# GE461 – Introduction to Data Science

Muhammad Arham Khan
21701848
Homework 5
15th May, 2021

---

**Part1: Dataset Generation**
Data generated and attached along with submission.

**Part 2: Data Stream Classification with Three Separate Online Single Classifiers: HT, KNN, MLP**
The temporal accuracy score outputs for the HT, KNN and MLP classifiers on all datasets (SEA Data, SEA Data 10, SEA Data 70) are as follows:

- HT | SEA Data: 0.9694
- HT | SEA Data 10: 0.8967
- HT | SEA Data 70: 0.3470

- KNN | SEA Data: 0.9710
- KNN | SEA Data 10: 0.8777
- KNN | SEA Data 70: 0.6058

- MLP | SEA Data: 0.9511
- MLP | SEA Data 10: 0.8813
- MLP | SEA Data 70: 0.6520

**Part 3: Data Stream Classification with Two Online Ensemble Classifiers: MV, WMV**
The temporal accuracy score outputs for the two online ensemble classifiers of majority voting (MV) and weighted majority voting (WMV) approach are as follows:

- MV | SEA Data : 0.99175
- MV | SEA Data 10: 0.89355
- MV | SEA Data 70: 0.66675
- WMV | SEA Data : 0.98885
- WMV | SEA Data 10: 0.89305
- WMV | SEA Data 70: 0.44055

**Part 4: Batch Classification with Three Separate Batch Single Classifiers: HT, KNN, MLP**
The temporal accuracy score outputs for the HT, KNN and MLP classifiers on all datasets (SEA Data, SEA Data 10, SEA Data 70) are as follows:

- HT | SEA Data: 0.983
- HT | SEA Data 10: 0.90275
- HT | SEA Data 70: 0.3495

- KNN | SEA Data: 0.97025
- KNN | SEA Data 10: 0.8755
- KNN | SEA Data 70: 0.60775

- MLP | SEA Data: 0.9955
- MLP | SEA Data 10: 0.9115
- MLP | SEA Data 70: 0.66825

**Part 5: Batch Classification with Two Batch Ensemble Classifiers: MV, WMV**

The temporal accuracy score outputs for the two batch ensemble classifiers of majority voting (MV) and weighted majority voting (WMV) approach are as follows:

- MV | SEA Data: 0.9925
- MV | SEA Data 10: 0.90675
- MV | SEA Data 70: 0.65875
- WMV | SEA Data: 0.99125
- WMV | SEA Data 10: 0.90725
- WMV | SEA Data 70: 0.66175

**Part 6: Comparison of Models**

- Figure analysis and comparison of method accuracies:
    - KNN approach provides the best accuracy for SEA Dataset at a temporal accuracy score of 0.9710. But upon analyzing the MV and WMV approaches in reference to this in online ensemble classification, both approaches outperform the KNN approach as MV's temporal accuracy score is 0.99175 and WMV's is 0.9885.

    - MLP approach provides the best accuracy for SEA Dataset 10 at a temporal accuracy score of 0.6520. Again, both MV and WMV approaches outperform as compared the to highest approach of MLP as MV is 0.89355 and WMV is 0.89305.

    - HT approach provides the best accuracy for SEA Dataset 70 at a temporal accuracy score of 0.8967. In this case although, HT performs better than the MV and the WMV approach as MV's temporal accuracy is 0.66675 and WMV's accuracy is 0.44055 which is lower than HT or MLP and KNN for that matter

- So, as identified by the accuracy data collected, except for the case of SEA Data 70, ensemble methods are better than the single individual models as they provide a higher accuracy than single methods. But for Data 70, considering the higher amount of noise in the dataset, the ensemble methods performed better because of their mediation logic.
- Talking about batch classifiers vs. the online counterparts, the batch classifier accuracies are lower than the online classifiers in the cases of individual classifiers. But in the case of ensemble methods, the batch classifiers have a higher accuracy than the online counterparts.

**Plots from SEA individual online classifications on Data, Data 10, Data 70**

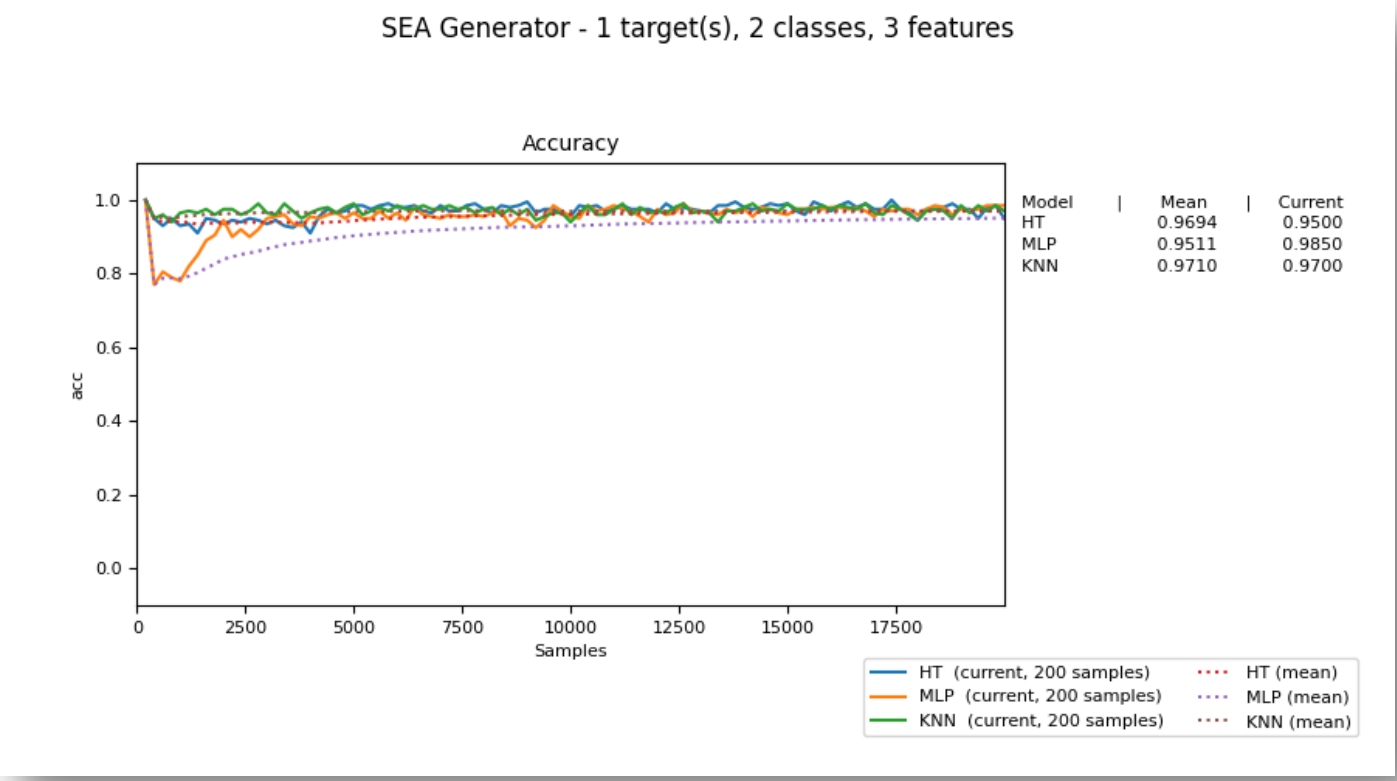


*Figure 1:Online single classifier from SEA Data*

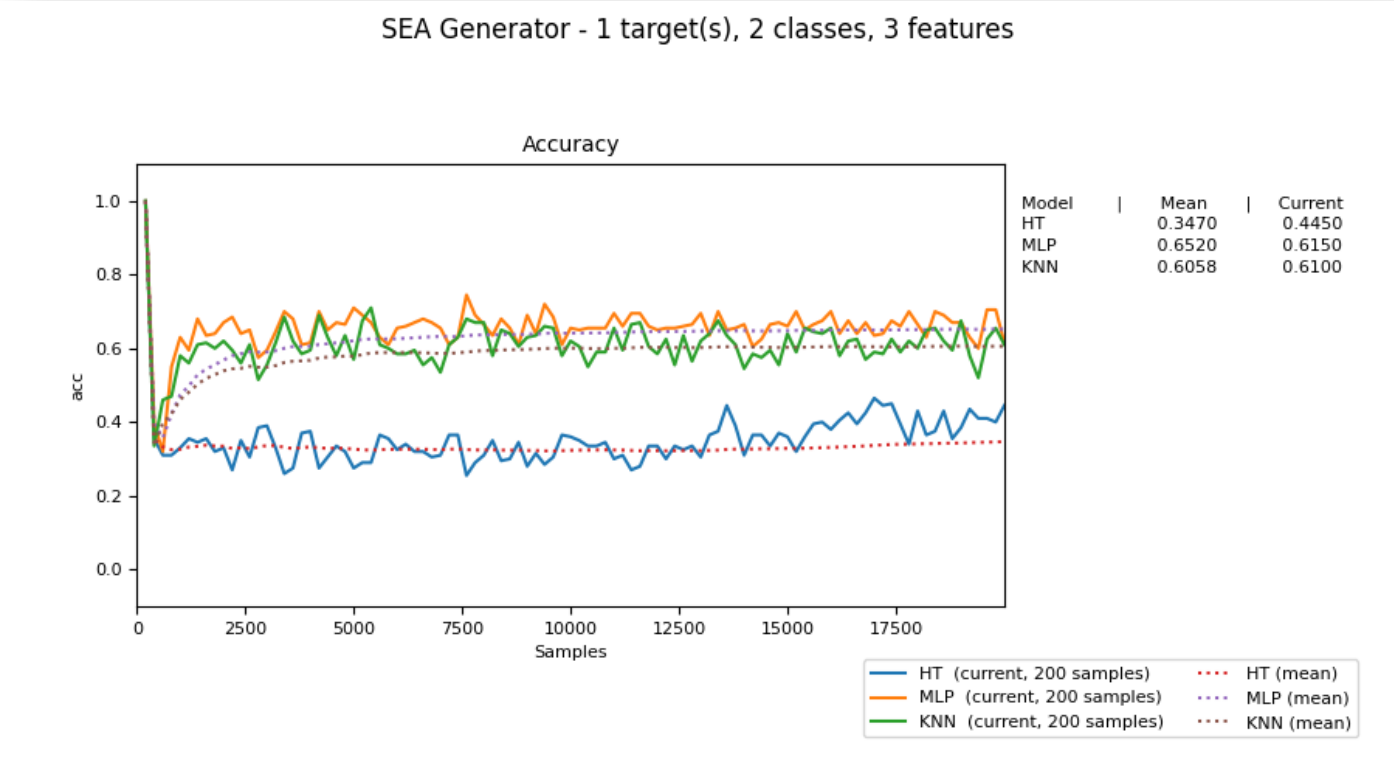


*Figure 2: Online single classifier from SEA Data 10*

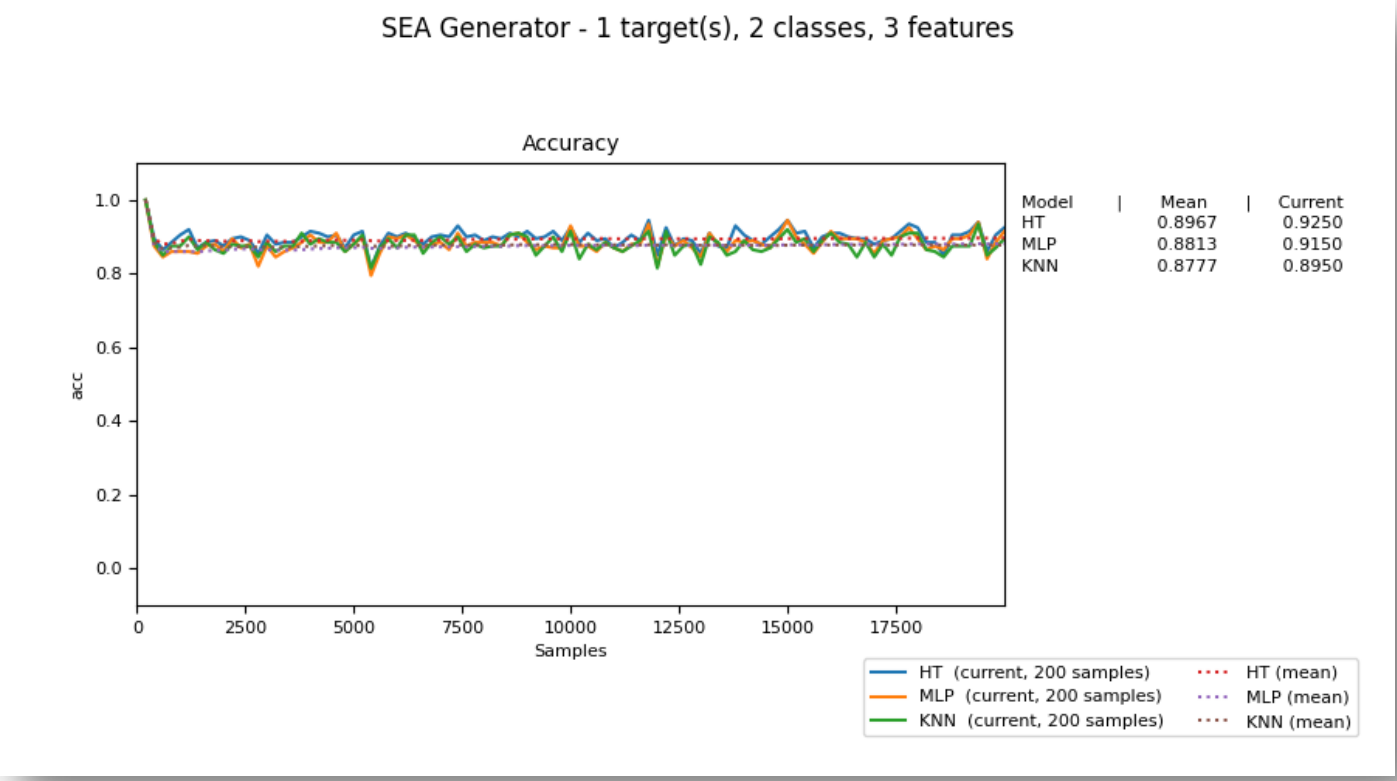


*Figure 3: Online single classifier from SEA Data70*