

# Estrapolazioni da fattura

## 1. Introduzione

Il seguente documento ha lo scopo di pianificare gli obiettivi e la definizione del problema che si vuole analizzare tramite le fatture elettroniche.

Verranno quindi inserite nei capitoli successivi tutte le idee possibili su come affrontare il problema, sui possibili nuovi dati estrapolabili da quelli già presenti e sugli insight che si vuole realizzare.

## 2. Struttura

Prima di intraprendere un ragionamento su le possibili ingegnerizzazioni delle caratteristiche, vorrei fare una breve panoramica su come è formattata una Fattura elettronica tramite il servizio Fex.

In questo capitolo verranno descritti solo i dati ritenuti importanti

### 2.1. Header

#### 2.1.1. Dati Trasmissione

Dati trasmissione contiene l'id del paese e del codice fiscale del fornitore all'interno dell'Id trasmittente. Altre informazioni utili sono il formato trasmissione che contiene un codice che identifica se la fattura è indirizzata verso una pubblica amministrazione o verso un privato.

#### 2.1.2. Cedente Prestatore e Cessario Committente

In questo campo vengono inserite la provenienza, il codice fiscale e la denominazione del fornitore. Vengono inoltre incluse informazioni come l'indirizzo della sede, l'iscrizione al repertorio delle notizie economiche ed amministrative e i relativi contatti. La struttura per il committente è la medesima.

### 2.2. Body

#### 2.2.1. Dati generali

In dati generali viene indicata la data della fattura, i riferimenti come l'identificativo del documento dell'ordine acquisto, la causale e l'importo totale.

#### 2.2.2. Dati Beni e Servizi

Il campo Dati Beni e Servizi viene diviso in Linee in base agli articoli in fattura.

Qui possiamo trovare il codice articolo, descrizione, la quantità e il prezzo totale.

Come in ogni Nodo maggiore, sono presenti più campi, molti dei quali ho deciso di trascurare perchè ritenuti poco utili allo scopo ultimo. Ad esempio l'unità di misura, lo sconto ecc... non mi è di utilità tenerlo in considerazione.

#### 2.2.3. Dati Pagamento

In questa sezione troviamo tutti i dettagli sul pagamento. Dal metodo e la modalità di pagamento, l'iban, la scadenza e l'importo. E' la sezione che ritengo più importante e con più informazioni utili da analizzare.

## 3. Insight pre-ML

Prima di procedere con algoritmi di Machine Learning, credo che sarebbe utile e interessante analizzare i dati puliti per ottenere dei grafici parametrici e semplici sui dati presenti così come

sono.

Questi insight verrebbero eseguiti con semplici funzioni python nelle quali si passa un parametro (esempio PIVA) per analizzare diversi aspetti delle aziende. I grafici che ritengo essere utili analizzare potrebbero essere:

- Totale delle vendite di un'azienda per mesi;
- Suddivisione dei tipi di pagamento come bonifici, assegni, contanti...;
- Comparazione vendite totali tra due aziende;
- Variazioni delle vendite per mesi o anni (dipende dalla quantità dei dati).

## **4. Ingegnerizzazioni delle caratteristiche e clusterizzazione**

I dati che estrapolerei per poi l'analisi finale e creare una clusterizzazione delle diverse aziende, potrebbero essere, il totale fatturato (non so possono bastare le fatture elettroniche o queste aziende hanno sedi fisiche per cui non eseguono solo fattura elettronica), la frequenza dei pagamenti, la quantità di articoli venduti.

Creare questi tre nuovi parametri per poi applicarci un modello di clustering potrebbe dare degli insight interessanti sulle aziende, come azienda che vende poche quantità perchè di nicchia ma con grande fatturato, azienda che vende molto ma beni di poco valore ecc...

Si potrebbe pensare anche, nel caso le varie aziende vendano anche in sedi senza fatture all'ingrosso, di estrapolare in qualche modo il fatturato annuo tramite PIVA, tramite API terze o attraverso web scraping. Chiaramente bisogna allineare le idee a poi i dati effettivi soprattutto in considerando l'arco temporale che questi occupano.

Il numero di cluster e la loro spiegazione non può essere fatta senza prima analizzare concretamente i dati, gli insight dei cluster vengono fatti dopo che si sono formati.

Anche in questo caso poi, sarebbe interessante estrapolarne dei grafici per visualizzare e rendere il tutto più chiaro.