

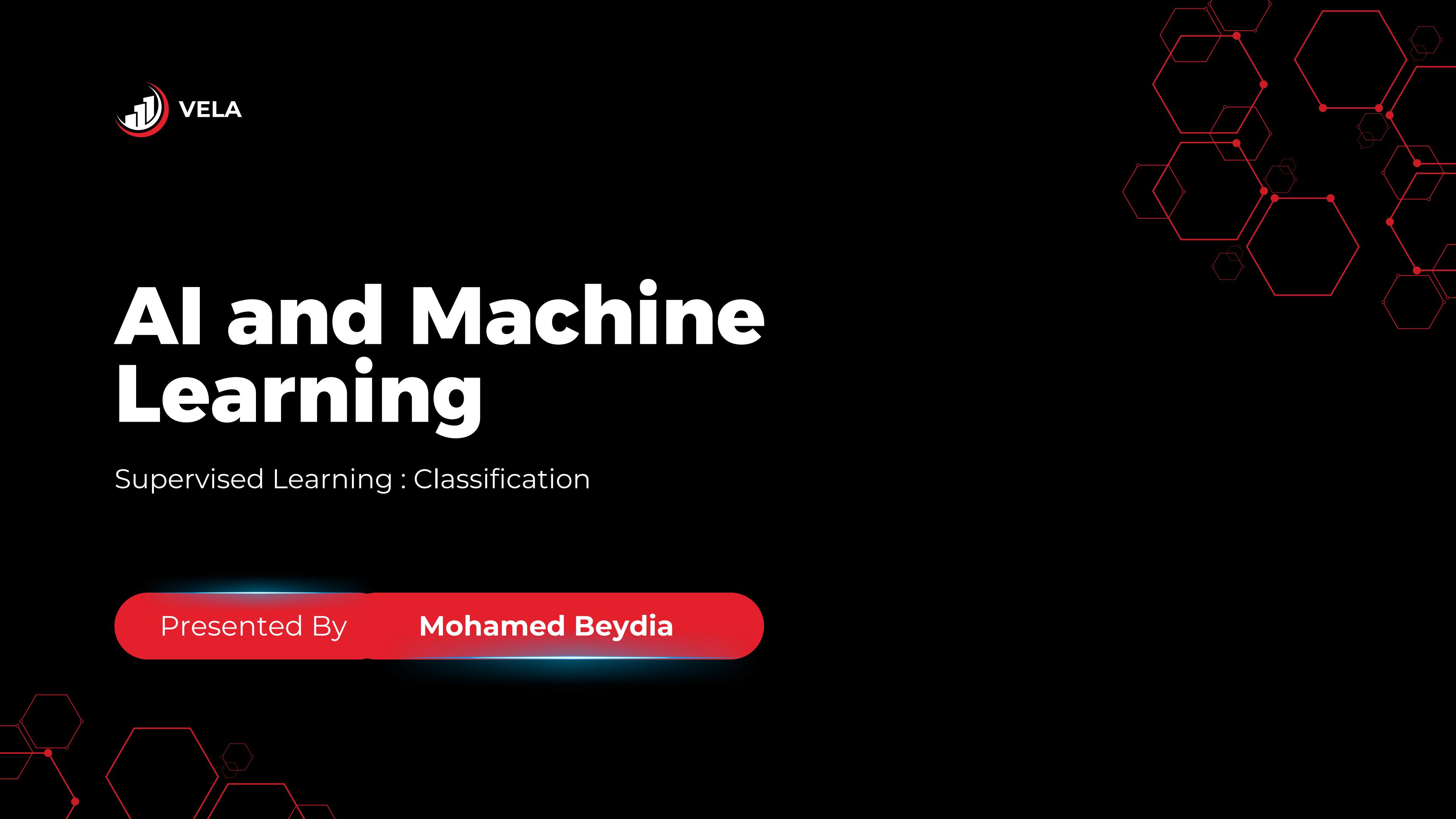


AI and Machine Learning

Supervised Learning : Classification

Presented By

Mohamed Beydia



- Machine Learning is using data to answer questions

Training

**using
data**

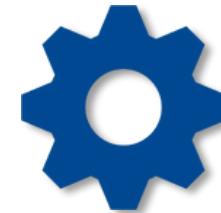
Prediction

**answer
questions**

Data



Training



Model



Predictions



**answer
questions**

- L'ensemble de données (échantillon) se compose d'exemples/instances (ligne dans un tableau).
- Les variables sont souvent appelées caractéristiques/attributs (colonne dans un tableau).
- Deux types :
 - **Variables catégorielles (discrètes)**
 - A uniquement des affinités ensemble de valeurs
 - Ordinal (élévé-moyen-faible, notes) ou
 - nominal (vrai-faux, couleur, profession)
 - **Variables numériques (continues)**
 - Contient des nombres réels comme valeurs (par exemple, température, taille, poids)
 - Ordonné, ne peut pas être énuméré facilement

Machine Learning Tasks –Dataset 1

Data about 860 recently deceased persons to study the effects of drinking, smoking, and body weight on the life expectancy

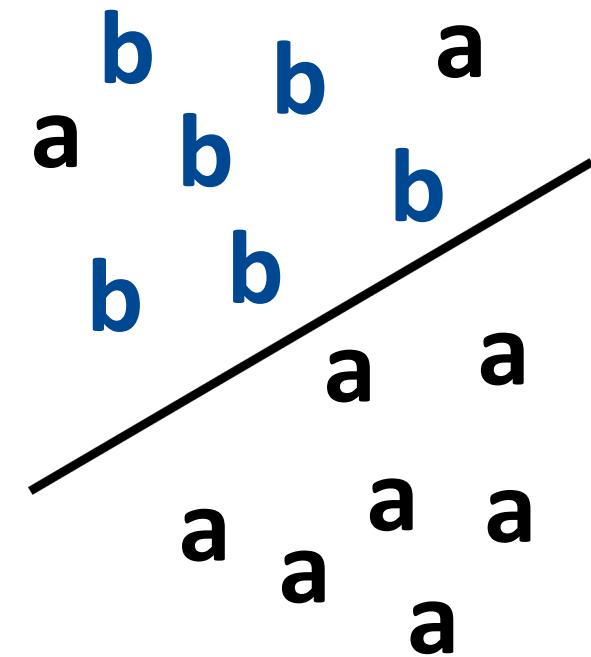
Drinker	Smoker	Weight	Age
Yes	Yes	120	44
No	No	70	96
Yes	No	72	88
Yes	Yes	55	52
No	Yes	94	56
No	No	62	93
...

- Exemple / Instance ?
- Caractéristiques / Attributs ?
- Données étiquetées / non étiquetées ?
- Variables discrètes / continues ?
- Variables nominales / ordinaires ?

Classification

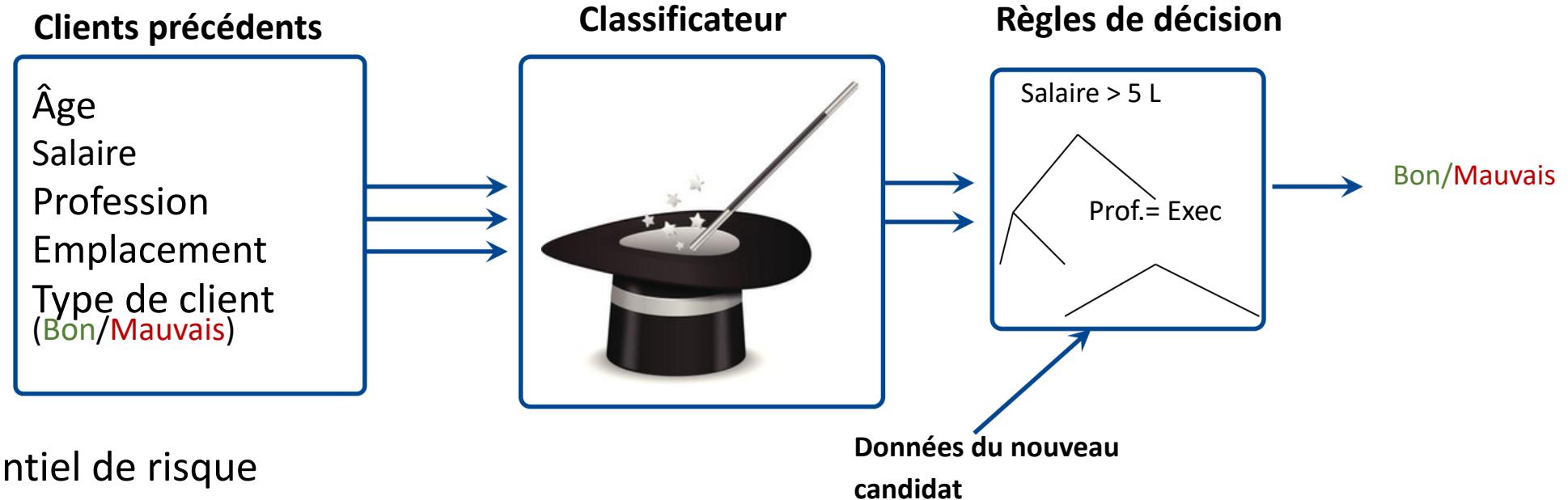
Classification

- Paramètre
 - Les étiquettes de classe sont connues pour un petit ensemble de « données d'entraînement »
- Tâche
 - Trouver des modèles/fonctions/règles qui peuvent
 - Décrire et distinguer les classes
 - Prédire l'appartenance à une classe pour les « nouveaux » objets
- Classification = apprentissage supervisé
 - Les Données d'entraînement contiennent des éléments étiquetés
 - Les nouvelles données sont classées en fonction des données d'entraînement.
 - Le classificateur prédit les étiquettes de classe.



Classification – Exemples

- Prédire l'éligibilité au prêt du nouveau demandeur



- Prédire le potentiel de risque

Données de formation			
ID	Âge	Type Voiture	Risk
1	23	Famille	Élevé
2	17	Sportive	Élevé
3	43	Sportive	Élevé
4	68	Famille	Faible
5	32	Camion	Faible

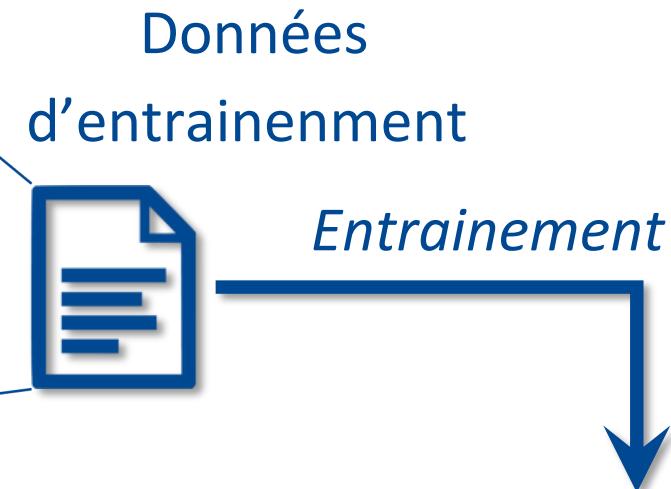
Classificateur simple

si l'âge > 50 alors le risque = faible ;
si l'âge ≤ 50 et le type de voiture = camion alors le risque = faible ;
si l'âge ≤ 50 et le type de voiture ≠ le camion alors le risque = élevé ;

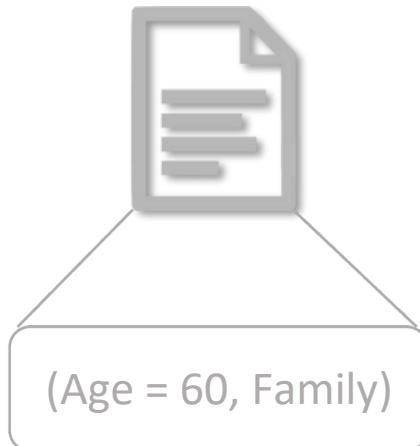
- Habituellement, l'ensemble de données donné est divisé en entraînement et de test
 - L'ensemble d'entraînement est utilisé pour entraîner le classificateur et construire le modèle
 - L'ensemble de tests est utilisé pour évaluer le classificateur
- Objectif : les données précédemment inédites doivent se voir attribuer une classe aussi précisément que possible
- Deux phases :
 - Phase d'entraînement/apprentissage (construction du modèle)
 - Phase de prédiction (inférence)

Classification – Phase D'entraînement

ID	Age	Type Voiture	Risk
1	23	Famille	Élevé
2	17	Sportive	Élevé
3	43	Sportive	Élevé
4	68	Famille	Faible
5	32	Camion	Faible



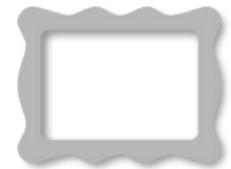
Données Inconnues



Classificateur



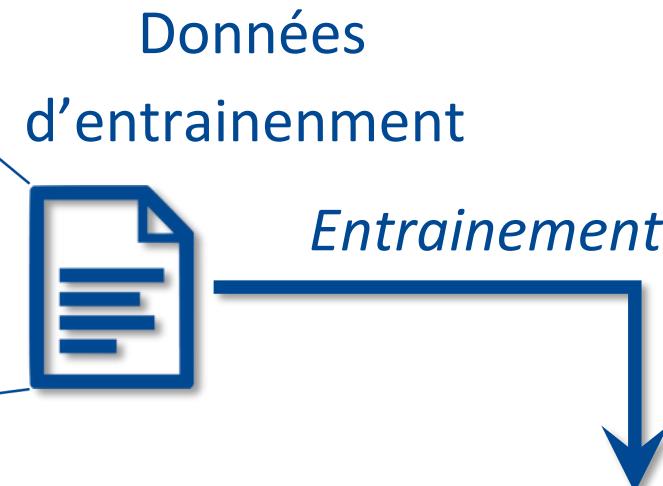
Class Label



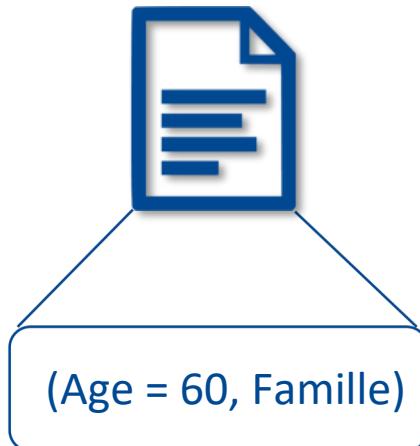
```
if Age > 50 then Risque = Low;  
if Age ≤ 50 and Type Voiture = Truck then Risque = Low;  
if Age ≤ 50 and Type Voiture ≠ Truck then Risque = High;
```

Classification – Phase de Prediction (Inference)

ID	Age	Type Voiture	Risk
1	23	Famille	Élevé
2	17	Sportive	Élevé
3	43	Sportive	Élevé
4	68	Famille	Faible
5	32	Camion	Faible



Données Inconnues

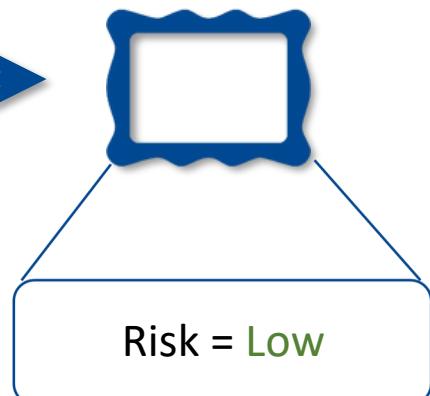


Classificateur



```
if Age > 50 then Risque = Low;  
if Age ≤ 50 and Type Voiture = Truck then Risque = Low;  
if Age ≤ 50 and Type Voiture ≠ Truck then Risque = High;
```

Class Label



- Principales méthodes de classification

- Classificateurs bayésiens
- Classificateurs d'arbre de décision
- Classificateurs du plus proche voisin
- Régression logistique
- Machines à vecteurs de support (SVM)
- Réseaux de neurones

Classificateur Bayésien

Classification

- Un cadre probabiliste pour résoudre les problèmes de classification
- Effectue une prédiction probabiliste, c'est-à-dire prédit les probabilités d'appartenance à une classe
- Fondement : basé sur le théorème de Bayes
- Performance: Un classificateur bayésien simple ; un classificateur bayésien naïf, a des performances comparables avec l'arbre de décision

- Théorie de Probabilité :

- Probabilité Conditionnelle $P(A|B) = \frac{P(A \wedge B)}{P(B)}$ (“probabilité de A sachant B”)
- Loi de Proba : $P(A \wedge B) = P(A|B) \cdot P(B)$

- Théorème de Bayes

- $P(A \wedge B) = P(A|B) \cdot P(B)$
- $P(B \wedge A) = P(B|A) \cdot P(A)$
- Since

$$P(A \wedge B) = P(B \wedge A) \Rightarrow$$
$$P(A|B) \cdot P(B) = P(B|A) \cdot P(A) \Rightarrow$$

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

Théorème de Bayes

Théorème de Bayes : Test Médical

ÉNONCÉ : Une maladie touche 1% de la population

Test de dépistage : 95% de précision (déetecte la maladie) + 5% de faux positifs

QUESTION : Si mon test est positif, quelle probabilité d'avoir la maladie ?

(Intuition naïve : 95% • Réalité avec Bayes : ?)

Population de 1000 personnes

MALADES
10 personnes
(1%)

SAINS
990 personnes (99%)



Résultats des tests :

VRAIS POSITIFS
9.5 personnes
(95% de 10)

FAUX NÉG 0.5

FAUX POSITIFS
49.5 personnes
(5% de 990)

VRAIS NÉGATIFS
940.5 personnes
(95% de 990)

TOTAL TESTS POSITIFS = 9.5 + 49.5 = 59 personnes

Probabilité d'être malade si test positif (Théorème de Bayes) :

$P(\text{Malade} | \text{Test}+) = \text{Vrais Positifs} / \text{Total Tests Positifs}$

$$P(\text{Malade} | \text{Test}+) = 9.5 / 59 \approx 16\%$$

Même avec un test positif, seulement 16% de chances d'être vraiment malade !

Théorème de Bayes en Classification

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)}$$

Définitions :

- **X** : un exemple de données ("evidence") dont la classe est inconnue
- **C** : une hypothèse selon laquelle X appartient à la classe C
- **Objectif** : déterminer $P(C|X)$, la **probabilité a posteriori**
→ probabilité que X appartienne à la classe C étant donné l'exemple observé X

P(C) (probabilité a priori) : probabilité initiale de la classe C

Ex : X achètera un ordinateur, indépendamment de l'âge, du revenu, etc.

Autres composants :

- **P(X)** : probabilité d'observer l'exemple (normalisation)
- **P(X|C) (vraisemblance)** : probabilité d'observer X sachant que l'hypothèse C est vraie
Ex : Sachant que X achètera un ordinateur, prob. que X ait 31-40 ans et un revenu moyen

Exemple concret : Classification d'achat d'ordinateur

Données observées (X) : Client âgé de 35 ans, revenu moyen, marié

Classes possibles (C) : "Achètera" ou "N'achètera pas"

Question : Quelle est $P(\text{"Achètera"} | \text{âge}=35, \text{revenu}=\text{moyen}, \text{marié})$?

→ Bayes utilise les probabilités a priori et les vraisemblances pour calculer cette probabilité a posteriori

Naïve Bayes Example

Outlook	Temp	Humidity	Wind	Play
Sunny	Hot	High	False	Yes
Sunny	Hot	High	False	No
Sunny	Hot	High	False	Yes
Sunny	Hot	High	False	No
Sunny	Mild	High	True	No
Overcast	Mild	High	False	Yes
Overcast	Mild	Normal	False	Yes
Overcast	Mild	Normal	False	Yes
Overcast	Mild	Normal	False	Yes
Rain	Cool	Normal	True	Yes
Rain	Cool	Normal	True	Yes
Rain	Mild	High	True	No
Rain	Cool	Normal	True	No
Rain	Cool	Normal	True	Yes

Naïve Bayes Example

Comment appliquer Bayes en pratique ? Prédisons si on peut "Jouer" dehors selon la météo !

Données d'entraînement : 14 exemples historiques avec 4 attributs météorologiques

Données d'Entraînement (Dataset)

Outlook	Temp	Humidity	Wind	Play
Sunny	Hot	High	False	Yes
Sunny	Hot	High	False	No

⋮⋮⋮⋮

(12 autres exemples)

Rain	Cool	Normal	True	Yes
------	------	--------	------	-----

Résumé des Données

- Total : 14 exemples
- Play = "Yes" : 9 cas (64%)
- Play = "No" : 5 cas (36%)

Valeurs possibles :

Outlook: Sunny, Overcast, Rain
Temp: Hot, Mild, Cool
Humidity: High, Normal

⌚ Nouveau Cas à Prédire

Outlook = Sunny, Temp = Cool, Humidity = High, Wind = True

Question : Peut-on jouer dehors ? (Play = Yes ou No ?)

Naïve Bayes va calculer $P(\text{Play}=\text{Yes}|\text{données})$ et $P(\text{Play}=\text{No}|\text{données})$ pour décider

⌚ Processus Naïve Bayes

- Probabilités a priori : $P(\text{Play}=\text{Yes}) = 9/14$, $P(\text{Play}=\text{No}) = 5/14$
- Vraisemblances : $P(\text{Outlook}=\text{Sunny}|\text{Play}=\text{Yes})$, $P(\text{Temp}=\text{Cool}|\text{Play}=\text{Yes})$, etc.
- Hypothèse "naïve" : Les attributs sont indépendants → multiplication des probabilités
- Classification : Choisir la classe avec la plus haute probabilité a posteriori

Formules du Classificateur Naïve Bayes

1. Théorème de Bayes (Base)

$$P(C|X) = P(X|C) \cdot P(C) / P(X)$$

2. Hypothèse "Naïve" d'Indépendance

Les attributs sont conditionnellement indépendants :

$$P(X|C) = P(x_1|C) \times P(x_2|C) \times P(x_3|C) \times \dots \times P(x_n|C)$$

3. Formule Naïve Bayes Complète

$$P(C|x_1, x_2, x_3, \dots, x_n) = P(C) \times \prod P(x_i|C) / P(x_1, x_2, x_3, \dots, x_n)$$

\prod = produit (multiplication) de tous les $P(x_i|C)$

En pratique, on ignore $P(X)$ car il est identique pour toutes les classes

4. Application à l'Exemple Météo

Pour : Outlook=Sunny, Temp=Cool, Humidity=High, Wind=True

$$P(\text{Play}=\text{Yes}|\text{données}) \propto P(\text{Play}=\text{Yes}) \times P(\text{Sunny}|\text{Yes}) \times P(\text{Cool}|\text{Yes}) \times P(\text{High}|\text{Yes}) \times P(\text{True}|\text{Yes})$$

$$P(\text{Play}=\text{No}|\text{données}) \propto P(\text{Play}=\text{No}) \times P(\text{Sunny}|\text{No}) \times P(\text{Cool}|\text{No}) \times P(\text{High}|\text{No}) \times P(\text{True}|\text{No})$$

proportionnel à signifie "proportionnel à" (on ignore le dénominateur commun)

5. Calcul des Probabilités (à partir des données)

Probabilités a priori : $P(\text{Play}=\text{Yes}) = 9/14 = 0.64$, $P(\text{Play}=\text{No}) = 5/14 = 0.36$

Vraisemblances : $P(\text{Sunny}|\text{Yes}) = (\text{nb de "Sunny" quand Play=Yes}) / (\text{nb total de Play=Yes})$

Décision : Choisir la classe avec la plus haute probabilité a posteriori

Decision Tree Classifiers

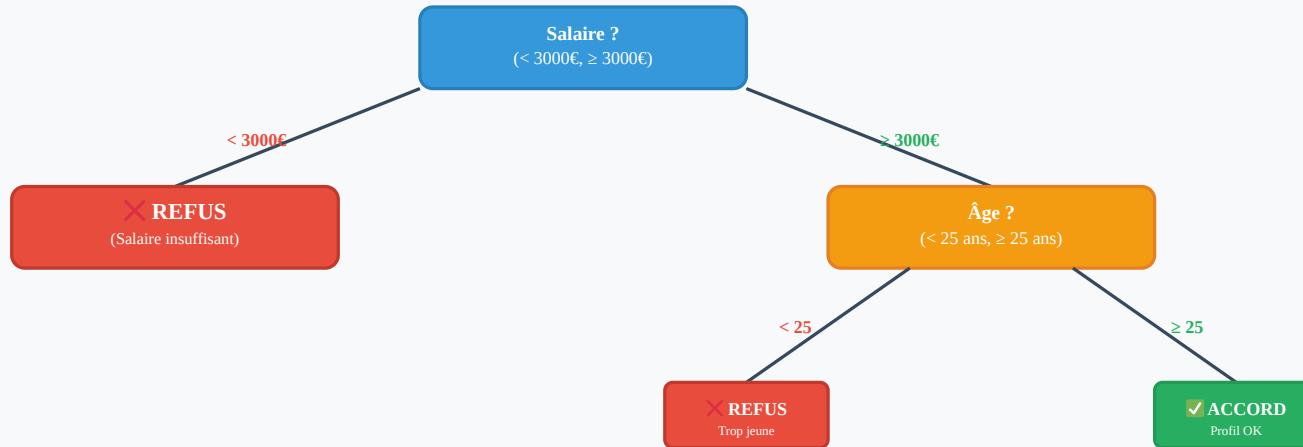
Classification

💡 INTUITION : Comment une banque décide-t-elle d'accorder un prêt ? Elle évalue le risque étape par étape !

Les arbres de décision imitent ce processus : une série de questions binaires pour arriver à une conclusion

"Si Salaire > 3000€ ? → Si Âge > 25 ? → Si Historique crédit bon ? → Accord ou Refus"

🏦 Exemple : Décision d'Octroi de Crédit Bancaire



🔑 Concepts Clés des Arbres de Décision

- **Nœud racine** : La première question (ici "Salaire?")
- **Nœuds internes** : Questions intermédiaires ("Âge?")
- **Feuilles** : Décisions finales ("Accord" ou "Refus")
- **Branches** : Réponses possibles à chaque question

⭐ Pourquoi les Arbres de Décision sont-ils Populaires ?

1. **Interprétabilité** : "Si... alors..." - logique claire et expliquable
2. **Pas de préparation** : Gère différents types de données (numériques, catégorielles)
3. **Sélection automatique** : L'algorithme choisit les meilleures questions à poser
4. **Non-linéaire** : Capture des relations complexes entre variables
5. **Rapide** : Prédiction = parcours de l'arbre (très efficace)

Arbre de Decision Intro

Comment l'Ordinateur Apprend à Poser les Bonnes Questions ?

💡 PROBLÈME : Face à 1000 questions possibles, laquelle choisir en premier ?

L'ordinateur va tester chaque question et garder celle qui classe le mieux nos clients !

🎯 Le Secret : Chercher les Groupes "Purs"

GROUPE PARFAIT ✓

Tous les riches → ACCORD

GROUPE MÉLANGÉ ✗

Moitié ACCORD, moitié REFUS

GROUPE PARFAIT ✓

Tous les pauvres → REFUS

📝 Comment Mesurer si un Groupe est "Pur" ?

L'ordinateur donne un "Score de Mélange" à chaque groupe

Groupe parfait :

Score = 0

(Rien n'est mélangé)

Groupe très mélangé :

Score = 1

(Mélange total)

Groupe un peu mélangé :

Score ≈ 0.9

(Un peu de mélange)

🏆 Le Concours : Quelle Question Nettoie le Mieux ?

Score de Nettoyage = Mélange AVANT - Mélange APRÈS

Plus le score est élevé, mieux la question sépare nos clients !
→ L'ordinateur garde la question qui a le meilleur score

📝 Exemple Concret : L'Ordinateur Hésite Entre Deux Questions

Au départ (20 clients mélangés) :

12 qui ont eu ACCORD, 8 qui ont eu REFUS
Score de mélange = 0.97 (assez mélangé !)

● Question "Salaire ≥ 3000€ ?"

- Groupe RICHES : 10 clients (9 ACCORD, 1 REFUS) → Score = 0.47
- Groupe PAUVRES : 10 clients (3 ACCORD, 7 REFUS) → Score = 0.88

Score de Nettoyage = 0.97 - 0.675 = 0.30 🌟

● Question "Âge ≥ 30 ans ?"

- Groupe VIEUX : 12 clients (8 ACCORD, 4 REFUS) → Score = 0.92
- Groupe JEUNES : 8 clients (4 ACCORD, 4 REFUS) → Score = 1.0

Score de Nettoyage = 0.97 - 0.95 = 0.02 🌟

🏆 GAGNANT : "Salaire" (Score 0.30) bat "Âge" (Score 0.02) !

L'ordinateur choisit "Salaire?" en premier car cette question sépare mieux nos clients !

Puis il recommence le même concours pour chaque sous-groupe jusqu'aux groupes parfaits

⌚ Récapitulatif : Comment l'Ordinateur Construit l'Arbre

- Tester toutes les questions possibles
- Garder celle qui nettoie le mieux
- Séparer les clients
- Recommencer pour chaque groupe
- Arrêter quand tous les groupes sont parfaits

Evaluation des modèles de classification

Erreur de classification

Correspond à la proportion des observations pour lesquelles les prédictions faites par le modèle sont erronées.

Erreur de classification

$$\frac{\text{nombre d'erreurs}}{\text{nombre d'obesrvations}}$$

Accuracy (taux de bonne classification)

$$\frac{\text{nombre de prédictions correctes}}{\text{nombre d'obesrvations}} = 1 - \text{Erreur de classification}$$

Types d'erreurs

Actual	Predicted
0	1
0	1
0	1
0	0
0	0
1	0
1	0
1	1
1	1
1	1

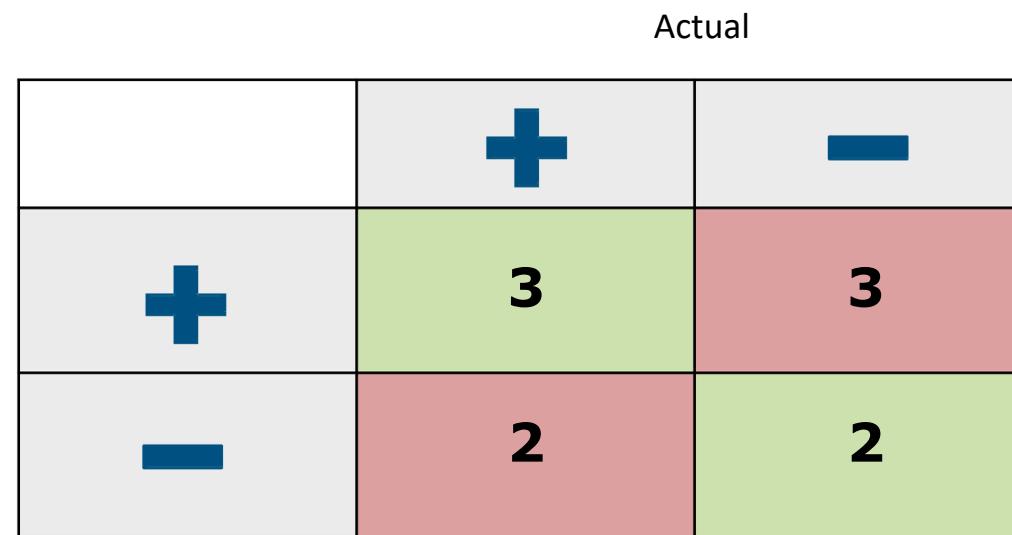
- le vrai label est $y = 1$ (dit positif) , alors que le modèle prédit $y_{\text{pred}} = 0$ (dit négatif)
- le vrai label est $y = 0$ (négatif), alors que le modèle prédit $y_{\text{pred}} = 1$ (positif)

- Actual : label réelle
- Predicted : label prédit par
le modèle de classification

Décision prise par le modèle de classification

Actual	Predicted
0	1
0	1
0	1
0	0
0	0
1	0
1	0
1	1
1	1
1	1

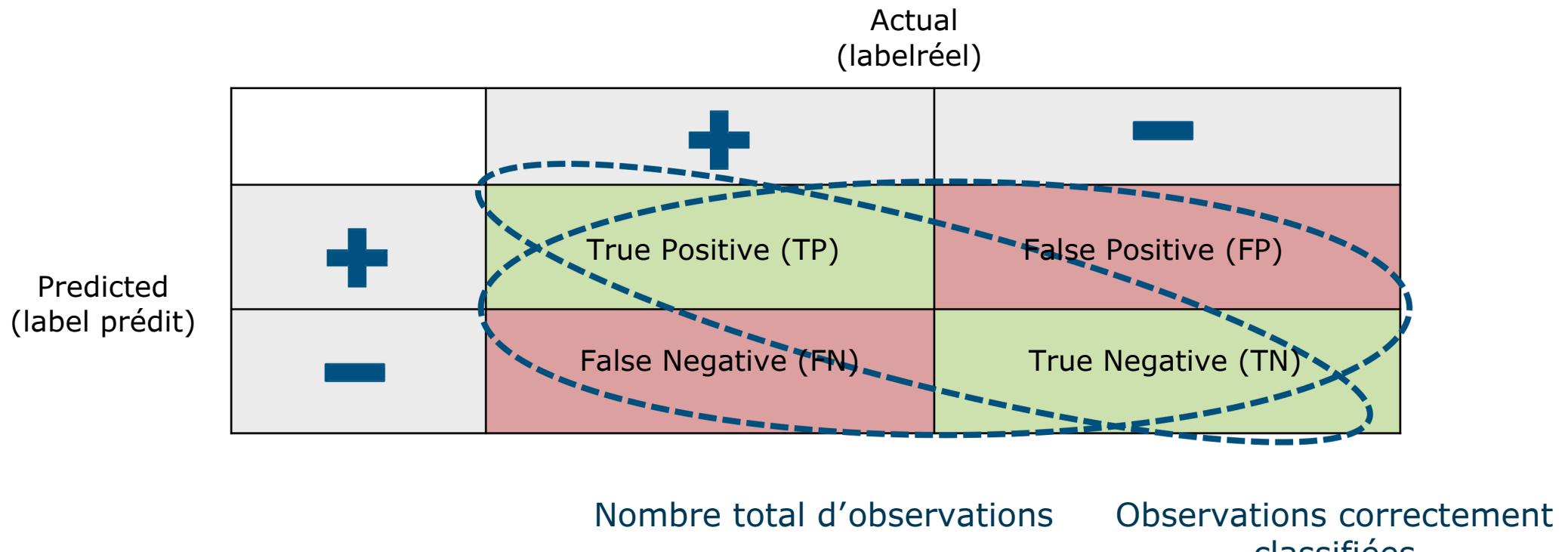
Predicted



Métriques d'évaluations

Accuracy

Quel est le taux de bonne classification ?



$$Accuracy = \frac{TP + TN}{TP+TN+FP+FN}$$

Détection des transaction fraudeuses

- La majorité des transactions correspondent à des transactions légales c.-à-d. non frauduleuses.
- Supposant que ces transaction représentent par exemple 90% des observations.
- Si on utilise un "Classificateur factice" qui prédit tout simplement la **classe majoritaire**.
- Ce modèle de classification ignore tout le temps les l'entrées et prédit la sortie comme non frauduleuse.

Actual	Predicted
LÉGALE	LÉGALE
FRAUDULEUSE	LÉGALE
LÉGALE	LÉGALE

$$\text{Accuracy} = \frac{9}{10} = 90\% !$$

Détection des transaction fraudeuses

- Ce mode assez simple peut avoir une très bonne accuracy
- Les données présentent un problème de déséquilibre de classe ou **class imbalance**.
- Il y a déséquilibre de classe lorsqu'une classe apparaît beaucoup plus fréquemment qu'une autre dans le dataset
- Cela pourrait suggérer que l' accuracy ne suffit pas toujours pour nous dire si le modèle de classification est un bon.

Métriques d'évaluations

Précision(Precision)

Parmi les instances que le modèle a prédit comme étant positifs, combien ont été correctement classifiées?

		Actual (label réel)
		+
Predicted (labelprédit)	+	True Positive (TP)
	-	False Negative (FN)
	-	True Negative (TN)

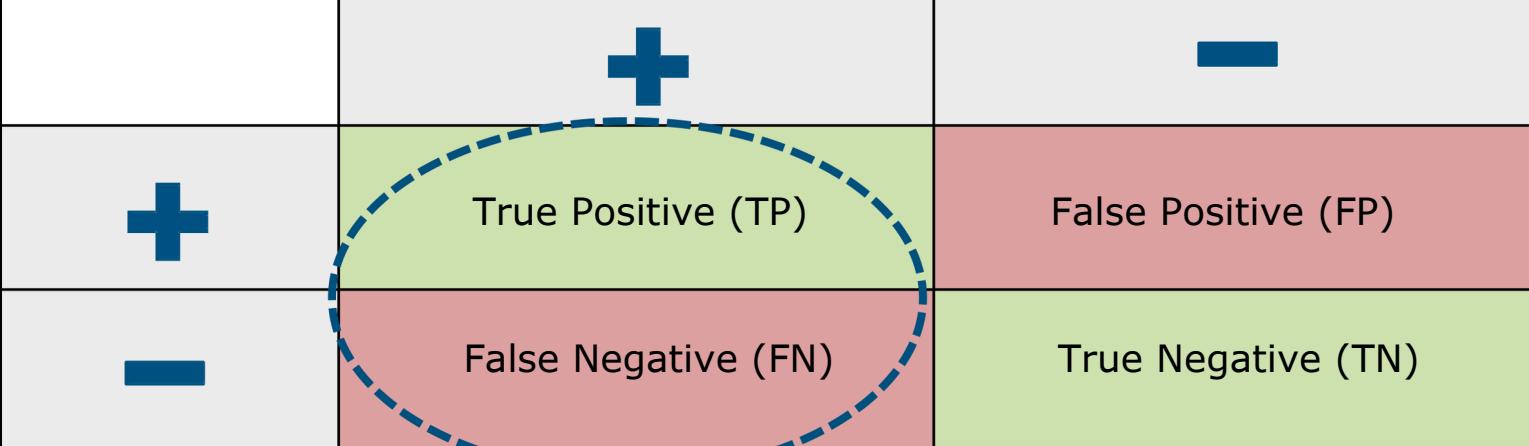
$$\text{Precision} = \frac{TP}{TP + FP}$$

Métriques d'évaluations

Rappel (Recall), sensibilité ou Taux des vrais positifs "

Parmi toutes les instances appartenant à la classe de type positif, combien ont été correctement classifiées ?

		Actual (label réel)
		+
Predicted (label prédict)	+	True Positive (TP)
	-	False Negative (FN)
	-	True Negative (TN)



$$\text{Recall} = \frac{TP}{TP + FN}$$

Capacité du classifieur à détecter toutes instances où la classe est de type 

Faux négatif ou faux positif

Qu'est-ce qui est plus mauvais : un faux négatif (False Negative) ou bien un faux positif (False positive)

Cela dépend entièrement du problème traité !

Détection de spams

- False Negative ↑ → Recall ↓ → *Capacité du classifieur à détecter les spams est faible*

Peut être gênant est fatigant car des emails qui sont réellement des spams sont classifiées comme non spam.

- False Positive ↑ → Precision ↓ → *Précision de classification est faible*

Beaucoup d'emails identifiés comme spam par le modèle , alors qu'ils ne le sont pas(emails perdus ou mal classés dans le dossier spam)

Which is Worse?

Qu'est-ce qui est plus mauvais : un faux négatif (False Negative) ou bien un faux positif (False positive)

Diagnostic médical

- False Negative ↑ → Recall ↓ → *Capacité du classifieur à détecter les cas (malades) est faible*
Maladie n'est pas traitée

- False Positive ↑ → Precision ↓ → *Précision de classification est faible*

Beaucoup de gens identifiés comme malade par le modèle , alors qu'ils ne le sont pas (gaspillage et traitement inutile)

Métriques d'évaluations

F-mesure (**F1score**)

Permet de prendre en considération à la fois le rappel et la précision en évaluant un compromis entre les deux, via le calcul de la moyenne harmonique:

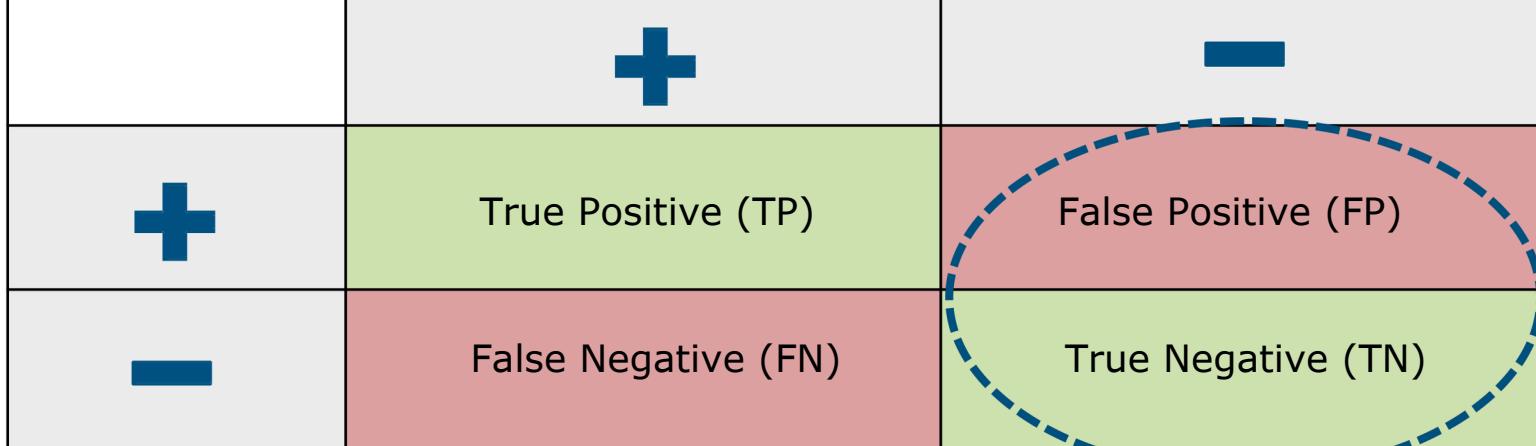
$$F - \text{mesure} = 2 \times \frac{\text{Précision} \times \text{Rappel}}{\text{Précision} + \text{Rappel}}$$

Métriques d'évaluations

Spécificité(Specificity) : taux des vrais négatifs

Parmi toutes les instances appartenant à la classe de type négatif , combien ont été correctement classifiées ?

		Actual (label réel)
		+
Predicted (label prédict)	+	True Positive (TP)
	-	False Negative (FN)
	-	True Negative (TN)



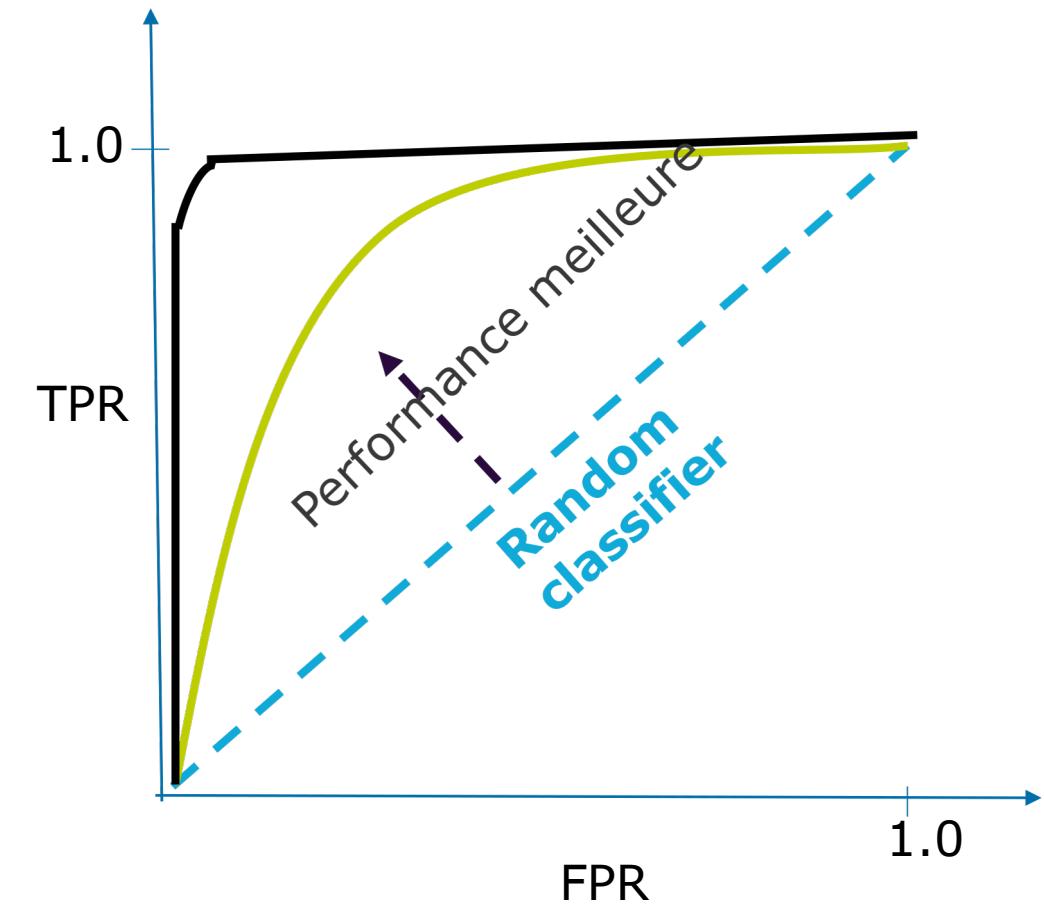
$$\text{Specificity} = \frac{TN}{FP + TN}$$

Capacité du classifieur à détecter toutes instances où la classe est de type -

Courbe ROC

Receiver Operating Characteristic

- Permet d'observer l'impact de la modification des seuils de classification sur les performances du modèle.
- Pour les modèles qui renvoient des probabilités pour les prédictions, le modèle est-il capable de faire la distinction entre les différents labels?



Pour chaque seuil de classification possible :

True Positive Rate

$$\text{TPR} = \frac{TP}{TP + FN}$$

Sensibilité

False Positive Rate

$$\text{FPR} = \frac{FP}{FP + TN}$$

1 - Specificité