Report CheckYourSmile

Tutorial Project

Viet Minh Thong Le, Dinh Triem Phan

May 2020

Tutors: Nadia Yassine-Diab, Sebastien Dejean

INSA Toulouse
Department GMM

Contents

Abstract

Check Your Smile (CYS) is a game-based tool to help students learn foreign languages. Currently, there are no results on the effectiveness of this tool on the progression of students' learning. In this project, we applied both linear and non-linear models to verify if this tool enables student progress in their vocabulary learning. We divided the data into two data sets (semester 3 and semester 4). In semester 3, ANOVA and ANCOVA showed that CYS was effective for students not in engineering classes (Engineering class as CMI). Furthermore, the decision tree, a non-linear model, reported three factors the most affecting (Pratical Work as TP, CMI, CYS) with a cross-validation error higher than that in the ANOVA/ANCOVA model . In general, CYS is effective for those not in engineering classes (non-CMI) and practicing in English. In semester 4, the models reported an uncertain result because of the missing values and variables.

1 Introduction

CheckYourSmile (CYS) is a web platform project for learning specialty vocabulary in foreign languages (e.g., IT English: networks / databases) through a set of "serious" games, founded by Dr. Nadia Yassine-Diab. The objective is to provide a complement to face-to-face language courses in higher education. Despite the importance of specialty vocabulary in integrating students into their professional careers, only a few hours per week are devoted to this teaching. Note that one of the innovations of CYS is to offer a collaborative system to propose and validate lexical entries: thus, everyone participates in constructing knowledge (cf. crowdsourcing).

The first CYS prototype was released in 2014 (currently online at www.checkyoursmile.fr). Initiative of Excellence (IDEX) funding from the University of Toulouse in 2016 made it possible to hire several developers, trainees and post-docs to develop the site; a new version was released in January 2017, including new games and new features. The platform is free and licensed under the Creative Commons license. The previous prototypes have already served us to demonstrate the concept and to propose a stable and functional version of the site which now includes six games that work on the four skills of learning a language (French and English as foreign languages).

Our objective is to obtain indicators on the plus-value of Check Your Smile in a university context and on the combinations of variables to obtain the best results and to improve the effects of the tool.

2 Materials

The subject aims to study a database acquired in three academic years (2016-7, 2017-8 and 2018-9). It contains the evaluation results of students of different UPS courses as well as details on their courses and the particularities of the language teaching they received (English or French TP, CMI engineering courses, use of CYS).

Semestre	Filière	snapshot.1	snapshot.2	CYS.S3	TP.S3	CMI
S3 2017-18: 38	EEA:181	Min. : 1.000	Min. : 1.75	non:120	FR:132	non:150
S3 2018-19:143		1st Qu.: 4.500	1st Qu.: 7.00	oui: 61	GB: 49	oui: 31
		Median : 6.750	Median : 8.75			
		Mean : 6.442	Mean : 8.82			
		3rd Qu.: 8.000	3rd Qu.:10.50			
		Max. :13.000	Max. :15.75			

Table 1: Summary of data semester 3

Table 1 shows that, in semester 3, among 181 students, 38 were in 2017-2018 and 143 were in 2018-2019, all 181 were in the sector EEA, 120 used the CYS tool while 41 did not use it, 132 had practical work in French while 49 used English, 160 were in CMI while 31 were not.

Semestre	Filiere	snapshot.1	snapshot.2	CYS.S4	TP.S4	CMI
S4 2017-18:54	BIOMIP :18	Min. : 3.500	Min. : 7.50	non:21	FR :20	non:62
S4 2018-9 :18	EEA :13	1st Qu.: 8.500	1st Qu.:12.00	oui:51	GB :11	oui:10
	Medecine:41	Median : 9.500	Median :16.00		non:41	
		Mean : 9.596	Mean :15.72			
		3rd Qu.:11.000	3rd Qu.:19.12			
		Max. :14.500	Max. :24.00			

Table 2: Summary of data semester 4

Table 2 shows that medical students do not have TP and they are not in engineering classes(CMI). Thus, we separated data into 2 groups: medical and non-medical.

Semestre	Filiere	snapshot.1	snapshot.2	CYS.S4	TP.S4	CMI
S4 2017-18:41	Medecine:41	Min. : 6.00	Min. :11.00	non:20	FR : 0	non:41
S4 2018-9 : 0		1st Qu.: 9.00	1st Qu.:16.00	oui:21	GB : 0	oui: 0
		Median :10.00	Median :19.00		non:41	
		Mean :10.11	Mean :18.34			
		3rd Qu.:11.00	3rd Qu.:20.50			
		Max. :14.50	Max. :24.00			

Table 3: Summary of data semester 4 for medical students

Table 3 shows that among 41 medical students in semester 4 , all 41 were in 2017-2018, 21 used the CYS tool while 20 did not use it, none had TP, none was in CMI.

Table 4 shows that among 31 non-medical students in semester 4 , 13 were in 2017-2018 while 18 were in 2018-2019, 18 were in BIOMIP while 13 were in EEA, 31 used the CYS tool while one did not use it, 20 had TP in French while 11 had TP in English, 10 were in CMI (all in EEA) while 21 were not.

Semestre	snapshot.1	snapshot.2	CYS.S4	TP.S4	CMI	filiere
S4 2017-18:13	Min. : 3.500	Min. : 7.50	non: 1	FR:20	non:21	BIOMIP :18
S4 2018-9 :18	1st Qu.: 8.290	1st Qu.:10.50	oui:30	GB:11	oui:10	EEA_CMI :10
	Median : 8.660	Median :12.00				EEA_non_CMI: 3
	Mean : 8.917	Mean :12.26				
	3rd Qu.:10.250	3rd Qu.:14.25				
	Max. :13.500	Max. :18.00				

Table 4: Summary of data semester 4 for non-medical students

3 Methods

We applied statistical tests on data from both semesters (3 and 4 respectively).

Firstly, descriptive statistics determined the most influential factors among the considered variables. Charts such as boxplot illustrated which variables are more important than others. Furthermore, "interaction.plot" showed how variables interacted mutually.

Secondly, linear regression (2) led to a linear formula to study how multiple variables affected on the student progression simultaneously, including their mutual interactions. We used AIC/BIC as a criterion to choose the model best fit the data. Model ANOVA indicated the effect of 3 qualitative variables on the student progression as a term of difference between 2 Snapshots and as a term of ratio. On the other hand, model ANCOVA showed the effect of 3 qualitative variables and Snapshot1 on Snapshot2. By comparing the R^2 values, we maintained the model whose R^2 value was higher.

Thirdly, decision trees (1)) were used to demonstrate how multiple variables affected the progressions of students.

According to cross-validated predictions, we kept the model which possessed the minimal complexity parameters. We applied the same method on linear regression, studying two cases: the effect of 3 qualitative variables on the student progression as a term of difference between 2 Snapshots; the effect of 3 qualitative variables and Snapshot1 on Snapshot2.

We calculated the cross-validation errors of these models to achieve the proper models.

4 Results

4.1 Descriptive statistics

4.1.1 Semester 3

Figures 1 and 2 suggest that a non-CMI student has a higher Snapshot 1 score when they used the CYS tool. Among non-CMI students, those who used CYS have higher Snapshot 2 scores than those who did not use it.

We observe that the CYS tool has a positive effect on non-CMI students.

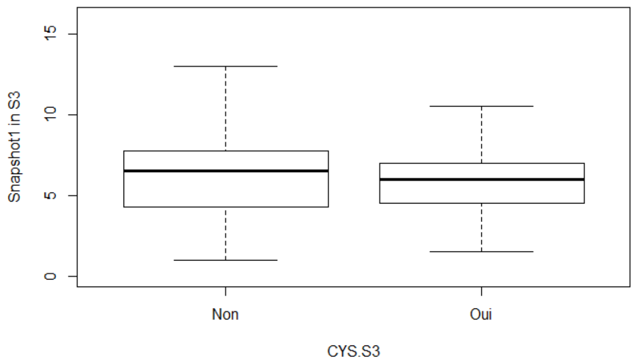

Figure 1: Snapshot1 of non-CMIs

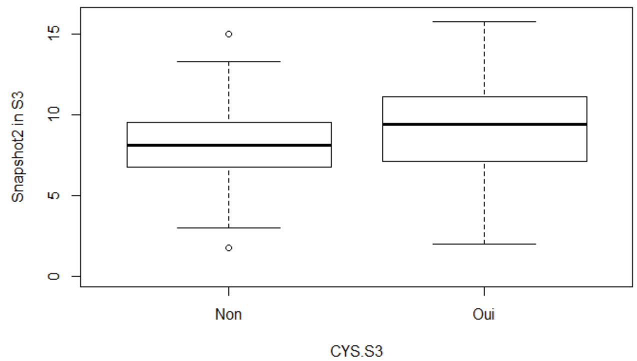

Figure 2: Snapshot2 of non-CMIs

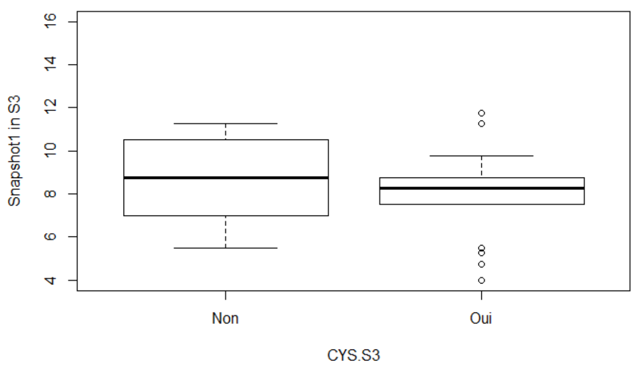

Figure 3: Snapshot1 of CMIs

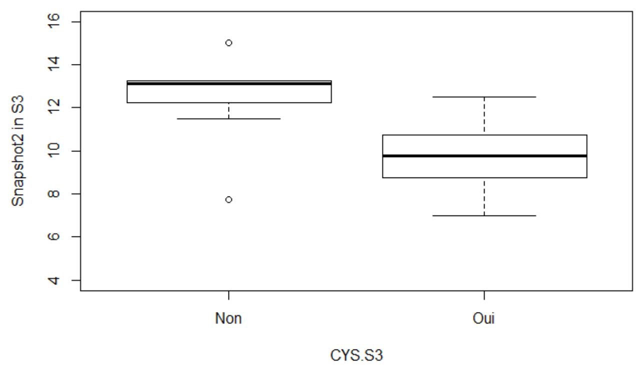

Figure 4: Snapshot2 of CMIs

Figures 3 and 4 show that a CMI student has a lower Snapshot 1 score when they did not use the CYS tool. Among non-CMI students, those who used CYS have lower Snapshot 2 scores than those who did not use it.

We study for this case the result evolution (Snapshot2-Snapshot1).

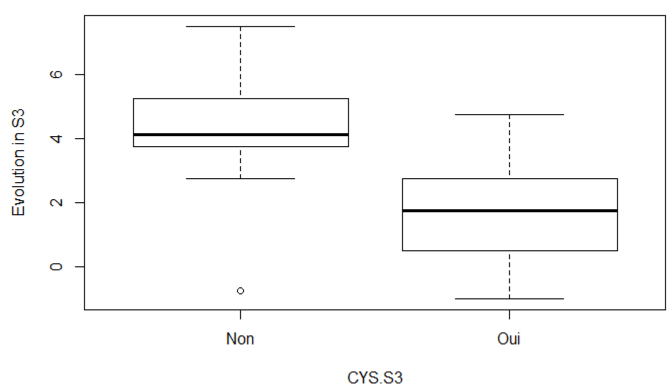

Figure 5: Evolution of CMIs

Figure 5 illustrates that CMI students progressed better when they did not use CYS.

4.1.2 Semester 4

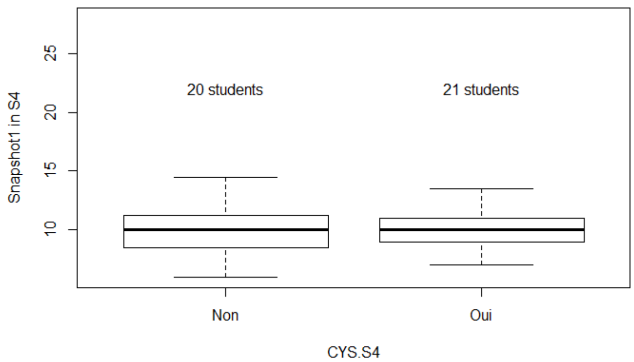

Figure 6: Snapshot1 of medical students

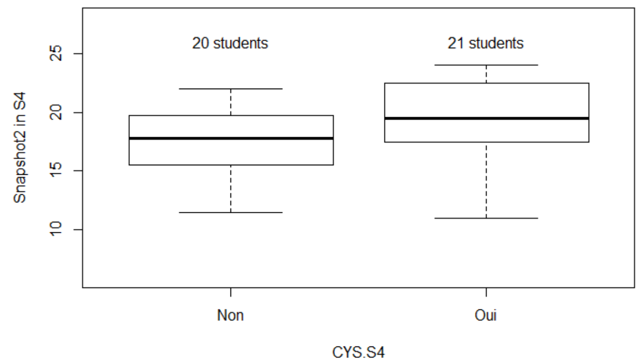

Figure 7: Snapshot2 of medical students

4.1.2.1 Medical

Figures 6 and 7 illustrate that there is no significant difference between medical students who used and did not use the CYS tool. However, among medical students, those who used CYS have slightly higher Snapshot 2 scores than those who did not use it.

We observe that the CYS tool has a positive effect on medical students.

We visualise the data of medical students based on the use of CYS (Figure 8) and show that the data is scattered.

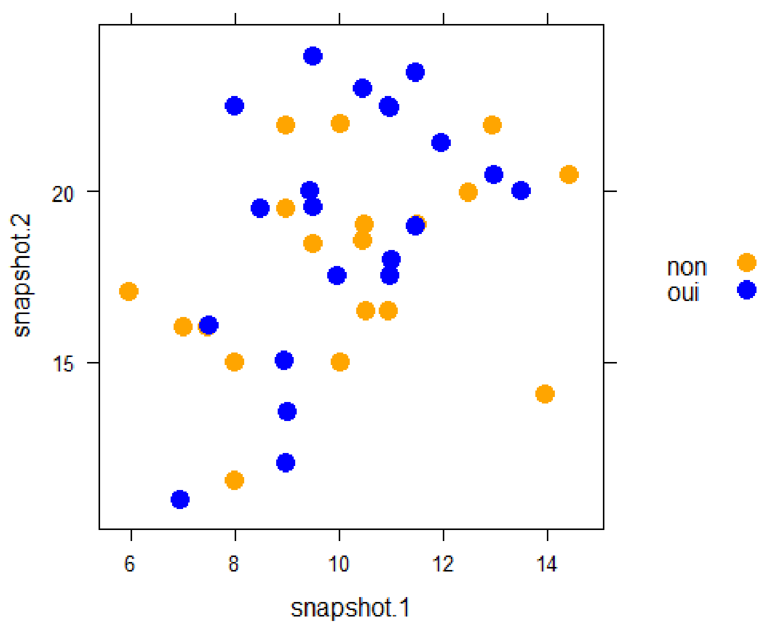

Figure 8: Scatter plot according to the use of CYS of medical students

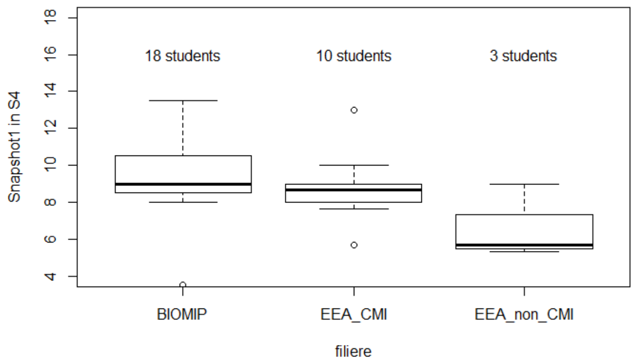

Figure 9: Snapshot 1 of non-medical students in Semester 4

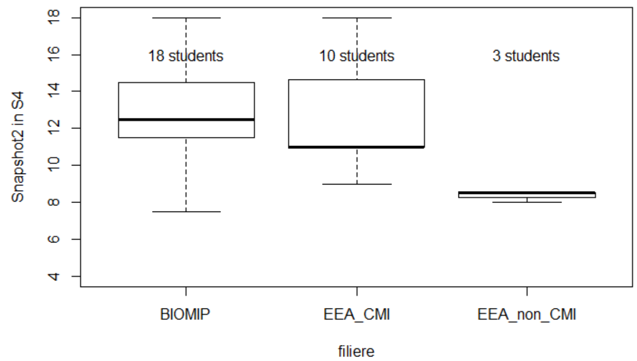

Figure 10: Snapshot 2 of non-medical students in Semester 4

4.1.2.2 Non-medical

Figures 9 and 10 show that a non-CMI student in EEA has the lowest Snapshot 1 score while BIOMIP students and CMI students in EEA have the same Snapshot 1 scores. A non-CMI student in EEA has the lowest Snapshot 2 score while a BIOMIP student has the greatest score.

We visualise the data of non-medical students as a function of department (Figure 11).

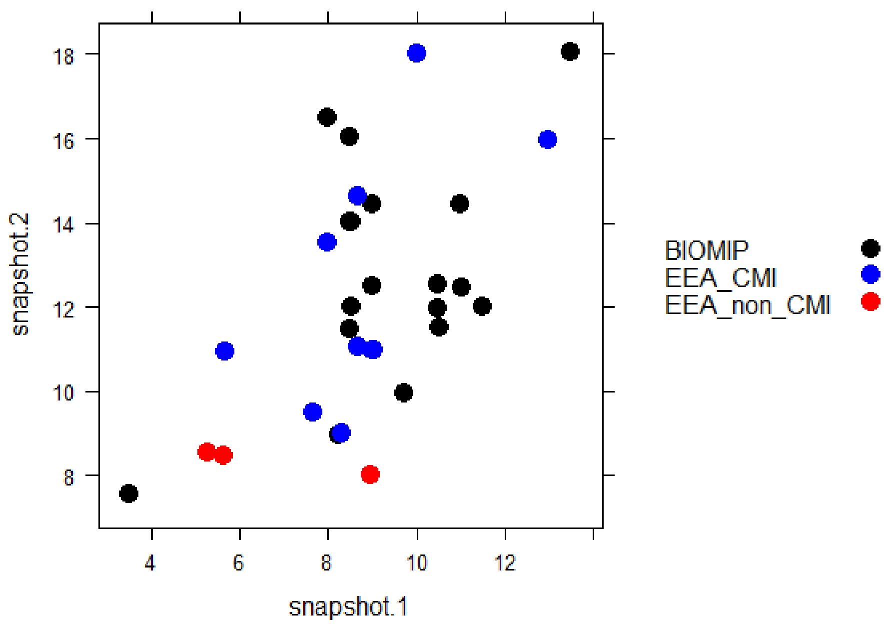

Figure 11: Scatter plot for non-medical as a function of department

We visualise the data of non-medical students based on the language of TP(Figure 12)

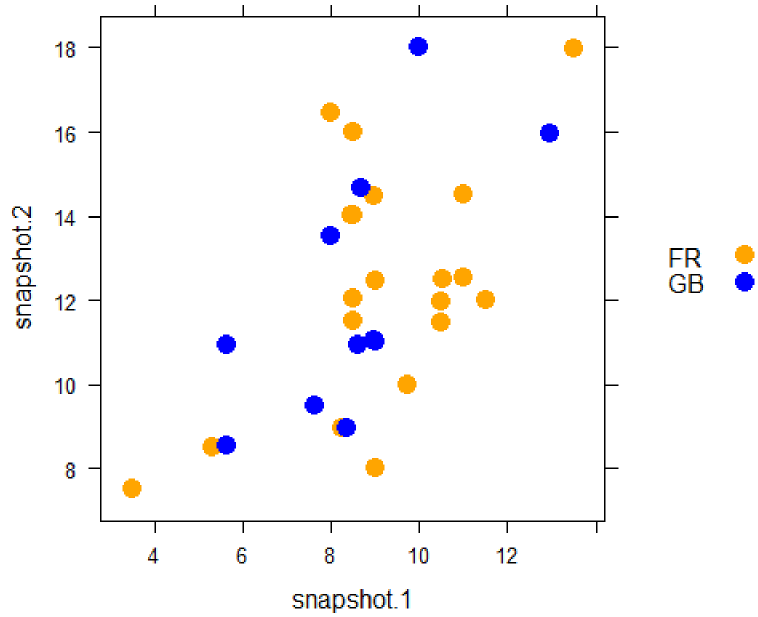

Figure 12: Scatter plot for non-medical students based on language of TP

Cross table			
Language	BIOMIP	EEA_CMI	EEA_non_CMI
FR	18	0	2
GB	0	10	1

Table 5: Number of non-medical students based on department and language of TP

Table 5 reports that only 1 among 31 non-medical students did not use CYS so we could eliminate the variable CYS in the model. Figure 17 shows that BIOMIP students always have TP in French while CMI-students in EEA have TP in English. The number of non-CMI students in EEA is just three.

Thus, there is no interaction between department and TP.

4.2 Linear model

We conducted ANOVA models to study the impact of CMI, CYS and TP language factors on the evolution of results between Snapshot 1 and Snapshot 2 scores in semester 3.

4.2.1 ANOVA

We applied ANOVA on the data in semester 3 because the R^2 value in semester 4 was far lower than ANCOVA.

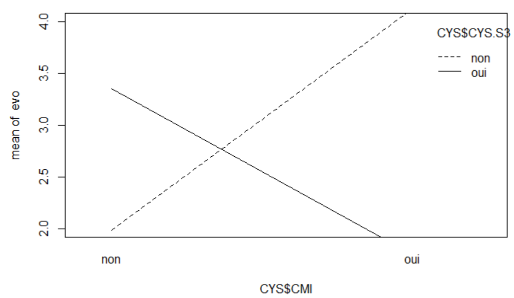

Figure 13: Interaction between CMI and CYS

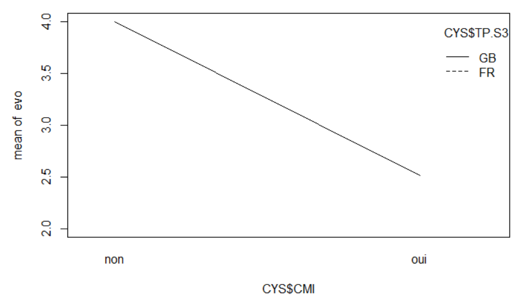

Figure 14: Interaction between CMI and TP

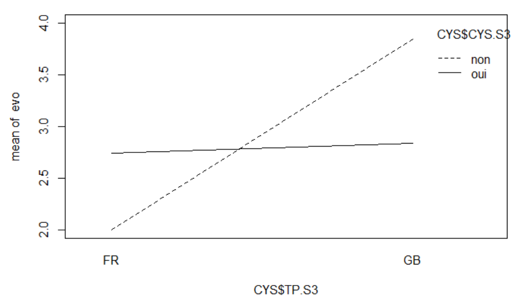

Figure 15: Interaction between CYS and TP

Figure 13, 14 and 15 show that the interaction between CYS and CMI is important. A student using CYS progresses less if he is in CMI. On the other hand, a student who does not use CYS progresses more if he is in CMI. There is no interaction between TP and CMI because no CMI student uses French for TP. The interaction between CYS and CMI exists: A student using CYS progresses more if he has TP in English. However, a student who does not use CYS progresses much more if he has TP in English.

We visualise the summary of ANOVA model (case difference=Snapshot2-Snapshot1)

```
Call:
lm(formula = dif_snap ~ CYS$CYS.S3 + CYS$CMI + CYS$CYS.S3:CYS$CMI,
    data = CYS)

Residuals:
   Min     1Q Median     3Q    Max
-4.875 -1.600  0.000  1.264  5.400

Coefficients:
```

```
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)                 1.9864     0.2068   9.606  < 2e-16 ***
CYS$CYS.S3oui               1.3636     0.4005   3.405 0.000818 ***
CYS$CMIoui                  2.1386     0.7164   2.985 0.003233 **
CYS$CYS.S3oui:CYS$CMIoui  -3.7386     0.9245  -4.044 7.84e-05 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 2.169 on 177 degrees of freedom
Multiple R-squared:  0.1009,Adjusted R-squared:  0.08569
F-statistic: 6.623 on 3 and 177 DF,  p-value: 0.000289
```

We visualise the summary of ANOVA model (case ratio=Snapshot2/Snapshot1)

```
Call:
lm(formula = ratio_snap ~ CYS$CMI, data = CYS)

Residuals:
    Min      1Q  Median      3Q     Max
-0.8666 -0.4505 -0.1435  0.1981  3.4334

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.56657    0.05769  27.153   <2e-16 ***
CYS$CMIoui  -0.20431    0.13941  -1.466    0.145
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.7066 on 179 degrees of freedom
Multiple R-squared:  0.01186,Adjusted R-squared:  0.006336
F-statistic: 2.148 on 1 and 179 DF,  p-value: 0.1445
```

By applying the BIC criterion on the complete model, we obtain the summary of ANOVA models. Under "modBIC2" model, we find no impact of the language of TP. When we consider the difference between 2 snapshots we observe the modBIC1 model where this difference depends on the use of CYS, the presence in engineering classes (CMI) and an interaction term between these two. When we consider the ratio between 2 snapshots, we observe the modBIC2 model where this ratio depends only on being in CMI.

However, the R-adjusted values were too small (less than 0.1) so we decided to apply the ANCOVA model.

4.2.2 ANCOVA

4.2.2.1 Semester 3

```
Call:
lm(formula = CYS$snapshot.2 ~ CYS$snapshot.1 + CYS$TP.S3 + CYS$CYS.S3 +
```

```
    CYS$CMI + CYS$CYS.S3:CYS$CMI, data = CYS)

Residuals:
    Min      1Q  Median      3Q     Max
-4.8965 -1.2725 -0.0205  1.1728  5.8232

Coefficients:
                         Estimate Std. Error t value Pr(>|t|)
(Intercept)               4.62384    0.44064  10.493  < 2e-16 ***
CYS$snapshot.1            0.56912    0.06511   8.740 1.85e-15 ***
CYS$TP.S3GB               1.90361    0.59802   3.183  0.00172 **
CYS$CYS.S3oui             0.43506    0.43744   0.995  0.32132
CYS$CMIoui                1.28154    0.86629   1.479  0.14084
CYS$CYS.S3oui:CYS$CMIoui -3.02652    0.85983  -3.520  0.00055 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 1.932 on 175 degrees of freedom
Multiple R-squared:  0.4764,Adjusted R-squared:  0.4615
F-statistic: 31.85 on 5 and 175 DF,  p-value: < 2.2e-16
```

We found the best model (*modbest*):

$$snapshot.2 \sim snapshot.1 + TP.S3 + CYS.S3 + CMI + CYS.S3 : CMI$$

We estimated the result of Snapshot2 by the model:

$$(modbest) : Snapshot2_{ijkl} = \mu + \alpha Snapshot1_{ijkl} + \beta_i + \gamma_j + \theta_k + \delta_{jk} + \varepsilon_{ijkl},\ \forall i = 1,\ 2,\ \forall j = 1, 2, \forall k = 1, 2$$

where:

i, j, k are modality indices of the qualitative variables TP.S3, CYS.S3 and CMI, respectively (1 for the answer *Non* and 2 for the answer *Oui* (*No* and *Yes*, respectively, in English), in the case of TP 1 for *FR* and 2 for *GB*).

The index *ijkl* is to indicate the l-th individual having modalities *i, j, k* for TP.S3, CYS.S3 and CMI, respectively. ε_{ijkl} is errors in the estimation of the individual with the index *ijkl*.

From where:

$$\mu = 4.62$$
$$\alpha = 0.57$$
$$\beta_1 = \gamma_1 = \theta_1 = \delta_{11} = \delta_{12} = \delta_{21} = 0$$
$$\beta_2 = 1.90$$
$$\gamma_2 = 0.44$$
$$\delta_{22} = -3.03$$

We decided to model the score of Snapshot2 according to Snapshot1, the use of the CYS tool, the language of TP and whether the student is in CMI.

Thus, under the ANCOVA model, we found that the three qualitative factors have an impact on the Snapshot2 result. However, Snaphot1 has a big effect on the result of Snapshot2. There is also an interaction term between the variable CYS.S3 and the variable CMI and the latter has a significant negative effect on Snapshot2.

The score of a CMI student using the Check Your Smile tool tends to decrease by about 2.6 points (-3.03 + 0.44) and the score of a non-CMI student using the Check Your Smile tool tends to progress around 0, 4 points (0.44)

The ANCOVA model gives us a higher R^2 value than the ANOVA model. (0.4615 >> 0.006 and 0.4615 >> 0.09)

We applied nonlinear models to try to improve the results.

4.2.2.2 Semester 4

Medical students

We visualise ANCOVA model for medical students in semester 4:

```
Call:
lm(formula = snapshot.2 ~ snapshot.1, data = med)

Residuals:
    Min      1Q  Median      3Q     Max
-6.9568 -1.9231 -0.2508  1.9046  6.0685

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  11.5448     2.6065   4.429 7.44e-05 ***
snapshot.1    0.6723     0.2532   2.656   0.0114 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 3.16 on 39 degrees of freedom
Multiple R-squared:  0.1531,Adjusted R-squared:  0.1314
F-statistic: 7.052 on 1 and 39 DF,  p-value: 0.01141
```

We found that snapshot2 only depends on snapshot1 for medical students. We saw no impact of CYS on the progression of medical students. However, the R^2-value is so small (0.1531) that we had to look for another model.

Non-medical students

We visualise ANCOVA model for non-medical students in semester 4:

```
Call:
lm(formula = snapshot.2 ~ snapshot.1, data = non_med)

Residuals:
    Min      1Q  Median      3Q     Max
-4.3288 -1.4694 -0.4584  2.0172  4.9658
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   5.1773     1.8965   2.730 0.010657 *
snapshot.1    0.7946     0.2072   3.834 0.000626 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 2.373 on 29 degrees of freedom
Multiple R-squared:  0.3364,Adjusted R-squared:  0.3135
F-statistic:  14.7 on 1 and 29 DF,  p-value: 0.0006261
```

We found that snapshot2 only depends on snapshot1 for non-medical students. We did not see the impact of CYS on the progression of non-medical students. However, the R^2-value is so small (0.3364) that we had to look for another model.

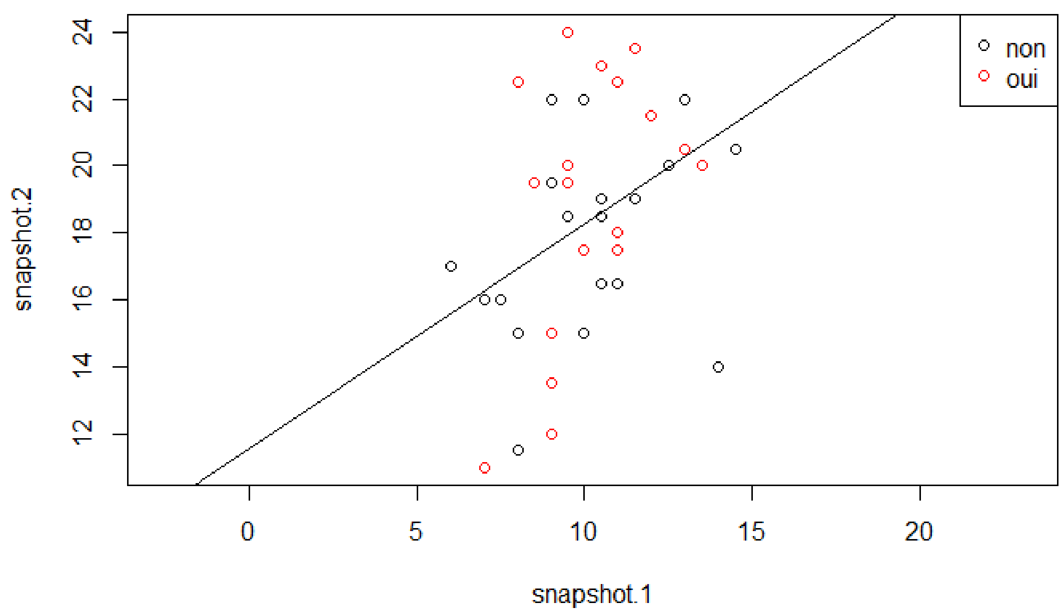

Figure 16: Scatter plot of medical students and regression line

Figure 16 shows that there is no impact of CYS tool on the student progression.

4.3 Decision tree

4.3.1 Semester 3

We applied the same method as linear regression, we studied 2 cases: the effect of 3 qualitative variables on the student progression as a term of difference between 2 Snapshots (first case); the effect of 3 qualitative variables and Snapshot1 on Snapshot2 (second case).

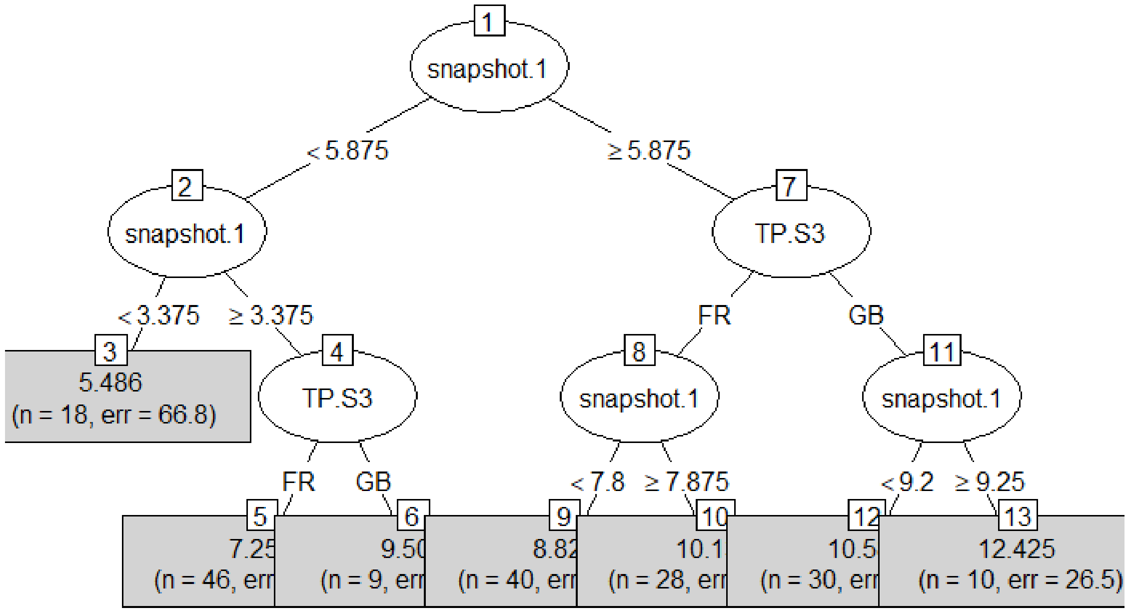

Figure 17: Regression tree of students in semester 3: Snapshot2 as a function of other variables

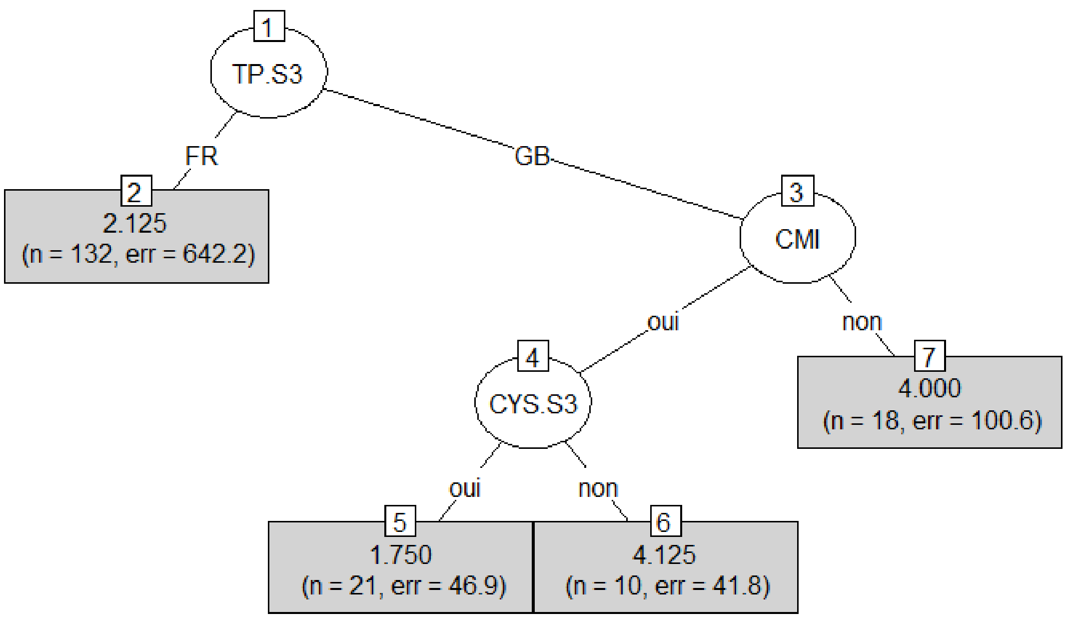

Figure 18: Regression tree of students in semester 3 :Snapshot2-Snapshot1 as a function of other variables

We tried several times to study the behaviour of cross-validation errors of these two types of the tree and we did not see the impact of CYS in the first case (Figure 17).

When we take the second, figure 18 shows that among students having TP in English, a student using the CYS tool progresses less than another not using CYS. (we conclude from sheets 5,6 and 7). Besides, among students having TP in English and using the CYS tool, a CMI student progresses less than another non-CMI. The binary regression tree concludes that the effect of the CYS tool in the student progression is not significant. The cross-validation error of the ANCOVA model is smaller than that of the decision tree. However, we find the same phenomenon for CMIs on the effect of CYS on student progression.

In terms of cross-validation error, the first tree gives us 122.96 while the second tree gives us 123.35

ANCOVA reports 100.79 as cross-validation error.

4.3.2 Semester 4

4.3.2.1 Medical students

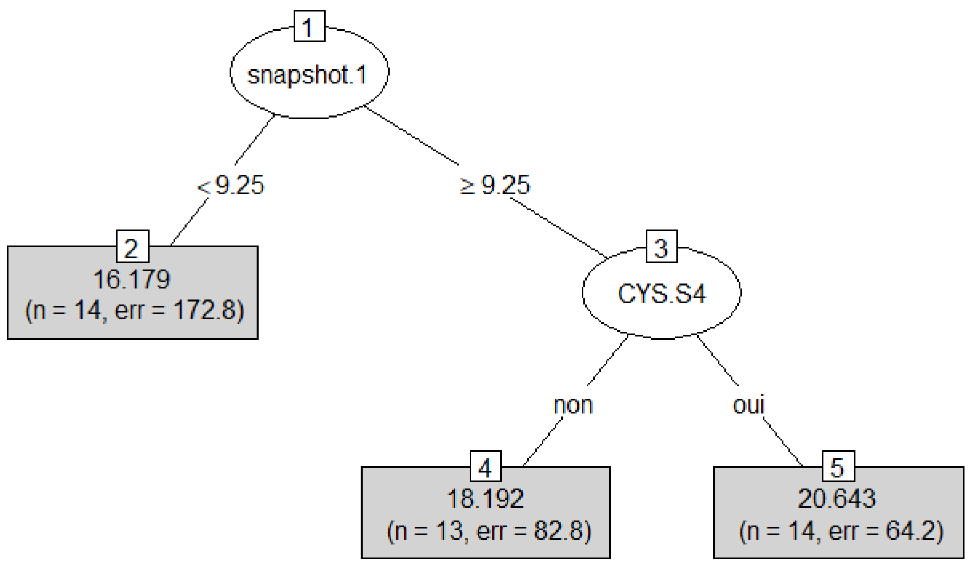

Figure 19: Regression tree of medical students

Figure 19 reports a positive impact of CYS on the scores of medical students. On average, the score of a medical student using CYS increases more than 2 points compared to that of a medical student not using CYS. In terms of cross-validation error, the decision tree of medical students gives us 159.01.

ANCOVA model reports 161.46 as cross-validation error.

4.3.2.2 Non-medical

Figure 20 that snapshot2 only depends on snapshot1 for non-medical students. Only 1 non-medical student did not use CYS so we can not conclude the impact of CYS on the student progression.

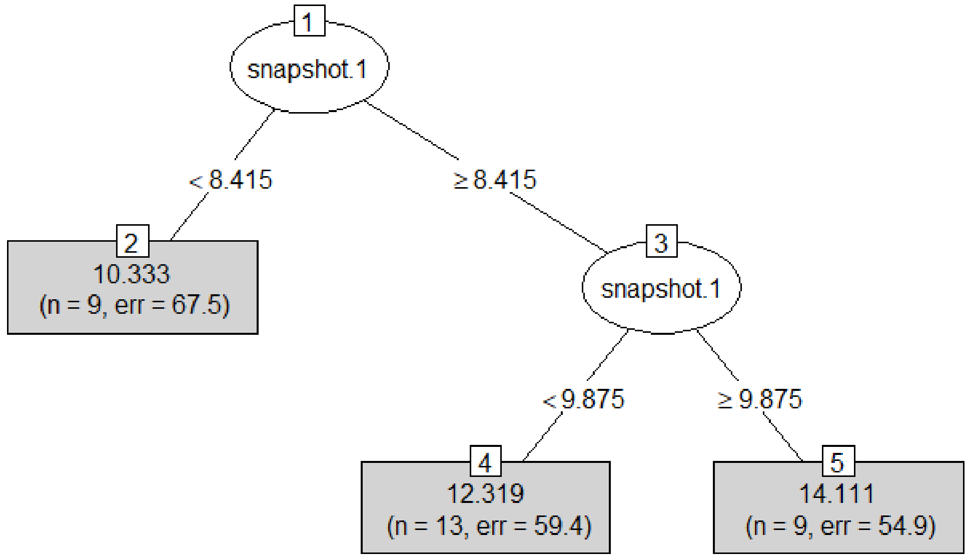

Figure 20: Regression tree of non-medical students

5 Discussion

In semester 3, both linear(ANCOVA) and non-linear(decision tree) models showed that CYS only had a positive impact on students who were not in CMI agrees with those found in the descriptive statistics.

However, the non-linear model showed us how the variable TP impacted on the result of students using CYS while the linear model did not.

In semester 4, the non-linear(decision tree) model showed that CYS had a positive impact on medical students. This result agrees with those found in the descriptive statistics.

In the case of non-medical, only 1 among them did not use CYS. Thus, we could not find the impact of CYS tool on the student progression.

In summary, modelling the score of snapshot2 from snapshot1 and other variables(TP, CYS, CMI) in semester 3 is not enough to claim the effectiveness of CYS. We have to add more variables because of small R^2-values. In semester 4, modelling the score of snapshot2 from snapshot1 and CYS does not ensure a good result.

6 Additional Analysis

As a complement, we also studied the answers of a survey on electronics students. We studied the relation among English level, electronics level and the capacity of learning in English. The levels in English were "bad", "sufficient", "good" and "very good". The levels in electronics were "bad", "sufficient", "good" and "very good". The levels of capacity of learning in English were "no", "maybe no", "yes but with difficulty" and "yes".

We considered two cases: Firstly, we modelled English level as a function of electronics level and the capacity of learning in English.

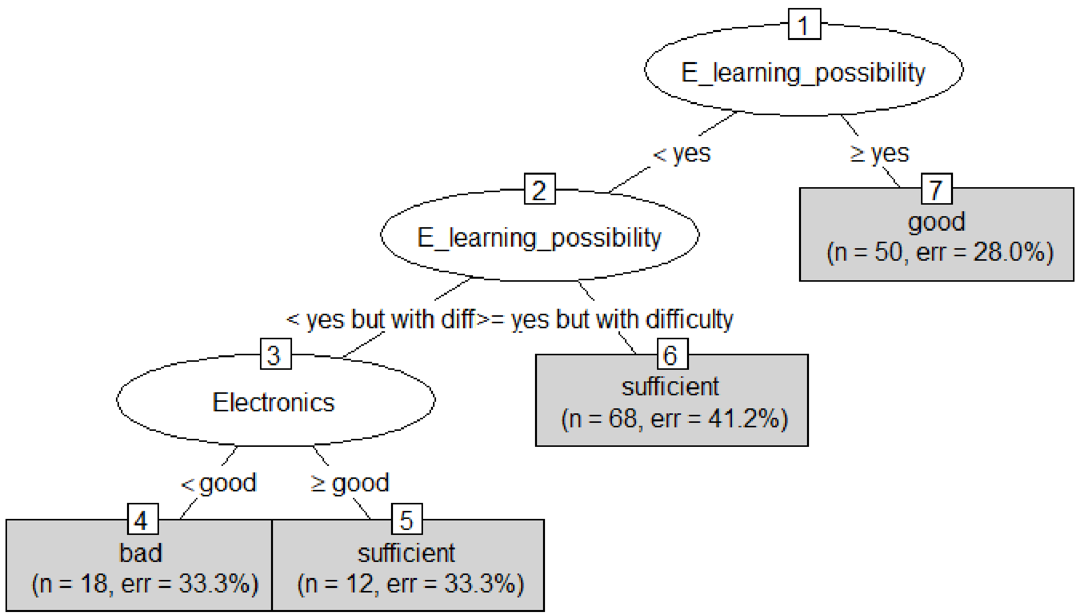

Figure 21: Regression tree of the English level of electronics students

Figure 21 shows that electronics students who think that they can study in English are good at English. Students who think that they can study in English (but with difficulty) have a sufficient English level. We looked over students thinking that they can not study in English (no/maybe no). Among them, students who are good/very good at electronics have a sufficient English level while others have a bad English level. We see that there is no relation between English level and electronics level.

This model reports 0.432 as the prediction error.

Figure 22 shows that electronics students who are good/very good at English think they can learn in English while others think that they can learn in English but with difficulty.

This model reports 0.27 as the prediction error.

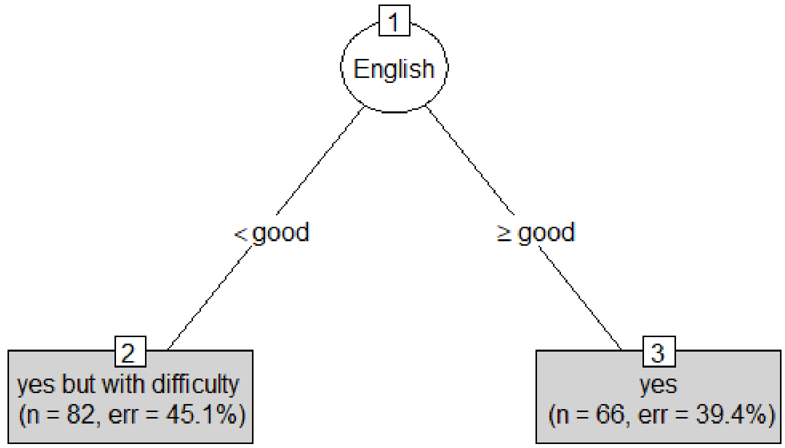

Figure 22: Regression tree of capacity of learning in English of electronics students

Statistics on games						
Index	Pendu	Audiowords	Check your motus	QCM	Check your taboo	Flashcards
Number of players	70	28	42	58	28	22
Performance	0.55	0.22	0.33	0.46	0.22	0.30

Table 6: Number of players based on the games

In table 2, *Performance* is an index of the number of players divided by the number of answers. We used it because there were some blank answers for the game "Flashcards". This table reports that "Pendu" is the game that attracts most of students.

We verified if the level of satisfaction depends on the type of game the student played and playing frequency.

Figure 23 shows that a student playing "Pendu" is satisfied with CYS (error of validation is 0.286). There are two cases for students playing other games. If they play once a week, they have no idea about satisfaction. In the case they play more frequently, if they play "QCM" the level of satisfaction is average, otherwise, they are satisfied with CYS. However, the error of validation of sheets 5, 6 and 7 is greater than 0.4 so we do not ensure the results on these sheets.

This model reports 0.385 as the prediction error.

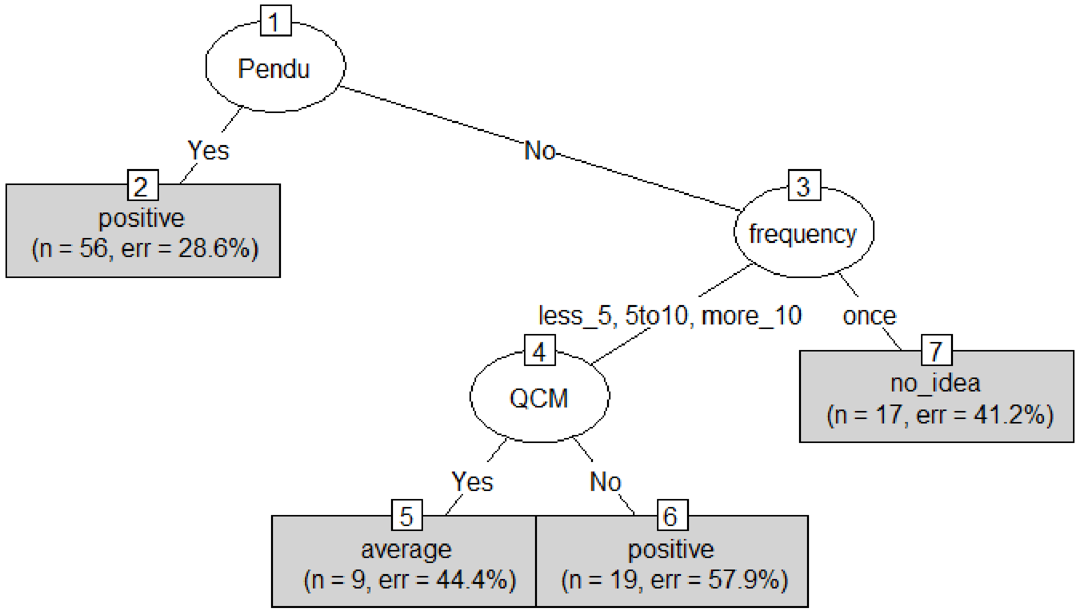

Figure 23: Regression tree of level of satisfaction

Figure 24 reports that most players spent less than four hours learning English vocabulary with CYS. There are several students who spent more than 10 hours on CYS.

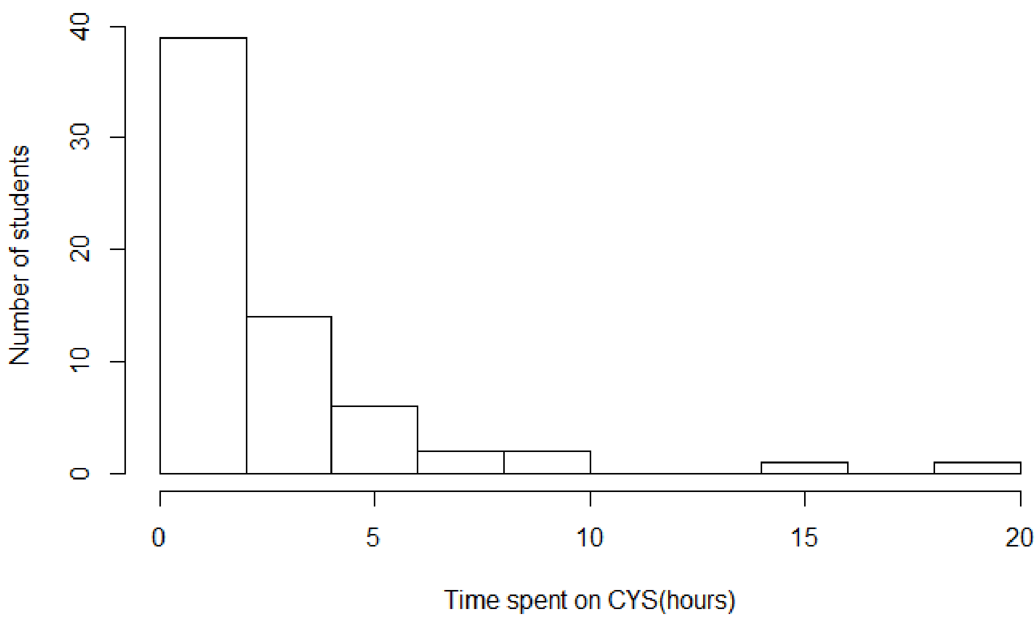

Figure 24: Histogram of times spent on CYS

Acknowledgements

We would like to express our sincere gratitude to our supervisors Dr. Nadia Yassine-Diab & Dr.Sebastien Dejean for providing their invaluable guidance, comments and suggestions throughout the course of the project. We would specially thank Dr. Sebastien Dejean for constantly motivating us to work harder and Dr. Nadia Yassine-Diab for getting us the data samples.

References

[1] R. J. Lewis, *An introduction to classification and regression tree (cart) analysis*, in Annual meeting of the society for academic emergency medicine in San Francisco, California, vol. 14, 2000.

[2] D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to linear regression analysis*, vol. 821, John Wiley & Sons, 2012.