



Choix de Modèle

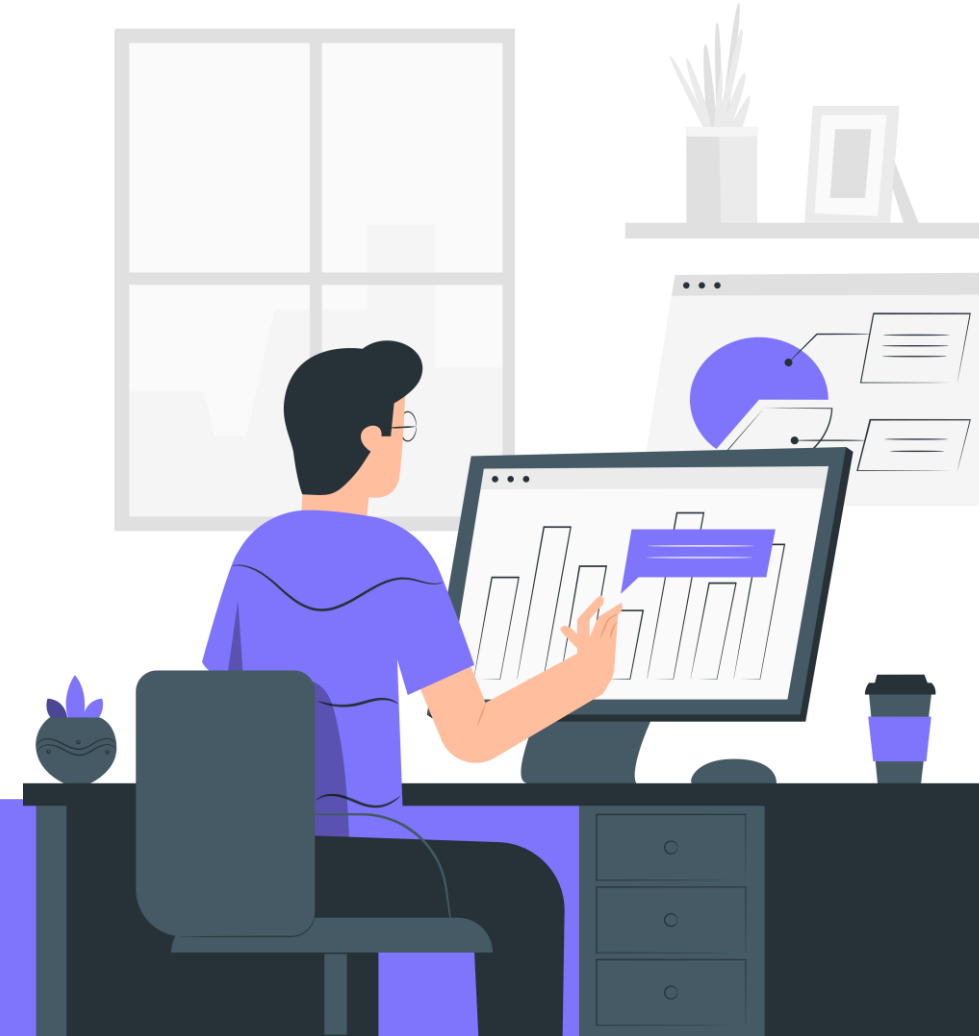
★ **Présenté par:**

MOKNIA Youssef

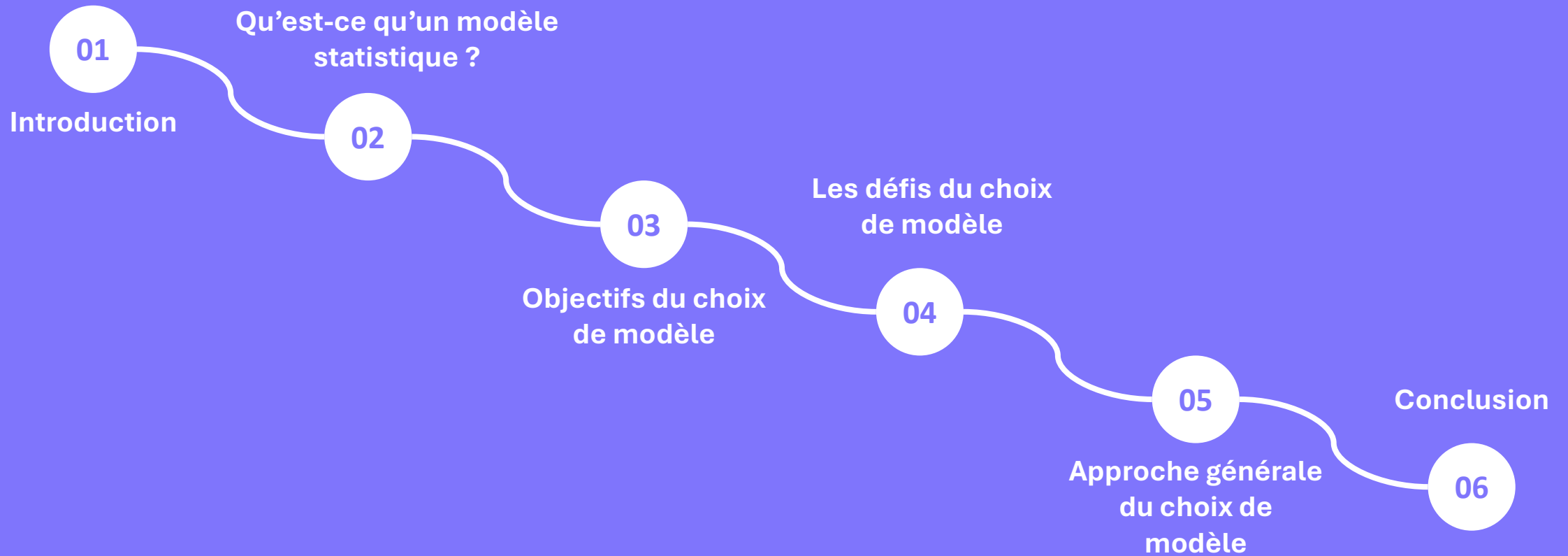
EL GOUDE Salma

★ **Encadré par:**

Mr. EL OUALI Mourad



PLAN



01

02

03

04

05

06

Introduction

Certains sont rapides
mais risqués



D'autres plus longs
mais sûrs

Introduction

Que choisiriez-vous ?

Et surtout, sur quels critères
baseriez-vous votre décision ?

Certains sont rapides
mais risqués

D'autres plus longs
mais sûrs



Introduction

Le choix d'un modèle statistique ou prédictif suit une logique similaire. Dans nos travaux en statistiques et en science des données, nous sommes confrontés à des données complexes et des objectifs variés, mais toujours avec une question centrale :

Quel modèle est le plus approprié ?

Ce choix n'est pas seulement technique, il a un impact direct sur :



La qualité de nos analyses



La précision de nos prédictions



La fiabilité des décisions que nous prenons

Qu'est-ce qu'un modèle statistique ?

- ★ **Un modèle statistique** est une tentative de simplifier le monde réel à l'aide de mathématiques. Il sert à représenter une relation entre des variables pour comprendre un phénomène ou faire des prédictions.
- ★ Lorsque plusieurs prédicteurs sont disponibles, différentes **combinaisons** de ces variables génèrent plusieurs modèles potentiels. **Par exemple** : Avec 3 prédicteurs(x_1, x_2, x_3), il y a $2^3 = 8$ modèles possibles. À mesure que le nombre de prédicteurs augmente (2^k), le nombre de modèles explose, rendant **le choix du modèle optimal** difficile.
- ★ Le véritable défi est alors de choisir **le modèle le plus adapté parmi plusieurs candidats**. Faut-il un modèle très simple pour aller à l'essentiel ? Ou un modèle plus complexe pour capturer toutes les nuances ? C'est là que le choix de modèle entre en jeu.

01

02

03

04

05

06

Objectifs du choix de modèle

Le choix d'un modèle n'est pas une démarche aléatoire. Il repose sur deux objectifs majeurs, chacun ayant des implications profondes sur la qualité et l'utilité de notre analyse.

1. Identifier les relations sous-jacentes

2. Prédire avec précision

01

02

03

04

05

06

Objectifs du choix de modèle

Le choix d'un modèle n'est pas une démarche aléatoire. Il repose sur deux objectifs majeurs, chacun ayant des implications profondes sur la qualité et l'utilité de notre analyse.

1. Identifier les relations sous-jacentes

L'un des principaux rôles d'un modèle est de révéler les relations sous-jacentes entre les variables. **Par exemple**, en prédisant les ventes d'un produit, le modèle cherche à **comprendre** comment le prix, la saisonnalité et d'autres facteurs influencent les ventes. Il ne s'agit pas seulement de s'ajuster aux données, mais de donner un sens aux phénomènes que nous observons.

2. Prédire avec précision

01

02

03

04

05

06

Objectifs du choix de modèle

Le choix d'un modèle n'est pas une démarche aléatoire. Il repose sur deux objectifs majeurs, chacun ayant des implications profondes sur la qualité et l'utilité de notre analyse.

1. Identifier les relations sous-jacentes

2. Prédire avec précision

Le modèle doit être capable de généraliser, c'est-à-dire de faire des **prédictions** fiables sur de nouvelles données. Un bon modèle est celui qui, après avoir appris sur un ensemble de données, reste performant sur des situations non observées, qu'il s'agisse de prévisions économiques, de comportements d'acheteurs, ou même de tendances futures.

Les défis du choix de modèle

Un **modèle sous-spécifié** constitue un défi important, car il ne prend pas en compte toutes les variables pertinentes pouvant affecter la variable dépendante. Cela signifie que le modèle est incomplet, ce qui entraîne un manque d'informations cruciales pour réaliser une analyse précise. En effet, lorsque des variables pertinentes sont omises, deux types d'erreurs peuvent survenir :

- ★ **Biais dans les réponses prédites** : mesure l'erreur systématique qu'un modèle commet lorsqu'il prédit une valeur, notamment lorsqu'il est trop simple ou sous-spécifié, produisant ainsi des erreurs constantes.
- ★ **Variation dans les réponses prédites** : indique à quel point les prédictions changent avec de nouvelles données. Un modèle trop complexe peut présenter une grande variation, ce qui peut mener à un surapprentissage, où le modèle devient trop sensible aux détails spécifiques des données d'entraînement et perd sa capacité à généraliser.

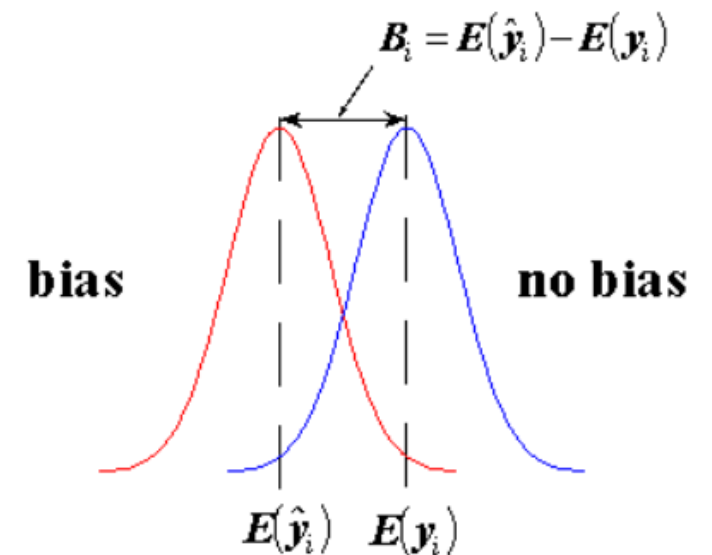
Biais dans les réponses prédites :

Le biais se produit lorsqu'il existe une différence systématique entre les valeurs réelles observées $E(y_i)$ et les valeurs prédites $E(\hat{y}_i)$ où \hat{y}_i est la prédiction faite par le modèle. Un modèle mal spécifié peut produire des prédictions biaisées.

- ★ **Sans Biais** : Si le modèle est bien spécifié, les valeurs prédites $E(\hat{y}_i)$ sont très proches de la vraie valeur moyenne $E(y_i)$. Un bon modèle vise à atteindre cet objectif.
- ★ **Avec Biais** : Si le modèle omet des informations cruciales ou fait de fausses hypothèses, il produira des prédictions incorrectes, créant une différence entre $E(\hat{y}_i)$ et $E(y_i)$ ce qui constitue un biais.

★ Formule du Biais :

$$B_i = E(\hat{y}_i) - E(y_i)$$



Variation des Réponses Prédites:

Lorsqu'on effectue des prédictions à l'aide d'un modèle de régression, il est essentiel de comprendre les différentes sources de variation susceptibles de les influencer. Ces variations se divisent en deux catégories principales :

- ★ **Variation Aléatoire** : La variation aléatoire est une variation naturelle qui se produit lorsque nous faisons des prédictions.
- ★ **Variation Due au Biais** : Si un modèle présente un biais, cela signifie que les prédictions qu'il fournit sont, en moyenne, incorrectes.
- ★ **la variation totale des prédictions** : Combinaison des Deux.

- **Avec biais :**

$$\Gamma_p = \frac{1}{\sigma^2} \left(\sum_{i=1}^n \sigma_{y^i}^2 + \sum_{i=1}^n [E(y^i) - E(y_i)]^2 \right)$$

- **Sans biais :**

$$\Gamma_p = \frac{1}{\sigma^2} \left\{ \sum_{i=1}^n \sigma_{y^i}^2 + 0 \right\} = k + 1$$

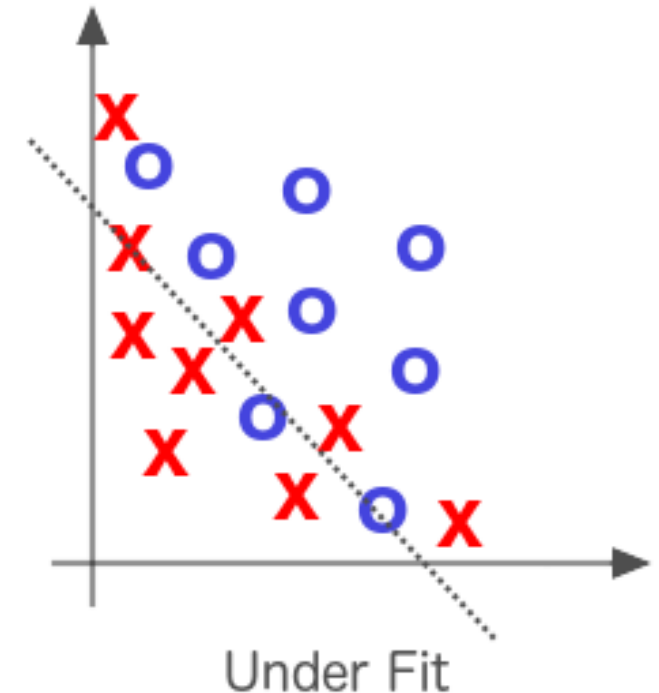
Les défis du choix de modèle

Maintenant que nous comprenons le **compromis entre biais et variance**, examinons comment ces deux phénomènes se manifestent concrètement dans les modèles à travers deux concepts importants :

Sous-ajustement (Underfitting):

Le sous-ajustement survient lorsque le modèle est trop simple pour capturer les relations complexes dans les données. Il en résulte un modèle qui présente :

- ★ **Biais élevé** : Le modèle échoue à identifier les tendances importantes.
- ★ **Faible performance** : Le modèle ne parvient pas à bien prédire, ni sur les données d'entraînement, ni sur de nouvelles données.



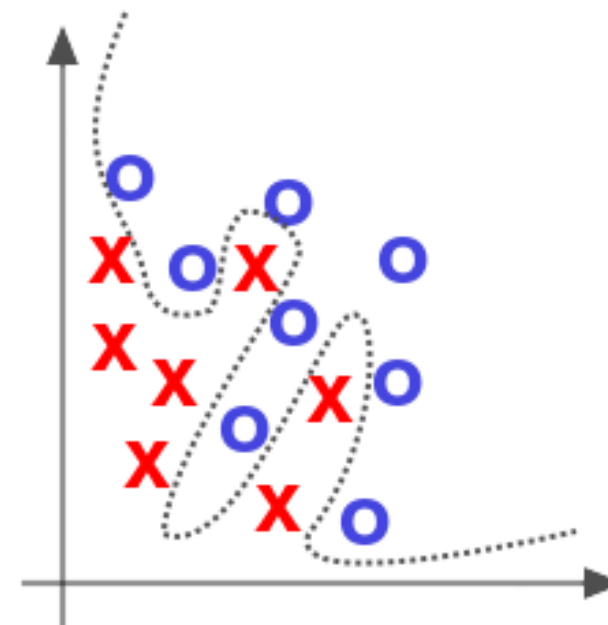
Les défis du choix de modèle

Maintenant que nous comprenons le **compromis entre biais et variance**, examinons comment ces deux phénomènes se manifestent concrètement dans les modèles à travers deux concepts importants :

Sur-ajustement (Overfitting):

Le sur-ajustement survient lorsque le modèle devient trop complexe et s'adapte parfaitement aux données d'entraînement, y compris au bruit et aux anomalies. Cela se traduit par un modèle avec :

- ★ **Variance élevée** : Le modèle est trop sensible aux spécificités des données d'entraînement.
- ★ **Mauvaise généralisation** : Bien que le modèle fonctionne bien sur les données d'entraînement, il échoue à prédire correctement sur de nouvelles données.



Over Fit

01

02

03

04

05

06

Approche générale du choix de modèle

Le choix d'un modèle repose sur une **approche** structurée qui répond à trois principes clés :

Parcimonie

Évaluation de la performance

Adaptation au contexte

- Un bon modèle doit être **aussi simple que possible**, tout en capturant les relations essentielles des données.
- La simplicité réduit les risques de **sur-ajustement** et facilite l'interprétation des résultats.

01

02

03

04

05

06

Approche générale du choix de modèle

Le choix d'un modèle repose sur une **approche** structurée qui répond à trois principes clés :

Parcimonie

Évaluation de la performance

Adaptation au contexte

- Mesurer la capacité du modèle à généraliser grâce à des outils comme la **validation croisée** et des critères tels que l'**AIC**, le **BIC** ou l'**erreur quadratique moyenne (MSE)**.
- **L'objectif** est de garantir des prédictions fiables sur des données nouvelles.

01

02

03

04

05

06

Approche générale du choix de modèle

Le choix d'un modèle repose sur une **approche** structurée qui répond à trois principes clés :

Parcimonie

Évaluation de la performance

Adaptation au contexte

- Le modèle doit répondre aux besoins spécifiques : interprétabilité, rapidité de calcul ou adéquation aux données disponibles.

01

02

03

04

05

06

Conclusion

En résumé, le choix de modèle est un processus délicat qui implique de trouver un équilibre entre **simplicité**, **performance** et **adaptation au contexte**. Bien que nous ayons abordé les principes généraux, il existe diverses méthodes spécifiques pour affiner ce choix.



Merci Pour Votre Attention

MOKNIA Youssef • EL GOUDE Salma