

MALATTIE AL FEGATO: IN BASE A QUALI INDICATORI E A QUALI METODI CLASSIFICARE I PAZIENTI COME MALATI?

Alessandro Fossati, Giorgio Nardi, Luca Pretini, Eric Spinelli
Università degli studi di Milano Bicocca, cdLMDataScience
(Dated: 16 febbraio 2020)

Il fegato è un organo vitale che svolge diverse importanti funzioni nell'organismo. I medici stanno cercando sempre più assiduamente di scoprire le cause che portano un individuo a contrarre qualche forma di malattia al fegato ed i segnali che possano indurre a pensare che un individuo possa essere malato.

In questo articolo in particolare verrà analizzato come l'innalzamento di alcuni valori di determinate proteine o enzimi del nostro corpo possono portare a contrarre malattie al fegato. Questo scopo può essere raggiunto anche grazie all'applicazione di differenti modelli di classificazione, utilizzati nell'ambito del Machine Learning (ML).

Solamente in seguito ad un attento confronto tra codesti metodi, basato sulle loro performance e finalizzato ad estrapolarne i migliori, si potrà determinare quali individui saranno classificabili come malati e quali come sani, basandosi sui valori presentati dalle analisi del sangue dei pazienti esaminati.

1. INTRODUZIONE

Le malattie al fegato sono sempre più frequenti e risultano essere spesso fatali per molti individui. Solitamente questi malanni sono causati dall'eccessivo consumo di alcool o da un'errata alimentazione, perciò sapere quali siano i valori da tenere sotto controllo potrebbe aiutare a prevenire queste patologie. All'interno di questo articolo si arriverà a classificare un paziente, attraverso sofisticati algoritmi di classificazione, come malato o come sano a partire da differenti variabili relative agli esami del sangue del soggetto.

Nello specifico, l'articolo è così costruito: inizialmente verrà presentato il set di dati di partenza utilizzato per svolgere le analisi; successivamente, dopo una sezione dedicata all'analisi esplorativa dei dati, verranno presentati i differenti algoritmi e metodi di classificazione utilizzati e le misure di performance che essi presentano in relazione ai dati.

Infine, viene effettuato un confronto tra i diversi algoritmi utilizzati al fine di selezionare i più adatti per gli scopi di classificazione prefissati.

2. PRESENTAZIONE DATASET

Il Dataset proviene dalla piattaforma Kaggle¹ ed è costituito da osservazioni riguardanti 583 pazienti degli ospedali indiani di Andhra Pradesh. Sono presenti i seguenti 11 attributi:

- **Age:** età del paziente.
- **Gender:** sesso del paziente.
- **Total_Bilirubin:** la bilirubina è un prodotto della degradazione dell'emoglobina che consegue alla distruzione dei globuli rossi da parte della milza. È costituita dalla somma di bilirubina diretta e indiretta.

- **Direct_Bilirubin:** la bilirubina diretta è quella parte di bilirubina che viene trasformata dai batteri in bilinogeno che concorre alla colorazione delle feci.

- **Alkaline_Phosphotase:** la fosfatasi alcalina (ALP) è un enzima presente in diversi tessuti dell'organismo, in particolare nelle ossa, nei reni, nell'intestino e nel fegato.

Nel fegato, essa si trova nelle cellule che rivestono i dotti biliari (piccoli tubi che trasportano la bile del fegato all'intestino).

- **Alamine_Aminotransferase:** l'alanina aminotransferasi (ALT) è un enzima che si trova principalmente nelle cellule del fegato e dei reni. Quando il fegato è danneggiato l'ALT viene rilasciata nel sangue.

- **Aspartate_Aminotransferase:** l'aspartato aminotransferasi (AST) è un enzima presente soprattutto nel cuore, nei muscoli, nel cervello, nel fegato e nei reni.

Si registra un rilascio di AST quando si è in presenza di un danno cellulare subito da questi tessuti.

- **Total_Protiens:** proteine totali nel sangue.

- **Albumin:** l'Albumina è una proteina sintetizzata dal fegato ed è contenuta soprattutto nei liquidi interstiziali e nel plasma, dove rappresenta, da sola, circa la metà delle proteine circolanti.

- **Albumin_and_Globulin_ratio:** rapporto tra Albumine e globuline. Le globuline sono delle proteine, il cui livello nel sangue si riscontra dalla differenza tra proteine totali e Albumine.

- **Dataset:** variabile binaria che identifica se il paziente ha una malattia al fegato.

Nella fase di preprocessing, dopo aver rinominato la variabile Dataset come *Disease*, si è proceduto con la categorizzazione della variabile Age. In particolare, i valori sono stati suddivisi in 3 livelli in base alle proporzioni presenti nel dataset: “GIOVANE” (meno di 33 anni), “ADULTO” (tra 33 e 56 anni), “ANZIANO” (più di 56 anni).

Nei dati di partenza sono stati individuati 4 valori mancanti, tutti relativi alla variabile *Albumin_and_Globulin_ratio*. Per tale ragione, si è deciso di procedere al trattamento degli NA con l'imputazione della media condizionata alla variabile *Disease*, in modo da usare una strategia che permettesse di non perdere alcuna osservazione.

Prima di cominciare con la classificazione, si è proceduto con una fase di analisi descrittiva che consentisse di captare qualche informazione d'insieme sui dati.

Innanzitutto, la variabile target *Disease* si distribuisce nel campione con le seguenti proporzioni: 71% di malati e 29% di non malati. Secondo la teoria, la classe da considerare positiva nella fase di classificazione sarebbe quella più rara, ovvero i non malati. Tuttavia, in questo caso, osservando gli scatterplot univariati condizionati all'attributo *Disease* si può constatare che non vi sono chiari pattern che caratterizzano i pazienti sani.

Al contrario, sembra che spesso valori alti di bilirubina o qualsiasi enzima siano una caratteristica dei pazienti malati. Per questo i malati costituiranno la classe positiva nell'opera di classificazione. Per quanto riguarda, invece, gli attributi riferiti alle proteine essi non sembrano essere discriminanti nei confronti dell'attributo target.

Uno strumento utile per fornire importanti indicazioni sui pattern dei dati di partenza è la cluster analysis. In particolare, si è deciso di utilizzare l'algoritmo di tipo Prototype-Based Fuzzy C-Means, in grado di produrre un clustering che fornisce un'indicazione del grado di appartenenza di ciascuna osservazione ad ogni cluster.

È stato scelto questo tipo di Clustering perché è più efficace nel gestire pattern vicino al bordo di due o più clusters e outliers.

Prima di procedere, sono state escluse le variabili categoriali (Age e Gender) e quelle meno discriminanti (*Total_Protiens*, *Albumin*, *Albumin_and_Globulin_ratio*). È stata effettuata una normalizzazione min-max per limitare il range di ciascun attributo tra 0 e 1.

È stato fissato un numero di gruppi pari a 3, scelto in base al fatto che tale numerosità massimizza la silhouette. Quest'ultimo è indice di bontà della clusterizzazione e assume valori tra -1 e 1, considerando i parametri di coesione all'interno dei clusters e separazione tra cluster. Il risultato del clustering porta quindi ad una silhouette media di 0.635, che risulta soddisfacente ai fini interpretativi. Il Grafico 1 rappresenta la suddivisione in clusters dei pazienti malati, la maggior parte di essi fanno parte del gruppo rosso, il quale, considerando i valori sui tre assi, potrebbe comprendere un'ampia varietà di malattie.

È evidente che i clusters di colore verde e viola risultano

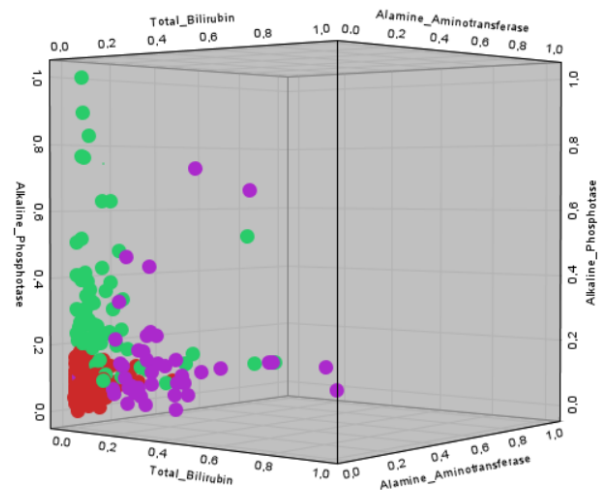


Figura 1: Risultato del clustering con algoritmo Fuzzy C-Means

molto più informativi e delineano la presenza di pazienti con diverse tipologie di malattie al fegato. In particolare, il gruppo verde include i pazienti con alto rilascio di ALP e bassa bilirubina. In aggiunta, alcuni di questi sono caratterizzati anche da alti valori di aspartato aminotransferasi. Consultando la letteratura medica, si può affermare che tali problematiche spesso caratterizzano i malati che presentano una malattia epatica che comporta l'ostruzione dei dotti biliari. Il gruppo di colore viola, invece, comprende in maggioranza pazienti con alta bilirubina totale e basso rilascio di enzimi. Tuttavia, in pochi record si notano alti valori di AST e ALP. Tale gruppo potrebbe quindi comprendere in grande maggioranza pazienti affetti da cirrosi, epatiti virali o danni epatici dovuti agli effetti collaterali di alcuni farmaci.

3. MODELLI E MISURE DI PERFORMANCE

3.1. MODELLI

Sono stati utilizzati all'interno di questa analisi modelli di differente tipologia, al fine di verificare quali tra questi si adattassero meglio ai dati presi in considerazione. Si è, pertanto, fatto uso di:

- **Modelli euristici:**

come modelli euristici sono stati utilizzati decision trees come il Weka J48, Random Forest e NBTree, utilizzabili per attributi categorici e numerici

- **Modelli di regressione:**

come modelli di regressione sono stati utilizzati il Logistic Regression Model ed il Simple Logistic Regression Model

- **Modelli di separazione:**

per la tipologia Support Vector Machine sono stati utilizzati i modelli SMO (Sequential Minimal Optimization), sia con kernel di tipo Puk che con kernel di tipo Polykernel ed il modello SPegasos con funzione di perdita di tipo ‘HingeLoss’.

Rispetto alla tipologia Artificial Neural Network invece, è stato utilizzato il Multilayer Perceptron

- **Modelli probabilistici:**

relativamente ai modelli di tipo probabilistico, sono stati selezionati, invece, quelli che sfruttano la regola di Bayes al fine di trovare una probabilità a posteriori che permette di classificare i record forniti come input.

In particolare sono stati utilizzati i modelli NaiveBayes e BayesNet, quest’ultimo con gli specifici algoritmi K2 e TAN

3.2. MISURE DI PERFORMANCE

All’interno dell’analisi vengono utilizzati differenti indici al fine di valutare la performance dei vari modelli di classificazione precedentemente presentati:

- Accuratezza:

$$Accuracy = \frac{TP + FN}{TP + TN + FP + FN}$$

Indice che permette di calcolare la percentuale di istanze correttamente classificate (True Positive TP e True Negative TN). Un valore vicino a 1 di accuracy indica che il classificatore utilizzato si adatta in maniera buona ai dati, tuttavia, una stima puntuale dell’accuracy potrebbe non essere sufficiente per decretare se un classificatore possa essere ritenuto performante o meno.

- Recall:

$$Recall = \frac{TP}{TP + FN}$$

Indice che permette di ottenere la percentuale di record positivi correttamente classificati. Come evidenziato in alcuni esempi successivi, spesso la Recall da sola non risulta essere un buon indicatore del fatto che un modello sia performante.

- Precision:

$$Precision = \frac{TP}{TP + FP}$$

Indice che permette di trovare la frazione di record effettivamente positivi tra tutti quelli che vengono classificati come positivi. Un valore vicino a 1 di

questo indice comporta che ci sia un numero basso di falsi positivi (FP). Anche questa misura, se considerata da sola, potrebbe risultare fuorviante sulla valutazione della performance di un modello di classificazione.

- F1-measure:

$$F1measure = \frac{2 * Precision * Recall}{Precision + Recall}$$

È la media armonica tra Recall e Precision, è un indice il cui eventuale valore elevato garantirebbe valori elevati sia di Recall che di Precision. È dunque una misura di sintesi dei due indici precedentemente presentati.

- ROC Curve:

Curva che presenta la percentuale di veri positivi sull’asse delle ordinate (%TP) e la percentuale dei falsi positivi sull’asse delle ascisse (%FP). Questo grafico permette di calcolare un indice per valutare la performance di un modello, denominato AUC (*area under curve*), che fornisce indicazioni riguardo all’area sottostante alla curva di ROC. Maggiore è l’AUC, maggiore è la probabilità di avere un modello performante.

4. CLASSIFICAZIONE CON DIFFERENTI APPROCCI

4.1. METODO HOLDOUT

Nel primo approccio di classificazione si è scelto di utilizzare il metodo dell’Holdout. Si è dunque eseguito un partizionamento con la tecnica dello stratified sampling. Si è scelto di partizionare il dataset iniziale in 67% di record per il training set e 33% per il test set.

In questo primo approccio di classificazione, si è inoltre deciso di non effettuare feature selection, come suggerito dal nodo Knime *Parameter Mutual Information*, il quale non individua attributi ridondanti. In aggiunta, la regressione logistica non consente in linea teorica l’utilizzo di variabili categoriali come Age e Gender che, pertanto, non sono state incluse nel modello.

In Tabella1 sono riportati i risultati delle performance degli algoritmi utilizzati:

Si evince innanzitutto che la percentuale di osservazioni correttamente classificate negli algoritmi più performanti si aggira intorno al 70%. Inoltre, si nota che nei modelli SVM il valore dell’indice di Recall risulta pari a 1, dunque massimo. Questo risultato è riconducibile al fatto che il numero di records classificati come FN (falsi negativi) è uguale a zero.

Tuttavia, tutte le osservazioni del test set vengono attribuite alla classe positiva dei malati, dunque, in questo

	Accuracy	Recall	Precision	F-measure
NBTree	0.663	0.768	0.762	0.765
RForest	0.642	0.775	0.737	0.756
J48	0.564	0.615	0.732	0.669
SPega	0.715	1.0	0.715	0.833
SMOpoly	0.715	1.0	0.715	0.833
SMOpuk	0.715	1.0	0.715	0.833
MLP	0.626	0.768	0.726	0.746
BN-K2	0.616	0.623	0.796	0.699
BN-TAN	0.663	0.804	0.744	0.773
NB	0.554	0.420	0.906	0.574
LOG	0.715	0.905	0.748	0.819
SLOG	0.699	0.891	0.740	0.809

Tabella I: Risultati del primo approccio

caso, la Recall non risulta essere un buon indicatore della performance dell'algoritmo SVM.

4.2. FEATURE SELECTION CON K-FOLDS CROSS VALIDATION

Al fine di selezionare le variabili più performanti e migliorare l'interpretabilità dei modelli è stata effettuata una selezione degli attributi con metodo CfsSubsetEval, tramite un approccio di tipo Filter Multivariate.

Quest'ultimo permette di selezionare gli attributi prima che il classificatore entri in gioco, escludendo nel contempo quelli ridondanti ed irrilevanti.

In questo secondo approccio di classificazione, il dataset iniziale viene partizionato mediante la tecnica del 7-Folds Cross Validation che, dividendo il dataset in 7 parti con metodo di campionamento stratificato, utilizza ogni parte come test set una sola volta, fornendo alla fine una stima dell'accuracy più efficace rispetto a quella fornita dal metodo holdout.

In Tabella2 sono riportati i risultati delle performance degli algoritmi utilizzati:

	Accuracy	Recall	Precision	F-measure
NBTree	0.688	0.913	0.722	0.807
RForest	0.71	0.947	0.728	0.823
J48	0.695	0.8	0.778	0.789
SPega	0.714	1.0	0.714	0.833
SMOpoly	0.714	1.0	0.714	0.833
SMOpuk	0.714	1.0	0.714	0.833
MLP	0.683	0.793	0.769	0.781
BN-K2	0.667	0.661	0.838	0.739
BN-TAN	0.686	0.76	0.792	0.775
NB	0.547	0.385	0.952	0.548
LOG	0.705	0.957	0.721	0.822
SLOG	0.707	0.99	0.712	0.82

Tabella II: Risultati del secondo approccio

Come anticipato nella sezione dedicata all'analisi esplorativa, le variabili *Total.Proteins*, *Albumin*, *Albu-*

min.and.Globulin.Ratio vengono valutate come irrilevanti dal Filter Multivariate e per questo vengono escluse dai modelli di classificazione.

Analizzando i valori della Tabella2, si evince che in tutti i classificatori migliora il valore di accuracy eccetto che per SMOpuk, SMOpoly e NaiveBayes (NB). Si può inoltre constatare che i classificatori RandomForest, MLP e BayesNet.K2 migliorano in tutti gli indici di performance. È evidente, inoltre, un miglioramento per il classificatore SimpleLogistic negli indici di performance eccetto che per la Precision. In particolare, è chiaro un miglioramento dell'indice Recall per questo algoritmo. Tuttavia, anche in questo caso ci si trova nella situazione in cui il modello ha difficoltà nel classificare correttamente i pazienti non malati (TN), classificandoli di fatto quasi tutti i pazienti come malati.

Basandosi su una visione di insieme dei quattro indici di performance forniti, si può concludere che il classificatore J48 risulta essere il migliore.

4.3. MATRICE DI COSTO

Si introduce una matrice di costo tale da cercare di penalizzare il più possibile la classificazione di pazienti malati come sani (falsi negativi), dal momento che, in medicina, classificare un paziente malato come sano risulterebbe molto più grave rispetto al classificarne uno sano come malato.

Vengono penalizzati comunque anche i falsi positivi, seppur in maniera minore. Consultando la letteratura, si è deciso di attribuire ai falsi negativi un costo pari a 1, mentre ai falsi positivi un costo pari a 0.4, ottenuto rapportando il numero di pazienti sani ed il numero di pazienti malati in riferimento al dataset iniziale.

Nel terzo ed ultimo approccio di classificazione, si è posto come obiettivo quello di istruire i modelli di classificazione tramite la matrice di costo mantenendo la feature selection e l'approccio 7-folds cross validation. I risultati ottenuti vengono riportati nella Tabella3:

	Accuracy	Recall	Precision	F-measure
NBTree	0.714	1	0.714	0.833
RForest	0.714	1	0.714	0.833
J48	0.71	0.901	0.746	0.816
SMOpoly	0.714	1	0.714	0.833
SMOpuk	0.714	1	0.714	0.833
MLP	0.714	1	0.714	0.833
BN-K2	0.696	0.779	0.714	0.785
BN-TAN	0.714	1	0.714	0.833
NB	0.554	0.397	0.948	0.559
LOG	0.712	0.998	0.713	0.832
SLOG	0.714	1	0.714	0.833

Tabella III: Risultati del terzo approccio

Si nota che la maggior parte dei modelli risultano essere molto sensibili alla matrice di costo, classificando tutti i pazienti come malati. I valori di Recall dei vari

modelli risultano essere molto alti, quasi tutti pari 1. Si può, dunque, affermare che gli algoritmi risultano risentire particolarmente della penalizzazione applicata ai falsi negativi.

È evidente, inoltre, come RForest e BayesNet_K2 risultino performanti anche dopo essere stati istruiti sulla cost matrix.

4.4. CONFRONTI FINALI

Una volta che gli algoritmi ed i relativi approcci, è bene decretare quali siano in grado di classificare meglio. Per selezionare i migliori modelli si è deciso di valutare i valori riscontrati per l'accuracy e per la F1-measure, selezionando gli algoritmi che presentano i valori più alti per entrambi gli indici.

Nella fase di scelta sono stati ignorati gli algoritmi che performano nello stesso modo del modello ZeroRule, poiché ritenuti poco informativi.

Considerando i suddetti indici, sono stati selezionati inizialmente i 6 modelli di classificazione più performanti, i cui valori di accuracy sono riportati nella Figura 2:

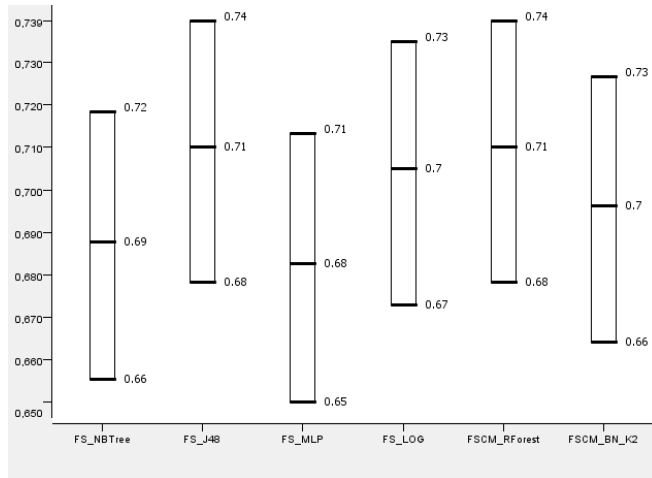


Figura 2: Valori di accuracy dei migliori modelli

Oltre a fornire i nomi degli algoritmi più performanti (asse delle ascisse), il Grafico2 risulta utile per comprendere che il valore di accuracy di ognuno dei sei algoritmi selezionati (asse delle ordinate) risulta essere non significativamente diverso dai valori di accuracy degli altri 5 presi in considerazione, in quanto ognuno dei valori ricade in tutti gli altri intervalli di confidenza per l'accuracy costruiti seguendo la formula di Wilson sotto riportata, considerando un livello di confidenza del 90%:

Dove con N si indica il numero di osservazioni totale, mentre con Z si indica il quantile della normale standard con $\alpha = 0.1$.

Preso in considerazione l'indice di accuracy, il line plot qui presentato riassume il miglioramento di essa conseguito ad ogni passaggio rispetto agli algoritmi

$$\left(\frac{\frac{acc + \frac{1}{2N} \pm \frac{1}{2} \sqrt{\frac{acc^2}{N} + \frac{1}{4N}}}{1 + \frac{1}{N}}, \frac{\frac{acc + \frac{1}{2N} \pm \frac{1}{2} \sqrt{\frac{acc^2}{N} + \frac{1}{4N}}}{1 + \frac{1}{N}} \right)$$

Figura 3: Intervallo di Wilson

prescelti.

Sull'asse delle ascisse si individuano gli step, in parti-

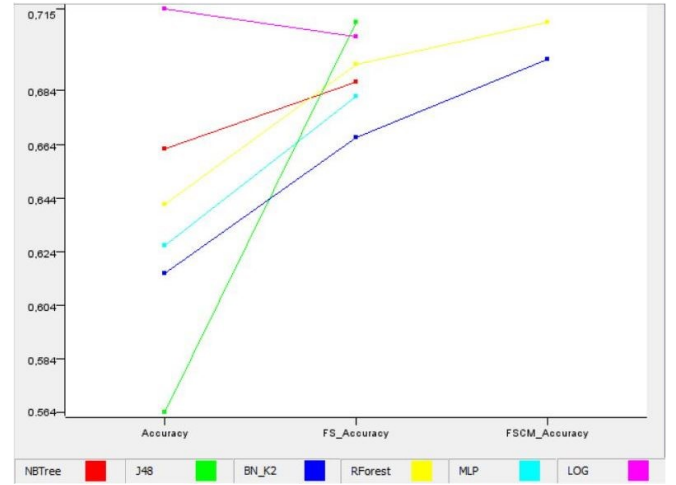


Figura 4: Line plot

colare con FS si specifica l'utilizzo di Feature Selection, mentre FSCM considera l'applicazione della Cost Matrix oltre alla Feature Selection; sull'asse delle ordinate è riportato il valore dell'accuracy.

Si può notare che al secondo passaggio tutti gli algoritmi si fermano eccetto Random Forest e Bayes Net-K2. Questo avviene dal momento in cui si è preferito escludere i casi Zero Rule, ovvero quelli in cui l'algoritmo tende a classificare tutte le osservazioni secondo la classe più frequente, risultando inutile ai fini della classificazione pur ottenendo l'accuracy più alta. In ultimo si osserva che il miglioramento più estremo lo ha ottenuto l'algoritmo J48 con Attribute Selection, mentre l'algoritmo Logistic ha subito un lieve peggioramento dell'indice al medesimo step.

Per un'ulteriore valutazione della performance dei classificatori, sono state confrontate le rispettive curve di ROC (Figura 5):

Innanzitutto, si nota che le AUC di tutti i modelli prescelti risultano essere maggiori rispetto a quella relativa al modello ZeroRule (AUC=0.5), ossia il modello che non aggiunge alcuna informazione.

A parità di accuracy, se si volesse selezionare il modello

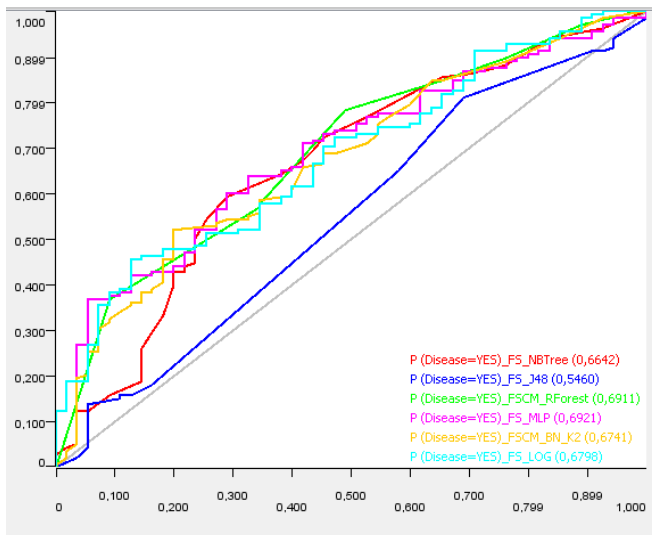


Figura 5: Confronto tramite curve ROC

migliore, ci si troverebbe in difficoltà, in quanto nessuna delle curve risulta essere sempre al di sopra delle altre, dunque tutti valori di AUC risultano essere molto simili, eccezion fatta per J48, il quale, a parità di percentuale di true positive (%TP) rispetto agli altri modelli considerati, restituisce più alte percentuali di false positive, risultando dunque meno performante. Infine potrebbe risultare interessante confrontare questi 6 modelli in base al costo che essi comportano (Figura 6):

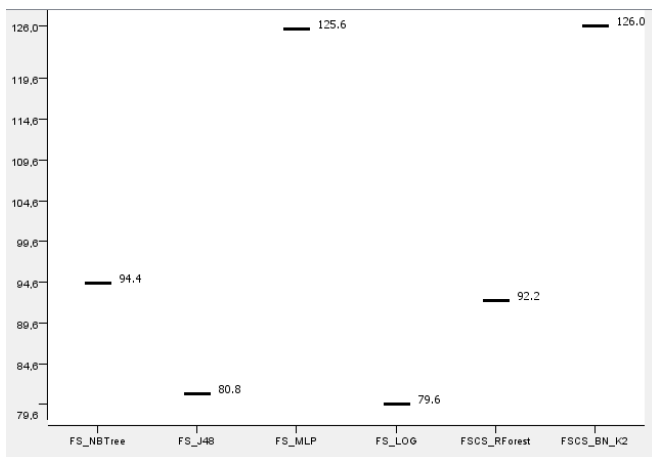


Figura 6: Confronto tra algoritmi basato sul costo

Dove il costo relativo ad ogni algoritmo viene calcolato a partire dalla rispettiva confusion matrix e dalla cost matrix precedentemente stabilita (Figura 7):

$$\text{COSTO} = 0 * \text{TN} + 0.4 * \text{FP} + 1 * \text{FN} + 0 * \text{TP}$$

ACTUAL/PREDICTED CLASS	-1	+1
-1	0	0.4
+1	1	0

Figura 7: Matrice di costo.

Si evince dal grafico che ci sono due algoritmi che comportano costi troppo elevati, MLP dopo feature selection e BayesNet_K2 istruito sulla matrice di costo. Inoltre, i modelli J48 dopo feature selection e LOG dopo feature selection risultano comportare costi più bassi rispetto a tutti gli altri classificatori, risultando dunque essere i migliori nella gestione del problema dei falsi negativi.

5. CONCLUSIONI

A parità di valori di accuracy, osservando i valori di AUC dei migliori modelli di classificazione selezionati, si nota come J48 non possa essere scelto come modello più performante, mentre la decisione potrebbe ricadere su MLP dopo feature selection.

Tuttavia, esso sembra non riuscire a gestire nella maniera adeguata il problema dei falsi negativi, comportando di fatto un costo troppo elevato da sostenere. La scelta del modello vincitore ricade dunque sulla regressione logistica dopo feature selection (FS.Log), in quanto alla luce di buoni valori di accuracy, F1-measure ed AUC, esso comporta il costo minore tra tutti gli algoritmi.

Nella Figura 8 vengono riportati i valori relativi agli odds ratio del modello selezionato come migliore, a fini interpretativi:

Odds Ratios	
Variable	Class
	NO
=====	
Total_Bilirubin	5.3527
Direct_Bilirubin	0.0409
Alkaline_Phosphotase	0.5443
Alamine_Aminotransferase	0.1969
Aspartate_Aminotransferase	0.204

Figura 8: Odds ratios.

Dapprima, si nota che nessuna variabile riguardante le proteine risulta essere inclusa nel modello, in quanto tali attributi sono stati ritenuti irrilevanti dal Filter Multivariate, confermando quanto affermato nella fase dedicata all'analisi esplorativa.

Inoltre, volendo fornire una prima interpretazione degli odds ratio (Figura 8), si nota come tutte le variabili eccetto la Bilirubina Totale risultano avere una associazione negativa con l'assenza della malattia.

Ad esempio, un incremento unitario del valore standardizzato dell'enzima Aspartato Aminotransferasi diminuisce l'attitudine ad essere sano di circa l'80%.

¹ "Dataset (kaggle.com)," <https://www.kaggle.com/uciml/indian-liver-patient-records>, accessed: 2020-02-14.

² "Matrice di costo," <http://frasca.di.unimi.it/MDSBF18/Lez5.pdf>, accessed: 2020-02-14.

³ "Valori del sangue," <https://www.my-personaltrainer.it/fegato/valori-fegato.html>, accessed: 2020-02-14.

⁴ "Malattie al fegato," <https://avisemiliaromagna.it/2016/03/16/5840/>, accessed: 2020-02-14.