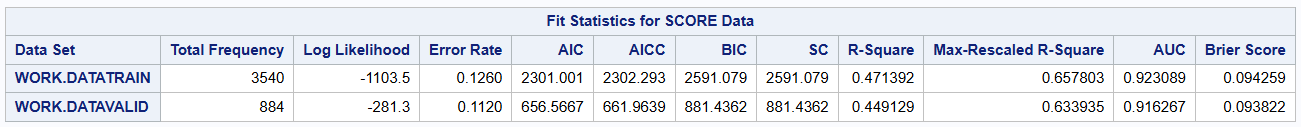
Logistic Regression Stats: Using Selection Methods (Backward, Forward and Stepwise)



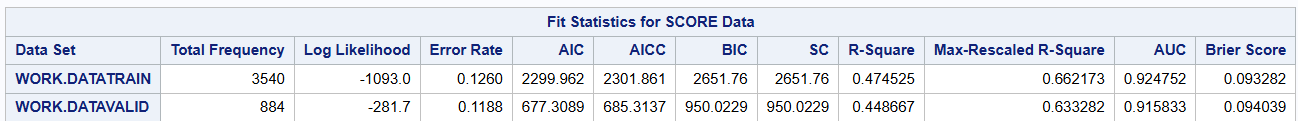
A screenshot of a computer screen

Description automatically generated A screenshot of a computer screen

Description automatically generated

backward

Forward



CART MODEL

Entropy

A screenshot of a graph

Description automatically generatedA graph of a curve

Description automatically generated

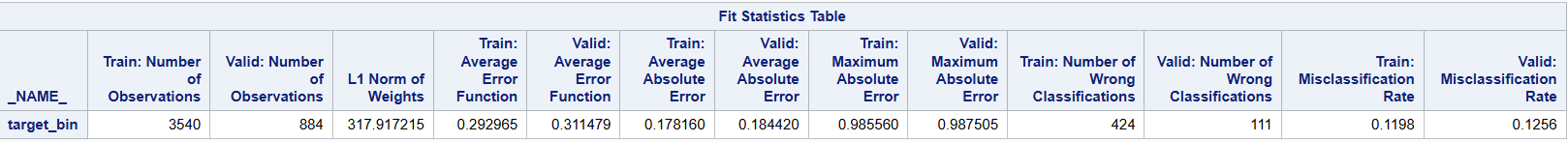
Gini

A screenshot of a computer screen

Description automatically generatedA graph of a training curve

Description automatically generated with medium confidence

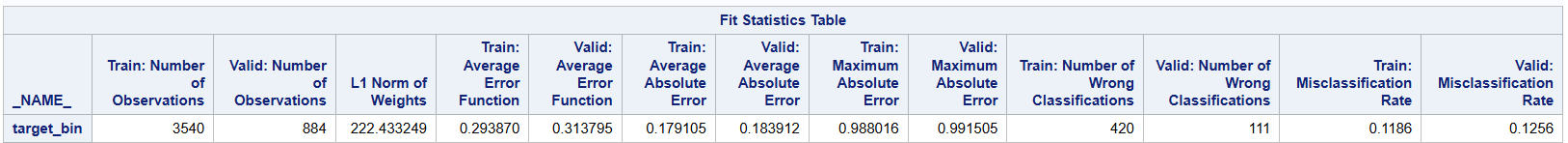
Neural Network



A screenshot of a computer

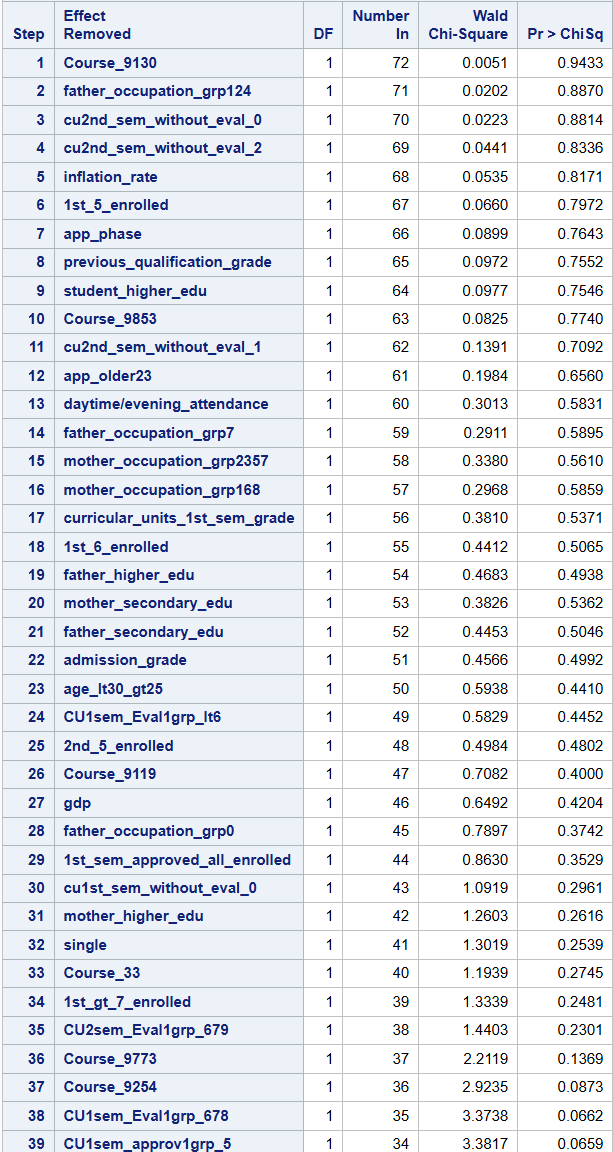
Description automatically generated A screenshot of a computer screen

Description automatically generated



Logistic Backward Selection:

* Variables Removed



Final model:

proc logistic data=data;

model target\_bin(event="1")= displaced debtor tuition\_fees\_up\_to\_date gender

scholarship\_holder age\_lt20 age\_lt25\_gt20 age\_lt30\_gt25

student\_secondary\_edu student\_higher\_edu

mother\_occupation\_grp168 mother\_occupation\_grp2357 mother\_occupation\_grp4910

father\_occupation\_grp124 father\_occupation\_grp3568 father\_occupation\_grp7 father\_occupation\_grp910

app\_phase app\_older23 app\_inter curricular\_units\_2nd\_sem\_grade

"2nd\_sem\_approved\_all\_enrolled"n

"1st\_5\_enrolled"n "1st\_6\_enrolled"n "1st\_7\_enrolled"n "1st\_gt\_7\_enrolled"n

"2nd\_5\_enrolled"n "2nd\_6\_enrolled"n

"2nd\_7\_enrolled"n "2nd\_8\_enrolled"n "2nd\_gt\_8\_enrolled"n CU1sem\_approv1grp\_5

CU1sem\_approv1grp\_lt5 CU2sem\_Eval1grp\_679 CU2sem\_Eval1grp\_8 CU2sem\_Eval1grp\_lt6 CU2sem\_approv1grp\_5 CU2sem\_approv1grp\_0 CU2sem\_approv1grp\_1234

cu1sem\_credit Course\_33 Course\_171 Course\_9003 Course\_9070 Course\_9085 Course\_9119 Course\_9130 Course\_9147\_9670 Course\_9238\_8014

Course\_9254 Course\_9500 Course\_9556 Course\_9773 Course\_9853 unemployment\_rate cu1st\_sem\_without\_eval\_1\_2 cu1st\_sem\_without\_eval\_0;

run;

Logistic Forward Selection:

* Variables Entered

A screenshot of a computer

Description automatically generated

* Interpreting variables, adding all the categories so it makes sense:

Variables to keep: curricular\_units\_2nd\_sem\_grade, tuition\_fees\_up\_to\_date, 2nd\_sem\_approved\_all\_enrolled, CU1sem\_approv1grp\_5, CU1sem\_approv1grp\_lt5, app\_phase, app\_older23, app\_inter, scholarship\_holder, Course\_33, Course\_171, Course\_9003, Course\_9070, Course\_9085, Course\_9119, Course\_9130, Course\_9147\_9670, Course\_9238\_8014, Course\_9254, Course\_9500, Course\_9556, Course\_9773, Course\_9853, gender, CU2sem\_Eval1grp\_679, CU2sem\_Eval1grp\_8, CU2sem\_Eval1grp\_lt6, CU2sem\_approv1grp\_5, CU2sem\_approv1grp\_0, CU2sem\_approv1grp\_1234, 2nd\_5\_enrolled, 2nd\_6\_enrolled, 2nd\_7\_enrolled, 2nd\_8\_enrolled, 2nd\_gt\_8\_enrolled, age\_lt20, age\_lt25\_gt20, age\_lt30\_gt25, father\_occupation\_grp124, father\_occupation\_grp3568, father\_occupation\_grp7, father\_occupation\_grp910, unemployment\_rate, displaced, debtor, mother\_occupation\_grp168 mother\_occupation\_grp2357 mother\_occupation\_grp4910, cu1st\_sem\_without\_eval\_1\_2 cu1st\_sem\_without\_eval\_0;

Final Model:

proc logistic data=data;

model target\_bin(event="1")=curricular\_units\_2nd\_sem\_grade tuition\_fees\_up\_to\_date "2nd\_sem\_approved\_all\_enrolled"n CU1sem\_approv1grp\_5

CU1sem\_approv1grp\_lt5 app\_phase app\_older23 app\_inter scholarship\_holder Course\_33 Course\_171 Course\_9003 Course\_9070 Course\_9085

Course\_9119 Course\_9130 Course\_9147\_9670 Course\_9238\_8014 Course\_9254 Course\_9500 Course\_9556 Course\_9773 Course\_9853 gender

CU2sem\_Eval1grp\_679 CU2sem\_Eval1grp\_8 CU2sem\_Eval1grp\_lt6 CU2sem\_approv1grp\_5 CU2sem\_approv1grp\_0 CU2sem\_approv1grp\_1234 "2nd\_5\_enrolled"n "2nd\_6\_enrolled"n

"2nd\_7\_enrolled"n "2nd\_8\_enrolled"n "2nd\_gt\_8\_enrolled"n age\_lt20 age\_lt25\_gt20 age\_lt30\_gt25 father\_occupation\_grp124

father\_occupation\_grp3568 father\_occupation\_grp7 father\_occupation\_grp910 unemployment\_rate displaced debtor cu1st\_sem\_without\_eval\_1\_2 cu1st\_sem\_without\_eval\_0;

run;

Logistic Stepwise Selection:

A screenshot of a computer

Description automatically generated

* Interpreting variables, adding all the categories so it makes sense:

Variables to keep: tuition\_fees\_up\_to\_date, 2nd\_sem\_approved\_all\_enrolled, CU1sem\_approv1grp\_5, CU1sem\_approv1grp\_lt5, app\_phase, app\_older23, app\_inter, scholarship\_holder, Course\_33, Course\_171, Course\_9003, Course\_9070, Course\_9085, Course\_9119, Course\_9130, Course\_9147\_9670, Course\_9238\_8014, Course\_9254, Course\_9500, Course\_9556, Course\_9773, Course\_9853, gender, CU2sem\_Eval1grp\_679, CU2sem\_Eval1grp\_8, CU2sem\_Eval1grp\_lt6, CU2sem\_approv1grp\_5, CU2sem\_approv1grp\_0, CU2sem\_approv1grp\_1234, 2nd\_5\_enrolled, 2nd\_6\_enrolled, 2nd\_7\_enrolled, 2nd\_8\_enrolled, 2nd\_gt\_8\_enrolled, age\_lt20, age\_lt25\_gt20, age\_lt30\_gt25, father\_occupation\_grp124, father\_occupation\_grp3568, father\_occupation\_grp7, father\_occupation\_grp910, unemployment\_rate, displaced, debtor, mother\_occupation\_grp168 mother\_occupation\_grp2357 mother\_occupation\_grp4910, cu1st\_sem\_without\_eval\_1\_2 cu1st\_sem\_without\_eval\_0;

Final Model:

proc logistic data=data;

model target\_bin(event="1")=tuition\_fees\_up\_to\_date "2nd\_sem\_approved\_all\_enrolled"n CU1sem\_approv1grp\_5

CU1sem\_approv1grp\_lt5 app\_phase app\_older23 app\_inter scholarship\_holder Course\_33 Course\_171 Course\_9003 Course\_9070 Course\_9085

Course\_9119 Course\_9130 Course\_9147\_9670 Course\_9238\_8014 Course\_9254 Course\_9500 Course\_9556 Course\_9773 Course\_9853 gender

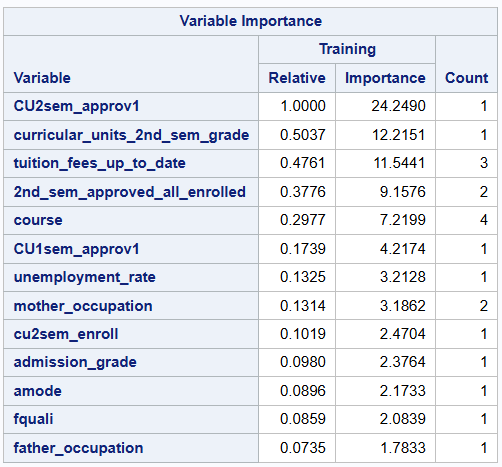
CU2sem\_Eval1grp\_679 CU2sem\_Eval1grp\_8 CU2sem\_Eval1grp\_lt6 CU2sem\_approv1grp\_5 CU2sem\_approv1grp\_0 CU2sem\_approv1grp\_1234 "2nd\_5\_enrolled"n "2nd\_6\_enrolled"n

"2nd\_7\_enrolled"n "2nd\_8\_enrolled"n "2nd\_gt\_8\_enrolled"n age\_lt20 age\_lt25\_gt20 age\_lt30\_gt25 father\_occupation\_grp124

father\_occupation\_grp3568 father\_occupation\_grp7 father\_occupation\_grp910 unemployment\_rate displaced debtor cu1st\_sem\_without\_eval\_1\_2 cu1st\_sem\_without\_eval\_0;

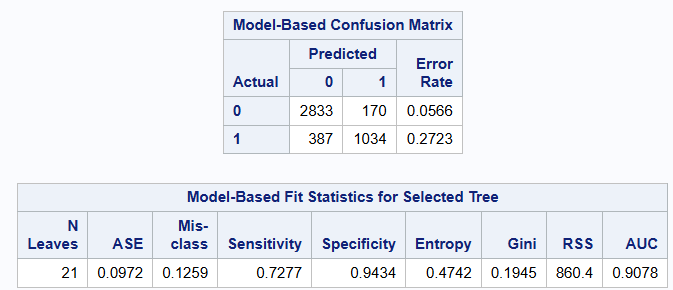
run;

CART Model Entropy:



We nee to take in consideration that if we remove a variable the output is probably going to change, I used only those variables in the table, and it gave me a completely different table after:

CU2\_approve1 curricular\_units\_2nd\_sem\_grade tuition\_fees\_up\_to\_date 2nd\_sem\_approved\_all\_enrolled course CU1sem\_approv1 unemployment\_rate mother\_occupation cu2sem\_enroll admission\_grade amode fquali father\_occupation;



Also, some measures like sensitivity and AUC decreased.

CART Model Gini:

A screenshot of a graph

Description automatically generated

With the Gini after I used only the model performed better (more nodes though)

Curricular\_units\_2nd\_sem\_grade tuition\_fees\_up\_to\_date course CU2sem\_approv1 unemployment\_rate debtor cu1sem\_approv1 cu2sem\_enroll admission\_grade father\_occupation cu2sem\_eval1

A screenshot of a computer

Description automatically generated

Here is the SAS code so you can try (Will also send a .sas file):

/\* Importing, remember to change the path\*/

proc import out= data datafile="/home/u64008123/CombineData\_1stestAttempt.csv"

dbms=csv replace;

run;

/\* Utilizing diferent models to measure the importance of our input variables \*/

/\*Backward\*/

proc logistic data=data;

model target\_bin(event="1")=previous\_qualification\_grade admission\_grade displaced debtor tuition\_fees\_up\_to\_date gender

scholarship\_holder curricular\_units\_1st\_sem\_grade curricular\_units\_2nd\_sem\_grade age\_lt20 age\_lt25\_gt20 age\_lt30\_gt25

mother\_occupation\_grp168 mother\_occupation\_grp2357 mother\_occupation\_grp4910

father\_occupation\_grp124 father\_occupation\_grp3568 father\_occupation\_grp7 father\_occupation\_grp910

single app\_phase app\_older23 app\_inter student\_secondary\_edu student\_higher\_edu mother\_secondary\_edu

mother\_higher\_edu father\_secondary\_edu father\_higher\_edu "1st\_sem\_approved\_all\_enrolled"n "2nd\_sem\_approved\_all\_enrolled"n

"1st\_5\_enrolled"n "1st\_6\_enrolled"n "1st\_7\_enrolled"n "1st\_gt\_7\_enrolled"n "2nd\_5\_enrolled"n "2nd\_6\_enrolled"n

"2nd\_7\_enrolled"n "2nd\_8\_enrolled"n "2nd\_gt\_8\_enrolled"n CU1sem\_Eval1grp\_678 CU1sem\_Eval1grp\_lt6 CU1sem\_approv1grp\_5

CU1sem\_approv1grp\_lt5 CU2sem\_Eval1grp\_679 CU2sem\_Eval1grp\_8 CU2sem\_Eval1grp\_lt6 CU2sem\_approv1grp\_5 CU2sem\_approv1grp\_0 CU2sem\_approv1grp\_1234

cu1sem\_credit cu2sem\_credit Course\_33 Course\_171 Course\_9003 Course\_9070 Course\_9085 Course\_9119 Course\_9130 Course\_9147\_9670 Course\_9238\_8014

Course\_9254 Course\_9500 Course\_9556 Course\_9773 Course\_9853 unemployment\_rate inflation\_rate gdp / selection=backward;

run;

/\* After selection method \*/

proc logistic data=data;

model target\_bin(event="1")= displaced debtor tuition\_fees\_up\_to\_date gender scholarship\_holder age\_lt20 age\_lt25\_gt20 age\_lt30\_gt25

mother\_occupation\_grp168 mother\_occupation\_grp2357 mother\_occupation\_grp4910 father\_occupation\_grp124

father\_occupation\_grp3568 father\_occupation\_grp7 father\_occupation\_grp910 app\_phase app\_older23 app\_inter

"1st\_sem\_approved\_all\_enrolled"n "2nd\_sem\_approved\_all\_enrolled"n "2nd\_5\_enrolled"n "2nd\_6\_enrolled"n

"2nd\_7\_enrolled"n "2nd\_8\_enrolled"n "2nd\_gt\_8\_enrolled"n CU1sem\_Eval1grp\_678 CU1sem\_Eval1grp\_lt6 CU1sem\_approv1grp\_5

CU1sem\_approv1grp\_lt5 CU2sem\_Eval1grp\_679 CU2sem\_Eval1grp\_8 CU2sem\_Eval1grp\_lt6 CU2sem\_approv1grp\_5 CU2sem\_approv1grp\_0 CU2sem\_approv1grp\_1234

cu1sem\_credit Course\_33 Course\_171 Course\_9003 Course\_9070 Course\_9085 Course\_9119 Course\_9130 Course\_9147\_9670 Course\_9238\_8014

Course\_9254 Course\_9500 Course\_9556 Course\_9773 Course\_9853 unemployment\_rate;

run;

/\*Foward\*/

proc logistic data=data;

model target\_bin(event="1")=previous\_qualification\_grade admission\_grade displaced debtor tuition\_fees\_up\_to\_date gender

scholarship\_holder curricular\_units\_1st\_sem\_grade curricular\_units\_2nd\_sem\_grade age\_lt20 age\_lt25\_gt20 age\_lt30\_gt25

mother\_occupation\_grp168 mother\_occupation\_grp2357 mother\_occupation\_grp4910

father\_occupation\_grp124 father\_occupation\_grp3568 father\_occupation\_grp7 father\_occupation\_grp910

single app\_phase app\_older23 app\_inter student\_secondary\_edu student\_higher\_edu mother\_secondary\_edu

mother\_higher\_edu father\_secondary\_edu father\_higher\_edu "1st\_sem\_approved\_all\_enrolled"n "2nd\_sem\_approved\_all\_enrolled"n

"1st\_5\_enrolled"n "1st\_6\_enrolled"n "1st\_7\_enrolled"n "1st\_gt\_7\_enrolled"n "2nd\_5\_enrolled"n "2nd\_6\_enrolled"n

"2nd\_7\_enrolled"n "2nd\_8\_enrolled"n "2nd\_gt\_8\_enrolled"n CU1sem\_Eval1grp\_678 CU1sem\_Eval1grp\_lt6 CU1sem\_approv1grp\_5

CU1sem\_approv1grp\_lt5 CU2sem\_Eval1grp\_679 CU2sem\_Eval1grp\_8 CU2sem\_Eval1grp\_lt6 CU2sem\_approv1grp\_5 CU2sem\_approv1grp\_0 CU2sem\_approv1grp\_1234

cu1sem\_credit cu2sem\_credit Course\_33 Course\_171 Course\_9003 Course\_9070 Course\_9085 Course\_9119 Course\_9130 Course\_9147\_9670 Course\_9238\_8014

Course\_9254 Course\_9500 Course\_9556 Course\_9773 Course\_9853 unemployment\_rate inflation\_rate gdp / selection=forward;

run;

/\* After selection method \*/

proc logistic data=data;

model target\_bin(event="1")=curricular\_units\_2nd\_sem\_grade tuition\_fees\_up\_to\_date "2nd\_sem\_approved\_all\_enrolled"n CU1sem\_approv1grp\_5

CU1sem\_approv1grp\_lt5 app\_phase app\_older23 app\_inter scholarship\_holder Course\_33 Course\_171 Course\_9003 Course\_9070 Course\_9085

Course\_9119 Course\_9130 Course\_9147\_9670 Course\_9238\_8014 Course\_9254 Course\_9500 Course\_9556 Course\_9773 Course\_9853 gender

CU2sem\_Eval1grp\_679 CU2sem\_Eval1grp\_8 CU2sem\_Eval1grp\_lt6 CU2sem\_approv1grp\_5 CU2sem\_approv1grp\_0 CU2sem\_approv1grp\_1234 "2nd\_5\_enrolled"n "2nd\_6\_enrolled"n

"2nd\_7\_enrolled"n "2nd\_8\_enrolled"n "2nd\_gt\_8\_enrolled"n age\_lt20 age\_lt25\_gt20 age\_lt30\_gt25 father\_occupation\_grp124

father\_occupation\_grp3568 father\_occupation\_grp7 father\_occupation\_grp910 unemployment\_rate displaced debtor;

run;

/\*Stepwise\*/

proc logistic data=data;

model target\_bin(event="1")=previous\_qualification\_grade admission\_grade displaced debtor tuition\_fees\_up\_to\_date gender

scholarship\_holder curricular\_units\_1st\_sem\_grade curricular\_units\_2nd\_sem\_grade age\_lt20 age\_lt25\_gt20 age\_lt30\_gt25

mother\_occupation\_grp168 mother\_occupation\_grp2357 mother\_occupation\_grp4910

father\_occupation\_grp124 father\_occupation\_grp3568 father\_occupation\_grp7 father\_occupation\_grp910

single app\_phase app\_older23 app\_inter student\_secondary\_edu student\_higher\_edu mother\_secondary\_edu

mother\_higher\_edu father\_secondary\_edu father\_higher\_edu "1st\_sem\_approved\_all\_enrolled"n "2nd\_sem\_approved\_all\_enrolled"n

"1st\_5\_enrolled"n "1st\_6\_enrolled"n "1st\_7\_enrolled"n "1st\_gt\_7\_enrolled"n "2nd\_5\_enrolled"n "2nd\_6\_enrolled"n

"2nd\_7\_enrolled"n "2nd\_8\_enrolled"n "2nd\_gt\_8\_enrolled"n CU1sem\_Eval1grp\_678 CU1sem\_Eval1grp\_lt6 CU1sem\_approv1grp\_5

CU1sem\_approv1grp\_lt5 CU2sem\_Eval1grp\_679 CU2sem\_Eval1grp\_8 CU2sem\_Eval1grp\_lt6 CU2sem\_approv1grp\_5 CU2sem\_approv1grp\_0 CU2sem\_approv1grp\_1234

cu1sem\_credit cu2sem\_credit Course\_33 Course\_171 Course\_9003 Course\_9070 Course\_9085 Course\_9119 Course\_9130 Course\_9147\_9670 Course\_9238\_8014

Course\_9254 Course\_9500 Course\_9556 Course\_9773 Course\_9853 unemployment\_rate inflation\_rate gdp / selection=stepwise;

run;

/\* After selection method \*/

proc logistic data=data;

model target\_bin(event="1")=curricular\_units\_2nd\_sem\_grade tuition\_fees\_up\_to\_date "2nd\_sem\_approved\_all\_enrolled"n CU1sem\_approv1grp\_5

CU1sem\_approv1grp\_lt5 app\_phase app\_older23 app\_inter scholarship\_holder Course\_33 Course\_171 Course\_9003 Course\_9070 Course\_9085

Course\_9119 Course\_9130 Course\_9147\_9670 Course\_9238\_8014 Course\_9254 Course\_9500 Course\_9556 Course\_9773 Course\_9853 gender

CU2sem\_Eval1grp\_679 CU2sem\_Eval1grp\_8 CU2sem\_Eval1grp\_lt6 CU2sem\_approv1grp\_5 CU2sem\_approv1grp\_0 CU2sem\_approv1grp\_1234 "2nd\_5\_enrolled"n "2nd\_6\_enrolled"n

"2nd\_7\_enrolled"n "2nd\_8\_enrolled"n "2nd\_gt\_8\_enrolled"n age\_lt20 age\_lt25\_gt20 age\_lt30\_gt25 father\_occupation\_grp124

father\_occupation\_grp3568 father\_occupation\_grp7 father\_occupation\_grp910 unemployment\_rate displaced debtor;

run;

/\*Score\*/

proc logistic data=data;

model target\_bin(event="1")=previous\_qualification\_grade admission\_grade displaced debtor tuition\_fees\_up\_to\_date gender

scholarship\_holder curricular\_units\_1st\_sem\_grade curricular\_units\_2nd\_sem\_grade age\_lt20 age\_lt25\_gt20 age\_lt30\_gt25

mother\_occupation\_grp168 mother\_occupation\_grp2357 mother\_occupation\_grp4910

father\_occupation\_grp124 father\_occupation\_grp3568 father\_occupation\_grp7 father\_occupation\_grp910

single app\_phase app\_older23 app\_inter student\_secondary\_edu student\_higher\_edu mother\_secondary\_edu

mother\_higher\_edu father\_secondary\_edu father\_higher\_edu "1st\_sem\_approved\_all\_enrolled"n "2nd\_sem\_approved\_all\_enrolled"n

"1st\_5\_enrolled"n "1st\_6\_enrolled"n "1st\_7\_enrolled"n "1st\_gt\_7\_enrolled"n "2nd\_5\_enrolled"n "2nd\_6\_enrolled"n

"2nd\_7\_enrolled"n "2nd\_8\_enrolled"n "2nd\_gt\_8\_enrolled"n CU1sem\_Eval1grp\_678 CU1sem\_Eval1grp\_lt6 CU1sem\_approv1grp\_5

CU1sem\_approv1grp\_lt5 CU2sem\_Eval1grp\_679 CU2sem\_Eval1grp\_8 CU2sem\_Eval1grp\_lt6 CU2sem\_approv1grp\_5 CU2sem\_approv1grp\_0 CU2sem\_approv1grp\_1234

cu1sem\_credit cu2sem\_credit Course\_33 Course\_171 Course\_9003 Course\_9070 Course\_9085 Course\_9119 Course\_9130 Course\_9147\_9670 Course\_9238\_8014

Course\_9254 Course\_9500 Course\_9556 Course\_9773 Course\_9853 unemployment\_rate inflation\_rate gdp/ selection=score;

run;

/\*

Disclaimer, We cannot use this selection method because we have to many variables and the way it works is that it provides

the full conbination of the variables (incrementally adding variables) with the Chi Score for each one of them so it eould be basically calculating the Score

for( i=1; i++; i<=total\_number\_variables ) {C = i Combinations total\_number\_variables; calculate Chi Score for each C}

That for our copious variables, is a lot and the SAS server will not run it.

\*/

/\* CART Entropy \*/

proc hpsplit data=data;

class target\_bin course displaced debtor tuition\_fees\_up\_to\_date gender scholarship\_holder father\_occupation mother\_occupation

single amode pquali fquali mquali "1st\_sem\_approved\_all\_enrolled"n "2nd\_sem\_approved\_all\_enrolled"n

cu1sem\_enroll cu1sem\_credit cu2sem\_credit cu2sem\_enroll age CU1sem\_Eval1 CU1sem\_approv1 CU2sem\_Eval1 CU2sem\_approv1;

model target\_bin(event="1")=previous\_qualification\_grade admission\_grade curricular\_units\_1st\_sem\_grade

curricular\_units\_2nd\_sem\_grade unemployment\_rate inflation\_rate gdp course displaced debtor

tuition\_fees\_up\_to\_date gender scholarship\_holder father\_occupation mother\_occupation

single amode pquali fquali mquali "1st\_sem\_approved\_all\_enrolled"n "2nd\_sem\_approved\_all\_enrolled"n

cu1sem\_enroll cu1sem\_credit cu2sem\_credit cu2sem\_enroll age CU1sem\_Eval1 CU1sem\_approv1 CU2sem\_Eval1 CU2sem\_approv1;

grow entropy;

prune cc;

run;

/\* Using only the variables of the importance table \*/

proc hpsplit data=data;

class target\_bin course tuition\_fees\_up\_to\_date gender scholarship\_holder father\_occupation mother\_occupation fquali

"2nd\_sem\_approved\_all\_enrolled"n cu1sem\_enroll cu2sem\_credit cu2sem\_enroll age CU1sem\_approv1 CU2sem\_approv1;

model target\_bin(event="1")=previous\_qualification\_grade admission\_grade curricular\_units\_2nd\_sem\_grade unemployment\_rate

inflation\_rate gdp course tuition\_fees\_up\_to\_date gender scholarship\_holder father\_occupation mother\_occupation fquali

"2nd\_sem\_approved\_all\_enrolled"n cu1sem\_enroll cu2sem\_credit cu2sem\_enroll age CU1sem\_approv1 CU2sem\_approv1;

grow entropy;

prune cc;

run;

/\* CART Gini \*/

proc hpsplit data=data;

class target\_bin course displaced debtor tuition\_fees\_up\_to\_date gender scholarship\_holder father\_occupation mother\_occupation

single amode pquali fquali mquali "1st\_sem\_approved\_all\_enrolled"n "2nd\_sem\_approved\_all\_enrolled"n

cu1sem\_enroll cu1sem\_credit cu2sem\_credit cu2sem\_enroll age CU1sem\_Eval1 CU1sem\_approv1 CU2sem\_Eval1 CU2sem\_approv1;

model target\_bin(event="1")=previous\_qualification\_grade admission\_grade curricular\_units\_1st\_sem\_grade

curricular\_units\_2nd\_sem\_grade unemployment\_rate inflation\_rate gdp course displaced debtor

tuition\_fees\_up\_to\_date gender scholarship\_holder father\_occupation mother\_occupation

single amode pquali fquali mquali "1st\_sem\_approved\_all\_enrolled"n "2nd\_sem\_approved\_all\_enrolled"n

cu1sem\_enroll cu1sem\_credit cu2sem\_credit cu2sem\_enroll age CU1sem\_Eval1 CU1sem\_approv1 CU2sem\_Eval1 CU2sem\_approv1;

grow gini;

prune cc;

run;

/\* Using only the variables of the importance table \*/

proc hpsplit data=data;

class target\_bin course tuition\_fees\_up\_to\_date father\_occupation mother\_occupation amode "1st\_sem\_approved\_all\_enrolled"n

"2nd\_sem\_approved\_all\_enrolled"n cu2sem\_enroll age CU1sem\_approv1 CU2sem\_approv1;

model target\_bin(event="1")=previous\_qualification\_grade admission\_grade curricular\_units\_2nd\_sem\_grade unemployment\_rate gdp course

tuition\_fees\_up\_to\_date father\_occupation mother\_occupation amode "1st\_sem\_approved\_all\_enrolled"n "2nd\_sem\_approved\_all\_enrolled"n

cu2sem\_enroll age CU1sem\_approv1 CU2sem\_approv1;

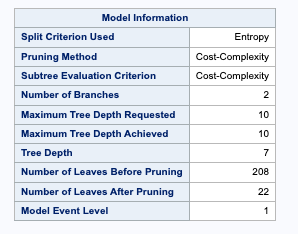
grow gini;

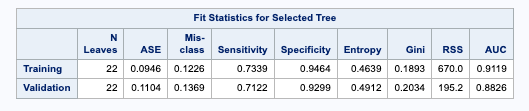
prune cc;

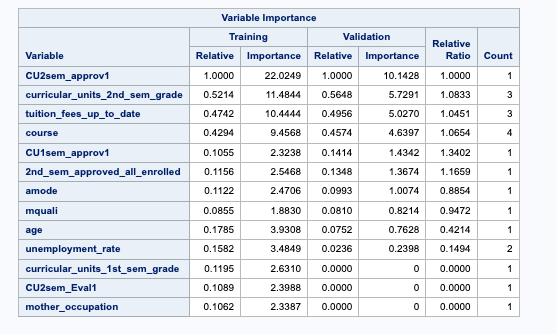
run;

CART Model

CART Entropy full vars







reduced list of vars: CU2sem\_approv1 curricular\_units\_2nd\_sem\_grade tuition\_fees\_up\_to\_date course CU1sem\_approv1 2nd\_sem\_approved\_all\_enrolled amode mquali age unemployment\_rate