

Modello di predizione del dissesto dei comuni

Framing del problema:

Lo stato degli enti locali si dividono in tre gruppi di sofferenza finanziaria. I comuni deficitari, quelli in dissesto e quelli in salute.

Sono **deficitari** quegli enti che sfiorano almeno cinque dei dieci parametri stabiliti dal decreto ministeriale del 18 febbraio 2013, per esempio un saldo negativo del risultato contabile di gestione superiore al 5% delle entrate correnti, oppure l'eccessiva quota di residui attivi o passivi in relazione a spese.

Il "default", meglio definito come "**dissesto finanziario**" si verifica quando un comune caratterizzato da gravi ed incontrovertibili condizioni di squilibrio finanziario. Più semplicemente: un comune in dissesto non riesce più a svolgere le sue funzioni di base e a fornire alla cittadinanza quei servizi considerati indispensabili.

Il dissesto è una cosa ben diversa dal fallimento di un'impresa privata in quanto non si può determinare l'estinzione del Comune proprio perché gli enti locali non possono cessare di esistere come una semplice impresa privata, ma bisogna garantire la continuità amministrativa.

Nel momento in cui viene dichiarato il dissesto del comune, sindaco, giunta e consiglio resterebbero in carica ma verrebbero coadiuvati da una commissione espressamente designata dal Ministero degli Interni.

La commissione si occuperebbe del disavanzo pregresso, mentre l'amministrazione gestirebbe il bilancio "risanato".

La sola ipotesi di commissariamento del Comune si verificherebbe nel caso in cui l'amministrazione non dovesse approvare il bilancio di previsione (la cui scadenza è alla fine del mese di maggio).

L'eventuale dichiarazione del dissesto di fatto congelerebbe invece la scadenza del bilancio stesso, mettendo in moto una procedura del tutto diversa per la definizione e l'approvazione del bilancio stesso; le conseguenze maggiori del dissesto finanziario si hanno sotto il profilo contabile.

Viene chiesto all'Ente locale di "contribuire" al risanamento attraverso l'adozione di provvedimenti eccezionali.

L'Ente dissestato è tenuto ad approvare un nuovo bilancio, basato principalmente sull'elevazione delle proprie entrate al livello massimo consentito dalla legge, vale a dire che tutte le tasse comunali (IMU, addizionale comunale, TARSU) saranno aumentate il più possibile fino ad arrivare al tetto massimo consentito dalla legge, basato, inoltre, sul contrasto all'evasione e sul contenimento di tutte le spese.

Spese comunali significa innanzitutto personale, la legge prevede che **gli impiegati comunali devono essere nella misura di 1 su 93**, pertanto da questa procedura scaturiranno esuberi di personale che verrà posto in mobilità.

Il comune è altresì tenuto a contribuire all'onere della liquidazione in particolare con **l'alienazione del patrimonio disponibile** non strettamente necessario all'esercizio delle funzioni istituzionali, la destinazione degli avanzi di amministrazione dei cinque anni a partire da quello del dissesto e delle entrate straordinarie, la contrazione di un mutuo a carico del proprio bilancio.

La dichiarazione di dissesto, in breve tempo, è parsa ai politici locali una negatività da evitare al fine di non essere costretti ad emanare provvedimenti così impopolari.

I provvedimenti da adottare in materia di personale e di tributi locali sono ritenuti così pesanti che gli enti arrivano il più delle volte alla dichiarazione di dissesto solo quando, a seguito delle azioni esecutive dei creditori che pignorano le somme della cassa comunale, non è più possibile pagare neppure gli stipendi al personale dipendente.

Obiettivo:

L'obiettivo è quello di costruire il miglior modello in grado di predire il dissesto o meno di un comune in base alle variabili finanziarie, sociopolitiche e geografiche di ciascuno di essi.

Costruzione del dataset:

Il dataset è stato realizzato aggregando i dati necessari da più fonti;

ISTAT:

- per quanto riguarda l'ottenimento dei dati demografici;

Ministero dell'Interni:

- tramite la piattaforma Open Data e il sito del Ministero, è stato possibile procurare informazioni che riguardano la composizione dei consigli comunali, in particolare i nominativi dei sindaci, il loro sesso, il livello di formazione e l'orientamento politico.
- Grazie alla Circolare 20/2020 dell'Ufficio I Consulenza e studi finanza locale, è stato possibile ottenere i comuni dissestati o deficitari fino all'anno 2019.

OpenCivitas:

- la politica di coesione del Fondo Sviluppo e Coesione (ex FAS) che, attraverso risorse nazionali, finanzia progetti strategici, sia di carattere infrastrutturale sia di carattere immateriale, di rilievo nazionale, interregionale e regionale.
- il fondo perequativo è uno strumento che mira a mitigare le diseguaglianze tra Regioni i cui abitanti presentano differenti capacità fiscali, al fine di garantire gli stessi standard di prestazione nell'erogazione dei servizi di competenza, nonostante gli squilibri economico-sociali: il fondo è istituito senza vincolo di destinazione ed è finanziato da quote di entrate tributarie.

MEF/Dipartimento delle finanze:

- ricavando i dati Irpef degli enti locali ha reso possibile aggiungere la variabile che teneva conto del reddito imponibile individuale per ciascuno di loro.

L'unione di queste informazioni ha reso possibile la creazione di un dataset con tutti i comuni italiani e per ciascuno vi sono stati riportati:

23 Indicatori finanziari:

1. grado di autonomia impositiva;
2. grado di dipendenza da contributi e trasferimenti correnti;
3. grado di finanziamento interno;
4. grado di autonomia finanziaria;
5. rigidità della spesa;
6. incidenza spese personale su spese correnti;
7. grado di dipendenza da finanziamento esterno;
8. incidenza spese personale su entrate correnti;
9. incidenza spese per rimborso prestiti su entrate correnti;
10. indice di consistenza iniziale dei residui passivi;
11. grado di copertura delle spese correnti e dei rimborsi prestiti con entrate correnti;
12. incidenza delle spese in conto capitale finanziate tramite mutui e prestiti obbligazionari;
13. capacità di riscossione;
14. capacità di spesa;
15. indice di smaltimento dei residui passivi;
16. avanzo / disavanzo di amministrazione in relazione alle entrate/correnti;
17. debiti fuori bilancio / entrate correnti;
18. rapporto tra i debiti di finanziamento finali e iniziali;
19. trasferimenti correnti / spese correnti;
20. trasferimenti in conto capitale / spese in conto capitale;
21. flussi finanziari per unità di spesa per il personale;
22. alienazione beni patrimoniali / spese correnti;
23. spese esterne per unità di risorse finanziarie.

Vengono considerati i valori medi di ciascun indice negli anni tra 2007 e 2017.

Informazioni sui sindaci:

- Titolo di studio
- Sesso
- Età

Informazione sulla popolazione:

- Media: è stata considerata la media della popolazione residente tra gli anni 2011 e 2017
- Diff: variabile che considera la variazione della popolazione nei 7 anni presi in analisi

Informazione sui singoli enti:

- Stato: dissesto=2, deficitarie=1, altro=0
- Nord: variabile binomiale dove le regioni Piemonte, Lombardia, Valle d'Aosta, Liguria, Veneto, Trentino-Alto Adige, Friuli-Venezia Giulia, Emilia-Romagna, Marche e Umbria, considerate come Nord le restanti come Sud
- Reddito imponibile individuale
- Fondo storico
- Fondo perequativo

Analisi Esplorativa:

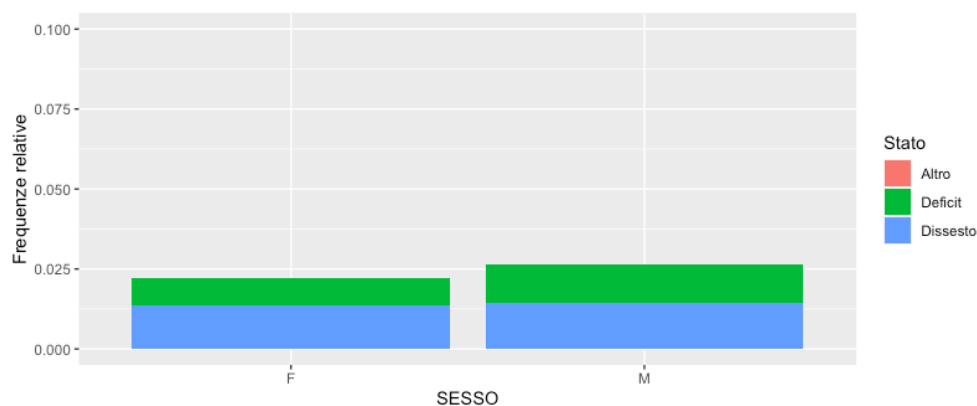
Dall'esplorazione dei dati si evince un forte sbilanciamento verso i comuni in buona salute rispetto a quelli deficitari e dissestati i quali sono settentrionali;

```
> summary(dati)
```

denominazione	Stato	Stato
Length:7545	Altro :7331	nord Altro Deficit Dissesto
Class :character	Deficit : 92	0 2566 63 98
Mode :character	Dissesto: 122	1 4628 23 7

Successivamente si sono fatte ricerche sulle possibili variabili che possano incidere maggiormente sullo stato economico dei comuni, come:

1) Il sesso, l'età e l'istruzione dei sindaci



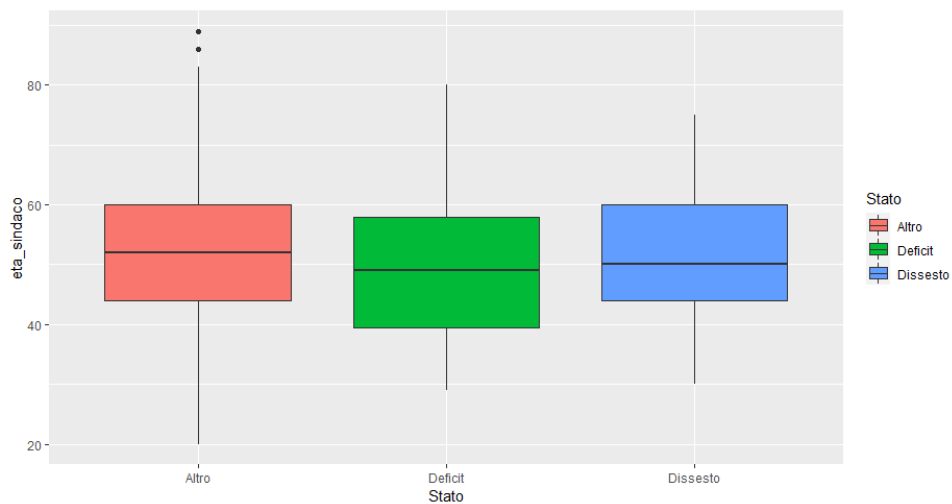
Dal grafico si evidenzia che non c'è sostanziale differenza tra i diversi stati di difficoltà finanziaria e il sesso del sindaco. Questo risultato è supportato dal test chi-quadro che non rifiuta l'ipotesi nulla di indipendenza statistica:

TEST sul sesso dei sindaci

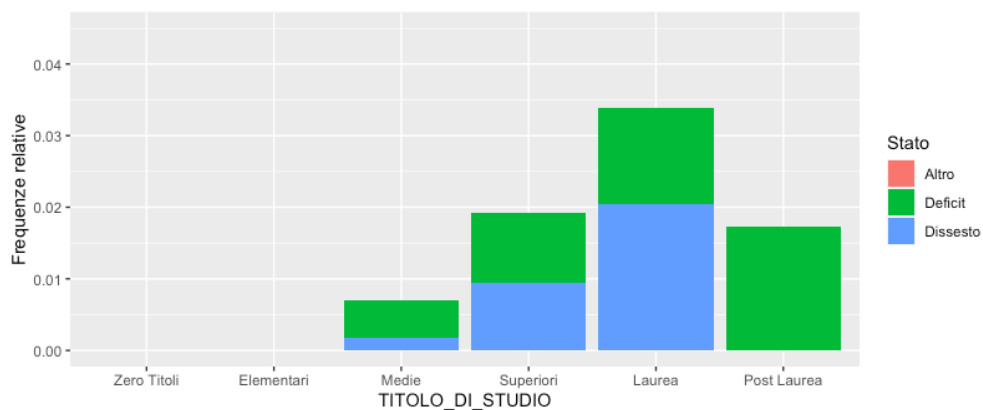
Pearson's Chi-squared test

data: tb

X-squared = 1.021, df = 2, p-value = 0.6002



Il boxplot mostra una maggiore dispersione nell'età del sindaco nel caso di comuni non in dissesto né deficitari. Questo risultato può trovare giustificazione nel fatto che il numero di comuni etichettati con "Altro" è notevolmente maggiore nel dataset.



Il grafico a barre è stato costruito considerando la frequenza relativa pari ad 1 per ogni categoria di titolo di studio del sindaco. Dai dati disponibili risulta che i sindaci dei comuni in dissesto e in deficit abbiamo una laurea o un diploma di scuola superiore con una frequenza più alta rispetto alle altre categorie. Questo risultato potrebbe essere giustificato nel fatto che i sindaci senza titolo o con titolo di studio elementare sono rari. Svolgendo un test chi-quadro risulta che c'è dipendenza tra titolo di studio del sindaco e stato di salute del comune che amministra.

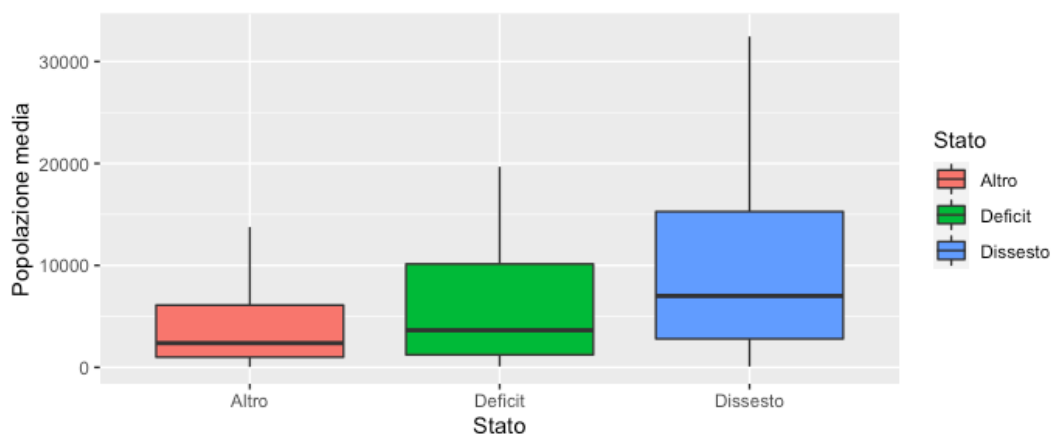
TEST sul titolo di studio dei sindaci

Pearson's Chi-squared test

data: tb1

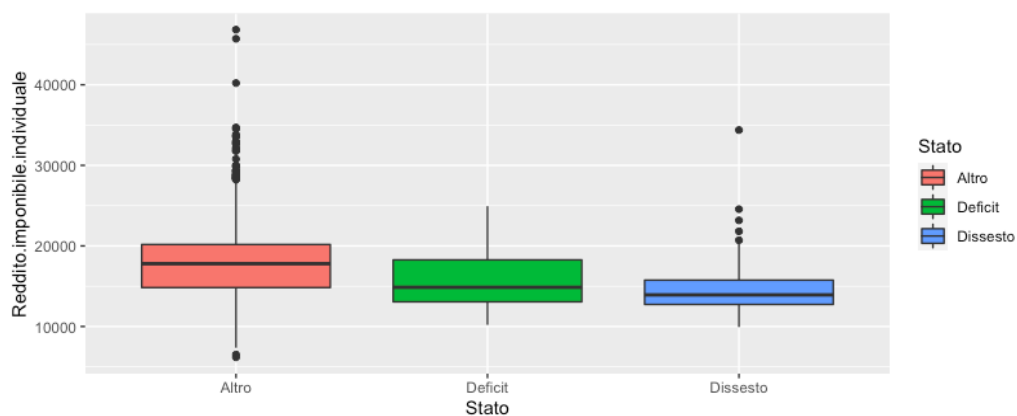
X-squared = 26.975, df = 10, p-value = 0.002629

2) Popolazione media dei comuni



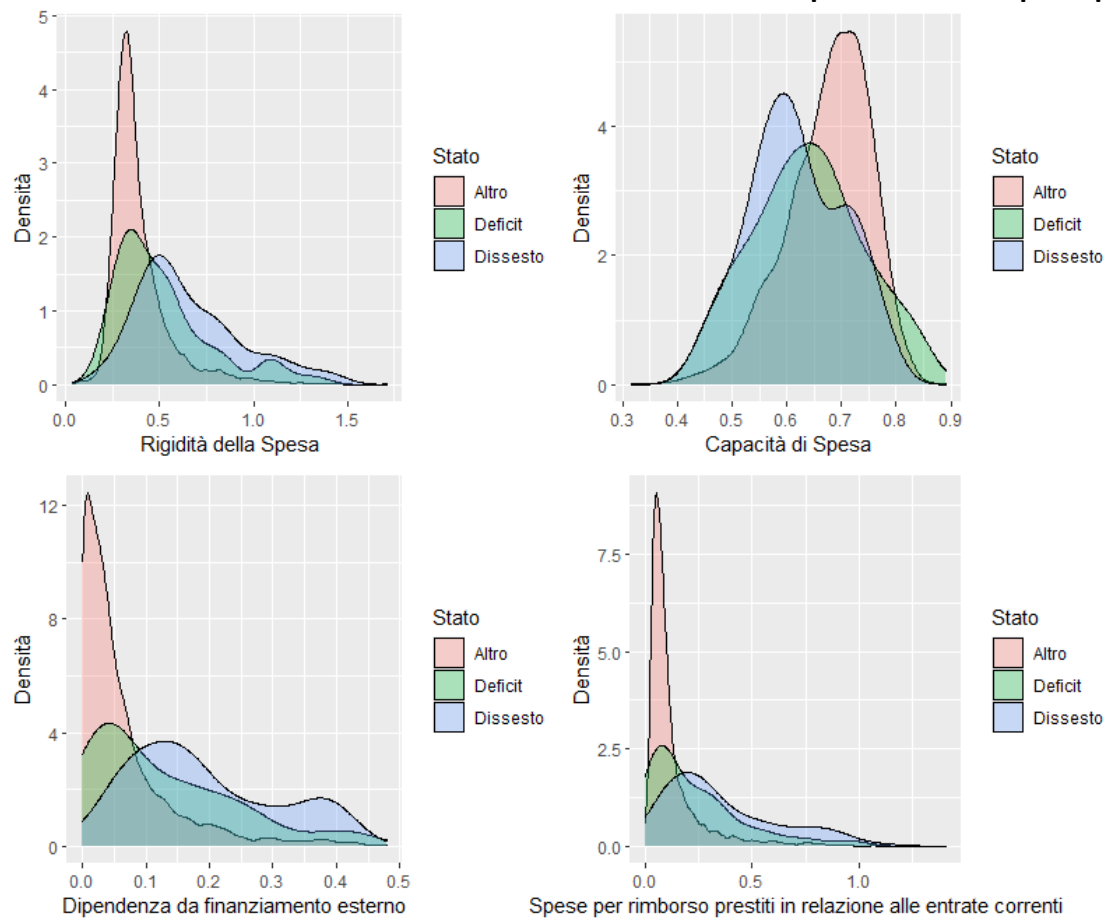
Il boxplot mostra che i comuni in deficit o in dissesto tra primo e terzo quartile sono tendenzialmente più popolati, con mediana e dispersione maggiore rispetto agli comuni con stato "Altro".

3) Reddito pro capite



Dal grafico emerge che i contribuenti dei comuni in deficit ed in particolare dei comuni in dissesto abbiano un reddito imponibile individuale minore rispetto agli altri comuni.

4) Distribuzione condizionata allo Stato finanziario del comune per 4 indicatori principali:



I comuni in “Altro” tendono ad avere una varianza ed asimmetria minore rispetto ai comuni deficitari o dissestati per questi 4 indicatori. Inoltre, i comuni in dissesto mostrano valori in media più alti per la Rigidità della Spesa, per il Grado di dipendenza da finanziamento esterno e per le Spese per rimborsare prestiti in relazione alle entrate correnti, un valore mediamente più basso in termini di Capacità di spesa.

Data Cleaning:

Sono state percorse diverse strade per l’eliminazione degli NA:

per prima cosa sono state rimosse colonne di alcuni indicatori finanziari ricche di NA (>60%);

in seguito, sono stati sostituiti con la media della provincia di appartenenza;

infine, qualora non fossero stati disponibili i dati provinciali, sono state utilizzate le media nazionali.

L’utilizzo di medie provinciali anziché nazionali ha ridotto l’errore medio sul dataset (dove erano presenti i dati) di un fattore 10, rivelandosi la strategia migliore.

Tutto ciò è stato necessario in virtù del fatto che le tecniche di Machine Learning che abbiamo implementato richiedono database privo di NA.

Preprocessing:

Il forte sbilanciamento del dataset impediva la corretta performance degli algoritmi di classificazione. L’algoritmo SMOTE si è rivelato superiore al semplice oversampling della classe minoritaria e ai metodi cost-based.

Modelli:

Nella selezione degli iperparametri si è data priorità nell'identificazione dei comuni in difficoltà finanziaria anche a costo di commettere errori, rispetto a parametri che avrebbero garantito un accuracy maggiore a scapito però dell'identificazione dei comuni problematici. Dunque, la metrica per valutare la bontà della previsione del modello sarà la Sensitività.

1. Multinomial Logit

Modello con tutte le variabili

Call:

```
multinom(formula = Stato ~ ., data =  
ds_smote)
```

Residual Deviance: 22184.03

AIC: **22268.03**

Modello zero

Call:

```
multinom(formula = Stato ~ 1, data =  
ds_smote)
```

Residual Deviance: 33821.88

AIC: **33825.88**

Confusion Matrix and Statistics

	Reference		
Prediction	0	1	2
0	1534	11	3
1	407	6	14
2	253	17	17

Accuracy : 0.6883

95% CI : (0.6688, 0.7074)

No Information Rate : 0.9699

P-Value [Acc > NIR] : 1

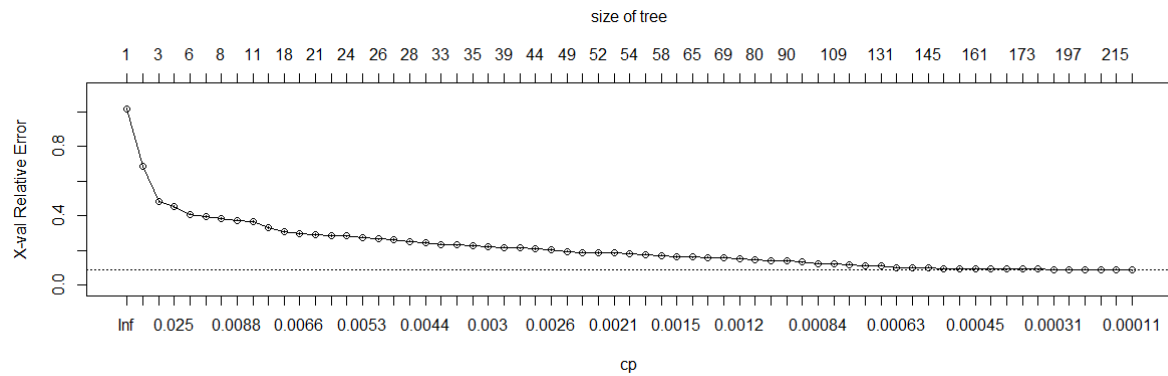
Kappa : 0.0598

Mcnemar's Test P-Value : <2e-16

Statistics by Class:

	Class: 0	Class: 1	Class: 2
Sensitivity	0.69918	0.176471	0.500000
Specificity	0.79412	0.811041	0.878815
Pos Pred Value	0.99096	0.014052	0.059233
Neg Pred Value	0.07563	0.984741	0.991392
Prevalence	0.96994	0.015031	0.015031
Detection Rate	0.67816	0.002653	0.007515
Detection Prevalence	0.68435	0.188771	0.126879
Balanced Accuracy	0.74665	0.493756	0.689408

2. Decision Tree



Bestcp = 1e-04

Confusion Matrix

```
predict_unseen
  0    1    2
0 2020  91   83
1   25    2    7
2   19    4   11
```

Accuracy = 0.8987622

Confusion Matrix and Statistics

```
Reference
Prediction    0     1     2
0      1991    19    19
1      111     4     9
2       97     6     6
```

Accuracy : 0.8846

95% CI : (0.8707, 0.8975)

No Information Rate : 0.9721

P-Value [Acc > NIR] : 1

Kappa : 0.0883

Mcnemar's Test P-Value : <2e-16

Statistics by Class:

	Class: 0	Class: 1	Class: 2
Sensitivity	0.9054	0.137931	0.176471
Specificity	0.3968	0.946261	0.953770
Pos Pred Value	0.9813	0.032258	0.055046
Neg Pred Value	0.1073	0.988307	0.986995
Prevalence	0.9721	0.012821	0.015031
Detection Rate	0.8802	0.001768	0.002653
Detection Prevalence	0.8970	0.054819	0.048187
Balanced Accuracy	0.6511	0.542096	0.565120

3. Support Vector Machine

Risultato

Confusion Matrix and Statistics

Reference

Prediction \ Reference	0	1	2
0	1464	6	0
1	520	9	8
2	222	12	21

Overall Statistics

Accuracy : 0.6605

95% CI : (0.6405, 0.68)

No Information Rate : 0.9752

P-Value [Acc > NIR] : 1

Kappa : 0.0619

McNemar's Test P-Value : <2e-16

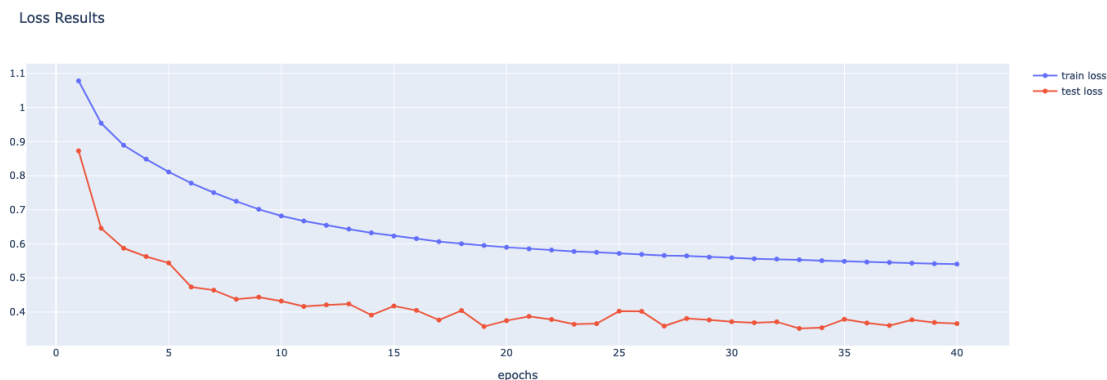
Statistics by Class:

	Class: 0	Class: 1	Class: 2
Sensitivity	0.66364	0.333333	0.724138
Specificity	0.89286	0.763758	0.895208
Pos Pred Value	0.99592	0.016760	0.082353
Neg Pred Value	0.06313	0.989565	0.996014
Prevalence	0.97524	0.011936	0.012821
Detection Rate	0.64721	0.003979	0.009284
Detection Prevalence	0.64987	0.237401	0.112732
Balanced Accuracy	0.77825	0.548546	0.809673

4. Rete Neurale:

Percettrone multistrato con due hidden layer di dimensioni 16 e 8 con funzione di attivazione relu, ed un output layer con 3 neuroni con funzione di costo Cross Entropy e adam come optimizer. Come learning rate si è optato per 0.001 come partenza con relativo scheduler con gamma uguale 0.99.

Come numero di epoche si è optato per 40 ma già dopo 19 il modello smette di avere aggiornamenti significativi nei parametri e smette quindi di aggiornarli.



	Pred: dissesto	Pred: deficit	Pred: sano
dissesto	9	7	9
deficit	13	18	6
sano	328	156	1716
Accuracy: 0.7705			
	Class: 0	Class: 1	Class: 2
Sensitivity:	0.78	0.486	0.36

5. Random Forest

Confusion Matrix and Statistics

Reference			
Prediction	0	1	2
0	2139	25	27
1	12	0	2
2	45	4	8

Overall Statistics

Accuracy : 0.9492

95% CI : (0.9393, 0.9578)

No Information Rate : 0.9708

P-Value [Acc > NIR] : 1.00000

Kappa : 0.1406

Mcnemar's Test P-Value : 0.02097

Statistics by Class:

	Class: 0	Class: 1	Class: 2
Sensitivity	0.9740	0.000000	0.216216
Specificity	0.2121	0.993730	0.977978
Pos Pred Value	0.9763	0.000000	0.140351
Neg Pred Value	0.1972	0.987100	0.986848
Prevalence	0.9708	0.012821	0.016357
Detection Rate	0.9456	0.000000	0.003537
Detection Prevalence	0.9686	0.006189	0.025199
Balanced Accuracy	0.5931	0.496865	0.597097

6. LDA

Actual			
Predicted	0	1	2
0	1489	10	2
1	469	11	9
2	247	10	15

Accuracy = 0.6698

	Class: 0	Class: 1	Class: 2
Sensitivity =	0,6752	0.3548	0.5769231

Analisi e conclusioni:

L'analisi si pone l'obiettivo di individuare i comuni a rischio deficit e dissesto.

Partendo da un dataset costruito ad hoc con indicatori finanziari, demografici e socio-politici.

Sono emerse difficoltà non trascurabili riguardo l'assenza di valori in maniera sparsa, risolte attraverso l'utilizzo di medie provinciali o rimuovendo variabili poco informative.

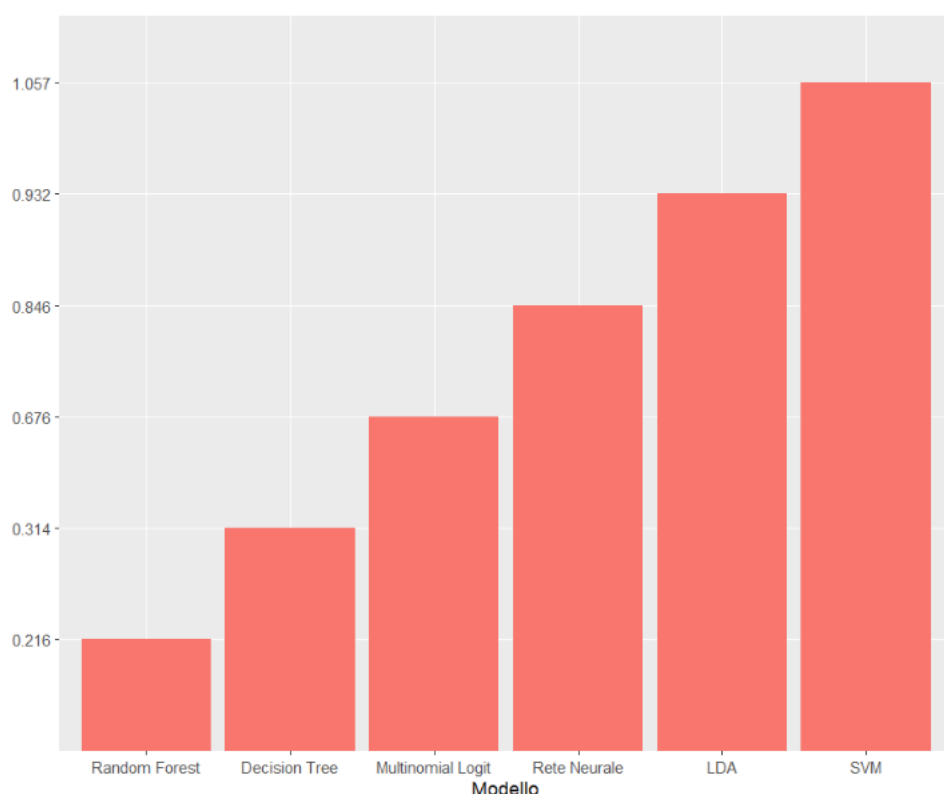
Abbiamo testato diversi modelli presenti in letteratura sul dataset una volta bilanciato con algoritmo SMOTE, mettendo enfasi sulla sensibilità al fine di rendere teoricamente utile il modello in produzione.

Gli alberi mantengono un accuracy elevata ma non sembrano adatti nella predizione dei "Dissesto", il loro limite è quello di predire tendenzialmente la classe "Sano".

La random forest accentua il problema aumentando l'accuracy ma perdendo in sensibilità.

La multinomial logit performa meglio in termini di sensibilità per le classi "Deficit" e "Dissesto", anche se il risultato migliore si ottiene utilizzando una rete neurale personalizzata coi parametri descritti sopra.

LDA e SVM sembrano essere i modelli migliori mostrando nel test set i valori migliori, la SVM in particolare risulta essere la più performante.



Sulle ordinate: Sensibilità della sofferenza finanziaria (deficit + dissesto)