Evaluating Different Classification Algorithms on SEA Dataset

İrem Ecem Yelkanat
İhsan Doğramacı Bilkent University
Ankara, Turkey
ecem.yelkanat@ug.bilkent.edu.tr

1 Dataset Generation

To evaluate the performance of different classification models, three different datasets consisting of 20000 instances was generated using SEA Generator provided by sckit-multiflow, and was written to corresponding files. Each data instance has three features and a corresponding feature which is either 0 or 1. First dataset that was generated, which is named “SEA Dataset”, has a noice percentage of 0. The other two datasets, which are named “SEA Dataset 10” and “SEA Dataset 70”, have noice percentages of 10 and 70 respectively. As mentioned before, the generated datasets are recorded in files to reuse them easily without creating a new dataset in each run.

2 Data Stream Classification with Three Separate Online Single Classifiers: HT, KNN, MLP

Three different online single classifiers were used to evaluate the performances in terms of accuracy, which are Hoeffding Tree (HT), K-Nearest Neighbor (KNN) and Multilayer Perceptron (MLP) which is composed of 4 hidden layers of 200 neurons as online learners. Interleaved-Test-Then-Train approach was utilized to evaluate the temporal accuracies of those online classifiers with a batch size of 1. The results of these three classifiers on three datasets can be seen in *Figure 1*, *Figure 2* and *Figure 3*.

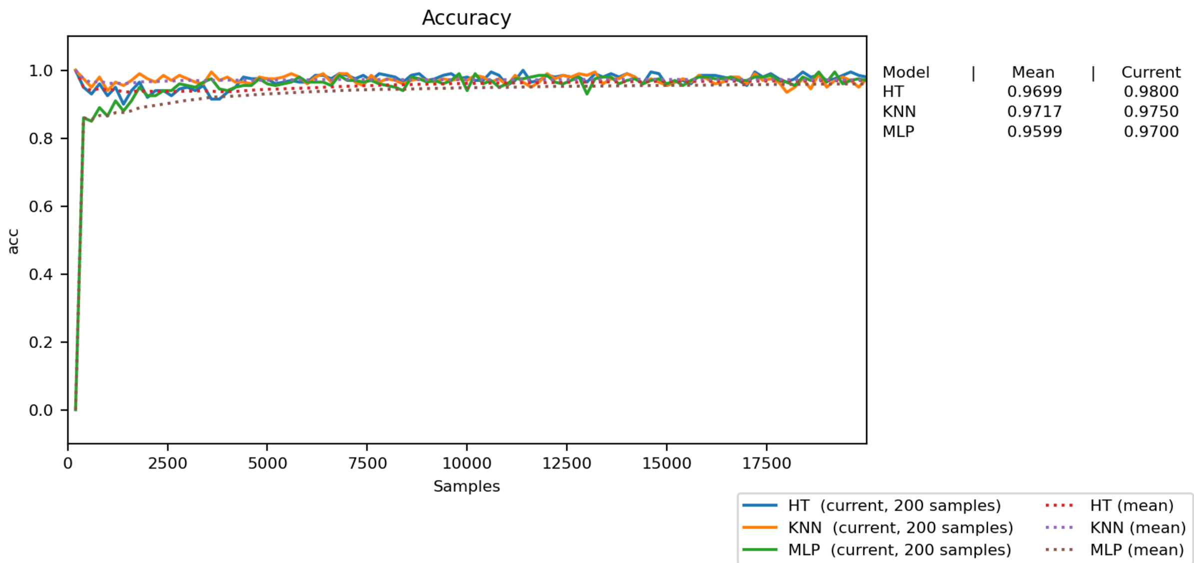

Figure 1: Temporal Accuracies of HT, KNN, MLP on SEA Dataset

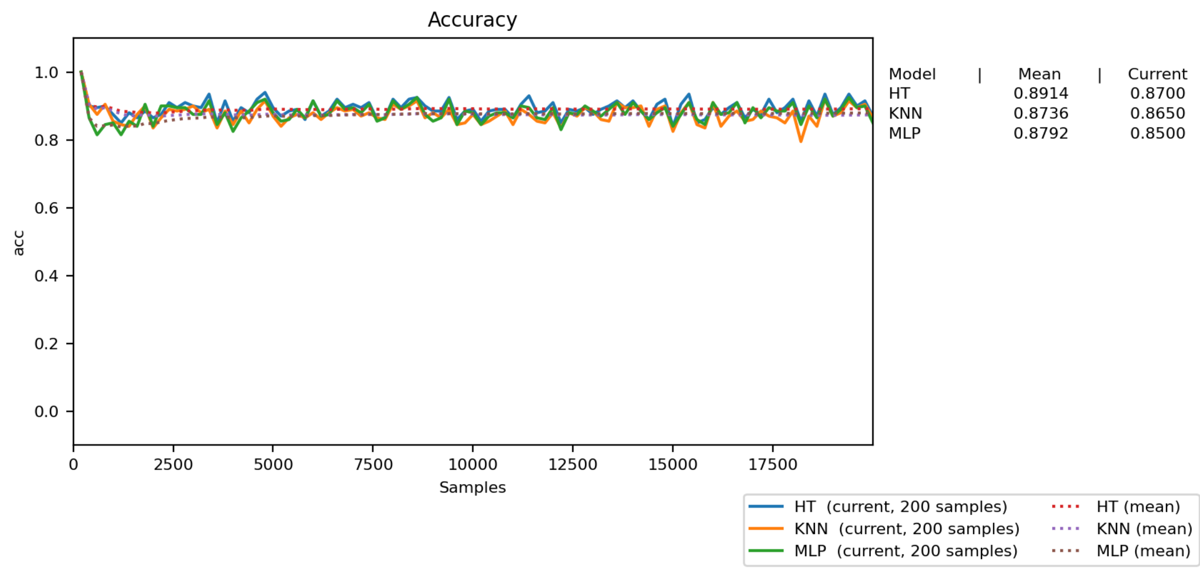

Figure 2: Temporal Accuracies of HT, KNN, MLP on SEA Dataset 10

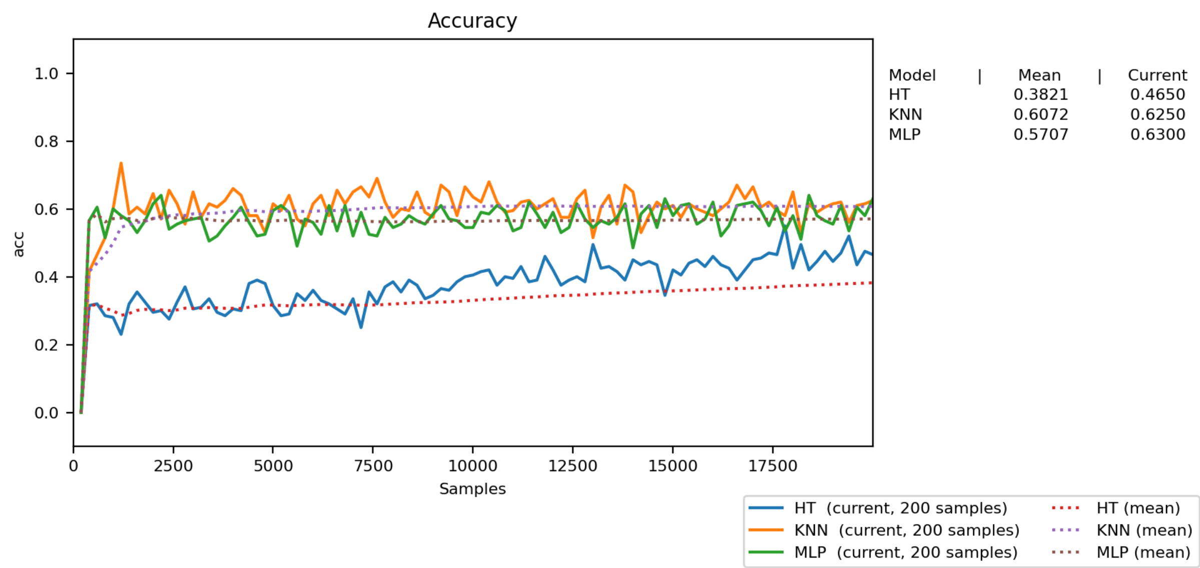

Figure 3: Temporal Accuracies of HT, KNN, MLP on SEA Dataset 70

For SEA Dataset, all classifiers performed similarly in terms of overall accuracy, with HT accuracy being 0.9699, KNN accuracy being 0.9717 and MLP accuracy being 0.9599. For SEA Dataset 10, all classifiers performed similarly in terms of overall accuracy, with HT accuracy being 0.8914, KNN accuracy being 0.8736 and MLP accuracy being 0.8792. For SEA Dataset 70, the results are considerably different compared to the results of SEA

Dataset and SEA Dataset 10. The overall accuracy for SEA Dataset 70 is 0.3821 for HT, 0.6072 for KNN and 0.5707 for MLP. These results can be seen more clearly in *Table 1*.

Classifier / Dataset	**SEA Dataset**	**SEA Dataset 10**	**SEA Dataset 70**
HT	0.9699	0.8914	0.3821
KNN	0.9717	0.8736	0.6072
MLP	0.9599	0.8792	0.5707

Table 1: Overall Accuracies of Online Single Classifiers on Datasets

When temporal accuracies are considered, the differences and similarities between the online classifiers can be seen in the figures *Figure 1*, *Figure 2* and *Figure 3*. In the case of SEA Dataset, until the 500th sample, temporal accuracies of HT and KNN are similar whereas the temporal accuracy of MLP is quite low compared to HT and KNN. Furthermore, after the 5000th sample, the temporal accuracies of HT, KNN and MLP are very similar to each other with minor differences. In the case of SEA Dataset 10, the temporal accuracies of HT, KNN and MLP are very similar to each other with minor differences, like SEA Dataset. In the case of SEA Dataset 70, the temporal accuracies of HT, KNN and MLP are very close to each other until the 400th sample, after this sample, the temporal accuracies differs considerably similar to overall accuracies. While the classifier of best temporal accuracy is shifted between KNN and MLP, the temporal accuracies of HT is below both KNN and MLP.

When comparing the accuracies for the three datasets, the classification algorithms performed best on SEA Dataset and worst for SEA Dataset 70. This is because, when the noice in the data is increased, concept drift is introduced to the data, and the capacity of the online classification algorithms are not sufficient to capture these differences in the data. When comparing the accuracies of three classifiers, the results are very similar for SEA Dataset and SEA Dataset 10, however, there is a considerable difference between the accuracies of the classifiers for SEA Dataset 70.

3 Data Stream Classification with Two Online Ensemble Classifiers: MV, WMV

Two different online ensemble classifiers were used to evaluate the performances in terms of accuracy, which are Majority Voting Rule (MV) and Weighted Majority Voting Rule (WMV). These ensemble classifiers combined the online single classifiers introduced in the previous section, which are Hoeffding Tree (HT), K-Nearest Neighbor (KNN) and Multilayer Perceptron (MLP) which is composed of 4 hidden layers of 200 neurons as online learners. Interleaved-Test-Then-Train approach was utilized to evaluate the temporal accuracies of those online ensemble classifiers with a batch size of 1. The results of these two classifiers on three datasets can be seen in *Figure 4*, *Figure 5* and *Figure 6*.

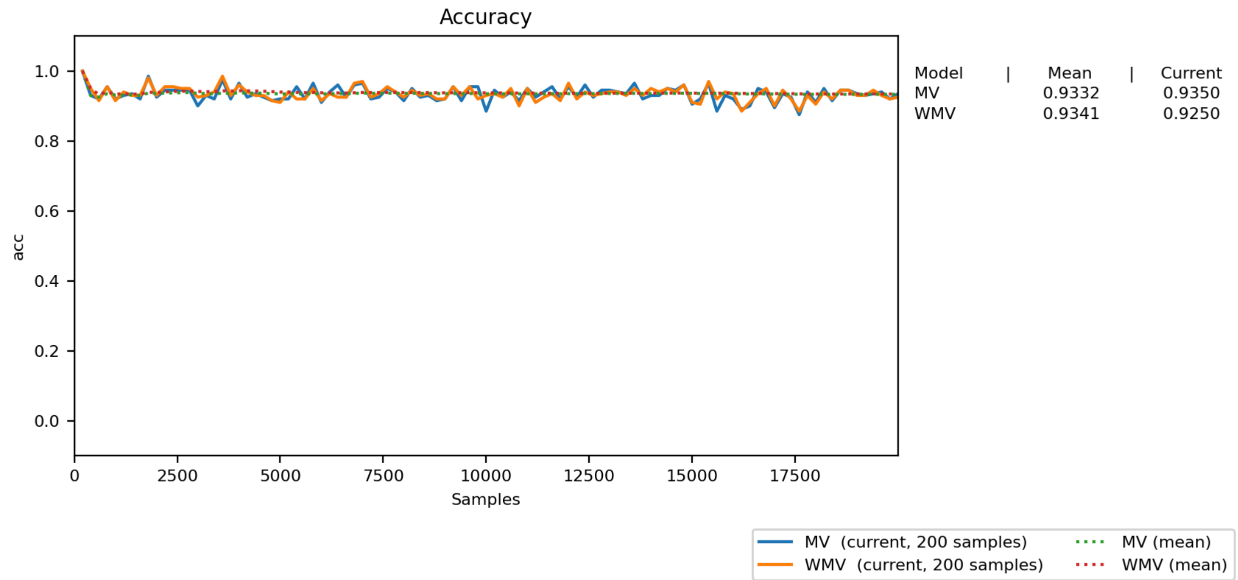

Figure 4: Temporal Accuracies of MV and WMV on SEA Dataset

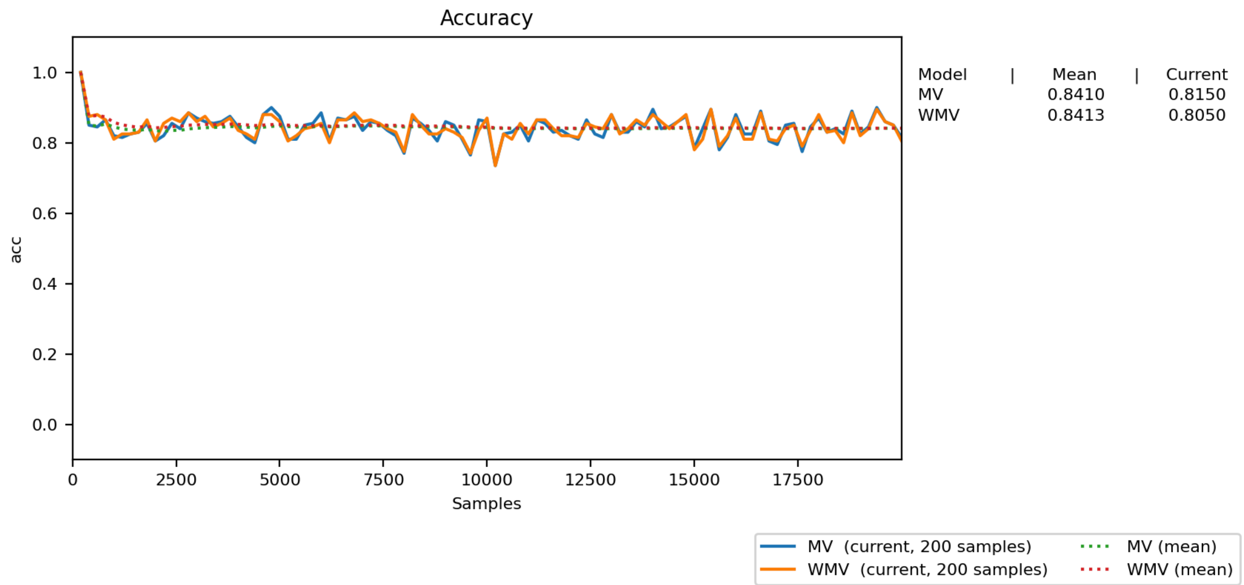

Figure 5: Temporal Accuracies of MV and WMV on SEA Dataset 10

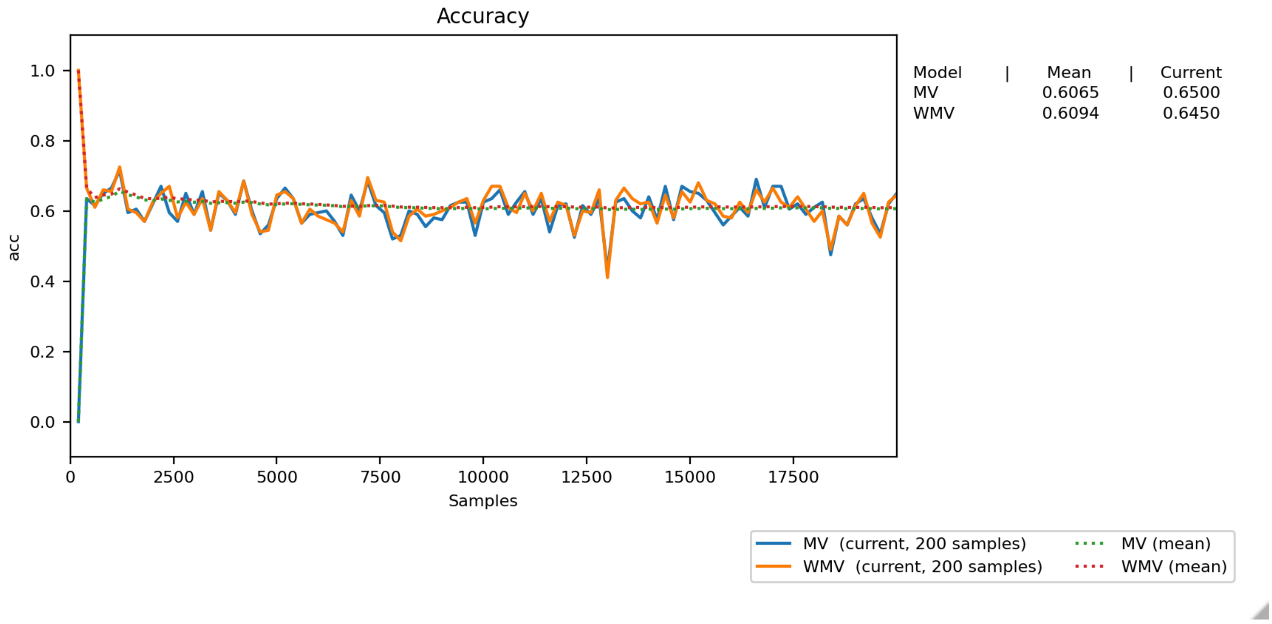

Figure 5: Temporal Accuracies of MV and WMV on SEA Dataset 70

For SEA Dataset, all classifiers performed similarly in terms of overall accuracy, with MV accuracy being 0.9332, and WMV accuracy being 0.9341. For SEA Dataset 10, all classifiers performed similarly in terms of overall accuracy, with MV accuracy being 0.8410 and WMV accuracy being 0.8413. For SEA Dataset 70, all classifiers performed similarly in terms of overall accuracy, with MV accuracy being 0.6065 and WMV accuracy being 0.6094. These results can be seen more clearly in *Table 2*.

Classifier / Dataset	SEA Dataset	SEA Dataset 10	SEA Dataset 70
MV	0.9332	0.8410	0.6065
WMV	0.9341	0.8413	0.6094

Table 2: Overall Accuracies of Online Ensemble Classifiers on Datasets

When temporal accuracies are considered, the differences and similarities between the online ensemble classifiers can be seen in the figures *Figure 4*, *Figure 5* and *Figure 6*. In the case of SEA Dataset, the temporal accuracies of VM and WMV are very similar to each other with minor differences. In the case of SEA Dataset 10, the temporal accuracies of VM and WMV are very similar to each other with minor differences, like SEA Dataset. In the case of SEA Dataset 70, until the 300^{th} sample, the temporal accuracy of MV is quite low compared to WMV. Furthermore, after the 300^{th} sample, the temporal accuracies of MV and WMV are very similar to each other with minor differences.

When comparing the accuracies for the three datasets, MV and WMV performed similarly in the context of the same datasets. In addition, the classification algorithms performed best on SEA Dataset and worst for SEA Dataset 70. This is because, when the noice in the data is increased, concept drift is introduced to the data, and the capacity of the online classification algorithms are not sufficient to capture these differences in the data. When comparing the accuracies of three classifiers, the results are very similar for SEA Dataset and SEA Dataset 10, however, there is a considerable difference between the accuracies of the classifiers for SEA Dataset 70. These results are consistent with the results obtained with online single classifiers.

3 Batch Size vs Accuracy for Online Classifiers

The online single classifiers and online ensemble classifiers were trained with different batch sizes to evaluate whether the batch size influences the performance of these methods. The results can be seen in *Table 3*.

Classifier	Dataset	Batch Size	Accuracy
HT	SEA Dataset	1	0.9699
HT	SEA Dataset	10	0.9698
HT	SEA Dataset	100	0.9690
HT	SEA Dataset	1000	0.9711
HT	SEA Dataset 10	1	0.8914
HT	SEA Dataset 10	10	0.8782
HT	SEA Dataset 10	100	0.8776
HT	SEA Dataset 10	1000	0.8781
HT	SEA Dataset 70	1	0.3821
HT	SEA Dataset 70	10	0.6652
HT	SEA Dataset 70	100	0.6644
HT	SEA Dataset 70	1000	0.6660

Table 3: Batch Sizes vs Accuracy for HT

As it can be seen in *Table 3*, changing the batch size does not affect the performance of the HT model significantly except for the SEA Dataset 70, in which the accuracy is nearly doubled between the batch sizes 1 and 10, however, after batch size of 10, the accuracy values does not change considerable either.

Classifier	Dataset	Batch Size	Accuracy
KNN	SEA Dataset	1	0.9717
KNN	SEA Dataset	10	0.9718
KNN	SEA Dataset	100	0.9714
KNN	SEA Dataset	1000	0.9711
KNN	SEA Dataset 10	1	0.8736
KNN	SEA Dataset 10	10	0.8729
KNN	SEA Dataset 10	100	0.8730
KNN	SEA Dataset 10	1000	0.8710
KNN	SEA Dataset 70	1	0.6072
KNN	SEA Dataset 70	10	0.6114
KNN	SEA Dataset 70	100	0.6105
KNN	SEA Dataset 70	1000	0.6083

Table 4: Batch Sizes vs Accuracy for KNN

As it can be seen in *Table 4*, changing the batch size does not affect the performance of the KNN model significantly.

Classifier	Dataset	Batch Size	Accuracy
MLP	SEA Dataset	1	0.9599
MLP	SEA Dataset	10	0.9386
MLP	SEA Dataset	100	0.7995
MLP	SEA Dataset	1000	0.8199
MLP	SEA Dataset 10	1	0.8792
MLP	SEA Dataset 10	10	0.8535
MLP	SEA Dataset 10	100	0.4717
MLP	SEA Dataset 10	1000	0.7470
MLP	SEA Dataset 70	1	0.5707
MLP	SEA Dataset 70	10	0.6517
MLP	SEA Dataset 70	100	0.6594
MLP	SEA Dataset 70	1000	0.6245

Table 5: Batch Sizes vs Accuracy for MLP

As it can be seen in *Table 5*, changing the batch size affects the performance of the MLP classifier. For SEA Dataset, the best performance is obtained when batch size is 1 and the worst performance is obtained when batch size is 100. For SEA Dataset 10, the best performance is obtained when batch size is 1 and the worst performance is obtained when batch size is 100, similar to SEA Dataset. For SEA Dataset 70, the best performance is obtained when batch size is 100 and the worst performance is obtained when batch size is 1. From these results, it cannot be concluded that there is a correlation between the batch size and the performance of the MLP model in terms of accuracy, since the accuracy values is not consistent within different datasets. Hence, there is no general conclusion about batch size vs performance.

Classifier	**Dataset**	**Batch Size**	**Accuracy**
MV	SEA Dataset	1	0.9332
MV	SEA Dataset	10	0.9347
MV	SEA Dataset	100	0.9326
MV	SEA Dataset	1000	0.9302
MV	SEA Dataset 10	1	0.8410
MV	SEA Dataset 10	10	0.8404
MV	SEA Dataset 10	100	0.8401
MV	SEA Dataset 10	1000	0.8473
MV	SEA Dataset 70	1	0.6065
MV	SEA Dataset 70	10	0.6062
MV	SEA Dataset 70	100	0.6091
MV	SEA Dataset 70	1000	0.6023

Table 6: Batch Sizes vs Accuracy for MV

As it can be seen in *Table 6*, changing the batch size does not affect the performance of the MV model significantly.

Classifier	Dataset	Batch Size	Accuracy
WMV	SEA Dataset	1	0.9341
WMV	SEA Dataset	10	0.9346
WMV	SEA Dataset	100	0.9326
WMV	SEA Dataset	1000	0.9318
WMV	SEA Dataset 10	1	0.8413
WMV	SEA Dataset 10	10	0.8395
WMV	SEA Dataset 10	100	0.8395
WMV	SEA Dataset 10	1000	0.8452
WMV	SEA Dataset 70	1	0.6094
WMV	SEA Dataset 70	10	0.6063
WMV	SEA Dataset 70	100	0.6108
WMV	SEA Dataset 70	1000	0.6105

Table 7: Batch Sizes vs Accuracy for WMV

As it can be seen in *Table 7*, changing the batch size does not affect the performance of the MV model significantly.

Considering all of the results, changing the batch size does not affect the performance of classifiers considerably for HT, KNN, MLP, MV and WMV, except for classification of SEA Dataset 70 with HT and MLP classifier. However, MLP classifier does not show any correlation. Hence, it can be said that there is no positive or negative correlation between the batch size used for the classifiers with the given data and obtained results.

4 Ensemble Methods vs Individual Models (Online)

With the given data and obtained results, in section 3 in *Table 3* through *Table 7*, it can be said that online ensemble methods are not better than individual models as when comparing the maximum accuracies of classifiers for a specific dataset, the accuracies for the ensemble methods are always lower than the individual methods. However, since the accuracies of ensemble models with different batch sizes are consistent and it is the opposite case for MLP, ensemble methods can be used instead of MLP model.

Finally, above conclusions may not apply to all datasets, since the results may depend on the content of the dataset and the problem that is tried to address.

5 Batch Classification

For the batch classification, the three datasets were split into two parts individually as training set and test set with a ratio of 20%, that is, 4000 samples were belong to test set and 16000 samples were belong to training set. The models were trained with training set and their performances were evaluated based on their accuracies on the test set.

6 Batch Classification with Three Separate Batch Single Classifiers: HT, KNN, MLP

Three different batch single classifiers were used to evaluate the performances in terms of accuracy, which are Hoeffding Tree (HT), K-Nearest Neighbor (KNN) and Multilayer Perceptron (MLP) which is composed of 4 hidden layers of 200 neurons as batch learners. Test set is used to evaluate the accuracies of those batch classifiers. The results of these three classifiers on three datasets can be seen in *Table 8*.

Classifier / Dataset	SEA Dataset	SEA Dataset 10	SEA Dataset 70
HT	0.9835	0.894	0.34175
KNN	0.9645	0.863	0.609
MLP	0.99725	0.9035	0.6825

Table 8: Overall Accuracies of Batch Single Classifiers on Datasets

As it can be seen in Table 8, for each dataset SEA Dataset, SEA Dataset 10 and SEA Dataset 70, MLP performed better compared to KNN and HT. The batch classifiers performed similarly on SEA Dataset. For SEA Dataset 10, the pairs of accuracies differ at most 4%, and the proportion of accuracy increased between the top first and top second classifiers is approximately %4. In SEA Dataset 70, the MLP performed approximately 13% better that its closest competitor, which is KNN, and MLP doubled the performance of HT, which is considerably large.

When comparing the accuracies for the three datasets, the classification algorithms performed best on SEA Dataset and worst for SEA Dataset 70. This is because, when the noice in the data is increased, concept drift is introduced to the data, and the capacity of the online classification algorithms are not sufficient to capture these differences in the data. When comparing the accuracies of three classifiers, the results are very similar for SEA Dataset and SEA Dataset 10, however, there is a considerable difference between the accuracies of the classifiers for SEA Dataset 70.

7 Batch Classification with Two Batch Ensemble Classifiers: MV, WMV

Two different batch ensemble classifiers were used to evaluate the performances in terms of accuracy, which are Majority Voting Rule (MV) and Weighted Majority Voting Rule (WMV). These ensemble classifiers combined the batch single classifiers introduced in the previous section, which are Hoeffding Tree (HT), K-Nearest Neighbor (KNN) and Multilayer Perceptron (MLP) which is composed of 4 hidden layers of 200 neurons as batch classifiers. The results of these two classifiers on three datasets can be seen in *Table 9*.

Classifier / Dataset	SEA Dataset	SEA Dataset 10	SEA Dataset 70
MV	0.9875	0.87125	0.678
WMV	0.9665	0.89175	0.68

Table 9: Overall Accuracies of Batch Ensemble Classifiers on Datasets

As it can be seen in Table 9, the overall accuracy of MV and WMV on SEA Dataset are very similar to each other with minor differences. For SEA Dataset 10, the overall accuracy of MV and WMV are very similar to each other with minor differences. Similarly, for SEA Dataset 70, the overall accuracy of MV and WMV are very similar to each other with minor differences.

When comparing the accuracies for the three datasets, MV and WMV performed similarly in the context of the same datasets. In addition, the classification algorithms performed best on SEA Dataset and worst for SEA Dataset 70. This is because, when the noice in the data is increased, concept drift is introduced to the data, and the capacity of the online classification algorithms are not sufficient to capture these differences in the data. When comparing the accuracies of three classifiers, the results are very similar for SEA Dataset and SEA Dataset 10, however, there is a considerable difference between the accuracies of the classifiers for SEA Dataset 70. These results are consistent with the results obtained with batch single classifiers.

8 Ensemble Methods vs Individual Models (Batch)

With the given data and obtained results, in section 7 in *Table 8* and *Table 9*, it can be said that batch ensemble methods are not better than individual models as when comparing the maximum accuracies of classifiers for a specific dataset, the accuracies for the ensemble methods are always lower than the individual methods. However, since the accuracies of ensemble models are always better than the minimum accuracies of single batch models, especially for noisy datasets, ensemble methods can be used instead of HT and KNN models, especially HT model.

Finally, above conclusions may not apply to all datasets, since the results may depend on the content of the dataset and the problem that is tried to address.

9 Comparing All Models In Terms of Accuracy

Up to this point, online single classifiers and online ensemble classifiers are compared within themselves and batch single classifiers and batch ensemble classifiers are compared within themselves. In *Table 10*, the overall accuracies of all models that are discussed can be seen.

Classifier / Dataset	SEA Dataset	SEA Dataset 10	SEA Dataset 70
HT (Online)	0.9699	0.8914	0.3821
KNN (Online)	0.9717	0.8736	0.6072
MLP (Online)	0.9599	0.8792	0.5707
MV (Online)	0.9332	0.8410	0.6065
WMV (Online)	0.9341	0.8413	0.6094
HT (Batch)	0.9835	0.894	0.34175
KNN (Batch)	0.9645	0.863	0.609
MLP (Batch)	0.99725	0.9035	0.6825
MV (Batch)	0.9875	0.87125	0.678
WMV (Batch)	0.9665	0.89175	0.68

Table 10: Overall Accuracies of All Models

As it can be seen in Table 10, batch models have better performance compared to online models with some exceptions. The differences among the performance can be see clearly for SEA Dataset 70, where the batch MLP model performed 18% better than online MLP model and batch MV model performed 12% better than online MV model. The differences between the online models and their corresponding batch models are more increased as the noice in the data is increased. Hence, it can be said that batch models perform much better on noisy data. The main reason behind these differences is that online models use Interleaved-Test-Then-Train approach, therefore data is fed into the model in small batches, and since the online model does not aware of the whole dataset unlike batch learning, in which all of the training data is fed into model at once, it is expected that online models performs poorly compared to batch models.

9 Models and Their Efficiencies

Classifier / Dataset	SEA Dataset	SEA Dataset 10	SEA Dataset 70
HT & KNN & MLP (Online)	34.82 s	33.17 s	34.07 s
MV & WMV (Online)	80.15 s	76.21 s	76.47 s
HT & KNN & MLP (Batch)	9.36 s	8.89 s	9.04 s
MV & WMV (Batch)	21.13 s	18.76 s	15.32 s

Table 11: Overall Accuracies of All Models

As it can be seen in *Table 11*, processing time of single online models is less than processing time of online ensemble models. Likewise, processing time of single batch models is less than processing time of batch ensemble models. This is as expected since ensemble models require more computational power than single models as in ensemble models three different single models are run and the predictions are compared to produce a final prediction. In addition, processing time of single batch models is less than processing time of single online models. Likewise, processing time of ensemble batch models is less than the processing time of online ensemble models. This is as expected since online models require more computational power that batch models as in online learning, for each new data the model is partially fit whereas in batch modes, data can be fitted fully or in small batches. Finally, there are no considerable differences between the models in terms of used datasets except for MV & WMV for batch learning, in which the running times differed 40% between SEA Dataset and SEA Dataset 70. However, it can be said that the dataset that is used does not have considerable affect on running time with the given data and results of observations.

10 Improving the Models

In the analysis of model performances, it can be seen that the performance of the models are decreasing when the noice in the data is increasing. As mentioned before, this is because, when the noice in the data is increased, concept drift is introduced to the data, and the capacity of the online classification algorithms are not sufficient to capture these differences in the data.

For online single classifiers, Bagging, Boosting and Random Forest models, which are ensemble models, can be used to improve the performance. The existing algorithms which are HT, KNN and MLP can be combined with ADWIN drift detection algorithm to overcome this problem of concept drift. For example, from sckit-multiflow library, KNNADWINClassifier class can be used as an extension to KNN classifier, In addition to extending the existing methods, newly developed ensemble methods can be used, such as Dynamic Weighted Majority (DWM).

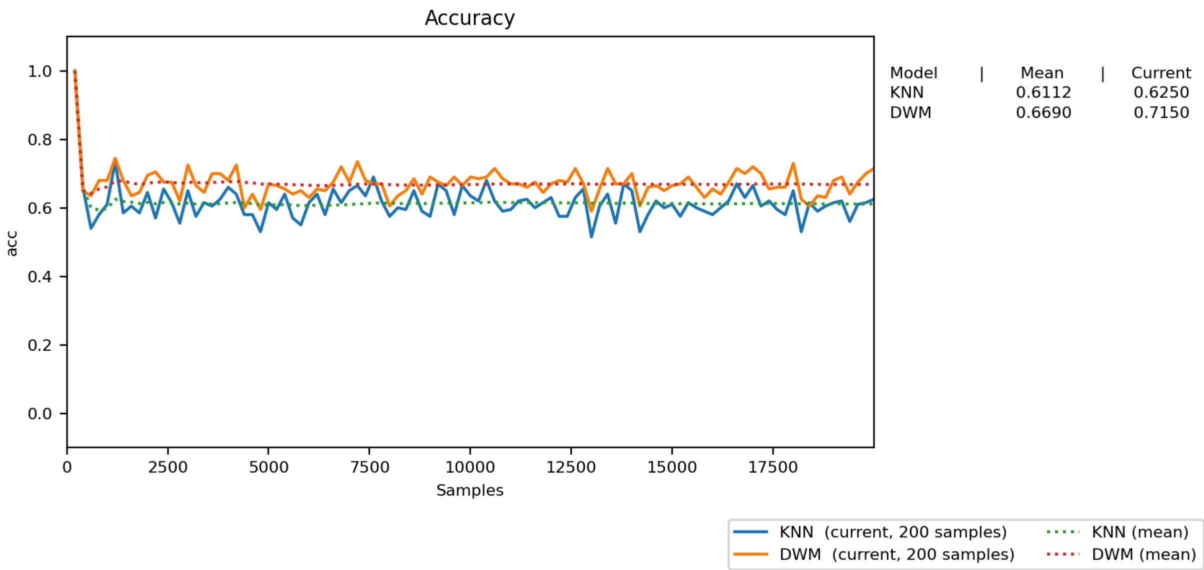

Figure 6: Performance of KNN vs DWM

To compare the performance and show the performance improvement clearly, KNN and DWN are compared in terms of temporal and overall accuracies. As it can be seen in *Figure 6*, using KNN as a base estimator for DWN improves the prediction accuracy compared to vanilla KNN.