

Progetto Supervised Learning

- Imperiale Angelo
- Turilli Francesco
- Acquaviva Sabino
- Pace Nicolò
- Mutzu Martis Massimiliano

Introduzione

La nostra azienda ha registrato negli ultimi periodi un numero altissimo di perdita di clienti.

Tali clienti hanno cessato il contratto con TTM per attivarne un altro con altri competitor: la causa di tale perdita di clienti è da imputare a logiche di mercato interne a TTM.

La join tra il team di Marketing Analysis e quello CRM ha permesso di approcciare al problema utilizzando nuovi punti di vista che finora non sono stati mai presi in considerazione.

Obiettivi

Gli obiettivi sono :

- ▶ Mettere in campo delle attività di retention Data Driven
- ▶ Mettere in campo un modello che permetta di identificare i probabili clienti pronti ad effettuare il Churn

Analisi preliminare del dataset

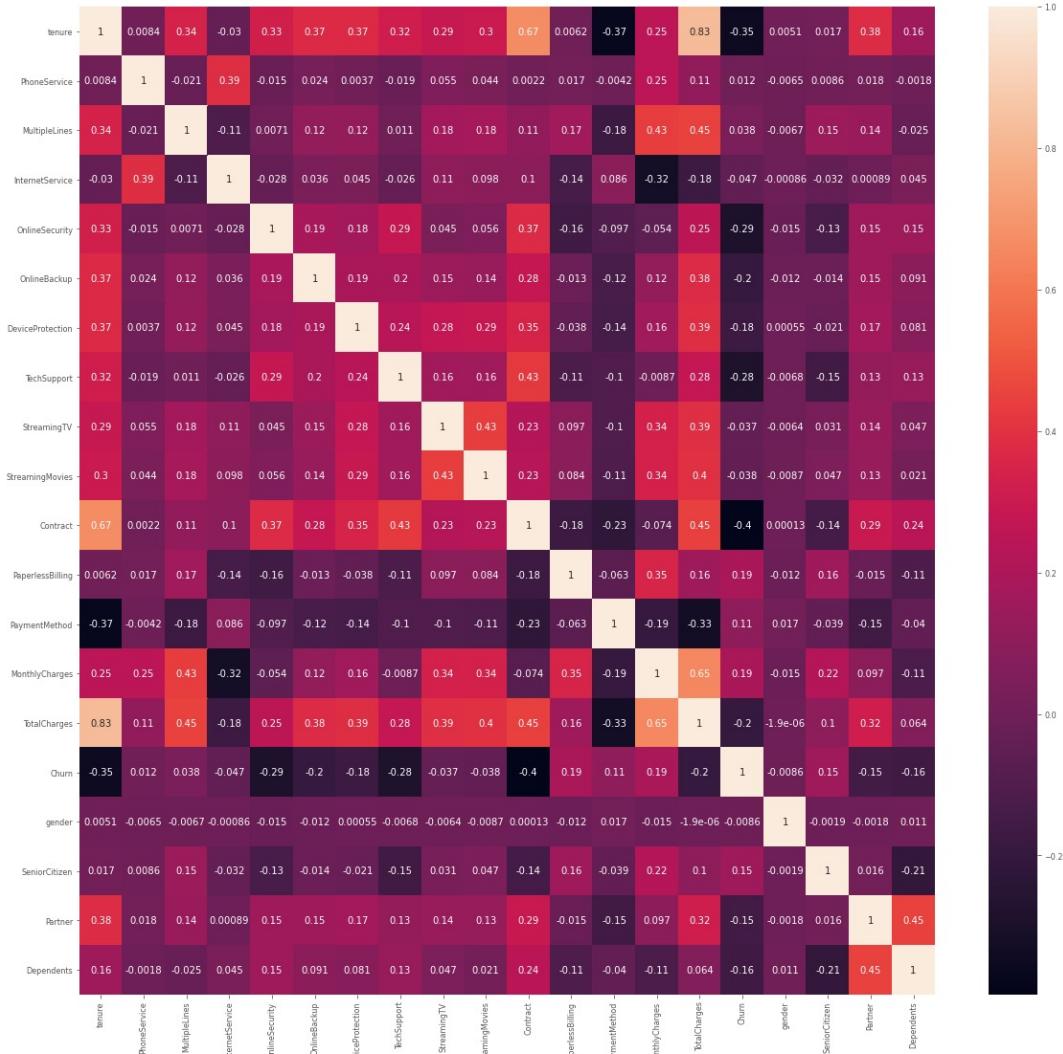
La prima procedura è lo studio dei dataset

Il progetto prevede l'analisi di due insiemi quindi preventivamente procediamo al merge dei dati e all'individuazione dei valori NA

Sono stati individuati 11 valori NA, che hanno una ratio sull'intero dataframe dello 0.002.

Secondo la nostra analisi i valori NA sono da attribuire a clienti nuovi che non hanno ancora fatturato.

Analisi preliminare del dataset

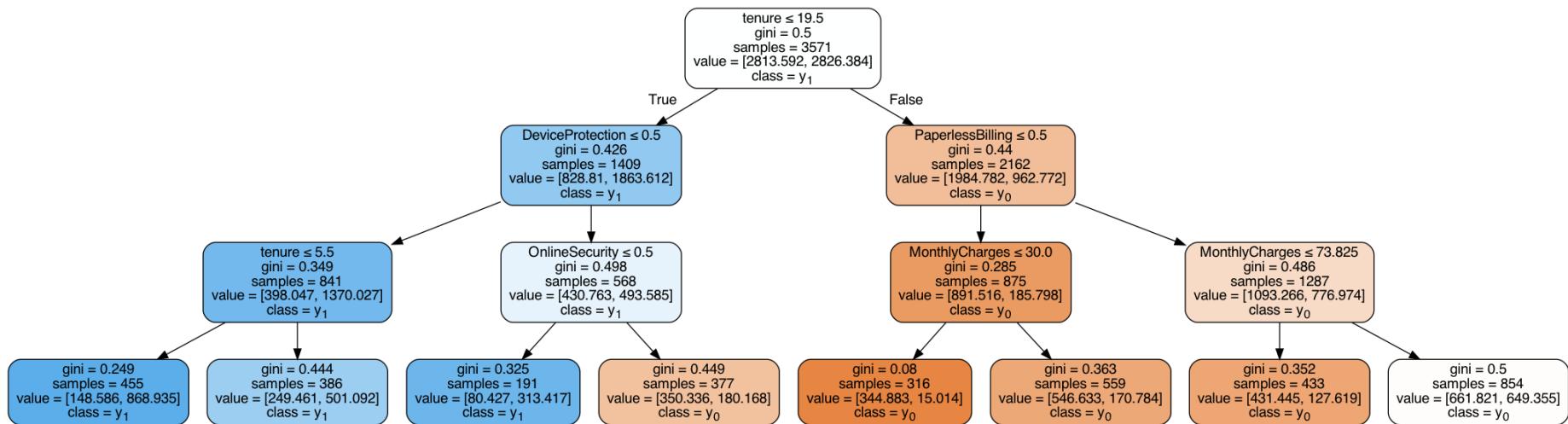


Dalla matrice di correlazione si nota scarsa correlazione fra tutte le variabili

Analisi dei modelli predittivi

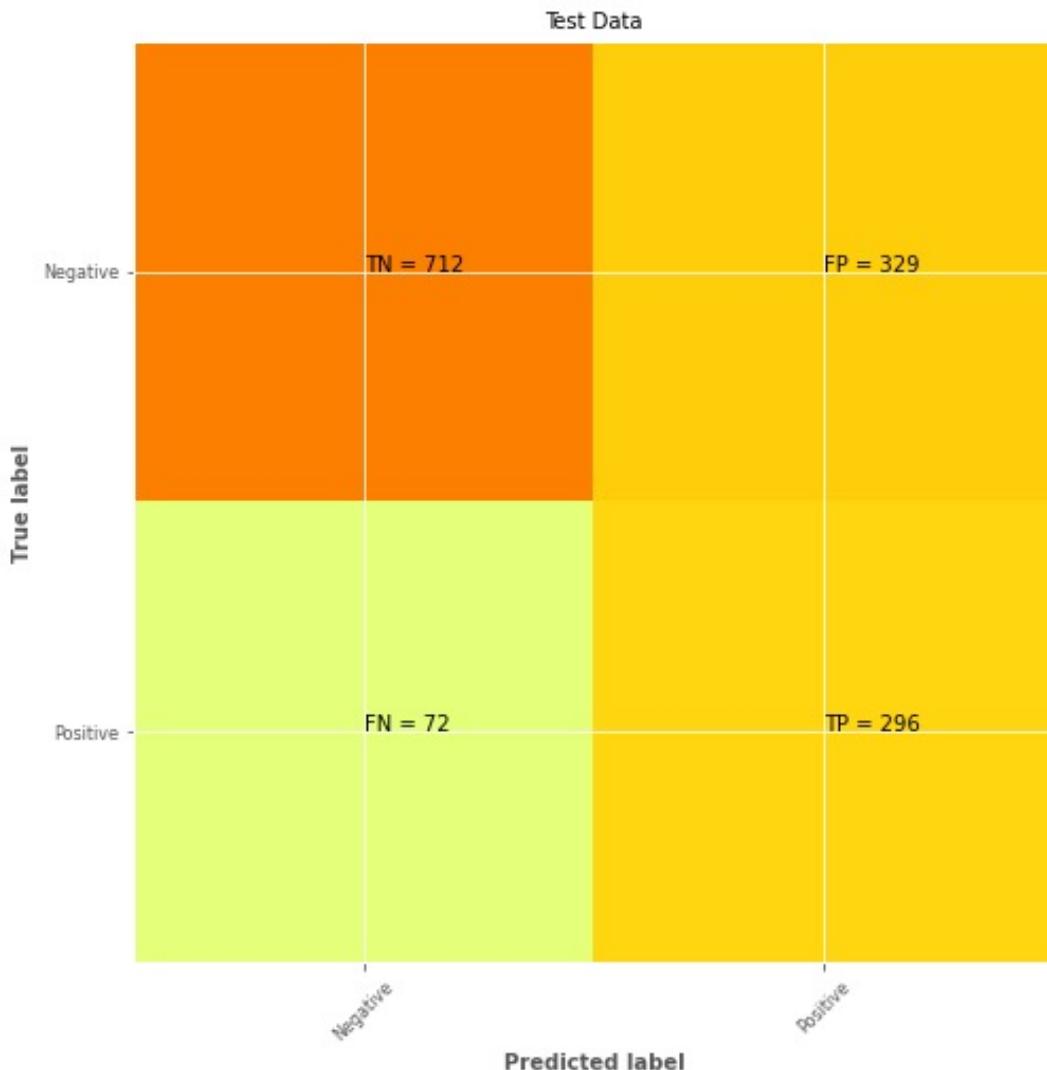
Il primo modello utilizzato per la nostra analisi è la Linear regression utilizzando un unico regressore.

- Ma l' R^2 del modello è < 0,3 ragione per cui il modello risulta inefficace.
- Vista la prevalenza di variabili qualitative si è proceduto ad utilizzare il Random Forest.



Random Forest

Confusion Matrix



Il False Negative Rate che è il parametro che a noi interessa per identificare i parametri di retention è pari a 19,5%, un valore ritenuto da noi accettabile per una corretta classificazione.

Questo rappresenta il risultato migliore raggiunto tra i diversi modelli addestrati

Parametri di bontà del modello

- L'accuracy score è superiore al 70%, ossia il modello ha una percentuale di affidabilità oltre il 70%

```
print(classification_report(y_test, y_pred))
```

	precision	recall	f1-score	support
0	0.91	0.68	0.78	1041
1	0.47	0.80	0.60	368
accuracy			0.72	1409
macro avg	0.69	0.74	0.69	1409
weighted avg	0.79	0.72	0.73	1409

Valutazioni (1)

Dal modello Random Forest si evince che:

- Dobbiamo intervenire nella fase iniziale (sotto i 20 mesi), subito dopo l'acquisizione del cliente (che è la fase temporale in cui si perdono la maggior parte).
- Se un cliente ha superato i 20 mesi la fatturazione digitale fidelizza maggiormente e un MonthlyCharges <=30 assicura una percentuale ottima di permanenza.
- I contratti che non prevedono il servizio di internet (pure attivati da poco tempo) risultano abbastanza fidelizzati.
- A chi necessita il servizio internet bisogna garantire la DeviceProtection e l'OnlineSecurity affinchè venga garantita la fidelizzazione.

Valutazioni (2)

Con il modello utilizzato abbiamo predetto gli eventuali clienti che potrebbero rescindere il contratto.

Nella slide successiva se ne può notare una preview.

La nostra previsione è motivata da una sensitività (la percentuale della previsione dei positivi corretti) del 80% come si evince dalla confusion matrix e dai parametri di bontà del modello.

Per confronto si noti che la specificità (la percentuale della previsione dei negativi corretti) del modello è del 68%.

Preview Predizioni

	customerID	tenure	MonthlyCharges	DeviceProtection	OnlineSecurity
4446	6161-UUUTA	3	79.90	No	No
5454	2506-TNFCO	1	69.55	No	No
5716	8903-XEBGX	2	55.30	No	No
134	2372-HWUHI	2	44.35	No	No
3731	2897-DOVND	2	69.80	No	No
...
1651	8580-AECUZ	1	79.10	No	No
5889	9693-XMUOB	59	50.25	No	No
2453	1197-BVMVG	4	74.45	No	No
2631	6861-XWTWQ	7	99.25	Yes	No
3065	5133-POWUA	1	45.80	No	No