

# Progetto Quality Treshold - Documentazione

A cura di: Vito Stefano Birardi

1 giugno 2025

## Indice

<b>1</b>	<b>Introduzione</b>	<b>3</b>
<b>2</b>	<b>Cos'è l'Algoritmo Quality Threshold</b>	<b>3</b>
<b>3</b>	<b>Principio di Funzionamento</b>	<b>3</b>
<b>4</b>	<b>Caratteristiche Distintive</b>	<b>3</b>
4.1	Vantaggi . . . . .	3
4.2	Svantaggi . . . . .	4
<b>5</b>	<b>Ambiti di Applicazione</b>	<b>4</b>
<b>6</b>	<b>Guida all'installazione</b>	<b>5</b>
6.1	Installazione JDK . . . . .	5
6.2	Variabili di ambiente . . . . .	5
6.3	Installazione MySQL . . . . .	6
6.4	Schermata iniziale . . . . .	6
6.5	Download dei file . . . . .	7
6.6	Configurazione del account MySQL . . . . .	7
6.7	Schermata finale . . . . .	8
<b>7</b>	<b>Eseguire il progetto base</b>	<b>9</b>
7.1	Importazione del database . . . . .	9
7.2	Normale funzionamento . . . . .	9
7.3	Struttura del progetto base . . . . .	10
<b>8</b>	<b>Avvio dell'applicazione</b>	<b>10</b>
8.1	Esecuzione di QTB_Server . . . . .	10
8.2	Esecuzione di QTB_Client . . . . .	10
8.3	Test QT_BASE . . . . .	11
<b>9</b>	<b>Eseguire il progetto esteso</b>	<b>14</b>
9.1	Premessa . . . . .	14
<b>10</b>	<b>Modifiche implementative</b>	<b>15</b>
10.1	Cambiamenti rispetto versione base . . . . .	15

## 1 Introduzione

Il progetto in questione utilizza l'algoritmo **Quality Threshold (QT)** per l'analisi dei dati. Tali dati possono essere estratti da un file o dalle tabelle di un database MySQL.

## 2 Cos'è l'Algoritmo Quality Threshold

L'algoritmo Quality Threshold è un algoritmo di **clustering** deterministico che appartiene alla famiglia degli algoritmi di raggruppamento gerarchico. A differenza degli algoritmi di partizionamento tradizionali, il QT non richiede di specificare a priori il numero di cluster da creare, ma si basa su un parametro di soglia di qualità (*threshold*) che definisce il raggio massimo consentito per ogni cluster.

## 3 Principio di Funzionamento

L'algoritmo si basa sulla **minimizzazione della distanza** tra i punti dati e il centroide del cluster di appartenenza. Il processo di clustering avviene attraverso i seguenti passaggi:

1. **Selezione del punto candidato:** Per ogni punto del dataset, viene valutata la possibilità di creare un nuovo cluster
2. **Formazione del cluster:** Vengono raggruppati tutti i punti che si trovano entro il raggio di soglia specificato
3. **Selezione del cluster ottimale:** Viene scelto il cluster che contiene il maggior numero di punti
4. **Iterazione:** Il processo si ripete sui punti rimanenti fino a quando tutti i punti sono stati assegnati a un cluster

## 4 Caratteristiche Distintive

L'algoritmo Quality Threshold presenta diverse caratteristiche che lo distinguono da altri approcci di clustering:

### 4.1 Vantaggi

- **Determinismo:** L'algoritmo restituisce sempre lo stesso risultato quando viene eseguito ripetutamente sullo stesso dataset, a differenza di algoritmi come il *k-means* che possono produrre risultati diversi a seconda dell'inizializzazione casuale
- **Numero di cluster automatico:** Non richiede di specificare il numero di cluster da creare, ma solo il raggio massimo (*threshold*), rendendo l'algoritmo più flessibile per dataset con strutture sconosciute
- **Robustezza al rumore:** È stato progettato per gestire efficacemente dati rumorosi e con distribuzioni non lineari

- **Approccio gerarchico:** Utilizza una strategia gerarchica per la creazione dei cluster, garantendo una maggiore stabilità dei risultati

## 4.2 Svantaggi

- **Complessità computazionale:** Richiede una potenza di calcolo maggiore rispetto ad algoritmi di partizionamento come il *k-means*, con una complessità temporale che può essere significativamente elevata per dataset di grandi dimensioni
- **Sensibilità al parametro threshold:** La scelta del raggio di soglia influenza notevolmente la qualità del clustering ottenuto

## 5 Ambiti di Applicazione

L'algoritmo Quality Threshold è particolarmente adatto per contesti in cui:

- I dati presentano rumore significativo
- La distribuzione dei dati è non lineare o presenta forme complesse
- Non si conosce a priori il numero ottimale di cluster
- È richiesta riproducibilità dei risultati
- La qualità del clustering è più importante dell'efficienza computazionale

## 6 Guida all'installazione

Prima di essere in grado di eseguire il programma, è necessario eseguire il file `risorse.bat` contenuto nella cartella Risorse.

Una volta eseguito il file, si aprirà una pagina di Powershell e seguirà un download.



Figura 1: Schermatta di download

Al termine del download del file compresso delle risorse, verranno estratti i file necessari per l'esecuzione del programma.

Per il progetto in questione è necessario installare il Java Developer Kit (JDK) nella versione 22.0.1 e il software di gestione del database MySQL nella sua versione 8.0.39.

La prima scheda di installazione che apparirà è quella del JDK.

Verrà chiesto se si vuole eseguire l'installazione del JDK tramite permessi di amministratore, cliccare su Sì.

### 6.1 Installazione JDK

Una volta confermata l'esecuzione come amministratore, si procederà all'installazione del JDK.



Cliccare su **Avanti**, nuovamente **Avanti** e infine, nel caso in cui l'installazione sia andata a buon fine, si dovrà cliccare sul tasto **Chiudi**.

### 6.2 Variabili di ambiente

è consigliato creare una variabile di ambiente per il JDK. Per farlo, si dovrà cercare la voce **variabili di ambiente** nella barra di ricerca di Windows e cliccare su Modifica le variabili di ambiente per il tuo account.

Successivamente si dovrà cliccare sul pulsante **Variabili di ambiente** e si aprirà una schermata con le variabili di ambiente.

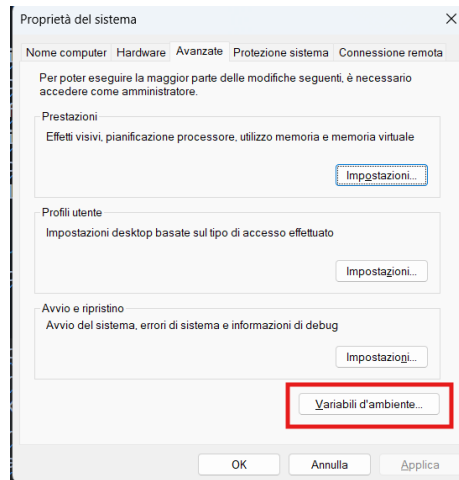


Figura 2: Schermata delle variabili di ambiente

Nella schermata, si dovrà cliccare sulla variabile **PATH** e modificarla in modo da aggiungere il percorso del JDK, inserendo il percorso in cui è stato installato il JDK.

Una volta completata l'installazione del JDK, si procederà automaticamente con l'installazione del MySQL.

### 6.3 Installazione MySQL

Anche qui sarà necessario fornire i permessi di amministratore, cliccare dunque su Sì.

### 6.4 Schermata iniziale

Dopodiché si aprirà la schermata di installazione di MySQL.

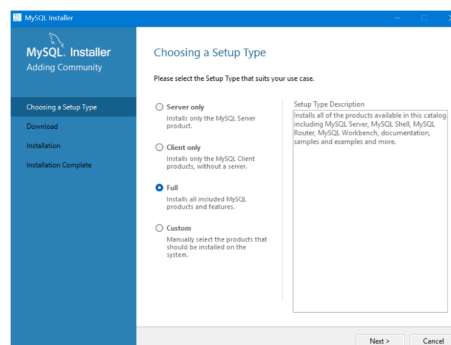


Figura 3: Scelta del setup

Selezionare il tipo di setup **Full** e cliccare su **Next**. Tale scelta permetterà di installare tutti i componenti necessari per l'esecuzione del programma.

## 6.5 Download dei file

Si passerà alla schermata di download dei file necessari per l'installazione, cliccare su **Execute**.

### Avvertenza

Tale schermata potrebbe richiedere un po' di tempo per il download dei file, a seconda della velocità della connessione internet.

Dopo aver scaricato i file, si dovrà cliccare nuovamente su **Execute**.

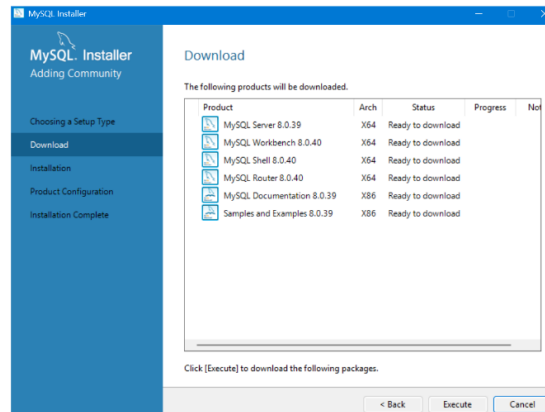


Figura 4: Download MySQL

## 6.6 Configurazione del account MySQL

Una volta terminato il download, si procederà con la configurazione dell'account MySQL. Si dovrà quindi scegliere una password, necessaria per accedere al database.

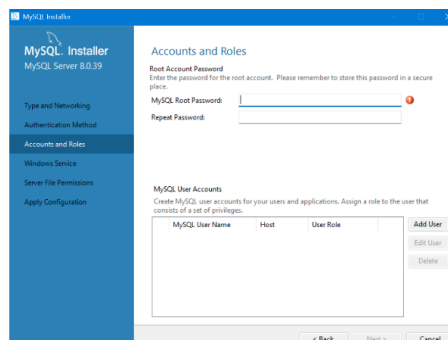


Figura 5: Configurazione account MySQL

Si giungerà infine alla schermata di applicazione della configurazione del database. Cliccare su **Execute** per applicare la configurazione.

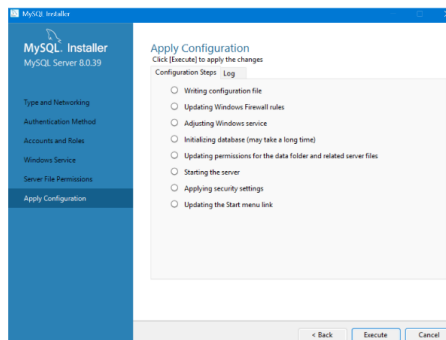


Figura 6: Completamento installazione

Una volta terminata anche l'installazione del MySQL, si dovrà cliccare su **Finish** per chiudere l'installler.

#### Avvertenza

Per un corretto funzionamento sia del JDK che del MySQL, è consigliato riavviare il computer.

## 6.7 Schermata finale

Al termine dell'installazione il programma `resources.bat` mostrerà la seguente schermata.



Figura 7: Completamento installazione



## 7 Eseguire il progetto base

### 7.1 Importazione del database

Per il corretto funzionamento del progetto base, è necessario, prima di eseguire il programma, importare il database `QT.sql` presente nella cartella `Progetto_MAP_2/risorse/` all'interno del proprio DBMS.

Siccome nella sezione precedente abbiamo fornito l'installer per Mysql, si consiglia di utilizzare quest'ultimo per importare il database. Per importare il database all'interno di Mysql, è necessario aprire il terminale, accedere con le credenziali di root a Mysql, e lanciare il seguente comando:

```
source *PERCORSO DEL FILE QT.sql*
```

#### Avvertenza

Si ricorda di cambiare il percorso del file `QT.sql` con quello corretto, effettuando un *drag and droop* del file all'interno del terminale, in modo da evitare errori di battitura.

Se il comando è stato eseguito correttamente, il terminale restituirà la seguente schermata:

### 7.2 Normale funzionamento

La base dell'algoritmo permette all'utente di scegliere un raggio massimo per il cluster.

Per ciascun punto interessato, ossia i punti con distanza inferiore al raggio impostato, l'algoritmo calcola i cluster candidati.

Una volta individuati i cluster, quello con il numero maggiore di punti verrà salvato nella cartella `./results/`, ossia nella root del progetto base. Il nome del file indicherà il database di partenza e il raggio massimo utilizzato per la ricerca, quindi scegliendo per esempio il database `gopicnic` con un raggio di 3, il file di output sarà `gopicnic_003.dmp`.

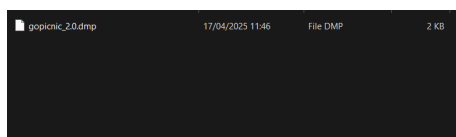


Figura 8: Risultati del salvataggio

Una volta creato il file di output, l'algoritmo provvederà ad eliminare anche i cluster non salvati, per evitare che vengano esaminati nuovamente.

Nel caso in cui più di un cluster abbia il numero massimo di punti, l'algoritmo ripeterà la procedura con il set ridotto di punti.

### 7.3 Struttura del progetto base

Il progetto base consiste in un'applicazione di tipo client/server. Di fatto, nella cartella **Jar+Bat** sono presenti i file **start\_server.bat** per l'esecuzione del server e **start\_client** per il client, che non sono altro che degli script batch per l'esecuzione dei file jar.

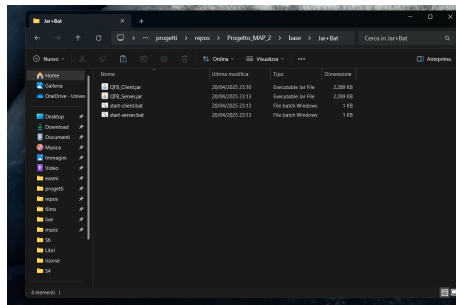


Figura 9: Jar del progetto base

#### Avvertenza

Per avviare il programma correttamente, è necessario eseguire prima il file **.bat** del server e poi quello del client.

## 8 Avvio dell'applicazione

### 8.1 Esecuzione di QTB\_Server

Avviando il file **start\_server.bat** verrà mostrata la seguente schermata, che indica che il server è in ascolto sulla porta 8080, in attesa di una richiesta da parte del client.

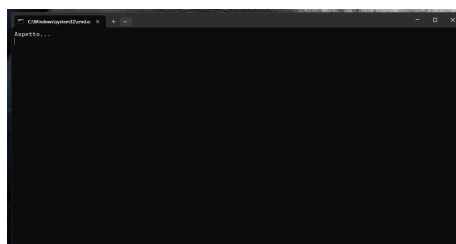


Figura 10: Server in ascolto

### 8.2 Esecuzione di QTB\_Client

Per quanto riguarda il client, una volta avviato, mostrerà la seguente schermata.

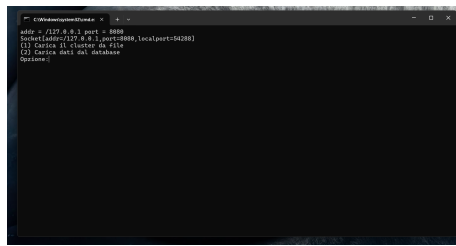
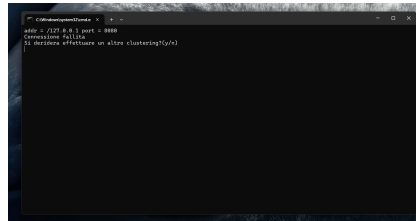


Figura 11: Schermata iniziale del client

#### Avvertenza

Nel caso in cui il server non si sia avviato correttamente, o sia stato avviato il client prima del server, il client mostrerà la seguente schermata, che indicherà appunto che la connessione al server non è avvenuta correttamente.



Una volta assicurati che il client comunichi correttamente con il server, si potrà decidere se caricare i dati da un file di un clustering precedente o da un database.

Se si sceglie di caricare i dati da una tabella del database, si procederà con l'inserimento del nome della tabella da cui si vogliono estrarre i dati. Nel caso in cui tale tabella non sia presente nel database, il client avviserà l'utente con un messaggio di errore e chiedendo di inserire un nome valido.

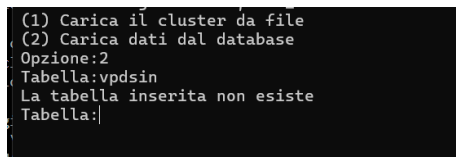


Figura 12: Errore tabella non presente

Una volta scelto correttamente una tabella, si procederà nel determinare il raggio del clustering e, automaticamente, il programma restituirà il risultato.

A fine esecuzione, il risultato del clustering verrà salvato nella cartella `./results/` con il nome della tabella e il raggio utilizzato per il clustering, come già spiegato in precedenza.

### 8.3 Test QT\_BASE

Durante l'esecuzione del programma, può capitare che l'utente inserisca dei valori che quest'ultimo non è in grado di gestire. Chiaramente nonostante la

```

Tabella:gopicnic
Raggio:2
Numero di cluster:5
1:Centroid=(rain 13.0 high weak no )
Examples:
[rain 13.0 high weak no ] dist=0.0
AvgDistance=0.0

2:Centroid=(sunny 0.1 normal weak no )
Examples:
[sunny 0.1 normal weak no ] dist=0.0
[sunny 40.0 low weak no ] dist=1.9975
AvgDistance=0.99875

3:Centroid=(sunny 30.3 high strong no )
Examples:
[sunny 30.3 high strong no ] dist=0.0
[overcast 12.5 high strong no ] dist=1.445
AvgDistance=0.7225

4:Centroid=(overcast 30.0 high weak yes )
Examples:
[sunny 30.3 high weak yes ] dist=1.0075
[overcast 30.0 high weak yes ] dist=0.0
[sunny 20.0 high weak yes ] dist=1.25
[overcast 29.21 normal weak yes ] dist=1.01975
[overcast 23.0 low weak yes ] dist=1.175
AvgDistance=0.8904500000000001

5:Centroid=(rain 0.0 normal strong no )
Examples:
[rain 40.0 high strong no ] dist=2.0
[rain 0.0 normal weak no ] dist=1.0
[rain 0.0 normal strong no ] dist=0.0
[overcast 0.1 normal strong no ] dist=1.0025
[rain 18.0 normal weak no ] dist=1.45
[sunny 12.5 normal strong no ] dist=1.3125
[rain 12.5 high strong no ] dist=1.3125
AvgDistance=1.1539285714285714

```

Figura 13: Risultato del clustering

presenza di queste situazioni, deve essere garantita la continuità di esecuzione del programma e la corretta amministrazione di queste situazioni estreme.

## Gestione del raggio

Nel caso in cui l'utente inserisca un valore del raggio maggiore del massimo raggio possibile, il programma restituirà un messaggio di errore e chiederà nuovamente il raggio.

Nel caso invece in cui l'utente inserisca un valore non numerico, il programma restituirà un messaggio di errore e chiederà nuovamente il raggio.

In entrambi i casi, il programma non si fermerà e continuerà a chiedere il raggio fino a quando non verrà inserito un valore corretto.

```

C:\Windows\system32\cmd.exe
addr = /127.0.0.1 port = 8880
Socket(addr=/127.0.0.1,port=8880,localport=55355)
(1) Carica il cluster da file
(2) Carica dati dal database
Opzione:2
Tabella:gopicnic
Raggio:5
Il raggio inserito è troppo grande
Si desidera effettuare un altro clustering?(y/n)

```

(a) Valore del raggio troppo grande

```

(2) Carica dati dal database
Opzione:2
Tabella:gopicnic
Raggio:d
Errore: inserisci un valore numerico!
Raggio:

```

(b) Valore del raggio non numerico

Figura 14: Errori nel valore del raggio

## Uscita dal programma

Una volta calcolato il raggio, il programma chiederà all'utente se ha intenzione di effettuare nuovamente un clustering: in caso positivo chiederà nuovamente la

misura del raggio, in caso opposto chiederà all'utente se vuole tornare al menù principale o chiudere il programma.



Figura 15: Gestione finale clustering

Domandando all'utente se si vuole ritornare al menù principale, si gestisce la chiusura del programma e, allo stesso tempo, il ritorno al menù principale, ossia alla situazione iniziale.

Inoltre, tale stratagemma previene lo spiacevole inconveniente per cui si è obbligati a forzare l'interruzione dell'esecuzione.

## 9 Eseguire il progetto esteso

### 9.1 Premessa

Rispetto al progetto base, si è reso necessario l'utilizzo di un gestore di dipendenze, in modo da poter installare le librerie necessarie per il progetto.

#### **Maven**

Il gestore di dipendenze scelto per questo scopo è **Maven**, il quale permette di scaricare le librerie necessarie e di gestire le versioni delle stesse.

Il vantaggio principale rispetto ad altri gestori di dipendenze, come il più blasonato Gradle, è la sua semplicità di utilizzo e la sua integrazione con gli IDE più diffusi.

Inoltre, al contrario di Gradle, Maven non richiede la creazione di un file di configurazione complesso, ma si basa su un semplice file `pom.xml` che contiene le informazioni sulle dipendenze del progetto.

Non richiede quindi la creazione di file di build o di scaricare i file delle dipendenze manualmente, ma il tutto viene gestito automaticamente.

#### **Telegram API**

Si è reso necessario l'utilizzo di un gestore delle dipendenze per via dell'utilizzo delle librerie **org.telegram.longpolling** e **org.telegram.telegrambots**, le quali permettono di interagire con l'API di Telegram.

## 10 Modifiche implementative

Come accennato nella premessa, è stato implementato un bot Telegram. Tale bot permette di interagire con il progetto direttamente dall'interfaccia di Telegram, rendendolo di fatto un client.

L'adozione di Maven per la gestione delle dipendenze, inoltre ha reso inutile l'utilizzo del file `risorse.bat`, in quanto le librerie vengono scaricate automaticamente e non è più necessario gestire manualmente le versioni delle stesse.

Di per se il funzionamento del progetto non è cambiato, è stato semplicemente trasformato il modo di interazione tra esso e l'utente, trasferendola quindi da un'interfaccia cli a una più interattiva.

### 10.1 Cambiamenti rispetto versione base

Dato che viene modificato solo la modalità di interazione con l'utente, la logica di funzionamento del processo non è stata modificata rispetto alla versione base.