

Raccolta e analisi dei requisiti per un progetto di
analisi dati

Analisi, Valutazione del modello di credit scoring
di Revolut Bank

Revolut



Giuseppe Schillaci

Epicode Project 1

30 Settembre 2022

<https://github.com/giuschil/epicode-data-analyst/tree/main/project1-week1>

Descrizione del caso di studio

Storia

Revolut è una piattaforma finanziaria globale, supportata da un'app che consente in ogni momento di gestire le tue finanze con un semplice smartphone. Revolut Ltd è una società fintech, con sede nel Regno Unito fondata nel 2015. Tuttavia, già nel 2018 ha raggiunto lo status di unicorno. Con ciò si intende che è stata valutata più di un miliardo di dollari. La società ha realizzato un conto corrente multivaluta.

A partire dal 2021 ha ottenuto la licenza bancaria anche in Italia dove viene riconosciuta a tutti gli effetti come una banca, dando quindi accesso ad un vero e proprio conto corrente. Tuttavia, esso non dà IBAN italiano, ma l'estensione LT è riconosciuta nell'area SEPA, cioè dei pagamenti europei. Revolut offre una carta di debito prepagata che consente di prelevare denaro (in valuta locale) in oltre 120 paesi del mondo.

Caso

Dal 2021 ha iniziato a concedere prestiti e credito al consumo direttamente online ai propri clienti direttamente tramite l'app proprietaria.

La suddetta analisi dei requisiti è necessaria per la progettazione di uno strumento per le analisi, valutazione e migrazione del modello già utilizzato da Revolut, in particolare si propone di:

1. Migliorare il modello di credit scoring di Revolut Bank per migliorare l'efficacia con cui concede credito ai suoi clienti
2. Poter analizzare e predire tramite dei modelli di ML quali dei suoi clienti potrebbe voler richiedere del credito.
3. Poter analizzare e predire il default rate dei clienti che ha già ottenuto il credito.
4. Predire quali e quando i clienti potrebbero aver necessità di credito per i loro acquisti e fornire i dati al team marketing per migliorare le leads performance

Analisi e raccolta dei requisiti

Il **modello di credit risk di Revolut** attuale utilizzato dalla banca calcola in base ai soli dati in possesso (età, sesso, disponibilità di denaro in conto corrente, se si è proprietari di immobile o in affitto, la professione che si svolge, la quantità di denaro richiesta, il tipo di acquisto che si vuole fare con il credito e la durata del credito) se concedere o meno il credito.

Per la creazione del progetto di analisi per la banca **Revolut** il **team Data & Technology** della società di consulenza **Epicode Consulting** ha raccolto i dati seguenti:

Dataset	dimensione dati	formato	link fonte	descrizione
dataset_bank.csv	10 clienti	csv	https://api.revolut.com/bankdata/	Dataset della banca sui clienti che hanno richiesto un credito e risultato del credit score
dataset_fiscalagency.json	10 contribuenti	json	https://api.agenziaentrate.gov.it/portale/	Dataset dell'agenzia delle entrate sui dieci clienti che hanno richiesto il credito
dataset_transactions.xml	10 transazioni di 1 cliente	xml	https://api.stripe.com/bank/37475852	Dataset creato da azienda di credito terza su tutti i pagamenti online e offline effettuati

Requisiti per la Banca

Solo i già clienti possono richiedere un prestito o credito al consumo.

I clienti possono richiedere un credito della durata massima 30 anni, di quantità non superiore a 300.000 €, non è importante sapere il tasso a cui è stato richiesto il prestito.

I clienti possono chiedere un credito per l'acquisto di una car, house, other. I dati forniti dalla banca sono in formato **CSV**.

Requisiti per le Transazioni

I dati sulle transazioni sono i dati raccolti dall'istituto di credito online per i pagamenti (esempio Stripe/Mastercard/Satispay) trasmessi tramite connessione API.

I dati sono in formato **XML** e racco

Requisiti per i dati dell'Agenzia delle entrate

I dati dell'agenzia dell'entrate sono rilasciati in formato **JSON** sono dati protetti e contengono l'anagrafica del contribuente e i redditi degli ultimi anni forniti dal contribuente

Team necessario Data & Technology

- Data Engineer
 - Progetta e struttura il Data Lake di dati non strutturati su server distribuiti tramite Hadoop e linguaggi di programmazione Java e Sql.
- Data Analyst
 - Analizza i dati tramite interrogazioni della base di dati, crea report e dashboard sulle spese maggiormente fatte, la media degli importi spesi, i luoghi e la categoria merceologica maggiormente acquistata.
- Data scientist

- Crea il modello Crea un modello per predire quale cliente potrebbe, in base alla sua capacità di spesa, voler richiedere un prestito, anticipando così le campagne di email marketing

Strutturazione dei requisiti in gruppi di frasi omogenee

Per i dati della banca

Per tutti Per i clienti (10), identificati da un codice numerico univoco, il codice fiscale, il codice cliente, l'età, il sesso, lo stato civile, il lavoro, se si vive in casa di proprietà o in affitto, la somma che è sul conto, l'ammontare del credito, la durata del credito richiesto, la decisione del modello.

Per i dati sulle transazioni

Le transazioni (10) o movimenti sono identificate da un codice numerico univoco. **Le transazioni sono di un solo cliente della banca.**

Per i dati sui contribuenti

I dati dei contribuenti (10) sono identificati da un codice, il nome, il cognome, il codice fiscale, l'indirizzo di residenza, il reddito dipendente degli ultimi 5 anni, altri tipi di redditi e il numero delle persone a carico.

Glossario dei termini

Termine	Descrizione	Sinonimi	Collegamenti
Cliente	Cliente della banca che richiede il credito	Utente, contribuente	Banca, Transazioni, Agenzia delle entrate
Prestito	Somma di denaro concessa al cliente dalla banca	Credito, Credito al consumo	Banca, Clienti
Banca	Azienda che permette la concessione di denaro ai suoi clienti o la possibilità di effettuare pagamenti	Istituto Bancario, Azienda di Credito	Clienti, transazioni
Agenzia delle entrate	L'Agenzia delle entrate è un'agenzia fiscale della pubblica amministrazione italiana svolge le funzioni relative ad accertamenti e controlli fiscali e alla gestione dei tributi.	Agenzia fiscale	Banca, contribuenti
Transazione	Transazione di denaro effettuata dal cliente della suddetta banca	Pagamento	banca, clienti
Credit Risk Model	Modello per stabilire la concessione del credito o meno		banca, agenzia delle entrate

Default Rate	Il default rate indica la probabilità di insolvenza del cliente	Tasso di insolvenza	Banca,Clienti
--------------	---	---------------------	---------------

Tipologia di dati

Dati Revolut Bank

Tipo di dataset strutturato in formato csv fornito dalla banca Revolut

feature	type	value	modify value
id	number		
fiscal_code	varchar		
banck_account_number	varchar		
age	number		
sex	boolean	M,F	0,1
personal_status	varchar	celibe,sposato,nubile	0,1,2,3,4
job	varchar	insegnante,operaio,avvocato	
housing	booeLAN	proprietà, affitto	0,1
saving_account	number		
credit_amount	number		
duration	number		
purpose	varchar	car,house,other	
decision	boolean	Yes,No	0,1

Dati transazioni-pagamenti online e offline

Tipo di dataset non strutturato in formato xml fornito dal sistema di pagamenti Online.

Le transazioni sono 10 e sono di un solo cliente della banca.

transaction id
category
date
time
company
url
title
price
currency
location
city
street

Dati Agenzia delle entrate

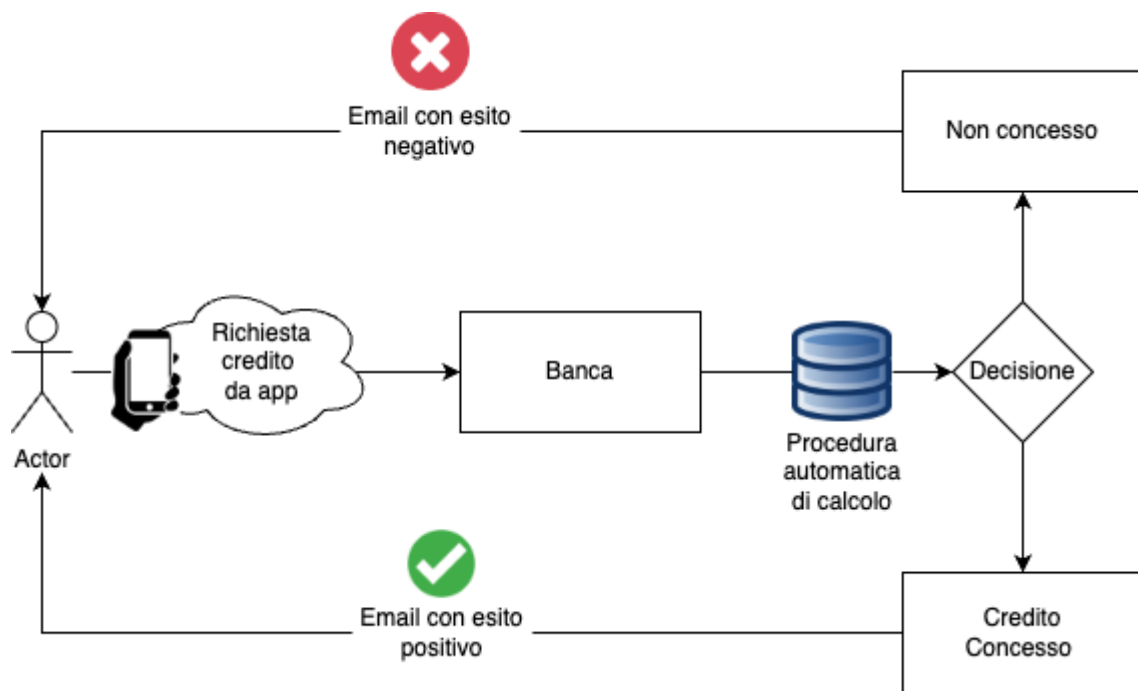
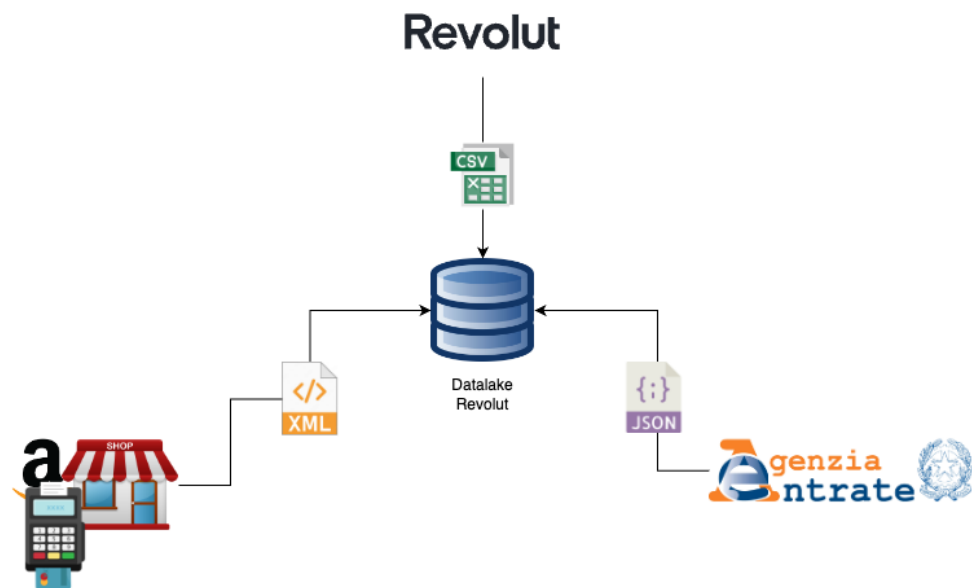
Tipo di dataset non strutturato in formato json fornito dall'agenzia delle entrate

codice
nome

cognome
cod_fiscale
indirizzo
citta
cap
numero
tipo
nome_zona
reddito
reddito dipendente
reddito altri
persone_a_carico

Implementazione

Di seguito viene illustrato come verrà implementato il data lake per la banca. Nel disegno possiamo vedere come le diverse origini di dati vengono convogliate in un data lake.



Supporto alle decisioni aziendali

Analisi di business

- Trend
 - Analisi dell'andamento delle richieste di prestito dai clienti
- Peak location
 - Quali Città hanno avuto più prenotazioni

- Quali tipi di strutture sono state maggiormente prenotate
- Target customers
 - Quali tipologie di utente ha effettuato più prenotazioni
- Campagne Marketing/Advertising
 - Aumentare la leads generation tramite pubblicità e email marketing ai clienti che potrebbero voler richiedere del credito.

Analisi predittiva

- Trend: Previsione nei prossimi 3 mesi
- Peak seasons
 - Periodo preferito dagli utenti
- Peak location
 - Quali Città saranno le preferite dagli utenti nei prossimi 3 mesi
- Target customers
 - Quali tipologie di utente effettuerà più prenotazioni nei prossimi 3 mesi
- Campagne Marketing/Advertising
 - Consigliare le strutture di quale città promuovere con campagne di advertising
 - Sapere l'effetto delle campagne marketing effettuate prima e dopo la data delle campagne

Storia/Infografica

1. Infografica sul cliente tipo per le città più grandi.
2. Infografica sul sole24ore.it sulla quantità di denaro concessa a prestito dall'azienda
3. Infografica sul portale revolut.com sui dati di genere ed età dei clienti

Analisi Aggiuntive Proposte

Sono richieste le seguenti analisi:

- Analisi statistica della concessione del credito ai clienti raggruppati per classe di età (18-25, 25-34, 35-44, 45-60).
- Analisi sul numero dei prestiti concessi
- Analisi sulla quantità di credito concesso
- Analisi sulla qualità del modello

Possibili implementazioni future

- 1) Anticipare la richiesta di credito da parte degli utenti tramite la creazione di modelli predittivi.
- 2) Migliorare il modello per abbassare il default rate dei clienti