



Projet de BBC NCI 60

Par Vincent Guidoux, Joël Kaufmann &
Nathan Flückiger



Sommaire

Introduction & Rappel du sujet

Utilisation de NCI 60 pour notre projet

Première approche Groupement des données

Travail effectué & Donnée Obtenue

Problème rencontré & Connaissances acquises

Aller Plus Loin

Questions



Rappel du sujet

Utilisation de la base de donnée NCI 60 pour la classification de cellule cancéreuse

Utilisation d'un modèle de MLG pour la classification de cellule cancéreuse



NCI 60 pour notre projet

Projet de recherche sur les cancers humains ayant pour but d'aider à tester les traitements aux cancer

Classification des types de cancers

- 9 types de cancers

- 60 lignées de cancers



Première approche Groupement des données

Regrouper par types (9 types) non concluant

5 catégories retenues:

- Adenocarcinoma : glandes

- Carcinoma : carcinomes

- Mélanomes (peau)

- Les différents types de leucémie

- Les cancers qui touchent le cerveau



Première approche Groupement des données

Validation des catégories en utilisant la variabilités des données

- Utilisation des premiers et derniers quartiles
- Pas de valeurs complètement aberrantes

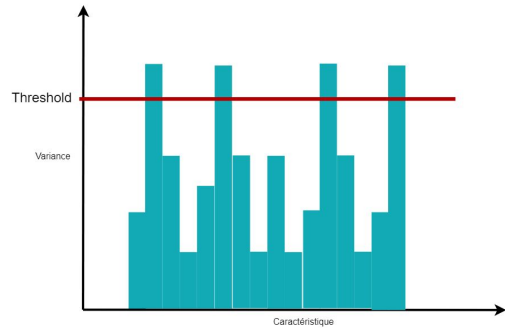
➡ aléatoire ou bonnes catégories?

Travail effectué & Données obtenues

Sélection des caractéristiques

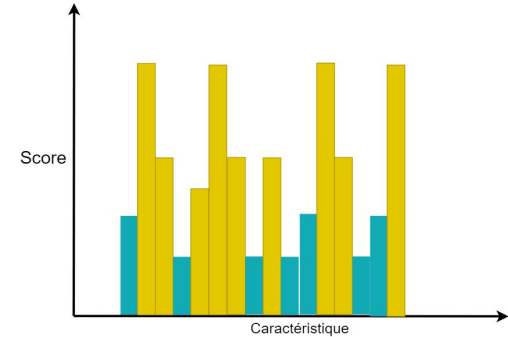
0.79

Seuil de Variance



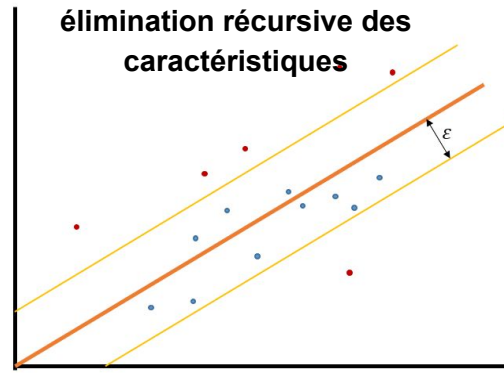
-60% des caractéristiques

Sélection univariée



0.868

élimination récursive des caractéristiques

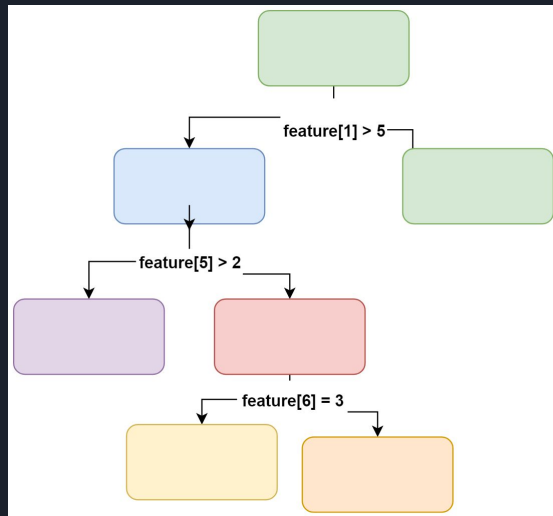


0.882

Travail effectué & Données obtenues

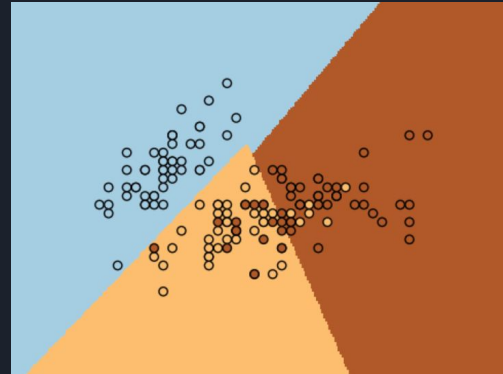
Classification

Arbres de décision



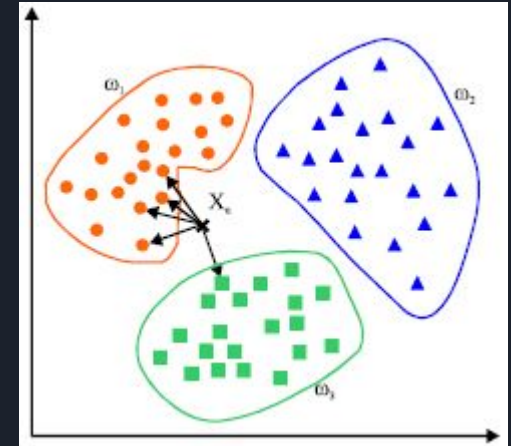
0.858

Support Vector Machine



0.868

Classification basée sur les voisins

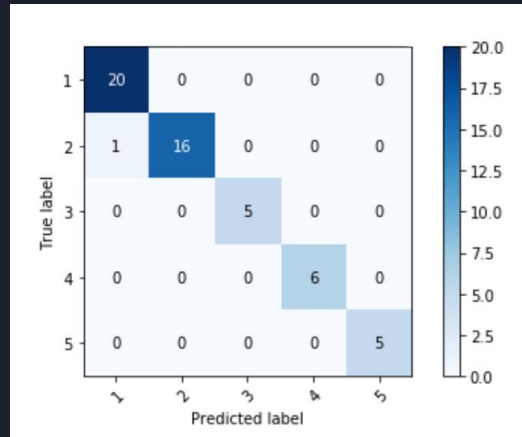


0.849

Travail effectué & Donnée Obtenue

Résultats

		Variance	Recursive features elimination			Unvariant Selection			
	Rien	Threshold	500	1000	2000	500	1000	2000	
Decision Tree	0.764	0.79	0.892	0.882	0.858	0.793	0.814	0.764	0.858
Support Vector Machine	0.604	0.849	0.981	0.981	0.962	0.811	0.849	0.868	0.8585
Knn	0.717	0.717	0.849	0.849	0.849	0.774	0.906	0.792	0.849
	0.6605	0.79	0.892	0.882	0.858	0.802	0.8775	0.868	





Problème rencontré & Connaissances acquises

Vérification des catégories

Python

Importance des caractéristiques

Pandas - Dataframe

Le set a-t-il de l'information ?

NCI60

Etats de l'art de notre sujet

Recherche de documents scientifiques



Aller plus loin

Utilisation d'autres échantillons et comparaison avec des cellules saines

Obtenir des profils types pour les différentes catégories

Chercher la cause de ces profils (expressions spécifiques des gènes) et des traitements possibles