

3.2 Évaluation des performances

Deux approches peuvent être utilisées pour évaluer correctement le succès d'un modèle DSUC et déterminer s'il atteint ou non ses objectifs commerciaux. La première approche consiste à évaluer le modèle en comparant ses résultats à l'aide d'une liste de métriques numériques bien établies. La deuxième approche consiste à évaluer les façons dont le modèle a influencé l'entreprise en l'aidant à améliorer et à atteindre ses objectifs.

Évaluation axée sur le modèle : métriques de performance

Le résultat du modèle de prédiction élaboré correspond soit à une classe ou une catégorie (modèle de classification), soit à un nombre discret ou une probabilité (modèle de régression). Nous discuterons des métriques qui sont régulièrement mises en œuvre pour évaluer la performance de chaque type de ces modèles.

Métriques d'évaluation pour un modèle de classification

Pour un modèle DSUC conçu avec uniquement deux sorties possibles {« oui », « non »}, la décision de sortie dépend d'un seuil attribué au modèle. Lorsque le modèle est appliqué à un enregistrement de données, uniquement quatre résultats sont possibles : vrai positif, vrai négatif, faux positif et faux négatif.

- Vrai positif (VP) : dans le cas d'une prédiction correcte, le classificateur associe la mention « oui » à l'enregistrement de données.
- Vrai négatif (VN) : dans le cas d'une prédiction correcte, le classificateur associe la mention « non » à l'enregistrement de données.
- Faux positif (FP) : dans le cas d'une prédiction incorrecte, le classificateur associe la mention « oui » à l'enregistrement de données.
- Faux négatif (FN) : dans le cas d'une prédiction incorrecte, le classificateur associe la mention « non » à l'enregistrement de données.

Ces quatre résultats possibles sont généralement présentés sous forme de matrice appelée « matrice de confusion », comme illustré ci-dessous.

Matrice de confusion			
		Sortie du modèle	
		OUI	NON
Sortie souhaitée	OUI	Nombre de VP	Nombre de FN
	NON	Nombre de FP	Nombre de VN

Pour les quatre sorties possibles, il existe trois indicateurs de performance pour mesurer la qualité du modèle. Il s'agit de la précision, de la justesse et du rappel, comme expliqué dans les équations suivantes.

$$\text{précision} = \frac{\text{nombre de VP}}{\text{nombre de VP} + \text{nombre de FP}}$$

$$\text{justesse} = \frac{\text{nombre de VP} + \text{nombre de VN}}{\text{nombre de VP} + \text{nombre de VN} + \text{nombre de FP} + \text{nombre de FN}}$$

$$\text{rappel} = \frac{\text{nombre de VP}}{\text{nombre de VP} + \text{nombre de FN}}$$

Pour faire la distinction entre les deux classes du modèle de classification {« oui », « non »}, un seuil doit être appliqué. Cette valeur limite peut être définie sur un certain pourcentage établi pendant l'analyse. Toute valeur de sortie dépassant cette valeur limite renverra « oui » tandis que toutes les sorties inférieures à cette valeur renverront « non ». Par conséquent, les performances du modèle dépendent de la valeur limite qui affecte le nombre de vrais positifs, de vrais négatifs, de faux positifs et de faux négatifs en conséquence.

La courbe caractéristique de la performance d'un test montre comment la modification de la valeur limite peut changer les taux de vrais positifs et de faux positifs. Un modèle idéal serait en mesure de terminer l'opération de classification avec une justesse de 100 %, ce qui signifie qu'il pourrait produire une vraie valeur positive de 100 % et une fausse valeur positive de 0 %. En réalité, aucun modèle ne peut être aussi précis. La courbe caractéristique de la performance d'un test aide ainsi à trouver une valeur seuil plus réaliste qui comprend le vrai positif à son taux le plus élevé et le faux positif à son taux le plus faible. Pour créer une courbe caractéristique de la performance d'un test, procédez comme suit :

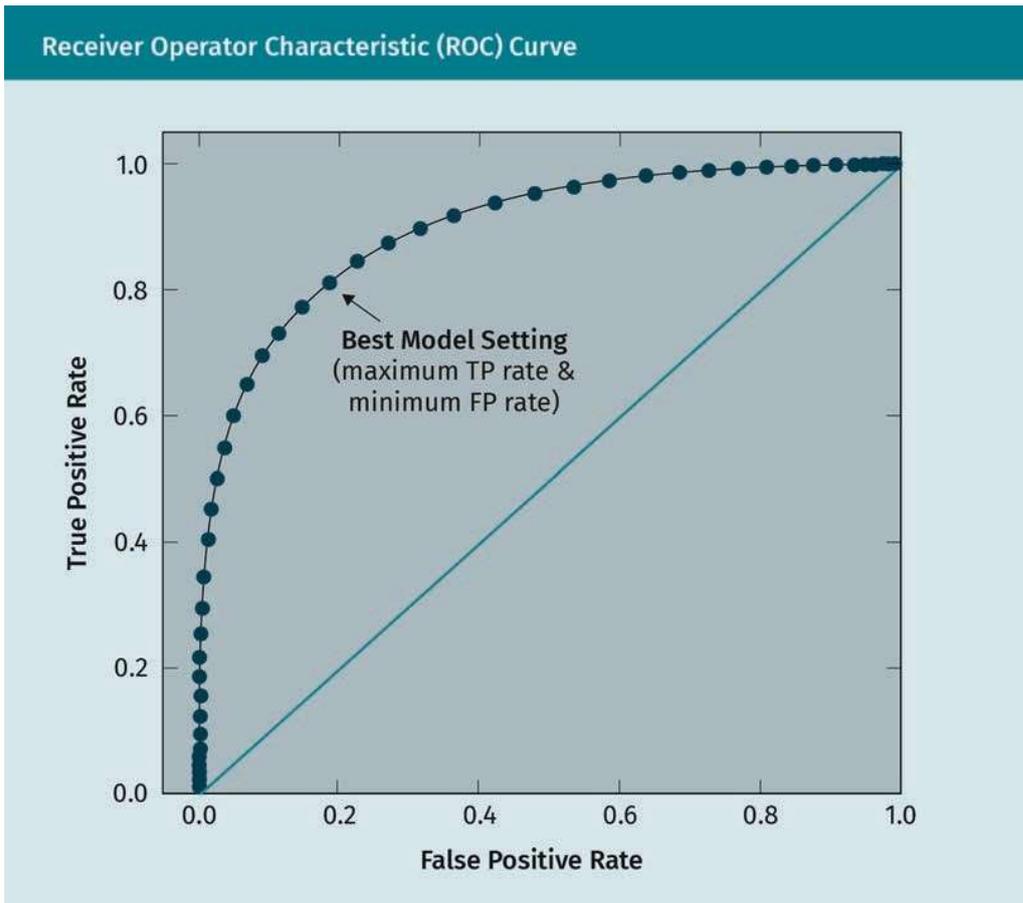
1. Une valeur limite doit être choisie entre 0 et 100.
2. Le modèle est appliqué à un ensemble de test et les valeurs VP, VN, FP et FN sont enregistrées.
3. Le calcul suivant est effectué :

$$\text{taux de faux positifs} = \frac{\text{nombre de FP}}{\text{nombre de FP} + \text{nombre de VN}}$$

et

$$\text{taux de vrais positifs} = \frac{\text{nombre de VP}}{\text{nombre de VP} + \text{nombre de FN}}$$

4. Chaque point sur la courbe caractéristique de la performance d'un test possède des coordonnées (taux de faux positifs, taux de vrais positifs).
5. Une autre fréquence limite est choisie, et les étapes 2 et 4 sont répétées, ce qui donne la courbe caractéristique de la performance d'un test indiquée ci-dessous.



Métriques d'évaluation pour un modèle de régression

L'objectif est de mesurer la proximité de la sortie d'un modèle de régression (y) par rapport à la sortie souhaitée (d). Il existe des métriques standard qui évaluent la justesse et les performances du modèle, à savoir l'erreur absolue, l'erreur relative, l'erreur absolue moyenne en pourcentage, l'erreur quadratique moyenne, l'erreur absolue moyenne et la racine carrée de l'erreur quadratique moyenne, comme indiqué dans les équations suivantes.

$$\text{Erreur absolue} \quad (\varepsilon) = |d - y|$$

$$\text{Erreur relative} \quad (\varepsilon^*) = \left| \frac{d-y}{d} \right| \cdot 100\%$$

$$\text{Erreur absolue moyenne en pourcentage} = \frac{1}{n} \sum_{i=1}^n \left| \frac{d_i - y_i}{d_i} \right| \cdot 100\%$$

$$\text{Erreur quadratique} \quad (\varepsilon^2) = (d - y)^2$$

$$\text{Erreur quadratique moyenne} = \frac{1}{n} \sum_{i=1}^n (d_i - y_i)^2$$

$$\text{Erreur absolue moyenne} = \frac{1}{n} \sum_{i=1}^n |d_i - y_i|$$

$$\text{Racine carrée de l'erreur quadratique moyenne} = \sqrt{\frac{1}{n} \sum_{i=1}^n (d_i - y_i)^2}$$

3.3 Décisions opérationnelles axées sur les données

Un aspect crucial de la pratique de la science des données réside dans l'opérationnalisation des statistiques issues des modèles analytiques utilisés. À cette fin, il est vital que les résultats des analyses soient communiqués et mis à disposition dans un format utile auprès des décisionnaires concernés au sein d'une entreprise. De plus, il est généralement bénéfique d'expliquer la logique de l'approche de modélisation afin que l'utilisateur final puisse faire une interprétation éclairée des résultats du modèle.

L'utilisateur final décide comment aligner les résultats du modèle afin de les adapter aux objectifs commerciaux. Par exemple, dans le cadre de la détection de fraude, l'utilisateur peut décider de la plage de pourcentages selon laquelle une transaction ou un comportement suspect est considéré comme une véritable fraude. Dans ce cas, le seuil sélectionné entraîne un compromis entre les faux négatifs et les faux positifs. Ces compromis doivent être pris en considération pour maximiser l'efficacité du modèle. L'utilisation de valeurs différentes pour le seuil permet aux responsables de l'entreprise de prendre en compte différents scénarios.

Dans certains cas, l'utilisateur final peut avoir à prendre une décision qui a un impact direct sur les enregistrements de données, ainsi qu'une décision concernant la valeur des seuils. Par exemple, une caractéristique possible dans un ensemble de données concerne le prix du produit. Dans certains cas, il peut être nécessaire de modifier le prix du produit. Dans ce cas, le modèle doit être capable de s'adapter à ces changements et être à nouveau entraîné.

L'objectif ultime d'un modèle intelligent réside dans l'automatisation des décisions des utilisateurs. Ces décisions dépendent souvent de la capacité de prédiction du modèle. Par exemple, un modèle peut être conçu pour analyser les avis laissés sur les hôtels et décider si ces avis sont fictifs ou non. Si les prédictions du modèle sont très précises, un avis peut être automatiquement accepté ou rejeté sans intervention humaine.

Évaluation centrée sur l'entreprise : rôle des indicateurs clés de performance

Une fois qu'un modèle a été évalué avec succès à l'aide des métriques d'évaluation mentionnées ci-dessus, il est prêt à être déployé. À ce stade, le modèle doit être en mesure de produire une valeur DSUC fiable pour la problématique métier associée. Les décisionnaires doivent alors être sûrs que le modèle DSUC est correctement mis en œuvre afin d'aider l'entreprise à atteindre ses objectifs commerciaux. La quantification du mérite du modèle est réalisée en définissant les indicateurs clés de performance. Il s'agit de mesures qui indiquent dans quelle mesure les objectifs commerciaux ont été atteints ou non. La plupart des indicateurs clés de performance se concentrent sur une efficacité accrue, une réduction des coûts, une amélioration des revenus et une meilleure satisfaction client.