



Università degli Studi di Milano-Bicocca

Dipartimento di Informatica, Sistemistica e Comunicazione

Corso di Laurea Magistrale in Data Science

Individuazione dei predittori più significativi del dropout sportivo in età adolescenziale

Relatore: Prof. Matteo PELAGATTI

Co-relatore: Prof.ssa Lucia DALLA PELLEGRINA

Tesi di Laurea Magistrale di:

Mauro Samarelli

Matricola 834196

Anno Accademico 2019-2020

SOMMARIO

Introduzione	5
Parte prima: CENNI INTRODUTTIVI	
Capitolo I: LO SPORT E IL DROPOUT SPORTIVO	
1. Dropout, burnout, engagement e Self-Determination Theory	6
2. Ruolo dello sport nell'infanzia e nell'adolescenza	8
3. Ruolo del senso di appartenenza alla comunità sportiva	11
Parte seconda: ANALISI STATISTICA	
Capitolo I: QUESTIONARIO PER RACCOLTA DATI DI ORANGOGO	
1. Descrizione delle macro-sezioni	13
2. Descrizione delle tipologie di domanda	17
3. Somministrazione e raccolta dati	20
Capitolo II: ANALISI DATI	
1. Fase di pre-processing	21
2. Analisi statistica descrittiva	24
3. Test del Chi-Quadro	35
Parte terza: INDIVIDUAZIONE DEI PREDITTORI	
Capitolo I: DATA PREPARATION	
1. Gestione dei missing values	38
2. Encoding delle variabili	47
Capitolo II: FEATURE SELECTION	
1. Decision Tree Learning	50
2. Feature Importance	56
3. Feature Selection tramite Random Forest	61

Parte quarta: ALGORITMI DI CLASSIFICAZIONE

Capitolo I: MODELLI PREDITTIVI

1. Cross-Validation con tuning dei parametri	66
2. Logistic Regression	68
3. Support Vector Machine	72
4. K-Nearest Neighbors	76
5. Multi-Layer Perceptron	78
6. Voting Classifier	82

Capitolo II: VALUTAZIONE PERFORMANCE

1. Metriche di performance	84
2. Confronto tra i modelli	89

Parte quinta: ANALISI DELLA CAUSALITA'

Capitolo I: TWO STAGE LEAST SQUARES

1. Variabili strumentali	96
2. Effetto della distanza da casa alla struttura sportiva	99

Parte sesta: QUESTIONARIO PREDITTIVO

Capitolo I: COSTRUZIONE DEL QUESTIONARIO

1. Descrizione delle domande	104
2. Casi di test	101

Conclusione	110
-------------	-----

Appendice A: Questionario per raccolta dati	113
---	-----

Appendice B: Questionario predittivo	152
--------------------------------------	-----

Bibliografia	163
--------------	-----

Ringraziamenti	165
----------------	-----

Introduzione

Il tema di questa tesi è la predizione del dropout sportivo dei minori, nella fascia di età dai 9 ai 17 anni, effettuata assegnando una probabilità di abbandono rispetto a dati anagrafici e una serie di informazioni relative a educazione familiare, sport praticato/i e frequenza, supporto dei genitori, rapporto con i compagni, rapporto con l'allenatore, autoefficacia e gestione del match, ricavati tramite un questionario predittivo.

Il questionario predittivo è il risultato del lavoro di analisi condotto in collaborazione con Orangogo, una startup innovativa a vocazione sociale conosciuta per aver dato vita al primo motore di ricerca degli sport in Italia.

Orangogo ha messo a disposizione i dati relativi a circa 3.000 questionari sottoposti a ragazzi nella fascia di età tra i 18 e i 20 anni, con lo scopo di costruire un questionario ad-hoc per predire, con una determinata probabilità, il rischio di abbandono sportivo negli adolescenti.

Il fine ultimo del questionario predittivo è individuare i fattori che maggiormente incidono sul rischio di abbandono e fornire supporto per prevenire che l'abbandono si realizzi effettivamente.

La prima parte della tesi introduce il ruolo dello sport nella società, con particolare enfasi sul periodo adolescenziale, il quale risulta essere molto critico per l'abbandono sportivo.

La seconda parte riguarda l'analisi statistica descrittiva del dataset contenente i dati dei questionari sottoposti a ragazzi di età compresa tra i 18 e i 20 anni, con lo scopo di derivare un quadro generale sui fattori di incidenza sul dropout.

La terza parte riguarda l'individuazione dei predittori più significativi del dropout sportivo sulla base dei dati a disposizione, mediante il task di feature selection sul dataset ottenuto dalle fasi di pre-processing e di data preparation.

Nella quarta parte vengono utilizzati e confrontati algoritmi predittivi di classificazione, con lo scopo di fornire una determinata probabilità di abbandono per nuovi questionari.

La quinta parte riguarda l'analisi della causalità delle variabili disponibili, effettuata attraverso tecniche tradizionali di statistica ed econometria, con lo scopo di individuare interessanti relazioni, ponendo particolare attenzione alle tematiche geografiche/spaziali.

La sesta parte descrive la costruzione del questionario predittivo dell'abbandono sportivo, avente come target i ragazzi tra i 9 e i 17 anni, sulla base dei modelli sviluppati e dei risultati ottenuti dalla fase di individuazione dei predittori più significativi.

CENNI INTRODUTTIVI

Capitolo Primo

Lo sport e il dropout sportivo

1. Dropout, burnout, engagement e Self-Determination Theory

Il *dropout sportivo* è definito "assenza prolungata di pratica sistematica e competizione, in uno sport (abbandono specifico dello sport) o in tutti gli sport (abbandono generale dello sport)".

Tenendo conto della sistematicità della pratica, non sono da considerare casi di dropout gli individui che abbandonano lo sport per una motivazione di forza maggiore, salvo poi riprendere la pratica sportiva nel momento in cui termina l'impedimento temporaneo.

In base ai fattori che determinano l'interruzione prolungata della pratica sportiva, all'interno dei casi di dropout è possibile rilevare casi di burnout.

Il *burnout sportivo* è definito "abbandono della pratica sportiva collegato a malessere psicofisico dell'atleta su cui incidono *vincoli intrapersonali* e *vincoli interpersonali*", quali stress derivante dalle esigenze associate all'allenamento e alla competizione, ridotto senso di realizzazione derivante dal sentimento di inefficacia e dalla tendenza di valutare negativamente la prestazione sportiva, svalutazione del ruolo dello sport derivante dalla mancanza di consapevolezza dei vantaggi di praticarlo ¹.

Tenendo conto di tale specifica, non sono da considerare casi di burnout gli individui che abbandonano lo sport per *vincoli strutturali* quali mancanza e inadeguatezza delle strutture sportive, costo economico, infortuni o disabilità permanenti.

In opposizione al burnout, l'*engagement* è definito come il "coinvolgimento nella pratica sportiva collegato al benessere psicofisico dell'atleta", riscontrabile da vigore (stato di resistenza mentale), dedizione (stato di alto coinvolgimento che si traduce in sentimenti di entusiasmo, ispirazione e orgoglio per il compito), assorbimento (stato di profonda concentrazione e attenzione alle attività) ¹.

¹ S. Scotto di Luzio, G. Martinent, E. Guillet-Descas, M. P. Daigle (2019), *Exploring the Role of Sport Sense of Community in Perceived Athlete Burnout, Sport Motivation, and Engagement*, Journal of Applied Sport Psychology

E' opportuno specificare che burnout ed engagement sono concetti indipendenti e distinti, ovvero l'assenza di gradi di burnout non indica necessariamente la presenza di gradi di engagement e viceversa.

La *Self-Determination Theory* (teoria delle motivazioni) interviene a supporto dell'analisi delle cause della riduzione di engagement e dell'incremento di burnout in un'atleta.

Questa teoria afferma che le forme di *motivazione autonoma* sono regolate da *processi interni di autodeterminazione*, mentre le forme di *motivazione controllata* sono regolate da *processi esterni di non autodeterminazione*.

Le motivazioni per cui un'atleta pratica o non pratica sport si classificano in: motivazioni intrinseche, motivazioni estrinseche e amotivation.

Le *motivazioni intrinseche* si riferiscono alla partecipazione sportiva per interesse o piacere nell'attività, risultato di processi con alta autodeterminazione.

Le *motivazioni estrinseche* sono basate su *regole integrate* (considerazione dell'attività in linea con i valori profondamente radicati nella persona, quindi processi con alta autodeterminazione), *regole identificate* (percezione del beneficio dello sport per la persona, quindi processi con alta autodeterminazione), *regole introiettate* (partecipazione all'attività per migliorare autostima e diminuire le difficoltà nei rapporti sociali, quindi processi con bassa autodeterminazione) e *regole esterne* (partecipazione all'attività per soddisfare richieste esterne, quindi processi con bassa autodeterminazione).

L'*amotivation* è definita come la mancanza di intenzione alla partecipazione nell'attività sportiva.

I risultati di diverse ricerche scientifiche affermano che il burnout è correlato negativamente con le forme di motivazione determinate da processi con alta autodeterminazione, mentre è correlato positivamente con le forme di motivazione determinate da processi con bassa autodeterminazione ².

² H. K. Larson, K. McFadden, T. F. McHugh, T. R. Berry, W. M. Rodgers (2018), *When you don't get what you want—and it's really hard: Exploring motivational contributions to exercise dropout*, Psychology of Sport & Exercise

2. Ruolo dello sport nell'infanzia e nell'adolescenza

La partecipazione sportiva durante l'infanzia e l'adolescenza ha importanti benefici per la salute fisica, psicologica e sociale.

Tra questi si riscontrano tassi più elevati di idoneità fisica, maggiore coinvolgimento nell'attività fisica nel tempo, tassi più bassi di comportamento sedentario, tassi più elevati di autostima, interazioni sociali più positive, diminuzione dei livelli di sintomi depressivi e tassi più bassi di obesità.

I benefici per la salute della partecipazione sportiva vanno oltre quelli attribuibili all'attività fisica e possono essere sostenuti dalla natura sociale della partecipazione sportiva.

Ciò è coerente con i maggiori benefici per la salute associati agli sport di squadra rispetto agli sport individuali ³.

D'altro canto, la pratica sportiva ad alti livelli durante l'adolescenza può essere un fattore negativo per il benessere psicofisico di un atleta, in termini di ansia da prestazione e infortuni frequenti o traumatici.

Gli adolescenti possono essere frequentemente esposti a crescenti richieste di allenamento, stress competitivo e pressioni sulla prestazione, con un costo considerevole in termini di energie mentali e fisiche.

Il verificarsi di suddette situazioni compromette l'engagement dell'atleta e aumenta il rischio di burnout.

Le azioni da intraprendere per ridurre il rischio di burnout e incrementare l'engagement si basano sulla ridefinizione degli obiettivi dell'attività sportiva: spostare lo scopo dalla vittoria verso il divertimento, equilibrare il coinvolgimento dei genitori, incoraggiare la pratica di più di uno sport, abilitare i bambini e gli adolescenti ad avere la propria autonomia durante le esperienze sportive, promuovere regole che permettono a tutti i bambini e adolescenti di avere la chance di gareggiare, diminuire la pressione della vittoria da parte dei genitori, sollecitare i genitori a non riflettere i loro sogni sportivi sui bambini e iniziare la pratica sportiva agonistica ad un'età appropriata ³.

³ P. A. Witt, T. B. Dangi (2018), *Why Children/Youth Drop Out of Sports*, Journal of Park and Recreation Administration

A tal proposito, il *Developmental Model of Sport Participation (DMSP)* di Cote e Hancock (2014) riveste particolare importanza a livello di letteratura ⁴.

Questo modello ha l'obiettivo di definire delle linee guida in grado di promuovere il benessere psicofisico del bambino o dell'adolescente impegnato nella pratica sportiva, riducendo il rischio di burnout.

Ecco nello specifico i punti fondamentali del DMSP attribuiti per il sano sviluppo a livello sportivo fino all'età di 13 anni, con particolare riferimento alla gestione delle attività da parte delle società sportive e degli allenatori:

- a) regolare la durata della stagione sportiva a 3 o 4 mesi, con durata massima 6 mesi;
- b) limitare viaggi lunghi per competizioni organizzate;
- c) introdurre programmi di base che mirano alla potenziale pratica di più sport;
- d) non sviluppare processi selettivi per gli atleti "più talentuosi" fino all'età di specializzazione;
- e) offrire opportunità competitive sane, senza dare troppa importanza alla vittoria e a competizioni a lungo termine;
- f) scoraggiare la specializzazione in un solo sport;
- g) permettere di provare tutti i ruoli disponibili nello sport praticato;
- h) promuovere l'interpretazione personale all'interno di un gioco organizzato;
- i) progettare situazioni in cui prevale il divertimento e l'ottenimento di un riconoscimento a breve termine;
- j) comprendere i bisogni dell'atleta.

Il DMSP prevede tre stage nella formazione dell'atleta:

- 1) *sampling years*: fascia 6-12 anni, in cui si incoraggia la partecipazione a più sport
- 2) *specializing years*: fascia 13-15 anni, in cui si riduce il numero di sport praticati
- 3) *investment years*: fascia 16+ anni, in cui si investe nello sport più significativo

La ricerca scientifica "*Socio-Ecological Predictors of Participation and Dropout in Organised Sports during Childhood*" di Vella, Cliff e Okely (2014) prende in esame circa 4.000 bambini australiani che praticano sport con età media di 8.25 anni, con lo scopo di individuare i predittori più significativi dell'abbandono entro il decimo anno di età, all'interno di un modello contenente variabili socio-ecologiche.

⁴ J. Cotè, D. J. Hancock (2014), *Evidence -based policies for youth sport programmes*, International Journal of Sport Policy

I risultati affermano che i predittori più significativi della partecipazione sportiva sono: sesso maschile, meno persone in casa, reddito familiare più elevato, inglese come lingua principale parlata in famiglia, educazione parentale superiore, bambino frequentemente accompagnato ad un evento sportivo e accesso ad un insegnante di educazione specialistica durante la scuola primaria.

I predittori più significativi del dropout, invece, sono: minore reddito familiare, lingua principale parlata a casa diversa dall'inglese, educazione parentale inferiore, bambino non accompagnato frequentemente ad un evento sportivo.

La ricerca scientifica *“Factors and actions to prevent dropout in youth sport: results of a rapid review and expert opinions”* di Murphy, Woods, Dijk e Seghers (2018) ha permesso di ottenere uno schema concettuale per sintetizzare l'effetto causa-conseguenza di diversi fattori rispetto al dropout sportivo.

I fattori presi in considerazione sono distinti in *fattori demografici*, *fattori di predisposizione*, quali autonomia, competenza percepita, aspettative e valore dello sport, *fattori di rinforzo*, quali relazione con l'allenatore, relazione con i compagni e supporto dei genitori, e *fattori di abilitazione*, quali costo, accessibilità, tipologia di sport, vincoli di tempo.

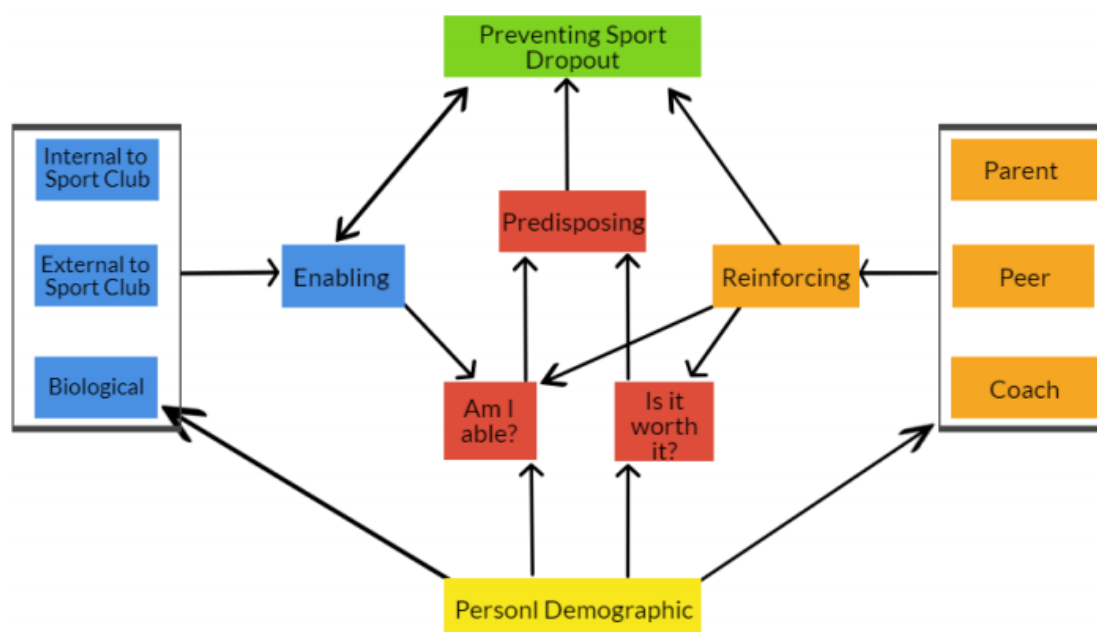


Fig. 1: Schema descrittivo dell'azione dei fattori incidenti sul dropout da *“Factors and actions to prevent dropout in youth sport: results of a rapid review and expert opinions”* di Murphy, Woods, Dijk e Seghers (2018)

3. Ruolo del senso di appartenenza alla comunità sportiva

Il *senso di appartenenza alla comunità sportiva* è definito “sentimento che hanno gli sportivi di contare l’uno per l’altro e per il gruppo e la fiducia condivisa sul fatto che i bisogni dei membri saranno soddisfatti attraverso il loro impegno ad essere un insieme”.

Sentirsi parte della comunità sportiva significa condividere i seguenti aspetti:

- a) *membership*: sensazione di far parte di una comunità avente chiari confini fisici o simbolici;
- b) *opportunità per influenza*: partecipazione attiva con la percezione di essere in grado di influenzare la comunità nello stesso tempo in cui si dipende da questa;
- c) *soddisfazione dei bisogni*: possibilità dello scambio di risorse per soddisfare esigenze personali e collettive;
- d) *connessione emotiva con la comunità*: sensazione di condividere una storia comune, credenze e valori, che rafforzano il legame tra i membri;
- e) *connessione emotiva con i compagni*: dimensione specifica per gli adolescenti che consiste nell’importanza di condividere lo stesso stato emotivo con i coetanei.

Il senso di appartenenza alla comunità sportiva e lo stato motivazionale vengono influenzati dall’ambiente sportivo e dalle relazioni con compagni e allenatore, rendendo il suo ruolo predittivo rispetto a burnout ed engagement maggiormente collegato al significato dei processi con bassa autodeterminazione.

Migliorare il “contorno” che vive l’atleta durante la pratica sportiva permette di affievolire l’incidenza potenzialmente negativa delle motivazioni controllate sull’engagement.

La ricerca scientifica “*Exploring the Role of Sport Sense of Community in Perceived Athlete Burnout, Sport Motivation, and Engagement*” di Scotto, Martinent, Guillet-Descas e Daigle (2019) prende in esame un campione di 250 atleti adolescenti francesi con livello competitivo nazionale aventi 8 anni in media di pratica e 13.5 ore settimanali in media di allenamento.

I risultati ottenuti evidenziano che il gruppo con un alto senso di appartenenza alla comunità sportiva ha un punteggio più basso di burnout rispetto al gruppo con un basso senso di appartenenza alla comunità sportiva.

Inoltre, il gruppo con un alto senso di appartenenza alla comunità sportiva ha un punteggio più alto di engagement rispetto al gruppo con un basso senso di appartenenza alla comunità sportiva.

Per quanto riguarda lo stato motivazionale, il senso di appartenenza alla comunità sportiva incide negativamente su motivazione controllata e non autodeterminata, mentre incide positivamente su motivazione autonoma e autodeterminata.

Inaspettatamente, il senso di appartenenza alla comunità sportiva incide positivamente sul ridotto senso di realizzazione e sulle regole introiettate delle motivazioni estrinseche, a distanza di 6 mesi.

Ciò potrebbe essere spiegato dalla possibilità che un senso di appartenenza alla comunità troppo forte può ridurre l'autonomia degli atleti, con conseguente sensazione di inefficacia delle prestazioni sportive.

Inaspettatamente, la connessione emotiva con i compagni incide positivamente sullo stress psicofisico, a distanza di 6 mesi.

Ciò potrebbe essere spiegato dal fatto che gli atleti adolescenti che si allenano in centri di formazione intensiva, e spesso sono lontani dalle loro famiglie, sono così tanto attaccati ai loro compagni al punto di perdere la loro autonomia durante la stagione sportiva, risultando anche più vulnerabili alla pressione derivante da problemi sportivi, scolastici e relazionali, derivandone stress psicofisico.

ANALISI STATISTICA

Capitolo Primo

Questionario per raccolta dati di Orangogo

1. Descrizione delle macro-sezioni

Il lavoro di ricerca e raccolta dati svolto da Orangogo si è basato sulla somministrazione di un questionario relativo al tema dell'abbandono sportivo agli studenti del quinto anno di scuola media superiore, con età compresa tra i 18 e 20 anni, distribuiti su territorio nazionale, per i quali è stato garantito l'anonimato.

E' opportuno specificare che le domande presenti nel questionario fanno riferimento al periodo pre-adolescenziale e adolescenziale (9-17 anni), in modo da poter utilizzare successivamente i dati a disposizione con fine predittivo.

Di seguito vengono descritte le macro-sezioni di domande che compongono il questionario.

La prima sezione riguarda i *dati anagrafici* ovvero sesso, anno di nascita, comune e provincia di residenza, nazionalità, tipologia di scuola superiore frequentata e informazioni su eventuali disabilità (temporanee o permanenti).

La seconda sezione riguarda i *dati generici* riferiti al contesto familiare ovvero titolo di studio di madre e padre, numero di fratelli e sorelle, numero di nonni che abitano vicino casa, figura familiare di riferimento dopo la scuola, eventuale cambio di residenza e le distanze temporali approssimative da casa e da scuola verso la struttura sportiva più vicina sia a piedi sia con i mezzi.

Terminate le prime due macro-sezioni, il questionario prevede la domanda "*Tra i 9 e i 17 anni ... Hai mai praticato sport?*" con possibilità di scelta di risposta tra *Si* e *No*.

Per coloro che affermano di non aver mai praticato sport nel periodo pre-adolescenziale e adolescenziale, il questionario prevede una serie di domande per analizzare la natura della sedentarietà, contesto diverso rispetto a quello collegato al fenomeno del dropout che presuppone la presenza di pratica sportiva.

Per coloro che affermano di aver praticato sport tra i 9 e i 17 anni, il questionario prevede macro-sezioni successive per analizzare l'eventuale fenomeno di dropout.

La terza sezione riguarda le *informazioni sullo sport* ovvero l'eventuale presenza di campioni sportivi che vivono o si allenano in città, lo/gli sport praticato/i e in caso di più sport qual è quello più significativo, eventuale pratica agonistica e frequenza settimanale della pratica.

Specificatamente allo sport più significativo, viene chiesto se questo sia stato l'unico praticato o il più significativo tra quelli praticati e, inoltre, se si sia mai pensato di abbandonarlo o se si sia abbandonato, sempre riferendosi al periodo tra i 9 e i 17 anni.

A seconda della risposta, le domande del questionario proseguono assumendo una forma semantica diversa per ogni caso, pur riferendosi agli stessi fattori di analisi.

Il *primo caso* raggruppa chi ha praticato un solo sport e, a sua volta, viene distinto in due sotto-casi: chi ha pensato di abbandonarlo senza effettivamente farlo (*caso 1.1*) e chi non ha mai pensato di abbandonarlo (*caso 1.2*); per entrambi i sotto-casi ricaviamo la conclusione che *non* si è verificato dropout, tenendo conto che i casi 1.1 sono da considerare "più critici" rispetto ai casi 1.2.

Il *secondo caso* raggruppa chi ha praticato più di un solo sport e, a sua volta, viene distinto in due sotto-casi: chi ha pensato di abbandonare lo sport più significativo senza effettivamente farlo (*caso 2.1*) e chi non ha mai pensato di abbandonarlo (*caso 2.2*); per entrambi i sotto-casi ricaviamo la conclusione che *non* si è verificato dropout, tenendo conto che i casi 2.1 sono da considerare "più critici" rispetto ai casi 2.2.

Il *terzo caso* raggruppa chi ha abbandonato l'unico sport che praticava o chi ha abbandonato lo sport più significativo tra quelli praticati (*caso 3*), ricavando la conclusione che si è verificato dropout.

Il *quarto caso* raggruppa chi non ha mai praticato sport tra i 9 e i 17 anni (*caso 4*), per i quali non è possibile analizzare le dinamiche del fenomeno di dropout, bensì si valuta la natura della condizione di sedentarietà.

La quarta sezione riguarda il *momento critico* in età adolescenziale, associato al pensiero di abbandonare lo sport più significativo praticato, ovvero età in cui si è pensato di abbandonare, stato d'animo prima di andare ad allenarsi, frequenza settimanale degli allenamenti e ore settimanali di attività "extra" (es. social network).

Dato il riferimento al momento critico, questa sezione viene inclusa nei questionari relativi ai casi 1.1, 2.1 e 3 (chi ha pensato all'abbandono e chi ha abbandonato), mentre viene esclusa nei questionari relativi ai casi 1.2 e 2.2 (chi non ha mai pensato all'abbandono).

Nella quinta sezione, riguardante i *fattori di protezione* della pratica, viene chiesto di selezionare i *tre principali motivi* che spingevano il ragazzo a continuare la pratica sportiva, fornendo un elenco di motivazioni possibili (34 item) relative a questioni tecniche e personali, rapporti interpersonali e contesto socio-sportivo.

Nella sesta sezione, riguardante i *fattori di rischio* dell'abbandono, viene chiesto di selezionare i *tre principali motivi* che, invece, spingevano il ragazzo ad interrompere in via definitiva la pratica sportiva, fornendo un elenco di motivazioni possibili (25 item) relative a questioni tecniche, rapporti interpersonali, contesto sociale e sportivo.

La settima sezione riguarda l'*impegno scolastico* ovvero frequenza giornaliera del recarsi a scuola, riuscire a conciliare l'impegno scolastico con l'attività sportiva e ore giornaliere di studio fuori dell'orario scolastico.

L'ottava sezione riguarda gli aspetti relazionali con *società, allenatore e compagni di squadra* ovvero il sentimento di sentirsi parte integrante della società sportiva frequentata, attenzioni e incoraggiamenti ricevuti dall'allenatore, definizione del carattere percepito dell'allenatore, stimolo o cattivo comportamento dei compagni di squadra sul ragazzo e il grado di divertimento durante l'allenamento.

La nona sezione riguarda il *supporto dei genitori* ovvero tipologia di lavoro di madre e padre, impegno richiesto dal lavoro di madre e padre, influenza di fratelli e sorelle nella scelta dello sport, rapporto che hanno i genitori con lo sport e il supporto emotivo percepito dal ragazzo.

La decima sezione riguarda l'*autoefficacia* e la *gestione del match* ovvero riuscire ad affrontare l'allenamento con preoccupazioni, stanchezza oppure essendo indaffarati e quali sentimenti predominano dopo una vittoria o una sconfitta di una gara sportiva.

L'undicesima sezione riguarda considerazioni personali sul *ruolo dello sport* per se stessi: viene chiesto di selezionare le *tre definizioni personali più appropriate* rispetto al concetto di praticare sport tra un elenco di possibili definizioni (10 item), il quale comprende sia potenziali definizioni "positive" (divertimento, passione, modo per essere liberi ...) sia potenziali definizioni "negative" (sacrificio, dovere, perdita di tempo ...).

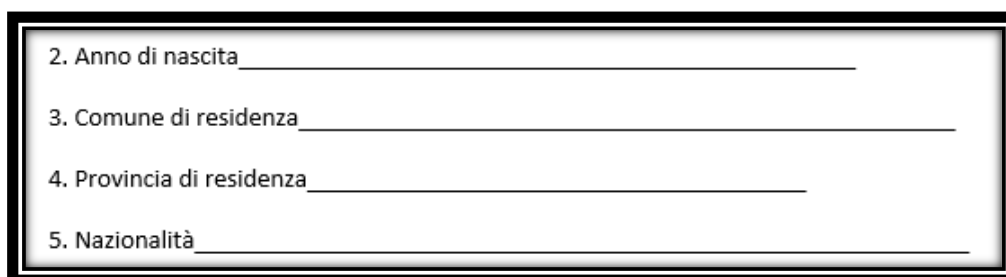
Infine, la dodicesima sezione riguarda le *caratteristiche della personalità*: viene chiesto al ragazzo quanto è d'accordo su un'affermazione che descrive se stesso in un determinato modo (affidabile, estroverso, ansioso, competitivo ...).

Le versioni complete dei questionari, differenziati per caso, sono allegate in *Appendice A*.

2. Descrizione delle tipologie di domanda

Di seguito vengono descritte le diverse tipologie di domanda utilizzate all'interno del questionario di raccolta dati, con lo scopo di introdurre le tecniche di processamento dei dati grezzi a disposizione (*Vedi Parte Seconda Capitolo 2*).

Le *domande a risposta aperta* consentono al ragazzo di fornire una risposta non soggetta a vincoli, ma allo stesso tempo rendono meno semplice la categorizzazione dei dati ottenuti.



2. Anno di nascita _____

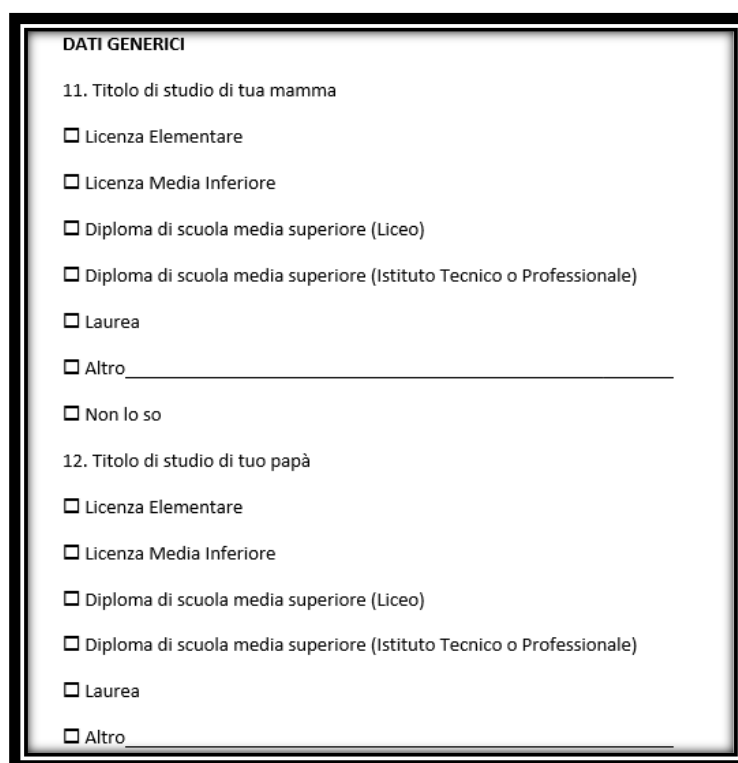
3. Comune di residenza _____

4. Provincia di residenza _____

5. Nazionalità _____

Fig. 2: Esempi di domande a risposta aperta presenti nel questionario di raccolta dati

Le *domande a risposta chiusa* consentono al ragazzo di selezionare *una sola* risposta tra una lista di quelle possibili, introducendo un restringimento degli scenari comunque contenuto se si fornisce l'opzione di risposta "*Altro*", ma allo stesso tempo favorendo la categorizzazione dei dati.



DATI GENERICI

11. Titolo di studio di tua mamma

- ☐ Licenza Elementare
- ☐ Licenza Media Inferiore
- ☐ Diploma di scuola media superiore (Liceo)
- ☐ Diploma di scuola media superiore (Istituto Tecnico o Professionale)
- ☐ Laurea
- ☐ Altro _____
- ☐ Non lo so

12. Titolo di studio di tuo papà

- ☐ Licenza Elementare
- ☐ Licenza Media Inferiore
- ☐ Diploma di scuola media superiore (Liceo)
- ☐ Diploma di scuola media superiore (Istituto Tecnico o Professionale)
- ☐ Laurea
- ☐ Altro _____

Fig. 3: Esempi di domande a risposta chiusa con una sola risposta selezionabile

Le *domande a risposta multipla* consentono al ragazzo di selezionare *massimo tre risposte* tra una lista di possibili, generalmente riferite al concetto di risposte “più appropriate” o “principali” secondo la propria esperienza.

Per questa tipologia di quesiti, la categorizzazione dei dati non può essere adottata con una singola variabile, bensì si rende necessario l'utilizzo di *N variabili booleane (Si/No)*, dove N è il numero di risposte presente nella lista di quelle selezionabili.

35. Quali fattori, invece, ti spingevano ad abbandonare? Tra le seguenti motivazioni, scegline al massimo 3.

Questioni tecniche	<input type="checkbox"/>	1. non mi divertivo più
	<input type="checkbox"/>	2. non mi sentivo abbastanza bravo
	<input type="checkbox"/>	3. troppa competizione
	<input type="checkbox"/>	4. infortunio/i sportivo/i
	<input type="checkbox"/>	5. eccessiva fatica
	<input type="checkbox"/>	6. incomprensioni con l'allenatore
	<input type="checkbox"/>	7. motivi di salute
	<input type="checkbox"/>	8. periodo critico nella vita personale
	<input type="checkbox"/>	9. sport troppo impegnativo
	<input type="checkbox"/>	10. Disabilità temporanea o permanente
Rapporti interpersonali:	<input type="checkbox"/>	11. non ero appoggiato dai miei genitori
	<input type="checkbox"/>	12. i miei genitori non mi potevano accompagnare
	<input type="checkbox"/>	13. rapporto difficile tra genitori e allenatore
	<input type="checkbox"/>	14. non mi trovavo bene con i compagni
	<input type="checkbox"/>	15. non avevo qualcuno con cui allenarmi
	<input type="checkbox"/>	16. sport troppo lontano
Contesto sociale:	<input type="checkbox"/>	17. non andavo bene a scuola
	<input type="checkbox"/>	18. non riuscivo a conciliare lo sport con l'impegno scolastico
	<input type="checkbox"/>	19. avevo altri interessi prevalenti
Contesto sportivo:	<input type="checkbox"/>	20. difficoltà a raggiungere l'impianto
	<input type="checkbox"/>	21. lo sport era troppo costoso
	<input type="checkbox"/>	22. palestra non adeguata
	<input type="checkbox"/>	23. materiali a disposizione in palestra carenti
	<input type="checkbox"/>	24. orari scomodi
	<input type="checkbox"/>	25. non ci sono sport che mi piacciono dove vivo io

Fig. 4: Esempio di domanda a risposta multipla con massimo tre risposte selezionabili

Le *domande a risposta su scala Likert* consentono al ragazzo di indicare il grado con il quale risulta essere d'accordo con l'affermazione proposta nel quesito, espresso tramite una scala che varia da 1 (*completamente in disaccordo*) a 5 o 7 (*completamente in accordo*).

Si tratta di quesiti che, oltre ad avere un insieme finito di opzioni, permettono di ricavare altre informazioni utili (es. media e deviazione standard delle risposte) in quanto forniscono *dati quantitativi*.

AUTOEFFICACIA

62. Torna al momento in cui hai pensato di abbandonare seriamente lo sport per te più significativo e rispondi.

1. Completamente in disaccordo
2. Abbastanza in disaccordo
3. Né in accordo né in disaccordo
4. Abbastanza in accordo
5. Completamente in accordo

"Mi sono sentito capace di affrontare l'allenamento..."

...anche quando avevo preoccupazioni e problemi personali	1	2	3	4	5
...anche quando ero stanco	1	2	3	4	5
...anche quando ero indaffarato	1	2	3	4	5

Fig. 5: Esempi di domande a risposta su scala Likert

Infine, le *domande a risposta alternativa (Si/No)* rappresentano un caso particolare delle domande a risposta chiusa con una sola possibile risposta tra due opzioni, consentendo di gestire i dati ottenuti tramite variabili booleane.

26. Riguardo al tuo sport, lo hai fatto a livello agonistico?

☐ Si

☐ No

Fig. 6: Esempio di domanda a risposta alternativa (Si/No)

3. Somministrazione e raccolta dati

Dopo aver definito e diviso le domande in sezioni specifiche, Orangogo ha caricato il questionario su Survey Monkey, un software di sviluppo di sondaggi online che opera come società di servizi, basato su cloud.

Successivamente, si è provveduto all'identificazione del campione a cui somministrare il questionario in modo stratificato rispetto al numero di ragazzi iscritti al quinto anno di scuola secondaria di secondo grado (età 18-20 anni) su territorio nazionale.

ITALIA		
iscritti – maschi	Liceo	451.337,00
iscritti – maschi	istituto tecnico	601.375,00
iscritti – maschi	istituto professionale	308.800,00
iscritti – femmine	Liceo	714.031,00
iscritti – femmine	istituto tecnico	339.424,00
iscritti – femmine	istituto professionale	232.090,00
TOTALE ISCRITTI		2.647.057,00
ISCRITTI MASCHI		1.361.512,00
ISCRITTI FEMMINE		1.285.545,00
Numero totale dei questionari 3000		

Fig. 7: Stratificazione degli studenti iscritti al quinto anno di scuola superiore di secondo grado su territorio nazionale

Sulla base di questa stratificazione, sono state ricercate le varie scuole sul territorio ed è stato inviato l'invito di partecipazione al progetto.

La somministrazione del questionario è avvenuta in un'aula informatica, in modo tale che i risultati dei questionari fossero automaticamente acquisiti dalla piattaforma Survey Monkey.

Al termine dell'intera fase di raccolta dati, Orangogo ha messo a disposizione dell'Università Bicocca il file CSV **"Abbandono_Sportivo.csv"** esportato dalla piattaforma Survey Monkey, contenente i risultati dei questionari sotto forma di dati grezzi.

Capitolo Secondo

Analisi dati

1. Fase di pre-processing

L'obiettivo della fase di pre-processing è quello di ottenere un dataset che raccoglie e descrive in forma tabulare i dati grezzi derivanti dal processo di raccolta dati svolto da Orangogo.

Tali dati grezzi sono immagazzinati in un unico file CSV "**Abbandono_Sportivo.csv**", in cui ogni singolo questionario viene individuato dal valore univoco della colonna "*respondent_id*" per ogni riga del file.

Le risposte date ad una singola domanda, invece, sono contenute in una specifica colonna avente come intestazione il testo della domanda stessa.

Sebbene i dati a disposizione siano in formato semistrutturato, essi necessitano di una corposa fase di pre-processing per due ragioni:

1. il questionario non è il medesimo per tutti i ragazzi che lo hanno compilato (ad eccezione delle sezioni "*Dati Anagrafici*" e "*Dati Generici*"), in quanto esso procede in modo differenziato nel momento in cui viene stabilita, durante il questionario stesso, l'appartenenza del ragazzo ad una determinata categoria: durante l'adolescenza, il ragazzo può dichiarare di aver praticato un solo sport e di aver pensato di abbandonarlo (*caso 1.1*), oppure di aver praticato un solo sport e di non aver mai pensato di abbandonarlo (*caso 1.2*), oppure di aver praticato più di uno sport e di aver pensato di abbandonare quello più significativo (*caso 2.1*), oppure di aver praticato più di uno sport e di non aver mai pensato di abbandonare quello più significativo (*caso 2.2*), oppure di aver abbandonato lo sport più significativo (*caso 3*), oppure di non aver mai praticato sport (*caso 4*).

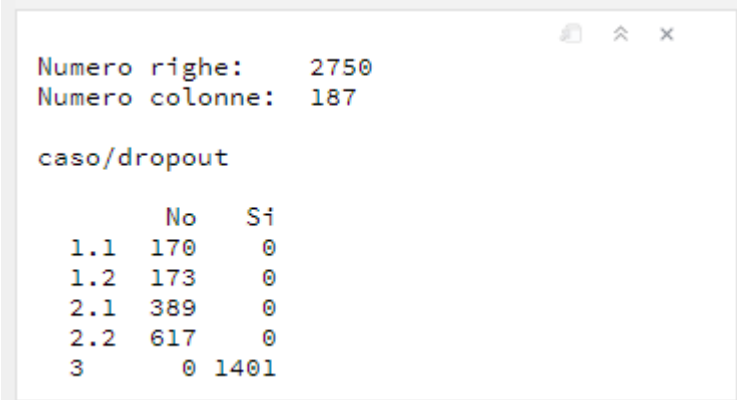
2. per le domande aperte è necessario processare le risposte affinché possano assumere una forma tale da poter essere suddivise in classi, mentre per le domande a risposta multipla è necessario riportare le possibili risposte in un'unica forma rispetto alle diverse forme per ogni caso descritto nel punto 1.

Al termine della fase di pre-processing, è stata assegnata una *label* ad ogni questionario, mediante la colonna “*dropout*”: viene assegnato valore “*Si*” per tutti i questionari relativi ai ragazzi appartenenti al *caso 3*, indicando che il dropout è effettivamente avvenuto; viene assegnato valore “*No*” per tutti i questionari relativi ai ragazzi appartenenti ai casi *1.1*, *1.2*, *2.1* e *2.2*, indicando che il dropout non è effettivamente avvenuto.

Le righe corrispondenti ai dati dei questionari dei ragazzi appartenenti al *caso 4*, cioè coloro che hanno dichiarato di non aver mai praticato sport in adolescenza, sono state rimosse in quanto non utili per la classificazione del dropout.

Tale scelta è motivata dal concetto espresso in letteratura, secondo il quale il dropout sportivo presuppone inizialmente che ci sia la pratica dell’attività sportiva.

Di seguito vengono riportate le dimensioni (numero righe e numero colonne) e le caratteristiche di stratificazione rispetto alle variabili “*caso*” e “*dropout*” del dataset ottenuto dalla fase di pre-processing.



```

Numero righe: 2750
Numero colonne: 187

caso/dropout

```

	No	Si
1.1	170	0
1.2	173	0
2.1	389	0
2.2	617	0
3	0	1401

Fig. 8: Caratteristiche del dataset ottenuto dalla fase di pre-processing

Il dataset risulta bilanciato rispetto alla variabile “*dropout*” (50,95% hanno abbandonato e 49,05% non hanno abbandonato), per cui è ragionevole utilizzarla come variabile target per gli algoritmi di classificazione.

Le **187** colonne del dataset possono essere classificate in:

1. *variabili booleane (135)* con valore *Si/No* riferite alle domande con risposta alternativa *Si/No* e alle domande a risposta multipla per le quali ogni opzione di risposta viene trasformata in una variabile con valore *Si/No* a seconda se è stata selezionata o meno;

2. *variabili categoriali (52)* con N possibili valori a seconda delle opzioni di risposta della domanda a risposta chiusa di riferimento oppure a seconda delle classi individuate nel processamento delle domande aperte.

	respondent_id	sex	scuola_superiore	disabilita	titolo_studio_madre	titolo_studio_padre	numero_fratelli_sorelle	numero_nonni_vicinanza_casa	riferime
0	6630885156	Maschio	Liceo	No	Diploma di scuola media superiore (Liceo)	Diploma di scuola media superiore (Liceo)	2	1	
1	6630885459	Femmina	Liceo	No	Diploma di scuola media superiore (Liceo)	Licenza Media Inferiore	1	2	
2	6630885495	Maschio	Liceo	No	Licenza Media Inferiore	Licenza Media Inferiore	1	0	
3	6630885564	Maschio	Liceo	No	Diploma di scuola media superiore (Istituto Te...	Licenza Media Inferiore	1	4	
4	6630885783	Maschio	Liceo	No	Diploma di scuola media superiore (Istituto Te...	Licenza Media Inferiore	1	4	

Fig. 9: Prime cinque righe del dataset ottenuto dalla fase di pre-processing

Il dataset ottenuto è stato esportato nel file CSV **"01 pre_processing.csv"**.

2. Analisi statistica descrittiva

L'analisi statistica descrittiva ha l'obiettivo di ispezionare i dati a disposizione per ricavare informazioni generali sul campione, dopo averli immagazzinati in un dataset e averli sottoposti al processo preliminare di "pulitura" (fase di pre-processing).

Di seguito, vengono riportate le analisi più significative relative alle variabili presenti nel dataset, soprattutto in relazione alle differenze tra il gruppo di osservazioni per cui si è verificato il dropout sportivo e il gruppo di osservazioni per cui non si è verificato.

Essendo il dataset privo di variabili quantitative e composto da variabili categoriali o booleane (variabili categoriali con 2 possibili valori), si è scelto di rappresentare graficamente le analisi sulla distribuzione delle variabili mediante *istogrammi* a scala percentuale.

E' doveroso specificare che tutte le conclusioni rilevate dall'analisi descrittiva rispetto al dropout si riferiscono al campione preso in considerazione, quindi non costituiscono conclusioni generali sulla popolazione.

Multisport

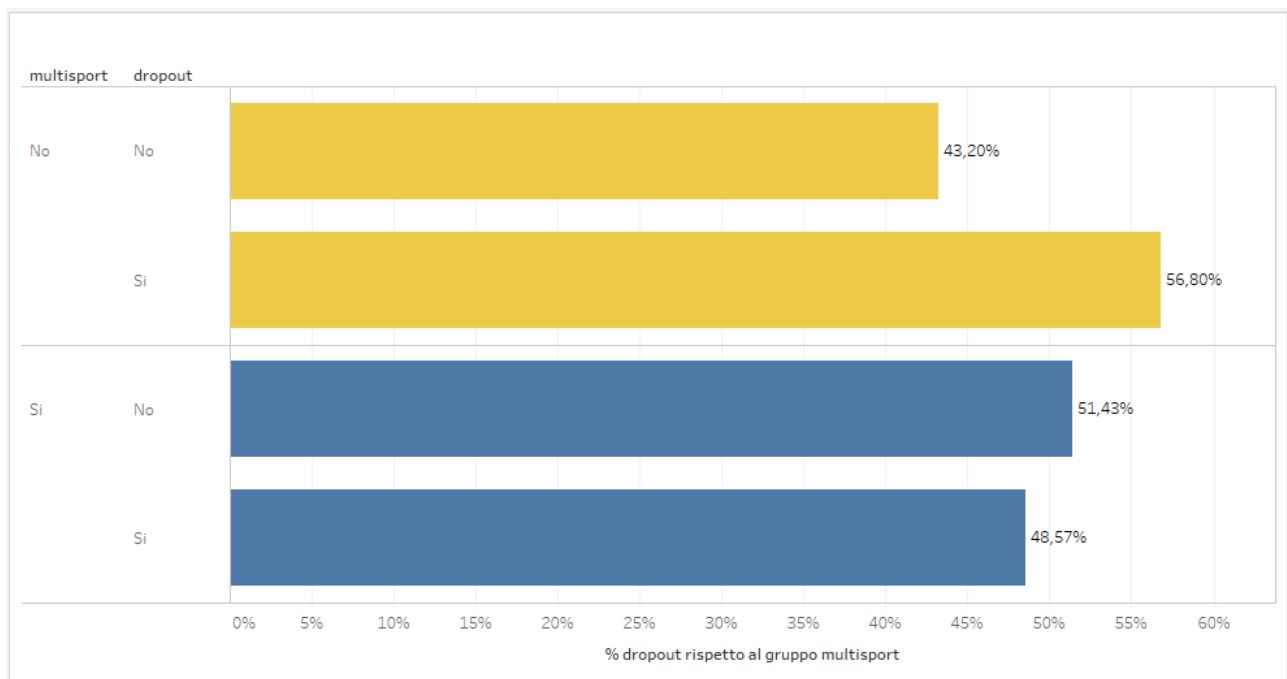


Fig. 10: Distribuzione dell'abbandono nel campione in base alla pratica di un solo sport o più sport

Nel campione, considerando chi ha praticato un solo sport durante l'adolescenza (*multisport* = "No"), la percentuale di chi ha abbandonato (56,80%) è superiore alla percentuale di chi non ha abbandonato (43,20%) di 13,5 punti percentuali.

Considerando, invece, chi ha praticato più di uno sport durante l'adolescenza (*multisport* = "Sì"), la percentuale di chi ha abbandonato (48,57%) è inferiore alla percentuale di chi non ha abbandonato (51,43%) di 2,86 punti percentuali.

Si può concludere che chi pratica più sport durante l'adolescenza è meno predisposto al dropout sportivo rispetto a chi pratica un solo sport.

Sesso

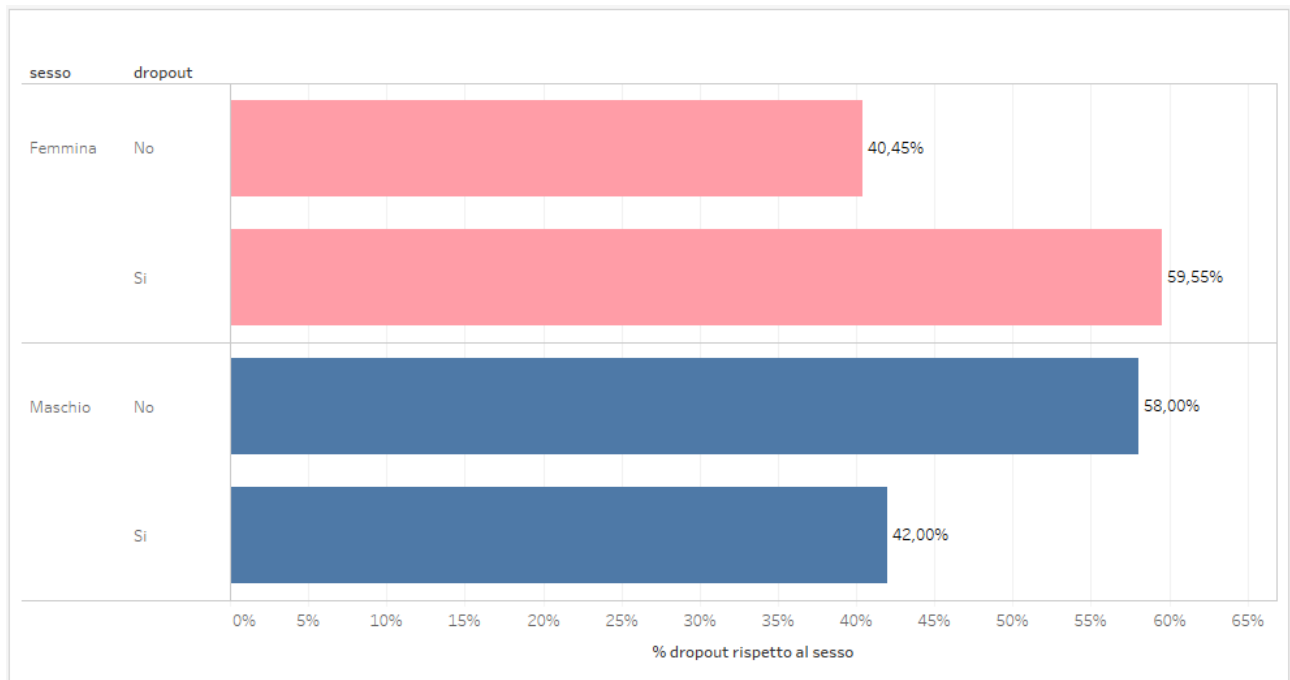


Fig. 11: Distribuzione dell'abbandono nel campione in base al sesso

Nel campione, considerando il sesso femminile, la percentuale di chi ha abbandonato (59,55%) è superiore alla percentuale di chi non ha abbandonato (40,45%) di 19,1 punti percentuali.

Considerando il sesso maschile, invece, la percentuale di chi ha abbandonato (42%) è inferiore alla percentuale di chi non ha abbandonato (58%) di 16 punti percentuali.

Si può concludere che le ragazze sono più predisposte al dropout sportivo rispetto ai ragazzi durante l'adolescenza.

Provenienza

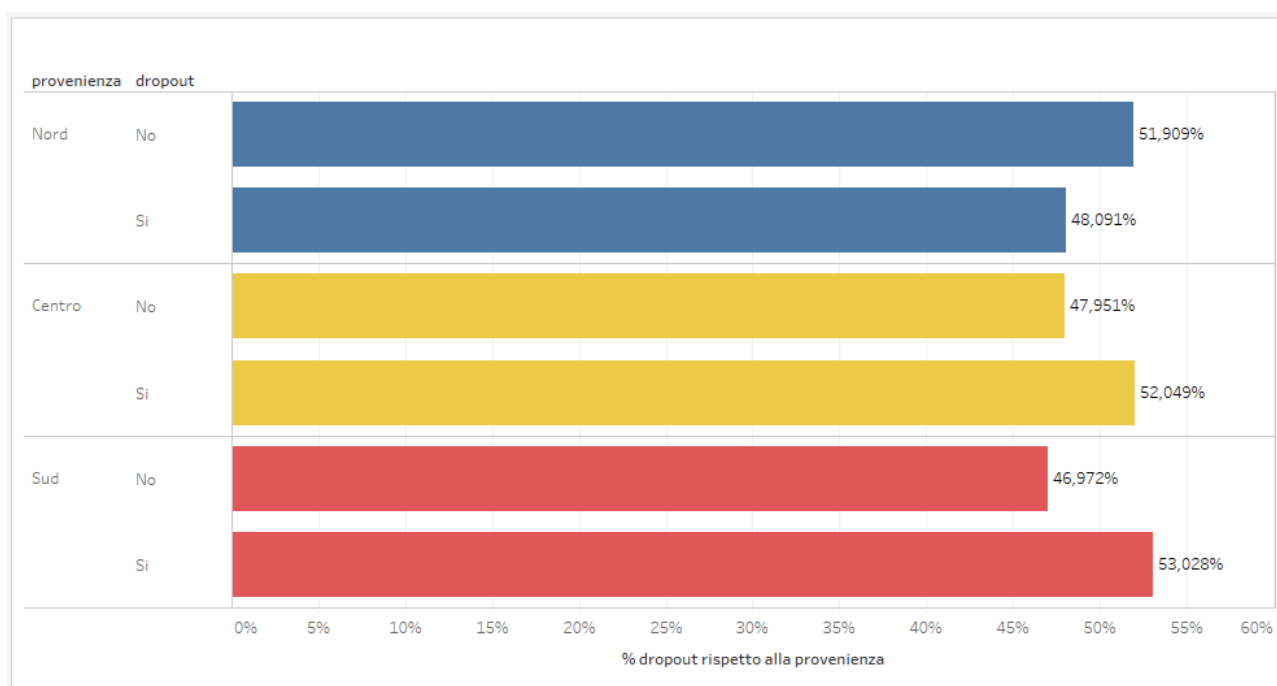


Fig. 12: Distribuzione dell'abbandono nel campione in base alla provenienza

Nel campione, considerando chi ha praticato sport al Nord, la percentuale di chi ha abbandonato (48,1%) è inferiore alla percentuale di chi non ha abbandonato (51,9%) di 3,8 punti percentuali.

Considerando, invece, chi ha praticato sport al Centro, la percentuale di chi ha abbandonato (52,05%) è superiore alla percentuale di chi non ha abbandonato (47,95%) di 4,1 punti percentuali.

Considerando, infine, chi ha praticato sport al Sud, la percentuale di chi ha abbandonato (53,03%) è superiore alla percentuale di chi non ha abbandonato (46,97%) di 6,06 punti percentuali.

Si può concludere che la predisposizione al dropout sportivo durante l'adolescenza aumenta progressivamente dal Nord verso il Sud.

Sport più significativo

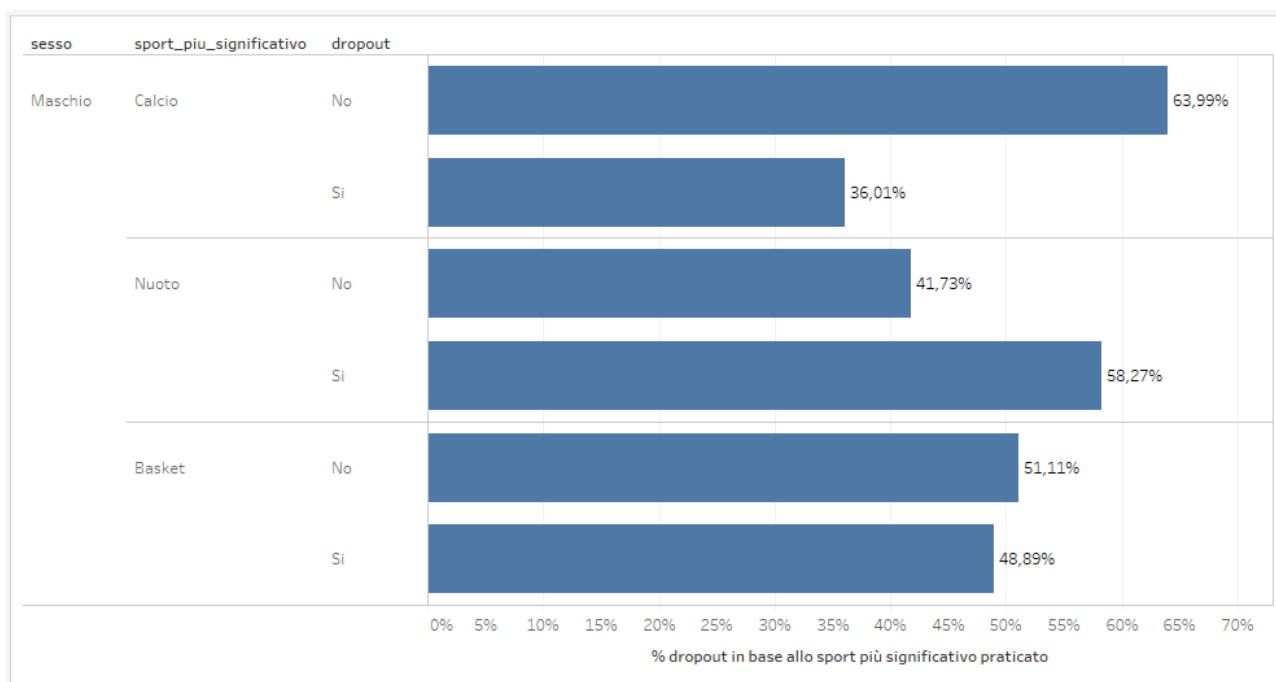


Fig. 13: Distribuzione abbandono nel campione in base ai tre sport più significativi maggiormente praticati dal sesso maschile

Nel campione, i tre sport più significativi maggiormente praticati dal sesso maschile durante l'adolescenza sono: calcio (40,24%), nuoto (10,44%) e basket (10,14%).

Considerando il sesso maschile che ha praticato il calcio come sport più significativo, la percentuale di chi ha abbandonato (36%) è inferiore alla percentuale di chi non ha abbandonato (64%) di 28 punti percentuali.

Considerando, invece, il sesso maschile che ha praticato il nuoto come sport più significativo, la percentuale di chi ha abbandonato (58,27%) è superiore alla percentuale di chi non ha abbandonato (41,73%) di 16,54 punti percentuali.

Considerando, infine, il sesso maschile che ha praticato il basket come sport più significativo, la percentuale di chi ha abbandonato (48,9%) è inferiore rispetto alla percentuale di chi non ha abbandonato (51,1%) di 2,2 punti percentuali.

Riguardo i tre sport significativi maggiormente praticati dal sesso maschile, quindi, si può concludere che chi pratica calcio è meno predisposto al dropout sportivo durante l'adolescenza rispetto a chi pratica basket o nuoto.

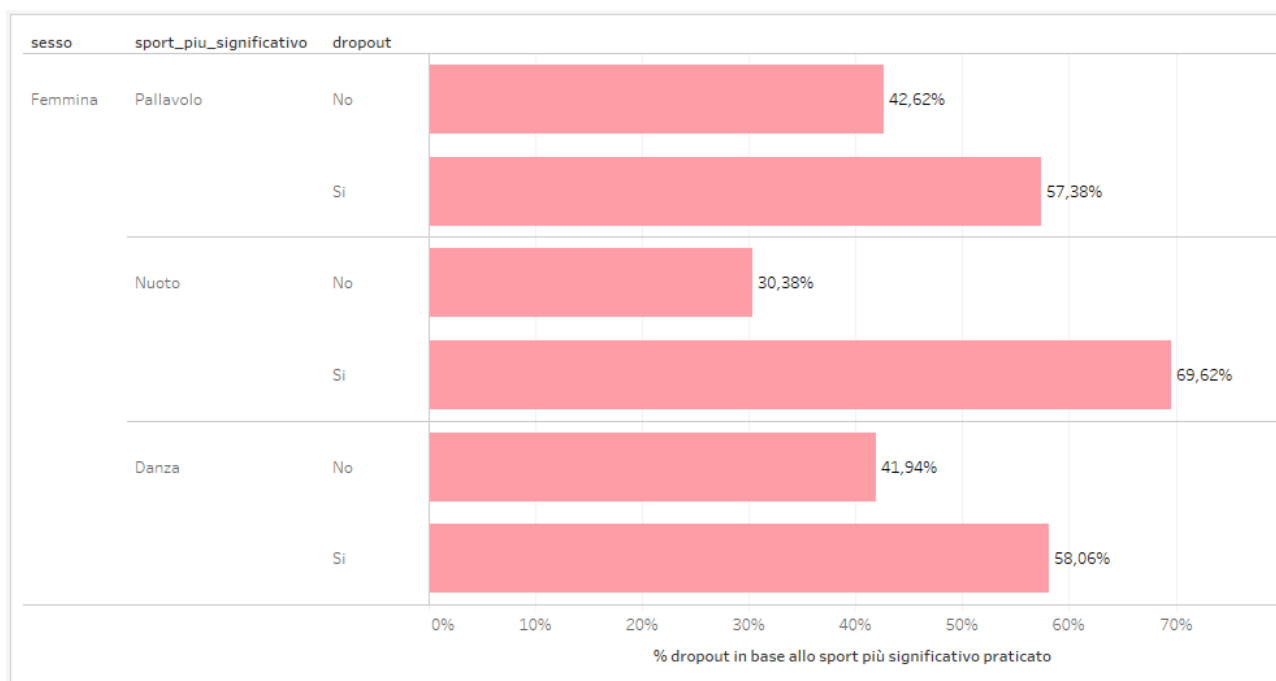


Fig. 14: Distribuzione abbandono nel campione in base ai tre sport più significativi maggiormente praticati dal sesso femminile

Nel campione, i tre sport più significativi maggiormente praticati dal sesso femminile durante l'adolescenza sono: pallavolo (17,82%), nuoto (17,31%) e danza (11,32%).

Considerando il sesso femminile che ha praticato la pallavolo come sport più significativo, la percentuale di chi ha abbandonato (57,38%) è superiore alla percentuale di chi non ha abbandonato (42,62%) di 14,76 punti percentuali.

Considerando, invece, il sesso femminile che ha praticato il nuoto come sport più significativo, la percentuale di chi ha abbandonato (69,62%) è superiore alla percentuale di chi non ha abbandonato (30,38%) di 39,24 punti percentuali.

Considerando, infine, il sesso femminile che ha praticato la danza come sport più significativo, la percentuale di chi ha abbandonato (58,06%) è superiore rispetto alla percentuale di chi non ha abbandonato (41,94%) di 16,12 punti percentuali.

Riguardo i tre sport significativi maggiormente praticati dal sesso femminile, quindi, si può concludere che chi pratica pallavolo o danza è meno predisposto al dropout sportivo durante l'adolescenza rispetto a chi pratica nuoto, pur rimanendo significativa la predisposizione al dropout data dal sesso.

Tipologia sport più significativo

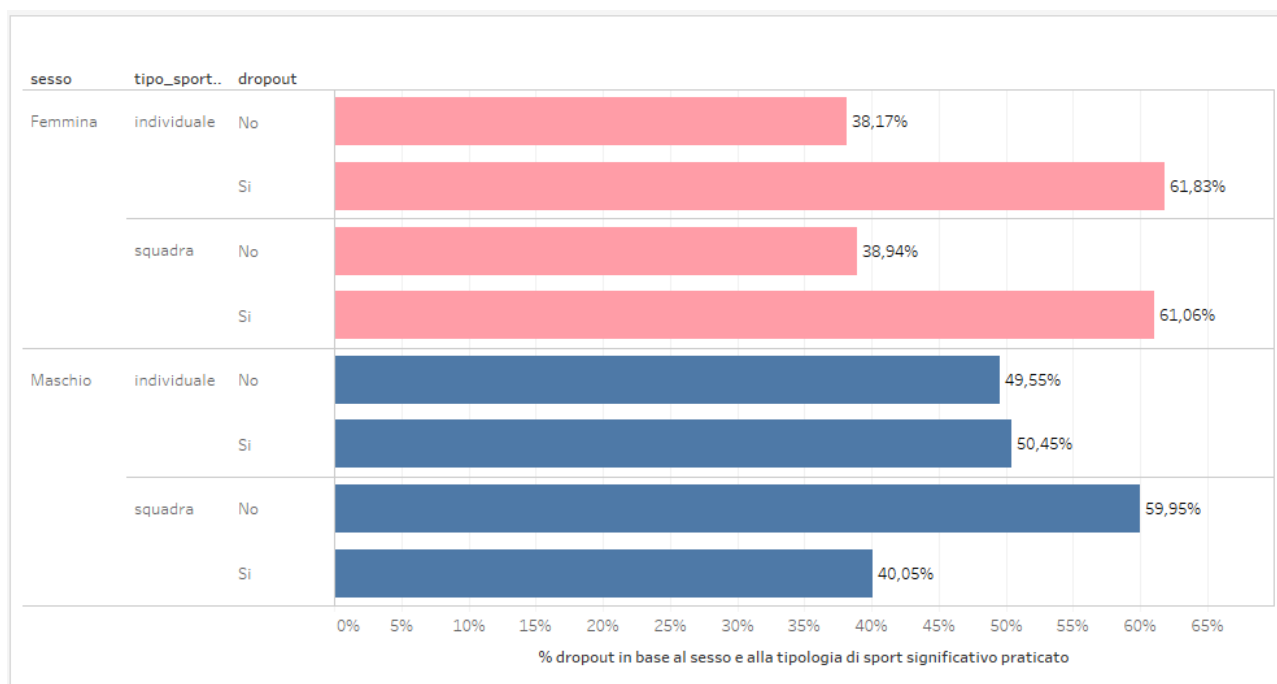


Fig. 15: Distribuzione abbandono nel campione in base al sesso e alla tipologia di sport più significativo praticato

Nel campione, considerando il sesso femminile che ha praticato lo sport più significativo di tipo individuale durante l'adolescenza, la percentuale di chi ha abbandonato (61,83%) è superiore alla percentuale di chi non ha abbandonato (38,17%) di 23,66 punti percentuali.

Considerando, invece, il sesso femminile che ha praticato lo sport più significativo di squadra durante l'adolescenza, la percentuale di chi ha abbandonato (61,06%) è superiore alla percentuale di chi non ha abbandonato (38,94%) di 22,12 punti percentuali.

Si può concludere, quindi, che per il sesso femminile la predisposizione al dropout sportivo rimane pressochè invariata in base alla tipologia di sport più significativo praticato, pur rimanendo significativa la predisposizione al dropout data dal sesso.

Nel campione, considerando il sesso maschile che ha praticato lo sport più significativo di tipo individuale durante l'adolescenza, la percentuale di chi ha abbandonato (50,45%) è superiore alla percentuale di chi non ha abbandonato (49,55%) di 0,9 punti percentuali.

Considerando, invece, il sesso maschile che ha praticato lo sport più significativo di squadra durante l'adolescenza, la percentuale di chi ha abbandonato (40,05%) è inferiore alla percentuale di chi non ha abbandonato (59,95%) di 19,9 punti percentuali.

Si può concludere, quindi, che per il sesso maschile chi pratica lo sport più significativo di squadra è meno predisposto al dropout sportivo rispetto a chi pratica lo sport più significativo di tipo individuale.

Agonismo

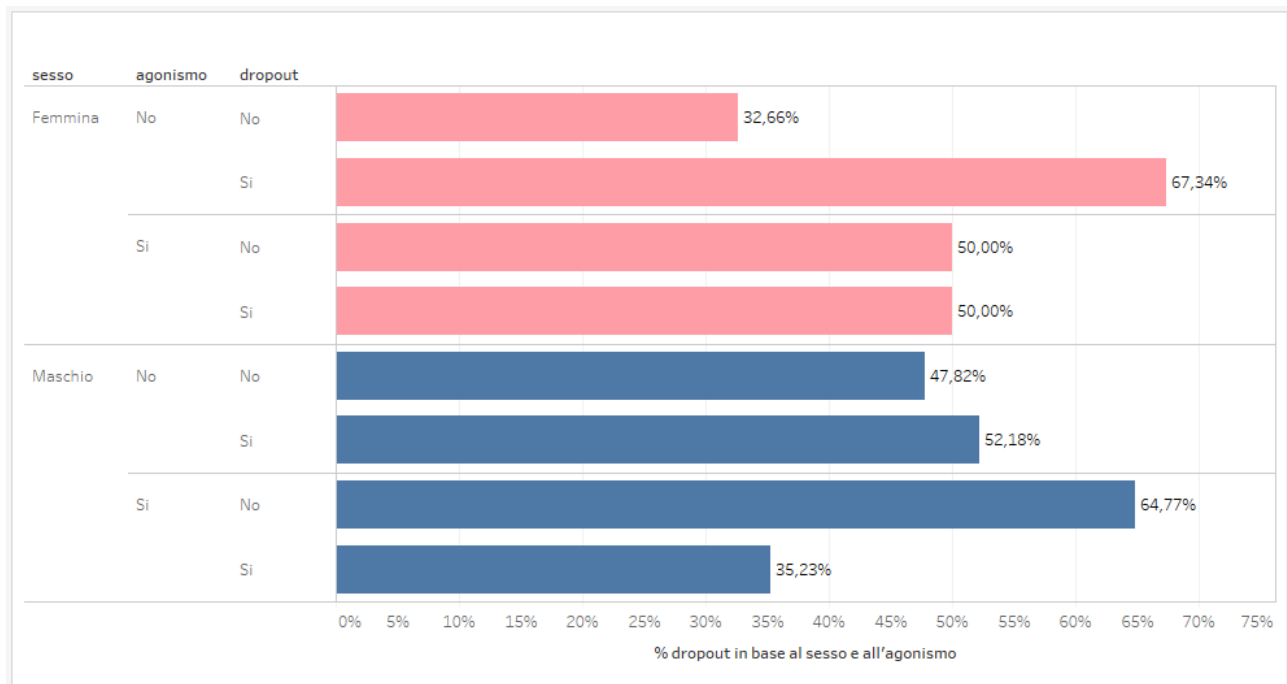


Fig. 16: Distribuzione dell'abbandono nel campione in base al sesso e alla pratica agonistica

Nel campione, considerando il sesso femminile che ha praticato sport a livello agonistico durante l'adolescenza, la percentuale di chi ha abbandonato (50%) è uguale alla percentuale di chi non ha abbandonato (50%).

Considerando, invece, il sesso femminile che non ha praticato sport a livello agonistico durante l'adolescenza, la percentuale di chi ha abbandonato (67,34%) è superiore alla percentuale di chi non ha abbandonato (32,66%) di 34,68 punti percentuali.

Si può concludere, quindi, che per il sesso femminile la predisposizione al dropout sportivo data dal sesso viene bilanciata dalla pratica agonistica, la quale risulta essere un *fattore di protezione* all'abbandono.

Nel campione, considerando il sesso maschile che ha praticato sport a livello agonistico durante l'adolescenza, la percentuale di chi ha abbandonato (35,25%) è inferiore alla percentuale di chi non ha abbandonato (64,77%) di 29,52 punti percentuali.

Considerando, invece, il sesso maschile che non ha praticato sport a livello agonistico durante l'adolescenza, la percentuale di chi ha abbandonato (52,18%) è superiore alla percentuale di chi non ha abbandonato (47,82%) di 4,36 punti percentuali.

Si può concludere, quindi, che per il sesso maschile chi pratica sport a livello agonistico è meno predisposto al dropout sportivo rispetto a chi non pratica sport a livello agonistico.

Motivazioni per continuare la pratica sportiva

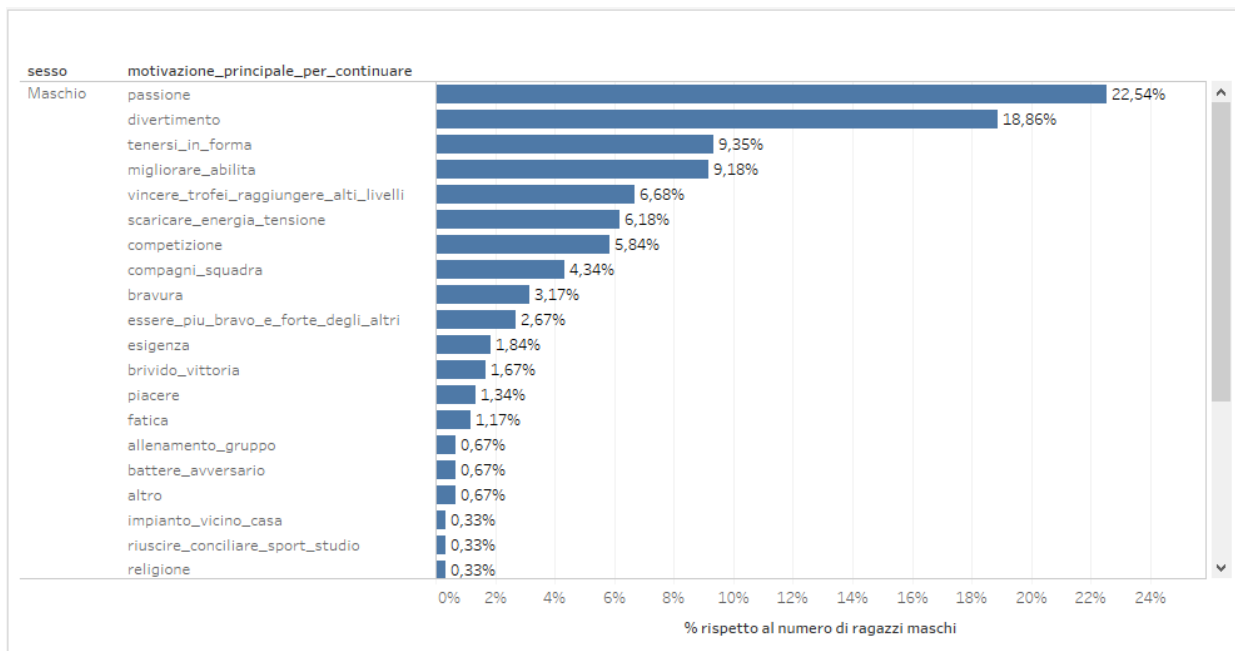


Fig. 17: Distribuzione dell'abbandono nel campione in base alla motivazione principale per continuare la pratica sportiva per il sesso maschile

Nel campione, considerando il sesso maschile, le cinque principali motivazioni per continuare la pratica sportiva sono: passione (22,54%), divertimento (18,86%), tenersi in forma (9,35%), migliorare le proprie abilità (9,18%) e vincere trofei e raggiungere alti livelli (6,68%).

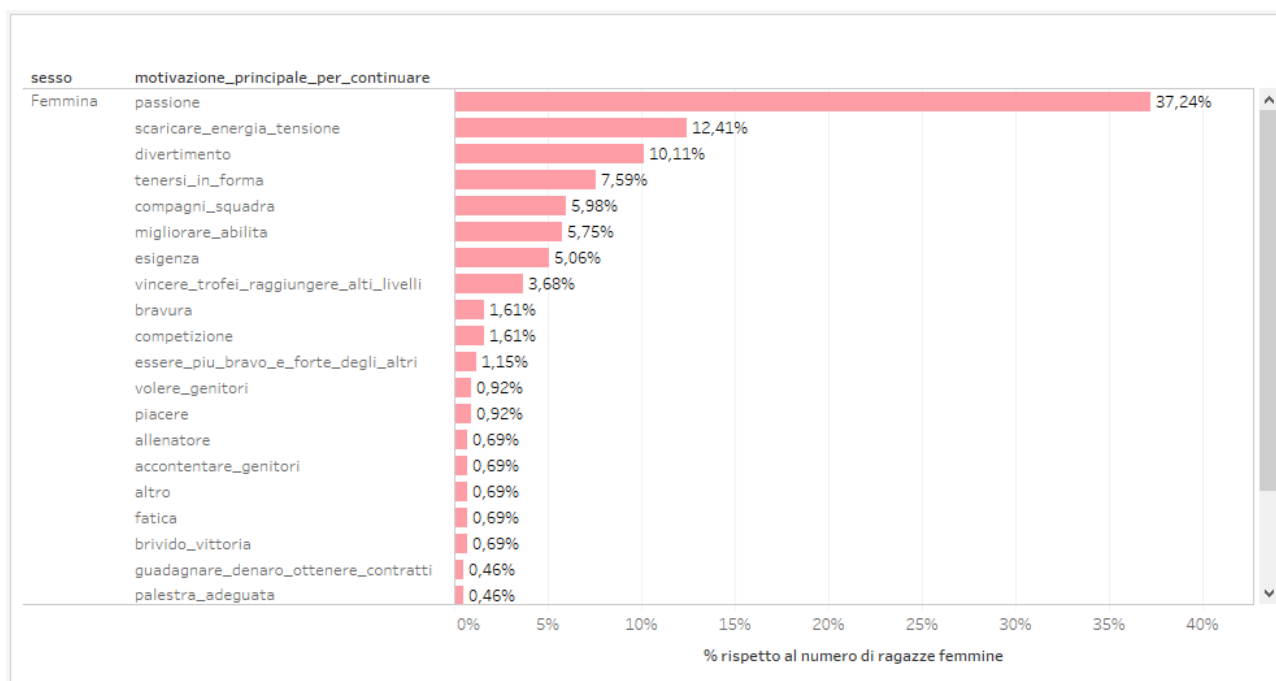


Fig. 18: Distribuzione dell'abbandono nel campione in base alle tre principali motivazioni per continuare la pratica sportiva per il sesso femminile

Nel campione, considerando il sesso femminile, le cinque principali motivazioni per continuare la pratica sportiva sono: passione (37,24%), scaricare energia e tensione (12,41%), divertimento (10,11%), tenersi in forma (7,59%) e il rapporto con le compagne di squadra (5,98%).

Motivazioni per abbandonare la pratica sportiva

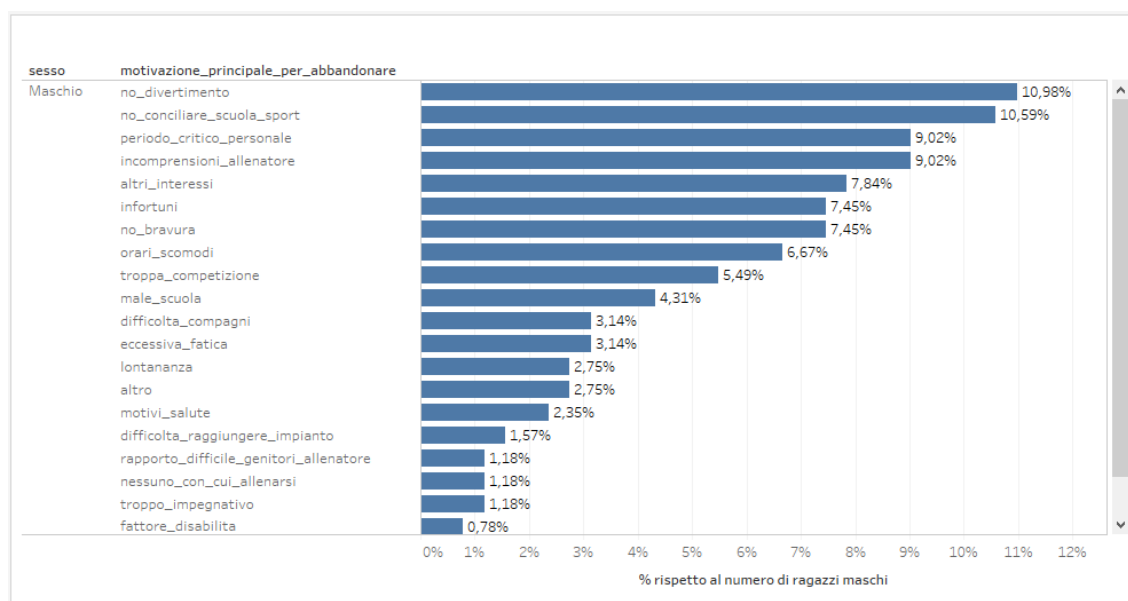


Fig. 19: Distribuzione dell'abbandono nel campione in base alle tre principali motivazioni per abbandonare la pratica sportiva per il sesso maschile

Nel campione, considerando il sesso maschile, le cinque principali motivazioni per abbandonare la pratica sportiva sono: non divertimento (10,98%), non riuscire a conciliare scuola e sport (10,59%), periodo critico personale (9,02%), incomprensioni con l'allenatore (9,02%) e altri interessi (7,84%).

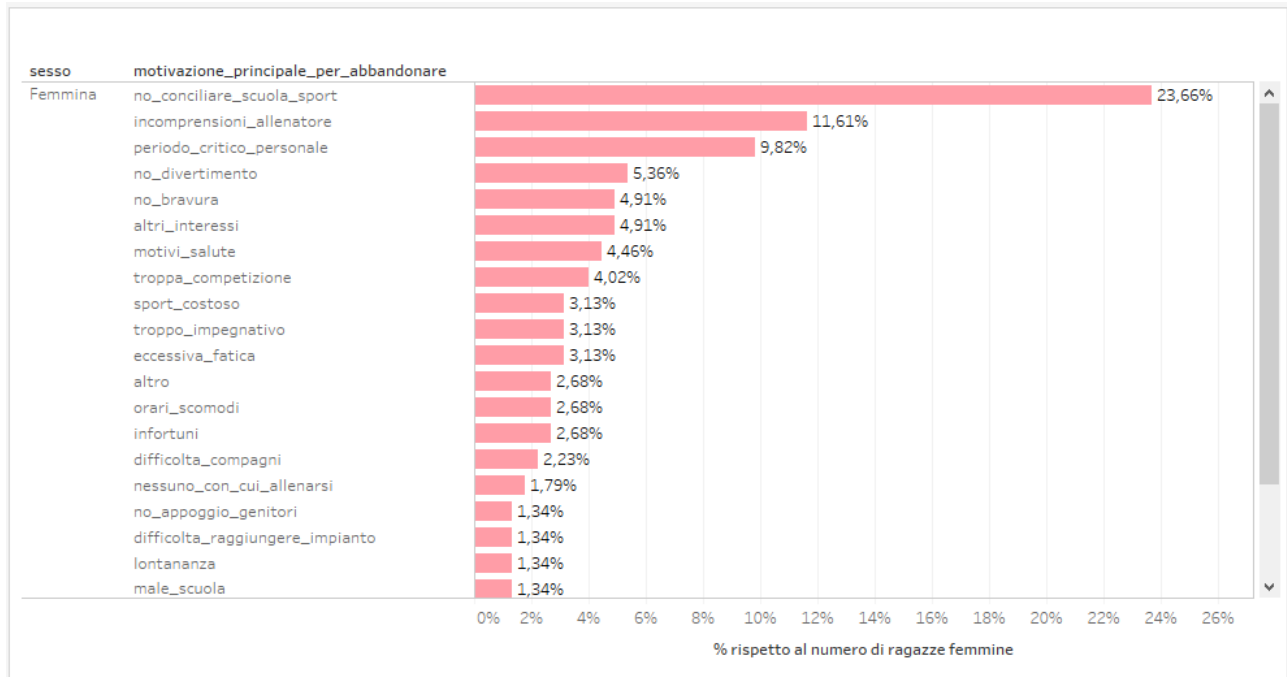


Fig. 20: Distribuzione dell'abbandono nel campione in base alle tre principali motivazioni per abbandonare la pratica sportiva per il sesso femminile

Nel campione, considerando il sesso femminile, le cinque principali motivazioni per abbandonare la pratica sportiva sono: non riuscire a conciliare scuola e sport (23,66%), incomprensioni con l'allenatore (11,61%), periodo critico personale (9,82%), non divertimento (5,36%) e non essere/sentirsi brava (4,91%).

Momento critico: età in cui si abbandona o si pensa di abbandonare

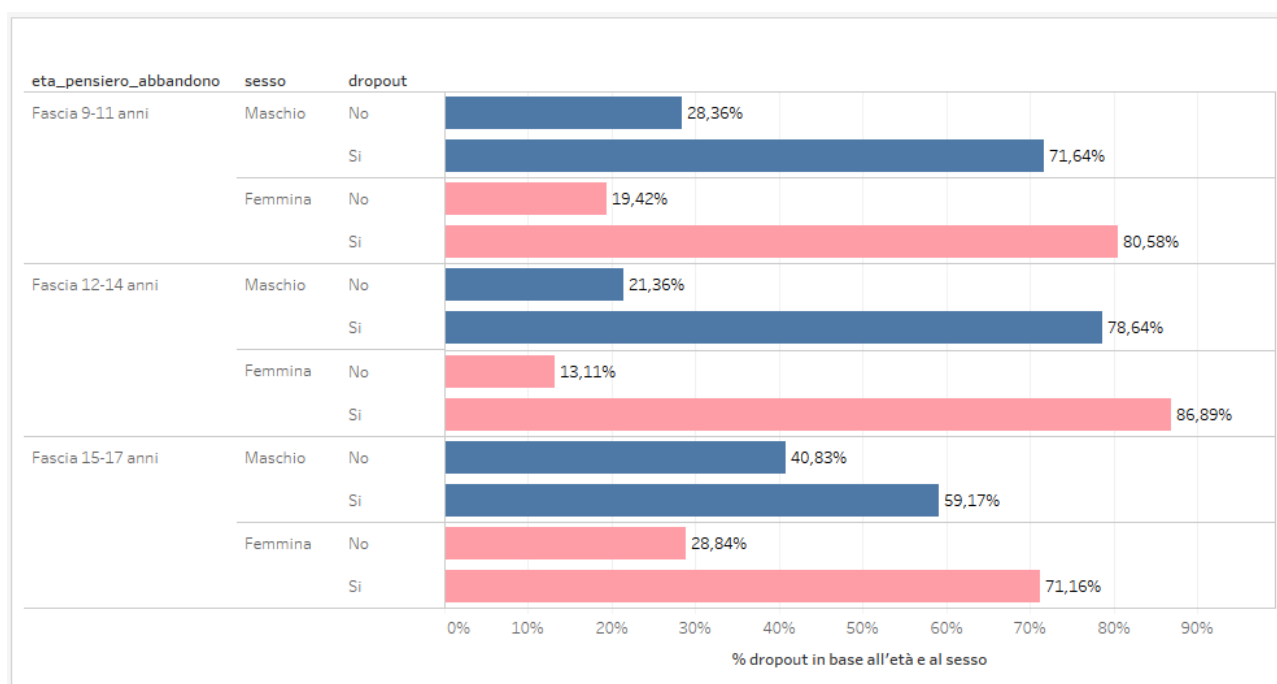


Fig. 21: Distribuzione dell'abbandono nel campione in base al sesso e all'età in cui si pensa di abbandonare

Nel campione, considerando chi ha attraversato un momento critico nella fascia di età dai 9 agli 11 anni, la percentuale di abbandono è pari al 71,64% per il sesso maschile, mentre è pari a 80,58% per il sesso femminile, con un incremento di 8,94 punti percentuali.

Considerando, invece, chi ha attraversato un momento critico nella fascia di età dai 12 ai 14 anni, la percentuale di abbandono è pari al 78,64% per il sesso maschile, mentre è pari a 86,89% per il sesso femminile, con un incremento di 8,25 punti percentuali.

Considerando, infine, chi ha attraversato un momento critico nella fascia di età dai 15 ai 17 anni, la percentuale di abbandono è pari al 59,17% per il sesso maschile, mentre è pari a 71,16% per il sesso femminile, con un incremento pari a 12 punti percentuali.

Si può concludere che, rispetto al sesso, la predisposizione delle ragazze al dropout sportivo rispetto ai ragazzi è maggiormente significativa nella fascia di età dai 15 ai 17 anni.

Inoltre, sia per il sesso maschile sia per il sesso femminile, la predisposizione al dropout raggiunge il picco nella fascia di età dai 12 ai 14 anni, per poi ridursi significativamente nella fascia di età dai 15 ai 17 anni, in cui risulta essere minore anche rispetto alla fascia di età dai 9 agli 11 anni.

3. Test del Chi-Quadro

Il test del *Chi-Quadro* è un test statistico che consente di determinare se due variabili di tipo categoriale sono indipendenti o correlate.

Si basa sulla formulazione di un'*ipotesi nulla* H_0 , secondo cui le due variabili sono indipendenti, e di un'*ipotesi alternativa* H_1 , secondo cui le due variabili non sono indipendenti.

Fissate le ipotesi, si sceglie un livello di significatività con il quale accettare o rifiutare l'ipotesi nulla: valori tipici di questo parametro, denominato α , sono 0.1, 0.05 e 0.01, che corrispondono rispettivamente ad un livello di significatività pari a 90%, 95% e 99%.

Per poter procedere al calcolo del valore del *Chi-Quadro*, è necessario ricavare la *tabella di contingenza* delle due variabili, ovvero il numero di valori osservati (*observed values*) per ogni possibile coppia determinata dalle due variabili.

	dropout	No	Si
eta_pensiero_abbandono			
Fascia 9-11 anni	39	133	
Fascia 15-17 anni	367	693	
Fascia 12-14 anni	92	480	

Fig. 22: Tabella di contingenza tra età pensiero abbandono e dropout (*observed values*)

Avendo a disposizione la *tabella di contingenza*, si procede a ricavare i valori attesi delle frequenze (*expected values*), nel seguente modo:

$$E = (\text{row total} * \text{col total}) / \text{grand total}$$

	dropout	No	Si
eta_pensiero_abbandono			
Fascia 9-11 anni	47.48	124.52	
Fascia 15-17 anni	292.62	767.38	
Fascia 12-14 anni	157.90	414.10	

Fig. 23: Tabella di contingenza tra età pensiero abbandono e dropout (*expected values*)

A questo punto, è possibile calcolare il valore del *Chi-Quadro*:

$$\text{Chi-Quadro} = \sum_i [(O_i - E_i) / E_i] \quad i = 1 \dots n$$

Dalla funzione di densità di probabilità del *Chi-Quadro*, la quale tiene conto dei gradi di libertà (*degrees of freedom*), si ottiene il valore del *p-value*.

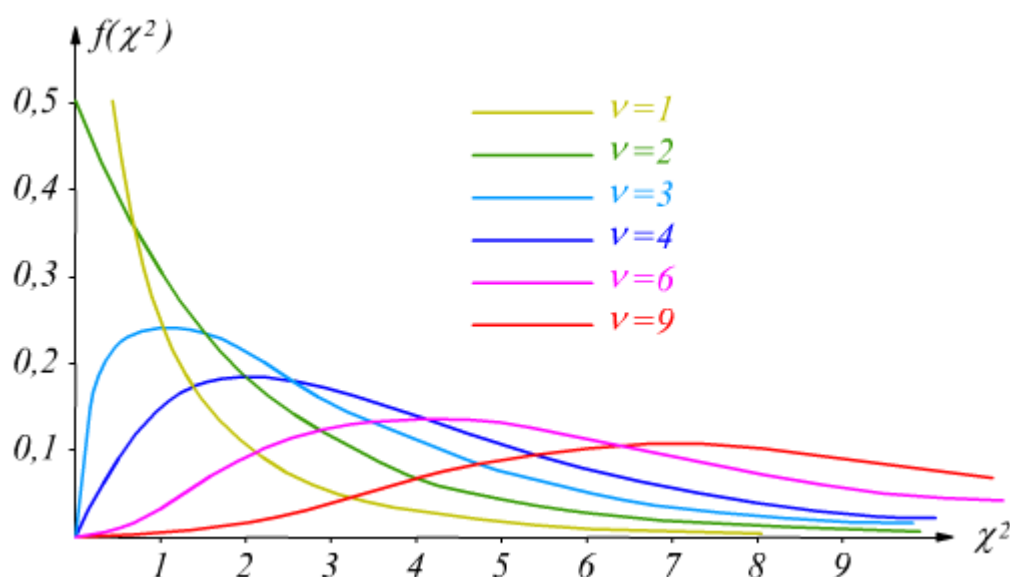


Fig. 24: Funzione di densità di probabilità del *Chi-Quadro*

Il valore del *p-value* determina l'esito del test: se il **p-value** è maggiore del parametro α , allora si accetta l'*ipotesi nulla* H_0 , concludendo che le due variabili sono indipendenti con un livello di significatività pari a $1 - \alpha$.

In caso contrario, ovvero se il **p-value** è minore del parametro α , allora si rifiuta l'*ipotesi nulla* H_0 , concludendo che le due variabili non sono indipendenti, perciò correlate tra loro, con un livello di significatività pari a $1 - \alpha$.

Nel lavoro di individuazione dei predittori più significativi del dropout sportivo durante l'adolescenza, è stato utilizzato il test del *Chi-Quadro* per testare l'indipendenza o la correlazione delle variabili con percentuale di *missing values* superiore al 25%, ovvero quelle variabili non presenti nei questionari dei ragazzi appartenenti ai casi 1.2 e 2.2.

Di seguito, viene riportato l'elenco delle variabili correlate al dropout sportivo (livello di significatività = 95% $\rightarrow \alpha = 0.05$), con relativi valori del *Chi-Quadro*, *p-value* e *gradi di libertà* (le variabili con un solo grado di libertà sono variabili booleane No/Sì):

	var	stat	p_value	dof
11	eta_pensiero_abbandono	66.204466	4.206133e-15	2
12	stato_animo_preAllenamento	41.010585	2.881796e-07	6
1	periodo_critico_personale	16.560994	4.711006e-05	1
0	incomprensioniAllenatore	15.531343	8.114856e-05	1
17	allenamento_essendo_indaffarato	22.285223	1.758480e-04	4
15	allenamento_con_preoccupazioni	21.605989	2.400553e-04	4
3	male_scuola	13.198868	2.801182e-04	1
16	allenamento_con_stanchezza	19.544143	6.142463e-04	4
2	nessuno_con_cuiAllenarsi	9.100323	2.555643e-03	1
13	attenzioniAllenatore	7.988451	4.707667e-03	1
5	altri_interessi	7.947190	4.816210e-03	1
9	all_riferimento	6.025693	1.409909e-02	1
4	no_conciliare_scuola_sport	5.780502	1.620491e-02	1
10	all_no_preparato	5.166741	2.302337e-02	1
14	stimolo_compagni	5.012909	2.515900e-02	1
18	sentimento_sconfitta	12.358645	3.019063e-02	5
19	sentimento_vittoria	11.946836	3.552356e-02	5
7	all_leader	4.315313	3.777083e-02	1
8	all_fissato	4.293308	3.826264e-02	1
6	all_competente	4.266177	3.887831e-02	1

Fig. 25: Variabili correlate al dropout sportivo, secondo il test *Chi-Quadro*, tra quelle non presenti nei questionari di chi non ha mai pensato di abbandonare

INDIVIDUAZIONE DEI PREDITTORI

Capitolo Primo

Data Preparation

1. Gestione dei missing values

La presenza di missing values all'interno del dataset ottenuto dalla fase di pre-processing può essere giustificata dalla mancanza di risposta a uno o più domande da parte dei ragazzi che si sono sottoposti alla compilazione del questionario.

Ciò può verificarsi in maniera volontaria, nel caso in cui il ragazzo ha scelto autonomamente di non rispondere, sia in maniera involontaria, nel caso in cui il ragazzo non si è accorto di non aver risposto a una o più domande, ipotizzando che il sistema telematico di acquisizione dati *Survey Monkey* non obblighi alla risposta per tutti i quesiti.

Non bisogna trascurare anche l'ipotesi di problemi tecnici durante la somministrazione del questionario che hanno impedito la corretta acquisizione di tutte le risposte.

La possibilità di rimuovere le osservazioni del dataset contenenti missing values non è praticabile, in quanto l'intero dataset (2750 osservazioni), già di per sé con dimensione inferiore rispetto alla maggior parte dei dataset usati in letteratura per testare le performance degli algoritmi predittivi, sarebbe ridotto più del 60%.

E' necessario specificare che alcune domande del questionario non erano previste per i ragazzi che, durante la compilazione, hanno dichiarato di appartenere ai *casi 1.2 e 2.2*, ovvero coloro che non hanno mai pensato di abbandonare lo sport più significativo durante l'adolescenza.

Si tratta di domande relative alla sezione "*Momento Critico*" del questionario, la quale è prevista per i *casi 1.1, 2.1* (coloro che hanno pensato di abbandonare lo sport più significativo senza effettivamente farlo) e per il *caso 3* (coloro che hanno abbandonato lo sport più significativo).

Per le variabili presenti nel dataset ottenuto dalla fase di pre-processing, contenenti i dati relativi alle risposte fornite per queste domande, il numero di missing values risulta essere maggiore del 25%.

Considerato l'obiettivo di ottenere un dataset volto all'individuazione dei predittori significativi del dropout e allo sviluppo di modelli predittivi, per queste variabili è stato scelto di adottare la *politica di rimozione* dal dataset.

L'analisi dell'incidenza sul dropout e l'individuazione di pattern significativi relativi a questo gruppo di variabili è riconducibile all'analisi statistica descrittiva e al *test del chi-quadro*, descritti nel capitolo precedente.

Per le variabili con numero di missing values minore del 25%, invece, i missing values sono stati gestiti adottando la *politica di sostituzione*, diversa per ogni variabile in base al suo significato concettuale.

Tra queste, per le variabili aventi un valore riferito ad una negazione (domande Si/No) o aventi un valore riferito all'assenza di risposta (es: "Nessuno ha influenzato la scelta dello sport più significativo"), è stato scelto questo valore per sostituire i missing values.

Per le restanti variabili, sono state adottate specifiche politiche di sostituzione che di seguito vengono riportate nel dettaglio: si tratta, come detto, di variabili con percentuale di missing values minore del 25%, ovvero variabili collegate alle domande comuni a tutti i casi di appartenenza di un ragazzo che svolge il questionario.

Dati Anagrafici

Variabile: sesso

Percentuale MV: 0,47%

Politica di sostituzione: valore più frequente rispetto allo sport più significativo praticato

```
dropout_final.loc[dropout_final.sport_piu_significativo == "Calcio",:].sesso.describe()
count          566
unique          2
top      Maschio
freq          536
Name: sesso, dtype: object
```

Fig. 26: Politica di sostituzione dei missing values per la variabile "sesso"

Variabile: regione

Percentuale MV: 0.22%

Politica di sostituzione: valore più frequente

```
dropout_final.regione.describe()
```

```
count      2744  
unique      20  
top        Campania  
freq       346  
Name: regione, dtype: object
```

Fig. 27: Politica di sostituzione dei missing values per la variabile “regione”

Variabile: nazionalità

Percentuale MV: 0,30%

Politica di sostituzione: valore più frequente

```
dropout_final.nazionalita.describe()
```

```
count      2742  
unique      2  
top        Italiana  
freq       2673  
Name: nazionalita, dtype: object
```

Fig. 28: Politica di sostituzione dei missing values per la variabile “nazionalità”

Variabile: scuola superiore

Percentuale MV: 0,22%

Politica di sostituzione: valore più frequente rispetto al sesso

```
dropout_final.loc[dropout_final.sesso == "Femmina",:].scuola_superiore.describe()
```

```
count      1395  
unique      4  
top        Liceo  
freq       999  
Name: scuola_superiore, dtype: object
```

Fig. 29: Politica di sostituzione dei missing values per la variabile “scuola superiore”

Informazioni sulla pratica sportiva

Variabile: grado divertimento allenamento

Percentuale MV: 15,34%

Politica di sostituzione: valore più frequente rispetto alla partecipazione alle gare

```
dropout_final.groupby('partecipazione_gare')['grado_divertimento_allenamento'].describe()
```

	count	unique	top	freq
partecipazione_gare				
No	507	5	Abbastanza	184
Si	1095	5	Abbastanza	419

Fig. 30: Politica di sostituzione dei missing values per la variabile “grado divertimento allenamento”

Variabile: frequenza settimanale in giorni

Percentuale MV: 8,94%

Politica di sostituzione: valore medio arrotondato rispetto al grado di divertimento in allenamento

```
dropout_final.groupby('grado_divertimento_allenamento')['frequenza_settimanale'].mean().round()
```

```
grado_divertimento_allenamento
Abbastanza    3.0
Molto         3.0
Neutrale      3.0
Per niente    3.0
Poco          3.0
Name: frequenza_settimanale, dtype: float64
```

Fig. 31: Politica di sostituzione dei missing values per la variabile “grado divertimento allenamento”

Variabile: sport più significativo tra quelli praticati

Percentuale MV: 1,31%

Politica di sostituzione: valore più frequente rispetto al sesso

```
dropout_final.groupby('sesso')['sport_piu_significativo'].describe()
```

	count	unique	top	freq
sesso				
Femmina	1379	31	Pallavolo	245
Maschio	1335	28	Calcio	537

Fig. 32: Politica di sostituzione dei missing values per la variabile “sport più significativo”

Variabile: tipologia sport più significativo (individuale/squadra)

Percentuale MV: 8,69%

Politica di sostituzione: valore più frequente rispetto al sesso

```
dropout_final.groupby('sesso')['tipo_sport_significativo'].describe()
```

	count	unique	top	freq
sesso				
Femmina	1300	2	individuale	960
Maschio	1211	2	squadra	767

Fig. 33: Politica di sostituzione dei missing values per la variabile “tipologia sport più significativo”

Impegno Scolastico

Variabile: grado conciliare scuola sport

Percentuale MV: 8,47%

Politica di sostituzione: valore medio arrotondato su scala Likert 1-5 rispetto alla scuola superiore

```
dropout_final.groupby('scuola_superiore')['grado_conciliare_scuola_sport'].mean().round()  
  
scuola_superiore  
Altro (specificare)      4.0  
Istituto Professionale   3.0  
Istituto Tecnico        3.0  
Liceo                   3.0  
Name: grado_conciliare_scuola_sport, dtype: float64
```

Fig. 34: Politica di sostituzione dei missing values per la variabile “grado conciliare scuola sport”

Variabile: ore studio giornaliera

Percentuale MV: 9,27%

Politica di sostituzione: valore medio arrotondato rispetto alla scuola superiore

```
dropout_final.groupby('scuola_superiore')['ore_studio_giornaliera'].mean().round()  
  
scuola_superiore  
Altro (specificare)      2.0  
Istituto Professionale   2.0  
Istituto Tecnico        2.0  
Liceo                   2.0  
Name: ore_studio_giornaliera, dtype: float64
```

Fig. 35: Politica di sostituzione dei missing values per la variabile “ore studio giornaliera”

Genitori

Variabile: lavoro madre

Percentuale MV: 14,18%

Politica di sostituzione: valore più frequente rispetto al titolo di studio della madre

```
dropout_final.groupby('titolo_studio_madre')['lavoro_madre'].describe()
```

	count	unique	top	freq
titolo_studio_madre				
Altro (specificare)	28	7	Altro	10
Diploma di scuola media superiore (Istituto Tecnico o Professionale)	633	7	Impiegata	280
Diploma di scuola media superiore (Liceo)	461	7	Impiegata	160
Laurea	584	7	Docente	173
Licenza Elementare	43	6	Casalinga	24
Licenza Media Inferiore	501	7	Casalinga	228
Non lo so	110	7	Casalinga	38

Fig. 36: Politica di sostituzione dei missing values per la variabile “lavoro madre”

Variabile: impegno lavoro madre

Percentuale MV: 16,18%

Politica di sostituzione: valore più frequente rispetto al lavoro della madre

```
dropout_final.groupby('lavoro_madre')['impegno_lavoro_madre'].describe()
```

	count	unique	top	freq
lavoro_madre				
Altro	195	4	Full time	80
Casalinga	570	4	Autonomo	334
Commerciante/Imprenditrice/Artigiana	166	4	Autonomo	82
Docente	251	4	Full time	161
Impiegata	707	4	Full time	373
Libera professionista	255	4	Autonomo	104
Operaia	157	4	Full time	80

Fig. 37: Politica di sostituzione dei missing values per la variabile “impegno lavoro madre”

Variabile: lavoro padre

Percentuale MV: 14,29%

Politica di sostituzione: valore più frequente rispetto al titolo di studio del padre

```
dropout_final.groupby('titolo_studio_padre')['lavoro_padre'].describe()
```

	count	unique	top	freq
titolo_studio_padre				
Altro (specificare)	24	5	Operaio	7
Diploma di scuola media superiore (Istituto Tecnico o Professionale)	734	7	Impiegato	250
Diploma di scuola media superiore (Liceo)	252	7	Impiegato	92
Laurea	534	7	Libero professionista	238
Licenza Elementare	72	5	Operaio	31
Licenza Media Inferiore	643	6	Operaio	263
Non lo so	98	7	Operaio	27

Fig. 38: Politica di sostituzione dei missing values per la variabile "lavoro padre"

Variabile: impegno lavoro padre

Percentuale MV: 14,47%

Politica di sostituzione: valore più frequente rispetto al lavoro del padre

```
dropout_final.groupby('lavoro_padre')['impegno_lavoro_padre'].describe()
```

	count	unique	top	freq
lavoro_padre				
Altro	277	4	Full time	171
Casalingo	11	4	Autonomo	6
Commerciante/Imprenditore/Artigiano	353	4	Full time	222
Docente	65	4	Full time	41
Impiegato	586	4	Full time	467
Libero professionista	547	4	Full time	331
Operaio	505	4	Full time	389

Fig. 39: Politica di sostituzione dei missing values per la variabile "impegno lavoro padre"

Variabile: carattere genitore al seguito

Percentuale MV: 13,82%

Politica di sostituzione: valore più frequente rispetto all'accompagnamento frequente

```
dropout_final.groupby('frequenza_accompagnamento_genitori')['def_genitore'].describe()
```

		count	unique	top	freq
frequenza_accompagnamento_genitori					
	No	579	10	Tranquillo	194
	Si	1780	10	Tranquillo	559

Fig. 40: Politica di sostituzione dei missing values per la variabile "carattere genitore al seguito"

Descrizione Personalità

Variabile: grado personalità competitiva

Percentuale MV: 15,53%

Politica di sostituzione: valore medio arrotondato scala Likert 1-7 rispetto al caso di appartenenza

```
dropout_final.groupby('caso')['personalita_competitiva'].mean().round()
```

```
caso
1.1    4.0
1.2    5.0
2.1    5.0
2.2    5.0
3.0    4.0
Name: personalita_competitiva, dtype: float64
```

Fig. 41: Politica di sostituzione dei missing values per la variabile "grado personalità competitiva"

Il dataset ottenuto dopo la fase di gestione dei missing values ha le seguenti dimensioni:

Numero Righe : 2750

Numero Colonne : 127

ed è stato esportato nel file CSV "02 cleaning_mv.csv".

2. Encoding delle variabili

La fase di encoding ha l'obiettivo di trasformare le variabili presenti nel dataset ottenuto dalle fasi precedentemente descritte (pre-processing e gestione missing values) in un formato opportuno per rendere possibile l'esecuzione degli algoritmi di classificazione forniti dalla libreria *Scikit-Learn* di *Python*, e poter confrontare le performance di classificazione delle osservazioni rispetto al dropout sportivo.

Considerando che il dataset da predisporre per la classificazione contiene variabili di tipo categoriale o booleano, si è scelto di utilizzare due diversi *Encoder* forniti da *Scikit-Learn*: *OneHotEncoder* per le variabili categoriali con valori non in scala ordinale e *LabelEncoder* per le variabili booleane e per le variabili categoriali con valori in scala ordinale.

OneHotEncoder

Tramite un *OneHotEncoder* si trasforma una variabile categoriale con N valori non in scala ordinale in N variabili booleane (0/1), ognuna delle quali fa riferimento ad un valore della variabile da trasformare.

Facendo riferimento al dataset a disposizione, un esempio di variabile categoriale con N valori non in scala ordinale è quella contenente i dati relativi alla provenienza geografica del ragazzo che si è sottoposto al questionario.

Si tratta di 3 possibili valori ("*Nord*", "*Centro*", "*Sud*") non in scala ordinale per cui, trasformando la variabile "*provenienza*" mediante *one-hot encoding*, si ottengono tre variabili "*provenienza_nord*", "*provenienza_centro*" e "*provenienza_sud*".

Se l'osservazione ha valore "*Nord*" per la variabile "*provenienza*", allora la configurazione *one-hot* è: "*provenienza_nord*"=1, "*provenienza_centro*"=0, "*provenienza_sud*"=0.

Se l'osservazione ha valore "*Centro*" per la variabile "*provenienza*", allora la configurazione *one-hot* è: "*provenienza_nord*"=0, "*provenienza_centro*"=1, "*provenienza_sud*"=0.

Se l'osservazione ha valore "*Sud*" per la variabile "*provenienza*", allora la configurazione *one-hot* è: "*provenienza_nord*"=0, "*provenienza_centro*"=0, "*provenienza_sud*"=1.

def_genitore_sostenitore_positivo	def_genitore_tranquillo	def_genitore_troppo_fissato	tipo_sport_significativo_individuale	tipo_sport_significativo_squadra
0	0	0	1	0
0	0	0	1	0
0	1	0	0	1
0	0	0	0	1
0	1	0	0	1

Fig. 42: Prime 5 righe del dataset ottenuto dalla fase di encoding (OneHotEncoding)

LabelEncoder

Tramite un *LabelEncoder* si trasforma una variabile booleana (*No/Sì*) o una variabile categoriale con N valori in scala ordinale in una variabile numerica intera.

Nel caso della trasformazione per una variabile booleana, la variabile ottenuta assume valore 0 per il valore “No” e valore 1 per il valore “Sì”.

Fa parte di questo caso la variabile indicante il dropout sportivo per un’osservazione, la quale sarà utilizzata come variabile target degli algoritmi di classificazione, assumendo valore 0 per chi non ha abbandonato la pratica dello sport più significativo durante l’adolescenza, e valore 1 per chi ha abbandonato.

Nel caso della trasformazione di una variabile categoriale con N valori in scala ordinale, la variabile ottenuta assume un valore tra 0 e N-1, preservando il significato ordinale dei valori e predisponendoli ad essere interpretati correttamente dagli algoritmi di classificazione.

Facendo riferimento al dataset a disposizione, un esempio di variabile categoriale con N valori in scala ordinale è quella riferita al grado di divertimento del ragazzo che pratica sport durante gli allenamenti.

Tali valori sono: “Per Niente”, “Poco”, “Neutrale”, “Abbastanza”, “Molto” che, mediante il *label encoding*, vengono trasformati rispettivamente in 0, 1, 2, 3, 4, con lo scopo di trasformare la variabile in formato numerico e, contemporaneamente, rendere possibile l’interpretazione della scala ordinale da parte degli algoritmi di classificazione.

	disabilita	numero_fratelli_sorelle	numero_nonni_vicinanza_casa	cambio_residenza	presenza_campioni_sportivi_citta	arrampicata	arti_marziali	atletica
0	0	2	1	1	0	0	0	0
1	0	1	2	0	0	0	1	1
2	0	1	0	0	0	0	0	1
3	0	1	4	1	1	0	0	0
4	0	1	4	0	0	0	0	0

5 rows × 187 columns

<

Fig. 43: Prime 5 righe del dataset ottenuto dalla fase di encoding (LabelEncoding)

Per effetto della fase di encoding si ottiene un dataset che sarà utilizzato per il task di feature selection e per gli algoritmi di classificazione, avente le seguenti dimensioni:

Numero Righe: 2750

Numero Colonne: 186

Si registra un aumento delle colonne del dataset rispetto alla fase di gestione dei missing values (da 127 a 186), in quanto l'*encoding* per le variabili categoriali non in scala ordinale sostituisce 1 variabile con N variabili rappresentanti N possibili valori della variabile da trasformare.

Questo dataset, contenente variabili in formato numerico e quindi processabili dagli algoritmi di classificazione forniti dalla libreria *Scikit-Learn* di *Python*, è stato esportato nel file CSV **"03 encoding.csv"**.

Capitolo Secondo

Feature Selection

1. Decision Tree Learning

L'individuazione dei predittori più significativi viene effettuata tramite il task di *feature selection*, il quale ha il compito di definire le variabili più "importanti" rispetto alla predizione del dropout sportivo.

Per fare ciò, è necessario introdurre i concetti teorici alla base del *Decision Tree Learning* (*albero decisionale di apprendimento*), che verrà utilizzato per ricavare le feature più significative per il dropout sportivo rispetto ai dati a disposizione, usando la misura dell'*importanza*.

L'apprendimento dell'albero decisionale è uno degli approcci di modellazione predittiva utilizzati in statistica, data mining e machine learning.

Esso può essere utilizzato sia come *algoritmo di regressione*, in cui la variabile da predire è di tipo continuo, sia come *algoritmo di classificazione*, in cui la variabile da predire è di tipo discreto (classi).

Avendo come scopo la predizione del dropout sportivo ($No/Si \rightarrow 0/1$), l'apprendimento dell'albero decisionale viene utilizzato come algoritmo di classificazione.

I componenti fondamentali che costituiscono un albero decisionale sono:

- *nodo (o split)*: rappresenta una coppia (k, tk) dove k è una variabile e tk è un valore di split; partendo dal nodo radice dell'albero (*root*), le osservazioni del campione (*sample*) sono soggette ad un test determinato dalla variabile k e dal valore di split tk : a seconda del risultato del test (*True/False*), vengono creati due subset, per ognuno dei quali è possibile distinguere due gruppi di osservazioni in base al valore della variabile target, oggetto della classificazione;
- *ramo*: rappresenta il percorso da seguire, all'interno dell'albero, in seguito al risultato ottenuto del test effettuato su un nodo;
- *foglia (o leaf)*: rappresenta il valore predetto della variabile target, tenendo conto del cammino (*path*) effettuato dal nodo radice al nodo foglia; il valore predetto corrisponde alla classe della variabile target con più osservazioni valutata nel nodo foglia.

L'algoritmo predittivo utilizzato per la costruzione dell'albero decisionale di apprendimento è il *CART (Classification and Regression Tree)*, il quale ha le seguenti caratteristiche:

- ogni coppia (k, tk) che definisce il test da effettuare all'interno di un nodo dell'albero è selezionata con lo scopo di ottenere la migliore suddivisione delle osservazioni, minimizzando il valore di una funzione di costo legata all'impurità dei nodi "figli";
- una volta selezionata la coppia (k, tk) e suddiviso un nodo in due, lo stesso processo viene applicato a ciascun nodo "figlio" (si tratta di una procedura ricorsiva);
- la divisione si interrompe quando *CART* rileva che non è possibile ottenere ulteriori guadagni dalla minimizzazione della funzione di costo legata all'impurità o vengono soddisfatte alcune regole di arresto predefinite.

Come detto, la coppia (k, tk) di un nodo viene selezionata con lo scopo di produrre i due sottoinsiemi più "puri" di osservazioni, pesati in base al numero di osservazioni.

Un nodo si definisce "*puro*" se tutte le osservazioni a cui viene applicato appartengono alla stessa classe della variabile target.

Traslando questo concetto al dropout sportivo, un nodo si definisce puro quando tutte le osservazioni a cui viene applicato sono relative a soggetti che hanno abbandonato, oppure tutte le osservazioni sono relative a soggetti che non hanno abbandonato.

In presenza di un nodo "puro", l'algoritmo si arresta, ovvero non è più necessario selezionare una coppia (k, tk) per effettuare lo split.

La funzione di costo da minimizzare, legata all'impurità dei nodi "figli" ottenuti dallo split del nodo "padre", è la seguente:

$$J(k, t_k) = \frac{m_{\text{left}}}{m} G_{\text{left}} + \frac{m_{\text{right}}}{m} G_{\text{right}}$$

where $\begin{cases} G_{\text{left/right}} \text{ measures the impurity of the left/right subset,} \\ m_{\text{left/right}} \text{ is the number of instances in the left/right subset.} \end{cases}$

Fig. 44: Funzione di costo da minimizzare dell'algoritmo CART (classificazione)

L'impurità dei nodi "figli" viene calcolata tramite due possibili metriche:

- *Indice Gini*

$$G_i = 1 - \sum_{k=1}^n p_{i,k}^2$$

- $p_{i,k}$ is the ratio of class k instances among the training instances in the i^{th} node.

Fig. 45: Formula di calcolo dell'Indice Gini

Individuato il nodo i all'interno dell'albero, il valore dell'*Indice Gini* si ottiene sottraendo all'unità la somma dei quadrati del *ratio* di una determinata classe della variabile target, considerando tutte le classi possibili della variabile target.

Ipotizzando uno split da applicare ad un sottoinsieme di 100 osservazioni, delle quali 70 relative a chi non ha abbandonato (*dropout* = "No") e 30 relative a chi ha abbandonato (*dropout* = "Si"), il valore dell'*Indice Gini* è dato da:

$$1 - 0.7^2 - 0.3^2 = 0.42$$

- *Entropia*

$$H_i = - \sum_{\substack{k=1 \\ p_{i,k} \neq 0}}^n p_{i,k} \log(p_{i,k})$$

Fig. 46: Formula di calcolo dell'Entropia

Individuato il nodo i all'interno dell'albero, il valore dell'*Entropia* ("disordine") si ottiene invertendo il segno della somma dei prodotti tra il *ratio* di una determinata classe della variabile target e il suo logaritmo, considerando tutte le classi possibili della variabile target.

Ipotizzando uno split da applicare ad un sottoinsieme di 100 osservazioni, delle quali 70 relative a chi non ha abbandonato (*dropout* = "No") e 30 relative a chi ha abbandonato (*dropout* = "Si"), il valore dell'*Entropia* è dato da:

$$- 0.7 * \log(0.7) - (0.3) * \log(0.3) = 0.25 + 0.36 = 0.61$$

Per entrambe le metriche, il valore 0 corrisponde ad un nodo “*puro*”, in presenza del quale tutte le osservazioni del nodo sono relative a chi ha abbandonato, oppure tutte le osservazioni del nodo sono relative a chi non ha abbandonato: l’algoritmo si arresta.

Per evitare il fenomeno dell’*overfitting*, ovvero l’ottenimento di un modello di apprendimento che fornisce performance elevate sui dati utilizzati per l’allenamento (*training data*) senza generalizzare per nuovi dati (*test data*), è necessario regolarizzare i parametri che determinano la conformazione dell’albero decisionale.

Con la regolarizzazione, l’algoritmo *CART* si arresta prima del raggiungimento di impurità nulla per i nodi terminali dell’albero, scongiurando l’*overfitting*.

I parametri scelti per la regolarizzazione dell’albero decisionale sono:

- *max_depth*: massima profondità dell’albero, ovvero il numero massimo di split per un ramo, partendo dal nodo radice (*root*);
- *min_samples_leaf*: numero minimo di osservazioni necessario per eseguire uno split.

Tramite la regolarizzazione dell’albero decisionale mediante questi parametri, l’algoritmo *CART* si arresta nel momento in cui l’albero raggiunge la profondità specificata dal parametro *max_depth* oppure nel momento in cui il numero di osservazioni dell’ultimo nodo considerato è inferiore a quello specificato dal parametro *min_samples_leaf*.

Di seguito, viene riportato in dettaglio l’*albero decisionale di apprendimento* per la predizione del dropout sportivo, ottenuto massimizzando la metrica “*recall*” mediante la procedura di *cross-validation* (10-Fold) e la regolarizzazione dei parametri del modello.

I valori dei parametri ottimizzati sono:

- *misura di impurità*: Indice Gini
- *max_depth* = 3
- *min_samples_leaf* = 0.05 (5% delle osservazioni totali)

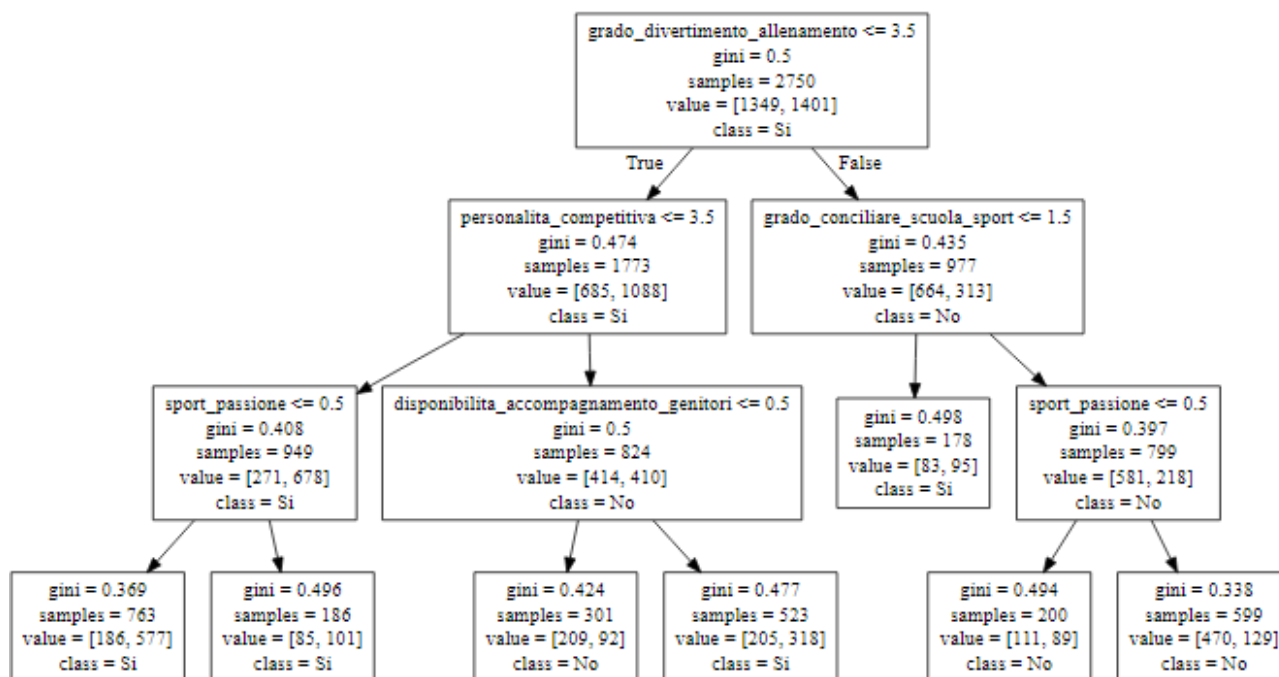


Fig. 47: Albero decisionale di apprendimento ottenuto con regolarizzazione dei parametri

Il nodo radice dell'albero definisce lo split delle osservazioni (2750) in base al test effettuato sulla variabile *grado_divertimento_allenamento*.

Nel ramo sinistro dell'albero vengono propagate le osservazioni per cui questa variabile assume valore inferiore a 3.5 (1773 obs), ovvero coloro che hanno scelto una tra le opzioni "Per niente", "Poco", "Neutrale" e "Abbastanza" (rispettivamente codificate con i valori da 0 a 3) alla domanda su quanto si divertono durante l'allenamento.

Nel ramo destro, invece, vengono propagate le osservazioni per cui questa variabile assume valore superiore a 3.5 (977 obs), ovvero coloro che hanno scelto l'opzione "Molto" (codificata con il valore 4) alla domanda su quanto si divertono durante l'allenamento.

Considerato che l'insieme delle osservazioni a cui viene applicato lo split (corrispondente all'intero dataset dato che si tratta del *root*), è composto da 1349 osservazioni di chi non ha abbandonato e 1401 di chi ha abbandonato, l'Indice Gini è calcolato in questo modo:

$$1 - (1349 / 2750)^2 - (1401 / 2750)^2 = 0.5$$

Nel nodo “figlio” presente sul ramo sinistro, le 1773 osservazioni sono distinte in 685 di chi non ha abbandonato e 1088 di chi ha abbandonato, con un Indice Gini pari a 0.474.

A livello predittivo, si deduce che la probabilità di abbandono è maggiore per chi non si diverte molto durante gli allenamenti.

Nel nodo “figlio” presente sul ramo destro, le 977 osservazioni sono distinte in 664 di chi non ha abbandonato e 313 di chi ha abbandonato, con un Indice Gini pari a 0.435.

A livello predittivo, si deduce che la probabilità di abbandono è minore per chi si diverte molto durante gli allenamenti.

Disponendo di queste informazioni, possiamo ricavare il valore della funzione di costo legata all’impurità dei nodi “figli”, minimizzata dall’algoritmo *CART* per selezionare la coppia (k, tk) del nodo radice.

$$\begin{aligned} J(k, tk) &= (m_{\text{left}} / m) * G_{\text{left}} + (m_{\text{right}} / m) * G_{\text{right}} \\ &= (1773 / 2750) * 0.474 + (977 / 2750) * 0.435 = 0.4595 \end{aligned}$$

Avendo scelto la coppia $(\text{grado_divertimento_allenamento}, 3.5)$, si deduce che qualsiasi altra coppia utilizzata come test per lo split produce un valore della funzione di costo legata all’impurità dei nodi “figli” maggiore di 0.4595.

Questo procedimento viene effettuato in modo ricorsivo fino all’arresto dell’algoritmo, il quale avviene nel momento in cui viene raggiunta la profondità massima di 3 nodi, settata dal parametro *max_depth*, oppure quando il numero di osservazioni del sottoinsieme considerato è inferiore al 5% delle osservazioni totali (inferiore a $2750 * 0.05 = 137.5$ obs), settato dal parametro *min_samples_leaf*.

2. Feature Importance

Utilizzando il *Decision Tree Learning* (albero decisionale di apprendimento) è possibile implementare un task di *feature selection filter-based*, ovvero all'interno dell'insieme delle *features* (variabili) del dataset a disposizione, vengono selezionate le *features* più significative rispetto alla predizione del dropout in base al valore di *importanza* che rivestono nell'albero decisionale.

La *feature importance* è calcolata come la diminuzione dell'impurità ottenuta da uno split relativo a quella *feature* ponderata dalla probabilità di raggiungere quello split.

La probabilità di raggiungere uno split viene calcolata tramite il rapporto tra il numero di osservazioni che raggiungono lo split e il numero totale di osservazioni.

Si può affermare, quindi, che una *feature* risulta più importante di un'altra nel caso in cui lo split per il quale viene scelta produce sottoinsieme più "puri" rispetto ai sottoinsiemi prodotti dall'altra *feature*, ricordando che la purezza di un sottoinsieme aumenta all'aumentare della differenza numerica tra osservazioni che appartengono a classi diverse della variabile target.

Maggiore è il valore della *feature importance*, più importante è il ruolo predittivo di quella *feature* rispetto alla variabile target dell'albero decisionale.

Preliminarmente al calcolo dell'importanza di una *feature*, è necessario definire il calcolo dell'importanza di un *nodo* (o *split*) all'interno dell'albero decisionale:

$$ni_j = w_j C_j - w_{left(j)} C_{left(j)} - w_{right(j)} C_{right(j)}$$

- $ni_{sub(j)}$ = the importance of node j
- $w_{sub(j)}$ = weighted number of samples reaching node j
- $C_{sub(j)}$ = the impurity value of node j
- $left(j)$ = child node from left split on node j
- $right(j)$ = child node from right split on node j

Fig. 48: Formula del calcolo dell'importanza di un nodo

Facendo riferimento alla Fig. 47, che descrive l'albero decisionale ottenuto per la classificazione del dropout, viene riportato il calcolo dell'importanza del nodo radice j:

- **W (j)** = 2750 / 2750 = 1 : numero pesato di osservazioni che raggiungono il nodo j (trattandosi del nodo radice, a tutte le osservazioni del dataset viene applicato lo split)
- **C (j)** = IndiceGini (j) = 0.5 : impurità del nodo j
- **W left (j)** = 1773 / 2750 = 0.645 : numero pesato di osservazioni che raggiungono il nodo "figlio" del nodo j presente nel ramo sinistro
- **C left (j)** = IndiceGini left (j) = 0.474 : impurità del nodo "figlio" del nodo j presente nel ramo sinistro
- **W right (j)** = 977 / 2750 = 0.355 : numero pesato di osservazioni che raggiungono il nodo "figlio" del nodo j presente nel ramo destro
- **C right (j)** = IndiceGini right (j) = 0.435 : impurità del nodo "figlio" del nodo j presente nel ramo destro

Node Importance ni (j) = $1 * 0.5 - 0.645 * 0.474 - 0.355 * 0.435 = 0.041$

Dopo aver calcolato l'importanza di ogni nodo presente nell'albero decisionale, è possibile calcolare l'importanza di ogni *feature* con la seguente formula:

$$fi_i = \frac{\sum_{j: \text{node } j \text{ splits on feature } i} ni_j}{\sum_{k \in \text{all nodes}} ni_k}$$

- $fi_{\text{sub}(i)}$ = the importance of feature i
- $ni_{\text{sub}(j)}$ = the importance of node j

Fig. 49: Formula del calcolo dell'importanza di una feature

Facendo riferimento alla figura X, che descrive l'albero decisionale ottenuto per la classificazione del dropout, viene riportato il calcolo dell'importanza della feature *grado_divertimentoAllenamento* (indicata con i), la quale viene utilizzata solo nel nodo radice (indicato con j):

- $\sum n_i(j) = 0.041$: somma delle importanze dei nodi j che utilizzano la feature i (se la feature i fosse presente in altri nodi, bisognerebbe calcolare la somma delle importanze dei nodi in cui è presente)
- $\sum n_i(k)$: somma delle importanze di tutti i k nodi dell'albero

Feature Importance $f_i(i) = 0.041 / \sum n_i(k)$

Il valore così ottenuto viene *normalizzato* nel range [0, 1], dividendolo per la somma delle importanze di tutte le *features*.

$$norm f_i = \frac{f_i}{\sum_{j \in \text{all features}} f_j}$$

Fig. 50: Formula della normalizzazione dell'importanza di una feature

Facendo riferimento alla feature *grado_divertimentoAllenamento*, il valore normalizzato della sua importanza è pari a:

$$norm f_i(i) = f_i(i) / \sum f_i(j)$$

La somma delle importanze normalizzate delle *features* è pari a 1: più il valore dell'importanza normalizzata di una *feature* è vicino a 1 più la variabile è importante per la classificazione della variabile target.

Di seguito, vengono riportati i valori di importanza delle *features* relativi all'albero decisionale di apprendimento utilizzato per la classificazione del dropout sportivo.

Numero variabili importanti: 5

	feature	importance
99	grado_divertimento_allenamento	0.463067
108	personalita_competitiva	0.177281
97	disponibilita_accompagnamento_genitori	0.149300
74	sport_passione	0.125684
90	grado_conciliare_scuola_sport	0.084667

Fig. 51: Elenco variabili più importanti dell'albero decisionale

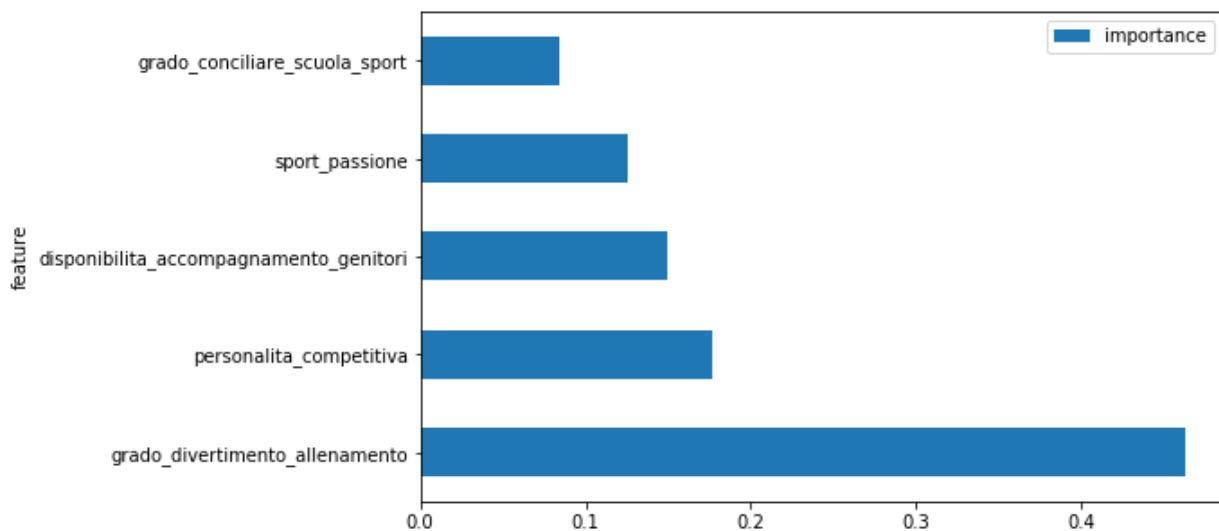


Fig. 52: Grafico variabili più importanti dell'albero decisionale

Il primo predittore più significativo del dropout sportivo è la variabile *grado_divertimento_allenamento* (*feature importance* = 0.46): nell'albero decisionale è presente un solo split che la utilizza, ossia il nodo radice.

Dalla distribuzione dell'abbandono nelle osservazioni dei nodi "figli", si deduce che la probabilità di abbandono è maggiore per chi si diverte poco durante gli allenamenti e minore per chi si diverte molto durante gli allenamenti.

Il secondo predittore più significativo del dropout sportivo è la variabile *personalita_competitiva* (*feature importance* = 0.18): nell'albero decisionale è presente un solo split a profondità 1 che la utilizza, subordinato al nodo del *grado_divertimento_allenamento*.

Supponendo, quindi, il poco divertimento durante gli allenamenti, si deduce che la probabilità di abbandono è maggiore per chi ha una bassa personalità competitiva e minore per chi ha un'elevata personalità competitiva.

Il terzo predittore più significativo del dropout sportivo è la variabile *disponibilita_accompagnamento_genitori* (*feature importance* = 0.15): nell'albero decisionale è presente un solo split a profondità 2 che la utilizza, subordinato al nodo del *grado_divertimentoAllenamento* (root) e al nodo della *personalita_competitiva* (depth=1).

Supponendo, quindi, il poco divertimento durante gli allenamenti e la bassa personalità competitiva tra chi si diverte poco agli allenamenti, si deduce che la probabilità di abbandono è maggiore per chi viene frequentemente accompagnato dai genitori ad allenarsi e minore per chi non viene frequentemente accompagnato.

Quest'ultima deduzione è riconducibile a casi in cui, accompagnare frequentemente ad allenarsi un ragazzo, che si diverte poco e con scarsa competizione, ha la conseguenza di amplificare la mancanza di indipendenza del ragazzo e di recepire l'allenamento come un obbligo, anziché di svolgere un ruolo di supporto.

Gli altri due predittori significativi del dropout sportivo sono: considerare o meno fare sport una passione (0.13) e il grado con il quale si riesce a conciliare l'impegno scolastico con l'impegno sportivo (0.08).

3. Feature Selection tramite Random Forest

L'individuazione dei predittori più significativi descritta nel paragrafo precedente deriva dall'apprendimento automatico basato su un singolo *Decision Tree Learning* con regolarizzazione dei parametri che conferiscono la struttura dell'albero.

Con lo scopo di generalizzare maggiormente il processo di apprendimento automatico e, quindi, determinare l'importanza delle variabili in modo più generalizzato, si utilizza un *algoritmo ensemble* chiamato *Random Forest (foresta casuale)*, che consiste nell'aggregare gli apprendimenti di più *Decision Tree Learning*.

La scelta di far uso di un *algoritmo ensemble* deriva dal concetto "*wisdom of the crowd*" (*saggezza della folla*), secondo cui la risposta aggregata ad una domanda complessa posta a migliaia di persone a caso è migliore rispetto ad una singola risposta di un esperto.

Allo stesso modo, aggregando le previsioni di un gruppo di predittori, è possibile ottenere una previsione migliore rispetto al miglior predittore individuale.

L'algoritmo *Random Forest* di classificazione è uno dei più potenti algoritmi di Machine Learning disponibili oggi, che consiste nell'ottenere le previsioni di tutti i singoli *Decision Tree Learning* che compongono la "foresta", per poi predire la classe della variabile target che ottiene il maggior numero di voti (*statistical mode*).

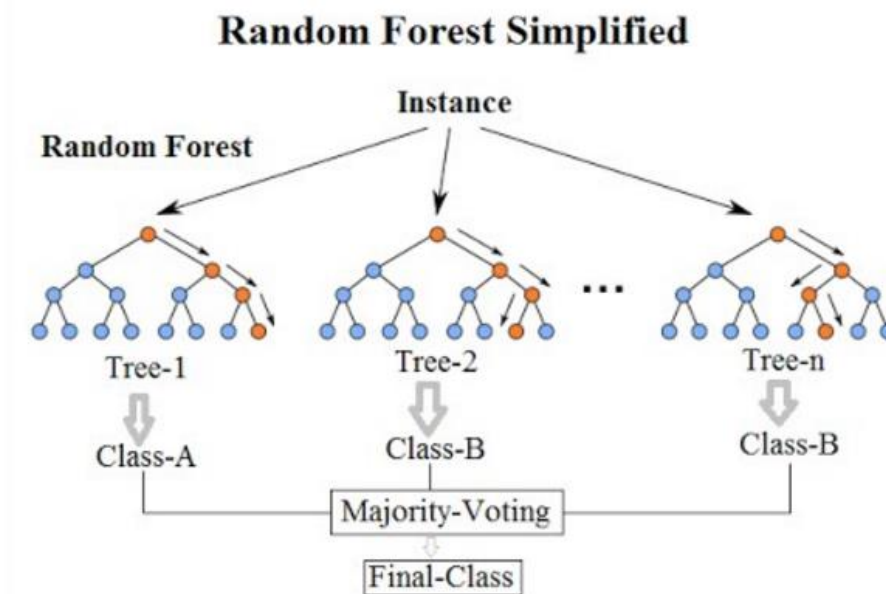


Fig. 53: Algoritmo Ensemble Random Forest

Il termine “casuale”, invece, è riferito alla composizione casuale del sottoinsieme dei dati su cui ogni *Decision Tree Learning* viene addestrato.

Quando il campionamento delle osservazioni viene eseguito con la tecnica “*with replacement*”, ovvero un’osservazione del set di training può far parte di più sottoinsiemi usati per addestrare i vari alberi decisionali della foresta, allora si tratta del metodo *Bagging* (*bootstrap aggregation*).

In caso contrario, il metodo di campionamento prende il nome di *Pasting*. Sia tramite il *Bagging* che il *Pasting*, è possibile anche campionare le features, nello stesso modo con cui vengono campionate le osservazioni.

Secondo questa logica, ogni *Decision Tree Learning* che compone la “foresta” viene addestrato su un sottoinsieme di dati derivante dal campionamento delle osservazioni e dal campionamento delle features (*Random Patch*), conferendo la randomizzazione dell’algoritmo di predizione.

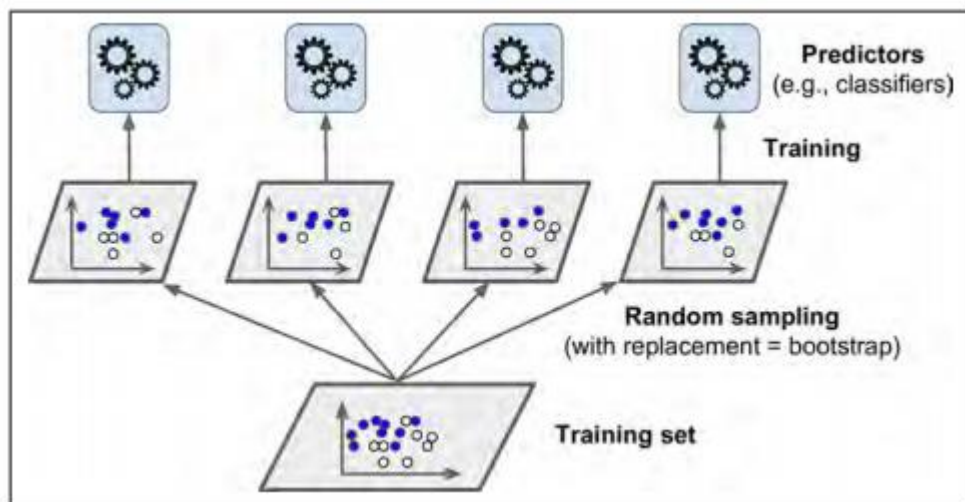


Fig. 54: Rappresentazione grafica dei metodi di Bagging/Pasting

Per quanto riguarda la scelta degli split all’interno dei vari *Decision Tree Learning* che compongono il *Random Forest*, rimane invariata la logica utilizzata dall’algoritmo *CART* nella scelta della variabile k e del valore di split tk , minimizzando la funzione di costo legata all’impurità dei nodi “figli” ottenuti dallo split.

Data la randomizzazione conferita dal *Bagging* o dal *Pasting*, sia per le osservazioni sia per le features, ogni *Decision Tree Learning* avrà split diversi a diverse profondità dell’albero.

Con lo scopo di individuare i predittori più significativi del dropout sportivo rispetto alle variabili a disposizione, è stato utilizzato un algoritmo *Random Forest* con i seguenti parametri, ottenuti mediante *cross-validation 10-Fold* con l'obiettivo di massimizzare la metrica "*recall*" sulle performance di predizione dell'abbandono:

Parametri di un singolo albero decisionale di apprendimento:

- *criterion* = "gini" (misura di impurità dei nodi)
- *max_depth* = 3 (massima profondità dell'albero)
- *min_samples_leaf* = 0.05 (minimo 5% delle osservazioni totali per un nodo foglia)

Parametri della "foresta":

- *n_estimators* = 500 (numero alberi decisionali addestrati)
- *bootstrap* = False (metodo di campionamento delle osservazioni → in questo caso *Pasting*)

La scelta di massimizzare la *recall* è dovuta alla scelta di limitare l'incidenza dei *falsi negativi* sulle performance, ovvero la classificazione delle osservazioni come "non abbandono" quando in realtà l'abbandono si è verificato.

Il procedimento di calcolo dell'importanza di una feature, descritto nel paragrafo 2, necessita di una variazione: per ogni feature, il valore di importanza viene calcolato facendo la media delle importanze relative agli alberi decisionali della foresta in cui quella feature viene scelta per uno split, pesando le diverse profondità.

Numero variabili importanti: 21

	feature	importance
99	grado_divertimentoAllenamento	0.153467
108	personalitaCompetitiva	0.111062
86	agonismo	0.088341
74	sportPassione	0.083153
90	gradoConciliareScuolaSport	0.052057
114	SessoMaschio	0.051652
11	calcio	0.048884
113	SessoFemmina	0.042952
109	personalitaTranquilla	0.031205
40	divertimento	0.029188
55	esigenza	0.027264
72	sportDivertimento	0.024484
111	maschioCalcio	0.021318
101	personalitaPolemica	0.019241
93	sportAgonisticoGenitori	0.017211
184	tipoSportSignificativoSquadra	0.016239
87	frequenzaSettimanale	0.015837
183	tipoSportSignificativoIndividuale	0.014784
75	sportBisognoCorpoMente	0.014495
104	personalitaAperta	0.011559
112	femminaPallavoloNuoto	0.011520

Somma valori di importanza: 0.8859129029470975

Fig. 55: Elenco variabili più importanti della foresta casuale

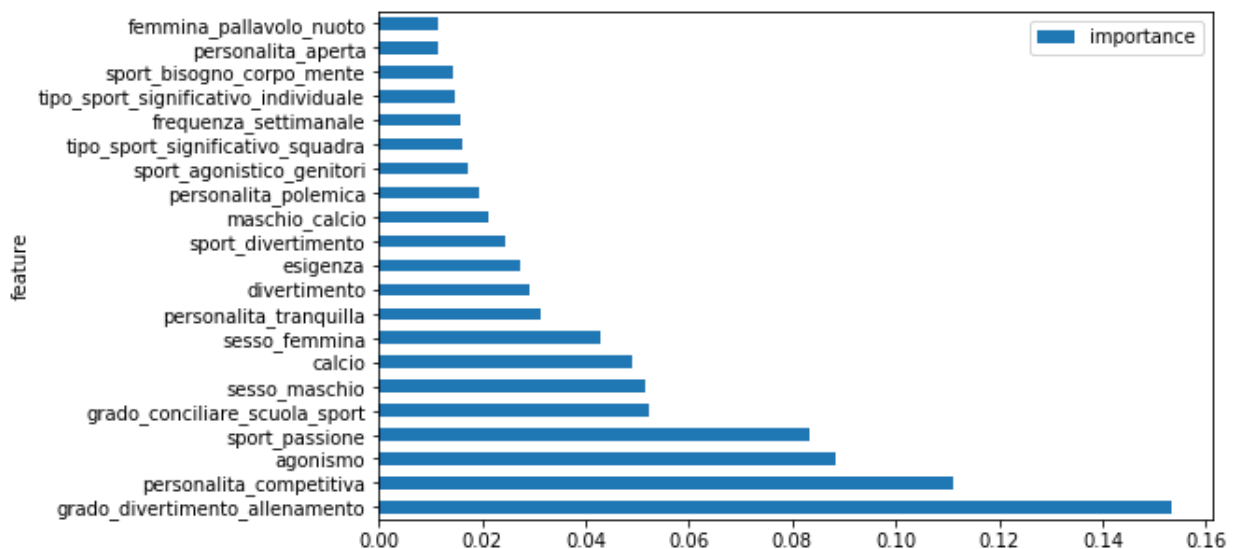


Fig. 56: Rappresentazione grafica variabili più importanti della foresta casuale

L'apprendimento dell'importanza delle variabili tramite *Random Forest* consente di individuare nuovi predittori significativi del dropout sportivo, oltre a quelli individuati tramite singolo *Decision Tree*, e di generalizzare l'importanza senza sovrastimarla o sottostimarla.

Il numero di variabili importanti ottenute con il *Random Forest* (21) è maggiore rispetto al numero di variabili importanti ottenute con l'apprendimento di un singolo *Decision Tree* (5).

Le 21 variabili selezionate come le più importanti dal *Random Forest* hanno un valore di importanza maggiore di 0.01 all'interno del range [0, 1] e rappresentano l'88% dell'importanza globale.

Il primo predittore più significativo del dropout sportivo resta la variabile che indica il *grado di divertimento durante gli allenamenti*: il suo valore di importanza si riduce, però, da 0.46 a 0.15, rilevando che l'apprendimento mediante singolo albero decisionale sovrastimava la sua importanza a livello predittivo.

Anche il secondo predittore più significativo resta invariato: si tratta del *livello di personalità competitiva*, con un valore di importanza ridotto da 0.18 a 0.11.

Il terzo predittore più significativo è *praticare sport a livello agonistico*, con valore di importanza pari a 0.09.

Procedendo, si individuano altri predittori significativi in relazione al dropout sportivo ma meno importanti: considerare lo sport come una passione (0.08), il grado con il quale si riesce a conciliare scuola e sport (0.05), il sesso maschile (0.05), praticare calcio (0.05) e altri in Fig. 55.

E' necessario specificare che valutare l'importanza di una variabile significa individuare i predittori più significativi del dropout sportivo, senza poter determinare se quella variabile è legata positivamente o negativamente al dropout.

Questa ulteriore conclusione può essere derivata ispezionando gli split che utilizzano la variabile nell'insieme degli alberi decisionali che compongono la foresta.

Il task di *feature selection* si conclude selezionando tutte le variabili con valore di importanza maggiore di 0.01 riportate in Fig. 55 (21 variabili), ovvero le variabili più importanti ottenute dall'apprendimento mediante *Random Forest*.

Il dataset ottenuto dalla fase precedente di *encoding* è stato filtrato, mantenendo solo le variabili ottenute dal task di *feature selection* ed è stato esportato nel file CSV "**04 important_fetures.csv**" (n° righe: 2750, n° colonne: 22).

Esso sarà utilizzato come dataset di input per gli ulteriori algoritmi di classificazione, i quali verranno confrontati tra loro, includendo anche il *Random Forest* descritto in questo paragrafo, per valutare le performance di classificazione.

ALGORITMI DI CLASSIFICAZIONE

Capitolo Primo

Modelli Predittivi

1. *Cross-Validation con tuning dei parametri*

In questo paragrafo viene descritta la tecnica di *Cross-Validation*, la quale in relazione alla predizione del dropout sportivo, viene utilizzata per addestrare alcuni *algoritmi di classificazione di tipo supervisionato*.

Si tratta di algoritmi che predicono una classe della variabile target per ogni osservazione del dataset, per poi essere soggetti alla valutazione delle performance sulla base del confronto tra valore predetto e valore vero (*label*).

L'apprendimento automatico di un algoritmo di Machine Learning (*addestramento*) viene effettuato su un sottoinsieme dell'intero set dei dati a disposizione, utilizzato con lo scopo di far "imparare" all'algoritmo i pattern nascosti nei dati: questo sottoinsieme è chiamato *Training Set* e costituisce il sottoinsieme più grande in cui viene diviso il dataset.

Dopo aver addestrato il modello sul *Training Set*, esso viene utilizzato per predire il valore della variabile target della classificazione per le osservazioni appartenenti all'altro sottoinsieme in cui è stato diviso l'intero set dei dati a disposizione, utilizzato con lo scopo di valutare le performance del modello su dati "non visti" durante l'allenamento: questo sottoinsieme è chiamato *Test Set* e costituisce il sottoinsieme più piccolo in cui viene diviso il dataset.

Un modello sviluppato con la tecnica precedentemente descritta (*Holdout*) non risulta essere un modello generalizzato, in quanto le sue performance dipendono sensibilmente dalla modalità con cui il dataset viene diviso in *Training Set* e *Test Set*, aumentando il rischio sia di *overfitting* (sovrastima della performance) sia di *underfitting* (sottostima della performance).

Per ovviare a questo problema, viene utilizzata una tecnica alternativa chiamata *Cross-Validation*, che consiste nel dividere l'intero set di dati in K sottoinsiemi contenenti lo stesso numero di osservazioni (*fold*) e nel selezionare, di volta in volta, una *k-esima fold (validation)* per

valutare le performance del modello addestrato sulle osservazioni contenute nelle altre $k-1$ *fold* (*training*).

In questa maniera, ogni osservazione dell'intero set di dati a disposizione viene utilizzata $k-1$ volte all'interno del *Training Set* e una volta anche all'interno del *Validation Set*, generalizzando il modello addestrato.

Si dimostra che valori ottimali del parametro K sono 3, 5 o 10.

Mediante la *Cross-Validation* con *K-Fold*, si ottengono K valori per ogni metrica di performance scelta per valutare le performance dell'algoritmo di apprendimento automatico, consentendo il confronto tra diversi algoritmi mediante valore medio, deviazione standard e costruzione degli intervalli di confidenza.

Un altro utile accorgimento da prendere, con lo scopo di generalizzare al meglio il modello predittivo, è quello di *stratificare* ogni fold in base alla distribuzione delle classi della variabile target nell'intero set di dati.

Il processo di *Cross-Validation* consente, anche, di ottenere la migliore configurazione dei parametri interni al modello predittivo (*hyperparameters*) all'interno di una griglia precedentemente definita (*tuning dei parametri*).

Ogni modello predittivo ottenuto con lo stesso algoritmo di apprendimento automatico ma con differente configurazione dei parametri è soggetto al processo di *Cross-Validation*: successivamente, si definisce il miglior modello (*best estimator*) e la migliore configurazione dei parametri (*best params*), ottenuti massimizzando una determinata metrica di performance.

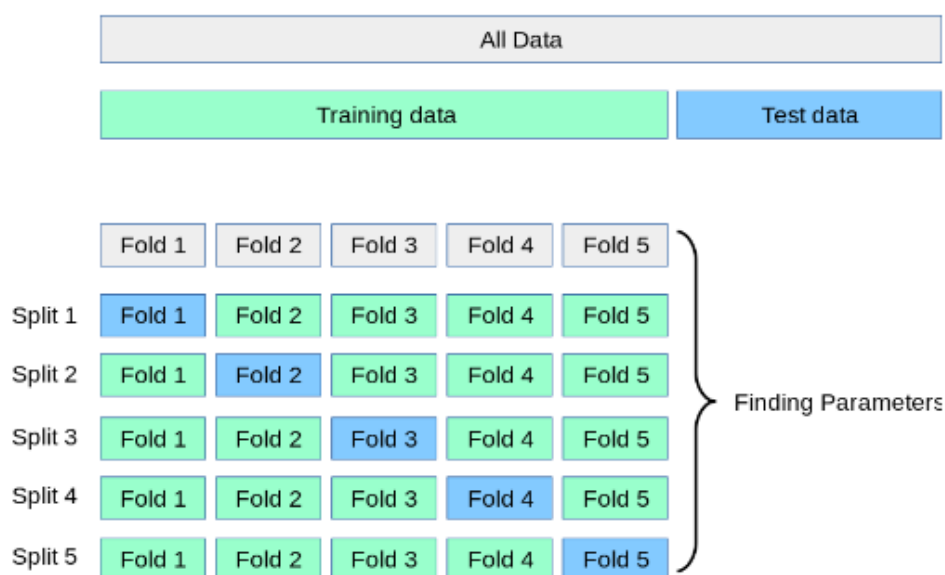


Fig. 57: Processo di *Cross-Validation* 5-Fold con *tuning dei parametri* di un modello

2. Logistic Regression

La *regressione logistica* (*Logistic Regression*), utilizzata come algoritmo di classificazione, effettua la stima della probabilità per un'osservazione di appartenere ad una determinata classe della *variabile target*, applicando la funzione “*logit*” alla somma pesata delle *features* di input, dette *regressori* o *variabili esplicative*.

Facendo riferimento alla classificazione del dropout sportivo tra gli adolescenti, questo modello costituisce un *classificatore binario* (*binary classifier*) che opera nel seguente modo:

- se la probabilità stimata è maggiore del 50%, si predice che l'osservazione appartiene alla “classe positiva”, ovvero viene predetto il dropout (1);
- se la probabilità stimata è minore del 50%, si predice che l'osservazione appartiene alla “classe negativa”, ovvero viene predetto il non dropout (0).

Di seguito viene riportata l'equazione per il calcolo della stima della probabilità in forma vettoriale:

$$\hat{p} = h_{\theta}(\mathbf{x}) = \sigma(\theta^T \cdot \mathbf{x})$$

Fig. 58: Stima probabilità del modello logistico in forma vettoriale

Di seguito, invece, viene riportata la funzione logistica $\sigma(\cdot)$ utilizzata nell'equazione in Fig. 58: si tratta di una *sigmoid function* applicata alla matrice dei *regressori* che fornisce un valore compreso tra 0 e 1, ovvero la stima della probabilità per un'osservazione di appartenere alla “classe positiva”

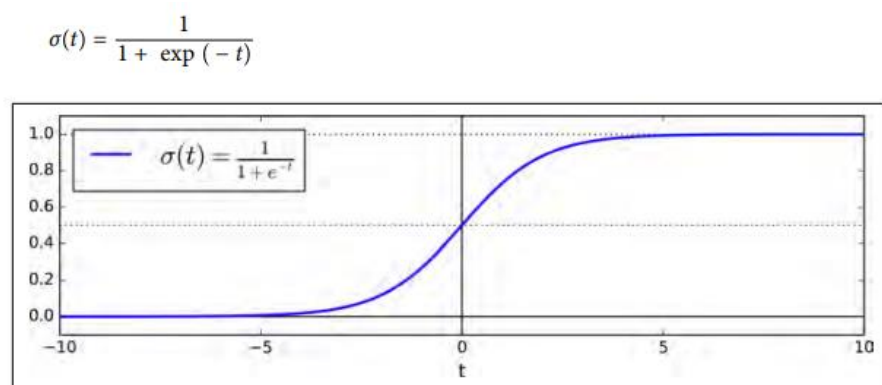


Fig. 59: Equazione e rappresentazione grafica della funzione logistica

Dopo aver determinato la stima della probabilità, si procede a predire la classe della variabile target per ogni osservazione nel seguente modo:

$$\hat{y} = \begin{cases} 0 & \text{if } \hat{p} < 0.5, \\ 1 & \text{if } \hat{p} \geq 0.5. \end{cases}$$

Fig. 60: Valore predetto tramite stima della probabilità

Il processo di apprendimento (*training*) del modello predittivo ha l'obiettivo di stimare il vettore θ , ovvero i valori dei parametri dei regressori nell'equazione in Fig. 58, con lo scopo di ottenere alti valori di stima della probabilità per le osservazioni appartenenti alla "classe positiva" ($y = 1$) e bassi valori di stima della probabilità per le osservazioni appartenenti alla "classe negativa" ($y = 0$).

Per questo viene utilizzata una *funzione di costo* $c(\theta)$ definita in questo modo per una singola osservazione:

$$c(\theta) = \begin{cases} -\log(\hat{p}) & \text{if } y = 1, \\ -\log(1 - \hat{p}) & \text{if } y = 0. \end{cases}$$

Fig. 61: Funzione di costo per una singola osservazione

Tenendo presente l'andamento decrescente della funzione $-\log(t)$ per valori di t compresi tra 0 e 1, si evince che la funzione di costo aumenta quando la stima della probabilità si avvicina allo 0 per le osservazioni appartenenti alla "classe positiva" ($y = 1$), oppure quando la stima della probabilità si avvicina a 1 per le osservazioni appartenenti alla "classe negativa" ($y = 0$).

La *funzione di costo* per l'intero set di allenamento (*Training Set*) si ottiene calcolando la media tra i valori della funzione di costo delle singole osservazioni:

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^m \left[y^{(i)} \log(\hat{p}^{(i)}) + (1 - y^{(i)}) \log(1 - \hat{p}^{(i)}) \right]$$

Fig. 62: Funzione di costo nel Training Set

Si dimostra che la funzione di costo $J(\theta)$ è convessa, per cui è possibile utilizzare un algoritmo di ottimizzazione per poter trovare il vettore dei parametri θ (coefficienti dei regressori) che rappresenta il minimo globale della funzione di costo $J(\theta)$.

Il modello predittivo *Logistic Regression* per il dropout sportivo è stato così ottenuto:

- *variabili esplicative*: 21 features derivanti dal task di *feature selection*;
- *variabile target*: dropout [“No” → classe negativa (0), “Sì” → classe positiva (1)];
- apprendimento mediante *Cross-Validation 10-Fold* (275 obs per fold) con tuning degli *hyperparameters*

Il tuning dei parametri interni al modello (*hyperparameters*) è stato effettuato considerando la seguente griglia di valori dei parametri:

- $C = [0.01, 0.1, 1, 10, 100]$: *inverse regularization parameter*
- **penalty** = [“l1”, “l2”]: termine di penalità utilizzato nella regolarizzazione

La configurazione migliore dei parametri ottenuta è [$C=0.1$, $penalty=“l1”$], sulla base della massimizzazione della metrica *recall* dell’algoritmo di classificazione: con questa scelta progettuale, è stato scelto di limitare l’incidenza dei *falsi negativi* sulle performance del modello, dove per *falsi negativi* si intendono le osservazioni predette come “non abbandonano” quando in realtà sono classificate come “abbandono” sportivo.

La valutazione delle performance del modello ottenuto è descritta nel capitolo successivo.

Di seguito, invece, viene riportata l’analisi delle stime della probabilità, tenendo conto di un modello di *regressione logistica* avente come regressori i tre predittori più significativi individuati dal task di *feature selection* del capitolo precedente:

- X_1 : grado di divertimento allenamento (*Per Niente, Poco, Neutrale, Abbastanza, Molto*)
- X_2 : personalità competitiva (*scala Likert da 1 a 7*)
- X_3 : pratica agonistica (*No/Sì*)

Si tratta di un modello in overfitting, utilizzato per comprendere l’incidenza di questi fattori all’interno del campione.

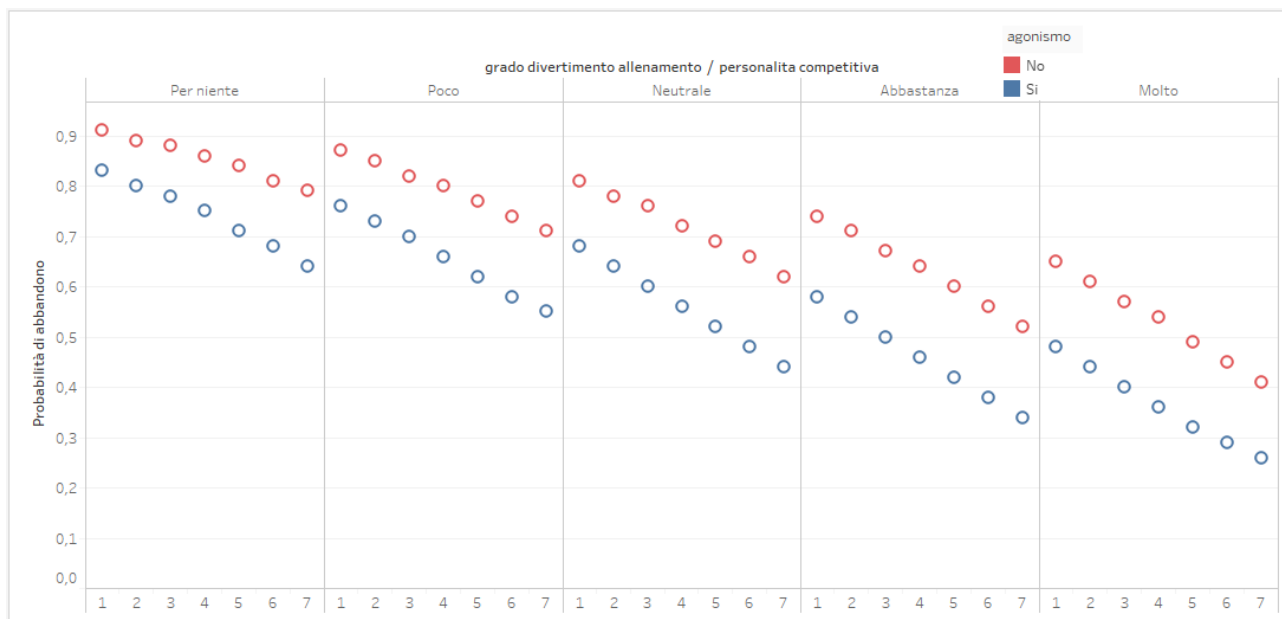


Fig. 63: Modello Logistico: confronto stime probabilità dropout sui tre predittori più significativi

Dalla rappresentazione grafica in Fig. 63 si evincono chiaramente tre pattern significativi:

- 1) la probabilità di dropout diminuisce all'aumentare del grado di divertimento durante l'allenamento; inoltre, fissato un grado di divertimento, diminuisce all'aumentare del livello di personalità competitiva ed è inferiore se si pratica sport a livello agonistico;
- 2) la probabilità di dropout diminuisce all'aumentare del livello di personalità competitiva; inoltre, fissato un livello di personalità competitiva, diminuisce all'aumentare del grado di divertimento durante l'allenamento ed è inferiore se si pratica sport a livello agonistico;
- 3) la probabilità di dropout è inferiore se si pratica sport a livello agonistico, per ogni grado di divertimento di allenamento e per ogni livello di competitività.

In riferimento a questi tre predittori, la probabilità di dropout maggiore, pari a 0.91, si ha quando il grado di divertimento allenamento è minimo ("Per Niente"), il livello di personalità competitiva è minimo (1) e non si pratica sport a livello agonistico.

Al contrario, la probabilità di dropout minore, pari a 0.25, si ha quando il grado di divertimento allenamento è massimo ("Molto"), il livello di personalità competitiva è massimo (7) e si pratica sport a livello agonistico.

3. Support Vector Machine

Il *Support Vector Machine* (SVM) è un algoritmo di Machine Learning versatile e potente, utilizzato sia per il task di regressione sia per il task di classificazione (lineare o non lineare).

Questo tipo di algoritmo risulta più efficiente di altri, quando viene applicato su dataset complessi ma di piccole o medie dimensioni: per questa ragione, nel caso del dataset a disposizione (2750 righe), si ipotizza che possa fornire performance migliori rispetto ad altri algoritmi.

L'idea fondamentale su cui si basa il SVM è, innanzitutto, la rappresentazione di un'osservazione in forma vettoriale, dove ogni componente del vettore rappresenta il valore che assume una determinata *feature*, con lo scopo di posizionare le osservazioni all'interno di uno spazio N-dimensionale, con N che specifica il numero di componenti.

A questo punto, tramite l'addestramento dell'algoritmo sul *Training Set*, viene definita un'opportuna funzione (lineare o non lineare), in grado di separare in modo ottimale le osservazioni nello spazio N-dimensionale per poterle identificare in base alla classe di appartenenza.

Quando il SVM utilizza il metodo "*large margin classification*", questa funzione ottimale massimizza la distanza dalle osservazioni localizzate sul confine della regione che delimita la classe di appartenenza.

Tali osservazioni prendono il nome di *support vectors* e tale metodo è possibile solo se la funzione di divisione è di tipo lineare: si afferma che le osservazioni sono linearmente separabili.

Quando, invece, il SVM utilizza il metodo "*soft margin classification*", possibile anche per funzioni non lineari, allora l'obiettivo è quello di rendere flessibile il modello, cercando un compromesso tra la distanza della funzione di divisione dai *support vectors* e la violazione del margine da parte degli outliers.

Questo equilibrio può essere controllato tramite l'hyperparameter *C*:

- bassi valori di *C* rendono più netta la funzione di divisione ma permettono più violazioni del margine;
- alti valori di *C* limitano le violazioni del margine, ma rendono più sottile la funzione di divisione.

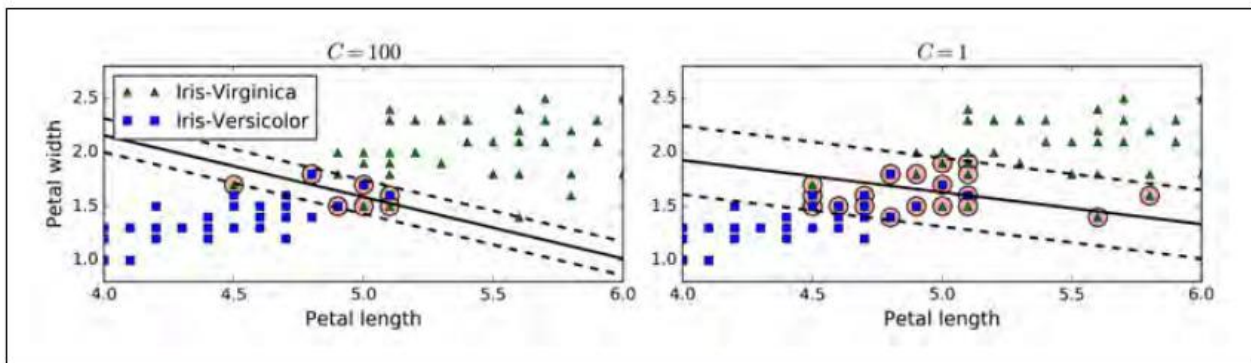


Fig. 64: Esempio della funzione lineare di divisione del SVM al variare di C (Iris Dataset)

Le funzioni di divisione di tipo non lineare, invece, sono utilizzate in altri metodi di addestramento dell'algoritmo SVM:

- *Polynomial Kernel*
- *Gaussian RBF Kernel*

Il modello predittivo SVM per il dropout sportivo è stato così ottenuto:

- *variabili esplicative*: 21 features derivanti dal task di *feature selection*;
- *variabile target*: dropout ["No" → classe negativa (0), "Sì" → classe positiva (1)];
- apprendimento mediante *Cross-Validation 10-Fold* (275 obs per fold) con tuning degli *hyperparameters*

Il tuning dei parametri interni al modello (*hyperparameters*) è stato effettuato considerando la seguente griglia di valori dei parametri:

- $C = [0.1, 1, 10]$: equilibrio tra "spessore" della divisione e violazione del margine
- **kernel** = ["linear", "poly", "rbf"]: tipologia della funzione di divisione

La configurazione migliore dei parametri ottenuta è $[C=0.1, \text{kernel}=\text{"poly"}]$, sulla base della massimizzazione della metrica *recall* dell'algoritmo di classificazione: con questa scelta progettuale, è stato scelto di limitare l'incidenza dei *falsi negativi* sulle performance del modello, dove per *falsi negativi* si intendono le osservazioni predette come "non abbandonano" quando in realtà sono classificate come "abbandono" sportivo.

La valutazione delle performance del modello ottenuto è descritta nel capitolo successivo.

Di seguito, invece, viene riportata graficamente la suddivisione dello spazio bidimensionale tramite kernel lineare, appresa da un modello predittivo *SVM* che riceve in input i due predittori più significativi individuati dal task di *feature selection* del capitolo precedente:

- X_1 : grado di divertimento allenamento
(0 → *Per Niente*, 1 → *Poco*, 2 → *Neutrale*, 3 → *Abbastanza*, 4 → *Molto*)
- X_2 : personalità competitiva (*scala Likert da 1 a 7* codificata in scala da 0 a 6)

Si tratta di un modello in overfitting, utilizzato per comprendere l'incidenza di questi fattori all'interno del campione.

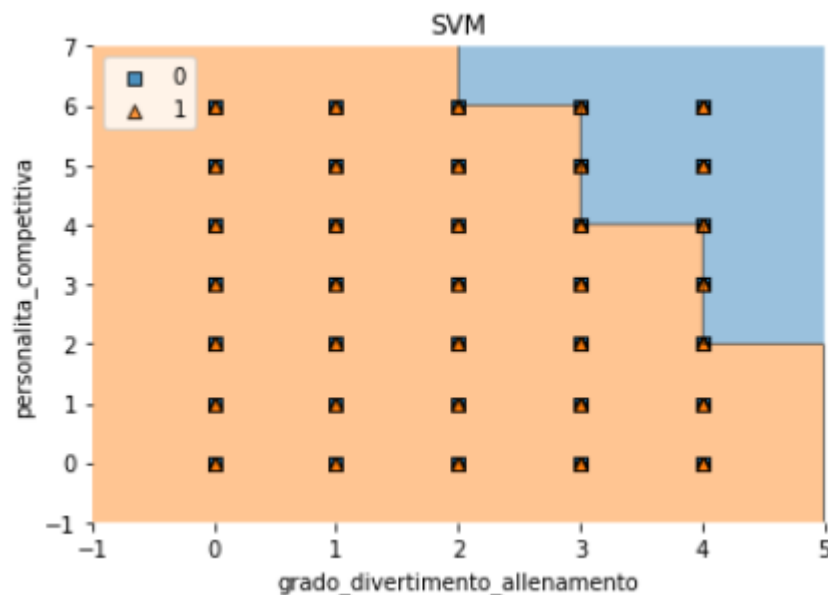


Fig. 65: Modello SVM lineare: classificazione in base ai due predittori più significativi

Prima di descrivere le deduzioni ricavate dal grafico in Fig. 65, si procede a valutare la predizione del dropout nei *support vector*, ossia nelle osservazioni posizionate sul margine della funzione lineare di divisione delle due classi.

```
print(mod.predict([[2, 6], [3, 6], [3, 5], [3, 4], [4, 4], [4, 3], [4, 2]]))
[0 0 0 0 0 0 0]
```

Fig. 66: Predizione dropout nei support vectors

Dai risultati ottenuti, si evince che tutti i *support vector* vengono classificati appartenenti alla “classe negativa” del dropout (0), ovvero viene predetto il “non abbandono”.

A questo punto si descrivono i pattern significativi appresi dal modello SVM:

- 1) per le osservazioni per cui il grado di divertimento è *“Per Niente”* o *“Poco”* viene predetta la “classe positiva” del dropout (1), ovvero viene predetto l’abbandono, indipendentemente dal livello di personalità competitiva;
- 2) per le osservazioni per cui il grado di divertimento è *“Neutrale”* viene predetta la “classe positiva” del dropout (1), ovvero viene predetto l’abbandono, ad eccezione del caso in cui il livello di personalità competitiva è massimo (7 su scala Likert);
- 3) per le osservazioni per cui il grado di divertimento è *“Abbastanza”* viene predetta la “classe positiva” del dropout (1), ovvero viene predetto l’abbandono, ad eccezione dei casi in cui il livello di personalità competitiva è compreso tra 5 e 7 della scala Likert;
- 4) per le osservazioni per cui il grado di divertimento è *“Molto”* viene predetta la “classe negativa” del dropout (0), ovvero viene predetto il non abbandono, ad eccezione dei casi in cui il livello di personalità competitiva è minimo (0 e 1 su scala Likert).

4. K-Nearest Neighbors

L'algoritmo *K-Nearest Neighbors (KNN)*, utilizzato per il task di classificazione, è un tipo di apprendimento automatico *instance-based*: non tenta di costruire un modello interno generale, ma archivia semplicemente le osservazioni dei dati di addestramento.

La classificazione viene effettuata tramite la tecnica del *majority vote*, secondo cui si indaga sulla classe delle osservazioni “più vicine” (*nearest neighbors*) a quella che si vuole classificare: viene assegnata la classe che ha il maggior numero di rappresentanti all'interno dei “vicini” identificati.

In questo modello gioca un ruolo fondamentale il parametro K , indicante il numero di osservazioni “vicine” da considerare e il cui valore ottimale dipende in modo sensibile dalla dimensione dei dati usati per l'addestramento.

L'individuazione dei “vicini” rispetto all'osservazione da classificare avviene utilizzando una metrica di distanza tra le osservazioni, rappresentate attraverso vettori posizionati all'interno di uno spazio N -dimensionale, dove N corrisponde al numero di *features* a disposizione.

Le metriche di distanza solitamente usate per vettori numerici sono: *distanza euclidea* e *distanza Manhattan*.

Il peso conferito ad un'osservazione “vicina” durante la tecnica di *majority vote* può essere valutato secondo due modalità:

- *pesi uniformi*: la classe da assegnare viene definita pesando allo stesso modo i “voti” dati dalle osservazioni “vicine”;
- *distanza*: la classe da assegnare viene definita pesando in modo diverso i “voti” dati dalle osservazioni “vicine”, utilizzando un peso proporzionale all'inverso della distanza dalla osservazione da classificare.

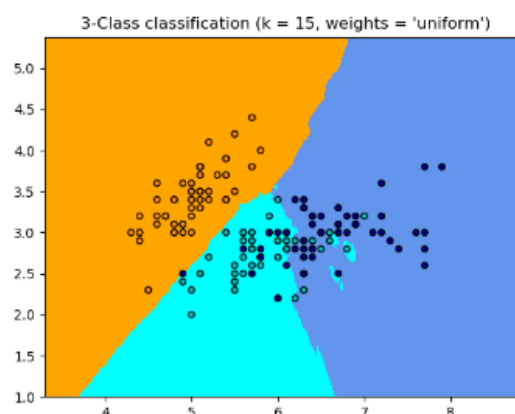


Fig. 67: Esempio di classificazione KNN nello spazio bidimensionale

Il modello predittivo *KNN* per il dropout sportivo è stato così ottenuto:

- *variabili esplicative*: 21 features derivanti dal task di *feature selection*;
- *variabile target*: dropout [“No” → classe negativa (0), “Sì” → classe positiva (1)];
- apprendimento mediante *Cross-Validation 10-Fold* (275 obs per fold) con tuning degli *hypeparameters*

Il tuning dei parametri interni al modello (*hypeparameters*) è stato effettuato considerando la seguente griglia di valori dei parametri:

- **n_neighbors** = [50, 100, 200]: numero dei “vicini” da considerare
- **p** = [1, 2]: metrica di distanza (p=1 → Manhattan, p=2 → Euclidea)
- **weights** = [“uniform”, “distance”]: valutazione peso dei “vicini” durante il *majority vote*

La configurazione migliore dei parametri ottenuta è [*n_neighbors*=50, *p*=2, *weights*=“distance”], sulla base della massimizzazione della metrica *recall* dell’algoritmo di classificazione: con questa scelta progettuale, è stato scelto di limitare l’incidenza dei *falsi negativi* sulle performance del modello, dove per *falsi negativi* si intendono le osservazioni predette come “non abbandonano” quando in realtà sono classificate come “abbandono” sportivo.

La valutazione delle performance del modello ottenuto è descritta nel capitolo successivo.

5. Multi-Layer Perceptron

Il *Multi-Layer Perceptron (MLP)* è la più semplice configurazione di *Artificial Neural Network (ANN)*, ovvero una rete di neuroni artificiale alla base degli algoritmi di Deep Learning.

Le *ANN* sono state sviluppate in relazione al concetto biologico di “neurone”: ogni unità presente nella rete riceve dati di input e produce un output.

Esse sono caratterizzate da elevata versatilità, potenza e scalabilità, tutti aspetti che le rendono ideali per approcciare task di Machine Learning grandi e complessi.

Il *Perceptron* è una delle architetture *ANN* più semplici, che si basa su un unico strato di neuroni, chiamati *unità di soglia lineare (LTU)*, in cui gli input e l'output sono numeri (anziché valori binari 0/1) e ad ogni connessione di input è associato un peso.

LTU calcola una somma ponderata dei suoi input ($z = w_1 x_1 + w_2 x_2 + \dots + w_n x_n = \mathbf{w}^T \cdot \mathbf{x}$), quindi applica una funzione step a quella somma e genera il risultato:

$$h_w(\mathbf{x}) = \text{step}(z) = \text{step}(\mathbf{w}^T \cdot \mathbf{x})$$

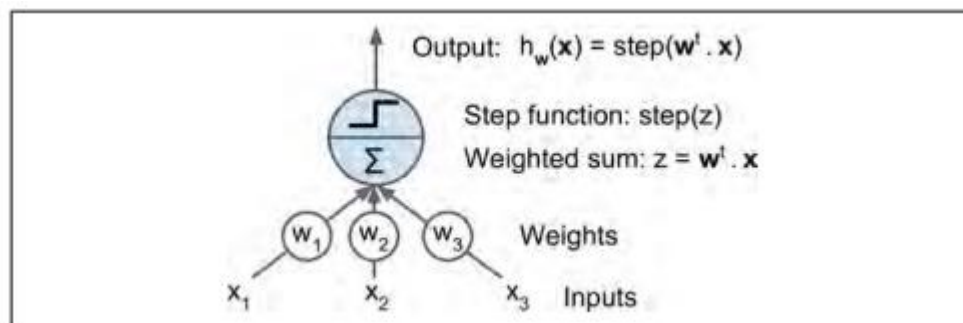


Fig. 68: Grafico della logica di funzionamento di una LTU

Le funzioni step più comuni sono le seguenti:

$$\text{heaviside}(z) = \begin{cases} 0 & \text{if } z < 0 \\ 1 & \text{if } z \geq 0 \end{cases} \quad \text{sgn}(z) = \begin{cases} -1 & \text{if } z < 0 \\ 0 & \text{if } z = 0 \\ +1 & \text{if } z > 0 \end{cases}$$

Fig. 69: Funzioni step di una LTU

Una singola *LTU* può essere utilizzata per una semplice classificazione binaria lineare. Essa opera nel seguente modo: calcola una combinazione lineare degli ingressi e se il risultato supera una soglia, produce come risultato la “classe positiva” (1), oppure restituisce la “classe negativa” (0).

Un *Perceptron* è semplicemente composto da un singolo strato di *LTU* (*input layer*), con ogni neurone collegato a tutti gli ingressi dello strato di uscita (*output layer*).

All'interno dell'*input layer* viene generalmente aggiunta una funzione di polarizzazione ($x_0 = 1$) che costituisce il cosiddetto *bias neuron*.

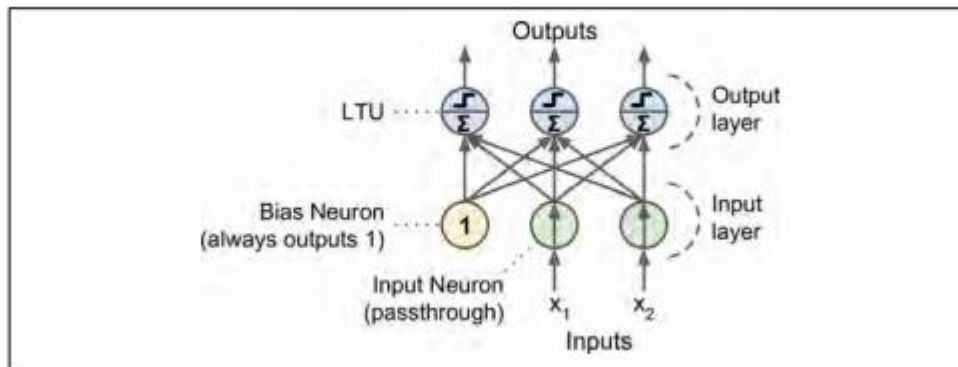


Fig. 70: Grafico della logica di funzionamento di un *Perceptron*

L'obiettivo dell'addestramento di un *Perceptron* è quello di determinare i giusti valori dei pesi dei neuroni presenti nell'*input layer* (w_0, w_1, \dots, w_n), seguendo l'*apprendimento Hebbiano*.

La regola alla base di questa idea consiste nell'aumentare il peso della connessione tra due neuroni ogni volta che producono lo stesso output, tenendo conto dell'errore commesso dalla rete ed evitando, quindi, di rafforzare le connessioni che producono un output errato.

Più specificamente, il *Perceptron* viene addestrato su un'osservazione del *Training Set* alla volta e per ognuna di questa effettua le sue previsioni.

Per ogni neurone in uscita che ha prodotto una previsione errata, viene rinforzata la connessione degli input che avrebbero contribuito alla previsione corretta, aumentandone il peso.

$$w_{i,j}^{(\text{next step})} = w_{i,j} + \eta(\hat{y}_j - y_j)x_i$$

- $w_{i,j}$ is the connection weight between the i^{th} input neuron and the j^{th} output neuron.
- x_i is the i^{th} input value of the current training instance.
- \hat{y}_j is the output of the j^{th} output neuron for the current training instance.
- y_j is the target output of the j^{th} output neuron for the current training instance.
- η is the learning rate.

Fig. 71: Formula di addestramento di un *Perceptron*

Avendo introdotto il funzionamento di un *Perceptron*, è possibile procedere nella descrizione della sua forma generalizzata: *Multi-Layer Perceptron (MLP)*.

Un *MLP* è composto da uno strato di input (*input layer*), uno o più strati intermedi di *LTU*, chiamati *hidden layers*, e uno strato finale di *LTU* chiamato *output layer*.

Ogni livello tranne quello di output include un neurone di polarizzazione ed è completamente collegato al livello successivo.

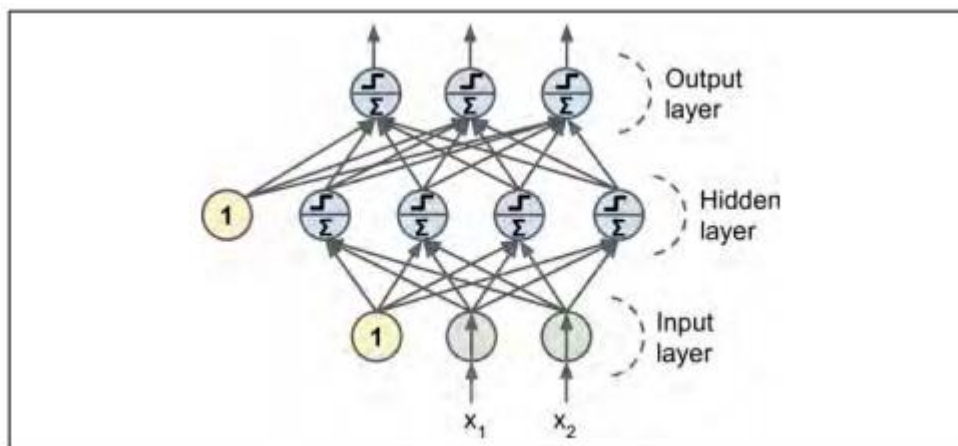


Fig. 72: Grafico della logica di funzionamento di un *MLP*

L'addestramento di un *MLP* avviene nel seguente modo: per ogni osservazione del *Training Set*, l'algoritmo calcola il valore di uscita di ogni neurone in ogni strato, quindi misura l'errore di output della rete, ovvero la differenza tra l'output desiderato e l'output effettivo della rete.

Viene stimato, inoltre, quanto ciascun neurone nell'ultimo *hidden layer* ha contribuito a generare questo errore, ripetendo il procedimento anche per i neuroni dei *layer* precedenti fino a quando si raggiunge l'*input layer*.

Questa tecnica, chiamata *backpropagation*, consente di calibrare i pesi delle connessioni per ciascuno strato della rete, tenendo conto delle previsioni errate effettuate precedentemente durante la fase di addestramento.

La funzione step, riferita ad una singola *LTU*, viene utilizzata da un intero strato della rete e prende il nome di *activation function* (funzione di attivazione): essa deve essere derivabile per permettere la valutazione della contribuzione all'errore in ogni strato.

Le principali *activation function* utilizzate sono:

- *hyperbolic tangent function*: $z = 2\sigma(2z) - 1$ con σ funzione logistica
- *ReLU function*: $z = \max(0, z)$

Un *MLP* viene spesso utilizzato per la classificazione binaria: a tal proposito è possibile anche introdurre all'interno dell'*output layer* una funzione, chiamata *softmax*, permettendo la stima della probabilità per un'osservazione di appartenere ad una classe.

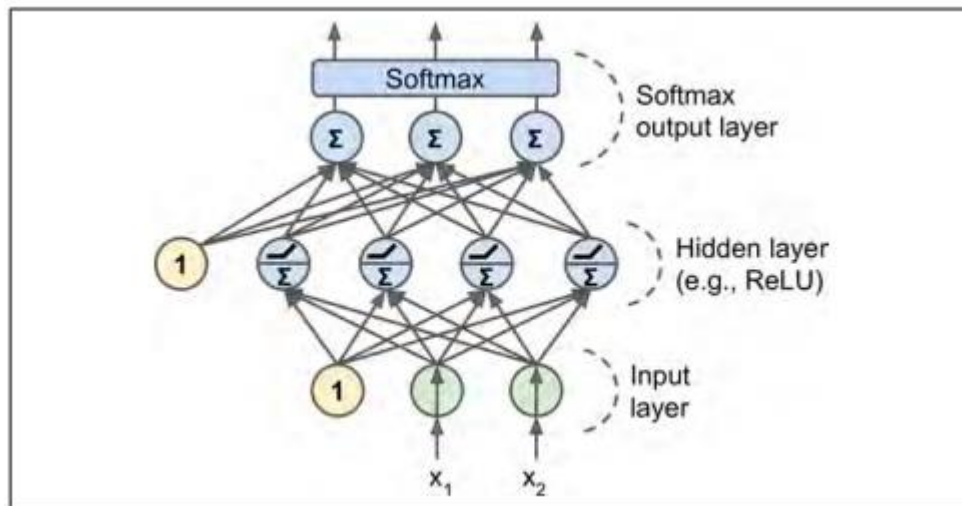


Fig. 73: Grafico della logica di funzionamento di un MLP con softmax

Il modello predittivo *MLP* per il dropout sportivo è stato così ottenuto:

- *variabili esplicative*: 21 features derivanti dal task di *feature selection*;
- *variabile target*: dropout ["No" → classe negativa (0), "Sì" → classe positiva (1)];
- apprendimento mediante *Cross-Validation 10-Fold* (275 obs per fold) con tuning degli *hypeparameters*

Il tuning dei parametri interni al modello (*hypeparameters*) è stato effettuato considerando la seguente griglia di valori dei parametri:

- **hidden_layer_sizes** = [(50,), (100,), (150,)]: numero di neuroni dell'*hidden layer*
- **activation** = ["tanh", "relu"]: tipo di funzione di attivazione

La configurazione migliore dei parametri ottenuta è [*hidden_layer_sizes*=(100,), *activation*="tanh"], sulla base della massimizzazione della metrica *recall* dell'algoritmo di classificazione: con questa scelta progettuale, è stato scelto di limitare l'incidenza dei *falsi negativi* sulle performance del modello, dove per *falsi negativi* si intendono le osservazioni predette come "non abbandono" quando in realtà sono classificate come "abbandono" sportivo.

La valutazione delle performance del modello ottenuto è descritta nel capitolo successivo.

6. Voting Classifier

Gli algoritmi di classificazione descritti finora possono essere aggregati tramite un *Voting Classifier*, con lo scopo di ottenere una previsione più accurata per ogni osservazione soggetta alla classificazione, migliorando di conseguenza le performance dei singoli modelli.

L'aggregazione delle previsioni può avvenire attraverso due diversi classificatori *ensemble*:

- *hard voting classifier*: ogni osservazione viene classificata assegnando la classe predetta dai singoli classificatori con il maggior numero di voti (*majority vote*)

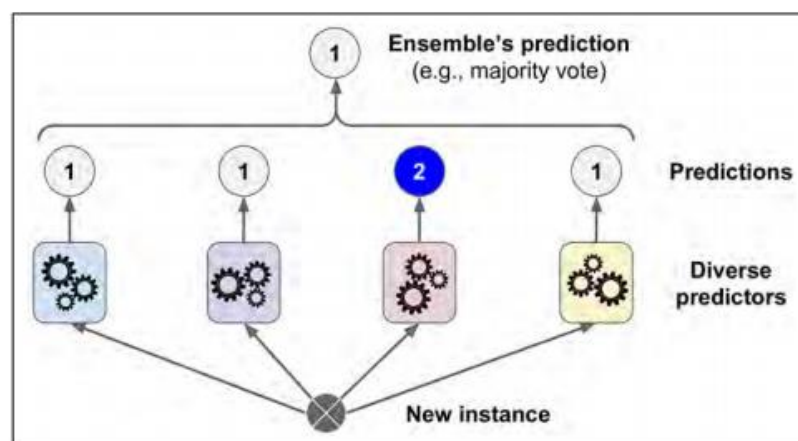


Fig. 74: Logica di classificazione dell'hard voting classifier

- *soft voting classifier*: ogni osservazione viene classificata calcolando la media delle probabilità stimate dai singoli classificatori, eventualmente pesata in base all'importanza dei classificatori, e assegnando la classe con probabilità stimata maggiore (ciò è possibile solo se tutti i classificatori sono in grado di fornire la stima della probabilità)

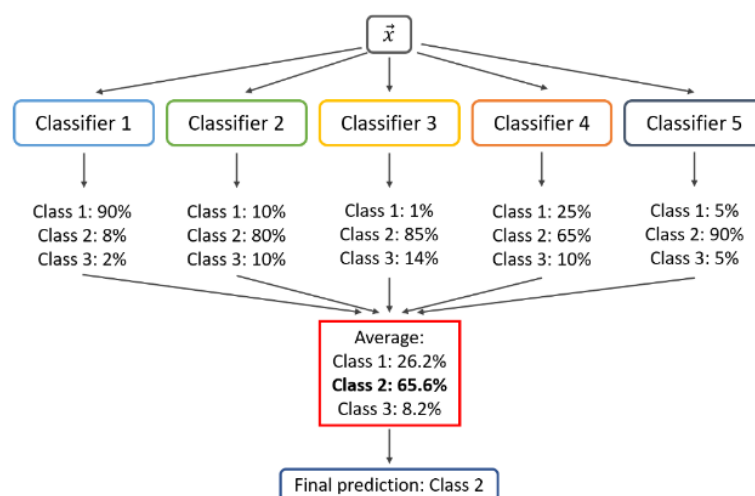


Fig. 75: Logica di classificazione del soft voting classifier

Si dimostra che se tutti i singoli classificatori sono perfettamente indipendenti tra loro, ovvero commettono errori di previsione non correlati, l'accuratezza della previsione aggregata è significativamente più alta rispetto a quella dei singoli classificatori.

Non è questo il caso degli algoritmi di classificazione descritti in questo capitolo, in quanto essi sono addestrati utilizzando lo stesso set di dati a disposizione.

Tuttavia, si ipotizza comunque di ottenere un leggero miglioramento delle performance, dato dal fatto che le tecniche di addestramento sono molto diverse tra loro.

Il *Voting Classifier*, che permette di ottenere la probabilità stimata del dropout secondo la tecnica del *soft voting*, è stato ottenuto dall'aggregazione dei seguenti classificatori:

- *Random Forest* (Parte Terza – Capitolo 2 – Paragrafo 3)
- *Logistic Regression* (Parte Quarta – Capitolo 1 – Paragrafo 2)
- *Support Vector Machine* (Parte Quarta – Capitolo 1 – Paragrafo 3)
- *K-Nearest Neighbors* (Parte Quarta – Capitolo 1 – Paragrafo 4)
- *Multi-Layer Perceptron* (Parte Quarta – Capitolo 1 – Paragrafo 5)

Ogni singolo classificatore, precedentemente ottimizzato mediante il tuning degli *hyperparameters* descritto nel rispettivo paragrafo, ha effettuato il task di classificazione nel seguente modo:

- *variabili esplicative*: 21 features derivanti dal task di *feature selection*;
- *variabile target*: dropout ["No" → classe negativa (0), "Sì" → classe positiva (1)];
- apprendimento mediante *Cross-Validation 10-Fold* (275 obs per fold)

Successivamente, mediante la tecnica del *Soft Voting*, sono state aggregate le probabilità stimate di "abbandono" e di "non abbandono" calcolando la media e pesando i singoli classificatori con peso uniforme.

A questo punto, la classe predetta per ogni osservazione risulta essere la classe con probabilità media maggiore.

La valutazione delle performance del modello ottenuto è descritta nel capitolo successivo.

Valutazione Performance

1. Metriche di performance

In questo paragrafo vengono descritte le metriche di performance, utilizzate per valutare e confrontare tra loro gli *algoritmi di classificazione binaria di tipo supervisionato*.

A tale scopo è necessario introdurre la *matrice di confusione (confusion matrix)*, che consiste nel raggruppamento delle osservazioni oggetto della classificazione in quattro diversi gruppi:

- *True Positive (TP)*: numero di osservazioni per cui è stata predetta correttamente la “classe positiva”, ovvero il numero di osservazioni per cui è stato predetto l’abbandono sportivo e nella realtà l’abbandono si è verificato;
- *True Negative (TN)*: numero di osservazioni per cui è stata predetta correttamente la “classe negativa”, ovvero il numero di osservazioni per cui è stato predetto il non abbandono sportivo e nella realtà l’abbandono non si è verificato;
- *False Positive (FP)*: numero di osservazioni per cui è stata predetta in modo errato la “classe positiva”, ovvero il numero di osservazioni per cui è stato predetto l’abbandono sportivo ma nella realtà l’abbandono non si è verificato;
- *False Negative (FN)*: numero di osservazioni per cui è stata predetta in modo errato la “classe negativa”, ovvero il numero di osservazioni per cui è stato predetto il non abbandono sportivo ma nella realtà l’abbandono si è verificato.

		PREDICTIVE VALUES	
		POSITIVE (1)	NEGATIVE (0)
ACTUAL VALUES	POSITIVE (1)	TP	FN
	NEGATIVE (0)	FP	TN

Fig. 76: Confusion Matrix per la classificazione binaria

Nel caso in cui il modello predittivo fornisce la stima della probabilità di appartenenza alla “classe positiva”, allora la classe predetta per un’osservazione è la “classe positiva” se tale probabilità è maggiore di 0.5, mentre è la “classe negativa” se tale probabilità è minore di 0.5.

In seguito a questa considerazione, si procede a collocare l’osservazione nel gruppo corretto tra quelli che costituiscono la matrice di confusione.

Avendo generato la matrice di confusione, è possibile valutare le performance di un modello predittivo di classificazione supervisionato tramite alcune metriche.

Accuracy

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN)$$

L’*accuracy* di un algoritmo di classificazione indica il rapporto tra il numero di osservazioni predette correttamente, ovvero la somma di *True Positive* e *True Negative*, e il numero di osservazioni totale utilizzate per il test.

Questa metrica, quindi, stabilisce il grado generale di correttezza dell’algoritmo senza indagare specificatamente su un tipo di errore di classificazione.

Precision

$$\text{Precision} = TP / (TP + FP)$$

La *precision* di un algoritmo di classificazione indica il rapporto tra il numero di osservazioni predette correttamente con “classe positiva”, ovvero il numero di *True Positive*, e il numero totale di osservazioni predette con “classe positiva”, ovvero la somma di *True Positive* e *False Positive*.

Questa metrica si occupa di indagare sull’incidenza dell’*errore di I tipo*, ovvero la predizione della “classe positiva” per osservazioni appartenenti alla “classe negativa”.

Nella classificazione del dropout sportivo, l’*errore di I tipo* viene commesso quando si predice l’abbandono per un’osservazione per la quale non si è verificato l’abbandono.

Recall

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$$

La *recall* di un algoritmo di classificazione indica il rapporto tra il numero di osservazioni predette correttamente con “classe positiva”, ovvero il numero di *True Positive*, e il numero totale di osservazioni appartenenti alla “classe positiva”, ovvero la somma di *True Positive* e *False Negative*.

Questa metrica si occupa di indagare sull’incidenza dell’*errore di II tipo*, ovvero la predizione della “classe negativa” per osservazioni appartenenti alla “classe positiva”.

Nella classificazione del dropout sportivo, l’*errore di II tipo* viene commesso quando si predice il non abbandono per un’osservazione per la quale si è verificato l’abbandono.

Per quanto riguarda la predizione del dropout sportivo, questo scenario ha un “costo” di errata classificazione più rilevante rispetto all’*errore di I tipo*: supponendo di commettere un errore nella previsione, si preferisce aver sbagliato a predire l’abbandono per un ragazzo che non ha abbandonato rispetto ad aver sbagliato a predire il non abbandono per un ragazzo che ha abbandonato.

Questa è la ragione per la quale tutti gli algoritmi di classificazione del dropout sportivo, descritti nel capitolo precedente, sono stati addestrati mettendo in atto il *tuning dei parametri* interni al modello (*hyperparameters*), con l’obiettivo di massimizzare la *recall*.

F1-Score

$$\text{F1-Score} = 2 * \text{precision} * \text{recall} / (\text{precision} + \text{recall})$$

L’indice *F1* è la metrica ottenuta calcolando la media armonica tra *precision* e *recall*, quindi ha l’obiettivo di “armonizzare” la valutazione dell’incidenza degli *errori di I e II tipo*.

ROC Curve

La *ROC Curve* (*receiver operating characteristic*) è un altro strumento comune utilizzato per la valutazione delle performance dei classificatori binari.

La curva ROC ha lo scopo di tracciare il grafico che riporta in ascissa il *False Positive Rate* (*FPR*) e in ordinata il *True Positive Rate* (*TPR*), variando la soglia (*threshold*) con la quale si classificano le osservazioni in base alla stima della probabilità dell'algoritmo.

Come deducibile dal nome, il *FPR* è il rapporto tra il numero di osservazioni appartenenti alla "classe negativa" classificate erroneamente, ovvero il numero di *False Positive*, e il numero totale di osservazioni appartenenti alla "classe negativa", ovvero la somma di *False Positive* e *True Negative*.

Il *FPR* viene calcolato anche come $1 - \textit{specificity}$, dove per *specificity* si intende il *FNR*.

$$\text{False Positive Rate (FPR)} = \text{FP} / (\text{FP} + \text{TN})$$

$$\text{FPR} = 1 - \text{FNR} = 1 - \text{Specificity}$$

Viceversa, il *TPR* è il rapporto tra il numero di osservazioni appartenenti alla "classe positiva" classificate correttamente, ovvero il numero di *True Positive*, e il numero totale di osservazioni appartenenti alla "classe positiva", ovvero la somma di *True Positive* e *False Negative*.

Il *TPR* viene chiamato anche *sensitivity*.

$$\text{True Positive Rate (TPR)} = \text{TP} / (\text{TP} + \text{FN})$$

$$\text{TPR} = \text{Specificity}$$

La *ROC Curve*, quindi, riporta graficamente in ascissa i valori di $1 - \textit{Specificity}$, mentre in ordinata i valori di *Sensitivity*, al variare della soglia (*threshold*) utilizzata dall'algoritmo di classificazione per la predizione della classe di appartenenza delle osservazioni.

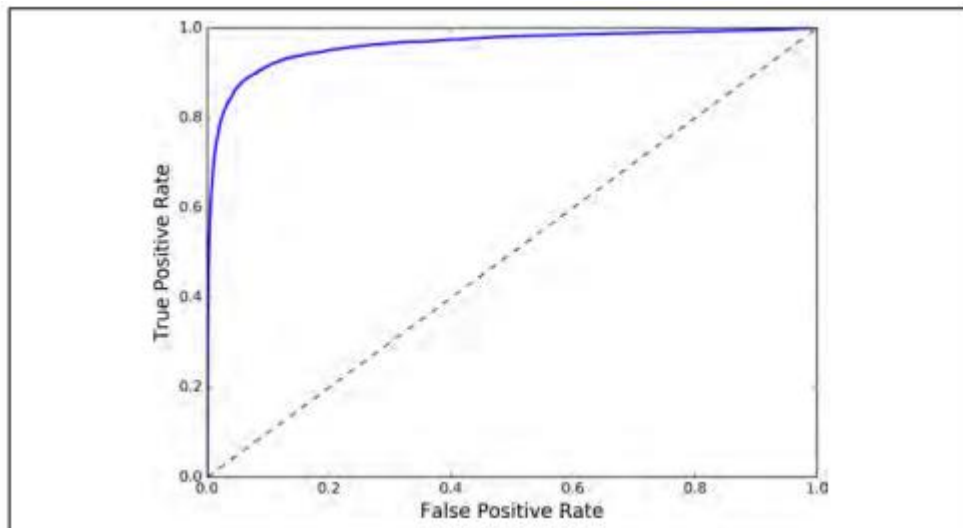


Fig. 77: Esempio di ROC Curve

La linea tratteggiata in Fig. 77 rappresenta la *curva ROC* di un classificatore casuale che fornisce valori di *TPR* e *FPR* uguali per ogni *threshold* di classificazione.

La *curva ROC* di un buon classificatore, invece, rimane il più lontano possibile da quella linea tratteggiata, coincidendo in due segmenti perpendicolari per un classificatore ideale.

Infatti, un classificatore ideale è un classificatore che, fissata una qualsiasi *threshold* di classificazione, ha *FPR* pari a 0 e *TPR* pari a 1, derivante dal fatto il numero di *False Positive* e il numero di *False Negative* sono nulli.

Un modo per confrontare diversi classificatori è misurare l'area sotto la curva (*AUC*). Secondo questa metrica, si distingue un classificatore casuale dal valore di *AUC* pari a 0.5, mentre un buon classificatore si distingue da un valore di *AUC* più vicino possibile a 1, ovvero il valore per il classificatore ideale.

La *curva ROC* e la metrica *AUC* sono strumenti utili per valutare e confrontare le performance di diversi classificatori applicati a un set di dati bilanciato (*balance dataset*), ovvero un set in cui la percentuale di osservazioni della “classe positiva” è più o meno uguale alla percentuale di osservazioni della “classe negativa”.

Ciò è verificato per il dataset a disposizione utilizzato per la classificazione del dropout sportivo, in cui la percentuale di chi ha abbandonato (50,95%) è sostanzialmente uguale alla percentuale di chi non ha abbandonato (49,05%).

Nel caso di *class imbalance*, invece, è più appropriato lo studio del *tradeoff* tra *precision* e *recall*.

2. Confronto tra i modelli

In questo paragrafo viene descritto il confronto tra le performance degli algoritmi di classificazione supervisionata, utilizzati per la predizione del dropout sportivo tra gli adolescenti.

Lo scopo del confronto è quello di determinare il modello predittivo migliore da usare per ricavare la probabilità di abbandono di un ragazzo a cui viene somministrato un questionario predittivo.

Tale questionario costituirà un'evoluzione del questionario di raccolta dati sviluppato da Orangogo, volto a predire l'abbandono su osservazioni "non viste" durante la fase di addestramento del modello.

Prima di procedere al confronto tra i modelli, viene esposta la tesi del teorema "*No Free Lunch*" (NFL) di David Wolpert e William Macready (1997):

"Se non viene fatta nessuna ipotesi sui dati, non c'è motivo di preferire a priori un modello rispetto a nessun altro."

In base a questo teorema, l'unico modo per valutare quale sia il miglior modello predittivo del dropout sportivo è confrontare le performance degli algoritmi di classificazione utilizzati nel capitolo precedente, ricordando che non è stata fatta nessuna ipotesi di distribuzione dei dati a disposizione.

Questo confronto è possibile grazie al fatto che tutti gli algoritmi di classificazione utilizzati sono stati addestrati con la stessa tecnica di suddivisione dei dati in *Training Set* e *Test Set*: è stata scelta la tecnica *Cross-Validation* dividendo l'intero set in *10 Fold* e utilizzando, di volta in volta, una fold diversa come *Test Set* (275 obs) e le altre nove come *Training Set* (2475 obs).

Per quanto riguarda le *features*, invece, tutti gli algoritmi di classificazione sono stati addestrati sull'insieme di 21 *features* derivanti dal task di *feature selection*: esse rappresentano le variabili più importanti per la predizione del dropout sportivo, in base ai valori di *feature importance* ottenuti applicando l'addestramento con *Random Forest*.

Per ogni metrica di valutazione della performance di un algoritmo di classificazione si ottengono 10 valori, uno per ogni fold utilizzata come *Test Set*.

E' possibile, quindi, ottenere valore medio, deviazione standard, intervalli di confidenza e altri parametri statistici quantitativi, interpretabili graficamente mediante *BoxPlot*.

Di seguito, viene riportato il confronto delle performance in base alla metrica *Accuracy*.

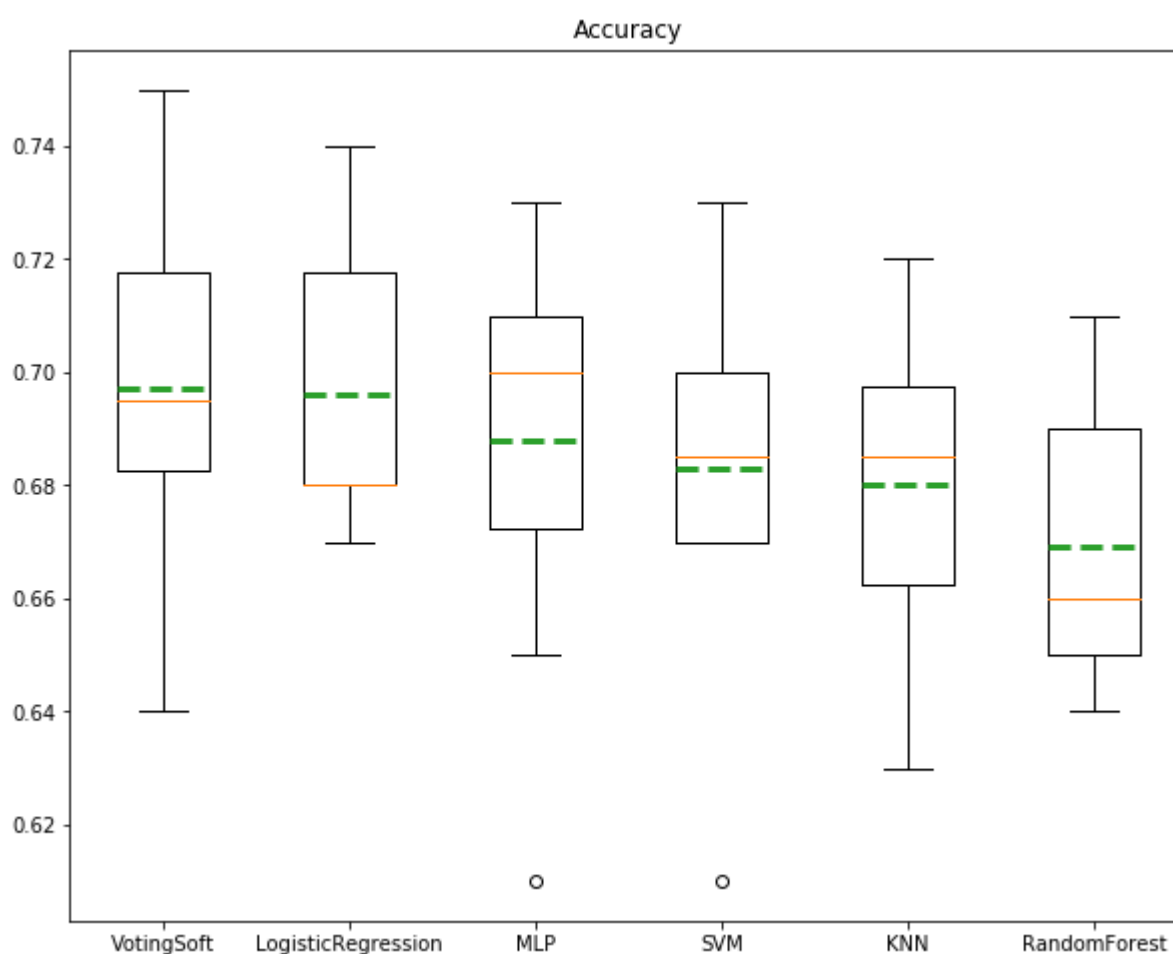


Fig. 78: Confronto accuracy degli algoritmi di classificazione del dropout sportivo

Il *Voting Soft* è il classificatore con *accuracy* media più alta (linea verde tratteggiata) pari a **0.70**: in termini pratici, ciò sta a significare che, sottoponendo nuovi ragazzi ad un questionario predittivo sul dropout sportivo, mediamente viene predetto in modo corretto l'abbandono (*True Positive*) o il non abbandono (*True Negative*) nel 70% dei casi.

AVG Accuracy VotingSoft:	0.7
AVG Accuracy LogisticRegression:	0.69
AVG Accuracy MLP:	0.69
AVG Accuracy SVM:	0.68
AVG Accuracy KNN:	0.68
AVG Accuracy RandomForest:	0.67

Fig. 79: Valori medi accuracy degli algoritmi di classificazione del dropout sportivo

Di seguito, viene riportato il confronto delle performance in base alla metrica *Precision*.

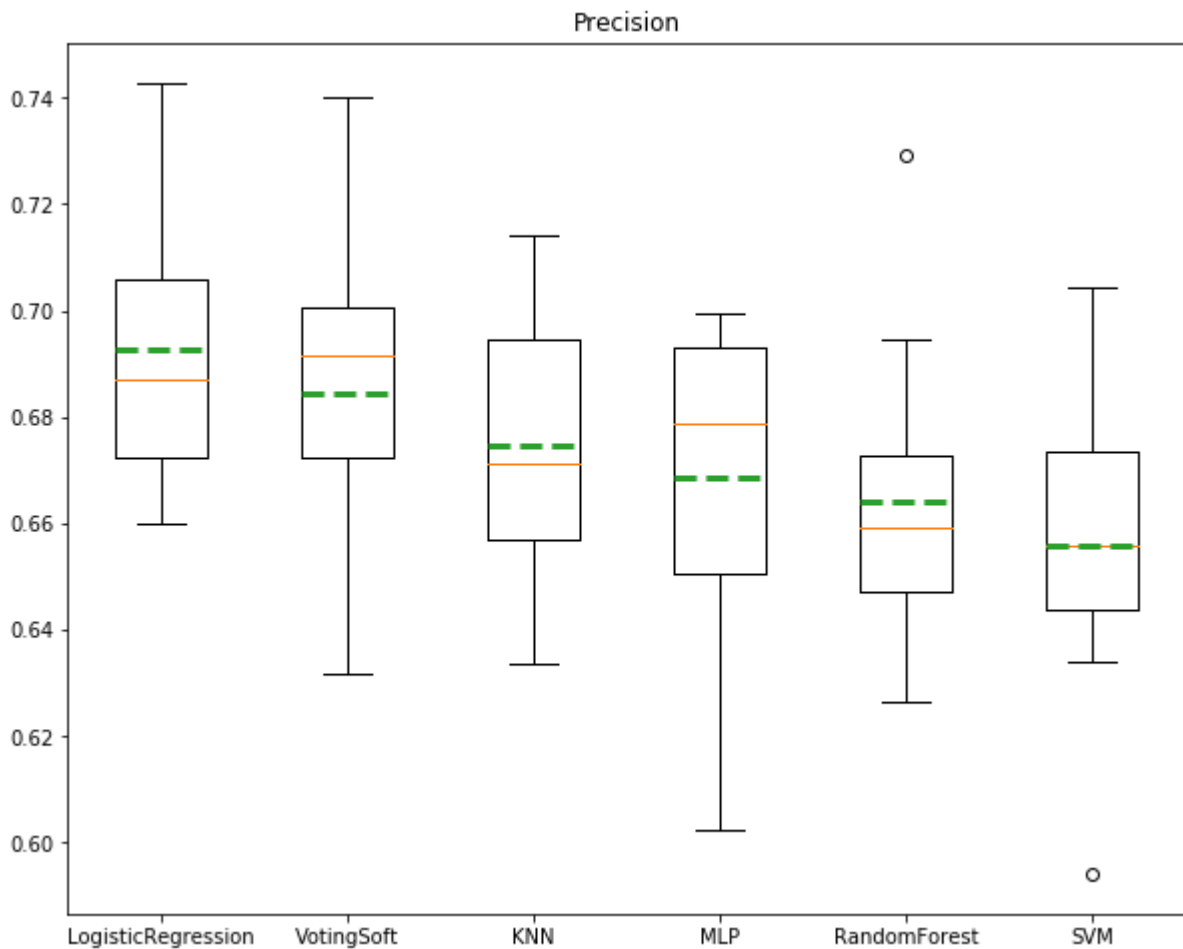


Fig. 80: Confronto precision degli algoritmi di classificazione del dropout sportivo

La *Logistic Regression* è il classificatore con *precision* media più alta (linea verde tratteggiata) pari a **0.69**.

In termini pratici, ciò sta a significare che, sottoponendo nuovi ragazzi ad un questionario predittivo sul dropout sportivo, mediamente la predizione della “classe positiva” indicante il verificarsi dell’abbandono è corretta nel 69% dei casi (*True Positive*) mentre è errata nel restante 31% (*False Positive*).

```
AVG Precision LogisticRegression: 0.69
AVG Precision VotingSoft:         0.68
AVG Precision KNN:                0.67
AVG Precision MLP:                0.67
AVG Precision RandomForest:        0.66
AVG Precision SVM:                0.66
```

Fig. 81: Valori medi precision degli algoritmi di classificazione del dropout sportivo

Di seguito, viene riportato il confronto delle performance in base alla metrica *Recall*.

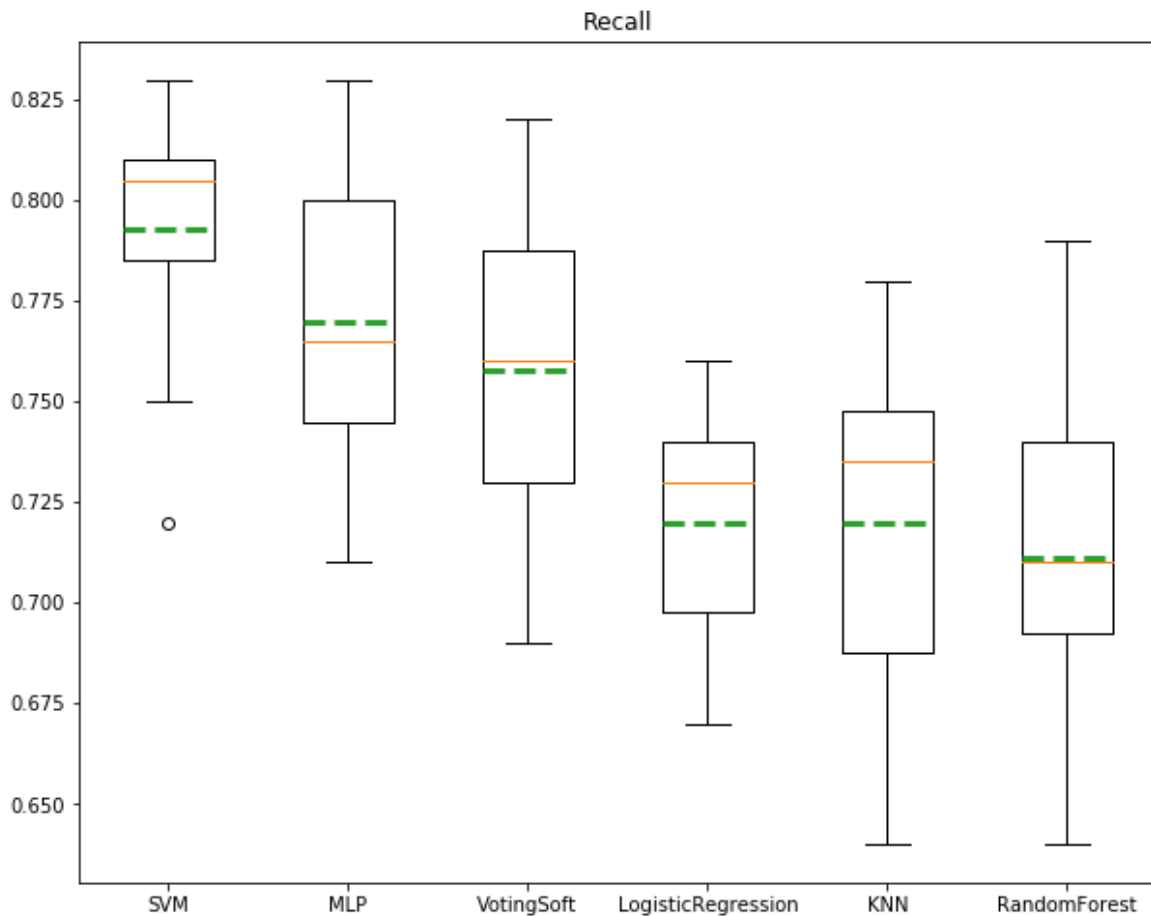


Fig. 82: Confronto recall degli algoritmi di classificazione del dropout sportivo

Il *Support Vector Machine* è il classificatore con *recall* media più alta (linea verde tratteggiata) pari a **0.79**: in termini pratici, ciò sta a significare che, sottoponendo nuovi ragazzi ad un questionario predittivo sul dropout sportivo, mediamente i casi reali di abbandono vengono predetti correttamente nel 79% dei casi (*True Positive*) mentre vengono predetti in modo errato nel 21% dei casi (*False Negative*).

Si ricorda che i parametri interni ad ogni modello (*hyperparameters*) sono stati ottimizzati nell'ottica di massimizzare la *recall*, ovvero per limitare i mancati riconoscimenti dell'abbandono quando si è verificato (falsi negativi).

AVG Recall SVM:	0.79
AVG Recall MLP:	0.77
AVG Recall VotingSoft:	0.76
AVG Recall LogisticRegression:	0.72
AVG Recall KNN:	0.72
AVG Recall RandomForest:	0.71

Fig. 83: Valori medi recall degli algoritmi di classificazione del dropout sportivo

Di seguito, viene riportato il confronto delle performance in base alla metrica *F1-Score*.

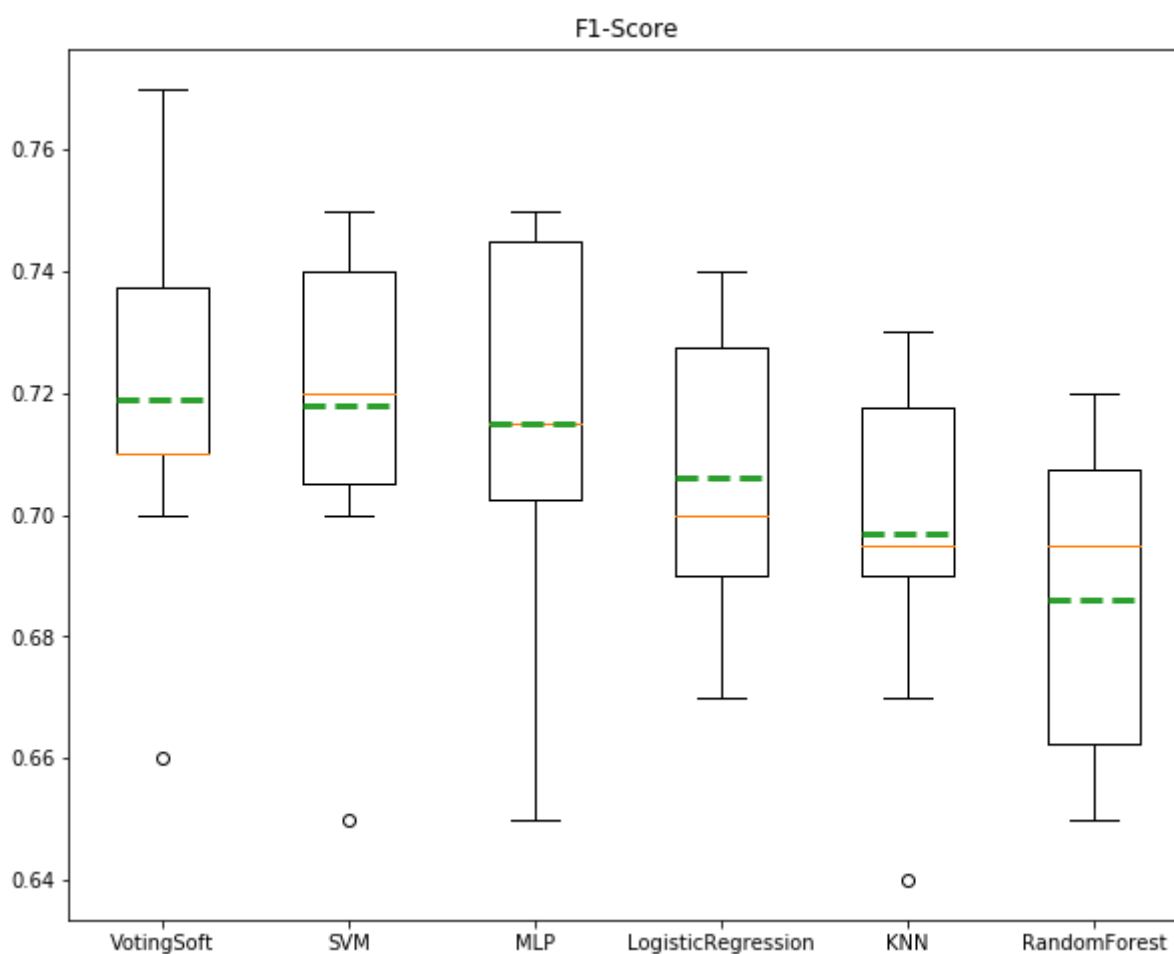


Fig. 84: Confronto *F1-Score* degli algoritmi di classificazione del dropout sportivo

Il *Voting Soft* è il classificatore con *F1-Score* medio più alto (linea verde tratteggiata) pari a **0.72**. Da questo confronto, si ricava la conclusione che questo classificatore è il migliore nell'armonizzare le metriche di *precision* e *recall*.

AVG F1-Score VotingSoft:	0.72
AVG F1-Score SVM:	0.72
AVG F1-Score MLP:	0.72
AVG F1-Score LogisticRegression:	0.71
AVG F1-Score KNN:	0.7
AVG F1-Score RandomForest:	0.69

Fig. 85: Valori medi *F1-score* degli algoritmi di classificazione del dropout sportivo

Di seguito, viene riportato il confronto tra le *curve ROC* e i valori *AUC* dei modelli.

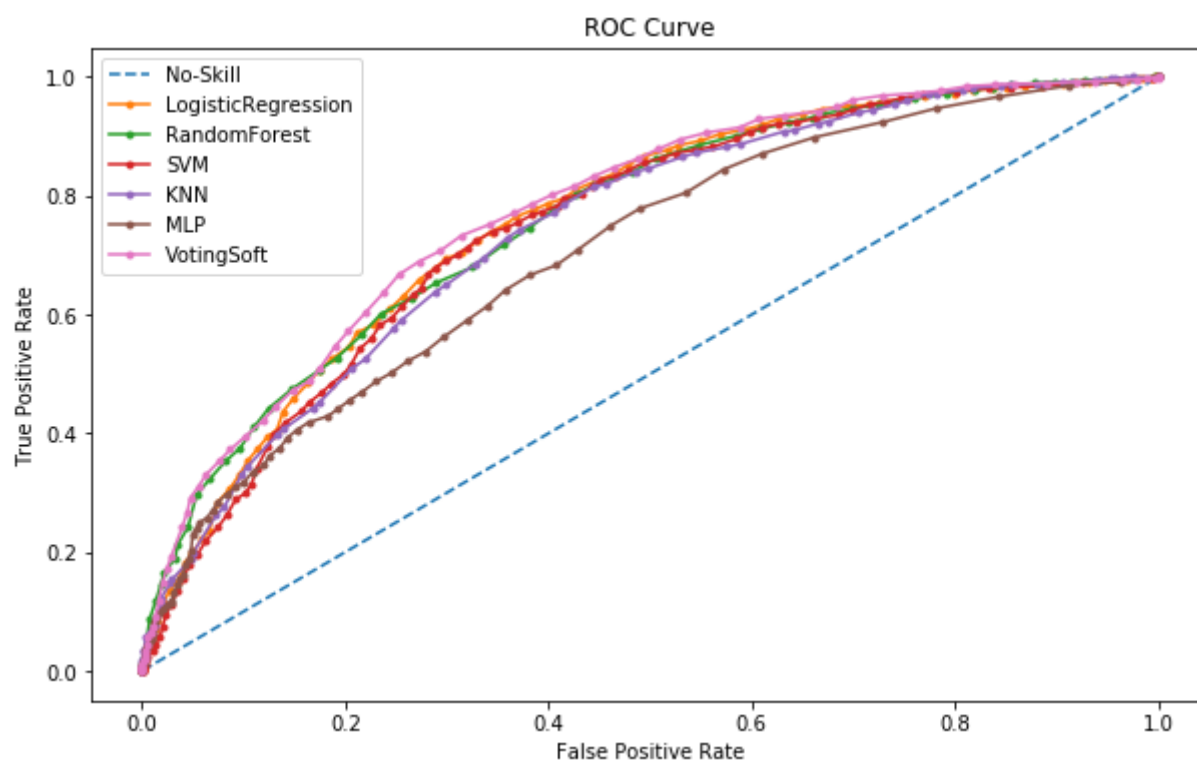


Fig. 86: ROC Curve degli algoritmi di classificazione del dropout sportivo

Il *Voting Soft* è il classificatore con valore *AUC* più alto, pari a **0.77**.

Da questo confronto, si ricava la conclusione che questo classificatore risulta il migliore in base alla relazione tra *True Positive Rate* ($TPR = \text{Sensitivity}$) e *False Positive Rate* ($FPR = 1 - \text{Specificity}$), al variare della soglia (*threshold*) con la quale si predice il dropout sportivo considerando la probabilità stimata dal modello.

AVG AUC VotingSoft:	0.77
AVG AUC MLP:	0.76
AVG AUC LogisticRegression:	0.75
AVG AUC RandomForest:	0.75
AVG AUC SVM:	0.75
AVG AUC KNN:	0.75

Fig. 87: Valori medi AUC dei classificatori del dropout sportivo

Ricapitolando, i migliori classificatori del dropout sportivo nel dataset a disposizione, in base a ciascuna metrica di valutazione delle performance sono:

- *Accuracy* → Voting Soft (0.70)
- *Precision* → Logistic Regression (0.69)
- *Recall* → Support Vector Machine (0.79)
- *F1-Score* → Voting Soft (0.72)
- *AUC* → Voting Soft (0.77)

Nel momento in cui un nuovo ragazzo si sottopone al questionario predittivo del dropout sportivo, verrà scelto un determinato modello predittivo tra quelli precedentemente elencati, in base alla metrica di performance che si vuole ottimizzare.

ANALISI DELLA CAUSALITA'

Capitolo Primo

Two Stage Least Squares

1. Variabili strumentali

In un modello di regressione $Y = \beta_0 + \sum_i \beta_i * X_i + u$, vi sono tre importanti minacce alla validità interna:

- errore da *variabile omessa* che è correlata con un regressore X_i , ma non essendo osservabile non può essere inclusa nella regressione;
- errore dovuto alla *causalità simultanea* (X_i determina Y , Y determina X_i);
- errori nelle variabili (X_i è misurata con errore – *errore di misura*).

Attraverso la regressione mediante *variabili strumentali* (*instrument variables – IV*) è possibile eliminare l'errore determinato da queste tre fonti e, quindi, ottenere le stime dei coefficienti β in maniera non distorta (*unbiased estimation*).

Ipotizzando di voler “strumentare” un singolo regressore X_i , la regressione IV ha lo scopo di isolare la parte non correlata con l'errore u , permettendo di ottenere una stima consistente del coefficiente β_i .

Questo processo è reso possibile dall'utilizzo di una variabile strumentale Z_i , detta anche semplicemente *strumento*, non correlata con l'errore e in grado di catturare i movimenti in X_i che non sono correlati con l'errore con lo scopo di utilizzarli per la stima di β_i .

All'interno di un modello econometrico, una variabile si dice *endogena* quando è correlata con l'errore, ovvero quando è “determinata nell'ambito di un sistema” e per questo soggetta a causalità simultanea con Y .

Al contrario, una variabile si dice *esogena* quando non è correlata con l'errore, ovvero quando è “indipendente nell'ambito del sistema”.

Utilizzando la regressione IV si risolve il problema di *endogeneità* di una variabile presente nel modello, isolando la parte non correlata con l'errore e ottenendo una stima consistente del relativo coefficiente di regressione.

Lo *strumento Z* deve essere opportunamente scelto, seguendo ipotesi di ricerca e affinché esso sia ritenuto valido, deve soddisfare due condizioni:

1. *rilevanza*: $\text{corr}(Z, X_i) \neq 0$
2. *esogeneità*: $\text{corr}(Z, e) = 0$

La condizione di *rilevanza* impone che il regressore endogeno e lo strumento siano significativamente correlati, mentre la condizione di *esogeneità* impone che lo strumento non sia correlato con l'errore.

In presenza di *variabili endogene* nel modello, la stima dei coefficienti mediante il *metodo OLS* tradizionale non è più valida, bensì si utilizza il *metodo dei minimi quadrati a due stadi*, conosciuto come *Two Stage Least Squares*, o semplicemente *TSLs*.

Come descritto dal termine, questo metodo è composto da due stadi, ognuno dei quali riferito ad una regressione.

(1) Nel primo stadio, si isola la parte del regressore X_i che non è correlata con l'errore: in termini pratici, si regredisce X_i sullo strumento Z utilizzando il metodo OLS tradizionale.

$$X_i = \pi_0 + \pi_1 Z_i + v_i$$

Fig. 88: Regressione del primo stadio del metodo Two Stage Least Squares

Stimando i coefficienti π_0 e π_1 , è possibile calcolare i valori predetti di X_i .

$$\hat{X}_i = \hat{\pi}_0 + \hat{\pi}_1 Z_i$$

Fig. 89: Valori predetti del primo stadio del metodo Two Stage Least Squares

(2) Nel secondo stadio, all'interno del modello di regressione di interesse, si sostituisce X_i con i valori predetti di X_i determinati nel primo stadio e si effettua la regressione utilizzando il metodo OLS tradizionale.

$$Y_i = \beta_0 + \beta_1 \hat{X}_i + u_i$$

Fig. 90: Regressione del secondo stadio del metodo Two Stage Least Squares

$\hat{\beta}_1^{TSLs}$ rappresenta lo stimatore dei *minimi quadrati a due stadi* e si tratta di una stima consistente, in quanto il regressore \hat{X}_i rappresenta la parte non correlata con l'errore, isolata mediante lo strumento Z durante il primo stadio.

Di seguito, viene ricapitolato il procedimento del metodo TSLs:

Supporre di conoscere uno strumento valido, Z_i .

Fase 1:

Regredire X_i su Z_i , e ottenere un valore predetto \hat{X}_i

Fase 2:

Regredire Y_i su \hat{X}_i ; il coefficiente su \hat{X}_i è lo stimatore TSLs, $\hat{\beta}_1^{TSLs}$.

Allora $\hat{\beta}_1^{TSLs}$ è uno stimatore consistente di β_1 .

Fig. 91: Descrizione sintetica del metodo Two Stage Least Squares

2. Effetto della distanza da casa alla struttura sportiva

In questo paragrafo si indaga su come incide sull'abbandono della pratica sportiva il vincolo della distanza da casa alla struttura sportiva presso la quale il ragazzo adolescente pratica l'attività sportiva, tenendo presente l'eventuale disponibilità di accompagnamento da parte dei genitori.

Avendo a disposizione i dati relativi alla *frequenza settimanale* della pratica sportiva da parte del ragazzo e i dati relativi all'eventuale *disponibilità ad essere accompagnati* alla struttura sportiva dai genitori, è possibile combinare queste due variabili per ottenere l'*ipotetica frequenza con la quale il ragazzo viene accompagnato dai genitori* presso la struttura sportiva.

Come ipotesi di ricerca, si vuole verificare la correlazione positiva con la distanza da casa alla struttura sportiva a piedi: ovvero, più è lontana la struttura sportiva più è necessario che il ragazzo venga accompagnato dai genitori tramite un mezzo di trasporto familiare.

frequenza_settimanale = {1: "Una volta", 2: "Due volte", 3: "Tre volte", 4: "Più di tre volte"}

disponibilita_accompagnamento_genitori = {0: "No", 1: "Si"}

frequenza settimanale	disponibilita accompagnamento genitori	ipotetica frequenza accompagnamento genitori = frequenza settimanale * disponibilita accompagnamento genitori
1	0	0
2	0	0
3	0	0
4	0	0
1	1	1
2	1	2
3	1	3
4	1	4

Si stima un *modello di regressione* per determinare l'abbandono sportivo in base alle seguenti variabili esplicative (*regressori*):

- *grado divertimento allenamento*: primo predittore più significativo (vedi Fig. 55)
- *livello personalità competitiva*: secondo predittore più significativo (vedi Fig. 55)
- *eventuale pratica a livello agonistico*: terzo predittore più significativo (vedi Fig. 55)
- *ipotetica frequenza accompagnamento genitori*

```
Call:
lm(formula = dropout ~ grado_divertimento_allenamento + personalita_competitiva +
    agonismo + frequenza_accompagnamento_genitori, data = data)

Residuals:
    Min       1Q   Median       3Q      Max
-1.0355 -0.4266  0.1124  0.4345  0.7628

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      1.148712   0.041736  27.523 < 2e-16 ***
grado_divertimento_allenamento -0.094836   0.008894 -10.662 < 2e-16 ***
personalita_competitiva -0.037624   0.005840  -6.443 1.38e-10 ***
agonismo        -0.173945   0.019322  -9.002 < 2e-16 ***
frequenza_accompagnamento_genitori  0.006423   0.005838   1.100  0.271
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4735 on 2745 degrees of freedom
Multiple R-squared:  0.1044,    Adjusted R-squared:  0.1031
F-statistic: 80.02 on 4 and 2745 DF,  p-value: < 2.2e-16
```

Fig. 92: Output modello di regressione di interesse

Dai risultati in Fig. 92, viene confermata la significatività dei coefficienti relativi ai tre predittori più significativi del dropout sportivo in adolescenza, ricavati tramite il calcolo delle features importance (vedi Parte Terza Capitolo II).

Dal segno negativo dei coefficienti, inoltre, viene confermato che il rischio di dropout sportivo diminuisce all'aumentare del grado di divertimento durante l'allenamento, all'aumentare del livello di personalità competitiva del ragazzo e se si pratica sport a livello agonistico.

Tuttavia, non risulta significativo il coefficiente relativo alla variabile "*frequenza accompagnamento genitori*", deducendo il rischio che si tratti di una *variabile endogena* all'interno del modello.

Per tale ragione, si procede a strumentarla con la variabile rappresentante il *tempo da casa alla struttura sportiva a piedi*, seguendo l'ipotesi di esistenza di una correlazione positiva.

Lo strumento “*tempo_casa_struttura_sportiva_piedi*” è una variabile numerica discreta con possibili valori da 0 a 4, i quali codificano rispettivamente “*Meno di un quarto d’ora*”, “*Un quarto d’ora*”, “*Tra un quarto d’ora e un’ora*”, “*Un’ora*” e “*Più di un’ora*”.

Si procede a verificare la *condizione di rilevanza* dello strumento, effettuando la regressione della variabile da strumentare (*frequenza accompagnamento genitori*) sullo strumento (*tempo da casa alla struttura sportiva a piedi*):

```
Call:
lm(formula = frequenza_accompagnamento_genitori ~ tempo_casa_struttura_sportiva_piedi,
    data = data)

Residuals:
    Min       1Q   Median       3Q      Max
-2.2013 -1.9068  0.7987  1.0932  2.0932

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)         1.90679    0.04306  44.279  <2e-16 ***
tempo_casa_struttura_sportiva_piedi > 1 ora  0.08055    0.13453   0.599  0.5494
tempo_casa_struttura_sportiva_piedi 1 ora    0.29451    0.13609   2.164  0.0305 *
tempo_casa_struttura_sportiva_piedi 15 minuti 0.18054    0.07905   2.284  0.0225 *
tempo_casa_struttura_sportiva_piedi tra 15 minuti e 1 ora 0.24640    0.08553   2.881  0.0040 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.602 on 2745 degrees of freedom
Multiple R-squared:  0.004744, Adjusted R-squared:  0.003294
F-statistic: 3.271 on 4 and 2745 DF, p-value: 0.011
```

Fig. 93: Output modello di regressione per verificare la condizione di rilevanza dello strumento

Come espresso nell’ipotesi di ricerca, la *frequenza di accompagnamento dei genitori* è significativamente correlata in modo positivo con il *tempo da casa alla struttura sportiva a piedi*.

Viene esclusa, comunque, la correlazione nel caso in cui la distanza temporale sia maggiore di un’ora: in questo caso, è ragionevole pensare che il ragazzo viva in un paese sprovvisto di strutture sportive e quindi che svolga attività sportiva in spazi pubblici.

In questo scenario non assume importanza la disponibilità all’accompagnamento da parte dei genitori.

A questo punto, si procede a verificare la *condizione di esogeneità* dello strumento: si include come regressore nella regressione di interesse in Fig. 92 e si valuta la significatività del coefficiente stimato.

```

Call:
lm(formula = dropout ~ grado_divertimentoAllenamento + personalita_competitiva +
    agonismo + frequenza_accompagnamento_genitori + tempo_casa_struttura_sportiva_piedi,
    data = data)

Residuals:
    Min       1Q   Median       3Q      Max
-1.0437 -0.4286  0.1161  0.4309  0.7701

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    1.138744   0.042466  26.815 < 2e-16 ***
grado_divertimentoAllenamento -0.094689   0.008894 -10.646 < 2e-16 ***
personalita_competitiva -0.037545   0.005839  -6.430 1.5e-10 ***
agonismo      -0.172593   0.019349  -8.920 < 2e-16 ***
frequenza_accompagnamento_genitori  0.005949   0.005850   1.017  0.309
tempo_casa_struttura_sportiva_piedi  0.009656   0.007620   1.267  0.205
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '.' 0.05 ' ' 1

Residual standard error: 0.4735 on 2744 degrees of freedom
Multiple R-squared:  0.1049,    Adjusted R-squared:  0.1033
F-statistic: 64.35 on 5 and 2744 DF,  p-value: < 2.2e-16

```

Fig. 94: Output modello di regressione per verificare la condizione di esogeneità dello strumento

Dai risultati ottenuti in Fig. 94, si evince che il coefficiente dello strumento “tempo_casa_struttura_sportiva_piedi” non è significativo, per cui viene sconsigliata l’accusa di utilizzare uno strumento endogeno.

Avendo dimostrato che lo strumento scelto è valido, ovvero soddisfa le condizioni di rilevanza ed esogeneità, si passa infine ad applicare il metodo della regressione mediante variabile strumentale utilizzando il metodo *Two Stage Least Squares*.

Per fare ciò viene utilizzata la procedura “IVREG” della libreria “AER” del software R.

Stage 1: regressione OLS “frequenza accompagnamento genitori” su “grado divertimento allenamento”, “personalita competitiva”, “agonismo”, “tempo casa struttura sportiva piedi”

Stage 2: regressione OLS “dropout” su “grado divertimento allenamento”, “personalita competitiva”, “agonismo”, “fitted(Stage 1)”

N.B.: “fitted(Stage 1)” rappresenta la parte isolata della variabile “frequenza accompagnamento genitori” strumentata con lo strumento “tempo casa struttura sportiva piedi”, quindi non correlata con l’errore del modello

```

Call:
ivreg(formula = dropout ~ grado_divertimentoAllenamento + personalita_competitiva +
      agonismo + frequenzaAccompagnamentoGenitori | grado_divertimentoAllenamento +
      personalita_competitiva + agonismo + tempoCasaStrutturaSportivaPiedi,
      data = data)

Residuals:
      Min       1Q   Median       3Q      Max
-1.39401 -0.48330  0.02999  0.41878  1.16889

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    1.002612   0.084251  11.900 < 2e-16 ***
grado_divertimentoAllenamento -0.104281  0.010970  -9.506 < 2e-16 ***
personalita_competitiva -0.052227  0.009591  -5.445 5.63e-08 ***
agonismo      -0.284507  0.057246  -4.970 7.11e-07 ***
frequenzaAccompagnamentoGenitori  0.163048  0.075326   2.165  0.0305 *

Diagnostic tests:
              df1  df2 statistic  p-value
Weak instruments  4 2742     5.238 0.000335 ***
Wu-Hausman       1 2744     5.508 0.019003 *
Sargan           3   NA     0.553 0.907044
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.532 on 2745 degrees of freedom
Multiple R-Squared:  -0.1304,    Adjusted R-squared:  -0.132
Wald test: 64.33 on 4 and 2745 DF, p-value: < 2.2e-16

```

Fig. 95: Output modello di regressione con metodo IV

Dai risultati ottenuti in Fig. 95, si evince che la variabile “*frequenza accompagnamento genitori*”, strumentata mediante il “*tempo casa struttura sportiva a piedi*” è significativa per l’abbandono, con *p-value* pari a **0.03**, contrariamente a quanto ottenuto per la regressione in Fig. 92, dove non veniva strumentata.

In particolare, valutando la stima del suo coefficiente, pari a **0.16**, si deduce che il rischio di abbandono aumenta del 16% per ogni volta che in una settimana il ragazzo viene accompagnato dai genitori presso la struttura sportiva, tenendo presente il vincolo della distanza da casa alla struttura sportiva.

Questo risultato è associabile all’interpretazione per la quale il fatto che il ragazzo necessita dell’accompagnamento dei genitori, a causa della difficoltà nel raggiungere la struttura sportiva, riduce il suo sentimento di indipendenza, portandolo ad abbandonare.

L’effetto di questo *vincolo strutturale* viene limitato dall’elevato grado di divertimento durante gli allenamenti, dall’elevato livello di personalità competitiva e dall’eventuale pratica sportiva a livello agonistico.

Dal test “*Weak instruments*” si rifiuta l’ipotesi nulla per la quale lo strumento sia debole e dal test di endogeneità “*Wu-Hausman*” si rifiuta l’ipotesi nulla per la quale la variabile “*frequenza accompagnamento genitori*”, senza essere strumentata, non sia correlata con l’errore del modello.

QUESTIONARIO PREDITTIVO

Capitolo Primo

Costruzione del questionario

1. Descrizione delle domande

Il questionario predittivo del dropout sportivo ha lo scopo di utilizzare un sottoinsieme delle domande che fanno parte del questionario usato per la raccolta dati, affinché si possano ricavare le informazioni dei predittori più significativi dell'abbandono in età adolescenziale e, contemporaneamente, fornire la probabilità di abbandono, per un ragazzo di età compresa tra i 9 e i 17 anni.

La forma verbale delle domande viene cambiata dal tempo passato al tempo presente e il sottoinsieme delle domande viene determinato congiungendo due diversi gruppi:

1. gruppo delle domande associate alle variabili correlate all'abbandono, escludendo il caso di chi non ha mai pensato di abbandonare: si tratta delle variabili significative ottenute dal test del *Chi-Quadro* (vedi Parte Seconda Capitolo II Paragrafo 3 – Fig. 25)
2. gruppo delle domande associate alle variabili importanti per la predizione dell'abbandono: si tratta dei predittori significativi ottenuti dal task di *Feature Selection* effettuato mediante la valutazione dell'importanza delle features secondo l'algoritmo *Random Forest* (vedi Parte Terza Capitolo II Paragrafo 3 – Fig. 55)

Le risposte ottenute per le domande del *gruppo 1* servono a ricavare indicazioni significative sul possibile abbandono e, contemporaneamente, incrementare la numerosità del campione a disposizione per analisi future, come ad esempio l'inserimento di queste variabili nella pipeline avente finalità predittiva.

Le risposte ottenute per le domande del *gruppo 2*, invece, costituiscono la base su cui applicare i modelli predittivi sviluppati nella *Parte Quarta*: infatti, dopo aver trasformato le risposte in modo opportuno (*encoding*) viene calcolata la probabilità di abbandono.

Ricapitolando, le domande del *gruppo 1* sono riferite alle seguenti variabili:

- età
- tre motivazioni più significative che spingono il ragazzo ad abbandonare la pratica sportiva
- stato d'animo prima dell'allenamento
- autoefficacia (capacità di allenarsi quando si è preoccupati, stanchi o indaffarati)
- percezione del carattere dell'allenatore
- rapporto con l'allenatore
- rapporto con i compagni di squadra
- sentimenti del ragazzo dopo una vittoria o una sconfitta sportiva

Le domande del *gruppo 2*, legate al calcolo della probabilità di abbandono, sono:

- sesso
- pratica agonistica
- tipologia dello sport più significativo praticato (individuale o squadra)
- praticare calcio per i ragazzi
- praticare nuoto o pallavolo per le ragazze
- frequenza settimanale della pratica sportiva
- tre motivazioni più significative che spingono il ragazzo a continuare la pratica sportiva
- grado di divertimento durante l'allenamento
- grado con il quale si riesce a conciliare l'impegno sportivo con l'impegno scolastico
- livello di pratica sportiva dei genitori (nessuna, per svago o agonistica)
- definizione personale di cosa significa praticare sport
- descrizione della personalità

Il questionario predittivo, allegato in *Appendice B*, è stato sviluppato in una forma pilota tramite la piattaforma *PythonAnywhere*, in modo da renderlo pubblico per testarne la funzionalità e valutare le probabilità di abbandono calcolate per questionari di test.

2. Casi di test

In questo paragrafo vengono testati tre casi specifici per valutare la probabilità di abbandono calcolata dal questionario predittivo:

1. probabilità di abbandono bassa;
2. probabilità di abbandono alta;
3. probabilità di abbandono nella fascia di incertezza (intorno al 50%)

Caso 1: Probabilità di abbandono bassa

Ipotizzando un ragazzo con basso rischio di abbandono, viene svolto il questionario predittivo fornendo le seguenti risposte alle domande sui due predittori più significativi del dropout (Fig. 55):

- *grado divertimento allenamento*: Molto
- *personalità competitiva*: Completamente d'accordo

Di seguito, viene riportata la tabella relativa al calcolo della probabilità di abbandono con la predizione della *label* (0 → No abbandono, 1 → Sì abbandono) per ogni modello predittivo sviluppato nella *Parte Quarta*.

Questionario Predittivo

Risultati							
Classificatore	Prob Dropout	Pred Dropout	Accuracy	Precision	Recall	F1-Score	AUC
Voting Soft	14.16	0	0.7	0.68	0.76	0.72	0.77
Logistic Regression	18.04	0	0.69	0.69	0.72	0.71	0.75
Random Forest	25.39	0	0.67	0.66	0.71	0.69	0.75
Support Vector Machine	8.32	0	0.68	0.66	0.79	0.72	0.75
K-Nearest Neighbors	13.07	0	0.68	0.67	0.72	0.7	0.75
Multi-Layer Perceptron	5.97	0	0.69	0.67	0.77	0.72	0.76

Fig. 96: Tabella risultati del caso di test con bassa probabilità di abbandono

Dai risultati della tabella in Fig. 92, si nota come la probabilità di abbandono assume valore minimo pari a **5.97%** per il *Multi-Layer Perceptron* e valore massimo per il *Random Forest*, pari a

25.39%; per il *Voting Soft*, ovvero il classificatore con più alta *accuracy*, la probabilità di abbandono è pari a **14.16%** e rappresenta la media delle probabilità calcolate dagli altri.

Tutti i classificatori stimano una probabilità di abbandono inferiore al 50%, di conseguenza, la *label* predetta per l'abbandono è "No" per tutti i classificatori.

Si può concludere che il ragazzo rappresentante questo caso di test non è a rischio per quanto riguarda il dropout sportivo durante il periodo adolescenziale.

Caso 2: Probabilità di abbandono alta

Ipotizzando un ragazzo con alto rischio di abbandono, viene svolto il questionario predittivo fornendo le seguenti risposte alle domande sui due predittori più significativi del dropout (Fig. 55):

- *grado divertimento allenamento*: Per niente
- *personalità competitiva*: Completamente in disaccordo

Di seguito, viene riportata la tabella relativa al calcolo della probabilità di abbandono con la predizione della *label* (0 → No abbandono, 1 → Sì abbandono) per ogni modello predittivo sviluppato nella *Parte Quarta*.

Questionario Predittivo

Risultati

Classificatore	Prob Dropout	Pred Dropout	Accuracy	Precision	Recall	F1-Score	AUC
Voting Soft	73.32	1	0.7	0.68	0.76	0.72	0.77
Logistic Regression	91.33	1	0.69	0.69	0.72	0.71	0.75
Random Forest	71.14	1	0.67	0.66	0.71	0.69	0.75
Support Vector Machine	71.53	1	0.68	0.66	0.79	0.72	0.75
K-Nearest Neighbors	69.94	1	0.68	0.67	0.72	0.7	0.75
Multi-Layer Perceptron	62.65	1	0.69	0.67	0.77	0.72	0.76

Fig. 97: Tabella risultati per il caso di test con alta probabilità di abbandono

Dai risultati della tabella in Fig. 93, si nota come la probabilità di abbandono assume valore minimo pari a **62.65%** per il *Multi-Layer Perceptron* e valore massimo per la *Logistic Regression*,

pari a **91.33%**; per il *Voting Soft*, ovvero il classificatore con più alta *accuracy*, la probabilità di abbandono è pari a **73.32 %** e rappresenta la media delle probabilità calcolate dagli altri.

Tutti i classificatori stimano una probabilità di abbandono superiore al 50%, di conseguenza, la *label* predetta per l'abbandono è "Sì" per tutti i classificatori.

Si può concludere che il ragazzo rappresentante questo caso di test è significativamente a rischio per quanto riguarda il dropout sportivo durante il periodo adolescenziale.

Caso 3: Probabilità di abbandono nella fascia di incertezza (intorno al 50%)

Ipotizzando un ragazzo con rischio medio di abbandono, ovvero nella fascia intorno al 50%, viene svolto il questionario predittivo fornendo le seguenti risposte alle domande sui due predittori più significativi del dropout (Fig. 55):

- *grado divertimento allenamento*: Neutrale
- *personalità competitiva*: Né in accordo né in disaccordo

Di seguito, viene riportata la tabella relativa al calcolo della probabilità di abbandono con la predizione della *label* (0 → No abbandono, 1 → Sì abbandono) per ogni modello predittivo sviluppato nella *Parte Quarta*.

Questionario Predittivo

Risultati							
Classificatore	Prob Dropout	Pred Dropout	Accuracy	Precision	Recall	F1-Score	AUC
Voting Soft	56.36	1	0.7	0.68	0.76	0.72	0.77
Logistic Regression	54.67	1	0.69	0.69	0.72	0.71	0.75
Random Forest	56.1	1	0.67	0.66	0.71	0.69	0.75
Support Vector Machine	56.53	1	0.68	0.66	0.79	0.72	0.75
K-Nearest Neighbors	46.67	0	0.68	0.67	0.72	0.7	0.75
Multi-Layer Perceptron	67.81	1	0.69	0.67	0.77	0.72	0.76

Fig. 98: Tabella risultati per il caso di test con probabilità media di abbandono

Dai risultati della tabella in Fig. 94, si nota come la probabilità di abbandono assume valore minimo pari a **46.67%** per il *K-Nearest Neighbors* e valore massimo per il *Multi-Layer Perceptron*, pari a **67.81%**; per il *Voting Soft*, ovvero il classificatore con più alta *accuracy*, la probabilità di abbandono è pari a **56.36 %** e rappresenta la media delle probabilità calcolate dagli altri.

Le probabilità di abbandono calcolate dai vari classificatori cadono all'interno della fascia di incertezza intorno al 50% per predire l'effettivo abbandono.

Per decidere se considerare a rischio abbandono o meno il seguente caso, è possibile seguire una linea guida basata sulla metrica di *recall*.

Facendo riferimento al *Support Vector Machine* che massimizza la *recall*, ovvero limita il potenziale rilevamento di un *falso negativo*, inteso come predizione di un non abbandono che invece si verifica, si nota che la probabilità di abbandono calcolata, pari a **56.53%**, è superiore al 50%.

Di conseguenza, si segue l'indicazione fornita dal *Support Vector Machine* e quindi si considera il ragazzo a rischio abbandono.

Conclusione

Il campione di dati messo a disposizione da Orangogo conferma che il dropout sportivo è un fenomeno critico nella fascia di età adolescenziale, infatti su 2750 questionari svolti, poco più della metà (50,95%) sono relativi a ragazzi che hanno abbandonato la pratica dello sport per loro più significativo durante questa fascia di età.

Conoscendo i benefici che conferisce l'attività sportiva in questo delicato periodo di sviluppo del ragazzo, sia fisico sia cognitivo, il dato che a livello nazionale 1 ragazzo su 2 abbandona la pratica sportiva è davvero preoccupante.

Congiungendo le evidenze ottenute dall'analisi statistica descrittiva e dal test del Chi-Quadro (*vedi Parte Seconda Capitolo Secondo Paragrafo 2 e 3*), sul campione a disposizione si traggono le seguenti conclusioni rispetto all'abbandono:

- la differenziazione della pratica sportiva risulta essere un fattore protettivo, mentre la specializzazione in un solo sport risulta essere un fattore di rischio;
- la predisposizione all'abbandono è maggiore per le ragazze rispetto ai ragazzi;
- dal punto di vista geografico, la predisposizione all'abbandono aumenta man mano che si scende dal Nord verso il Sud;
- per quanto riguarda i ragazzi, praticare sport di squadra risulta essere un fattore protettivo, mentre praticare sport a livello individuale risulta essere un fattore di rischio;
- sia per i ragazzi sia per le ragazze, praticare sport a livello agonistico risulta essere un fattore protettivo;
- per i ragazzi, il calcio risulta essere lo sport più praticato, seguito da nuoto e basket, ma a differenza di questi ultimi, la percentuale di chi abbandona è inferiore alla percentuale di chi non abbandona;
- per le ragazze, invece, la pallavolo risulta essere lo sport più praticato seguito da nuoto e danza: tenendo conto della predisposizione all'abbandono data dal sesso, l'abbandono è meno frequente per pallavolo e danza rispetto al nuoto;
- per i ragazzi, le tre principali motivazioni per le quali si è spinti a continuare la pratica sportiva sono: passione (23%), divertimento (19%) e tenersi in forma (9%);
- per le ragazze, le tre principali motivazioni per le quali si è spinti a continuare la pratica sportiva sono: passione (37%), scaricare energia e tensione (12%) e divertimento (10%);

- per i ragazzi, le tre principali motivazioni per le quali si è spinti ad abbandonare la pratica sportiva sono: non divertimento (11%), non riuscire a conciliare l'impegno scolastico con l'impegno sportivo (11%) e un periodo critico personale (9%);
- per le ragazze, le tre principali motivazioni per le quali si è spinti ad abbandonare la pratica sportiva sono: non riuscire a conciliare l'impegno scolastico con l'impegno sportivo (24%), incomprensioni con l'allenatore (12%) e un periodo critico personale (10%);
- considerando chi attraversa un momento critico, l'abbandono sportivo risulta significativamente correlato con l'età in cui si abbandona o si pensa di abbandonare: infatti, sia per i ragazzi sia per le ragazze, la percentuale di abbandono aumenta dalla fascia 9-11 anni alla fascia 12-14 anni, nella quale raggiunge il picco massimo; successivamente, decresce nella fascia 15-17 raggiungendo il picco minimo.

La fase di individuazione dei predittori più significativi del dropout sportivo, svolta tramite il task di Feature Selection mediante valutazione della misura di importanza delle features conferita dall'algoritmo Random Forest (*vedi Parte Terza Capitolo Secondo Paragrafo 3*), ha permesso di conoscere che i tre predittori più significativi dell'abbandono sono:

- il grado con il quale il ragazzo si diverte durante gli allenamenti;
- il livello di personalità competitiva da lui descritto;
- l'eventuale pratica sportiva a livello agonistico.

In particolare, facendo riferimento al modello predittivo ottenuto con la *regressione logistica* applicata a queste tre variabili (*vedi Parte Quarta Capitolo Primo Paragrafo 2*), si evince che il rischio di abbandono diminuisce all'aumentare del grado di divertimento durante gli allenamenti, diminuisce all'aumentare del livello di personalità competitiva ed è inferiore se si pratica sport a livello agonistico.

Valutando le performance di classificazione dei vari algoritmi addestrati e, tenendo conto che la non previsione di un reale abbandono, ovvero un *falso negativo*, è più grave rispetto alla previsione di un abbandono che non si verifica, ovvero un *falso positivo*, è stato ottenuto un modello predittivo basato sull'algoritmo *Support Vector Machine*, in grado di fornire una *recall* media pari a 0.79 (*vedi Parte Quarta Capitolo Secondo Paragrafo 2*).

Ciò sta a significare che, considerando 100 ragazzi adolescenti che abbandoneranno la pratica dello sport più significativo, il modello sviluppato è in grado di predire correttamente l'abbandono per 79 di loro.

Tramite il questionario predittivo sviluppato, inoltre, è possibile valutare non solo la previsione della classe [*Abbandono Si / Abbandono No*], ma soprattutto di avere a disposizione una probabilità associata all'abbandono e l'insieme dei dati relativi alle variabili più significative che la determinano (*vedi Parte Sesta Capitolo Primo Paragrafo 2*).

L'insieme di queste informazioni consentiranno ad Orangogo di supportare il ragazzo e di poterlo orientare nel giusto modo con lo scopo di evitare che l'abbandono si realizzi effettivamente.

Appendice A

Questionario per raccolta dati

QUESTIONARIO CASO 1

DATI ANAGRAFICI

1. Sei maschio o femmina? M ☐ F ☐
2. Anno di nascita _____
3. Comune di residenza _____
4. Provincia di residenza _____
5. Nazionalità _____
6. Scuola superiore che frequenti o hai frequentato:
- ☐ Liceo
- ☐ Istituto Tecnico
- ☐ Istituto Professionale
- ☐ Altro _____
7. Titolo di studio raggiunto?
- ☐ Licenza Elementare
- ☐ Licenza Media Inferiore
- ☐ Diploma di scuola media superiore (Liceo)
- ☐ Diploma di scuola media superiore (Istituto Tecnico o Professionale)
- ☐ Laurea
- ☐ Altro _____
8. Attualmente...
- ☐ Lavori
- ☐ Studi
9. In base alla risposta precedente, indica qual è il tuo lavoro e/o cosa stai studiando?
- _____
10. Tra i 9 e i 17 anni... hai avuto una disabilità temporanea (es. gamba ingessata) o permanente?
- Sì, scrivi quale _____
- No. Segna con una X ☐

DATI GENERICI

11. Titolo di studio di tua mamma
- ☐ Licenza Elementare
- ☐ Licenza Media Inferiore
- ☐ Diploma di scuola media superiore (Liceo)

- ☐ Diploma di scuola media superiore (Istituto Tecnico o Professionale)
- ☐ Laurea
- ☐ Altro _____
- ☐ Non lo so

12. Titolo di studio di tuo papà

- ☐ Licenza Elementare
- ☐ Licenza Media Inferiore
- ☐ Diploma di scuola media superiore (Liceo)
- ☐ Diploma di scuola media superiore (Istituto Tecnico o Professionale)
- ☐ Laurea
- ☐ Altro _____
- ☐ Non lo so

13. Quanti fratelli o sorelle hai?

- ☐ Non ne ho
- ☐ 1
- ☐ 2
- ☐ 3
- ☐ Più di tre

14. Dai tuoi 9 ai 17 anni... Quanti dei tuoi nonni in media hanno abitato vicino a te (a circa un quarto d'ora da casa tua)?

- ☐ Nessuno
- ☐ 1
- ☐ 2
- ☐ 3
- ☐ 4

15. Dopo la scuola, in genere, chi si occupava di te?

- ☐ Mamma
- ☐ Papà
- ☐ Fratello o sorella
- ☐ Nonni
- ☐ Baby sitter
- ☐ Zii
- ☐ Nessuno
- ☐ Altro _____

16. Hai cambiato comune di residenza o quartiere durante il periodo delle scuole elementari e medie?

Si. Scrivi in quali anni _____

No. Metti una X _____

17. Tra i 9 e 17 anni... Pensa alla struttura sportiva più vicina a casa tua, **indipendentemente** dal fatto che vi praticassi il tuo sport: quanto tempo impiegavi per raggiungerla?

A piedi...	Meno di un quarto d'ora	Un quarto d'ora	Più di un quarto d'ora	Un'ora	Più di un'ora
In auto/autobus/altro mezzo di trasporto ...	Meno di un quarto d'ora	Un quarto d'ora	Più di un quarto d'ora	Un'ora	Più di un'ora

18. Qual era la distanza approssimativa in chilometri? _____

19. Tra i 9 e 17 anni... Pensa alla struttura sportiva più vicina alla tua scuola, **indipendentemente** dal fatto che vi praticassi il tuo sport: quanto tempo impiegavi per raggiungerla?

A piedi...	Meno di un quarto d'ora	Un quarto d'ora	Più di un quarto d'ora	Un'ora	Più di un'ora
In auto/autobus/altro mezzo di trasporto ...	Meno di un quarto d'ora	Un quarto d'ora	Più di un quarto d'ora	Un'ora	Più di un'ora

20. Qual era la distanza approssimativa in chilometri?

21. Tra i 9 e 17 anni...Hai mai praticato sport?

- ☐ Sì (salta al caso 1, 2 o 3)
- ☐ No (salta al caso 4)

INFORMAZIONI SULLO SPORT (casi 1,2,3)

22. Tra i 9 e 17 anni... Nel tuo quartiere (se vivevi in città) o nel tuo comune (se vivevi in un centro abitato di piccole dimensioni) vivevano o si allenavano dei campioni dello sport?

Sì. Scrivi in quale sport _____

No. Segna con una X _____

23. Quale/i sport hai praticato dai 9 ai 17 anni?

Arrampicata	Equitazione	Ping pong
Arti marziali	Football americano	Rugby
Atletica	Ginnastica Artistica	Scherma
Balli da sala	Ginnastica Ritmica	Sci
Baseball	Golf	Skateboard
Basket	Hockey su ghiaccio	Sport da combattimento
Calcio	Hockey su prato	Tennis
Canottaggio	Nuoto	Tiro Con L'arco
Ciclismo	Nuoto sincronizzato	Tuffi
Danza classica	Pallavolo	Yoga
Danza contemporanea	Pattinaggio a Rotelle	Altro _____
Danza moderna	Pattinaggio su ghiaccio	

24. Scrivi il nome dello sport per te più significativo tra i 9 e i 17 anni _____

25. Crocetta la risposta che ti rappresenta.

“Lo sport più significativo per me...

- ☐ ...È l'unico che ho fatto tra i 9 e i 17 anni compresi” (caso 1)
- ☐ ...È l'unico che ho fatto tra i 9 e 17 anni ma l'ho abbandonato prima dei 17 anni” (caso 3)
- ☐ ...è il più importante tra gli sport che ho praticato e a 17 anni lo praticavo ancora” (caso 2)
- ☐ ... è il più importante tra gli sport che ho praticato e l'ho abbandonato prima dei 17 anni” (caso 3)

CASO 1. Se ha fatto solo uno sport tra i 9 e 17 anni (che continua a fare o ha smesso dopo i 17 anni)

26. Riguardo al tuo sport, lo hai fatto a livello agonistico?

- ☐ Sì
- ☐ No

27. Quante volte alla settimana, in media, praticavi il tuo sport significativo tra i 9 e i 17 anni? Considera anche le partite (o le gare o i saggi).

- ☐ 1 volta
- ☐ 2 volte
- ☐ 3 volte
- ☐ Più di tre

28. Tra i tuoi 9 e 17 anni...Hai mai pensato seriamente di abbandonare lo sport per te più significativo?

NB. Per “seriamente” si intende che stavi per lasciarlo o che l'hai lasciato per poco tempo (massimo 6 mesi) e poi hai ripreso.

- ☐ Sì (continua)
- ☐ No (vai alla domanda 70)

MOMENTO CRITICO

29. A quanti anni hai pensato seriamente di abbandonare lo sport per te più significativo?

- | | |
|-----------------------------|-----------------------------|
| <input type="checkbox"/> 9 | <input type="checkbox"/> 13 |
| <input type="checkbox"/> 10 | <input type="checkbox"/> 14 |
| <input type="checkbox"/> 11 | <input type="checkbox"/> 15 |
| <input type="checkbox"/> 12 | <input type="checkbox"/> 16 |
| <input type="checkbox"/> 13 | <input type="checkbox"/> 17 |
| <input type="checkbox"/> 14 | |
| <input type="checkbox"/> 15 | |

30. Quando hai pensato di abbandonare lo sport per te più significativo, come ti sentivi prima di andare ad allenarti?

- ☐ Stanco
- ☐ Felice
- ☐ Carico

- ☐ Svogliato
- ☐ Demotivato
- ☐ Ansioso
- ☐ Altro (specificare) _____

31. Quando hai pensato di abbandonare lo sport per te più significativo, con che frequenza andavi ad allenarti?

- ☐ Sempre (3 allenamenti su 3)
- ☐ La maggior parte delle volte (2 allenamenti su 3)
- ☐ Raramente (1 allenamento su 3)
- ☐ Mai

32. Nell'ultimo mese, prima di abbandonare il tuo sport, quante ore alla settimana dedicavi ad attività extra-scolastiche e/o extra-sportive (es. arte, social network ecc.)?

- ☐ Nessuna
- ☐ 1 ora
- ☐ 2 ore
- ☐ 3 ore
- ☐ 4 ore
- ☐ Più di 4

33. Quando hai pensato di abbandonare lo sport per te più significativo, quali fattori ti hanno spinto a continuare? Tra le seguenti motivazioni, **sceglie al massimo 3**.

Questioni tecniche
e personali

- ☐ 1. Mi divertivo
- ☐ 2. Mi sentivo bravo
- ☐ 3. Mi piaceva la competizione
- ☐ 4. Non mi sono mai infortunato/a
- ☐ 5. Mi piaceva faticare
- ☐ 6. Mi piaceva migliorare le mie abilità
- ☐ 7. Volevo essere più bravo e/o forte degli altri nel mio sport
- ☐ 8. Volevo tenermi in forma
- ☐ 9. Avevo un obiettivo (es. vincere trofei o raggiungere alti livelli)
- ☐ 10. Mi piaceva provare il brivido della vittoria
- ☐ 11. Mi piaceva battere l'avversario
- ☐ 12. Mi serviva a scaricare l'energia o la tensione
- ☐ 13. Volevo guadagnare denaro e ottenere contratti
- ☐ 14. Mi piacevo quando facevo sport
- ☐ 15. Motivi religiosi
- ☐ 16. Sentivo l'esigenza di fare sport
- ☐ 17. Era una passione
- ☐ 18. Era uno sport alla moda

Rapporti interpersonali:

- ☐ 19. Ero appoggiato dai miei genitori
- ☐ 20. Per far contenti i miei genitori (famiglia)
- ☐ 21. I miei genitori volevano che io facessi sport
- ☐ 22. Per far contento il mio allenatore
- ☐ 23. Bel rapporto tra i miei genitori e l'allenatore
- ☐ 24. Mi trovavo bene con i miei compagni di squadra
- ☐ 25. Avevo qualcuno con cui allenarmi
- ☐ 26. Avevo un allenatore che mi piaceva

Contesto socio-sportivo:

- ☐ 27. Riuscivo a conciliare lo sport con lo studio/lavoro
- ☐ 28. Ho sempre trovato il tempo di fare sport
- ☐ 29. L'impianto era facile da raggiungere
- ☐ 30. L'impianto era vicino casa
- ☐ 31. Lo sport non era troppo costoso
- ☐ 32. Palestra adeguata
- ☐ 33. Orari comodi

☐ 34. Altro _____

34. Tra le motivazioni che hai scelto scrivi il numero di quella che ha maggiormente determinato la tua voglia di continuare

35. Quali fattori, invece, ti spingevano ad abbandonare? Tra le seguenti motivazioni, scegline al massimo 3.

- | | | |
|--------------------------|--------------------------|---|
| Questioni tecniche | <input type="checkbox"/> | 1. non mi divertivo più |
| | <input type="checkbox"/> | 2. non mi sentivo abbastanza bravo |
| | <input type="checkbox"/> | 3. troppa competizione |
| | <input type="checkbox"/> | 4. infortunio/i sportivo/i |
| | <input type="checkbox"/> | 5. eccessiva fatica |
| | <input type="checkbox"/> | 6. incomprensioni con l'allenatore |
| | <input type="checkbox"/> | 7. motivi di salute |
| | <input type="checkbox"/> | 8. periodo critico nella vita personale |
| | <input type="checkbox"/> | 9. sport troppo impegnativo |
| | <input type="checkbox"/> | 10. Disabilità temporanea o permanente |
| Rapporti interpersonali: | <input type="checkbox"/> | 11. non ero appoggiato dai miei genitori |
| | <input type="checkbox"/> | 12. i miei genitori non mi potevano accompagnare |
| | <input type="checkbox"/> | 13. rapporto difficile tra genitori e allenatore |
| | <input type="checkbox"/> | 14. non mi trovavo bene con i compagni |
| | <input type="checkbox"/> | 15. non avevo qualcuno con cui allenarmi |
| | <input type="checkbox"/> | 16. sport troppo lontano |
| Contesto sociale: | <input type="checkbox"/> | 17. non andavo bene a scuola |
| | <input type="checkbox"/> | 18. non riuscivo a conciliare lo sport con l'impegno scolastico |
| | <input type="checkbox"/> | 19. avevo altri interessi prevalenti |
| Contesto sportivo: | <input type="checkbox"/> | 20. difficoltà a raggiungere l'impianto |
| | <input type="checkbox"/> | 21. lo sport era troppo costoso |
| | <input type="checkbox"/> | 22. palestra non adeguata |
| | <input type="checkbox"/> | 23. materiali a disposizione in palestra carenti |
| | <input type="checkbox"/> | 24. orari scomodi |
| | <input type="checkbox"/> | 25. non ci sono sport che mi piacciono dove vivo io |
| | <input type="checkbox"/> | 26. altro _____ |

36. Tra le motivazioni che hai scelto scrivi il numero di quella che maggiormente ti spingeva ad abbandonare

IMPEGNO SCOLASTICO

37. Quando hai pensato di abbandonare lo sport per te più significativo ... Frequentavi...

- ☐ Scuola primaria (elementare)
- ☐ Scuola media inferiore
- ☐ Scuola media superiore

38. Quando hai pensato di abbandonare lo sport per te più significativo ... Frequentavi la scuola...

- ☐ Solo al mattino
- ☐ Mattino e pomeriggio
- ☐ Tutte le mattine e qualche pomeriggio

39. Quando hai pensato di abbandonare lo sport per te più significativo ... Riuscivi a conciliare l'impegno scolastico con l'impegno sportivo

(1= completamente in disaccordo; 2= abbastanza in disaccordo; 3= né in disaccordo né in accordo; 4= abbastanza d'accordo; 5=completamente d'accordo)

1	2	3	4	5
---	---	---	---	---

40. Quando hai pensato di abbandonare lo sport per te più significativo ... Quante ore studiavi in media al giorno al di fuori dell'orario scolastico?

☐ Nessuna ☐ 1 ora ☐ 2 ore ☐ 3 ore ☐ 4 ore ☐ più di 4

SOCIETÀ E ALLENATORE

41. Quando hai pensato di abbandonare lo sport per te più significativo ... Ti sentivi parte integrante della società sportiva a cui appartenevi?

☐ Sì ☐ No

42. Quando hai pensato di abbandonare lo sport per te più significativo ... Il tuo allenatore ha influito sulla tua scelta di continuarlo?

Sì. Scrivi il perché _____

No. Segna con una X _____

43. Quando hai pensato di abbandonare lo sport per te più significativo ... Il tuo allenatore ti dava le giuste attenzioni e importanza (es. ti seguiva negli esercizi quando avevi bisogno)

☐ Sì ☐ No

44. Quando hai pensato di abbandonare lo sport per te più significativo ... Il tuo allenatore ti incoraggiava quando perdevi o sbagliavi

☐ Sì ☐ No

45. Quando hai pensato di abbandonare lo sport per te più significativo ... Il tuo allenatore era (scegline al massimo 3)

<input type="checkbox"/> Severo	<input type="checkbox"/> Autorevole	<input type="checkbox"/> Un amico o confidente
<input type="checkbox"/> Competente	<input type="checkbox"/> Un buon leader	<input type="checkbox"/> Troppo fissato
<input type="checkbox"/> Un punto di riferimento al di fuori della palestra	<input type="checkbox"/> Non in grado di allenare	
<input type="checkbox"/> Avvilente	<input type="checkbox"/> Irritabile	<input type="checkbox"/> Un motivatore
<input type="checkbox"/> Non era sufficientemente preparato tecnicamente	<input type="checkbox"/> Altro _____	

COMPAGNI DI SQUADRA

NB: per "compagni di squadra" si intendono le persone con cui hai praticato il tuo sport (sia esso individuale o di squadra).

Es. le persone con cui hai fatto lezione di danza, o lezione di tennis, o con cui ti sei allenato a basket ecc.

46. Quando hai pensato di abbandonare lo sport per te più significativo ... Avevi compagni di squadra che ti stimolavano a impegnarti in allenamento e/o durante le gare/partite?

☐ Sì ☐ No

47. Quando hai pensato di abbandonare lo sport per te più significativo ... I tuoi compagni di squadra ti hanno influenzato sulla scelta di continuarlo?

Si. Scrivi perché _____

No. Segna con una X _____

48. Quando hai pensato di abbandonare lo sport per te più significativo ... I tuoi compagni di squadra ogni tanto non si comportavano bene con te (es. ti prendevano in giro, non ti calcolavano ecc.)

☐ Si ☐ No

GENITORI

49. Quando hai pensato di abbandonare lo sport per te più significativo ... Tua mamma lavorava?

☐ Si

☐ No

50. Quando hai pensato di abbandonare lo sport per te più significativo ... Che professione svolgeva tua mamma?

☐ Libera professione (avvocato, medico, commercialista, artista, carriera sportiva ecc.)

☐ Commerciante/Imprenditrice/Artigiana

☐ Impiegata

☐ Operaia

☐ Casalinga

☐ Docente

☐ Altro _____

51. Quando hai pensato di abbandonare lo sport per te più significativo ... Tua mamma era impegnata per lavoro...

☐ Full time

☐ Part time

☐ Si gestiva autonomamente

☐ Altro _____

52. Quando hai pensato di abbandonare lo sport per te più significativo ... Tuo papà lavorava?

☐ Si

☐ No

53. Quando hai pensato di abbandonare lo sport per te più significativo ... Che professione svolgeva tuo papà?

☐ Libera professione (avvocato, medico, commercialista, artista, carriera sportiva ecc.)

☐ Commerciante/Imprenditore/Artigiano

☐ Impiegato

☐ Operaio

☐ Casalingo

☐ Docente

☐ Altro _____

54. Quando hai pensato di abbandonare lo sport per te più significativo ... Tuo papà era impegnato per lavoro...

☐ Full time

☐ Part time

☐ Si gestiva autonomamente

☐ Altro _____

55. Quanti dei tuoi fratelli hanno fatto il tuo stesso sport significativo?

- ☐ Nessuno
- ☐ 1
- ☐ 2
- ☐ 3
- ☐ Più di tre

56. Chi ti ha influenzato in modo determinante nella scelta del tuo sport significativo?

- ☐ Genitori ☐ Fratelli/ sorelle ☐ Amici ☐ Qualcun altro _____
- ☐ Nessuno

57. Almeno uno dei tuoi genitori ha praticato nella sua vita il tuo sport significativo (puoi rispondere a più di un'opzione)

- ☐ Sì, a livello agonistico. Scrivi chi: _____
- ☐ Sì, per svago. Scrivi chi: _____
- ☐ No, segna con una X _____
- ☐ Non lo so, segna con una X _____

58. Quando hai pensato di abbandonare lo sport per te più significativo ... I tuoi genitori (o uno di essi) ti portavano la maggior parte delle volte (es. 2 volte su 3) agli allenamenti del tuo sport significativo

- ☐ Sì ☐ No

59. Quando hai pensato di abbandonare lo sport per te più significativo ... I tuoi genitori (o uno di essi) ti hanno spinto a continuare a fare proprio quello sport

- ☐ Sì ☐ No

60. Tra i tuoi 9 e 17 anni... Quando non avevi voglia di andare ad allenarti nel tuo sport, i tuoi genitori (o uno di essi) ti convincevano o obbligavano ad andare

- ☐ Sì ☐ No

61. Pensa a quale dei tuoi genitori ti ha seguito di più nello sport per te più significativo e rispondi:

Lui/lei, rispetto al tuo sport, lo definiresti ...

- ☐ Tranquillo ☐ Troppo fissato ☐ Si arrabbiava alle gare ☐ Sostenitore positivo
- ☐ Incoraggiante ☐ Indifferente ☐ Pesante ☐ Ansiogeno
- ☐ Fastidioso ☐ Altro _____

AUTOEFFICACIA

62. Torna al momento in cui hai pensato di abbandonare seriamente lo sport per te più significativo e rispondi.

1. Completamente in disaccordo
2. Abbastanza in disaccordo
3. Né in accordo né in disaccordo
4. Abbastanza in accordo
5. Completamente in accordo

“Mi sono sentito capace di affrontare l'allenamento...”

...anche quando avevo preoccupazioni e problemi personali	1	2	3	4	5
...anche quando ero stanco	1	2	3	4	5
...anche quando ero indaffarato	1	2	3	4	5

63. Quando hai pensato di abbandonare lo sport per te più significativo ...Partecipavi a gare/partite/saggi?

- ☐ Sì (continua)
- ☐ No (salta a domanda: 66)

GESTIONE DEL MATCH

NB. Per chi ha praticato danza: per "gestione del match" si intende anche "gestione del saggio".

Quindi, nelle due domande successive traduci "perdere una gara" con "quando la tua esibizione non veniva bene" e "vincere una partita" con "quando la tua esibizione veniva bene"

64. Perdere una partita (o una gara) ...

- ☐ Era indifferente
- ☐ Era poco rilevante
- ☐ Era un dramma
- ☐ Era uno stimolo per far meglio
- ☐ Ti faceva venir voglia di smettere
- ☐ Altro _____

65. Vincere una partita (o una gara) ...

- ☐ Era indifferente
- ☐ Era poco rilevante
- ☐ Era una gioia
- ☐ Era uno stimolo per far meglio
- ☐ Ti faceva venir voglia di continuare
- ☐ Altro _____

SPORT E DIVERTIMENTO

66. Quando hai pensato di abbandonare lo sport per te più significativo, ti divertivi ad allenamento...

- ☐ Molto ☐ Abbastanza ☐ Neutrale ☐ Poco ☐ Per niente

67. Quando hai pensato di abbandonare lo sport per te più significativo ... Il tuo sport era ... (scegline al massimo 3)

- ☐ Un sacrificio
- ☐ Un divertimento
- ☐ Un dovere
- ☐ Una passione
- ☐ Un bisogno del tuo corpo e della tua mente

- ☐ Un modo per essere liberi
- ☐ Un modo ricaricarmi
- ☐ Un modo per staccare dalla vita quotidiana
- ☐ Un modo per scaricare l'energia
- ☐ Una perdita di tempo
- ☐ Altro (specificare) _____

CONCLUSIONE

68. Per favore, leggi le seguenti caratteristiche di personalità e indica quanto ti senti descritto da ogni coppia, anche se pensi che una delle due caratteristiche ti descriva più dell'altra.

Ti ritieni una persona...

Completamente in disaccordo	Molto in disaccordo	Un po' in disaccordo	Né d'accordo né in disaccordo	Un po' d'accordo	Molto d'accordo	Completamente d'accordo
1	2	3	4	5	6	7

Sono una persona...

1.	Estroversa, esuberante	1	2	3	4	5	6	7
2.	Polemica, litigiosa	1	2	3	4	5	6	7
3.	Affidabile, auto-disciplinata	1	2	3	4	5	6	7
4.	Ansiosa, che si agita facilmente	1	2	3	4	5	6	7
5.	Aperta alle nuove esperienze, con molti interessi	1	2	3	4	5	6	7
6.	Riservata, silenziosa	1	2	3	4	5	6	7
7.	Comprensiva, affettuosa	1	2	3	4	5	6	7
8.	Disorganizzata, distratta	1	2	3	4	5	6	7
9.	Competitiva	1	2	3	4	5	6	7
10.	Tranquilla, emotivamente stabile	1	2	3	4	5	6	7
11.	Tradizionalista, abitudinaria	1	2	3	4	5	6	7

69. Ti è piaciuto questo test?

- ☐ Sì
- ☐ No

INFO SULLO SPORT (caso 1, che non ha mai abbandonato né ci ha pensato)

70. Perché non hai mai pensato di abbandonare il tuo sport? Tra le seguenti motivazioni, scegline al massimo 3.

- | | | |
|--------------------------------|--------------------------|---|
| Questioni tecniche e personali | <input type="checkbox"/> | 1. Mi divertivo |
| | <input type="checkbox"/> | 2. Mi sentivo bravo |
| | <input type="checkbox"/> | 3. Mi piaceva la competizione |
| | <input type="checkbox"/> | 4. Non mi sono mai infortunato/a |
| | <input type="checkbox"/> | 5. Mi piaceva faticare |
| | <input type="checkbox"/> | 6. Mi piaceva migliorare le mie abilità |
| | <input type="checkbox"/> | 7. Volevo essere più bravo e/o forte degli altri nel mio sport |
| | <input type="checkbox"/> | 8. Volevo tenermi in forma |
| | <input type="checkbox"/> | 9. Avevo un obiettivo (es. vincere trofei o raggiungere alti livelli) |
| | <input type="checkbox"/> | 10. Mi piaceva provare il brivido della vittoria |
| | <input type="checkbox"/> | 11. Mi piaceva battere l'avversario |
| | <input type="checkbox"/> | 12. Mi serviva a scaricare l'energia o la tensione |
| | <input type="checkbox"/> | 13. Volevo guadagnare denaro e ottenere contratti |
| | <input type="checkbox"/> | 14. Mi piacevo quando facevo sport |
| | <input type="checkbox"/> | 15. Motivi religiosi |
| | <input type="checkbox"/> | 16. Sentivo l'esigenza di fare sport |
| | <input type="checkbox"/> | 17. Era una passione |
| | <input type="checkbox"/> | 18. Era uno sport alla moda |
| Rapporti interpersonali: | <input type="checkbox"/> | 19. Ero appoggiato dai miei genitori |
| | <input type="checkbox"/> | 20. Per far contenti i miei genitori (famiglia) |
| | <input type="checkbox"/> | 21. I miei genitori volevano che io facessi sport |
| | <input type="checkbox"/> | 22. Per far contento il mio allenatore |
| | <input type="checkbox"/> | 23. Bel rapporto tra i miei genitori e l'allenatore |
| | <input type="checkbox"/> | 24. Mi trovavo bene con i miei compagni di squadra |
| | <input type="checkbox"/> | 25. Avevo qualcuno con cui allenarmi |
| | <input type="checkbox"/> | 26. Avevo un allenatore che mi piaceva |
| Contesto socio-sportivo: | <input type="checkbox"/> | 27. Riuscivo a conciliare lo sport con lo studio/lavoro |
| | <input type="checkbox"/> | 28. Ho sempre trovato il tempo di fare sport |
| | <input type="checkbox"/> | 29. L'impianto era facile da raggiungere |
| | <input type="checkbox"/> | 30. L'impianto era vicino casa |
| | <input type="checkbox"/> | 31. Lo sport non era troppo costoso |
| | <input type="checkbox"/> | 32. Palestra adeguata |
| | <input type="checkbox"/> | 33. Orari comodi |
| | <input type="checkbox"/> | 34. Altro _____ |

71. Tra le motivazioni che hai scelto scrivi il numero di quella che ha maggiormente determinato la tua voglia di continuare

72. In media, tra i tuoi 9 e 17 anni, quante ore alla settimana dedicavi ad attività extra-scolastiche e/o extra-sportive (es. arte, social network ecc.)?

☐ Nessuna ☐ 1 ora ☐ 2 ore ☐ 3 ore ☐ 4 ore ☐ più di 4

IMPEGNO SCOLASTICO

73. Tra i tuoi 9 e i 17 anni ... Sei in media sempre riuscito a conciliare l'impegno scolastico con l'impegno sportivo?

(1= completamente in disaccordo; 2= abbastanza in disaccordo; 3= né in disaccordo né in accordo; 4= abbastanza d'accordo; 5=completamente d'accordo)

1	2	3	4	5
---	---	---	---	---

74. Tra i 9 e i 17 anni...Quante ore studiavi in media al giorno al di fuori dell'orario scolastico?

- ☐ Nessuna ☐ 1 ora ☐ 2 ore ☐ 3 ore ☐ 4 ore ☐ Più di 4

GENITORI

75. Tra i tuoi 9 e i 17 anni... tua mamma lavorava?

- ☐ Sì
☐ No

76. Tra i tuoi 9 e 17 anni... Che professione svolgeva tua mamma?

- ☐ Libera professione (avvocato, medico, commercialista, artista, carriera sportiva ecc.)
☐ Commerciante/Imprenditrice/Artigiana
☐ Impiegata
☐ Operaia
☐ Casalinga
☐ Docente
☐ Altro _____

77. Tra i tuoi 9 e 17 anni... Tua mamma era impegnata per lavoro...

- ☐ Full time
☐ Part time
☐ Si gestiva autonomamente
☐ Altro _____

78. Tra i tuoi 9 e 17 anni... Tuo papà lavorava?

- ☐ Sì
☐ No

79. Tra i tuoi 9 e 17 anni... Che professione svolgeva tuo papà?

- ☐ Libera professione (avvocato, medico, commercialista, artista, carriera sportiva ecc.)
☐ Commerciante/Imprenditore/Artigiano
☐ Impiegato
☐ Operaio
☐ Casalingo
☐ Docente
☐ Altro _____

80. Tra i tuoi 9 e 17 anni... Tuo papà era impegnato per lavoro...

- ☐ Full time
☐ Part time

☐ Si gestiva autonomamente

☐ Altro _____

81. Tra i tuoi 9 e 17 anni... Quanti dei tuoi fratelli hanno fatto il tuo stesso sport significativo?

☐ Nessuno

☐ 1

☐ 2

☐ 3

☐ Più di tre

82. Tra i tuoi 9 e 17 anni... Chi ti ha influenzato in modo determinante nella scelta del tuo sport significativo?

☐ Genitori ☐ Fratelli/ sorelle ☐ Amici ☐ Qualcun altro _____

☐ Nessuno

83. Almeno uno dei tuoi genitori ha praticato nella sua vita il tuo sport significativo?

☐ Sì, a livello agonistico. Scrivi chi: _____

☐ Sì, per svago. Scrivi chi: _____

☐ No. Segna con una X _____

84. Tra i tuoi 9 e 17 anni...I tuoi genitori (o uno di essi) in media ti hanno portato la maggior parte delle volte (es. 2 volte su 3) agli allenamenti del tuo sport significativo

☐ Sì ☐ No

85. Tra i tuoi 9 e 17 anni...I tuoi genitori (o uno di essi) ti hanno sostenuto nel continuare a fare lo sport per te più significativo

☐ Sì ☐ No

86. Tra i tuoi 9 e 17 anni...Quando non avevi voglia di andare ad allenarti, i tuoi genitori (o uno di essi) ti convincevano o obbligavano ad andare

☐ Sì ☐ No

87. Pensa a quale dei tuoi genitori ti ha seguito di più nel tuo sport significativo e rispondi: il più influente tra i tuoi genitori, rispetto al tuo sport, lo definiresti...

☐ Tranquillo ☐ Troppo fissato ☐ Si arrabbiava alle gare ☐ Sostenitore positivo

☐ Incoraggiante ☐ Indifferente ☐ Pesante ☐ Ansiogeno

☐ Fastidioso ☐ Altro _____

SPORT E DIVERTIMENTO

88. Tra i tuoi 9 e 17 anni ... In media, ti divertivi nel praticare il tuo sport significativo...

☐ Molto ☐ Abbastanza ☐ Neutrale ☐ Poco ☐ Per niente

89. Tra i 9 e i 17 anni...il tuo sport per te è stato... **(scegline al massimo 3)**

☐ Un sacrificio

☐ Un divertimento

☐ Un dovere

☐ Una passione

☐ Un bisogno del tuo corpo e della tua mente

- ☐ Un modo per essere liberi
- ☐ Un modo ricaricarmi
- ☐ Un modo per staccare dalla vita quotidiana
- ☐ Un modo per scaricare l'energia
- ☐ Una perdita di tempo
- ☐ Altro (specificare) _____

90. Per favore, leggi le seguenti caratteristiche di personalità e indica quanto ti senti descritto da ogni coppia, anche se pensi che una delle due caratteristiche ti descriva più dell'altra.

Ti ritieni una persona...

		Completamente in disaccordo		Molto in disaccordo	Un po' in disaccordo	Né d'accordo né in disaccordo	Un po' d'accordo	Molto d'accordo	Completamente d'accordo			
		1	2	3	4	5	6	7				
1.	Estroversa, esuberante					1	2	3	4	5	6	7
2.	Polemica, litigiosa					1	2	3	4	5	6	7
3.	Affidabile, auto-disciplinata					1	2	3	4	5	6	7
4.	Ansiosa, che si agita facilmente					1	2	3	4	5	6	7
5.	Aperta alle nuove esperienze, con molti interessi					1	2	3	4	5	6	7
6.	Riservata, silenziosa					1	2	3	4	5	6	7
7.	Comprensiva, affettuosa					1	2	3	4	5	6	7
8.	Disorganizzata, distratta					1	2	3	4	5	6	7
9.	Competitiva					1	2	3	4	5	6	7
10.	Tranquilla, emotivamente stabile					1	2	3	4	5	6	7
11.	Tradizionalista, abitudinaria					1	2	3	4	5	6	7

91. Ti è piaciuto questo test?

- ☐ Sì
- ☐ No

QUESTIONARIO CASO 2

DATI ANAGRAFICI

Uguale a caso 1

DATI GENERICI

Uguale a caso 1

INFORMAZIONI SULLO SPORT

Uguale a caso 1 fino a 25

25. Crocetta la risposta che ti rappresenta.

“Lo sport più significativo per me...”

- ☐ ...È l'unico che ho fatto tra i 9 e i 17 anni compresi” (caso 1)
- ☐ ...È l'unico che ho fatto tra i 9 e 17 anni ma l'ho abbandonato prima dei 17 anni” (caso 3)
- ☐ ...è il più importante tra gli sport che ho praticato e a 17 anni lo praticavo ancora” (caso 2)
- ☐ ... è il più importante tra gli sport che ho praticato e l'ho abbandonato prima dei 17 anni” (caso 3)

CASO 2. Se ha fatto più di uno sport tra i 9 e 17 anni, e quello più a 17 anni lo praticava ancora (quindi o lo continua a fare o lo ha abbandonato dopo i 17 anni)

26. Riguardo al tuo sport più significativo, lo hai fatto a livello agonistico?

- ☐ Sì
- ☐ No

27. Quante volte alla settimana, in media, praticavi il tuo sport significativo prima dei 17 anni? Considera anche le partite (o le gare o i saggi).

- ☐ 1 volta
- ☐ 2 volte
- ☐ 3 volte
- ☐ Più di tre

28. Prima dei 17 anni...Hai mai pensato seriamente di abbandonare lo sport per te più significativo?

NB. Per “seriamente” si intende che stavi per lasciarlo o che l'hai lasciato per poco tempo (massimo 6 mesi) e poi hai ripreso.

- ☐ Sì (continua)
- ☐ No (salta alla domanda 70)

MOMENTO CRITICO

29. A quanti anni hai pensato seriamente di abbandonare lo sport per te più significativo?

- | | |
|-----------------------------|-----------------------------|
| <input type="checkbox"/> 9 | <input type="checkbox"/> 13 |
| <input type="checkbox"/> 10 | <input type="checkbox"/> 14 |
| <input type="checkbox"/> 11 | <input type="checkbox"/> 15 |

- | | |
|-----------------------------|-----------------------------|
| <input type="checkbox"/> 12 | <input type="checkbox"/> 16 |
| <input type="checkbox"/> 13 | <input type="checkbox"/> 17 |
| <input type="checkbox"/> 14 | |
| <input type="checkbox"/> 15 | |

30. Quando hai pensato di abbandonare lo sport per te più significativo come ti sentivi prima di andare ad allenarti?

- ☐ Stanco
☐ Felice
☐ Carico
☐ Svogliato
☐ Demotivato
☐ Ansioso
☐ Altro _____

31. Quando hai pensato di abbandonare lo sport per te più significativo, con che frequenza andavi ad allenarti?

- ☐ Sempre (3 allenamenti su 3)
☐ La maggior parte delle volte (2 allenamenti su 3)
☐ Raramente (1 allenamento su 3)
☐ Mai

32. Nell'ultimo mese, prima di abbandonare lo sport, quante ore alla settimana dedicavi ad attività extra-scolastiche e/o extra-sportive (es. arte, social network ecc.)? **SBAGLIATA!**

- ☐ Nessuna ☐ 1 ora ☐ 2 ore ☐ 3 ore ☐ 4 ore ☐ Più di 4

33. Quando hai pensato di abbandonare lo sport per te più significativo, quali fattori ti hanno spinto a continuare? Tra le seguenti motivazioni, **sceglina al massimo 3.**

Questioni tecniche
e personali

- | | |
|--------------------------|---|
| <input type="checkbox"/> | 1. Mi divertivo |
| <input type="checkbox"/> | 2. Mi sentivo bravo |
| <input type="checkbox"/> | 3. Mi piaceva la competizione |
| <input type="checkbox"/> | 4. Non mi sono mai infortunato/a |
| <input type="checkbox"/> | 5. Mi piaceva faticare |
| <input type="checkbox"/> | 6. Mi piaceva migliorare le mie abilità |
| <input type="checkbox"/> | 7. Volevo essere più bravo e/o forte degli altri nel mio sport |
| <input type="checkbox"/> | 8. Volevo tenermi in forma |
| <input type="checkbox"/> | 9. Avevo un obiettivo (es. vincere trofei o raggiungere alti livelli) |
| <input type="checkbox"/> | 10. Mi piaceva provare il brivido della vittoria |
| <input type="checkbox"/> | 11. Mi piaceva battere l'avversario |
| <input type="checkbox"/> | 12. Mi serviva a scaricare l'energia o la tensione |
| <input type="checkbox"/> | 13. Volevo guadagnare denaro e ottenere contratti |
| <input type="checkbox"/> | 14. Mi piacevo quando facevo sport |
| <input type="checkbox"/> | 15. Motivi religiosi |
| <input type="checkbox"/> | 16. Sentivo l'esigenza di fare sport |
| <input type="checkbox"/> | 17. Era una passione |
| <input type="checkbox"/> | 18. Era uno sport alla moda |

Rapporti interpersonali:

- | | |
|--------------------------|---|
| <input type="checkbox"/> | 19. Ero appoggiato dai miei genitori |
| <input type="checkbox"/> | 20. Per far contenti i miei genitori (famiglia) |
| <input type="checkbox"/> | 21. I miei genitori volevano che io facessi sport |
| <input type="checkbox"/> | 22. Per far contento il mio allenatore |
| <input type="checkbox"/> | 23. Bel rapporto tra i miei genitori e l'allenatore |
| <input type="checkbox"/> | 24. Mi trovavo bene con i miei compagni di squadra |

- ☐ 25. Avevo qualcuno con cui allenarmi
- ☐ 26. Avevo un allenatore che mi piaceva

Contesto socio-sportivo:

- ☐ 27. Riuscivo a conciliare lo sport con lo studio/lavoro
- ☐ 28. Ho sempre trovato il tempo di fare sport
- ☐ 29. L'impianto era facile da raggiungere
- ☐ 30. L'impianto era vicino casa
- ☐ 31. Lo sport non era troppo costoso
- ☐ 32. Palestra adeguata
- ☐ 33. Orari comodi
- ☐ 34. Altro _____

34. Tra le motivazioni che hai scelto scrivi il numero di quella che ha maggiormente determinato la tua voglia di continuare

35. Quali i fattori, invece, ti spingevano ad abbandonare? Tra le seguenti motivazioni, scegline al massimo 3.

Questioni tecniche

- ☐ 1. non mi divertivo più
- ☐ 2. non mi sentivo abbastanza bravo
- ☐ 3. troppa competizione
- ☐ 4. infortunio/i sportivo/i
- ☐ 5. eccessiva fatica
- ☐ 6. incomprensioni con l'allenatore
- ☐ 7. motivi di salute
- ☐ 8. periodo critico nella vita personale
- ☐ 9. sport troppo impegnativo
- ☐ 10. Disabilità temporanea o permanente

Rapporti interpersonali:

- ☐ 11. non ero appoggiato dai miei genitori
- ☐ 12. i miei genitori non mi potevano accompagnare
- ☐ 13. rapporto difficile tra genitori e allenatore
- ☐ 14. non mi trovavo bene con i compagni
- ☐ 15. non avevo qualcuno con cui allenarmi
- ☐ 16. Sport troppo lontano

Contesto sociale:

- ☐ 17. non andavo bene a scuola
- ☐ 18. non riuscivo a conciliare lo sport con l'impegno scolastico
- ☐ 19. avevo altri interessi prevalenti

Contesto sportivo:

- ☐ 20. difficoltà a raggiungere l'impianto
- ☐ 21. lo sport era troppo costoso
- ☐ 22. palestra non adeguata
- ☐ 23. materiali a disposizione in palestra carenti
- ☐ 24. orari scomodi
- ☐ 25. non ci sono sport che mi piacciono dove vivo io
- ☐ 26. altro _____

36. Tra le motivazioni che hai scelto scrivi il numero di quella che maggiormente ti spingeva ad abbandonare

IMPEGNO SCOLASTICO

37. Quando hai pensato di abbandonare lo sport per te più significativo ... Frequentavi...

- ☐ Scuola primaria (elementare)

- ☐ Scuola media inferiore
- ☐ Scuola media superiore

38. Quando hai pensato di abbandonare lo sport per te più significativo ... Frequentavi la scuola...

- ☐ Solo al mattino
- ☐ Mattino e pomeriggio
- ☐ Tutte le mattine e qualche pomeriggio

39. Quando hai pensato di abbandonare lo sport per te più significativo ... Riuscivi a conciliare l'impegno scolastico con l'impegno sportivo

(1= completamente in disaccordo; 2= abbastanza in disaccordo; 3= né in disaccordo né in accordo; 4= abbastanza d'accordo; 5=completamente d'accordo)

1	2	3	4	5
---	---	---	---	---

40. Quando hai pensato di abbandonare lo sport per te più significativo ...Quante ore studiavi in media al giorno al di fuori dell'orario scolastico?

- ☐ Nessuna ☐ 1 ora ☐ 2 ore ☐ 3 ore ☐ 4 ore ☐ Più di 4

SOCIETÀ E ALLENATORE

41. Quando hai pensato di abbandonare lo sport per te più significativo ... Ti sentivi parte integrante della società sportiva a cui appartenevi?

- ☐ sì ☐ No

42. Quando hai pensato di abbandonare lo sport per te più significativo ... Il tuo allenatore ha influito sulla tua scelta di continuarlo?

Sì. Scrivi il perché _____

No. Segna con una X _____

43. Quando hai pensato di abbandonare lo sport per te più significativo ... il tuo allenatore ti dava le giuste attenzioni e importanza (es. ti seguiva negli esercizi quando avevi bisogno)?

- ☐ Sì ☐ No

44. Quando hai pensato di abbandonare lo sport per te più significativo ... Il tuo allenatore ti incoraggiava quando perdevi o sbagliavi?

- ☐ Sì ☐ No

45. Quando hai pensato di abbandonare lo sport per te più significativo ... Il tuo allenatore era (scegline al massimo 3)

- | | | |
|---|---|--|
| <input type="checkbox"/> Severo | <input type="checkbox"/> Autorevole | <input type="checkbox"/> Un amico o confidente |
| <input type="checkbox"/> Competente | <input type="checkbox"/> Un buon leader | <input type="checkbox"/> Troppo fissato |
| <input type="checkbox"/> Un punto di riferimento al di fuori della palestra | <input type="checkbox"/> Non in grado di allenare | |
| <input type="checkbox"/> Avvilente | <input type="checkbox"/> Irritabile | <input type="checkbox"/> Un motivatore |
| <input type="checkbox"/> Non era sufficientemente preparato tecnicamente | <input type="checkbox"/> Altro _____ | |

COMPAGNI DI SQUADRA

NB: per "compagni di squadra" si intendono le persone con cui hai praticato il tuo sport (sia esso individuale o di squadra).

Es. le persone con cui hai fatto lezione di danza, o lezione di tennis, o con cui ti sei allenato a basket ecc.

46. Quando hai pensato di abbandonare lo sport per te più significativo ... Avevi compagni di squadra che ti stimolavano a impegnarti in allenamento e/o durante le gare/partite?

☐ Sì ☐ No

47. Quando hai pensato di abbandonare lo sport per te più significativo ... I tuoi compagni di squadra ti hanno influenzato sulla scelta di continuarlo?

Sì. Scrivi perché _____

No. Segna con una X _____

48. Quando hai pensato di abbandonare lo sport per te più significativo ... I tuoi compagni di squadra ogni tanto non si comportavano bene con te (ti prendevano in giro, non ti calcolavano ecc.)

☐ Sì ☐ No

GENITORI

49. Quando hai pensato di abbandonare lo sport per te più significativo ...Tua mamma lavorava?

☐ Sì

☐ No

50. Quando hai pensato di abbandonare lo sport per te più significativo ... Che professione svolgeva tua mamma?

☐ Libera professione (avvocato, medico, commercialista, artista, carriera sportiva ecc.)

☐ Commerciante/Imprenditrice/Artigiana

☐ Impiegata

☐ Operaia

☐ Casalinga

☐ Docente

☐ Altro _____

51. Quando hai pensato di abbandonare lo sport per te più significativo ...Tua mamma era impegnata per lavoro...

☐ Full time

☐ Part time

☐ Si gestiva autonomamente

☐ Altro _____

52. Quando hai pensato di abbandonare lo sport per te più significativo ...Tuo papà lavorava?

☐ Sì

☐ No

53. Quando hai pensato di abbandonare lo sport per te più significativo ...Che professione svolgeva tuo papà?

☐ Libera professione (avvocato, medico, commercialista, artista, carriera sportiva ecc.)

☐ Commerciante/Imprenditore/Artigiano

- ☐ Impiegato
- ☐ Operaio
- ☐ Casalingo
- ☐ Docente
- ☐ Altro _____

54. Quando hai pensato di abbandonare lo sport per te più significativo ...Tuo papà era impegnato per lavoro...

- ☐ Full time
- ☐ Part time
- ☐ Si gestiva autonomamente
- ☐ Altro _____

55. Quanti dei tuoi fratelli hanno fatto il tuo stesso sport significativo?

- ☐ Nessuno
- ☐ 1
- ☐ 2
- ☐ 3
- ☐ Più di tre

56. Chi ti ha influenzato in modo determinante nella scelta dello tuo sport significativo?

- ☐ Genitori ☐ fratelli/sorelle ☐ Amici ☐ Qualcun altro _____
- ☐ Nessuno

57. Almeno uno dei tuoi genitori ha praticato nella sua vita il tuo sport significativo (puoi rispondere a più di un'opzione)?

Sì, a livello agonistico. Scrivi chi _____

Sì, per svago. Scrivi chi _____

No. Segna con una X _____

Non lo so. Segna con una X _____

58. Quando hai pensato di abbandonare lo sport per te più significativo ... I tuoi genitori (o uno di essi) ti portavano la maggior parte delle volte (es. 2 volte su 3) agli allenamenti del tuo sport significativo?

- ☐ Sì ☐ No

59. Quando hai pensato di abbandonare lo sport per te più significativo ... I tuoi genitori (o uno di essi) ti hanno spinto a continuare a fare proprio quello sport?

- ☐ Sì ☐ No

60. Tra i tuoi 9 e 17 anni... Quando non avevi voglia di andare ad allenarti nel tuo sport significativo, i tuoi genitori (o uno di essi) ti convincevano o obbligavano ad andare?

- ☐ Sì ☐ No

61. Pensa a quale dei tuoi genitori ti ha seguito di più nello sport per te più significativo e rispondi:

Lui/lei, rispetto al tuo sport, lo definiresti...

- ☐ Tranquillo ☐ Troppo fissato ☐ Si arrabbiava alle gare ☐ Sostenitore positivo
- ☐ Incoraggiante ☐ Indifferente ☐ Pesante ☐ Ansiogeno
- ☐ Fastidioso ☐ Altro _____

AUTOEFFICACIA

62. Torna al momento in cui hai pensato di abbandonare seriamente lo sport per te più significativo e rispondi.

1. Completamente in disaccordo
2. Abbastanza in disaccordo
3. Né in accordo né in disaccordo
4. Abbastanza in accordo
5. Completamente in disaccordo

“Mi sono sentito capace di affrontare l’allenamento...

...anche quando avevo preoccupazioni e problemi personali	1	2	3	4	5
...anche quando ero stanco	1	2	3	4	5
...anche quando ero indaffarato	1	2	3	4	5

63. Quando hai pensato di abbandonare lo sport per te più significativo ...Partecipavi a gare/partite?

- ☐ Sì (continua)
- ☐ No (salta alla domanda 66)

GESTIONE DEL MATCH

NB. Per chi ha praticato danza: per "gestione del match" si intende anche "gestione del saggio".

Quindi, nelle due domande successive traduci "perdere una gara" con "quando la tua esibizione non veniva bene" e "vincere una partita" con "quando la tua esibizione veniva bene"

64. Perdere una partita (o una gara) ...

- ☐ Era indifferente
- ☐ Era poco rilevante
- ☐ Era un dramma
- ☐ Era uno stimolo per far meglio
- ☐ Ti faceva venir voglia di smettere
- ☐ Altro _____

65. Vincere una partita (o una gara) ...

- ☐ Era indifferente
- ☐ Era poco rilevante
- ☐ Era una gioia
- ☐ Era uno stimolo per far meglio
- ☐ Ti faceva venir voglia di continuare
- ☐ Altro _____

SPORT E DIVERTIMENTO

66. Quando hai pensato di abbandonare lo sport per te più significativo, ti divertivi ad allenamento...

☐ Molto ☐ Abbastanza ☐ Neutrale ☐ Poco ☐ Per niente

67. Quando hai pensato di abbandonare lo sport per te più significativo ... Il tuo sport era... **(scegline al massimo tre)**

- ☐ Un sacrificio
- ☐ Un divertimento
- ☐ Un dovere
- ☐ Una passione
- ☐ Un bisogno del tuo corpo e della tua mente
- ☐ Un modo per essere liberi
- ☐ Un modo ricaricarmi
- ☐ Un modo per staccare dalla vita quotidiana
- ☐ Un modo per scaricare l'energia
- ☐ Una perdita di tempo
- ☐ Altro _____

68. Per favore, leggi le seguenti caratteristiche di personalità e indica quanto ti senti descritto da ogni coppia, anche se pensi che una delle due caratteristiche ti descriva più dell'altra.

Ti ritieni una persona...

Completamente in disaccordo	Molto in disaccordo	Un po' in disaccordo	Né d'accordo né in disaccordo	Un po' d'accordo	Molto d'accordo	Completamente d'accordo
1	2	3	4	5	6	7

Sono una persona...

1.	Estroversa, esuberante	1	2	3	4	5	6	7
2.	Polemica, litigiosa	1	2	3	4	5	6	7
3.	Affidabile, auto-disciplinata	1	2	3	4	5	6	7
4.	Ansiosa, che si agita facilmente	1	2	3	4	5	6	7
5.	Aperta alle nuove esperienze, con molti interessi	1	2	3	4	5	6	7
6.	Riservata, silenziosa	1	2	3	4	5	6	7
7.	Comprensiva, affettuosa	1	2	3	4	5	6	7
8.	Disorganizzata, distratta	1	2	3	4	5	6	7
9.	Competitiva	1	2	3	4	5	6	7
10.	Tranquilla, emotivamente stabile	1	2	3	4	5	6	7
11.	Tradizionalista, abitudinaria	1	2	3	4	5	6	7

69. Ti è piaciuto questo test?

- ☐ Sì
☐ No

INFO SULLO SPORT (caso 2, che non ha abbandonato il suo sport né ci ha pensato)

70. Perché non hai mai pensato di abbandonare il tuo sport? Tra le seguenti motivazioni, **sceglie al massimo 3**.

Questioni tecniche
e personali

- ☐ 1. Mi divertivo
- ☐ 2. Mi sentivo bravo
- ☐ 3. Mi piaceva la competizione
- ☐ 4. Non mi sono mai infortunato/a
- ☐ 5. Mi piaceva faticare
- ☐ 6. Mi piaceva migliorare le mie abilità
- ☐ 7. Volevo essere più bravo e/o forte degli altri nel mio sport
- ☐ 8. Volevo tenermi in forma
- ☐ 9. Avevo un obiettivo (es. vincere trofei o raggiungere alti livelli)
- ☐ 10. Mi piaceva provare il brivido della vittoria
- ☐ 11. Mi piaceva battere l'avversario
- ☐ 12. Mi serviva a scaricare l'energia o la tensione
- ☐ 13. Volevo guadagnare denaro e ottenere contratti
- ☐ 14. Mi piacevo quando facevo sport
- ☐ 15. Motivi religiosi
- ☐ 16. Sentivo l'esigenza di fare sport
- ☐ 17. Era una passione
- ☐ 18. Era uno sport alla moda

Rapporti interpersonali:

- ☐ 19. Ero appoggiato dai miei genitori
- ☐ 20. Per far contenti i miei genitori (famiglia)
- ☐ 21. I miei genitori volevano che io facessi sport
- ☐ 22. Per far contento il mio allenatore
- ☐ 23. Bel rapporto tra i miei genitori e l'allenatore
- ☐ 24. Mi trovavo bene con i miei compagni di squadra
- ☐ 25. Avevo qualcuno con cui allenarmi
- ☐ 26. Avevo un allenatore che mi piaceva

Contesto socio-sportivo:

- ☐ 27. Riuscivo a conciliare lo sport con lo studio/lavoro
- ☐ 28. Ho sempre trovato il tempo di fare sport
- ☐ 29. L'impianto era facile da raggiungere
- ☐ 30. L'impianto era vicino casa
- ☐ 31. Lo sport non era troppo costoso
- ☐ 32. Palestra adeguata
- ☐ 33. Orari comodi
- ☐ 34. Altro _____

71. Tra le motivazioni che hai scelto, scrivi il numero di quella che ha maggiormente determinato la tua voglia di continuare

72. In media, tra i tuoi 9 e 17 anni, quante ore alla settimana dedicavi ad attività extra-scolastiche e/o extra-sportive (es. arte, social network ecc.)?

- ☐ Nessuna ☐ 1 ora ☐ 2 ore ☐ 3 ore ☐ 4 ore ☐ Più di 4

IMPEGNO SCOLASTICO

73. Tra i tuoi 9 e i 17 anni ... Sei in media sempre riuscito a conciliare l'impegno scolastico con l'impegno sportivo?
(1= completamente in disaccordo; 2= abbastanza in disaccordo; 3= né in disaccordo né in accordo; 4= abbastanza d'accordo; 5=completamente d'accordo)

1	2	3	4	5
---	---	---	---	---

74. Tra i 9 e i 17 anni...Quante ore studiavi in media al giorno al di fuori dell'orario scolastico?

☐ Nessuna ☐ 1 ora ☐ 2 ore ☐ 3 ore ☐ 4 ore ☐ Più di 4

GENITORI

75. Tra i tuoi 9 e i 17 anni... Tua mamma lavorava?

☐ Sì

☐ No

76. Tra i tuoi 9 e i 17 anni... Che professione svolgeva tua mamma?

☐ Libera professione (avvocato, medico, commercialista, artista, carriera sportiva ecc.)

☐ Commerciante/Imprenditrice/Artigiana

☐ Impiegata

☐ Operaia

☐ Casalinga

☐ Docente

☐ Altro _____

77. Tra i tuoi 9 e i 17 anni... Tua mamma era impegnata per lavoro...

☐ Full time

☐ Part time

☐ Si gestiva autonomamente

☐ Altro _____

78. Tra i tuoi 9 e i 17 anni... Tuo papà lavorava?

☐ Sì

☐ No

79. Tra i tuoi 9 e i 17 anni... Che professione svolgeva tuo papà?

☐ Libera professione (avvocato, medico, commercialista, artista, carriera sportiva ecc.)

☐ Commerciante/Imprenditore/Artigiano

☐ Impiegato

☐ Operaio

☐ Casalingo

☐ Docente

☐ Altro _____

80. Tra i tuoi 9 e i 17 anni... Tuo papà era impegnato per lavoro...

☐ Full time

☐ Part time

☐ Si gestiva autonomamente

☐ Altro _____

81. Tra i tuoi 9 e i 17 anni... Quanti dei tuoi fratelli hanno fatto il tuo stesso sport significativo?

☐ Nessuno

☐ 1

☐ 2

☐ 3

☐ Più di tre

82. Tra i tuoi 9 e i 17 anni... Chi ti ha influenzato in modo determinante nella scelta del tuo sport significativo?

☐ Genitori ☐ Fratelli/sorelle ☐ Amici ☐ Qualcun altro _____

☐ Nessuno

83. Almeno uno dei tuoi genitori ha praticato nella sua vita il tuo sport significativo?

Sì, a livello agonistico. Scrivi chi _____

Sì, per svago. Scrivi chi _____

No. Segna con una X _____

Non lo so. Segna con una X _____

84. Prima dei 17 anni... I tuoi genitori (o uno di essi) in media ti hanno portato la maggior parte delle volte (es. 2 volte su 3) agli allenamenti del tuo sport significativo?

☐ Sì ☐ No

85. Tra i tuoi 9 e i 17 anni... I tuoi genitori (o uno di essi) ti hanno sostenuto nel continuare a fare lo sport per te più significativo?

☐ Sì ☐ No

86. Prima dei 17 anni... Quando non avevi voglia di andare ad allenarti (nel tuo sport significativo), i tuoi genitori (o uno di essi) ti convincevano o obbligavano ad andare?

☐ Sì ☐ No

87. Pensa a quale dei tuoi genitori ti ha seguito di più nel tuo sport significativo e rispondi:

Lui/lei, rispetto al tuo sport, lo definiresti...

☐ Tranquillo ☐ Troppo fissato ☐ Si arrabbiava alle gare ☐ Sostenitore positivo

☐ Incoraggiante ☐ Indifferente ☐ Pesante ☐ Ansiogeno

☐ Fastidioso ☐ Altro _____

SPORT E DIVERTIMENTO

88. Prima dei 17 anni ... In media, ti divertivi nel praticare il tuo sport significativo...

☐ Molto ☐ Abbastanza ☐ Neutrale ☐ Poco ☐ Per niente

89. Il tuo sport significativo, fino ai 17 anni, per te è stato... (scegline al massimo tre)

☐ Un sacrificio

- ☐ Un divertimento
- ☐ Un dovere
- ☐ Una passione
- ☐ Un bisogno del tuo corpo e della tua mente
- ☐ Un modo per essere liberi
- ☐ Un modo ricaricarmi
- ☐ Un modo per staccare dalla vita quotidiana
- ☐ Un modo per scaricare l'energia
- ☐ Una perdita di tempo
- ☐ Altro _____

90. Per favore, leggi le seguenti caratteristiche di personalità e indica quanto ti senti descritto da ogni coppia, anche se pensi che una delle due caratteristiche ti descriva più dell'altra.

Ti ritieni una persona...

				Né							
Completamente in disaccordo		Molto in disaccordo	Un po' in disaccordo	d'accordo né in disaccordo	Un po' d'accordo	Molto d'accordo	Completamente d'accordo				
1		2	3	4	5	6	7				
1.	Estroversa, esuberante				1	2	3	4	5	6	7
2.	Polemica, litigiosa				1	2	3	4	5	6	7
3.	Affidabile, auto-disciplinata				1	2	3	4	5	6	7
4.	Ansiosa, che si agita facilmente				1	2	3	4	5	6	7
5.	Aperta alle nuove esperienze, con molti interessi				1	2	3	4	5	6	7
6.	Riservata, silenziosa				1	2	3	4	5	6	7
7.	Comprensiva, affettuosa				1	2	3	4	5	6	7
8.	Disorganizzata, distratta				1	2	3	4	5	6	7
9.	Competitiva				1	2	3	4	5	6	7
10.	Tranquilla, emotivamente stabile				1	2	3	4	5	6	7
11.	Tradizionalista, abitudinaria				1	2	3	4	5	6	7

91. Ti è piaciuto questo test?

- ☐ Sì
- ☐ No

QUESTIONARIO CASO 3

DATI ANAGRAFICI

Uguale a caso 1

DATI GENERICI

Uguale a caso 1

INFORMAZIONI SULLO SPORT

Uguale a caso 1 fino a 25

25. Crocetta la risposta che ti rappresenta.

“Lo sport più significativo per me...

- ☐ ...È l'unico che ho fatto tra i 9 e i 17 anni compresi” (caso 1)
- ☐ ...È l'unico che ho fatto tra i 9 e 17 anni ma l'ho abbandonato prima dei 17 anni” (caso 3)
- ☐ ...è il più importante tra gli sport che ho praticato e a 17 anni lo praticavo ancora” (caso 2)
- ☐ ... è il più importante tra gli sport che ho praticato e l'ho abbandonato prima dei 17 anni” (caso 3)

CASO 3. Se ha abbandonato quello che per lui è stato il suo sport significativo

26. Riguardo al tuo sport più significativo, lo hai fatto a livello agonistico?

- ☐ Sì
- ☐ No

27. Quante volte alla settimana, in media, praticavi il tuo sport significativo tra i 9 e i 17 anni? Considera anche le partite (o le gare o i saggi).

- ☐ 1 volta
- ☐ 2 volte
- ☐ 3 volte
- ☐ Più di tre

28. A quanti anni hai abbandonato lo sport per te più significativo?

- | | |
|-----------------------------|-----------------------------|
| <input type="checkbox"/> 9 | <input type="checkbox"/> 13 |
| <input type="checkbox"/> 10 | <input type="checkbox"/> 14 |
| <input type="checkbox"/> 11 | <input type="checkbox"/> 15 |
| <input type="checkbox"/> 12 | <input type="checkbox"/> 16 |
| <input type="checkbox"/> 13 | <input type="checkbox"/> 17 |
| <input type="checkbox"/> 14 | |
| <input type="checkbox"/> 15 | |

29. Nell'ultimo mese, prima di abbandonare il tuo sport, come ti sei sentito prima di andare ad allenarti?

- ☐ Stanco
- ☐ Felice

- ☐ Carico
- ☐ Svogliato
- ☐ Demotivato
- ☐ Ansioso
- ☐ Altro _____

30. Nei mesi prima di abbandonare lo sport per te più significativo, con che frequenza andavi ad allenarti?

- ☐ Sempre (3 allenamenti su 3)
- ☐ La maggior parte delle volte (2 allenamenti su 3)
- ☐ Raramente (1 allenamento su 3)
- ☐ Mai

31. Nell'ultimo mese, prima di abbandonare il tuo sport, quante ore alla settimana dedicavi ad attività extra-scolastiche e/o extra-sportive (es. arte, social network ecc.)?

- ☐ Nessuna ☐ 1 ora ☐ 2 ore ☐ 3 ore ☐ 4 ore ☐ Più di 4

32. Perché hai smesso di praticare il tuo sport significativo? Tra le seguenti motivazioni, **sceglie al massimo 3**.

- | | | |
|--------------------------|--------------------------|---|
| Questioni tecniche | <input type="checkbox"/> | 1. non mi divertivo più |
| | <input type="checkbox"/> | 2. non mi sentivo abbastanza bravo |
| | <input type="checkbox"/> | 3. troppa competizione |
| | <input type="checkbox"/> | 4. infortunio/i sportivo/i |
| | <input type="checkbox"/> | 5. eccessiva fatica |
| | <input type="checkbox"/> | 6. incomprensioni con l'allenatore |
| | <input type="checkbox"/> | 7. motivi di salute |
| | <input type="checkbox"/> | 8. periodo critico nella vita personale |
| | <input type="checkbox"/> | 9. sport troppo impegnativo |
| | <input type="checkbox"/> | 10. Disabilità temporanea o permanente |
| Rapporti interpersonali: | <input type="checkbox"/> | 11. non ero appoggiato dai miei genitori |
| | <input type="checkbox"/> | 12. i miei genitori non mi potevano accompagnare |
| | <input type="checkbox"/> | 13. rapporto difficile tra genitori e allenatore |
| | <input type="checkbox"/> | 14. non mi trovavo bene con i compagni |
| | <input type="checkbox"/> | 15. non avevo qualcuno con cui allenarmi |
| | <input type="checkbox"/> | 16. Sport troppo lontano |
| Contesto sociale: | <input type="checkbox"/> | 17. non andavo bene a scuola |
| | <input type="checkbox"/> | 18. non riuscivo a conciliare lo sport con l'impegno scolastico |
| | <input type="checkbox"/> | 19. avevo altri interessi prevalenti |
| Contesto sportivo: | <input type="checkbox"/> | 20. difficoltà a raggiungere l'impianto |
| | <input type="checkbox"/> | 21. lo sport era troppo costoso |
| | <input type="checkbox"/> | 22. palestra non adeguata |
| | <input type="checkbox"/> | 23. materiali a disposizione in palestra carenti |
| | <input type="checkbox"/> | 24. orari scomodi |
| | <input type="checkbox"/> | 25. non ci sono sport che mi piacciono dove vivo io |
| | <input type="checkbox"/> | 26. altro _____ |

33. Tra le motivazioni che hai scelto scrivi il numero di quella che ha maggiormente determinato il tuo abbandono

34. Prima di abbandonare, cosa ti spingeva invece a praticare? Tra le seguenti motivazioni, sceglie al massimo 3.

- Questioni tecniche e personali
- ☐ 1. Mi divertivo
 - ☐ 2. Mi sentivo bravo
 - ☐ 3. Mi piaceva la competizione
 - ☐ 4. Non mi sono mai infortunato/a
 - ☐ 5. Mi piaceva faticare
 - ☐ 6. Mi piaceva migliorare le mie abilità
 - ☐ 7. Volevo essere più bravo e/o forte degli altri nel mio sport
 - ☐ 8. Volevo tenermi in forma
 - ☐ 9. Avevo un obiettivo (es. vincere trofei o raggiungere alti livelli)
 - ☐ 10. Mi piaceva provare il brivido della vittoria
 - ☐ 11. Mi piaceva battere l'avversario
 - ☐ 12. Mi serviva a scaricare l'energia o la tensione
 - ☐ 13. Volevo guadagnare denaro e ottenere contratti
 - ☐ 14. Mi piacevo quando facevo sport
 - ☐ 15. Motivi religiosi
 - ☐ 16. Sentivo l'esigenza di fare sport
 - ☐ 17. Era una passione
 - ☐ 18. Era uno sport alla moda

- Rapporti interpersonali:
- ☐ 19. Ero appoggiato dai miei genitori
 - ☐ 20. Per far contenti i miei genitori (famiglia)
 - ☐ 21. I miei genitori volevano che io facessi sport
 - ☐ 22. Per far contento il mio allenatore
 - ☐ 23. Bel rapporto tra i miei genitori e l'allenatore
 - ☐ 24. Mi trovavo bene con i miei compagni di squadra
 - ☐ 25. Avevo qualcuno con cui allenarmi
 - ☐ 26. Avevo un allenatore che mi piaceva

- Contesto socio-sportivo:
- ☐ 27. Riuscivo a conciliare lo sport con lo studio/lavoro
 - ☐ 28. Ho sempre trovato il tempo di fare sport
 - ☐ 29. L'impianto era facile da raggiungere
 - ☐ 30. L'impianto era vicino casa
 - ☐ 31. Lo sport non era troppo costoso
 - ☐ 32. Palestra adeguata
 - ☐ 33. Orari comodi
 - ☐ 34. Altro _____

35. Tra le motivazioni che hai scelto scrivi il numero di quella che ha maggiormente determinato la tua voglia di continuare

IMPEGNO SCOLASTICO

36. Gli ultimi mesi prima di abbandonare lo sport per te più significativo ... Frequentavi...

- ☐ Scuola primaria (elementare)
- ☐ Scuola media inferiore
- ☐ Scuola media superiore

37. Gli ultimi mesi prima di abbandonare lo sport per te più significativo ... Frequentavi la scuola...

- ☐ Solo al mattino
- ☐ Mattino e pomeriggio
- ☐ Tutte le mattine e qualche pomeriggio

38. Gli ultimi mesi prima di abbandonare lo sport per te più significativo ... Riuscivi a conciliare l'impegno scolastico con l'impegno sportivo

(1= completamente in disaccordo; 2= abbastanza in disaccordo; 3= né in disaccordo né in accordo; 4= abbastanza d'accordo; 5=completamente d'accordo)

1	2	3	4	5
---	---	---	---	---

39. Gli ultimi mesi prima di abbandonare lo sport per te più significativo ... Quante ore studiavi in media al giorno al di fuori dell'orario scolastico?

☐ Nessuna ☐ 1 ora ☐ 2 ore ☐ 3 ore ☐ 4 ore ☐ Più di 4

SOCIETÀ E ALLENATORE

40. Gli ultimi mesi prima di abbandonare il tuo sport significativo ... Ti sentivi parte integrante della società sportiva a cui appartenevi?

☐ Sì ☐ No

41. Gli ultimi mesi prima di abbandonare il tuo sport significativo ... Il tuo allenatore ha influito sulla tua scelta di abbandonare lo sport che praticavi?

Sì. Scrivi il perché _____

No. Segna con una X _____

42. Gli ultimi mesi prima di abbandonare il tuo sport significativo ... Il tuo allenatore ti dava le giuste attenzioni e importanza (es. ti seguiva negli esercizi quando avevi bisogno)?

☐ Sì ☐ No

43. Gli ultimi mesi prima di abbandonare il tuo sport significativo ... Il tuo allenatore ti incoraggiava quando perdevi o sbagliavi?

☐ Sì ☐ No

44. Gli ultimi mesi prima di abbandonare il tuo sport significativo ... Il tuo allenatore era ... (scegline al massimo 3)

<input type="checkbox"/> Severo	<input type="checkbox"/> Autorevole	<input type="checkbox"/> Un amico o confidente
<input type="checkbox"/> Competente	<input type="checkbox"/> Un buon leader	<input type="checkbox"/> Troppo fissato
<input type="checkbox"/> Un punto di riferimento al di fuori della palestra	<input type="checkbox"/> Non in grado di allenare	
<input type="checkbox"/> Avvilente	<input type="checkbox"/> Irritabile	<input type="checkbox"/> Un motivatore
<input type="checkbox"/> Non era sufficientemente preparato tecnicamente	<input type="checkbox"/> Altro _____	

COMPAGNI DI SQUADRA

NB: per "compagni di squadra" si intendono le persone con cui hai praticato il tuo sport (sia esso individuale o di squadra).

Es. le persone con cui hai fatto lezione di danza, o lezione di tennis, o con cui ti sei allenato a basket ecc.

45. Gli ultimi mesi prima di abbandonare il tuo sport... Avevi compagni di squadra che ti stimolavano a impegnarti in allenamento e/o durante le gare/partite?

☐ Sì ☐ No

46. Gli ultimi mesi prima di abbandonare il tuo sport ... I tuoi compagni di squadra hanno influito sulla tua scelta di abbandonare lo sport che praticavi?

Si. Scrivi il perché _____

No. Segna con una X _____

47. Gli ultimi mesi prima di abbandonare il tuo sport significativo... I tuoi compagni di squadra ogni tanto non si comportavano bene con te (es. ti prendevano in giro, non ti calcolavano ecc.)

☐ Si ☐ No

GENITORI

48. Gli ultimi mesi prima di abbandonare il tuo sport significativo ... Tua mamma lavorava?

☐ Si

☐ No

49. Gli ultimi mesi prima di abbandonare il tuo sport significativo ... Che professione svolgeva tua mamma?

☐ Libera professione (avvocato, medico, commercialista, artista, carriera sportiva ecc.)

☐ Commerciante/Imprenditrice/Artigiana

☐ Impiegata

☐ Operaia

☐ Casalinga

☐ Docente

☐ Altro _____

50. Gli ultimi mesi prima di abbandonare il tuo sport significativo ... Tua mamma era impegnata per lavoro...

☐ Full time

☐ Part time

☐ Si gestiva autonomamente

☐ Altro _____

51. Gli ultimi mesi prima di abbandonare il tuo sport significativo ... Tuo papà lavorava?

☐ Si

☐ No

52. Gli ultimi mesi prima di abbandonare il tuo sport significativo ... Che professione svolgeva tuo papà?

☐ Libera professione (avvocato, medico, commercialista, artista, carriera sportiva ecc.)

☐ Commerciante/Imprenditore/Artigiano

☐ Impiegato

☐ Operaio

☐ Casalingo

☐ Docente

☐ Altro _____

53. Gli ultimi mesi prima di abbandonare il tuo sport significativo ... Tuo papà era impegnato per lavoro...

☐ Full time

☐ Part time

☐ Si gestiva autonomamente

☐ Altro _____

54. Quanti dei tuoi fratelli hanno fatto il tuo stesso sport (che poi hai abbandonato)?

- ☐ Nessuno
- ☐ 1
- ☐ 2
- ☐ 3
- ☐ Più di tre

55. Chi ti ha influenzato in modo determinante nella scelta dello sport (che poi hai abbandonato)?

- ☐ Genitori ☐ Fratelli/sorelle ☐ Amici ☐ Qualcun altro _____
- ☐ Nessuno

56. Almeno uno dei tuoi genitori ha praticato nella sua vita lo sport che hai abbandonato (puoi rispondere a più di un'opzione)

- ☐ Sì, a livello agonistico. Scrivi chi dei due _____
- ☐ Sì, per svago. Scrivi chi dei due _____
- ☐ No. Segna con una X _____
- ☐ Non lo so, segna con una X _____

57. Gli ultimi mesi prima di abbandonare il tuo sport significativo ... I tuoi genitori (o uno di essi) ti portavano la maggior parte delle volte (es. 2 volte su 3) agli allenamenti dello sport che poi hai abbandonato?

- ☐ Sì ☐ No

58. Quando hai deciso di abbandonare il tuo sport significativo...I tuoi genitori (o uno di essi) ti hanno spinto a continuare a fare comunque sport, qualunque esso fosse?

- ☐ Sì ☐ No

59. Quando non avevi voglia di andare ad allenarti nel tuo sport significativo, i tuoi genitori (o uno di essi) ti convincevano o obbligavano ad andare?

- ☐ Sì ☐ No

60. Pensa a quale dei tuoi genitori ti ha seguito di più nel tuo sport e rispondi: quando hai abbandonato il tuo sport, lui/lei era...

- ☐ Molto dispiaciuto ☐ Dispiaciuto ☐ Indifferente ☐ Contento ☐ Molto contento

61. Pensa a quale dei tuoi genitori ti ha seguito di più nello sport che hai abbandonato e rispondi:

Lui/lei, rispetto al tuo sport, lo definiresti...

- ☐ Tranquillo ☐ Troppo fissato ☐ Si arrabbiava alle gare ☐ Sostenitore positivo
- ☐ Incoraggiante ☐ Indifferente ☐ Pesante ☐ Ansiogeno
- ☐ Fastidioso ☐ Altro _____

AUTOEFFICACIA

62. Torna al momento in cui hai pensato di abbandonare seriamente lo sport per te più significativo e rispondi.

1. Completamente in disaccordo
2. Abbastanza in disaccordo
3. Né in accordo né in disaccordo

4. Abbastanza in accordo

5. Completamente in disaccordo

"Mi sono sentito capace di affrontare l'allenamento..."

...anche quando avevo preoccupazioni e problemi personali	1	2	3	4	5
...anche quando ero stanco	1	2	3	4	5
...anche quando ero indaffarato	1	2	3	4	5

63. Nei mesi prima di abbandonare il tuo sport più significativo ... Partecipavi a gare/partite/saggi?

☐ Sì (continua)

☐ No (salta alla domanda 66)

GESTIONE DEL MATCH

NB. Per chi ha praticato danza: per "gestione del match" si intende anche "gestione del saggio".

Quindi, nelle due domande successive traduci "perdere una gara" con "quando la tua esibizione non veniva bene" e "vincere una partita" con "quando la tua esibizione veniva bene"

64. Nei mesi prima di abbandonare il tuo sport più significativo ... Perdere una partita (o una gara) ...

☐ Era indifferente

☐ Era poco rilevante

☐ Era un dramma

☐ Era uno stimolo per far meglio

☐ Ti faceva venir voglia di smettere

☐ Altro _____

65. Nei mesi prima di abbandonare il tuo sport più significativo ... Vincere una partita (o una gara) ...

☐ Era indifferente

☐ Era poco rilevante

☐ Era una gioia

☐ Era uno stimolo per far meglio

☐ Ti faceva venir voglia di continuare

☐ Altro _____

SPORT E DIVERTIMENTO

66. Nei mesi prima di abbandonare il tuo sport... Ti divertivi ad allenamento...

☐ Molto ☐ Abbastanza ☐ Neutrale ☐ Poco ☐ Per niente

67. Nell'ultimo mese prima di abbandonare il tuo sport significativo ... Il tuo sport era... **(scegline al massimo 3)**

☐ Un sacrificio

☐ Un divertimento

- ☐ Un dovere
- ☐ Una passione
- ☐ Un bisogno del tuo corpo e della tua mente
- ☐ Un modo per essere liberi
- ☐ Un modo ricaricarmi
- ☐ Un modo per staccare dalla vita quotidiana
- ☐ Un modo per scaricare l'energia
- ☐ Una perdita di tempo
- ☐ Altro _____

68. Per favore, leggi le seguenti caratteristiche di personalità e indica quanto ti senti descritto da ogni coppia, anche se pensi che una delle due caratteristiche ti descriva più dell'altra.
Ti ritieni una persona...

Completamente in disaccordo	Molto in disaccordo	Un po' in disaccordo	Né d'accordo né in disaccordo	Un po' d'accordo	Molto d'accordo	Completamente d'accordo
1	2	3	4	5	6	7

Ti ritieni una persona...

1.	Estroversa, esuberante	1	2	3	4	5	6	7
2.	Polemica, litigiosa	1	2	3	4	5	6	7
3.	Affidabile, auto-disciplinata	1	2	3	4	5	6	7
4.	Ansiosa, che si agita facilmente	1	2	3	4	5	6	7
5.	Aperta alle nuove esperienze, con molti interessi	1	2	3	4	5	6	7
6.	Riservata, silenziosa	1	2	3	4	5	6	7
7.	Comprensiva, affettuosa	1	2	3	4	5	6	7
8.	Disorganizzata, distratta	1	2	3	4	5	6	7
9.	Competitiva	1	2	3	4	5	6	7
10.	Tranquilla, emotivamente stabile	1	2	3	4	5	6	7
11.	Tradizionalista, abitudinaria	1	2	3	4	5	6	7

69. Ti è piaciuto questo test?

Si/No

QUESTIONARIO CASO 4

DATI ANAGRAFICI

Uguale a caso 1

DATI FAMILIARI

Uguale a caso 1

INFORMAZIONI SULLO SPORT

18. Hai mai praticato sport?

☐ Si

☐ No

19. Tra i tuoi 9 e 17 anni... Nel tuo quartiere (se vivevi in città) o nel tuo comune (se vivevi in un centro abitato di piccole dimensioni) vivevano o si allenavano dei campioni dello sport?

Si. Scrivi in quale sport _____

No. Segna con una X _____

20. Se ne avessi l'opportunità, ti piacerebbe praticare qualche sport?

☐ Si

☐ No

21. Perché non hai mai praticato sport? Tra le motivazioni, scegline al massimo 3

Questioni tecniche

- ☐ 1. non mi divertivo
- ☐ 2. non mi sentivo abbastanza bravo
- ☐ 3. troppa competizione
- ☐ 4. eccessiva fatica
- ☐ 5. motivi di salute
- ☐ 6. Disabilità temporanea o permanente

Rapporti interpersonali:

- ☐ 7. non ero appoggiato dai miei genitori
- ☐ 8. i miei genitori non mi potevano accompagnare
- ☐ 9. non avevo qualcuno con cui allenarmi
- ☐ 10. sport troppo lontano

Contesto sociale:

- ☐ 11. non andavo bene a scuola
- ☐ 12. non riuscivo a conciliare lo sport con l'impegno scolastico
- ☐ 13. avevo altri interessi prevalenti

Contesto sportivo:

- ☐ 14. difficoltà a raggiungere l'impianto
- ☐ 15. lo sport era troppo costoso
- ☐ 16. palestra non adeguata
- ☐ 17. materiali a disposizione in palestra carenti
- ☐ 18. orari scomodi
- ☐ 19. non ci sono sport che mi piacciono dove vivo io
- ☐ 20. altro _____

22. Tra le motivazioni che hai scelto scrivi il numero di quella che maggiormente ti ha spinto a non fare sport

IMPEGNO SCOLASTICO

23. Tra i tuoi 9 e 17 anni ... il pensiero di dover conciliare l'impegno scolastico con quello sportivo è stato un motivo che ti ha spinto a non fare sport

(1= completamente in disaccordo; 2= abbastanza in disaccordo; 3= né in disaccordo né in accordo; 4= abbastanza d'accordo; 5=completamente d'accordo)

1	2	3	4	5
---	---	---	---	---

24. Tra i 9 e i 17 anni...Quante ore studiavi in media al di fuori dell'orario scolastico?

☐ Nessuna ☐ 1 ora ☐ 2 ore ☐ 3 ore ☐ 4 ore ☐ Più di 4

GENITORI

25. Tra i tuoi 9 e i 17 anni... Tua mamma lavorava?

☐ Sì

☐ No

26. Tra i tuoi 9 e 17 anni... Che professione svolgeva tua mamma?

☐ Libera professione (avvocato, medico, commercialista, artista, carriera sportiva ecc.)

☐ Commerciante/Imprenditrice/Artigiana

☐ Impiegata

☐ Operaia

☐ Casalinga

☐ Docente

☐ Altro _____

27. Tra i tuoi 9 e 17 anni... Tua mamma era impegnata per lavoro...

☐ Full time

☐ Part time

☐ Si gestiva autonomamente

☐ Altro _____

28. Tra i tuoi 9 e 17 anni... Tuo papà lavorava?

☐ Sì

☐ No

29. Tra i tuoi 9 e 17 anni... Che professione svolgeva tuo papà?

☐ Libera professione (avvocato, medico, commercialista, artista, carriera sportiva ecc.)

☐ Commerciante/Imprenditore/Artigiano

☐ Impiegato

☐ Operaio

☐ Casalingo

☐ Docente

☐ Altro _____

30. Tra i tuoi 9 e 17 anni... Tuo papà era impegnato per lavoro...

☐ Full time

☐ Part time

☐ Si gestiva autonomamente

☐ Altro _____

31. I tuoi genitori (o uno di essi) ti hanno influenzato in modo determinante nella scelta di non fare sport?

☐ Sì ☐ No

32. Tra i tuoi 9 e 17 anni... Quanti dei tuoi fratelli hanno fatto sport?

☐ Nessuno

☐ 1

☐ 2

☐ 3

☐ Più di tre

33. Qualcuno ha influenzato la tua scelta di non praticare sport al di fuori della scuola?

☐ Sì. Scrivi chi è stato _____

☐ No. Metti una X _____

34. I tuoi genitori tra i 9 e i 17 anni praticavano sport?

☐ Sì, papà. Scrivi lo sport che praticava _____

☐ Sì, mamma. Scrivi lo sport che praticava _____

☐ Sì, entrambi. Scrivi lo sport che praticavano _____

☐ No. Segna con una X _____

☐ Non lo so. Segna con una X _____

35. Tra i tuoi 9 e 17 anni... lo sport per te era... (scegline al massimo 3)

☐ Un sacrificio

☐ Un divertimento

☐ Un dovere

☐ Una passione

☐ Un bisogno del tuo corpo e della tua mente

☐ Un modo per essere liberi

☐ Un modo ricaricarmi

☐ Un modo per staccare dalla vita quotidiana

☐ Un modo per scaricare l'energia

☐ Una perdita di tempo

☐ Altro _____

36. Per favore, leggi le seguenti caratteristiche di personalità e indica quanto ti senti descritto da ogni coppia, anche se pensi che una delle due caratteristiche ti descriva più dell'altra.

Ti ritieni una persona...

Completamente in disaccordo	Molto in disaccordo	Un po' in disaccordo	Né d'accordo né in disaccordo	Un po' d'accordo	Molto d'accordo	Completamente d'accordo
1	2	3	4	5	6	7

Sono una persona...

1.	Estroversa, esuberante	1	2	3	4	5	6	7
2.	Polemica, litigiosa	1	2	3	4	5	6	7
3.	Affidabile, auto-disciplinata	1	2	3	4	5	6	7
4.	Ansiosa, che si agita facilmente	1	2	3	4	5	6	7
5.	Aperta alle nuove esperienze, con molti interessi	1	2	3	4	5	6	7
6.	Riservata, silenziosa	1	2	3	4	5	6	7
7.	Comprensiva, affettuosa	1	2	3	4	5	6	7
8.	Disorganizzata, distratta	1	2	3	4	5	6	7
9.	Competitiva	1	2	3	4	5	6	7
10.	Tranquilla, emotivamente stabile	1	2	3	4	5	6	7
11.	Tradizionalista, abitudinaria	1	2	3	4	5	6	7

37. Ti è piaciuto questo test?

- ☐ Sì
- ☐ No

Il tuo questionario finisce qui. Grazie per aver partecipato!

Questionario Predittivo

1) Sei maschio o femmina?

- ☐ Maschio
☐ Femmina

2) Età

9 ▼

3) Quale/i sport pratichi?

- ☐ Aikido
☐ Alpinismo
☐ Armi sportive da caccia

Specifica la disciplina

- ☐ Arrampicata

Specifica la disciplina

- ☐ Arti circensi

Specifica la disciplina

- ☐ Arti marziali

Specifica la disciplina

- ☐ Atletica leggera

Specifica la disciplina

- ☐ Attività subacquee (es. nuoto pinnato, apnea ecc.)

Specifica la disciplina

- ☐ Automobilismo

Specifica la disciplina

☐ **Badminton**☐ **Baseball**☐ **Beach sport (es. soccer, tennis, volley ecc.)****Specifica la disciplina**☐ **Biathlon**☐ **Biliardo sportivo****Specifica la disciplina**☐ **Bob**☐ **Bocce****Specifica la disciplina**☐ **Body building**☐ **Bowling**☐ **Bridge**☐ **Calcio (a 5, a 7, a 8, a 11)****Specifica la disciplina**☐ **Calisthenics**☐ **Canoa****Specifica la disciplina**☐ **Canottaggio****Specifica la disciplina**☐ **Cheerleading**☐ **Ciclismo****Specifica la disciplina**☐ **Cricket**☐ **Dama**☐ **Danza (es. classica, caraibiche, valzer, hip hop ecc.)****Specifica la disciplina**

☐ **Fitness****Specifica la disciplina**☐ **Football americano**☐ **Freccette**☐ **Ginnastica (es. artistica, acrobatica, parkour ecc.)****Specifica la disciplina**☐ **Golf**☐ **Hitball**☐ **Judo**☐ **Ju-Jitsu**☐ **Karate**☐ **Ginnastica (es. artistica, acrobatica, parkour ecc.)**☐ **Kickboxing****Specifica la disciplina**☐ **Lacrosse**☐ **Lotta****Specifica la disciplina**☐ **Motociclismo****Specifica la disciplina**☐ **Motonautica****Specifica la disciplina**☐ **Muay Thai**☐ **Nuoto****Specifica la disciplina**☐ **Nuoto sincronizzato**☐ **Orienteering**

☐ Pallacanestro☐ Pallamano☐ Pallapugno☐ Pallanuoto☐ Pallatamburello☐ Pallavolo☐ Pentathlon moderno

Specifica la disciplina

☐ Pesca sportiva☐ Pesistica☐ Pole dance☐ Pugilato☐ Rafting☐ Rugby

Specifica la disciplina

☐ Scacchi☐ Scherma

Specifica la disciplina

☐ Sci

Specifica la disciplina

☐ Sci nautico

Specifica la disciplina

☐ Slittino☐ Softball☐ Sport a rotelle (es. pattinaggio artistico, skateboard, hockey ecc.)

Specifica la disciplina

☐ Sport dell'aria (es. paracadutismo, volo ecc.)

Specifica la disciplina

- ☐ Sport da combattimento (es. boxe, kickboxing ecc.)

Specifica la disciplina

- ☐ Sport equestri

Specifica la disciplina

- ☐ Sport su ghiaccio (es. pattinaggio di velocità, hockey, curling ecc.)

Specifica la disciplina

- ☐ Squash

- ☐ Surfing

- ☐ Taekwondo

- ☐ Tennis (es. tennis, paddle ecc.)

Specifica la disciplina

- ☐ Tennis tavolo

- ☐ Tiro a segno

Specifica la disciplina

- ☐ Tiro a volo

Specifica la disciplina

- ☐ Tiro con l'arco

Specifica la disciplina

- ☐ Tiro dinamico sportivo

Specifica la disciplina

- ☐ Triathlon

Specifica la disciplina

- ☐ Tuffi

Specifica la disciplina

☐ **Ultimate frisbee**

☐ **Vela**

Specifica la disciplina

☐ **Yoga**

Specifica la disciplina

☐ **Altro**

Specifica la disciplina

4) Qual è lo sport per te più significativo? Per favore scegli uno tra quelli che hai selezionato prima altrimenti non ha senso!

5) Come pratichi il tuo sport più significativo?

☐ **individuale**

☐ **coppia**

☐ **squadra**

6) Riguardo il tuo sport significativo, lo pratichi a livello agonistico?

☐ **No**

☐ **Sì**

7) Quante volte alla settimana, in media, pratichi il tuo sport significativo? Considera anche le partite (o le gare o i saggi)

☐ **1 volta**

☐ **2 volte**

☐ **3 volte**

☐ **Più di tre**

8) Come ti senti prima di andare ad allenarti?

☐ **Stanco/a**

☐ **Felice**

☐ **Carico/a**

☐ **Svegliato/a**

☐ **Demotivato/a**

☐ **Ansioso/a**

☐ Altro

Specifica

9) Quali fattori ti spingono a continuare a praticare lo sport significativo? Tra le seguenti motivazioni, scegline massimo 3.

- ☐ Mi diverto
- ☐ Mi sento bravo
- ☐ Mi piace la competizione
- ☐ Non mi sono mai infortunato/a
- ☐ Mi piace faticare
- ☐ Mi piace migliorare le mie abilità
- ☐ Voglio essere più bravo e/o forte degli altri nel mio sport
- ☐ Voglio tenermi in forma
- ☐ Ho un obiettivo (es. vincere trofei o raggiungere alti livelli)
- ☐ Mi piace provare il brivido della vittoria
- ☐ Mi piace battere l'avversario
- ☐ Mi serve a scaricare l'energia o la tensione
- ☐ Voglio guadagnare denaro e ottenere contratti
- ☐ Mi piace quando faccio sport
- ☐ Motivi religiosi
- ☐ Sento l'esigenza di fare sport
- ☐ E' una passione
- ☐ E' uno sport alla moda
- ☐ Sono appoggiato dai miei genitori
- ☐ Per far contenti i miei genitori (famiglia)
- ☐ I miei genitori vogliono che io faccia sport
- ☐ Per far contento il mio allenatore
- ☐ Bel rapporto tra i miei genitori e l'allenatore
- ☐ Mi trovo bene con i miei compagni di squadra
- ☐ Ho qualcuno con cui allenarmi
- ☐ Ho un allenatore che mi piace
- ☐ Riesco a conciliare lo sport con lo studio/lavoro
- ☐ Trovo sempre il tempo per fare sport
- ☐ L'impianto è facile da raggiungere
- ☐ L'impianto è vicino casa
- ☐ Lo sport non è troppo costoso
- ☐ Palestra adeguata
- ☐ Orari comodi
- ☐ Altro

Specifica

10) Quali fattori, invece, ti spingerebbero ad abbandonare? Tra le seguenti motivazioni, scegline massimo 3.

- ☐ Mancanza di divertimento
- ☐ Mancanza di abbastanza bravura
- ☐ Troppa competizione
- ☐ Infortunio/i sportivo/i
- ☐ Eccessiva fatica
- ☐ Incomprensioni con l'allenatore
- ☐ Motivi di salute
- ☐ Periodo critico nella vita personale
- ☐ Sport troppo impegnativo
- ☐ Disabilità temporanea o permanente
- ☐ Mancanza di appoggio dei genitori
- ☐ Mancanza di accompagnamento da parte dei genitori
- ☐ Rapporto difficile tra genitori e allenatore
- ☐ Mancanza di un buon rapporto con i compagni
- ☐ Mancanza di qualcuno con cui allenarmi
- ☐ Sport troppo lontano
- ☐ Basso rendimento scolastico
- ☐ Non riuscire a conciliare lo sport con l'impegno scolastico
- ☐ Altri interessi prevalenti
- ☐ Difficoltà a raggiungere l'impianto
- ☐ Sport troppo costoso
- ☐ Palestra non adeguata
- ☐ Materiale a disposizione in palestra carente
- ☐ Orari scomodi
- ☐ Mancanza di sport preferiti dove vivo
- ☐ Altro

Specifica

11) Riesci a conciliare l'impegno scolastico con l'impegno sportivo

- ☐ Completamente in disaccordo
- ☐ Abbastanza in disaccordo
- ☐ Nè in disaccordo nè in accordo
- ☐ Abbastanza d'accordo
- ☐ Completamente d'accordo

12) Il tuo allenatore ti dà le giuste attenzioni e importanza (es. ti segue negli esercizi quando hai bisogno)?

- ☐ No
☐ Sì

13) Il tuo allenatore è (scegline al massimo 3)

- ☐ Severo
☐ Autorevole
☐ Un amico o confidente
☐ Competente
☐ Un buon leader
☐ Troppo fissato
☐ Un punto di riferimento al di fuori della palestra
☐ Non in grado di allenare
☐ Avvilente
☐ Irritabile
☐ Un motivatore
☐ Non è sufficientemente preparato tecnicamente
☐ Altro

Specifica

14) Hai compagni che ti stimolano a impegnarti in allenamento e/o durante le gare/partite?

- ☐ No
☐ Sì

15) Almeno uno dei tuoi genitori ha praticato nella sua vita il tuo sport significativo? (puoi rispondere a più di un'opzione)

- ☐ Sì, a livello agonistico
☐ Sì, per svago
☐ No
☐ Non lo so

16) Quanto ti diverti durante gli allenamenti?

- ☐ Per niente
☐ Poco
☐ Neutrale
☐ Abbastanza
☐ Molto

17) Il tuo sport significativo per te è ... (scegline massimo 3)

- ☐ Un sacrificio
☐ Un divertimento

- ☐ Un dovere
- ☐ Una passione
- ☐ Un bisogno del tuo corpo e della tua mente
- ☐ Un modo per essere liberi
- ☐ Un modo per ricaricarmi
- ☐ Un modo per staccare dalla vita quotidiana
- ☐ Un modo per scaricare l'energia
- ☐ Una perdita di tempo

18) Leggi le seguenti caratteristiche di personalità e indica quanto ti senti descritto da ogni coppia, anche se pensi che una delle due caratteristiche ti descriva più dell'altra

Sono una persona polemica, litigiosa

- ☐ Completamente in disaccordo
- ☐ Molto in disaccordo
- ☐ Un po' in disaccordo
- ☐ Nè d'accordo nè in disaccordo
- ☐ Un po' d'accordo
- ☐ Molto d'accordo
- ☐ Completamente d'accordo

Sono una persona aperta alle nuove esperienze, con molti interessi

- ☐ Completamente in disaccordo
- ☐ Molto in disaccordo
- ☐ Un po' in disaccordo
- ☐ Nè d'accordo nè in disaccordo
- ☐ Un po' d'accordo
- ☐ Molto d'accordo
- ☐ Completamente d'accordo

Sono una persona competitiva

- ☐ Completamente in disaccordo
- ☐ Molto in disaccordo
- ☐ Un po' in disaccordo
- ☐ Nè d'accordo nè in disaccordo
- ☐ Un po' d'accordo
- ☐ Molto d'accordo
- ☐ Completamente d'accordo

Sono una persona tranquilla, emotivamente stabile

- ☐ Completamente in disaccordo

2/5/2020

Dropout Sportivo

- ☐ Molto in disaccordo
- ☐ Un po' in disaccordo
- ☐ Nè d'accordo nè in disaccordo
- ☐ Un po' d'accordo
- ☐ Molto d'accordo
- ☐ Completamente d'accordo

Submit

Bibliografia

- [1] S. Scotto di Luzio, G. Martinent, E. Guillet-Descas, M. P. Daigle (2019), *Exploring the Role of Sport Sense of Community in Perceived Athlete Burnout, Sport Motivation, and Engagement*, Journal of Applied Sport Psychology
- [2] H. K. Larson, K. McFadden, T. F. McHugh, T. R. Berry, W. M. Rodgers (2018), *When you don't get what you want—and it's really hard: Exploring motivational contributions to exercise dropout*, Psychology of Sport & Exercise
- [3] P. A. Witt, T. B. Dangi (2018), *Why Children/Youth Drop Out of Sports*, Journal of Park and Recreation Administration
- [4] J. Cotè, D. J. Hancock (2014), *Evidence -based policies for youth sport programmes*, International Journal of Sport Policy
- [5] J. J. Murphy, C. B. Woods, D. Dijk, J. Seghers (2018), *Factors and actions to prevent dropout in youth sport: results of a rapid review and expert opinions*, 7th International Society for Psychological Activity and Health Congress
- [6] S. Vella, D. P. Cliff, A. D. Okely (2014), *Socio-Ecological Predictors of Participation and Dropout in Organised Sports during Childhood*, International Journal of Behavioral Nutrition and Physical Activity
- [7] J. Tribojevic, J. Petrovic (2018), *Dropping out of sports in adolescence: The role of gender and amotivation for sports*, University of Novi Sad
- [8] O. Molinero, A. Salguero, S. Marquez (2009), *Reasons for dropout in youth soccer: a comparison with other team sport*, Universidad de Leon
- [9] F. Vitali (2012), *L'abbandono nello sport giovanile: le strategie contro*, Università degli Studi di Verona, *I giovani e lo sport della montagna: 4° Convegno Nazionale di Formazione*

[10] C. Ratti (2019), *Analisi del drop out sportivo in adolescenza: l'esperienza di Sport Grand Tour Srl*, Università degli Studi di Milano Bicocca, Tesi di Laurea in Scienze Statistiche ed Economiche

[11] A. Geron (2017), *Hands-On Machine Learning with Scikit-Learn and TensorFlow*, O'Reilly Media Inc

[12] <https://scikit-learn.org/>, Libreria Open Source di apprendimento automatico per il linguaggio di programmazione Python, David Cornapeau, BSD License

[13] J. M. Wooldridge (2016), *Introductory econometrics: A modern approach*, Nelson Education

[14] G. W. Imbens (2014), *Instrumental variables: an econometrician's perspective*, National Bureau of Economic Research

Ringraziamenti

Ringrazio i miei relatori, il Prof. Pelagatti e la Prof.ssa Dalla Pellegrina, per avermi dato l'opportunità di approfondire una tematica che applica le metodologie di Data Science al mondo dello sport, per cui nutro da sempre tanta passione e che desidero sia al centro della mia carriera professionale.

Li ringrazio, inoltre, per avermi supportato durante il periodo di stage pre-laurea e nella scrittura di questa tesi.

Ringrazio i miei genitori, per avermi supportato durante il mio percorso di studi e per aver appoggiato le mie scelte di vita, soprattutto nei periodi più critici personali, da cui sono uscito più forte che mai.

Ringrazio il mio caro amico Abderazzak, che ho conosciuto pochi giorni dopo il mio trasferimento a Milano e con cui ho condiviso esperienze di lavoro, di studio e di tempo libero.