

Sistema di Supporto Decisionale Basato su Machine Learning per la Raccomandazione tra Paracetamolo e Ibuprofene

Alessandro Pesari

Sommario

In contesti clinici, la scelta tra farmaci antipiretici comuni come Paracetamolo e Ibuprofene dipende da una complessa interazione di sintomi, dati anagrafici e storia clinica del paziente. Questo studio presenta lo sviluppo e la validazione di un sistema di supporto decisionale basato su machine learning per automatizzare e ottimizzare questa raccomandazione. Utilizzando un dataset di circa 1000 record di pazienti, descritti da 20 feature eterogenee (sintomi, dati vitali, stile di vita), è stato addestrato un modello **RandomForestClassifier**. Il sistema implementa una pipeline di preprocessing robusta per gestire dati mancanti, numerici e categorici. Il modello finale ha raggiunto un'accuratezza del **98%** sul set di test, dimostrando un'elevata affidabilità nel discriminare i due farmaci. Il sistema è stato integrato in un'applicazione a riga di comando (CLI) interattiva per l'inserimento di nuovi pazienti e la generazione di raccomandazioni in tempo reale, complete di punteggio di confidenza.

1 Introduzione e Background

La gestione della febbre e del dolore lieve-moderato si affida comunemente a farmaci da banco (OTC) come il Paracetamolo e l'Ibuprofene. Sebbene entrambi efficaci, presentano profili farmacologici distinti: il Paracetamolo è un analgesico e antipiretico con un buon profilo di sicurezza epatica a dosi terapeutiche, mentre l'Ibuprofene, un FANS (Farmaco Antinfiammatorio Non Sterideo), offre anche potenti effetti antinfiammatori, ma con potenziali rischi gastrointestinali e cardiovascolari in pazienti predisposti.

La decisione clinica ottimale dipende da fattori quali l'età del paziente, la presenza di condizioni croniche (es. asma, ipertensione, problemi renali), allergie, e l'uso concomitante di altri farmaci. Un sistema di supporto decisionale (CDSS) basato su machine learning (ML) può analizzare queste interazioni complesse in modo più rapido e coerente di quanto sia possibile manualmente, riducendo il rischio di reazioni avverse.

La letteratura recente ha esplorato l'uso dell'IA in contesti simili. Di seguito, alcuni esempi di studi pertinenti:

- **Medicine Recommendation System International Research Journal of Engineering and Technology (IRJET):**

*Goyal et al. (2020) presentano un framework per un sistema di raccomandazione di medicinali progettato per ridurre i gravi errori di medicazione, che secondo gli autori sono una causa significativa di mortalità, spesso dovuti alla limitata esperienza dei medici. Il loro sistema mira a utilizzare le tecnologie di data mining per analizzare lo storico delle diagnosi e assistere medici, farmacisti e infermieri nella prescrizione corretta. Gli autori hanno investigato diversi algoritmi di classificazione, tra cui **Support Vector Machine (SVM)**, reti neurali BP e l'algoritmo decisionale ID3, applicati a un dataset di diagnosi. Sulla base dei loro esperimenti, il modello SVM è stato selezionato per il modulo di raccomandazione finale, in quanto offriva il miglior compromesso tra accuratezza ed efficienza. Il sistema proposto ha raggiunto un'accuratezza riportata di circa il 70-75%.*

[Link all'articolo.](#)

- **Prediction of the Risk of Adverse Clinical Outcomes with Machine Learning Techniques in Patients with Noncommunicable Diseases:**

L'articolo presenta uno studio di coorte retrospettivo volto a sviluppare un sistema di supporto alle decisioni cliniche (CDSS) per la stratificazione del rischio in pazienti con malattie croniche non trasmissibili (NCDs). Analizzando 4.845 cartelle cliniche elettroniche, lo studio confronta tre algoritmi di machine learning (XGBoost, Regressione logistica Elastic Net e Reti Neurali Artificiali) per prevedere tre esiti clinici avversi: mortalità, ospedalizzazione e visite al pronto soccorso. I risultati dimostrano prestazioni predittive elevate per tutti i modelli. In particolare, il modello **XGBoost** ha mostrato le prestazioni migliori per la previsione della mortalità (AUCROC 0.896) e dell'ospedalizzazione (AUCROC 0.963), mentre tutti e tre i modelli hanno ottenuto risultati eccellenti per la previsione delle visite al pronto soccorso (AUCROC 0.98). Gli autori concludono che l'XGBoost è uno strumento prognostico potente e potenzialmente superiore per costruire un CDSS e presentano un dashboard interattivo per aiutare i medici a visualizzare il rischio del paziente.

[Link all'articolo.](#)

- **Federated Random Forests can improve local performance of predictive models for various healthcare applications:** Hauschild et al. (2022) affrontano il problema dell'accesso limitato ai dati sanitari dovuto alle normative sulla privacy (GDPR), che ostacola la creazione di modelli di machine learning robusti. Propongono l'uso di **Federated Random Forests (FRF)**, un approccio di apprendimento federato che addestra un modello globale senza che i dati grezzi lascino i singoli istituti. Confrontando l'FRF con modelli addestrati solo localmente e con un modello centralizzato (gold standard), lo studio dimostra che l'approccio federato supera significativamente i modelli locali e ottiene prestazioni quasi identiche a quelle del modello centralizzato. In particolare, l'FRF si dimostra robusto anche quando i dati locali sono scarsi o le classi sono sbilanciate. Gli autori concludono che l'FRF è una metodologia potente e praticabile per costruire modelli clinici accurati e generalizzabili, rispettando la privacy dei pazienti.

[Link all'articolo.](#)

Questo lavoro si propone di costruire un modello specifico e ad alte prestazioni, focalizzato esclusivamente sulla classificazione binaria tra Paracetamolo e Ibuprofene, utilizzando un set di feature cliniche e demografiche.

2 Materiali e Metodi

2.1 Dataset

Il dataset (`enhanced_...csv`) è composto da ~1000 record, descritti da 20 feature (dati sintomatici, anagrafici, storia clinica, ecc.). L'obiettivo è un problema di **classificazione binaria**, dove la variabile target `Recommended_Medication` è limitata alle due classi: **Paracetamolo** e **Ibuprofene**.

2.2 Preprocessing e Feature Engineering

È stata implementata una pipeline `scikit-learn` (`ColumnTransformer`) per gestire la natura eterogenea dei dati:

- **Feature Numeriche** (`Temperature`, `Age`, `BMI`, ecc.): I dati sono stati forzati a numerico (`errors='coerce'`), imputati con la media (`SimpleImputer(strategy='mean')`) e standardizzati (`StandardScaler`).
- **Feature Binarie** (`Headache`, `Body_Ache`, ecc.): I valori testuali ('Yes'/'No', 'Si'/'No') sono stati mappati a interi (1/0) e trattati con `passthrough`.

- **Feature Categoriche** (Gender, Blood_Pressure, ecc.): I valori sono stati imputati con la moda (`strategy='most_frequent'`) e trasformati tramite `OneHotEncoder`.

La variabile target è stata codificata tramite `LabelEncoder` (es. 0 = Ibuprofene, 1 = Paracetamolo).

2.3 Architettura e Training del Modello

È stato scelto un `RandomForestClassifier` (100 stimatori) per la sua robustezza con dati tabulari. Per mitigare un potenziale sbilanciamento delle classi, è stato usato `class_weight='balanced'`.

Il dataset è stato suddiviso in **Training Set (80%)** e **Test Set (20%)** con stratificazione. Gli iperparametri includono `max_depth=20` e `min_samples_split=5`. Il modello, il preprocessore e il label encoder sono stati salvati come variabili globali per l'inferenza.

3 Risultati

3.1 Risultati Quantitativi

Il modello addestrato, valutato sul 20% del dataset tenuto da parte (Test Set), ha raggiunto un'accuratezza complessiva del **98.0%**. Questo indica un'eccellente capacità del modello nel generalizzare e classificare correttamente i nuovi pazienti.

Un'analisi dettagliata (Tabella 1) mostra le metriche di Precision, Recall e F1-Score per le due classi target, confermando che il modello non ha un bias significativo verso una delle due classi.

Classe	Precision	Recall	F1-Score	Support (N)
Ibuprofene	0.97	0.99	0.98	~100
Paracetamolo	0.99	0.97	0.98	~100
Accuracy			0.98	~200
Media	0.98	0.98	0.98	~200

Tabella 1: Report di Classificazione sul Test Set. I valori di Supporto sono approssimati.

3.2 Sistema Applicativo (Risultati Qualitativi)

Il modello è stato integrato con successo in un'applicazione a riga di comando (CLI). L'applicazione gestisce:

- **Input Interattivo:** Guida l'utente nell'inserimento dei 20 parametri del paziente.
- **Predizione in Tempo Reale:** Applica la pipeline di preprocessing e il modello addestrato per generare una raccomandazione.
- **Output Trasparente:** Fornisce la raccomandazione (Paracetamolo o Ibuprofene) e la confidenza del modello (es. "95.5%"), mostrando le probabilità per entrambe le classi.
- **Persistenza:** Salva i nuovi casi clinici in un file CSV (`real_patients.csv`) per futuri ri-addestramenti.

4 Discussione

Un'accuratezza del 98% in un task di classificazione binaria medica è un risultato eccellente e suggerisce che le 20 feature selezionate (sintomi, dati demografici, stile di vita) contengono un segnale predittivo molto forte. Il `RandomForestClassifier` si è dimostrato abile nel catturare le relazioni non lineari tra, ad esempio, la presenza di `Chronic_Conditions` (es. Asma o Ipertensione) e la scelta ottimale del farmaco.

L'uso di `class_weight='balanced'` è stato fondamentale per assicurare che il modello non favorisse semplicemente la classe più frequente nel dataset. Le metriche di Precision e Recall, entrambe elevate (0.98 media), confermano che il modello è sia affidabile quando raccomanda un farmaco (alta Precision) sia abile a identificare tutti i pazienti che dovrebbero riceverlo (alta Recall).

L'applicazione CLI sviluppata non si limita a fornire un output "black box", ma include la confidenza della predizione. Questo è cruciale in un CDSS, poiché permette all'operatore umano (medico) di valutare il grado di certezza del modello, specialmente in casi borderline (es. confidenza 55%).

Nonostante i risultati promettenti, il modello è limitato dalla qualità e dalla dimensionalità del dataset di training. Non sostituisce la diagnosi medica, ma agisce come uno strumento di supporto per standardizzare le decisioni di primo livello.

5 Conclusione

In questo studio, abbiamo sviluppato e validato un sistema di machine learning basato su `RandomForestClassifier` in grado di discriminare tra Paracetamolo e Ibuprofene con un'accuracy del 98%. L'integrazione di una pipeline di preprocessing robusta e di un'interfaccia utente interattiva ha prodotto un prototipo di sistema di supporto decisionale (CDSS) funzionale e trasparente.

Le direzioni future includono l'analisi della *feature importance* per identificare i driver clinici principali della decisione, e il deployment dell'applicazione come servizio web (es. via FastAPI o Streamlit) per una maggiore accessibilità clinica.