# Paper: A reliable ensemble based approach to semi-supervised learnin

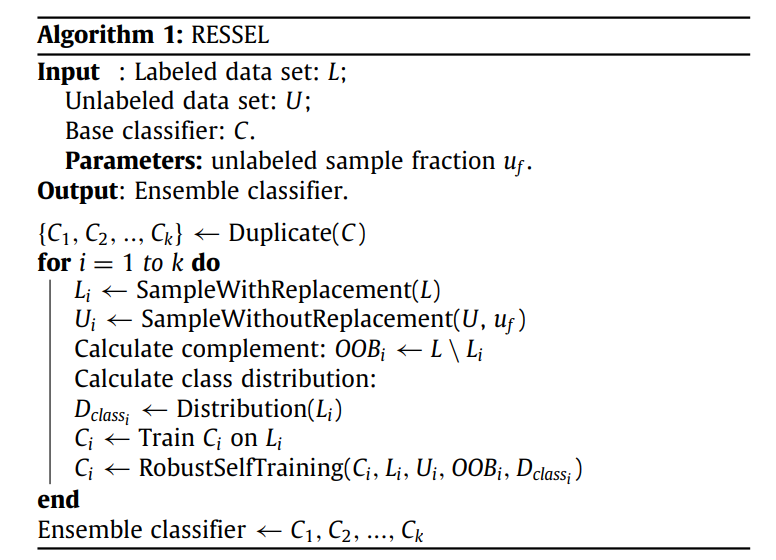
URL: [A reliable ensemble based approach to semi-supervised learning - ScienceDirect](https://www.sciencedirect.com/science/article/pii/S0950705121000010)

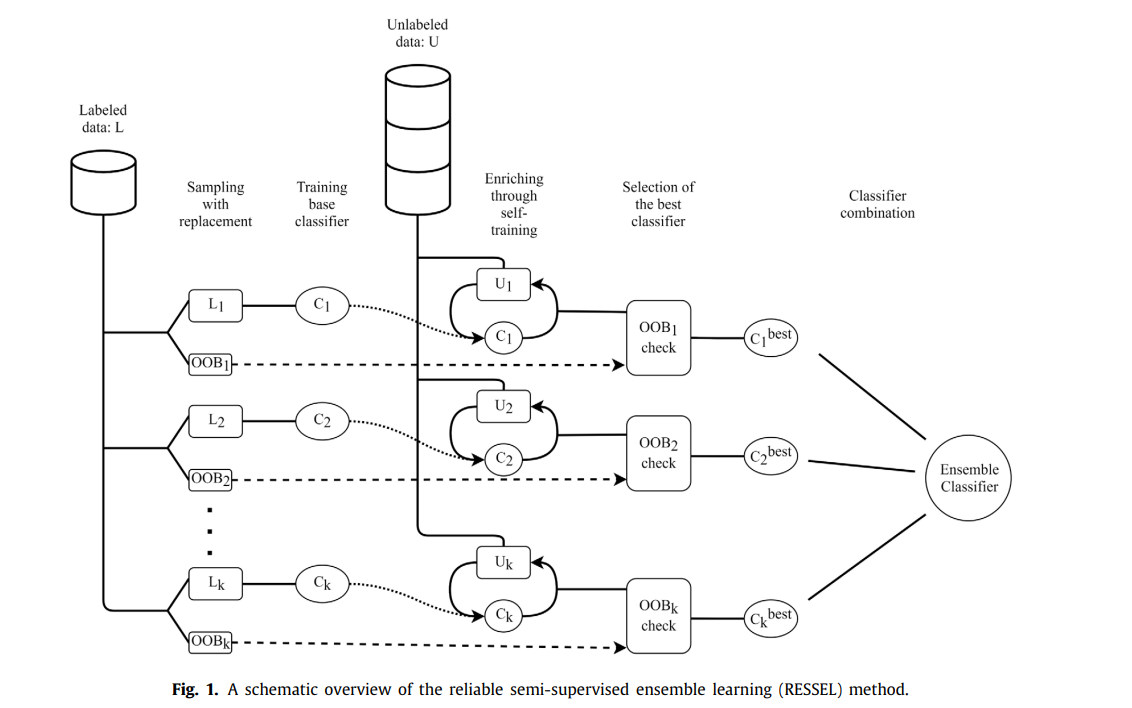
PDF: [main.pdf (sciencedirectassets.com)](https://pdf.sciencedirectassets.com/271505/1-s2.0-S0950705121X00020/1-s2.0-S0950705121000010/main.pdf?X-Amz-Security-Token=IQoJb3JpZ2luX2VjEEMaCXVzLWVhc3QtMSJIMEYCIQD1T8N3eIicN8KdUFVSW6OX8ooSiaWgwDkHhS8WE2rk%2FgIhAI5cC%2B4ytvv410wAY%2BKr2z%2BMXeM05wpwAEn3eqDN5gn0KrMFCGwQBRoMMDU5MDAzNTQ2ODY1IgxnojB3sYMwPt0byGEqkAXK2hJtVfoKVJqdesUC4hHjmCBt36cyf9m%2F6iKntnOjOrk4EcXOweQTV0nAWlbLtVg8O8jAqWTv8ilASUi4FHCHSgJIsHJe9RPLRIdepOq854oQJuKxdqdLSPwuKGC9AvT00%2FUbyZRolSZ1uTUwch4nYGG5gjZSyqkkwhdr2zd9qC%2F7IzY1WevQR02z7CIVEWoD3s5J5sD6q1IZtrH%2F7AgTSwiV3JY%2BAjTE17s0XMxp%2FcvFYmbyfT69%2BR90FnKDRmV3PjICQIHiN9gMAlF%2FVxxdlRq57M7z8Rk72mRklU0Ko%2FpzwtgI%2B94hy0KokQ5RptZU1DWKDIlKbL2%2Fw1GYDrxBhYls8V8Np1pjwLfizpRjAdvl5ycjViFuwzPE39CKMSxbyyr8QRW9VGV57zGn2X4wEQ562YiDvVxWwrpX1yRIiPe5WdhPL11gEQVwEB1DOR9RYkIRhfr8gVweLI4j7pCdwk%2FceM7Haa%2B%2BeBfLYKE6PXERAFbETXMIZv%2BlRVs%2FC8KONcijb285JkrTdeLuLRcEOIkIajPvdKaQGPJrY3AGCoGZ3QhWq1ISsClfRBMxsKIWZAGkr1fCrhuDWDQ4V2FWYe26WS83ntM8NtCWW8DJIYGI3nTbLuN4I4f0XjrBfklI5NYeTEO3IaPaglCLTLkDmOxWvkTIgEYdZOvkcDV3YhhwDWPZwor%2BsI0dTfjFwF0J961cl1uHsCysf871CYhbBMxeKfcwLKHaWWhreUW4SH9S35ZesdjTTXoeRb%2BlhRFjJUhmTrmR9SOIYM%2FS798e7qlCoIKrvs9Hsw46t0ggwvGGQuHmrgi0CCkhRJ95dn2ubh1llUa4tbtxUwFtIGliV05IqsECls4X8hx1Zka1ezD32%2F62BjqwAb4sztlcSWOEVtxud%2B5vpvR3uEOauxKzUbrZ0bzWujY5tYiDMCQStzFPx82%2BjkSPeV%2BBH9KZ0197Wje9K1xnmZAlLzOQTkQzEews8cQFFKsS9KTMS8lxWkNN0ETbVArDbupdYcj0afa62xzdTefdinaf8M%2BmspxqMQ%2BFlLGCHehusJlF9g0Ca9J9NSt1FRj%2BusVq7IDH1yiifYMNt5iu4AFe2%2BzRUgdtgXpbaf%2FsJheX&X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Date=20240910T032116Z&X-Amz-SignedHeaders=host&X-Amz-Expires=300&X-Amz-Credential=ASIAQ3PHCVTY5QBEEPNJ%2F20240910%2Fus-east-1%2Fs3%2Faws4_request&X-Amz-Signature=cf890095a5705097386fbf3a9d17c9850cb08f68484c919670214fb76ac853cf&hash=10bfc643673abe86d82730446452fecdb6b107b838826b4f747fbf3204d1b0a5&host=68042c943591013ac2b2430a89b270f6af2c76d8dfd086a07176afe7c76c2c61&pii=S0950705121000010&tid=spdf-cd26153a-6708-45c6-9387-c80526163157&sid=cf24761191be0448637970c99c74c08104a9gxrqb&type=client&tsoh=d3d3LnNjaWVuY2VkaXJlY3QuY29t&ua=16075f01520c5058515358&)

PDF Was downloaded in this directory

Category: Semi Supervised, Ensemble

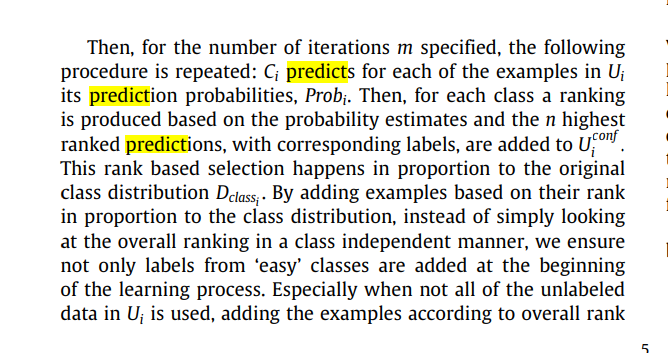
## Study:

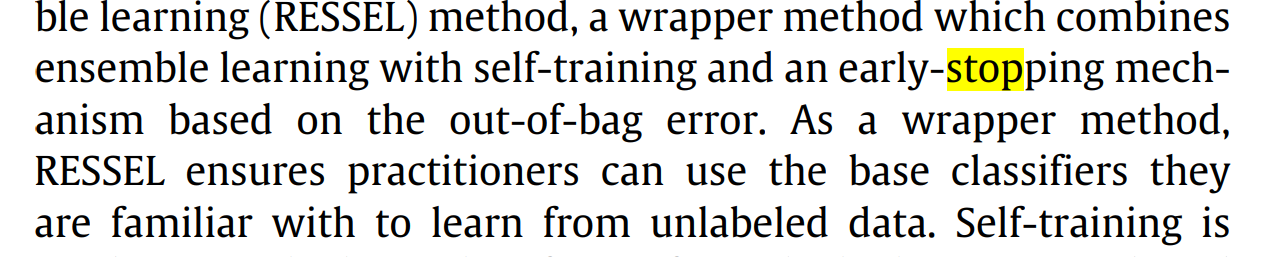


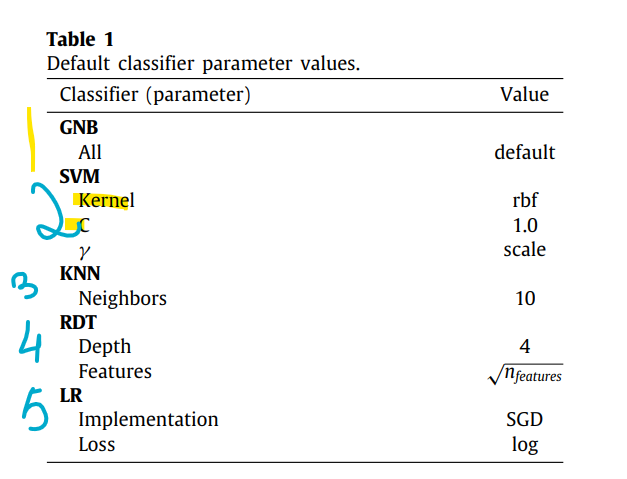


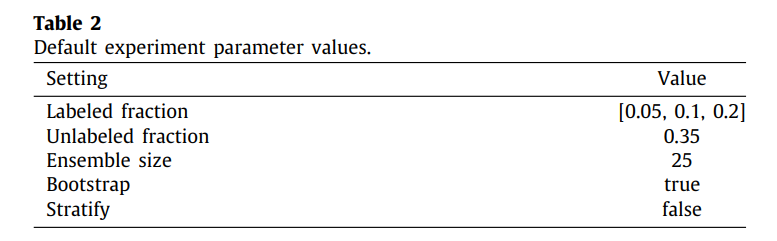
Ok so first we train the base classifier on the Labeled data, Then , when trying to enrich though self training we add the UnLabeled

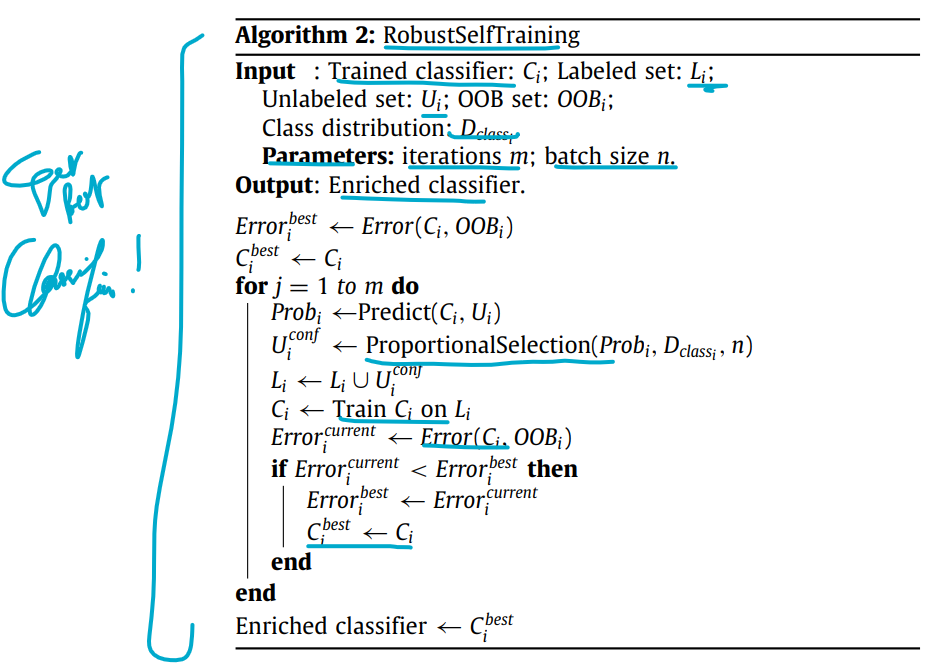
So the ensemble classufuer should technically have all the classifiers trained, How Are we supposed to use it to predict the data correctly.

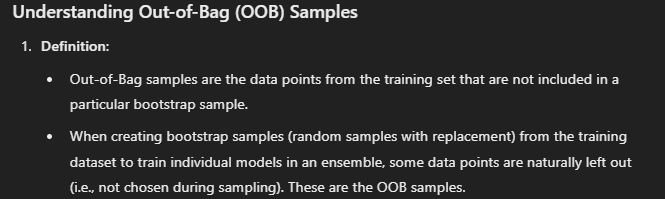


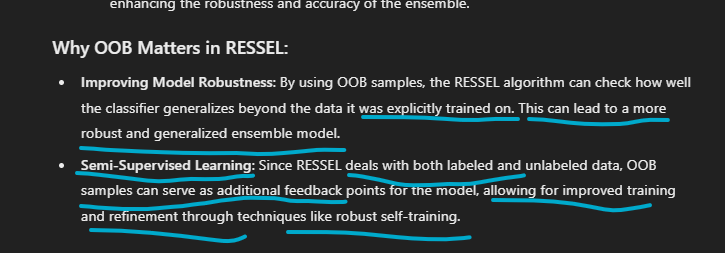










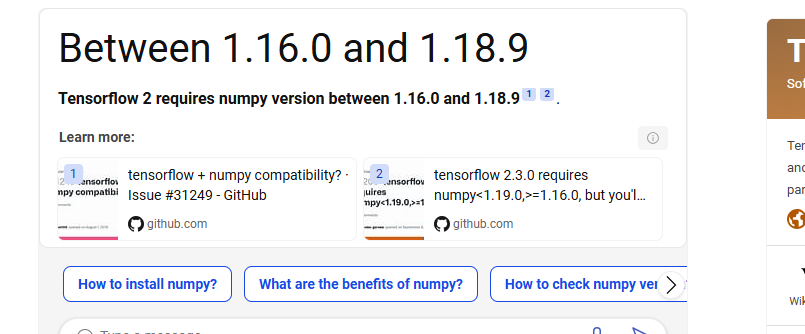


We will first test with One classifier with is Decision tree, just to make sure things work correctly

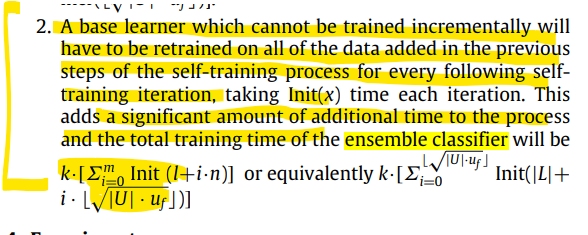
Then we will add the same c;assofiers mentioned in the paper, This way, We followed it correctly.

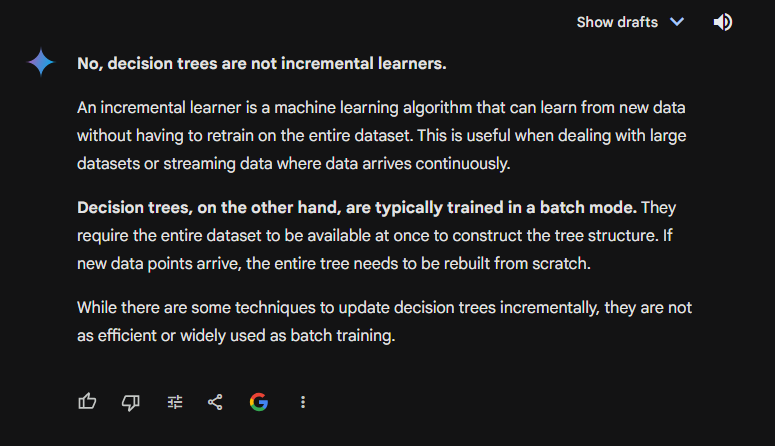
We will set the parameters correctly same as set in the paper as well!

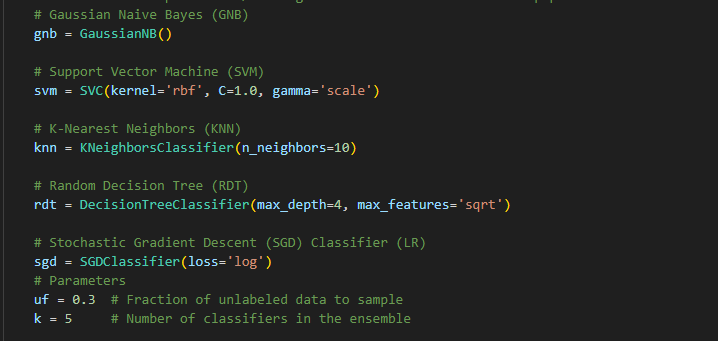
Also, This will be the only code used with the Semi Supervised Ensemble learning, we will add another for the

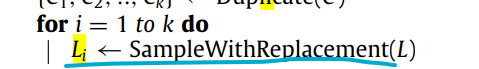


Installed numpy 1.18.4





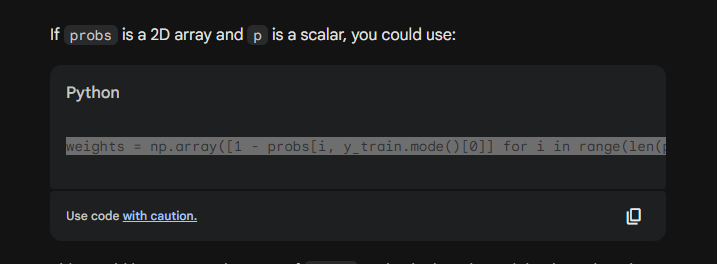
Classifiers I used



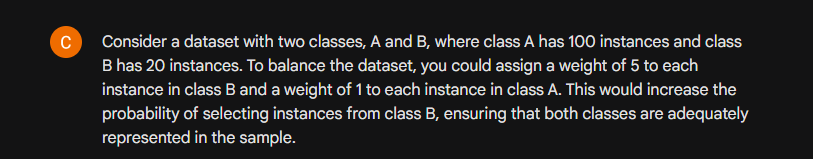
We should sample, with number = size?

I think enno no, Bas rah etreka metel l paper actually, does that take same size? Eh



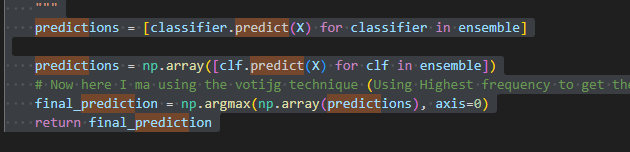


Ok for now the algorithm cannot be more correct than this!



Class weights:

[How to Improve Class Imbalance using Class Weights in ML? (analyticsvidhya.com)](https://www.analyticsvidhya.com/blog/2020/10/improve-class-imbalance-class-weights/)



Argmax btredelle l index!

Laheik aande masgkal ano eende 3 bel output

Lets regenerate the output on #8

Then check if the noisy items are still the same, if the same no issue

If gher, yaane aande mashkal shwey

**mode()**: This function from scipy.stats returns the most frequent prediction along the axis (per sample).

Even After Splitting the data, the output contains 0 noisy items.

So Khalas Last try, check if the indexes exist or no

Else khalas I will take into consideration that its actual output is 0

Yaane hye mawjoude aal axis, yaane in other words, eza hye aala l axis, Yaane ma ken fii te2sir aala l system

Yaane, maybe mneeteber ano this algorithm doesn’t affect our work, hence we keep it

It seems #3 and #4, don’t have anything in common, but they did an output on NF, unlike 0 and 2 and 1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| GaussianNB | SVC | KNeighborsClassifier | DecisionTreeClassifier | SGDClassifier |
| **Classifier\_0** | **Classifier\_1** | **Classifier\_2** | **Classifier\_3** | **Classifier\_4** |
|  |  |  |  |  |

I will test with a variation of only these last two

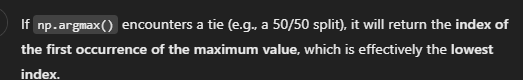
Variation 2 That gave us some answers



I think bel dataset li shtaghaalna aalaya, akhadna l 3 bas into consideration (yemken)

So maybe, maybe, we can test, maa baed Wahad fi 3 algorithms, and we have 2 RDT 1 SGD

Oke wel first variation will be the algorithms, but with the third algo as the target, yaane Decision tree lahala



Aande l awal jeb natije l tene lae