COMP 6970 Fall 2020 Project 4
Names - Kushagra Kushagra, Chase Townson, Mousumi Akter

Part 1: (output file is SEDA_svmr_25_0.4_0.0001.txt)

Assignment Part = 1 using SEDA
Population size = 25
Selection = Binary Tournament Selection
Crossover = Uniform Crossover ((PopSize/2)-parent)
Mutation Rate = 0.4
Replacement = Replace the Worst
method = svmr
epsilon = 0.0001

Label: 0.0, Predicted: 1, Probs: [0.5 0.5], Evaluation Number: 60565, CI Number: 1
Label: 0.0, Predicted: 1, Probs: [0.5 0.5], Evaluation Number: 119557, CI Number: 2
Label: 0.0, Predicted: 1, Probs: [0.5 0.5], Evaluation Number: 121939, CI Number: 3
Label: 0.0, Predicted: 1, Probs: [0.5 0.5], Evaluation Number: 249862, CI Number: 4
Label: 0.0, Predicted: 1, Probs: [0.5 0.5], Evaluation Number: 406731, CI Number: 5
Label: 0.0, Predicted: 1, Probs: [0.5 0.5], Evaluation Number: 419379, CI Number: 6
Label: 0.0, Predicted: 1, Probs: [0.5 0.5], Evaluation Number: 461557, CI Number: 7
Label: 0.0, Predicted: 1, Probs: [0.5 0.5], Evaluation Number: 489863, CI Number: 8
Label: 0.0, Predicted: 1, Probs: [0.5 0.5], Evaluation Number: 507173, CI Number: 9
Label: 0.0, Predicted: 1, Probs: [0.5 0.5], Evaluation Number: 630394, CI Number: 10

CIs are -
[0.0, 0.1386750490563073, 0.1386750490563073, 0.1386750490563073,
0.1386750490563073, 0.0, 0.1386750490563073, 0.1386750490563073, 0.0, 0.0, 0.0,
0.1386750490563073, 0.1386750490563073, 0.1386750490563073, 0.0,
0.1386750490563073, 0.1386750490563073, 0.0, 0.0, 0.1386750490563073,
0.1386750490563073, 0.1386750490563073, 0.1386750490563073, 0.0,
0.1386750490563073, 0.0, 0.0, 0.1386750490563073, 0.1386750490563073, 0.0,
0.1386750490563073, 0.1386750490563073, 0.1386750490563073, 0.1386750490563073,
0.1386750490563073, 0.0, 0.0, 0.1386750490563073, 0.1386750490563073,
0.1386750490563073, 0.0, 0.1386750490563073, 0.1386750490563073,
0.1386750490563073, 0.1386750490563073, 0.0, 0.1386750490563073, 0.0, 0.0, 0.0,
0.1386750490563073, 0.1386750490563073, 0.0, 0.1386750490563073, 0.0, 0.0, 0.0,
0.1386750490563073, 0.1386750490563073, 0.1386750490563073, 0.0, 0.0,
0.1386750490563073, 0.0, 0.0, 0.1386750490563073, 0.1386750490563073, 0.0, 0.0,
0.1386750490563073, 0.0, 0.0, 0.1386750490563073, 0.1386750490563073, 0.0,
0.1386750490563073, 0.0, 0.0, 0.1386750490563073, 0.0, 0.0, 0.0, 0.1386750490563073, 0.0,

0.0, 0.0, 0.1386750490563073, 0.0, 0.1386750490563073, 0.1386750490563073, 0.0,
0.1386750490563073, 0.1386750490563073, 0.1386750490563073, 0.0]

[0.1336306209562122, 0.1336306209562122, 0.0, 0.1336306209562122, 0.0, 0.0,
0.1336306209562122, 0.0, 0.0, 0.1336306209562122, 0.1336306209562122,
0.1336306209562122, 0.1336306209562122, 0.1336306209562122, 0.1336306209562122,
0.1336306209562122, 0.1336306209562122, 0.1336306209562122, 0.0, 0.0,
0.1336306209562122, 0.1336306209562122, 0.1336306209562122, 0.1336306209562122,
0.0, 0.0, 0.0, 0.0, 0.1336306209562122, 0.1336306209562122, 0.0, 0.1336306209562122,
0.1336306209562122, 0.1336306209562122, 0.1336306209562122, 0.1336306209562122,
0.1336306209562122, 0.0, 0.0, 0.1336306209562122, 0.1336306209562122, 0.0, 0.0,
0.1336306209562122, 0.1336306209562122, 0.1336306209562122, 0.1336306209562122,
0.0, 0.1336306209562122, 0.1336306209562122, 0.0, 0.0, 0.1336306209562122,
0.1336306209562122, 0.0, 0.1336306209562122, 0.0, 0.1336306209562122,
0.1336306209562122, 0.0, 0.0, 0.0, 0.1336306209562122, 0.0, 0.1336306209562122,
0.1336306209562122, 0.0, 0.0, 0.0, 0.0, 0.1336306209562122, 0.0, 0.1336306209562122, 0.0,
0.1336306209562122, 0.1336306209562122, 0.0, 0.0, 0.1336306209562122,
0.1336306209562122, 0.0, 0.1336306209562122, 0.1336306209562122, 0.0, 0.0,
0.1336306209562122, 0.1336306209562122, 0.1336306209562122, 0.1336306209562122,
0.0, 0.1336306209562122, 0.0, 0.0, 0.1336306209562122, 0.1336306209562122]

[0.14002800840280097, 0.14002800840280097, 0.14002800840280097,
0.14002800840280097, 0.0, 0.0, 0.0, 0.0, 0.14002800840280097, 0.0, 0.0, 0.0,
0.14002800840280097, 0.14002800840280097, 0.0, 0.0, 0.0, 0.14002800840280097, 0.0,
0.14002800840280097, 0.0, 0.14002800840280097, 0.0, 0.14002800840280097, 0.0, 0.0, 0.0,
0.14002800840280097, 0.0, 0.14002800840280097, 0.0, 0.14002800840280097, 0.0,
0.14002800840280097, 0.14002800840280097, 0.14002800840280097, 0.0, 0.0, 0.0, 0.0,
0.14002800840280097, 0.14002800840280097, 0.0, 0.14002800840280097,
0.14002800840280097, 0.0, 0.14002800840280097, 0.14002800840280097, 0.0,
0.14002800840280097, 0.14002800840280097, 0.14002800840280097, 0.0, 0.0, 0.0, 0.0,
0.14002800840280097, 0.0, 0.14002800840280097, 0.14002800840280097,
0.14002800840280097, 0.14002800840280097, 0.14002800840280097, 0.0,
0.14002800840280097, 0.0, 0.0, 0.14002800840280097, 0.14002800840280097, 0.0,
0.14002800840280097, 0.14002800840280097, 0.0, 0.0, 0.14002800840280097,
0.14002800840280097, 0.0, 0.14002800840280097, 0.14002800840280097,
0.14002800840280097, 0.14002800840280097, 0.14002800840280097, 0.0,
0.14002800840280097, 0.14002800840280097, 0.0, 0.14002800840280097,
0.14002800840280097, 0.0, 0.14002800840280097, 0.0, 0.0, 0.0, 0.14002800840280097,
0.14002800840280097]

[0.1414213562373095, 0.1414213562373095, 0.0, 0.0, 0.1414213562373095, 0.0,
0.1414213562373095, 0.1414213562373095, 0.0, 0.1414213562373095,
0.1414213562373095, 0.1414213562373095, 0.1414213562373095, 0.0,
0.1414213562373095, 0.1414213562373095, 0.1414213562373095, 0.1414213562373095,

0.0, 0.1414213562373095, 0.0, 0.1414213562373095, 0.0, 0.0, 0.1414213562373095,
0.1414213562373095, 0.0, 0.1414213562373095, 0.1414213562373095, 0.0,
0.1414213562373095, 0.0, 0.1414213562373095, 0.0, 0.1414213562373095, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.1414213562373095, 0.1414213562373095, 0.1414213562373095,
0.1414213562373095, 0.1414213562373095, 0.1414213562373095, 0.0,
0.1414213562373095, 0.1414213562373095, 0.0, 0.0, 0.0, 0.1414213562373095, 0.0,
0.1414213562373095, 0.1414213562373095, 0.1414213562373095, 0.0, 0.0, 0.0,
0.1414213562373095, 0.0, 0.0, 0.0, 0.1414213562373095, 0.0, 0.0, 0.1414213562373095, 0.0,
0.1414213562373095, 0.1414213562373095, 0.0, 0.0, 0.1414213562373095,
0.1414213562373095, 0.0, 0.1414213562373095, 0.1414213562373095, 0.0,
0.1414213562373095, 0.1414213562373095, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.1414213562373095,
0.0, 0.1414213562373095, 0.1414213562373095, 0.1414213562373095,
0.1414213562373095, 0.0]

[0.14586499149789456, 0.0, 0.14586499149789456, 0.0, 0.0, 0.0, 0.0, 0.14586499149789456,
0.14586499149789456, 0.14586499149789456, 0.14586499149789456,
0.14586499149789456, 0.0, 0.14586499149789456, 0.14586499149789456,
0.14586499149789456, 0.0, 0.0, 0.0, 0.0, 0.14586499149789456, 0.0, 0.14586499149789456,
0.14586499149789456, 0.14586499149789456, 0.0, 0.0, 0.14586499149789456,
0.14586499149789456, 0.14586499149789456, 0.0, 0.14586499149789456, 0.0,
0.14586499149789456, 0.14586499149789456, 0.0, 0.0, 0.14586499149789456, 0.0, 0.0, 0.0,
0.14586499149789456, 0.0, 0.14586499149789456, 0.14586499149789456, 0.0,
0.14586499149789456, 0.0, 0.0, 0.14586499149789456, 0.14586499149789456,
0.14586499149789456, 0.0, 0.0, 0.14586499149789456, 0.14586499149789456, 0.0, 0.0,
0.14586499149789456, 0.0, 0.0, 0.14586499149789456, 0.0, 0.14586499149789456, 0.0, 0.0,
0.0, 0.0, 0.14586499149789456, 0.0, 0.14586499149789456, 0.14586499149789456,
0.14586499149789456, 0.0, 0.14586499149789456, 0.14586499149789456, 0.0, 0.0, 0.0,
0.14586499149789456, 0.14586499149789456, 0.14586499149789456, 0.0, 0.0,
0.14586499149789456, 0.14586499149789456, 0.0, 0.14586499149789456,
0.14586499149789456, 0.0, 0.0, 0.0, 0.0, 0.14586499149789456, 0.0]

[0.0, 0.0, 0.15075567228888181, 0.15075567228888181, 0.15075567228888181, 0.0, 0.0,
0.15075567228888181, 0.0, 0.0, 0.0, 0.0, 0.0, 0.15075567228888181, 0.0,
0.15075567228888181, 0.15075567228888181, 0.15075567228888181, 0.0, 0.0, 0.0, 0.0, 0.0,
0.15075567228888181, 0.0, 0.0, 0.15075567228888181, 0.0, 0.15075567228888181,
0.15075567228888181, 0.0, 0.0, 0.15075567228888181, 0.15075567228888181, 0.0, 0.0,
0.15075567228888181, 0.0, 0.15075567228888181, 0.15075567228888181,
0.15075567228888181, 0.15075567228888181, 0.15075567228888181, 0.0, 0.0, 0.0,
0.15075567228888181, 0.15075567228888181, 0.0, 0.15075567228888181,
0.15075567228888181, 0.15075567228888181, 0.0, 0.0, 0.15075567228888181, 0.0,
0.15075567228888181, 0.0, 0.0, 0.15075567228888181, 0.0, 0.15075567228888181,
0.15075567228888181, 0.15075567228888181, 0.15075567228888181,
0.15075567228888181, 0.15075567228888181, 0.0, 0.0, 0.0, 0.0, 0.0, 0.15075567228888181,
0.15075567228888181, 0.15075567228888181, 0.15075567228888181, 0.0,

0.15075567228888181, 0.0, 0.0, 0.0, 0.15075567228888181, 0.0, 0.0, 0.0, 0.0,
0.15075567228888181, 0.0, 0.15075567228888181, 0.15075567228888181, 0.0,
0.15075567228888181, 0.0, 0.0, 0.0]

[0.0, 0.14433756729740646, 0.0, 0.14433756729740646, 0.0, 0.0, 0.14433756729740646, 0.0,
0.14433756729740646, 0.14433756729740646, 0.0, 0.0, 0.14433756729740646, 0.0, 0.0,
0.14433756729740646, 0.14433756729740646, 0.14433756729740646, 0.0,
0.14433756729740646, 0.14433756729740646, 0.14433756729740646,
0.14433756729740646, 0.14433756729740646, 0.0, 0.0, 0.14433756729740646, 0.0, 0.0, 0.0,
0.0, 0.14433756729740646, 0.0, 0.0, 0.14433756729740646, 0.0, 0.0, 0.0, 0.0, 0.0,
0.14433756729740646, 0.14433756729740646, 0.14433756729740646, 0.0,
0.14433756729740646, 0.0, 0.14433756729740646, 0.14433756729740646,
0.14433756729740646, 0.0, 0.14433756729740646, 0.0, 0.0, 0.14433756729740646, 0.0, 0.0,
0.14433756729740646, 0.14433756729740646, 0.14433756729740646,
0.14433756729740646, 0.14433756729740646, 0.14433756729740646, 0.0,
0.14433756729740646, 0.0, 0.14433756729740646, 0.0, 0.0, 0.14433756729740646,
0.14433756729740646, 0.14433756729740646, 0.14433756729740646, 0.0,
0.14433756729740646, 0.0, 0.0, 0.14433756729740646, 0.14433756729740646,
0.14433756729740646, 0.0, 0.14433756729740646, 0.14433756729740646, 0.0, 0.0, 0.0, 0.0,
0.14433756729740646, 0.0, 0.14433756729740646, 0.14433756729740646, 0.0, 0.0,
0.14433756729740646, 0.0, 0.0]

[0.0, 0.1414213562373095, 0.1414213562373095, 0.0, 0.1414213562373095, 0.0,
0.1414213562373095, 0.1414213562373095, 0.0, 0.0, 0.0, 0.1414213562373095,
0.1414213562373095, 0.1414213562373095, 0.0, 0.0, 0.0, 0.1414213562373095, 0.0, 0.0, 0.0,
0.1414213562373095, 0.1414213562373095, 0.0, 0.0, 0.0, 0.0, 0.1414213562373095,
0.1414213562373095, 0.1414213562373095, 0.0, 0.1414213562373095, 0.0, 0.0,
0.1414213562373095, 0.1414213562373095, 0.1414213562373095, 0.0,
0.1414213562373095, 0.1414213562373095, 0.1414213562373095, 0.1414213562373095,
0.1414213562373095, 0.1414213562373095, 0.1414213562373095, 0.0,
0.1414213562373095, 0.0, 0.1414213562373095, 0.1414213562373095,
0.1414213562373095, 0.1414213562373095, 0.0, 0.1414213562373095,
0.1414213562373095, 0.1414213562373095, 0.1414213562373095, 0.0, 0.0, 0.0, 0.0,
0.1414213562373095, 0.1414213562373095, 0.1414213562373095, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.1414213562373095, 0.0, 0.1414213562373095, 0.1414213562373095,
0.1414213562373095, 0.0, 0.0, 0.1414213562373095, 0.0, 0.1414213562373095, 0.0, 0.0, 0.0,
0.0, 0.1414213562373095, 0.1414213562373095, 0.1414213562373095, 0.0, 0.0,
0.1414213562373095, 0.1414213562373095, 0.0, 0.1414213562373095, 0.0,
0.1414213562373095]

[0.0, 0.1414213562373095, 0.1414213562373095, 0.1414213562373095,
0.1414213562373095, 0.0, 0.1414213562373095, 0.0, 0.1414213562373095,
0.1414213562373095, 0.0, 0.0, 0.1414213562373095, 0.1414213562373095, 0.0,
0.1414213562373095, 0.1414213562373095, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,

0.1414213562373095, 0.1414213562373095, 0.0, 0.0, 0.1414213562373095, 0.1414213562373095, 0.1414213562373095, 0.1414213562373095, 0.1414213562373095, 0.0, 0.0, 0.1414213562373095, 0.0, 0.1414213562373095, 0.0, 0.1414213562373095, 0.0, 0.1414213562373095, 0.1414213562373095, 0.0, 0.0, 0.0, 0.0, 0.0, 0.1414213562373095, 0.1414213562373095, 0.0, 0.0, 0.1414213562373095, 0.0, 0.1414213562373095, 0.0, 0.1414213562373095, 0.1414213562373095, 0.0, 0.1414213562373095, 0.0, 0.1414213562373095, 0.1414213562373095, 0.1414213562373095, 0.1414213562373095, 0.1414213562373095, 0.0, 0.1414213562373095, 0.1414213562373095, 0.1414213562373095, 0.1414213562373095, 0.0, 0.1414213562373095, 0.1414213562373095, 0.0, 0.0, 0.0, 0.0, 0.1414213562373095, 0.1414213562373095, 0.0, 0.0, 0.0, 0.0, 0.0, 0.1414213562373095, 0.0, 0.1414213562373095, 0.1414213562373095, 0.1414213562373095, 0.1414213562373095, 0.1414213562373095, 0.0, 0.0]

[0.0, 0.0, 0.0, 0.15249857033260467, 0.15249857033260467, 0.15249857033260467, 0.0, 0.0, 0.0, 0.15249857033260467, 0.0, 0.15249857033260467, 0.0, 0.15249857033260467, 0.0, 0.0, 0.0, 0.0, 0.0, 0.15249857033260467, 0.0, 0.15249857033260467, 0.15249857033260467, 0.15249857033260467, 0.15249857033260467, 0.15249857033260467, 0.0, 0.0, 0.0, 0.0, 0.15249857033260467, 0.15249857033260467, 0.0, 0.15249857033260467, 0.0, 0.0, 0.15249857033260467, 0.0, 0.0, 0.15249857033260467, 0.15249857033260467, 0.0, 0.15249857033260467, 0.15249857033260467, 0.15249857033260467, 0.0, 0.15249857033260467, 0.0, 0.0, 0.15249857033260467, 0.0, 0.0, 0.0, 0.0, 0.15249857033260467, 0.0, 0.15249857033260467, 0.0, 0.0, 0.15249857033260467, 0.0, 0.15249857033260467, 0.0, 0.0, 0.0, 0.0, 0.15249857033260467, 0.15249857033260467, 0.0, 0.0, 0.0, 0.0, 0.15249857033260467, 0.15249857033260467, 0.0, 0.0, 0.15249857033260467, 0.15249857033260467, 0.0, 0.15249857033260467, 0.15249857033260467, 0.15249857033260467, 0.0, 0.15249857033260467, 0.15249857033260467, 0.15249857033260467, 0.0, 0.15249857033260467, 0.0, 0.0, 0.15249857033260467, 0.0, 0.0, 0.0, 0.0, 0.0, 0.15249857033260467, 0.0, 0.15249857033260467, 0.15249857033260467, 0.15249857033260467, 0.15249857033260467, 0.15249857033260467]

Average number of function evaluations (over ten runs) = 63039.4

Execution time = 1287.7425323833334 minutes

Other run outputs are in files (all included) -
SEDA_svmr_10_0.2.txt
SEDA_svmr_10_0.005.txt
SEDA_svmr_15_0.25.txt
SEDA_svmr_50_0.5.txt

We notice that with fewer runs, it was difficult to get with final labels = 0.0. Later, I felt that by changing gamma = “scale” to gamma=”auto” of SVM with RBF, we could have obtained label = 0.0 in fewer runs.

Part 2:

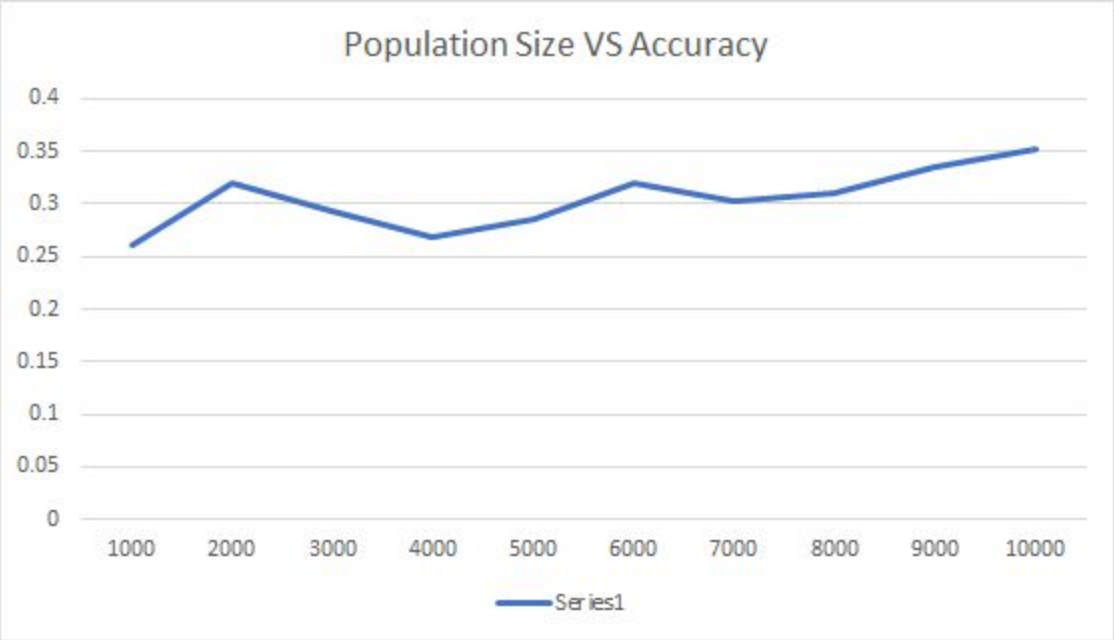
a) Result taken from file RE_SVM_svmr_1000.txt -

Assignment Part = 2 using RE_SVM
Population size = 1000
method = svmr
Accuracy of First Dataset: 0.7394957983193278
Accuracy of Second Dataset: 0.2605042016806723
Execution time = 1.4214263833333334 minutes

The program for comparing the performance of the SVM of the first dataset and the SVM of the second of the dataset is based on input data. The first dataset is the original dataset of HTML_malware_dataset.csv and using feature masking, we obtained SVM by invoking Scikit's SVM for RBF. Whereas, for the second dataset, we created 95 tupled chromosomes by randomly selecting the value of each tuple using uniform random variable and setting lower bound and upper bound separately for each tuple. No non-linear transformation such as setting the value of tuple to either 0 or 1 by comparing with any threshold value was used. Initially, I was not able to get fitness values as -1 even for 1000 random inputs then I changed the gamma parameter of Scikit's SVM from scale to auto and that solved the issue.

Results from Probe_SVM_svmr_data.txt
Probe Dataset Accuracy = Highly volatile, so far I've gotten as low as 0.795 to as high as 0.99
This result was acquired by randomly generating the 1000 probes by finding the max value of each column from the original dataset and for each data point, picking a random number between 0 and the max of that column. Most final labels were -1 with the SVM scale set to auto but there were a few 1s. This high volatility is not very useful for making a good SVM. Also, the higher accuracies, anything over 90, were probably gotten when the majority of labels were the same. I'm guessing that depending on the random probe dataset, that the new SVM could be learning or just guessing -1 every time and getting good accuracy due to the low amount of 1 labels.

b) No, it is not enough to have 1000 random probes. Yes, the accuracy does increase with increase in random probes or population size as we can see in the graph below -



Population Size	Accuracy of the Second Dataset
1000	0.260504202
2000	0.319327731
3000	0.294117647
4000	0.268907563
5000	0.285714286
6000	0.319327731
7000	0.302521008
8000	0.31092437
9000	0.336134454
10000	0.352941176

Part 3:
In this part we worked with SEDA because from our previous assignment we inspect SEDA takes less time.

- For label 1, best fitness for epsilon= 0.05 with candidate fitness around 0.83
- For label -1, best fitness for epsilon= 0.005 with candidate fitness around -0.52
- For both labels euclidean distance for the closest instance is around ~0.9 which implies that candidate instances are not similar with the original dataset.
- The reason can be because of the randomized samples of training and testing split and it's less likely that two samples will be closest.

Label	Fitness	Euclidean Distance	Population Size	Mutation Rate	Epsilon	Time (minute)
1	0.83	0.913	15	0.25	0.05	3.19
-1	-0.52	0.935	15	0.25	0.005	38.533

To find the best candidate instances please review:

- SEDA_L+1_svmr_15_0.25.txt (For Label 1)
- SEDA_L-1_svmr_15_0.25.txt (For Label -1)