

Vous travaillez pour une entreprise ayant une forte présence dans le commerce électronique. Vous disposez d'un ensemble de données décrivant le comportement de navigation de différentes personnes sur la base d'une analyse de leurs traces Web et indiquant si elles ont généré des revenus pour votre entreprise ou non. Vous devez analyser le jeu de données afin de déterminer quels sont les attributs les plus prédictifs permettant de savoir si une personne génère des revenus pour votre entreprise ou non. Finalement, vous devriez créer et évaluer un modèle prédictif qui, compte tenu de la description d'une personne, déterminera si cette personne générera ou non des revenus pour l'entreprise.

Le jeu de données a les attributs suivants:

*"Administrative"*, *"Administrative Duration"*, *"Informational"*, *"Informational Duration"*, *"Product Related"* and *"Product Related Duration"* : représente le nombre de types différents de pages visitées par le visiteur au cours de cette session et le temps total passé dans chacune de ces catégories de pages. Les valeurs de ces fonctionnalités sont dérivées des informations d'URL des pages visitées par l'utilisateur et mises à jour en temps réel lorsqu'un utilisateur effectue une action, par ex. passer d'une page à une autre.

*"Bounce Rate"*, *"Exit Rate"* and *"Page Value"* représentent les statistiques mesurées par *"Google Analytics"* pour chaque page du site de commerce électronique.

- La valeur de la fonctionnalité *"Bounce Rate"* pour une page Web fait référence au pourcentage de visiteurs qui entrent sur le site à partir de cette page, puis quittent ("rebond") sans déclencher aucune autre demande au serveur d'analyse au cours de cette session.
- La valeur de la fonctionnalité *"Exit Rate"* pour une page Web spécifique est calculée comme pour toutes les pages visualisées sur la page, le pourcentage correspondant au dernier de la session.
- La fonctionnalité *"Page Value"* représente la valeur moyenne d'une page Web visitée par un utilisateur avant la fin d'une transaction de commerce électronique.

*"Special day"* indique que l'heure de visite du site est proche d'une journée spéciale (par exemple, la fête des mères, la Saint-Valentin) au cours de laquelle les sessions ont plus de chances d'être finalisées lors d'une transaction. La valeur de cet attribut est déterminée en tenant compte de la dynamique du commerce électronique, telle que la durée entre la date de la commande et la date de livraison. Par exemple, pour la fête de St. Valentine, cette valeur prend une valeur non nulle entre le 2 février et le 12 février, zéro avant et après cette date sauf si elle est proche d'un autre jour spécial et sa valeur maximale de 1 le 8 février.

Le jeu de données comprend également *"operating system"*, *"browser"*, *"region"*, *"traffic type"*, *"visitor type"* visiteur renvoyé ou le nouveau visiteur, une valeur booléenne indiquant si la date de la visite est le week-end et le mois de l'année.