



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 d'Informatique  
Département Informatique

---

# L'IA au service de l'apprentissage

## (Cas : Etudiants Portugais)

---

M1 SII 2018/2019

Module : Apprentissage Automatique et Réseaux de Neurones

**ADLI Housseem** 201500008444 ad.ho@yahoo.com

**DJAIDJA Taki Eddine Toufik** 201500008653 d.taqei13@gmail.com

Groupe 02.

# 1 Introduction

Si nous définissons de façon simpliste, l'IA est un moyen de développer des programmes efficaces pouvant remplacer l'activité humaine sans chercher à modéliser l'esprit humain. Boîte noire intelligente découlant des domaines des systèmes experts (réseaux néroniens ou simulation du comportement interne) et des heuristiques [2].

Le concept *smart home* ou maison intelligente (thermostat, caméra de surveillance et éclairage ..) et l'analyse des fameuses *Big Data* permettant d'obtenir une image très précise du trafic routier en cours et de réaliser, par un processus de *Machine Learning*, des prévisions du trafic à venir, ne sont que des exemples parmi de très nombreux cas réalisés.

Des applications similaires en matière d'éducation et d'apprentissage ont été traitées et qui ont pour finalité l'amélioration, l'assistance et le calibrage de la qualité de l'apprentissage à grande échelle.

Dans cette vision nous nous sommes inspirés pour l'élaboration d'une application dont l'objectif : orientations et conseils aux étudiants portugais.

L'exposé comprend :

- Position de la problématique ainsi que les données nécessaires ;
- Traitement des données et modélisation ;
- Résultats et leurs exploitations ;
- Brève conclusion.

## 2 Problématique

Dans le but d'améliorer les performances des étudiants qui ont réussi leurs examens du 2ème cycle, on veut élaborer un système de conseils dédiée aux étudiants portugais .

## 3 Données

Afin de résoudre la problématique, on doit d'abords prédire les notes des examens finals du 3eme cycle selon le profil et les informations personnelles de l'étudiant. Pour cela, on a utilisé une étude faite sur les étudiants des établissements (Gabriel Pereira et Mousinho da Silveira) contenant les différentes informations sur l'étudiant ainsi que les notes obtenues dans les examens finals de mathématique et de langue portugaise (Student Performance) [4].

### 3.1 Prétraitement des données

Ces données sont séparées en deux parties (input et target) dont target représente la valeur qu'on cherche à prédire (les notes des examens en 3eme cycle).

Après vérification de la cohérence des données dans les deux datasets, on procède à la codification. Un script a été établi (**encode.m**) permet de codifier et de séparer les input et le target à partir du dataset(.csv).

Pour le target, une décritisation de la note (0..20) a été faite selon la réglementation en vigueur du ministere de l'éducation portugais [3].

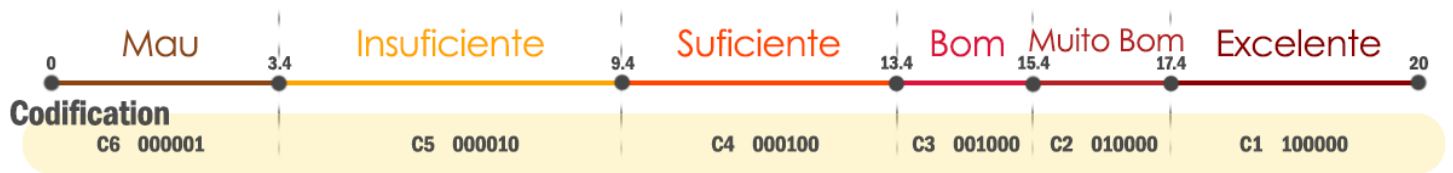


FIGURE 1 – Système de notation portugais

Le fichier (dataset/**dataset.txt**) contient la description des attributs ainsi que leur codification ; Les dossiers (datasets/**mat/** et dataset/**por/**) contiennent les datasets sources ainsi que la codification des dataset.

**Distribution des classes** (le nombre d'instances par classe)

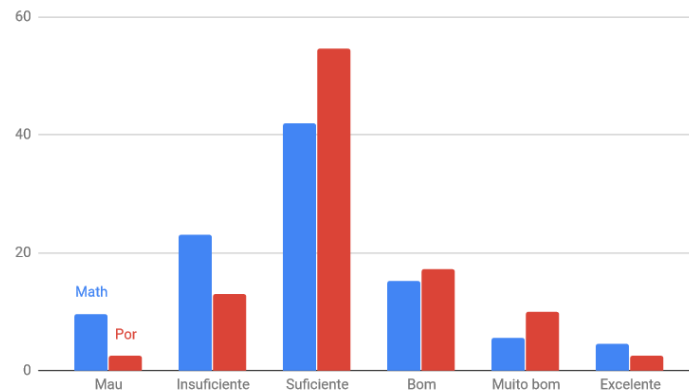


FIGURE 2 – Distribution des classes

## 4 Apprentissage

Après avoir fini le pré traitement des données, on passe à l'apprentissage ,on cherche à trouver la meilleur architecture qui donne des résultats satisfaisants. Une architecture d'un réseau de neurones avec propagation arrière comprend :

- la fonction d'activation du réseau.
- le nombre de couches cachées ainsi que le nombre de neurones dans chaque couche.
- la fonction de transfert de la couche output.

L'espace de recherche contient les différentes combinaisons que peut prendre une architecture, dans le but d'optimiser cette recherche, le recours à une méthode de résolution méta-heuristique est préférable.

### 4.1 BBO Algorithmme

Biogeography-Based Optimization est un algorithme évolutionnaire inspiré d'un processus naturel portant l'appellation : biogéographie. Les modèles mathématiques de biogéographie étudient la répartition et la migration des espèces biologiques dans le domaine spatial entre les îles.

Les îles avec un HSI (Habitat Suitability Index) élevé accueilleraient de nombreuses espèces, donc le HSI représente la **fitness** [5]. On a choisi cette approche pour deux raisons :

- Vouloir exploiter de nouveau algorithmes évolutionnaires.
- BBO a prouvé une convergence plus rapide que les algorithmes génétiques et PSO [1].

## 4.2 La forme d'une architecture

Une architecture est représentée par un vecteur de 7 cases : 

FctAct	FctTrns	NC	$C_1$	$C_2$	$C_3$	$C_4$
--------	---------	----	-------	-------	-------	-------

- Fonction d'activation : {1 : "trainrp", 2 : "trainscg", 3 : "traincgp"}
- Fonction de transfert : {1 : "logsig", 2 : "tansig", 3 : "purelin"}
- NC : le nombre de couches cachées {1,2,3,4}
- $C_i$  : le nombre de neurones dans la couche  $i = [1,150]$ .

## 4.3 HSI

La fonction HSI permet d'évaluer la qualité d'une solution (architecture) :

$$HSI_{SOL} = 1 - (bestPerf + 2 * |bestPerf - bestTPperf|)$$

Une meilleur solution est celle dont le taux d'erreur est minime ( $\simeq 0$ ) et qui ne cause pas un sur-apprentissage (overfitting), le but est de trouver une solution où  $HSI \simeq 1$ .

## 4.4 Pseudo-Code

---

### Algorithme 1 : Algorithme d'Apprentissage

---

```

1 population ← GenererDesSolutionAléatoires() ;
2 HSI(population) // 3 retrain pour chaque solution
3 tant que  $\neg$  ConditionArret faire
4   | Migration ;
5   | Mutation ;
6   | Calculer HSI // retrain pour chaque solution
7 fin
8 Slectionner BestSolution
9 Calculer HSI // 100 retrain
10 retourner BestSolution ;

```

---

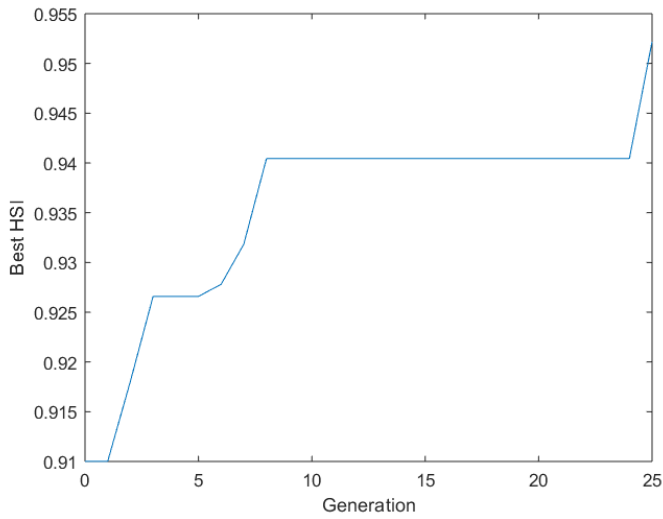
La description des différents scripts implémentés est dans l'annexe B.

Le schéma suivant illustre le processus d'apprentissage :

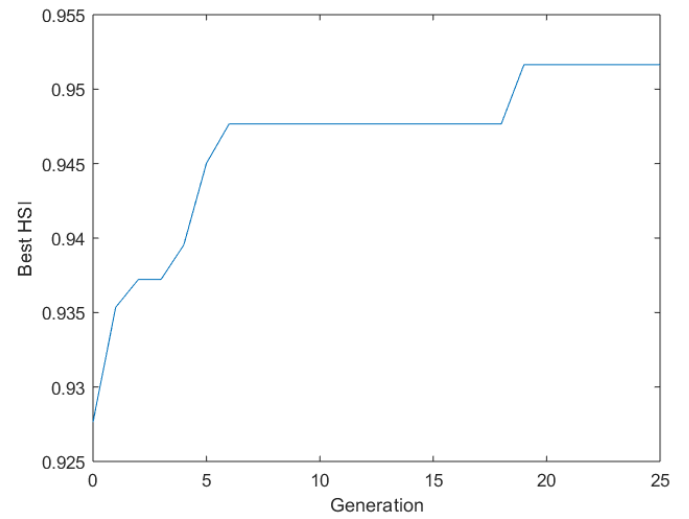


FIGURE 3 – Processus d'apprentissage

Le script **bbo** développe les solutions où le HSI est souhaitable ( $\simeq 1$ ), les deux graphes suivants montrent le HSI obtenu dans chaque génération de l'algorithme (itération).



(a) Math Dataset



(b) Por Dataset

FIGURE 4

Les différentes architectures explorées sont enregistrées dans les dossiers **mathtests** et **portests**. Après avoir trouvé la meilleure architecture, on a rajouté une nouvelle couche qui a pour but de trouver le résultat finale d'un *input*, le *output* sera la classe qui a la grande probabilité d'être sélectionnée (**output.m**).

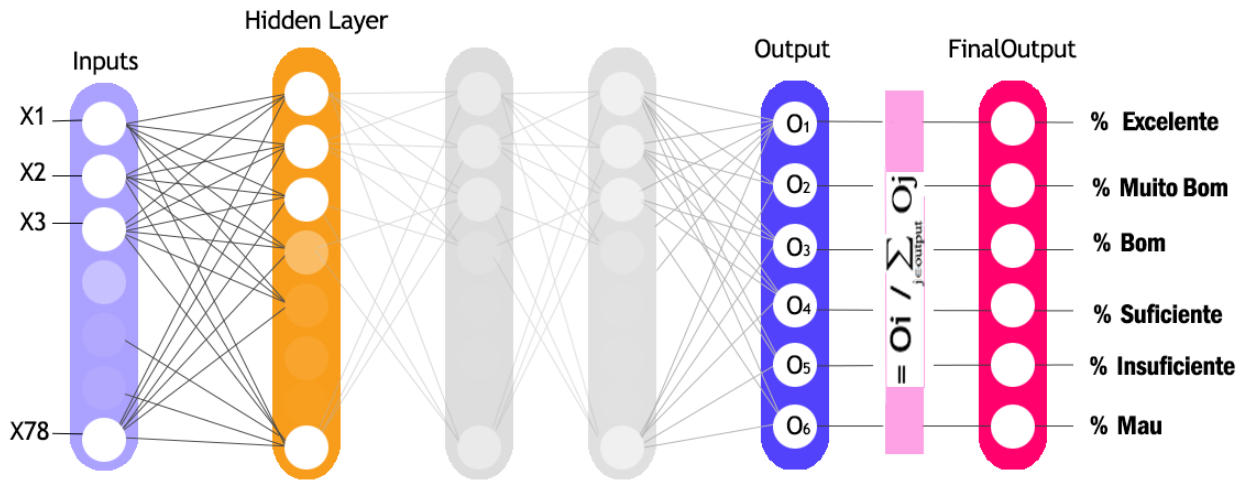


FIGURE 5 – Output d'un RN

## 4.5 Résultats

**Nombre d'architectures parcourues** (MathDataSet) : 304, (PorDataSet) : 280.

les matrices (matbest/**matbests.mat**) et (porbest/**porbests.mat**) contient les meilleures 30 architectures trouvées, dans l'annexe A on cite quelque architectures. On a fait 100 **retrain** pour la meilleure architecture trouvée :

DataSet	Architecture	train perf	vperf	Regression	Exactitude	HSI
Math	trainscg,tansig,[13]	0.0468	0.0664	0.9035	81.27%	0.9522
Por	trainscg,tansig,[73]	0.0473	0.0664	0.86658	79.82%	0.9516

## 4.6 Validation du modèle

On juge les taux d'exactitude (81.27% pour math, 81.27% pour Por) bons mais insuffisants, on cherche une exactitude  $\simeq 100\%$ , les faiblesses des taux est compte tenu de notre méconnaissance du problème étudié, Les graphes suivants montrent que en enrichissant notre base de connaissance le taux d'exactitude (**accuracy**) continue à s'accroître.

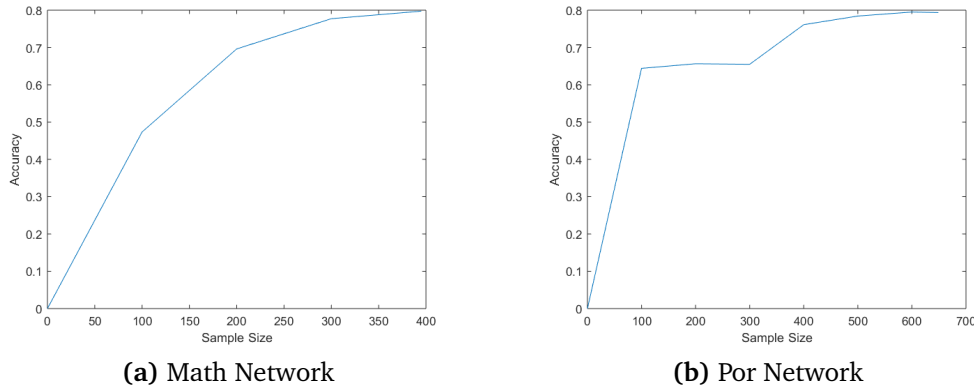


FIGURE 6 – Accuracy

## 5 Application

L'objectif de notre application est d'élaborer un système de recommandations dédié aux étudiants portugais. L'idée, d'une part, est de prédire les résultats dans les examens du **3ème** cycle en introduisant les informations initiales sur l'étudiant en question et d'une autre part, de trouver les attributs (comportements) à maximiser ou à minimiser Notre vision est d'améliorer les résultats à prédire et de garantir une meilleur performance. Pour optimiser la recherche des attributs à modifier, on a utilisé une métaheuristique (BBO [5]) , où :

$$HSI_{Comportement} = \frac{(rate\_math * (6 - class\_math)) + (rate\_por * (6 - class\_por))}{distance(Comportement_0, Comportement)}$$

Remarques :

- La structure **Comportement** est un vecteur de taille 5 où chaque case représente un comportement (attribut) modifiable : 

StudyTime	SchoolSup	GoOut	Dalc	Walc
-----------	-----------	-------	------	------
- **distance(Comportement<sub>0</sub>, Comportement)** est la distance cartésien entre les deux point *Comportement<sub>0</sub>* et *Comportement* , où dans *Comportement<sub>0</sub>* on trouve les informations initiales de l'étudiant ; une recommandation doit être raisonnable et faisable.

## 6 Conclusion

L'IA est domaine très vaste et pluridisciplinaires. Le cas traité nous a permis de découvrir, comprendre et d'appliquer nos connaissances en la matière. Les résultats obtenus grâce au modèle proposé (architecture, algorithme ...) semblent acceptable compte tenu de notre méconnaissance des détails et de la réalité de l'apprentissage au Portugal.

# A Meilleures Architectures

## A.1 Mathématique

Les meilleures architectures trouvées pour l'examen de mathématiques sont :

Fonction d'Activation	Fonction de transfert	Couches	HSI
trainscg	tansig	[13]	0.9522
trainscg	tansig	[13 112 127]	0.9404
trainscg	purelin	[13]	0.9397
traincgp	tansig	[13]	0.9372
trainscg	tansig	[13 112]	0.9255

TABLE 1 – Math : meilleures architectures

La meilleure architecture :

- Fonction d'apprentissage : trainscg.
- Fonction de transfert : tansig.
- une seule couche cachée avec 13 neurones.

La régression est calculée après l'amélioration du résultat obtenu par le réseau de neurones 4.4.

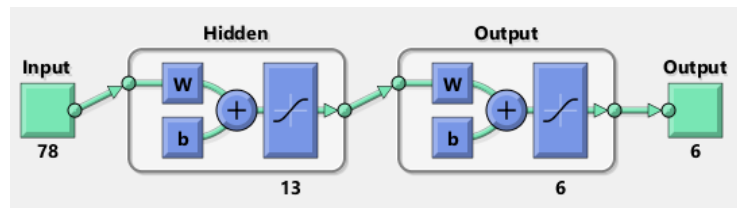
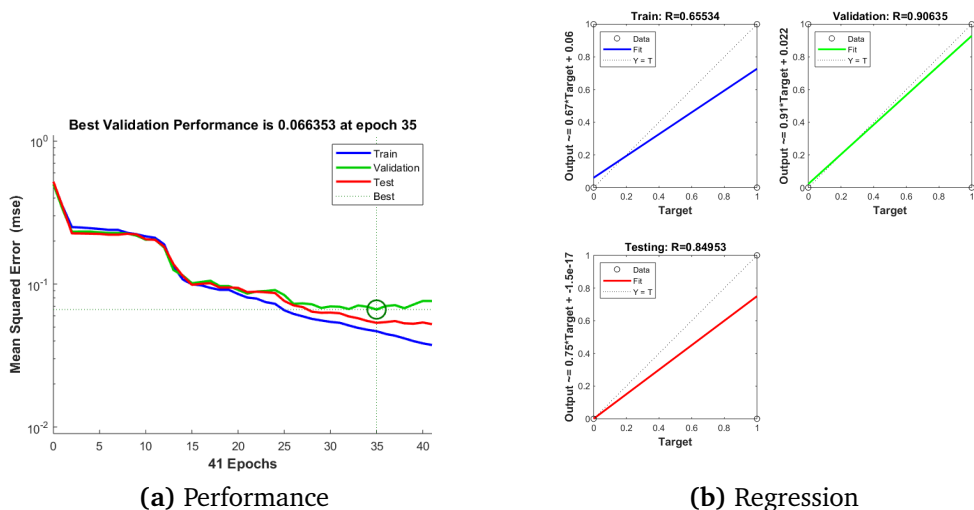


FIGURE 7 – Math Network



## A.2 Langue Portugaise

Les meilleures architectures trouvées pour l'examen de langue portugaise sont :

Fonction d'Activation	Fonction de transfert	Couches	HSI
trainscg	tansig	[73]	0.9516
trainscg	tansig	[133]	0.9452
trainscg	tansig	[73 25]	0.9261
trainscg	tansig	[96]	0.9171
trainscg	tansig	[73 25 12 17]	0.9143

TABLE 2 – Por : meilleures architectures

La meilleure architecture :

- Fonction d'apprentissage : trainscg.
- Fonction de transfert : tansig.
- une seule couche cachée avec 73 neurones.

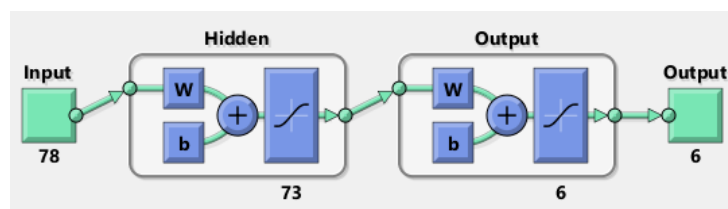
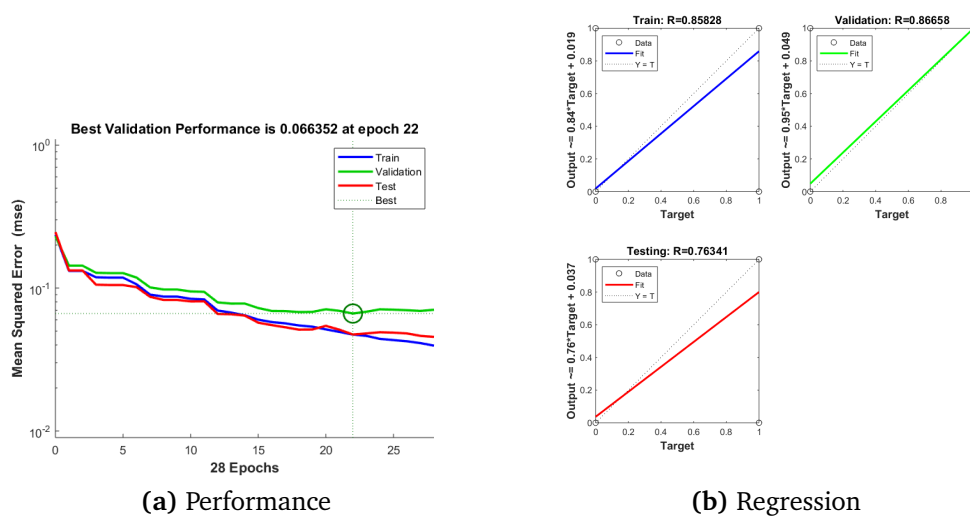


FIGURE 8 – Por Network



## B Scripts

### encode.m

**function [input,target] = encode(mat)**

**Entrée :** mat , est une matrice de chaines de caractères générée lors l'importation du dataset. **Sortie :** les deux matrices input et target, après codification des différents champs.



## **bbo.m**

**function [bests]=bbo(input,target)**

**Entrée** : les deux matrices input et target. **Sortie** : une matrice contenant les meilleurs architectures trouvées.

L'algorithme de **BBO** est écrit par [1] , on l'a implémenté et l'adapter à notre problème.

## **trainASol.m**

**function [bestnet,besttr,besthsi]=trainASol(solution,input,target,n,f)**

**Entrée** : Ce script crée un réseau de neurone ayant les l'architecture "**solution**", n représente le nombre de **retrain** et f représente le dossier où le réseau est enregistré.

**Sortie** : retourne la meilleure architecture trouvée.

## **trainANet.m**

**function [bestnet,besttr,besthsi] = trainANet(net,input,target,n)**

le rôle de **trainANet** est de faire le training et de chercher les meilleures paramètres (valeur des poids) pour une architecture donnée.

## **output.m**

**function [out] = output(net,input)** Cette fonction implémente **FinalLayer** ( Paragraphe 4.4).

## **courbeLearn.m**

**function courbeLearn(sol,input,target)** courbeLearn est une procédure qui génère le graphe de validation (Paragraphe 4.6).

## **classify.m**

**function [m,rm,p,rp,BetterComportement,mb,rmb,pb,rpb] = classify(x)**

Classify est la fonction utilisée dans l'application, elle calcule le output pour un étudiant donné et cherche le meilleur comportement ; où :

- m : la classe pour l'examen de mathématique.
- rm : rate\_mate est la probabilité de certitude.
- m : la classe pour l'examen de langue portugaise.
- rm : rate\_mate est la probabilité de certitude.
- BetterComportement : un vecteur contenant les différentes modifications afin d'améliorer le résultat (voir paragraphe 5).
- bm (better m) : la classe pour l'examen de mathématique après amélioration.
- brm : rate\_mate est la probabilité de certitude.
- bm : la classe pour l'examen de langue portugaise après amélioration.
- brm : rate\_mate est la probabilité de certitude.

## C Captures d'écran

**Educational System** ADMINISTRATION

**SOCIAL** **EDUCATIONAL**

School: GABRIEL PEREIRA MOUSINHO DA SILVEIRA

Age: ☒ Male ☐ Female

Sex: ☐ Male ☒ Female

Address: Urban

Family Size: ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10

Parent's Status: Together

Mother's education: none

Father's education: none

Mother's job: teacher

Father's job: teacher

Reason to choose this school: Home

Guardian: Mother

**Predictions**

CLASSES

PREDICT

Math Score: \_\_\_\_\_

Portugais Score: \_\_\_\_\_

**AMELIORATIONS**

Math Score: \_\_\_\_\_

Portugais Score: \_\_\_\_\_

School Sup: \_\_\_\_\_

GoOut: \_\_\_\_\_

Dalc: \_\_\_\_\_

Walc: \_\_\_\_\_

Study Time: \_\_\_\_\_

(c) SOCIAL

**Educational System** ADMINISTRATION

**SOCIAL** **EDUCATIONAL**

Travel Time: ☐ 1 ☐ 2 ☒ 3

Absences: ☐ 1 ☐ 2 ☒ 3

School Sup: ☐ Yes ☒ No

Extra Classes: ☐ Yes ☒ No

Nursery: ☐ Yes ☒ No

Internet: ☐ Yes ☒ No

Famrel: ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10

GoOut: ☐ 1 ☐ 2 ☒ 3

Health: ☐ Good ☒ Bad

workday alcohol consumption: ☐ Low ☐ Medium ☐ High

weekend alcohol consumption: ☐ Low ☐ Medium ☐ High

G1: ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10

G2: ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10

**Predictions**

CLASSES

PREDICT

Math Score: \_\_\_\_\_

Portugais Score: \_\_\_\_\_

**AMELIORATIONS**

Math Score: \_\_\_\_\_

Portugais Score: \_\_\_\_\_

School Sup: \_\_\_\_\_

GoOut: \_\_\_\_\_

Dalc: \_\_\_\_\_

Walc: \_\_\_\_\_

Study Time: \_\_\_\_\_

(d) EDUCATIONAL

FIGURE 9 – IHM : écran d'accueil

L'étape suivante est le remplissage des différent champs en introduisant les information de l'étudiant en question.

**Educational System** ADMINISTRATION

**SOCIAL** **EDUCATIONAL**

Travel Time: ☐ 1 ☐ 2 ☒ 3

Absences: ☐ 1 ☐ 2 ☒ 3

School Sup: ☐ Yes ☒ No

Extra Classes: ☐ Yes ☒ No

Nursery: ☐ Yes ☒ No

Internet: ☐ Yes ☒ No

Famrel: ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10

GoOut: ☐ 1 ☐ 2 ☒ 3

Health: ☐ Good ☒ Bad

workday alcohol consumption: ☐ Low ☐ Medium ☐ High

weekend alcohol consumption: ☐ Low ☐ Medium ☐ High

G1: ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10

G2: ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10

**Predictions**

CLASSES

PREDICT

Math Score: **Muito Bom**

Portugais Score: **Excelente**

**AMELIORATIONS**

Math Score: **Insuficiente**

Portugais Score: **Excelente**

School Sup: yes=====>yes

GoOut:3=====>2

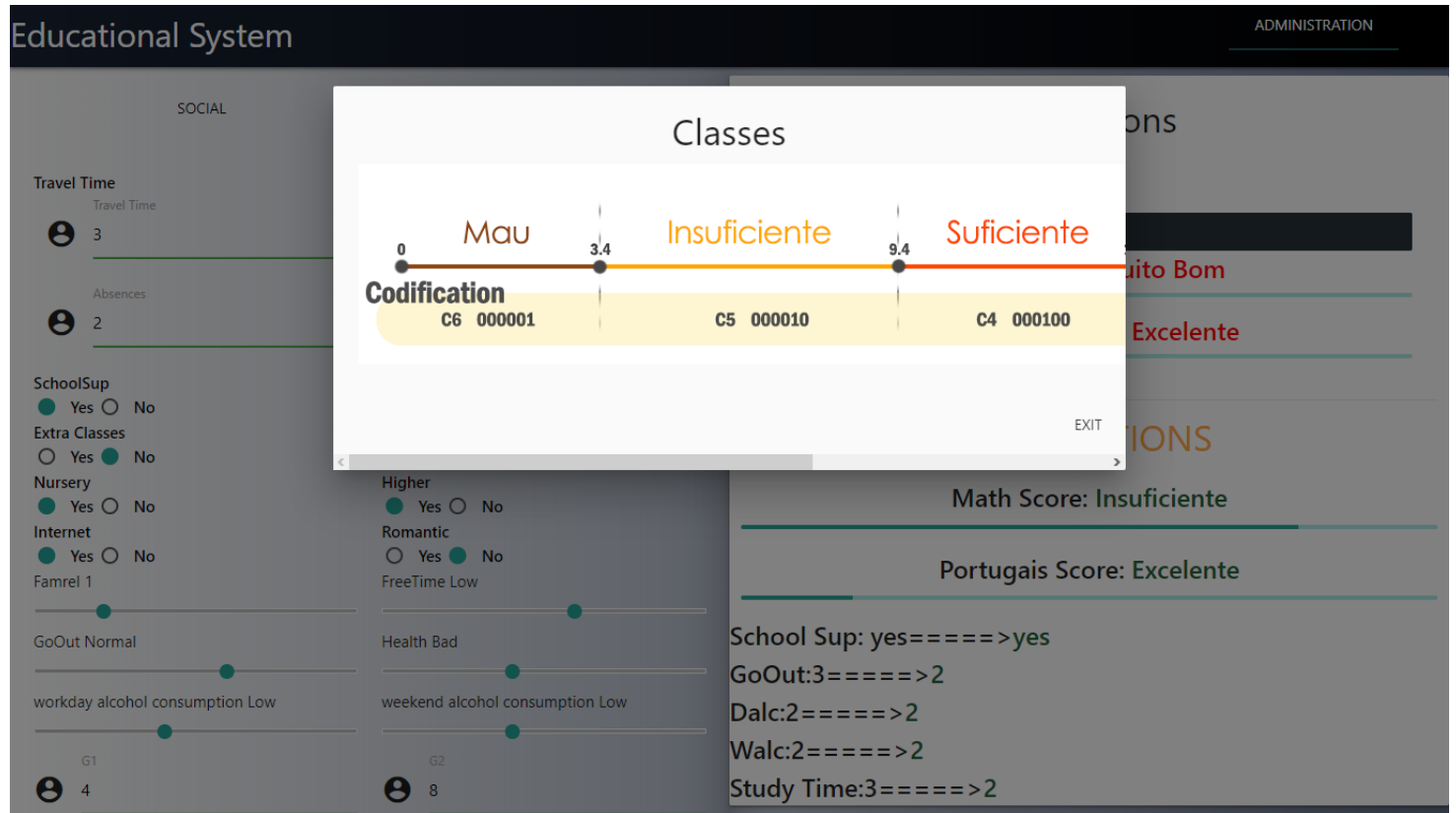
Dalc:2=====>2

Walc:2=====>2

Study Time:3=====>2

FIGURE 10 – IHM : Prédiction

Dans ce volet, on trouve les prédiction ainsi que les recommandations proposées afin d'améliorer ce résultat. Les informations concernant la notation portugaise et la note équivalente :



**Exécution** Pour exécuter l'application il suffit de lancer le fichier exécutable : Executable/**projet.exe**

## D Répartition des tâches

- Djaidja Taki Eddine Toufik : Partie Apprentissage.
- Adli Housseem : Partie Application.

# Table des matières

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Problématique</b>	<b>1</b>
<b>3</b>	<b>Données</b>	<b>1</b>
3.1	Prétraitement des données . . . . .	1
<b>4</b>	<b>Apprentissage</b>	<b>2</b>
4.1	BBO Algorithme . . . . .	2
4.2	La forme d'une architecture . . . . .	3
4.3	HSI . . . . .	3
4.4	Pseudo-Code . . . . .	3
4.5	Résultats . . . . .	4
4.6	Validation du modèle . . . . .	5
<b>5</b>	<b>Application</b>	<b>5</b>
<b>6</b>	<b>Conclusion</b>	<b>5</b>
<b>A</b>	<b>Meilleures Architectures</b>	<b>6</b>
A.1	Mathématique . . . . .	6
A.2	Langue Portugaise . . . . .	6
<b>B</b>	<b>Scripts</b>	<b>7</b>
<b>C</b>	<b>Captures d'écran</b>	<b>9</b>
<b>D</b>	<b>Répartition des tâches</b>	<b>10</b>

# Liste des sigles et acronymes

<b>IA</b>	Intelligence Artificielle
<b>BBO</b>	Biogeography-Based Optimization
<b>HSI</b>	Habitat Suitability Index
<b>PSO</b>	Particle Swarm Optimization
<b>RN</b>	Réseau de Neurones
<b>Math</b>	Mathématique
<b>Por</b>	Langue Portugaise
<b>SchoolSup</b>	extra educational support
<b>DALC</b>	workday alcohol consumption
<b>WALC</b>	weekend alcohol consumption

## Références

- [1] Biogeography-based optimization. [https://en.wikipedia.org/wiki/Biogeography-based\\_optimization](https://en.wikipedia.org/wiki/Biogeography-based_optimization).
- [2] Introduction à l'intelligence artificielle. Pierre De Loor – ENIB.
- [3] Ministério da Ciência, Inovação e Ensino Superior. Decreto-lei nº 42/2005 de 22 de fevereiro.
- [4] P. Cortez and A. Silva. Using data mining to predict secondary school student performance. <http://www3.dsi.uminho.pt/pcortez/student.pdf>, 2008.
- [5] Ali Nazari and Amin Hadidi. Biogeography based optimization algorithm for economic load dispatch of power system. *AMERICAN JOURNAL OF ADVANCED SCIENTIFIC RESEARCH*, 1 :99–105, 09 2012.