



LABORATORIO 10

REALIZZAZIONE DI UN SW PER GESTIRE ALBERI DI DECISIONE

Algoritmi e strutture dati 2018-2019

DIFFERENZE RISPETTO AI LABO PRECEDENTI

- Maggiore libertà nell'implementazione
 - non viene fornita nessuna traccia
 - non viene imposta una struttura dati
- Esiste però una descrizione **‘abbastanza precisa’** (ma non completa) di quello che dovete implementare e dei vincoli/suggerimenti che vanno rispettati
- Il laboratorio deve essere svolto e consegnato in modo individuale!
- Il laboratorio dà diritto all'acquisizione di due punti e mezzo sull'esame di ASD

GLI ALBERI DI DECISIONE (DECISION TREES)

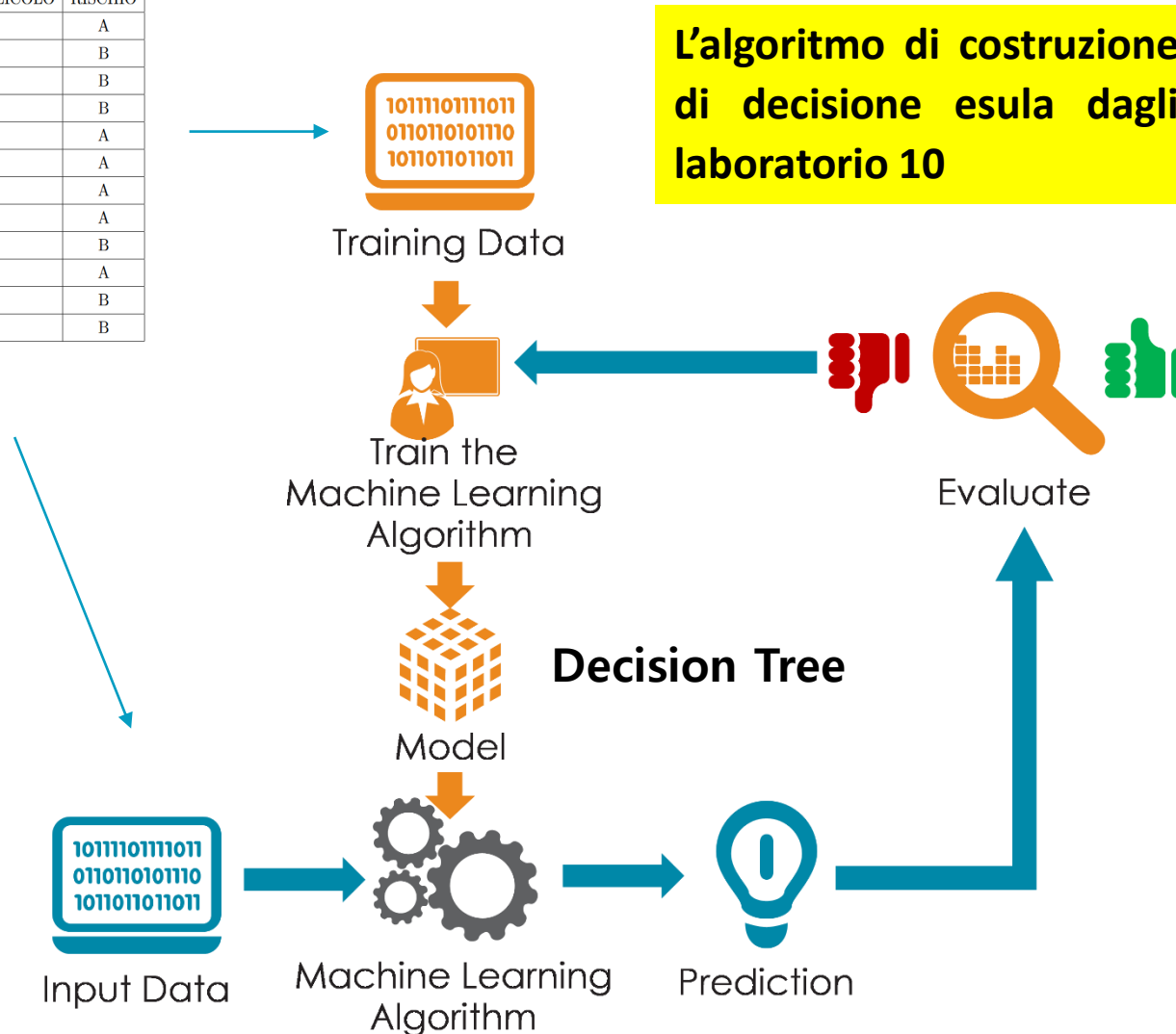
- Un albero di decisione è un **modello predittivo** utile per differenti scopi e spesso viene utilizzato come strumento per il supporto alle decisioni
- Ad esempio un albero di decisione potrebbe essere utilizzato per
 - stimare il fatturato di un punto vendita sulla base delle sue caratteristiche
 - dimensioni del punto vendita,
 - città, numero di impiegati,
 -
 - classificare alcuni movimenti bancari in ordinari o fraudolenti



COSTRUZIONE DI UN DECISION TREE

Insieme di dati (data set) di cui si conosce la predizione

	ETÀ	TIPO DI AUTOVEICOLO	RISCHIO
Utente 1	17	S	A
Utente 2	43	FA	B
Utente 3	68	FA	B
Utente 4	32	AU	B
Utente 5	23	FA	A
Utente 6	18	FA	A
Utente 7	20	FA	A
Utente 8	45	S	A
Utente 9	50	AU	B
Utente 10	64	AU	A
Utente 11	46	FA	B
Utente 12	40	FA	B



L'algoritmo di costruzione dell'albero di decisione esula dagli scopi del laboratorio 10

ESEMPIO: COMPAGNIA DI ASSICURAZIONI (1)

- Supponiamo che dati del training set provengano da una compagnia di assicurazioni, nella quale un esperto ha assegnato ad ogni utente un livello di rischio (A=alto, B=basso) basandosi su incidenti effettivamente avvenuti

	ETÁ	TIPO DI AUTOVEICOLO	RISCHIO
Utente 1	17	S	A
Utente 2	43	FA	B
Utente 3	68	FA	B
Utente 4	32	AU	B
Utente 5	23	FA	A
Utente 6	18	FA	A
Utente 7	20	FA	A
Utente 8	45	S	A
Utente 9	50	AU	B
Utente 10	64	AU	A
Utente 11	46	FA	B
Utente 12	40	FA	B

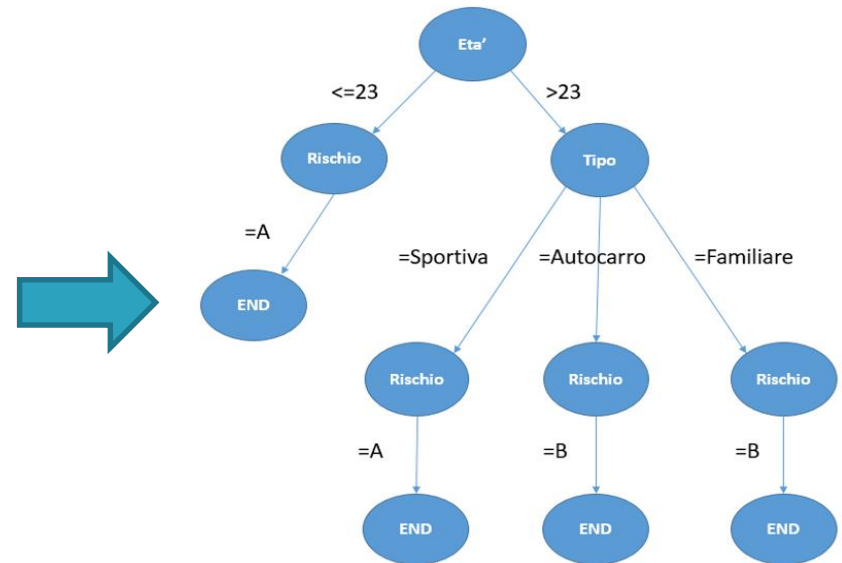
S sta per sportivo,
FA sta per familiare,
AU sta per autocarro

- La compagnia di assicurazioni vorrebbe una procedura automatica che sia in grado di segnalare se **un nuovo cliente** può essere rischioso oppure no

ESEMPIO: COMPAGNIA DI ASSICURAZIONI (2)

- L'esempio rappresenta proprio un tipico problema di classificazione e/o predizione che può essere risolto con la costruzione di un albero di decisione associato al training set di cui sopra

	ETÁ	TIPO DI AUTOVEICOLO	RISCHIO
Utente 1	17	S	A
Utente 2	43	FA	B
Utente 3	68	FA	B
Utente 4	32	AU	B
Utente 5	23	FA	A
Utente 6	18	FA	A
Utente 7	20	FA	A
Utente 8	45	S	A
Utente 9	50	AU	B
Utente 10	64	AU	A
Utente 11	46	FA	B
Utente 12	40	FA	B



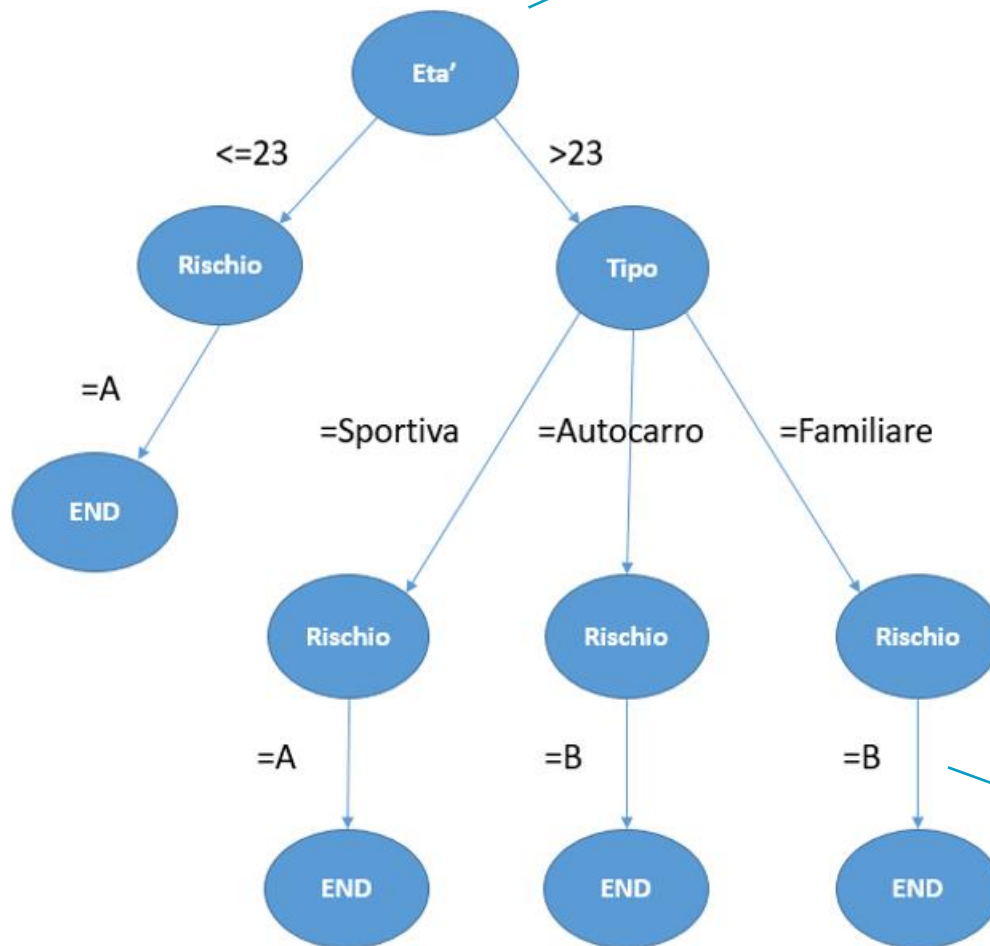
STRUTTURA DI UN ALBERO DI DECISIONE

ogni nodo rappresenta una variabile (escluso END)

due tipi di variabili:

- Quantitative ($<$, $>$, \leq , \geq , $=$, \neq)
- Categorie ($=$, \neq)

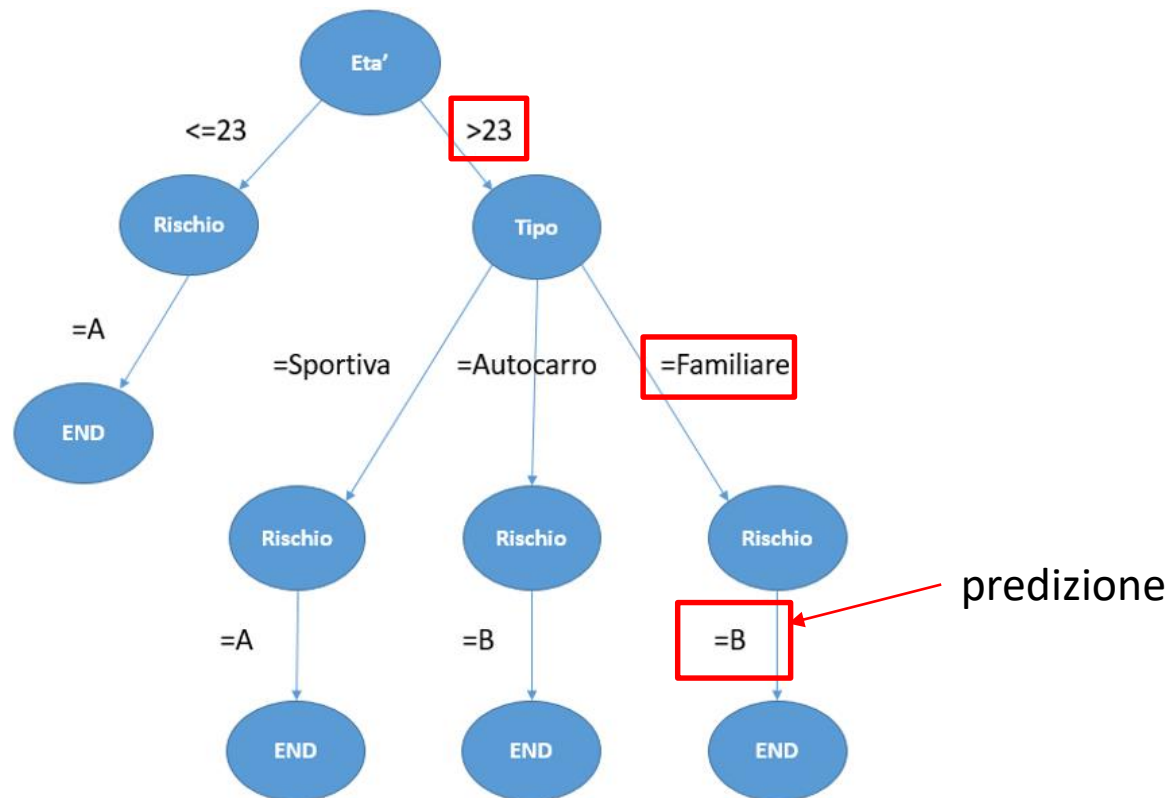
un arco verso un nodo figlio rappresenta un possibile valore (o un insieme di valori) per quella variabile



l'arco che precede la foglia rappresenta il valore predetto dall'albero di decisione a partire dai valori delle altre variabili

ESEMPIO DI PREDIZIONE

- Se abbiamo un cliente di 49 anni con una macchina familiare, il manager dell'assicurazione potrebbe inferire che il cliente presenta un rischio basso di incidente



SCOPO DEL LABORATORIO 10



- Realizzare un programma C++ che sia in grado di:
 - 1) leggere da file un albero di decisione e modificarlo con le operazioni cancella nodo, aggiungi nodo, modifica nodo;
 - 2) visualizzare **in modo testuale** l'albero di decisione;
 - 3) inferire e visualizzare le variabili dell'albero di decisione (nell'esempio in figura il programma dovrebbe visualizzare Età, Rischio e Tipo);
 - 4) effettuare una predizione a partire da un albero di decisione precedentemente inserito. In particolare il programma dovrebbe chiedere all'utente, uno alla volta, i valori da associare alle variabili durante il path che porta alla predizione;
 - 5) effettuare una predizione a partire da un albero di decisione precedentemente inserito e da un insieme di valori delle variabili. Il programma deve chiedere un insieme di coppie (variabile, valore)

SEMPLICE INTERFACCIA UTENTE (O MENÙ)

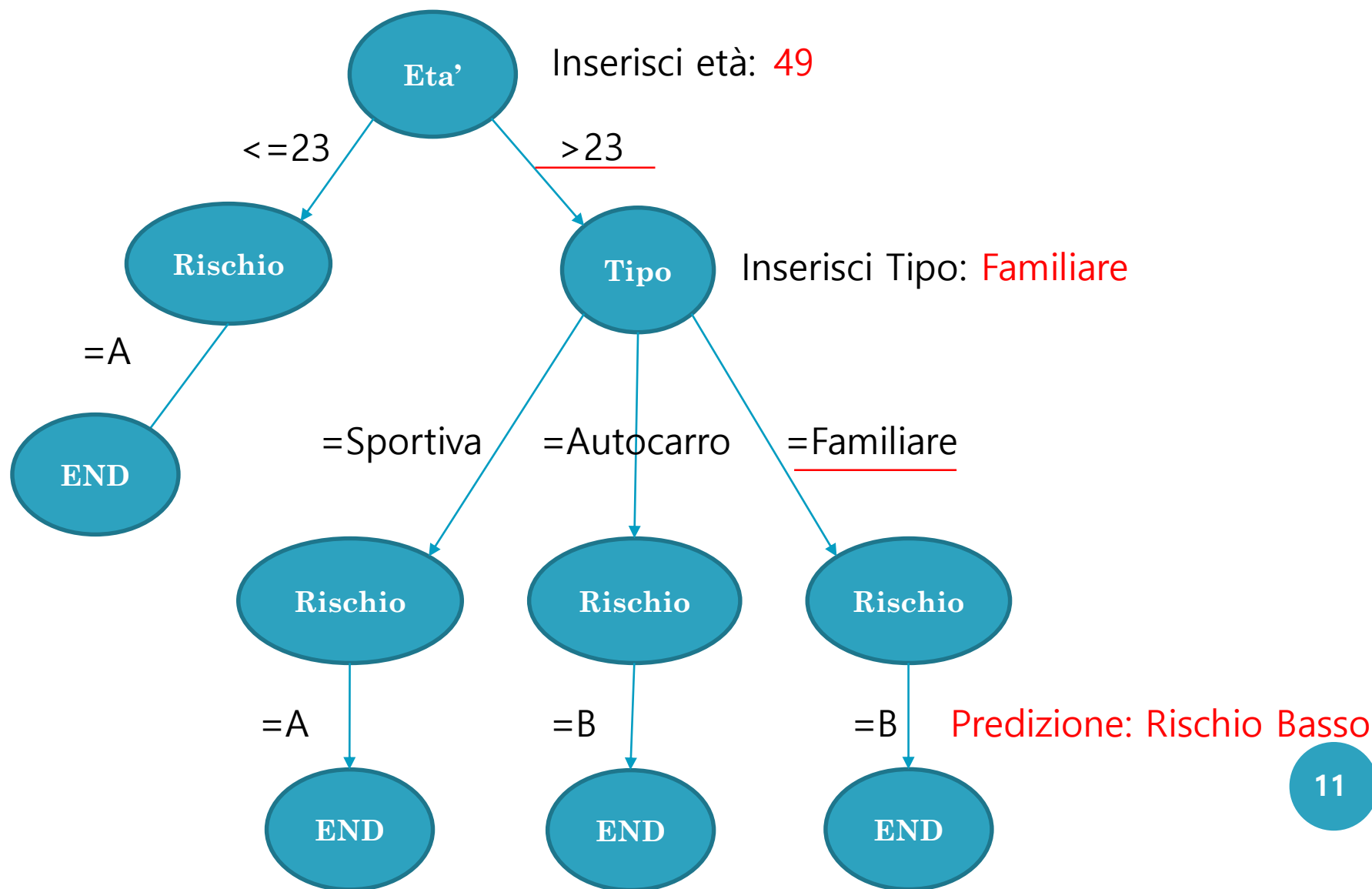
MENU

1. Lettura albero di decisione da file
2. Inserimento di un nodo etichettato labelFiglio attaccato a un padre etichettato labelPadre
3. Cancellazione di un nodo dall'albero;
4. Modifica di un nodo dall'albero
5. Visualizzazione dell'albero di decisione
6. Stampa variabili dell'albero di decisione
7. Effettua predizione inserendo i valori uno alla volta
8. Effettua predizione inserendo tutti i valori all'inizio
0. Uscita

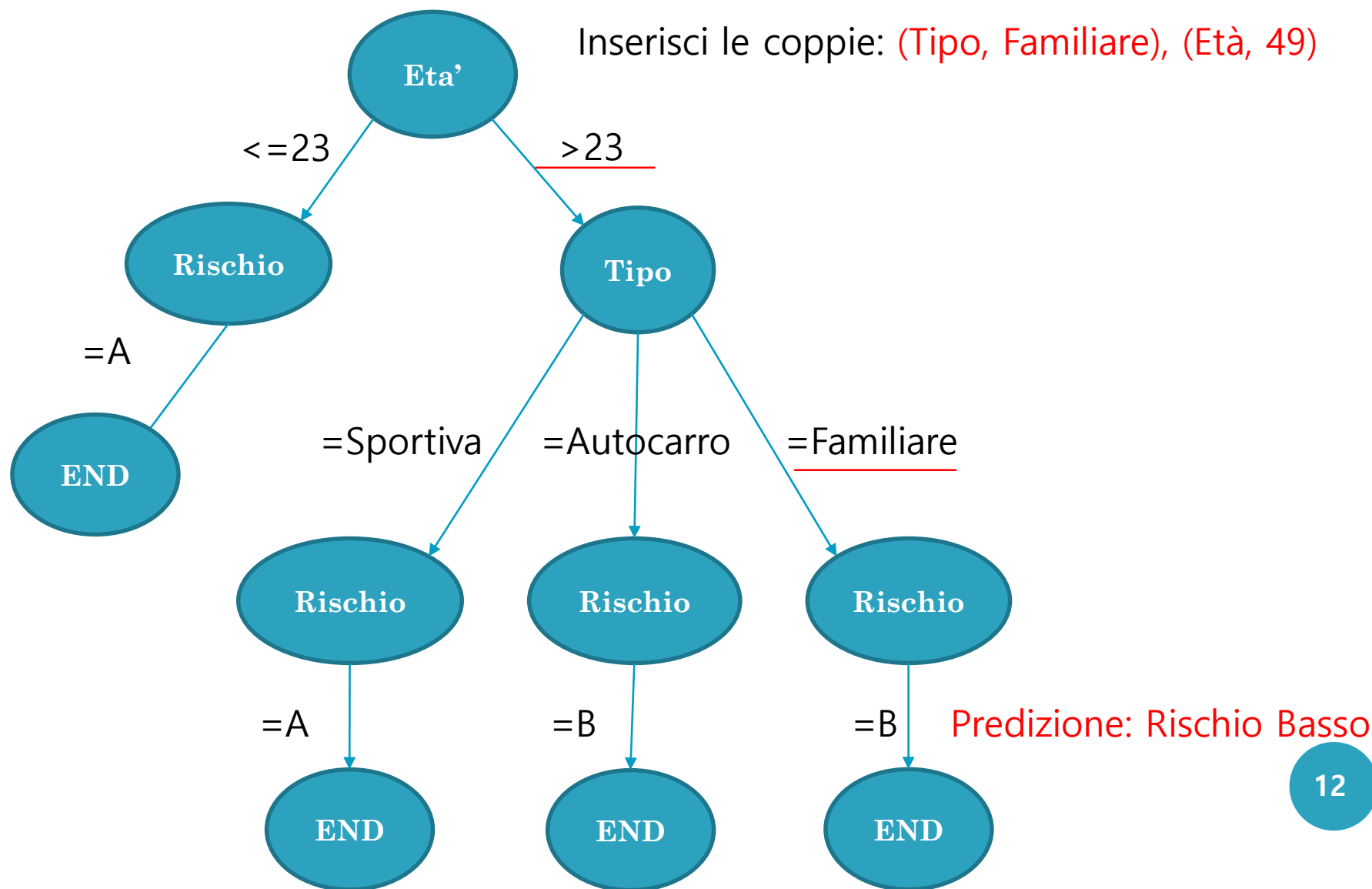
Fornisci la tua scelta --->

Non è richiesto che il menù sia identico a quello mostrato

EFFETTUA PREDIZIONE (INSERENDO I VALORI UNO ALLA VOLTA)



EFFETTUA PREDIZIONE (INSERENDO TUTTI I VALORI ALL'INIZIO)



‘CASI PARTICOLARI’

- Potrebbero esistere casi in cui si ha più di una condizione vera sugli archi
 - ad esempio se abbiamo due condizioni $a=5$ e $a>4$ sui due archi e il valore di ‘a’ è uguale a ‘5’
- oppure casi in cui nessuna condizione è vera
 - ad esempio quando abbiamo le condizioni $a=5$ e $a>5$ sugli unici due archi e il valore di ‘a’ è ‘4’
- Nel primo caso il programma dovrà scegliere in modo casuale uno degli archi che hanno la condizione vera
- Nel secondo caso il programma dovrà stampare il testo: “la predizione non può avere luogo in quanto esiste un nodo per il quale non c’è un arco percorribile”

LETTURA DI UN ALBERO DI DECISIONE DA FILE

- Il formato del file di input è molto simile a quello usato a lezione per gli **alberi generici**
- Ovvero:
 - prima riga del file deve contenere l'etichetta della radice
 - le righe seguenti devono contenere come prima etichetta quella di un nodo (che deve già essere stato elencato prima) seguita dalle coppie etichetta di uno dei suoi figli e la corrispondente l'etichetta che rappresenta la condizione dell'arco

radice

radice nodo1 cond1 nodo2 cond2 nodo3 cond3

nodo1 nodo4 cond4 nodo5 cond5 nodo6 cond6

nodo2 nodo7 cond7 nodo8 cond8 nodo9 cond9

ESEMPIO DI FILE DI INPUT

Età_1

Età_1 Rischio_1 <=23 Tipo_1 >23

Rischio_1 END_1 =A

Tipo_1 Rischio_2 =Sportiva

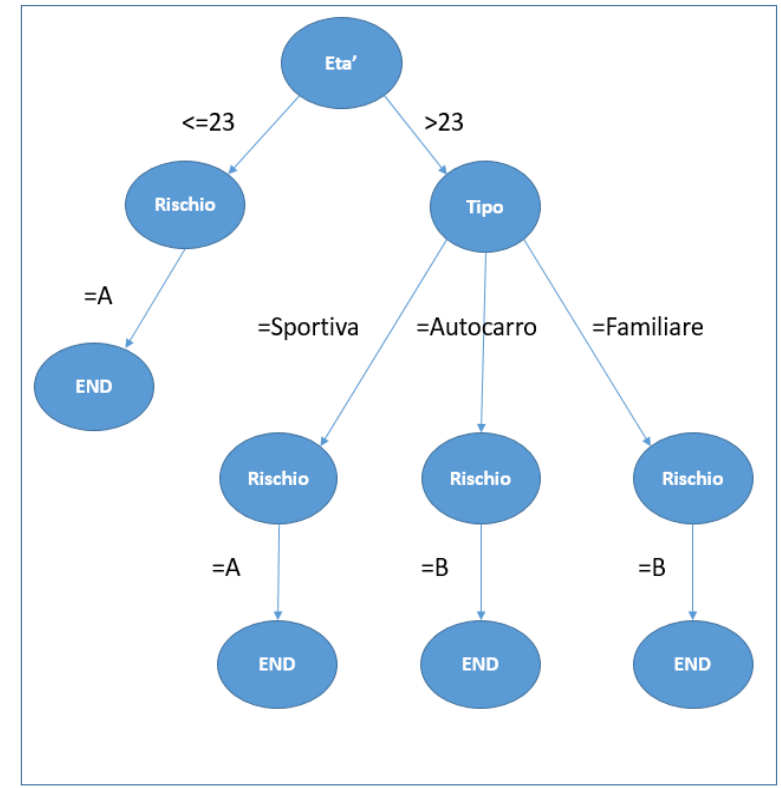
Rischio_3 =Autocarro Rischio_4

=Familiare

Rischio_2 END_2 =A

Rischio_3 END_3 =B

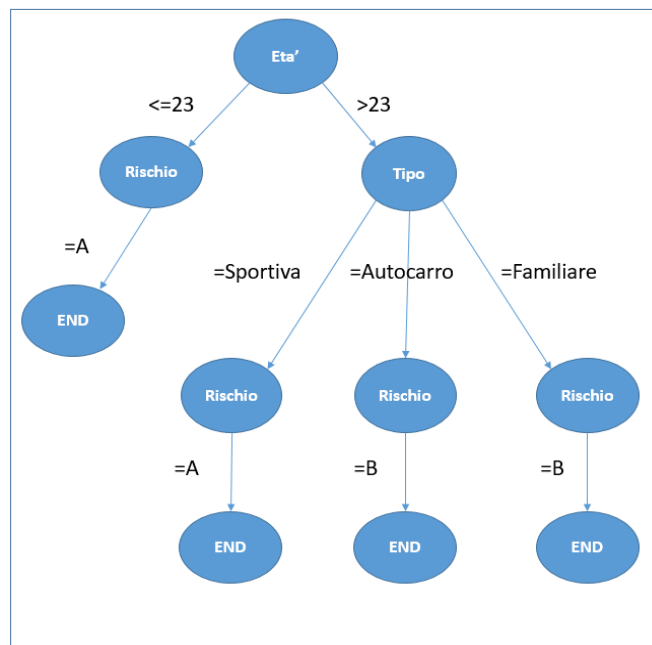
Rischio_4 END_4 =B



- le etichette dei nodi e le condizioni sono stringhe che non contengono il carattere di spazio
- le condizioni sono stringhe che iniziano con un'operatore relazionale tra i seguenti {=, ≠, <, >, <=, >=} seguito da un valore di una variabile
- underscore seguito da un numero serve per disambiguare le label

STAMPA DI UN ALBERO DI DECISIONE

- La stampa dovrà contenere **‘abbastanza informazione’** da poter permettere ad un utente del vostro software di ricostruire l'albero di decisione (ad esempio usando penna e carta)



```
eta_1
--(tipo_1, >23)
----(rischio_5, =Familiare)
----- (end_5, =B)
----(rischio_4, =Autocarro)
----- (end_4, =B)
----(rischio_3, =Sportiva)
----- (end_3, =A)
--(rischio_2, <23)
---- (end_2, =A)
--(rischio_1, =23)
---- (end_1, =A)
```


CONSIGLI E VINCOLI

- **Non è consentito utilizzare librerie specifiche per gli alberi o per i grafi** (ad esempio, The Boost Graph Library - BGL) così come non è consentito utilizzare i Vector
- Si consiglia di riusare il più possibile il codice sviluppato e testato durante l'anno

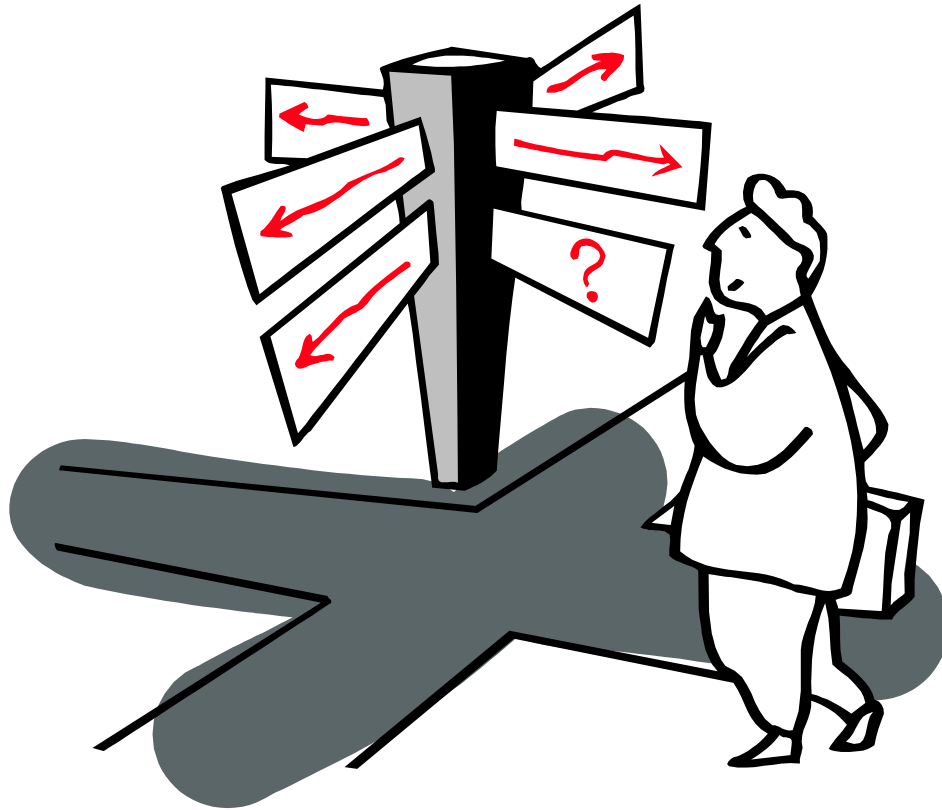
CONSEGNA

- Si richiede di consegnare i file .cpp e .h in un unico file zip con il formato **CognomePrimaLetteraNome.zip**
 - ad esempio RiccaF.zip
- Il laboratorio deve essere svolto e consegnato in modo individuale!
- Data di consegna: **24 Giugno 2018**

PUNTI E VALUTAZIONE

- Il punteggio massimo di **2.5 punti** si raggiunge non solo se le funzioni e le strutture dati sono corrette, ma anche se sono implementate in modo efficiente e **non ci sono problemi “di stile”**
- In particolare, sono elementi apprezzati ai fini della valutazione (l'elenco non è esaustivo):
 - Struttura del progetto in: main, header files (.h), code files (.cpp)
 - corretta indentazione del codice
 - identificatori significativi
 - introduzione di funzioni ausiliarie quando appropriato
 - commenti significativi

THE END ...



Domande?