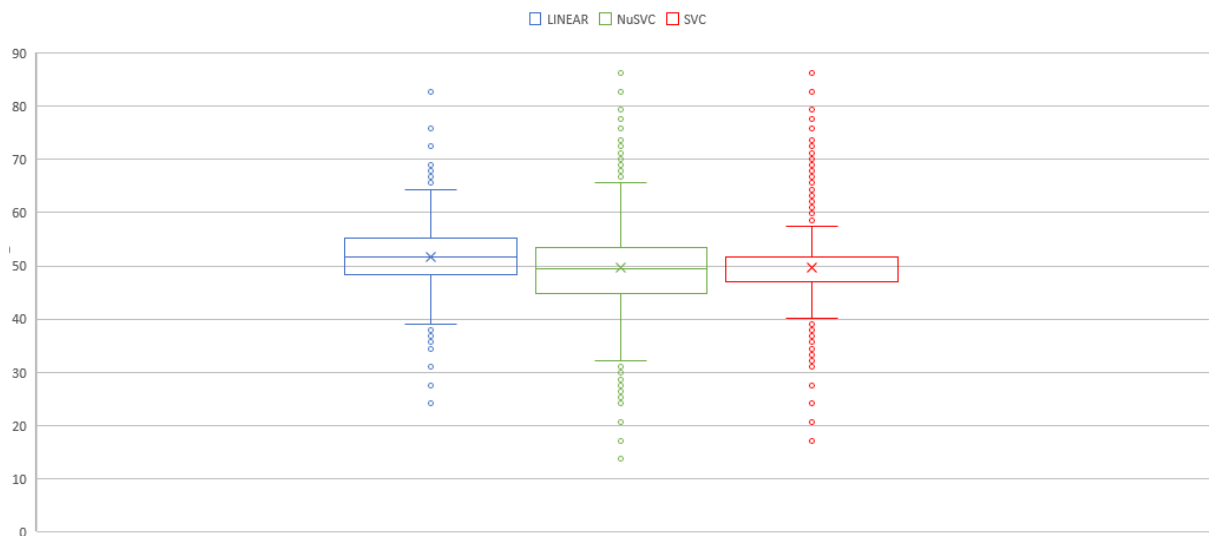


Analyse des SVM et réseau de neurones

L'entraînement est réalisé avec un pool de tous les échantillons fournis, les neurones sont tirés au hasard pour former les 25% d'entraînement et 75% de test et ainsi de suite.

SVM :

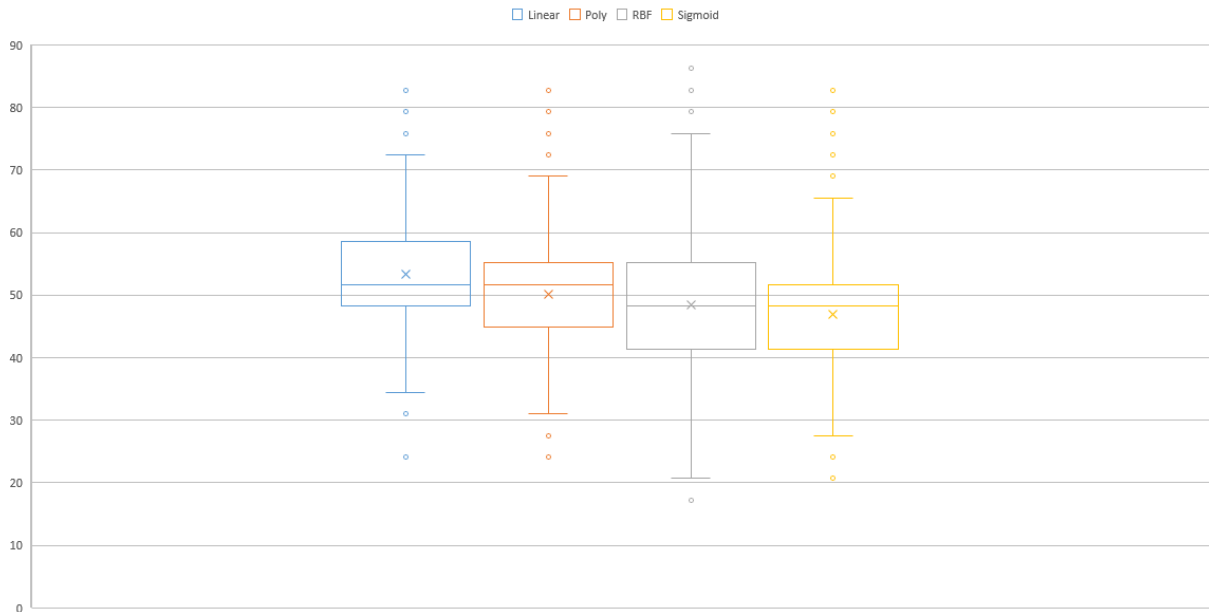
COMPARAISON DES DIFFÉRENTES CLASSES DES SVM DE SCIKIT-LEARN



	Taille du jeu de données	MAX	MIN	MOYENNE	Classement de la vitesse
LINEAR	9937	82,75	24,13	51,70	1
NuSVC	166995	86,20	13,79	49,71	3
SVC	337609	86,20	17,24	49,71	2

Ici on va choisir la classe de SVM : **SVC**, en effet c'est une classe complète qui fournit le plus de résultats positifs, le plus haut score. Cependant au niveau de la moyenne et du minimum on observe que celles-ci ne font pas parties des plus élevées, ce qui peut s'expliquer en raison du grand nombre de données générées.

ANALYSE DES MÉTHODES POUR LA CLASSE SVC

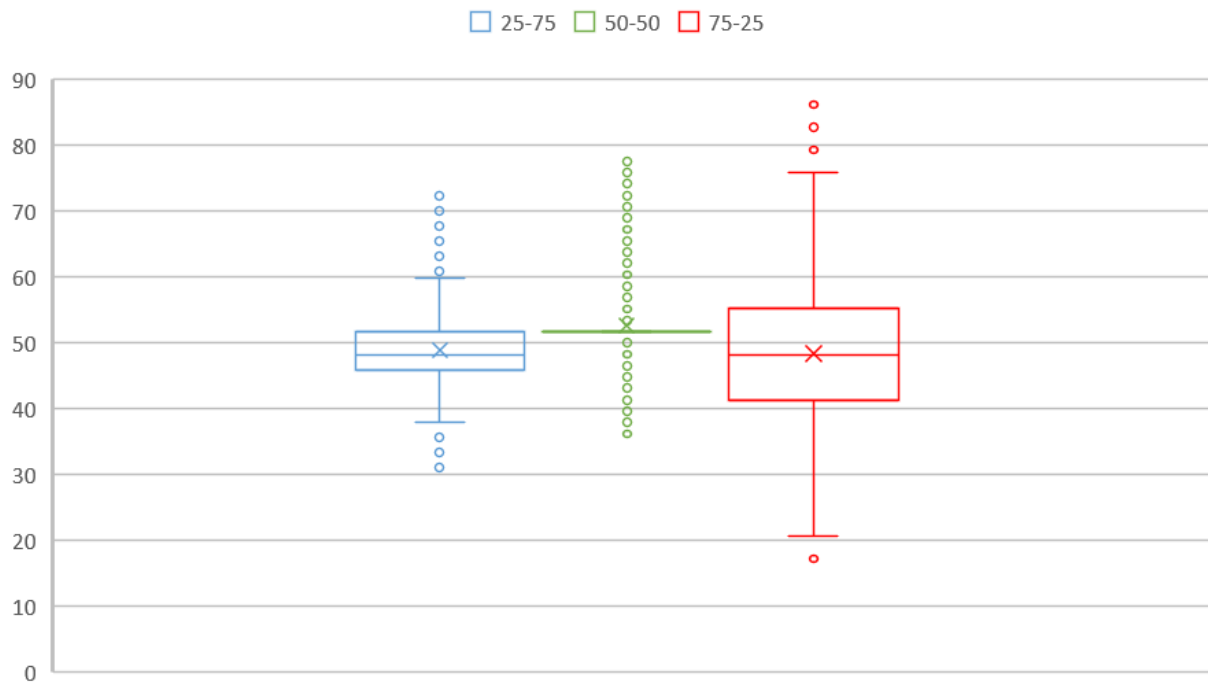


Ici on remarque que le plus grand nombre de haut score est fourni par **la méthode RBF** parmi les méthodes analysées, c'est donc celle qui sera retenue.

Pourcentage	C	Gamma	IR	RMP	RH	ST	DTFS	SA	SD	fAHP
86,2068966	1E+5	1E-6			RH				SD	
86,2068966	1E+6	1E-8			RH				SD	
86,2068966	1E+7	1E-5	IR		RH	ST			SD	
86,2068966	1E+7	1E-8	IR			ST	DTFS	SA	SD	fAHP
82,7586207	1E+4	1E-6			RH				SD	
82,7586207	1E+5	1E-7			RH				SD	
82,7586207	1E+6	1E-6			RH				SD	
82,7586207	1E+7	1E-4				ST			SD	fAHP
82,7586207	1E+6	1E-4	IR		RH	ST			SD	
82,7586207	1E+7	1E-6	IR	RMP		ST			SD	fAHP
82,7586207	1E+7	1E-7			RH	ST		SA	SD	fAHP
82,7586207	1E+6	1E-7	IR	RMP		ST	DTFS	SA	SD	
82,7586207	1E+7	1E-7			RH	ST	DTFS	SA	SD	fAHP

Sachant que nous faisons varier **C** de 1E+7 jusqu'à 1E-5 et **gamma** de 1E+3 à 1E-12.

ANALYSE DE L'ÉCHANTILLONNAGE

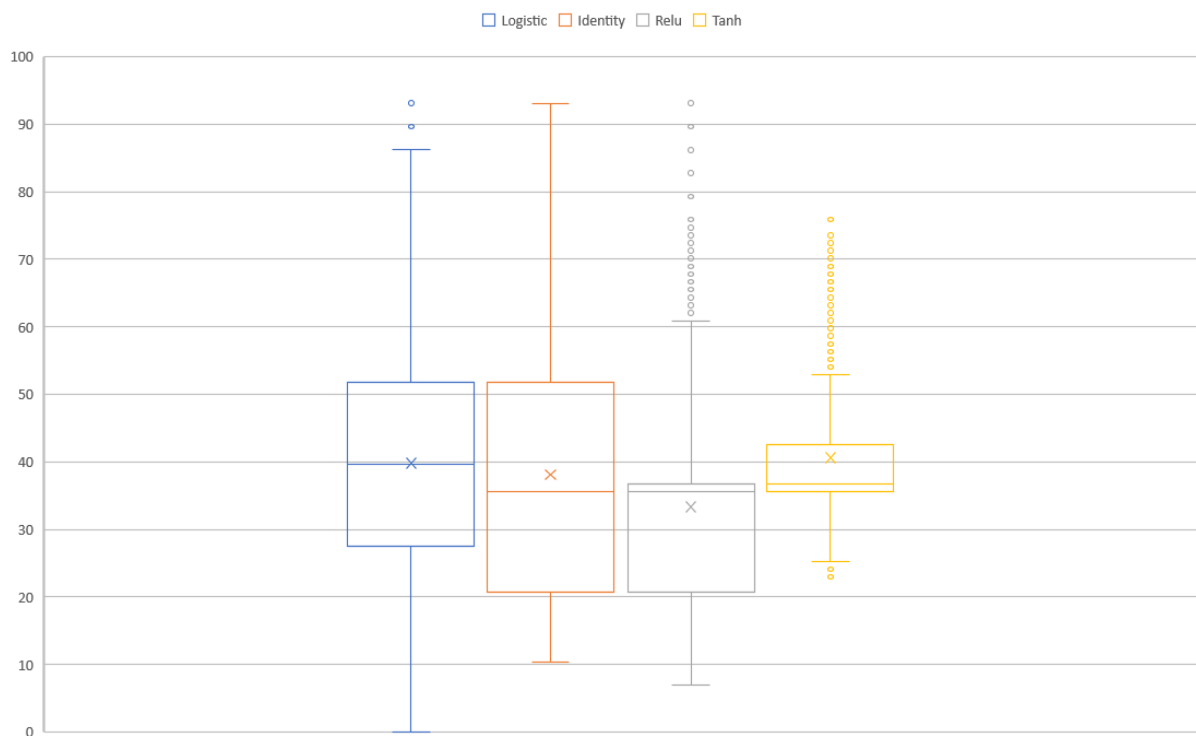


On remarque ici un net effet de l'échantillonnage, en effet même si cela n'est pas significatif on remarque qu'entraîner avec 75% des données et tester avec 25% permet d'obtenir les meilleurs résultats.

➔ On va donc réessayer les combinaisons de paramètres vu dans le tableau pour un $C[1E+7 : 1E+4]$ et un $\gamma[1E-4 : 1E-8]$, afin de vérifier que ces combinaisons soient convenables pour l'entraînement du modèle.

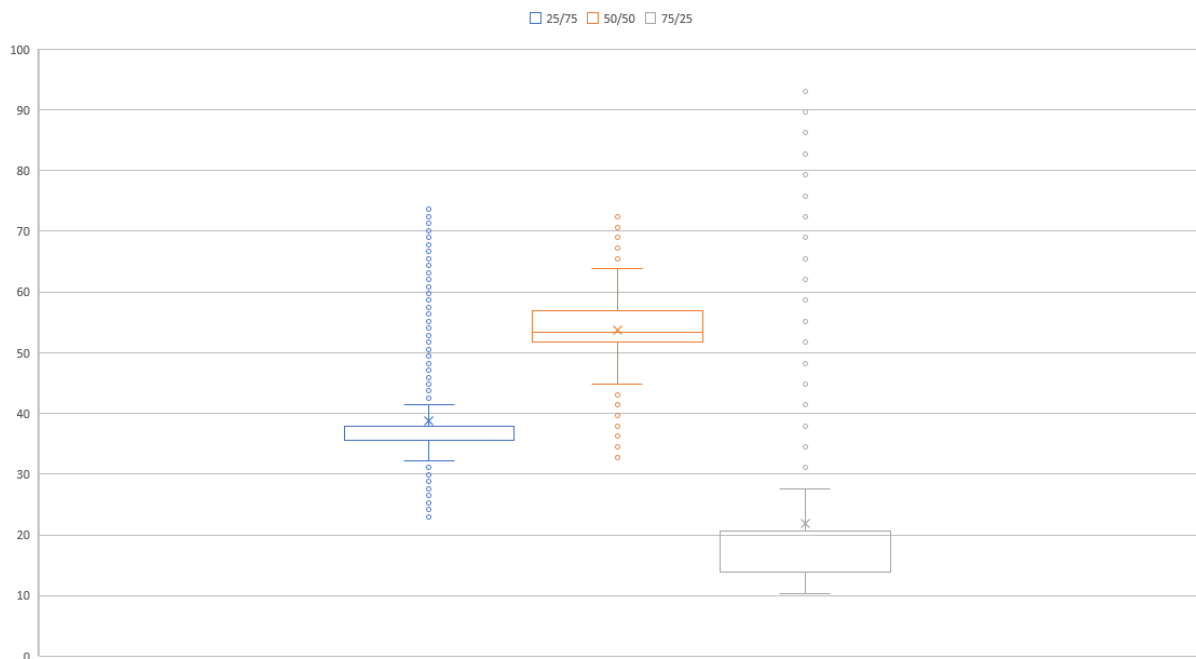
Réseau de neurones :

COMPARAISON DES MÉTHODES



Ici on va préférer la méthode **Identity**, en effet elle est la méthode la plus lente mais elle fournie le plus fort aux de réussite parmi les réseaux de neurones.

COMPARAISON DE L'ÉCHANTILLONNAGE SUR LA MÉTHODE IDENTITY



Ici on remarque une structure particulière, en effet les résultats sont les plus hauts pour l'échantillonnage selon un motif 75% entraînement et 25 % test, or on observe que la plupart des résultats pour cet échantillonnage se retrouvent être des taux de réussite relativement faible.

Cependant le fait qu'il fournisse de très hauts taux de réussite nous oriente vers ce choix, on gardera alors la méthode Identity avec 75% d'entraînement et 25% de test.

<i>Pourcentage</i>	<i>alpha</i>	<i>tall</i>	<i>Validation fraction</i>	<i>IR</i>	<i>RMP</i>	<i>RH</i>	<i>ST</i>	<i>DTFS</i>	<i>SA</i>	<i>SD</i>	<i>fAHP</i>
93,1034483	1E-5	100	0,3								fAHP
89,6551724	1E-4	1	0							SD	
89,6551724	1E+1	1	0,8							SD	
89,6551724	1E-4	10	0								fAHP
89,6551724	1E+1	10	0								fAHP
89,6551724	1E-5	10	0,1								fAHP
89,6551724	1E-3	1	0,2								fAHP
89,6551724	1E-4	1	0,8								fAHP
89,6551724	1E-2	1	0,3							SD	fAHP
89,6551724	1E-3	100	0,4	IR		RH	ST			SD	
89,6551724	1E-3	100	0,9	IR	RMP	RH	ST	DTFS			
89,6551724	1	100	0,9	IR	RMP	RH	ST			SD	
89,6551724	1E-5	100	0,7	IR	RMP		ST	DTFS	SA	SD	
89,6551724	1E-5	100	0	IR		RH	ST	DTFS	SA	SD	
89,6551724	1E-4	100	0,5	IR	RMP	RH		DTFS		SD	fAHP
89,6551724	1E+2	100	0,5	IR	RMP	RH		DTFS		SD	fAHP
89,6551724	1E-4	100	0,5	IR		RH	ST		SA	SD	fAHP

➔ On va donc réessayer les combinaisons de paramètres vu dans le tableau pour un alpha compris entre [1E-5:1E+2], un Validation-fraction compris entre [0:1] et un tall compris entre [1:1E+2], afin de vérifier que ces combinaisons soient convenables pour l'entraînement du modèle.