

La réussite universitaire

Méthodes et outils d'analyse

Rabat, 14-16 avril 2004



VANDAMME Jean-Philippe

MESKENS Nadine

ARTIBA Abdelhakim



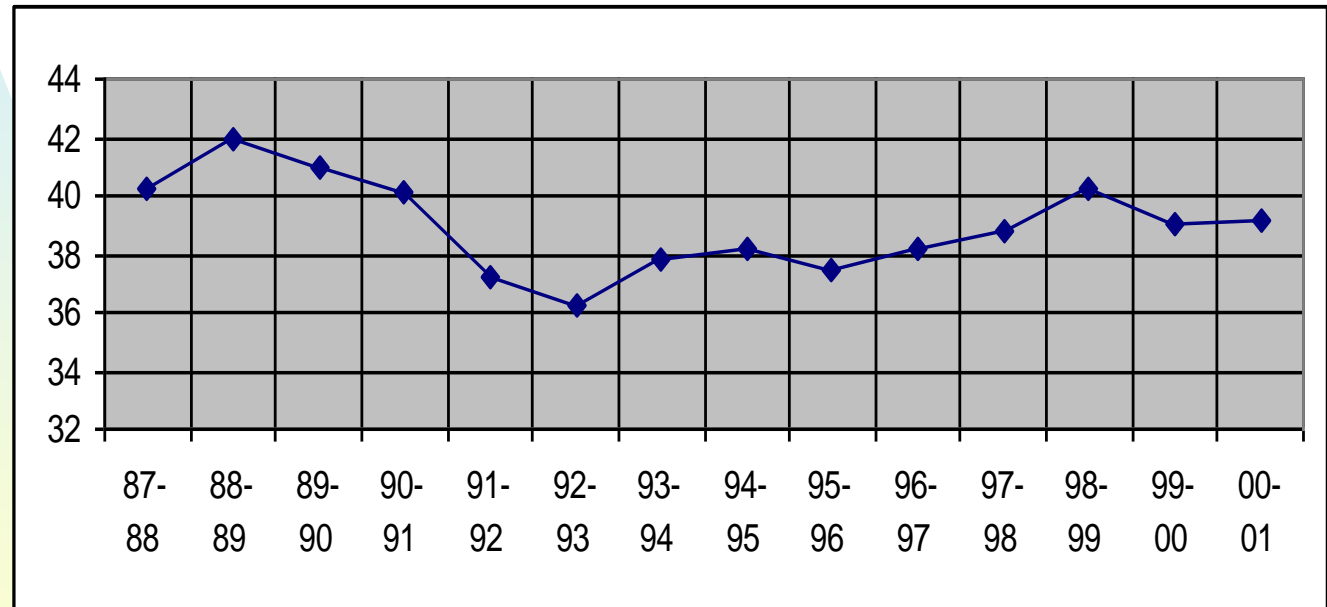
Plan de la présentation

- 1) Contexte de l'étude
- 2) Méthodologie de l'étude
- 3) Jeux de données
- 4) Méthodes utilisées
- 5) Résultats
- 6) Conclusions



Contexte de l'étude

➤ Réussite universitaire



➤ Prédiction

Identité et passé socio-familial
Raisons choix université/études
Passé scolaire

Pouvoir apprendre

Perceptions
Approches d'étude
Confiance en soi

Vouloir apprendre

Comportements
d'implication

**Performances
académiques
de l'étudiant**



Deux problèmes

➤ Données complètes non disponibles

➤ Choix de la méthode



2 phases à notre étude

Méthodologie

- Objectifs
- Données
- Prétraitements
- Méthodes
- Validation croisée
- Comparaison des résultats



Jeux de données

- Questionnaire
- Données supplémentaires
- 2 années
- 273 étudiants
- 143 variables
- Réussite ou échec



Prétraitements

? Valeurs manquantes

? Variables continues

? Variables inutiles

? Nombre de variables



5 jeux de données



Jeux de données

ACPtot

clusACP

clusBIN

INFLU

TABLE

- Analyse en composantes principales
- 143 variables continues
- Pas de valeurs manquantes



Jeux de données

ACPtot

clusACP

clusBIN

INFLU

TABLE

- Classification numérique
(algorithme de Ward, 5 groupes de variables, $R^2 \cong 89\%$)
- Analyse en composantes principales
(max 10 variables ou $R^2 \cong 80\%$)
- 32 variables continues
- Pas de valeurs manquantes



Jeux de données

ACPtot

clusACP

clusBIN

INFLU

TABLE

- Classification numérique
(algorithme de Ward, 10 groupes de variables, $R^2 \cong 92\%$)
- Sélection d'une variable par groupe
(sur base de la corrélation avec la variable de décision)
- 10 variables essentiellement binaires
- 35 valeurs manquantes



Jeux de données

ACPtot

clusACP

clusBIN

INFLU

TABLE

- Test du chi-carré
- 19 variables essentiellement binaires
- 17 valeurs manquantes



Jeux de données

ACPtot

- Données initiales sans prétraitement

clusACP

- 143 variables

clusBIN

- 164 valeurs manquantes

INFLU

TABLE



Les méthodes

- Les analyses discriminantes
 - ▣ linéaire sans sélection de variables
 - ▣ quadratique
 - ▣ plus proches voisins
 - ▣ noyau
 - ▣ linéaire avec sélection de variables
- Les régressions
- Les ensembles approximatifs

Les méthodes

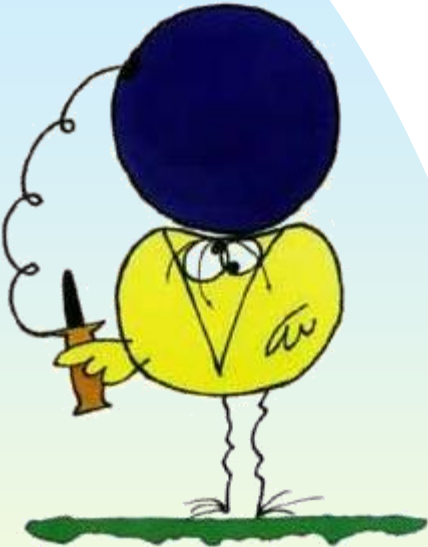
- Les analyses discriminantes
- Les régressions
 - ▣ multiple sans sélection de variables
 - ▣ multiple avec sélection de variables
 - ▣ logistique sans sélection de variables
 - ▣ logistique avec sélection de variables
- Les ensembles approximatifs

Les méthodes

- Les analyses discriminantes
- Les régressions
- Les ensembles approximatifs
 - sans les extensions
(modèle de PAWLAK [1991])
 - avec les extensions
(modèle de GRECO et al. [2001])

Validation

- 11 méthodes
- 5 jeux de données
- Ensembles d'apprentissage et de validation



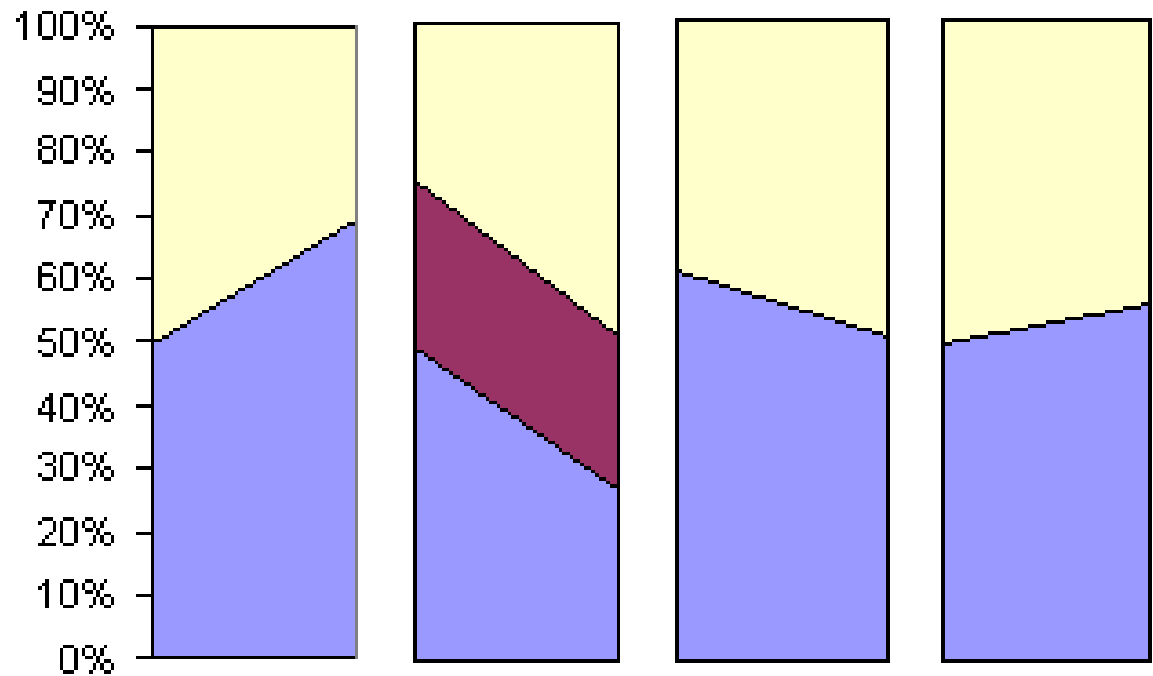
En essayant continuellement, on finit par réussir. Donc: plus ça rate, plus on a de chances que ça marche.

(Devise Shadok)

➤ **Aléatoirement**
Selon l'année académique

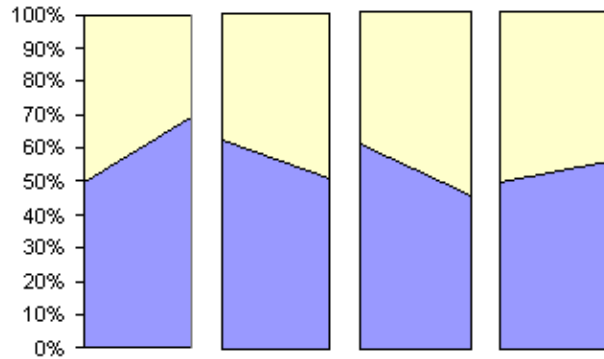


Résultats

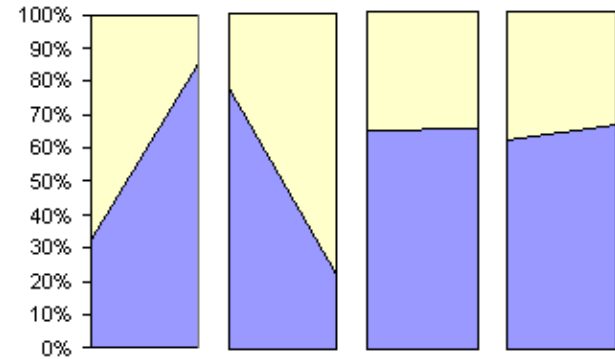


Meilleur jeu de données

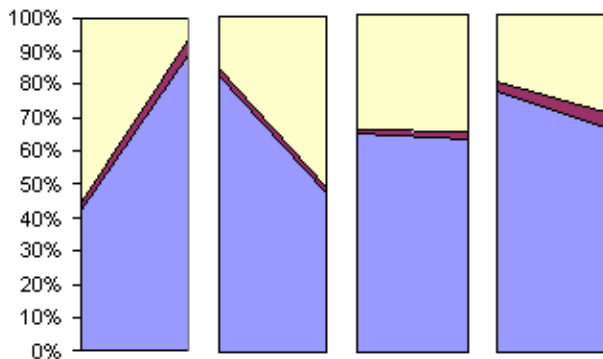
ACPtot



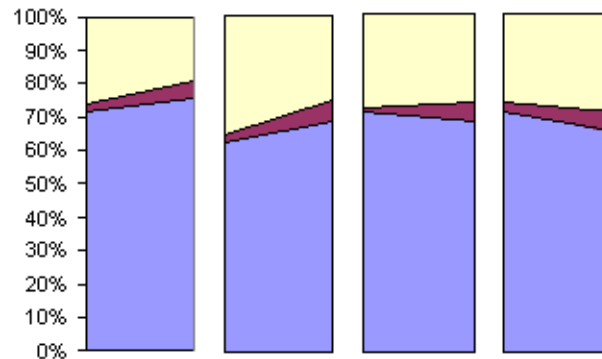
clusACP



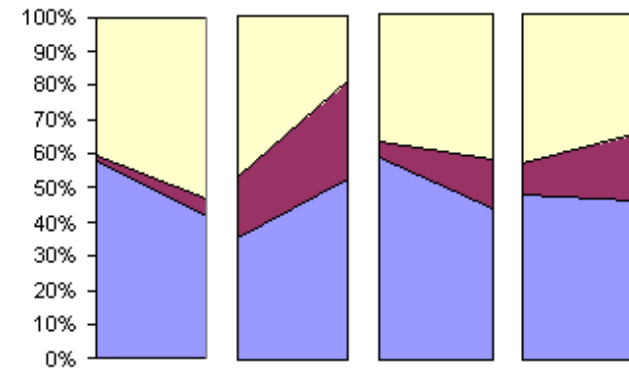
clusBIN



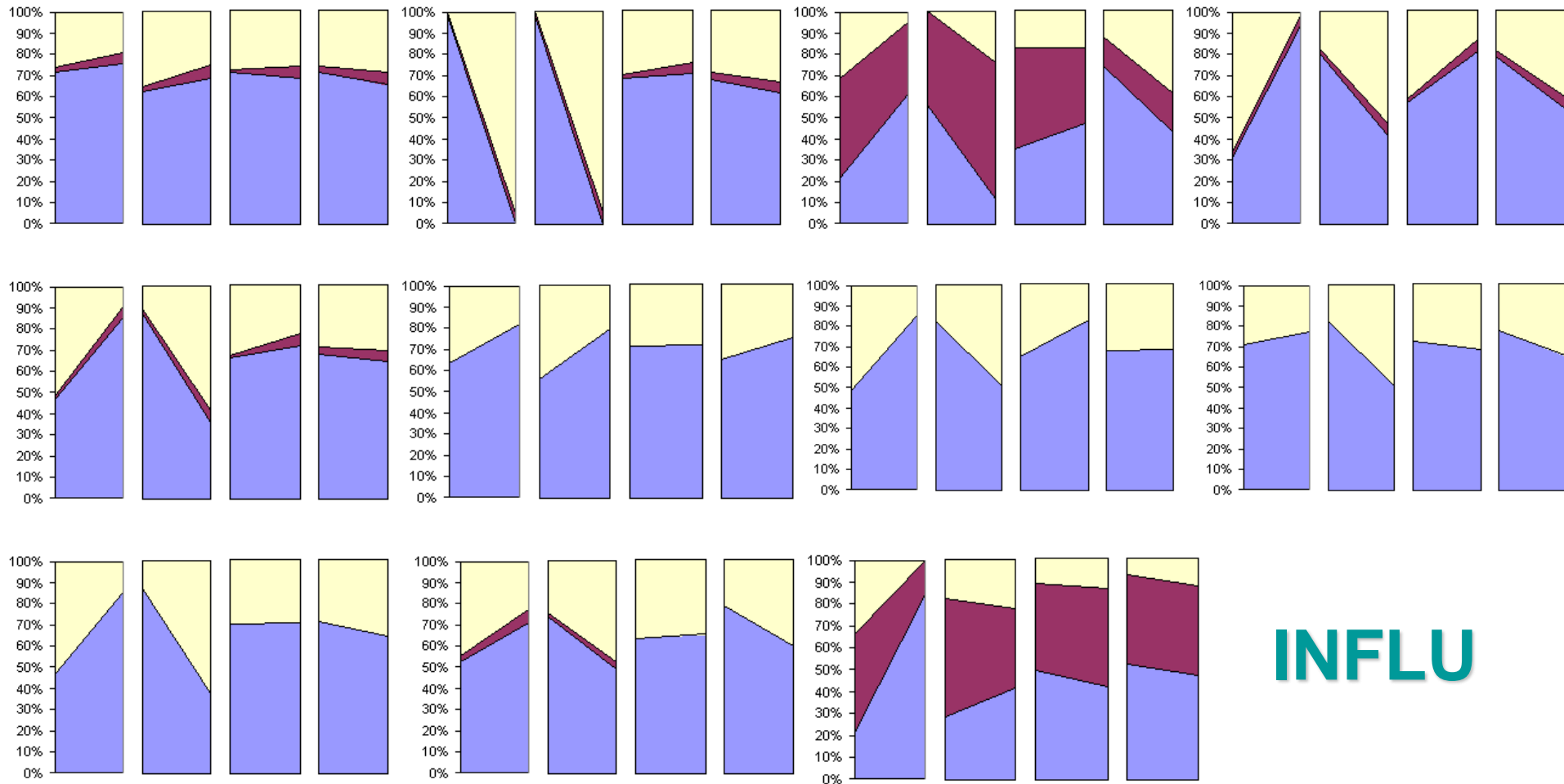
INFLU



TABLE



Meilleure méthode



Conclusions

- Analyse discriminante sans sélection de variables
- Jeu de données constitué sur base de tests du chi-carré
- Hétérogénéité des données préjudiciables



Perspectives

- ? Données plus complètes
- ? Données plus homogènes
- ? Données plus nombreuses
- ? Variable de décision
- ? Etudiants incorrectement classés
- ? Combinaison de méthodes
- ? Autres méthodes



Questions



S'il n'y a pas de solution, c'est qu'il n'y a pas de problème.

(Devise Shadok)

