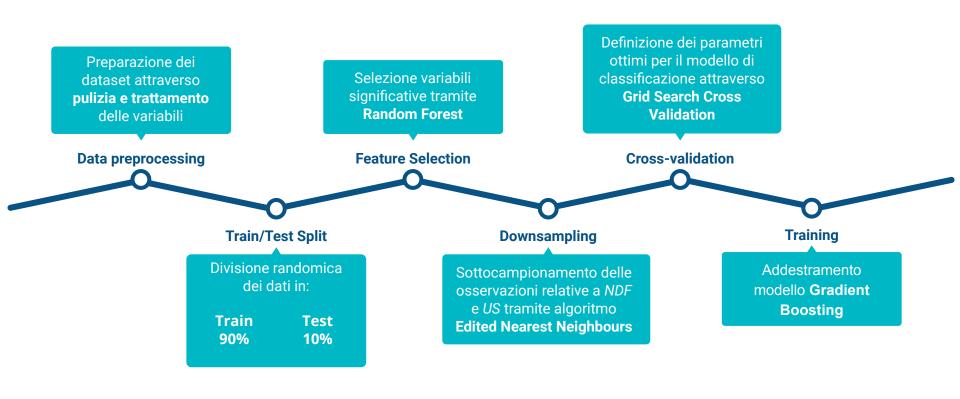
ConTe.it

First user booking

Creazione dei dataset

Variabili estratte Datasets ottenuti Dataset "users_no_sessions" Categoriali - sesso, lingua, primo device usato.... Contiene tutti gli utenti presenti nel file "train users" con le Numeriche - età e signup_flow Files colonne estrapolabili solo dallo stesso file Data - data di iscrizione, di primo accesso alla piattaforma e distanza temporale tra le train_users.csv due Quantitative - numero totale azioni per Dataset "users sessions" sessions.csv utente, numero totale azioni per tipologia, numero totale device utilizzati, ... Contiene tutti gli utenti le cui sessioni online siano state Di Durata - tempo totale speso per singolo tracciate, con le variabili di device, per singola tipologia di azione, durata entrambi i file di origine media per singolo device, ...

Approccio al problema



Risultati

Matrice di confusione del Test Set



Il modello indovina il 66% delle sue previsioni NDF



Il modello intercetta l' 89% delle osservazioni NDF reali

		Previsti				NDF			US	other	
0	0	0	0	0	0	0	34	0	0	15	0
0	0	0	0	1	0	0	80	0	0	44	2
0	0	0	0	0	0	0	67	0	0	41	1
0	0	0	0	1	0	0	145	0	0	89	0
0	0	0	0	2	0	0	332	0	0	168	8
0	0	0	0	2	0	0	168	0	0	75	2
0	0	0	0	2	0	2	185	0	0	79	1
0	0	0	0	7	0	1	11062	0	0	1324	40
0	0	0	0	0	0	0	50	0	0	32	1
0	0	0	0	0	0	0	12	0	0	7	1
1	0	0	0	7	0	0	3871	0	0	2302	55
0	0	0	1	2	0	0	667	0	0	336	21
	0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0	0 0 0 0 1 0 0 0 0 0 0 0 0 0 1 0 0 0 0 2 0 0 0 0 2 0 0 0 0 7 0 0 0 0 0 0 0 0 0 0 1 0 0 0 7	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 2 0 0 0 0 0 2 0 0 0 0 0 2 0 0 0 0 0 7 0 0 0 0 0 0 0 1 0 0 0 7 0	0 0	0 0 0 0 0 34 0 0 0 0 0 34 0 0 0 0 0 80 0 0 0 0 0 67 0 0 0 0 0 0 67 0 0 0 0 0 0 145 0 0 0 0 2 0 0 332 0 0 0 0 2 0 0 168 0 0 0 0 2 0 2 185 0 0 0 0 0 0 50 0 0 0 0 0 0 12 1 0 0 0 0 0 3871	0 0 0 0 0 34 0 0 0 0 0 0 34 0 0 0 0 0 0 0 0 0 0 0 0 0 67 0 0 0 0 0 0 67 0 0 0 0 0 0 0 145 0 0 0 0 0 2 0 0 332 0 0 0 0 2 0 0 168 0 0 0 0 2 0 2 185 0 0 0 0 0 0 0 50 0 0 0 0 0 0 0 50 0 0 0 0 0 0 0 12 0 1 0	0 0 0 0 0 0 34 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 67 0 0 0 0 0 0 0 0 67 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 2 0 0 145 0 0 0 0 0 2 0 0 332 0 0 0 0 0 2 0 168 0 0 0 0 0 2 0 1 11062 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 34 0 0 15 0 0 0 0 0 34 0 0 15 0 0 0 0 0 80 0 0 44 0 0 0 0 0 67 0 0 41 0 0 0 1 0 0 67 0 0 41 0 0 0 1 0 0 145 0 0 89 0 0 0 2 0 0 332 0 0 168 0 0 0 2 0 0 168 0 0 75 0 0 0 2 0 2 185 0 0 79 0 0 0 0 0 1 11062 0 0 1324

Drovisti

Risultati

User	P1	P2	P3	P4	True
q9p###	US	other	NDF	GB	other
9bz###	NDF	us	other	GB	US
kux###	NDF	US	other	AU	US
qqn###	NDF	US	other	FR	NDF
nw9###	NDF	US	other	FR	NDF
oop###	NDF	US	other	FR	IT
a62###	NDF	other	US	FR	NDF
611###	NDF	other	US	FR	FR
qbs###	NDF	us	FR	other	US
f2k###	NDF	us	other	FR	US

Ranking

La tabella contiene:



User - nome utente (parzialmente oscurato)



P1_P4 - le prime 4 previsioni per singolo utente in ordine di probabilità, partendo dalla più probabile (P1)



True - la destinazione effettiva

Nel campione di test estratto a sinistra, per **9 utenti su 10** il modello attribuisce una buona probabilità alla destinazione effettiva. Nella popolazione totale di test, tale rapporto si attesta a **9,4 su 10**.

Il modello, quindi, si dimostra essere in grado di fornire una **lista accurata** di (poche) possibili destinazioni.

Conclusioni

Summary

Previsione Ranking Linee guida Il modello addestrato non si è Tuttavia, quando si considera il Se l'obiettivo è una campagna ranking attribuito dal modello marketing mirata, **si** dimostrato in grado distinguere tra le osservazioni base alle probabilità. suggerisce di focalizzare più frequenti (NDF, US e other) e osserviamo come lo stesso sia sforzi e risorse verso i primi 4 gli altri Paesi di destinazione... in grado di fornire una lista Paesi previsti per ogni molto accurata di possibili utente. destinazioni.

Possibili miglioramenti

In fase di analisi è sembrato evidente come l'impossibilità di ottenere le informazioni delle attività online di ogni utente prima del 2014 sia stato un grosso limite per l'addestramento. Se tale limitatezza è dettata dall'esigenza di contenere le dimensioni dei file, si potrebbe contenere la selezione degli utenti ad un solo anno ma rendere disponibili le informazioni sulle sessioni online per tutti.