Projet Données qualitatives: Fifa19

Dorra Bennour; Hadrien Boulanger

Février 2021

1 Introduction

Dans ce projet, on a choisi de travailler sur des données relatives au jeu Fifa19. Le but étant de prédire la note générale ("overall") des joueurs à partir des différentes variables explicatives (caractéristiques de joueurs de foot) présentées dans la partie suivante, on cherche alors, dans un premier temps à analyser les différentes variables par une étude exploratoire, puis à appliquer une analyse de correspondance multiple suivie par une classification par arbres de décisions pour la prédiction.

2 Présentation des données

Le dataset de départ est la base de données de tous les joueurs de football du jeu Fifa 19 (https://www.kaggle.com/karangadiya/fifa19). Les variables associées y sont nombreuses : elles représentent des informations sur ces les capacités des joueurs, leur age, leur nationalité par exemple.

2.1 Prétraitement

Nous avons sélectionné 21 variables qui nous intéressaient. Certaines étaient déja qualitatives comme le pied favori ou simplement pouvaient être considérées comme qualitatives (ex : les notes sur 5 de réputation internationale).

2.1.1 transformation des variables quantitatives

Mais d'autres étaient quantitatives. Pour l'âge par exemple, nous avons donc découpé l'étendue des âges en 10 intervales.

Mais surtout, la majorité des attributs des joueurs étaient des notes de 0 à 100 sur différentes qualités possibles d'un joueur de football. Nous avons donc fait en sorte de donner pour tous les joueurs et sur chacun de ces attributs une modalité nul, moyen ou bon selon sa note. Mais comment choisir les frontières entre chaque modalité? A partir de quelle note un joueur est-il bon et au dessous de quelle note un joueur est il considéré comme mauvais? Nous avons donc choisi de dire ceci : un joueur est considéré mauvais dans un domaine s'il possède pour ce domaine une des 20% moins bonnes notes, et un joueur est considéré bon dans un domaine s'il est dans le top 20% dans ce domaine en terme de note. Sinon, il est considéré comme moyen dans ce domaine. Voici par exemple un histogramme de la note générale (Overall) des joueurs :

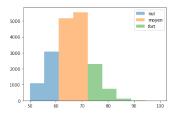


FIGURE 1 – Histogramme de fréquence de la variable relative à la note globale "Overall" et leur transformation qualitative

Ainsi les joueurs correspondants aux batons bleus se verront attribuer comme modalité pour la note générale Gnul, les joueurs correspondant aux batons oranges se verront attribuer la modalité Gmoy et les batons verts correspondent à la modalité Gbon.

2.1.2 inventaire définitif des variables

Voici donc la liste des variables après traîtement que nous avons choisi d'étudier :

- Leur note générale ("**overall**", modalités : Gnul, Gmoy, Gbon), attribut que l'on souhaitera en finalité prédire à partir des autres.
- Le pied favori ("**preferred foot**", left ou right)
- La note de reputation internationale ("international reputation" entre RI1 et RI5)
- La note de qualité du pied faible ("weak foot" allant de WF1 à WF5)
- La note technique ("skill moves" entre SK1 et SK5)
- La morphologie ("bodytype" Normal, Lean, Socky, Ronaldo, Messi, Neymar)
- Le **Crossing** (3 modalités : CRnul, CRmoy ou CRfort)
- Le **Finishing** (3 modalités : FINnul, FINmoy ou FINfort)
- Le **HeadingAccuracy** (3 modalités : HEAnul, HEAmoy ou HEAfort)
- Le **ShortPassing** (3 modalités : SHPnul, SHPmoy ou SHPfort)
- Le **LongPassing** (3 modalités : LOPnul, LOPmoy ou LOPfort)
- Le **Agility** (3 modalités : AGInul, AGImoy ou AGIfort)
- Le **BallControl** (3 modalités : BALnul, BALmoy ou BALfort)
- Le **Dribbling** (3 modalités : DRInul, DRImoy ou DRIfort)
- Le **Stamina** (3 modalités : STAnul, STAmoy ou STAfort)
- Le **Aggression** (3 modalités : AGGnul, AGGmoy ou AGGfort)
- Le **BallControl** (3 modalités : BALnul, BALmoy ou BALfort)
- L'**Age** (10 modalités d'intervales entre 15.971 et 42.1 ans)
- La note de plongeons de gardien **GKDiving** (3 modalités : GKDnul, GKDmoy ou GKDfort)
- La note de prise ferme à la main de gardien **GKHandling** (3 modalités : GKHnul, GKHmoy ou GKHfort)
- La note de dégagement au pied de gardien **GKKicking** (3 modalités : GKKnul, GKKmoy ou GKKfort)
- La note de Positionnement de gardien **GKPositioning** (3 modalités : GKPnul, GKPmoy ou GKPfort)

2.1.3 Sélection des joueurs

La base de donnée contenant au départ plus de 18000 joueurs, nous avons fait le choix de ne sélectionner que 1000 joueurs. Pour être certains d'avoir une bonne représentativité de niveau général des joueurs (Overall), nous avons choisi au hasard 200 joueurs de modalité Gnul, 600 joueurs de modalité Gmoy, et 200 joueurs de modalité Gfort.

3 Analyse exploratoire des données

3.1 Analyse univariée

On commence par réaliser, comme première exploration, les histogrammes de fréquence de chaque variable par la procédure freq

Comme on s'y attend, les effectifs de joueurs sur les modalités de Overall sont bien ceux choisis au préalable (200, 600, 200). Les proportions restent à peu près les mêmes pour toutes les variables de modalités type "X"nul (~ 200) "X"moy (~ 600) "X"bon (~ 200).

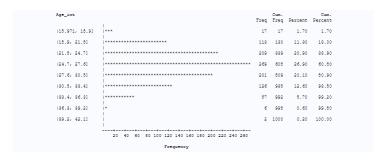


FIGURE 2 – Histogramme catégoriel des ages

On remarque que la répartition des âges a une allure plutôt Gaussienne autour de la catégorie 24.7 - 27.6 ans.

Une bonne réputation internationale est quelque chose de rare, la majorité des joueurs (87%) ont 1 (RI 1) il n'y a même aucun joueur possédant la note de 5.

Les joueurs gauchers de pied représentent environ un quart de l'échantillon.

3.2 Analyse bivariée

On étudie de même les différentes relations entre la variable cible "Overall" relative à la note générale d'un joueur en fonction des autres caractéristiques. On représente par exemple la table de statistiques de la variable cible en fonction de la variable "Ball control" dans la figure ci-dessous :

| Statistic | DF | Value | Prob |
|-----------------------------|-----|----------|--------|
| Chi-Square | 4 | 322.9891 | <.0001 |
| Likelihood Ratio Chi-Square | 4 | 297.7574 | <.0001 |
| Mantel-Haenszel Chi-Square | - 1 | 164.2305 | <.0001 |
| Phi Coefficient | | 0.5683 | |
| Contingency Coefficient | | 0.4941 | |
| Cramer's V | | 0.4019 | |

FIGURE 3 – Table de statistiques de la variable cible en fonction de la variable "Ball control"

La variable "Overall" est très corrélée à la variable "Ball control", en effet, on a une mesure de chi2 qui très élevée (322.98). On constate aussi d'après les autres distances de chi2 en fonction des autres variables, que la variable cible est plus corrélée/ dépendante des variables caractéristiques des joueurs qui ne sont pas des gardiens. Concernant les variables sous jacentes aux notes de gardiens (GKD, GKH, GKK, GKP) elles sont plutôt moins corrélée à la note générale (distance de chi2 entre 1.47 et 9.49. Ceci peut se remarquer aussi d'après les p-valeurs (p-value) qui sont supérieure à 0.01.

4 Analyse factorielle

4.1 Analyse des correspondances multiples

Pour saisir les realtions entre les différentes variables, aussi bien que pour étudier la relation de la variable cible "Overall" en fonction des différentes caractéristiques des joueurs on procède dans un premier temps, par une analyse de correspondance multiples sur toutes les variables, vu qu'on ne traite que des données qualitatives. L'ACM réalisée alors par la procédure correspons donne les résultats suivants après avoir appliqué la correction de BENZECRI afin de mieux comprendre les taux d'inertie/information. La variable "Overall" est considérée comme variable supplémentaire.

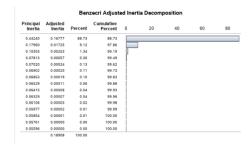


FIGURE 4 – Inertie des composantes trouvées par ACM

En s'appuyant sur le critère de coude, on retient les deux premières composantes pour interpréter la proximation des variables sur les plans factoriels. On a eu un pourcentage d'inertie égale à 88.73% pour le premier axe et 97.86% pour les deux premières composantes.

4.2 Contribution aux axes factoriels

4.2.1 Contribution et qualité de représentation des variables

En analysant maintenant la table des coordonnées, contributions et cosinus carrées des modalités actives par rapport au deux premiers axes factoriels, on constate que les modalités les mieux représentées sont celles qui ont un cosinus au carré proche de 1, et donc il s'agit des modalités des variables relatifs aux notes de "crossing", "finishing", "heading", "shortpassing", et "longpassing".

Les modalités ayant contribuées le plus au premier axe sont : "GKPbon", "GKDbon", "STAnul", "DRImoy", "BALnul", "AGInul", "LOPnul", "FINnul", "FINmoy", "CRnul".

Les modalités ayant contribuées le plus au deuxième axe sont : "DRImoy", "DRIbon", "BALmoy", "BALbon", "AGIbon", "LOPmoy", "LOPbon", "SHPbon", "FINbon", "CRmoy", et "WF4".

4.2.2 Variable Supplémentaire

Les modalités de la variable supplémentaire "overall" sont bien représentées sur l'axe factoriel. La projection des variables sur le plan factoriel est alors la suivante :

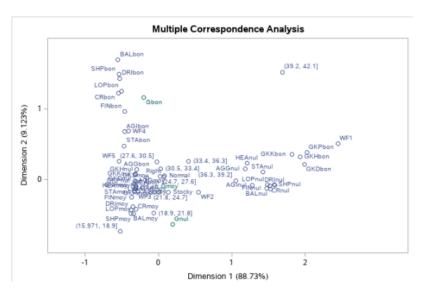


FIGURE 5 – Projection des variables sur le plan factoriel 1-2

Par soucis de visibilité, on a enlevé la variable relative à la réputation internationale. Les modalités de la variable cible "overall" sont représentées en vert. Projection avec les individus.

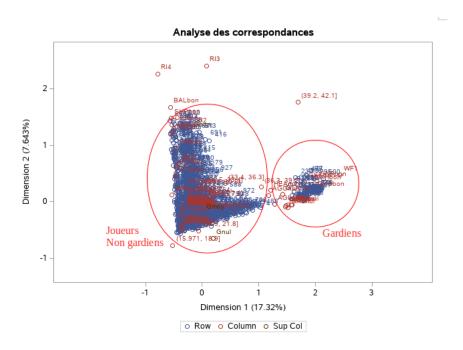


FIGURE 6 – Projection sur le plan factoriel 1-2

Sur cette projection on a les modalités en rouge, les individus en bleu, les modalités de la variable cible "overall" en marron. On remarque que l'on a 2 types de populations : à droite les gardiens et à gauche les autres joueurs.

Interprétation

On constate que les joueurs qui ont une note générale moyenne ont tendance à être moyen partout. Par contre pour ceux qui ont une bonne note en globale, sont plutôt excellents sur quelques aspects (contrôle du ballon, drible, finition et les passes, agilité), et moyens sur d'autres. On remarque aussi que les gardiens forment une sorte de classe particulière (groupe à droite), comme seulement les modalités relatives aux bonnes notes de gardiens, on ne peut pas interpréter les notes faibles et moyennes. Cependant, on tire que les bons gardiens (ceux qui sont bons dans les attributs correspondants au poste), peuvent ne pas avoir de bonnes notes pour d'autres caractéristiques.

5 Selection des variables

Avant d'appliquer tout type d'algorithme, il est intéressant de commencer par une sélection des variables les plus pertinentes (parmi les dimensions de l'acm). En effet, on dipose de 46 dimensions (65 modalités - 19 variables), ce qui est considéré comme un grand nombre de variables. On procède alors par la sélection de variables par la procédure STEPDISC qui étant donné une variable de classification et plusieurs variables quantitatives, la procédure STEPDISC effectue une analyse discriminante par étapes afin de sélectionner un sous-ensemble de variables quantitatives à utiliser pour faire la distinction entre les classes. On obtient alors 28 dimensions les plus discriminantes au seuil 5% qui sont : Dim1, Dim2, Dim3, Dim4, Dim6, Dim7, Dim8, Dim9, Dim10, Dim11, Dim14, Dim15, Dim19, Dim20, Dim24, Dim26, Dim27, Dim28, Dim29, Dim30, Dim31, Dim32, Dim33, Dim34, Dim35, Dim37, Dim39, Dim41. Les résultats de cette procédure est présenté par la table ci-dessous :

| | | | | The ST | EPDISC P | rocedure | • | | | |
|------|--------|----------|---------|-----------|-----------|----------|------------|--------|---------------------|-------|
| | | | | Ste pwise | Selection | n Summ | агу | | | |
| | | | | | | | | | A verage Squared | |
| | Number | | | Partial | | | Wilks | Pr< | Canonical | Pr > |
| Step | In | Entere d | Removed | R-Square | F Value | Pr > F | Lambda | Lambda | Correlation | ASC |
| 1 | 1 | Dim2 | | 0,4232 | 365,76 | <.0001 | 0,57679200 | <.0001 | 0,21160400 | <.000 |
| 2 | 2 | Dim3 | | 0,0934 | 51,32 | <.0001 | 0,52290682 | <.0001 | 0,23915088 | <.000 |
| 3 | 3 | Dim32 | | 0,0512 | 26,83 | <.0001 | 0,49614883 | <.0001 | 0,25842942 | <.000 |
| 4 | 4 | Dim1 | | 0,0308 | 15,67 | <.0001 | 0,48098811 | <.0001 | 0,26865262 | <.000 |
| 5 | 5 | Dim14 | | 0,0240 | 12,21 | <.0001 | 0,46944324 | <.0001 | 0,27415374 | <.000 |
| 6 | 6 | Dim11 | | 0,0214 | 10,85 | <.0001 | 0,45939687 | <.0001 | 0,27961004 | <.000 |
| 7 | 7 | Dim34 | | 0,0182 | 9,18 | 0,0001 | 0,45103984 | <.0001 | 0,28786810 | <.000 |
| 8 | 8 | Dim8 | | 0,0165 | 8,30 | 0,0003 | 0,44380548 | <.0001 | 0,29178910 | <.000 |
| 9 | 9 | Dim33 | | 0,0161 | 8,07 | 0,0003 | 0,43648134 | <.0001 | 0,29596645 | <.000 |
| 10 | 10 | Dim6 | | 0,0149 | 7,49 | 0,0006 | 0,42996123 | <.0001 | 0,30294019 | <.000 |
| 11 | 11 | Dim19 | | 0,0147 | 7,38 | 0,0007 | 0,42384352 | <.0001 | 0,30716724 | <.000 |
| 12 | 12 | Dim9 | | 0,0131 | 6,54 | 0,0015 | 0,41809649 | <.0001 | 0,31108650 | <.000 |
| 13 | 13 | Dim7 | | 0,0115 | 5,71 | 0,0034 | 0,41330481 | <.0001 | 0,31580881 | <.000 |
| 14 | 14 | Dim29 | | 0,0097 | 4,80 | 0,0085 | 0,40931472 | <.0001 | 0,31926821 | <.000 |
| 15 | 15 | Dim31 | | 0,0098 | 4,75 | 0,0088 | 0,40539289 | <.0001 | 0,32320985 | <.000 |
| 16 | 16 | Dim28 | | 0,0093 | 4,61 | 0,0102 | 0,40162091 | <.0001 | 0,32591395 | <.000 |
| 17 | 17 | Dim20 | | 0,0091 | 4,50 | 0,0114 | 0,39797160 | <.0001 | 0,32857833 | <.000 |
| 18 | 18 | Dim10 | | 0,0088 | 4,24 | 0,0147 | 0,39455764 | <.0001 | 0,33083323 | <.000 |
| 19 | 19 | Dim4 | | 0,0085 | 4,18 | 0,0156 | 0,39122021 | <.0001 | 0,33445864 | <.000 |
| 20 | 20 | Dim39 | | 0,0080 | 3,97 | 0,0193 | 0,38807276 | <.0001 | 0,33715971 | <.000 |
| 21 | 21 | Dim24 | | 0,0077 | 3,77 | 0,0233 | 0,38509857 | <.0001 | 0,34029923 | <.000 |
| 22 | 22 | Dim26 | | 0,0075 | 3,70 | 0,0250 | 0,38219744 | <.0001 | 0,34298923 | <.000 |
| 23 | 23 | Dim15 | | 0,0074 | 3,63 | 0,0268 | 0,37936914 | <.0001 | 0,34460773 | <.000 |
| 24 | 24 | Dim35 | | 0,0060 | 2,96 | 0,0524 | 0,37707887 | <.0001 | 0,34811888 | <.000 |
| 25 | 25 | Dim37 | | 0,0055 | 2,68 | 0,0690 | 0,37501210 | <.0001 | 0,34828443 | <.000 |
| 26 | 26 | Dim30 | | 0,0052 | 2,53 | 0,0803 | 0,37307124 | <.0001 | 0,34939842 | <.000 |
| 27 | 27 | Dim27 | | 0,0048 | 2,25 | 0,1080 | 0,37135074 | <.0001 | 0,35079324 | <.000 |
| 28 | 28 | Dim41 | | 0,0044 | 2,16 | 0,1154 | 0,36970101 | <.0001 | 0,35191779 | <.000 |

FIGURE 7 – Résultats de la sélection des variables

6 Analyse factorielle discriminante DISQUAL

6.1 Rappel de la méthode

Comme nous ne disposons que de variables qualitatives dans notre modèle, on opte alors, à des fins de discrimination de la variable cible Y='="Overall", à une analyse discriminante dite DISQUAL. Cette méthode consiste à la combinaison de deux méthodes : Analyse des correspondances multiples (ACM) et l'analyse factorielle discriminante (AFD),on obtient ainsi des mêmes propriétés bénéfiques qui sont : la possibilité de conserver plus de variables explicatives dans le modèle, des poids correctement réparties entre ces dernières et l'affranchissement des méfaits de la multicolinéarité

6.2 Résultats de la méthode

On applique alors l'analyse discriminante sur la base de données des nouvelles variables de l'ACM (variables sélectionnées par STEPDISC). On obtient deux axes canoniques discriminants (car on a 3 groupes et donc k-1 axes discriminants) dont les coefficients en fonction des dimensions sont les suivants :

| The CANDISC Procedure | | | | | | | | | |
|-----------------------|------------------|--------------|--|--|--|--|--|--|--|
| Rav | v Canonical Coef | ficients | | | | | | | |
| Variable | Can1 | Can2 | | | | | | | |
| Dim2 | 3,055742726 | 0,038366032 | | | | | | | |
| Dim3 | -1,449803220 | 0,482870993 | | | | | | | |
| Dim32 | 1,393446963 | 2,661813024 | | | | | | | |
| Dim1 | -0,380111639 | 0,185087500 | | | | | | | |
| Dim14 | 0,788641005 | 1,026231788 | | | | | | | |
| Dim11 | 0,824270369 | -0,355852286 | | | | | | | |
| Dim34 | 0,428649987 | -2,295952634 | | | | | | | |
| Dim8 | 0,703205191 | 0,059597072 | | | | | | | |
| Dim33 | 0,999653048 | 0,764346727 | | | | | | | |
| Dim6 | -0,126883555 | 1,355989715 | | | | | | | |
| Dim19 | -0,640549385 | -0,790766659 | | | | | | | |
| Dim9 | 0,537450938 | -0,699920985 | | | | | | | |
| Dim7 | -0,270433955 | 1,061253912 | | | | | | | |
| Dim29 | 0,468688476 | 1,167753690 | | | | | | | |
| Dim31 | 0,316179413 | 1,439794405 | | | | | | | |
| Dim28 | -0,584318566 | 0,755648618 | | | | | | | |
| Dim20 | -0,486932086 | 0,665590423 | | | | | | | |
| Dim10 | -0,452225362 | -0,463599387 | | | | | | | |
| Dim4 | 0,146783272 | -0,909492894 | | | | | | | |
| Dim39 | -0,624384732 | -1,390613571 | | | | | | | |
| Dim24 | -0,231690082 | 1,076442197 | | | | | | | |
| Dim26 | 0,363827782 | -0,928938673 | | | | | | | |
| Dim15 | 0,487673603 | -0,134975627 | | | | | | | |
| Dim35 | 0,635747494 | -0,565409091 | | | | | | | |
| Dim37 | 0,317870451 | -1,343143121 | | | | | | | |
| Dim30 | 0,529307947 | -0,069454679 | | | | | | | |
| Dim27 | -0,361827316 | 0,623019157 | | | | | | | |
| Dim41 | -0,632479234 | -0,627233919 | | | | | | | |

 ${\tt Figure~8-Coefficients~des~axes~canoniques~en~fonction~des~dimensions~s\'electionn\'ees~de~l'ACM}$

Le premier axe d'après la figure 9 permet d'expliquer presque la totalité de l'information, en effet il en représente 90.29~%.

| | The CANDISC Procedure | | | | | | | | | | | | |
|---|-----------------------|--|----------|-------------|------------|---|------------|------------|------------|-------------|--------|--------|--------|
| | | Eigenvalues of invEi/HI Test of Hit: The canonical correlations in the current row and all that follow are Adjusted Approximate Squared = CastRegit CastReg) 22 are | | | | | | | | | | | |
| | Canonical | Canonical | Standard | Canonical | | | | | Likelihood | Approximate | | | |
| | Correlation | Correlation | Error | Correlation | Eigenvalue | Difference | Proportion | Cumulative | Ratio | FValue | Num DF | Den DF | Pr > F |
| 1 | 0,759069 | 0,751359 | 0,013408 | 0,576216 | 1,3597 | 1,2134 | 0,9029 | 0,9029 | 0,36970101 | 22,33 | 56 | 1940 | <.0001 |
| 2 | 0,3 57239 | 0,325912 | 0,027601 | 0,127620 | 0,1460 | 0,1463 0,0974 1,0000 0,87238010 5,26 27 971 | | | | | | | <.0001 |

FIGURE 9 - Axes discriminants de l'AFD

Et donc pour obtenir les coefficients de ces axes, il suffit de remplacer la combinaison linéaire des dimensions en fonction des modalités (Voir annexe) dans les équations des axes canoniques.

Afin de mieux saisir les différentes relations entre les axes discriminants et les modalités, on explore davantage les corrélations entre ces axes et les dimensions retenus de l'ACM.

| The CANDISC Procedure | | | | | | | | | |
|-----------------------|----------------|-----------|--|--|--|--|--|--|--|
| Total | Canonical Stru | ucture | | | | | | | |
| Variable | Can1 | Can2 | | | | | | | |
| Dim2 | 0,856976 | 0,015438 | | | | | | | |
| Dim3 | -0,301679 | 0,144161 | | | | | | | |
| Dim32 | 0,158496 | 0,434397 | | | | | | | |
| Dim1 | -0,160492 | 0,112124 | | | | | | | |
| Dim14 | 0,121214 | 0,226307 | | | | | | | |
| Dim11 | 0,132118 | -0,081836 | | | | | | | |
| Dim34 | 0,045119 | -0,346735 | | | | | | | |
| Dim8 | 0,116469 | 0,014162 | | | | | | | |
| Dim33 | 0,106995 | 0,117377 | | | | | | | |
| Dim6 | -0,021356 | 0,327460 | | | | | | | |
| Dim19 | -0,093041 | -0,164798 | | | | | | | |
| Dim9 | 0,087363 | -0,163238 | | | | | | | |
| Dim7 | -0,045300 | 0,255057 | | | | | | | |
| Dim29 | 0,055990 | 0,200150 | | | | | | | |
| Dim31 | 0,036175 | 0,236353 | | | | | | | |
| Dim28 | -0,072974 | 0,135401 | | | | | | | |
| Dim20 | -0,070667 | 0,138591 | | | | | | | |
| Dim10 | -0,072741 | -0,106991 | | | | | | | |
| Dim4 | 0,026078 | -0,231830 | | | | | | | |
| Dim39 | -0,053614 | -0,171322 | | | | | | | |
| Dim24 | -0,031704 | 0,211336 | | | | | | | |
| Dim26 | 0,048302 | -0,176944 | | | | | | | |
| Dim15 | 0,074129 | -0,029437 | | | | | | | |
| Dim35 | 0,062088 | -0,079226 | | | | | | | |
| Dim37 | 0,028677 | -0,173852 | | | | | | | |
| Dim30 | 0,061939 | -0,011661 | | | | | | | |
| Dim27 | -0,045372 | 0,112090 | | | | | | | |
| Dim41 | -0,051912 | -0,073864 | | | | | | | |

FIGURE 10 - Corrélations entre les axes discriminants et les dimensions de l'ACM

Le premier axe canonique est fortement corrélé à la Dimension 2 et négativement corrélé à la Dimension 3. Tandis qu'il est faiblement corrélé aux autres dimensions. Concernant les deuxième axe discriminant, il est plutôt corrélé positivement aux dimensions 6,32 et 7 et négativement corrélé à la dimension 34. On rappelle que : Les coefficients les plus importants correspondent aux modalités :

- RI3, RI4, CRbon, SHPbon, LOPbon, BALbon, DRIbon et la tranche d'âge 39-42 pour la dimension 2.
- RI2, RI4, WF5, HEAbon, AGGbon, tranche d'âge entre 15.9 et 18 ans, et tranche d'âge entre 39 et 42 ans pour la dimension 3.
- RI4, WF1, tranche d'âge entre 15.9 et 18 ans, tranche d'âge entre 36 et 39 ans, et tranche d'âge entre 39 et 42 ans pour la dimension 6.
- RI2, RI3, RI4, WF5, tranche d'âge entre 33 et 36 ans, et tranche d'âge entre 39 et 42 ans pour la dimension 7.
- RI2, RI3 pour la dimension 32

Et donc les variables qui contribuent au plus aux axes sont : Réputation internationale RI, le crossing CR, les passes, le contrôl du ballon et l'âge.

La projection de nos données alors selon les 3 classes de la variables cible "Overall" sur le plan formé par les deux facteurs canoniques est la suivante :

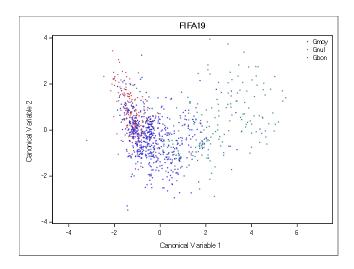


FIGURE 11 – Projection des données sur les deux axes canoniques

On remarque que ces deux axes canoniques ne discriminent pas assez les classes. En effet, les deux classes "Gnul" et "Gmoy" sont mêlés. Ceci peut être dû soit à la capacité discriminante des axes, soit à la proximité des deux classes en termes de modalités.

7 Classification

Dans cette partie on utilisera 2 méthodes de classification : un arbre de décision simple, puis la méthode random forest. Nous découpons notre ensemble de 1000 individus en 2 ensembles échantillonnés qui respectent les proportions de note générale ("overall") telles qu'on les a définies dans la partie 2.1.1 (page 1) : 20% de Gnul, 80% de Gmoy, 20% de Gbon. On a donc :

- Un ensemble d'entraînement composé de 750 individus (75% de la population) : 150 Gnul, 450 Gmoy, 150 Gbon.
- Un ensemble de test composé de 250 individus (25% de la population) : 50 Gnul, 150 Gmoy, 50 Gbon. On utilise pour la classification les 28 variables retenues lors de la sélection (méthode stepdisc). On obtient les résultats suivants.

7.1 Arbre de décision

On réalise un arbre de décision (voir annexe) explicatif de "Overall". L'arbre est construit à chaque étape de manière à discriminer la population en des sous ensembles les plus homogènes possibles. Sur l'ensemble de test, on obtient la matrice de confusion suivante :

FIGURE 12 – Matrice de confusion sur l'ensemble de test

On obtient une proportion de 66.8% d'individus bien classés. Voyons les résultats obtenus avec l'algorithme des forêts aléatoires.

7.2 Forêts aléatoires

Toujours avec les mêmes ensembles, voyons la matrice de confusion sur l'ensemble de tests :

FIGURE 13 - Matrice de confusion sur l'ensemble de test

On obtient une proportion de 69.2% d'individus bien classés. On s'attendait à obtenir une meilleure performance.

8 Conclusion

Tout au long de ce projet nous avons traité des bases de données de variables qualitatives. Sachant que le traitement des données qualitatives n'est pas le même que celui des données quantitatives, nous avons mené en un premier temps une étude exploratoire nous permettant d'étudier les relations entre les différentes variables. L'objectif de notre étude étant de prédire la note générale d'un joueur donné sachant ses notes dans différentes compétences, on a alors procédé à un classement des individus à l'aide de l'analyse factorielle discriminante en se basant sur les dimensions retenues par l'ACM après sélection de variables. Une classification par les arbres de décisions et des forêts aléatoires a été ensuite effectuée. Les résultats obtenus (entre 65 et 70 % de bons classements) sont corrects mais pourraient être grandement améliorés. En effet, l'analyse des correspondances multiples a mis en évidence l'existence de 2 grands groupes d'individus dans la population : les gardiens et les joueurs des autres postes. En reproduisant les mêmes étapes sur les 2 populations séparément il est probable que nous puissions obtenir d'excellents résultats en classification.

Annexe

| Obs | _NAME_ | Dim1 | Dim2 | Dim3 | Dim4 | Dim6 | Dim7 | Dim8 | Dim9 | Dim10 | Dim11 |
|-----|--------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| 1 | Left | -0.22221 | -0.10794 | 0.25182 | 0.23108 | -0.25912 | 0.52410 | -0.69633 | 0.16162 | 0.23100 | 0.98052 |
| 2 | Right | 0.07407 | 0.03598 | -0.08394 | -0.07703 | 0.08637 | -0.17470 | 0.23211 | -0.05387 | -0.07700 | -0.32684 |
| 3 | RI1 | 0.02108 | -0.21785 | 0.10976 | -0.01577 | -0.07050 | 0.05408 | 0.02298 | -0.03028 | -0.00706 | -0.07459 |
| 4 | RI2 | -0.16907 | 1.32212 | -1.02166 | -0.05678 | 0.43541 | -1.12496 | -0.55109 | -0.33153 | -0.01344 | 0.92154 |
| 5 | RI3 | 0.07598 | 2.40137 | 0.46641 | 0.62358 | 0.31215 | 1.80893 | 0.87248 | 2.77288 | 0.54434 | -1.40151 |
| 6 | RI4 | -0.78527 | 2.25989 | -1.68889 | 1.11226 | 2.63269 | 4.15453 | 2.71169 | -0.52346 | -0.89846 | 1.55682 |
| 7 | WF1 | 2.43927 | 0.47554 | 0.93156 | 2.52471 | 1.37003 | 0.62576 | 3.23271 | -3.02945 | 0.76171 | 1.63240 |
| 8 | WF2 | 0.54746 | -0.17576 | -0.15890 | 0.03101 | -0.92860 | 0.37046 | -0.47340 | 0.14548 | 0.43564 | 0.44489 |
| 9 | WF3 | -0.05985 | -0.16926 | -0.08282 | 0.06792 | 0.25704 | 0.10790 | 0.05577 | 0.08302 | -0.19858 | -0.13644 |
| 10 | WF4 | -0.41691 | 0.67092 | 0.32973 | -0.39010 | 0.01404 | -0.66152 | 0.08772 | -0.33913 | -0.02127 | -0.08489 |
| 11 | WF5 | -0.53889 | 0.11993 | 1.01988 | 1.08613 | 0.47296 | -1.34332 | 1.43611 | 1.60321 | 4.61833 | 0.38380 |

| Obs | Dim14 | Dim15 | Dim19 | Dim20 | Dim24 | Dim26 | Dim27 | Dim28 | Dim29 | Dim30 | Dim31 |
|-----|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| 1 | 0.21868 | -0.18740 | 0.05554 | -0.11360 | -0.12117 | -0.00993 | 0.32816 | 0.34348 | 0.12175 | 0.53732 | -0.01225 |
| 2 | -0.07289 | 0.06247 | -0.01851 | 0.03787 | 0.04039 | 0.00331 | -0.10939 | -0.11449 | -0.04058 | -0.17911 | 0.00408 |
| 3 | -0.03643 | -0.01268 | 0.04992 | 0.02516 | 0.00374 | 0.03540 | -0.04850 | -0.03866 | -0.00815 | 0.01294 | -0.09638 |
| 4 | 0.33078 | 0.23184 | -0.27178 | -0.21492 | -0.26905 | -0.34776 | 0.13885 | 0.30427 | -0.50454 | -0.14050 | 0.73213 |
| 5 | 0.34493 | 0.59018 | -0.82996 | -1.34151 | 1.10379 | -0.33419 | 0.61895 | 0.27114 | 2.42713 | 0.14420 | 0.38210 |
| 6 | -1.47816 | -4.82488 | 0.11360 | 5.61559 | -0.34717 | 1.93802 | 3.11735 | -0.25323 | 0.43779 | -0.20643 | 1.16750 |
| 7 | -2.57427 | 0.29050 | 2.16912 | 1.09573 | -2.08006 | 0.59858 | -0.69981 | -0.30412 | 4.14483 | -0.42123 | 0.50333 |
| 8 | 0.26014 | 0.40969 | 0.21314 | 0.29958 | 0.52569 | -0.13732 | -0.78970 | -0.14751 | 0.17283 | -0.42419 | 0.19353 |
| 9 | -0.00021 | -0.26813 | -0.17693 | -0.07339 | -0.17498 | 0.04220 | -0.01673 | -0.07026 | -0.11442 | -0.00390 | 0.00967 |
| 10 | -0.11988 | 0.40822 | 0.24418 | -0.09006 | 0.02596 | -0.19216 | 0.82027 | 0.35068 | 0.03036 | 0.40216 | -0.24232 |
| 11 | -0.55156 | 0.52828 | 0.31243 | -0.33617 | 2.43892 | 4.06825 | 0.43173 | 0.53430 | -0.26034 | 0.82573 | 0.14291 |

| Obs | Dim32 | Dim33 | Dim34 | Dim35 | Dim37 | Dim39 | Dim41 |
|-----|----------|----------|----------|----------|----------|----------|----------|
| 1 | -0.13876 | -0.10068 | -0.06026 | 0.09422 | 0.04225 | 0.01355 | 0.01621 |
| 2 | 0.04625 | 0.03356 | 0.02009 | -0.03141 | -0.01408 | -0.00452 | -0.00540 |
| 3 | -0.15992 | -0.11646 | 0.02309 | -0.00216 | -0.01604 | -0.01061 | -0.00188 |
| 4 | 1.10640 | 0.80040 | -0.13339 | 0.02499 | 0.08493 | 0.05824 | 0.03099 |
| 5 | 1.39934 | 0.82121 | -0.30797 | -0.00797 | 0.23909 | 0.04234 | -0.00115 |
| 6 | 0.64990 | 1.44604 | -0.13309 | -0.06653 | 0.13139 | 0.55561 | -0.25945 |
| 7 | 0.03869 | 0.25487 | 0.27971 | -0.22167 | 0.06527 | -0.47567 | 0.05025 |
| 8 | -0.08067 | 0.10435 | 0.04342 | -0.01768 | 0.05714 | 0.00587 | -0.00914 |
| 9 | 0.07550 | 0.03352 | 0.02899 | -0.00088 | 0.00222 | 0.00948 | 0.01015 |
| 10 | -0.18181 | -0.24303 | -0.13089 | 0.01903 | -0.06105 | -0.01079 | -0.01800 |
| 11 | 0.54511 | 0.68204 | -0.30376 | 0.23783 | -0.08090 | -0.11527 | -0.16752 |

| Obs | _NAME_ | Dim1 | Dim2 | Dim3 | Dim4 | Dim6 | Dim7 | Dim8 | Dim9 | Dim10 | Dim11 |
|-----|--------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| 12 | Lean | -0.17183 | -0.09394 | 0.50060 | -0.34019 | 0.06216 | 0.17092 | 0.14076 | -0.23821 | -0.39843 | 0.01040 |
| 13 | Normal | 0.07914 | 0.07380 | -0.23241 | 0.19526 | 0.06949 | -0.06976 | -0.08834 | 0.13394 | 0.12688 | 0.04259 |
| 14 | Stocky | 0.13526 | -0.15328 | -0.38027 | -0.01936 | -0.77925 | -0.20127 | 0.06421 | 0.00720 | 0.73506 | -0.36088 |
| 15 | CRbon | -0.55644 | 1.19836 | 0.18607 | 0.16660 | -0.37655 | 0.31378 | -0.40505 | -0.05528 | 0.18802 | 0.26130 |
| 16 | CRmoy | -0.34777 | -0.37923 | 0.05621 | 0.08969 | 0.14013 | -0.18112 | 0.09119 | -0.00067 | -0.07914 | -0.06574 |
| 17 | CRnul | 1.51387 | -0.05930 | -0.33575 | -0.41230 | -0.04085 | 0.21664 | 0.12498 | 0.05428 | 0.04642 | -0.06103 |
| 18 | FINbon | -0.46448 | 0.94488 | 0.28465 | -0.73181 | 0.18393 | -0.55340 | 0.26539 | 0.15362 | 0.29793 | 0.30636 |
| 19 | FINmoy | -0.36834 | -0.25345 | -0.00096 | 0.35784 | -0.03269 | 0.04944 | -0.04592 | -0.06953 | -0.10333 | -0.12579 |
| 20 | FINnul | 1.46944 | -0.09079 | -0.24527 | -0.39660 | -0.06580 | 0.33933 | -0.09853 | 0.06709 | 0.03906 | 0.09663 |
| 21 | HEAbon | -0.16428 | 0.14559 | -1.28069 | -0.40544 | 0.15915 | -0.35639 | 0.65904 | 0.40246 | -0.12971 | 0.52154 |
| 22 | HEAmoy | -0.35038 | -0.11176 | 0.13642 | -0.02290 | -0.08101 | 0.19301 | -0.23336 | -0.22214 | 0.12627 | -0.24258 |

| Obs | Dim14 | Dim15 | Dim19 | Dim20 | Dim24 | Dim26 | Dim27 | Dim28 | Dim29 | Dim30 | Dim31 |
|-----|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| 12 | 0.53527 | 0.21732 | -0.35956 | -0.28988 | -0.45817 | 0.37473 | -0.08831 | 0.15120 | 0.13934 | -0.04015 | 0.28953 |
| 13 | -0.39698 | -0.03929 | -0.00900 | 0.29793 | 0.26344 | -0.18320 | 0.02787 | -0.03865 | -0.04444 | 0.01807 | -0.16852 |
| 14 | 0.69846 | -0.62360 | 1.58260 | -0.99556 | -0.02950 | -0.21599 | 0.16479 | -0.34964 | -0.25655 | 0.03475 | 0.03385 |
| 15 | -0.08426 | -0.00028 | -0.14760 | -0.02462 | -0.40614 | 0.00271 | 0.34734 | 0.00845 | -0.02666 | -0.80345 | -0.53279 |
| 16 | 0.03410 | -0.06986 | 0.05998 | 0.00711 | 0.17114 | -0.04290 | -0.19115 | -0.04491 | 0.06142 | 0.38505 | 0.21227 |
| 17 | -0.01691 | 0.19847 | -0.03034 | 0.00314 | -0.10080 | 0.11914 | 0.21331 | 0.11941 | -0.14900 | -0.33134 | -0.09754 |
| 18 | -0.04932 | -0.27800 | 0.14164 | -0.34330 | 0.35646 | 0.44639 | -0.20786 | 0.05730 | -0.00370 | -0.09434 | -0.01520 |
| 19 | 0.03289 | 0.04095 | -0.00055 | 0.11513 | -0.06876 | -0.15166 | 0.02186 | -0.00954 | 0.02525 | -0.03617 | 0.05491 |
| 20 | -0.05209 | 0.12389 | -0.12183 | -0.03360 | -0.11187 | 0.04937 | 0.11796 | -0.02235 | -0.06975 | 0.18676 | -0.14548 |
| 21 | 0.26438 | 0.12443 | -0.17276 | -0.07646 | -0.01646 | 0.21767 | -0.25872 | 0.08621 | 0.16547 | 0.07309 | -0.41337 |
| 22 | -0.14551 | -0.02366 | 0.01801 | -0.04224 | 0.01216 | -0.03650 | 0.12559 | -0.04965 | -0.05526 | -0.02812 | 0.20124 |

| Obs | Dim32 | Dim33 | Dim34 | Dim35 | Dim37 | Dim39 | Dim41 |
|-----|----------|----------|----------|----------|----------|----------|----------|
| 12 | -0.16434 | 0.01674 | 0.08151 | 0.00462 | -0.01616 | 0.06661 | 0.01126 |
| 13 | 0.04347 | 0.02456 | -0.04539 | -0.00066 | -0.00101 | -0.03028 | -0.00699 |
| 14 | 0.36917 | -0.25335 | -0.00577 | -0.01453 | 0.07565 | -0.05541 | 0.00455 |
| 15 | 0.05822 | -0.09116 | 0.04354 | 0.08672 | -0.20705 | -0.05097 | 0.06123 |
| 16 | -0.00807 | -0.01598 | -0.04651 | -0.02229 | 0.00310 | 0.02354 | -0.03563 |
| 17 | -0.03225 | 0.13169 | 0.09073 | -0.01892 | 0.18736 | -0.01849 | 0.04310 |
| 18 | -0.42973 | 0.27198 | 0.35941 | -0.02100 | -0.21928 | 0.02907 | 0.18272 |
| 19 | 0.07344 | -0.11210 | -0.27873 | 0.14318 | 0.04771 | 0.00454 | 0.09460 |
| 20 | 0.16220 | 0.08701 | 0.49247 | -0.39556 | 0.05319 | -0.03844 | -0.43265 |
| 21 | 0.10548 | -0.48074 | -0.32155 | 0.20325 | 0.22842 | -0.22061 | 0.08756 |
| 22 | -0.17839 | 0.26280 | -0.00361 | 0.06588 | 0.03295 | -0.08341 | 0.00074 |

| Obs | _NAME_ | Dim1 | Dim2 | Dim3 | Dim4 | Dim6 | Dim7 | Dim8 | Dim9 | Dim10 | Dim11 |
|-----|--------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| 23 | HEAnul | 1.20513 | 0.20299 | 0.76283 | 0.44042 | 0.09800 | -0.25440 | 0.09853 | 0.29984 | -0.26120 | 0.25226 |
| 24 | SHPbon | -0.55135 | 1.47493 | -0.20788 | 0.18459 | 0.07646 | 0.29796 | -0.02857 | -0.32899 | -0.10279 | -0.24893 |
| 25 | SHPmoy | -0.33848 | -0.47519 | 0.08877 | 0.03761 | -0.04362 | -0.17579 | 0.00795 | 0.14587 | -0.01104 | 0.15319 |
| 26 | SHPnul | 1.57745 | -0.05701 | -0.05768 | -0.29985 | 0.05428 | 0.22907 | 0.00488 | -0.10751 | 0.13711 | -0.21043 |
| 27 | LOPbon | -0.51041 | 1.24588 | -0.40228 | 0.32521 | 0.06503 | 0.29509 | -0.02901 | -0.45763 | -0.05616 | -0.27384 |
| 28 | LOPmoy | -0.30689 | -0.41120 | 0.11058 | -0.02733 | -0.03657 | -0.12962 | -0.05899 | 0.24021 | -0.03716 | 0.15176 |
| 29 | LOPnul | 1.48399 | -0.07002 | 0.09056 | -0.26323 | 0.04261 | 0.08207 | 0.21155 | -0.24727 | 0.17367 | -0.17253 |
| 30 | AGIbon | -0.46644 | 0.60473 | 1.03736 | 0.03753 | -0.15380 | -0.10098 | 0.25683 | 0.91179 | -0.30524 | -0.05320 |
| 31 | AGImoy | -0.28158 | -0.15939 | -0.13585 | 0.07770 | 0.05849 | -0.01913 | -0.14960 | -0.32733 | 0.11879 | 0.03053 |
| 32 | AGInul | 1.27620 | -0.04745 | -0.50294 | -0.27136 | -0.04296 | 0.14810 | 0.23085 | 0.19510 | -0.09353 | -0.04640 |
| 33 | BALbon | -0.56331 | 1.66942 | 0.14397 | -0.23866 | 0.04119 | 0.13017 | 0.05146 | -0.10160 | 0.07547 | 0.01318 |

| Obs | Dim14 | Dim15 | Dim19 | Dim20 | Dim24 | Dim26 | Dim27 | Dim28 | Dim29 | Dim30 | Dim31 |
|-----|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| 23 | 0.19572 | -0.04281 | 0.10408 | 0.19720 | -0.02151 | -0.08959 | -0.14095 | 0.07043 | 0.01471 | 0.01767 | -0.22696 |
| 24 | 0.20297 | -0.10412 | 0.09576 | -0.07243 | 0.17637 | -0.14037 | -0.28808 | -0.05470 | 0.05095 | 0.25514 | -0.04270 |
| 25 | -0.07623 | 0.03419 | -0.02959 | -0.01707 | -0.14528 | 0.07497 | 0.04229 | -0.10537 | -0.01562 | -0.04626 | -0.13602 |
| 26 | 0.02482 | 0.00206 | -0.00751 | 0.12463 | 0.25991 | -0.08425 | 0.16346 | 0.37295 | -0.00436 | -0.11824 | 0.45324 |
| 27 | 0.21640 | -0.06658 | 0.16274 | -0.07850 | 0.03613 | -0.17346 | -0.21932 | 0.30573 | -0.01998 | 0.24897 | -0.20794 |
| 28 | -0.12095 | 0.04862 | -0.08012 | 0.02121 | -0.12989 | 0.06164 | 0.04963 | -0.32090 | 0.01021 | 0.02814 | 0.00036 |
| 29 | 0.13948 | -0.07782 | 0.07172 | 0.01880 | 0.35912 | -0.00369 | 0.08199 | 0.65632 | -0.00995 | -0.35173 | 0.22069 |
| 30 | -0.10702 | 0.08147 | 0.43976 | 0.63140 | -0.34109 | -0.17218 | -0.30472 | 0.47097 | -0.38210 | 0.05393 | 0.27598 |
| 31 | 0.05305 | -0.06128 | -0.21498 | -0.19523 | 0.20683 | 0.08140 | 0.14282 | -0.24358 | 0.09193 | -0.07508 | 0.06950 |
| 32 | -0.06773 | 0.11559 | 0.26910 | 0.03875 | -0.33156 | -0.09686 | -0.16762 | 0.32908 | 0.05687 | 0.18228 | -0.45753 |
| 33 | -0.01248 | -0.05430 | -0.13241 | 0.01641 | -0.04589 | -0.04408 | -0.22246 | -0.27106 | 0.04769 | 0.12807 | 0.01574 |

| Obs | Dim32 | Dim33 | Dim34 | Dim35 | Dim37 | Dim39 | Dim41 |
|-----|----------|----------|----------|----------|----------|----------|----------|
| 23 | 0.44028 | -0.35051 | 0.30550 | -0.38453 | -0.30846 | 0.45318 | -0.08246 |
| 24 | 0.03173 | -0.00839 | 0.04625 | 0.02670 | -0.24929 | 0.08346 | -0.02850 |
| 25 | -0.01893 | 0.13182 | 0.03260 | -0.08489 | 0.19173 | -0.07657 | 0.01536 |
| 26 | 0.02503 | -0.38898 | -0.14500 | 0.22900 | -0.32628 | 0.14657 | -0.01751 |
| 27 | -0.19681 | 0.04766 | 0.21956 | -0.41631 | 0.45934 | -0.07300 | 0.04651 |
| 28 | 0.05677 | 0.03806 | -0.05868 | 0.02650 | -0.20390 | -0.00179 | 0.03416 |
| 29 | 0.03612 | -0.16736 | -0.05454 | 0.36293 | 0.13427 | 0.08336 | -0.15421 |
| 30 | 0.13671 | -0.06674 | 0.27233 | 0.13675 | 0.06521 | -0.36249 | 0.05076 |
| 31 | 0.08241 | -0.18148 | 0.05156 | -0.09458 | 0.01207 | 0.00756 | -0.00951 |
| 32 | -0.37370 | 0.61525 | -0.39931 | 0.16865 | -0.09478 | 0.29805 | -0.01584 |
| 33 | -0.03813 | -0.08730 | -0.11470 | 0.14696 | -0.28328 | -0.36263 | 0.07909 |

| Obs | _NAME_ | Dim1 | Dim2 | Dim3 | Dim4 | Dim6 | Dim7 | Dim8 | Dim9 | Dim10 | Dim11 |
|-----|--------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| 34 | BALmoy | -0.37133 | -0.47719 | 0.06462 | 0.22956 | 0.01428 | -0.15087 | 0.01359 | 0.02291 | -0.04731 | 0.03328 |
| 35 | BALnul | 1.52099 | -0.10480 | -0.30422 | -0.43528 | -0.07528 | 0.30903 | -0.08222 | 0.02350 | 0.06709 | -0.10421 |
| 36 | DRIbon | -0.53687 | 1.36311 | 0.65413 | -0.40842 | -0.09477 | -0.14469 | 0.08349 | 0.23848 | 0.14924 | 0.13946 |
| 37 | DRImoy | -0.37389 | -0.40989 | -0.08698 | 0.26386 | 0.03979 | -0.03272 | -0.00914 | -0.10914 | -0.07780 | -0.01548 |
| 38 | DRInul | 1.58195 | -0.05712 | -0.34705 | -0.39375 | -0.02893 | 0.22811 | -0.05001 | 0.09904 | 0.08966 | -0.08292 |
| 39 | STAbon | -0.47164 | 0.45652 | -0.05279 | 0.48283 | -0.88327 | 0.05605 | -0.13703 | 0.15245 | -0.67894 | -0.65404 |
| 40 | STAmoy | -0.32928 | -0.17012 | -0.01147 | -0.19032 | 0.12632 | 0.03465 | 0.14177 | -0.06748 | 0.20822 | 0.20112 |
| 41 | STAnul | 1.42346 | 0.13229 | 0.08092 | 0.17202 | 0.37061 | -0.15531 | -0.31991 | 0.07715 | -0.05837 | -0.05789 |
| 42 | AGGbon | -0.30421 | 0.24766 | -1.20200 | 0.60639 | -0.18729 | -0.02716 | 0.32974 | 0.46704 | -0.29665 | -0.04525 |
| 43 | AGGmoy | -0.30351 | -0.11424 | 0.08928 | -0.30385 | 0.01038 | 0.09759 | -0.19294 | -0.17892 | 0.10421 | -0.06347 |
| 44 | AGGnul | 1.17767 | 0.11291 | 0.83168 | 0.34803 | 0.14006 | -0.26465 | 0.27145 | 0.10463 | -0.03847 | 0.22953 |

| Obs | Dim14 | Dim15 | Dim19 | Dim20 | Dim24 | Dim26 | Dim27 | Dim28 | Dim29 | Dim30 | Dim31 |
|-----|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| 34 | 0.07927 | -0.05177 | 0.04169 | -0.03887 | 0.05709 | 0.02071 | 0.01183 | 0.14581 | 0.00049 | -0.12007 | -0.06188 |
| 35 | -0.21048 | 0.19119 | -0.00240 | 0.09435 | -0.11983 | -0.01987 | 0.15842 | -0.17367 | -0.04240 | 0.22488 | 0.15913 |
| 36 | -0.10633 | 0.05356 | -0.00887 | 0.13534 | -0.07069 | -0.07138 | 0.03399 | -0.62384 | -0.01680 | -0.05395 | 0.15244 |
| 37 | 0.10720 | -0.07970 | 0.03542 | -0.06725 | 0.09002 | 0.02442 | -0.04225 | 0.26263 | 0.04789 | -0.05972 | -0.09569 |
| 38 | -0.21467 | 0.18301 | -0.09503 | 0.07166 | -0.19737 | -0.00561 | 0.09188 | -0.19245 | -0.12407 | 0.22348 | 0.13882 |
| 39 | -0.07937 | 0.11707 | -0.30986 | 0.03321 | 0.26478 | 0.71844 | 0.13704 | 0.19173 | -0.25369 | 0.09693 | -0.20226 |
| 40 | -0.05530 | 0.04762 | 0.06897 | 0.02766 | -0.08398 | -0.20536 | -0.00570 | -0.04115 | 0.10189 | -0.00539 | 0.08153 |
| 41 | 0.23920 | -0.24794 | 0.05386 | -0.11406 | 0.03136 | 0.01553 | -0.10045 | -0.03804 | -0.09623 | -0.06684 | -0.07764 |
| 42 | 0.60556 | 0.40070 | 0.10977 | 0.18044 | 0.07535 | -0.09325 | 0.48449 | -0.33480 | 0.04681 | 0.03528 | 0.21101 |
| 43 | -0.30576 | -0.11481 | -0.02396 | -0.05973 | -0.10566 | 0.07669 | -0.24151 | 0.10071 | -0.01295 | 0.09011 | -0.12217 |
| 44 | 0.35446 | -0.02502 | -0.02907 | 0.01256 | 0.24463 | -0.14240 | 0.27433 | 0.00670 | -0.00428 | -0.29945 | 0.16985 |

| Obs | Dim32 | Dim33 | Dim34 | Dim35 | Dim37 | Dim39 | Dim41 |
|-----|----------|----------|----------|----------|----------|----------|----------|
| 34 | -0.04586 | 0.09432 | 0.01070 | 0.02129 | 0.11359 | 0.20034 | -0.09847 |
| 35 | 0.16080 | -0.18811 | 0.06885 | -0.18585 | -0.07324 | -0.24707 | 0.20675 |
| 36 | -0.19119 | -0.12917 | -0.34138 | 0.06386 | 0.39711 | 0.50397 | -0.31599 |
| 37 | 0.05622 | 0.10203 | 0.11421 | 0.02813 | -0.17898 | -0.16566 | 0.02208 |
| 38 | 0.01171 | -0.17867 | -0.01937 | -0.14057 | 0.15689 | 0.02000 | 0.22571 |
| 39 | 0.24006 | -0.10851 | -0.02075 | 0.06974 | 0.20625 | 0.08576 | 0.08704 |
| 40 | 0.01899 | -0.02641 | 0.21662 | 0.11892 | 0.06778 | 0.01766 | 0.03405 |
| 41 | -0.26546 | 0.17504 | -0.65128 | -0.42744 | -0.38705 | -0.12842 | -0.18017 |
| 42 | -0.49622 | 0.06352 | 0.44344 | 0.17849 | -0.25940 | 0.08933 | -0.14171 |
| 43 | 0.23171 | -0.01914 | -0.03574 | 0.08903 | -0.01436 | 0.06494 | -0.05432 |
| 44 | -0.23456 | -0.00117 | -0.29851 | -0.42687 | 0.27921 | -0.27408 | 0.29036 |

| Obs | _NAME_ | Dim1 | Dim2 | Dim3 | Dim4 | Dim6 | Dim7 | Dim8 | Dim9 | Dim10 | Dim11 |
|-----|--------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| 45 | GKDbon | 1.98537 | 0.20436 | 0.26363 | 0.38508 | -0.14101 | -0.27088 | -0.12210 | -0.12182 | -0.13246 | -0.00724 |
| 46 | GKDmoy | -0.31197 | -0.01779 | -0.11370 | -0.08808 | 0.06224 | 0.16661 | 0.03145 | 0.11389 | -0.25227 | 0.18766 |
| 47 | GKDnul | -0.33933 | -0.06644 | 0.11401 | -0.00513 | -0.06412 | -0.22670 | -0.00611 | -0.18769 | 0.62362 | -0.40924 |
| 48 | GKHbon | 1.92345 | 0.31785 | 0.26063 | 0.46696 | -0.15346 | -0.29027 | -0.13402 | -0.13104 | -0.14275 | 0.09866 |
| 49 | GKHmoy | -0.30903 | -0.11894 | -0.02678 | -0.02732 | -0.16194 | 0.03666 | 0.33854 | -0.13356 | 0.01492 | 0.09792 |
| 50 | GKHnul | -0.31367 | 0.09126 | -0.07433 | -0.17673 | 0.41843 | 0.06837 | -0.64649 | 0.34735 | 0.04018 | -0.25596 |
| 51 | GKKbon | 1.81557 | 0.34823 | 0.25752 | 0.44257 | -0.05358 | -0.39252 | -0.05351 | -0.19336 | -0.15533 | 0.06563 |
| 52 | GKKmoy | -0.31401 | -0.13499 | -0.04167 | 0.00325 | -0.06947 | 0.07777 | 0.15924 | 0.18446 | 0.23572 | -0.15662 |
| 53 | GKKnul | -0.31815 | 0.09009 | -0.05092 | -0.23883 | 0.16854 | 0.04880 | -0.29378 | -0.27137 | -0.39494 | 0.28212 |
| 54 | GKPbon | 2.01791 | 0.37463 | 0.18003 | 0.39303 | -0.09638 | -0.25727 | -0.13242 | -0.22596 | -0.07344 | 0.10685 |
| 55 | GKPmoy | -0.31373 | -0.07452 | -0.04847 | -0.16697 | -0.21922 | -0.09001 | 0.12402 | 0.13206 | 0.00304 | 0.04987 |

| Obs | Dim14 | Dim15 | Dim19 | Dim20 | Dim24 | Dim26 | Dim27 | Dim28 | Dim29 | Dim30 | Dim31 |
|-----|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| 45 | -0.09303 | -0.15079 | -0.17330 | -0.13756 | -0.02400 | 0.13057 | -0.08531 | -0.36314 | 0.04317 | 0.22918 | -0.02837 |
| 46 | 0.03764 | -0.03347 | 0.29258 | -0.11605 | 0.00714 | 0.06202 | 0.02265 | -0.01245 | -0.04809 | -0.12198 | 0.06410 |
| 47 | -0.03477 | 0.15158 | -0.55436 | 0.32647 | -0.00332 | -0.20397 | -0.00577 | 0.21505 | 0.08353 | 0.15003 | -0.12642 |
| 48 | 0.08983 | -0.12084 | 0.00620 | -0.16888 | -0.03566 | 0.13155 | -0.06988 | -0.16578 | 0.09401 | 0.22007 | -0.09578 |
| 49 | -0.02456 | -0.11467 | -0.09847 | -0.11031 | -0.01187 | -0.22450 | -0.00944 | 0.06077 | 0.02916 | -0.07621 | -0.03824 |
| 50 | 0.00670 | 0.30239 | 0.20450 | 0.31732 | 0.04293 | 0.40730 | 0.05497 | -0.04493 | -0.10865 | 0.05023 | 0.12868 |
| 51 | 0.01726 | -0.11183 | -0.21207 | -0.13782 | -0.15160 | 0.02800 | -0.07591 | -0.24087 | 0.05318 | 0.05656 | -0.06858 |
| 52 | -0.03240 | 0.06188 | 0.11585 | -0.13958 | -0.17844 | -0.15607 | 0.05961 | 0.02392 | -0.08599 | 0.01656 | 0.03283 |
| 53 | 0.05643 | -0.06638 | -0.12288 | 0.35445 | 0.44023 | 0.30076 | -0.08064 | 0.07806 | 0.14589 | -0.06315 | -0.03036 |
| 54 | 0.02522 | -0.11925 | -0.08164 | -0.09183 | 0.09085 | 0.04262 | -0.05051 | -0.15508 | 0.02172 | 0.20670 | -0.18951 |
| 55 | -0.05450 | -0.28736 | -0.23902 | 0.05960 | 0.07162 | -0.14746 | 0.05285 | 0.07171 | -0.03320 | -0.02146 | 0.08572 |

| Obs | Dim32 | Dim33 | Dim34 | Dim35 | Dim37 | Dim39 | Dim41 |
|-----|----------|----------|----------|----------|----------|----------|----------|
| 45 | -0.03281 | 0.15670 | 0.19095 | 0.49253 | 0.16510 | 0.12351 | 0.37971 |
| 46 | 0.03167 | -0.05490 | -0.06758 | -0.04158 | -0.02101 | -0.02294 | -0.05981 |
| 47 | -0.05274 | 0.03985 | 0.05005 | -0.16299 | -0.03908 | -0.01333 | -0.06457 |
| 48 | 0.11674 | 0.21139 | 0.00425 | 0.10502 | 0.18925 | -0.22329 | -0.29797 |
| 49 | 0.01434 | -0.05151 | -0.01548 | -0.03755 | -0.03729 | 0.06059 | 0.05487 |
| 50 | -0.08882 | 0.00252 | 0.03050 | 0.02648 | -0.01635 | -0.01570 | 0.03384 |
| 51 | -0.06690 | -0.09225 | 0.08358 | 0.55540 | 0.06707 | -0.42258 | -0.58815 |
| 52 | 0.03901 | 0.04860 | -0.00470 | -0.09926 | -0.04326 | 0.08846 | 0.10612 |
| 53 | -0.04373 | -0.04982 | -0.03436 | -0.09085 | 0.05223 | 0.04298 | 0.09417 |
| 54 | 0.00195 | 0.17685 | 0.25186 | 0.21635 | -0.08143 | 0.29285 | 0.61808 |
| 55 | -0.08240 | -0.01421 | 0.00965 | -0.03762 | 0.00112 | -0.06143 | -0.11404 |

| Obs | _NAME_ | Dim1 | Dim2 | Dim3 | Dim4 | Dim6 | Dim7 | Dim8 | Dim9 | Dim10 | Dim11 |
|-----|----------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| 56 | GKPnul | -0.32604 | -0.02548 | 0.01501 | 0.16446 | 0.51988 | 0.32151 | -0.20133 | -0.17221 | 0.02992 | -0.16042 |
| 57 | (15.971, 18.9] | -0.51824 | -0.76775 | 1.36027 | 0.84813 | 1.56820 | -0.28475 | 0.28680 | -0.55920 | 0.50723 | 0.66118 |
| 58 | (18.9, 21.8] | -0.06056 | -0.51130 | 0.73199 | -0.93786 | 0.81239 | 0.39439 | -0.37236 | 0.16360 | -0.32183 | -0.34137 |
| 59 | (21.8, 24.7] | -0.09843 | -0.20282 | 0.14170 | -0.15301 | -0.37014 | 0.02324 | 0.56276 | -0.38470 | 0.41123 | 0.10353 |
| 60 | (24.7, 27.6] | 0.02084 | 0.03339 | 0.12369 | 0.07938 | -0.16337 | 0.22900 | 0.05269 | 0.32178 | -0.30324 | 0.06872 |
| 61 | (27.6, 30.5] | -0.02481 | 0.23345 | -0.29768 | -0.08021 | -0.68560 | -0.43087 | -0.13225 | -0.44313 | -0.55129 | 0.00328 |
| 62 | (30.5, 33.4] | 0.02360 | 0.26903 | -0.57812 | 0.51416 | 0.62654 | 0.36830 | -0.02632 | 0.36317 | 0.50443 | 0.37164 |
| 63 | (33.4, 36.3] | 0.40382 | 0.32149 | -0.53390 | 0.75455 | 0.45480 | -1.11149 | -0.98233 | 0.48696 | 1.40314 | -1.15510 |
| 64 | (36.3, 39.2] | 1.03982 | 0.26134 | -1.63289 | -0.23353 | 4.32500 | -2.32267 | -3.16935 | 1.22388 | -2.13105 | 0.29320 |
| 65 | (39.2, 42.1] | 1.68784 | 1.76074 | 2.08871 | 5.95724 | 4.91680 | 5.65423 | 5.15816 | -3.46286 | 1.71339 | 1.90414 |

| Obs | Dim14 | Dim15 | Dim19 | Dim20 | Dim24 | Dim26 | Dim27 | Dim28 | Dim29 | Dim30 | Dim31 |
|-----|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| 56 | 0.10484 | 0.67796 | 0.55520 | -0.08276 | -0.19932 | 0.29637 | -0.08873 | -0.07744 | 0.06071 | -0.05640 | -0.09052 |
| 57 | 0.03827 | 4.18736 | -1.16022 | 0.34975 | 0.52951 | -0.86378 | 0.31491 | 0.09064 | -0.28495 | -0.02384 | -0.63914 |
| 58 | 1.09989 | -0.45415 | 0.08060 | 0.27965 | 0.47713 | -0.74324 | 0.11525 | -0.40942 | 0.12659 | 0.01963 | -0.61264 |
| 59 | 0.46927 | -0.08337 | 0.25422 | 0.42002 | 0.05457 | 0.53336 | 0.21305 | -0.11835 | -0.15301 | 0.09668 | -0.15593 |
| 60 | -0.73295 | 0.01683 | -0.12011 | -0.72821 | 0.11576 | -0.00485 | 0.04277 | 0.08519 | 0.06291 | -0.03232 | 0.09262 |
| 61 | -0.24285 | 0.24869 | 0.28552 | 0.34701 | -0.06572 | -0.13355 | -0.03333 | 0.02844 | 0.15941 | 0.05583 | 0.28471 |
| 62 | 0.18163 | -0.01250 | -0.68957 | -0.01500 | -0.25567 | 0.01775 | -0.55920 | 0.19292 | -0.33466 | 0.01619 | 0.15404 |
| 63 | 0.43833 | -0.90718 | 0.00630 | 0.29182 | -1.30686 | 0.27567 | -0.01385 | 0.24360 | 0.45288 | -0.31638 | 0.18694 |
| 64 | -3.34206 | -0.76416 | 2.70930 | -1.28962 | 1.65559 | -0.16684 | 0.71159 | 0.85739 | 1.36857 | -1.48059 | 0.01935 |
| 65 | -2.42884 | 0.45799 | 1.33670 | -3.09745 | 2.25992 | -0.80252 | -0.36550 | -1.25500 | -9.15199 | 0.16380 | 0.18098 |

| Obs | Dim32 | Dim33 | Dim34 | Dim35 | Dim37 | Dim39 | Dim41 |
|-----|----------|----------|----------|----------|----------|----------|----------|
| 56 | 0.17647 | -0.05719 | -0.14578 | -0.02638 | 0.03800 | -0.01307 | -0.06122 |
| 57 | -0.05981 | 0.17726 | 0.29201 | 0.37427 | -0.18559 | -0.02965 | -0.14235 |
| 58 | 0.23169 | 0.23617 | 0.13841 | 0.33994 | 0.15495 | -0.10143 | 0.08047 |
| 59 | 0.28001 | 0.11388 | -0.07238 | -0.01329 | -0.06372 | -0.05363 | -0.02319 |
| 60 | -0.04261 | 0.02460 | -0.02702 | -0.03386 | -0.04761 | -0.03866 | -0.03437 |
| 61 | -0.02416 | 0.05409 | -0.06585 | -0.16089 | -0.04215 | 0.06548 | -0.00388 |
| 62 | -0.36060 | -0.30312 | -0.05353 | -0.06515 | 0.04504 | 0.08375 | 0.04248 |
| 63 | -0.24591 | -0.43833 | 0.20086 | 0.06803 | 0.16129 | 0.12214 | 0.04911 |
| 64 | -0.99187 | -1.08880 | 1.53324 | 0.16539 | 0.75006 | 0.26496 | 0.08185 |
| 65 | -0.98303 | -0.64066 | 0.56150 | 1.39350 | 0.43637 | 0.65439 | -0.22168 |

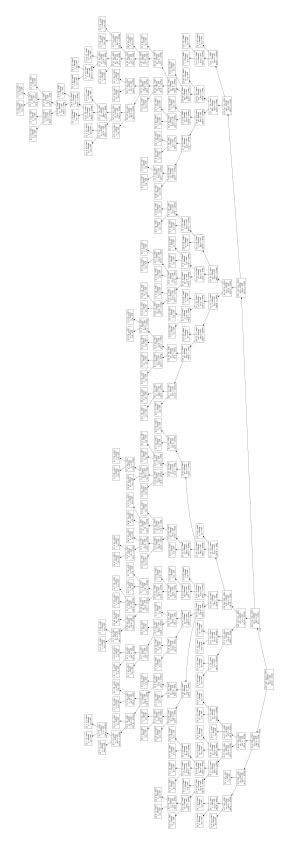


FIGURE 14 – Arbre de décision