

Esercizio banca - Trasformazione e predizione

1. Carica in un dataframe il file “bank.csv” (preso da kaggle - <https://www.kaggle.com/datasets/mathchi/churn-for-bank-customers>) che contiene dati relativi ai clienti di una banca. Lo scopo è predire se il cliente abbandonerà la banca (feature `Exited`). Quante righe e quante colonne sono presenti? Ci sono valori nulli? Il dataset è bilanciato rispetto alla classe da predire?
2. Converti i valori di `Gender` in numerici (`Male` = 0, `Female` = 1).
3. Trasforma la feature `Geography` con `OneHotEncoder`.
4. Elimina le feature `RowNumber`, `CustomerId`, `Surname` e dividi il dataset in train (75% del dataset) e test (25%). Usa `random_state=0`.
5. Allena un decision tree per predire l’abbandono dei clienti, scegli tu i valori dei parametri e calcola l’accuratezza della predizione.
6. Aggiungi una nuova feature nel dataset con il valore di $(\text{EstimatedSalary} * \text{Tenure} + \text{Balance}) / 2$. L’accuratezza della predizione migliora?
7. Applica due trasformazioni diverse a tutto il dataset e/o a feature specifiche, valuta la predizione.

Sfidatevi ad ottenere il valore più alto di accuracy! Il mio è 79,6%