

POLITECNICO DI TORINO

Corso di Laurea Magistrale in
Ingegneria Informatica

Tesi di Laurea Magistrale

**Laboratori didattici su SQL: analisi
del comportamento degli utenti con
tecniche di data mining**



Relatori

prof. Laura Farinetti
prof. Luca Cagliero

Candidato

Andrea Lupo

A.A. 2017-2018

Indice

Elenco delle tabelle	4
Elenco delle figure	5
1 Framework di analisi	10
1.1 Raccolta dei dati	11
1.2 Formattazione dei dati	12
1.3 Enumerazione delle sequenze frequenti	14
1.3.1 Scelta dell'algoritmo	16
1.3.2 cSpade	16
1.4 Categorizzazione delle sequenze	19
2 Funzionamento del programma di analisi	20
2.1 Visualizzazione delle sequenze	22
2.2 Finestre	23
2.2.1 Main Window	24
2.2.2 Results Window	25
2.2.3 Search Window	27
2.2.4 Categories Window	28
2.2.5 Join Categories	30
3 Analisi dei risultati ottenuti	31
3.1 Analisi esplorativa	32
3.2 Categorizzazione delle sequenze	41
3.2.1 Accessi	41
3.2.2 Approccio	41
3.2.3 Assistenza	45
3.2.4 Soluzione	48
3.3 Analisi temporale	56

4 Conclusioni	69
4.1 Spunti di approfondimento	71
Bibliografia	72

Elenco delle tabelle

1.1	Descrizione degli eventi	17
2.1	Descrizione degli eventi (con colori)	22
2.2	Elenco delle finestre	23
2.3	Elenco delle categorie	28
3.1	Struttura dei turni di laboratorio	38

Elenco delle figure

1.1	Rappresentazione del framework di analisi	10
1.2	Interfaccia per gli studenti	11
1.3	Processo di formattazione del file	13
1.4	Rappresentazione del framework di analisi specializzato	19
2.1	Esempio di sequenza	22
2.2	Esempio di sequenza con frequenza	23
2.3	Finestra principale	24
2.4	Finestra di visualizzazione dei risultati	25
2.5	Finestra di ricerca di una sequenza	27
2.6	Finestra delle categorie	29
2.7	Finestra di join delle categorie	30
3.1	Numero di accessi per esercizio	32
3.2	Media di errori per esercizio	33
3.3	Richieste e interventi	33
3.4	Esercizi corretti	34
3.5	Percentuale di esercizi corretti	35
3.6	Numero di esercizi corretti per ciascuno studente	35
3.7	Numero di errori per esercizio	36
3.8	Percentuale di errori per esercizio	37
3.9	Percentuale di esercizi corretti rispetto agli accessi, divisi per turno	39
3.10	Interventi degli esercitatori, divisi per turno	40
3.11	Studenti che hanno fatto accesso all'esercizio precedente	45
3.12	Percentuale di errori rispetto al totale dei corretti	50
3.13	Studenti che hanno risolto esercizi al primo tentativo	51
3.14	Percentuale di studenti che hanno fatto 3 errori consecutivi su un esercizio	53
3.15	Percentuale di studenti che hanno fatto 6 errori consecutivi su un esercizio	55
3.16	Tempo impiegato a risolvere l'esercizio 1	59

3.17	Tempo impiegato a risolvere l'esercizio 2	59
3.18	Tempo impiegato a risolvere l'esercizio 3	60
3.19	Tempo impiegato a risolvere l'esercizio 4	61
3.20	Tempo impiegato a risolvere l'esercizio 5	61
3.21	Errore entro 3 minuti dopo l'accesso all'esercizio	63

Introduzione

Le attività di laboratorio sono un momento fondamentale dell'apprendimento di uno studente di ingegneria, perchè permettono di mettere in pratica le conoscenze trattate in teoria durante le lezioni frontali. Inoltre nonostante sul web abbondino materiale di supporto spesso le spiegazioni sono poco chiare o vengono fornite delle soluzioni ad hoc per una particolare situazione senza ulteriori precisazioni o una contestualizzazione che dia un background teorico.

Questo tipo di approccio mina fortemente le capacità di problem solving dello studente, che si trova ad essere un mero esecutore e che, di fronte a un nuovo problema, non è in grado di orientarsi in maniera autosufficiente.

I laboratori tenuti nel contesto di un corso universitario invece permettono di mettere gli studenti che cercano di risolvere gli esercizi a confronto diretto tra loro, migliorando le capacità di relazionarsi e lavorare in gruppo, un aspetto che non può mancare tra le competenze di un ingegnere.

In questa tesi è stato analizzato l'andamento del secondo laboratorio di un corso di Basi di Dati orientato all'apprendimento del linguaggio SQL da parte di studenti del corso di laurea di Ingegneria Gestionale.

Successivamente vengono utilizzati i principi indicati in [2] per descrivere la struttura dell'attività di laboratorio.

Quest'attività è stata sviluppata scegliendo esercizi mirati a sviluppare le competenze di base relative al linguaggio SQL; ogni esercizio introduce una differente richiesta o l'utilizzo di un diverso operatore in modo da affrontare gradualmente e con difficoltà crescente la composizione delle query. È stata definita una struttura del laboratorio che riprende gli argomenti trattati in aula, ma che al tempo stesso negli ultimi esercizi cerca di stimolare gli studenti con richieste più articolate e complesse. Lo studente deve riuscire a unire le nozioni imparate in aula con l'esperienza ottenuta dagli esercizi precedenti per procedere con successo nel laboratorio. Dalle successive analisi si potrà capire se la struttura del laboratorio abbia rispettato una crescita della complessità graduale o se invece prevedesse una curva di apprendimento troppo ripida, inficiando negativamente l'efficacia del lavoro.

L'esercitazione è costituita da 13 esercizi. I primi esercizi richiedevano l'utilizzo di semplici *SELECT* combinate con *JOIN* di due tabelle in modo da reperire le informazioni richieste. Successivamente venivano proposti esercizi che prevedevano l'uso di *nested queries*, argomento più complesso; seguivano esercizi che richiedevano l'uso dell'operatore *GROUP BY* e dell'operatore *HAVING*, che solitamente comportano difficoltà di comprensione. Si valuterà quindi se la necessità di questi operatori abbia costituito un problema per gli studenti nella soluzione dell'esercizio relativo.

In questo laboratorio è prevista la presenza di esercitatori, in parte docenti e in parte studenti che hanno superato con ottimi risultati il corso e hanno potuto quindi offrire un aiuto ai propri colleghi. Questi assistenti hanno il compito di fornire allo studente un supporto diretto che è complesso avere con un approccio orientato all'autoapprendimento. Uno degli aspetti da approfondire è quindi se e in che misura l'aiuto degli esercitatori sia stato efficace e se questo tipo di approccio abbia un risvolto positivo sull'apprendimento degli studenti.

Il numero di esercitatori presenti in laboratorio però è solitamente insufficiente per sopperire a tutte le richieste di aiuto, che spesso non vengono portate a termine impedendo agli studenti di proseguire. Questo può anche portare lo studente ad abbandonare il laboratorio prima della conclusione.

Da un punto di vista delle interazioni degli studenti inoltre è previsto che, oltre a richiedere aiuto agli esercitatori, gli studenti possano confrontarsi tra loro in modo da sviluppare la capacità di lavorare in gruppo.

Il laboratorio è stato organizzato in 6 turni in cui gli studenti sono stati suddivisi: due turni erano costituiti da un numero di studenti minore rispetto agli altri. Una parte del lavoro sarà orientata allo studio di questi turni presi singolarmente, per verificare in che modo il numero di studenti influisca sulle prestazioni di questi ultimi.

Per questo esperimento è stata utilizzata un'applicazione sviluppata in un precedente lavoro di tesi, attraverso la quale gli studenti hanno potuto eseguire le query richieste e vedere se il risultato coincidesse con la soluzione o meno. In questo modo è stato possibile registrare in maniera automatica tutte le azioni compiute dagli studenti, in modo da tenere traccia dei tentativi falliti e andati a buon fine. Questo ha permesso di raccogliere i dati necessari per l'analisi successiva.

L'applicazione dotava gli esercitatori di un'interfaccia che in tempo reale forniva per ogni postazione un indice di difficoltà riscontrata dagli studenti, basato sul numero di errori sull'esercizio corrente, sullo svolgimento degli esercizi precedenti e

sul tempo impiegato per l'esercizio. Questo ha permesso loro di scegliere di intervenire in presenza di situazioni considerate critiche in cui lo studente doveva essere indirizzato. D'altra parte l'applicazione permetteva anche agli studenti di fare una richiesta di aiuto in maniera automatica agli esercitatori; questo ha consentito di gestire in maniera più efficiente la distribuzione degli aiuti da parte degli assistenti, che hanno potuto aiutare prima le persone con più difficoltà.

I dati raccolti sono quindi stati utilizzati in questa tesi per analizzare l'andamento degli studenti e valutare gli aspetti evidenziati in precedenza. Per ottenere informazioni più specifiche è stato utilizzato cSpade, un algoritmo di sequence mining per estrarre le sequenze frequenti tra le sequenze di azioni compiute dagli studenti. In questo modo è stato possibile comprendere quali siano stati i comportamenti frequenti, evidenziando problemi comuni nell'approccio al laboratorio e dando uno spunto per indirizzare la didattica a una migliore assimilazione dei concetti fondamentali da parte degli studenti. Per mostrare ed elaborare le sequenze frequenti ottenute da cSpade infine è stata sviluppata un'applicazione desktop che permette di visualizzare le sequenze frequenti in maniera più intuitiva e permette di classificare le sequenze come appartenenti a categorie predefinite.

Questa tesi è divisa in 4 capitoli:

- Nel primo capitolo viene introdotto il framework di lavoro e descritto l'algoritmo cSpade, utilizzato per estrarre le sequenze frequenti.
- Nel secondo capitolo è spiegato il funzionamento del programma sviluppato al fine di visualizzare le sequenze degli studenti e ottenere informazioni più specifiche, utili all'analisi.
- Nel terzo capitolo vengono raccolti i risultati ottenuti tramite l'utilizzo di cSpade e il programma di analisi, raggruppando i risultati in maniera organica e traendo alcune valutazioni.
- Nel quarto capitolo vengono tratte le conclusioni basate sui risultati ottenuti. Inoltre sono indicate alcune possibili espansioni di questo lavoro.

Capitolo 1

Framework di analisi

In questa tesi è stato sviluppato un framework di analisi costituito da alcuni building blocks fondamentali per l'estrapolazione di informazioni interessanti da parte del docente.



Figura 1.1: Rappresentazione del framework di analisi

Come rappresentato in figura 1.1, il framework è costituito da 4 building blocks:

1. Raccolta dei dati: i dati relativi alle azioni effettuate dagli studenti devono essere registrati in maniera automatica.

2. Formattazione dei dati: i dati grezzi ottenuti dal punto precedente devono essere formattati in modo da poter essere utilizzati dall'algoritmo di sequence mining prescelto.
3. Enumerazione delle sequenze frequenti: si utilizza un algoritmo di sequence mining per estrarre le sequenze frequenti nel dataset.
4. Categorizzazione delle sequenze: vengono definite delle categorie in cui raggruppare le sequenze con caratteristiche affini in modo da reperire informazioni utili.

1.1 Raccolta dei dati

La raccolta dei dati non è stata oggetto di questa tesi; per completezza viene descritto l'ambiente di lavoro.

Per questo compito è stata sviluppata un'architettura client/server che permettesse di raccogliere tutti i tentativi da parte degli studenti.

Exercise 3

Go to exercise 2 Go to exercise 4

Find the names and initials of the deliverers that have serviced requests from all of the firms in table COMPANIES.

Input data HELP!!!

Query

Submit query

Query errors

Your result Expected result

NAME	INITIALS
Reiley	SP
Bishop	D
Miller	IF
Drone	M
Everett	R
Wise	GWG
Parmenter	J
Hope	PK
Baker	E
Newcastle	D

Figura 1.2: Interfaccia per gli studenti

L'applicazione client consiste in un'interfaccia web sviluppata in PHP e collegata al DBMS Oracle, il quale è eseguito localmente su ciascun computer. Tramite questa applicazione gli studenti affrontano ciascun esercizio sottomettendo la loro interpretazione della query richiesta e i risultati computati localmente dal

DBMS Oracle vengono mostrati nell'interfaccia utente. L'applicazione quindi mostra eventuali errori rilevati da Oracle, oltre alla comparazione tra risultato ottenuto dalla query immessa dallo studente e quello della query corretta. Inoltre è presente un bottone per richiedere aiuto ad un esercitatore.

Gli esercitatori invece sono provvisti di un'applicazione dedicata con la quale verificare l'andamento di ciascun gruppo di studenti tramite un'interfaccia grafica che assegna un colore ad ogni postazione in base a come procede nel laboratorio. In questo modo gli assistenti sono in grado di identificare studenti che sono in difficoltà e fornire un aiuto, anche se non richiesto esplicitamente dallo studente.

L'applicazione server è costituita da un'applicazione PHP connessa a un server MySQL e fornisce un insieme di API REST per ottenere i dati da ciascun client. Il server quindi colleziona tutte le informazioni necessarie, tra cui l'id dello studente, il timestamp, la tipologia di evento (ad esempio accesso a un esercizio, esercizio corretto).

1.2 Formattazione dei dati

I dati raccolti tramite l'applicazione citata precedentemente contenevano informazioni ridondanti rispetto a quelli utili per il passo successivo, ossia l'enumerazione delle sequenze frequenti tramite un algoritmo di sequence mining.

È evidente che questo passaggio è influenzato dalla scelta dell'algoritmo; per questo motivo in questa sezione viene definito il procedimento da adottare; i dettagli strettamente correlati all'algoritmo scelto saranno affrontati nella sezione successiva.

L'applicazione per la raccolta dei dati ha raggruppato i dati separandoli in tre file:

1. file relativo alle query sottomesse dagli studenti
2. file relativo agli accessi ai singoli esercizi da parte di ciascuno studente
3. file relativo agli interventi da parte degli esercitatori

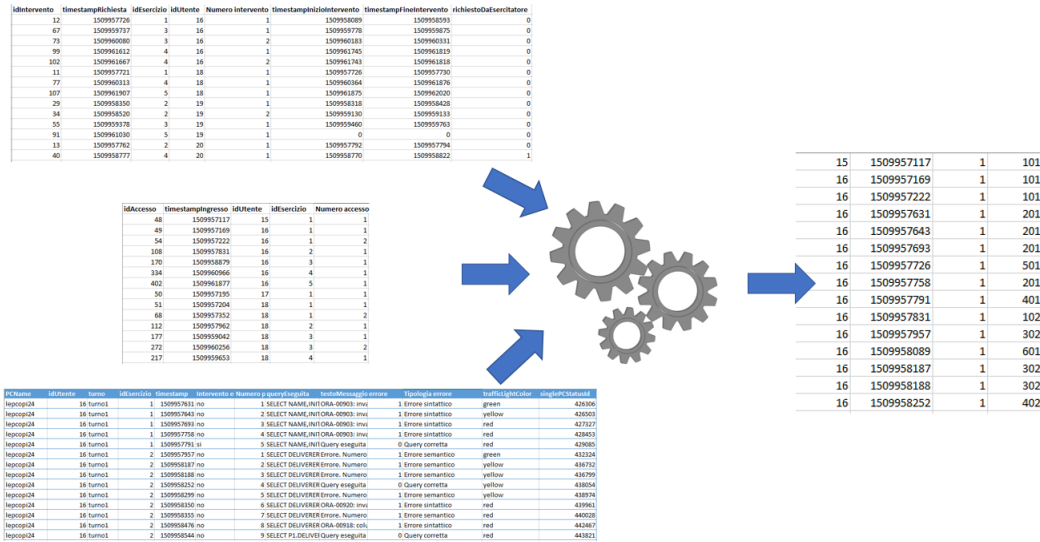


Figura 1.3: Processo di formattazione del file

Ciascun file riporta molte informazioni interessanti che però non sono utili per l'analisi eseguita in questo lavoro. A titolo d'esempio, nel file relativo alle query sottomesse sono stati riportati il testo della query eseguita e l'output di Oracle che non sono indicativi per l'estrazione delle sequenze frequenti o per la definizione degli eventi. Inoltre la descrizione dell'evento è stata riportata tramite una descrizione testuale, mentre l'algoritmo utilizzato richiede una descrizione dell'evento in formato numerico.

I tre file sono stati uniti tramite una semplice applicazione C# (seguendo il processo indicato in figura 1.3) che da ciascun file ha estrapolato per ciascun evento solo 3 informazioni:

- identificativo dello studente
- timestamp in cui è avvenuto l'evento
- descrizione dell'evento

Infine le tuple relative a ciascun evento sono state ordinate per identificativo dello studente e, in caso di eventi relativi allo stesso studente, per timestamp crescenti.

1.3 Enumerazione delle sequenze frequenti

L'estrazione di sequenze di eventi è uno strumento utilizzato in molti ambiti, come ad esempio lo studio delle sequenze del DNA, l'analisi delle pagine web visitate per fini pubblicitari, lo studio dei pattern di acquisto da parte dei consumatori ed altro.

La definizione delle sequenze frequenti è un task molto importante che implica uno spazio di ricerca di notevoli dimensioni: l'efficienza dell'algoritmo utilizzato risulta fondamentale per permettere di ottenere risultati in tempi ragionevoli. A volte lo spazio di ricerca risulta talmente ampio da rendere intrattabile il problema.

Indipendentemente dall'algoritmo utilizzato, sono necessarie alcune definizioni per poter formulare il problema (tratte da [4]).

Sia dato un insieme di m item distinti, chiamato I . Un evento o itemset è un insieme non ordinato di item.

Una sequenza è una lista ordinata di eventi; ha una lunghezza che è data dal numero di eventi. Una sequenza di k elementi è detta k -sequenza.

Ogni sequenza all'interno di un database ha un identificativo chiamato *sid*, mentre ogni evento ha un identificativo chiamato *eid*. Di solito per quest'ultimo si utilizza il timestamp (o istante temporale) in cui avviene l'evento.

Un evento inoltre è legato alla sequenza di appartenenza tramite il *sid*. La formalizzazione di un singolo evento risulta quindi la seguente:

`sid eid #item itemset`

dove

- *sid*: identificativo della sequenza del soggetto che esegue l'itemset
- *eid*: timestamp in cui avvengono gli eventi dell'itemset
- *#item*: numero di item che costituiscono l'itemset e che avvengono in quell'istante
- *itemset*: elenco degli eventi che avvengono in quell'istante

Per rendere più chiaro questo insieme di definizioni si può considerare il seguente esempio. Si consideri come item l'acquisto di un oggetto su un negozio online; si considerino ad esempio l'acquisto di una penna, una matita e una gomma (d'ora in avanti indicati rispettivamente con P , M e G). P , M e G rappresentano degli item. Si supponga che una persona abbia acquistato P ed M nello stesso ordine (al tempo $t1$) e in un ordine successivo al tempo $t2$ abbia comprato G . La sequenza corrispondente è la seguente:

$$P \ M \rightarrow G$$

PM è quindi un evento (costituito da due item), mentre l'insieme ordinato di eventi costituisce la sequenza di eventi relativi alla persona considerata.

Una sequenza s è una sottosequenza di un'altra sequenza t se esiste una funzione che preservi l'ordine degli eventi e garantisca che gli itemset di s siano contenuti negli itemset di t .

Ad esempio, la sequenza

$$P \rightarrow G$$

è una sottosequenza della sequenza indicata in precedenza.

Il supporto o frequenza di una sequenza è il numero totale di sequenze all'interno del database che contiene questa sequenza. L'utente può inserire una soglia, chiamata minimum support (minsup), compresa tra 0 e 1 che indica la percentuale di sequenze in cui una particolare sequenza deve comparire per essere considerata frequente.

Sia dato un database con m sequenze. Una sequenza è frequente se compare come sottosequenza più di $\text{minsup} * m$ volte. Preso ad esempio un database con 100 sequenze e fissato il valore di minsup pari a 0.2, una sequenza è frequente se ha un supporto minimo pari a 20: almeno 20 sequenze devono contenere una sottosequenza perché questa sia frequente.

Valori molto elevati di minsup permettono di comprendere quale siano i comportamenti più frequenti da parte degli studenti.

Si deve inoltre sottolineare un'importante considerazione: se la sequenza s è una sottosequenza della sequenza t e t è una sequenza frequente, allora anche s è una sequenza frequente e il supporto di s sarà maggiore o uguale al supporto di t .

Si prendano in considerazione le sequenze introdotte precedentemente e si supponga che per la prima si abbia il seguente supporto:

$$P \ M \rightarrow G \ - \ 50$$

Questo significa che 50 clienti del sito hanno acquistato una penna e una matita insieme e successivamente una gomma. È evidente che un numero maggiore di persone avrà acquistato una penna e successivamente una gomma, dato che sono incluse sia le persone che hanno acquistato la matita che quelle che non l'hanno fatto. Ci si può aspettare ad esempio che il supporto sia il seguente:

$$P \rightarrow G \ - \ 60$$

1.3.1 Scelta dell'algoritmo

Gli algoritmi presi in considerazione per l'estrazione delle sequenze frequenti sono stati spade, cSpade e PRISM. Le considerazioni riguardanti le prestazioni sono state tratte da [4], [6] e [8]

cSpade non è altro che un'evoluzione di spade, quindi a livello di performance sono equivalenti; nel caso in cui però nel primo vengano fissati i vincoli temporali lo spazio di ricerca viene notevolmente ridotto, diminuendo notevolmente i tempi di esecuzione.

PRISM ha migliori prestazioni sia a livello di tempi di esecuzione che a livello di memoria impiegata; rispetto a spade impiega circa 4 volte meno tempo e richiede da 3 a 5 volte minore memoria. Al tempo stesso però non permette di aggiungere vincoli sulle tempistiche tra gli eventi appartenenti a una sequenza; dato che questo è un requisito molto importante ai fini dell'analisi, è stato scelto l'algoritmo cSpade per estrarre le sequenze frequenti.

1.3.2 cSpade

Aspetti teorici

cSpade è un algoritmo che ha lo scopo di generare sequenze frequenti utilizzando un insieme di vincoli [6]. Questi vincoli permettono di ridurre lo spazio di ricerca e migliorare le prestazioni, oltre ad ottenere informazioni più specifiche relative alla caratteristica delle sequenze.

I vincoli introdotti da cSpade riguardano la lunghezza delle sequenze, intervalli di tempo minimi tra elementi consecutivi che costituiscono una sequenza, intervalli di tempo massimi, finestre temporali dell'intera sequenza.

Questi vincoli permettono inoltre di ottenere informazioni interessanti che, nel caso specifico, permettono di delineare con maggiore precisione come gli studenti abbiano proceduto nel laboratorio. Ad esempio permettono di evidenziare eventuali difficoltà degli studenti tramite l'accesso ad esercizi consecutivi molto distanziati temporalmente o, al contrario, evidenziare l'eccellenza tramite la soluzione di un esercizio poco tempo dopo aver fatto l'accesso.

Applicazione

Nel caso di studio di questa tesi, un evento è un'azione compiuta da uno studente (o da un esercitatore). Questa azione è caratterizzata da:

- un id che identifica lo studente
- il timestamp che indica l'istante in cui l'evento è avvenuto

Dato che in ogni istante uno studente può compiere una sola azione, ogni itemset è costituito da un singolo evento, che corrisponde all'azione stessa.

In questo contesto, un evento può avere la seguente forma:

18 1529478218 1 2

dove il significato è il seguente: lo studente con sid 18 nell'istante di tempo 1529478218 ha eseguito l'evento codificato con il numero 2. È stato quindi necessario formalizzare il problema definendo, nel caso di studio di questa tesi, la forma di un evento. In particolare ci si è concentrati su due aspetti: l'azione compiuta dallo studente e l'esercizio su cui è stata compiuta.

Poiché l'algoritmo richiede per la rappresentazione di un evento l'uso di numeri è stata utilizzata la seguente codifica per un evento:

xyy

dove x rappresenta un codice assegnato a ciascuna azione, mentre yy rappresenta il numero dell'esercizio su due cifre a cui l'azione si riferisce.

Codice	Evento	Descrizione
1	Accesso	accesso a un esercizio
2	Errore	errore su un esercizio
3	Corretto	esercizio corretto
4	Richiesta	richiesta di aiuto a un assistente
5	Aiuto	aiuto da parte di un assistente

Tabella 1.1: Descrizione degli eventi

Nella tabella 1.1 vengono riportati i codici relativi a ciascuna azione e la sua descrizione.

Un esempio di sequenza può essere la seguente:

101 -> 301 -> 102 -> 402

Dove si può desumere che lo studente ha fatto accesso al primo esercizio, lo ha risolto correttamente, ha quindi fatto accesso al secondo esercizio e infine ha richiesto aiuto ad un assistente per tale esercizio.

Una semplice sequenza può essere costituita dall'accesso ad uno specifico esercizio e la sua soluzione. La sequenza è costituita da due eventi: ha quindi lunghezza pari a 2. Infine la frequenza di una sequenza è indicata dopo due trattini orizzontali, come segue:

301 -> 102 - - 189

La sequenza indicata ad esempio ha frequenza 189.

Per utilizzare cSpade è stato sviluppato un script bash che automaticamente applica l'algoritmo sui turni di laboratorio distinti e sull'intero laboratorio, in modo da considerare l'influenza del singolo turno sui risultati finali. Lo script si occupa di salvare i risultati su file; questi vengono presi come input dal programma di analisi per visualizzare le sequenze.

1.4 Categorizzazione delle sequenze

Sono state create delle categorie per catalogare le sequenze in modo che sia più semplice ricavare informazioni relative a un particolare contesto.

Nel caso di studio di questa tesi, in cui si è scelto di utilizzare l'algoritmo cSpade per i motivi visti precedentemente, è necessario estendere il framework per aggiungere le informazioni relative ai vincoli temporali:

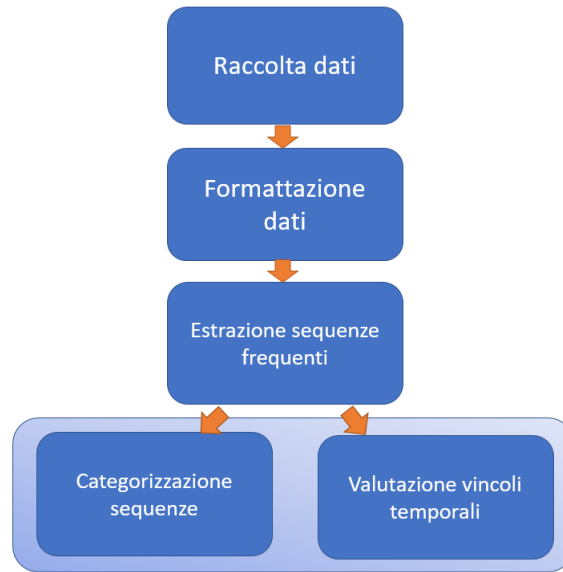


Figura 1.4: Rappresentazione del framework di analisi specializzato

Si può quindi vedere che è stato necessario aggiungere l'analisi dei vincoli temporali.

Questi devono essere trattati separatamente in quanto dipendono direttamente dai vincoli imposti a cSpade dalla particolare esecuzione e possono essere visti come un'estensione delle categorie viste in precedenza.

La scelta specifica delle categorie dipende dalla tipologia di esercizi da affrontare nel laboratorio e può essere specializzata per particolari contesti. L'appartenenza di una sequenza a una categoria dipende dalla tipologia di eventi che la costituiscono e dall'ordine relativo in cui avvengono questi eventi.

Capitolo 2

Funzionamento del programma di analisi

L'utilizzo dell'algoritmo cSpade è stato fondamentale per poter analizzare i dati e ottenere informazioni riguardo all'andamento generale del laboratorio in questione.

Il limite dell'algoritmo è che i risultati testuali sono di difficile interpretazione, soprattutto nel caso in cui sia necessaria una codifica articolata.

Per questo motivo è stata sviluppata un'applicazione che prende in input i risultati dell'algoritmo cSpade e permette di visualizzarli graficamente in modo da essere più facilmente interpretabili.

Come indicato in precedenza, cSpade permette di estrarre le sequenze frequenti; non tutte però portano informazioni utili.

A titolo di esempio, si valuti la seguente sequenza:

301 -> 102 - - 189

Questa sequenza non fornisce informazioni interessanti; infatti è evidente che tutti coloro che hanno risolto correttamente l'esercizio 1 hanno poi fatto accesso all'esercizio 2. È dunque fondamentale poter effettuare una ricerca all'interno dei dati. Il limite del risultato testuale è che la ricerca manuale di una particolare sequenza può risultare complessa dato il numero spesso elevato di sequenze contenute in un file. Inoltre la semplice ricerca implementata da qualunque editor di testo è complicata dalla forma complessa che assume la sequenza in forma testuale. Per questo motivo è stata implementata una ricerca che permetta di inserire l'evento ricercato semplicemente cliccando su uno degli elementi contenuti all'interno di un pannello di ricerca. Questo ha permesso di rendere la ricerca molto più rapida ed intuitiva.

Infine per migliorare ulteriormente l'interpretabilità dei risultati raccolti sono state aggiunte delle categorie che raggruppano le sequenze estratte in modo da ottenere più rapidamente alcune informazioni utili.

2.1 Visualizzazione delle sequenze

Come indicato in precedenza, una sequenza è una lista di eventi, ordinata secondo il timestamp in cui l'evento è avvenuto.

Dato che la codifica utilizzata è poco intuitiva, si è scelto di rappresentare un evento indicando solo il numero dell'esercizio e utilizzando un colore di sfondo per rappresentare la tipologia di evento.






Codice	Evento	Colore	Descrizione
1	Accesso		accesso a un esercizio
2	Errore		errore su un esercizio
3	Corretto		esercizio corretto
4	Richiesta		richiesta di aiuto a un assistente
5	Aiuto		aiuto da parte di un assistente

Tabella 2.1: Descrizione degli eventi (con colori)

Seguono due esempi di utilizzo.

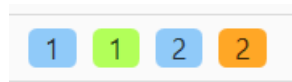


Figura 2.1: Esempio di sequenza

Nella figura 2.1 si può individuare un accesso all'esercizio 1 seguito dalla sua soluzione corretta, poi un accesso all'esercizio 2 per cui lo studente ha richiesto aiuto ad un esercitatore di laboratorio.

Un altro aspetto importante tra le caratteristiche di una sequenza è la frequenza, che indica il numero di studenti che hanno effettuato tale sequenza di eventi.

Nell'applicazione vengono riportate per ciascuna sequenza la relativa frequenza e la percentuale di studenti che ha compiuto tale sequenza. La percentuale è stata introdotta in quanto fornisce un'indicazione più immediata dell'importanza della sequenza stessa. Le sequenze vengono ordinate per percentuale decrescente: sono mostrate prima le sequenze più frequenti, poi quelle meno frequenti.

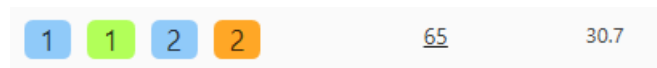


Figura 2.2: Esempio di sequenza con frequenza

Nella figura 2.2 sono riportate, oltre alla sequenza, la frequenza e la percentuale di studenti rispetto al totale. Si può quindi facilmente desumere che circa il 30% ha eseguito questa sequenza di azioni.

2.2 Finestre

La seguente trattazione si occupa di analizzare ogni finestra dell'applicazione, la sua funzionalità e indica la navigazione tra una finestra e l'altra.

Viene inoltre riportata una tabella delle finestre con una breve descrizione per avere un quadro d'insieme:

Finestra	Descrizione
Main window	permette di inserire i risultati da visualizzare
Results window	visualizza i risultati
Search Window	permette di fare una ricerca tra le sequenze frequenti
Categories Window	visualizza le sequenze che appartengono alla categoria selezionata
Join Categories	visualizza le sequenze che appartengono alle categorie scelte dall'utente

Tabella 2.2: Elenco delle finestre

2.2.1 Main Window

La finestra principale permette all'utente di inserire i risultati ottenuti da cSpade all'interno dell'applicazione per visualizzarli.



Figura 2.3: Finestra principale

L'utente deve trascinare i file che vuole visualizzare sulla finestra, quindi cliccare il pulsante *Show* per mostrare i risultati in una versione più interpretabile all'interno della finestra **Result Window**.

Nella parte centrale della finestra vengono elencati i file che l'utente ha rilasciato sulla finestra.

Nel menù *Utilities* è possibile visualizzare il testo degli esercizi, una loro possibile implementazione e le sequenze di ciascun utente selezionando l'id desiderato.

2.2.2 Results Window

In questa finestra l'utente può visualizzare i risultati ottenuti da cSpade in maniera semplice ed intuitiva.

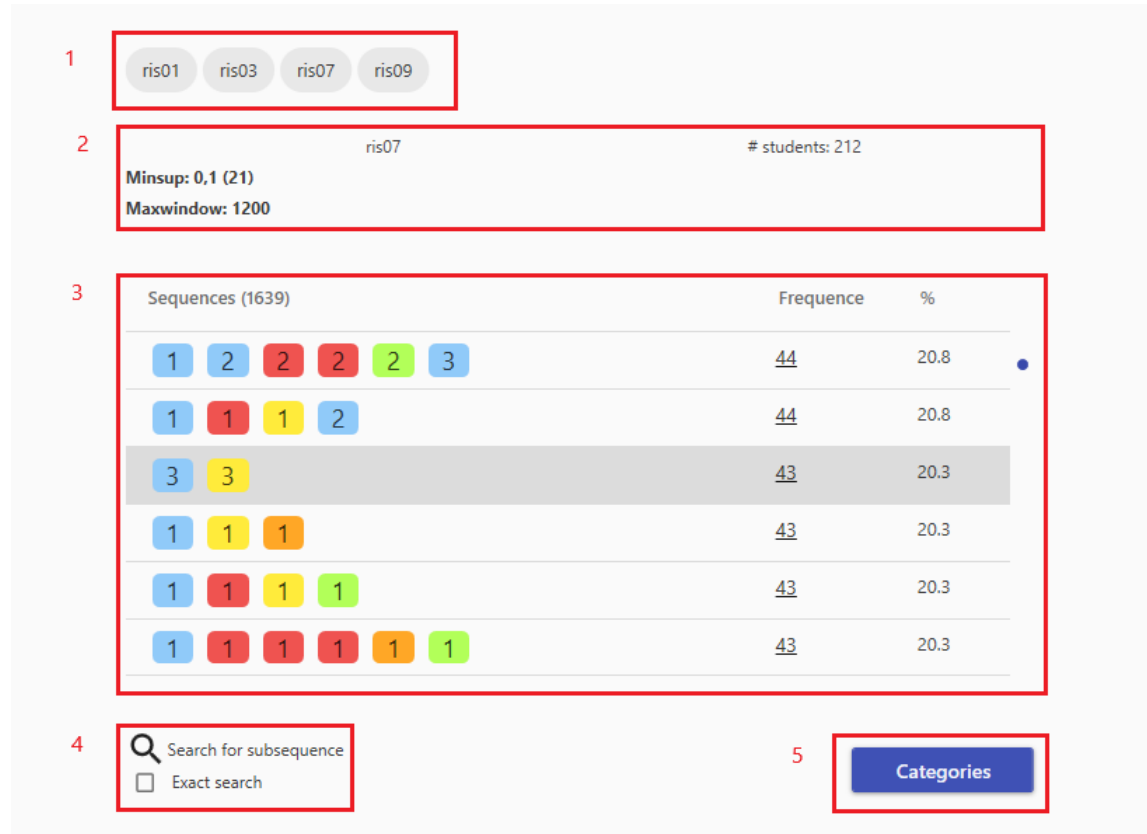


Figura 2.4: Finestra di visualizzazione dei risultati

In figura 2.4 viene riportato l'aspetto della finestra. Si possono individuare 5 sezioni, evidenziate in figura dal contorno rosso:

1. sezione in cui vengono elencati i file da visualizzare. Cliccando sul nome di un file viene visualizzato il suo contenuto.
2. elenco dei parametri del file. Vengono riportati (se presenti come parametri nel file):
 - nome del file selezionato
 - numero di studenti in totale presenti nel file
 - valore del parametro *minsup*

- valore del parametro *mingap*
 - valore del parametro *maxgap*
 - valore del parametro *maxwindow*
3. sezione in cui vengono visualizzate le sequenze contenute nel file. Sono elencate le sequenze frequenti che cSpade ha estratto utilizzando i parametri elencati nella sezione 2. Per ogni sequenza è indicata la frequenza e la percentuale rispetto al totale di studenti. Cliccando la frequenza inoltre è possibile visualizzare l'elenco degli id degli studenti che hanno compiuto quella sequenza.
 4. sezione che permette di ricercare una sottosequenza. Dato che il numero di sequenze contenute in un file può raggiungere valori molto elevati (anche centinaia di migliaia) questa sezione risulta molto utile. Permette di aprire la finestra **Search Window** e ricercare tutte le sequenze di cui la sequenza inserita dall'utente risulta essere una sottosequenza. Aggiungendo l'opzione *Exact search* si cerca solamente la sequenza inserita.
 5. questo bottone permette di aprire la finestra **Categories Window** che permette di utilizzare le categorie per filtrare le sequenze.

L'utente quindi dopo aver aperto questa finestra seleziona il file da visualizzare; quest'operazione gli permette di mostrare le sequenze frequenti. L'utente può successivamente ricercare una sequenza contenuta all'interno di quelle generate o approfondire l'analisi utilizzando le categorie.

2.2.3 Search Window

Questa finestra serve per permettere all'utente di selezionare la sequenza di eventi da ricercare all'interno di una finestra di visualizzazione di risultati.



Figura 2.5: Finestra di ricerca di una sequenza

In figura 2.5 si può vedere l'aspetto della finestra. Sono presenti 3 sezioni:

1. sequenza che l'utente vuole cercare. L'utente può cliccare sul bottone contenente la X per cancellare la sequenza e inserirla nuovamente.
2. Lista degli eventi disponibili nelle sequenze del file analizzato attualmente. L'utente può aggiungere un evento alla sequenza ricercata cliccando sull'evento corrispondente in questa sezione.
3. Il bottone di ricerca apre una nuova finestra con tutte le sequenze che contengono la sottosequenza introdotta dall'utente nella sezione 1. Si noti che se l'opzione *Exact search* in **Result Window** non è attiva, possono esserci degli ulteriori eventi all'interno di una sequenza.

Questa finestra è richiamata da **Results Window**, **Categories Window**.

2.2.4 Categories Window

Macro	Categoria	Descrizione
Accesso	cat1	Raggruppa gli studenti per numero di esercizi accede a pochi esercizi
	cat2	accede a molti esercizi
Approccio	cat1	Analizza come gli studenti si avvicinano agli esercizi non va all'esercizio successivo finchè non termina quello corrente
	cat2	passa all'esercizio successivo senza aver terminato quello corrente (almeno due volte)
	cat3	procede senza tornare a un esercizio precedente
	cat4	torna a un esercizio precedente almeno n volte
Assistenza	cat1	Si focalizza sull'aiuto prestato dagli esercitatori ha fatto più di n richieste
	cat2	non ha mai fatto richiesta d'aiuto
	cat3	ha richiesto aiuto per tutti gli esercizi a cui ha fatto accesso
	cat4	ha richiesto aiuto due volte per lo stesso esercizio
	cat5	dopo l'aiuto ha risolto immediatamente l'esercizio (non ci sono eventi tra l'aiuto e la soluzione corretta)
	cat6	dopo l'aiuto ha commesso immediatamente un errore su quell'esercizio (non ci sono eventi tra l'aiuto e l'errore)
	cat7	l'aiuto ha permesso di risolvere l'esercizio
Soluzione	cat1	Soluzioni da parte degli studenti ha risolto correttamente i primi n esercizi
	cat2	ha risolto tutti gli esercizi a cui ha fatto accesso al primo tentativo
	cat3	ha risolto almeno due esercizi al primo tentativo
	cat4	ha fatto almeno un errore su tutti gli esercizi a cui ha fatto accesso
	cat5	ha fatto almeno n errori su tutti gli esercizi a cui ha fatto accesso
	cat6	ha fatto almeno 3 errori su un esercizio prima di risolverlo
	cat7	ha fatto almeno n errori

Tabella 2.3: Elenco delle categorie

Questa finestra permette di visualizzare le sequenze che hanno determinate caratteristiche, in modo da ottenere più agevolmente informazioni.

In tabella 2.3 è riportato un elenco delle categorie con una breve descrizione. Le categorie sono state raggruppate in quattro macrocategorie: Accesso, Approccio, Assistenza e Soluzione. La Tabella 2.3 contiene anche una breve descrizione per ciascuna macrocategoria.

Per ogni categoria in cui è indicato il parametro n si intende che l'utente deve inserire il valore all'interno della finestra che appare dopo aver selezionato la categoria.



Figura 2.6: Finestra delle categorie

In figura 2.6 si vede l'aspetto della finestra. Sono presenti 6 sezioni:

1. Nome del file selezionato attualmente.
2. Elenco dei file che l'utente ha inserito. Questi file sono nuovamente i file di output di cSpade.
3. Pannello in cui l'utente può selezionare la categoria che sta cercando. Nella parte superiore c'è il path della categoria corrente, che può anche essere utilizzato per navigare più velocemente tra le categorie.
4. Bottone per unire diverse categorie, ossia estrarre le sequenze che fanno match su tutte le categorie selezionate.

5. Lista delle sequenze che rispondono ai vincoli imposti dalla categoria.
6. Bottone per la ricerca, come in Result Window.

Inizialmente l'utente deve selezionare il file desiderato nella sezione 2; dopo deve selezionare una categoria dalla sezione 3.

Quando viene raggiunta una categoria terminale verrà mostrato nella sezione 4 l'elenco di tutte le sequenze che appartengono alla categoria selezionata.

Ad esempio, il path *categories/Approach/cat1* corrisponde a una categoria terminale e viene mostrata la lista di sequenze per quella categoria.

2.2.5 Join Categories

Questa finestra viene raggiunta a partire dalla finestra *Categories Window*.

Select categories that you want to join.

	Name	Description
<input type="checkbox"/>	Access/cat1	Student who access to few exercises (less or equal to *input*).
<input type="checkbox"/>	Access/cat2	Students who access to many exercises (more or equal to *input*).
<input type="checkbox"/>	Approach/cat1	Student does not go to next exercise until he solves the current one.
<input type="checkbox"/>	Approach/cat2	Student go to the next exercise even if he does not solve the current one (at least twice
<input type="checkbox"/>	Approach/cat3	Students proceedes without going back to previous exercises.

Join

Figura 2.7: Finestra di join delle categorie

L'utente deve selezionare l'elenco di categorie (2 o più) che vuole unire cliccando sul checkbox corrispondente alla categoria. Per ogni categoria è riportata la descrizione in modo da avere un'immediato riscontro. Per procedere si deve cliccare sul pulsante *Join*.

Nella finestra che viene aperta si trovano tutte le sequenze che compaiono in tutte le categorie selezionate dall'utente.

Capitolo 3

Analisi dei risultati ottenuti

Le analisi sono state condotte tenendo conto degli accessi a ciascun esercizio e delle richieste di aiuto.

In quest'analisi sono presenti 212 studenti.

Nella prima sezione ci si focalizza sulle informazioni generali per avere un quadro d'insieme; si fa anche riferimento ai turni di laboratorio, mettendo a confronto alcuni aspetti.

Nella seconda sezione si vanno a raccogliere i risultati secondo le categorie, evidenziando alcuni aspetti più specifici.

Nella terza sezione infine si studiano alcuni parametri temporali per andare cercare di capire l'effettivo andamento degli studenti e come si sono approcciati al laboratorio.

3.1 Analisi esplorativa

La prima parte dell'analisi ha carattere esplorativo. Vengono quindi utilizzate informazioni che non coinvolgono sequenze, ma che hanno grande importanza per capire in generale quale sia stato l'andamento di questo laboratorio. In figura 3.1

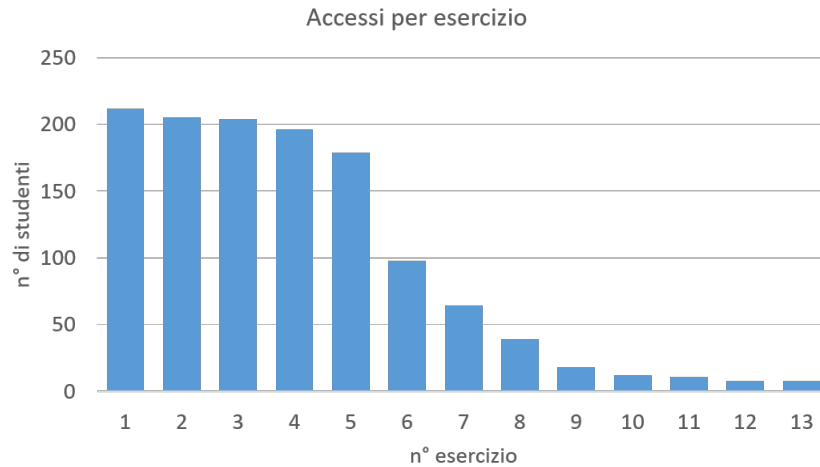
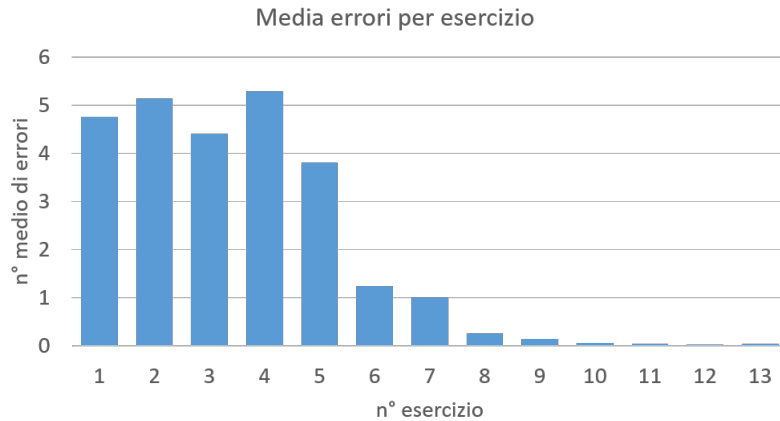


Figura 3.1: Numero di accessi per esercizio

è rappresentato il numero di studenti che hanno fatto accesso all'esercizio.

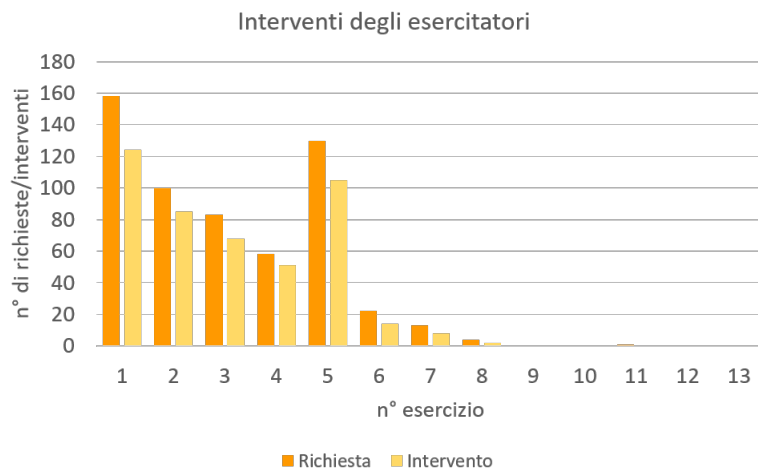
Si può notare che la maggior parte degli studenti ha fatto accesso fino al quinto esercizio. Questo indica che gran parte degli studenti ha avuto problemi sui primi esercizi.

Pochi studenti hanno fatto accesso ad un esercizio successivo al numero 8. In particolare si nota che pochissimi studenti sono arrivati al termine del laboratorio.

**Figura 3.2:** Media di errori per esercizio

In figura 3.2 viene riportato il numero di errori compiuti in media dagli studenti per ogni esercizio. Si può vedere che c'è un grande numero di errori sui primi 5 esercizi, aspetto che rafforza l'ipotesi che gli studenti abbiano avuto difficoltà all'inizio del laboratorio. C'è da sottolineare però che la media è un dato indicativo che non avrà diretta correlazione con le sequenze che risultano frequenti.

Gli esercizi dal numero 9 in avanti presentano una media di errori prossima a zero: questo è sicuramente influenzato dal fatto che poche persone hanno fatto accesso ad essi. Allo stesso tempo però queste persone in media non hanno fatto errori, ad indicare che non hanno avuto problemi per risolvere gli esercizi.

**Figura 3.3:** Richieste di aiuto e interventi da parte degli esercitatori

In figura 3.3 si mettono a confronto le richieste da parte degli studenti rispetto agli interventi da parte degli esercitatori. Si può vedere che le richieste hanno avuto un buon tasso di risposta. Inoltre le richieste sono concentrate principalmente sui primi 5 esercizi, in particolare sugli esercizi 1 e 5.

Per diversi esercizi, in particolare gli esercizi 9, 10, 12, 13, non risulta alcuna richiesta di aiuto. Questo è da mettere in relazione al basso numero di accessi a questi esercizi.

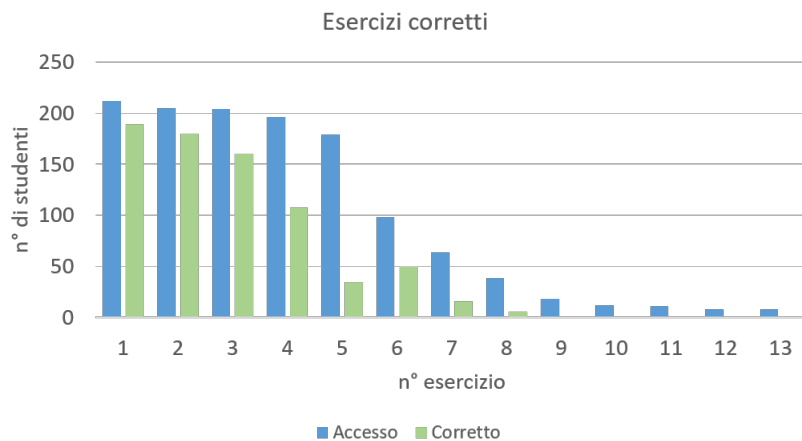


Figura 3.4: Esercizi corretti

In figura 3.4 si mettono a confronto l'accesso dell'esercizio con lo svolgimento corretto da parte dei singoli studenti. Si può notare che l'esercizio che ha dato più problematiche sia il 5, dato che in proporzione poche persone che vi hanno fatto accesso sono riuscite a risolverlo correttamente. Al contrario, sui primi esercizi nonostante il numero elevato di errori la maggior parte degli studenti è riuscita a portare a termine positivamente l'esercizio. Molto probabilmente gli interventi degli esercitatori hanno avuto un ruolo importante.

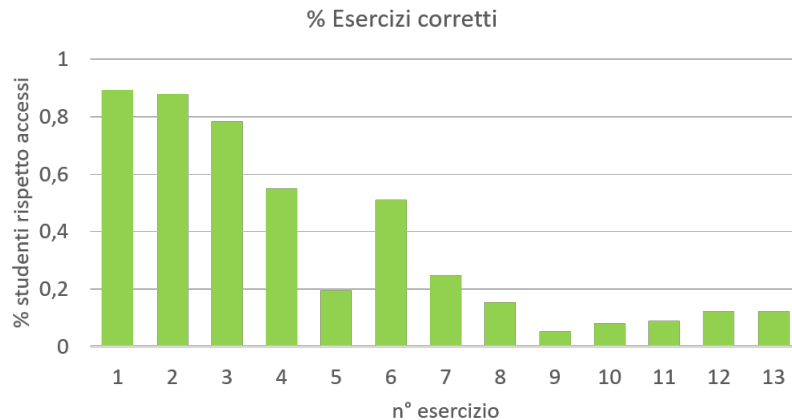


Figura 3.5: Percentuale di esercizi corretti

In figura 3.5 si rappresenta la percentuale di studenti che hanno risolto correttamente l'esercizio rispetto al totale di studenti che vi hanno fatto accesso. Per i primi 3 esercizi la percentuale è molto alta, superiore al 75%.

Un dato rilevante è relativo all'esercizio 5, per cui solo il 20% degli studenti che vi hanno fatto accesso lo hanno risolto.

Negli esercizi successivi all'8 poi la percentuale scende sotto il 15%, segno che i pochi che hanno fatto accesso all'esercizio non sono stati in grado di risolverlo: si può ipotizzare che le richieste fossero troppo complesse per il primo laboratorio.

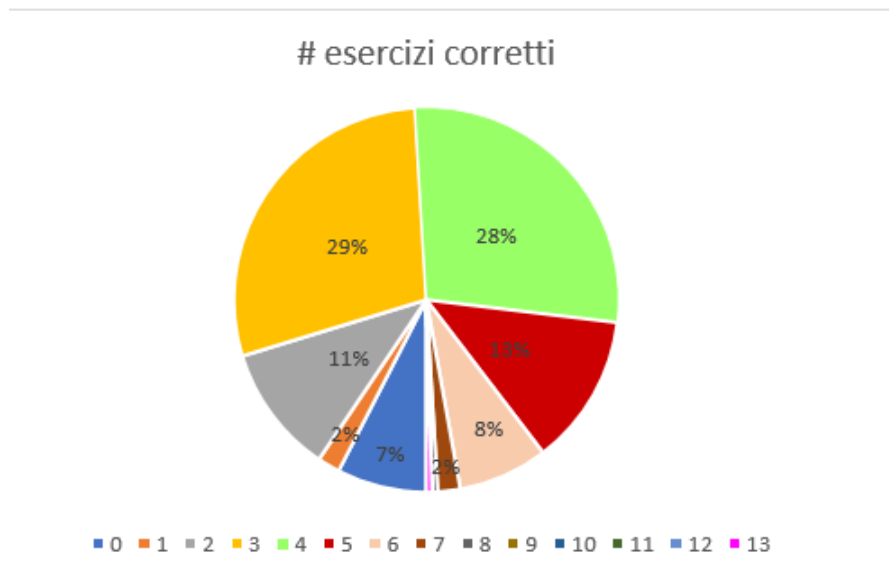


Figura 3.6: Numero di esercizi corretti per ciascuno studente

In figura 3.6 viene indicato il numero di esercizi svolti correttamente da parte di ciascuno studente. Si può vedere che il 57% degli studenti ha svolto tra 3 e 4 esercizi. La maggior parte degli studenti ha svolto un basso numero di esercizi. Il 7% degli studenti non ha svolto nessun esercizio corretto (o ha avuto grandissime difficoltà o non si è effettivamente impegnato dato che l'aiuto degli esercitatori è stato puntuale quando richiesto). Il 21% degli studenti ha svolto correttamente da 5 a 6 esercizi, quindi ci sono persone che hanno trovato meno difficoltà; in ogni caso è un numero molto basso rispetto al numero totale di esercizi. Solo uno studente è riuscito a svolgere correttamente tutti gli esercizi.

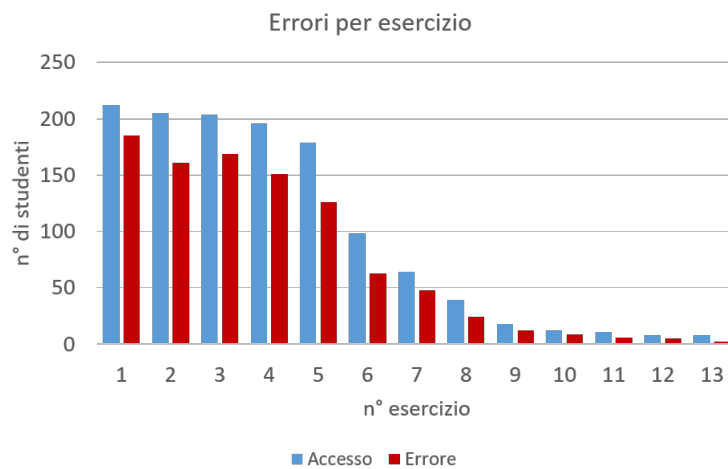


Figura 3.7: Numero di errori per esercizio

In figura 3.7 è riportato il numero di studenti che ha commesso almeno un errore sull'esercizio. Si può notare che il numero di errori è molto elevato nei primi esercizi, mentre diminuisce sui successivi. In realtà quest'informazione non tiene conto del fatto che è diminuito il numero di accessi.

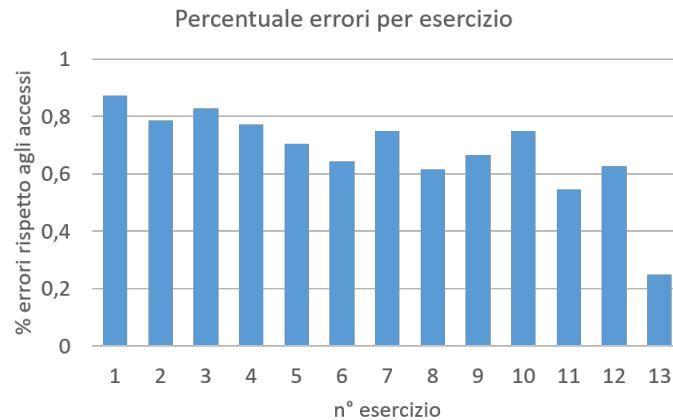


Figura 3.8: Percentuale di errori per esercizio

In figura 3.8 invece viene riportata la percentuale di studenti che hanno commesso almeno un errore rispetto al numero di accessi. Si vede che la situazione è abbastanza bilanciata nei diversi esercizi.

Spiccano tra tutti gli esercizi 1 e 3, in cui più dell'80% degli studenti ha commesso almeno un errore.

Dall'altra parte invece l'esercizio 13 è l'unico con una percentuale minore del 50%, ossia circa 25% di studenti che hanno fatto accesso. Evidentemente solo i più bravi hanno raggiunto questo esercizio.

Nella parte successiva ci si concentra sull'analisi dei distinti turni di laboratorio.

In particolare si vuole evidenziare la capacità di risolvere gli esercizi, l'efficacia dell'aiuto degli esercitatori in funzione del numero di studenti all'interno del turno. Si vuole verificare se effettivamente un turno con meno studenti implichi un migliore rendimento da parte degli studenti, in quanto possono essere più facilmente seguiti da parte degli esercitatori.

Sono presenti 6 turni e sono organizzati come segue:

Turno	Studenti	Esercitatori
1	35	4
2	46	3
3	27	5
4	39	4
5	26	4
6	39	3

Tabella 3.1: Struttura dei turni di laboratorio

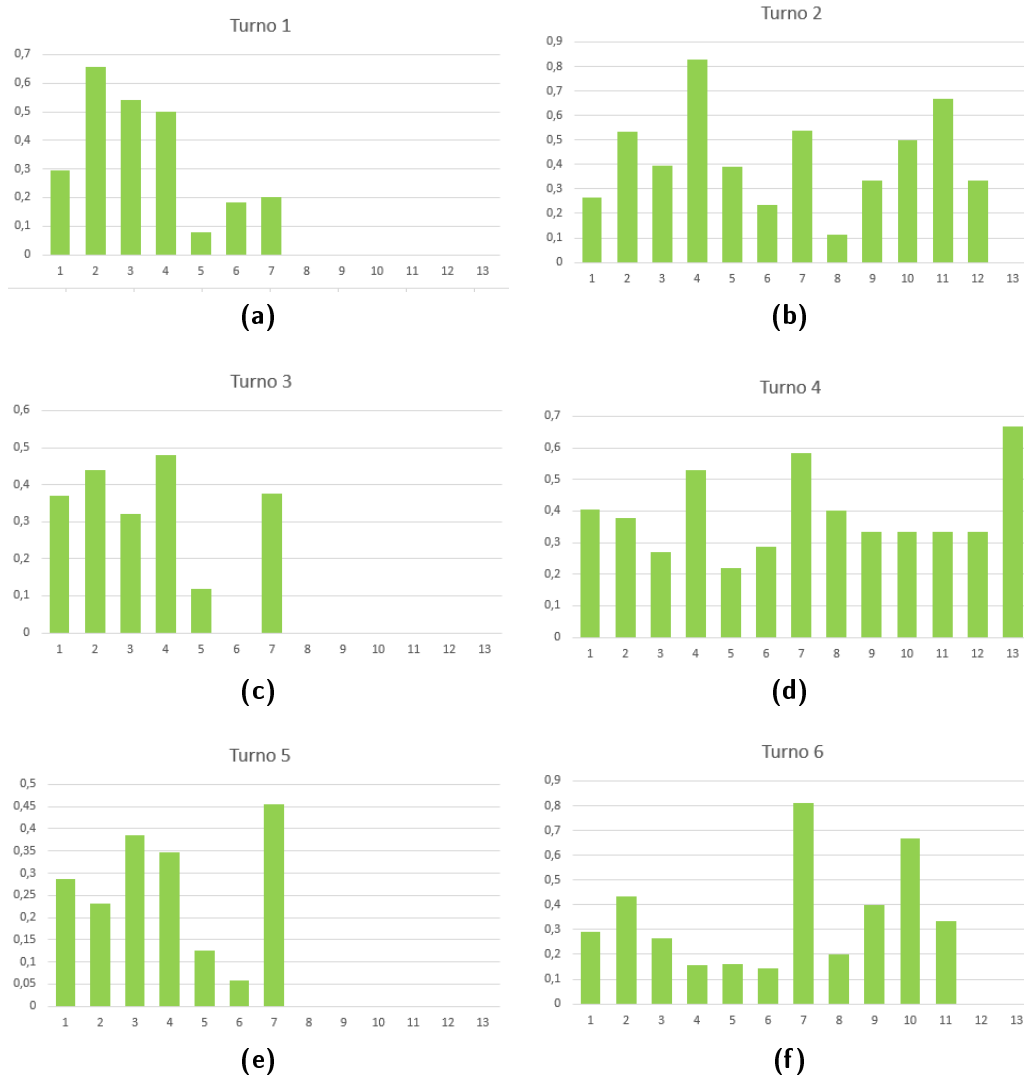


Figura 3.9: Percentuale di esercizi corretti rispetto agli accessi, divisi per turno

In questi grafici è riportato per ogni esercizio la percentuale di studenti che hanno svolto l'esercizio correttamente rispetto a coloro che hanno fatto accesso all'esercizio. Si può vedere come nei turni 1, 3, 5 gli studenti siano riusciti a svolgere solamente fino all'esercizio 7, mentre negli altri turni ci sono stati studenti che hanno raggiunto l'ultimo esercizio. Bisogna tenere conto del fatto che si tratta di percentuali rispetto agli accessi: ad esempio nel turno 4 la percentuale di soluzione dell'esercizio 13 è maggiore rispetto agli altri esercizi, ma il numero di accessi è molto basso.

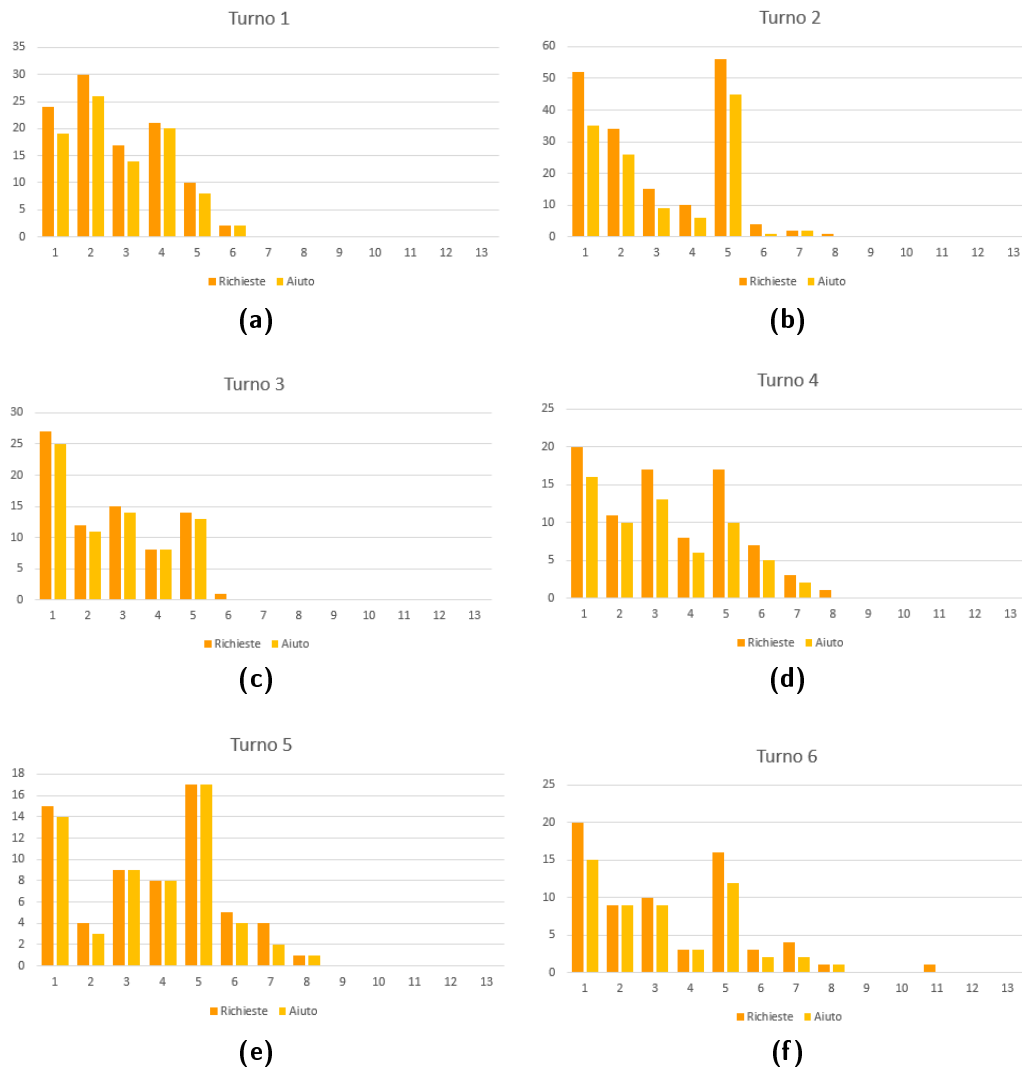


Figura 3.10: Interventi degli esercitatori, divisi per turno

In questi grafici invece vengono riportate le informazioni relative agli interventi degli esercitatori. Si può vedere che il tasso maggiore di intervento rispetto alle richieste si trova nei turni 3 e 5. Chiaramente questo fatto è dovuto al minor numero di studenti in questi turni.

Anche negli altri turni però gli esercitatori sono riusciti a rispondere con efficienza alle necessità degli studenti.

3.2 Categorizzazione delle sequenze

In questa parte si vanno ad analizzare le sequenze raggruppandole per categorie.

3.2.1 Accessi

In questa categoria ci si focalizza sul numero di accessi che sono stati fatti dagli studenti. Si noti che l'applicazione utilizzata dagli studenti durante il laboratorio era tale per cui per accedere a un esercizio successivo al primo fosse necessario accedere prima a tutti gli esercizi precedenti. Ad esempio, chi ha fatto accesso all'esercizio 5 avrà necessariamente fatto accesso anche ai primi 4 esercizi.

- 204 studenti hanno fatto accesso ai primi 3 esercizi, mentre 196 hanno fatto accesso ai primi 4 esercizi. Il 3.7% degli studenti non ha fatto accesso all'esercizio 4.

1	2	3				<u>204</u>	96.2
1	2	3	4			<u>196</u>	92.5

- 179 studenti hanno fatto accesso agli esercizi da 1 a 5 (84.4%).

1	2	3	4	5		<u>179</u>	84.4
---	---	---	---	---	--	------------	------

- 98 studenti hanno fatto accesso ai primi 6 esercizi (46.2%). Come visto in precedenza, il numero di studenti che ha fatto accesso a tutti i primi 6 esercizi è molto più basso rispetto ai primi 5.

1	2	3	4	5	6	<u>98</u>	46.2
---	---	---	---	---	---	-----------	------

3.2.2 Approccio

In questa categoria si analizza come gli studenti si avvicinano al laboratorio. In particolare se tendono a rimanere su un esercizio fino a quando non lo terminano oppure se in caso di difficoltà passano all'esercizio successivo.

- 147 studenti hanno svolto correttamente gli esercizi da 1 a 3 senza passare a un esercizio successivo prima di averlo terminato. (69.3%) – inoltre non risultano frequenti sequenze che contengano tra un accesso a un esercizio e il

suo svolgimento l'accesso a un esercizio successivo (ponendo una soglia pari al 30% degli studenti).



- Dato che 204 studenti hanno fatto accesso ai primi 3 esercizi e solo 146 studenti hanno svolto correttamente i primi 3 esercizi, 58 studenti (tra quelli che hanno fatto accesso agli esercizi) non sono stati in grado di risolvere uno o più esercizi tra i primi 3 (27%).
- 91 studenti hanno svolto correttamente gli esercizi da 1 a 4 (42.9%). Non risultano accessi ad esercizi successivi tra l'accesso a un esercizio e il corretto svolgimento.



- Il 92.5% degli studenti ha fatto accesso ai primi 4 esercizi; di questi, il 42.9% li ha svolti tutti e 4 correttamente. Inoltre:
 - 157 studenti hanno svolto correttamente gli esercizi 1 e 3
 - 175 studenti hanno svolto correttamente gli esercizi 1 e 2
 - 94 hanno svolto correttamente gli esercizi 3 e 4
 - 92 hanno svolto correttamente gli esercizi 1, 3 e 4
 - 102 hanno svolto correttamente gli esercizi 1,2 e 4
 - 93 hanno svolto correttamente gli esercizi 2, 3, 4
 - 104 studenti hanno svolto correttamente gli esercizi 2 e 4

Quindi almeno $104 - 93 = 11$ studenti hanno saltato due esercizi. È un numero basso di studenti.

1	2	3	4	<u>196</u>	92.5
1	1	3	3	<u>157</u>	74.1
1	1	2	2	<u>175</u>	82.5
3	3	4	4	<u>94</u>	44.3
1	1	3	3	<u>92</u>	43.4
1	1	2	2	<u>102</u>	48.1
2	2	3	3	<u>93</u>	43.9
2	2	4	4	<u>104</u>	49.1

Nei successivi commenti ci si focalizza invece sul fatto che lo studente abbia fatto accesso a un esercizio precedente a quello corrente. Questo comportamento può essere dovuto al confronto tra l'esercizio corrente e il precedente in modo da avere uno spunto per la soluzione dell'esercizio.

- Non risulta frequente che gli studenti siano tornati ad un esercizio precedente più di una volta.
- 43 studenti sono tornati indietro all'esercizio 1 dopo aver fatto accesso all'esercizio 2 (20.3%).

1	2	1	<u>43</u>	20.3
---	---	---	-----------	------

- 39 studenti sono tornati indietro all'esercizio 2 dopo aver fatto accesso all'esercizio 3 (18.4%).

2	3	2	<u>39</u>	18.4
---	---	---	-----------	------

- 42 studenti sono tornati indietro all'esercizio 3 dopo aver fatto accesso all'esercizio 4 (19.8%).

3	4	3	<u>42</u>	19.8
---	---	---	-----------	------

- 42 studenti sono tornati indietro all'esercizio 4 dopo aver fatto accesso all'esercizio 5 (19.8%).

4 5 4 42 19.8

- 36 studenti sono tornati indietro all'esercizio 5 dopo aver fatto accesso all'esercizio 6 (17%).

5 6 5 36 17

I risultati precedenti possono essere rappresentati nel seguente grafico:

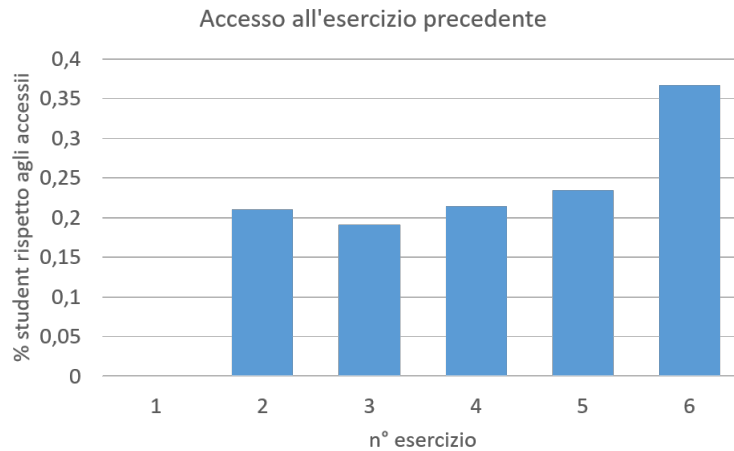


Figura 3.11: Studenti che hanno fatto accesso all'esercizio precedente

Nel grafico 3.11 viene rappresentata per ogni esercizio la percentuale di studenti che è tornata all'esercizio precedente rispetto al totale degli accessi all'esercizio. Ad esempio, circa il 21% degli studenti ha fatto accesso all'esercizio 2 e successivamente ha fatto accesso all'esercizio 1 (per questo motivo la colonna relativa all'esercizio 1 ha valore 0).

Si può vedere quindi che per l'esercizio 6 è particolarmente alta questa percentuale; questo significa che gran parte degli studenti è tornato all'esercizio 5 per prendere spunto per risolvere l'esercizio 6.

Per gli altri esercizi invece si nota che un numero esiguo di studenti è tornato a un esercizio precedente: non è un comportamento adottato dalla maggior parte degli studenti, ma essendo circa il 20% è un numero sufficiente da essere preso in considerazione.

3.2.3 Assistenza

In questa categoria si analizza se sia frequente che gli studenti richiedano aiuto. Inoltre si controlla l'efficacia degli aiuti prestati da parte degli esercitatori.

- Non risulta frequente più di una richiesta d'aiuto per la maggior parte degli studenti.
- La maggior parte delle richieste d'aiuto è stata fatta sull'esercizio 1 ed è stata fatta da 93 studenti (43.9%).

- 80 studenti che hanno fatto una richiesta d'aiuto sull'esercizio 1 lo hanno poi svolto correttamente (86% di chi ha fatto richiesta d'aiuto).

1 1 1 80 37.7

- Sono presenti 2 richieste d'aiuto solamente per il 16.5% degli studenti o una percentuale minore. Le due richieste d'aiuto sono sempre focalizzate su due esercizi differenti, a meno del primo esercizio che è l'unico per cui risultano essere fatte 2 richieste da parte dello stesso studente (16.5%).

1 1 1 35 16.5

Ci si concentra ora sull'efficacia degli aiuti da parte degli esercitatori, dove un aiuto è considerato efficace se ha portato in seguito alla soluzione dell'esercizio.

- 61 studenti dopo l'aiuto di un esercitatore sull'esercizio 1 hanno risolto correttamente l'esercizio (28.8%).

1 1 1 61 28.8

Si potrebbe erroneamente pensare che 61 studenti abbiano risolto correttamente l'esercizio 1 immediatamente dopo l'aiuto, ma la seguente sequenza

1 1 1 1 48 22.6

fa capire che in realtà solo $61 - 48 = 13$ persone hanno risolto correttamente l'esercizio 1 immediatamente dopo l'aiuto.

1 1 76 35.8

Dato che 76 persone hanno ricevuto un aiuto da parte di un esercitatore per l'esercizio 1, l'aiuto non è stato utile per $76 - 61 = 15$ persone (19.7% di chi ha richiesto aiuto su questo esercizio).

- 52 studenti dopo l'aiuto di un esercitatore sull'esercizio 2 hanno risolto correttamente l'esercizio (24.5%).

2	2	2	<u>52</u>	24.5
---	---	---	-----------	------

Dato che 61 persone hanno ricevuto un aiuto da parte di un esercitatore per l'esercizio 2, l'aiuto non è stato utile per $61 - 52 = 9$ persone (14.7% di chi ha richiesto aiuto su questo esercizio). Il risultato è migliore del primo esercizio.

2	2	<u>61</u>	28.8
---	---	-----------	------

- 36 studenti dopo l'aiuto di un esercitatore sull'esercizio 3 hanno risolto correttamente l'esercizio (17%).

3	3	3	<u>36</u>	17
---	---	---	-----------	----

Dato che 52 persone hanno ricevuto un aiuto da parte di un esercitatore per l'esercizio 3, l'aiuto non è stato utile per $52 - 36 = 16$ persone (30.7% di chi ha richiesto aiuto su questo esercizio). Il risultato è peggiore degli esercizi precedenti.

3	3	<u>52</u>	24.5
---	---	-----------	------

Nei casi evidenziati l'aiuto è stato utile agli studenti. Nel terzo esercizio però sono stati riscontrati problemi nonostante l'aiuto in un numero considerevole di casi.

Ci si concentra ora sugli aiuti a cui è seguito un errore.

- 52 studenti dopo l'aiuto di un esercitatore sull'esercizio 1 hanno fatto un errore (24.5%).

1	1	1	<u>52</u>	24.5
---	---	---	-----------	------

- 49 studenti dopo l'aiuto di un esercitatore sull'esercizio 2 hanno fatto un errore (23.1%).

2	2	2	<u>49</u>	23.1
---	---	---	-----------	------

- 32 studenti dopo l'aiuto di un esercitatore sull'esercizio 3 hanno fatto un errore (15.1%).

3	3	3	<u>32</u>	15.1
---	---	---	-----------	------

Si può vedere che in alcuni casi all'aiuto è seguito un errore. Questo può essere dato o da una mancata comprensione da parte dello studente o da un approccio dell'esercitatore che non si limita a dare la soluzione, ma fornisce un indizio per far avvicinare lo studente alla soluzione.

3.2.4 Soluzione

- Dei 147 studenti che hanno svolto correttamente i primi 3 esercizi, 128 hanno fatto un errore sull'esercizio 1, 116 sull'esercizio 2 e 117 sull'esercizio 3. Quindi al massimo 21 studenti hanno svolto correttamente i primi 3 esercizi al primo tentativo senza commettere un errore.

1	1	1	2	2	3	3	<u>128</u>	60.4
1	1	2	2	2	3	3	<u>116</u>	54.7
1	1	2	2	3	3	3	<u>117</u>	55.2

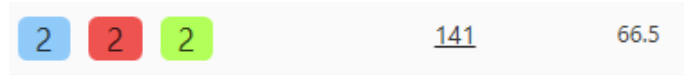
Nei punti successivi si valuta quante persone abbiano fatto un errore su un esercizio prima di farlo corretto.

- Su 189 studenti che hanno svolto correttamente l'esercizio 1, 169 hanno commesso un errore (79.7%).

1	1	1	<u>169</u>	79.7
---	---	---	------------	------

Quasi tutti gli studenti che hanno fatto corretto questo esercizio hanno prima commesso un errore (89.4% di coloro che hanno svolto l'esercizio correttamente). Inoltre 20 studenti hanno svolto l'esercizio 1 al primo tentativo (9.43% del totale degli studenti).

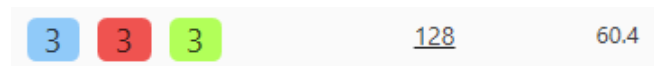
- Su 180 studenti che hanno svolto correttamente l'esercizio 2, 141 hanno commesso un errore (66.5%).



In questo caso buona parte degli studenti ha commesso un errore prima di risolvere l'esercizio correttamente, ma ci sono stati meno errori rispetto all'esercizio precedente (78.3% di coloro che hanno svolto l'esercizio correttamente). Questo può essere dato dal fatto che l'esercizio non introduceva grandi difficoltà rispetto al precedente e buona parte degli studenti è stata in grado di ricondursi alla soluzione proposta precedentemente per scrivere la query richiesta.

Ne consegue che 39 studenti siano riusciti a risolvere l'esercizio 2 al primo tentativo, quasi il doppio del caso precedente.

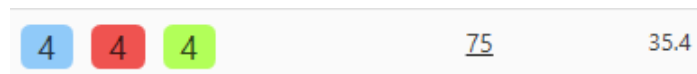
- Su 160 studenti che hanno svolto correttamente l'esercizio 3, 128 hanno commesso un errore (60.4%).



Il risultato è paragonabile all'esercizio precedente (80% di coloro che hanno svolto l'esercizio correttamente).

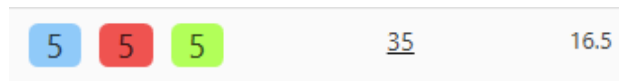
Si desume che 42 studenti siano riusciti a risolvere l'esercizio 3 al primo tentativo.

- Su 108 studenti che hanno svolto correttamente l'esercizio 4, 75 hanno prima commesso un errore (35.4%).



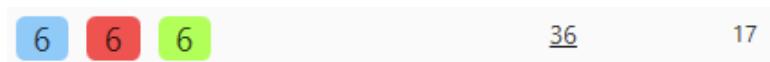
Per l'esercizio 4 la situazione è migliorata: infatti il 70% delle persone che ha svolto correttamente l'esercizio ha commesso un errore. Questo significa che 33 studenti sono riusciti a risolvere l'esercizio al primo tentativo. È da sottolineare il fatto che il numero di accessi a questo esercizio è calato.

- Tutti i 35 studenti che hanno svolto correttamente l'esercizio 5 hanno commesso un errore su questo esercizio (16.5%).



A differenza degli esercizi precedenti, si può vedere che tutti gli studenti in grado di risolvere l'esercizio hanno commesso almeno un errore. Tra tutti gli esercizi visti finora quindi risulta essere l'esercizio che ha portato più problematiche.

- Su 50 studenti che hanno risolto correttamente l'esercizio 6, 36 studenti hanno commesso un errore (17%).



In questo esercizio la situazione torna a migliorare. Il 72% degli studenti che hanno risolto correttamente l'esercizio ha fatto un errore; 14 studenti sono riusciti a risolvere l'esercizio al primo tentativo. Inoltre rispetto all'esercizio precedente il numero totale di studenti che ha risolto l'esercizio è tornato ad aumentare, segno che l'esercizio è risultato più semplice rispetto al numero 5.

I risultati vengono riassunti in figura 3.12:

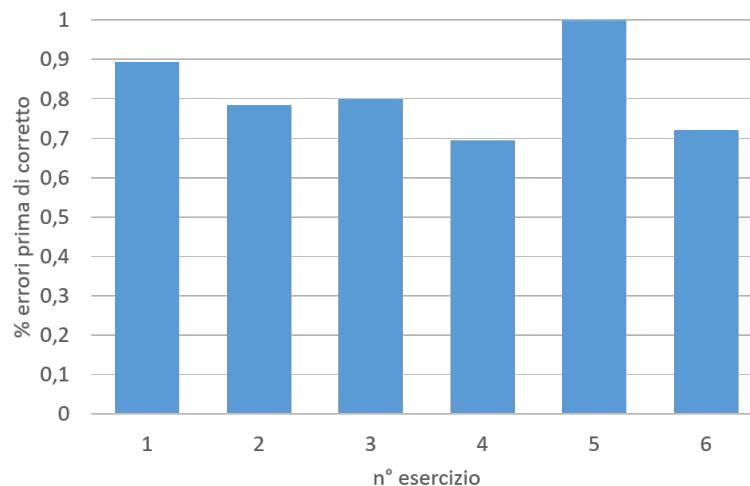


Figura 3.12: Percentuale di errori rispetto al totale dei corretti

In questo grafico per i 6 esercizi analizzati si riporta la percentuale di studenti che hanno fatto un errore su un esercizio prima di risolverlo rispetto al totale degli

studenti che hanno risolto l'esercizio.

Si vede immediatamente che tutti gli studenti che hanno risolto l'esercizio 5 hanno fatto almeno un errore.

Inoltre è immediato vedere che l'esercizio 4 risulta l'esercizio in cui il numero di studenti che non ha compiuto errori prima di risolverlo è massimo.

Dall'analisi del dataset inoltre è possibile ottenere il seguente grafico.

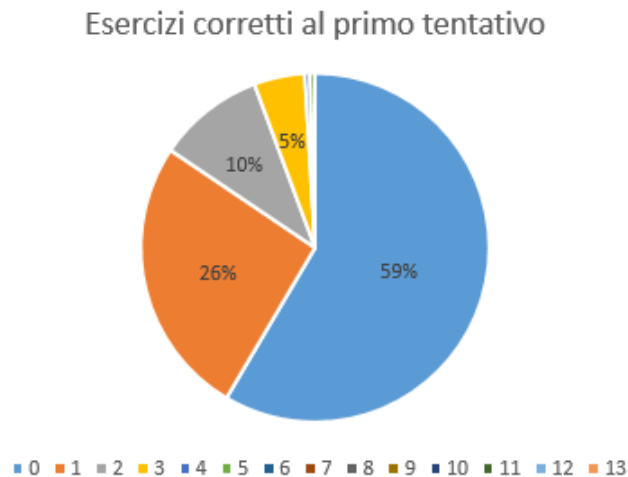


Figura 3.13: Studenti che hanno risolto esercizi al primo tentativo

In figura 3.13 viene riportato il numero di studenti che ha risolto al primo tentativo un numero di esercizi pari al valore corrispondente al colore riportato in legenda.

Si vede che la maggior parte degli studenti non è riuscita a risolvere un esercizio senza sbagliarlo.

Il 36% degli studenti però è riuscito a risolvere 1 o 2 esercizi senza fare errori. Considerando l'andamento degli esercizi e il gran numero di errori su ciascun esercizio, è un buon risultato.

Una percentuale minima di studenti è stata in grado di risolvere più di 3 esercizi senza fare errori.

Ora ci si concentra sulle sequenze di errori su esercizi differenti.

- 103 studenti hanno commesso almeno un errore sugli esercizi 1 e 2; 95 hanno commesso almeno un errore sugli esercizi 2 e 3; 110 hanno commesso un errore sugli esercizi 1 e 3; 83 hanno commesso almeno un errore sui primi 3 esercizi. Questo significa che al massimo $147 - 110 = 37$ studenti hanno svolto correttamente almeno due esercizi al primo tentativo. Dato che solo 83 hanno fatto un errore su tutti e 3 gli esercizi, il numero di questi studenti è certamente inferiore a 37.
- 128 studenti hanno fatto un errore su tutti gli esercizi da 1 a 3 (60.4%). I primi 3 esercizi quindi contengono molti errori, considerando che vi ha fatto accesso il 96.2% degli studenti. D'altra parte però il 40% degli studenti non ha commesso errori su almeno uno di questi 3 esercizi.



- 110 studenti hanno fatto 2 errori sui primi 2 esercizi (51.9%).



- 103 studenti hanno fatto un errore su tutti gli esercizi da 1 a 4 (48.6%). Dato che 196 studenti (92.5%) hanno fatto accesso ai primi 4 esercizi, il 52.5% di questi hanno fatto un errore su tutti i primi esercizi; è un numero molto elevato.

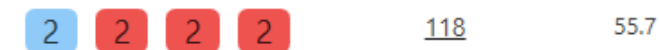


Ci si concentra quindi su errori ripetuti sullo stesso esercizio.

- 135 studenti hanno fatto 3 errori sull'esercizio 1 (63.7%).



- 118 studenti hanno fatto 3 errori sull'esercizio 2 (55.7%).



- 90 studenti hanno fatto 3 errori sull'esercizio 5 (42.5%).

5 5 5 5 90 42.5

- 102 studenti hanno fatto 3 errori sull'esercizio 3 (48.1%).

3 3 3 3 102 48.1

- 81 studenti hanno fatto 3 errori sull'esercizio 4 (38.2%).

4 4 4 4 81 38.2

- 72 studenti hanno fatto 3 errori sull'esercizio 1 e 5 errori sull'esercizio 2 (34%).

1 1 1 1 2 2 2 2 2 2 72 34

Si può vedere che fare errori ripetuti sullo stesso esercizio è un comportamento molto frequente, indipendentemente dall'esercizio.

È interessante confrontare quest'informazione con il totale degli studenti che hanno fatto accesso a questi esercizi:



Figura 3.14: Percentuale di studenti che hanno fatto 3 errori consecutivi su un esercizio

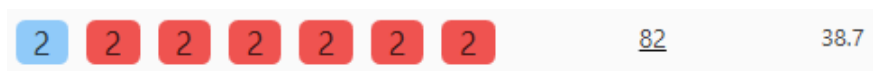
Si può vedere che una percentuale altissima di studenti ha fatto 3 errori sul primo esercizio; solamente l'esercizio 4 ha una percentuale minore del 50%. Fare errori consecutivi è quindi un comportamento ricorrente.

Raddoppiando la soglia a 6 errori consecutivi sullo stesso esercizio si ottengono i seguenti risultati:

- 70 studenti hanno fatto 6 errori sull'esercizio 1 (33%).



- 82 studenti hanno fatto 6 errori sull'esercizio 2 (38.7%).



- 49 studenti hanno fatto 6 errori sull'esercizio 5 (23.1%).



- 41 studenti hanno fatto 6 errori sull'esercizio 3 (19.3%).



- 55 studenti hanno fatto 6 errori sull'esercizio 4 (25.9%).



Viene fatto nuovamente il confronto con il totale di studenti che hanno fatto accesso agli esercizi:



Figura 3.15: Percentuale di studenti che hanno fatto 6 errori consecutivi su un esercizio

Per quanto le percentuali siano chiaramente più basse del caso precedente, si può vedere come raddoppiando il numero di errori le percentuali siano scese all'incirca del 20% per ciascun esercizio.

È indicativo inoltre che il 40% degli studenti che ha fatto accesso all'esercizio 2 abbia fatto almeno 6 errori su questo esercizio.

3.3 Analisi temporale

In questa parte si va a sfruttare l'aspetto più interessante dell'algoritmo cSpade, ossia l'utilizzo di vincoli temporali per comprendere come effettivamente il laboratorio sia stato affrontato dagli studenti. Le tempistiche inoltre permettono di evidenziare la difficoltà che è stata riscontrata dagli studenti nell'affrontare gli esercizi.

Successivamente le analisi verranno suddivise aggiungendo i parametri passati a cSpade per ogni particolare caso. Questo è necessario per evidenziare il significato delle sequenze che seguono.

In tutte le analisi è stato utilizzato un valore di *Minsup* pari a 0.1, ossia 21 studenti; per questo motivo il parametro non viene riportato. I valori dei parametri sono riportati in secondi, come richiesto dall'algoritmo cSpade. Nell'analisi si fa invece riferimento ai minuti per una interpretazione più diretta.

Risultato 1

Parametri:

- Mingap = 60
- Maxgap = 300

Con questi parametri si cercano sequenze in cui ogni evento è distanziato temporalmente dal successivo da poco tempo, compreso tra 1 e 5 minuti. Può essere interessante per valutare se sono presenti studenti che hanno risolto in poco tempo alcuni esercizi.

- 46 studenti hanno svolto correttamente l'esercizio 2 in meno di 5 minuti (21.7%).

2	2	<u>46</u>	21.7
---	---	-----------	------

- Non risulta frequente che siano stati svolti correttamente né l'esercizio 1 né l'esercizio 5 in meno di 5 minuti. Questo supporta l'ipotesi che su questi esercizi ci siano state più difficoltà rispetto ad altri.
- 38 studenti hanno svolto correttamente l'esercizio 4 in meno di 5 minuti (17.9%).

4	4	<u>38</u>	17.9
---	---	-----------	------

- 30 studenti hanno svolto correttamente l'esercizio 3 in meno di 5 minuti (14.2%). Rispetto all'esercizio precedente c'è una percentuale decisamente minore.

3	3	<u>30</u>	14.2
---	---	-----------	------

- 23 studenti hanno svolto correttamente l'esercizio 6 in meno di 5 minuti (10.8%).

6	6	<u>23</u>	10.8
---	---	-----------	------

- Non risulta frequente che gli studenti abbiano svolto correttamente più di un esercizio nell'arco di 5 minuti ciascuno.
- 47 studenti dopo hanno commesso un errore sull'esercizio 1 e poi lo hanno svolto correttamente (22%).

1	1	1	<u>47</u>	22.2
---	---	---	-----------	------

È da notare che l'intera sequenza è avvenuta in un arco di tempo che va da 2 a 10 minuti a seconda degli studenti, dato che i parametri sono relativi alla distanza temporale tra due eventi consecutivi. In ogni caso, circa il 20% degli studenti ha svolto correttamente l'esercizio 1 in un tempo tra 5 e 10 minuti.

Risultato 2

Parametri:

- $\text{Mingap} = 10$
- $\text{Maxgap} = 60, 120, 180, 240$

I file in esame differiscono nei valori di *maxgap*. In questo modo è stato possibile definire per i primi esercizi all'incirca quanto sia stato il tempo necessario per risolverli.

- Non risulta frequente che un esercizio sia stato risolto in meno di due minuti.

- 28 studenti hanno svolto correttamente l'esercizio 2 in meno di 3 minuti (13.2%). Per queste persone deve essere risultato particolarmente semplice.



- 43 studenti hanno svolto correttamente l'esercizio 2 in meno di 4 minuti (20.3%).



Di conseguenza $43 - 28 = 15$ studenti hanno risolto l'esercizio 2 in meno di 3 minuti (7%).

- 31 studenti hanno svolto correttamente l'esercizio 4 in meno di 4 minuti (14.6%).



Si può analizzare inoltre il tempo che è stato impiegato dagli studenti per risolvere ciascun esercizio, concentrandosi sui primi 5 esercizi.

Nei successivi grafici in ascissa viene indicata la fascia temporale espressa in minuti che indica il tempo tra l'accesso all'esercizio e la sua soluzione. Sono state scelte 5 fasce temporali:

- 0 - 5 minuti: l'esercizio non ha costituito problemi e la soluzione è stata quasi immediata
- 5 - 10 minuti: l'esercizio è stato risolto in poco tempo, richiedendo un minimo di ragionamento
- 10 - 15 minuti: c'è stato qualche problema di comprensione che ha richiesto probabilmente diversi tentativi
- 15 - 30 minuti: l'esercizio è risultato per lo studente molto arduo da affrontare
- > 30 minuti: l'esercizio era decisamente al di sopra delle capacità dello studente

In ordinata invece è riportata la percentuale di studenti che ha impiegato un tempo corrispondente alla fascia temporale.

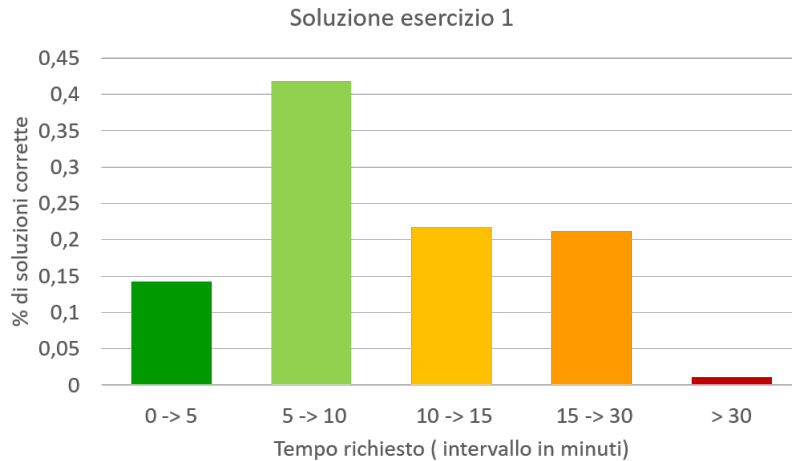


Figura 3.16: Tempo impiegato a risolvere l'esercizio 1

Nel grafico 3.16 viene rappresentato il tempo impiegato dagli studenti per risolvere l'esercizio 1. Si può vedere come la maggior parte degli studenti abbia impiegato tra 5 e 10 minuti per risolvere l'esercizio. Tanti studenti però hanno impiegato un tempo tra 15 e 30 minuti, che per il primo esercizio è un tempo decisamente elevato.

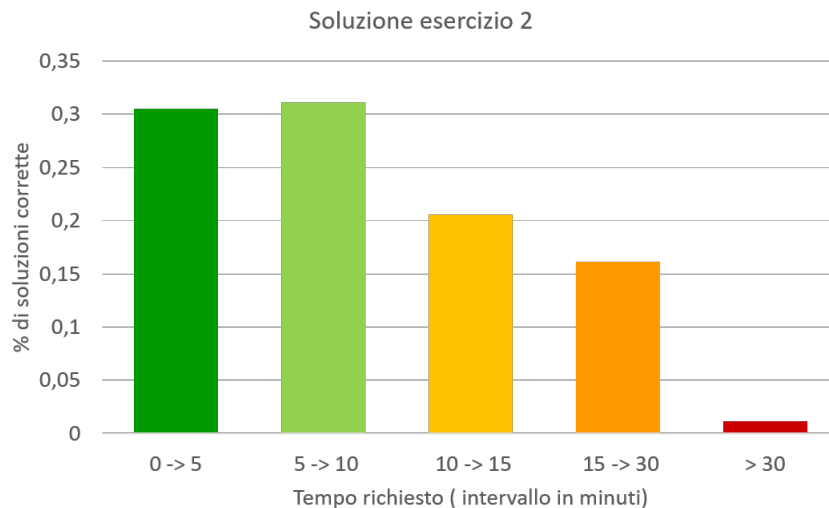


Figura 3.17: Tempo impiegato a risolvere l'esercizio 2

Nel grafico 3.17 viene rappresentato il tempo impiegato dagli studenti per risolvere l'esercizio 2. Rispetto all'esercizio precedente moltissimi studenti hanno risolto l'esercizio in meno di 5 minuti e altrettanti hanno impiegato meno di 10 minuti. Probabilmente sono riusciti a ricondursi al primo esercizio per risolvere più velocemente l'esercizio. Anche se in minor numero, molti studenti hanno impiegato un tempo considerevole a risolvere l'esercizio. Questi non sono stati in grado di collegarsi all'esercizio precedente per trovare uno schema da seguire.

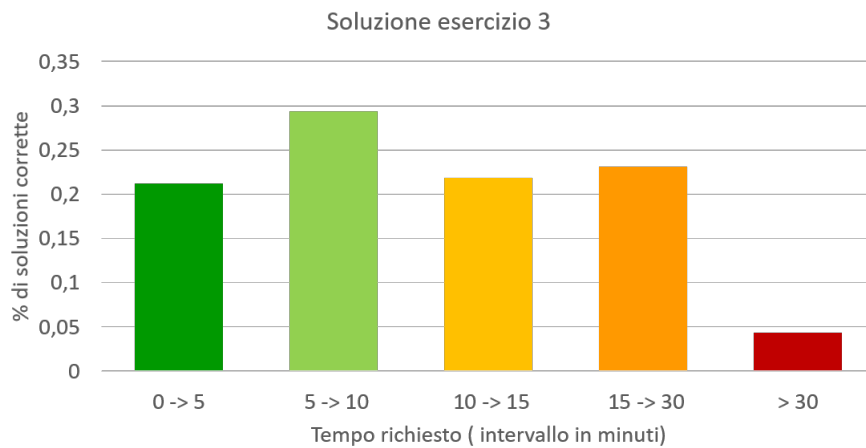


Figura 3.18: Tempo impiegato a risolvere l'esercizio 3

Nel grafico 3.18 viene rappresentato il tempo impiegato dagli studenti per risolvere l'esercizio 3. Il numero di studenti che è riuscito a risolvere l'esercizio in meno di 5 minuti è di nuovo diminuito, tornando ai livelli del primo esercizio. C'è di nuovo un buon numero di studenti che ha impiegato un tempo considerevole a risolvere l'esercizio. Su questo esercizio inizia ad essere considerevole il numero di studenti che hanno impiegato più di mezz'ora: l'esercizio 3 si conferma una fonte di problemi per gli studenti.

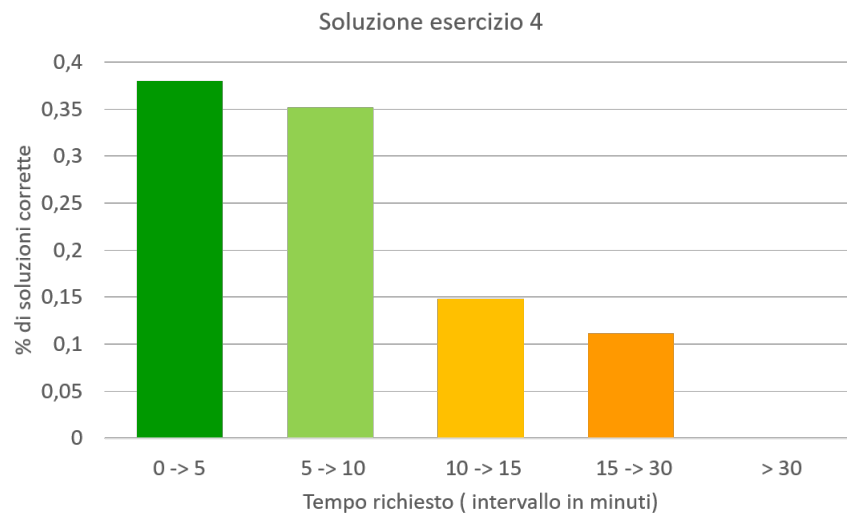


Figura 3.19: Tempo impiegato a risolvere l'esercizio 4

Nel grafico 3.19 viene rappresentato il tempo impiegato dagli studenti per risolvere l'esercizio 4. È evidente come in questo esercizio non ci siano stati particolari problemi: la gran maggioranza degli studenti l'ha risolto in meno di 10 minuti (e tra questi più della metà in meno di 5 minuti), mentre un numero più contenuto ha impiegato più tempo.

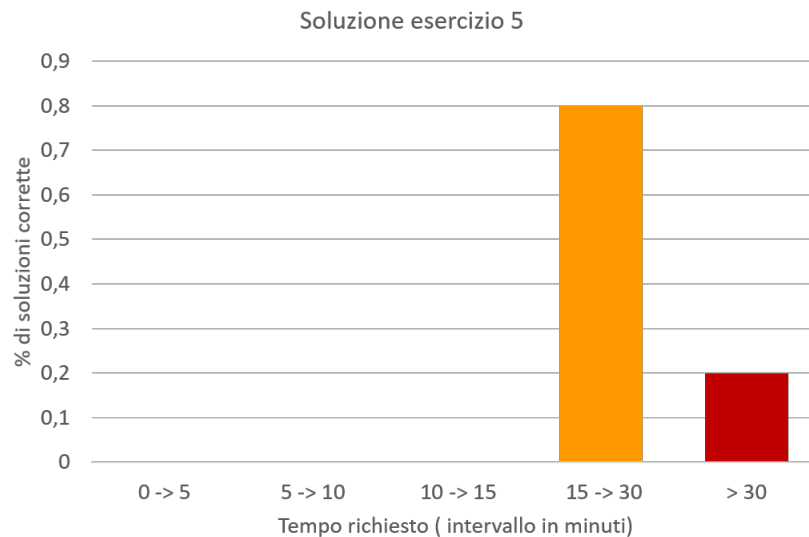


Figura 3.20: Tempo impiegato a risolvere l'esercizio 5

Nel grafico 3.20 viene rappresentato il tempo impiegato dagli studenti per risolvere l'esercizio 5. La situazione è opposta rispetto a tutti gli esercizi precedenti. Nessuno studente è infatti riuscito a risolvere l'esercizio in meno di 15 minuti. La maggior parte degli studenti ha impiegato tra i 15 e i 30 minuti per risolverlo. Inoltre il numero totale di studenti che è riuscito a risolvere l'esercizio è molto inferiore rispetto agli esercizi precedenti, quindi quest'esercizio è risultato essere un vero punto d'arresto nel laboratorio.

Attraverso gli stessi parametri è possibile anche studiare quanto tempo è intercorso tra l'accesso e il primo tentativo dello studente, che può essere ad esempio un errore. In questo modo si può valutare se il comportamento più comune degli studenti sia fare immediatamente dei tentativi o cercare di ragionare sull'esercizio per poi testare la propria idea.

- 30 studenti hanno fatto accesso all'esercizio 1 e in meno di un minuto hanno fatto un errore sull'esercizio (14.2%). Questi studenti non si sono fermati a ragionare, ma hanno probabilmente testato la prima idea che hanno avuto.

1	1	30	14.2
---	---	----	------

- Aumentando la soglia massima a 2 minuti si riscontra che sono frequenti tentativi anche sugli esercizi 2 e 3. In particolare si arriva intorno al 15% degli studenti per questi due esercizi, mentre per il primo esercizio si supera il 20%.

1	1	47	22.2
2	2	35	16.5
3	3	27	12.7

- Ponendo infine la soglia massima a 3 minuti si può vedere che compaiono tutti i primi 7 esercizi. Questo indica che sugli esercizi dal 4 in poi gli studenti hanno trovato difficoltà anche solo a formulare una query che potesse soddisfare la richiesta.

1	1	<u>59</u>	27.8
2	2	<u>56</u>	26.4
3	3	<u>38</u>	17.9
4	4	<u>32</u>	15.1
6	6	<u>30</u>	14.2
5	5	<u>24</u>	11.3
7	7	<u>22</u>	10.4

È possibile fare il confronto rispetto al numero di accessi per l'esercizio:

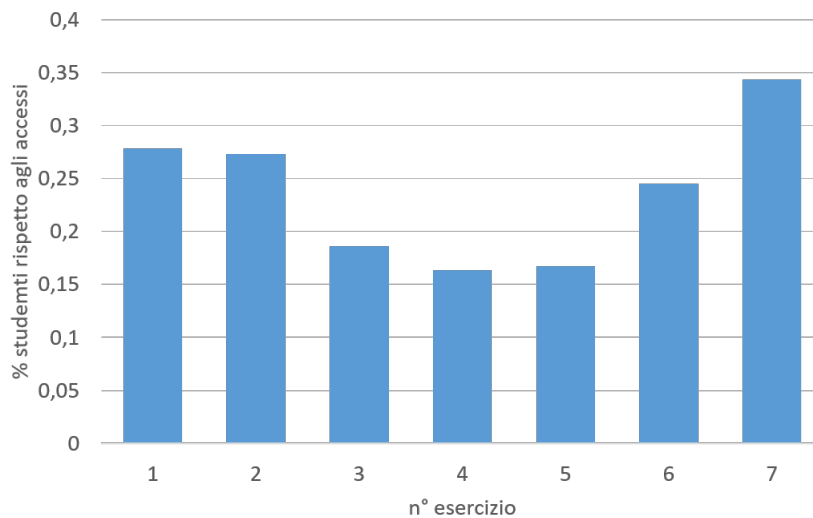


Figura 3.21: Errore entro 3 minuti dopo l'accesso all'esercizio

Come si può vedere, in media il 20% degli studenti ha compiuto un errore entro 3 minuti dall'accesso. La percentuale sale molto per l'esercizio 7. Per l'esercizio 5 invece la percentuale è più bassa: può darsi che gli studenti abbiano impiegato più tempo a fare un tentativo, viste le difficoltà rilevate nelle precedenti analisi.

Risultato 3

Parametri:

- Mingap = 1800

Ci si concentra principalmente su un'ampia distanza temporale tra due accessi (30 minuti) per vedere se ci sono state difficoltà nel procedere nel laboratorio.

- 158 studenti hanno fatto accesso all'esercizio 1 e dopo almeno 30 minuti hanno fatto accesso all'esercizio 5 (74.5%). La maggior parte degli studenti ha quindi speso almeno 30 minuti sui primi 4 esercizi.

1	5	<u>158</u>	74.5
---	---	------------	------

Questo significa che, dato che 179 studenti hanno fatto accesso all'esercizio 5, 21 studenti sono riusciti a fare accesso all'esercizio 5 dopo meno di 30 minuti (10% del totale, 12% di coloro che hanno fatto accesso all'esercizio 5). Inoltre sono presenti persone che hanno impiegato più di 30 minuti a far accesso all'esercizio 5.

- 82 studenti hanno fatto accesso all'esercizio 3 e dopo almeno 30 minuti hanno fatto accesso all'esercizio 5 (38.7%). Circa un terzo degli studenti ha avuto grandi difficoltà concentrate sugli esercizi 3 e 4.

3	5	<u>82</u>	38.7
---	---	-----------	------

- 38 studenti hanno fatto accesso all'esercizio 1 e dopo almeno 30 minuti hanno fatto accesso all'esercizio 8 (17.9%). Si vede che circa il 20% ha fatto accesso a questo esercizio (non si può desumere molto di più dato che non c'è un limite superiore al tempo tra i due eventi in quest'analisi).

1	8	<u>38</u>	17.9
---	---	-----------	------

- 34 studenti hanno fatto accesso all'esercizio 1 e dopo almeno 30 minuti hanno fatto accesso all'esercizio 2 (16%). Questi studenti hanno avuto grandi difficoltà sul primo esercizio ed è un numero considerevole.

1	2	<u>34</u>	16
---	---	-----------	----

- 24 studenti hanno fatto accesso all'esercizio 1, dopo almeno 30 minuti all'esercizio 4 e dopo altri 30 minuti all'esercizio 6 (11.3%). Si può vedere in questa sequenza come effettivamente i primi esercizi abbiano richiesto almeno un'ora.

1	4	6	<u>24</u>	11.3
---	---	---	-----------	------

- Non risulta frequente che gli studenti abbiano impiegato 30 minuti per svolgere correttamente un esercizio.
- Non risulta frequente che gli studenti abbiano fatto un errore su un esercizio 30 minuti dopo avervi fatto accesso.

Risultato 4

Parametri:

- Mingap = 60
- Maxgap = 600

Ci si concentra su eventuali richieste d'aiuto fatte dopo l'accesso in un arco di 1 – 10 minuti.

- 58 studenti hanno fatto accesso all'esercizio 1 e entro 10 minuti hanno fatto una richiesta d'aiuto (27.4%).

1	1	<u>58</u>	27.4
---	---	-----------	------

- 44 studenti hanno fatto accesso all'esercizio 2 e entro 10 minuti hanno fatto una richiesta d'aiuto (20.8%).

2	2	<u>44</u>	20.8
---	---	-----------	------

Se si abbassa Maxgap a 300, non risulta frequente che siano state fatte richieste d'aiuto per l'esercizio 1, mentre per l'esercizio 2 è vero solo per 22 studenti, ossia la metà rispetto a un intervallo di tempo massimo di 10 minuti.

Queste informazioni fanno desumere che tendenzialmente gli studenti hanno fatto dei tentativi o si sono fermati a ragionare su una possibile soluzione piuttosto che

chiedere immediatamente l'aiuto di un esercitatore.

Con gli stessi parametri si valuta la rapidità con cui gli interventi hanno seguito la richiesta d'aiuto. Vengono analizzati i primi 5 esercizi essendo questi gli esercizi su cui sono state frequenti le richieste.

- 39 studenti hanno ricevuto assistenza meno di 10 minuti dopo aver fatto richiesta sull'esercizio 1.

1	1	<u>39</u>	18.4
---	---	-----------	------

- 31 studenti hanno ricevuto assistenza meno di 10 minuti dopo aver fatto richiesta sull'esercizio 2.

2	2	<u>31</u>	14.6
---	---	-----------	------

- 24 studenti hanno ricevuto assistenza meno di 10 minuti dopo aver fatto richiesta sull'esercizio 3.

3	3	<u>24</u>	11.3
---	---	-----------	------

- 31 studenti hanno ricevuto assistenza meno di 10 minuti dopo aver fatto richiesta sull'esercizio 4.

4	4	<u>31</u>	14.6
---	---	-----------	------

- 32 studenti hanno ricevuto assistenza meno di 10 minuti dopo aver fatto richiesta sull'esercizio 5.

5	5	<u>32</u>	15.1
---	---	-----------	------

Questi dati sembrano indicare che gli studenti spesso hanno dovuto aspettare un pò di tempo prima di ricevere l'aiuto da parte degli esercitatori; d'altra parte il numero di esercitatori è limitato rispetto al numero di studenti, per cui in una situazione in cui tutti gli studenti richiedono aiuto sullo stesso esercizio è da considerarsi

normale avere un'attesa di 10 minuti.

Inoltre si deve aggiungere anche che gli esercitatori avevano la possibilità di fornire un supporto basandosi sul colore del semaforo di ogni postazione visualizzabile sull'app dedicata agli assistenti. È possibile quindi che per un numero considerevole di interventi da parte di un esercitatore non sia presente in precedenza la richiesta di aiuto da parte dello studente.

Risultato 5

Parametri:

- Mingap = 600
- Maxgap = 900

È stato scelto un intervallo tra 10 e 15 minuti tra un evento e il successivo per capire se ci siano stati esercizi che hanno richiesto un po' di tempo da parte degli studenti o che abbiano portato difficoltà.

- 97 studenti hanno commesso un errore sull'esercizio 1 10-15 minuti dopo l'accesso (45.8%).

1	1	<u>97</u>	45.8
---	---	-----------	------

- 81 studenti hanno commesso un errore sull'esercizio 3 10-15 minuti dopo l'accesso (38.2%).

3	3	<u>81</u>	38.2
---	---	-----------	------

- 74 studenti hanno commesso un errore sull'esercizio 5 10-15 minuti dopo l'accesso (34.9%).

5	5	<u>74</u>	34.9
---	---	-----------	------

Dai precedenti punti si avvalora l'ipotesi che gli studenti abbiano trovato molte difficoltà sugli esercizi 1, 3 e 5. Infatti tra 10 e 15 minuti dopo l'accesso hanno continuato a sbagliare su questi esercizi.

- 63 studenti hanno svolto correttamente l'esercizio 1 nell'arco di 10-15 minuti (29.7%).

1	1	<u>63</u>	29.7
---	---	-----------	------

- 48 studenti hanno svolto correttamente l'esercizio 2 nell'arco di 10-15 minuti (22.6%).

2	2	<u>48</u>	22.6
---	---	-----------	------

- 47 studenti hanno svolto correttamente l'esercizio 3 nell'arco di 10-15 minuti (22.2%).

3	3	<u>47</u>	22.2
---	---	-----------	------

Dai precedenti punti si nota che una buona parte degli studenti ha impiegato un certo tempo per risolvere i primi tre esercizi.

Capitolo 4

Conclusioni

Dalle analisi viste precedentemente è emerso che il laboratorio era troppo complesso per la preparazione degli studenti.

La difficoltà è evidenziata dal fatto che la maggior parte degli studenti non è riuscita a fare accesso a un esercizio successivo al 5. Infatti l'85% degli studenti ha fatto accesso all'esercizio 5, mentre questa percentuale scende al 46% per l'esercizio 6 e al 18% per l'esercizio 8. Inoltre il numero di errori e il tempo utilizzato sull'esercizio 5 fanno pensare che questo esercizio dovesse trovarsi più avanti nel laboratorio.

Ponendo l'attenzione sugli errori compiuti dagli studenti, già nell'analisi esplorativa riportata in figura 3.2 si era evidenziato che sui primi esercizi ci fossero in media tra 4 o 5 errori da parte di uno studente. Questa tesi risulta rafforzata dalle figure 3.14 e 3.15 le quali mostrano che fare diversi errori sullo stesso esercizio sia un comportamento frequente. Questo può essere dovuto da un approccio mirato a comprendere per tentativi, ma anche dall'effettiva difficoltà riscontrata. Tra le due alternative però i risultati relativi alle tempistiche degli accessi a esercizi successivi suggeriscono che la causa di questi errori sia la problematicità delle richieste. Il gran numero di errori anche ripetuti sugli esercizi 3 e 5 inoltre suggerisce che gli studenti abbiano trovato difficoltà nella comprensione dell'utilizzo dell'operatore *GROUP BY*.

Come si è indicato in precedenza, la presenza di errori potrebbe essere data da un approccio di tipo *trial-and-error*, in cui gli studenti fanno tentativi per capire come approcciarsi a un esercizio.

Di conseguenza è importante tenere conto anche degli esercizi risolti correttamente: come si evince dalla figura 3.5 sui primi 3 esercizi nonostante l'elevato numero di errori c'è anche un'elevata percentuale di studenti che ha svolto correttamente gli esercizi. D'altro canto invece sull'esercizio 5 solo il 20% degli studenti è riuscito

a risolvere l'esercizio, fatto che rafforza la considerazione che l'esercizio dovrebbe essere spostato più avanti nel laboratorio.

Bisogna inoltre tenere conto del grafico in figura 3.12 che riporta l'elevata frequenza di un errore da parte degli studenti prima di risolvere correttamente l'esercizio. I risultati relativi agli errori sono quindi da mettere in correlazione con quelli relativi alle soluzioni corrette. Per quanto ad esempio sia l'esercizio 1 che l'esercizio 5 abbiano percentuale di errore rispetto all'esercizio corretto molto alte (rispettivamente 90% e 100%), il grafico in figura 3.5 indica che per l'esercizio 1 è stato frequente compiere un errore, ma non comporta necessariamente grandi difficoltà, mentre per l'esercizio 5 la bassa percentuale di soluzioni corrette sottolinea quanto questo esercizio sia stato complesso per gli studenti.

Per quanto riguarda l'assistenza di laboratorio da parte degli esercitatori, si è evidenziato come questa sia stata efficace. In particolare il tasso di risposta alle richieste è molto alto, com'è evidenziato in figura 3.3. Solo negli esercizi 1 e 5, che sono quelli su cui sono stati riscontrati più problemi, il tasso di risposta è un po' più basso. Si è visto poi nella Categorizzazione delle Sequenze che in gran percentuale gli aiuti degli esercitatori hanno portato a risolvere correttamente l'esercizio. Questo porta ad affermare che l'approccio che coinvolge sia studenti che assistenti più esperti sia stato positivo. Inoltre rafforza l'efficacia dell'ambiente di lavoro sviluppato per la raccolta dei dati, che facilita l'interazione tra esercitatori e studenti e permette di evitare agli studenti lunghe attese senza ottenere supporto. Questo risultato trova riscontro anche nelle prime impressioni evidenziate da studenti ed esercitatori dopo il laboratorio, come evidenziato in [1].

Come si evince dal Risultato 02 dell'Analisi temporale, gran parte degli studenti ha avuto grandi difficoltà a risolvere gli esercizi e ha impiegato molto tempo a terminarli. Un fatto ancora più evidente è che nessuno è stato in grado di risolvere l'esercizio 5 in meno di 15 minuti. Questo indica che l'esercizio richiede almeno un chiarimento o un'indicazione specifica in modo da prendere la direzione corretta.

Queste considerazioni possono essere utili per i docenti per rivedere la struttura del laboratorio, che dovrebbe includere esercizi più semplici; inoltre l'aumento di difficoltà graduale dovrebbe essere accompagnato da un richiamo all'esercizio precedente, in modo che gli studenti possano trovare continuità tra un esercizio e il successivo. Quest'accortezza potrebbe portare a un miglioramento dei risultati ottenuti dagli studenti: in questo modo si sentirebbero più motivati a procedere nel laboratorio e non si lascerebbero scoraggiare dalla mole di lavoro rimanente.

4.1 Spunti di approfondimento

Sono presenti ancora diversi aspetti che possono fornire materiale interessante di studio. Un'estensione di questo lavoro può essere non limitarsi alla raccolta di dati di un singolo laboratorio, ma acquisire i dati di ogni laboratorio settimanalmente. In questo modo si può valutare in maniera concreta se ci sia un miglioramento da parte degli studenti e dinamicamente riprogrammare le esercitazioni in aula in modo da colmare le lacune precedenti.

Questo tipo di lavoro richiederebbe una modifica dell'applicazione client in modo da supportare non solo gli esercizi del primo laboratorio ma anche i successivi. D'altra parte l'applicazione non deve essere modificata ulteriormente; si potrebbe aggiungere un menù a tendina da cui selezionare il numero del laboratorio in modo che chi non ha avuto modo di terminare il laboratorio precedente possa concluderlo.

Inoltre sarebbe possibile anche comprendere effettivamente quanti studenti smettano a un certo punto del corso di frequentare i laboratori e si potrebbe cercare di valutare il motivo. In particolare se lo studente aveva riscontrato difficoltà nei laboratori precedenti con buona probabilità pensa di non essere sufficientemente preparato; d'altra parte uno studente con buoni risultati nel laboratorio può ritenere di non aver problemi e svolgere gli esercizi per conto proprio.

La modifica dell'applicazione quindi permetterebbe di avere maggiori informazioni. Si aggiunge però un vincolo: se si vuole mantenere la privacy dello studente salvando solo un'identificativo numerico per il gruppo, è necessario che lo stesso identificativo venga utilizzato per le sessioni successive alla prima.

Data quest'analisi generale su come gli studenti hanno affrontato gli esercizi, potrebbe essere istruttivo svolgere una ricerca di text mining sugli esercizi sottomessi dagli studenti in modo da studiare quali siano gli errori più comuni ed evidenziarli in aula in modo da prevenirli.

Un altro aspetto interessante da approfondire sarebbe la clusterizzazione degli studenti in modo da capire se sono presenti gruppi di studenti molto diversi tra loro e quali siano le caratteristiche che li differenziano. Questo task sarebbe fondamentale soprattutto per individuare i gruppi con maggiori difficoltà e per consigliare un piano di recupero per colmare le lacune.

Bibliografia

- [1] Luca Cagliero, Luigi De Russis, Laura Farinetti, Teodoro Montanaro, *Improving the effectiveness of SQL learning practice: a data-driven approach*, 2018.
- [2] Deborah Knox, Ursula Wolz, Daniel Joice, Elliot Koffman, Joan Krone, Atika Laribi, J. Paul Myers, Viera K. Proulx, Kenneth A. Reek, *Use of laboratories in Computer Science education: guidelines for good practice*, ITiCSE '96 Proceedings of the 1st conference on Integrating technology into computer science education, 1996.
- [3] Rakesh Agrawal, Ramakrishnan Srikant, *Mining Sequential Patterns*, 11th ICDE Conf, 1995.
- [4] Mohammed J. Zaki, *SPADE: An Efficient Algorithm for Mining Frequent Sequences*, Machine Learning Journal, 2001.
- [5] Mohammed J. Zaki, *Efficient Enumeration of Frequent Sequences*, 9th ACM International Conference on Information and Knowledge Management, 1998.
- [6] Mohammed J. Zaki, *Sequences Mining in Categorical Domains: Incorporating Constraints*, 9th ACM International Conference on Information and Knowledge Management, 2000.
- [7] Karam Gouda, Mosab Hassaan, Mohammed J. Zaki, *Sequences Mining in Categorical Domains: Incorporating Constraints*, Journal of Computer and Systems Sciences, 2010.
- [8] Karam Gouda, Mosab Hassaan, Mohammed J. Zaki, *PRISM: A Prime-Encoding Approach for Frequent Sequence Mining*, 7th IEEE International Conference on Data Mining, 2007.