

Analisi dei clienti di una banca

Descrizione del Progetto

L'azienda **Banking Intelligence** vuole sviluppare un modello di machine learning supervisionato per prevedere i comportamenti futuri dei propri clienti, basandosi sui dati transazionali e sulle caratteristiche del possesso di prodotti. Lo scopo del progetto è creare una tabella denormalizzata con una serie di indicatori (feature) derivati dalle tabelle disponibili nel database, che rappresentano i comportamenti e le attività finanziarie dei clienti.

Obiettivo

Il nostro obiettivo è creare una tabella di feature per il training di modelli di machine learning, arricchendo i dati dei clienti con vari indicatori calcolati a partire dalle loro transazioni e dai conti posseduti. La tabella finale sarà riferita all'**ID cliente** e conterrà informazioni sia di tipo quantitativo che qualitativo.

Valore Aggiunto

La tabella denormalizzata permetterà di estrarre **feature comportamentali avanzate** per l'addestramento di modelli di **machine learning supervisionato**, fornendo numerosi vantaggi per l'azienda:

- **Predizione del comportamento dei clienti:** Analizzando le transazioni e il possesso di prodotti, si possono identificare pattern di comportamento utili per prevedere azioni future come l'acquisto di nuovi prodotti o la chiusura di conti.
- **Riduzione del tasso di abbandono:** Utilizzando gli indicatori comportamentali, si può costruire un modello per identificare i clienti a rischio di abbandono, permettendo interventi tempestivi da parte del team di marketing.
- **Miglioramento della gestione del rischio:** La segmentazione basata su comportamenti finanziari consente di individuare clienti ad alto rischio e ottimizzare le strategie di credito e rischio.
- **Personalizzazione delle offerte:** Le feature estratte possono essere utilizzate per personalizzare offerte di prodotti e servizi in base alle abitudini e preferenze dei singoli clienti, aumentando così la **customer satisfaction**.
- **Prevenzione delle frodi:** Attraverso l'analisi delle transazioni per tipologia e importi, il modello può rilevare anomalie comportamentali indicative di frodi, migliorando le strategie di sicurezza e prevenzione.

Questi vantaggi porteranno un miglioramento complessivo delle operazioni aziendali, consentendo una maggiore efficienza nella gestione dei clienti e una crescita sostenibile del business.

Struttura del Database

Il database (che puoi [scaricare da qui](#)) è costituito dalle seguenti tabelle:

1. **Cliente**: contiene informazioni personali sui clienti (ad esempio, età).
2. **Conto**: contiene informazioni sui conti posseduti dai clienti.
3. **Tipo_conto**: descrive le diverse tipologie di conti disponibili.
4. **Tipo_transazione**: contiene i tipi di transazione che possono avvenire sui conti.
5. **Transazioni**: contiene i dettagli delle transazioni effettuate dai clienti sui vari conti.

Indicatori Comportamentali da Calcolare

Gli indicatori saranno calcolati per ogni singolo cliente (riferiti a `id_cliente`) e includono:

Indicatori di base

1. **Età del cliente** (da tabella `cliente`).

Indicatori sulle transazioni

2. **Numero di transazioni in uscita** su tutti i conti.
3. **Numero di transazioni in entrata** su tutti i conti.
4. **Importo totale transato in uscita** su tutti i conti.
5. **Importo totale transato in entrata** su tutti i conti.

Indicatori sui conti

6. **Numero totale di conti posseduti**.
7. **Numero di conti posseduti per tipologia** (un indicatore per ogni tipo di conto).

Indicatori sulle transazioni per tipologia di conto

8. **Numero di transazioni in uscita per tipologia di conto** (un indicatore per tipo di conto).
9. **Numero di transazioni in entrata per tipologia di conto** (un indicatore per tipo di conto).
10. **Importo transato in uscita per tipologia di conto** (un indicatore per tipo di conto).
11. **Importo transato in entrata per tipologia di conto** (un indicatore per tipo di conto).

Piano per la Creazione della Tabella Denormalizzata

1. Join delle Tabelle

Per costruire la tabella finale, sarà necessario eseguire una serie di **join** tra le tabelle disponibili nel database.

2. Calcolo degli Indicatori

Gli indicatori comportamentali verranno calcolati utilizzando operazioni di aggregazione (SUM, COUNT) per ottenere i totali richiesti.

Modalità di consegna: File SQL