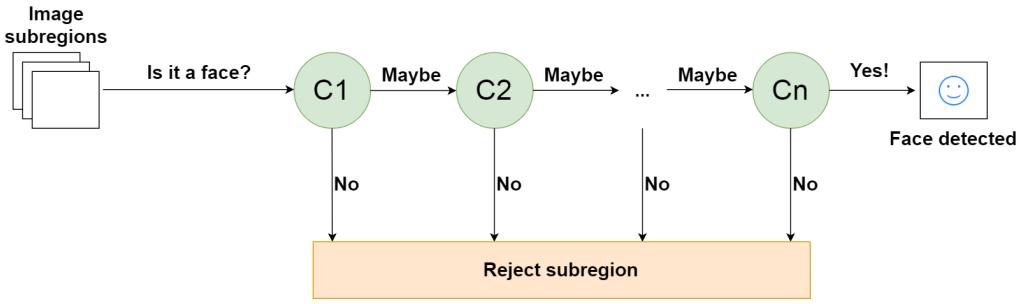
Machine Learning - Assignment 2

Maor Sagi

Introduction

In this assignment we requested to implement a basic cascading meta-learner structure and measure its performance using the Log-loss metric. Later try to improve it by using different engineering methods.

The Cascading ensemble technique is used when we want to be absolutely sure about the accuracy of the result. The architecture of this kind of model contains a sequence of models, while each model should be confident in the result more than some threshold that is determined. Otherwise, the process would propagate to the next model, until we reach a confident prediction or get to the last model in the sequence.



A cascade of n classifiers for face detection. Image taken from here.

Meta-Learner Construction

In this assignment I built a sequence of classifiers starting from the weakest low-resource classifier and as long as it went deeper it would use more complex classifiers. The prediction process would stop when it reaches a confident prediction, or has already used the most resource-intensive classifier.

My Cascading Classifier was constructed in a few phases. In the first phase I used basic configuration:

1. My base classifier is a simple decision tree. First tree depth is 1, and the max depth of the next trees would increase as long as we go deeper.
2. The threshold is set to 95%.
3. The cascading depth is 15.

The cascading classifier evaluated using the Log-loss metric.

Each Classifier could be trained in 2 ways, one is by classifying the predictions of the model by confidence higher than 95% or lower. If the model classifies a prediction with high confidence, he won't pass this record to the next complex classifier. In the class we learned that sometimes this amount of data is not enough, so I created a flag to execute training in an additional way - passing forward part of the confident tagged records in addition to the non-confidence ones.

The Base classifier built with the base configurations I mentioned above. In addition I added a mechanism that preserves all the target labels for all the classifiers created, I pass forward records with classes that vanished because of certain predictions.

After Building the Base classifier I improved the model by a mechanism of giving different subset of data for each model. As we learned in class we can improve the model that way because we are trying to create different classifiers as much as we can.

This method didn't improve the model drastically. The next Improvement I made provided significantly better results but required more resources. The second trial of improving the model was to replace all Decision Tree Classifiers with Random Forest Classifiers. This might be overkill for some domains. It is important to consider the advantages and disadvantages for each domain, as can be seen in the evaluation section.

Data and Preprocessing

Wine Quality:
I used the red wine dataset, 1599 instances. There are 11 attributes not including the output attribute. The target attribute is quality, 0-10 score.
https://archive.ics.uci.edu/ml/datasets/Wine+Quality

Data Set Characteristics:	Multivariate	**Number of Instances:**	4898	**Area:**	Business
Attribute Characteristics:	Real	**Number of Attributes:**	12	**Date Donated**	2009-10-07
Associated Tasks:	Classification, Regression	**Missing Values?**	N/A	**Number of Web Hits:**	1518895

Drug Consumption:

Classify type of drug consumer by personality data. 13 first non-id attribute used to predict the last attribute - volatile substance abuse consumption (6 target classes).

https://archive.ics.uci.edu/ml/datasets/Drug+consumption+%28quantified%29

Data Set Characteristics:	Multivariate	**Number of Instances:**	1885	**Area:**	Social
Attribute Characteristics:	Real	**Number of Attributes:**	32	**Date Donated**	2016-10-17
Associated Tasks:	Classification	**Missing Values?**	N/A	**Number of Web Hits:**	152373

Poker Hand:

The task is to predict the poker hands based on 10 attributes - suits of cards and ranking. The target class range is 0-9 while each number represents a different hand.

https://archive.ics.uci.edu/ml/datasets/Poker+Hand

Data Set Characteristics:	Multivariate	**Number of Instances:**	1025010	**Area:**	Game
Attribute Characteristics:	Categorical, Integer	**Number of Attributes:**	11	**Date Donated**	2007-01-01
Associated Tasks:	Classification	**Missing Values?**	No	**Number of Web Hits:**	639298

HCV Egyptian:

Predict stage of HCV (out of 5) based on 28 attributes. The dataset contains Egyptian patients data.

https://archive.ics.uci.edu/ml/datasets/Hepatitis+C+Virus+%28HCV%29+for+Egyptian+patients

Data Set Characteristics:	Multivariate	**Number of Instances:**	1385	**Area:**	Life
Attribute Characteristics:	Integer, Real	**Number of Attributes:**	29	**Date Donated**	2019-09-30
Associated Tasks:	Classification	**Missing Values?**	N/A	**Number of Web Hits:**	59421

Dry Beans:

This dataset contains 16 attributes + target attribute. The task is a classification of different dry beans to 7 classes.

https://archive.ics.uci.edu/ml/datasets/Dry+Bean+Dataset

Data Set Characteristics:	Multivariate	**Number of Instances:**	13611	**Area:**	Computer
Attribute Characteristics:	Integer, Real	**Number of Attributes:**	17	**Date Donated**	2020-09-14
Associated Tasks:	Classification	**Missing Values?**	N/A	**Number of Web Hits:**	36647

Workers Productivity:

This dataset contains 14 different attributes to predict a worker productivity out of 4 classes. The original dataset target class is the percentage of productivity and I discriticize it to: bad, average, good and excellent. I change all the different types of data to numerical data (for example - date, day etc). There are empty cells in the dataset so I chose to fill it with Iterative Imputer.

https://archive.ics.uci.edu/ml/datasets/Productivity+Prediction+of+Garment+Employees

Data Set Characteristics:	Multivariate, Time-Series	**Number of Instances:**	1197	**Area:**	Business
Attribute Characteristics:	Integer, Real	**Number of Attributes:**	15	**Date Donated**	2020-08-03
Associated Tasks:	Classification, Regression	**Missing Values?**	Yes	**Number of Web Hits:**	7294

Clothing Shopping:

The given data contains information about clickstream from online stores. It contains 13 non-id attributes and I chose to predict the price of a product. I created prices bins: very low, low, mid, high and very high and splitted all the prices accordingly.

https://archive.ics.uci.edu/ml/datasets/clickstream+data+for+online+shopping

Data Set Characteristics:	Multivariate, Sequential	**Number of Instances:**	165474	**Area:**	Business
Attribute Characteristics:	Integer, Real	**Number of Attributes:**	14	**Date Donated**	2019-12-09
Associated Tasks:	Classification, Regression, Clustering	**Missing Values?**	N/A	**Number of Web Hits:**	28186

Obesity:

The classification task in this dataset is to predict the Obesity level out of 7 levels. The data is of individuals from Mexico, Peru and Colombia. Contains info about eating habits and physical condition. As part of preprocessing I mapped the categorical values to numbers.

https://archive.ics.uci.edu/ml/datasets/Estimation+of+obesity+levels+based+on+eating+habits+and+physical+condition+

Data Set Characteristics:	Multivariate	**Number of Instances:**	2111	**Area:**	Life
Attribute Characteristics:	Integer	**Number of Attributes:**	17	**Date Donated**	2019-08-27
Associated Tasks:	Classification, Regression, Clustering	**Missing Values?**	N/A	**Number of Web Hits:**	35138

Internet Firewall:

Traffic data records, collected on a university's firewall. There are 12 attributes in the dataset and the target I chose is the Action, there are 3 options: allow, deny, drop.

https://archive.ics.uci.edu/ml/datasets/Internet+Firewall+Data

Data Set Characteristics:	Multivariate	**Number of Instances:**	65532	**Area:**	Computer
Attribute Characteristics:	N/A	**Number of Attributes:**	12	**Date Donated**	2019-02-04
Associated Tasks:	Classification	**Missing Values?**	N/A	**Number of Web Hits:**	10841

Avila:

Data collected from an XII century giant latin copy of the Bible. The task is predicting which copyist relevant for each pattern, the records are the patterns, characterized by 9 features. There are 12 copyists classes.

https://archive.ics.uci.edu/ml/datasets/Avila

Data Set Characteristics:	Multivariate	**Number of Instances:**	20867	**Area:**	Computer
Attribute Characteristics:	Real	**Number of Attributes:**	10	**Date Donated**	2018-06-20
Associated Tasks:	Classification	**Missing Values?**	N/A	**Number of Web Hits:**	41977

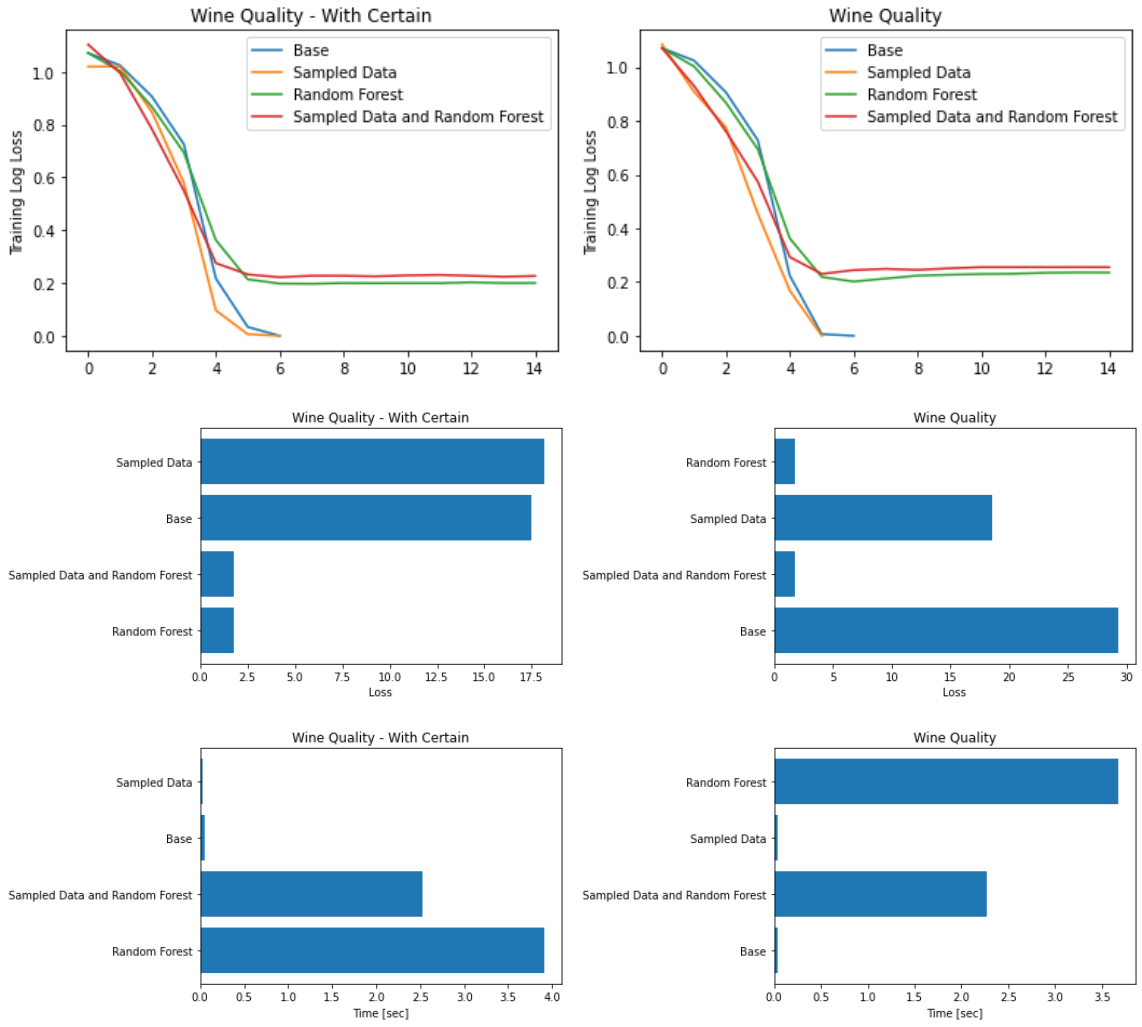
Evaluation Results and Discussion

The experiments configurations:

1. Max Depth is: [1, 2, 4, 6, 10, 14, 18, 24, 30, 36, 48, 60, 72, 86,100], while the first classifier depth is 1, and the 15th depth is 100.
2. Threshold = 0.95 as requested.
3. Split of train-test was according to the given split, otherwise by stratified split while test size is 20%.
4. Sampled Data - when this option is on - 30% of the original data would be randomly sampled for each classifier until there are less than 30% records in the train set.
5. While passing forward certain predictions, the model would randomly sample and pass 50% of the certained records to the next classifier.
6. For vanishing labels - 30% of the records would be randomly sampled from the missing label records and would pass forward to the next classifier.

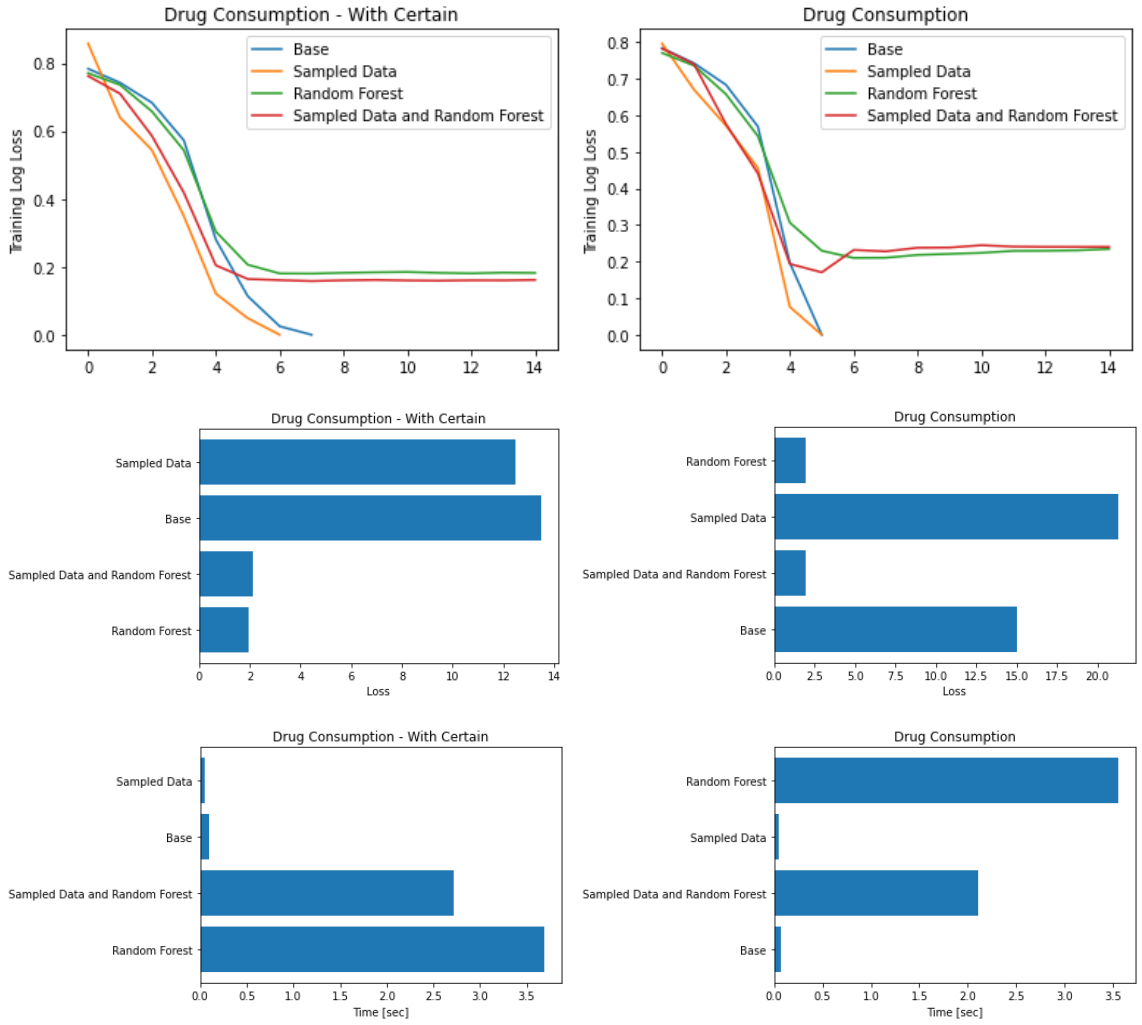
I will discuss the results for each dataset separately. For each dataset there are 6 figures. On the left you will see the “With Certain” plot, this plot is a result of experiments that sample data from the points the model was confident in more than 95%. On the right, would be the “Witout Certain” results, meaning only the points confidence was less than 95% passed to the next model. First the training loss figures, later the loss evaluation and then the total time. All the results appear in the appendices, if needed.

Wine Quality:



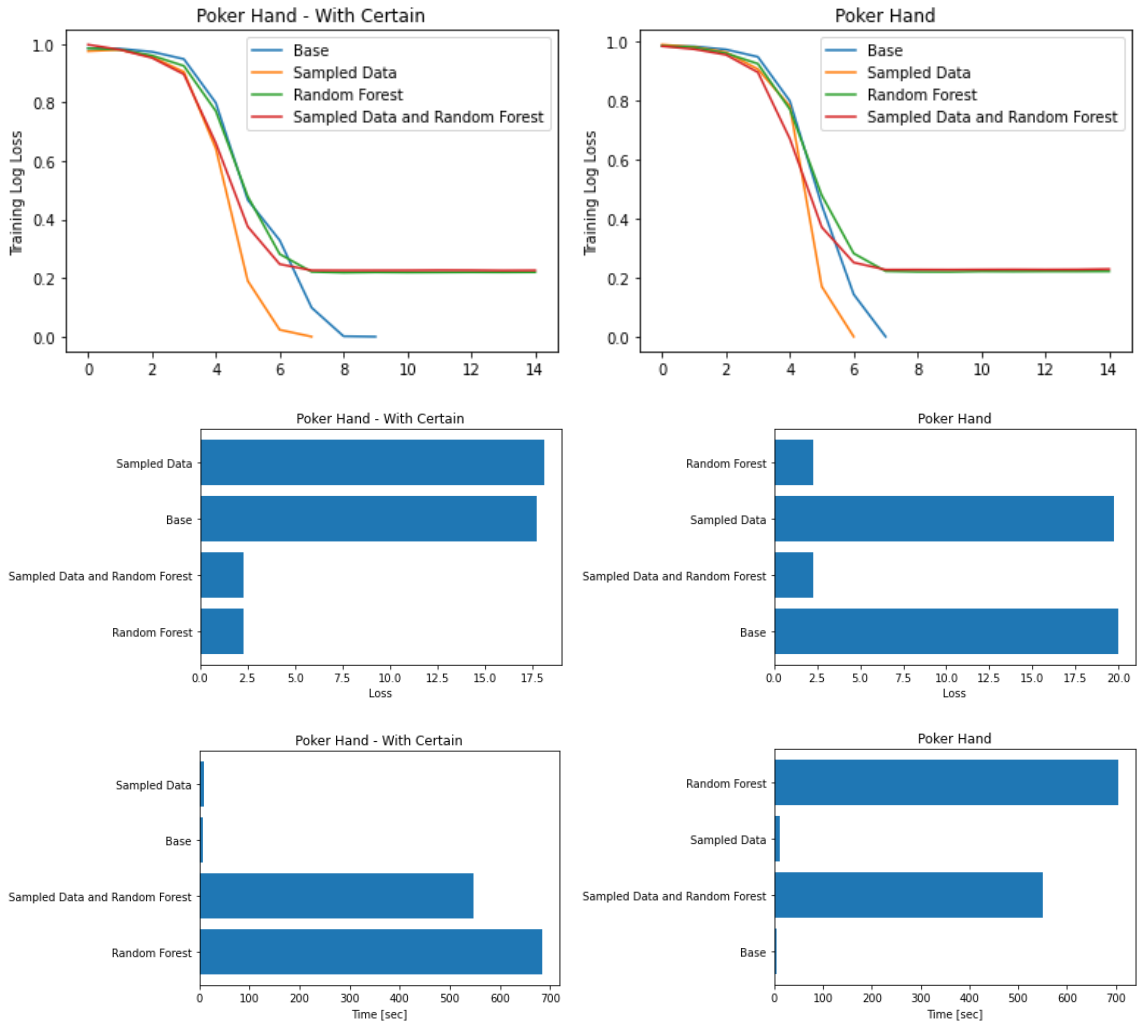
We can see by the Training Loss figures that in both (certain and uncertain) cases, the training included less classifiers with the Random Forest classifiers over the Decision Trees. This phenomenon happens for most of the experiments, probably because it allows more divisions and branchings and more “opinions” of different trees. The Loss with confidence records passed forward, is lower (30 to 17.5) for the Base, the other types return similar results. The total time results look very similar. We can see complicated implementations’ total time is much higher than basics, and the loss much lower.

Drug Consumption:



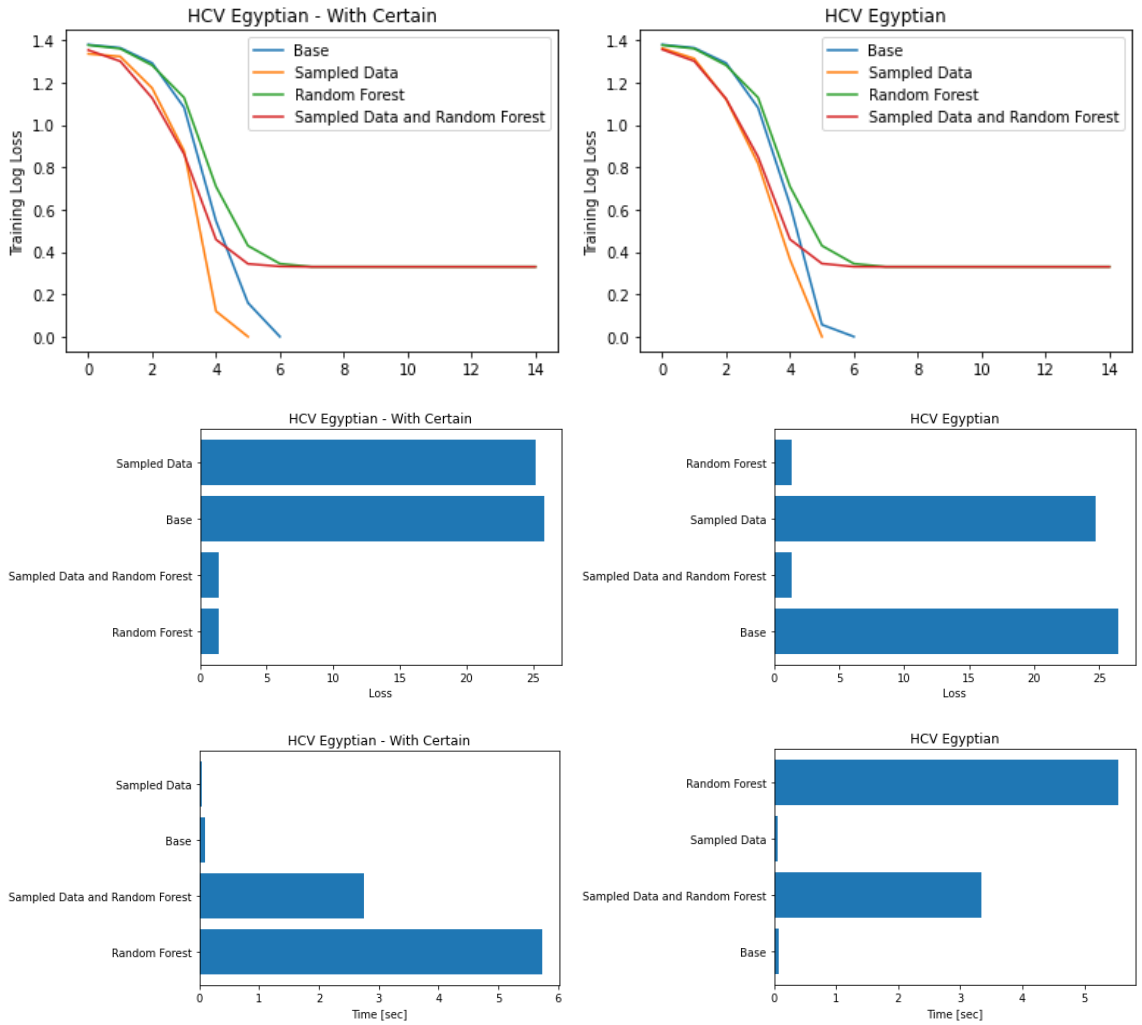
Like in the previous dataset, the Random Forest improvements required more classifiers, complicated models imply more time and better loss performance. In addition, adding confidence records provides lower loss, especially for models with subset sampled data (20 to 13). Note that the addition of these records costs time.

Poker Hand:



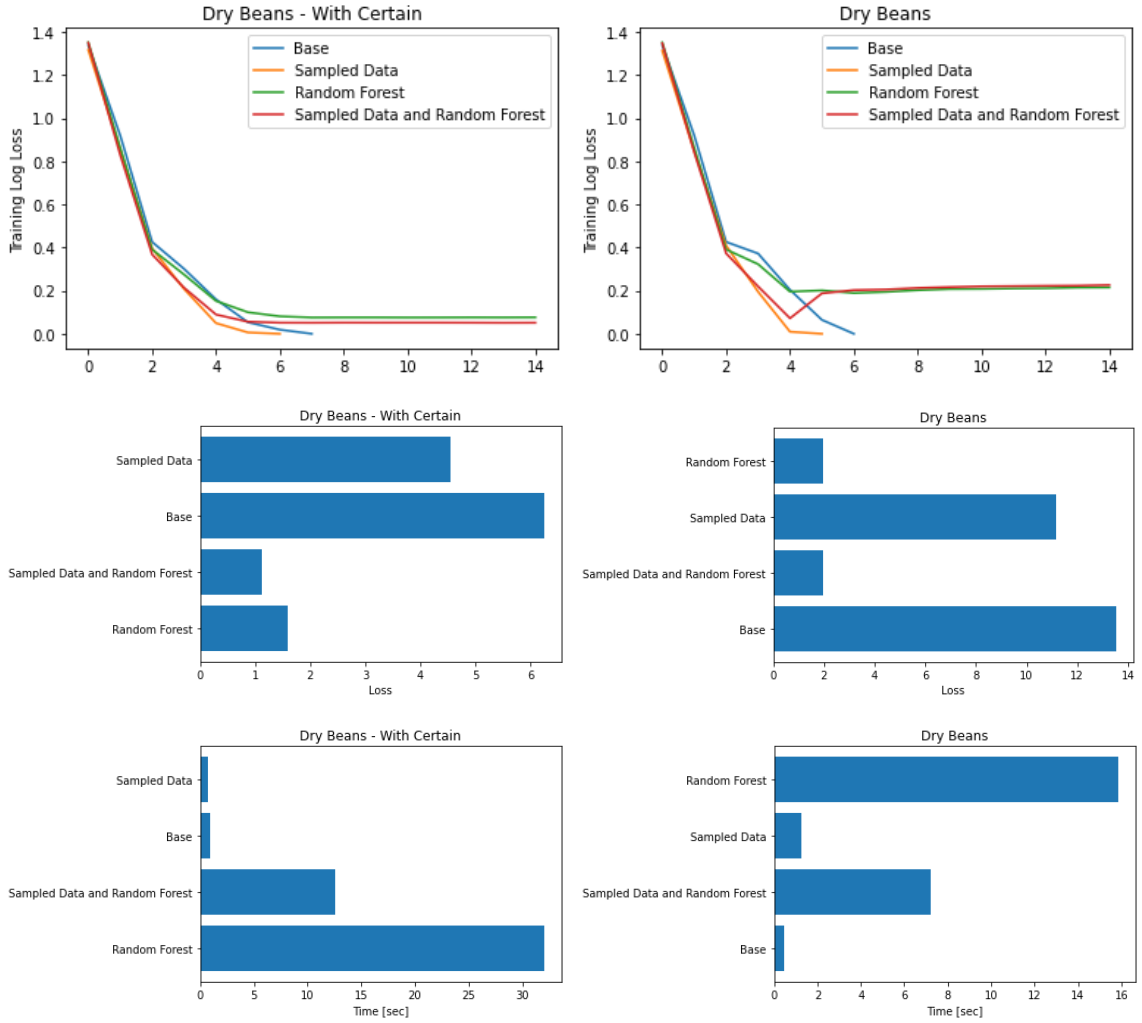
The results description for this dataset is similar to the previous ones.

HCV Egyptian:



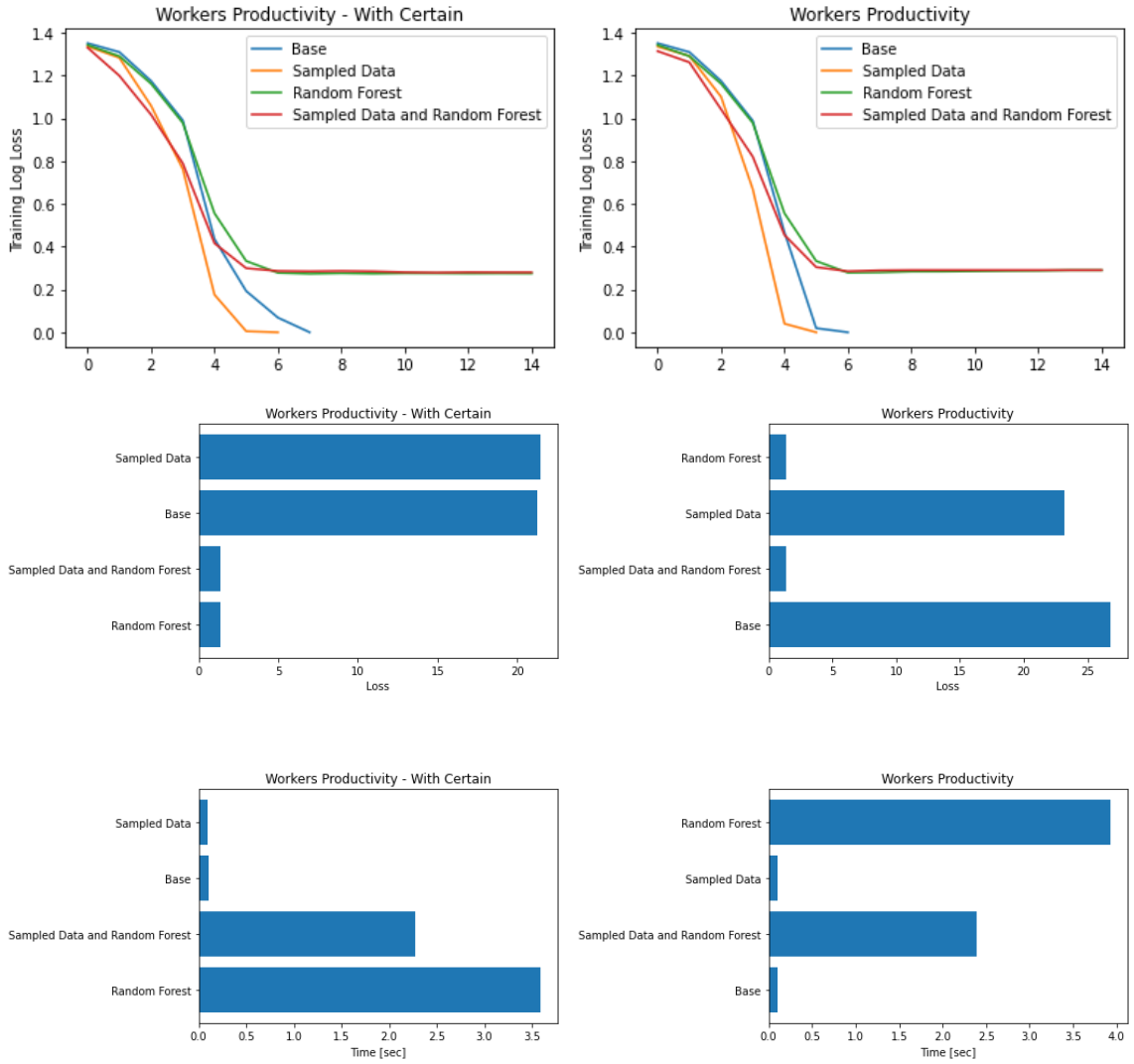
The results description for this dataset is similar to the previous ones. You can see that use of subset sampling of data improves the loss and here is twice faster.

Dry Beans:



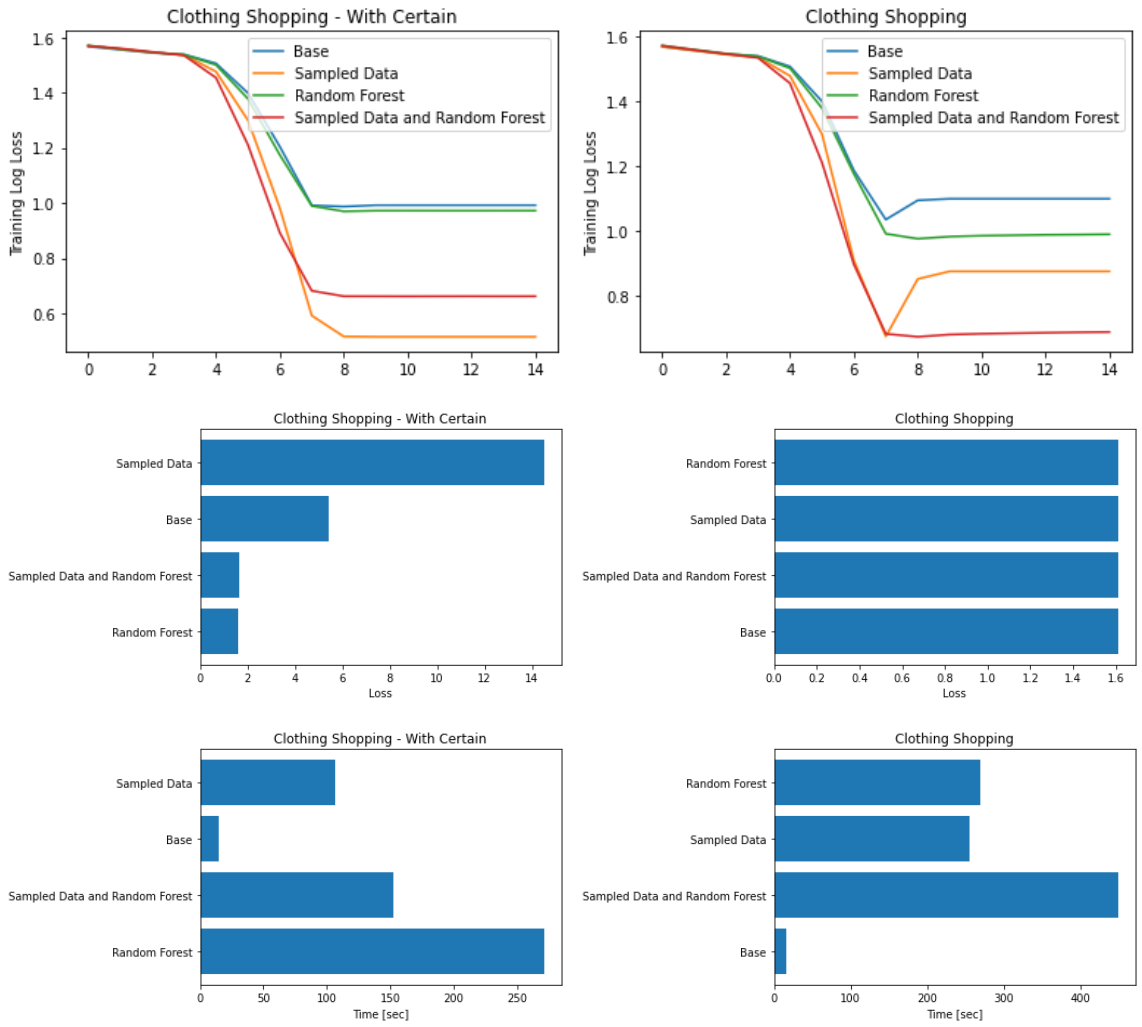
The results description for this dataset is similar to the previous ones.

Workers Productivity:



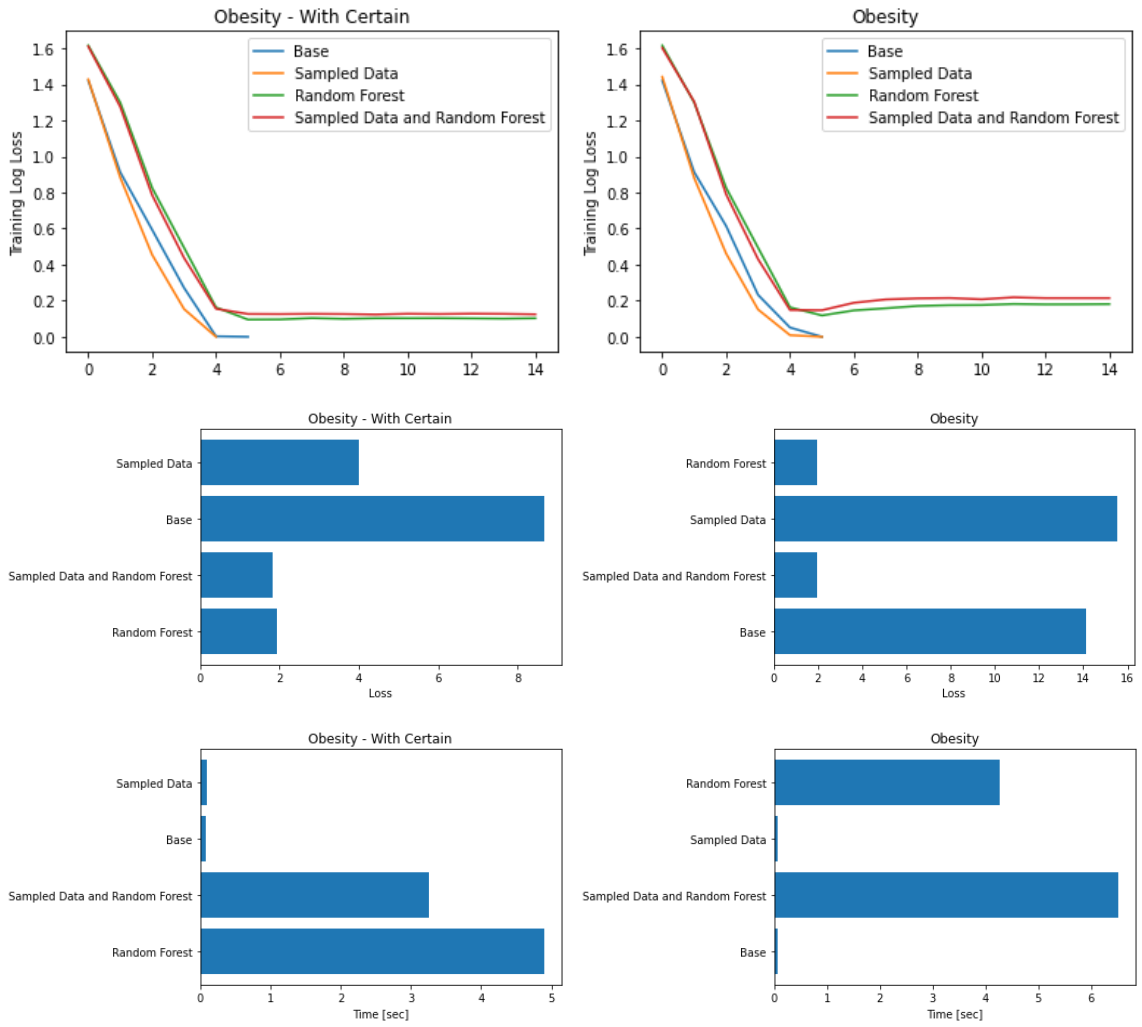
The results description for this dataset is similar to the previous ones.

Clothing Shopping:



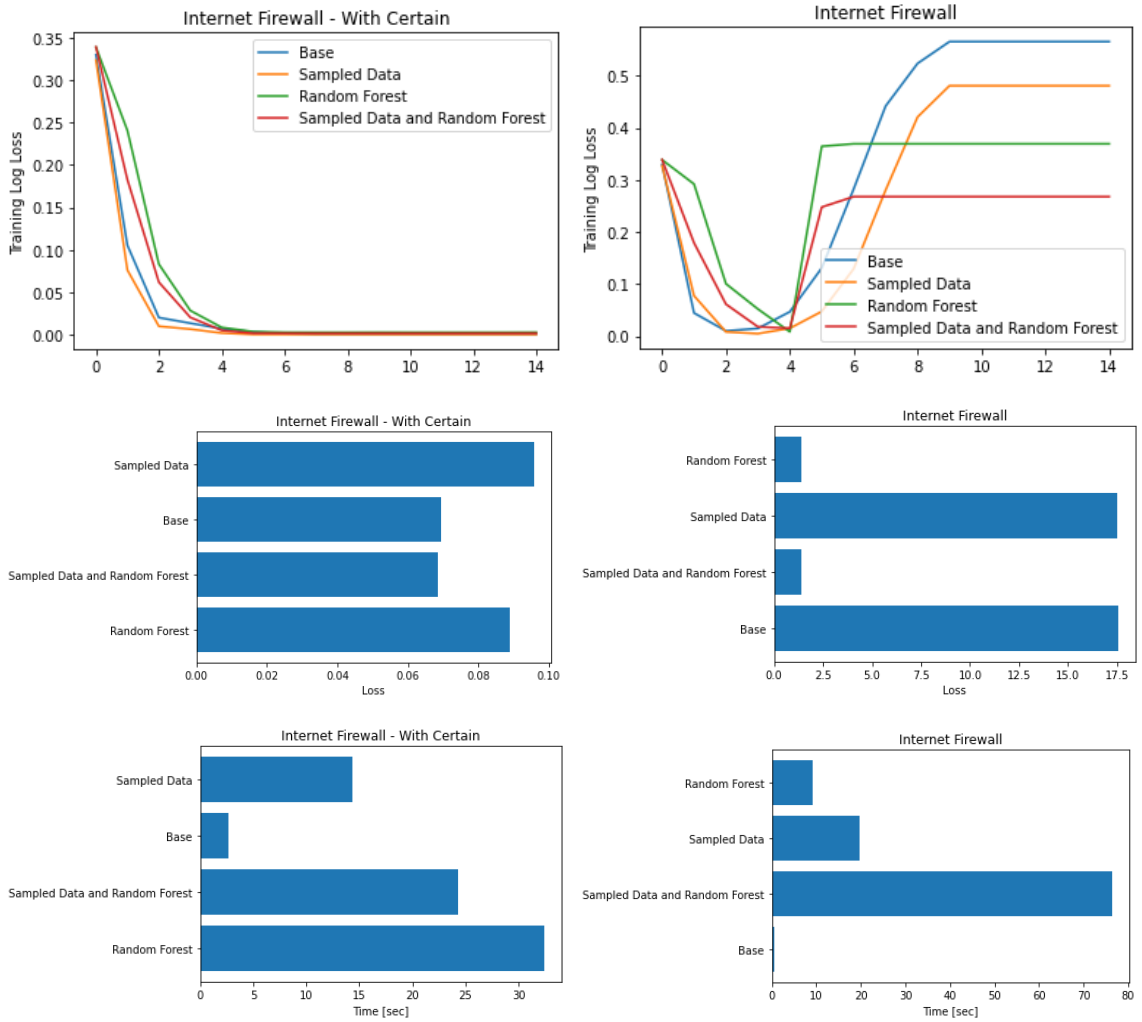
Here something new happens, you can see that the training for all model variants required all the models (got to the most complex one). Probably because executing the task with the data was very hard for the model, with the need to be 95% sure of the result.
There is a phenomenon that visualizes in the uncertain data passing experiments, the loss is identical for all models. Complex classifiers could not improve the model, you can see in the left figure the minimum loss is the same. If you look at the right Time figure, you will notice that for the same loss, the time costs were different.

Obesity:



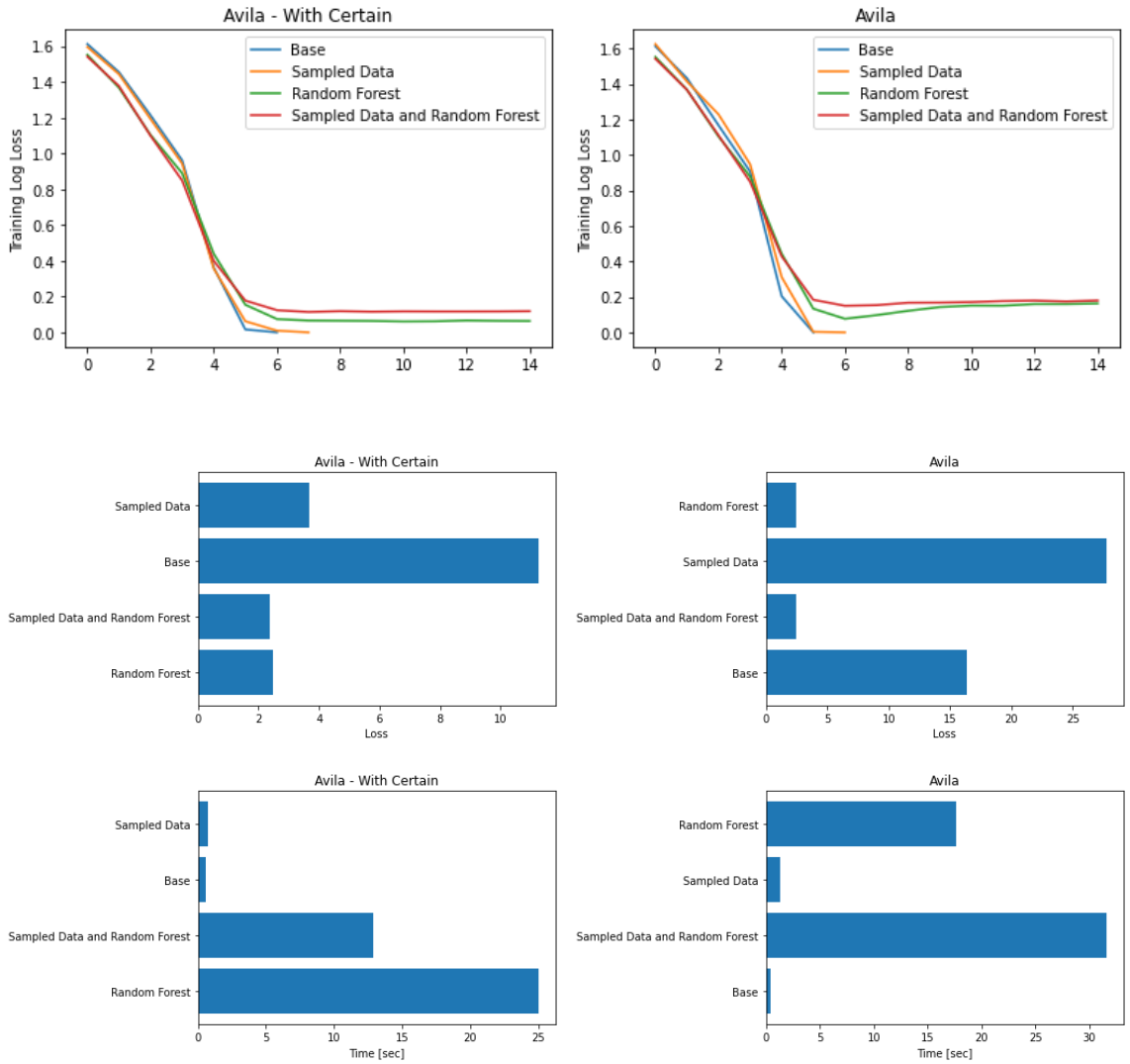
The results description for this dataset is similar to the previous ones. You can see the sampled data and random forest model on the left was faster than the right one.

Internet Firewall:



Here something odd happens in the 4th model on the right training loss figure. The model stops convergencing, and by getting to the deepest and most complex classifier we can realize that the best most complex model to use is the 4th.

Avila:



The results description for this dataset is similar to the previous ones.

Conclusion

We saw different outputs for the same configurations in different domains. There are some cases where we should sample data from the confident predictions of the model, There are other domains where we should sample subsets of data for few classifiers similar to Bagging algorithm. We can understand that sometimes it is better to use Random Forest as the classifiers the model consists of, and sometimes it is not necessary and waste a lot of resources over a small improvement of the loss results. As expected for all the domains the complex models provide better loss results, and for most of them the behaviour of the training phase was alike.

Appendices

With Certain Results:

Base_with_certain

Name: wine_quality
[1.0709310375863847, 1.0245122327719582, 0.9061583994806335, 0.7257090843617247, 0.21730157176729123, 0.033288887429593565, 5.302668873102972e-15]
Log Loss: 17.48525554992354
Time: 0.052796363830566406 sec

****

Name: drug_consumption
[0.7834636698269986, 0.7423610381513577, 0.6833175979270237, 0.5724027308631873, 0.28053456135215105, 0.11366680244446974, 0.02500564467385433, 6.531473940170786e-15]
Log Loss: 13.467374350270216
Time: 0.09226226806640625 sec

****

Name: poker_hand
[0.9851805729488852, 0.9831759012563497, 0.9727986128656043, 0.947820498889363, 0.7984105188906019, 0.4668190858403366, 0.3291277359118501, 0.09888942710729393, 0.001396850453657823, 9.214833367780093e-15]
Log Loss: 17.741187883009847
Time: 6.634655475616455 sec

****

Name: hcv_egyptian
[1.3794484453153923, 1.3643190254480575, 1.2925304694808974, 1.0825279607645553, 0.5464290559788108, 0.15966801797097185, 3.160363017670839e-15]
Log Loss: 25.810565753597515
Time: 0.0866243839263916 sec

****

Name: dry_beans

[1.338929476644076, 0.9206017188119336, 0.42608667203628314, 0.3012068869857585, 0.15973950982788987, 0.0532900649581498, 0.01900738851744314, 6.3162468709785226e-15]
Log Loss: 6.240572158022795
Time: 0.8876371383666992 sec

****

Name: workers_productivity
[1.350445890991119, 1.3090464546129459, 1.1741332143829757, 0.990164213220605, 0.4355214852842793, 0.19255624305672406, 0.06838891309140283, 3.1758570983301345e-15]
Log Loss: 21.29891211019492
Time: 0.10522007942199707 sec

****

Name: clothing_shopping
[1.566788189965672, 1.5552880729302687, 1.5439519758176639, 1.5379325857231607, 1.50550004309818, 1.3982037661005724, 1.2039498638106219, 0.9914410096262986, 0.9871119024142484, 0.9919351621660564, 0.9919844467882244, 0.9919844467882244, 0.9919844467882244, 0.9919844467882244, 0.9919844467882244]
Log Loss: 5.4149047878065595
Time: 14.67401933670044 sec

****

Name: obesity
[1.4207359079280375, 0.9130089977084789, 0.5936630828714605, 0.2712734859059566, 0.0027276184731490636, 6.33532442193162e-15]
Log Loss: 8.655107087140742
Time: 0.08603262901306152 sec

****

Name: internet_firewall
[0.3295625202449027, 0.10525107418006253, 0.020206144744281458, 0.01361171517011615, 0.007119969356123284, 0.0032921853160322655, 0.0015322366663413263, 0.0008560066544894957, 0.0008560066544894957, 0.0008560066544894957, 0.0008560066544894957, 0.0008560066544894957, 0.0008560066544894957, 0.0008560066544894957]
Log Loss: 0.06957130768893859

Time: 2.659590482711792 sec

****

Name: avila
[1.611448380425713, 1.4544246850701399, 1.2132239466798767, 0.9620253238979497, 0.3623360362982117, 0.01606538544949341, 1.1430256319660606e-14]
Log Loss: 11.238256648185956
Time: 0.5600826740264893 sec

****

Sampled_Data_with_certain

Name: wine_quality
[1.0193291894080718, 1.0200603651044666, 0.8475075546088481, 0.5793289232506519, 0.09612634387241728, 0.005872951902029421, 5.320664129763654e-15]
Log Loss: 18.132857607328113
Time: 0.03314685821533203 sec

****

Name: drug_consumption
[0.8576719347874254, 0.6398533038660491, 0.5439169758454154, 0.3492462870045303, 0.12115523861368978, 0.048789070787798254, 6.539017115391817e-15]
Log Loss: 12.459611643787413
Time: 0.053656578063964844 sec

****

Name: poker_hand
[0.9748757541638039, 0.9796513796999023, 0.9554178359130301, 0.90408774667585, 0.6422438100281519, 0.19045309733741494, 0.023492629013078966, 9.214703133840443e-15]
Log Loss: 18.119042096770173
Time: 9.774953126907349 sec

****

Name: hcv_egyptian
[1.3360532307407, 1.3233526320825426, 1.1742595534532956, 0.8773904096333559, 0.12018389717232343, 3.158785147773869e-15]
Log Loss: 25.187122136360863

Time: 0.04988741874694824 sec

****

Name: dry_beans
[1.3167927922194436, 0.8548569371999148, 0.39600266417193414, 0.20775092230064732, 0.04939931485456211, 0.006095559683661, 6.308487106856622e-15]
Log Loss: 4.55358822099631
Time: 0.7450094223022461 sec

****

Name: workers_productivity
[1.3354932519977607, 1.2820918098174103, 1.0581492214708488, 0.7622000824232865, 0.17555706576013658, 0.004830293941187452, 3.1716789473525e-15]
Log Loss: 21.44282367850705
Time: 0.09661293029785156 sec

****

Name: clothing_shopping
[1.5694859791377083, 1.5571705915394691, 1.5455046616802748, 1.5358792177123042, 1.4758241687716793, 1.2998209363333018, 0.9811255817177624, 0.5920912515131679, 0.5165698589417297, 0.5155257465894588, 0.5155257465894587, 0.5155257465894587, 0.5155257465894587, 0.5155257465894587, 0.5155257465894587]
Log Loss: 14.494416214561964
Time: 106.20648717880249 sec

****

Name: obesity
[1.4269869523068508, 0.8833552347560812, 0.45479090955891976, 0.15382060080355006, 6.3651324000742875e-15]
Log Loss: 4.00094572896129
Time: 0.09945917129516602 sec

****

Name: internet_firewall
[0.32325157043448094, 0.0760998671659617, 0.009979584947972298, 0.006665303758131972, 0.0020092196374945034, 0.0003921060020394109, 0.00039210600203941104, 0.00039210600203941104, 0.00039210600203941104,

0.00039210600203941104, 0.00039210600203941104, 0.00039210600203941104,
0.00039210600203941104, 0.00039210600203941104, 0.00039210600203941104]
Log Loss: 0.0957111699935748
Time: 14.33869194984436 sec

****

Name: avila
[1.595193804977226, 1.442915950617314, 1.192583065595933, 0.9435334908065723,
0.35629750542783545, 0.06207384544728189, 0.009446343998059475,
1.141102765703184e-14]
Log Loss: 3.669972503780405
Time: 0.7592916488647461 sec

****

Random_Forest_with_certain

Name: wine_quality
[1.0711438798014166, 1.002713451710904, 0.8675708560920333, 0.6940550240380704,
0.3625727977908777, 0.2136438876733034, 0.19768200967692218, 0.1969335655552624,
0.19965401838039318, 0.1991408445925595, 0.19942192445860535, 0.19935395044505158,
0.20248340723389294, 0.19985945776586647, 0.199997346805425]
Log Loss: 1.791759469228055
Time: 3.9050240516662598 sec

****

Name: drug_consumption
[0.770318756427451, 0.7358339011956438, 0.6576926584135211, 0.5430630012773994,
0.3040227991732358, 0.20633631074901818, 0.18059531253943528, 0.18013443375403326,
0.18247449632969218, 0.18402842890544624, 0.18500494106039686,
0.18232448794129397, 0.1808746259249392, 0.18305055177392254, 0.1821041585878483]
Log Loss: 1.9459101490553141
Time: 3.6843509674072266 sec

****

Name: poker_hand
[0.9848414690166561, 0.9802135313552015, 0.9599909368934506, 0.9244715563507608,
0.7702754103721362, 0.47877295884770127, 0.2819543559974356, 0.22137215064050655,

0.2185028851583997, 0.21980809771980428, 0.21914322510000545, 0.2195005114937326, 0.21999115984380552, 0.21973367395111645, 0.22014410607467355]
Log Loss: 2.3025850929940415
Time: 683.1901443004608 sec

****

Name: hcv_egyptian
[1.3772680665815733, 1.360515156788658, 1.2818135373022732, 1.1292719620837468, 0.7090692076165355, 0.4296120539657819, 0.3445968133748683, 0.33014560876203364, 0.33018268660832295, 0.33018268660832295, 0.33018268660832295, 0.33018268660832295, 0.33018268660832295, 0.33018268660832295, 0.33018268660832295]
Log Loss: 1.3862943611198906
Time: 5.7309510707855225 sec

****

Name: dry_beans
[1.3512330004533766, 0.8638107677751291, 0.3910970074818029, 0.2746905383341342, 0.1523277770973654, 0.09981442242548319, 0.08098929556371258, 0.07561891710515914, 0.07594438601999398, 0.07599807094726069, 0.07559343331925764, 0.07574055746601731, 0.07613628466639262, 0.07590040213864668, 0.07634375574868617]
Log Loss: 1.5852044810464525
Time: 31.995985746383667 sec

****

Name: workers_productivity
[1.3419771423189366, 1.2896080963581953, 1.1611608369876265, 0.9781079518511405, 0.5559158407561586, 0.3328040035827939, 0.27771617987708674, 0.27361534913368013, 0.27611295956724463, 0.27414406055089086, 0.2759473093696982, 0.2756655415652329, 0.2749349120147949, 0.27535994699789873, 0.2750249876686718]
Log Loss: 1.3862943611198906
Time: 3.58528470993042 sec

****

Name: clothing_shopping
[1.570349368988388, 1.557591510581727, 1.545415168044432, 1.5368996963416648, 1.5001238917778497, 1.377503794212717, 1.1737342730148053, 0.989401134753265,

0.9696976564071106, 0.9721350374373264, 0.972225338992956, 0.9721168930645602, 0.9721015858022481, 0.9721611587577421, 0.9721920387381582]
Log Loss: 1.6171060822599106
Time: 271.2509341239929 sec

****

Name: obesity
[1.6151242090249693, 1.2993238283862627, 0.8259252337283962, 0.49383809625262076, 0.16309137504318455, 0.09541990730178264, 0.09613247351236551, 0.10374707699171389, 0.09951542778365709, 0.10283972518024469, 0.10289599381090693, 0.10311552994319492, 0.1020478297416385, 0.1005672235900919, 0.10280015843596488]
Log Loss: 1.9229088470570246
Time: 4.891226053237915 sec

****

Name: internet_firewall
[0.33846609848504183, 0.240400987693538, 0.08312544546545346, 0.028479793374677355, 0.008590164378249044, 0.0036551688127661294, 0.0029032958200717767, 0.0028245066576732306, 0.0028273695387157466, 0.0028996957463337685, 0.0028583826185785907, 0.002843424148808101, 0.0028461409233444027, 0.002840068941440083, 0.002915325834140398]
Log Loss: 0.08888711980394157
Time: 32.382041454315186 sec

****

Name: avila
[1.5514415535592374, 1.3672671995325631, 1.1033980019819045, 0.8898145652915611, 0.43911962399054116, 0.15428189416410543, 0.07436614369628397, 0.0659501180982258, 0.06484979171522402, 0.06370249361948625, 0.060790938637903316, 0.06167053952386167, 0.06635308056415559, 0.06420427717274096, 0.06330338902524593]
Log Loss: 2.46585974627329
Time: 24.990662097930908 sec

****

Sampled_Data_and_Random_Forest_with_certain

Name: wine_quality
[1.103337712931403, 0.9955746750513048, 0.7828364380900439, 0.5495499653043242,
0.27535379913914854, 0.23230071788022166, 0.22175959064885398, 0.2275343165559076,
0.22729072803741462, 0.2250118482137431, 0.22887593022336883, 0.23055382339374073,
0.22730194993525518, 0.2236605486061926, 0.22655814480077935]
Log Loss: 1.791759469228055
Time: 2.5255370140075684 sec

****
Name: drug_consumption
[0.7620088139402842, 0.7099886400184545, 0.5856543934875655, 0.4183347736810386,
0.20460311306220638, 0.16420128591012995, 0.16083750472442548, 0.15811594102384965,
0.1603038606421095, 0.16151517634696969, 0.15994109833932274, 0.1592140303198429,
0.1603672838486474, 0.16022066387121708, 0.16149706609282236]
Log Loss: 2.1188166012879686
Time: 2.718571662902832 sec

****
Name: poker_hand
[0.9971230215498844, 0.9798457427644305, 0.951951525274937, 0.8959037088774726,
0.6608953590535649, 0.3755694378288393, 0.247528151868023, 0.22637294626359267,
0.2261217146294935, 0.22621767189282022, 0.22636852615739383, 0.2268498075602332,
0.22668434956660208, 0.22585565706847924, 0.22622071296135698]
Log Loss: 2.3025850929940415
Time: 547.9668483734131 sec

****
Name: hcv_egyptian
[1.3537870413096662, 1.301077050363637, 1.1272427290688247, 0.8625305000207562,
0.45871543679519955, 0.34390747882768685, 0.3318061105406407, 0.3298674750012581,
0.3298674750012581, 0.3298674750012581, 0.3298674750012581, 0.3298674750012581,
0.3298674750012581, 0.3298674750012581, 0.3298674750012581]
Log Loss: 1.3862943611198906
Time: 2.762906312942505 sec

****
Name: dry_beans

[1.350121864078291, 0.8295524676708184, 0.36788860796962947, 0.21354109953301426, 0.0892346029844146, 0.055785627073650264, 0.051373415197718486, 0.05109466179528983, 0.05164781601795045, 0.051486643248172516, 0.05159647808723906, 0.05156511947392256, 0.05133598178293876, 0.05091506592600626, 0.051216123220373416]
Log Loss: 1.1296334279710205
Time: 12.527148008346558 sec

****

Name: workers_productivity
[1.3293804029580123, 1.1983521921791964, 1.0163913321153701, 0.7892182283188117, 0.4157787141066447, 0.299302004678683, 0.2862517071612965, 0.28460842217361215, 0.28632638371832664, 0.2844973957448178, 0.28032571693076325, 0.27882813536622253, 0.28040671472732387, 0.2799796473212494, 0.2799796473212494]
Log Loss: 1.3862943611198906
Time: 2.272759199142456 sec

****

Name: clothing_shopping
[1.5689739432451082, 1.559755913707783, 1.547009062384318, 1.5342600456600228, 1.4535625967755235, 1.210773453846992, 0.8919720244425867, 0.6820805923948302, 0.6623560400435164, 0.6622364012161314, 0.6620374047160319, 0.6622326266005334, 0.6623235966009664, 0.6621652046868971, 0.6623364493262528]
Log Loss: 1.6424633738067127
Time: 152.4908983707428 sec

****

Name: obesity
[1.6099994243527278, 1.275588627169807, 0.7856378549709513, 0.4372144050505812, 0.15530350616583402, 0.12647523691450735, 0.12580773437689918, 0.12728543001308285, 0.1260958825638093, 0.12337040017457278, 0.1277061280227199, 0.12623649790214925, 0.1281968375303181, 0.12717809571986036, 0.12444766834694214]
Log Loss: 1.8217031182645496
Time: 3.263861894607544 sec

****

Name: internet_firewall

[0.3388230174051152, 0.18252212560441183, 0.06187702932253319, 0.02043856279367542, 0.005145802917877821, 0.0017667930765525273, 0.0017049958529522387, 0.0016450123404790327, 0.0016419533902825591, 0.001657756163732076, 0.00163970648187593 5, 0.0016395752390634972, 0.0016720953325535695, 0.00166372552124514 1, 0.0016733371664936336]
Log Loss: 0.06847399615275224
Time: 24.325860261917114 sec

****

Name: avila
[1.5411393103044864, 1.3769709876789988, 1.0988366627300614, 0.848664090140018, 0.3991911067667775, 0.17749114304215505, 0.1239605096693427, 0.1138858217962377, 0.11845721341656959, 0.11532203405510395, 0.11713209643680544, 0.11656934116451839, 0.11652302813033855, 0.11694392312981233, 0.1183141823293519]
Log Loss: 2.3789582489874204
Time: 12.887636661529541 sec

****

Without Certain Results:

Base

Name: wine_quality
[1.0709310375863847, 1.0245122327719582, 0.9061583994806335, 0.7275161922183909, 0.22368896887396567, 0.006134045845668517, 5.292063084046594e-15]
Log Loss: 29.250026259439995
Time: 0.03902387619018555 sec

****

Name: drug_consumption
[0.7834636698269986, 0.7423610381513577, 0.6833175979270237, 0.5692770085129734, 0.1968244062066333, 6.523813747281262e-15]
Log Loss: 15.024825805743642
Time: 0.06615805625915527 sec

****

Name: poker_hand

[0.9851805729488852, 0.9831759012563497, 0.9727986128656043, 0.947820498889363, 0.7984105188906019, 0.44547633678228915, 0.143332556061481, 9.214772448843348e-15]
Log Loss: 19.955330682581938
Time: 5.072678804397583 sec

****

Name: hcv_egyptian
[1.3794484453153923, 1.3643190254480575, 1.2925304694808974, 1.0811914831175728, 0.6238162965142771, 0.05653848381207112, 3.187265266528059e-15]
Log Loss: 26.43400937083417
Time: 0.08134245872497559 sec

****

Name: dry_beans
[1.338929476644076, 0.9206017188119336, 0.42608667203628314, 0.3723331395409033, 0.20396374597784803, 0.06442716999933393, 6.331292255396405e-15]
Log Loss: 13.521239675716048
Time: 0.48353052139282227 sec

****

Name: workers_productivity
[1.350445890991119, 1.3090464546129459, 1.1741332143829757, 0.9884963691288563, 0.46697390909955994, 0.01914688341642904, 3.1513253545129493e-15]
Log Loss: 26.767551706055773
Time: 0.09894394874572754 sec

****

Name: clothing_shopping
[1.566788189965672, 1.5552880729302687, 1.5439519758176639, 1.5379325857231607, 1.505559301898696, 1.3983015969779764, 1.1850700029191539, 1.033459758137109, 1.0927822856300917, 1.0975796832946612, 1.0975796832946612, 1.0975796832946612, 1.0975796832946612, 1.0975796832946612, 1.0975796832946612]
Log Loss: 1.6094379124340998
Time: 15.745813131332397 sec

****

Name: obesity

[1.4207359079280375, 0.9130089977084789, 0.6141614097700766, 0.23230088810840865, 0.05127956416410479, 6.276229536084107e-15]
Log Loss: 14.125787981842906
Time: 0.07583212852478027 sec

****

Name: internet_firewall
[0.3295625202449027, 0.044515464257014546, 0.010064865123214339, 0.014846193662972473, 0.046799921745522166, 0.13086644784395438, 0.2826715273429377, 0.44167426147333844, 0.5234657913758078, 0.5653430546858722, 0.5653430546858722, 0.5653430546858722, 0.5653430546858722, 0.5653430546858722, 0.5653430546858722]
Log Loss: 17.540215305604463
Time: 0.6180019378662109 sec

****

Name: avila
[1.611448380425713, 1.4316826658806372, 1.1663872241564295, 0.9056856608746225, 0.20337123735514917, 1.1432397953728482e-14]
Log Loss: 16.33452144153294
Time: 0.4510488510131836 sec

****

Sampled_Data

Name: wine_quality
[1.084260694694017, 0.9084322480168117, 0.7756246537678116, 0.4537626707808448, 0.16795418387740976, 5.298791708438261e-15]
Log Loss: 18.564592312264494
Time: 0.036673784255981445 sec

****

Name: drug_consumption
[0.7961623743148508, 0.6701783058677319, 0.5718198297822642, 0.4558795326700292, 0.07631536094127733, 6.493014011759405e-15]
Log Loss: 21.25463162763735
Time: 0.053118228912353516 sec

****

Name: poker_hand
[0.9891737479360091, 0.9813600265869724, 0.9636091472051047, 0.9061084699003789, 0.7808778409802475, 0.16986721483475528, 9.21460034820033e-15]
Log Loss: 19.75521301214983
Time: 10.852378129959106 sec

****

Name: hcv_egyptian
[1.3626455408316223, 1.3135424510635196, 1.1211479751297584, 0.8174051748792708, 0.3641802889835002, 3.1559296239127328e-15]
Log Loss: 24.688367242571534
Time: 0.05857563018798828 sec

****

Name: dry_beans
[1.3126976267161898, 0.8479175545449259, 0.4093102132824883, 0.19494499978242896, 0.00947982855977677, 6.379639768368456e-15]
Log Loss: 11.174683071581462
Time: 1.2632737159729004 sec

****

Name: workers_productivity
[1.3346435306450475, 1.2906982675011511, 1.1001843340515662, 0.6684879856516276, 0.04008264961313386, 3.208544541166707e-15]
Log Loss: 23.16976249825258
Time: 0.10152459144592285 sec

****

Name: clothing_shopping
[1.5662614290526393, 1.5546333273750126, 1.5426050242346847, 1.5335674564792956, 1.476727092995565, 1.2974518390291734, 0.9076185212049378, 0.6729716237585078, 0.8501466170464694, 0.8737750022841004, 0.8737750022841004, 0.8737750022841004, 0.8737750022841004, 0.8737750022841004, 0.8737750022841004]
Log Loss: 1.6094379124340998
Time: 254.96734619140625 sec

****

Name: obesity
[1.4413934814502756, 0.8801526161278811, 0.46141835155742167, 0.1511852524866681, 0.008886502314877366, 6.342149028171227e-15]
Log Loss: 15.513871193931514
Time: 0.07549452781677246 sec

****

Name: internet_firewall
[0.3271590283121629, 0.0775680662516407, 0.008231554838196146, 0.0048342848224954636, 0.0150899476907937, 0.047395187817556424, 0.12942532057870618, 0.2804215279205263, 0.42063229188078777, 0.48072261929232857, 0.48072261929232857, 0.48072261929232857, 0.48072261929232857, 0.48072261929232857]
Log Loss: 17.52440446724561
Time: 19.594242095947266 sec

****

Name: avila
[1.6239366877570056, 1.4135138563354968, 1.227983897538276, 0.9468320553763683, 0.31076993369250844, 0.0033431536039983, 1.1417069611443839e-14]
Log Loss: 27.71176712953744
Time: 1.316098928451538 sec

****

Random_Forest

Name: wine_quality
[1.0711438798014166, 1.002713451710904, 0.8675708560920333, 0.6940550240380704, 0.3625727977908777, 0.2188598337752636, 0.20193443366636254, 0.21307146582873945, 0.22392500170123922, 0.22719472723847808, 0.22982793879555163, 0.23095765262658402, 0.23448295718333487, 0.2356873645459831, 0.23537246772014367]
Log Loss: 1.791759469228055
Time: 3.669628143310547 sec

****

Name: drug_consumption

[0.770318756427451, 0.7358339011956438, 0.6576926584135211, 0.5430630012773994, 0.3066502786926423, 0.2297524732348999, 0.21003260523940673, 0.2105551421827308, 0.21826268315672082, 0.22111468657920075, 0.22388925881083804, 0.22928107643387416, 0.22960550122626575, 0.230790658160946, 0.2345097672282243]
Log Loss: 1.9459101490553141
Time: 3.5506832599639893 sec

****

Name: poker_hand
[0.9848414690166561, 0.9802135313552015, 0.9599909368934506, 0.9244715563507608, 0.7702754103721362, 0.47877295884770127, 0.2819543559974356, 0.22170863770734348, 0.219793944587034, 0.219594243351362, 0.22105542181019291, 0.22065677986864188, 0.22103539358722168, 0.2209312119139711, 0.22114646506805102]
Log Loss: 2.3025850929940415
Time: 704.1169028282166 sec

****

Name: hcv_egyptian
[1.3772680665815733, 1.360515156788658, 1.2818135373022732, 1.1292719620837468, 0.7090692076165355, 0.4296120539657819, 0.3445968133748683, 0.33014560876203364, 0.33018268660832295, 0.33018268660832295, 0.33018268660832295, 0.33018268660832295, 0.33018268660832295, 0.33018268660832295, 0.33018268660832295]
Log Loss: 1.3862943611198906
Time: 5.537571430206299 sec

****

Name: dry_beans
[1.3512330004533766, 0.8638107677751291, 0.3910970074818029, 0.3229885732655471, 0.19529483547397541, 0.2012019645132239, 0.18901292667346578, 0.19422608337147862, 0.20234289861903354, 0.20752931426452553, 0.20811895830209615, 0.21017388328412223, 0.21125073535851058, 0.21437893558594387, 0.21502194769393487]
Log Loss: 1.945910149055314
Time: 15.839182376861572 sec

****

Name: workers_productivity

[1.3419771423189366, 1.2896080963581953, 1.1611608369876265, 0.9781079518511405, 0.5559158407561586, 0.3328040035827939, 0.2784514163614044, 0.27974035479652, 0.283680974513318, 0.2838527311872788, 0.2852857458365523, 0.28684699898248145, 0.2876578997272331, 0.29011255672398806, 0.29011255672398806]
Log Loss: 1.3862943611198906
Time: 3.923978328704834 sec

****

Name: clothing_shopping
[1.570349368988388, 1.557591510581727, 1.545415168044432, 1.5368996963416648, 1.5001238917778497, 1.377503794212717, 1.1745323503050178, 0.9897016364280015, 0.974295275338064, 0.9810028470900155, 0.9836769559747116, 0.9849676293164635, 0.986377422475108, 0.9870506173123487, 0.9880782821234269]
Log Loss: 1.6094379124340998
Time: 269.48245763778687 sec

****

Name: obesity
[1.6151242090249693, 1.2993238283862627, 0.8259252337283962, 0.49383809625262076, 0.16344917613486792, 0.11820943792136852, 0.14615093047850566, 0.1580088029288613, 0.17096644478166934, 0.1752449808838281, 0.176111686053014, 0.18182264969486964, 0.1796566075684368, 0.18004015863737483, 0.1811909872189091]
Log Loss: 1.9459101490553143
Time: 4.268782377243042 sec

****

Name: internet_firewall
[0.33846609848504183, 0.2919789835726049, 0.10056853651440484, 0.052076951903090044, 0.008425465331053672, 0.3644352964973918, 0.36920610509300467, 0.36920610509300467, 0.36920610509300467, 0.36920610509300467, 0.36920610509300467, 0.36920610509300467, 0.36920610509300467, 0.36920610509300467]
Log Loss: 1.3862943611198906
Time: 9.20642900466919 sec

****

Name: avila

[1.5514415535592374, 1.3672671995325631, 1.1033980019819045, 0.8776384907745395, 0.44262443779936805, 0.13334328367395498, 0.07633832134117037, 0.09681483390606112, 0.12132252613815021, 0.14232175803519345, 0.1508153623649195, 0.15014622247007972, 0.15942987229000022, 0.15944309191801492, 0.16286535890108172]
Log Loss: 2.484906649788001
Time: 17.640121698379517 sec

****

Sampled_Data_and_Random_Forest

Name: wine_quality
[1.0702875154716536, 0.9302681245789622, 0.760781451226233, 0.5730066240856992, 0.29281093142085, 0.23036942458985019, 0.24440516751695854, 0.24909931747997255, 0.2453367095117727, 0.2511436275357598, 0.25547198576047264, 0.25547198576047264, 0.25547198576047264, 0.25547198576047264, 0.25547198576047264]
Log Loss: 1.791759469228055
Time: 2.272571563720703 sec

****

Name: drug_consumption
[0.7822734454956559, 0.7398679168034246, 0.5761044462944014, 0.4409840413161322, 0.19375478541791127, 0.17065519512849642, 0.23188107070116704, 0.22797904885541934, 0.23786409822969193, 0.2384846428998727, 0.2447483703338726, 0.24103776808297186, 0.24029245148901343, 0.24029245148901343, 0.24029245148901343]
Log Loss: 1.9459101490553141
Time: 2.1125857830047607 sec

****

Name: poker_hand
[0.9837959415388975, 0.9737811050774042, 0.9547567383102814, 0.8953496670020582, 0.6706081589155157, 0.37070263235273077, 0.25112541385119064, 0.22669219195920073, 0.22687330999942318, 0.2266939087713965, 0.22719395352754407, 0.2275072593621874, 0.22717831017039425, 0.22752615655455868, 0.2290209489580982]
Log Loss: 2.3025850929940415
Time: 551.0552361011505 sec

****

Name: hcv_egyptian

[1.356377836368595, 1.3016289197784443, 1.1231698817472957, 0.8492910128880036, 0.459567823741087, 0.34515021872320323, 0.3305576804859686, 0.3298674750012581, 0.3298674750012581, 0.3298674750012581, 0.3298674750012581, 0.3298674750012581, 0.3298674750012581, 0.3298674750012581]
Log Loss: 1.3862943611198906
Time: 3.347541093826294 sec

****

Name: dry_beans
[1.3475499712336685, 0.8430154629408311, 0.372614923370648, 0.21983822357287974, 0.07186440713569005, 0.18751538010171087, 0.202476454110034, 0.20545677893106404, 0.21251582117337267, 0.2164933509834753, 0.22007350010731797, 0.22119719626414883, 0.22241774773875694, 0.22322999450354056, 0.22595547150648188]
Log Loss: 1.945910149055314
Time: 7.194616794586182 sec

****

Name: workers_productivity
[1.3132253064038955, 1.2611679982480222, 1.041696786155979, 0.8203033721352961, 0.4544060149486432, 0.30437038771569513, 0.28502520524150887, 0.28920308087063834, 0.290432892317294, 0.290432892317294, 0.290432892317294, 0.290432892317294, 0.290432892317294, 0.290432892317294, 0.290432892317294]
Log Loss: 1.3862943611198906
Time: 2.38621187210083 sec

****

Name: clothing_shopping
[1.5695799569428441, 1.5566877625168372, 1.5442724142659825, 1.5322164673545795, 1.4537711703795615, 1.2096447653202624, 0.8952952280366177, 0.6809190181448417, 0.6719084085196056, 0.6787246189402695, 0.6812097517276785, 0.6827855010482363, 0.6843456351953172, 0.685688158865071, 0.6865339829105311]
Log Loss: 1.6094379124340998
Time: 449.36956906318665 sec

****

Name: obesity
[1.6012431617881653, 1.3032950544552693, 0.7894926098197355, 0.43031410590875113, 0.1483505081321017, 0.14716914979559942, 0.18855698005223306, 0.20709394529822353,

0.21304641482666745, 0.21491724472392612, 0.20787663670928427, 0.2188883501559632, 0.2144967850431263, 0.2144967850431263, 0.2144967850431263]
Log Loss: 1.9459101490553143
Time: 6.509314060211182 sec

****

Name: internet_firewall
[0.33871680655670566, 0.17911931489912658, 0.061366698216501175, 0.018012769798727322, 0.015090494870556775, 0.2476922205971883, 0.2677641532091376, 0.2677641532091376, 0.2677641532091376, 0.2677641532091376, 0.2677641532091376, 0.2677641532091376, 0.2677641532091376, 0.2677641532091376, 0.2677641532091376]
Log Loss: 1.3862943611198906
Time: 76.32226300239563 sec

****

Name: avila
[1.541310479180877, 1.3660854047802986, 1.1098342086289696, 0.8483133885629588, 0.42746008361348603, 0.18352866444228144, 0.1496725084706544, 0.15323038868986205, 0.16684186781052937, 0.16755689869740975, 0.1706895304102552, 0.17636870929343168, 0.17902218511966264, 0.1742167767635949, 0.1794788692359631]
Log Loss: 2.484906649788001
Time: 31.552831649780273 sec

****