



UNIVERSITÀ DEGLI STUDI DI PADOVA
DIPARTIMENTO DI PSICOLOGIA GENERALE

Corso di laurea Magistrale in
Psicologia Cognitiva Applicata

Tesi di Laurea Magistrale

**Instagram ed immagine corporea:
un'analisi delle immagini profilo basata sul data-mining**

Instagram and body image:
a profile pictures analysis based on data-mining

Relatore
Prof. Antonio Calcagni

Laureando: Amos Torretta
Matricola: 1202558

Anno Accademico 2020/2021

Sommario

Introduzione	5
1. Psicologia e social media	7
1.1 L'impronta digitale	7
1.2 Social media e psicologia, alcune ricerche	10
1.3 Instagram	12
1.3.1 Motivazioni all'uso di Instagram	14
1.3.2 Instagram e psicologia, alcune ricerche	16
1.3.3 Instagram, immagine corporea e auto-oggettivazione.....	18
2. Ricerca e campione	21
2.1 Ricerca ed obiettivi	21
2.1.1 Aspetti etici.....	22
2.2 I dati	23
2.2.1 Instagram Addiction Scale	23
2.2.2 Instagram Activity Scale	24
2.2.3 SATAQ.....	24
2.2.4 SOBBS	25
2.2.5 Informazioni demografiche	25
2.2.6 I dati di Instagram.....	26
2.3 Il campione.....	27
3. Costruzione delle variabili	29
3.1 Variabili dipendenti.....	29
3.2 Le immagini di Instagram.....	29
3.2.1 Data-mining	30
3.2.2 Object Detection	33

3.2.3 Data-mining in questa ricerca	35
3.2.4 Creazione degli indici.....	39
3.3 Variabili indipendenti.....	40
3.3.1 Immagini caricate sul profilo	40
3.3.2 Immagine del profilo	43
4. Analisi dati	47
4.1 Alcune statistiche descrittive	47
4.2 Aspetti teorici dell'analisi.....	49
4.2.1 Predire o spiegare i dati?.....	49
4.2.2 Alberi di regressione.....	51
4.2.3 Random Forest	54
4.3 Gli step dell'analisi	55
5. Risultati.....	57
5.1 Immagini del profilo.....	57
5.2 Immagini caricate sul profilo	58
5.2.1 Instagram Addiction Scale	60
5.2.2 Instagram Activity Scale: Interaction	62
5.2.3 Instagram Activity Scale: Browsing	64
5.2.4 Instagram Activity Scale: Broadcasting	66
5.2.5 Immagine corporea (SATAQ).....	68
5.2.6 Auto-oggettivazione (SOBBS)	70
5.2.7 Un esempio di un albero di regressione	72
Conclusioni	75
Bibliografia	79
Appendice.....	85

Introduzione

L'obiettivo di questo lavoro è stato quello di mettere in relazione i dati di Instagram di 250 giovani donne con alcuni costrutti psicologici, tra cui l'immagine corporea, l'auto-oggettivazione e la dipendenza da questo *social media*. La novità proposta risiede nell'utilizzo di un metodo di *data-mining* in grado di estrarre informazione dai dati di Instagram e più nello specifico dalle immagini. Le partecipanti hanno infatti scaricato ed inviato ai ricercatori i propri dati Instagram, e grazie al *software* di programmazione Python sono stati implementati degli algoritmi in grado di estrarre le informazioni di interesse da questi file. Le oltre 75000 immagini caricate su questo *social media* dalle partecipanti sono state analizzate tramite delle reti neurali già addestrate, arrivando così a conoscerne il contenuto. È stato evidenziato come in letteratura siano veramente poche le ricerche che hanno preso in considerazione questa tipologia di dati e che li abbiano analizzati con delle procedure automatiche e completamente replicabili. I codici utilizzati sono disponibili in una *repository* su GitHub¹. Successivamente, basandosi principalmente sui dati estratti dalle immagini, sono stati utilizzati dei modelli non parametrici, i *random forest*, che hanno permesso di creare degli algoritmi predittivi per poter risalire ai punteggi dei questionari indaganti i costrutti psicologici presi in esame. L'utilizzo di modelli predittivi è stato motivato da delle considerazioni metodologiche circa alcuni aspetti negativi delle più comuni tipologie di analisi nella ricerca in psicologia, basate generalmente su modelli descrittivi. I risultati hanno messo in luce come il conoscere i contenuti delle immagini pubblicate su Instagram permetta di predire i costrutti di interesse. Specialmente per quanto riguarda il costrutto dell'immagine corporea e dell'auto-oggettivazione, è stato evidenziato come i contenuti delle immagini riguardanti le persone e come queste vengano ritratte siano i più importanti per tali predizioni. Inoltre sono emerse delle caratteristiche positive di queste tipologie di analisi che hanno permesso di riconoscere delle relazioni complesse tra i dati estratti, laddove tipologie di analisi più tradizionali avrebbero condotto a risultati di difficile interpretazione.

¹ https://github.com/AmosTorretta/Datamining_on_Instagram_images

1. Psicologia e social media

1.1 L'impronta digitale

Uno degli obiettivi che la psicologia cognitiva si è da sempre proposta è stato quello di studiare il comportamento umano per poterlo comprendere, predire ed influenzare. Per provare a raggiungere questo obiettivo, il primo passo da compiere è sicuramente quello di raccogliere dati circa i fenomeni di interesse, e quindi circa determinate caratteristiche psicologiche o più in generale circa il comportamento di un gruppo di individui. Tradizionalmente per raccogliere informazioni che rispecchino il comportamento di un soggetto viene svolto uno studio in laboratorio dove sono osservate e misurate le variabili di interesse. Un altro metodo molto comune in psicologia cognitiva è quello di somministrare dei questionari nei quali sono gli stessi soggetti a rispondere alle domande e fornire così i dati circa il proprio comportamento. Negli ultimi anni però è nata una nuova possibilità per accedere ad informazioni circa il comportamento di un individuo, e sta prendendo sempre più spazio nella ricerca psicologica.

Con l'arrivo delle nuove tecnologie, e nello specifico la possibilità di avere sempre accesso ad internet in qualsiasi luogo grazie agli smartphone, sono nate delle piattaforme dove le persone hanno la possibilità di interfacciarsi tra di loro. Si pensi per esempio ai *social media*. L'aspetto interessante da notare subito è che le azioni che gli utenti compiono utilizzando queste piattaforme lasciano delle tracce quasi indelebili di ciò che fanno, e quindi del loro comportamento. Questa impronta delle loro azioni viene chiamata *digital footprint* (Azucar et al. 2018). Certamente la gamma dei possibili comportamenti che si possono svolgere su queste piattaforme è limitata rispetto a quella che potrebbe avvenire in un contesto naturale. Le *digital footprint* contengono una gran varietà di dati, che possono essere testuali, immagini, video ed audio. Alcune di queste rappresentano le specifiche azioni che si possono svolgere all'interno delle varie piattaforme, e quindi le interazioni tra gli utenti. Analizzarle direttamente non è possibile e sono necessarie delle competenze di programmazione per trasformare questi dati in informazioni utilizzabili. Le impronte digitali costituiscono una grande mole di dati circa il comportamento di un individuo e non sono

soggette a distorsioni tipiche di altre procedure di raccolta dati come per esempio i questionari, sia auto che etero somministrati. I dati ci sono e sono “puri”, sta ai ricercatori capire come usarli. Per poter rendere analizzabili questo tipo di dati viene utilizzato un approccio chiamato di *data-mining*. Tramite algoritmi, scritti grazie a software di programmazione, e metodi automatici, come per esempio metodi di *machine learning*, vengono estratte informazioni da questa grande mole di dati. La metafora è quella della miniera (i dati grezzi) all’interno della quale tramite varie procedure vengono estratti i minerali di interesse (dati analizzabili).

Nel mondo si stima ci siano 3.96 miliardi di utenti attivi sui social, 35 milioni dei quali in Italia. Globalmente il tempo medio giornaliero trascorso sui *social media* è di 2h e 22min e Italia di 1h e 57min². Il 98% degli Italiani inoltre accedono ai *social media* da *smartphone*. Da *computer* i *social* più visitati in Italia sono in ordine: Facebook, YouTube e Instagram. La durata media della visita è rispettivamente di circa 10 minuti, 21 minuti e 8 minuti. Il numero di pagine per visita invece di 8, 8 e 12³. Maggiori informazioni circa questi dati possono essere trovate in Tabella 1 e in Tabella 2.

	<i>Nel Mondo</i>	<i>In Italia</i>
Popolazione	7790 milioni	60 milioni
Utenti attivi sui <i>social media</i>	3960 milioni	35 milioni
Incremento utenti <i>social media</i> (tra luglio 2019 e luglio 2020)	376 milioni	4 milioni
Tempo trascorso su internet (al giorno)	6 h 42 min	6 h 0 min
Tempo trascorso sui <i>social media</i> (al giorno)	2 h 22 min	1 h 57 min
Utenti che accedono ai <i>social media</i> da <i>smartphone</i>	99%	98%

Tabella 1: *Utilizzo dei social media nel mondo ed in Italia*

² (2021, 17 gennaio) <https://archivio.lucapoma.info/comunicazione/comunicazione-non-convenzionale/lo-stato-dei-social-network-in-italia-nel-2020/>

³ (Ibidem)

	<i>Durata della visita</i>	<i>Numero di pagine per visita</i>
Facebook	10 min 48 sec	8.88
YouTube	21 min 54 sec	8.9
Instagram	8 min 14 sec	12.61
Twitter	11 min 15 sec	12.07
WhatsApp	2 min 46 sec	1.72

Tabella 2: *Utilizzo medio dei social media più visitati in Italia*

Negli ultimi anni i *social media* sono diventati parte integrante della vita delle persone ed hanno acquisito sempre più importanza. Per questo motivo sono stati condotti molti studi in ambito psicologico per analizzare come questi stiano influenzando la vita delle persone. Raramente, tuttavia, in questi studi vengono presi in considerazione i veri e propri dati dei *social media*. Questo accade sia per le complessità tecniche dell'analizzarli descritte precedentemente, sia per la difficoltà nella raccolta di questi dati, in quanto sono gli stessi soggetti che si devono occupare di recuperarli. Non è facile infatti convincere le persone a fornire queste informazioni, in quanto sono molto personali e le procedure per scaricare i propri dati sono lente e macchinose per l'utente medio. Le norme sulla privacy e la protezione dei dati di queste piattaforme inoltre hanno creato dei sistemi di sicurezza che impediscono di accedere a questi dati se non tramite l'utilizzo delle proprie credenziali e com'è facile intuire, non è una strada percorribile quella di chiedere nome utente e *password* ai soggetti di una ricerca. Per questi motivi quindi le ricerche che si sono occupate di *social media* hanno spesso analizzato dati provenienti da questionari somministrati ai partecipanti, e non i veri e propri dati dei *social media*.

1.2 Social media e psicologia, alcune ricerche

Come appena accennato sono state condotte varie ricerche in ambito psicologico per provare ad indagare le relazioni tra i *social media* e dei particolari costrutti. Ne viene ora presentata una breve rassegna per provare a delineare un quadro generale sui principali ambiti indagati ed i metodi utilizzati.

Inizialmente molte ricerche sono state condotte allo scopo di predire le caratteristiche di personalità dei soggetti in relazione ai dati dei loro *social media*. Già in una ricerca del 2011 condotta all'università del Maryland, Washington, Golbeck e colleghi crearono un modello che permetteva di stimare i punteggi al *Big Five Questionnaire* (BFQ), uno dei più diffusi ed utilizzati questionari di personalità, partendo dai dati pubblici dei profili Twitter (Golbeck et al., 2011). Con delle tecniche di analisi testuale presero in considerazione gli ultimi 2000 *tweet* di ognuno dei 50 soggetti che aveva aderito alla ricerca compilando il BFQ. Sulla base di questi ed altri dati pubblici estraibili dai profili Twitter dei partecipanti, come per esempio il numero di follower o il numero di risposte ai *tweet*, e grazie all'utilizzo di tecniche di *machine learning* si arrivò alla creazione del modello (Golbeck et al., 2011). Un'altra ricerca condotta qualche anno dopo, ebbe il medesimo scopo. Questa volta venne preso in considerazione un numero decisamente più elevato di partecipanti e, cosa più importante, i soggetti fornirono ai ricercatori i loro dati Facebook. Il modello creato partendo direttamente dai dati degli utenti diede dei risultati superiori rispetto a quello di Golbeck (Markovikj, Gievska, Kosinski e Stillwell, 2013). Per arrivare alla raccolta di questi dati infatti nel 2007 alcuni ricercatori dell'università di Cambridge crearono un sito internet (myPersonality.org) dove ogni utente che lo visitava aveva la possibilità di compilare dei questionari di personalità ed avere successivamente accesso ai risultati. Se gli utenti avessero voluto, infine, avrebbero potuto lasciare i propri dati Facebook ai ricercatori con pochi click. Sei milioni di utenti visitarono il sito e compilarono almeno uno dei questionari e il 40% decise generosamente di lasciare i propri dati. Markovikj e colleghi attinsero da questo database per la loro ricerca. I dati di myPersonality vennero infatti condivisi e furono svolte diverse ricerche e creati vari modelli in grado di predire accuratamente le caratteristiche di personalità e non solo (Stillwell e Kosinski, 2004). Successivamente però il sito fu chiuso e i dati raccolti cessarono di essere condivisi con il resto della comunità scientifica, molto probabilmente a causa dello scandalo di Facebook e

Cambridge Analytica. I legami tra queste ricerche e la società di consulenza britannica restano però torbidi e si trovano notizie contrastanti a riguardo (Prichard, 2021).

Altre ricerche che presero in analisi i *social media* ebbero scopi differenti. Savci e Tekin (2020), per esempio, condussero uno studio per cercare di predire la PSU (*Problematic social media use*) utilizzando un approccio di analisi basato sul *machine learning*. Vennero indagate quindici variabili predittive, tra cui le abitudini dei soggetti nell'utilizzo dei *social media*: frequenza di utilizzo giornaliera, frequenza di controllo dei propri profili, numero di condivisioni e altri. Fu identificata una correlazione di 0.63 tra il modello e la variabile di risposta e vennero identificati i migliori predittori di questa tra cui: la frequenza di utilizzo giornaliera, la frequenza di controllo dei propri profili e la *Fear of missing out* (FOMO), un costrutto che indaga la paura di essere tagliato fuori dall'attività dei propri pari (Savci e Tekin, 2020). Tuttavia, nonostante le metodologie di analisi avanzate, i dati restano raccolti tramite l'utilizzo di questionari *self report*, cosa che viene identificata dagli stessi autori come uno dei principali limiti della ricerca.

Altri studi invece presero analizzarono le relazioni con dei costrutti patologici legati alla salute mentale. Partendo da studi di psicolinguistica sulla depressione, Schwartz e colleghi esaminarono i *post* di 28000 utenti Facebook, che precedentemente avevano anche compilato dei questionari di personalità, e trovarono che alcune caratteristiche estratte dall'analisi dei *post* sono modestamente correlate alla severità della depressione (Schwartz et al. 2014). Tra queste furono identificate per esempio la mancanza di speranza, la mancanza di aiuto o la solitudine. Similmente un'altra ricerca, condotta invece partendo dai dati di Twitter, mise in luce come il livello di depressione negli utenti correli ad alcune caratteristiche nell'utilizzo del *social media* come l'orario di pubblicazione dei *post*, la frequenza dei *post* ma anche ad altre caratteristiche degli stessi *post* come l'uso del pronome di prima persona (De Choudhury, 2013). Coppersmith e colleghi invece esaminarono il profilo Twitter di utenti che avevano tentato il suicidio analizzando caratteristiche e pattern nel linguaggio dei *post* (Coppersmith et al. 2016). Furono trovati quantificabili segnali di rischio suicidario e venne creato un semplice modello di classificazione allo scopo di utilizzarlo come strumento di screening non invasivo. Una ricerca svolta invece da Ive e colleghi prese in esame invece i dati provenienti da Reddit per analizzarli in relazione a vari costrutti della salute mentale (Ive et. al. 2018).

Questi erano undici problemi di salute mentale tra cui: schizofrenia, ansia, depressione e disturbo bipolare. Per condurre le analisi vennero usate delle reti neurali ricorrenti, un approccio più interessante rispetto alle più diffuse reti neurali convoluzionali, per le loro peculiarità nell'analisi testuale (Ive et. al. 2018). I dati dei *social media* analizzati infatti erano ancora di tipo testuale ricavati dai *post* degli utenti.

Come si può notare da questa breve rassegna la maggior parte degli studi che hanno preso in esame i vari *social media* ha analizzato principalmente due tipi di dati. I primi sono inerenti al come i *social media* vengono utilizzati, il che si traduce in variabili come il tempo medio giornaliero trascorso sui *social* o anche il numero di *post* pubblicati da un utente. I secondi invece riguardano più i contenuti proposti dagli stessi utenti. Sono stati studiati quasi esclusivamente dati di tipo testuale e non immagini, per una difficoltà tecnica maggiore nell'estrazione di informazioni analizzabili da queste ultime. Come espongono Guntuku e colleghi in una rassegna sui principali metodi di analisi dei *social media* in relazione alle malattie mentali, per creare un modello predittivo si inizia estraendo delle informazioni (*features*), che diventano poi le variabili predittive, dai dati dei *social media* (Guntuku et. al 2017). Vengono quindi individuate le caratteristiche nei contenuti che potrebbero essere interessanti in relazione al fenomeno in esame e vengono sviluppati i metodi specifici per poterle estrarre. Questo è il passaggio cruciale per condurre una buona ricerca, comporta infatti un *expertise* in ambedue gli ambiti menzionati, in quanto: l'estrazione di variabili non inerenti al costrutto indagato difficilmente porterebbe a dei risultati interessanti ed interpretabili; mentre l'incapacità di estrarre queste informazioni impedirebbe il proseguimento della ricerca.

1.3 Instagram

Instagram è un *social media* sviluppato da Systorm e Krieger e lanciato per la prima volta nell'ottobre del 2010. Ha guadagnato velocemente popolarità arrivando ad un milione di utenti in due mesi, a dieci milioni in un anno e a maggio del 2019 ha raggiunto il miliardo di iscritti. In Italia attualmente è il terzo *social* più visitato, dopo Facebook e YouTube. Nella nostra penisola nell'aprile del 2019 erano 19 milioni gli utenti mensilmente attivi, il 36.6% tra

i 19 e i 29 anni: 6.95 milioni⁴. Secondo i dati ISTAT aggiornati al 2020, la popolazione italiana tra i 20 e i 29 anni è di 6.15 milioni. Si può quindi concludere che in pratica la totalità dei giovani appartenente a questa fascia d'età sia attiva su Instagram.

Nasce inizialmente con l'obiettivo di condividere digitalmente con gli amici in maniera istantanea delle fotografie appena scattate, con la possibilità di utilizzare dei filtri, che agiscono sul colore, per rendere questi scatti amatoriali più professionali. L'idea deriva dalla popolare fotografia istantanea della Polaroid tipica degli anni Ottanta, che permetteva di osservare immediatamente con gli amici le immagini scattate, senza dover aspettare il tempo per lo sviluppo della pellicola. Dopo soli due anni dal lancio Instagram viene comprata dal colosso dei *social media* Facebook per un miliardo di dollari. Col tempo vengono lentamente e costantemente aggiunte delle nove funzionalità, senza però rivoluzionare l'app e lasciandola sostanzialmente diversa rispetto agli altri *social media*. Viene prima introdotto e migliorato un servizio di messaggistica interna all'app chiamato "direct message" e successivamente aggiunta la possibilità di condividere anche video e non solo immagini. Qualche anno dopo, prendendo l'idea da un altro social chiamato Snapchat, nascono le "Stories". Grazie a questa nova funzionalità possono essere condivise immagini e video disponibili unicamente per ventiquattr'ore, una volta scaduto questo periodo di tempo infatti vengono archiviate automaticamente e solamente chi le ha pubblicate ha la possibilità di rivederle. Un'altra interessante novità rispetto ad altri *social media* è l'introduzione della sezione "esplora". Normalmente sono gli utenti a scegliere quali sono i contenuti che vogliono visualizzare, possono infatti seguire altri account e visualizzare quello che essi pubblicano nella "bacheca", una sezione tipica dei *social media* e creata appositamente per svolgere questa funzione. Nella sezione "esplora" invece, si possono trovare contenuti di account simili a quelli che già si seguono. L'algoritmo che gestisce questa sezione prende quindi in esame il comportamento di ogni singolo utente, mostrando *post* simili con quelli con cui ha in precedenza interagito. Sicuramente ciò che viene visualizzato in questa sezione e come l'utente vi naviga può dire molto circa i suoi interessi. Un'altra caratteristica, tipica di Instagram, è la possibilità di utilizzare gli *hashtag*, una funzione precedentemente introdotta da Twitter. Quando un utente carica un'immagine o una storia ha la possibilità di aggiungere dei *tag* in qualche modo

⁴ (2021, 17 gennaio) <https://blog.axura.com/2019/04/instagram-in-italia-numeri-statistiche-e-curiosita-aggiornati-ad-aprile-2019/>

relazionati al contenuto dell'immagine, mettendo prima della parola di interesse il simbolo "#". Instagram dà in fatti la possibilità di cercare e seguire degli specifici hashtag, i *post* e le *stories* contrassegnate da questi hanno quindi la possibilità di essere visualizzate da un maggior numero di utenti, dando per riflesso maggior visibilità al profilo che le ha pubblicate. Col tempo inoltre sono state introdotte specifiche funzionalità per utilizzare il *social media* a fini commerciali, ma non verranno esposte in questo lavoro in quanto non sono inerenti ai nostri scopi. Nonostante questi cambiamenti ed aggiunte, Instagram resta un *social media* incentrato sulle immagini, il che lo rende più fluido e immediato rispetto alla concorrenza ed è probabilmente uno dei motivi del suo successo.

1.3.1 Motivazioni all'uso di Instagram

Da un punto di vista più strettamente psicologico si possono ipotizzare vari motivi per cui le persone sono portate ad usare Instagram. Una ricerca dell'università dell'Alabama che ha preso in considerazione un campione di 239 studenti, raccogliendo informazioni tramite la somministrazione di un questionario, ha evidenziato come tra le maggiori motivazioni all'uso di Instagram le due più comuni siano: la *Surveillance/Knowledge about Others* e cioè il desiderio di rimanere informati riguardo ciò che gli altri fanno; e la *Documentation* cioè il desiderio di documentare la propria vita per poter ricordare e celebrare momenti speciali e per condividerli con gli altri (Sheldon e Bryant, 2016). Quest'ultima ragione sembra essere tipica di questo *social media* in quanto non era stata evidenziata in studi precedenti condotti prendendo in esame altri *social media* (Sheldon e Bryant, 2016). Un altro studio simile condotto da un gruppo di ricercatori coreani ha identificato due motivazioni centrali nell'utilizzo di Instagram: il desiderio di auto-esprimersi e il desiderio di interazione sociale (Lee et. al, 2015). I ricercatori hanno ipotizzato che gli utenti usino le immagini per presentare l'idea che hanno di sé stessi e mantenere relazioni sociali (Lee et. al, 2015). Tra altre motivazioni, tipiche anche di altri *social media*, vi è indubbiamente la possibilità di entrare e restare in contatto con degli amici o familiari, soddisfacendo i bisogni sociali e sperimentando emozioni positive. In uno studio che mirava ad esplorare l'associazione tra la solitudine e le varie attività di Instagram, è stato evidenziato come un maggior utilizzo di Instagram è positivamente correlato a bassi livelli di solitudine (Yang, 2016). Mediando però questa relazione per la *Social Comparison Orientation* (SCO), un costrutto che quantifica la tendenza

di un soggetto a confrontarsi con gli altri, si è scoperto come questa relazione non sia sempre valida. La correlazione positiva tra uso di Instagram e solitudine infatti viene evidenziata unicamente in presenza di bassi livelli di SCO (Yang, 2016). Instagram inoltre permette facilmente di entrare in contatto anche con persone che non appartengono alla cerchia dei propri conoscenti, come personaggi famosi o brand, ma anche semplicemente sconosciuti. A differenza di come previsto inizialmente da Facebook infatti, Instagram dà la possibilità di seguire un utente senza che questi ci segua a sua volta, facilitando una sorta di contatto unidirezionale con personaggi famosi come per esempio attori o cantanti. Il fatto di seguire questi personaggi può influenzare il confronto sociale che gli utenti tendono a mettere in atto mentre usano questo social, e di conseguenza avere delle ripercussioni sul loro benessere (Farduoly e Bryde 2017).

Date le sue peculiarità Instagram permette ai propri utenti di raccontare sé stessi e condividere le proprie esperienze con i propri *follower*. Viene quindi condivisa in rete la propria vita privata, cosa che aiuta la costruzione e definizione della propria identità, specialmente in giovane età. Gli utenti tramite questo processo di auto-presentazione tendono quindi a dare la miglior visione di sé stessi, postando immagini che rappresentano gli aspetti più positivi della loro vita (Lup et al. 2015). Questo viene fatto tramite la pubblicazione di *post* e storie. Per entrambe i propri *follower* hanno la possibilità di interagire con dei *feedback* in relazione ai contenuti che vengono condivisi. Possono farlo pubblicamente, mettendo *like* ai *post* e commentandoli, o privatamente, rispondendo alle storie tramite il servizio di *direct message*. Applicando la teoria degli usi e delle gratificazioni nell'ambito dei *social media*, si può vedere come gli utenti siano spinti verso la ricerca di questi *feedback*, mettendo in atto comportamenti allo scopo di aumentarli (Wu et al. 2010). Una ricerca di Bell e colleghi ha messo in luce come la frequenza di pubblicazione dei *post* su Instagram può essere predetta dai contenuti dei *post* precedenti e dai *feedback* a questi, cioè il numero *like* e commenti (Bell et. al, 2018). I contenuti nello specifico erano i *selfie* presenti, e quali di questi mostrassero tratti di auto-oggettivazione. La ricerca mette quindi in luce come il numero di *feedback* sia correlato all'uso che un utente fa di questo *social media* e alla tipologia di contenuti che promuove (Bell et. al, 2018). Lo scopo, a volte inconscio, è quindi quello di aumentare le interazioni con i propri *follower*, che si ricorda possono rispecchiare o meno la cerchia dei propri amici e conoscenti. Come conferma infatti uno studio di Sheldon e Bryant

(2016), uno dei fattori che spinge all'utilizzo di Instagram è il desiderio di approvazione ed apprezzamento da parte degli altri. Gli autori hanno anche evidenziato come il ricevere apprezzamenti da altri utenti influenzi indirettamente la propria autostima (Sheldon e Bryant, 2016).

1.3.2 Instagram e psicologia, alcune ricerche

Il modo in cui Instagram viene utilizzato dai vari utenti e le possibili conseguenze del suo utilizzo da un punto di vista psicologico hanno ricevuto sempre più interesse dalla ricerca negli ultimi anni. Si è ancora lontani tuttavia dall'averne un quadro ampio e dettagliato di questo fenomeno. Alcune delle prime ricerche condotte si sono basate, al pari di altri *social media*, sul provare ad identificare come si differenzi l'utilizzo di Instagram in relazione alla personalità dei vari utenti. In uno studio svolto da un gruppo di ricercatori austriaci nel 2015, sono state prese in esame le ultime 20 immagini pubblicate su Instagram da ogni partecipante e i rispettivi tratti di personalità. Le immagini sono state analizzate con delle procedure computerizzate automatiche, per estrarre informazioni riguardanti i colori e la luminosità. Inoltre, allo scopo di ottenere altre informazioni, alcuni giudici hanno osservato ogni immagine singolarmente contando il numero di persone e volti per ognuna (Ferwerda et. al, 2015). I risultati hanno evidenziato alcune correlazioni interessanti tra i tratti di personalità e le proprietà delle immagini. Va notato che questo studio propone una novità nel panorama sulle ricerche nei *social media*, nella rassegna iniziale infatti è stato evidenziato come siano veramente pochi gli studi che hanno analizzato anche le immagini, in quanto la maggior parte si sono concentrati principalmente su dati di tipo testuale.

Vari studi invece hanno preso in esame il fenomeno dell'*addiction* ai *social media*. Fra questi uno studio condotto da Kircaburun e Griffiths (2018) ha messo in relazione le caratteristiche di personalità di 752 studenti, rilevate tramite il *Big Five Inventory*, con l'*addiction* ad Instagram, misurata tramite la *Instagram Addiction Scale*. I risultati hanno evidenziato come due tratti, amicalità e coscienziosità, siano negativamente associati alla dipendenza da Instagram.

Nonostante questo *social media* possa provocare diverse gratificazioni positive per il benessere psicologico, come la possibilità di esprimersi o passare il tempo intrattenendosi, una indagine promossa dalla UK Royal Society for Public Health ha messo in evidenza come Instagram sia la peggior piattaforma per la salute mentale tra i giovani tra i 14 e i 24 anni (Yurdagül et. al, 2019). Nella ricerca erano stati presi in esame anche altri quattro tra i più diffusi *social media*: Facebook, Twitter, YouTube e Snapchat. Sempre nello stesso studio inoltre è stato messo in luce come le femmine, rispetto ai partecipanti di sesso maschile, riportavano punteggi significativamente più elevati in variabili quali la depressione, solitudine, ansia generale ed ansia derivante dall'insoddisfazione per il proprio aspetto fisico (Yurdagül et. al, 2019). Un'altra ricerca, svolta da Sherlock e Wagstaff, ha mostrato come la frequenza nell'uso di Instagram sia correlata con sintomi depressivi, autostima e insoddisfazione corporea. Inoltre, ha messo in evidenza come le relazioni tra l'uso di Instagram ed ognuna di queste variabili sia mediata dalla tendenza al confronto sociale, cioè la propensione che un soggetto ha di paragonarsi agli altri. (Sherlock e Wagstaff, 2019). Anche una ricerca di Lup e colleghi (2015) ha evidenziato risultati simili mettendo in luce come l'uso di Instagram abbia una associazione negativa con il confronto sociale ed i sintomi depressivi. Questa associazione tuttavia si verifica unicamente per i soggetti che seguono molti sconosciuti, mentre diventa positiva per i soggetti che invece ne seguono pochi. Lo stesso studio ha inoltre evidenziato le implicazioni negative di un uso passivo di Instagram, al pari di come riscontrato da studi precedenti su altri *social media*. Un gruppo di ricercatori Coreani ha invece preso in considerazione il ruolo del narcisismo in relazione all'utilizzo di Instagram (Moon et al. 2016). I risultati hanno mostrato come soggetti con alti livelli di narcisismo tendano a postare più immagini di sé stessi, aggiornare la propria immagine del profilo più frequentemente e spendere più tempo sul *social media* (Moon et al. 2016). I dati dello studio, tuttavia, son stati interamente raccolti tramite questionari in cui si chiedeva ai soggetti di stimare il numero di *selfie* che avevano nel proprio profilo e il tempo medio giornaliero trascorso su Instagram, non è stato quindi analizzato nessun dato proveniente direttamente dal *social media*.

1.3.3 Instagram, immagine corporea e auto-oggettivazione

Viene ora presentato un costrutto psicologico particolarmente interessante in relazione all'uso di Instagram: l'immagine corporea. Schilder nel 1935 definì l'immagine corporea come: "l'immagine e l'apparenza del corpo umano che ci forniamo nella mente, e cioè il modo in cui il nostro corpo ci appare". Negli anni questo costrutto è andato specificandosi sempre più, venendo descritto da vari esperti come un costrutto multidimensionale. Ad esso possono essere attribuite quattro dimensioni principali: una percettiva, determinata da come una persona percepisce sensorialmente il proprio corpo in relazione per esempio alle sue forme e peso; una di carattere emotivo, e quindi relativa ai sentimenti e alle sensazioni che suscita la rappresentazione del proprio corpo; una di carattere cognitivo, che comprende i pensieri, schemi e credenze che un soggetto sviluppa circa il proprio corpo; ed una comportamentale, inerente ai comportamenti messi in atto nei confronti del proprio corpo (Slade, 1994; Gleaves et al., 1995).

Com'è stato esposto precedentemente, Instagram è un *social media* che viene usato per mostrare sé stessi, giocando inevitabilmente un ruolo nella costruzione della propria identità, e di riflesso anche nello sviluppo della propria immagine corporea. In questo processo di influenza sicuramente svolge un ruolo fondamentale il confronto sociale (Tiggemann et al. 2020). I soggetti che usano questo *social media* tendono a valutare sé stessi in relazione ai contenuti che possono trovare all'interno di Instagram, i quali molto spesso hanno un focus quasi esclusivo sull'aspetto fisico. Data la propensione a condividere immagini in cui si ha l'impressione di apparire particolarmente bene, e a quella di utilizzare filtri e inquadrature che migliorano da un punto di vista estetico le immagini, vengono promossi indirettamente dei canoni di bellezza media più distanti dalla realtà e sicuramente più idealizzati. I *social media* tendono infatti ad esporre gli utenti, e soprattutto le donne, ad una costante analisi del proprio aspetto fisico in relazione ai canoni di bellezza che vengono promossi e, vista la irrealtà e la forzatura di questi, è facile intuire come il confronto sociale messo in atto possa portare a sperimentare sentimenti negativi in relazione al proprio corpo, ed in generale un peggioramento del benessere psicologico. Una meta-analisi condotta da Grabe e colleghi sul ruolo dei media in relazione all'immagine corporea tra le donne, ha messo in luce come l'esposizione a vari tipi di media sia collegata ad una generalizzata insoddisfazione corporea

(Grabe, Hyde e Ward, 2008). Inoltre, come evidenziato in uno studio esposto precedentemente, la propensione a confrontare la propria immagine corporea con i contenuti che vengono visualizzati nei *social media* è correlata a sintomi depressivi ed ansia (Sherlock e Wagstaff, 2019). Un altro modo in cui il confronto sociale si esplicita durante l'utilizzo di Instagram è tramite l'interazione con i propri *follower*. Questo accade fondamentalmente attraverso i *feedback* che gli utenti ricevono in conseguenza ai propri *post* e alle proprie storie. La quantità e tipologia di questi *feedback* funge da parametro per l'autovalutazione dei contenuti caricati che sono collegati alla valutazione di sé, visto che spesso su questo *social media* si tende a presentare immagini di sé stessi.

Parlando di immagine corporea e *social media*, non si può non citare la teoria dell'auto-oggettivazione di Fredrickson e Roberts. Secondo questa teoria l'oggettivazione si verifica quando un soggetto viene valutato esclusivamente in base al proprio corpo ed in relazione alle sue caratteristiche, come se questo fosse rappresentativo di tutta la persona, che viene di conseguenza considerata come un oggetto. Secondo la teoria il corpo, soprattutto quello femminile, tende ad essere oggettivato sessualmente, considerato come un oggetto e valutato a seconda del grado di utilità per gli altri (Fredrickson e Roberts, 1997). Le donne tendono ad interiorizzare gli ideali di bellezza promossi dai pari e dalla società e a valutarsi in relazione a questi da un punto di vista esterno, considerando il proprio corpo unicamente come un oggetto e auto oggettivandosi. In uno studio di Vandenbosch ed Eggermont (2012) è stata analizzata la relazione tra vari media, tra cui videoclip musicali sessualmente oggettivanti e alcuni *social media*, e l'interiorizzazione di ideali di bellezza e l'auto-oggettivazione. I risultati hanno mostrato come l'esposizione ai media sessualmente oggettivanti abbia una relazione indiretta con l'interiorizzazione di ideali di bellezza, ed un'associazione indiretta con l'auto-oggettivazione. Risulta chiaro quindi come vedere immagini rappresentanti dei corpi sessualmente oggettivati porti a sviluppare un'abitudine a valutare gli altri come oggetti, e di conseguenza induce il soggetto ad auto valutarsi allo stesso modo, oggettivandosi maggiormente (Vandenbosch ed Eggermont, 2012). Nello studio di Farper e Tiggemann (2008), si è voluto indagare se l'esposizione a immagini rappresentanti modelli idealizzati di bellezza presenti sui *social media* potesse aumentare l'auto-oggettivazione, l'umore negativo e l'insoddisfazione corporea in giovani donne. I risultati hanno messo in luce come vi sia un significativo aumento dell'auto-oggettivazione in seguito all'esposizione a questo tipo

immagini. Inoltre, è stata riscontrata un'associazione con un declino generale dell'umore e del livello di insoddisfazione corporea, e ad un aumento dell'ansia rispetto al proprio peso. Infine è stato evidenziato come vi siano diversi fattori che possono influenzare l'intensità di questa auto-oggettivazione, per esempio il contesto sociale, caratterizzato dalle interazioni con i pari, ma anche i mass media (Farper e Tiggemann, 2008). Naturalmente quando nel 1997 Fredrikson sviluppò questa teoria non erano ancora presenti i *social media* ed egli infatti parlò in di *mass media*, tuttavia è facile intuire come questa relazione possa essere considerata valida, se non addirittura intensificata, in un *social media* come Instagram. Una meta-analisi di Karsay e colleghi (2018) volta ad indagare la relazione tra l'oggettivazione sessuale in vari media e l'auto-oggettivazione, ha dimostrato come i media *on-line* e i videogiochi portino ad un effetto di auto-oggettivazione più forte rispetto alla televisione. Uno studio di Farduly e Bryde (2017) ha esplorato proprio la relazione tra immagine corporea, auto-oggettivazione ed Instagram in un campione di giovani ragazze americane tra i 18 e i 25 anni. I risultati hanno mostrato come un maggior uso di Instagram sia associato ad una maggiore auto-oggettivazione e come questa relazione sia mediata attraverso il confronto con le *celebrities* seguite. Questi ed altri risultati presentati nello studio suggeriscono come l'utilizzo di Instagram possa influenzare negativamente le credenze e le preoccupazioni riguardo il proprio aspetto.

Risulta quindi chiaro come le caratteristiche specifiche di Instagram lo rendano il *social media* per eccellenza dove si possono innescare dei meccanismi che coinvolgono l'auto-oggettivazione e il confronto sociale, e che possono avere esiti negativi sul benessere delle persone. Gli autori della teoria dell'auto-oggettivazione, infatti, mettono in luce come l'auto-oggettivarsi abbia degli effetti psicologici negativi nella vita delle persone. Diversi studi infatti hanno evidenziato come questo costrutto aumenti la probabilità di sperimentare sentimenti negativi in contesti sociali, come per esempio ansia. Questa deriva dal fatto di pensare che gli altri utenti valutino il proprio corpo come lo stesso soggetto fa e, visto che il soggetto non si sente in linea con i canoni di bellezza a cui tende, si sente negativamente valutato. L'auto-oggettivazione inoltre può essere considerata come un fattore di rischio anche per altri disturbi psicologici come: disturbi depressivi, disfunzioni sessuali o disturbi legati all'alimentazione (Fredrickson e Roberts, 1997).

2. Ricerca e campione

2.1 Ricerca ed obiettivi

Questo lavoro si snoda all'interno di una più ampia ricerca condotta in un campione di giovani donne allo scopo di indagare e meglio comprendere le relazioni tra immagine corporea ed auto-oggettivazione, e l'utilizzo di Instagram. Sui dati che ora verranno presentati sono stati condotti già alcuni studi, mentre altri sono ancora in via di sviluppo. Ai soggetti che hanno deciso di contribuire è stato somministrato un questionario diviso in varie sezioni e scale. Nella parte iniziale del questionario veniva chiesto al partecipante se avesse o meno un *account* Instagram e quanto tempo al giorno dedicasse all'utilizzo di questo *social media*. Successivamente venivano proposte alcune scale: *Instagram addiction scale*, *Instagram activity scale*, *SATAQ (Sociocultural Attitudes Toward Appearance Scale)* e *SOBBS (Self-Objectification Beliefs and Behaviors Scale)*. Vi era poi una scala di controllo sull'ambientalismo, una sezione contenente alcune informazioni demografiche e qualche altra domanda, inerente per esempio all'attività fisica svolta dal soggetto o la sua altezza e peso. Infine, vi era una sezione che spiegava come scaricare i propri dati Instagram per inviarli ai ricercatori.

Come evidenziato negli studi presentati precedentemente, la maggior parte delle ricerche che sono state condotte utilizzando i dati dei *social media*, e quindi escludendo quelle che invece raccolgono dati unicamente tramite questionari, segue una struttura abbastanza tipica. Le variabili dipendenti vengono misurate con uno strumento somministrato ai partecipanti, i quali forniscono i propri dati del *social media* in analisi. Le variabili indipendenti vengono invece estratte tramite delle procedure di *data-mining* dai dati forniti dai partecipanti. A questo punto vi sono due possibilità: la più comune è quella di analizzare la relazione tra le variabili indipendenti e quelle dipendenti cercando delle correlazioni tra queste, un'altra è invece quella di provare a creare dei modelli che permettano di stimare le variabili dipendenti partendo da quelle dipendenti. I risultati nel primo caso sono presentati in termini di correlazioni significative riscontrate, mentre nel secondo caso i modelli vengono valutati in relazione alla loro capacità di predire i dati.

L'obiettivo di questo lavoro è stato quello di analizzare in ottica esplorativa le possibili relazioni tra alcuni costrutti psicologici di nostro interesse e una parte specifica dei dati estratti da Instagram: le immagini. I costrutti sono stati l'Instagram *addiction* e l'immagine corporea, declinata nella teoria dell'auto-oggettivazione. Il primo è stato misurato tramite l'*Instagram Addiction Scale* e l'*Instagram Activity Scale*, mentre il secondo tramite il SATAQ e il SOBBS. La novità di questo studio sta proprio nel provare a prendere in esame direttamente le immagini del *social media*. Com'è stato mostrato precedentemente sono molto poche le ricerche che hanno analizzato questo tipo di dati, preferendo invece dati di tipo testuale. Per estrarre le informazioni dalle immagini sono stati utilizzati degli specifici *software*, creando così una procedura completamente automatizzata, replicabile ed indipendente dai ricercatori. L'obiettivo dell'analisi è stato quello di provare a creare un modello che permettesse di stimare i punteggi alle variabili di interesse partendo dai dati estratti dalle immagini, in un'ottica più predittiva che esplicativa ed in linea con la natura esplorativa e non confermativa di questo lavoro (Grafico 1). Nonostante questo taglio più predittivo il modello, seppur di non facile interpretazione, può darci delle informazione circa l'influenza reciproca delle variabili predittrici nello spiegare quelle di risposta.

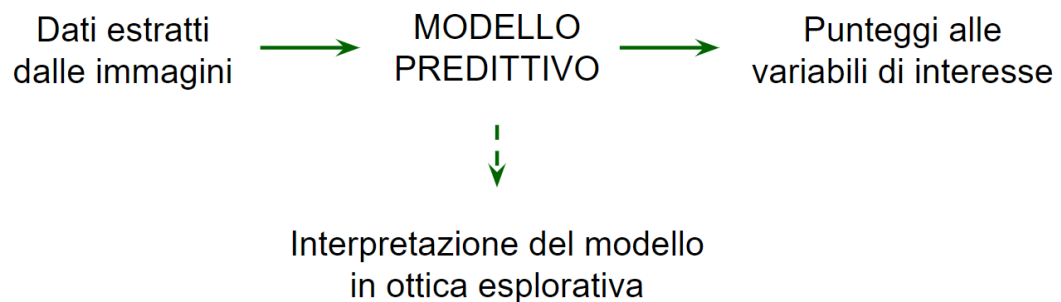


Grafico 1: Schema degli obiettivi della ricerca

2.1.1 Aspetti etici

Vista la natura sensibile di questa tipologia di dati il presente studio è stato esaminato ed approvato da un comitato etico ed è stato spiegato ai partecipanti quali dati sarebbero stati esaminati. Naturalmente all'inizio del questionario era presente un consenso informato, per spiegare ai partecipanti come sarebbero stati utilizzati i dati da loro forniti.

2.2 I dati

Le partecipanti, dopo essere state reclutate, hanno compilato il questionario tramite una piattaforma *on-line* e successivamente inviato, seguendo la procedura, i propri dati Instagram agli sperimentatori. Vengono ora brevemente presentate le varie scale di interesse per la ricerca che erano presenti nel questionario:

2.2.1 Instagram Addiction Scale

L'*Instagram Addiction Scale* è stata sviluppata da Kirkaburun e Griffiths (2018) partendo dall'*Internet Addiction Test* di Young, cambiando la parola *internet* con la parola "Instagram". È stata tradotta in italiano dal gruppo di ricerca che si è occupato della raccolta dati, con una procedura di *back-translation*. Ha lo scopo di misurare la presenza e severità della dipendenza da Instagram tra gli adulti. La scala è composta da quindici *item* divisi in due sottoscale. La prima sottoscala, chiamata "Social Effect", comprende *item* come "Quanto spesso preferisce utilizzare Instagram piuttosto che passare del tempo con i tuoi amici perché ti dà più piacere?" ed indaga i possibili effetti negativi che l'utilizzo di Instagram può avere sulla propria vita sociale. La seconda sottoscala, chiamata "Compulsion", comprende *item* come "Quanto spesso provi a ridurre il tempo che trascorri su Instagram senza successo?" ed indaga la crescita nel bisogno dell'uso di Instagram, il non rendersi conto dello scorrere del tempo mentre si usa questo *social media* e l'utilizzo di Instagram per evadere dai problemi della vita di tutti i giorni. Ogni domanda è riferita ai sei mesi precedenti la compilazione del questionario e prevede una risposta da uno a sei punti su una scala Likert dove 1 equivale a "mai", 2 a "raramente", 3 ad "occasionalmente", 4 a "frequentemente", 5 a "spesso" e 6 a "sempre".

2.2.2 Instagram Activity Scale

L'Instagram Activity scale è stata costruita e validata da Yang per uno studio che aveva l'obiettivo di esplorare l'associazione tra la solitudine e l'uso di Instagram (Yang, 2016). È stata tradotta in italiano dal gruppo di ricerca che si è occupato della raccolta dati, con una procedura di *back-translation*. Lo scopo della scala è quello di misurare la frequenza di specifiche attività svolte durante l'utilizzo di Instagram. La scala è composta da sei *item* ed è stata divisa in tre sottoscale a seguito di un'analisi fattoriale svolta dagli autori (Yang, 2016). Ogni sottoscala, composta da due *item*, rispecchia una tipologia di attività diversa: IG *interaction*, indaga aspetti relativi alla comunicazione con altri utenti durante l'utilizzo di Instagram; IG *browsing*, misura la frequenza degli utenti nell'aggiornare il *feed* e controllare i profili degli altri; IG *broadcasting* che invece si concentra sulle attività che vengono svolte dagli utenti senza interagire direttamente con altri *account*. Gli *item* si riferiscono al periodo dei sei mesi precedenti e prevedono una risposta su una scala Likert che va da 1 (mai) a 5 (molto) che ne determina il grado di accordo.

2.2.3 SATAQ

Il *Sociocultural Attitudes Appearance Questionnaire-3* (SATAQ) è uno strumento la cui terza versione è stata sviluppata in inglese da Thompson e colleghi, allo scopo di rilevare gli atteggiamenti verso le influenze socioculturali sull'immagine corporea e l'interiorizzazione di norme di bellezza (Thompson et al., 2004). La sua versione in italiano è stata tradotta e validata da Stefanile e colleghi (Stefanile et al., 2011). Per la raccolta dati di questa ricerca la versione di Stefanile è stata leggermente modificata, escludendo una sottoscala e cambiando alcuni termini degli *item* per dare un'enfasi sull'utilizzo dei *social media*. Per esempio, l'*item* "ho sentito la pressione della tv e delle riviste ad avere un corpo perfetto" è stato sostituito con l'*item* "ho sentito la pressione della tv e dei *social media* ad avere un corpo perfetto". Nel questionario somministrato in questa ricerca conteneva quindi ventisette *item* divisi in tre sottoscale: "interiorizzazione-generale", "pressione" ed "informazione". Queste sottoscale indagano rispettivamente: come sono state interiorizzate le norme di bellezza a cui si è venuti in contatto attraverso i vari media; la pressione percepita che spinge verso l'adeguamento alle norme di bellezza; ed infine quanto un soggetto attinge dai media le informazioni per costruire

i propri ideali di bellezza (Stefanile et al., 2011). Il SATAQ prevede per ogni *item* una risposta su una scala Likert che ne indica il grado di accordo, dove 1 equivaleva a “molto in disaccordo” e 5 a “molto d’accordo”.

2.2.4 SOBBS

La *Self-Objectification Beliefs and Behaviors Scale* (SOBBS) è una scala creata da Lindner e Tantleff-Dunn, basandosi direttamente sulla teoria dell’auto-oggettivazione di Fredrikson e Roberts, allo scopo di misurare l’auto-oggettivazione di tratto in un soggetto (Lindner e Tantleff-Dunn, 2017). È stata tradotta in italiano dal gruppo di ricerca che si è occupato della raccolta dati, con una procedura di *back-translation*. È una scala composta da quattordici *item* divisi in due fattori: il primo è volto a misurare il grado di interiorizzazione di una prospettiva esterna nell’osservare il proprio corpo (“penso a come il mio corpo appare agli altri anche quando sono da sola”), mentre il secondo rappresenta la misura in cui il corpo di un soggetto contribuisca a definirne il sé, mettendo enfasi sulle emozioni e i pensieri che il soggetto esperisce a riguardo (“come appaio fisicamente è per me più importante di quello che penso o che provo”) (Lindner e Tantleff-Dunn, 2017). Ogni *item* prevede una risposta su una scala Likert a cinque livelli che indica il grado di accordo, 1 equivaleva a “molto in disaccordo” e 5 a “molto d’accordo”.

2.2.5 Informazioni demografiche

Ai fini di indagare le caratteristiche del campione preso in esame, sono state considerate alcune informazioni non appartenenti ad una specifica scala di misurazione: l’età; il genere; l’orientamento sessuale; il livello di istruzione; l’occupazione; il corso di studi e l’anno di università, nel caso la partecipante fosse iscritta ad una facoltà universitaria; ed infine l’altezza ed il peso, per poter calcolare per ogni partecipante l’indice di massa corporea (BMI, *body mass index*), una misura che si è pensato potesse in qualche modo essere collegata al costrutto di immagine corporea ed all’auto-oggettivazione.

2.2.6 I dati di Instagram

L'ultima sezione del questionario era composta dalle istruzioni per poter scaricare ed inviare agli sperimentatori i propri dati di Instagram. Per fare ciò infatti i partecipanti dovevano accedere al proprio profilo Instagram, andare alla sezione "Privacy e Sicurezza" delle impostazioni e premere il comando "scarica dati", dopo aver inserito l'indirizzo *e-mail* a cui volevano inviare i dati. Il *social media* provvedeva così a creare ed inviare dei *link* alla *e-mail* specificata che restavano attivi per quattro giorni ed attraverso i quali era possibile scaricare i propri dati compressi. I dati comprendevano varie cartelle e *file*, in formato JSON (JavaScript Object Notation), che potevano essere annidati in varie sottocartelle, chiamate "part1", "part2", e così via. Al loro interno vi erano delle cartelle contenenti i *file* multimediali, immagini e video: la cartella "direct" conteneva tutti i file multimediali inviati nelle chat private con gli altri *account*, la cartella "photos" tutte le immagini pubblicate sul proprio profilo; la cartella "profile" conteneva l'attuale immagine del profilo di un *account*; la cartella "stories" conteneva invece tutti i video e le immagini pubblicate dai partecipanti nelle proprie *Instagram Stories*; ed infine la cartella "videos" conteneva tutti i video pubblicati da un partecipante sul proprio profilo. All'interno di ognuna di queste cartelle inoltre i *file* erano organizzati in sottocartelle ognuna indicate il mese e l'anno di caricamento di quegli specifici contenuti sul *social media*. I file JSON, presenti all'interno della cartella principale, invece contenevano informazioni circa l'*account*, come: il nome del profilo, l'indirizzo *e-mail* associato, i contatti o alcune altre impostazioni, oppure potevano contenere la traccia di alcune azioni compiute, come: i *like* messi ad altri *account*, i commenti o i messaggi inviati tramite il servizio *direct*. Come si può notare, alcuni di questi dati erano di natura privata, durante la raccolta dati dunque si è deciso di invitare i partecipanti ad eliminare alcuni di questi dati prima di inviarli agli sperimentatori, indicandolo nell'ultima sezione del questionario insieme alle procedure per il *download* dei propri dati. Nello specifico ai partecipanti veniva chiesto di eliminare le informazioni circa i messaggi privati mandati ad altri utenti, e quindi il file "messages.json" e la cartella "direct". Prima di procedere all'invio dei propri dati agli sperimentatori inoltre, veniva chiesto ai partecipanti di rinominare la cartella principale contenente i propri dati con le ultime quattro cifre del proprio numero di cellulare, per proteggere così la *privacy* di ogni partecipante e creare un codice identificativo casuale.

2.3 Il campione

Il campione di questa ricerca è composto da 242 soggetti tutti di sesso femminile, reclutati su base volontaria. Le partecipanti hanno un'età che va dai 18 ai 33 anni, con un'età media di 23 anni ($sd = 2.4$). Nel Grafico 2 si può notare la distribuzione dell'età nel campione. La maggior parte delle partecipanti ha affermato di essere iscritta ad un corso di studi universitario, e poco meno della metà delle studentesse ha già conseguito una laurea triennale. Delle iscritte all'università inoltre circa il 40% risulta frequentare la facoltà di Psicologia. Per informazioni più dettagliate si rimanda alla Tabella 3, nella quale si trovano altre due informazioni circa il campione: l'orientamento sessuale e la situazione sentimentale. Da notare che quello di questa ricerca non è un campione probabilistico.

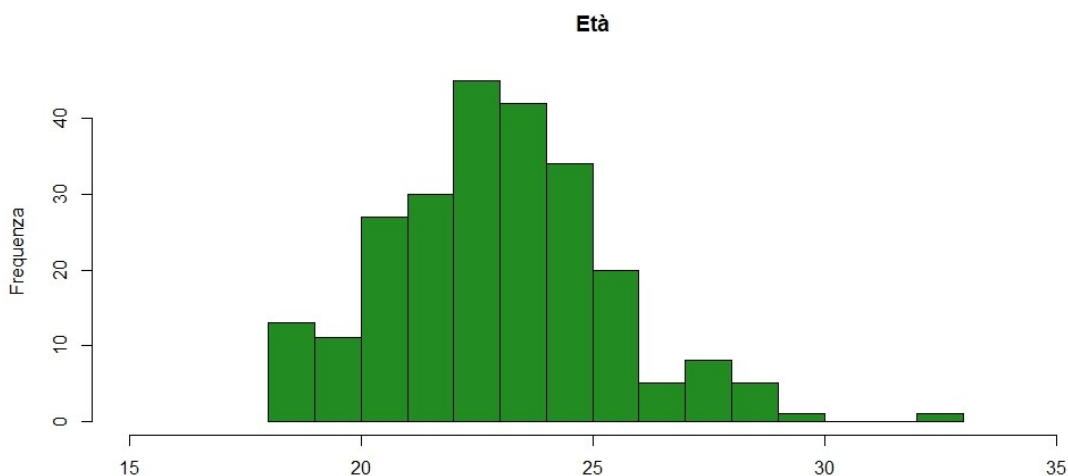


Grafico 2: Distribuzione dell'età nel campione

<i>Genere</i>	Maschile	0 /242	0%
	Femminile	241 /242	99.59%
	Non precisato	1 /242	0.41%
<i>Orientamento sessuale</i>	Eterosessuale	220 /235	93.61%
	Omosessuale	5 /235	2.12%
	Bisessuale	9 /235	3.82%
	Altro	1 /235	0.42%
<i>Situazione sentimentale</i>	Single	82 /242	33.88%
	Appuntamenti casuali	5 /242	2.00%
	Impegnato in una relazione	142 /242	58.67%
	Preferisco non rispondere	13 /242	5.37%
<i>Livello di istruzione</i>	Licenza media	4 /242	1.65%
	Diploma superiore	98 /242	40.50%
	Laurea triennale	106 /242	43.80%
	Laurea magistrale	32 /242	13.22%
	Altro	2 /242	0.83%
<i>Occupazione</i>	Studente	144 /242	59.50%
	Lavoratore	48 /242	19.83%
	Studente/Lavoratore	42 /242	17,36%
	Disoccupato	8 /242	3.31%
<i>Corso di studi</i>	Psicologia	87 /227	38.32%
	Altri	140 /227	61.67%
<i>Anno di università</i>	Primo	21 /195	10.77%
	Secondo	27 /195	13.85%
	Terzo	42 /195	21.54%
	Quarto	31 /195	15.90%
	Quinto	38 /195	19.49%
	Sesto	9 /195	4.62%
	Preferisco non rispondere	27 /195	13.85%

Tabella 3: Informazioni circa il campione

3. Costruzione delle variabili

3.1 Variabili dipendenti

Per quanto riguarda le variabili dipendenti, le risposte dei soggetti ai questionari sono state condensate nei punteggi alle varie scale e sottoscale.

- Il punteggio per l'*Instagram Addiction Scale* è stato calcolato facendo una media delle risposte ad ogni *item*, senza tenere in considerazione le due sottoscale. (ADDICTION)
- Per quanto riguarda l'*Instagram Activity Scale* invece sono stati calcolati tre punteggi, ognuno riferito ad una delle sottoscale, facendo la media degli *item* per ogni sottoscala. (ACTRIVITY1, ACTIVITY2, ACTIVITY3)
- Per ogni sottoscala del SATAQ invece è stato calcolato l'alpha di Cronbach, un indicatore statistico basato sulla covarianza degli *item*, e la media di ogni sottoscala è stata moltiplicata per questo alpha, in modo da dare maggior peso nel punteggio finale alla scala più consistente. È stato quindi ottenuto un unico punteggio a tutta la scala calcolando la media di tutte le sottoscale. (SATAQ)
- Il punteggio del SOBBS è stato ottenuto in modo del tutto analogo a quello del SATAQ, sempre cercando di soppesare il punteggio di una sottoscala con la propria consistenza. (SOBBS)

Sono state così ottenute sei variabili di riposta per ogni soggetto.

3.2 Le immagini di Instagram

In questo lavoro si è deciso di prendere in esame solamente una parte dei dati di Instagram inviatici dalle partecipanti: le immagini pubblicate sul proprio profilo. Sono state tenute separate per ogni partecipante l'immagine del profilo dalle immagini caricate sul profilo. La prima è quella che viene visualizzata quando si cerca un utente all'interno del *social media*, ed è possibile vederla indipendentemente se questi abbia un profilo pubblico o privato. La stessa immagine inoltre viene visualizzata in miniatura in alto a sinistra sulla pagina personale di un utente, nella sezione "direct" degli utenti con cui è stata iniziata una

conversazione, e nella miniatura delle *Instagram Stories* che appare nella parte alta della “bacheca”. Questa immagine può essere aggiornata in qualsiasi momento dal proprietario del profilo, e una volta cambiata non resta traccia dell’immagine precedente. Le immagini caricate sul profilo invece sono quelle che ogni utente può decidere di postare sul *social media* in diversi momenti; possono venir visualizzate nella “bacheca” e restano disponibili nel profilo dell’utente che le ha caricate. Da notare che il profilo di un utente può essere “pubblico” o “privato”, ed è lo stesso utente che lo decide. Per poter vedere le immagini caricate da un utente con un profilo privato bisogna mandare una richiesta di *following* ed aspettare che il proprietario del profilo la confermi, se questo non avviene non possono essere visualizzate le immagini caricate sul suo profilo. Nell’eventualità l’utente abbia un profilo pubblico invece le immagini caricate sul suo profilo possono essere visualizzate senza necessariamente aver mandato precedentemente a quest’ultimo una richiesta di *following*.

In totale le immagini del profilo e quelle caricate sul profilo di tutti i partecipanti allo studio erano circa 75000.

3.2.1 Data-mining

Per poter comprendere il processo di *data-mining* che ha permesso l’estrazione di informazioni analizzabili dalle immagini di Instagram vanno introdotti alcuni concetti.

Il *Machine Learning*, conosciuto in italiano come apprendimento automatico, è una branca dell’intelligenza artificiale che raccoglie vari metodi: statistica computazionale, riconoscimento di *pattern*, reti neurali artificiali, *data-mining*, elaborazione delle immagini, ed altri ancora. In generale questi sono metodi statistici che permettono di creare un algoritmo in grado di riconoscere dei *pattern* all’interno di certi dati, e successivamente di migliorarne sempre più le *performance* di riconoscimento. Il concetto base è quindi quello di fornire ad un algoritmo una grande quantità di dati dello stesso tipo, come testi, immagini o audio, e fare in modo che impari a riconoscere all’interno di questi dati delle caratteristiche ricorrenti. A ciò si può arrivare in due modi: o attraverso l’“Apprendimento Supervisionato”, o attraverso l’“Apprendimento Non-Supervisionato”. Nel primo caso i dati vengono precedentemente etichettati a seconda delle loro caratteristiche. Tipicamente questo viene fatto manualmente

da coloro i quali si occupano di addestrare l'algoritmo, ma si possono anche usare uno o più algoritmi precedentemente addestrati per associare ad ogni dato la sua etichetta corrispondente. L'algoritmo, addestrandosi, cercherà di apprendere a suddividere i dati nelle varie categorie. Da notare che questa tipologia di modelli effettuano ipotesi induttive, arrivando a fornire risposte probabilistiche. Ad ogni categorizzazione infatti viene associato un numero che va da zero ad uno che indica la percentuale di accuratezza. Le etichette inoltre possono anche essere numeri, e non unicamente variabili categoriali, in questo caso si parlerà di un algoritmo di regressione e non di uno di classificazione. L'algoritmo di regressione apprenderà quindi a stimare un valore per ogni dato, al posto di associarlo ad una determinata categoria. L'apprendimento non-supervisionato invece si basa su dati che non sono stati in alcun modo precedentemente codificati, lo scopo dell'algoritmo è quindi quello di creare delle categorie e suddividere i dati in queste, a seconda di caratteristiche comuni che riesce a trovare. Questo tipo di apprendimento si può dividere in due categorie: *dimensionality reduction*, come la PCA (*principal component analysis*); e *clustering*. Quest'ultimo è una delle tecniche più comuni basate su questa metodologia, e viene utilizzato allo scopo di rilevare similarità strutturali nei dati durante l'analisi statistica e di suddividerli in gruppi tra loro omogenei. Normalmente viene gestito da due parametri: un errore che l'algoritmo ha lo scopo di diminuire il più possibile, ed un parametro di semplicità che permette di preferire una soluzione meno complessa anche se con un errore maggiore. Di fatto la soluzione con l'errore minore è quella in cui ogni dato è diviso in una categoria separata, la complessità di questa soluzione però è massima, mentre la soluzione con minore complessità prevede un'unica categoria all'interno della quale vengono messi tutti i dati, l'errore però in questo caso sarebbe massimo.

Il *Deep Learning* è una sottocategoria degli algoritmi di *machine learning*, la sua peculiarità è quella di occuparsi di elaborare dati molto complessi ed estrarre quelle che vengono definite caratteristiche ad alto livello. Con questa metodologia si addestrano degli specifici algoritmi che vengono chiamati reti neurali, per analogia di funzionamento con i neuroni umani. Queste reti neurali infatti sono divise in vari livelli: l'informazione accede alla rete tramite il livello di *input*, successivamente passa per vari livelli interni alla rete i quali la analizzano e modificano ad ogni passaggio, per poi arrivare al livello finale dove viene prodotto un *output*. Una sottocategoria di rete neurale profonda (*deep neural network*) sono le

Convolutional Neural Networks, che vengono spesso usate per l'analisi di specifici dati complessi: le immagini.

La *Computer Vision*, nota in italiano col nome di visione artificiale, è un campo scientifico interdisciplinare che si occupa di come i computer possano sviluppare delle buone capacità di comprensione di immagini digitali o video. Lo scopo principale è quello di provare a riprodurre la vista umana, cercando di estrarre delle informazioni da immagini e video. I principali compiti della *computer vision* sono tre: *image classification*, *object recognition* e *object detection*. Vengono presentati in dettaglio nella Tabella 4.

<i>Image Classification</i>	L'immagine, analizzata olisticamente da un algoritmo, viene assegnata ad una classe corrispondente, con un'associata percentuale di correttezza. L'algoritmo, tipicamente una rete neurale, è stato precedentemente allenato e sono già state definite tutte le possibili classi. Si può quindi capire qual è il contenuto dell'immagine.
<i>Object Recognition</i>	Partendo da un'immagine contenente vari oggetti, l'algoritmo è in grado di identificare dove sono questi oggetti, e delimitarne i contorni, tipicamente tramite un rettangolo.
<i>Object Detection</i>	Un algoritmo individua la posizione di tutti gli oggetti presenti in un'immagine e cerca di capire cosa essi siano, classificandoli. In pratica l'immagine viene scomposta in varie sottoimmagini tramite algoritmi di <i>object recognition</i> , e successivamente ognuna di queste sottoimmagini viene classificata da un algoritmo di <i>image classification</i> . Va notato che ad ogni oggetto viene associata una percentuale di correttezza, che corrisponde a quanto il modello è sicuro di aver trovato quello specifico oggetto. Le reti neurali che si occupano di <i>object detection</i> sono le più complesse e sono in grado di riconoscere un gran numero di oggetti diversi. Questi modelli permettono quindi di rispondere alla domanda: "dove sono gli oggetti e quali sono?" (Zou et al., 2019). Per l'elaborazione dei dati di questa ricerca è stato utilizzato questo metodo. Nell'Immagine 1 viene mostrato un esempio di <i>object detection</i> .

Tabella 4: I principali compiti della computer vision



Immagine 1: Esempio di object detection
Fonte: EfficientDet⁵

3.2.2 Object Detection

Come si può immaginare vi sono moltissimi modelli per l'*object detection*, questi possono essere divisi in due macrocategorie: *two-stage* e *one-stage detectors*. Nel primo caso il modello divide il processo di ricerca di oggetti in varie fasi, mentre nel secondo il riconoscimento di oggetti viene completato in un unico passaggio tramite l'utilizzo di una unica rete (Zou et al., 2019). Tra i più utilizzati ed implementati appartenenti alla seconda tipologia si trova: YOLO, *You Only Look Once*, diventato celebre grazie alla velocità di elaborazione delle immagini, 155 *frame* per secondo, che ha permesso di riconoscere oggetti in tempo reale e quindi anche durante la stessa riproduzione di un video (Redmon et al., 2016); SSD, *Single Shot Detector*, che ha implementato la fase di riconoscimento degli oggetti rispetto a YOLO, arrivando ad un riconoscimento di oggetti più veloce e preciso (Liu et al., 2016); ed anche *Retina-net*, creato allo scopo di migliorare ancora la predizione in questi modelli estremamente veloci (Lin et al., 2017).

⁵ (2021, 17 gennaio) <https://tfhub.dev/tensorflow/efficientdet/d3/1>

Va tenuto in considerazione comunque che la velocità massima raggiunta da questi modelli dipende anche dalle specifiche *hardware* della macchina da cui sono analizzati i dati. Per questo tipo di compiti infatti l'utilizzo di una comune CPU (*Central Processing Unit*) richiederebbe molto tempo, quasi indipendentemente dalla potenza di questa. Per spiegare in gergo non tecnico questo fenomeno, si può dire che le CPU sono in grado di compiere molto velocemente calcoli complessi, tuttavia hanno un processamento dell'informazione di tipo seriale, cosa che permette di fare pochi calcoli simultaneamente, a seconda del numero di *core*. Perciò per l'applicazione di questi modelli vengono tipicamente utilizzate le GPU (*Graphics Processing Units*), che nonostante siano più lente delle CPU a svolgere un singolo calcolo, sono in grado di fare molti calcoli in parallelo. Sotto questo punto di vista si può notare come il funzionamento delle GPU sia più simile al funzionamento del cervello umano. Non è quindi un caso il fatto che le GPU abbiano prestazioni migliori nell'utilizzo di questi modelli, la cui architettura, come nel caso delle *Deep Neural Network*, si basa sul funzionamento del cervello umano. Sono stati inoltre sviluppati da Google degli specifici processori per l'utilizzo di reti neurali, e più nello specifico per la libreria *open source* TensorFlow: i TPU (*Tensor Processing Units*), che hanno delle prestazioni superiori anche alle GPU per questi compiti. Non tutti i pc comunque hanno una scheda grafica, e probabilmente non sono in commercio pc che montano delle TPU, per questo motivo utilizzare queste reti neurali su dei comuni pc può risultare più dispendioso a livello di tempo. Google ha sviluppato anche un ambiente *on-line*, chiamato Colaboratory (o più semplicemente Colab), che permette di scrivere programmi in linguaggio di Python, salvando i *notebook* su GoogleDrive. Ciò che rende questo ambiente molto interessante è che il codice viene eseguito sui server di Google, e non in locale sulla macchina, cosa che permette di utilizzare gli *hardware* dei server, naturalmente con certe limitazioni. Per l'analisi delle immagini di questo lavoro è stato utilizzato, come vedremo più avanti, questo ambiente *on-line* e le GPU dei server di Google.

I modelli citati precedentemente rappresentano la struttura portante su cui si basa il compito dell'*object detection*, ma tutti devono essere allenati su un *dataset*, cioè un insieme di decine di migliaia di immagini già etichettate. Le caratteristiche di questi *dataset* dipendono dallo scopo per cui le reti neurali vengono create, sono molti e svariati infatti gli ambiti in cui iniziano ad essere utilizzate. Per esempio vi sono *dataset* specifici per l'agricoltura, lo sport,

immagine mediche o di videosorveglianza⁶. Uno dei *dataset* più utilizzati è COCO (*Common Objects in COntext*), composto da oltre trecentomila immagini contenenti uno o più oggetti appartenenti ad ottanta categorie diverse⁷. Proprio su questo *dataset* è stata allenata una delle reti neurali utilizzate in questo lavoro per analizzare le immagini estratte da Instagram.

Le reti neurali e i modelli di *object detection* vengono usati in vari ambiti. Ne vengono elencati ora alcuni per dare al lettore una panoramica generale sulle interessanti applicazioni di questa tecnologia. Sono utilizzati per esempio per compiti di videosorveglianza, per la quale viene adoperata la tecnologia del *object tracking*; sono utilizzati nell'astronomia, sia per il riconoscimento di oggetti astronomici come galassie o stelle, ma anche per la elaborazione di dati complessi. Uno dei più celebri esempi è stato quello della creazione della prima immagine di un buco nero, per lo sviluppo della quale sono stati utilizzati anche dei modelli di *deep learning*. Vengono utilizzati inoltre anche in ambito medico, come aiuto per la diagnosi di varie malattie, come per esempio i tumori; ed anche per lo sviluppo di sistemi complessi come quello della guida autonoma.

3.2.3 Data-mining in questa ricerca

Com'è stato specificato precedentemente, i dati di Instagram inviati dai partecipanti alla ricerca comprendevano varie informazioni oltre a quelle di interesse per questo lavoro. Tramite l'utilizzo di Python, un *software* di programmazione, sono stati creati dei programmi in grado di estrarre, dai dati annidati in varie cartelle e sottocartelle, le immagini e qualche informazione in più inerente a queste. La struttura delle cartelle contenenti i dati di ogni partecipante non era sempre la stessa ma dipendeva dalla quantità dei dati Instagram del soggetto. Inoltre, alcuni partecipanti, nell'eliminare i dati privati da non inviare agli sperimentatori, hanno alterato la struttura dei propri dati. Per riuscire a trovare all'interno delle varie cartelle tutte le immagini di interesse per questo lavoro sono quindi stati utilizzati i file JSON, dei file di tipo testuale di cui si è parlato precedentemente. All'interno del file "media.json" sono presenti alcune informazioni tra cui il nome di ogni immagine caricata sul

⁶ (2021, 17 gennaio) <https://pub.towardsai.net/50-object-detection-datasets-from-different-industry-domains-1a53342ae13d>

⁷ (2021, 17 gennaio) <https://cocodataset.org>

profilo, con associato l'orario e la data di caricamento, e la *caption*, cioè la didascalia che si può allegare all'immagine quando viene postata sul *social media*. Per ogni partecipante è stata quindi creata una lista contenente tutti i nomi delle immagini caricate sul profilo, e tramite questa lista le immagini sono state individuate nelle varie sottocartelle. Anche per recuperare l'immagine del profilo per ogni soggetto è stata utilizzata una procedura analoga. Infine ad ogni immagine caricata sul profilo è stata associata la data di caricamento sul *social media*, informazione chiave per l'analisi, come si vedrà in seguito.

Tutte queste immagini sono state processate da reti neurali, già allenate, per l'*object detection*. Per questa ricerca, dopo alcuni tentativi con varie reti neurali, ne sono state selezionate due: EfficientDet e NudeNet.

- EfficientDet⁸ è una rete neurale per l'*object detection* con una architettura SSD, sviluppata da Google e disponibile su TensorFlow. È allenata sul *dataset* COCO 2017, ed è in grado di riconoscere circa ottanta oggetti diversi. In *input* riceve un'immagine trasformata in un tensore, che sostanzialmente è una matrice a più dimensioni, e fornisce in *output* una lista di oggetti trovati, con associata una percentuale di accuratezza e la posizione dello specifico oggetto all'interno dell'immagine.
- NudeNet⁹ invece è una rete neurale disponibile su GitHub, un sito dove gli sviluppatori condividono i loro *software* ed interagiscono con la *community* per implementarli. È stata sviluppata allo scopo di identificare immagini contenenti nudità e censurarle, per questo è in grado di riconoscere varie parti del corpo, sia coperte che esposte. È stata allenata con centosessantamila immagini, classificate precedentemente tramite varie tecniche ibride e tramite l'utilizzo di altre reti neurali. In *input* ed *output* riceve e fornisce oggetti del tutto analoghi alla rete neurale descritta precedentemente. Tuttavia, vista la natura pornografica delle immagini di addestramento, gli autori specificano come il modello identifichi le varie parti del corpo coperte, unicamente in posizioni definite come "provocatorie".

Queste reti neurali e le immagini sono state caricate tramite Google Colaboratory, un'ambiente di programmazione online in linguaggio Python, che ha permesso di eseguire il

⁸ (2021 17 gennaio) <https://tfhub.dev/tensorflow/efficientdet/d3/1>

⁹ (2021, 17 gennaio) <https://github.com/notAI-tech/NudeNet>

codice sul *Cloud* utilizzando dei potenti *hardware*. L'insieme delle 75000 immagini utilizzate in questo lavoro è stato analizzato usando entrambi i modelli e sono stati così ricavati i risultati dell'*object detection*. Dalla lista di oggetti che possono essere identificati tramite EfficientDet sono stati eliminati quelli che non sono mai stati trovati all'interno delle immagini di Instagram analizzate in questo lavoro. I rimanenti oggetti sono poi stati raggruppati in varie categorie. Alcuni oggetti trovati da NudeNet sono stati scartati, visto che per la natura dei dati di questa ricerca questi risultavano essere dei "falsi positivi". Provando per esempio a cercare certe parti del corpo esposte all'interno di questi dati, dove chiaramente non ci si aspetta di trovarne, quando l'algoritmo troverà qualcosa sarà unicamente dovuto ad un errore. Infine la percentuale di accuratezza degli oggetti trovati per entrambe le reti neurali è stata scelta arbitrariamente a 0.6, soglia che corrisponde all'opzione di *default* della rete EfficientDet. Tutti gli oggetti trovati dalle reti con una precisione inferiore a questa soglia sono stati quindi scartati. Nella Tabella 5 si possono vedere i vari oggetti identificati da entrambe le reti, con le categorie a cui appartengono.

Ad ogni immagine è stata associata una variabile conteggio corrispondente al numero di volte in cui uno specifico oggetto è stato individuato. Successivamente tale variabile è stata trasformata sommando tutti gli oggetti appartenenti alla stessa categoria, ad eccezione delle parti del corpo trovate da NudeNet, per le quali si è deciso di tenere sia la somma di queste nella categoria "Body Part", sia ogni oggetto separatamente. Si è passati così da una variabile relativa a 79 oggetti trovati, ad una di 21 oggetti per ogni immagine.

EfficientDet	Person	Person
	Accessories	Backpack, Handbag, Suitcase, Tie
	Animal	Bear, Bird, Cat, Cow, Dog, Elephant, Giraffe, Horse, Sheep
	Food	Apple, Banana, Broccoli, Cake, Carrot, Donut, Hot Dog, Orange, Pizza, Sandwich
	House	Bed, Chair, Couch, Dining Table, Potted Plant, Toilet
	Kitchen	Bottle, Bowl, Cup, Fork, Knife, Spoon
	Other	Book, Hair Drier, Microwave Oven, Refrigerator, Scissors, Sink, Teddy Bear, Toaster, Toothbrush
	Sport	Baseball Bat, Baseball Glove, Frisbee, Kite, Skateboard, Skis, Snowboard, Sports Ball, Surfboard, Tennis Racket
	Street	Bench, Fire Hydrant, Parking Meter, Stop Signal
	Tech	Cell Phone, Keyboard, Laptop, Mouse, Remote
	Transport	Airplane, Bicycle, Boat, Bus, Car, Motorcycle
NudeNet	Body Part	Covered Belly, Covered Breast F, Covered Feet, Covered Genitalia F, Exposed Armpits, Exposed Belly, Exposed Feet, Face F, Face M

Tabella 5: *Classi ed oggetti dell'object detection*

3.2.4 Creazione degli indici

Prendendo spunto da una ricerca condotta da Hatton e Trautner (2011) che aveva l'obiettivo di indagare la sessualizzazione di uomini e donne nelle copertine del *Rolling Stones*, famosa rivista statunitense di cultura *pop*, sono stati creati degli indici, pensando che tipo di relazione potessero avere la presenza o assenza di certi oggetti all'interno delle immagini di Instagram. Una volta definiti questi indici sono stati calcolati per ognuna delle immagini esaminate.

- *index_personORnot*: indice che può assumere valore zero o uno. Assume valore uno quando viene individuata almeno una persona all'interno dell'immagine, e valore zero quando non ve ne è presente nemmeno una. Il motivo per cui è stato creato questo indice è che per ogni soggetto ci sono varie immagini, e per ogni immagine vari oggetti trovati; dato che l'unità statistica di questo lavoro sono i partecipanti, i risultati delle immagini dovranno essere condensati tramite la media. In questo caso allora la media delle persone trovate nel profilo di una partecipante avrà probabilmente un risultato diverso rispetto alla media di questo indice, e verranno così rispecchiate due informazioni diverse.
- *index_onebodyNOface*: anche in questo caso un indice dicotomico, che può assumere valore zero ed uno. Assume valore uno quando è presente nell'immagine un'unica persona ma non viene trovato un volto, ed assume valore zero se questa condizione non si verifica. Questo indice è stato creato tenendo in considerazione la teoria dell'auto-oggettivazione di Fredrickson e Roberts. Ci si potrebbe infatti aspettare che mettere in mostra il proprio corpo, non mostrando il volto possa avere una certa relazione con questo costrutto.
- *index_mirrorselfie*: un indice che assume valore uno quando sono presenti da una a tre persone nell'immagine ed un cellulare, ed assume valore zero nel caso questa condizione non si verifichi. Questo indice riesce ad identificare quando l'immagine contiene un *selfie* scattato allo specchio, in quanto questa è la circostanza più comune in cui si trova sia un cellulare che un numero ristretto di persone in questi dati. Anche in questo caso l'indice è stato creato per una presunta possibile relazione con le variabili dipendenti.

3.3 Variabili indipendenti

Dopo aver estratto le informazioni dalle immagini sono stati costruiti due *dataset* contenenti quelle che diventeranno le variabili predittive dei modelli: uno per le immagini caricate sul profilo, ed uno per le immagini del profilo. A fine capitolo si può trovare in uno schema i principali passaggi svolti per la creazione delle variabili indipendenti (Grafico 3).

3.3.1 Immagini caricate sul profilo

Per ogni soggetto sono stati tenuti in considerazione i risultati delle immagini pubblicate su Instagram nei dodici mesi precedenti alla compilazione del questionario. Questa scelta è stata motivata dalle caratteristiche delle variabili di risposta. Com'è stato precedentemente evidenziato la propria immagine corporea si crea in adolescenza e può cambiare nel tempo a seconda di vari fattori, interni ed esterni alla persona. Per questo motivo per prendere retrospettivamente in esame un comportamento che potrebbe essere collegato a questo costrutto, è necessario scegliere un arco di tempo adeguato. Inizialmente si era pensato di prendere i dati inerenti ai sei mesi precedenti alla compilazione del questionario, gli *item* dell'*Instagram Activity Scale* e dell'*Instagram Addiction Scale* per esempio fanno riferimento a questo specifico arco di tempo. Tuttavia visto che la raccolta dei dati è stata svolta in un periodo di circa quattro mesi, tra ottobre 2019 e febbraio 2020, i sei mesi precedenti alla compilazione del questionario sarebbero stati diversi per ogni soggetto. È facile intuire come le caratteristiche delle immagini caricate sul *social media* possano dipendere dal mese e dalla stagione in cui vengono scattate e postate, perciò si è deciso di tenere in considerazione i dati dei dodici mesi precedenti alla compilazione del questionario, annullando così le possibili differenze dovute ad una variazione stagionale nei contenuti delle immagini di Instagram. I risultati di ognuna delle immagini pubblicate nei dodici mesi precedenti alla compilazione del questionario sono stati condensati in un unico vettore calcolando la media per ogni variabile estratta, e quindi la media per ogni oggetto presente nelle immagini di Instagram. Successivamente questi valori sono stati standardizzati per una maggior semplicità di interpretazione dei risultati. A queste variabili inoltre ne sono state aggiunte altre due: il numero di immagini pubblicate negli ultimi dodici mesi e il BMI (indice di

massa corporea), di cui si era interessati a vedere la possibile influenza con le variabili dipendenti. Si è arrivati così ad ottenere 25 variabili indipendenti per ciascun soggetto (Tabella 6).

Variabili indipendenti per ciascun soggetto	
<i>Oggetti</i>	Accessories, Animal, Food, House, Kitchen, Other, Sport, Street, Tech, Transport
<i>Parti del corpo</i>	Person, Body Part, Covered Belly, Covered Breast F, Covered Feet, Covered Genitalia F, Exposed Armpits, Exposed Belly, Exposed Feet, Face F, Face M, index_personORnot, index_onebodyNOface, index_mirrorselfie
<i>Altre</i>	Number of Images, BMI

Tabella 6: Classi ed oggetti dell'object detection

Per ridurre ulteriormente il numero di variabili e semplificare l'interpretazione dei risultati si è deciso di fare una PCA (*principal component analysis*) con rotazione *varimax*. L'analisi delle componenti principali è una tecnica statistica utilizzata per la semplificazione dei dati. Lo scopo è quello di ridurre il numero di variabili osservate mediante un numero inferiore di variabili di sintesi. Le variabili sulle quali è stata condotta la PCA sono: Covered Breast F, Covered Feet, Covered Genitalia F, Exposed Armpits, Exposed Belly, Exposed Feet, Face F, Face M, Body Part, Person, index_onebodyNOface, index_mirrorselfie e index_personORnot. Sono state così ridotte a 5 che spiegano il 73% della varianza osservata. La prima è stata rinominata "pca_person", in quanto era principalmente composta da: Face M, Face F, Body Part, Person, Index_personORnot, tutte variabili collegate alla presenza o meno di una persona all'interno dell'immagine. La seconda è la terza sono state chiamate "pca_nudity1" e "pca_nudity2", in quanto erano composte dalle parti del corpo trovate nell'immagine. "pca_nudity1" da Covered Genitalia F, Exposed Belly ed Exposed Feet. Invece "pca_nudity2" da Covered Breast F ed Exposed Armpits. Sarebbe che questa divisioni rispecchi la parte alta e la parte bassa del corpo. La quarta variabile estratta dalla PCA è stata rinominata "pca_selfie", in quanto composta principalmente da un'unica variabile,

l'index_mirrorselfie. Mentre la quinta è stata chiamata "pca_noface", ed analogamente alla precedente è composta principalmente da index_onebodyNOface. Nella Tabella 7 vengono mostrate più nel dettaglio le composizioni delle variabili estratte tramite la PCA. Al fine di semplificare la comprensione della tabella, sono stati riportati unicamente i valori dei coefficienti del modello maggiori per ogni variabile. Si è passati così da venticinque a diciassette variabili indipendenti, si rimanda alla Tabella 9 a fine capitolo per una visione globale di quali siano.

	pca_person	pca_nudity1	pca_nudity2	pca_selfie	pca_noface
<i>Person</i>	0.90				
<i>Body Part</i>	0.91				
<i>Face F</i>	0.91				
<i>Face M</i>	0.66				
<i>Exposed Armpits</i>			0.87		
<i>Covered Breast</i>			0.65		
<i>Exposed Belly</i>		0.56			
<i>Covered Genitalia F</i>		0.77			
<i>Covered Feet</i>	0.37				
<i>Exposed Feet</i>		0.71			
<i>Index_personORnot</i>	0.81				
<i>Index_mirrorselfie</i>				0.90	
<i>Index_onebodyNOface</i>					0.93
<i>Varianza cumulativa</i>	0.29	0.42	0.54	0.64	0.73

Tabella 7: Risultati della PCA per le immagini caricate sul profilo

3.3.2 Immagine del profilo

Per le immagini del profilo la procedura è stata più semplice. Ogni partecipante ha infatti un'unica immagine del profilo, quelle precedenti non vengono salvate nei dati di Instagram, e quindi non è stato necessario fare una media tra i dati estratti dalle varie immagini o tenerne in considerazione solo quelle di un certo periodo di tempo. Inoltre, non è stato possibile risalire alla data di caricamento dell'immagine del profilo sul *social media*, quindi non sarebbe comunque stato possibile filtrarle per un certo periodo di tempo. I risultati dell'*object detection* non sono stati standardizzati visto che rappresentano direttamente il numero degli oggetti trovati nell'immagine e non una media come per le immagini caricate sul profilo. Anche su questi dati è stata svolta un PCA che ha portato a dei risultati leggermente diversi a quella precedente, riassumendo tredici variabili in cinque e spiegando il 68% della variabilità. La prima variabile è stata chiamata "pca_nudity1", è del tutto analoga a quella descritta precedentemente e sembra comprendere principalmente le parti del corpo che si trovano nella zona bassa del corpo: ventre, genitali e piedi. Una seconda variabile è stata chiamata "pca_faceF", composta principalmente dalla variabile Face F, e quindi dai volti femminili, e da index_onebodyNOface la quale però è legata negativamente. La terza variabile è stata chiamata "pca_nudity2" ed è anche questa analoga a quella descritta nel paragrafo precedente ma con una differenza: in questa sono state condensate anche le informazioni inerenti alla variabili Body Part ed index_mirrorselfie. La quarta variabile è stata chiamata "pca_person" ed è leggermente diversa da quella descritta per le immagini caricate sul profilo, infatti comprende unicamente le variabili Person e Index_personORnot. L'ultima variabile è stata chiamata invece "pca_faceM", in quanto comprende principalmente la variabile Face M. In Tabella 8 sono riportati maggiori dettagli. Anche in questo caso per ogni variabile è stato riportato il valore dei coefficienti maggiore, al fine di semplificarne la comprensione. Il numero di variabili predittive è stato ridotto così a quindici. Si rimanda alla Tabella 9 a fine capitolo per una visione globale di quali siano. Da notare che la categoria "Tech" è stata sostituita con la categoria "Tech_Cell" in quanto l'unico oggetto appartenente a questa categoria trovato tra le immagini del profilo è stato appunto il telefono cellulare.

	pca_nudity1	pca_faceF	pca_nudity2	pca_person	pca_faceM
<i>Person</i>				0.77	
<i>Body Part</i>			0.60		
<i>Face F</i>		0.86			
<i>Face M</i>					0.94
<i>Exposed Armpits</i>			0.76		
<i>Covered Breast</i>			0.71		
<i>Exposed Belly</i>	0.68				
<i>Covered Genitalia F</i>	0.85				
<i>Covered Feet</i>				0.30	
<i>Exposed Feet</i>	0.76				
<i>Index_personORnot</i>				0.77	
<i>Index_mirrorselfie</i>			0.58		
<i>Index_onebodyNOface</i>		- 0.92			
<i>Varianza cumulativa</i>	0.18	0.34	0.49	0.60	0.68

Tabella 8: Risultati della PCA per le immagini del profilo

	<i>Immagini caricate sul profilo</i>	<i>Immagine del profilo</i>
Oggetti	Accessories	Accessories
	Animal	Animal
	Food	Food
	House	House
	Kitchen	Kitchen
	Other	Other
	Sport	Sport
	Street	Street
	Tech	Tech_Cell
	Transport	Transport
PCA	pca_nudity1	pca_nudity1
	pca_nudity2	pca_nudity2
	pca_person	pca_person
	pca_selfie	pca_faceF
	pca_noface	pca_faceM
Altri	num_images	
	BMI	

Tabella 9: Variabili indipendenti

Data-Mining

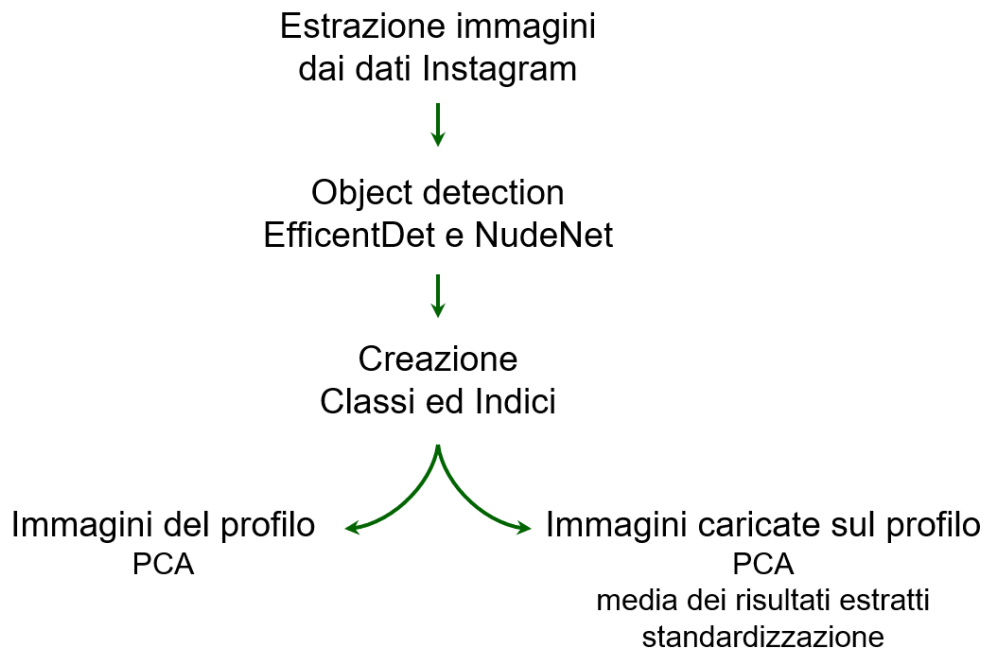


Grafico 3: Processo di costruzione delle variabili indipendenti

4. Analisi dati

4.1 Alcune statistiche descrittive

Vengono ora proposti tre grafici per descrivere la distribuzione di alcune variabili. Per quanto riguarda le variabili di riposta è stato creato un grafico tramite la funzione `pairs.panels` della libreria `psych` di R (Grafico 4). Si può notare, in linea con le considerazioni teoriche esposte precedentemente, una correlazione tra la variabile `SATAQ` e la variabile `SOBBS`. Si possono inoltre notare le distribuzioni delle variabili dipendenti: la variabile `ADDICTION` sembra avere una distribuzione asimmetrica negativa; le variabili `ACTIVITY2` e `ACTIVITY3` una distribuzione asimmetrica positiva; le restanti variabili invece sembra abbiano una distribuzione abbastanza simmetrica.

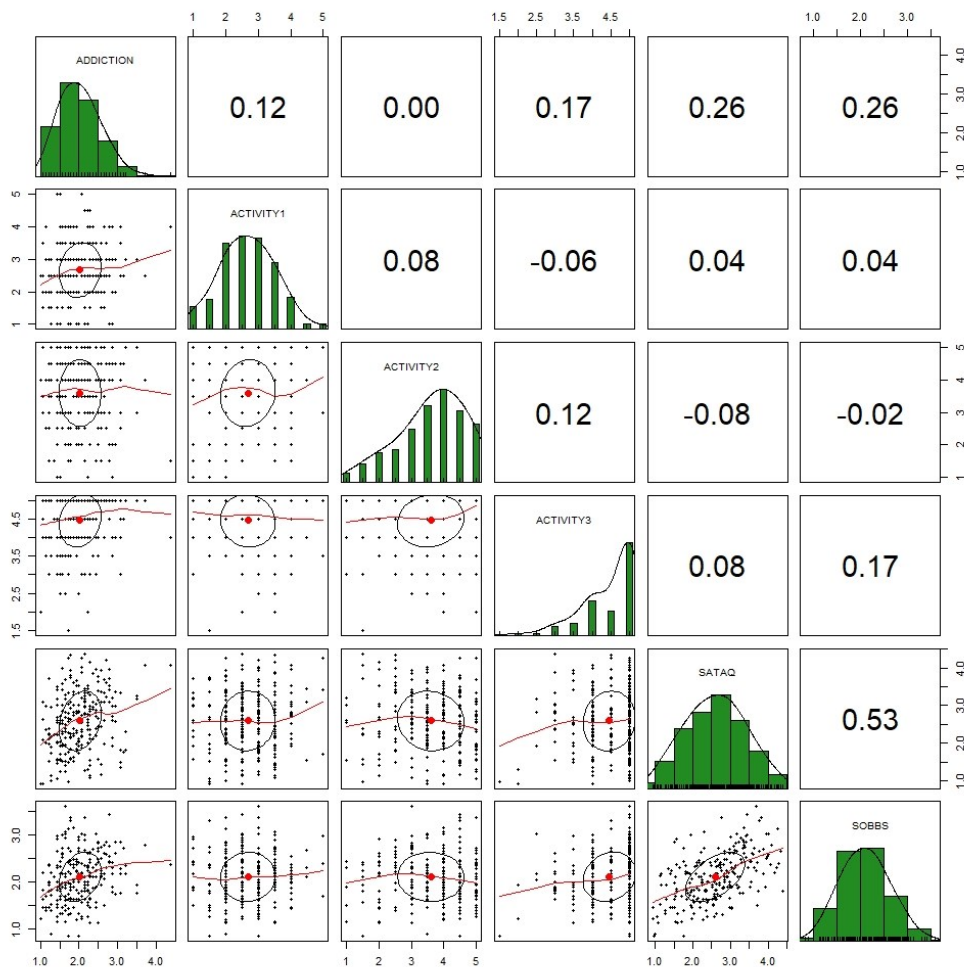


Grafico 4: Distribuzione delle variabili dipendenti

Per quanto riguarda invece le variabili predittrici sono stati creati due grafici: uno per la variabile indice di massa corporea (Grafico 5), ed uno per la variabile numero di immagini (Grafico 6). A differenza delle altre variabili estratte tramite le procedure di object detection, queste non sono state normalizzate. Conoscere la loro distribuzione quindi può aiutare ad interpretare meglio i risultati ed i modelli.

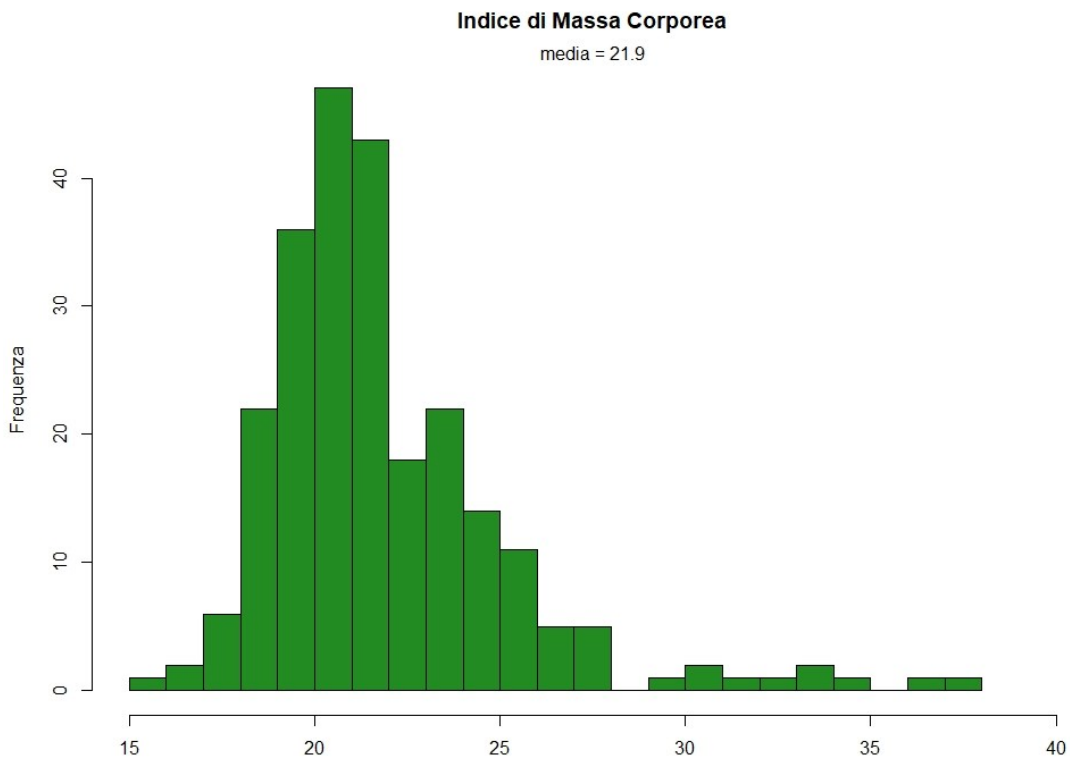


Grafico 5: Distribuzione della variabile "BMI"

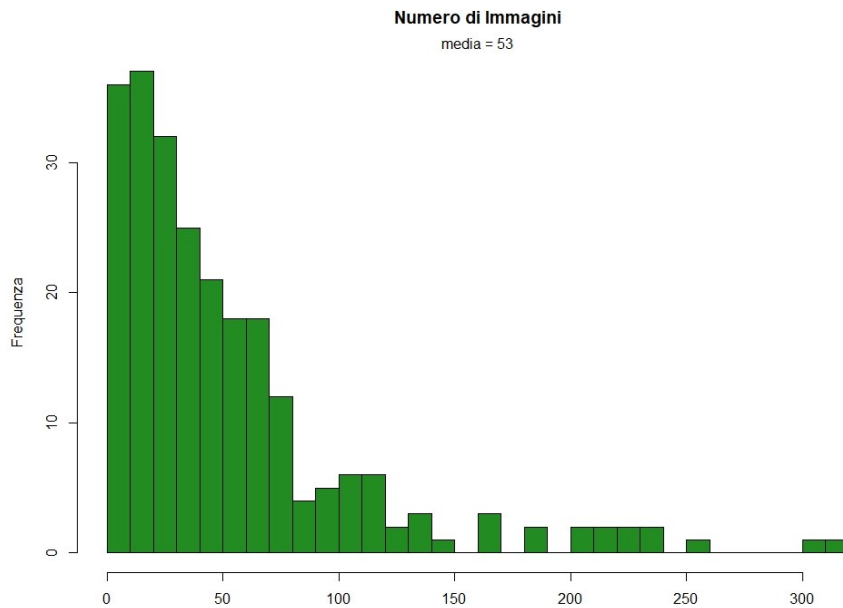


Grafico 6: Distribuzione della variabile "num_images"

4.2 Aspetti teorici dell'analisi

4.2.1 Predire o spiegare i dati?

Dalla ricerca condotta da Yarkoni e Westfall (2017) dell'università di Austin dal titolo *"Choosing Prediction Over Exoplanation in Psychology: Lessons From Machine Learning"* sono emerse delle considerazioni interessanti riguardanti il modo in cui tradizionalmente viene condotta la ricerca in ambito psicologico. Per comprendere meglio la scelta delle analisi svolte in questo lavoro vengono ora descritte brevemente queste considerazioni. Storicamente la ricerca in psicologia si è occupata di capire il comportamento umano, provando a spiegarlo e predirlo, due obiettivi che raramente vengono considerati separatamente (Yarkoni e Westfall, 2017). Da un punto di vista statistico tuttavia si può già notare come questi due obiettivi siano separati, un modello che descrive in modo estremamente preciso i dati infatti non sarà il modello migliore nel predirne dei nuovi. Questo sarà dovuto ad un fenomeno conosciuto come *overfitting*, chiamato in italiano "sovradattamento". Per maggiori informazioni riguardo questo fenomeno si rimanda a Hawkins (2004). Gli autori mettono in evidenza come i fenomeni studiati in psicologia non è detto che siano sufficientemente semplici da essere

spiegati da un modello comprensibile. I modelli infatti potrebbero essere o troppo complessi per poi risultare comprensibili, o molto semplici ed eleganti nello spiegare un comportamento senza tuttavia essere in grado di predirlo (Yarkoni e Westfall, 2017). Per questo motivo nella pratica della ricerca in psicologia si cerca di scegliere una via di mezzo tra questi due obiettivi, con un focus più o meno accentuato su spiegazione o predittività. Gli autori inoltre sostengono come storicamente il focus quasi esclusivo sullo spiegare le cause di un comportamento, abbia portato una buona parte del campo psicologico ad essere popolato da programmi di ricerca che sviluppano complesse teorie di meccanismi psicologici, ma che hanno una scarsa o sconosciuta capacità nel predire comportamenti futuri (Yarkoni e Westfall, 2017). Prendendo in considerazione alcuni aspetti dal *machine learning* vengono proposte delle osservazioni sull'importanza di dare un taglio predittivo oltre che esplicativo nella ricerca in psicologia, e viene evidenziato come grazie alle recenti scoperte in quest'ambito si possano ottenere sempre migliori risultati da un punto di vista predittivo (Yarkoni e Westfall, 2017). L'osservazione degli autori non vuole essere in alcun modo una critica al passato, ma una riflessione circa la possibilità di condurre nuove tipologie di analisi con questi mezzi, arrivando anche ad ottenere risultati interessanti laddove, con i metodi di analisi più tradizionali, non se ne otterrebbero. Inoltre, vi sono ambiti in psicologia dove un focus sulla predizione è più auspicabile che quello esplicativo, come per esempio in certi ambiti della psicologia del lavoro o della psicologia clinica (Yarkoni e Westfall, 2017). Il non tenere in adeguata considerazione il focus sulla predizione ha contribuito, secondo gli autori, a molti dei problemi riscontrati nella ricerca psicologica negli ultimi anni. Secondo Yarkoni e Westfall quindi anche nei casi in cui il focus principale sia quello di provare a spiegare certi comportamenti, i metodi e concetti estrapolati dal *machine learning* possono contribuire a costruire modelli migliori e risultati più solidi e replicabili. Un modello perciò dovrebbe essere innanzitutto valutato per la sua capacità di predire nuovi risultati e successivamente interpretato per poter teorizzare la relazione tra le variabili da cui è composto.

Per i motivi descritti fin ora, in questo lavoro si è deciso di utilizzare dei modelli di analisi complessi con un focus principale sulla predizione. Naturalmente va tenuto in considerazione la natura esplorativa di questo lavoro, dovuta anche alle caratteristiche del campione, non adatto a poter generalizzare dei risultati. La scelta di queste analisi è suggerita anche dalla tipologia dei dati presi in esame, ovvero quelli di un *social media*. Come si evince

dalle varie ricerche presentate nel primo capitolo, le analisi con un focus predittivo sono molto utilizzate in questo ambito.

4.2.2 Alberi di regressione

Gli alberi di classificazione e di regressione sono dei metodi di *machine learning* utilizzati per costruire modelli predittivi (Loh, 2011). Sono chiamati così poiché vengono solitamente rappresentati con dei grafici che ricordano degli alberi. Si parte da una radice, ogni biforcazione dei rami rappresenta un valore di una certa variabile predittiva, fino ad arrivare alle foglie che corrispondono ai valori della funzione. Tra i metodi non parametrici per stimare una funzione, uno dei più semplici è sicuramente quello della funzione a tratti, e su questo si basano gli alberi di regressione. Lo spazio della variabile indipendente viene suddiviso in vari intervalli e ad ognuno di questi viene assegnato un certo valore della variabile dipendente. Il ragionamento è analogo per quando si hanno più variabili indipendenti (Immagine 2). Quando si stima una funzione a tratti vanno prese tre decisioni: in quante aree bisogna suddividere lo spazio delle variabili indipendenti, come scegliere i punti di divisione di tali variabili, e che valore della variabile dipendente assegnare a quel sottospazio (Azzalini e Scarpa, 2012). L'ultima scelta da compiere è la più semplice, normalmente infatti il valore che viene assegnato alla variabile dipendente è la sua media in quello specifico intervallo. Le prime due invece risultano più complesse e vengono determinate da una funzione di costo-complexità influenzata da un parametro non negativo. La funzione prende in considerazione il numero di sottoinsiemi delle variabili indipendenti e la varianza della variabile dipendente che viene spiegata dalla funzione a tratti, e sceglie una funzione a tratti che spiega maggiormente la variabile dipendente, suddividendo il meno possibile lo spazio delle variabili indipendenti (Azzalini e Scarpa, 2012).

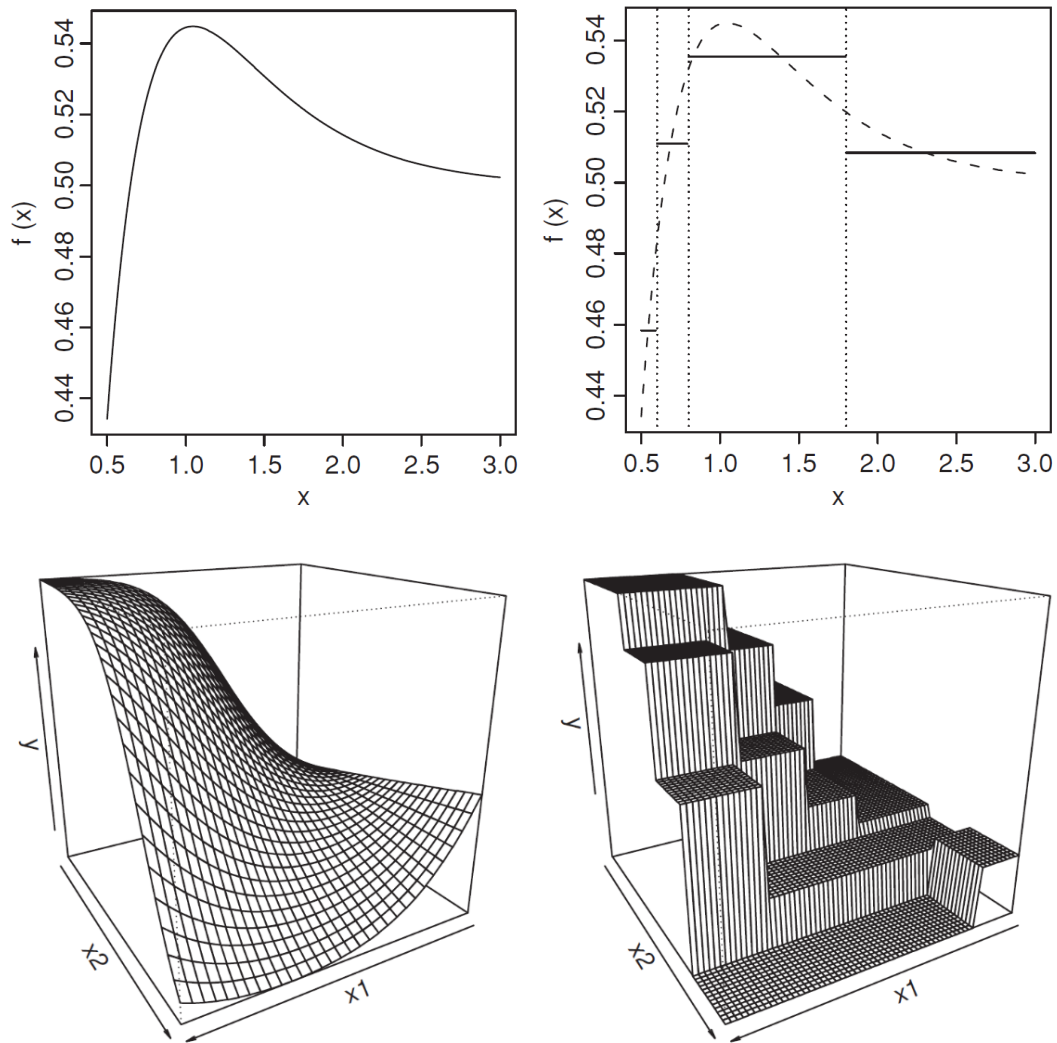


Immagine 2: Funzioni continue con una e due variabili indipendenti, e loro approssimazioni tramite una funzione a tratti.

Fonte: Azzalini e Scarpa (2012)

Una funzione a tratti può essere anche rappresentata con un albero binario, cosa che diventa particolarmente comoda quando si ha a che fare con più di due variabili indipendenti. Le componenti dell'albero vengono chiamate nodi: si inizia ad esaminare l'albero dal primo nodo più in alto, chiamato "radice", se la disuguaglianza espressa nel nodo è vera si prosegue verso destra, altrimenti a sinistra. Reiterando questo processo per il numero di volte necessario si arriva all'ultimo nodo, chiamato "foglia", che corrisponde al valore della variabile dipendente per il sottospazio descritto dalle varie disuguaglianze (Azzalini e Scarpa, 2012). Nell'Immagine 3 si può notare un esempio di albero binario, che rappresenta la stessa funzione a due variabili indipendenti presentata precedentemente nell'Immagine 2.

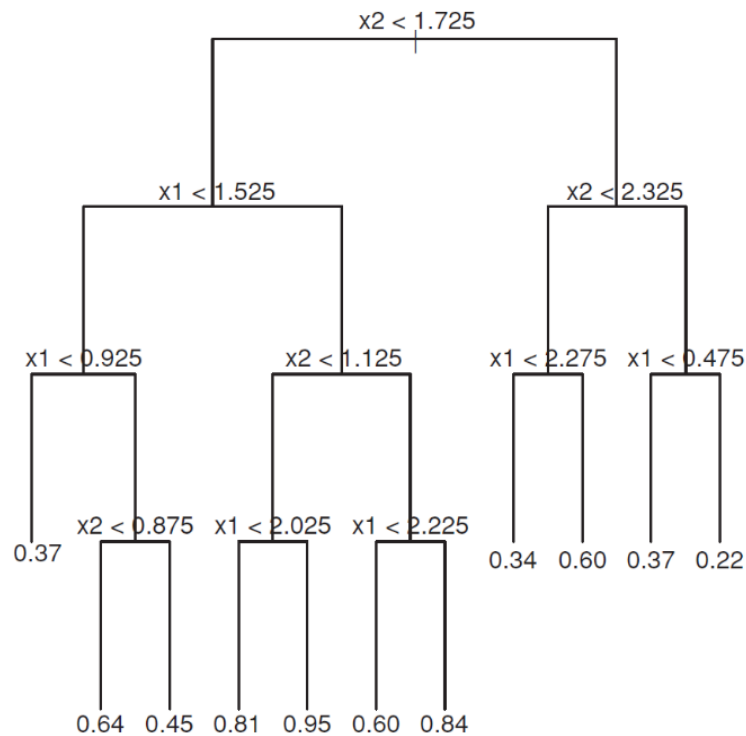


Immagine 3: Esempio di un albero di regressione con due variabili indipendenti
 Fonte: Azzalini e Scarpa (2012)

Questi alberi possono quindi essere utilizzati per stimare un modello utile a predire una variabile dipendente avendo più variabili indipendenti con relazioni tra loro complesse. Da notare che, avendo raccolto i dati da un campione di numerosità n , ci si aspetta che il numero di foglie dell'albero sia di gran lunga inferiore alla numerosità campionaria (Azzalini e Scarpa, 2012). Un albero con esattamente n foglie sarebbe in grado di spiegare in modo perfetto i dati raccolti ma non avrebbe grandi capacità di predire dei nuovi risultati, questo fenomeno è conosciuto come *overfitting*. Per ovviare questo problema il *dataset* viene diviso in due parti, una per il *training* ed una per *testing*. Il modello viene quindi stimato sulla parte di *training* ma valutato su quella di *testing*, valutando quindi fin da subito la capacità predittiva del modello. Può anche essere utilizzata una procedura chiamata *k-fold cross validation*, sostanzialmente simile a quella ora descritta, ma più raffinata in quanto il campione viene suddiviso in k sottogruppi cui quali vengono stimati k modelli utilizzando le $k-1$ parti del campione in qualità di *training* e l'ultima parte in qualità di *testing*. Un'altra soluzione comune per il problema dell'*overfitting* è utilizzare un approccio di analisi chiamato "random forests" (Cox et al., 2020). Di questo approccio parleremo in dettaglio in seguito.

Una considerazione finale che è opportuno specificare è che la posizione delle variabili, che nell'albero appaiono associate ad ogni nodo, non ne indica l'importanza: una variabile importante è quella la cui presenza può contribuire a ridurre significativamente la varianza non spiegata della variabile dipendente, tuttavia questa informazione non può essere estratta direttamente da un grafico di un albero di regressione (Azzalini e Scarpa, 2012). Se una variabile non appare mai in un albero di regressione si può però dire che non contribuisca in alcun modo a predire i dati (Azzalini e Scarpa, 2012).

4.2.3 Random Forest

Come accennato precedentemente uno dei vantaggi degli alberi di regressione e di classificazione è che sono in grado di catturare delle interazioni complesse tra variabili (Hastie et al., 2009). Sono inoltre abbastanza semplici da creare, utilizzare ed interpretare, seppur a volte comprendere in che modo le variabili si influenzino non sia per nulla facile. Hanno però uno svantaggio, rispetto ad altri metodi sono meno accurati nella predizione, per quanto lavorino bene con i loro dati di creazione. I *random forest*, in italiano "foresta casuale", al contrario invece sono molto più flessibili, ed hanno una capacità predittiva maggiore. Una foresta casuale combina al suo interno molti alberi decisionali e li consulta tutti per predire un risultato, facendo una media delle risposte nel caso di un compito di regressione, oppure scegliendo la classe più rappresentata in un compito di classificazione. Il fatto interessante è che seppur singolarmente le predizioni di ogni singolo albero della *random forest* potrebbero essere meno accurate, rispetto a quelle di un singolo albero di regressione, combinate insieme daranno dei risultati migliori (Hastie et al., 2009). Ogni albero infatti è diverso, creato partendo da un campione di dati derivante da un ricampionamento ottenuto tramite una tecnica statistica chiamata *bootstrap*. In pratica consiste nel selezionare casualmente n osservazioni dal campione iniziale, dove n è la numerosità campionaria, creando così un nuovo campione in cui saranno presenti un sottoinsieme di osservazioni del campione originale. In questo modo nel nuovo campione non saranno presenti tutte le osservazioni del campione di partenza, normalmente infatti circa una osservazione su tre viene esclusa, mentre alcune si ripeteranno. L'albero creato partendo da questo nuovo *dataset* inoltre viene costruito prendendo in considerazione solo un certo numero di variabili scelte casualmente per ogni nodo. Questo numero di variabili è fisso per tutta la foresta, ma vengono create varie foreste

variando questo numero, per poi tenere quella con una miglior capacità predittiva. Per valutare la capacità predittiva vengono utilizzate le osservazioni escluse durante il processo di creazione dei campioni. Il numero di alberi da creare invece viene scelto precedentemente, normalmente aumentando questo numero le predizioni dovrebbero diventare sempre migliori, anche se superata una certa soglia si arriva ad una sorta di asintoto orizzontale, dove all'aumentare del numero di alberi la capacità predittiva della *random forest* non migliora.

Va tenuto in considerazione che le foreste casuali sono di più difficile interpretazione rispetto agli alberi di regressione. Permettono di capire quali sono le variabili che hanno contribuito maggiormente a predire i dati, quantificando la loro importanza in termini relativi rispetto alle altre. Tuttavia, comprendere la direzionalità delle interazioni tra queste variabili non è per nulla semplice, va ricordato infatti che questi metodi riescono a cogliere delle relazioni complesse e non lineari tra le variabili predittrici e quelle di riposta.

4.3 Gli step dell'analisi

Le analisi di questo lavoro sono state condotte tramite il software R e le foreste casuali sono state implementate grazie all'utilizzo del pacchetto *randomForest*. Per entrambi i *dataset*, quello delle immagini del profilo e quello delle immagini caricate sul profilo, e per ognuna delle sei variabili dipendenti sono stati condotti i seguenti passaggi:

- Il *dataset* iniziale è stato suddiviso in due parti, una di *training* ed una di *testing*, assegnando casualmente il 75% delle osservazioni al *dataset* di *training* e il restante 25% al *dataset* di *testing*.
- Prendendo come base il *dataset* di *training*, sono state create tante foreste casuali quante le variabili indipendenti, ognuna con cinquecento alberi di regressione. Ogni foresta aveva un numero diverso di variabili estratte casualmente ad ogni nodo, come nella procedura descritta precedentemente.
- È stata selezionata la foresta casuale con la capacità predittiva migliore, calcolata quantificando l'errore di predizione sui dati del *dataset* di *test*, mentre sono state scartate le restanti.

- Per ogni modello di predizione, cioè per ogni foresta casuale, è stata valutata l'importanza di ogni variabile indipendente nel predire la variabile di risposta, calcolando per ogni variabile l'indice di purezza del nodo.
- In seguito è stata esplorata graficamente la possibile presenza di un'associazione lineare tra la variabile dipendente e le nove variabili predittive considerate più importanti all'interno del modello.
- Per entrambi i *dataset* è stato creato un grafico che mette in relazione i valori predetti con i valori osservati per ogni variabile dipendente. Successivamente per ogni modello si è cercato di quantificare la bontà di predizione calcolando: il *mean absolute percentage error* (MAPE) ed il *mean absolute error* (MAE).
- Infine, a scopo illustrativo, per ogni *dataset* e variabile di risposta è stato selezionato all'interno della foresta casuale l'albero più semplice e con un minor errore di predizione, ed è stato visualizzato graficamente.

5. Risultati

5.1 Immagini del profilo

Per valutare la bontà di predizione dei modelli per ogni variabile dipendente sono stati creati dei grafici di dispersione mettendo in ascissa i valori osservati della variabile di risposta e in ordinata i valori predetti dal modello. È stato calcolato per questi il MAPE e il MAE per avere anche numericamente un indicatore della bontà di predizione. In una situazione ideale, se il modello predicesse perfettamente i dati, questi si distribuirebbero sulla bisettrice e quindi su una linea retta inclinata a quarantacinque gradi passante per l'origine del piano. Come si può vedere dal Grafico 7 i punti si distribuiscono su una retta orizzontale, ciò è dovuto al fatto che il modello tende a predire lo stesso valore per ogni soggetto, e che quindi non è stata trovata alcuna relazione tra le variabili estratte dalle immagini del profilo e le variabili di risposta.

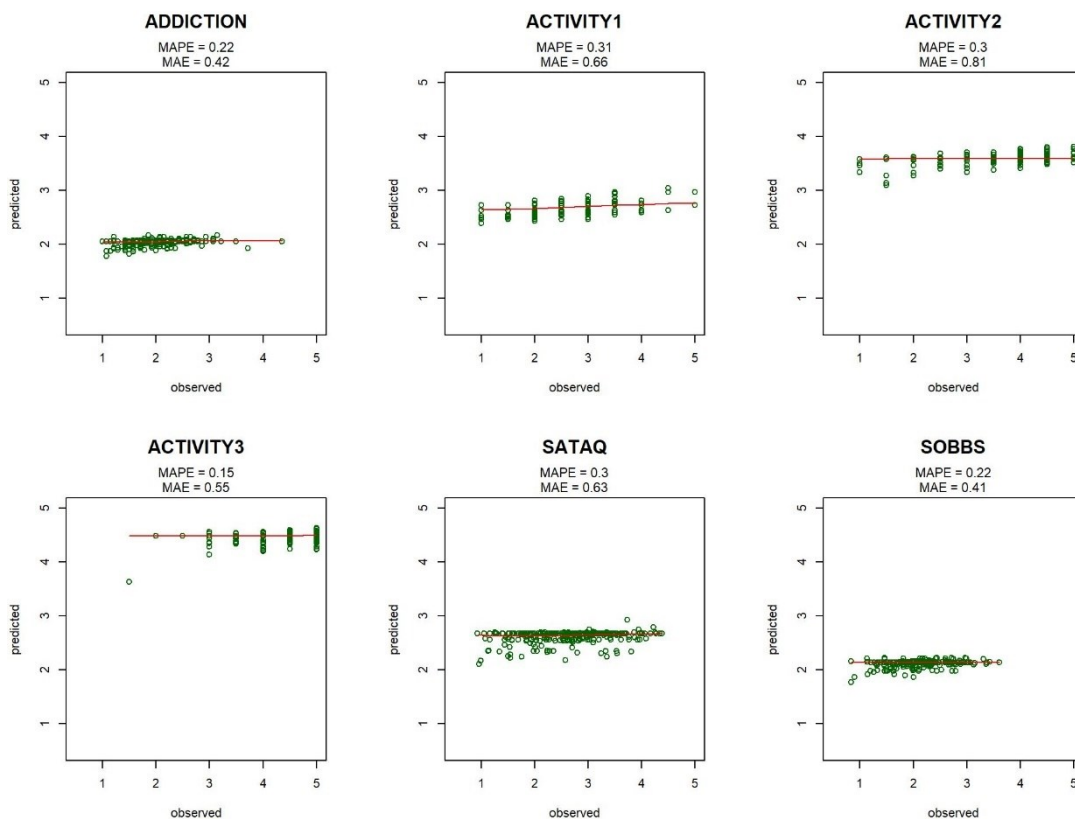


Grafico 7: Immagini del profilo:
valori osservati vs valori predetti per le sei variabili dipendenti

Si può di conseguenza concludere che ciò che i partecipanti hanno pubblicato come propria immagine del profilo di Instagram non abbia alcuna relazione con ciò che riportano sul proprio utilizzo del *social media*, con la propria immagine corporea e l'interiorizzazione delle norme di bellezza, e con quali ritengono essere i propri livelli di auto-oggettivazione.

5.2 Immagini caricate sul profilo

Come per i dati estratti dalle immagini del profilo, sono stati creati dei grafici di dispersione per mettere in relazione i valori osservati delle variabili dipendenti con quelli predetti dal modello, ed è stato calcolato MAPE ed MAE (Grafico 8). A differenza di quanto però osservato prima, ora si può notare come la nuvola di punti si distribuisca seguendo una retta con una certa inclinazione, anche se in tutti i casi inferiore ai quarantacinque gradi. Si può quindi dedurre come il modello riesca a stimare i punteggi alle variabili di riposta, cosa che indica una certa relazione tra le scale misurate e i contenuti delle immagini che i partecipanti hanno caricato sul proprio profilo nei dodici mesi precedenti alla compilazione del questionario. Confrontando i valori del MAPE per le diverse variabili si è in grado di capire quale tra i modelli ha la miglior capacità di predizione dei dati. Tramite il MAE invece, meno adatto per il confronto tra modelli, si può valutare quale sia l'errore di predizione medio in termini di punteggio alla scala. In pratica quindi di quanti punti in media si discosta il valore predetto da quello osservato.

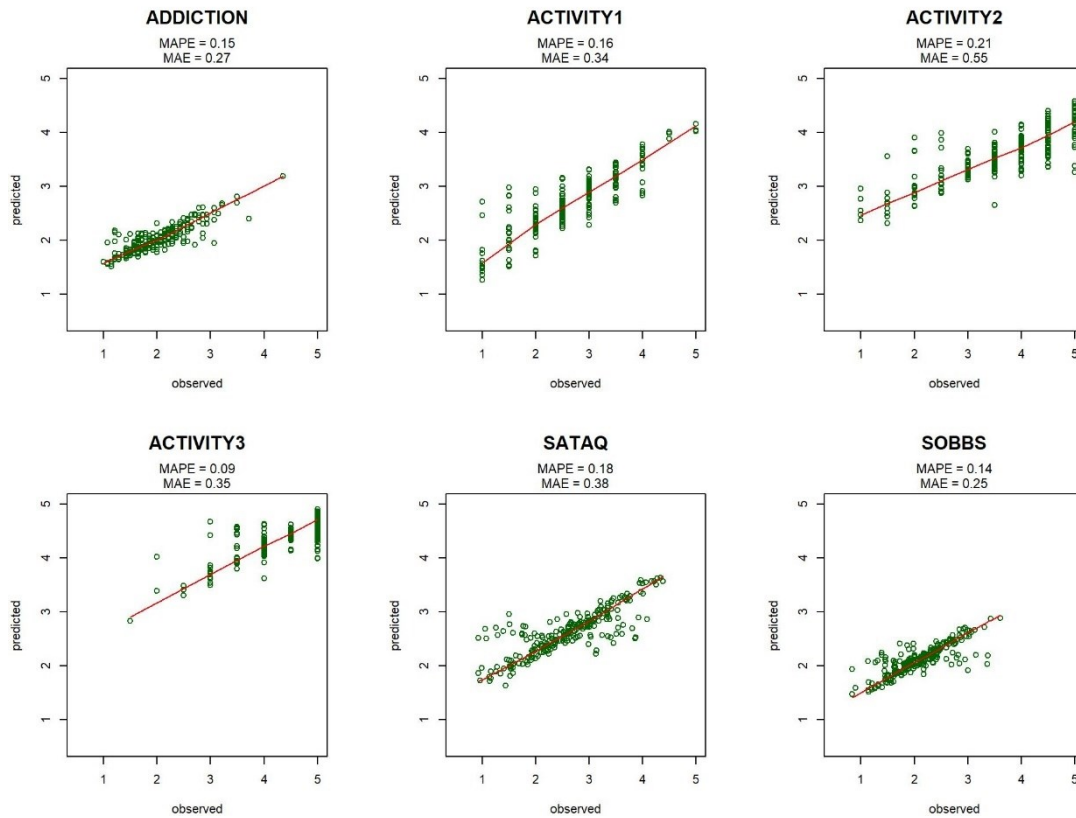


Grafico 8: Immagini caricate sul profilo:
valori osservati vs valori predetti per le sei variabili dipendenti

Ora verranno esplorate più in dettaglio le relazioni individuate dal modello tra le variabili predittrici e quelle di risposta.

5.2.1 Instagram Addiction Scale

Si ricorda che questa scala ha l'obiettivo di misurare la dipendenza da Instagram di un soggetto, ed indaga nelle sue due sottoscale i possibili effetti negativi dell'uso di Instagram e la crescita del bisogno dell'uso di questo *social media*. Dal Grafico 9 si può notare come non vi sia una variabile indipendente il cui impatto sul modello sia significativamente maggiore rispetto alle altre. Da un punto di vista teorico non ci si aspettava ci fosse una relazione tra i contenuti delle immagini postate e la dipendenza da Instagram, questo potrebbe spiegare perché effettivamente nel modello non vi sia una specifica variabile o un gruppo di queste la cui importanza emerga rispetto alle altre. Tra le variabili comunque si trova al primo posto "pca_selfie", che si ricorda essere principalmente composta da un indice che rappresenta il numero di autoscatti davanti allo specchio; al secondo invece "pca_nudity2", una variabile che indica la nudità nella parte alta del corpo; mentre al terzo si trova "tech", che rappresenta una specifica categoria di oggetti tra cui cellulare, laptop e tastiere. Quest'ultima categoria è sicuramente in qualche modo collegata alla prima, in quanto l'indice su cui si basa "pca_selfie" prende in considerazione la presenza di volti di persone e quella di un cellulare. L'aspetto interessante da notare è come in assenza della variabile "pca_selfie" sarebbe difficile capire perché la categoria "tech" risulti essere importante nel predire i punteggi all'*Instagram Addiction Scale*, mentre grazie a questa si sa che è a causa della presenza dei cellulari nelle immagini. Questo sottolinea l'importanza della creazione degli indici, che permettono di sviluppare una comprensione del modello da un punto di vista psicologico e non unicamente statistico. Al quarto e quinto posto invece si trovano rispettivamente le variabili "pca_person", che indica la presenza o meno di persone all'interno delle immagini, e "BMI" cioè l'indice di massa corporea. Va comunque ricordato, come anche per le successive variabili di risposta, il fatto che il modello dica quanto una variabile indipendente sia importante nel predire la variabile di risposta, ma non dia informazioni circa una eventuale direzionalità di questa relazione. Ciò significa che non necessariamente all'aumentare del numero di *selfie* presenti nel profilo aumenti il punteggio di dipendenza da Instagram. Nel Grafico 10 inoltre si può notare come non vi sia una relazione lineare tra nessuna delle variabili predittrici e la variabile dipendente. Il modo in cui i dati estratti dalle immagini concorrono a spiegare la dipendenza da Instagram è probabilmente complesso.

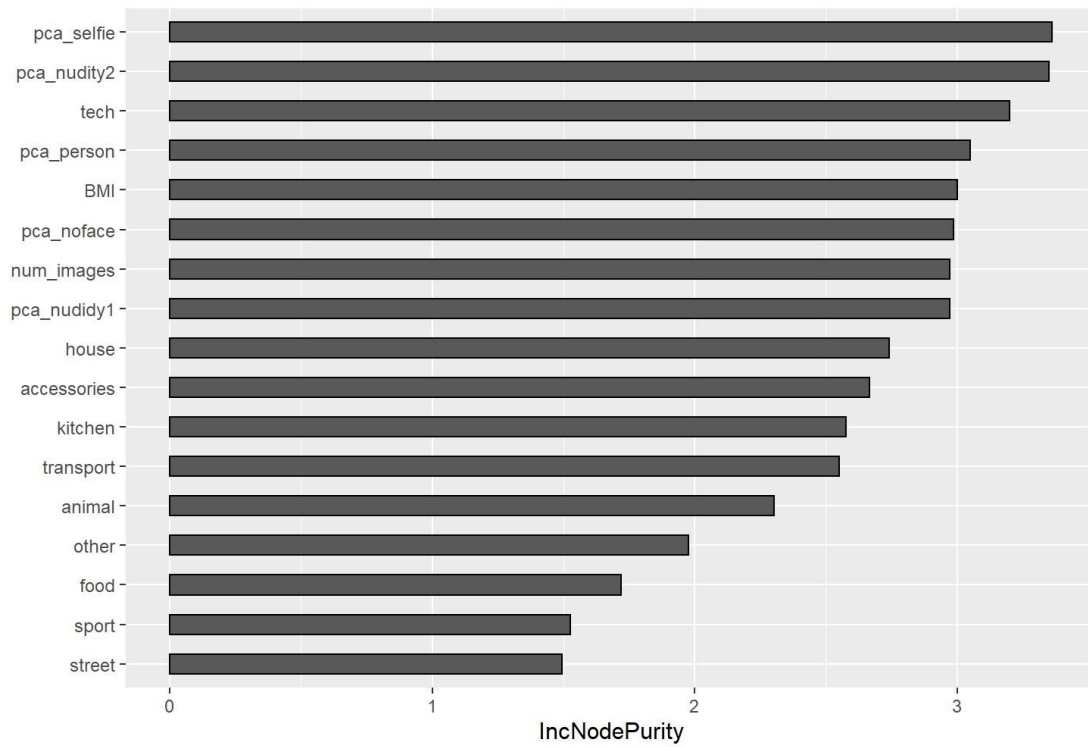


Grafico 9: *Indice di importanza delle variabili*

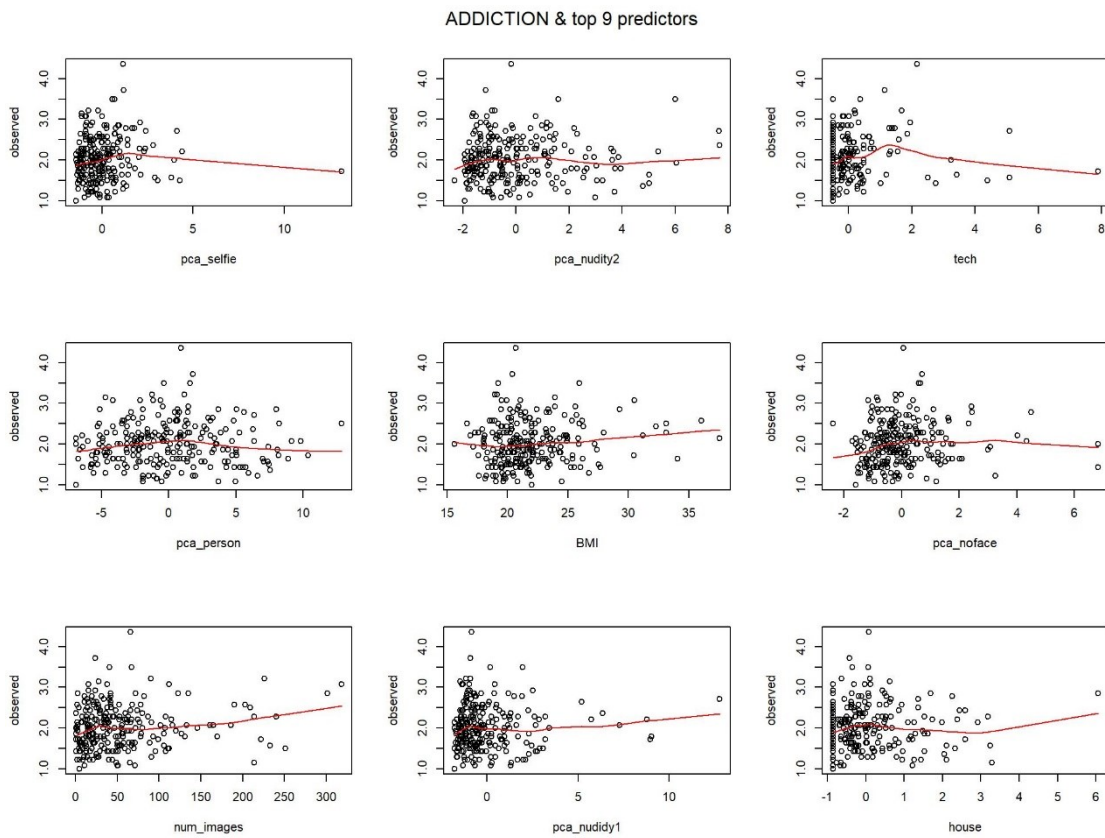


Grafico 10: *Grafici di dispersione tra la variabile dipendente e le nove migliori variabili predittrici*

5.2.2 Instagram Activity Scale: Interaction

La sottoscala dell'*Interaction*, nell'*Instagram Addiction Scale*, ha l'obiettivo di misurare in che modo il soggetto comunichi con altri utenti durante l'utilizzo di Instagram. Dal Grafico 11 si può vedere come vi siano quattro variabili indipendenti che sembrano aiutare a predire i punteggi in maniera superiore rispetto alle altre. Queste sono: "transport", "pca_nudity2", "kitchen" e "num_images". Anche per questa variabile di risposta non ci si aspettava una specifica relazione con certi contenuti delle immagini caricate su Instagram. Non è facile quindi provare a capire perché il modello faccia principalmente affidamento su queste variabili per predire l'interazione dei partecipanti con gli altri utenti. Si ricorda che la categoria "transport" è composta dai seguenti oggetti: aeroplano, bicicletta, barca, autobus, automobile e motociclo. Mentre all'interno della categoria "kitchen" comprende: bottiglia, ciotola, tazza, forchetta, coltello e cucchiaio. Il Grafico 12 inoltre mostra come non ci siano delle possibili relazioni lineari tra l'*Interaction* e le nove variabili dipendenti considerate le più importanti dal modello. La strana distribuzione della variabile *Interaction* è dovuta dal fatto che questa sottoscala sia composta unicamente da due *item*, e la media tra questi due ha prodotto un range di valori limitato. Questa considerazione sarà analoga anche per le prossime due variabili dipendenti presentate.

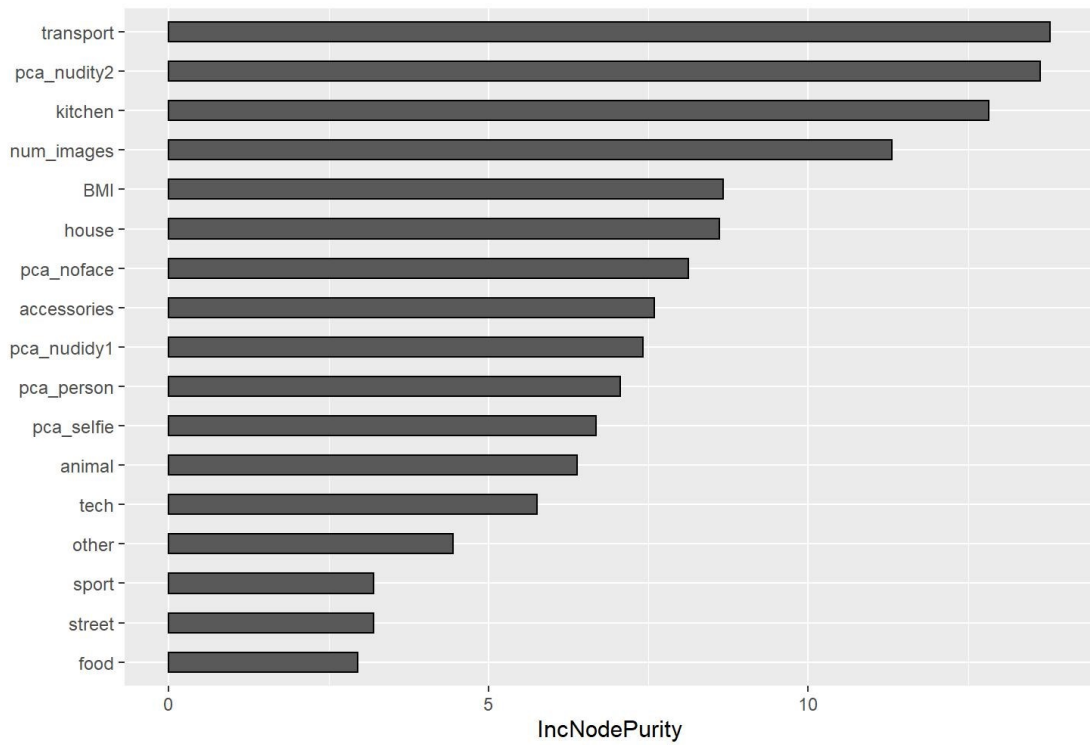


Grafico 11: Indice di importanza delle variabili

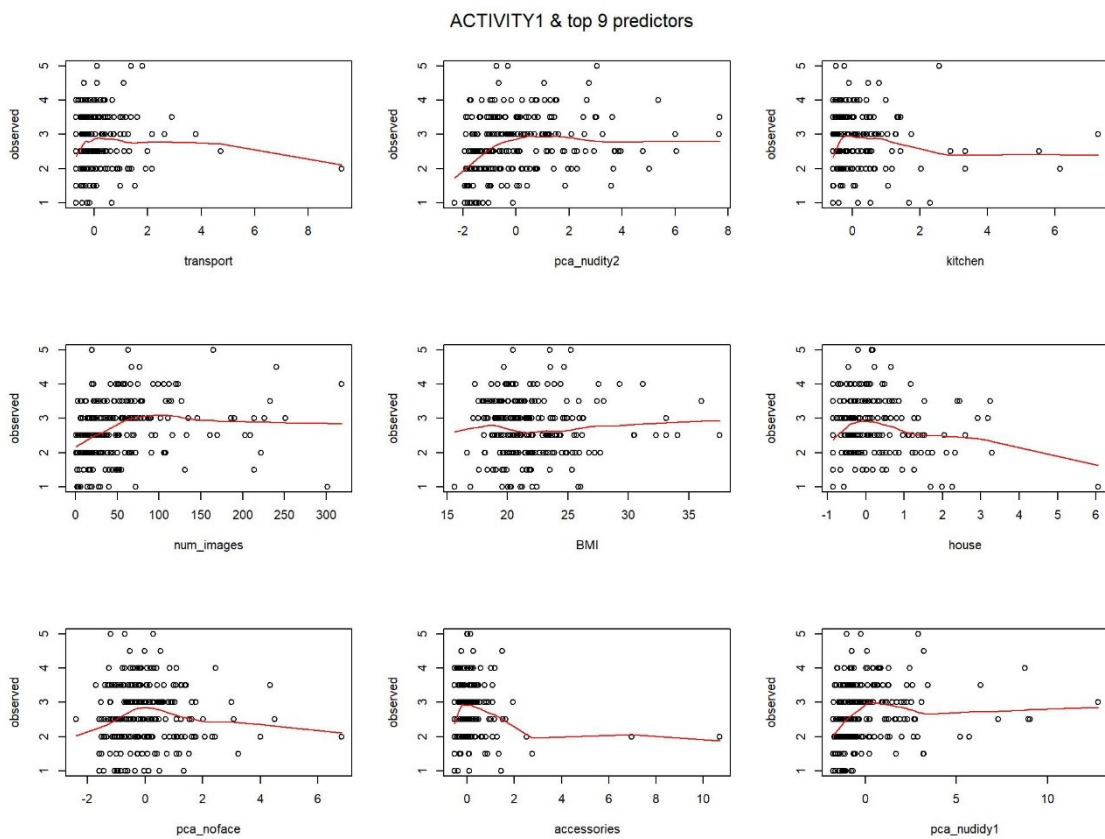


Grafico 12: Grafici di dispersione tra la variabile dipendente e le nove migliori variabili predittrici

5.2.3 Instagram Activity Scale: Browsing

La sottoscala *Browsing* nell'*Instagram Addiction Scale*, ha l'obiettivo di misurare la frequenza degli utenti nell'aggiornare il feed e controllare il profilo degli altri utenti. Anche per questa variabile di risposta non ci si aspettava una specifica relazione con certi contenuti delle immagini caricate su Instagram. È interessante tuttavia notare come al primo posto tra le variabili più importanti si trova il numero di immagini, forse l'unica variabile che riflette un certo comportamento degli utenti nell'uso di Instagram, più che effettivamente i contenuti delle immagini che vengono pubblicate (Grafico 13). Al secondo e terzo posto si trovano rispettivamente la variabile "pca_selfie" e "pca_person". Si può anche notare come non sembra ci sia una variabile indipendente o un gruppo di queste che emerga rispetto alle altre nello spiegare questo specifico utilizzo di Instagram. Anche in questo caso, come si può vedere nel Grafico 14, non sembra esserci una possibile relazione lineare tra principali variabili indipendenti e la sottoscala in esame.

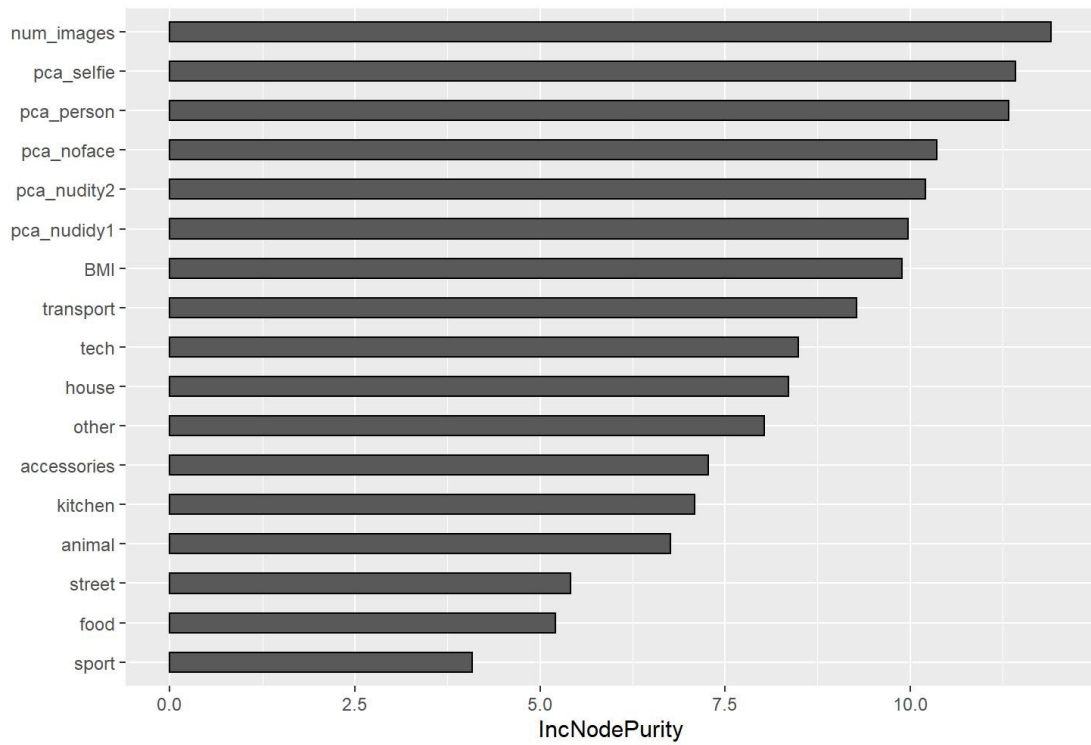


Grafico 13: Indice di importanza delle variabili

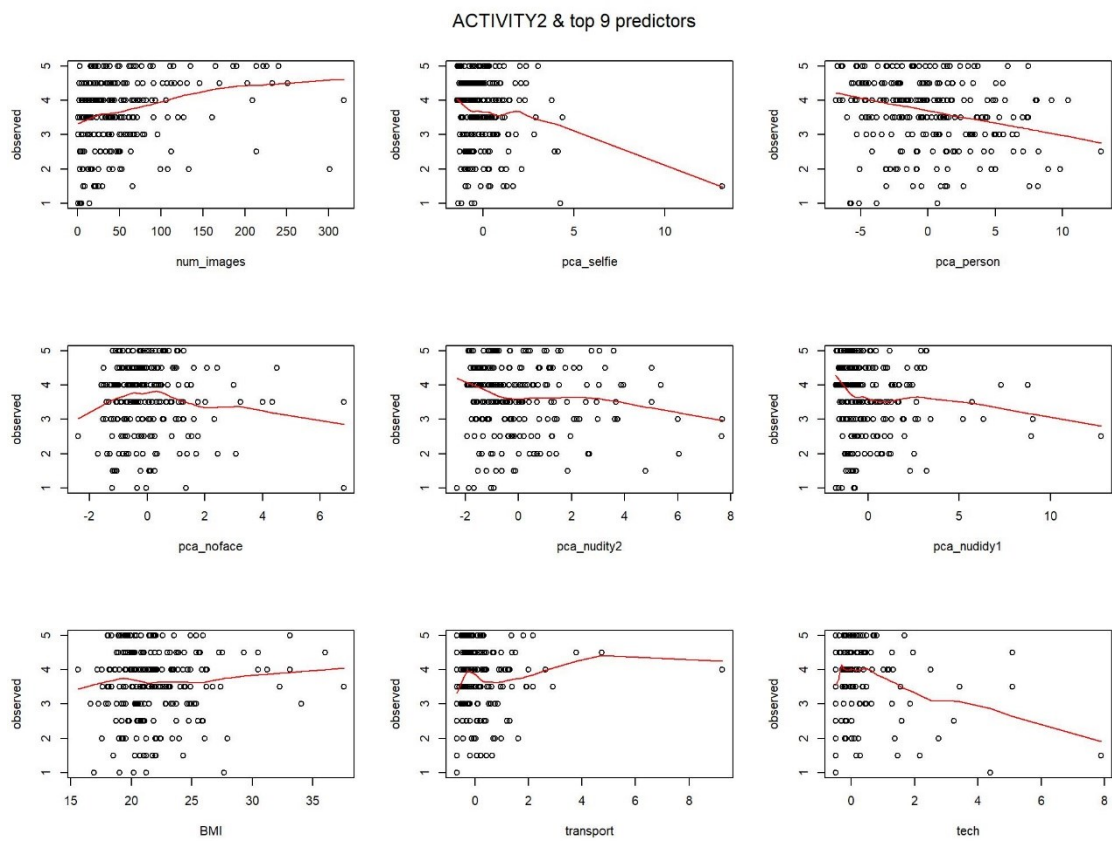


Grafico 14: Grafici di dispersione tra la variabile dipendente e le nove migliori variabili predittrici

5.2.4 Instagram Activity Scale: Broadcasting

La sottoscala *Broadcasting* nell'*Instagram Addiction Scale*, ha l'obiettivo di misurare le attività che vengono svolte dagli utenti senza interagire con altri account. Come per le precedenti sottoscale non ci si aspettava una specifica relazione tra questa e le variabili indipendenti estratte dalle immagini di Instagram. Dal Grafico 15 si può notare come il modello identifichi un gruppo di tre variabili come le principali nel contribuire al predire questa sottoscala. Le prime due, che sono state già trovate precedentemente sono: "pca_selfie" e "num_images". La terza invece è "pca_nudity1", una variabile che indica la nudità nella parte bassa del corpo. Mentre non è facile capire come quest'ultima variabile e "pca_selfie" possano contribuire in modo significativo alla predizione del modo in cui gli utenti utilizzano Instagram senza interagire con altri account, è interessante notare come al secondo posto si trova il numero di immagini. Questa variabile indipendente a differenza dalle altre estratte dalle immagini non rispecchia direttamente ciò che i soggetti pubblicano su Instagram, ma il quanto lo fanno, e quindi un comportamento nell'uso di questo *social media*. Dal Grafico 16 si può invece notare come non vi sia una possibile relazione lineare tra nessuna delle prime nove variabili predittrici e la variabile di risposta indagata.

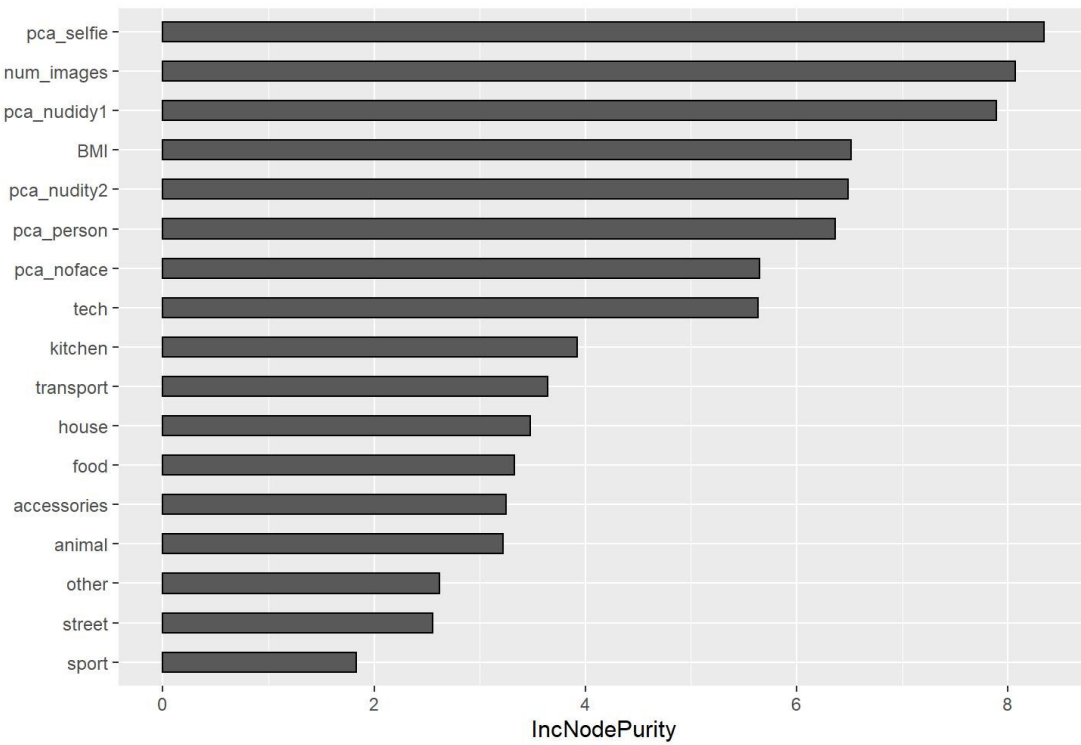


Grafico 15: Indice di importanza delle variabili

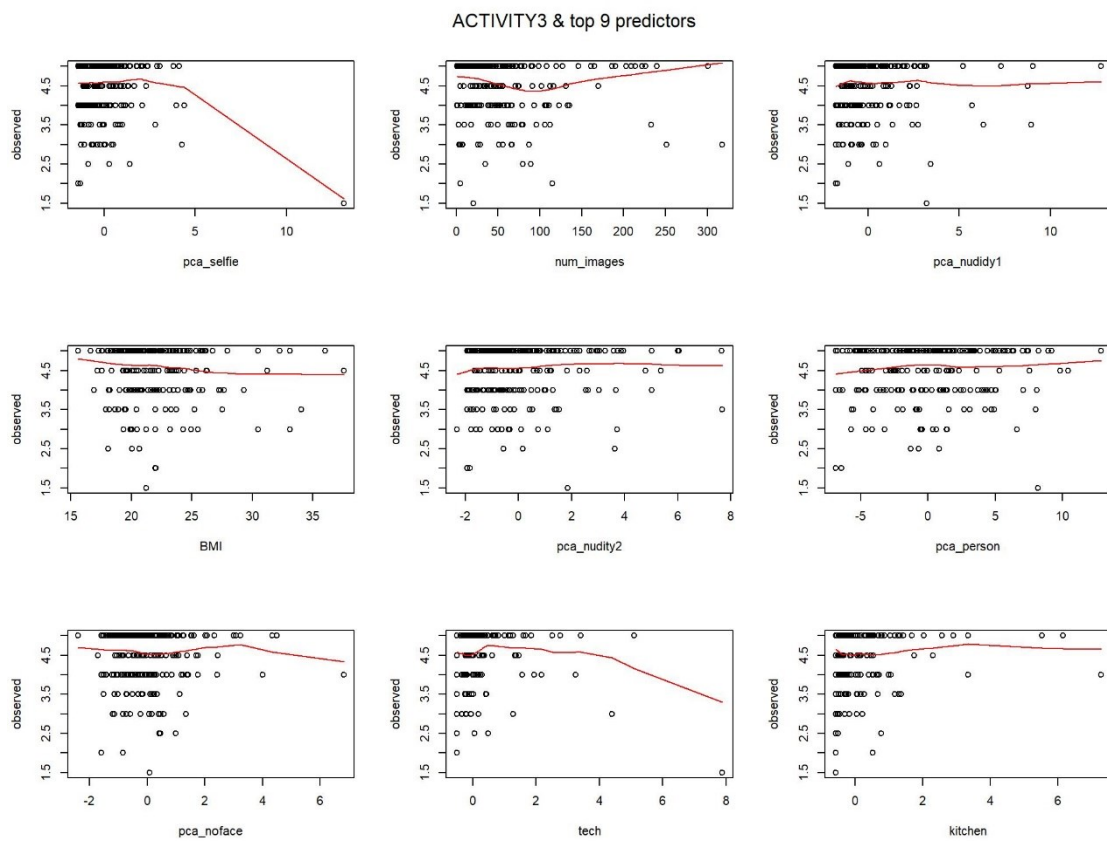


Grafico 16: Grafici di dispersione tra la variabile dipendente e le nove migliori variabili predittrici

5.2.5 Immagine corporea (SATAQ)

Per quanto riguarda gli atteggiamenti verso le influenze socioculturali sull'immagine corporea e l'interiorizzazione di norme di bellezza, ci si poteva aspettare una relazione con i contenuti delle immagini riguardanti il corpo, a differenza delle variabili dipendenti esposte precedentemente. Come si può notare dal Grafico 17, al primo posto tra le variabili considerate dal modello come le più importanti si trova "BMI", e quindi l'indice di massa corporea, che sembra aiutare nella predizione in modo cospicuo rispetto alle altre variabili. A seguire si trova un blocco di sei variabili che sembrano avere circa la stessa importanza. Va notato che, ad esclusione del numero di immagini, queste sono tutte variabili che hanno a che fare con contenuti delle immagini riguardanti il corpo, mentre le successive variabili sono tutte categorie di vari oggetti. Anche in questo caso, come si può vedere nel Grafico 18, non si notano delle possibili relazioni lineari tra le prime nove variabili predittrici e la variabile dipendente. È difficile quindi riuscire a fare un'interpretazione più approfondita di questo risultato. La cosa interessante da notare è che, nonostante non vi siano delle relazioni lineari tra l'immagine corporea ed ogni variabile indipendente presa singolarmente, si può affermare che vi sia una relazione complessa tra i contenuti delle immagini riguardanti i corpi e la variabile dipendente presa in esame.

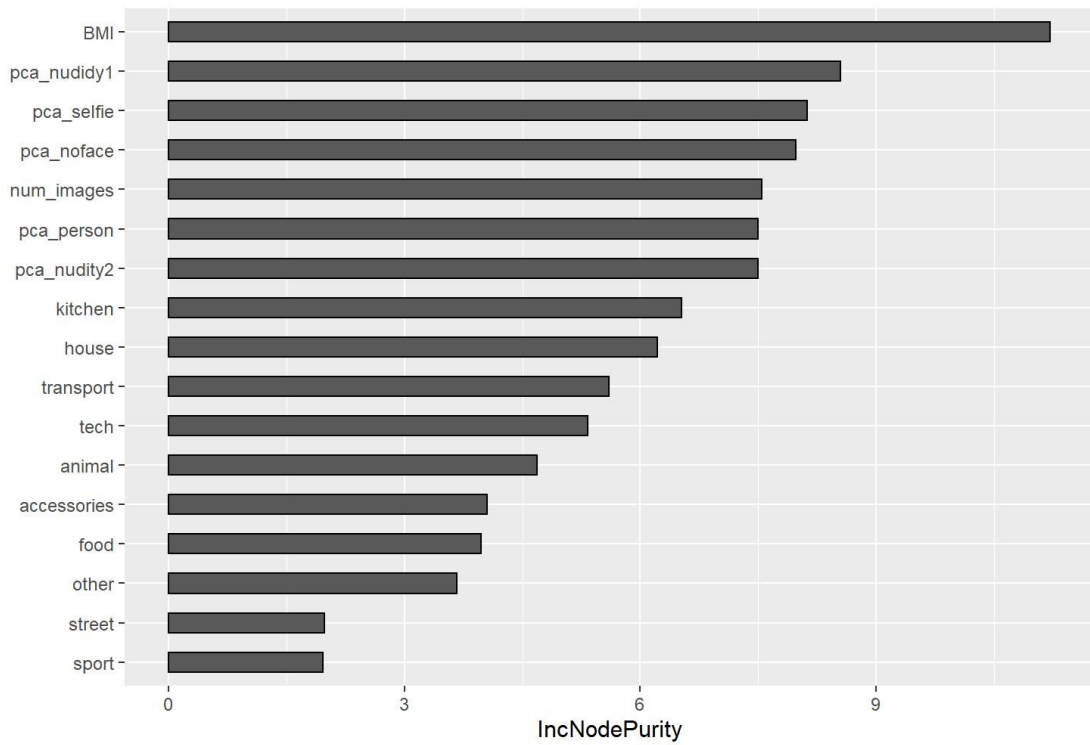


Grafico 17: *Indice di importanza delle variabili*

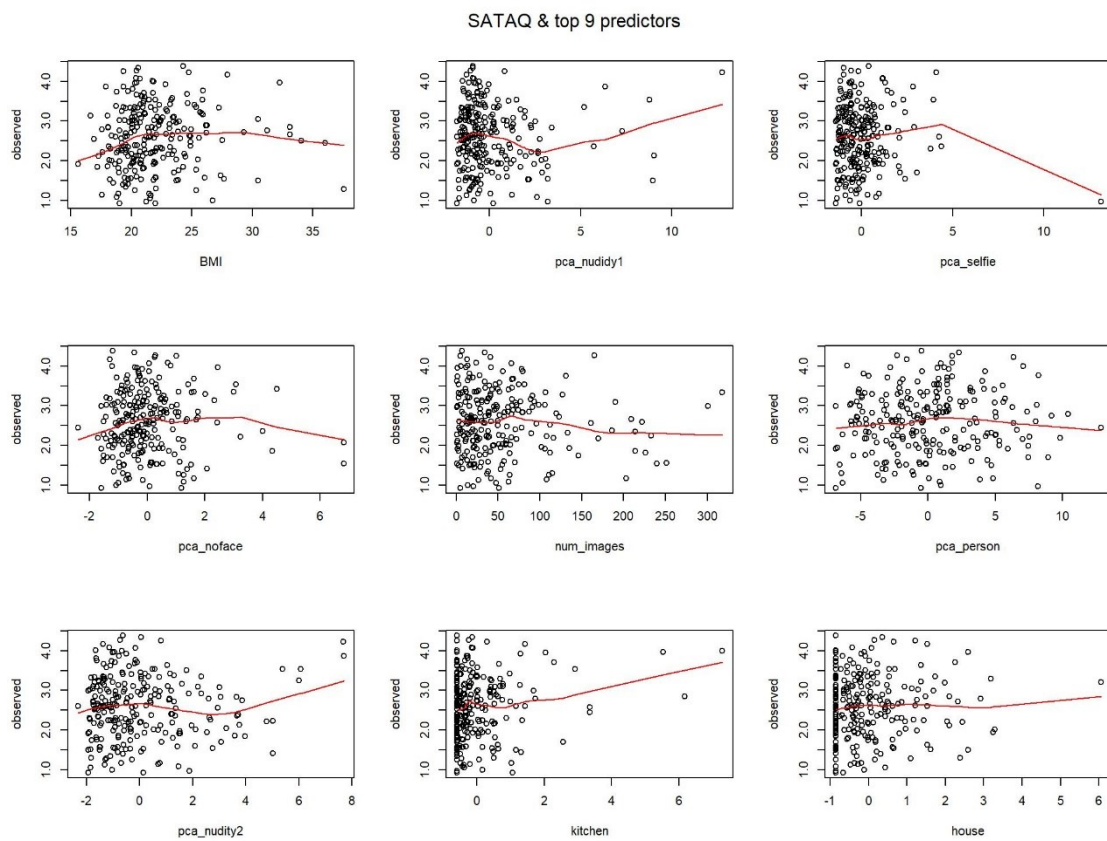


Grafico 18: *Grafici di dispersione tra la variabile dipendente e le nove migliori variabili predittrici*

5.2.6 Auto-oggettivazione (SOBBS)

L'auto-oggettivazione di tratto, misurata tramite la *Self-Objectification Beliefs and Behaviors Scale*, ha prodotto dei risultati simili a quelli dell'immagine corporea. Anche in questo caso ci si aspettava che i contenuti delle immagini riguardanti il corpo potessero avere delle relazioni con questo costrutto. Tra le variabili più considerate più importanti dal modello, perfettamente in linea con le ipotesi, si trova al primo posto "pca_noface", che sembra aiutare a predire i risultati in modo significativamente superiore rispetto alle altre (Grafico 19). Questa variabile deriva principalmente da un indice che valuta la presenza di una sola persona in una immagine, nella quale però non viene mostrato il volto. Per quanto non vi sia un relazione lineare tra questa variabile e l'auto-oggettivazione, è sicuramente interessante notare come il modello sottolinei l'importanza del mettere in mostra il proprio corpo senza mostrare il volto nel determinare i punteggi al SOBBS. Questo fenomeno è conosciuto in letteratura come *face-ism*, ed è identificato come la tendenza a mostrare unicamente il corpo di una donna in riviste e mass media tagliandone fuori dall'immagine il volto (Fredrickson e Roberts, 1997). Le successive cinque variabili più importanti, ad esclusione dell'indice di massa corporea, si riferiscono a contenuti delle immagini inerenti al corpo, ed aiutano a predire la variabile dipendente in modo superiore rispetto agli altri oggetti trovati nelle immagini. Come anche per le altre variabili di risposta non sono state trovate delle possibili relazioni lineari tra i nove migliori predittori e l'auto-oggettivazione di tratto (Grafico 20).

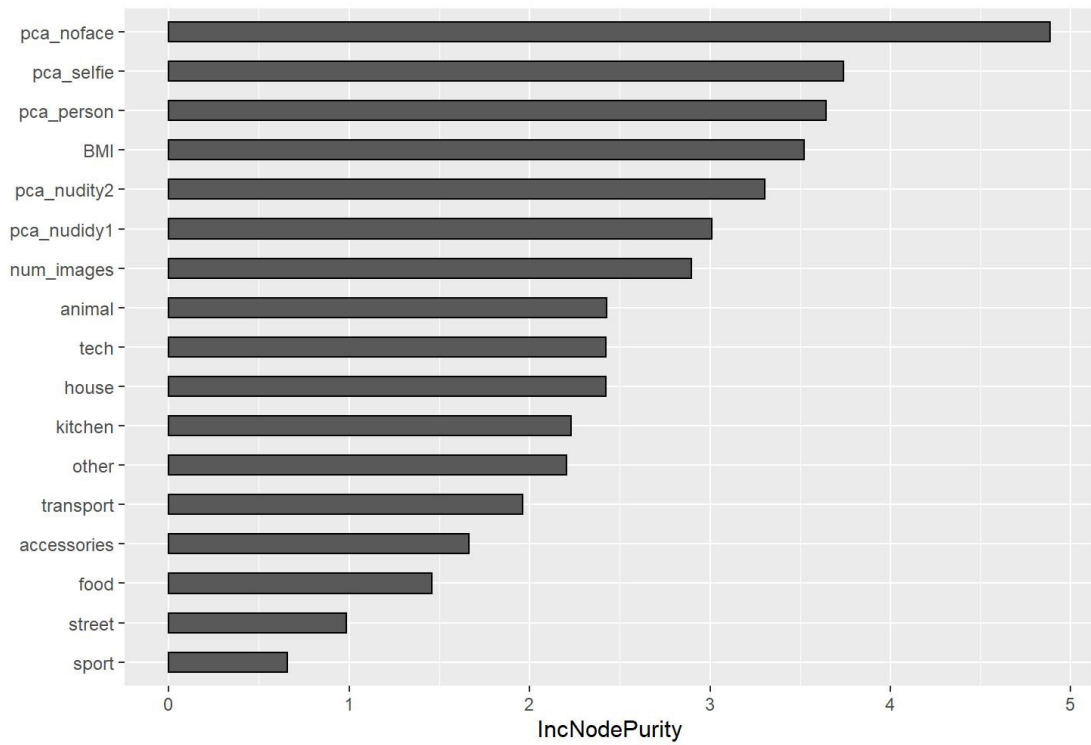


Grafico 19: Indice di importanza delle variabili

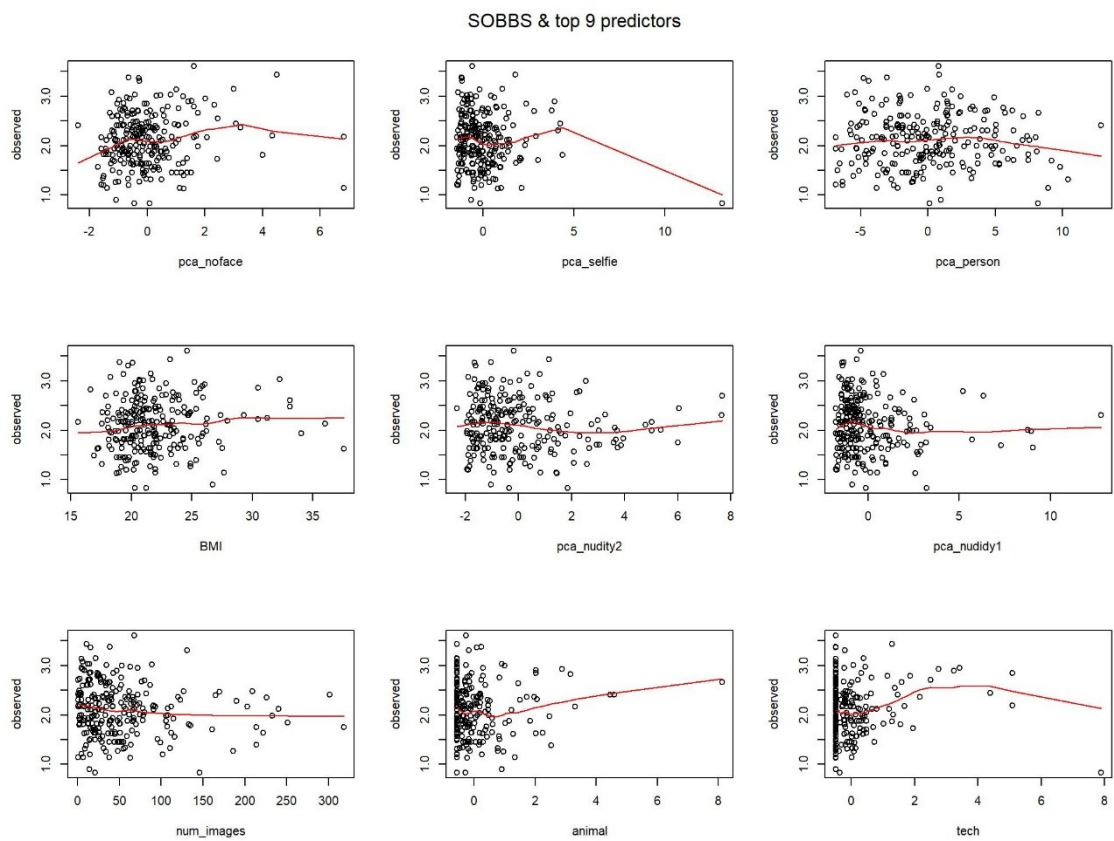


Grafico 20: Grafici di dispersione tra la variabile dipendente e le nove migliori variabili predittrici

5.2.7 Un esempio di un albero di regressione

Nel Grafico 21 si può vedere un esempio di albero di regressione. Questo è stato selezionato all'interno della *random forest* creata per predire i punteggi al SATAQ, tra i cinquecento alberi infatti quello mostrato è il più semplice, e quindi quello con meno nodi, tra quelli con una capacità predittiva migliore. Ai nodi si trova il punteggio di una certa variabile predittrice, partendo da questi e seguendo l'albero verso destra si troveranno tutte le osservazioni che hanno un punteggio maggiore di quella soglia, mentre verso sinistra quelle aventi un punteggio inferiore. Nella parte inferiore dell'immagine invece si trovano le foglie e quindi i punteggi stimati della variabile dipendente. Prima di commentare questo albero di regressione si ricorda come venga costruita una foresta di regressione, e quindi le considerazioni che si possono fare circa questo non sono necessariamente vere anche per gli altri alberi ed in generale per il modello. Alla radice dell'albero si trova un valore alto della variabile "pca_nudity1", identificata dal modello come la seconda più importante nel predire la variabile di risposta. È interessante notare come per i soggetti che hanno un valore maggiore di questa soglia l'albero porti direttamente ad una foglia, che ha il valore più alto osservato tra tutte le altre predizioni. In pratica per i soggetti che tendono a postare immagini in cui è presente la parte bassa del corpo, indipendentemente da tutte le altre variabili, si può dire che il punteggio al SATAQ sia alto. Si può vedere l'impatto della variabile "pca_nudity1" anche in altri punti dell'albero e come questa tenda a portare a diversi punteggi della variabile dipendente. L'interpretazione di altre foglie nel grafico non è altrettanto semplice, in quanto va tenuto in considerazione che i soggetti ai quali è stato assegnato quello specifico valore, fanno parte di una categoria descritta da più di una variabile. Per fare un esempio i soggetti che hanno un punteggio inferiore a 1.73 alla variabile "pca_nudity1" hanno un punteggio stimato al SATAQ di 1.9, mentre per i soggetti con un punteggio maggiore di 1.73 il punteggio stimato è 1.5. Questa considerazione è comunque vera unicamente per i soggetti descritti in quella ramificazione dell'albero, e quindi per quelli che hanno un punteggio inferiore a 8.09 alla variabile "pca_nudity1", inferiore a 19.02 alla variabile "BMI", maggiore di -0.99 alla variabile "pca_nudity2", inferiore di -0.12 alla variabile "pca_selfie" e minore di 0.1 alla variabile "food". Si capisce quindi che l'interpretazione di un albero di regressione di queste dimensioni non è per nulla semplice.

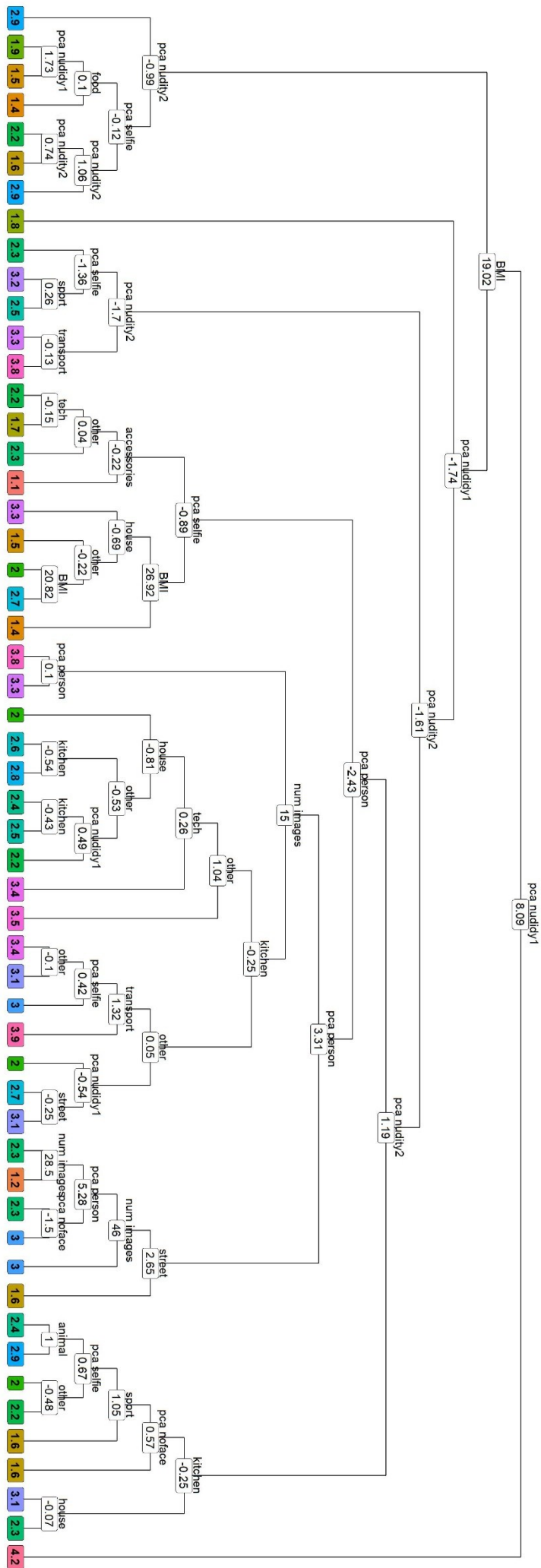


Grafico 21:
 esempio di un
 albero di
 regressione per
 la variabile
 SATAQ

Conclusioni

Lo scopo di questa ricerca è stato quello di indagare a livello esplorativo le relazioni tra alcune variabili psicologiche e i contenuti delle immagini condivise su Instagram. Per fare ciò sono state raccolte tramite un questionario auto-somministrato sei variabili indipendenti, inerenti a: la dipendenza e le diverse modalità di utilizzo di questo *social media*; l'immagine corporea e l'interiorizzazione delle norme di bellezza; l'auto-oggettivazione. I partecipanti alla ricerca alla fine del questionario hanno inviato ai ricercatori i propri dati Instagram. Tramite una procedura automatica e grazie all'utilizzo di algoritmi di *object detection*, dai dati di ogni soggetto sono state estratte ed analizzate le immagini pubblicate sul *social media*. In questo processo l'immagine del profilo è stata analizzata separatamente da quelle caricate sul profilo. Infine, dopo essere state opportunamente modificate, le informazioni estratte dalle immagini sono state messe in relazione alle variabili di interesse grazie a modelli predittivi.

Dai modelli si è potuto dedurre che le immagini del profilo non sono indicative di nessuna variabile di interesse, mentre dai contenuti delle immagini caricate sul profilo si può predire con un certo errore il punteggio alle variabili esposte in precedenza. Da un punto di vista esplicativo non è stato possibile capire in che modo i contenuti delle immagini siano collegati alla dipendenza da Instagram e alle diverse modalità di utilizzo del *social media* indagate tramite i questionari. Per quanto riguarda invece il costrutto dell'immagine corporea e quello dell'auto-oggettivazione, in linea con le ipotesi iniziali della ricerca, è stato evidenziato come queste possano essere predette dai contenuti delle immagini inerenti alle persone e a come queste vengano ritratte. Anche per questi costrutti tuttavia non è stata evidenziata una direzionalità precisa tra le variabili indipendenti e quelle di risposta, e quindi non è chiaro come queste si influenzino per arrivare a predire le variabili dipendenti. In Tabella 10 si possono trovare ulteriori informazioni che riassumono i principali risultati.

	<i>MAPE</i>	<i>MAE</i>	<i>Variabili più importanti</i>	<i>Interpretazione</i>
Addiction	0.15	0.27	pca_selfie, pca_nudity2, tech.	Non è possibile capire da un punto di vista teorico perché il modello abbia scelto queste variabili come le migliori predittrici.
Activity1	0.16	0.34	transport, pca_nudity2, kitchen, num_images.	Per tutte è presente la variabile “numero delle immagini”, unico indicatore del comportamento degli utenti, più che del contenuto delle immagini pubblicate. Per le altre variabili non è possibile capire da un punto di vista teorico perché siano state scelte dal modello come le migliori.
Activity2	0.21	0.55	num_images, pca_selfie, pca_person.	
Activity3	0.09	0.35	pca_selfie, num_images, pca_nudity1.	
SATAQ	0.18	0.38	BMI, pca_nudity1, pca_selfie, pca_noface, pca_person, pca_nudity2.	Le migliori variabili predittrici sono quelle che hanno a che fare con i contenuti delle immagini inerenti alle persone e a come queste vengano ritratte.
SOBBS	0.14	0.25	pca_noface, pca_selfie, pca_person, BMI, pca_nudity2, pca_nudity1.	

Tabella 10: Principali risultati per le immagini caricate sul profilo

Uno dei limiti principali di questo lavoro è stata l'impossibilità appena descritta del riuscire a comprendere quali siano le caratteristiche delle variabili indipendenti che portino a certi punteggi delle variabili dipendenti, e di riflesso quindi l'impossibilità di sviluppare delle teorie esplicative di queste interazioni. Un modo per superare questo limite potrebbe essere quello di utilizzare dei metodi statistici per provare a descrivere più nel dettaglio i modelli e comprendere l'influenza di ogni variabile predittrice nello stimare le variabili di risposta. Un altro aspetto da tenere in considerazione è il fatto che potrebbe essere scelta una diversa soglia di precisione degli oggetti trovati nelle immagini, ottenendo quindi dei nuovi modelli che potrebbero avere delle capacità predittive migliori. I risultati inoltre hanno evidenziato l'importanza della costruzione degli indici per ottenere una migliore interpretazione dei modelli, per questo motivo una possibilità per implementare questo lavoro potrebbe essere quella di creare altri indici partendo dai dati raccolti tramite il *data-mining*, per vedere se questi possono contribuire a spiegare le variabili dipendenti. Sicuramente infine sarebbe interessante provare a creare degli ulteriori modelli di predizione che prendano in considerazione altre variabili estratte dai dati di Instagram, e non unicamente quelle delle immagini. Un modello basato anche su altri indicatori potrebbe portare a dei risultati interessanti e a delle migliori predizioni, soprattutto per le variabili che da un punto di vista teorico sembrerebbe abbiano meno a che fare con i contenuti delle immagini postate, ossia per la dipendenza da Instagram e per le diverse modalità di utilizzo di questo *social media*.

Bibliografia

Azucar, D., Marengo, D., & Settanni, M. (2018). Predicting the Big 5 personality traits from digital footprints on social media: A meta-analysis. *Personality and individual differences, 124*, 150-159.

Azzalini, A., & Scarpa, B. (2012). *Data analysis and data mining: An introduction*. OUP USA.

Bell, B. T., Cassarly, J. A., & Dunbar, L. (2018). Selfie-objectification: Self-objectification and positive feedback ("likes") are associated with frequency of posting sexually objectifying self-images on social media. *Body image, 26*, 83-89.

Colantonio, S., & Salvetti, O. (2005). Categorizzazione automatica di immagini mediante algoritmi neurali.

Coppersmith, G., Ngo, K., Leary, R., & Wood, A. (2016). Exploratory analysis of social media prior to a suicide attempt. In *Proceedings of the third workshop on computational linguistics and clinical psychology* (pp. 106-117).

Cox, C., Moscardini, E. H., Cohen, A. S., & Tucker, R. P. (2020). Machine learning for suicidology: A practical review of exploratory and hypothesis-driven approaches. *Clinical Psychology Review, 101940*.

De Choudhury, M., Gamon, M., Counts, S., & Horvitz, E. (2013). Predicting depression via social media. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 7, No. 1).

Fardouly, J., Willburger, B. K., & Vartanian, L. R. (2018). Instagram use and young women's body image concerns and self-objectification: Testing mediational pathways. *New Media & Society, 20*(4), 1380-1395.

Ferwerda, B., Schedl, M., & Tkalcic, M. (2015). Predicting personality traits with instagram pictures. In *Proceedings of the 3rd Workshop on Emotions and Personality in Personalized Systems 2015* (pp. 7-10).

Fredrickson, B. L., & Roberts, T. A. (1997). Objectification theory: Toward understanding women's lived experiences and mental health risks. *Psychology of women quarterly*, 21(2), 173-206.

Gleaves, D. H., Williamson, D. A., Eberenz, K. P., Sebastian, S. B., & Barker, S. E. (1995). Clarifying body-image disturbance: Analysis of a multidimensional model using structural modeling. *Journal of Personality Assessment*, 64(3), 478-493.

Golbeck, J., Robles, C., Edmondson, M., & Turner, K. (2011). Predicting personality from twitter. In 2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing (pp. 149-156). IEEE.

Grabe, S., Ward, L. M., & Hyde, J. S. (2008). The role of the media in body image concerns among women: a meta-analysis of experimental and correlational studies. *Psychological bulletin*, 134(3), 460.

Guntuku, S. C., Yaden, D. B., Kern, M. L., Ungar, L. H., & Eichstaedt, J. C. (2017). Detecting depression and mental illness on social media: an integrative review. *Current Opinion in Behavioral Sciences*, 18, 43-49.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media.

Hatton, E., & Trautner, M. N. (2011). Equal opportunity objectification? The sexualization of men and women on the cover of Rolling Stone. *Sexuality & culture*, 15(3), 256-278.

Hawkins, D. M. (2004). The problem of overfitting. *Journal of chemical information and computer sciences*, 44(1), 1-12.

Ho, T. K. (1995). Random decision forests. In *Proceedings of 3rd international conference on document analysis and recognition* (Vol. 1, pp. 278-282). IEEE.

Ive, J., Gkotsis, G., Dutta, R., Stewart, R., & Velupillai, S. (2018). Hierarchical neural model with attention mechanisms for the classification of social media text related to mental health. In *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic* (pp. 69-77).

Karsay, K., Knoll, J., & Matthes, J. (2018). Sexualizing media use and self-objectification: A meta-analysis. *Psychology of Women Quarterly*, *42*(1), 9-28.

Kircaburun, K., & Griffiths, M. D. (2018). Instagram addiction and the Big Five of personality: The mediating role of self-liking. *Journal of behavioral addictions*, *7*(1), 158-170.

Lee, E., Lee, J. A., Moon, J. H., & Sung, Y. (2015). Pictures speak louder than words: Motivations for using Instagram. *Cyberpsychology, behavior, and social networking*, *18*(9), 552-556.

Lindner, D., & Tantleff-Dunn, S. (2017). The development and psychometric evaluation of the Self-Objectification Beliefs and Behaviors Scale. *Psychology of Women Quarterly*, *41*(2), 254-272.

Loh, W. Y. (2011). Classification and regression trees. *Wiley interdisciplinary reviews: data mining and knowledge discovery*, *1*(1), 14-23.

Lup, K., Trub, L., & Rosenthal, L. (2015). Instagram# instasad?: exploring associations among instagram use, depressive symptoms, negative social comparison, and strangers followed. *Cyberpsychology, Behavior, and Social Networking*, *18*(5), 247-252.

Markovikj, D., Gievska, S., Kosinski, M., & Stillwell, D. (2013). Mining facebook data for predictive personality modeling. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 7, No. 1).

Mitchell, M., Hollingshead, K., & Coppersmith, G. (2015). Quantifying the language of schizophrenia in social media. In *Proceedings of the 2nd workshop on Computational linguistics and clinical psychology: From linguistic signal to clinical reality* (pp. 11-20).

Mohr, D. C., Zhang, M., & Schueller, S. M. (2017). Personal sensing: understanding mental health using ubiquitous sensors and machine learning. *Annual review of clinical psychology, 13*, 23-47.

Moon, J. H., Lee, E., Lee, J. A., Choi, T. R., & Sung, Y. (2016). The role of narcissism in self-promotion on Instagram. *Personality and individual Differences, 101*, 22-25.

Orabi, A. H., Buddhitha, P., Orabi, M. H., & Inkpen, D. (2018). Deep learning for depression detection of twitter users. In *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic* (pp. 88-97).

Prichard, E. C. (2021). Is the Use of Personality Based Psychometrics by Cambridge Analytical Psychological Science's "Nuclear Bomb" Moment?. *Frontiers in Psychology, 12*.

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779-788).

Savci, M., Tekin, A., & Elhai, J. D. (2020). Prediction of problematic social media use (PSU) using machine learning approaches. *Current Psychology, 1-10*.

Schwartz, H. A., Eichstaedt, J. C., Dziurzynski, L., Kern, M. L., Blanco, E., Kosinski, M., ... & Ungar, L. H. (2013). Toward personality insights from language exploration in social media. In *2013 AAAI Spring Symposium Series*.

Schwartz, H. A., Eichstaedt, J., Kern, M., Park, G., Sap, M., Stillwell, D., ... & Ungar, L. (2014). Towards assessing changes in degree of depression through facebook. In *Proceedings of the workshop on computational linguistics and clinical psychology: from linguistic signal to clinical reality* (pp. 118-125).

Sheldon, P., & Bryant, K. (2016). Instagram: Motives for its use and relationship to narcissism and contextual age. *Computers in human Behavior, 58*, 89-97.

Sherlock, M., & Wagstaff, D. L. (2019). Exploring the relationship between frequency of Instagram use, exposure to idealized images, and psychological well-being in women. *Psychology of Popular Media Culture, 8*(4), 482.

Slade, P. D. (1994). What is body image?. *Behaviour research and therapy, 32*(5), 497-502.

Stefanile, C., Matera, C., Nerini, A., & Pisani, E. (2011). Validation of an Italian version of the Sociocultural Attitudes Towards Appearance Questionnaire-3 (SATAQ-3) on adolescent girls. *Body Image, 8*(4), 432-436.

Stillwell, D. J., & Kosinski, M. (2004). myPersonality project: Example of successful utilization of online social networks for large-scale social research. *American Psychologist, 59*(2), 93-104.

Thompson, J. K., Van Den Berg, P., Roehrig, M., Guarda, A. S., & Heinberg, L. J. (2004). The sociocultural attitudes towards appearance scale-3 (SATAQ-3): Development and validation. *International journal of eating disorders, 35*(3), 293-304.

Tiggemann, M., & Anderberg, I. (2020). Social media is not real: The effect of 'Instagram vs reality' images on women's social comparison and body image. *New Media & Society, 22*(12), 2183-2199.

Vaes, J., Paladino, P., & Puvia, E. (2011). Are sexualized women complete human beings? Why men and women dehumanize sexually objectified women. *European Journal of Social Psychology, 41*(6), 774-785.

Vandenbosch, L., & Eggermont, S. (2012). Understanding sexual objectification: A comprehensive approach toward media exposure and girls' internalization of beauty ideals, self-objectification, and body surveillance. *Journal of Communication, 62*(5), 869-887.

Ward, L. M. (2016). Media and sexualization: State of empirical research, 1995–2015. *The Journal of Sex Research, 53*(4-5), 560-577.

Wu, J. H., Wang, S. C., & Tsai, H. H. (2010). Falling in love with online games: The uses and gratifications perspective. *Computers in Human Behavior, 26*(6), 1862-1871.

Yang, C. C. (2016). Instagram use, loneliness, and social comparison orientation: Interact and browse on social media, but don't compare. *Cyberpsychology, Behavior, and Social Networking, 19*(12), 703-708.

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science, 12*(6), 1100-1122.

Yurdagül, C., Kircaburun, K., Emirtekin, E., Wang, P., & Griffiths, M. D. (2019). Psychopathological consequences related to problematic Instagram use among adolescents: The mediating role of body image dissatisfaction and moderating role of gender. *International Journal of Mental Health and Addiction, 1-13*.

Zou, Z., Shi, Z., Guo, Y., Ye, J. (2019). Object detection in 20 years: A survey. *arXiv preprint arXiv:1905.05055*.

Appendice

Viene ora riportato parte del questionario somministrato ai partecipanti in fase di raccolta dati. Nello specifico le domande riguardanti le variabili dipendenti.

Instagram Addiction Scale

Istruzioni:

Di seguito sono riportate una serie di affermazioni riguardanti il tuo uso di Instagram negli ultimi 6 mesi. Pensando alla tua personale esperienza, ti preghiamo di indicare quanto sei d'accordo con ciascuna affermazione su una scala di valutazione da 1 a 6. La scala di valutazione è la seguente: 1 = Mai, 2 = Raramente, 3 = Occasionalmente, 4 = Frequentemente, 5 = Spesso, 6 = Sempre.

Item:

1. Quanto spesso preferisci utilizzare Instagram piuttosto che passare del tempo con i tuoi amici perché ti dà più piacere?
2. Quanto spesso crei nuove relazioni con altri utenti di Instagram?
3. Quanto spesso ti capita di essere riservata o sulla difensiva quando ti viene chiesto cosa fai su Instagram?
4. Quanto spesso i tuoi voti o la tua prestazione accademica risentono negativamente del tempo che trascorri su Instagram?
5. Quanto spesso scatti, urli o ti infastidisci se qualcuno ti disturba mentre sei su Instagram?
6. Quanto spesso cerchi di nascondere quanto tempo trascorri su Instagram?
7. Quanto spesso scegli di dedicare più tempo a stare su Instagram invece di uscire con altre persone?
8. Quanto spesso ti senti depressa, nervosa o con sbalzi di umore se non puoi accedere a Instagram, situazione che migliora quando hai la possibilità di accedervi?
9. Quanto spesso provi a ridurre il tempo che trascorri su Instagram senza successo?
10. Quanto spesso controlli il tuo account Instagram prima di qualcos'altro che devi fare?

11. Quanto spesso pensi che Instagram sia di aiuto per smettere di pensare a cose spiacevoli?
12. Quanto spesso pensi al quando potrai accedere di nuovo a Instagram?
13. Quanto spesso temi che la tua vita sarebbe noiosa, vuota e priva di gioie se non avessi Instagram?
14. Quanto spesso perdi sonno perché utilizzi Instagram fino a tarda notte?
15. Quanto spesso dici a te stessa "ancora qualche minuto" quando usi Instagram?

Instagram Activity Scale

Istruzioni:

Di seguito sono riportate una serie di affermazioni riguardanti il tuo uso di Instagram negli ultimi 6 mesi. Pensando alla tua personale esperienza, ti preghiamo di indicare quanto sei d'accordo con ciascuna affermazione su una scala da 1 (Mai) a 5 (Molto).

Item:

1. Commenti o rispondi ai post di altre persone
2. Tagghi altre persone nei tuoi post o commenti
3. Pubblichiamo / Carichi qualcosa sul tuo profilo senza taggare nessuno
4. Pubblichiamo qualcosa che non è diretto a persone specifiche profilo senza taggare nessuno
5. Navighi sulla home page / news feed (dove vedi i nuovi aggiornamenti) senza lasciare commenti
6. Guardi i profili degli altri senza commentare

Sottoscala Interaction: item 1 e 2.

Sottoscala Browsing: item 3 e 4.

Sottoscala Broadcasting: item 5 e 6.

SATAQ

Istruzioni:

Indica il tuo grado di accordo o di disaccordo per ciascuna delle seguenti affermazioni. Ricorda non ci sono risposte giuste o sbagliate, il tuo contributo sincero è per noi molto prezioso.

Item:

1. I programmi televisivi sono una fonte importante d'informazione su ciò che è di moda e su come essere attraente.
2. Ho sentito la pressione della tv o dei social media a perdere peso.
3. Vorrei che il mio corpo assomigliasse a quello dei personaggi televisivi.
4. Confronto il mio corpo con quello delle dive della tv e del cinema.
5. Le pubblicità sono una fonte importante d'informazione su ciò che è di moda e su come essere attraente.
6. Ho sentito la pressione della tv o dei social media ad apparire carina.
7. Vorrei che il mio corpo assomigliasse a quello delle modelle che appaiono sulle riviste e i social media.
8. Confronto il mio aspetto con quello delle dive della tv e del cinema.
9. I video musicali sono una fonte importante d'informazione su ciò che è di moda e su come essere attraente.
10. Ho sentito la pressione della tv e dei social media ad essere magra.
11. Vorrei che il mio corpo assomigliasse a quello delle protagoniste dei film.
12. Confronto il mio corpo con quello delle ragazze che appaiono nelle riviste e nei social media.
13. Le riviste e i social media sono una fonte importante di informazione su ciò che è di moda e su come essere attraente.
14. Ho sentito la pressione della TV o dei social media ad avere un corpo perfetto.
15. Desidero assomigliare alle modelle presenti nei video musicali.
16. Confronto il mio aspetto con quello delle persone nei social media.

17. Le pubblicità sono una fonte importante d'informazione su ciò che è di moda e su come essere attraente.
18. Ho sentito la pressione della TV o dei social media a seguire una dieta.
19. Confronto il mio corpo con quello delle persone "in forma".
20. Le immagini sui social media sono una fonte importante d'informazione su ciò che è di moda e su come essere attraente.
21. I film sono una fonte importante di informazione su ciò che è di moda e su come essere attraente.
22. Ho sentito la pressione della tv o delle social media a fare esercizio fisico.
23. Ho sentito la pressione della tv o dei social media a cambiare il mio aspetto.
24. Cerco di assomigliare ai personaggi televisivi.
25. Le dive del cinema sono una fonte importante d'informazione su ciò che è di moda e su come essere attraente.
26. Le persone famose sono una fonte importante d'informazione su ciò che è di moda e su come essere attraente.
27. Desidero assomigliare alle modelle presenti nei video musicali.

SOBBS

Istruzioni:

Indica il tuo grado di accordo o di disaccordo per ciascuna delle seguenti affermazioni.

Item:

1. Apparire attraente agli altri è per me più importante che essere contenta di come sono dentro.
2. Cerco di immaginare come gli altri vedono il mio corpo (come se mi stessi guardando dall'esterno).
3. Come appaio fisicamente è per me più importante di quello che penso o che provo.
4. Scelgo i miei abiti o accessori sulla base di come fanno apparire il mio corpo agli altri.
5. Il mio aspetto fisico è più importante della mia personalità.

6. Quando mi guardo allo specchio, noto delle aree del mio aspetto fisico che penso le altre persone vedranno in modo critico.
7. Penso a come il mio corpo apparirà alle altre persone con i vestiti che indosso.
8. Penso spesso a come il mio corpo appare agli altri.
9. Il mio aspetto fisico mi rappresenta di più rispetto al mio intelletto.
10. Quanto le persone mi trovano sessualmente attraente dice qualcosa di come sono come persona.
11. Il mio aspetto fisico è più importante delle mie abilità fisiche.
12. Cerco di immaginare le reazioni delle altre persone al mio aspetto fisico.
13. Il mio corpo è ciò che mi dà valore agli occhi delle altre persone.
14. Penso a come il mio corpo appare agli altri anche quando sono da sola.

Istruzioni download dati Instagram

Istruzioni:

Adesso ti chiediamo di darci un piccolo aiuto. Occorrerebbe che tu scaricassi direttamente dall'app di Instagram i tuoi dati. Ti diamo una mano:

1. Entra sull'app di Instagram;
2. Apri le impostazioni;
3. Vai alla voce "Privacy e sicurezza".
4. Vai alla voce "Scarica i dati": inserendo la tua mail, riceverai presso il tuo indirizzo di posta elettronica i dati inerenti il tuo profilo Instagram.

Una volta ricevuto la mail da Instagram (vedi immagini), non ti resta che aprire il link e scaricare sul tuo computer i dati richiesti.



Ecco il file che hai richiesto con le foto, i commenti, le informazioni del profilo e gli altri dati che hai condiviso su Instagram come



Questo link funzionerà solo per i prossimi 4 giorni. Dato che potrebbe contenere informazioni personali, assicurati di conservarlo in modo sicuro e di scaricare l'archivio solo sul tuo computer.

[Scarica i dati](#)

I tuoi dati di Instagram

Ecco il file che hai richiesto con le foto, i commenti, le informazioni del profilo e gli altri dati che hai condiviso su Instagram come



Questo link funzionerà solo per i prossimi 4 giorni. Dato che potrebbe contenere informazioni personali, assicurati di conservarlo in modo sicuro e di scaricare l'archivio solo sul tuo computer.

[Scarica i dati - Parte 1](#)

[Scarica i dati - Parte 2](#)

[Scarica i dati - Parte 3](#)

[Scarica i dati - Parte 4](#)