1.62	KNN	3.33	1.71	126%	Different
XGB	1.62	LGBM	3.42	1.80	133%	Different
KNN	3.33	LGBM	3.42	0.09	7%	Equal

Table E.1: Comparison of classification algorithms using data preprocessing techniques in multiclass models of specific Group 1 attacks from the perspective of **Precision metric** for the Nemenyi test.

Recall (Table E.2): the XGBoost algorithm presented a significant statistical difference when compared to the LightGBM algorithm, with the first greater distance, 35% above the critical limit margin. The kNN with the second largest distance showed a significant difference when compared to LightGBM, with 29% above the critical distance value. The third largest distance was with RF compared to LightGBM, which showed a significant difference with only 5% above the critical distance value. In the other comparisons, all values were within the critical limit margin, varying between 6% and 30% of the critical distance value.

Critical Distance: 1.35						
Group 1	**Ranking**	**Group 2**	**Ranking**	**Distance**	**(%)**	**Hypothesis**
XGB	1.92	KNN	2.00	0.08	6%	Equal
XGB	1.92	RF	2.33	0.41	30%	Equal
XGB	1.92	LGBM	3.75	1.83	135%	Different
KNN	2.00	RF	2.33	0.33	24%	Equal
KNN	2.00	LGBM	3.75	1.75	129%	Different
RF	2.33	LGBM	3.75	1.42	105%	Different

Table E.2: Comparison of classification algorithms using data preprocessing techniques in multiclass models of specific Group 1 attacks from the perspective of **Recall metric** for the Nemenyi test.

F1-Score (Table E.3): the XGBoost algorithm presented a significant statistical difference compared to the LightGBM algorithm, with the first greater distance, with 60% above the critical limit margin. The second and third largest distances were within the critical limit margin. They did not present significant differences, with 92% of the critical distance value, when comparing XGBoost with kNN and RF with LightGBM, respectively. In the other comparisons, all values were within the critical limit margin, varying between 25% and 67% of the critical distance value.

Critical Distance: 1.35						
Group 1	**Ranking**	**Group 2**	**Ranking**	**Distance**	**(%)**	**Hypothesis**
XGB	1.42	RF	2.33	0.91	0,67	Equal
XGB	1.42	KNN	2.67	1.25	0,92	Equal
XGB	1.42	LGBM	3.58	2.16	1,60	Different
RF	2.33	KNN	2.67	0.34	0,25	Equal
RF	2.33	LGBM	3.58	1.25	0,92	Equal
KNN	2.67	LGBM	3.58	0.91	0,67	Equal

Table E.3: Comparison of classification algorithms using data preprocessing techniques in multiclass models of specific Group 1 attacks from the perspective of **F1-Score metric** for the Nemenyi test.

FAR (Table E.4): the XGBoost and RF algorithms showed significant statistical differences when compared to the LightGBM and kNN algorithms, showing the same distances concerning the critical limit margin, with 32% and 14% above the critical distance value, for the first and second largest distances, respectively. In the other comparisons, all values were within the critical limit margin, varying between 0% and 18% of the critical distance value.

Critical Distance: 1.35						
Group 1	**Ranking**	**Group 2**	**Ranking**	**Distance**	**(%)**	**Hypothesis**
XGB	1.54	RF	1.79	0.25	18%	Equal
XGB	1.54	LGBM	3.33	1.79	132%	Different
XGB	1.54	KNN	3.33	1.79	132%	Different
RF	1.79	LGBM	3.33	1.54	114%	Different
RF	1.79	KNN	3.33	1.54	114%	Different
LGBM	3.33	KNN	3.33	0.00	0%	Equal

Table E.4: Comparison of classification algorithms using data preprocessing techniques in multiclass models of specific Group 1 attacks from the perspective of **FAR metric** for the Nemenyi test.

Appendix F. Comparison of classification algorithms using data preprocessing techniques in multiclass models of specific Group 2 attacks from the perspective of evaluation metrics for the Nemenyi test.

In general, XGBoost (eXtreme Gradient-Boosting) was the algorithm that presented the best results, obtaining the first three most considerable distances from the perspective of the Recall, F1-Score, and FAR metrics. RF (Random Forest) had the second most significant distance in overall performance, standing out in the Precision metric. kNN (k-Nearest Neighbors) performed better than LightGBM (Light Gradient-Boosting Model) in detecting specific DoS (UNSW-NB15) and DoS/DDoS (CIC-IDS2017) attacks. LightGBM was the algorithm that presented the worst results. Next, the classification algorithms' detailed results are presented from each analyzed metric's perspective.

Precision (Table F.1): the RF algorithm showed a significant statistical difference when compared to the LightGBM algorithm, with the first greater distance, 41% above the critical limit margin. The XGBoost, with the second largest distance, showed a difference compared to LightGBM, with 26% above the critical distance value. The third largest distance is kNN compared to LightGBM, which presented a distance within the critical limit margin, with 96% of the critical distance value. In the other comparisons, all values were within the critical limit margin, varying between 15% and 45% of the critical distance value.

Critical Distance: 1.10						
Group 1	**Ranking**	**Group 2**	**Ranking**	**Distance**	**(%)**	**Hypothesis**
RF	1.94	XGB	2.11	0.17	15%	Equal
RF	1.94	KNN	2.44	0.50	45%	Equal
RF	1.94	LGBM	3.50	1.56	141%	Different
XGB	2.11	KNN	2.44	0.33	30%	Equal
XGB	2.11	LGBM	3.50	1.39	126%	Different
KNN	2.44	LGBM	3.50	1.06	96%	Equal

Table F.1: Comparison of classification algorithms using data preprocessing techniques in multiclass models of specific Group 2 attacks from the perspective of **Precision metric** for the Nemenyi test.

Recall (Table F.2): the XGBoost algorithm presented a significant statistical difference when compared to the LightGBM algorithm, with the first greater distance, with 46% above the critical limit margin. The kNN with the second largest distance showed a difference compared to LightGBM, with 41% above the critical distance value. The third largest distance is with the RF, compared to the LGBM, which presented a distance within the critical limit margin, with 96% of the critical distance value. In the other comparisons, all values were within the critical limit margin, varying between 5% and 51% of the critical distance value.

Critical Distance: 1.10						
Group 1	**Ranking**	**Group 2**	**Ranking**	**Distance**	**(%)**	**Hypothesis**
XGB	1.94	KNN	2.00	0.06	5%	Equal
XGB	1.94	RF	2.50	0.56	51%	Equal
XGB	1.94	LGBM	3.56	1.62	146%	Different
KNN	2.00	RF	2.50	0.50	45%	Equal
KNN	2.00	LGBM	3.56	1.56	141%	Different
RF	2.50	LGBM	3.56	1.06	96%	Equal

Table F.2: Comparison of classification algorithms using data preprocessing techniques in multiclass models of specific Group 2 attacks from the perspective of **Recall metric** for the Nemenyi test.

F1-Score (Table F.3): the XGBoost algorithm presented a significant statistical difference compared to the LightGBM algorithm, with the first greater distance and 51% above the critical limit margin. The kNN with the second largest distance showed a difference compared to LightGBM, with 40% above the critical distance value. The third largest distance is with the RF compared to LightGBM, which showed a statistically significant difference, with 10% above the value of the critical distance. In the other comparisons, all values were within the critical limit margin, varying between 11% and 41% of the critical distance value.

Critical Distance: 1.10						
Group 1	**Ranking**	**Group 2**	**Ranking**	**Distance**	**(%)**	**Hypothesis**
XGB	1.94	KNN	2.06	0.12	11%	Equal
XGB	1.94	RF	2.39	0.45	41%	Equal
XGB	1.94	LGBM	3.61	1.67	151%	Different
KNN	2.06	RF	2.39	0.33	30%	Equal
KNN	2.06	LGBM	3.61	1.55	140%	Different
RF	2.39	LGBM	3.61	1.22	110%	Different

Table F.3: Comparison of classification algorithms using data preprocessing techniques in multiclass models of specific Group 2 attacks from the perspective of **F1-Score metric** for the Nemenyi test.

FAR (Table F.4): the XGBoost algorithm presented the most significant statistical differences compared to the LightGBM algorithm, having the first, second, and third largest distances, with 86%, 51%, and 26% above the critical limit margin, respectively. In the other comparisons, all values were within the critical limit margin, varying between 25% and 61% of the critical distance value.

Critical Distance: 1.10						
Group 1	**Ranking**	**Group 2**	**Ranking**	**Distance**	**(%)**	**Hypothesis**
XGB	1.22	RF	2.61	1.39	126%	Different
XGB	1.22	KNN	2.89	1.67	151%	Different
XGB	1.22	LGBM	3.28	2.06	186%	Different
RF	2.61	KNN	2.89	0.28	25%	Equal
RF	2.61	LGBM	3.28	0.67	61%	Equal
KNN	2.89	LGBM	3.28	0.39	35%	Equal

Table F.4: Comparison of classification algorithms using data preprocessing techniques in multiclass models of specific Group 2 attacks from the perspective of **FAR metric** for the Nemenyi test.