

Telco 4.0: analisi del traffico Internet

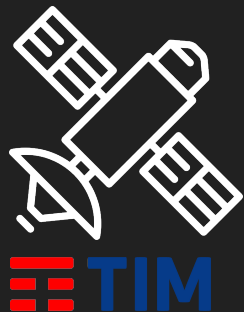
Manutenzione predittiva per l'identificazione
di anomalie nel traffico internet

Agenda

- **Introduzione**
 - Business Case
 - Obiettivo
 - Dataset
- **Data Exploration & Preprocessing**
 - Caratteristiche Training e Test set
 - Problemi dei Dati
 - Preprocessing
 - Risultati Preprocessing
- **Approccio metodologico**
 - Manutenzione Condition Based vs Predittiva
 - Il nostro approccio
 - Risultati
- **Conclusioni**

Introduzione

Business Case



- Parte della rete di **FASTWEB** è in comune con **TIM**
- Grazie all'analisi dei dati di traffico si vuole predire possibili rotture sull'infrastruttura per prevenire le lamentele dei clienti

Introduzione

Obiettivo

Obiettivo è quello anticipare un possibile crollo dell'utilizzo del traffico dati collegato ad un malfunzionamento dell'infrastruttura



Introduzione

Dataset



- Training-set
 - 1977 serie temporali:
 - Lunghezza minima 1440
 - Massima 8636



- Test-set
 - 2121 serie temporali:
 - Lunghezza minima 1143
 - Massima 8048

Introduzione

Dataset



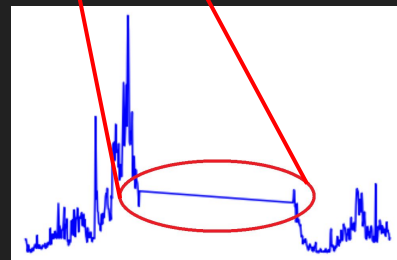
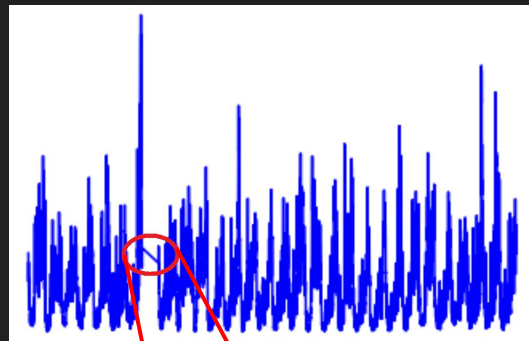
Frequenza dati ogni 5 min



Missing temporali non segnalati nel dataset

Variabili:

- TS
- USAGE
- KIT_ID
- AVG_SPEED_DW
- NUM_CLI
- VAR_CLASS (target)



Data Exploration & Preprocessing

Caratteristiche Training-set

osservazioni:

16.522.034



472 con valore 2



36 con valore 1



Solo 3 serie hanno almeno un valore diverso da 0

Data Exploration & Preprocessing

Caratteristiche Test-set

osservazioni:

16.926.846



626 con valore 2



60 con valore 1

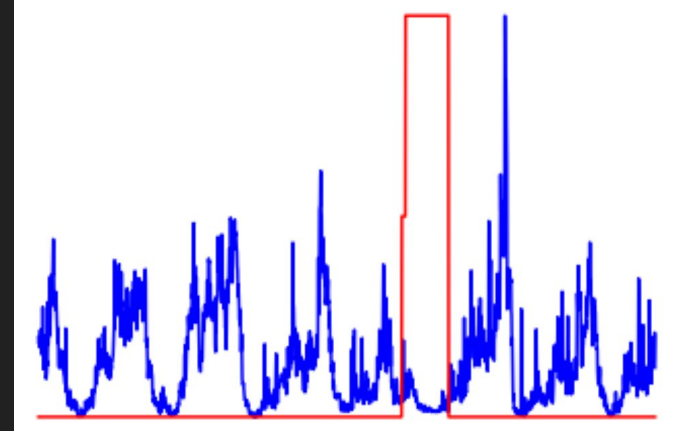
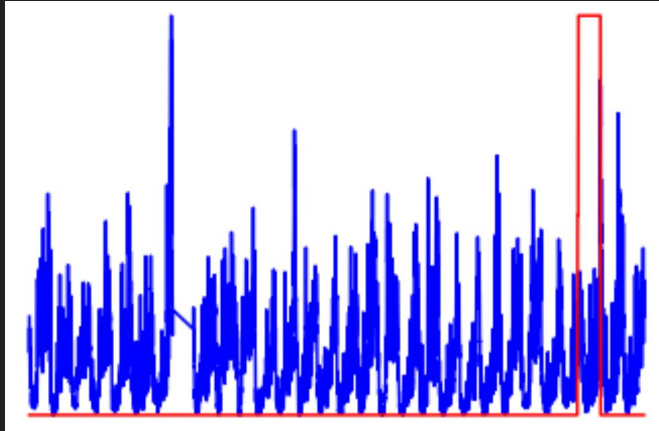


Solo 5 serie hanno almeno un valore diverso da 0

Data Exploration & Preprocessing

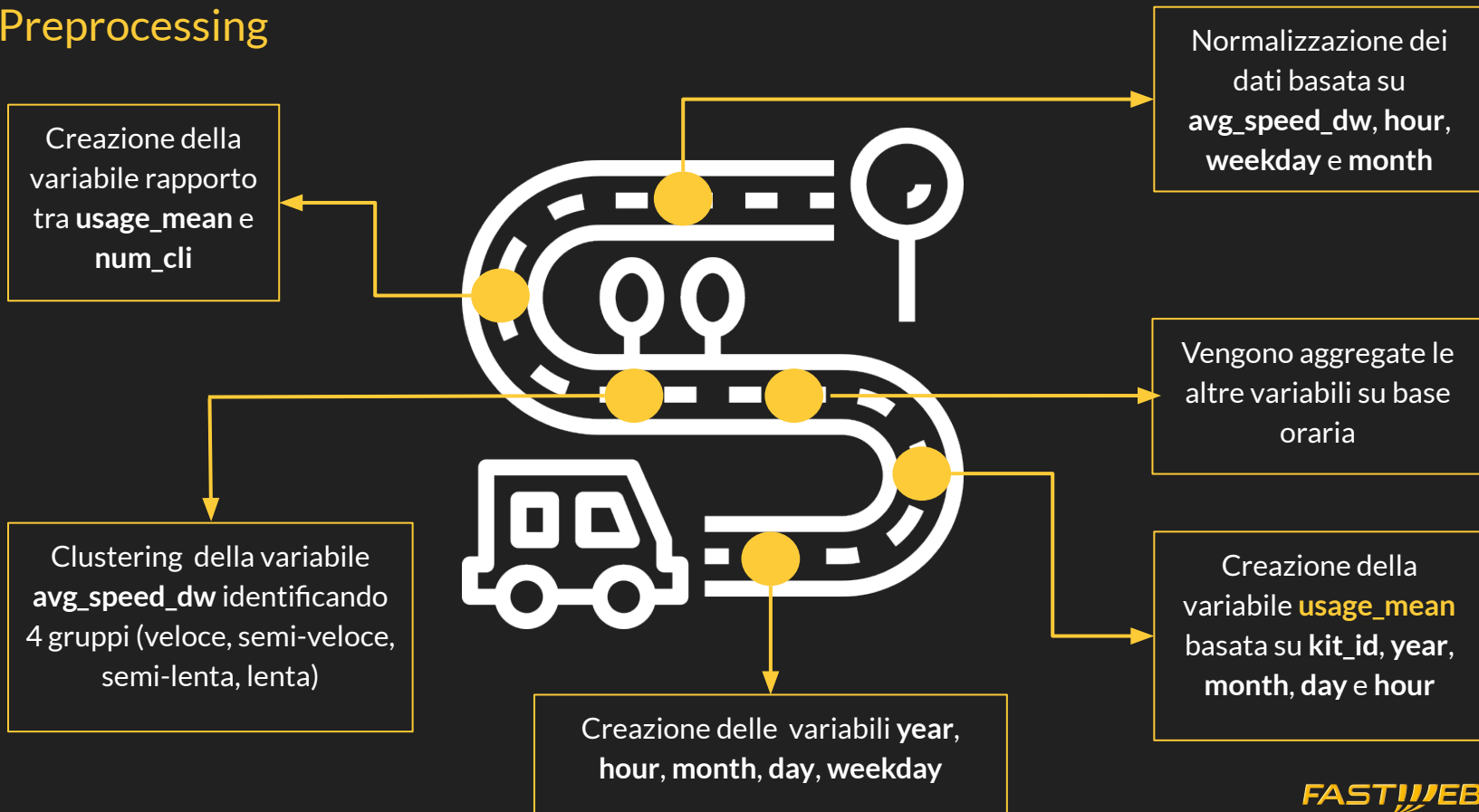
Problemi dei Dati

! Dai dati reali **non sempre è possibile** identificare dei pattern anomali in presenza di una failure



Data Exploration & Preprocessing

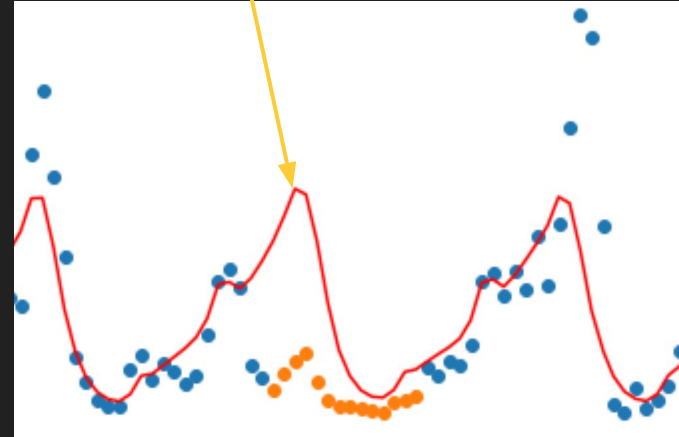
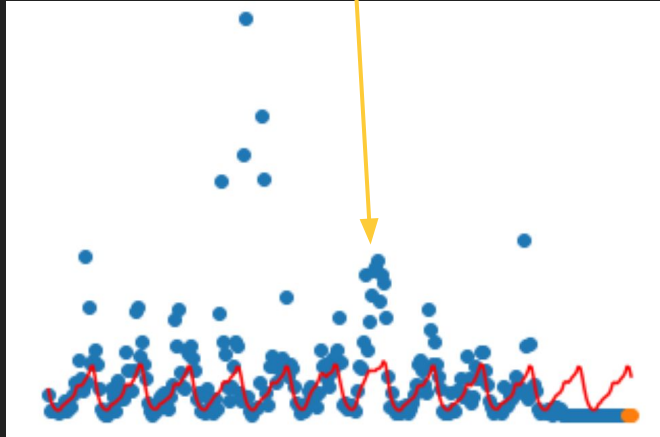
Preprocessing



Data Exploration & Preprocessing

Risultati Preprocessing

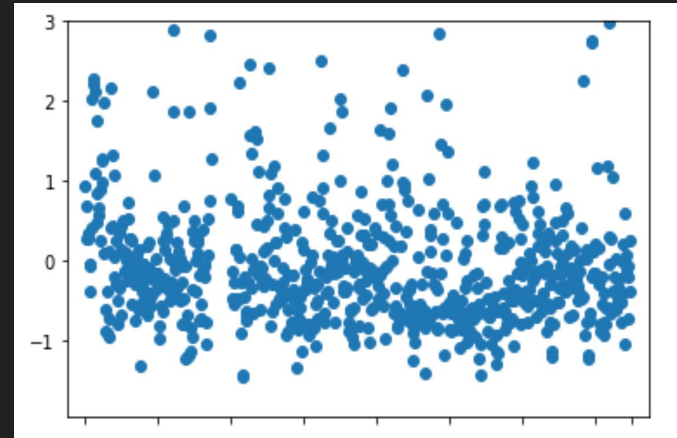
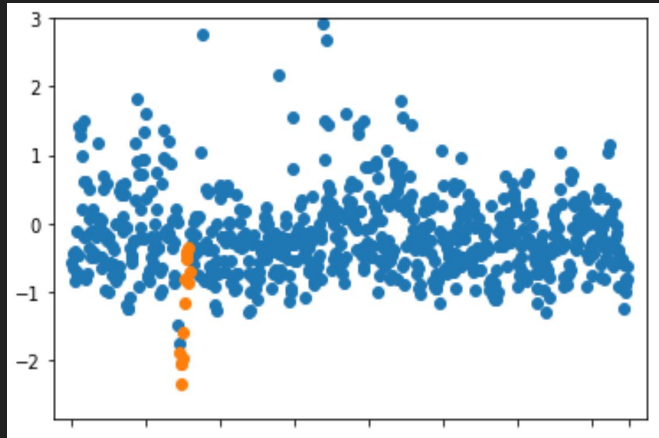
Dati reali confrontati con l'andamento medio



Data Exploration & Preprocessing

Risultati Preprocessing

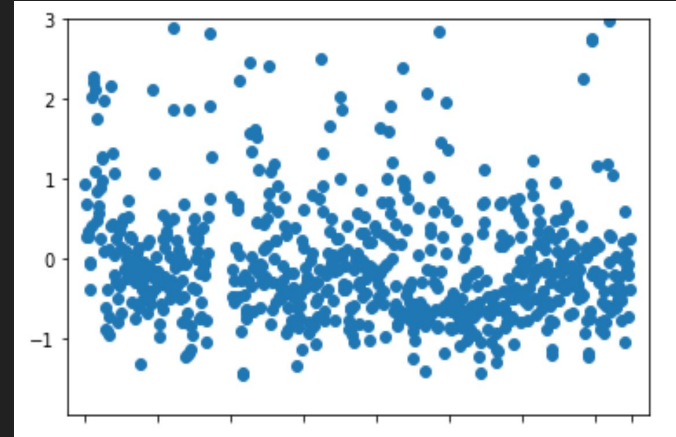
Dati **normalizzati**



Data Exploration & Preprocessing

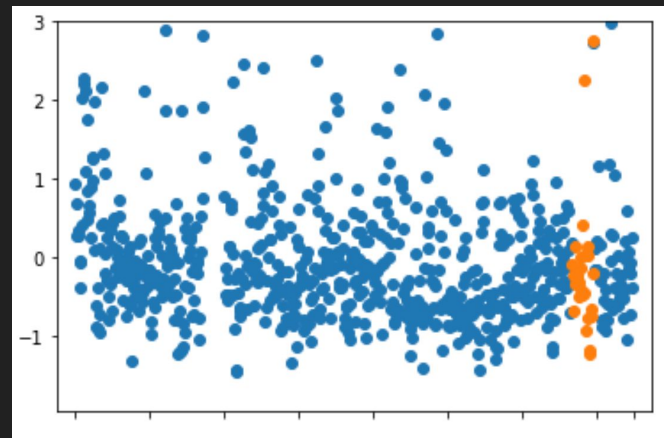
Risultati Preprocessing

Dov'è l'**anomalia** ?



Data Exploration & Preprocessing

Risultati Preprocessing



Approccio Metodologico

Manutenzione Condition Based vs Predittiva

Anche applicando tecniche di oversampling / undersampling è **complicato utilizzare dati così sbilanciati**



Inoltre, le variabili esplicative **non sembrano essere sufficienti** a spiegare completamente il fenomeno

Approccio Metodologico

Manutenzione Condition Based vs Predittiva

L'utilizzo di tecniche più semplici ispirate alla manutenzione **condition-based** potrebbero essere più efficienti rispetto alla modellistica tipica della manutenzione predittiva



Possibile **soluzione**:

utilizzo di soglie per
l'identificazione di
comportamenti anomali



Approccio Metodologico

Il nostro approccio

2

serie anomale su

3

identificate nella fase di training



Utilizzando una **soglia** pari a
-2 sui dati **standardizzati**

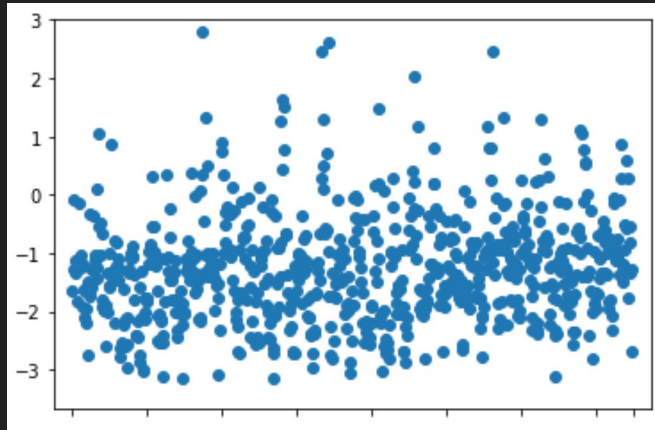


Purtroppo, data la
natura dei dati,
vengono trovati
anche molti **False
Positive**

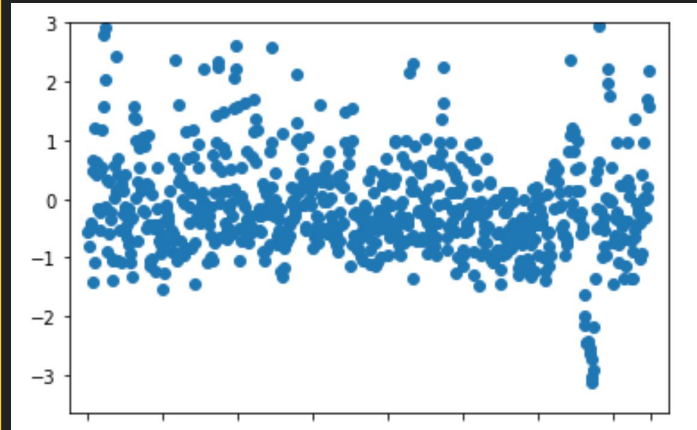
Approccio Metodologico

Il nostro approccio

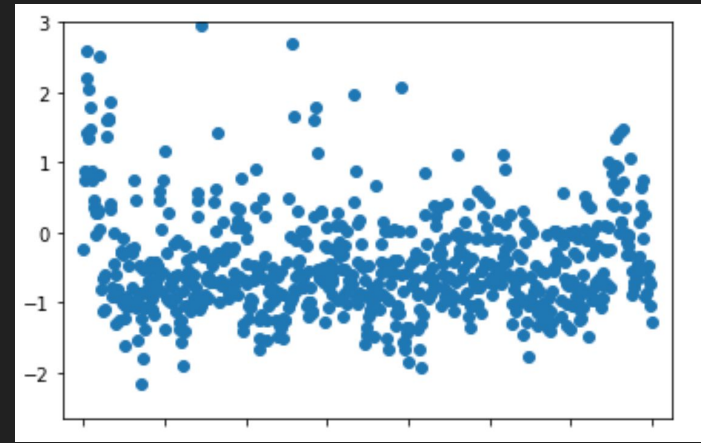
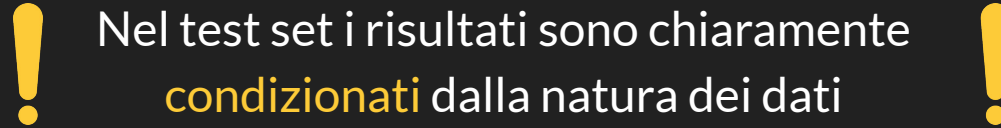
KIT_ID con comportamenti
anomali costanti



Comportamento anomalo
NON classificato come failure



Risultati

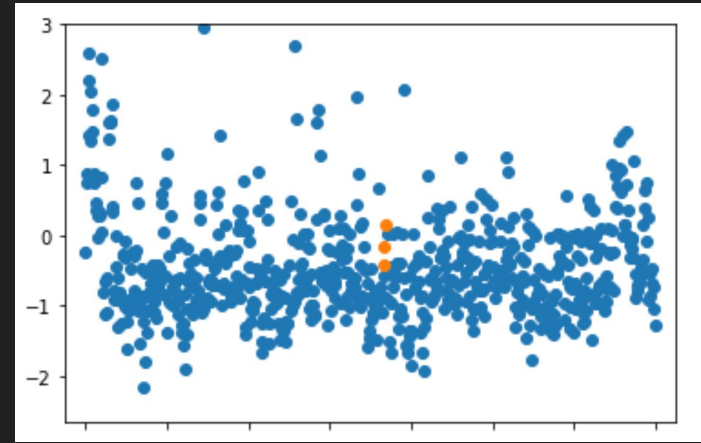
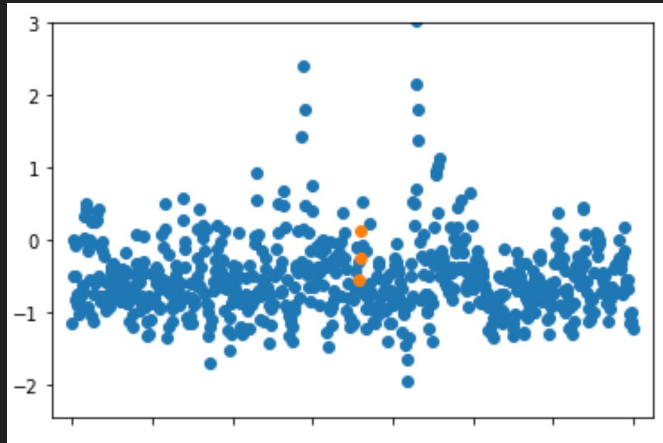


Approccio Metodologico

Risultati



Nel test set i risultati sono chiaramente
condizionati dalla natura dei dati



Conclusioni

Possibili **Miglioramenti / Soluzioni:**

Clusterizzazione KIT_ID



Integrare con dati clienti



Più dati con anomalie per
applicare tecniche predittive

Grazie!

FAST!!!WEB