**GUIDA ALL’UTENTE**

**CAPITOLO I – INTRODUZIONE DEL SOFTWARE**

Con la presente guida si vuole illustrare il funzionamento del software realizzato, il suo scopo e mostrare come utilizzare il software mediante interfaccia grafica o a riga di comando.

Alla base del progetto vi è l’algoritmo di regressione **k-nearest neighbors**, più comunemente noto come ***KNN;*** esso è un algoritmo che utilizza la prossimità per effettuare previsioni sul raggruppamento di un singolo punto dati.

L’algoritmo, per funzionare, si basa su un training set contenente una serie di dati che verranno utilizzati per prevenire il dato numerico mancante. Più nello specifico, si rapportano i dati forniti ad una distanza k, la quale indica il numero di attributi che possono differire dalla query dell’utente, e si effettua una media dei valori nei vicini selezionati, in base al valore ignoto richiesto dall’utente nella query.

Il calcolo della distanza *k* si differenzia in base alla natura dell’attributo:

1. Per attributi **discreti**, si calcola la distanza di Hamming tra i due attributi
2. Per attributi **continui,** si effettua un processo di scalatura sfruttando l’algoritmo di *min-max scaler* (utilizzando la seguente formula: ) per poi effettuare la seguente differenza:

(Per min e max si intende il minimo ed il massimo della variabile da scalare)

È necessario applicare un processo di scalatura per non avere risultati falsati in caso due valori siano troppo distanti numericamente tra loro.

**CAPITOLO II – FUNZIONAMENTO DEL SOFTWARE**

Il software può essere utilizzato usando un’interfaccia grafica oppure mediante linea di comando. Nella seguente guida verranno mostrati entrambi i funzionamento inserendo delle opportune immagini.

Si procede indicando gli step dell’esecuzione del programma mediante linea di comando:

1. Scelta della sorgente del Training Set

Una volta avviato il software, all’utente verrà richiesto la sorgente da cui acquisire il Training Set e il nome del file (in caso di file/file binario) o tabella (in caso di database).

L’utente dovrà inserire la propria scelta esprimendo il numero desiderato e premendo il tasto “Invio”

*Immagine che contiene testo

Descrizione generata automaticamente*

Dopo aver inserito la scelta, l’utente deve procedere digitando il nome del file/tabella seguita dal tasto “*Invio”:*

*Immagine che contiene testo

Descrizione generata automaticamente*

1. Acquisizione degli attributi

Dopo che il KNN è stato caricato correttamente nel Server, l’utente deve inserire gli attributi digitando, al termine dell’inserimento del valore, il tasto invio:

Immagine che contiene testo

Descrizione generata automaticamente

1. Acquisizione del valore di K

Terminata la fase di acquisizione degli attributi, bisogna digitare la distanza *k* che l’algoritmo dovrà considerare per il calcolo della predizione:

Immagine che contiene testo

Descrizione generata automaticamente

1. Ripetizione della predizione

Dopo aver ottenuto la predizione l’utente può decidere se:

* Ripetere la predizione con lo stesso Training Set
* Ripetere la predizione con un altro Training Set
* Terminare l’esecuzione

Per ripetere la predizione con lo stesso Training Set è sufficiente digitare “*Y”:*

Immagine che contiene testo

Descrizione generata automaticamente

All’utente verrà richiesto di inserire gli attributi e il valore di K come descritto precedentemente.

Per ripetere la predizione con un altro Training Set è sufficiente digitare qualsiasi tasto che non sia “Y” e dopo alla richiesta di esecuzione con nuovo oggetto KNN digitare “Y”:

Immagine che contiene testo

Descrizione generata automaticamente

Per terminare l’esecuzione del programma è sufficiente digitare due volte qualsiasi stringa che non sia Y:

Immagine che contiene testo

Descrizione generata automaticamente

Interfaccia grafica: