



République Algérienne Démocratique et populaire
Ministère de l'Enseignement Supérieur et de la Recherche
Scientifique
Université des Sciences et de la Technologie Houari Boumediene
Faculté d'Electronique et Informatique
Département Informatique.

Rapport de projet

Module : Apprentissage automatique et réseaux de neurones
Master 1 SII

Intitulé du projet :

*Implémentation d'un réseau de neurones sur le dataset
Student's Academic Performance*

Réalisé par :

Dahimene Ouassil 201400008363

Messaouali Abdeslem 201500008151

Groupe1

Table des matières

1. Introduction :	3
2. Problématique et objectif :	3
3. Objectif :	3
4. Solution proposée :	3
a. Encodage du dataset :	3
b. Partitionnement des données :	3
c. Apprentissage et analyse des résultats :	3
i. Détails à propos du paramétrage des hyper-paramètres :	3
ii. Apprentissage :	4
iii. Essai de prédiction :	5
d. Application :	5
5. Conclusion général :	5

1. Introduction :

De nos jours, l'intelligence artificielle est devenue indispensable pour l'évolution de l'homme dans tous les domaines, notamment l'un de ses piliers est l'apprentissage automatique, qu'on peut le définir par un système qui simule le raisonnement d'un cerveau à l'aide de réseau de neurones artificiels qui est inspiré d'un réseau de neurone biologique, l'objectif de ce mini projet est d'appliquer ces notions théoriques de l'apprentissage automatique dans un réseau de neurone développé en utilisant un ensemble de données spécifiques au domaine donné.

2. Problématique :

Les performances des étudiants diffèrent de l'un à l'autre et qu'on peut les classer dans des catégories selon leurs notes finales. Celles-ci, sont étroitement liées à plusieurs aspects de la vie de l'étudiant comme sa proximité à l'école ou situation financière ou familiale.

Connaitre approximativement la performance de l'étudiant selon sa situation pourrait aider à améliorer les conditions de travail et de vie de celui-ci et pour cela nous allons essayer de concevoir une application qui permet de prédire la note finale de l'étudiant par rapport à ces circonstances.

3. Objectif :

Objectif de ce mini projet est de :

- Proposer un encodage du data pour donner un meilleur résultat d'apprentissage.
- Choisir l'architecture de réseau de neurone la mieux adaptée pour ce dataset et configurer ses paramètres.
- Comparer les différents résultats des différentes configurations et choisir la meilleure d'entre elles.
- Implémenter l'interface de l'application pour utiliser le modèle choisit.

4. Solution proposée :

Afin de développer cette application on doit concevoir un réseau de neurone vu la complexité de l'algorithme, donc nous avons décidé de choisir le dataset étudiant de la langue portugais et le langage de programmation python en utilisant 2 bibliothèques tensorflow[1] et keras[2] ensuite après l'apprentissage on essaie de générer un code JS et l'implémenter dans une application web.

a. Encodage du dataset :

Pour cette étape nous avons décidé de travailler avec le dataset 'student-por' qui contient 649 entrées car c'est celui qui contient le plus d'entrées mais aussi car si on le fusionne avec le dataset 'student-mat' il y'aurait des entrées qui n'auront pas de valeurs chez l'un ou l'autre. On a choisi les 32 premiers attributs comme input et la note finale (G3) comme target.

L'encodage a été choisi selon les valeurs des attributs et sa description complète est disponible dans le fichier "description-Attribut-Codification.txt" dans le dossier 'Informations'.

b. Partitionnement des données :

Dans cette étape On a partitionné les données en 3 ensembles :

- 85% de l'ensemble de données utilisé pour l'apprentissage.
- 15% de l'ensemble de données utilisé pour la validation.
- 15% de l'ensemble de données utilisé pour le test.

Remarque : On mélange d'abord la distribution du dataset aléatoirement.

c. Apprentissage et analyse des résultats :

i.Détails à propos du paramétrage des hyper-paramètres :

Pour trouver l'architecture la plus adaptée pour ce réseau de neurone on a décidé de varier :

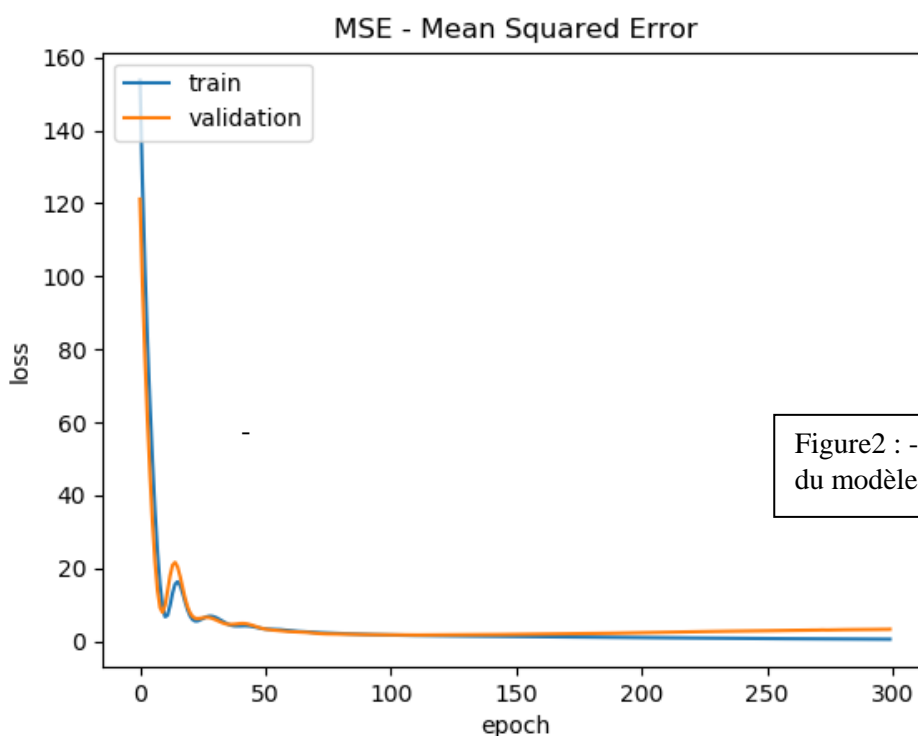
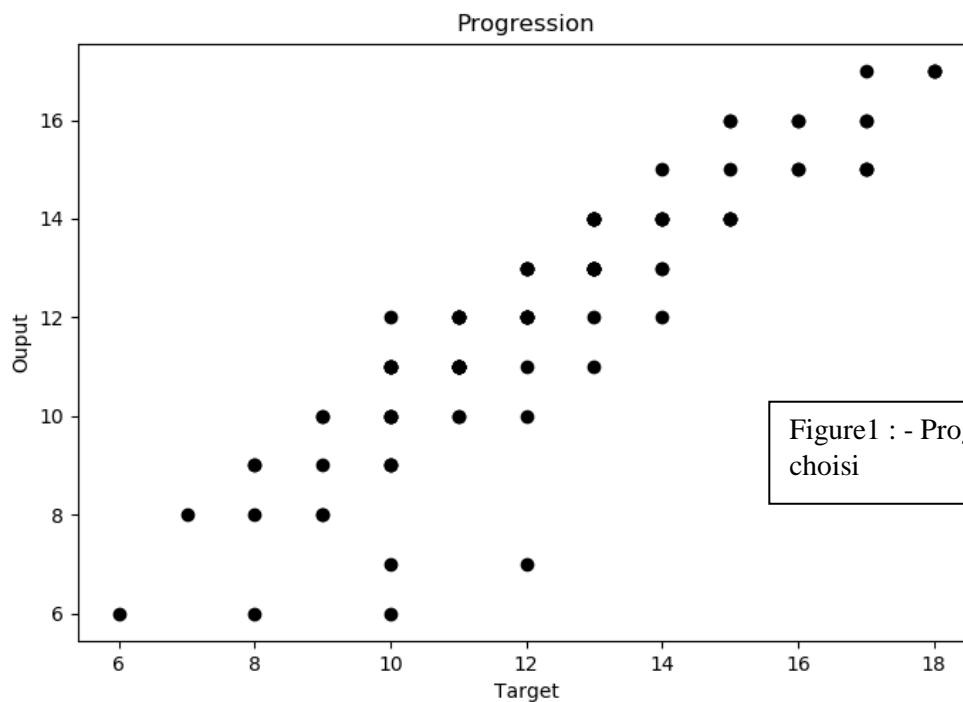
- Le nombre des couches de 1 à 4.
- Le nombre de neurones dans une couche de 10 à 1000 neurones.

- Les fonctions d'activation: RELU « $f(x) = \max(0, x)$ » [3] et linière « $y=x$ » [3].
- Les fonctions d'optimisation: Adam [4], Adamax [5], FTRL [6] et Adagrad [4].
- Le nombre d'epochs de 100 à 1500.

ii.Apprentissage :

Après des heures d'entrainement de plusieurs configurations que nous avons inséré dans un tableau excel (fichier disponible dans le dossier 'Informations'), nous avons choisi l'une des meilleures configurations qu'on a pu avoir en les comparant avec leurs performances (MSE 'mean squared error' « $f(x) = \frac{1}{n} \sum_1^n (y_i - \tilde{y}_i)^2$ ». [7]) et en faisant attention à ne pas choisir un modèle qui a fait un overfit en comparant le taux d'erreur dans l'apprentissage avec le taux d'erreur dans les tests.

La configuration choisi est : 4 couches, 50 neurones, 300 epochs,



iii.Essai de prédiction :

Nous avons utilisé des données de la partie Test pour prédire leurs notes finales en utilisant le script test-script.py (disponible dans le dossier dataset + scripts + model) et voici un petit échantillon du résultat proposé :

TARGET																			
17.0	16.0	12.0	11.0	10.0	11.0	10.0	13.0	14.0	13.0	11.0	13.0	6.0	14.0	16.0	12.0	12.0	16.0	11.0	13.0
Ouput																			
15.	16.	13.	11.	10.	11.	9.	13.	14.	14.	11.	13.	6.	14.	15.	13.	12.	15.	12.	14.

d. Application :

Voici un screenshot qui montre l'interface de l'application qui sous forme d'une page web html avec un framework javascript "TensorflowJs" [8] qui sera utilisé pour importer le modèle entraîné pour l'utiliser.

Prediction de la note d'examen de la langue portugaise selon la situation de l'élève

Ecole	Sexe	Age	Adresse	Nb Membres famille	Situation des parents	Niveau de la mere
Domaine de travail du pere	Domaine de travail de la mere	Domaine de travail du pere	Raison de choix de l'école	Responsable		
Temps de transport	Temps d'étude par semaine	Années redoublées	Cours sup a l'école	Aide familiale en etude		
Cours sup payants	Activités en plus	Attented nursery	Vouloir s'améliorer	Internet a la maison	En couple	
Relation avec la famille	Temps Libre	Sorties entre Amis	Consommation d'alchool le weekend la semaine			
Consommation d'alchool le weekend	Santé	Nombre d'absences	Note G1	Note G2		

Predire Note de l'examen =

-Figure 3 : - Interface de l'application

5. Conclusion général :

Durant ce projet, nous avons approfondi nos connaissances dans le domaine de l'apprentissage automatique et réseaux de neurones en général, et cela nous a permis de passer de la théorie à la pratique mais aussi a utilisé différents outils et méthodes.

Répartition des taches :

- Messaouali Abdeslem :
 - Recherche bibliographique et rédaction du rapport.
 - Cleaning et encodage du dataset
- Dahimene Ouassil :
 - Programmation du script et interface.
 - Apprentissage sur le dataset.

Références :

- [1] <https://www.tensorflow.org/>
- [2] <https://keras.io/>
- [3] [https://en.wikipedia.org/wiki/Rectifier_\(neural_networks\)](https://en.wikipedia.org/wiki/Rectifier_(neural_networks))
- [4] http://www2.ift.ulaval.ca/~pgiguere/cours/DeepLearning/03-Deep_Learning_Optimization.pdf
- [5] <https://towardsdatascience.com/10-gradient-descent-optimisation-algorithms-86989510b5e9>
- [6] <https://courses.cs.washington.edu/courses/cse599s/12sp/scribes/Lecture8.pdf>
- [7] <https://www.freecodecamp.org/news/machine-learning-mean-squared-error-regression-line-c7dde9a26b93/>
- [8] <https://www.tensorflow.org/js>