SMOClust: Synthetic Minority Oversampling based on Stream Clustering for Evolving Data Streams - Supplementary Docuement

Chun Wai Chiu[1*] and Leandro L. Minku[1*]

[1*]School of Computer Science, University of Birmingham, Edgbaston, Birmingham, B15 2TT, United Kingdom.

*Corresponding author(s). E-mail(s): cxc1015@student.bham.ac.uk; l.l.minku@bham.ac.uk;

1 General

This is a supplementary document to the paper of "SMOClust: Synthetic Minority Oversampling based on Stream Clustering for Evolving Data Streams" and it is organised as follows:

- Section 2 presents the pseudo-code of Synthetic Minority Oversampling based on stream Clustering (SMOClust) written in a lower level of abstraction.
- Sections 3 and 4 presents the comprehensive results of the predictive performance of approaches on artificial data streams and real-world data streams respectively.

2 Proposed Approach

This section presents the pseudo-code of SMOClust written in a lower level of abstraction. Algorithm 1 presents the pseudo-code over-viewing SMOClust. Algorithm 2 presents the pseudo-code of this method. The details of generating a synthetic minority class example using micro-clusters can be described as follows. Algorithm 3 presents the details of how to combine a set of micro-clusters into one. Algorithm 4 presents the algorithm of sampling from a hypersphere based on a skewed Gaussian where the maximum of the probability density function is predefined.

Algorithm 1 Synthetic Minority Oversampling based on Stream Clustering - SMOClust

Hyper-parameters: Base Learner(b), Stream Clustering Method(sc), Class Size Fading Factor(θ), Gaussian Noise Variance(v), Categorical Change Probability(P_c), k-Nearest neighbour(k), Data Stream(S)
Variables: Base Learner(B), Stream Clustering Methods array($SC[]$)

$SC[] \leftarrow createStreamClusteringMethods(sc, 2)$ ▷ "2" refers to the number of SC to create.
for $s_t \in S$ **do**
  $drift\_level \leftarrow DriftDetection(B, s_t)$
  if $drift\_level == DRIFT$ **then**
    $B \leftarrow createNewBaseLearner(b)$; $B.resetClassSize()$
  end if
  $B.trainOnInstance(s_t)$; $B.updateClassSize(s_t, \theta)$
  $last\_inst[s_t.classValue()] \leftarrow s_t$ ▷ Store the last seen example of each class
  $c_{maj} \leftarrow getMajorityClass()$; $c_{min} \leftarrow getMinorityClass()$
  while $(B.rawClassSize(c_{min}) < B.rawClassSize(c_{maj}))$ **and** $((SC[0].isReady()$ **and** $SC[1].isReady())$ **or** $last\_inst[c_{min}] \neq NULL)$ **do**
    if $SC[0].isReady()$ **and** $SC[1].isReady()$ **then**
      $mCluster_{anchor} \leftarrow weightedRandomDrawByAvgTimeStamp(SC[c_{min}])$
      if $mCluster_{anchor}.surroundedBySameClass(SC)$ **then**
        $synthInst^{Bin} \leftarrow genSynthInstBykNN(SC[c_{min}], mCluster_{anchor}, c_{min}, k)$ ▷ Alg. 2
      else
        $synthInst^{Bin} \leftarrow genSynthInstByGauSampling(mCluster_{anchor})$
      end if
      $synthInst \leftarrow binaryToNominal(synthInst^{Bin}.copy())$
      $SC[c_{min}].trainOnInstance(synthInst^{Bin}.deleteClassAttribute())$
      $B.trainOnInstance(synthInst)$; $B.updateClassSize(synthInst, \theta)$
    else ▷ **if** $last\_inst[c_{min}] \neq NULL$
      $synthInst \leftarrow addGaussianNoiseToInstance(last\_inst[c_{min}], v, P_c)$
      $synthInst^{Bin} \leftarrow nominalToBinary(synthInst.copy())$
      $SC[c_{min}].trainOnInstance(synthInst^{Bin}.deleteClassAttribute())$
      $B.trainOnInstance(synthInst)$; $B.updateClassSize(synthInst, \theta)$
    end if
  end while
  $s_t^{Bin,noClass} \leftarrow nominalToBinary(s_t.copy())$
  $SC[s_t.classValue()].trainOnInstance(s_t^{Bin,noClass}.deleteClassAttribute())$
end for

Algorithm 2 Generate Synthetic Instance with k-NN Micro-Clusters

function GENSYNTHINSTBYKNN($SC[c_{min}]$, $mCluster_{anchor}$, c_{min}, k)
  $kNNmClusters \leftarrow SC[c_{min}].getkNNmClusters(mCluster_{anchor})$
  $sphere\_cluster \leftarrow createSphereCluster(mCluster_{anchor}, kNNmClusters)$
  $synthInst \leftarrow sphere\_cluster.sample\_around\_target(anchor_mCluster.getCentre())$ ▷ Alg. 4
  $synthInst.setClassValue(s_t.classValue())$
  return $synthInst$
end function

Algorithm 3 Combining a set of micro-clusters into one

function COMBINE($mClusters[]$)
  $dimensions \leftarrow mClusters[0].numOfDimensionss()$
  for $i \in range(0..mClusters.length)$ **do**
    $all\_centres[i] \leftarrow mClusters[i].getCentre()$
    $all\_weights[i] \leftarrow mClusters[i].getWeight()$
    $all\_radius[i] \leftarrow mClusters[i].getRadius()$
  end for
  $newCentre \leftarrow createArrayWithSize(dimensions)$
  for $i \in range(0..dimenstions)$ **do** ▷ Weighted sum of centres, by dimension.
    $result\_by\_dim \leftarrow 0$
    for $j \in range(0..mClusters.length)$ **do**
      $result\_by\_dim \leftarrow result\_by\_dim + all\_centres[j][i] * all\_weights[j]$
    end for
    $newCentre[i] = result\_by\_dim/sum(all\_weights)$
  end for
  $r_n \leftarrow createArrayWithSize(all\_radius.length)$
  for $i \in range(0..all\_radius.length)$ **do** ▷ Find the distance from $newCentre$ to farthest hull.
    $distance\_to\_newCentre \leftarrow euclidean\_distance(all\_centres[i], newCentre)$
    $r_n \leftarrow r_n \cup (all\_radius[i] + abs(distance\_to\_newCentre))$
  end for
  $r_n \leftarrow r_n.sort(descending);\ new\_radius \leftarrow r_n[0]$
  return createMicroCluster(newCentre, newRadius)
end function

Algorithm 4 Sampling from a Hyper-Sphere by Skewed Gaussian with the Maximum of the Probability Density Function at a Designated Location

function SAMPLE_AROUND_TARGET($\alpha^{(1)}$, $sphere\_cluster$)
  $\beta \leftarrow sphere\_cluster.getCentre()$
  $r \leftarrow sphere\_cluster.getRadius()$
  $dimensions \leftarrow \beta.numOfDimensions()$
  $\delta \leftarrow createArrayWithSize(dimensions)$
  $\gamma \leftarrow createArrayWithSize(dimensions)$
  $\alpha^{(2)} \leftarrow sample\_random\_from\_hypersphere(\alpha^{(1)}, 1)$ ▷ By Muller's Method [?]
  $A \leftarrow 0;\ B \leftarrow 0;\ C \leftarrow 0$
  for $i \in range(0..dimensions)$ **do**
    $\delta[i] \leftarrow \alpha^{(2)}[i] - \alpha^{(1)}[i]$
    $\gamma[i] \leftarrow \beta[i] - \alpha^{(1)}[i]$
    $A \leftarrow A + (\delta[i] * \delta[i])$ ▷ $A = \sum_{i=0}^{n} \delta_i^2$
    $B \leftarrow B + (\delta[i] * \gamma[i])$ ▷ $\sum_{i=0}^{n} \delta_i \gamma_i$
    $C \leftarrow C + (\gamma[i] * \gamma[i])$ ▷ $\sum_{i=0}^{n} \gamma_i^2$
  end for
  $B \leftarrow B * -2$ ▷ $B = -2(\sum_{i=0}^{n} \delta_i \gamma_i)$
  $C \leftarrow C - (r * r)$ ▷ $C = (\sum_{i=0}^{n} \gamma_i^2) - r^2$
  return $(-B + sqrt(B * B - 4 * A * C))/(2 * A)$ ▷ $\frac{-B+\sqrt{B^2-4AC}}{2A}$
end function

3 Results with Artificial Data Streams

This section resents the comprehensive results of the predictive performance of approaches on artificial data streams.

- Correspond to Figure 4 in the paper:
  - Figure 1 presents the difference of the thirty runs average G-Mean of the compared approaches against SMOClust on **five**-dimensional class imbalanced artificial data streams.
  - Table 3 presents the thirty runs average G-Mean of all approaches on **five** dimensional class imbalanced artificial data streams and the A12 effect size results of comparing existing approaches against SMOClust.
- Correspond to Figure 5 in the paper:
  - Figure 2 presents the difference of the thirty runs average G-Mean of the compared approaches against SMOClust on **five**-dimensional severely class imbalanced artificial data streams.
  - Table 4 presents the thirty runs average G-Mean of all approaches on **five** dimensional severely class imbalanced artificial data streams and the A12 effect size results of comparing existing approaches against SMOClust.
- Correspond to Figure 14 in the paper:
  - Figure 3 presents the difference of the thirty runs average G-Mean of the compared approaches against SMOClust on **two**-dimensional class imbalanced artificial data streams.
  - Table 5 presents the thirty runs average G-Mean of all approaches on **two** dimensional class imbalanced artificial data streams and the A12 effect size results of comparing existing approaches against SMOClust.
- Correspond to Figure 15 in the paper:
  - Figure 4 presents the difference of the thirty runs average G-Mean of the compared approaches against SMOClust on **two**-dimensional severely class imbalanced artificial data streams.
  - Table 6 presents the thirty runs average G-Mean of all approaches on **two** dimensional severely class imbalanced artificial data streams and the A12 effect size results of comparing existing approaches against SMOClust.
- Correspond to Tables 7 in the paper:
  - Table 1 presents the thirty runs average G-Mean of the approaches on the two-dimensional version of StaticIm1_Move7 stream and the A12 effect size results of comparing existing approaches against SMOClust.
- Correspond to Tables 8 in the paper:
  - Table 2 presents the thirty runs average G-Mean of the approaches on the two-dimensional version of StaticIm10_Rare100 stream and the A12 effect size results of comparing existing approaches against SMOClust.

- Correspond to Figure 6 in the paper:
  - Figure 5 compares the predictive performance of SMOClust with that of $\mathrm{OOB_d}$, $\mathrm{UOB_d}$, $\mathrm{oOS_d}$, and $\mathrm{oUnderOverB_d}$ in the StaticIm1_Move7 stream. The comparison is made over time steps in the median run[1] of the approaches.
- Correspond to Figure 10 in the paper:
  - Figure 6 compares the predictive performance of SMOClust with that of $\mathrm{OOB_d}$, $\mathrm{UOB_d}$, $\mathrm{oOS_d}$, and $\mathrm{oUnderOverB_d}$ in the StaticIm10_Rare100 stream. The comparison is made over time steps in the median run of the approaches.

[1]Median run refers to the run that leads to the median of predictive performances averaged across time steps.

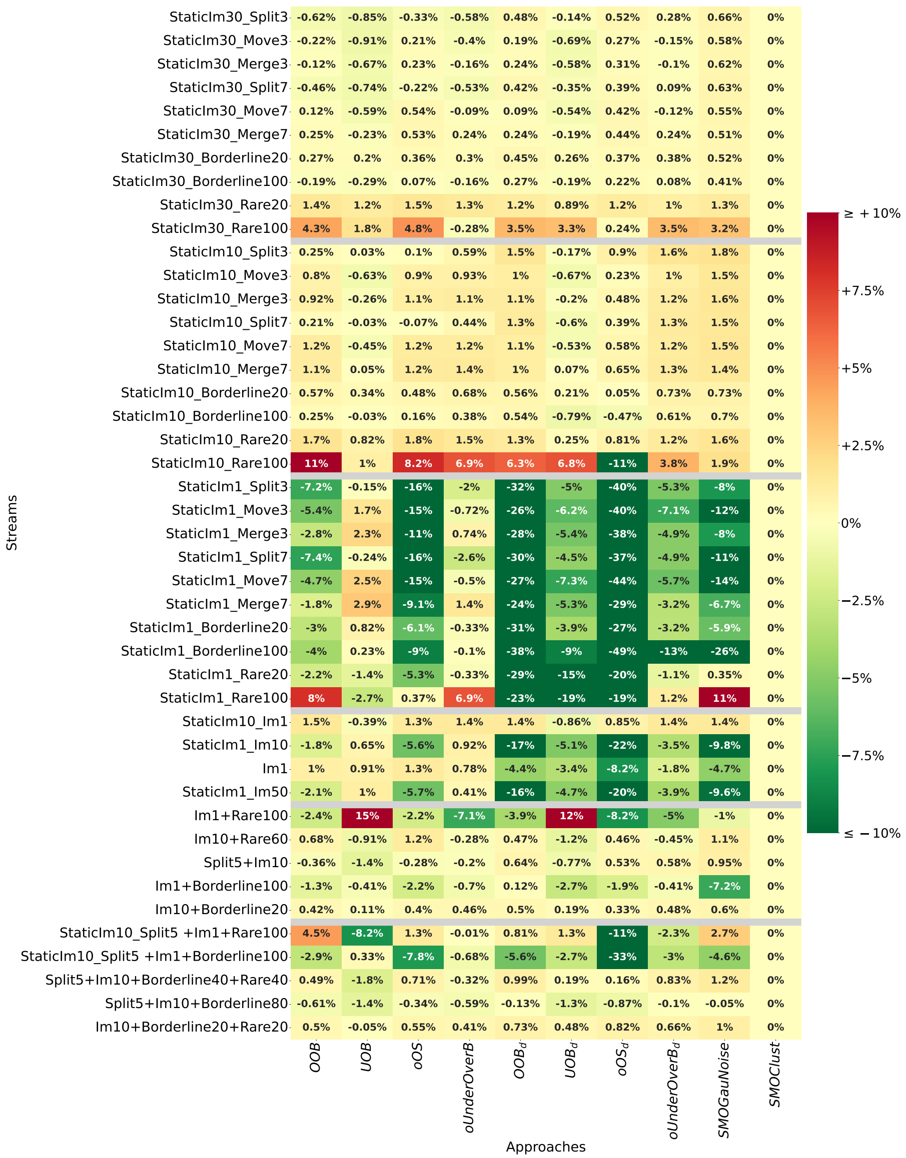

Fig. 1: 30 Runs Average G-Mean on Class Imbalanced Artificial Data Streams (Difference against SMOClust; Green cells indicate SMOClust performed better; Red cells indicate SMOClust performed worse)

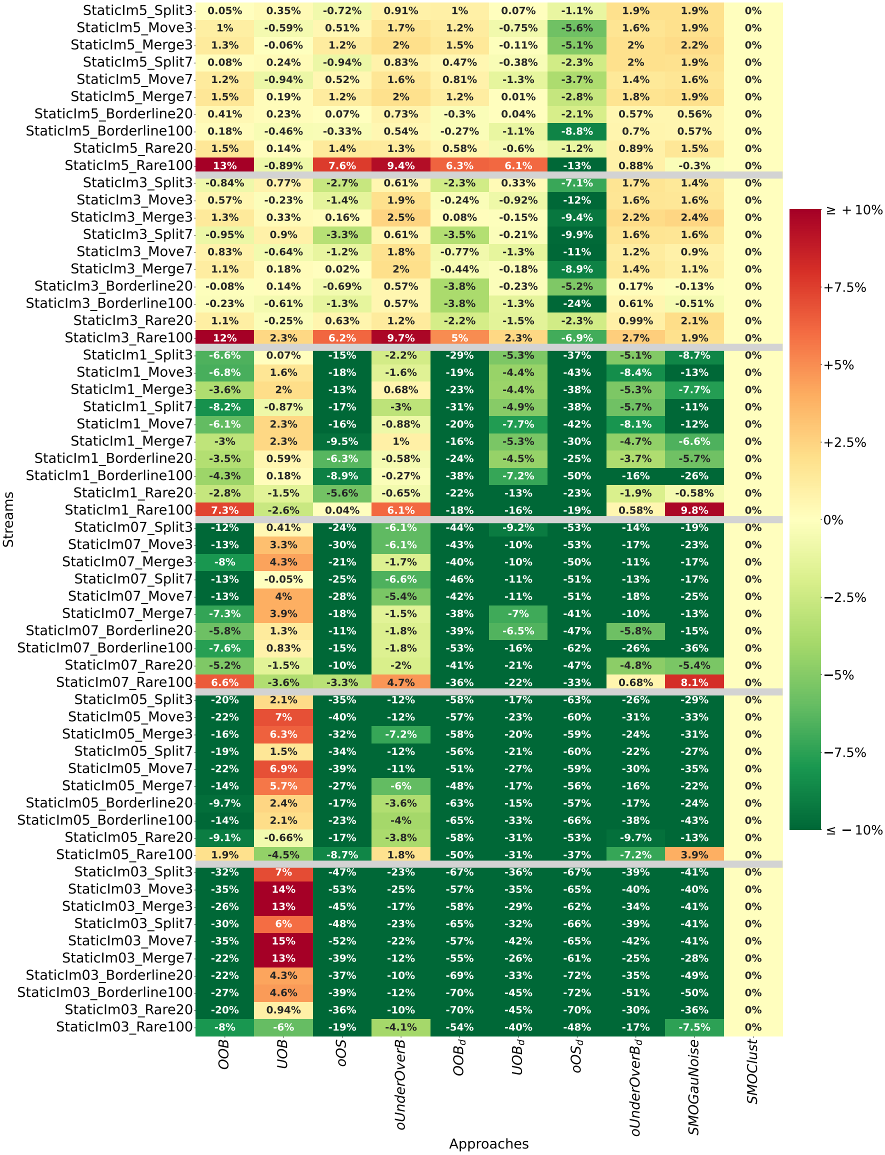

Fig. 2: 30 Runs Average G-Mean on Severely Class Imbalanced Artificial Data Streams (Difference against SMOClust; Green cells indicate SMOClust performed better; Red cells indicate SMOClust performed worse)

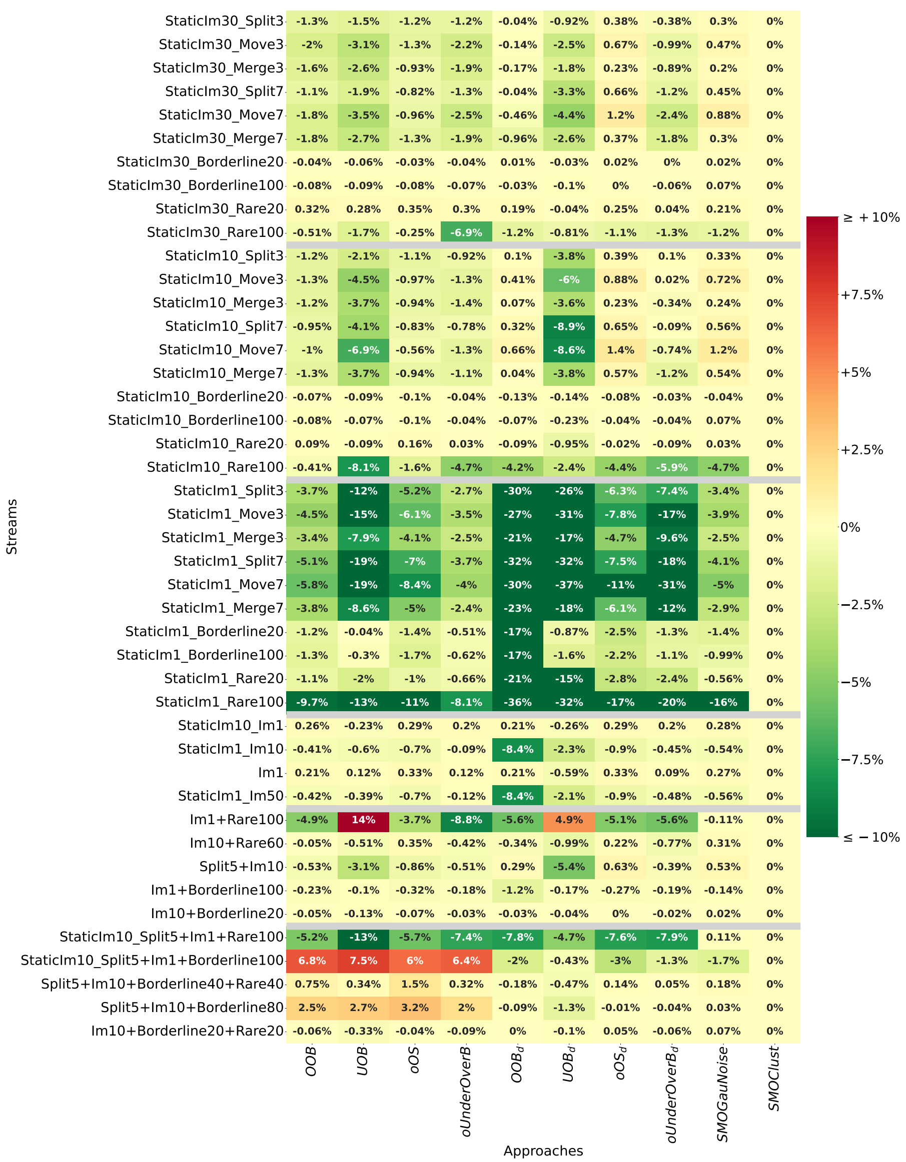

Fig. 3: 30 Runs Average G-Mean on 2D Class Imbalanced Artificial Data Streams (Difference against SMOClust; Green cells indicate SMOClust performed better; Red cells indicate SMOClust performed worse)

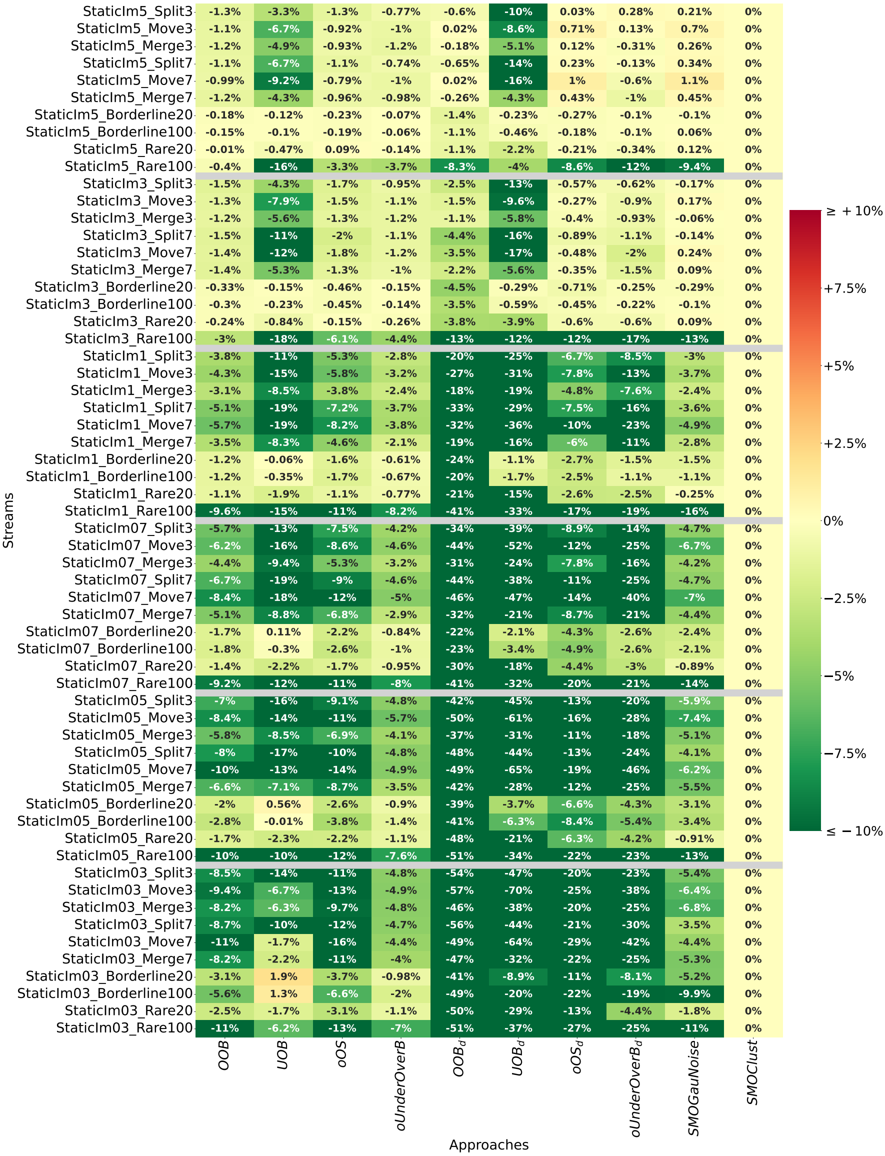

Fig. 4: 30 Runs Average G-Mean on 2D Severely Class Imbalanced Artificial Data Streams (Difference against SMOClust; Green cells indicate SMOClust performed better; Red cells indicate SMOClust performed worse)

Table 1: 30 Runs Average G-Mean on Two Dimensional Version of Representative Artificial Data Streams where SMOClust Performed Better (A12 SMOClust vs Others)

Stream	OOB	UOB	oOS	oUnderOverB	OOB_d
StaticIm1_Move7	82.11%[-b]	76.3%[-b]	79.46%[-b]	85.26%[-b]	53.45%[-b]
Stream	**UOB_d**	**oOS_d**	**$oUnderOverB_d$**	**SMOGauNoise**	**SMOClust**
StaticIm1_Move7	56.94%[-b]	76.88%[-b]	45.12%[-b]	82.94%[-b]	91.23%

- Based on the average G-Mean, cells are highlighted in lime / light grey when SMOClust performed better than the corresponding approach and cells are highlighted in orange / dark grey cells when SMOClust performed worse than the corresponding approach. The colour intensity scales with the absolute difference of average G-Mean between the SMOClust and the approach of the column and the intensity reaches the maximum when such difference is $\geq$ 10%.
- Symbols [*], [s], [m] and [b] represent insignificant, small, medium and large A12 effect size against SMOClust. Presence/absence of the sign "-" in the effect size means that the corresponding approach was worse/better than SMOClust.

Table 2: 30 Runs Average G-Mean on Two-Dimensional Version of Representative Artificial Data Streams where SMOClust Performed Worse (A12 SMOClust vs Others)

Stream	OOB	UOB	oOS	oUnderOverB	OOB_d
StaticIm10_Rare100	70.61%	63.65%	69.14%	68.19%	65.49%
Stream	**UOB_d**	**oOS_d**	**$oUnderOverB_d$**	**SMOGauNoise**	**SMOClust**
StaticIm10_Rare100	68.17%	65.04%	64.98%	64.56%	70.32%

- Based on the average G-Mean, cells are highlighted in lime / light grey when SMOClust performed better than the corresponding approach and cells are highlighted in orange / dark grey cells when SMOClust performed worse than the corresponding approach. The colour intensity scales with the absolute difference of average G-Mean between the SMOClust and the approach of the column and the intensity reaches the maximum when such difference is $\geq$ 10%.

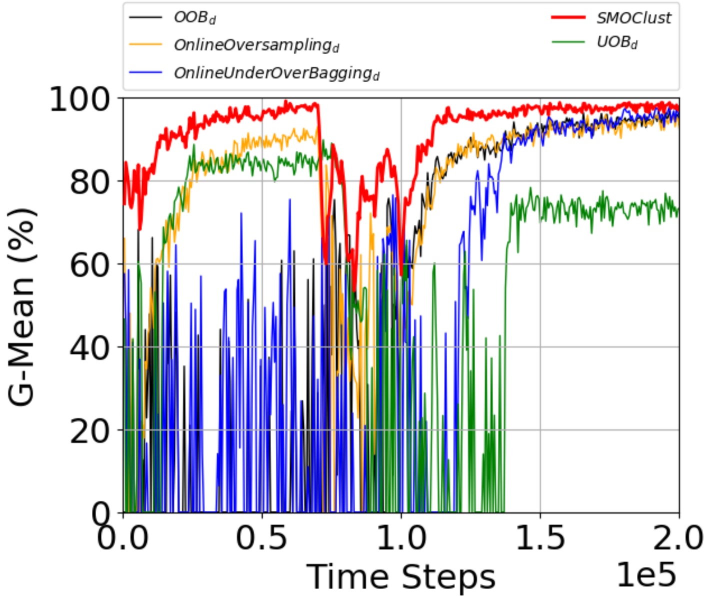

Fig. 5: Periodic Class Balanced Holdout Test G-Mean Against Time Steps in Two-Dimensional StaticIm1_Move7

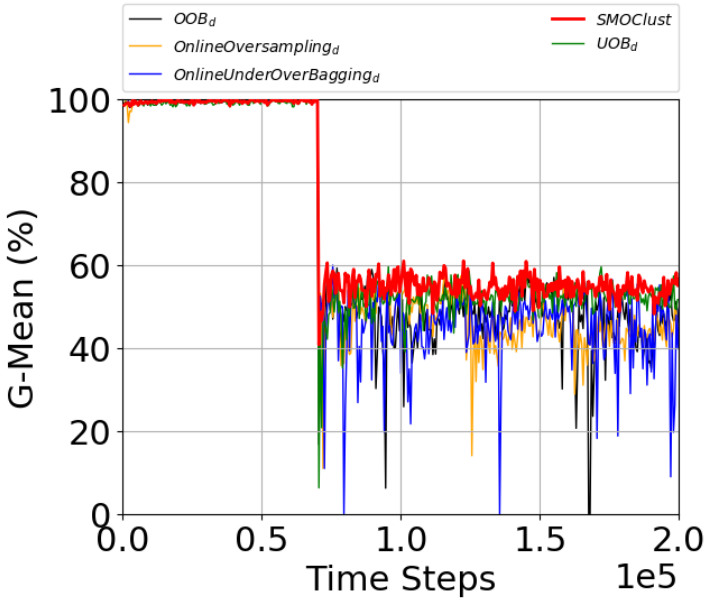

Fig. 6: Periodic Class Balanced Holdout Test G-Mean Against Time Steps in Two-Dimensional StaticIm10_Rare100

Table 3: 30 Runs Average G-Mean on Five Dimensional Artificial Data Streams (A12 SMOClust vs Others)

Stream	OOB	UOB	oOS	oUnder-OverB	OOB$_d$	UOB$_d$	oOS$_d$	oUnder-OverB$_d$	SMO-GauNoise	SMOClust
StaticIm30										
StaticIm30_Split3	96.56%[-b]	96.33%[-b]	96.85%[-b]	96.6%[-b]	97.66%[b]	97.04%[*]	97.7%[b]	97.46%[b]	97.84%[b]	97.18%
StaticIm30_Move3	96.39%[-s]	95.7%[-b]	96.82%[m]	96.21%[-b]	96.8%[s]	95.92%[-b]	96.88%[m]	96.46%[-s]	97.19%[b]	96.61%
StaticIm30_Merge3	96.96%[-s]	96.41%[-b]	97.31%[m]	96.92%[-s]	97.32%[m]	96.5%[-b]	97.39%[b]	96.98%[*]	97.7%[b]	97.08%
StaticIm30_Split7	96.9%[-b]	96.62%[-b]	97.14%[-m]	96.83%[-b]	97.78%[b]	97.01%[-m]	97.75%[b]	97.45%[s]	97.99%[b]	97.36%
StaticIm30_Move7	96.41%[s]	95.7%[-b]	96.83%[b]	96.2%[*]	96.38%[s]	95.75%[-b]	96.71%[b]	96.17%[-s]	96.84%[b]	96.29%
StaticIm30_Merge7	97.34%[b]	96.86%[-s]	97.62%[b]	97.33%[m]	97.33%[b]	96.9%[-s]	97.53%[b]	97.33%[m]	97.6%[b]	97.09%
StaticIm30_Borderline20	97.17%[m]	97.1%[s]	97.26%[b]	97.2%[m]	97.35%[b]	97.16%[m]	97.27%[b]	97.28%[b]	97.42%[b]	96.9%
StaticIm30_Borderline100	95.7%[-s]	95.6%[-s]	95.96%[*]	95.73%[-s]	96.16%[s]	95.7%[-s]	96.11%[s]	95.97%[*]	96.3%[m]	95.89%
StaticIm30_Rare20	93.15%[b]	92.95%[b]	93.27%[b]	93.05%[b]	92.93%[b]	92.61%[b]	92.94%[b]	92.73%[b]	92.99%[b]	91.72%
StaticIm30_Rare100	71.6%[b]	69.1%[b]	72.11%[b]	67.03%[-s]	70.83%[b]	70.61%[b]	67.55%[s]	70.79%[b]	70.48%[b]	67.31%
StaticIm10										
StaticIm10_Split3	96.06%[s]	95.84%[*]	95.91%[*]	96.4%[b]	97.3%[b]	95.64%[-s]	96.71%[b]	97.43%[b]	97.57%[b]	95.81%
StaticIm10_Move3	96.16%[b]	94.73%[-b]	96.26%[b]	96.29%[b]	96.39%[b]	94.69%[-b]	95.59%[m]	96.37%[b]	96.82%[b]	95.36%
StaticIm10_Merge3	96.79%[b]	95.61%[-s]	96.96%[b]	97.02%[b]	97.02%[b]	95.67%[*]	96.35%[b]	97.04%[b]	97.43%[b]	95.87%
StaticIm10_Split7	96.36%[s]	96.12%[*]	96.08%[*]	96.59%[b]	97.43%[b]	95.55%[-b]	96.54%[b]	97.42%[b]	97.62%[b]	96.15%
StaticIm10_Move7	96.25%[b]	94.65%[-m]	96.33%[b]	96.34%[b]	96.15%[b]	94.57%[-m]	95.68%[b]	96.29%[b]	96.62%[b]	95.1%
StaticIm10_Merge7	97.21%[b]	96.16%[*]	97.36%[b]	97.47%[b]	97.13%[b]	96.18%[*]	96.76%[b]	97.43%[b]	97.52%[b]	96.11%
StaticIm10_Borderline20	97.21%[b]	96.98%[b]	97.12%[b]	97.32%[b]	97.2%[b]	96.85%[s]	96.69%[s]	97.37%[b]	97.37%[b]	96.64%
StaticIm10_Borderline100	95.73%[s]	95.45%[*]	95.64%[s]	95.86%[s]	96.02%[b]	94.69%[-b]	95.01%[-s]	96.09%[b]	96.18%[b]	95.48%
StaticIm10_Rare20	93.41%[b]	92.56%[b]	93.53%[b]	93.19%[b]	93.05%[b]	91.99%[m]	92.55%[b]	92.99%[b]	93.35%[b]	91.74%
StaticIm10_Rare100	72.91%[b]	62.94%[m]	70.19%[b]	68.8%[b]	68.27%[b]	68.76%[b]	51.42%[-b]	65.76%[b]	63.83%[b]	61.94%
StaticIm1										
StaticIm1_Split3	82.88%[-b]	89.88%[*]	74.12%[-b]	87.98%[-b]	58.16%[-b]	85.0%[-b]	50.0%[-b]	84.73%[-b]	82.01%[-b]	90.03%
StaticIm1_Move3	84.22%[-b]	91.28%[m]	74.16%[-b]	88.87%[-m]	63.31%[-b]	83.36%[-b]	49.59%[-b]	82.48%[-b]	77.1%[-b]	89.59%
StaticIm1_Merge3	88.14%[-b]	93.23%[b]	80.01%[-b]	91.68%[s]	62.83%[-b]	85.54%[-b]	53.14%[-b]	86.08%[-b]	82.98%[-b]	90.94%
StaticIm1_Split7	82.88%[-b]	90.08%[*]	74.62%[-b]	87.72%[-b]	60.18%[-b]	85.78%[-b]	53.32%[-b]	85.46%[-b]	79.56%[-b]	90.32%
StaticIm1_Move7	84.99%[-b]	92.21%[b]	74.64%[-b]	89.17%[-s]	62.19%[-b]	82.4%[-b]	46.07%[-b]	83.93%[-b]	75.46%[-b]	89.67%
StaticIm1_Merge7	89.39%[-b]	94.12%[b]	82.11%[-b]	92.6%[m]	67.24%[-b]	85.9%[-b]	62.51%[-b]	87.98%[-b]	84.47%[-b]	91.19%
StaticIm1_Borderline20	92.55%[-b]	96.34%[b]	89.44%[-b]	95.19%[-s]	64.7%[-b]	91.6%[-b]	68.54%[-b]	92.32%[-b]	89.66%[-b]	95.52%
StaticIm1_Borderline100	89.3%[-b]	93.52%[*]	84.25%[-b]	93.19%[*]	55.61%[-b]	84.27%[-b]	43.86%[-b]	80.26%[-b]	67.17%[-b]	93.29%
StaticIm1_Rare20	90.37%[-b]	91.24%[-b]	87.27%[-b]	92.27%[-s]	63.16%[-b]	77.91%[-b]	73.08%[-b]	91.46%[-b]	92.95%[s]	92.6%
StaticIm1_Rare100	68.79%[b]	58.12%[-b]	61.17%[s]	67.67%[b]	37.73%[-b]	41.75%[-b]	42.28%[-b]	62.02%[m]	71.92%[b]	60.8%
Imbalance Ratio Drift										
StaticIm10_Im1	98.11%[b]	96.27%[-s]	97.98%[b]	98.06%[b]	98.05%[b]	95.8%[-s]	97.51%[b]	98.06%[b]	98.01%[b]	96.66%
StaticIm1_Im10	93.08%[-b]	95.51%[s]	89.22%[-b]	95.78%[m]	78.05%[-b]	89.72%[-b]	73.33%[-b]	91.38%[-b]	85.05%[-b]	94.86%
Im1	97.98%[b]	97.88%[b]	98.25%[b]	97.75%[b]	92.58%[*]	93.59%[-b]	88.8%[*]	95.15%[s]	92.31%[-m]	96.97%

StaticIm1_Im50	93.27%[-b]	96.39%[b]	89.72%[-b]	95.78%[s]	79.14%[-b]	90.66%[-b]	75.44%[-b]	91.47%[-b]	85.77%[-b]	95.37%
Double Factors										
Im1+Rare100	51.35%[-b]	68.42%[b]	51.5%[-b]	46.6%[-b]	49.82%[-b]	65.26%[b]	45.57%[-b]	48.76%[-b]	52.74%[-b]	53.74%
Im10+Rare60	79.26%[b]	77.67%[-b]	79.8%[b]	78.3%[-b]	79.05%[b]	77.34%[-b]	79.04%[b]	78.13%[-b]	79.67%[b]	78.58%
Split5+Im10	96.58%[-m]	95.55%[-b]	96.66%[-m]	96.74%[-s]	97.58%[b]	96.17%[-b]	97.47%[b]	97.52%[b]	97.89%[b]	96.94%
Im1+Borderline100	93.93%[-b]	94.87%[-s]	93.1%[-b]	94.58%[-m]	95.4%[*]	92.61%[-b]	93.39%[-m]	94.87%[*]	88.1%[*]	95.28%
Im10+Borderline20	97.19%[b]	96.88%[s]	97.17%[b]	97.23%[b]	97.27%[b]	96.96%[s]	97.1%[m]	97.25%[b]	97.37%[b]	96.77%
Complex Factors										
StaticIm10_Split5 +Im1+Rare100	61.46%[b]	48.75%[-b]	58.21%[b]	56.93%[*]	57.75%[b]	58.28%[*]	46.11%[-b]	54.66%[-b]	59.65%[b]	56.94%
StaticIm10_Split5 +Im1+Borderline100	84.66%[-b]	87.89%[*]	79.73%[-b]	86.88%[-m]	81.97%[-b]	84.9%[-s]	54.87%[-b]	84.59%[-b]	82.97%[-b]	87.56%
Split5+Im10 +Borderline40+Rare40	80.21%[b]	77.88%[-b]	80.43%[b]	79.4%[-m]	80.71%[b]	79.91%[s]	79.88%[b]	80.55%[b]	80.96%[b]	79.72%
Split5+Im10 +Borderline80	90.84%[-b]	90.06%[-b]	91.11%[-s]	90.86%[-b]	91.32%[-s]	90.17%[-b]	90.58%[-m]	91.35%[*]	91.4%[*]	91.45%
Im10+Borderline20 +Rare20	91.01%[b]	90.46%[*]	91.06%[b]	90.92%[b]	91.24%[b]	90.99%[b]	91.33%[b]	91.17%[b]	91.51%[b]	90.51%

- Based on the average G-Mean, cells are highlighted in lime / light grey when SMOClust performed better than the corresponding approach and cells are highlighted in orange / dark grey cells when SMOClust performed worse than the corresponding approach. The colour saturation scales with the absolute difference of average G-Mean between the SMOClust and the approach of the column and the saturation reaches the maximum when such difference is $\geq$ 10%.
- Symbols [*], [s], [m] and [b] represent insignificant, small, medium and large A12 effect size against SMOClust. Presence/absence of the sign "-" in the effect size means that the corresponding approach was worse/better than SMOClust.

Table 4: 30 Runs Average G-Mean on Five Dimensional Severely Class Imbalanced Artificial Data Streams (A12 SMOClust vs Others)

Stream	OOB	UOB	oOS	oUnder-OverB	OOB_d	UOB_d	oOS_d	oUnder-$OverB_d$	SMO-GauNoise	SMOClust
StaticIm5										
StaticIm5_Split3	95.24%[*]	95.54%[s]	94.47%[-b]	96.1%[b]	96.2%[b]	95.26%[*]	94.09%[-s]	97.11%[b]	97.11%[b]	95.19%
StaticIm5_Move3	95.26%[b]	93.64%[-m]	94.74%[s]	95.93%[b]	95.46%[b]	93.48%[-b]	88.6%[-b]	95.88%[b]	96.11%[b]	94.23%
StaticIm5_Merge3	96.01%[b]	94.64%[*]	95.88%[b]	96.67%[b]	96.18%[b]	94.59%[*]	89.64%[-b]	96.65%[b]	96.87%[b]	94.7%
StaticIm5_Split7	95.07%[*]	95.23%[s]	94.05%[-b]	95.82%[b]	95.46%[m]	94.61%[-s]	92.68%[-s]	96.94%[b]	96.86%[b]	94.99%
StaticIm5_Move7	95.56%[b]	93.37%[-b]	94.83%[m]	95.93%[b]	95.12%[b]	93.02%[-b]	90.63%[-b]	95.74%[b]	95.93%[b]	94.31%
StaticIm5_Merge7	96.72%[b]	95.4%[s]	96.42%[b]	97.16%[b]	96.36%[b]	95.22%[*]	92.4%[-b]	97.0%[b]	97.15%[b]	95.21%
StaticIm5_Borderline20	97.1%[b]	96.92%[s]	96.76%[*]	97.42%[b]	96.39%[-s]	96.73%[*]	94.59%[-b]	97.26%[b]	97.25%[b]	96.69%
StaticIm5_Borderline100	95.63%[s]	94.99%[-s]	95.12%[-s]	95.99%[m]	95.18%[-s]	94.31%[-b]	86.67%[-b]	96.15%[b]	96.02%[m]	95.45%
StaticIm5_Rare20	93.68%[b]	92.29%[s]	93.51%[b]	93.46%[b]	92.73%[b]	91.55%[-b]	90.99%[-b]	93.04%[b]	93.66%[b]	92.15%
StaticIm5_Rare100	73.34%[b]	59.64%[-s]	68.17%[b]	69.93%[b]	66.84%[b]	66.64%[b]	47.73%[-b]	61.41%[s]	60.23%[*]	60.53%
StaticIm3										
StaticIm3_Split3	93.67%[-b]	95.28%[m]	91.81%[-b]	95.12%[m]	92.24%[-b]	94.84%[s]	87.39%[-b]	96.2%[b]	95.94%[b]	94.51%
StaticIm3_Move3	93.84%[s]	93.04%[-s]	91.87%[-b]	95.21%[b]	93.03%[*]	92.35%[-m]	80.77%[-b]	94.86%[b]	94.83%[b]	93.27%
StaticIm3_Merge3	95.04%[b]	94.03%[*]	93.86%[*]	96.19%[b]	93.78%[*]	93.55%[*]	84.34%[-b]	95.86%[b]	96.09%[b]	93.7%
StaticIm3_Split7	93.2%[-b]	95.05%[b]	90.86%[-b]	94.76%[m]	90.64%[-b]	93.94%[*]	84.29%[-b]	95.8%[b]	95.76%[b]	94.15%
StaticIm3_Move7	94.33%[b]	92.86%[-m]	92.31%[-b]	95.33%[b]	92.73%[-s]	92.23%[-b]	82.03%[-b]	94.74%[b]	94.4%[b]	93.5%
StaticIm3_Merge7	95.93%[b]	94.97%[*]	94.81%[*]	96.79%[b]	94.35%[*]	94.61%[-s]	85.91%[-b]	96.23%[b]	95.85%[b]	94.79%
StaticIm3_Borderline20	96.58%[*]	96.8%[s]	95.97%[-b]	97.23%[b]	92.88%[-b]	96.43%[-s]	91.48%[-b]	96.83%[s]	96.53%[*]	96.66%
StaticIm3_Borderline100	94.97%[-s]	94.59%[-m]	93.93%[-b]	95.77%[m]	91.38%[-b]	93.86%[-b]	70.99%[-b]	95.81%[m]	94.69%[*]	95.2%
StaticIm3_Rare20	93.43%[b]	92.09%[-m]	92.97%[b]	93.59%[b]	90.17%[-b]	90.87%[-b]	90.02%[-b]	93.33%[b]	94.44%[b]	92.34%
StaticIm3_Rare100	71.72%[b]	62.22%[b]	66.12%[b]	69.63%[b]	64.93%[b]	62.23%[b]	53.0%[-b]	62.63%[b]	61.83%[s]	59.92%
StaticIm1										
StaticIm1_Split3	84.29%[-b]	90.98%[*]	75.65%[-b]	88.67%[-b]	61.71%[-b]	85.62%[-b]	54.23%[-b]	85.79%[-b]	82.17%[-b]	90.91%
StaticIm1_Move3	82.76%[-b]	91.14%[m]	71.59%[-b]	87.92%[-b]	70.48%[-b]	85.17%[-b]	46.61%[-b]	81.17%[-b]	76.86%[-b]	89.55%
StaticIm1_Merge3	86.61%[-b]	92.27%[b]	77.68%[-b]	90.9%[s]	67.28%[-b]	85.85%[-m]	52.57%[-b]	84.94%[-b]	82.54%[-b]	90.22%
StaticIm1_Split7	82.03%[-b]	89.38%[*]	73.04%[-b]	87.2%[-b]	58.85%[-b]	85.31%[-b]	52.58%[-b]	84.55%[-b]	79.12%[-b]	90.25%
StaticIm1_Move7	83.3%[-b]	91.7%[b]	73.53%[-b]	88.48%[-m]	69.71%[-b]	81.62%[-b]	47.2%[-b]	81.25%[-b]	76.87%[-b]	89.36%
StaticIm1_Merge7	88.16%[-b]	93.49%[b]	81.61%[-b]	92.15%[m]	74.91%[-b]	85.81%[-b]	60.94%[-b]	86.48%[-b]	84.54%[-b]	91.15%
StaticIm1_Borderline20	92.42%[-b]	96.49%[m]	89.57%[-b]	95.32%[-m]	72.26%[-b]	91.37%[-b]	70.68%[-b]	92.25%[-b]	90.22%[-b]	95.9%
StaticIm1_Borderline100	89.38%[-b]	93.84%[*]	84.72%[-b]	93.39%[-s]	55.97%[-b]	86.42%[-b]	43.44%[-b]	77.92%[-b]	67.67%[-b]	93.66%
StaticIm1_Rare20	90.09%[-b]	91.35%[-b]	87.3%[-b]	92.24%[-m]	71.2%[-b]	80.35%[-b]	70.04%[-b]	90.97%[-b]	92.31%[*]	92.89%
StaticIm1_Rare100	68.79%[b]	58.9%[-b]	61.54%[*]	67.6%[b]	43.52%[-b]	45.74%[-b]	42.37%[-b]	62.08%[*]	71.26%[b]	61.5%
StaticIm07										
StaticIm07_Split3	77.51%[-b]	89.86%[*]	65.19%[-b]	83.34%[-b]	45.48%[-b]	80.25%[-b]	36.62%[-b]	74.97%[-b]	70.32%[-b]	89.45%

StaticIm07_Move3	74.0%[-b]	90.55%[b]	57.69%[-b]	81.13%[-b]	43.83%[-b]	76.82%[-b]	34.75%[-b]	70.38%[-b]	63.8%[-b]	87.26%
StaticIm07_Merge3	79.16%[-b]	91.48%[b]	66.0%[-b]	85.44%[-s]	47.07%[-b]	77.02%[-b]	37.31%[-b]	75.7%[-b]	70.53%[-b]	87.18%
StaticIm07_Split7	75.59%[-b]	88.17%[*]	63.64%[-b]	81.65%[-b]	41.83%[-b]	77.0%[-b]	37.01%[-b]	74.84%[-b]	70.73%[-b]	88.22%
StaticIm07_Move7	74.2%[-b]	91.4%[b]	59.56%[-b]	82.08%[-b]	45.21%[-b]	76.38%[-b]	36.29%[-b]	69.28%[-b]	62.5%[-b]	87.43%
StaticIm07_Merge7	81.7%[-b]	92.91%[b]	71.34%[-b]	87.48%[-m]	50.87%[-b]	81.99%[-b]	48.24%[-b]	78.85%[-b]	75.74%[-b]	88.97%
StaticIm07_Borderline20	89.33%[-b]	96.39%[b]	84.19%[-b]	93.29%[-b]	56.19%[-b]	88.59%[-b]	48.37%[-b]	89.35%[-b]	80.2%[-b]	95.1%
StaticIm07_Borderline100	84.99%[-b]	93.42%[s]	77.45%[-b]	90.82%[-b]	39.67%[-b]	76.94%[-b]	30.87%[-b]	66.34%[-b]	56.31%[-b]	92.59%
StaticIm07_Rare20	87.34%[-b]	90.98%[-b]	82.05%[-b]	90.52%[-b]	51.72%[-b]	71.26%[-b]	45.94%[-b]	87.73%[-b]	87.15%[-b]	92.51%
StaticIm07_Rare100	68.59%[b]	58.39%[-b]	58.71%[-b]	66.68%[b]	25.99%[-b]	40.18%[-b]	28.49%[-b]	62.65%[m]	70.08%[b]	61.97%
StaticIm05										
StaticIm05_Split3	67.41%[-b]	89.06%[m]	51.8%[-b]	75.26%[-b]	28.85%[-b]	70.03%[-b]	24.15%[-b]	61.4%[-b]	58.01%[-b]	86.92%
StaticIm05_Move3	60.73%[-b]	89.92%[b]	42.99%[-b]	71.27%[-b]	25.71%[-b]	60.37%[-b]	22.91%[-b]	52.3%[-b]	49.96%[-b]	82.97%
StaticIm05_Merge3	68.16%[-b]	90.63%[b]	52.0%[-b]	77.04%[-b]	26.29%[-b]	63.85%[-b]	25.45%[-b]	60.63%[-b]	53.63%[-b]	84.28%
StaticIm05_Split7	66.73%[-b]	87.47%[s]	52.25%[-b]	74.33%[-b]	29.97%[-b]	64.71%[-b]	26.36%[-b]	64.25%[-b]	59.18%[-b]	85.98%
StaticIm05_Move7	61.93%[-b]	91.08%[b]	44.91%[-b]	72.82%[-b]	32.82%[-b]	57.0%[-b]	25.03%[-b]	54.0%[-b]	49.63%[-b]	84.18%
StaticIm05_Merge7	72.74%[-b]	92.48%[b]	59.63%[-b]	80.73%[-b]	39.25%[-b]	69.5%[-b]	31.25%[-b]	70.66%[-b]	64.57%[-b]	86.78%
StaticIm05_Borderline20	84.19%[-b]	96.29%[b]	76.48%[-b]	90.35%[-b]	30.95%[-b]	79.13%[-b]	36.96%[-b]	77.36%[-b]	69.86%[-b]	93.92%
StaticIm05_Borderline100	77.54%[-b]	93.15%[b]	68.43%[-b]	87.03%[-b]	26.19%[-b]	57.71%[-b]	25.39%[-b]	52.65%[-b]	47.62%[-b]	91.05%
StaticIm05_Rare20	82.38%[-b]	90.85%[-m]	74.39%[-b]	87.71%[-b]	33.53%[-b]	60.14%[-b]	38.56%[-b]	81.84%[-b]	78.82%[-b]	91.51%
StaticIm05_Rare100	65.62%[m]	59.16%[-b]	55.02%[-b]	65.53%[b]	14.14%[-b]	32.84%[-b]	26.25%[-b]	56.47%[-b]	67.64%[m]	63.7%
StaticIm03										
StaticIm03_Split3	49.77%[-b]	88.35%[b]	33.95%[-b]	58.57%[-b]	14.32%[-b]	45.41%[-b]	13.85%[-b]	42.25%[-b]	39.95%[-b]	81.31%
StaticIm03_Move3	40.57%[-b]	88.81%[b]	22.43%[-b]	49.93%[-b]	18.07%[-b]	40.61%[-b]	10.55%[-b]	35.38%[-b]	35.13%[-b]	75.16%
StaticIm03_Merge3	49.77%[-b]	89.5%[b]	31.09%[-b]	59.18%[-b]	17.83%[-b]	46.84%[-b]	14.48%[-b]	41.79%[-b]	35.16%[-b]	76.22%
StaticIm03_Split7	50.45%[-b]	86.36%[b]	32.87%[-b]	57.4%[-b]	15.86%[-b]	48.09%[-b]	14.5%[-b]	41.15%[-b]	39.14%[-b]	80.4%
StaticIm03_Move7	41.14%[-b]	90.27%[b]	23.33%[-b]	53.4%[-b]	18.43%[-b]	33.73%[-b]	10.78%[-b]	33.98%[-b]	34.82%[-b]	75.72%
StaticIm03_Merge7	56.06%[-b]	91.5%[b]	38.89%[-b]	65.82%[-b]	23.33%[-b]	52.59%[-b]	16.89%[-b]	52.72%[-b]	49.98%[-b]	78.12%
StaticIm03_Borderline20	70.16%[-b]	96.09%[b]	55.01%[-b]	81.64%[-b]	23.08%[-b]	58.81%[-b]	19.34%[-b]	56.61%[-b]	43.26%[-b]	91.8%
StaticIm03_Borderline100	61.02%[-b]	92.72%[b]	49.14%[-b]	76.31%[-b]	18.34%[-b]	43.35%[-b]	15.68%[-b]	37.21%[-b]	37.93%[-b]	88.13%
StaticIm03_Rare20	69.26%[-b]	90.63%[s]	53.43%[-b]	79.4%[-b]	19.93%[-b]	44.64%[-b]	19.35%[-b]	59.5%[-b]	53.38%[-b]	89.69%
StaticIm03_Rare100	56.62%[-b]	58.64%[-b]	45.25%[-b]	60.56%[-b]	10.95%[-b]	24.78%[-b]	16.73%[-b]	47.49%[-b]	57.16%[-b]	64.62%

- Based on the average G-Mean, cells are highlighted in lime / light grey when SMOClust performed better than the corresponding approach and cells are highlighted in orange / dark grey cells when SMOClust performed worse than the corresponding approach. The colour saturation scales with the absolute difference of average G-Mean between the SMOClust and the approach of the column and the saturation reaches the maximum when such difference is $\geq$ 10%.
- Symbols [*], [s], [m] and [b] represent insignificant, small, medium and large A12 effect size against SMOClust. Presence/absence of the sign "-" in the effect size means that the corresponding approach was worse/better than SMOClust.

Table 5: [2D] 30 Runs Average G-Mean on Artificial Data Streams (A12 SMOClust vs Others)

Stream	OOB	UOB	oOS	oUnder-OverB	OOB_d	UOB_d	oOS_d	oUnder-$OverB_d$	SMO-GauNoise	SMOClust
StaticIm30										
StaticIm30_Split3	96.93%[-b]	96.76%[-b]	97.08%[-b]	97.06%[-b]	98.19%[*]	97.31%[-b]	98.61%[b]	97.85%[-b]	98.53%[b]	98.23%
StaticIm30_Move3	95.46%[-b]	94.33%[-b]	96.15%[-b]	95.22%[-b]	97.28%[-s]	94.87%[-b]	98.09%[b]	96.43%[-b]	97.89%[b]	97.42%
StaticIm30_Merge3	96.99%[-b]	95.98%[-b]	97.69%[-b]	96.71%[-b]	98.45%[-b]	96.87%[-b]	98.85%[b]	97.73%[-b]	98.82%[b]	98.62%
StaticIm30_Split7	96.07%[-b]	95.33%[-b]	96.36%[-b]	95.87%[-b]	97.14%[*]	93.9%[-b]	97.84%[b]	96.0%[-b]	97.63%[b]	97.18%
StaticIm30_Move7	93.6%[-b]	91.96%[-b]	94.45%[-b]	92.88%[-b]	94.95%[-s]	90.97%[-b]	96.57%[b]	92.99%[-b]	96.29%[b]	95.41%
StaticIm30_Merge7	96.12%[-b]	95.24%[-b]	96.64%[-b]	96.04%[-b]	96.94%[-b]	95.32%[-b]	98.27%[b]	96.11%[-b]	98.2%[b]	97.9%
StaticIm30_Borderline20	99.47%[-m]	99.45%[-m]	99.48%[-s]	99.47%[-s]	99.52%[s]	99.48%[-s]	99.53%[s]	99.51%[*]	99.53%[s]	99.51%
StaticIm30_Borderline100	99.34%[-b]	99.33%[-b]	99.34%[-b]	99.35%[-b]	99.39%[-s]	99.32%[-b]	99.42%[*]	99.36%[-m]	99.49%[b]	99.42%
StaticIm30_Rare20	93.79%[b]	93.75%[b]	93.82%[b]	93.77%[b]	93.66%[b]	93.43%[*]	93.72%[b]	93.51%[b]	93.68%[b]	93.47%
StaticIm30_Rare100	68.44%[-b]	67.27%[-b]	68.7%[-b]	62.09%[-b]	67.72%[-b]	68.14%[-b]	67.85%[-b]	67.69%[-b]	67.8%[-b]	68.95%
StaticIm10										
StaticIm10_Split3	96.69%[-b]	95.74%[-b]	96.75%[-b]	96.96%[-b]	97.98%[s]	94.08%[-b]	98.27%[b]	97.98%[s]	98.21%[b]	97.88%
StaticIm10_Move3	95.52%[-b]	92.32%[-b]	95.86%[-b]	95.52%[-b]	97.24%[b]	90.83%[-b]	97.71%[b]	96.85%[s]	97.55%[b]	96.83%
StaticIm10_Merge3	97.21%[-b]	94.71%[-b]	97.51%[-b]	97.05%[-b]	98.52%[s]	94.86%[-b]	98.68%[b]	98.11%[-b]	98.69%[b]	98.45%
StaticIm10_Split7	95.75%[-b]	92.61%[-b]	95.87%[-b]	95.92%[-b]	97.02%[m]	87.8%[-b]	97.35%[b]	96.61%[*]	97.26%[b]	96.7%
StaticIm10_Move7	93.61%[-b]	87.73%[-b]	94.1%[-b]	93.36%[-b]	95.32%[b]	86.08%[-b]	96.03%[b]	93.92%[-b]	95.9%[b]	94.66%
StaticIm10_Merge7	96.2%[-b]	93.82%[-b]	96.54%[-b]	96.36%[-b]	97.52%[s]	93.64%[-b]	98.05%[b]	96.28%[-b]	98.02%[b]	97.48%
StaticIm10_Borderline20	99.44%[-b]	99.42%[-b]	99.41%[-b]	99.47%[-m]	99.38%[-b]	99.37%[-b]	99.43%[-b]	99.48%[-s]	99.47%[-s]	99.51%
StaticIm10_Borderline100	99.29%[-b]	99.3%[-b]	99.27%[-b]	99.33%[-s]	99.3%[-m]	99.14%[-b]	99.33%[-s]	99.33%[-s]	99.44%[m]	99.37%
StaticIm10_Rare20	93.85%[b]	93.67%[-b]	93.92%[b]	93.79%[b]	93.67%[-b]	92.81%[-b]	93.74%[*]	93.67%[-b]	93.79%[b]	93.76%
StaticIm10_Rare100	68.69%[-b]	61.0%[-b]	67.54%[-b]	64.38%[-b]	64.93%[-b]	66.67%[-b]	64.68%[-b]	63.16%[-b]	64.37%[-b]	69.1%
StaticIm1										
StaticIm1_Split3	92.62%[-b]	84.82%[-b]	91.17%[-b]	93.61%[-b]	65.97%[-b]	70.84%[-b]	90.03%[-b]	88.95%[-b]	92.94%[-b]	96.35%
StaticIm1_Move3	91.09%[-b]	81.03%[-b]	89.5%[-b]	92.16%[-b]	68.97%[-b]	64.54%[-b]	87.83%[-b]	78.87%[-b]	91.72%[-b]	95.61%
StaticIm1_Merge3	94.54%[-b]	90.06%[-b]	93.79%[-b]	95.41%[-b]	77.41%[-b]	80.93%[-b]	93.25%[-b]	88.34%[-b]	95.42%[-b]	97.93%
StaticIm1_Split7	89.73%[-b]	75.71%[-b]	87.88%[-b]	91.15%[-b]	63.24%[-b]	63.2%[-b]	87.41%[-b]	76.54%[-b]	90.78%[-b]	94.87%
StaticIm1_Move7	87.09%[-b]	73.68%[-b]	84.53%[-b]	88.9%[-b]	63.32%[-b]	55.73%[-b]	82.0%[-b]	62.05%[-b]	87.95%[-b]	92.92%
StaticIm1_Merge7	93.08%[-b]	88.23%[-b]	91.87%[-b]	94.47%[-b]	73.99%[-b]	79.08%[-b]	90.75%[-b]	85.28%[-b]	93.94%[-b]	96.87%
StaticIm1_Borderline20	98.19%[-b]	99.31%[-s]	97.97%[-b]	98.84%[-b]	81.89%[-b]	98.48%[-b]	96.83%[-b]	98.05%[-b]	97.92%[-b]	99.35%
StaticIm1_Borderline100	97.95%[-b]	99.0%[-b]	97.61%[-b]	98.68%[-b]	82.35%[-b]	97.73%[-b]	97.1%[-b]	98.15%[-b]	98.31%[-b]	99.3%
StaticIm1_Rare20	93.69%[-b]	92.78%[-b]	93.77%[-b]	94.12%[-b]	73.79%[-b]	79.71%[-b]	92.01%[-b]	92.36%[-b]	94.22%[-m]	94.78%
StaticIm1_Rare100	62.82%[-b]	59.84%[-b]	61.6%[-b]	64.39%[-b]	36.31%[-b]	40.2%[-b]	55.35%[-b]	52.08%[-b]	56.94%[-b]	72.51%
Imbalance Ratio Drift										
StaticIm10_Im1	99.71%[b]	99.22%[-b]	99.74%[b]	99.65%[b]	99.66%[b]	99.19%[-b]	99.74%[b]	99.65%[b]	99.73%[b]	99.45%
StaticIm1_Im10	98.96%[-b]	98.77%[-b]	98.67%[-b]	99.28%[-s]	90.94%[-b]	97.04%[-b]	98.47%[-b]	98.92%[-b]	98.83%[-b]	99.37%
Im1	99.65%[b]	99.56%[b]	99.77%[b]	99.56%[b]	99.65%[b]	98.85%[*]	99.77%[b]	99.53%[b]	99.71%[b]	99.44%

StaticIm1_Im50	98.98%[-b]	99.01%[-b]	98.7%[-b]	99.28%[-m]	91.05%[-b]	97.34%[-b]	98.5%[-b]	98.92%[-b]	98.84%[-b]	99.4%
Double Factors										
Im1+Rare100	48.84%[-b]	67.45%[b]	50.07%[-b]	44.91%[-b]	48.1%[-b]	58.6%[m]	48.67%[-b]	48.17%[-b]	53.62%[-s]	53.73%
Im10+Rare60	78.57%[*]	78.11%[-b]	78.97%[b]	78.2%[-b]	78.28%[-b]	77.63%[-b]	78.84%[b]	77.85%[-b]	78.93%[b]	78.62%
Split5+Im10	96.97%[-b]	94.44%[-b]	96.64%[-b]	96.99%[-b]	97.79%[b]	92.14%[-b]	98.13%[b]	97.11%[-m]	98.03%[b]	97.5%
Im1+Borderline100	99.17%[-b]	99.3%[-b]	99.08%[-b]	99.22%[-b]	98.16%[-b]	99.23%[-m]	99.13%[-b]	99.21%[-b]	99.26%[*]	99.4%
Im10+Borderline20	99.45%[-m]	99.37%[-b]	99.43%[-m]	99.47%[-s]	99.47%[-s]	99.46%[-s]	99.5%[*]	99.48%[-s]	99.52%[*]	99.5%
Complex Factors										
StaticIm10_Split5 +Im1+Rare100	55.24%[-b]	47.75%[-b]	54.77%[-b]	53.1%[-b]	52.7%[-b]	55.73%[-s]	52.86%[-b]	52.52%[-b]	60.56%[*]	60.45%
StaticIm10_Split5 +Im1+Borderline100	85.08%[b]	85.79%[b]	84.35%[b]	84.72%[b]	76.32%[-b]	77.87%[m]	75.3%[-b]	77.03%[-m]	76.59%[-m]	78.3%
Split5+Im10 +Borderline40+Rare40	79.65%[b]	79.24%[s]	80.43%[b]	79.22%[s]	78.72%[-s]	78.43%[-s]	79.04%[*]	78.95%[*]	79.08%[s]	78.9%
Split5+Im10 +Borderline80	89.55%[b]	89.74%[b]	90.28%[b]	89.1%[b]	86.98%[*]	85.78%[-m]	87.06%[*]	87.03%[*]	87.1%[*]	87.07%
Im10+Borderline20 +Rare20	93.23%[-m]	92.96%[-b]	93.25%[-s]	93.2%[-b]	93.29%[*]	93.19%[-b]	93.34%[m]	93.23%[-m]	93.36%[b]	93.29%

- Based on the average G-Mean, cells are highlighted in lime / light grey when SMOClust performed better than the corresponding approach and cells are highlighted in orange / dark grey cells when SMOClust performed worse than the corresponding approach. The colour saturation scales with the absolute difference of average G-Mean between the SMOClust and the approach of the column and the saturation reaches the maximum when such difference is $\geq$ 10%.
- Symbols [*], [s], [m] and [b] represent insignificant, small, medium and large A12 effect size against SMOClust. Presence/absence of the sign "-" in the effect size means that the corresponding approach was worse/better than SMOClust.

Table 6: [2D] 30 Runs Average G-Mean on Severely Class Imbalanced Artificial Data Streams (A12 SMOClust vs Others)

Stream	OOB	UOB	oOS	oUnder-OverB	OOB_d	UOB_d	oOS_d	oUnder-$OverB_d$	SMO-GauNoise	SMOClust
StaticIm5										
StaticIm5_Split3	96.26%[-b]	94.21%[-b]	96.23%[-b]	96.76%[-b]	96.93%[-b]	87.48%[-b]	97.56%[*]	97.81%[m]	97.74%[m]	97.53%
StaticIm5_Move3	95.3%[-b]	89.73%[-b]	95.5%[-b]	95.38%[-b]	96.44%[*]	87.8%[-b]	97.13%[b]	96.55%[s]	97.12%[b]	96.42%
StaticIm5_Merge3	97.1%[-b]	93.38%[-b]	97.32%[-b]	97.08%[-b]	98.07%[-s]	93.19%[-b]	98.37%[s]	97.94%[-m]	98.51%[b]	98.25%
StaticIm5_Split7	95.25%[-b]	89.69%[-b]	95.21%[-b]	95.6%[-b]	95.69%[-m]	81.97%[-b]	96.57%[s]	96.21%[*]	96.68%[b]	96.34%
StaticIm5_Move7	93.15%[-b]	84.98%[-b]	93.35%[-b]	93.1%[-b]	94.16%[s]	78.3%[-b]	95.19%[b]	93.54%[-s]	95.24%[b]	94.14%
StaticIm5_Merge7	96.23%[-b]	93.15%[-b]	96.45%[-b]	96.43%[-b]	97.15%[-s]	93.06%[-b]	97.84%[b]	96.38%[-b]	97.86%[b]	97.41%
StaticIm5_Borderline20	99.33%[-b]	99.39%[-b]	99.28%[-b]	99.44%[-b]	98.16%[-b]	99.28%[-b]	99.24%[-b]	99.41%[-b]	99.41%[-b]	99.51%
StaticIm5_Borderline100	99.2%[-b]	99.25%[-b]	99.16%[-b]	99.29%[-m]	98.28%[-b]	98.89%[-b]	99.17%[-b]	99.25%[-b]	99.41%[m]	99.35%
StaticIm5_Rare20	94.0%[*]	93.54%[-b]	94.1%[b]	93.87%[-b]	92.87%[-b]	91.79%[-b]	93.8%[-b]	93.67%[-b]	94.13%[b]	94.01%
StaticIm5_Rare100	68.78%[-b]	53.22%[-b]	65.87%[-b]	65.5%[-b]	60.87%[-b]	65.21%[-b]	60.57%[-b]	57.55%[-b]	59.83%[-b]	69.18%
StaticIm3										
StaticIm3_Split3	95.75%[-b]	93.0%[-b]	95.55%[-b]	96.31%[-b]	94.8%[-b]	84.39%[-b]	96.69%[-b]	96.64%[*]	97.09%[-s]	97.26%
StaticIm3_Move3	94.92%[-b]	88.37%[-b]	94.69%[-b]	95.09%[-b]	94.72%[-b]	86.62%[-b]	95.96%[-s]	95.33%[-b]	96.4%[s]	96.23%
StaticIm3_Merge3	96.92%[-b]	92.56%[-b]	96.9%[-b]	97.0%[-b]	97.02%[-b]	92.4%[-b]	97.76%[-b]	97.23%[-b]	98.1%[-s]	98.16%
StaticIm3_Split7	94.66%[-b]	84.86%[-b]	94.08%[-b]	95.02%[-b]	91.72%[-b]	79.73%[-b]	95.24%[-b]	94.99%[-m]	95.99%[-s]	96.13%
StaticIm3_Move7	92.6%[-b]	82.01%[-b]	92.2%[-b]	92.8%[-b]	90.52%[-b]	76.74%[-b]	93.49%[-m]	91.97%[-b]	94.21%[s]	93.97%
StaticIm3_Merge7	95.97%[-b]	92.05%[-b]	96.05%[-b]	96.38%[-b]	95.2%[-b]	91.75%[-b]	97.04%[-m]	95.94%[-b]	97.48%[s]	97.39%
StaticIm3_Borderline20	99.18%[-b]	99.36%[-b]	99.05%[-b]	99.36%[-b]	95.01%[-b]	99.22%[-b]	98.8%[-b]	99.26%[-b]	99.22%[-b]	99.51%
StaticIm3_Borderline100	99.07%[-b]	99.14%[-b]	98.92%[-b]	99.23%[-b]	95.87%[-b]	98.78%[-b]	98.92%[-b]	99.15%[-b]	99.27%[-m]	99.37%
StaticIm3_Rare20	94.05%[-b]	93.45%[-b]	94.14%[-b]	94.03%[-b]	90.51%[-b]	90.39%[-b]	93.69%[-b]	93.69%[-b]	94.38%[m]	94.29%
StaticIm3_Rare100	67.01%[-b]	52.07%[-b]	63.98%[-b]	65.69%[-b]	56.76%[-b]	58.21%[-b]	58.45%[-b]	53.06%[-b]	57.32%[-b]	70.06%
StaticIm1										
StaticIm1_Split3	92.71%[-b]	85.74%[-b]	91.24%[-b]	93.65%[-b]	76.82%[-b]	71.12%[-b]	89.85%[-b]	88.04%[-b]	93.45%[-b]	96.5%
StaticIm1_Move3	91.27%[-b]	81.03%[-b]	89.76%[-b]	92.32%[-b]	68.52%[-b]	64.46%[-b]	87.73%[-b]	82.06%[-b]	91.85%[-b]	95.55%
StaticIm1_Merge3	94.82%[-b]	89.41%[-b]	94.13%[-b]	95.51%[-b]	79.99%[-b]	79.37%[-b]	93.15%[-b]	90.39%[-b]	95.57%[-b]	97.95%
StaticIm1_Split7	89.82%[-b]	75.83%[-b]	87.8%[-b]	91.21%[-b]	61.89%[-b]	65.49%[-b]	87.41%[-b]	78.94%[-b]	91.36%[-b]	94.95%
StaticIm1_Move7	87.15%[-b]	73.77%[-b]	84.62%[-b]	89.05%[-b]	60.65%[-b]	57.18%[-b]	82.42%[-b]	69.44%[-b]	87.99%[-b]	92.87%
StaticIm1_Merge7	93.45%[-b]	88.61%[-b]	92.3%[-b]	94.84%[-b]	77.92%[-b]	80.93%[-b]	90.99%[-b]	86.07%[-b]	94.13%[-b]	96.94%
StaticIm1_Borderline20	98.17%[-b]	99.3%[-s]	97.8%[-b]	98.75%[-b]	75.25%[-b]	98.24%[-b]	96.63%[-b]	97.85%[-b]	97.83%[-b]	99.36%
StaticIm1_Borderline100	98.09%[-b]	98.96%[-b]	97.61%[-b]	98.64%[-b]	79.41%[-b]	97.63%[-b]	96.84%[-b]	98.17%[-b]	98.19%[-b]	99.31%
StaticIm1_Rare20	93.68%[-b]	92.89%[-b]	93.69%[-b]	94.05%[-b]	73.67%[-b]	80.05%[-b]	92.23%[-b]	92.28%[-b]	94.57%[-s]	94.82%
StaticIm1_Rare100	62.68%[-b]	57.72%[-b]	61.24%[-b]	64.07%[-b]	31.76%[-b]	39.77%[-b]	55.4%[-b]	52.83%[-b]	56.63%[-b]	72.29%
StaticIm07										
StaticIm07_Split3	90.14%[-b]	82.64%[-b]	88.33%[-b]	91.65%[-b]	61.98%[-b]	56.83%[-b]	86.9%[-b]	81.45%[-b]	91.14%[-b]	95.84%
StaticIm07_Move3	88.4%[-b]	79.1%[-b]	86.04%[-b]	90.04%[-b]	50.73%[-b]	43.13%[-b]	83.11%[-b]	69.94%[-b]	87.97%[-b]	94.64%
StaticIm07_Merge3	93.08%[-b]	88.02%[-b]	92.13%[-b]	94.23%[-b]	66.29%[-b]	73.48%[-b]	89.69%[-b]	81.7%[-b]	93.21%[-b]	97.45%

StaticIm07_Split7	86.43%[-b]	74.12%[-b]	84.09%[-b]	88.5%[-b]	49.38%[-b]	54.93%[-b]	82.33%[-b]	68.52%[-b]	88.42%[-b]	93.09%
StaticIm07_Move7	82.27%[-b]	73.03%[-b]	78.95%[-b]	85.64%[-b]	45.01%[-b]	43.87%[-b]	76.71%[-b]	50.16%[-b]	83.67%[-b]	90.64%
StaticIm07_Merge7	90.95%[-b]	87.27%[-b]	89.29%[-b]	93.12%[-b]	64.4%[-b]	75.23%[-b]	87.39%[-b]	74.94%[-b]	91.68%[-b]	96.06%
StaticIm07_Borderline20	97.42%[-b]	99.23%[m]	96.97%[-b]	98.28%[-b]	76.78%[-b]	97.06%[-b]	94.82%[-b]	96.54%[-b]	96.71%[-b]	99.12%
StaticIm07_Borderline100	97.31%[-b]	98.86%[-b]	96.57%[-b]	98.14%[-b]	76.49%[-b]	95.77%[-b]	94.28%[-b]	96.57%[-b]	97.1%[-b]	99.16%
StaticIm07_Rare20	93.36%[-b]	92.53%[-b]	93.08%[-b]	93.81%[-b]	64.74%[-b]	76.9%[-b]	90.37%[-b]	91.8%[-b]	93.87%[-b]	94.76%
StaticIm07_Rare100	63.08%[-b]	60.16%[-b]	60.81%[-b]	64.27%[-b]	30.95%[-b]	39.92%[-b]	52.3%[-b]	51.42%[-b]	58.15%[-b]	72.27%
StaticIm05										
StaticIm05_Split3	86.93%[-b]	78.45%[-b]	84.9%[-b]	89.2%[-b]	52.38%[-b]	49.15%[-b]	80.64%[-b]	73.54%[-b]	88.08%[-b]	93.97%
StaticIm05_Move3	84.0%[-b]	78.24%[-b]	81.57%[-b]	86.62%[-b]	42.25%[-b]	31.16%[-b]	76.55%[-b]	64.53%[-b]	84.97%[-b]	92.36%
StaticIm05_Merge3	90.39%[-b]	87.69%[-b]	89.25%[-b]	92.11%[-b]	59.16%[-b]	65.55%[-b]	85.31%[-b]	78.09%[-b]	91.12%[-b]	96.19%
StaticIm05_Split7	81.56%[-b]	72.69%[-b]	79.46%[-b]	84.77%[-b]	41.73%[-b]	45.58%[-b]	76.15%[-b]	65.27%[-b]	85.44%[-b]	89.57%
StaticIm05_Move7	75.06%[-b]	72.68%[-b]	71.52%[-b]	80.41%[-b]	36.53%[-b]	20.72%[-b]	66.29%[-b]	39.12%[-b]	79.14%[-b]	85.35%
StaticIm05_Merge7	87.33%[-b]	86.85%[-b]	85.27%[-b]	90.43%[-b]	51.63%[-b]	65.48%[-b]	82.04%[-b]	69.28%[-b]	88.46%[-b]	93.96%
StaticIm05_Borderline20	96.61%[-b]	99.13%[b]	95.93%[-b]	97.67%[-b]	59.51%[-b]	94.88%[-b]	91.99%[-b]	94.27%[-b]	95.46%[-b]	98.57%
StaticIm05_Borderline100	95.99%[-b]	98.77%[*]	94.97%[-b]	97.42%[-b]	58.26%[-b]	92.52%[-b]	90.37%[-b]	93.4%[-b]	95.41%[-b]	98.78%
StaticIm05_Rare20	92.83%[-b]	92.21%[-b]	92.33%[-b]	93.43%[-b]	46.1%[-b]	73.74%[-b]	88.25%[-b]	90.36%[-b]	93.63%[-b]	94.54%
StaticIm05_Rare100	61.79%[-b]	61.89%[-b]	59.97%[-b]	64.29%[-b]	21.15%[-b]	37.6%[-b]	49.57%[-b]	49.17%[-b]	58.63%[-b]	71.94%
StaticIm03										
StaticIm03_Split3	79.64%[-b]	74.54%[-b]	76.82%[-b]	83.34%[-b]	33.86%[-b]	40.94%[-b]	67.75%[-b]	64.86%[-b]	82.72%[-b]	88.15%
StaticIm03_Move3	74.18%[-b]	76.9%[-b]	70.89%[-b]	78.76%[-b]	26.98%[-b]	13.98%[-b]	58.82%[-b]	45.28%[-b]	77.24%[-b]	83.63%
StaticIm03_Merge3	83.59%[-b]	85.47%[-b]	82.05%[-b]	86.97%[-b]	45.54%[-b]	53.7%[-b]	71.45%[-b]	66.63%[-b]	84.98%[-b]	91.76%
StaticIm03_Split7	72.97%[-b]	71.19%[-b]	69.91%[-b]	76.93%[-b]	25.94%[-b]	38.05%[-b]	60.16%[-b]	51.67%[-b]	78.2%[-b]	81.65%
StaticIm03_Move7	62.67%[-b]	71.82%[-s]	57.33%[-b]	69.13%[-b]	24.09%[-b]	9.58%[-b]	44.13%[-b]	31.13%[-b]	69.21%[-b]	73.56%
StaticIm03_Merge7	79.72%[-b]	85.68%[-m]	76.69%[-b]	83.89%[-b]	40.91%[-b]	55.71%[-b]	66.38%[-b]	63.1%[-b]	82.59%[-b]	87.89%
StaticIm03_Borderline20	93.86%[-b]	98.9%[b]	93.33%[-b]	96.0%[-b]	56.21%[-b]	88.05%[-b]	85.82%[-b]	88.9%[-b]	91.81%[-b]	96.98%
StaticIm03_Borderline100	91.62%[-b]	98.54%[b]	90.6%[-b]	95.28%[-b]	47.88%[-b]	76.88%[-b]	75.17%[-b]	78.49%[-b]	87.32%[-b]	97.23%
StaticIm03_Rare20	90.76%[-b]	91.57%[-b]	90.13%[-b]	92.13%[-b]	43.55%[-b]	64.12%[-b]	80.52%[-b]	88.84%[-b]	91.43%[-b]	93.24%
StaticIm03_Rare100	58.56%[-b]	63.55%[-b]	56.97%[-b]	62.83%[-b]	19.02%[-b]	33.07%[-b]	43.29%[-b]	45.14%[-b]	58.47%[-b]	69.79%

- Based on the average G-Mean, cells are highlighted in lime / light grey when SMOClust performed better than the corresponding approach and cells are highlighted in orange / dark grey cells when SMOClust performed worse than the corresponding approach. The colour saturation scales with the absolute difference of average G-Mean between the SMOClust and the approach of the column and the saturation reaches the maximum when such difference is $\geq 10\%$.
- Symbols [*], [s], [m] and [b] represent insignificant, small, medium and large A12 effect size against SMOClust. Presence/absence of the sign "-" in the effect size means that the corresponding approach was worse/better than SMOClust.

4 Results with Real-world Data Streams

This section resents the comprehensive results of the predictive performance of approaches on real-world data streams.

- Correspond to Figure 4 in the paper:
  - Figure 7 presents the difference of the thirty runs average G-Mean of the compared approaches against SMOClust on real-world data streams.
  - Table 7 presents the thirty runs average G-Mean of all approaches on real-world data streams and the A12 effect size results of comparing existing approaches against SMOClust.

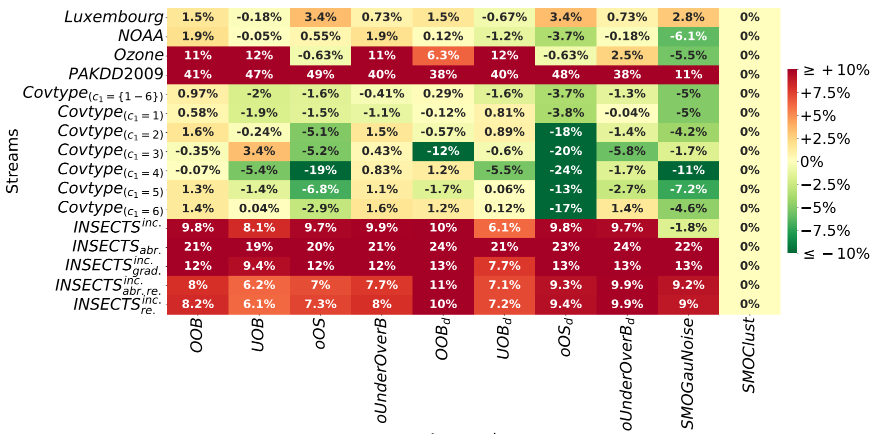

Fig. 7: 30 Runs Average Prequential G-Mean on Real-World Data Streams (Difference against SMOClust; Green cells indicate SMOClust performed better; Red cells indicate SMOClust performed worse)

Table 7: 30 Runs Average Prequential G-Mean on Real-World Data Streams (A12 SMOClust vs Others)

Groups	OOB	UOB	oOS	oUnderOverB	$\mathrm{OOB_d}$	$\mathrm{UOB_d}$	$\mathrm{oOS_d}$	$\mathrm{oUnderOverB_d}$	SMOGauNoise	SMOClust
Luxembourg	93.29%[b]	91.58%[*]	95.15%[b]	92.49%[s]	93.29%[b]	91.09%[-m]	95.15%[b]	92.49%[s]	94.6%[b]	91.76%
NOAA	71.44%[b]	69.47%[s]	70.07%[b]	71.39%[b]	69.64%[*]	68.28%[-b]	65.85%[-b]	69.34%[*]	63.4%[-b]	69.52%
Ozone	65.7%[b]	66.49%[b]	54.03%[-m]	65.58%[b]	60.92%[b]	66.44%[b]	54.03%[-m]	57.12%[b]	49.17%[-b]	54.66%
PAKDD2009	50.84%[b]	56.8%[b]	57.91%[b]	49.64%[b]	47.56%[b]	49.46%[b]	57.2%[b]	47.8%[b]	20.84%[b]	9.36%
$\mathrm{Covtype}_{(c_1=\{1\text{-}6\})}$	91.49%[b]	88.55%[-b]	88.93%[-b]	90.11%[-b]	90.81%[b]	88.91%[-b]	86.84%[-b]	89.19%[-b]	85.47%[-b]	90.52%
$\mathrm{Covtype}_{(c_1=1)}$	90.59%[b]	88.1%[-b]	88.51%[-b]	88.91%[-b]	89.89%[-b]	90.82%[b]	86.2%[-b]	89.97%[*]	85.0%[-b]	90.01%
$\mathrm{Covtype}_{(c_1=2)}$	67.14%[b]	65.32%[-b]	60.45%[-b]	67.08%[b]	64.99%[-b]	66.45%[b]	47.65%[-b]	64.16%[-b]	61.38%[-b]	65.56%
$\mathrm{Covtype}_{(c_1=3)}$	56.88%[-m]	60.58%[b]	52.02%[-b]	57.66%[b]	45.26%[-b]	56.63%[-b]	37.73%[-b]	51.46%[-b]	55.49%[-b]	57.23%
$\mathrm{Covtype}_{(c_1=4)}$	90.05%[-s]	84.73%[-b]	71.16%[-b]	90.95%[b]	91.31%[b]	84.67%[-b]	65.71%[-b]	88.47%[-b]	79.16%[-b]	90.12%
$\mathrm{Covtype}_{(c_1=5)}$	65.98%[b]	63.22%[-b]	57.8%[-b]	65.74%[b]	62.93%[-b]	64.7%[m]	51.36%[-b]	61.95%[-b]	57.47%[-b]	64.64%
$\mathrm{Covtype}_{(c_1=6)}$	68.76%[b]	67.44%[s]	64.53%[-b]	69.01%[b]	68.65%[b]	67.52%[b]	50.34%[-b]	68.84%[b]	62.78%[-b]	67.4%
$\mathrm{INSECTS}^{\mathrm{inc.}}$	74.91%[b]	73.22%[b]	74.8%[b]	74.95%[b]	75.33%[b]	71.2%[b]	74.87%[b]	74.77%[b]	63.31%[-b]	65.09%
$\mathrm{INSECTS}_{\mathrm{abr.}}$	73.21%[b]	70.56%[b]	71.65%[b]	72.91%[b]	75.98%[b]	72.91%[b]	74.76%[b]	75.68%[b]	73.54%[b]	52.0%
$\mathrm{INSECTS}^{\mathrm{inc.}}_{\mathrm{grad.}}$	76.96%[b]	74.43%[b]	77.19%[b]	76.78%[b]	77.93%[b]	72.73%[b]	77.53%[b]	78.14%[b]	77.65%[b]	65.0%
$\mathrm{INSECTS}^{\mathrm{inc.}}_{\mathrm{abr.\ re.}}$	72.11%[b]	70.31%[b]	71.07%[b]	71.81%[b]	74.78%[b]	71.23%[b]	73.35%[b]	74.01%[b]	73.29%[b]	64.09%
$\mathrm{INSECTS}^{\mathrm{inc.}}_{\mathrm{re.}}$	72.75%[b]	70.64%[b]	71.78%[b]	72.49%[b]	75.0%[b]	71.67%[b]	73.88%[b]	74.4%[b]	73.49%[b]	64.51%

- Based on the average G-Mean, cells are highlighted in lime / light grey when SMOClust performed better than the corresponding approach and cells are highlighted in orange / dark grey cells when SMOClust performed worse than the corresponding approach. The colour saturation scales with the absolute difference of average G-Mean between the SMOClust and the approach of the column and the saturation reaches the maximum when such difference is $\geq 10\%$.
- Symbols [*], [s], [m] and [b] represent insignificant, small, medium and large A12 effect size against SMOClust. Presence/absence of the sign "-" in the effect size means that the corresponding approach was worse/better than SMOClust.