# Méthode Générale

Compte tenu de votre objectif de prédire si un étudiant est de haut niveau ou de bas niveau à la fin de l'année en prenant en compte la sériation et les deux périodes, voici comment vous pourriez procéder pour créer une colonne cible appropriée :

Calcul du niveau global : Commencez par calculer le niveau global de l'étudiant en utilisant les performances des deux périodes d'études. Vous pouvez utiliser une métrique de performance agrégée, comme la moyenne des notes des deux périodes.

Classification binaire : Ensuite, définissant une règle ou un seuil pour diviser les étudiants en deux catégories : "haut niveau" et "bas niveau". Par exemple, si vous considérez que les étudiants ayant un niveau global supérieur à un certain score sont "haut niveau" et les autres sont "bas niveau", vous pouvez utiliser cette règle pour attribuer une étiquette binaire à chaque étudiant.

Considération de la sériation : Tenez compte de la sériation dans votre modèle en tant que caractéristique ou variable d'entrée. La sériation peut être une catégorie variable, et vous pouvez l'encoder de manière appropriée, par exemple en utilisant l'encodage one-hot.

Entraînement du modèle : Entraînez un modèle de machine learning, tel qu'une régression logistique ou un modèle d'arbre de décision, en utilisant les performances globales calculées, les données sur la sériation et d'autres caractéristiques pertinentes que vous pourriez avoir.

Évaluation du modèle : Évaluez la performance de votre modèle en utilisant des métriques de classification binaire, telles que l'exactitude (accuracy), la précision (precision), le rappel (recall) et le score F1, sur un ensemble de test distinct.

Réglage du seuil : Si nécessaire, ajustez le seuil de classification en fonction de vos objectifs et des performances du modèle. Un seuil plus élevé rendra la classification "haut niveau" plus stricte, tandis qu'un seuil plus bas sera plus inclusif.

Cette approche prend en compte à la fois les performances globales des étudiants sur les deux périodes d'études et la série en tant que caractéristique importante pour la classification. Elle vous permet de classer les étudiants en "haut niveau" ou "bas niveau" en fonction de votre seuil défini tout en prenant en compte leur parcours académique. Assurez-vous de choisir judicieusement votre seuil en fonction de vos objectifs spécifiques et des performances du modèle.

## Plus de catégories de classification

Si vous avez besoin de classifier les étudiants en cinq catégories différentes (faible, insuffisant, moyen, correct et bien) en fonction de leurs performances, de la sériation et des deux périodes d'études, voici comment vous pourriez procéder :

Calcul du niveau global : Comme précédemment, commencez par calculer le niveau global de l'étudiant en utilisant les performances des deux périodes d'études, par exemple en calculant la moyenne des notes des deux périodes.

Classification en cinq catégories : Ensuite, défini des plages ou des seuils pour attribuer chaque étudiant à l'une des cinq catégories. Par exemple, vous pourriez définir des plages de scores, par exemple 0-40 pour "faible", 41-60 pour "insuffisant", 61-80 pour "moyen", 81-90 pour "correct" et 91-100 pour " bien". En fonction du score global de chaque étudiant, attribuez-lui l'une de ces cinq catégories.

Considération de la sériation : Comme précédemment, incluez la sériation en tant que caractéristique ou variable d'entrée dans votre modèle.

Entraînement du modèle : Entraînez un modèle de machine learning, tel qu'une régression logistique multinomiale ou un modèle d'arbre de décision multiclasse, en utilisant les performances globales calculées, les données sur la sériation et d'autres caractéristiques pertinentes que vous pourriez avoir .

Évaluation du modèle : Évaluez la performance de votre modèle en utilisant des métriques de classification multiclasse, telles que la matrice de confusion, le taux de classification correcte (exactitude), la précision, le rappel, et le score F1 pour chaque catégorie.

Réglage des plages ou seuils : Si nécessaire, ajustez les plages de scores pour chaque catégorie en fonction de vos objectifs spécifiques et des performances du modèle. Assurez-vous que les plages sont cohérentes avec votre interprétation des catégories (par exemple, ce que signifie "bien" par rapport à "correct").

Cette approche vous permet de classifier les étudiants en cinq catégories différentes en fonction de leurs performances globales tout en prenant en compte leur parcours académique et la sériation. Assurez-vous de choisir judicieusement les plages de partitions pour chaque catégorie en fonction de la signification que vous attribuez à ces catégories dans le contexte de votre étude.

# Inclure plus de paramètres ou non ?

Pour déterminer le niveau global d'un étudiant de manière précise, il est généralement préférable d'utiliser des caractéristiques supplémentaires en plus des notes finales, car les notes finales peuvent ne pas refléter pleinement la performance d'un étudiant. Les notes de contrôle continu, de contrôle final et d'autres indicateurs peuvent fournir des informations importantes sur la manière dont un étudiant a performé tout au long du semestre ou de l'année.

Voici quelques raisons pour lesquelles il peut être bénéfique d'inclure ces caractéristiques supplémentaires :

Évolution de la performance : Les notes de contrôle continu peuvent montrer comment la performance de l'étudiant a évolué au fil du temps. Par exemple, un étudiant ayant des notes en constante pourrait être évalué différemment de celui ayant des notes en déclin.

Équilibre entre les évaluations : Les notes de contrôle continu et de contrôle final peuvent aider à équilibrer la contribution de différentes évaluations tout au long du semestre ou de l'année. Cela peut être important si certaines évaluations sont plus difficiles ou plus importantes que d'autres.

Validation de crédit : Les notes de validation de crédit peuvent indiquer si l'étudiant a réussi à valider tous les crédits pour une période donnée, ce qui peut être un indicateur important de la réussite globale.

En incluant ces caractéristiques supplémentaires, vous pouvez créer un modèle de prédiction du niveau global qui prend en compte les aspects importants de la performance de l'étudiant tout au long de l'année académique. Cependant, il est essentiel de normaliser ou de normaliser les différentes notes ou indicateurs pour garantir qu'elles contribuent de manière équitable à la prédiction du niveau global.

En résumé, il est recommandé d'inclure des caractéristiques supplémentaires telles que les notes de contrôle continu, de contrôle final et de validation de crédit pour créer un modèle de prédiction du niveau global plus robuste et informatif.

# Norme sur d’étapes à suivre :

La préparation des données, y compris la création de la colonne cible, doit généralement être effectuée avant l'entraînement du modèle d'apprentissage. Vous devez avoir un jeu de données bien structuré qui contient à la fois les caractéristiques (features) et la colonne cible (la variable à prédire) pour entraîner efficacement votre modèle.

Voici les étapes générales que vous devez suivre :

Collecte de données : Collectez toutes les données pertinentes, y compris les notes de contrôle continu, de contrôle final, de validation de crédit, la sériation, etc.

Prétraitement des données : Effectuez des opérations de prétraitement sur vos données, telles que la gestion des valeurs manquantes, la normalisation des données si nécessaire, l'encodage des variables catégorielles, etc.

Calcul de la colonne cible : À ce stade, vous devrez calculer la colonne cible qui contient les catégories de niveau (faible, insuffisant, moyen, correct, bien) pour chaque étudiant en utilisant les informations de notes et d'autres critères que vous avez défini.

Création du jeu de données final : Une fois que vous avez calculé la colonne cible, assurez-vous que votre jeu de données contient toutes les caractéristiques (features), y compris la sériation, ainsi que la colonne cible.

Division des données : Divisez ensuite votre jeu de données en un ensemble d'entraînement et un ensemble de test pour évaluer la performance de votre modèle.

Entraînement du modèle : Entraînez votre modèle sur l'ensemble d'entraînement en utilisant les caractéristiques et la colonne cible.

Évaluation du modèle : Évaluez la performance de votre modèle sur l'ensemble de test en utilisant des métriques appropriées.

La création de la colonne cible (dans votre cas, les catégories de niveau) est une étape cruciale de la préparation des données, et elle doit être effectuée avant l'entraînement du modèle. Une fois que vous avez préparé votre jeu de données de cette manière, vous pouvez passer à l'entraînement de différents modèles de machine learning pour la classification des niveaux d'étudiants en fonction de leurs caractéristiques.