Title: The new face of ETL in Data Mining and Machine Learning for a Big Data Business Analysis

# CONTENTS

**DEDICATION - ABSTRACT - ACKNOWLEDGEMENTS**

**LIST OF TABLES & FIGURES**

**Introduction (3 pag)**

Problem Description, Methodology & Scope and Limitation

**Chapter 1 Background (25 pag)**

Cosa sono i Big Data?

Barriere All’utilizzo Dei Big Data

Tecniche Per L’analisi Dei Big Data

Progetti Di Big Data Nel Marketing

Direct E Digital Marketing

Customer Micro-Segmentation

Price Optimization

Location-Based Marketing

In-Store Analysis

Cross-Selling / Up-Selling

Architettura concettuale per l’analisi dei Big Data

Focus on Social- Sentiment Analytics

OLTP vs OLAP

On-Line Transaction Processing (OLTP)

On-Line Analytical Processing (OLAP)

Knowledge Discovery in Databases (KDD)

**Chapter 2 Social Analytics & ETL REAL TIME (25 pag)**

Business Intellingence(BI)

Data Warehouse

ETL

Streaming ETL-> ottima per la presentazione dinamica alla discussione

Kafka-Confluent ( dati da TWITTER)

What is K-SQL?

SinkConnectors (oracle, Elastic)

DataMart

Facts model (DFM)

Star schema

Snowflake

**Chapter 3 Data Mining** **- Machine Learning (15 pag)**

Definition Data Mining & Machine Learning

Data Mining Algorithms

Association Rules

Clustering

Classification

Decision Tree

Artificial Neural Networks

Bayesian Networks

Instance-Based Learning (kNN)

Support Vector Machines

Error of Machine Learning (Word Error Rate, Translation Error Rate, NIST, METEOR)

Data Mining Applications

**Chapter 4 – PROJECT (40 pag)**

Top-Down ANALYSIS

Datagrip

Connection to DB

Importing Done Datawarehouse of Furla

Creation of Data Mart

Star schema -> fare schema su Furla

Snowflake -> fare schema su Furla

Talend

Metadata by connection to db

ELT traditional with components

t-maps

Facts model (DFM) -> fare schema su Furla

Creation of Business Datawarehouse OracleDB

Real time ETL

Python code -> Codice per collegare twitter db a real time ETL

Kafka-Confluent ( dati da TWITTER) -> ottima per la presentazione dinamica alla discussione

K-SQL

Creation of Table

Creation of Stream

SinkConnectors (oracle, Elastic)

Grafana -> data visualization

SPSS?

Association Rules (borsa, portafoglio) (Zaino,portafoglio) (borsa,zaino,portafoglio)

Clustering (Cliente, Prodotto)

Classification (sneakers, sesso, Tipo di negozio)

**Chapter 5 – Result by Data Visualization (10 pag)**

Microsoft Power BI -> pag1: KPI + Geographical location

pag2: Sales flux

pag3: Sneakers

pag4: Italy Benchmarketing

pag5: TOP-WORST

**Conclusion & What reserve the future (3 pag)**

Conclusion

What reserve the future

**REFERENCES**

**ANALISI TOP-DOWN TESI FURLA - LUCA BREGATA**

**Attributi aziendali di successo nel campo della moda:**

* **posizione strategica**
* **qualità/prezzo**
* **diversificazione dei prodotti**
* **innovazione**
* **comodità**

**Obiettivo aziendale:**

* **Guadagnare il piu possibile avendo il minimo inventario in magazzino = costi**
* **Aumentare le vendite**

**Competitors:**

* **Gucci**
* **Prada**
* **Michael Kors**
* **Burburry**
* **Dolce & Gabbana**
* **Moschino**
* **Luis Vitton**
* **Versace**
* **Fendi**
* **Ridurrei a 5: Gucci, Burburry, Michael Kors, Fendi, Prada**

**Regole di associazione**

* **(Borsa, Portafoglio)**
* **(Zaino, Portafoglio)**
* **(Borsa, Zaino, Portafoglio)**

**Cluster:**

* **Prodotto**
* **Country**
* **Cliente**
* **Sneakers**

**Classificazione**

* **Tipo di negozio (CANALE\_DESC)**
* **Sesso**

**DOMANDE DA FARMI PER AUMENTARE L’EFFICIENZA AZIENDALE**

**ANALISI 1: NEGOZI**

1. **meglio puntare nell’e-commerce o retail?**

**Quali oggetti una persona preferisce vedere prima di comprare?**

Valutazione in base ai prodotti piu venduti dell’ecommerce e sui piu venduti nel retail.

1. **Meglio una vetrina elegante o stravagante per attirare i clienti?**

Osservo i top 10 negozi e tramite un backward analisi guardo se son negozi piu eleganti e molto statici oppure son negozi piu “colorati e vivaci” nell’esporre i propri prodotti

1. **Dopo quanto devo spostare la merce da un negozio retail ad un outlet?**

Valuto il tempo che un oggetto rimane nello store, calcolato dal tempo di consegna al tempo attuale/ tempo di vendita

1. **Quante volte un mio cliente ritorna a comprare? Quante volte invece torna a comprare ma attraverso l’e-commerce?**

Valutazione Negozi. Utilizzo il count della gente che entra ed esce e associa ad essi le vendite

1. **Quanto magazzino può permettersi questo negozio? Quanto posso rischiare di lasciare invenduto?**

Magazzino= costi. Calcolare la percentuale merce venduta/merce arrivata e sfilare una classifica dei WORST 10. In questi negozi bisogna spedire meno merce.

1. **Quale particolare mi differenzia e mi fa subito riconoscere da competitors?**

Bisogna far si che l’azienda si focalizzi su pochi prodotti che innalzano la Brand Loyalty, valutazione tramite pareto

1. **Meglio aprire negozi di nicchia (corner) in Grandi città o aprire negozi negli outlet?**

Calcolare il rapporto vendite profitto per tipo di shop

calcolare # visitatori nei negozi outlet e nei corner per vedere se l’affluenza e le vendite mi posso far cambiare strategia

Preferisco visibilità o fatturato?

Quali Outlet hanno un numero maggiore di visibilità?

Analisi territori tramite dati istat

In quali zone l’azienda è debole in termini di mercato?

**ANALISI 1: SCONTRINI / CLIENTI**

1. **Quale è la fascia di età dei miei clienti? E come faccio ad espanderla?**

**Quali età del cliente NON coprono i competitors?**

in anagrafica non c'è l'età dei clienti, ma si potrebbe cercare di simularli

1. **Quali colori sono prediletti nella mia visione di moda?**

Fare un cluster sui colori ed evidenziare il piu amato dai clienti

1. **Quando devo creare una promozione e per chi?**

**Meglio a breve o lungo termine?**

**Mi conviene fare promozioni su un singolo prodotti o basandomi sulle regole di associazione?**

**Quando il mio numero delle vendite è sotto la media?**

Come faccio a riequilibrare questo dato? con delle promozioni singolari per determinati oggetti o promozioni tramite regole di associazione in questo modo determino anche il periodo dell’offerta.

1. **Quanto cresce/diminuisce il mio marchio rispetto al settore?**

**Dove il mio marchio fa parlare e coinvolge la gente?**

Tramite real time tweets con l’uso di python e confluent inserendo i dati in un db grazie ai sink connettors. osservazione dei dati e dei tweet inerenti al mio marchio e a quelli dei competitors. Quale è il feedback dei miei clienti?

Il numero di citazioni sta salendo o decrescendo rispetto ad un periodo passato?

In quale luogo il mio marchio è piu conosciuto? In quali paesi invece devo investire per aumentare la mia visibilità? -> **dati real time** **by twitter**

1. **Come si stanno muovendo i miei competitors?**

Continua osservazione dei puri dati delle vendite di ogni competitors con relativo simbolo di trading positivo o negativo per evidenziare l’andamento odierno della azienda rispetto ai vari avversari del settore...il tutto riassunto in una tabella

1. **Il mio cliente è premium?**

Offerte speciali per utenti premium individuati tramite creazione di un boolean

1. **Mi conviene puntare piu su una clientela puramente femminile o anche maschile?**

Calcolare la percentuale di clientela femminile sul totale

Calcolare la percentuale delle vendite di un prodotto femminile

**ANALISI 3: SNEAKERS**🡪 nuovo prodotto febbraio 2019 nella fashion week

1. **Quale impatto ha il nuovo prodotto nell’economia dell’azienda? È un prodotto basato sul retail o sul e-commerce?**

Fatturato prima e dopo febbraio 2019

Fatturato specifico sneakers

1. **Cattura piu un mercato rispetto ai prodotti precedenti?**

Analisi sul numero di clienti prima e dopo la nascita della linea scarpe

1. **Il fatturato è così efficace da pensare in un recente futuro ad inizializzare anche una linea di produzione dedicata al vestiario per furla?**

Analisi costi/benefici

1. **tasferisce/aumenta la visibilità e le vendite in uno specifico continente?**

Tramite real time tweets con l’uso di python e confluent inserendo i dati in un db grazie ai sink connettors. osservazione dei dati e dei tweet inerenti al mio marchio rispetto a quel specifico tipo di prodotto e a quelli dei competitors. **dati real time** **by twitter**

1. **In Italia?**

Fare filtro in linguaggio Python ed estrarre i dati in una tabella

**DATA VISUALIZATION**

**1 pagina KPI + POSIZIONE GEOGRAFICA**

* **andamento vendite** **e costi/benefici retail rispetto ad anno precedente**
* **andamento vendite e-commerce e costi/benefici rispetto ad anno precedente**
* **prodotto piu venduto negozio**
* **prodotto piu venduto e-commerce**
* **vendite in real time aggiornate 🡪 simulare**
* **andamento generale rispetto al settore moda**
* **andamento in twitter rispetto ad ultimi mesi**
* **scostamento negozi ufficiali/ retail in base al numero vendite e al fatturato**
* **scostamento mercato femminile rispetto a quello maschile: numero utenti e fatturato**
* **Quante volte un mio cliente ritorna a comprare da furla?**
* **Controllo Qualità annuale ( rapporto Reso/Prodotti venduti)**
* **Posizione geografica con scelta differenza colori nuovi/vecchi negozi e ufficiali/retail**

**2 pagina FLUSSI**

* **Andamento flusso di gente in negozio per fascia oraria…per ottimizzare costi di personale e ottimizzare le ore e i giorni di consegna. FILTRO: GIORNO SETTIMANALE**
* **fascia oraria che richiede piu personale**
* **fascia oraria ideale per la consegna della merce**
* **Differenza vendite prodotti in saldo rispetto ai prodotti non scontati**
* **Miglior paese per fatturato: ASIA**
* **andamento vendite con opzioni di filtro**
* **mese, trimestre, semestre, anno**
* **categoria prodotto**
* **Country: EUROPA vs ASIA**

**3 pagina ITALIA**

* **vendite in real time aggiornate**
* **fatturato totale e quale percentuale occupa del totale**
* **andamento generale rispetto al totale numero di vendite**
* **andamento in twitter…ne parla la gente? Filtro (language: Italian)**
* **andamento vendite con opzioni di filtro mese, trimestre, semestre, anno**
* **andamento previsto prossimi 5 anni**
* ​**TOP-WORST 10 negozi per il mondo per fatturato INDICANDO TIPO**

**4 pagina SNEAKERS**

* **vendite in real time aggiornate**
* **fatturato totale derivante solo dalle sneakers**
* **andamento generale rispetto al totale numero d vendite**
* **quale percentuale occupa del totale**
* **andamento in twitter…ne parla la gente?**
* **andamento vendite con opzioni di filtro mese, trimestre, semestre**
* **andamento previsto prossimi 5 anni**
* **scostamento negozi ufficiali/ retail in base al numero vendite e al fatturato**
* **scostamento mercato femminile rispetto a quello maschile: numero utenti e fatturato**
* **TOP-WORST 5 paesi**

**5 pagina BEST – WORST**

* ​**TOP-WORST 10 negozi per il mondo per fatturato INDICANDO TIPO**
* **Categoria prodotti Migliore (come crescita annuale)**
* **TOP-WORST 5 prodotti venduti**
* **TOP-WORST 5 paesi**
* **Pareto**

**DATI CHE MI SERVONO:**

1. Prodotto (id, descrizione, categoria, data di uscita, stagionalità)
2. Cliente (id, nome ,cognome, eta, sesso, nazionalità, count acquisti, count visite, premiumBoolean)
3. Negozi (id, lat, long, descrizione, tipo)
4. FACT Scontrini (id, id\_data, prodotto, id\_cliente, id\_magazzino, valore, quantità)
5. Data (id, giorno, mese, trimestre, semestre, anno)
6. Magazzino (id, descrizione, movimenti, scambio tra negozi differenti, merce restituita (resi)) -> INGLOBARE IN UNA UNICA TABELLA

DC STORE CLOSE (giacenza)

MARKDOWN (sconti),

MOTIVA SCONTO RIGA (causale\_sconti)

RTV (scarti)

1. on order to DC (forecast per magazzino)
2. on order to store (forecast per store)
3. Vendite e-commerce
4. Real time vendite retail
5. Real time tweet in base alla localizzazione (Italia principalmente) che parlano di furla
6. Real time tweet di furla rispetto ai competiors
7. Dati pubblici commerciali: bilancio, Dati ISTAT