

Titolo della tesi: Analisi integrata delle risultanze di un framework controlli nel contesto Big Financial Data di Intesa Sanpaolo

Autore: Chiara Perri

Abstract

Ogni giorno la vita ci porta continuamente ad effettuare scelte. In base alla situazione da affrontare, le potenziali alternative sono vagliate da ciascuno di noi con un differente grado di profondità, dipendente dalle evidenze a nostra disposizione. Senza alcuna sorpresa ne consegue che, a parità di processo decisionale, la possibilità di fare la scelta giusta sia legata fortemente alla qualità delle informazioni utilizzate.

La tesi si inserisce nel contesto del modello di Data Governance di Intesa Sanpaolo, focalizzandosi sul framework dei controlli che vengono applicati ai dati utilizzati e prodotti dalla Banca. La *mission* assegnata prevede di analizzare le difettosità che emergono a seguito delle verifiche effettuate, con l'intento di individuare pattern stagionali significativi nelle stesse.

A partire dallo studio del Database, sia da un punto di vista tecnico, sia nei contenuti, si è svolta una significativa attività di Data Preparation. Con l'utilizzo del prodotto di mercato Irion-EDM, è stato disegnato un Dataflow direttamente alimentato dalle tabelle in cui risiedono i dati oggetto di studio. Il processo consta di un primo layer di Data Quality, personalizzato sulla base delle necessità, e successivamente di uno strato di Data Transformation volto alla predisposizione del subset di dati ritenuto affidabile per l'analisi finale. La ri-esecuzione del Dataflow disegnato permette di aggiornare in maniera automatica la base dati di partenza per le analisi, garantendo sempre la possibilità di disporre di informazioni correttamente predisposte e di certificata qualità.

In prima battuta l'analisi ha riguardato lo studio di possibili correlazioni esistenti tra le difettosità riscontrate e gli attributi che definiscono le caratteristiche dei controlli. Lo studio preliminare del Dizionario Controlli ha permesso di individuare gli attributi su cui operare un'opportuna transcodifica da variabili categoriche a variabili ordinali. Nella tesi è presentato un subset dell'analisi per cui sono state individuate evidenze significative a diversi livelli, sia sul totale delle difettosità, sia in base alla tipologia del controllo (ad esempio tecnico o di merito).

Successivamente lo studio è proseguito con l'analisi temporale delle difettosità nel periodo di riferimento individuato e sul subset di controlli definito

attraverso la Data Preparation. Anche in questo caso l'analisi ha preso in considerazione sia i risultati a livello globale, differenziando le diverse periodicità di elaborazione del controllo stesso, sia quanto ottenuto da sottoinsiemi basati sulla differente tipologia di controllo. Questa analisi ha permesso, da un lato, di individuare alcuni risultati significativi e, dall'altro, di confermare la relativa immaturità della base dati a disposizione per poter procedere con un'attività di forecasting. Si tratta infatti di una base dati il cui popolamento è avvenuto gradualmente a partire dal 2016 inoltrato, e per cui la prima annualità stabile individuata corrisponde al 2018.

Non trascurabile è infine lo studio di aspetti legati alla Data Governance e alla loro effettiva applicazione nel modello aziendale Intesa Sanpaolo. A questi argomenti sono stati dedicati i primi due Capitoli della tesi, per comprendere appieno l'ambito in cui il lavoro svolto si colloca.