

The Business Game for Data Scientists

Istruzioni

Career Service UniPD &
Bee-Viva

13 Aprile 2017
Dipartimento di Scienze Statistiche
Università degli Studi di Padova



Sommario

① BusinessGame @UniPD

② Forecast Ticket Closing Time

③ I Dati

④ Real time feedback



Chi

UniPD

- Career Service
- Bee Viva (spin-off UniPD)

10 Aziende

- 2C solution
- Accenture
- Business Integration Partners
- Cognitive Technology
- Infocamere
- I.T. Euroconsulting
- KPMG
- Méthode
- Pam Panorama
- Unox.



Come funziona

- *Mattina:*
 - Presentazione del Business problem
 - L'intera mattina per presentare il vostro miglior modello di predizione
 - Valutazione del vostro operato da parte di bee-viva
- *Pomeriggio:* colloqui con le aziende per i migliori classificati



Sommario

① BusinessGame @UniPD

② Forecast Ticket Closing Time

③ I Dati

④ Real time feedback



Business Game

I business games sono delle competizioni caratterizzate da un contesto simulato di natura aziendale dove i partecipanti si confrontano con problematiche manageriali.

The Business Game for Data Scientist si distingue nell'affrontare il business problem con un approccio data driven.



Forecast Ticket Closing Time

Il perfezionamento dei processi aziendali è sempre stato un tema rilevante per il management.

La sfida che vi proponiamo è quella di stimare il **tempo necessario per la chiusura di un intervento** di assistenza da parte di un noto gruppo automobilistico.

La Business Unit After Sales gestisce queste richieste attraverso un sistema di ticketing. Il tempo necessario al loro completamento può richiedere da una giornata fino ad oltre un mese.

Ci sono molti fattori in gioco: la gravità dell'anomalia e le difficoltà dell'intervento, i feedback del committente (un ticket può essere riaperto perché la soluzione non è ritenuta soddisfacente dal cliente), le capacità tecniche e di gestione dell'unità organizzativa preposta, ecc. . .

Gli interventi possono essere effettuati su scala mondiale. Ragione per cui prevedere tempi e costi permette al management di valutare ed eventualmente riorganizzare l'attività delle unità locali.



Sommario

① BusinessGame @UniPD

② Forecast Ticket Closing Time

③ **I Dati**

④ Real time feedback



I Dati

I dati, anonimizzati, descrivono oltre 100.000 interventi d'assistenza (ticket):

- 80825 costituiscono il **training set**;
- i restanti 34639 costituiscono il **test set**; i dati su cui dovrete stimare il tempo necessario per la chiusura dell'intervento di assistenza.

L'accuratezza delle vostre previsioni verrà valutata utilizzando il **MAE** (mean absolute error - errore medio assoluto).



Legenda (1/2)

CASEID	Id Ticket
TARGET	Tempo necessario per la chiusura dell'intervento di assistenza (ore)
TKT_START	Timestamp attivazione del Ticket (espresso in ore)
MERCATO	Area in cui viene risolto il Ticket (20 modalità)
DEA	Dealer/Committente Ticket (9029 modalità)
COD_PR	Tipo Dealer/Committente (11 modalità)
MCALL	Codice motivo apertura Ticket (16 modalità)
TKT_TYPE	Tipologia del Ticket (4 modalità)
GRAVITA	Gravità del Ticket (4 modalità)
DINTERV	Difficoltà Intervento (4 modalità)
MARCA	Marca del veicolo (8 modalità)
MODELLO	Modello del veicolo (135 modalità)
SERIE	Serie del veicolo (19 modalità)
AVARROLE	Livello Risorse incluse nel processo (2 modalità)
AVARUSER	Risorse incluse nel processo (2481 modalità)



Legenda (2/2)

Variant (conteggio dei seguenti eventi per ogni Ticket)

Apertura	Evento Apertura
Assegnato	Evento Assegnazione
Attesa_Ricambi	Evento Attesa Ricambi
Attesa_di_Conferma	Evento Attesa Conferma
Attivazione_Specialista	Evento Attivazione Specialista
Caso_Riaperto	Evento Caso Riaperto
Caso_Singolo	Evento Caso Singolo
Cet	Evento Cet
Confirmed	Evento Confirmed
Escalation	Evento Escalation
Feedback_Negativo	Evento Feedback Negativo
Prima_Activazione_Secondo_Livello	Evento 1° Attivazione Secondo Livello
Riapertura	Evento Riapertura
Secondo_Livello_in_Uscita	Evento Secondo Livello in Uscita
Soluzione_Non_Efficace	Evento Soluzione Non Efficace



Sommario

① BusinessGame @UniPD

② Forecast Ticket Closing Time

③ I Dati

④ Real time feedback



Real time feedback

- Ogni submission produce un **MAE parziale**, basato su un campione casuale predeterminato di 12.000 clienti presenti nello score set
- Questi punteggi creano una classifica parziale consultabile online dai partecipanti
- Questo fornisce un feedback sulla qualità delle soluzioni prodotte, ma . . . attenti al demone dell'overfitting!!



Data Scientist

A data scientist is someone who is better at statistics than any software engineer, and better at software engineering than any statistician.

