

K-MEANS

K-MEANS è un progetto svolto durante il corso di Metodi Avanzati di Programmazione e consiste nell'implementazione dell'algoritmo K-means.

L'algoritmo K-means è un metodo di apprendimento non supervisionato utilizzato per raggruppare i dati.

Questo algoritmo divide iterativamente i punti dati in K cluster minimizzando la varianza all'interno di ciascun cluster. L'algoritmo K-Means esegue la divisione degli oggetti in cluster che condividono somiglianze e sono dissimili rispetto agli oggetti appartenenti ad un altro cluster.

Funzionamento:

1. Si specifica il numero desiderato di cluster K.
2. L'algoritmo K-means assegna poi ciascuna osservazione a esattamente uno dei K cluster.

Essendo una tecnica di apprendimento non supervisionata non ci sono dati etichettati per questo raggruppamento.

Note di Progetto

Il progetto è stato interamente sviluppato utilizzando l'IDE IntelliJ, con JDK 19. Per connettersi al database si è utilizzato il connettore JDBC di MySQL nella versione 8.0.29. Il Database è stato creato utilizzando un script MySQL. La versione di MySQL utilizzata è la 8.0. Per lo sviluppo dell'interfaccia grafica si è fatto utilizzo della JavaFX 20.

Guida Utente

Prerequisiti

Prima di poter utilizzare l'applicazione è necessario soddisfare i seguenti prerequisiti:

- Java JRE: versione 19 o superiore
 - il progetto è stato scritto con la JDK 19 versioni inferiori di JRE potrebbero causare problemi di compatibilità
 - [Guida all'installazione][<https://docs.oracle.com/goldengate/1212/gg-winux/GDRAD/java.htm#BGBFJHAB>]
- MySQL versione 5.7 o maggiore
 - download: [MySQL :: Download MySQL Community Server](#)

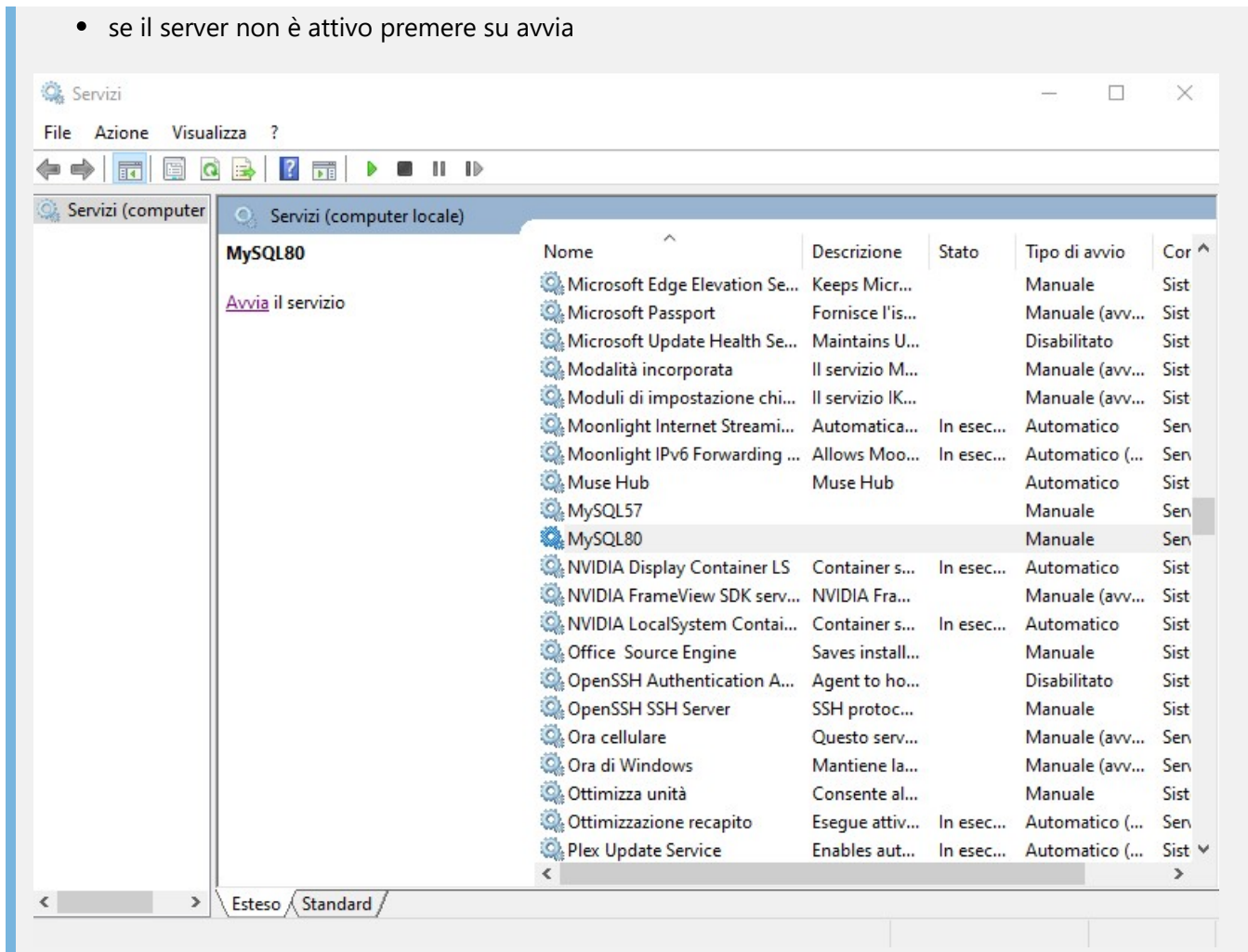
Avviare l'applicazione

Per prima cosa verificare che il server MySQL sia attivo:

Se su Windows:

- premere: **shift + r**
- digitare nella finestra aperta: **services.msc** e premere ok
- scorrere i servizi fino a MySQL

- se il server non è attivo premere su avvia



Se su Linux:

- verificare lo stato del server: `sudo systemctl status mysql`
- attivare il server: `sudo systemctl stop mysql`

Attivato il server MySQL si può creare il database:

- eseguire lo script SQL `createDB.SQL` presente nella cartella `scpritSQL`

A questo punto si può iniziare ad utilizzare l'applicazione.

1. Per prima cosa va sempre avviato il server ovvero il file `KmeansServer.jar`.

1. Per avviarlo si può:

1. Doppio click sul file `1.server.bat`
2. Aprendo il terminale nella cartella contenente il file e inserendo il comando: `java -jar KmeansServer.jar`

2. A questo punto si può avviare il file `KmeansClient.jar` allo stesso modo:

1. Doppio click sul file `2.client.bat`
2. Digitando nel terminale aperto nella cartella contenente il file: `java -jar KmeansClient.jar localhost 8080`

Utilizzo

Avviato prima il server poi il client si saranno aperte due finestre, una mostrerà il server:

```
Server avviato...  
Connessione con Client: 127.0.0.1
```

l'altra invece mostrerà il client:

```
addr = localhost/127.0.0.1  
Socket[addr=localhost/127.0.0.1,port=8080,localport=54877]  
Scegli una opzione  
(1) Carica Cluster da File  
(2) Carica Dati  
Risposta:
```

L'interazione con l'utente viene effettuata attraverso il client quindi proseguiremo utilizzando solo quest'ultimo (ATTENZIONE: il Server deve comunque rimanere aperto)

L'interfaccia del client mostra due possibilità:

1. Caricare i cluster da File
2. Caricare invece i dati dal Database e calcolare i cluster sui quei dati

Caricare i cluster da File

Inserendo 1 dovremo inserire il nome della tabella nel Database e il numero di iterate, in questo modo il Server aprirà il file `nomeTabella_iterate.bin` e verranno mostrati a schermo sul client i cluster memorizzati in quel file:

```
Scegli una opzione  
(1) Carica Cluster da File  
(2) Carica Dati  
Risposta:1  
Nome tabella:playtennis  
Numero iterate:2  
0: T: overcast, 29.605, normal, weak, yes  
1: T: overcast, 13.975, high, strong, yes  
2: T: rain, 6.25, normal, strong, no  
3: T: sunny, 11.316666666666668, normal, weak, yes  
  
Vuoi scegliere una nuova operazione da menu?(y/n)
```

A questo punto l'applicazione permette di scegliere se effettuare una nuova operazione e quindi tornare alla scelta precedente, oppure terminare l'esecuzione.

Calcolare i Cluster

Inserendo 2 invece, verrà richiesto il nome della tabella nel Database dalla quale si voglio recuperare i dati. Poi viene richiesto il numero di cluster da calcolare, quindi il client manderà la richiesta al server ed il server risponderà con i cluster calcolati:

```
Scegli una opzione
(1) Carica Cluster da File
(2) Carica Dati
Risposta:2
Nome tabella:playtennis
Numero di cluster:4
Clustering output:
  Numero di Iterazioni: 2
0: T: overcast, 29.605, normal, weak, yes
1: T: overcast, 13.975, high, strong, yes
2: T: rain, 6.25, normal, strong, no
3: T: sunny, 11.316666666666668, normal, weak, yes

Vuoi ripetere l'esecuzione?(y/n)
```

A questo punto l'applicazione chiede se si vuole ripetere l'esecuzione, quindi verrà di nuovo chiesto il numero di cluster. Se si risponde no invece verrà chiesto se si vuole tornare alla scelta precedente oppure terminare l'esecuzione:

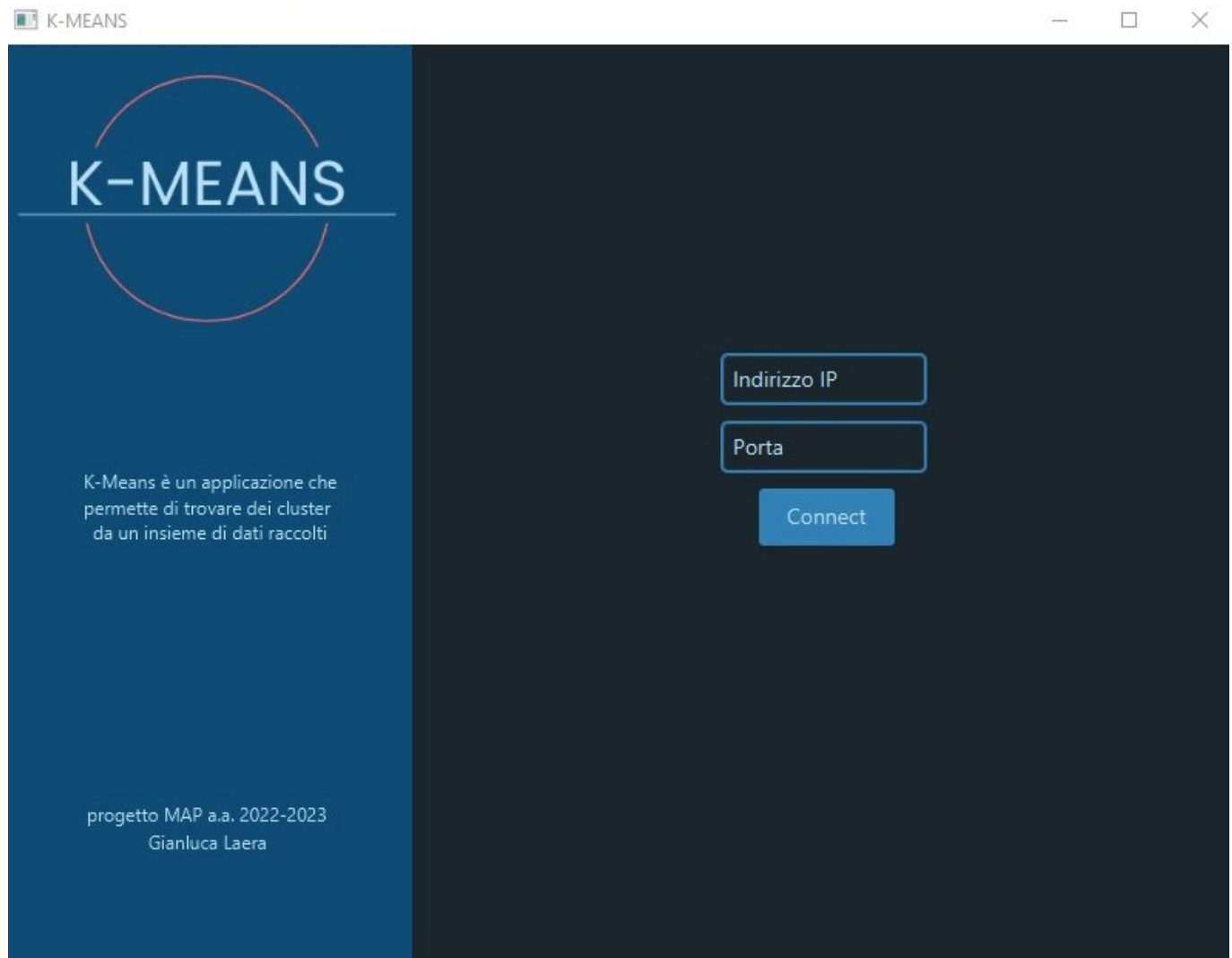
```
Vuoi scegliere una nuova operazione da menu?(y/n)y
Scegli una opzione
(1) Carica Cluster da File
(2) Carica Dati
Risposta:2
Nome tabella:playtennis
Numero di cluster:3
Clustering output:
  Numero di Iterazioni: 3
0: T: overcast, 16.862000000000002, normal, strong, yes
1: T: sunny, 21.525, high, strong, no
2: T: rain, 5.0200000000000005, normal, weak, yes

Vuoi ripetere l'esecuzione?(y/n)y
Numero di cluster:4
Clustering output:
  Numero di Iterazioni: 3
0: T: rain, 7.5200000000000005, normal, strong, yes
1: T: rain, 6.275, normal, weak, yes
2: T: overcast, 29.605, normal, weak, yes
3: T: sunny, 24.533333333333333, high, weak, no

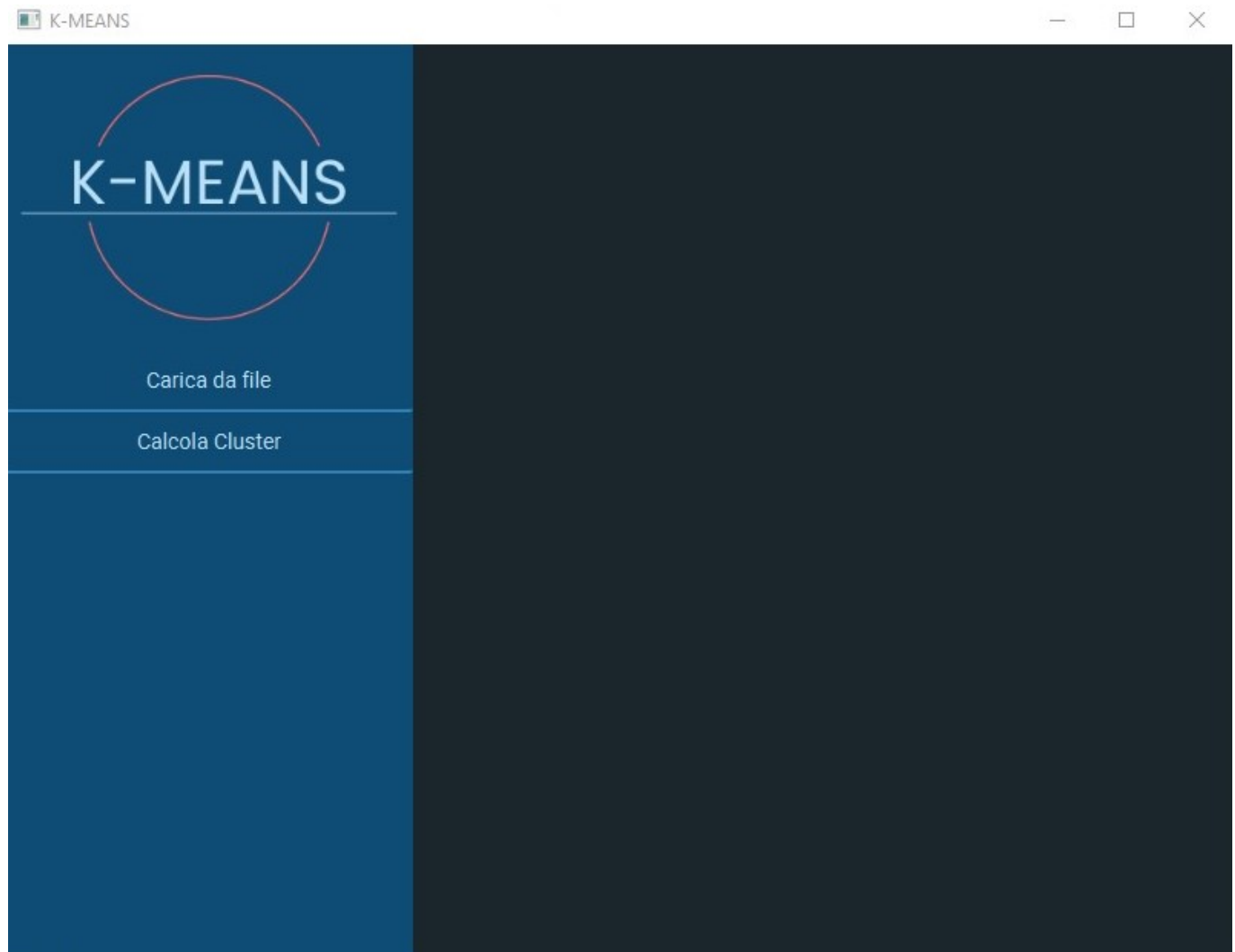
Vuoi ripetere l'esecuzione?(y/n)n
Vuoi scegliere una nuova operazione da menu?(y/n)n
```

Client con interfaccia grafica

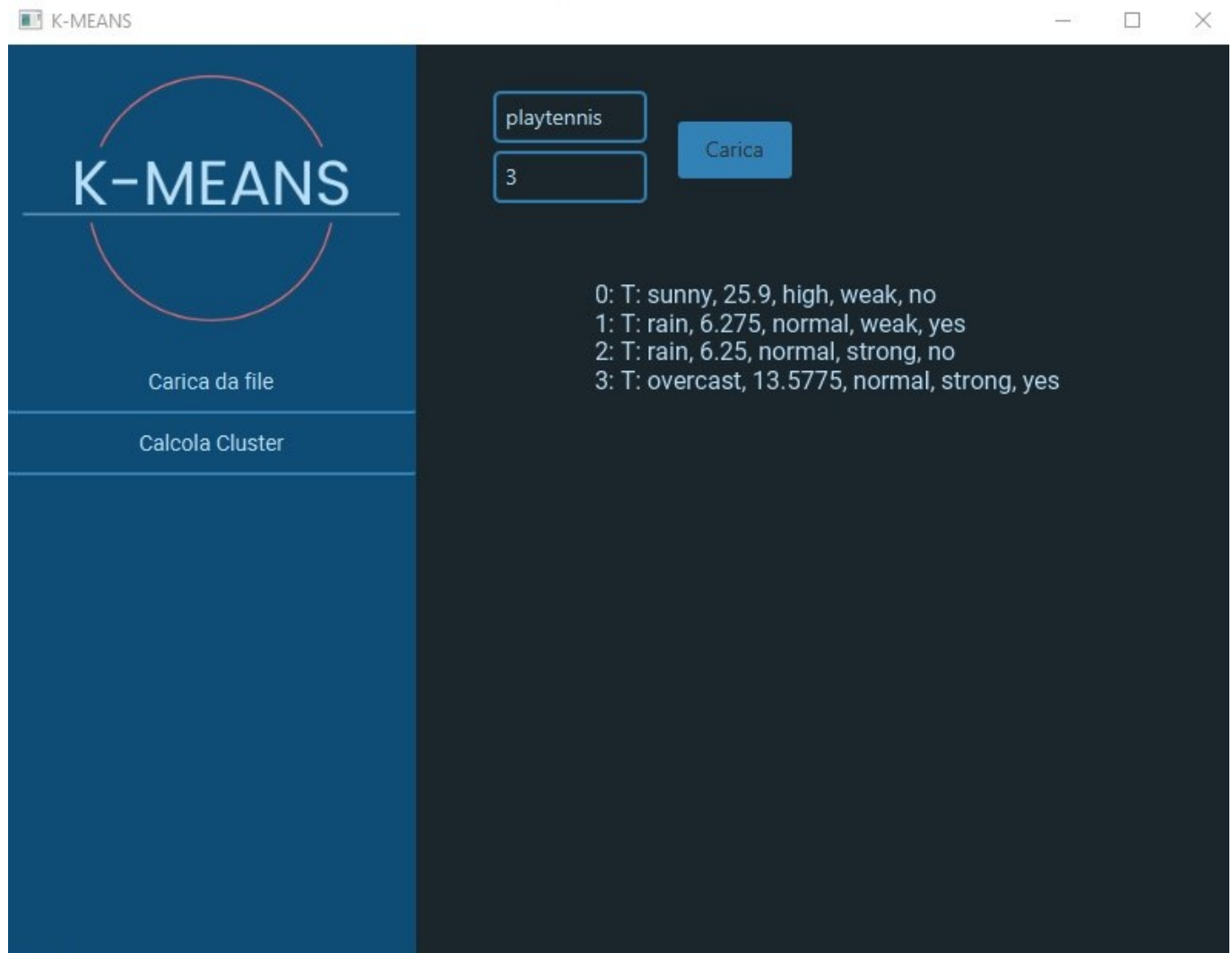
All'avvio viene presentata la finestra di connessione in cui si devono inserire indirizzo e porta a cui ci si intende connettere:



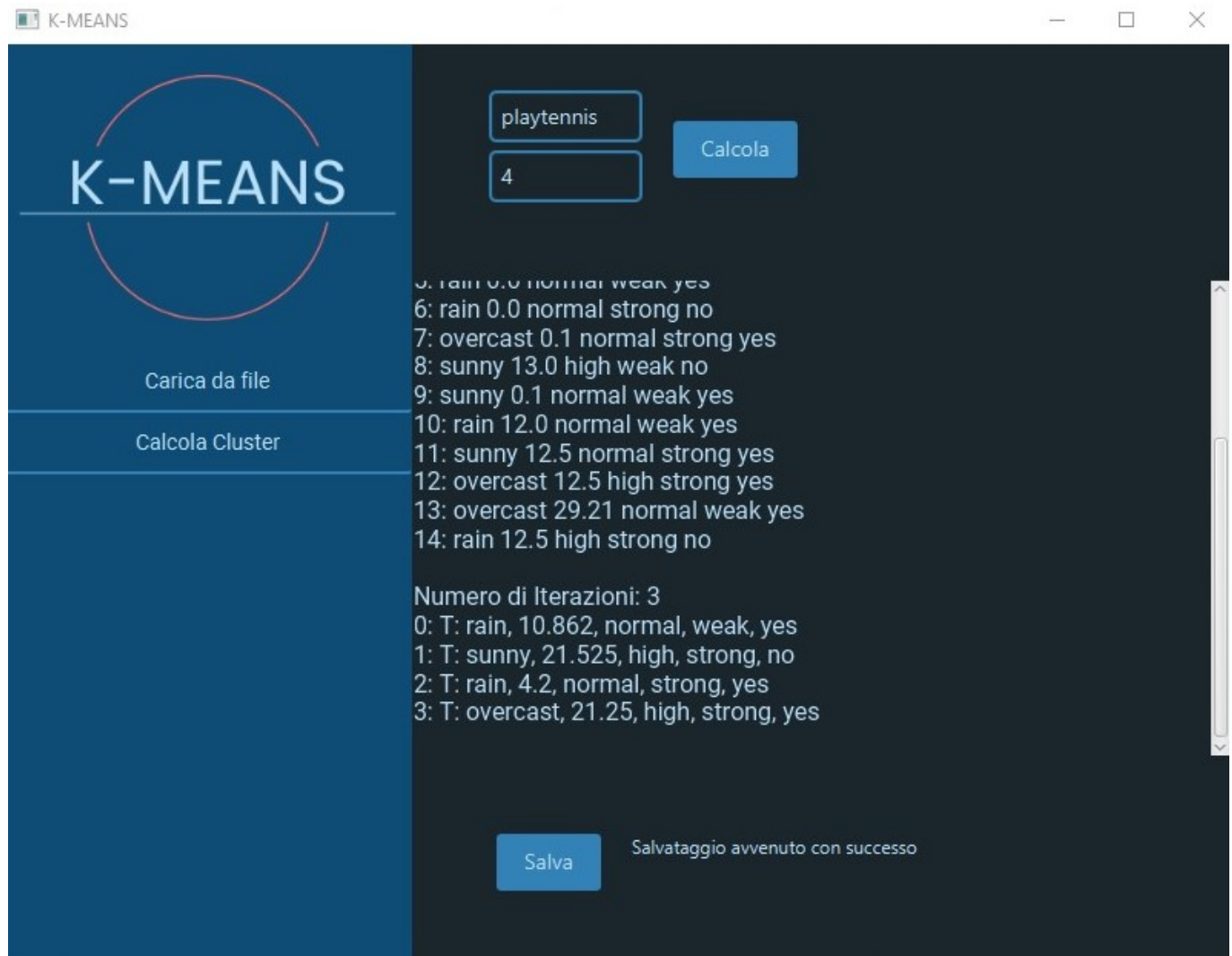
A questo punto si aprirà una nuova schermata con 2 pulsanti sulla sinistra che permettono di eseguire le diverse operazioni:



Caricare da file permette di caricare i cluster da un file e poi vengono mostrati a schermo:



Calcola cluuster invece permette di calcolare i cluster su una taballe recuperata dal database, e quindi mostra schermo la tabella e i cluster trovati:

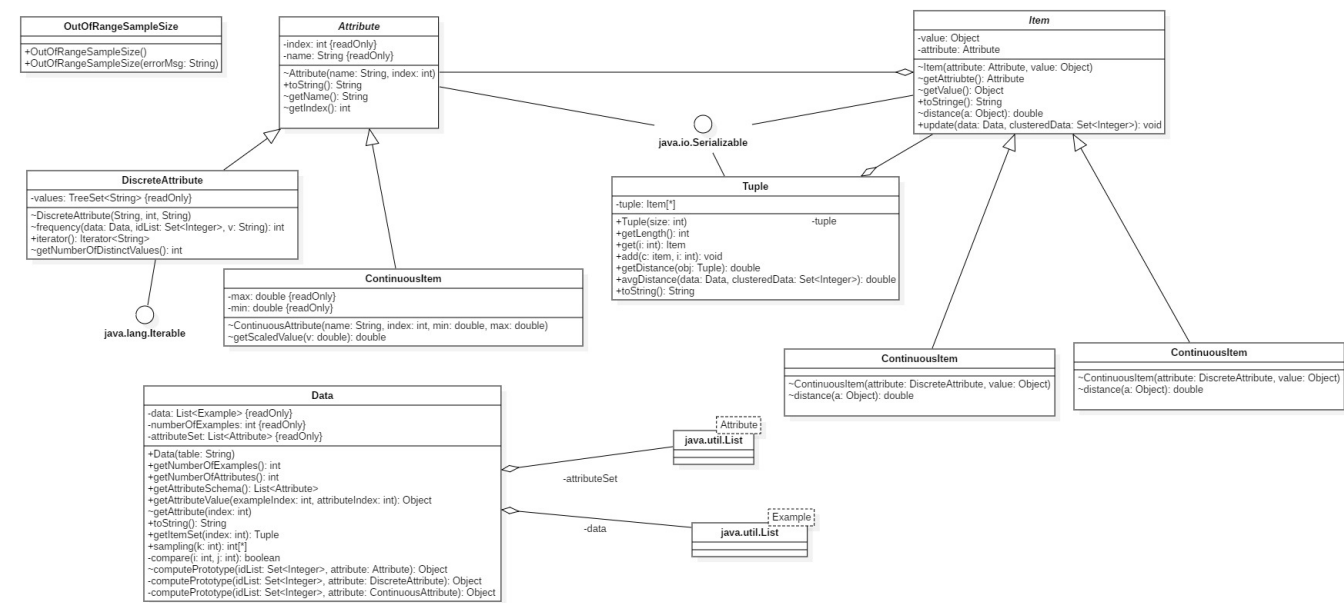


I cluster trovati poi possono essere salvati tramite il pulsante salva.

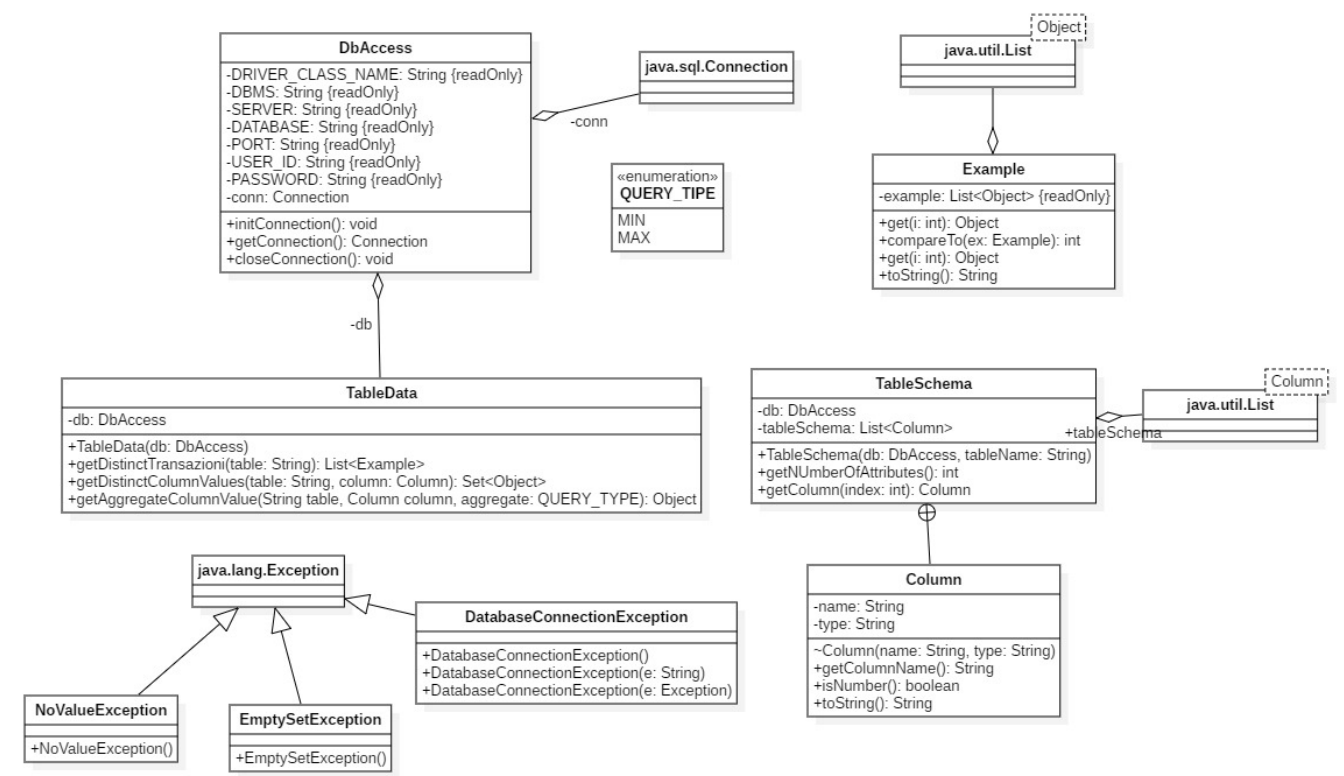
Diagramma UML delle classi

Di seguito vengono presentati i diagrammi UML delle classi del Server suddivisi per Package:

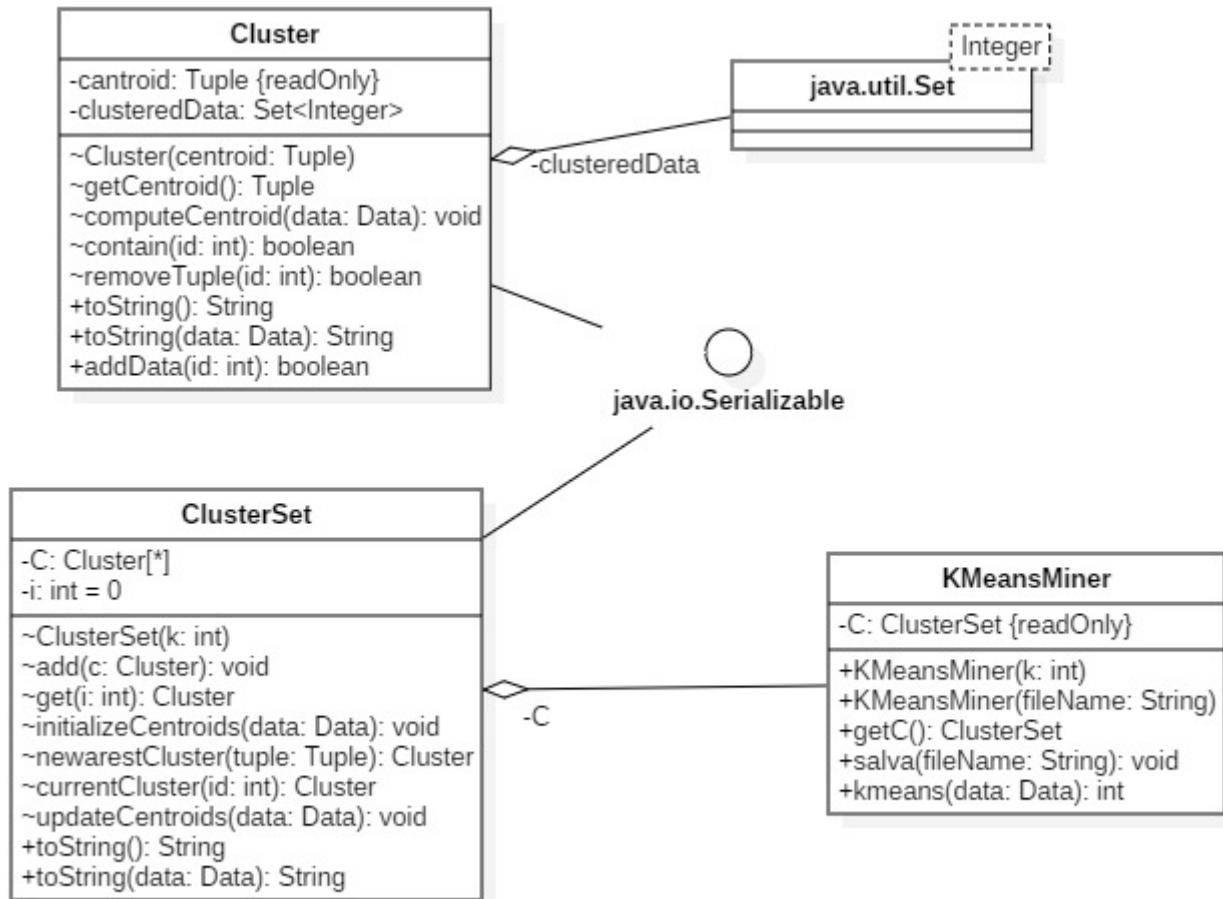
Package: data



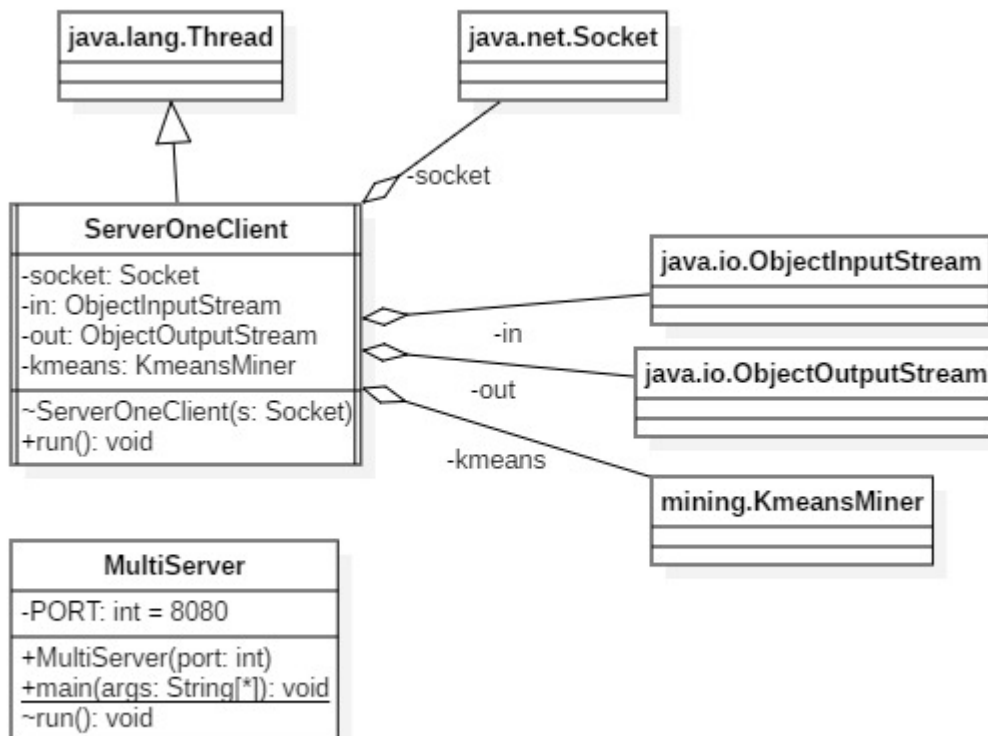
Package: database



Package: mining



Package: default



Di qui si passa ai diagrammi UML per il client:

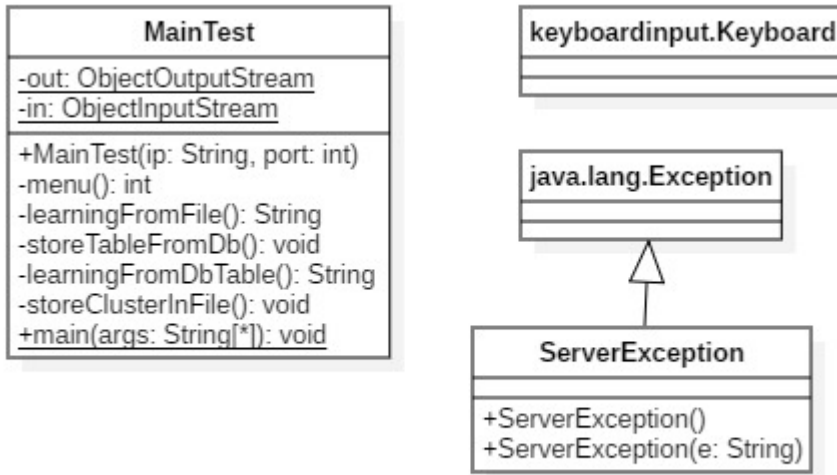


Diagramma UML delle classi (estensione)

Ecco invece il diagramma UML del client che implementa l'interfaccia grafica:

