



Caso di Studio : Data Mining

estrazione di conoscenza nascosta in basi di dati voluminose

Autori:

Dibisceglia Michele

Finiguerra Alessia

Sommario

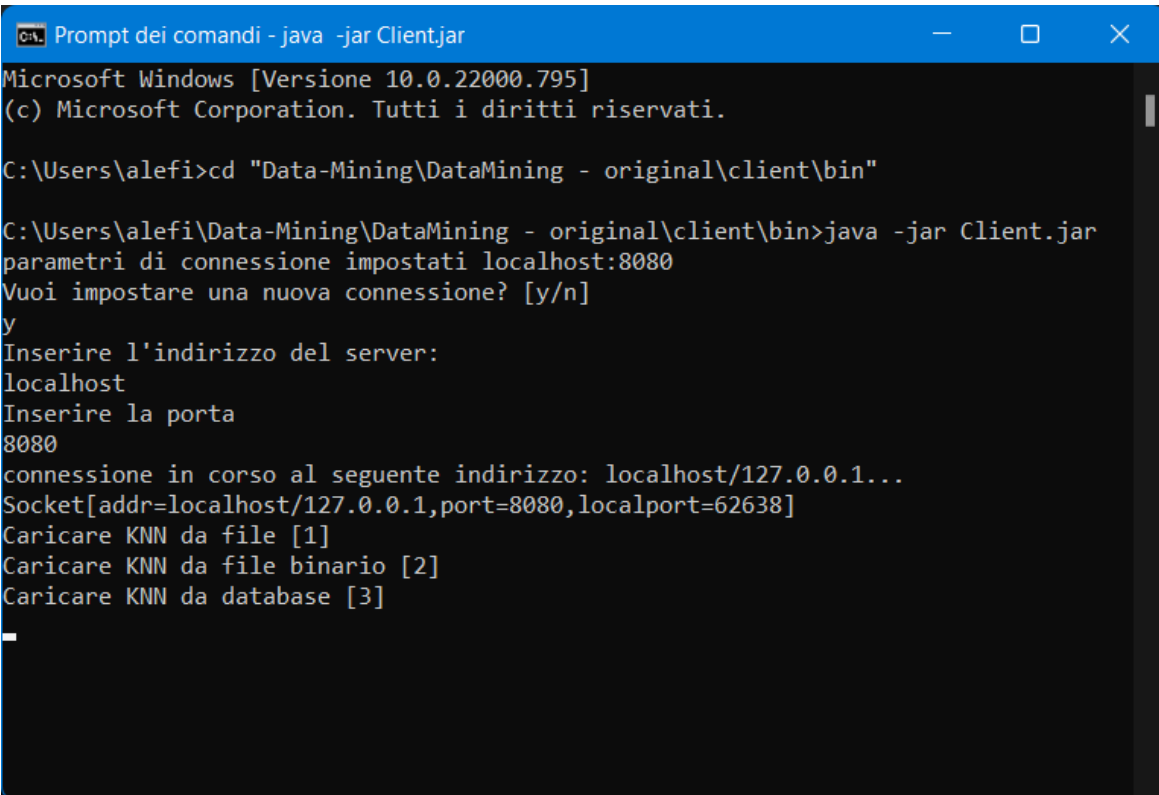
Versione Base	2
Input: file di testo	3
Input: file binario	7
Input: database	9
Versione Estensione	9
Input: file di testo	12
Input: file binario	13
Input: database	14

Guida utente

Versione Base

Una volta avviati sia l'applicativo server sia l'applicativo client, viene richiesto all'utente in quale modo caricare il training set già predisposto:

- ◆ tramite **file di testo**;
- ◆ tramite **file binario**;
- ◆ tramite **database**.



```
Prompt dei comandi - java -jar Client.jar
Microsoft Windows [Versione 10.0.22000.795]
(c) Microsoft Corporation. Tutti i diritti riservati.

C:\Users\alefi>cd "Data-Mining\DataMining - original\client\bin"

C:\Users\alefi\Data-Mining\DataMining - original\client\bin>java -jar Client.jar
parametri di connessione impostati localhost:8080
Vuoi impostare una nuova connessione? [y/n]
y
Inserire l'indirizzo del server:
localhost
Inserire la porta
8080
connessione in corso al seguente indirizzo: localhost/127.0.0.1...
Socket[addr=localhost/127.0.0.1,port=8080,localport=62638]
Caricare KNN da file [1]
Caricare KNN da file binario [2]
Caricare KNN da database [3]
_
```

Input: file di testo

Se vogliamo caricare un file di testo è sufficiente **digitare “1”** e **incollare il percorso del file** denominato *“provaC.dat”* (presente nella cartella *server\bin\provaC.dat*).

```

C:\Users\alefi>cd "Data-Mining\DataMining - original\client\bin"

C:\Users\alefi\Data-Mining\DataMining - original\client\bin>java -jar Client.jar
parametri di connessione impostati localhost:8080
Vuoi impostare una nuova connessione? [y/n]
y
Inserire l'indirizzo del server:
localhost
Inserire la porta
8080
connessione in corso al seguente indirizzo: localhost/127.0.0.1...
Socket[addr=localhost/127.0.0.1,port=8080,localport=62823]
Caricare KNN da file [1]
Caricare KNN da file binario [2]
Caricare KNN da database [3]
1
Inserire il path di un file contenente un training set valido:
C:\Users\alefi\Data-Mining\DataMining - original\server\bin\provaC.dat
KNN caricato con successo. E' stato effettuato un salvataggio nel file C:\Users\alefi\Data-Mining\
DataMining - original\server\bin\provaC.dmp
Inserisci valore discreto X[0]:
-
```

In caso di path del file errato, l'applicativo ci segnalerà un **errore** e ci chiederà di inserire nuovamente il percorso di un file di testo.

```

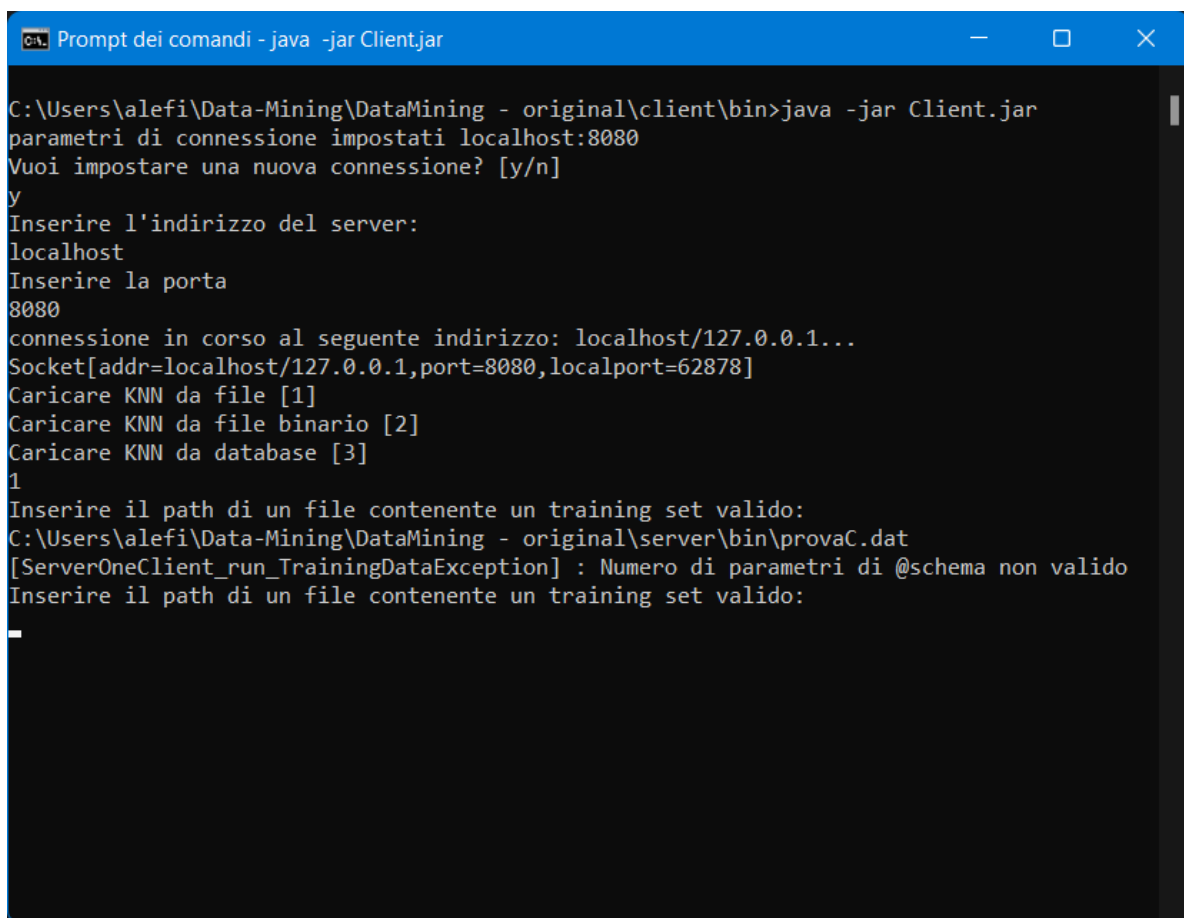
C:\Users\alefi>cd "Data-Mining\DataMining - original\client\bin"

C:\Users\alefi\Data-Mining\DataMining - original\client\bin>java -jar Client.jar
parametri di connessione impostati localhost:8080
Vuoi impostare una nuova connessione? [y/n]
y
Inserire l'indirizzo del server:
localhost
Inserire la porta
8080
connessione in corso al seguente indirizzo: localhost/127.0.0.1...
Socket[addr=localhost/127.0.0.1,port=8080,localport=62865]
Caricare KNN da file [1]
Caricare KNN da file binario [2]
Caricare KNN da database [3]
1
Inserire il path di un file contenente un training set valido:
C:\Users\alefi\Data-Mining\DataMining - original\server\bin\palla.dat
[ServerOneClient_run_FileNotFoundException] : File non trovato o e' una directory
Inserire il path di un file contenente un training set valido:
```

Il **training set** contenuto nel file di testo deve essere formattato nel seguente modo:

- ♦ nel primo rigo del file si indica il **numero di variabili** presenti nel training set (@schema);
- ♦ si procede con la **descrizione di ogni variabile**, specificando se si tratta di una variabile discreta o di una variabile numerica (@desc);
- ♦ si dichiara la **variabile target** (@target);
- ♦ si procede con l'**elenco degli esempi** presenti nel training set dopo aver indicato il **numero** di essi.

Se il training set non è formattato nel modo sopra citato possono generarsi alcuni errori:



```
Prompt dei comandi - java -jar Client.jar

C:\Users\alefi\Data-Mining\DataMining - original\client\bin>java -jar Client.jar
parametri di connessione impostati localhost:8080
Vuoi impostare una nuova connessione? [y/n]
y
Inserire l'indirizzo del server:
localhost
Inserire la porta
8080
connessione in corso al seguente indirizzo: localhost/127.0.0.1...
Socket[addr=localhost/127.0.0.1,port=8080,localport=62878]
Caricare KNN da file [1]
Caricare KNN da file binario [2]
Caricare KNN da database [3]
1
Inserire il path di un file contenente un training set valido:
C:\Users\alefi\Data-Mining\DataMining - original\server\bin\provaC.dat
[ServerOneClient_run_TrainingDataException] : Numero di parametri di @schema non valido
Inserire il path di un file contenente un training set valido:
-
```

Segnalazione della mancanza del numero di variabili

```

Prompt dei comandi - java -jar Client.jar
Microsoft Windows [Versione 10.0.22000.795]
(c) Microsoft Corporation. Tutti i diritti riservati.

C:\Users\alefi>cd "Data-Mining\DataMining - original\client\bin"

C:\Users\alefi\Data-Mining\DataMining - original\client\bin>java -jar Client.jar
parametri di connessione impostati localhost:8080
Vuoi impostare una nuova connessione? [y/n]
y
Inserire l'indirizzo del server:
localhost
Inserire la porta
8080
connessione in corso al seguente indirizzo: localhost/127.0.0.1...
Socket[addr=localhost/127.0.0.1,port=8080,localport=62914]
Caricare KNN da file [1]
Caricare KNN da file binario [2]
Caricare KNN da database [3]
1
Inserire il path di un file contenente un training set valido:
C:\Users\alefi\Data-Mining\DataMining - original\server\bin\provaC.dat
[ServerOneClient_run_TrainingDataException] : Il numero dei parametri di un esempio non e' coerente con
quanto dichiarato, riga: 1
Inserire il path di un file contenente un training set valido:

```

Segnalazione della mancanza della descrizione di qualche variabile

```

Prompt dei comandi - java -jar Client.jar
Microsoft Windows [Versione 10.0.22000.795]
(c) Microsoft Corporation. Tutti i diritti riservati.

C:\Users\alefi>cd "Data-Mining\DataMining - original\client\bin"

C:\Users\alefi\Data-Mining\DataMining - original\client\bin>java -jar Client.jar
parametri di connessione impostati localhost:8080
Vuoi impostare una nuova connessione? [y/n]
y
Inserire l'indirizzo del server:
localhost
Inserire la porta
8080
connessione in corso al seguente indirizzo: localhost/127.0.0.1...
Socket[addr=localhost/127.0.0.1,port=8080,localport=62929]
Caricare KNN da file [1]
Caricare KNN da file binario [2]
Caricare KNN da database [3]
1
Inserire il path di un file contenente un training set valido:
C:\Users\alefi\Data-Mining\DataMining - original\server\bin\provaC.dat
[ServerOneClient_run_TrainingDataException] : Parametro @target non dichiarato
Inserire il path di un file contenente un training set valido:

```

Segnalazione della mancanza dell'attributo target

Se, invece, il training set contenuto nel file è formattato nel modo corretto, esso verrà caricato sul server e, a questo punto, verrà chiesto al client di inserire **l'esempio** di cui si vuole **calcolare il target**.

```
Prompt dei comandi - java -jar Client.jar

C:\Users\alefi\Data-Mining\DataMining - original\client\bin>java -jar Client.jar
parametri di connessione impostati localhost:8080
Vuoi impostare una nuova connessione? [y/n]
y
Inserire l'indirizzo del server:
localhost
Inserire la porta
8080
connessione in corso al seguente indirizzo: localhost/127.0.0.1...
Socket[addr=localhost/127.0.0.1,port=8080,localport=62932]
Caricare KNN da file [1]
Caricare KNN da file binario [2]
Caricare KNN da database [3]
1
Inserire il path di un file contenente un training set valido:
C:\Users\alefi\Data-Mining\DataMining - original\server\bin\provaC.dat
KNN caricato con successo. E' stato effettuato un salvataggio nel file C:\Users\alefi\
Data-Mining\DataMining - original\server\bin\provaC.dmp
Inserisci valore discreto X[0]:
A
Inserisci valore continuo X[1]:
5
Inserisci valore k>=1:8_
```

Chiaramente, in caso di input sbagliati da parte dell'utente, saranno generati degli errori:

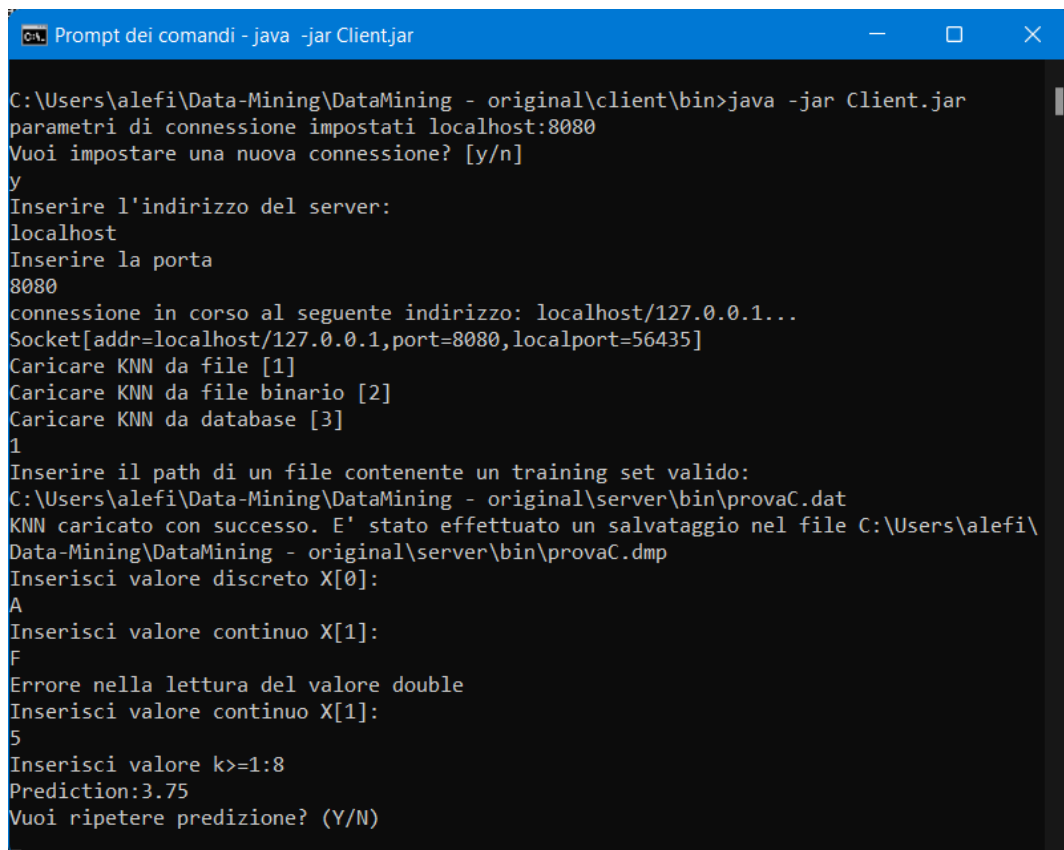
```
Prompt dei comandi - java -jar Client.jar

C:\Users\alefi>cd "Data-Mining\DataMining - original\client\bin"

C:\Users\alefi\Data-Mining\DataMining - original\client\bin>java -jar Client.jar
parametri di connessione impostati localhost:8080
Vuoi impostare una nuova connessione? [y/n]
y
Inserire l'indirizzo del server:
localhost
Inserire la porta
8080
connessione in corso al seguente indirizzo: localhost/127.0.0.1...
Socket[addr=localhost/127.0.0.1,port=8080,localport=56435]
Caricare KNN da file [1]
Caricare KNN da file binario [2]
Caricare KNN da database [3]
1
Inserire il path di un file contenente un training set valido:
C:\Users\alefi\Data-Mining\DataMining - original\server\bin\provaC.dat
KNN caricato con successo. E' stato effettuato un salvataggio nel file C:\Users\alefi\
Data-Mining\DataMining - original\server\bin\provaC.dmp
Inserisci valore discreto X[0]:
A
Inserisci valore continuo X[1]:
F
Errore nella lettura del valore double
Inserisci valore continuo X[1]:
5
Inserisci valore k>=1:_
```

Segnalazione errore nell'inserimento del valore continuo

Una volta inseriti i parametri in modo corretto, il server procederà ad eseguire i calcoli necessari e invierà al client l'output contenente il **valore predetto**; chiederà, infine, all'utente se vuole eseguire un'altra predizione.



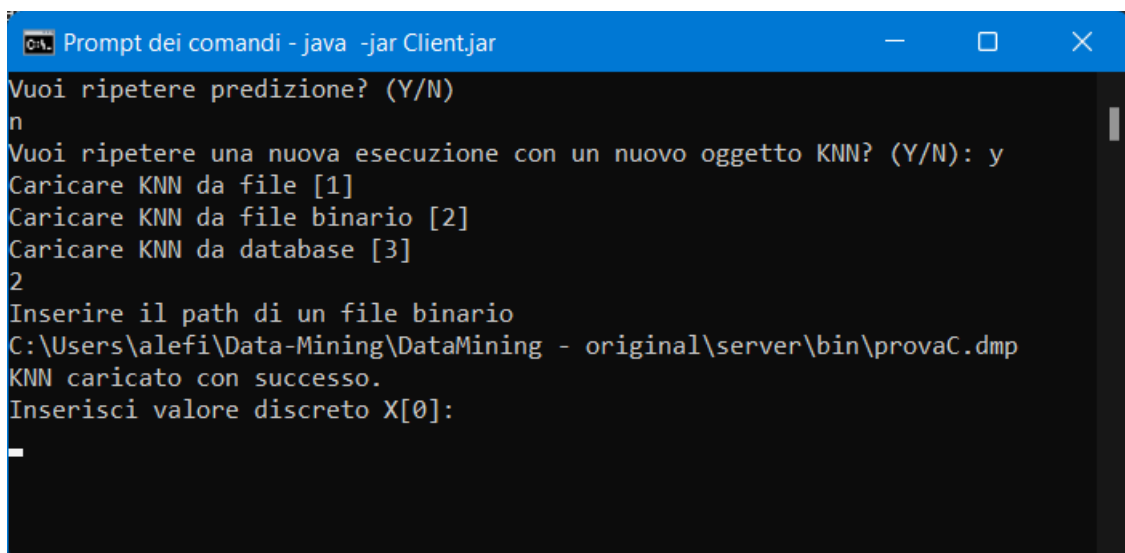
```
Prompt dei comandi - java -jar Client.jar

C:\Users\alefi\Data-Mining\DataMining - original\client\bin>java -jar Client.jar
parametri di connessione impostati localhost:8080
Vuoi impostare una nuova connessione? [y/n]
y
Inserire l'indirizzo del server:
localhost
Inserire la porta
8080
connessione in corso al seguente indirizzo: localhost/127.0.0.1...
Socket[addr=localhost/127.0.0.1,port=8080,localport=56435]
Caricare KNN da file [1]
Caricare KNN da file binario [2]
Caricare KNN da database [3]
1
Inserire il path di un file contenente un training set valido:
C:\Users\alefi\Data-Mining\DataMining - original\server\bin\provaC.dat
KNN caricato con successo. E' stato effettuato un salvataggio nel file C:\Users\alefi\
Data-Mining\DataMining - original\server\bin\provaC.dmp
Inserisci valore discreto X[0]:
A
Inserisci valore continuo X[1]:
F
Errore nella lettura del valore double
Inserisci valore continuo X[1]:
5
Inserisci valore k>=1:8
Prediction:3.75
Vuoi ripetere predizione? (Y/N)
```

L'applicativo funziona allo stesso modo se l'utente decide di inserire in input un **file binario** o un **database**.

Input: file binario

Per inserire in input un file binario è necessario **digitare "2"** e **incollare il percorso** del file "*provaC.dmp*" (contenuto nella cartella *server\bin\provaC.dmp*).



```
Prompt dei comandi - java -jar Client.jar

Vuoi ripetere predizione? (Y/N)
n
Vuoi ripetere una nuova esecuzione con un nuovo oggetto KNN? (Y/N): y
Caricare KNN da file [1]
Caricare KNN da file binario [2]
Caricare KNN da database [3]
2
Inserire il path di un file binario
C:\Users\alefi\Data-Mining\DataMining - original\server\bin\provaC.dmp
KNN caricato con successo.
Inserisci valore discreto X[0]:
A
```

Se viene inserito un path errato, l'applicativo genererà un errore e richiederà l'input

```
Prompt dei comandi - java -jar Client.jar
Microsoft Windows [Versione 10.0.22000.795]
(c) Microsoft Corporation. Tutti i diritti riservati.

C:\Users\alefi>cd "Data-Mining\DataMining - original\client\bin"

C:\Users\alefi\Data-Mining\DataMining - original\client\bin>java -jar Client.jar
parametri di connessione impostati localhost:8080
Vuoi impostare una nuova connessione? [y/n]
y
Inserire l'indirizzo del server:
localhost
Inserire la porta
8080
connessione in corso al seguente indirizzo: localhost/127.0.0.1...
Socket[addr=localhost/127.0.0.1,port=8080,localport=56505]
Caricare KNN da file [1]
Caricare KNN da file binario [2]
Caricare KNN da database [3]
2
Inserire il path di un file binario
C:\Users\alefi\Data-Mining\DataMining - original\server\bin\palla.dmp
[ServerOneClient_run_IOException | ClassNotFoundException]: File non trovato o e' una directory
Inserire il path di un file binario
```

Dopo aver inserito un file corretto, il training set verrà caricato sul server e verranno eseguiti gli stessi passaggi che si sono succeduti avendo un file di testo come metodo di caricamento dell'input.

```
Prompt dei comandi - java -jar Client.jar
Microsoft Windows [Versione 10.0.22000.795]
(c) Microsoft Corporation. Tutti i diritti riservati.

C:\Users\alefi>cd "Data-Mining\DataMining - original\client\bin"

C:\Users\alefi\Data-Mining\DataMining - original\client\bin>java -jar Client.jar
parametri di connessione impostati localhost:8080
Vuoi impostare una nuova connessione? [y/n]
y
Inserire l'indirizzo del server:
localhost
Inserire la porta
8080
connessione in corso al seguente indirizzo: localhost/127.0.0.1...
Socket[addr=localhost/127.0.0.1,port=8080,localport=56521]
Caricare KNN da file [1]
Caricare KNN da file binario [2]
Caricare KNN da database [3]
2
Inserire il path di un file binario
C:\Users\alefi\Data-Mining\DataMining - original\server\bin\provaC.dmp
KNN caricato con successo.
Inserisci valore discreto X[0]:
A
Inserisci valore continuo X[1]:
5
Inserisci valore k>=1:8
Prediction:3.75
Vuoi ripetere predizione? (Y/N)
```


Input: database

Se si vuole usare un database come metodo di caricamento è sufficiente **digitare “3”** e **indicare il nome della tabella** (*provaC* nel nostro caso) dalla quale prelevare i valori del training set.

Attenzione: ciò è possibile soltanto se prima è stato eseguito lo script SQL contenuto nella cartella *server\bin\script.sql*

```
C:\> Prompt dei comandi - java -jar Client.jar
Caricare KNN da file [1]
Caricare KNN da file binario [2]
Caricare KNN da database [3]
3
Inserire il nome della tabella presente nel database
provaC
KNN caricato con successo. E' stato effettuato un salvataggio nel file provaC.dmp
Inserisci valore discreto X[0]:
A
Inserisci valore continuo X[1]:
5
Inserisci valore k>=1:8
Prediction:3.75
Vuoi ripetere predizione? (Y/N)
-
```

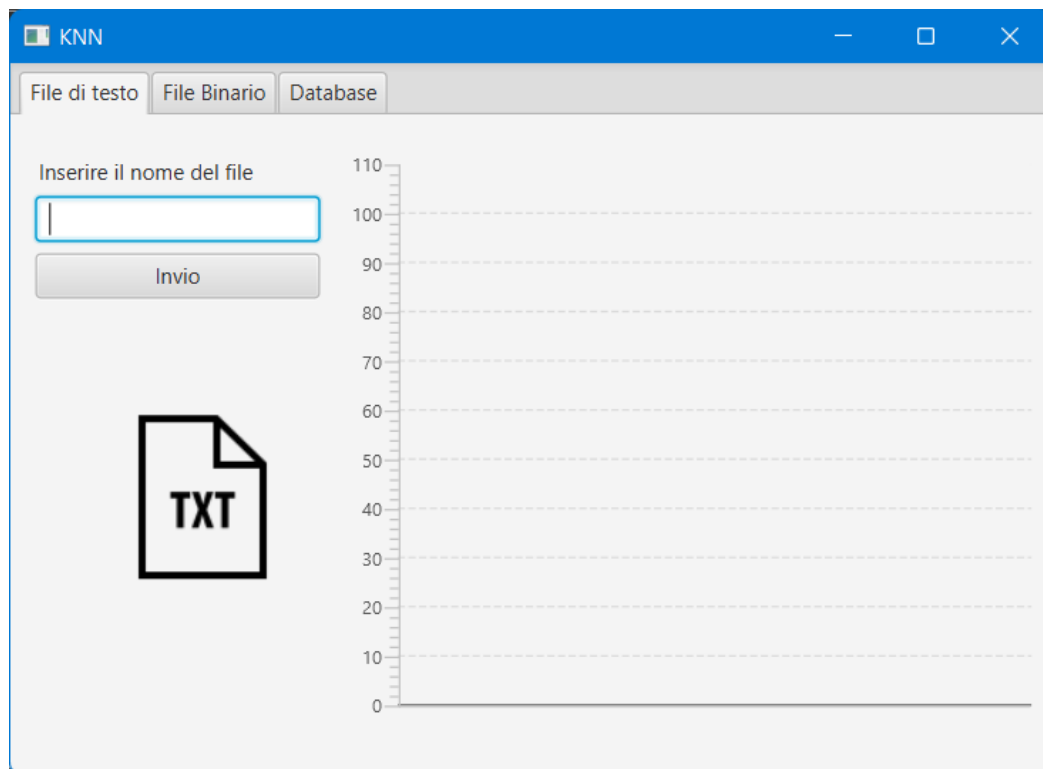
Versione Estensione

Nella versione estesa è stato deciso di mantenere lo stesso server, implementando invece **un'interfaccia grafica** per il client.

Alla prima apertura dell'applicativo client apparirà una schermata iniziale che richiederà all'utente di inserire i parametri da impostare per la connessione al server.

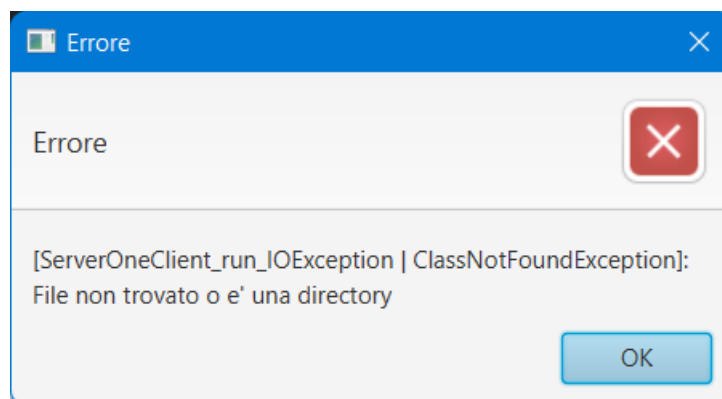
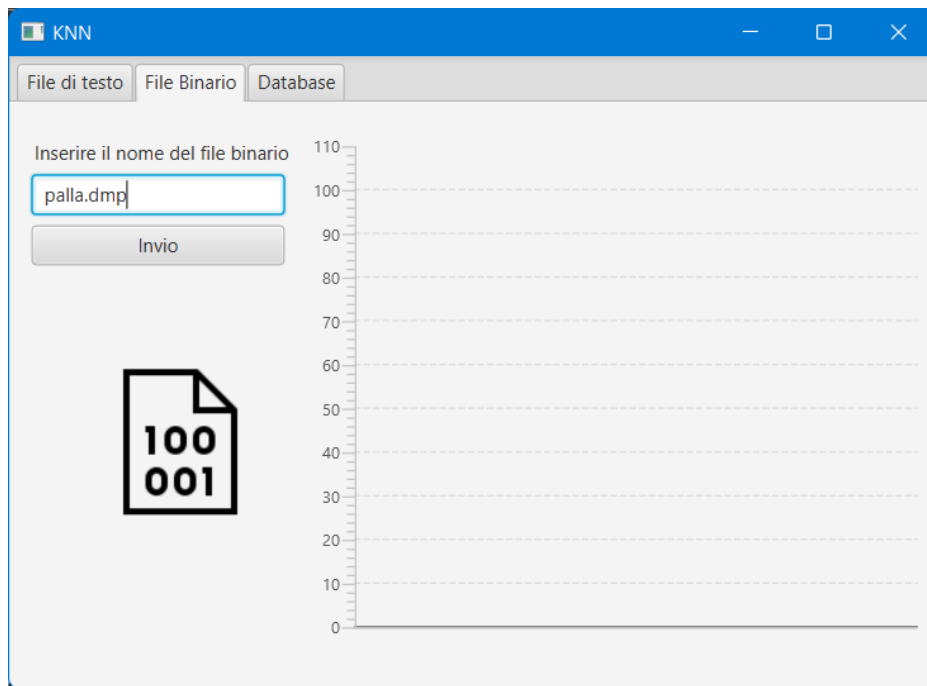


Una volta eseguita la connessione con il server, l'applicativo client chiederà all'utente il **metodo di caricamento** del training set (che può essere nuovamente un file di testo, un file binario o un database)



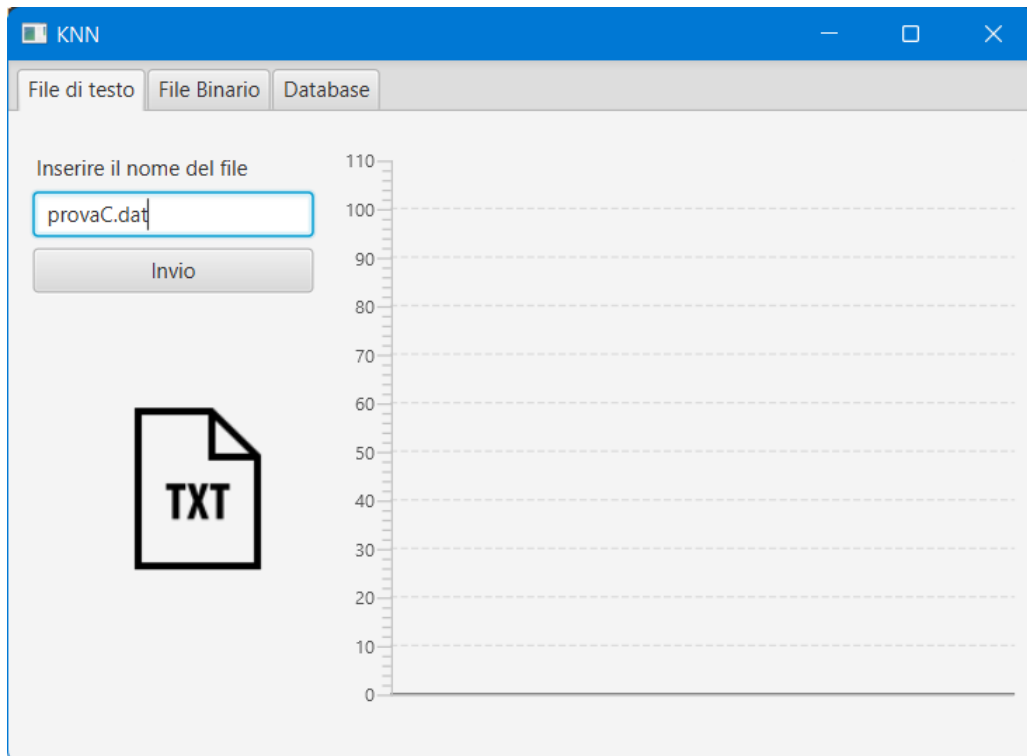
- ◆ se si vuole caricare il training set da un **file di testo**, basterà digitare "*provaC.dat*" nella sezione relativa al file di testo;
- ◆ se lo si vuole caricare mediante **file binario**, basterà digitare "*provaC.dmp*" nella sezione relativa al file binario;
- ◆ se, invece, lo si vuole caricare mediante **database**, basterà digitare "*provaC*" nella sezione relativa al database.

Anche in questo caso, se per il metodo di caricamento del training set scelto dall'utente si inserisce un parametro sbagliato, si avranno degli errori:



Una volta caricato il training set sul server, l'applicativo client procederà nello stesso modo previsto per la versione base del progetto, ovvero richiedendo in input all'utente **l'esempio** di cui si vuole **calcolare il target**.

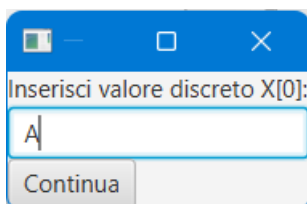
Input: file di testo



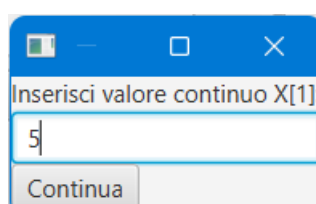
The screenshot shows the KNN application window with the 'File di testo' tab selected. The input field contains the filename 'provaC.dat'. Below the input field is an 'Invio' button. To the left of the input field is a 'TXT' file icon. To the right is a vertical axis with a scale from 0 to 110.

Dopo aver inserito l'esempio in input, il server esegue i calcoli e mostrerà l'output in un grafico contenente sia gli **esempi** presenti nel training set (rappresentati con un **cerchio rosso**) sia l'**esempio predetto** (rappresentato con un **quadrato giallo**).

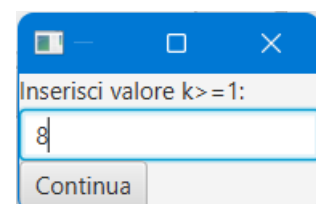
L'esempio predetto viene anche segnalato nella legenda del grafico ("**esempio calcolato**").



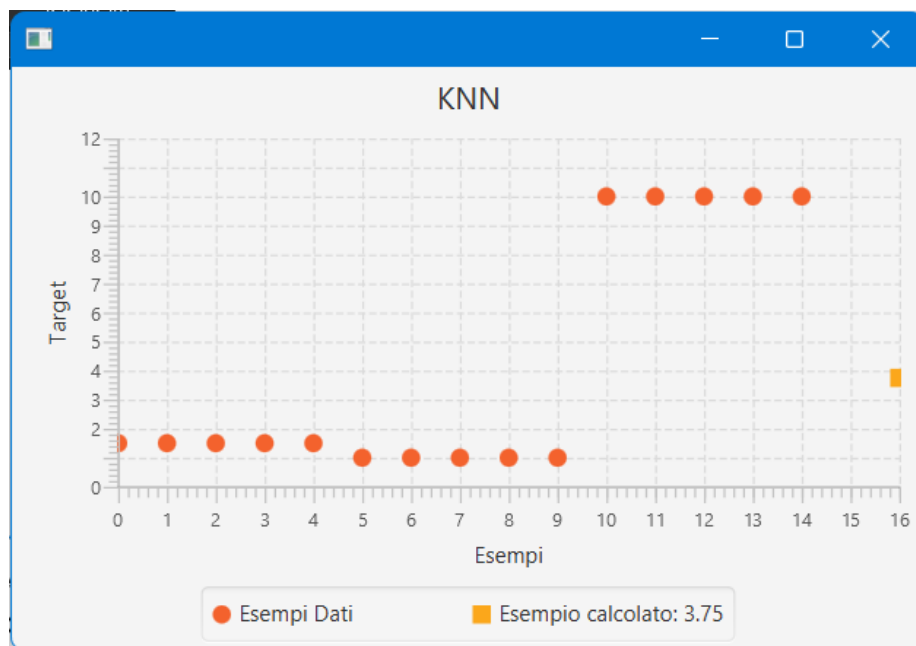
The dialog box prompts for a discrete value for X[0]. The input field contains the letter 'A'. Below the input field is a 'Continua' button.



The dialog box prompts for a continuous value for X[1]. The input field contains the number '5'. Below the input field is a 'Continua' button.

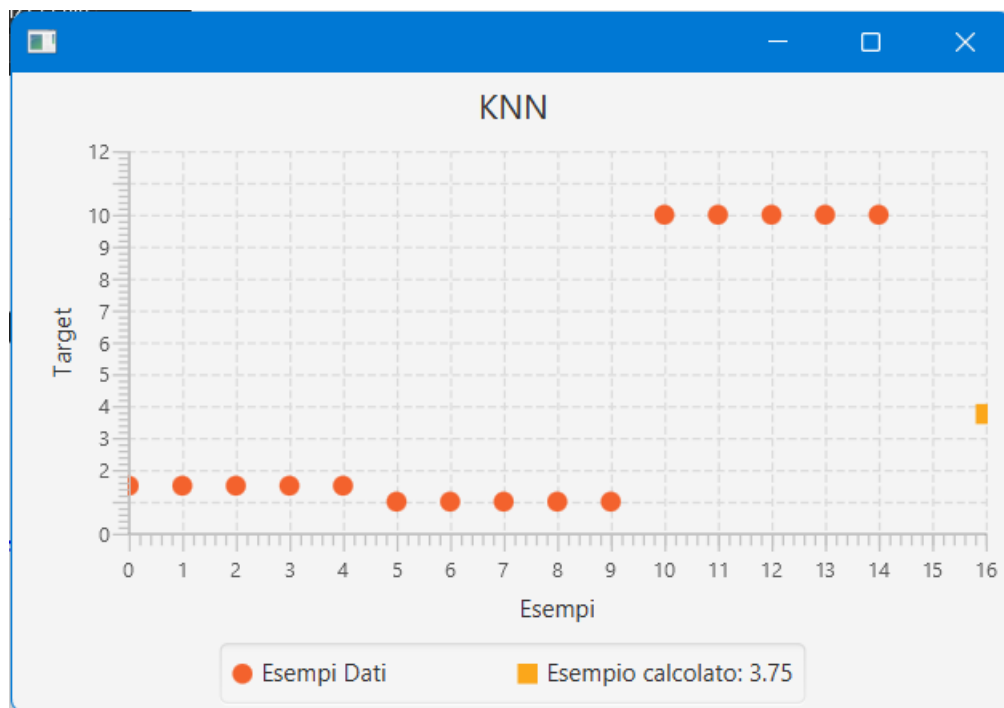


The dialog box prompts for a value k greater than or equal to 1. The input field contains the number '8'. Below the input field is a 'Continua' button.



Input: file binario

The figure shows the "KNN" application window with the "File Binario" tab selected. The window has three tabs: "File di testo", "File Binario", and "Database". The "File Binario" tab contains a text input field labeled "Inserire il nome del file binario" with the text "provaC.dmp" entered. Below the input field is an "Invio" button. To the left of the input field is a file icon with the text "100 001". To the right of the input field is a vertical axis ranging from 0 to 110.



Input: database

The interface is titled "KNN" and has three tabs: "File di testo", "File Binario", and "Database". The "Database" tab is active. It contains a text input field labeled "Inserire il nome della tabella" with the value "provaC". Below the input field is an "Invio" button. To the left of the input field is a database icon (cylinder). To the right is a large empty plot area with a y-axis ranging from 0 to 110.

