

Soddisfazione e impegno per il lavoro nelle Cooperative Sociali

Paola Zuccolotto

Dipartimento Metodi Quantitativi
Università degli Studi di Brescia
zuk@eco.unibs.it

Sommario: La soddisfazione per il lavoro nel suo complesso è una percezione che risente della soddisfazione maturata dal lavoratore riguardo molteplici aspetti dell'esperienza lavorativa, che possono riguardare la sfera relazionale, quella della realizzazione personale, quella più concretamente retributiva, e molte altre. Naturalmente non è detto che tutti questi fattori godano della stessa considerazione da parte del lavoratore, che può essere maggiormente influenzato nella sua soddisfazione complessiva da alcuni aspetti piuttosto che da altri. In questo contributo si cerca di valutare quali siano gli ambiti di soddisfazione sul lavoro che maggiormente influiscono sull'appagamento complessivo del lavoratore e quali abbiano maggior impatto sull'impegno che egli dispensa nello svolgimento dei propri compiti. L'attenzione è focalizzata sui lavoratori appartenenti alle Cooperative Sociali oggetto dell'indagine ICSI2007. Dal punto di vista statistico il problema viene affrontato mediante l'utilizzo di recenti tecniche di tipo algoritmico basate sugli alberi decisionali.

Keywords: Settore nonprofit, *Job Satisfaction*, *Effort*, TreeBoost, misure di importanza, Analisi delle Corrispondenze.

1. Introduzione

Lo studio della soddisfazione per il lavoro (*Job Satisfaction*, JS) presenta implicazioni di notevole interesse sia dal punto di vista economico, poiché si ritiene che a più elevati livelli di soddisfazione corrisponda in genere maggior impegno nel lavoro (*Effort*) e quindi migliori rendimenti nelle prestazioni, sia dal punto di vista statistico-metodologico, in quanto svariate tecniche possono essere impiegate per l'analisi e la misurazione della JS.

I metodi più diffusi per la rilevazione tramite questionari di giudizi che consentano di misurare la JS sono sostanzialmente di due tipologie: l'utilizzo di un'unica domanda che chieda al rispondente di esprimere un giudizio sulla propria soddisfazione complessiva (*overall JS*) oppure la preparazione di una batteria di *item*, attraverso la quale richiedere giudizi di soddisfazione su molteplici aspetti distinti del lavoro come stipendio, relazioni con i colleghi, realizzazione personale, etc., (*facet JS*). Quando si utilizza il secondo approccio vengono poi tipicamente impiegate procedure statistiche che consentono di ottenere una misura globale di JS come sintesi delle singole misure rilevate tramite i vari *item*. Tali procedure possono essere molto semplici (e talvolta grossolane, come quelle che consistono nell'effettuare una semplice media dei singoli giudizi) o più raffinate, come ad esempio quelle che consentono di ottenere una riduzione della dimensionalità che tiene conto della natura di variabili qualitative in scala ordinale dei giudizi espressi (CATPCA,

si veda ad esempio Manisera, 2005) o quelle che si rifanno a modellistiche introdotte in ambito psicometrico, come il modello di Rash (si veda ad esempio Brentari e Golia, 2008).

Spesso uno stesso questionario è predisposto per la rilevazione di giudizi sia *overall* che *facet*. In questo caso la misura globale ottenuta come sintesi dei singoli *item* può essere confrontata con il giudizio di tipo *overall* espresso dal rispondente. Le due misure potrebbero paradossalmente essere scarsamente correlate poiché, come puntualizzato da alcuni autori (Smith *et al.*, 1969; Scarpello e Campbell, 1983), il giudizio complessivo viene formulato dal rispondente tenendo conto di una grande varietà di fattori, che talvolta esulano dallo stretto ambito lavorativo. Un altro motivo di possibile divergenza tra le due misure consiste nella diversa importanza che i rispondenti attribuiscono ai singoli *item* nel processo inconscio di formazione del giudizio complessivo. Tale importanza può essere misurata attraverso tecniche statistiche, con l'obiettivo di individuare le determinanti di fondo della JS (*driver*). Tale aspetto assume particolare rilevanza nel contesto in cui è ambientato lo studio applicativo qui presentato, focalizzato sui lavoratori delle Cooperative Sociali oggetto dell'indagine ICSI2007 (per i dettagli sull'indagine si veda Borzaga C., 2007). Infatti alcune teorie economiche sostengono che in tale ambiente la JS dei lavoratori segua dinamiche particolari, che coinvolgono la percezione dell'utilità del lavoro svolto più che, ad esempio, gli aspetti economici. Seguendo questa impostazione alcuni autori hanno suggerito di utilizzare i singoli *item* come variabili esogene in modelli di regressione lineare e non, dove la *overall* JS compare come variabile dipendente (Aldag and Brief, 1978; Ferratt, 1981; Allen *et al.*, 2004). Carpita e Zuccolotto (2008) hanno utilizzato, in alternativa ai modelli di regressione tradizionali, modelli algoritmici basati sull'aggregazione di alberi di regressione come le Random Forest (Breiman, 2001a) e il TreeBoost (Friedman, 2001), estrapolando da essi delle misure di importanza dei singoli predittori (*Predictive Importance* e *Constructive Importance*), utilizzate per misurare l'impatto dei singoli *item* sulla *overall* JS.

In questo lavoro, seguendo l'impostazione di Carpita e Zuccolotto (2008), la *Constructive Importance* ottenuta tramite l'impiego del TreeBoost viene utilizzata per misurare l'importanza dei singoli *item* di JS e individuare i *driver* della JS nelle Cooperative Sociali analizzate. Successivamente lo stesso approccio di analisi è utilizzato per riconoscere gli *item* di JS che maggiormente influiscono sull'*Effort*, esplorando così anche l'ipotesi secondo la quale maggiori livelli di soddisfazione contribuirebbero a migliorar le prestazioni del lavoratore.

Il Capitolo è organizzato come segue: nel paragrafo 2 sono brevemente richiamati alcuni aspetti metodologici riguardanti i modelli algoritmici utilizzati; nei paragrafi 3 e 4 sono illustrati i risultati dell'applicazione; alcune considerazioni conclusive sono tratte nel paragrafo 5.

2. Richiami metodologici: le *tree-based learning ensemble*

Negli ultimi anni la ricerca nell'ambito del *data mining* si è orientata verso un nuovo approccio di analisi, che si basa sulla considerazione della realtà come un complesso e imperscrutabile insieme di meccanismi, impossibili da descrivere con modelli semplificati (Breiman, 2001b). Tale *black box* richiede l'utilizzo di una nuova classe di modelli, detti algoritmici, in grado di gestire la presenza di un gran numero di variabili interagenti se-

condo meccanismi complicati. Esempi di modelli di questo tipo sono le reti neurali e una classe di predittori di recente introduzione, detti *learning ensembles* (Breiman, 1996b; Friedman e Popescu, 2005), basati sull'aggregazione di un gran numero di funzioni di previsione distinte (*ensemble members*) e sulla loro combinazione lineare per ottenere una previsione unica. *Learning ensembles* possono essere costruite impiegando diversi tipi di funzioni di previsione come *ensemble members*, tuttavia le proposte più diffuse sono quelle che utilizzano gli alberi decisionali CART come funzioni di previsione di base (Friedman et al., 2001). In quest'ambito si possono annoverare le Random Forest (Breiman, 2001a) e l'algoritmo TreeBoost (Friedman, 2001).

I CART (Classification And Regression Trees) furono introdotti da Breiman, Friedman, Olshen e Stone nel 1984 come strumento per la classificazione e per la regressione. Data una variabile Y da prevedere in funzione di un insieme di predittori X (variabili esogene), il meccanismo alla base del funzionamento dei CART consiste nell'utilizzare le variabili esogene per partizionare ripetutamente il campione in due sottocampioni disgiunti, all'interno dei quali la variabile Y presenti minore eterogeneità. Per valutare l'eterogeneità si può fare riferimento a diversi criteri. Tipicamente si utilizza l'indice di Gini o l'entropia di Shannon se Y è categoriale, la varianza o lo scarto quadratico medio se Y è quantitativa. Formalmente la procedura prende avvio da un nodo iniziale contenente l'intero campione, che viene suddiviso in due nodi-figli per mezzo della variabile esogena che assicura la migliore bipartizione, nel senso che genera i due sottocampioni più omogenei con riferimento alla variabile Y . I nodi risultanti vengono a loro volta bipartiti secondo lo stesso criterio e così via fino al raggiungimento di un predefinito criterio per il termine della procedura, che può consistere ad esempio nella definizione di un numero minimo di unità, oltre il quale un nodo non può ulteriormente essere bipartito. A questo punto è normalmente richiesta una procedura di "potatura" dell'albero la cui struttura risulterebbe altrimenti scarsamente interpretabile. Tale operazione viene in genere effettuata ottimizzando una funzione di costo-complessità che mira alla riduzione dell'errore di previsione (possibilmente calcolato su un test set, vale a dire su un insieme di dati non utilizzato per costruire l'albero stesso), tenendo conto di una penalità collegata alla dimensione dell'albero.

I CART hanno costituito un passo importante nella direzione della costruzione di modelli statistici che superassero i limiti di quelli "tradizionali", come ad esempio la regressione lineare multipla (RLM) nel caso di una variabile Y quantitativa. Utilizzando i CART, infatti, si possono modellare anche relazioni di tipo non lineare tra la variabile Y e i predittori, tenendo conto inoltre di eventuali interazioni tra gli stessi, che nei modelli statistici tradizionali vengono in genere trascurate.

Tuttavia l'utilizzo dei CART nelle applicazioni reali ha messo in luce un difetto fondamentale, l'instabilità (Breiman, 1996a), che consiste nell'esistenza, a fronte di un certo *dataset*, di una molteplicità di alberi, molto diversi tra loro in termini di interpretazione, ma molto simili dal punto di vista dell'errore di previsione. Capita sovente che piccole modifiche al *dataset* comportino la costruzione di alberi completamente diversi.

Un altro problema relativo all'utilizzo dei CART è insito nella procedura di potatura, che in genere risente di un certo grado di soggettività e non garantisce completamente dalla naturale tendenza degli alberi a cadere in situazioni di *overfitting*.

L'aggregazione di molteplici CART in una *learning ensemble* consente di superare il problema dell'instabilità e dell'*overfitting* tipici degli alberi, traendo vantaggio allo stesso tempo dai loro principali punti di forza, derivanti dal meccanismo di base della bipartizione del campione.

Le Random Forest (RF) sono costituite dall'aggregazione di una molteplicità di CART diversi, ottenuti tramite l'introduzione di una componente di casualità nel meccanismo della loro costruzione. Ad ogni nodo di ogni albero viene selezionato casualmente un piccolo sottoinsieme di predittori e solo tra di essi avviene la "competizione" per la scelta di quello che assicura la miglior bipartizione. Gli alberi vengono sviluppati fino alla loro dimensione massima e non potati.

A questa procedura viene di solito affiancata la tecnica del *bagging* (Breiman, 1996b), che consiste nell'utilizzo ad ogni albero di un sottoinsieme del *dataset* originale, selezionato casualmente.

La previsione della RF si ottiene come media delle previsioni dei singoli alberi.

La filosofia che sta alla base di tale procedura è semplice: si cerca di perturbare al massimo la costruzione dei CART in modo da consentire la massima espressione della loro instabilità, che dovrebbe poi essere neutralizzata dalla combinazione di tutte le previsioni nella loro media. Un complesso apparato probabilistico sottostante a questa semplice idea di base ha consentito di dimostrare alcuni importanti teoremi che assicurano il funzionamento di tale procedura, che risulta inoltre esente da problemi di *overfitting* (Breiman, 2001a).

La tecnica del Gradient Boosting Machine (GBM – Friedman, 2001) è una particolare *learning ensemble* basata sull'accostamento sequenziale di modelli di previsione (*ensemble members*) ai cosiddetti pseudo-residui, rappresentati dal gradiente della funzione di perdita prescelta. Quando tale funzione di perdita è di tipo quadratico, il suo gradiente è costituito dai residui di previsione. Nel caso particolare in cui gli alberi decisionali CART siano utilizzati come *ensemble members*, si ha l'algoritmo chiamato TreeBoost (TB) o, nel caso di una variabile Y quantitativa, Multiple Additive Regression Trees (MART). Con la tecnica del TB associata a una funzione di perdita quadratica, quindi, si parte accostando un CART (tipicamente di piccole dimensioni) al *dataset* iniziale, si ricavano i residui di previsione e si accosta a questi un secondo CART, e così via iterando il procedimento un gran numero di volte. La funzione di previsione finale risulta costituita dalla somma pesata delle previsioni effettuate ad ogni passo. Anche con il TB è possibile applicare una tecnica analoga al *bagging*, che consente di randomizzare di volta in volta il *training set* migliorando la *performance* complessiva del modello. Quando si utilizza l'algoritmo TB, come accade con molte altre tecniche algoritmiche per il *data mining*, vi è il rischio di cadere in situazioni di *overfitting*, per cui è utile controllare l'andamento dell'errore di previsione su un *dataset* differente da quello impiegato per l'accostamento del modello, facendo ricorso all'utilizzo di un *test set*, oppure a procedure *out-of-bag* (Breiman, 1996b) o di *cross-validation*.

2.1 Misure di importanza delle variabili esogene

Lo svantaggio principale connesso all'utilizzo di modelli di tipo algoritmico consiste nel loro scarso potere interpretativo della natura dei fenomeni: essi sono in genere delle *black box*, capaci magari di fornire previsioni accurate, ma il cui interno contiene meccanismi completamente indecifrabili. Le *learning ensembles* sono state originariamente pensate in modo da superare questo problema, mediante procedure in grado di estrapolare dall'interno della *black box* alcune informazioni aggiuntive di supporto per l'interpretazione del fenomeno. Nel caso delle RF e del TB vengono calcolate alcune grandezze che misurano l'importanza delle varie variabili esogene nella previsione di Y

(*variable importance measures*, VIM). Le due fondamentali VIM di una generica variabile esogena X_h sono ottenute nel modo seguente (Breiman, 2002):

- M1 – *Predictive Importance*: vengono casualmente permutati i valori assunti da X_h e calcolate le previsioni con questo nuovo *dataset*, in cui l'eventuale associazione tra X_h e Y è completamente distrutta. L'incremento dell'errore di previsione ottenuto con il nuovo *dataset* misura l'importanza di X_h nella previsione di Y .
- M2 – *Constructive Importance*: vengono sommate le diminuzioni di eterogeneità ottenute lungo tutti i nodi di tutti gli alberi che sono stati bipartiti in base alla variabile X_h .

Alcuni recenti studi hanno mostrato che la misura M2 è affetta da una distorsione che tende ad attribuire maggiore importanza alle variabili che consentono per loro natura un maggior numero di possibili bipartizioni del campione (ad esempio le variabili quantitative o le qualitative in scala nominale aventi un gran numero di modalità) o che sono affette da un maggior numero di valori mancanti (Strobl, 2005). L'utilizzo di M2 impone pertanto una preventiva depurazione da tale distorsione, finalità per la quale sono state proposte in letteratura alcune tecniche (Strobl *et. al.*, 2007a, 2007b; Sandri e Zuccolotto, 2008).

La possibilità di calcolare le VIM fa sì che le RF e il TB possano essere utilizzati per la *variable selection*, vale a dire l'identificazione dei predittori più importanti di un dato fenomeno. Svariati studi applicativi mostrano le notevoli potenzialità delle *learning ensembles* in quest'ambito, specialmente in casi in cui pochi importanti predittori si trovano "nascosti" in mezzo a un gran numero di variabili "di disturbo" (si veda ad esempio Sandri e Zuccolotto, 2006, Archer K.J. e Kimes R.V., 2008).

In questo lavoro la misura M2 calcolata con il TB e depurata con la tecnica algoritmica proposta da Sandri e Zuccolotto (2008) viene utilizzata per misurare l'importanza dei singoli *item* della *facet* JS nel processo di formazione del giudizio complessivo rilevato tramite la *overall* JS, e successivamente per valutare il legame esistente tra soddisfazione ed *Effort*.

3. I driver della *Job Satisfaction*

Nel questionario lavoratori remunerati di ICSI2007 erano presenti diversi *item* di *facet* JS e un *item* di *overall* JS. Per ognuno di essi era richiesto al lavoratore di valutare in una scala crescente da 1 a 7 il proprio livello di soddisfazione riguardo il particolare aspetto considerato. La regressione effettuata con l'algoritmo TB utilizza come variabile dipendente la *overall* JS e come variabili esogene gli *item* di *facet* JS, sintetizzati in Tab. 1, unitamente all'indicazione della sfera di riferimento (relazionale, R, estrinseca, E) del tipo di soddisfazione.

Sono state condotte due diverse analisi, la prima (A1) riguardante tutti i 4134 lavoratori intervistati, la seconda (A2) limitata ai 2823 lavoratori che dichiarano di dedicarsi *spesso* o *sempre*, durante il lavoro, alle relazioni con gli utenti e i loro familiari.

È apparso utile effettuare tale distinzione in quanto è verosimile che il meccanismo di formazione della soddisfazione complessiva sia differente a seconda che il lavoratore svolga un lavoro a diretto contatto con gli utenti oppure mansioni amministrative, tecnico-auxiliarie, etc.

In entrambe le analisi è stato utilizzato l'80% del campione come *training set* e il restante 20% come *test set* per valutare il numero di alberi ottimale per costruire la *learning ensemble* e, successivamente, la bontà del modello. I risultati, che evidenziano una bontà di accostamento di medio livello, sono sintetizzati in Tab. 2.

Tabella 1. *Item* di JS utilizzati nel modello come variabili esogene

DESCRIZIONE ITEM	NOME	TIPO
Organizzazione dell'orario di lavoro	Orario	E
Flessibilità dell'orario di lavoro	Flex	E
Stabilità del posto di lavoro	Stab	E
Ambiente fisico di lavoro	Amb	E
Tutele previdenziali e sociali garantite	Tutele	E
Retribuzione complessiva (compresi benefici e varie)	Retrib	E
Coinvolgimento nelle decisioni della Cooperativa	Coinv	E
Trasparenza dei rapporti con la Cooperativa	Trasp	E
Riconoscimento del lavoro da parte della Cooperativa	Ricon Coop	R
Crescita e formazione garantite dalla Cooperativa	Crescita	E
Autonomia e indipendenza nel Suo lavoro	Indip	E
Avanzamenti conseguiti e prospettive di carriera	Carriera	E
Realizzazione personale	Realizz	R
Relazioni con i colleghi	Colleghi	R
Relazioni con i volontari	Volont	R
Relazioni nel gruppo di lavoro	Team	R
Relazioni con i superiori	Super	R
Relazioni con utenti e loro familiari	Utenti	R
Varietà e creatività	Variet	R
Riconoscimento dei colleghi	Ricon Coll	R
Riconoscimento sociale	Ricon Sociale	R
Riconoscimento degli utenti	Ricon Utenti	R
Utilità del lavoro per utenti e loro familiari	Utilit	R
Coerenza tra lavoro e formazione complessiva	Coerenza	E

Tabella 2. Risultati della regressione con la tecnica TB

ANALISI	NUMERO DI ALBERI	VARIANZA SPIEGATA (%)
A1	1649	46,29
A2	1224	45,41

Il risultato più interessante è dato dal calcolo delle misure di importanza dei predittori, che fornisce un'indicazione riguardo i *driver* della JS. Nella Fig. 1 sono rappresentate graficamente le graduatorie dei predittori secondo la misura di importanza M2, depurata con la tecnica algoritmica proposta da Sandri e Zuccolotto (2008), che consiste nell'introduzione nel *dataset* di un insieme di predittori "fittizi", indipendenti dalla variabile risposta, in grado di misurare l'entità della distorsione.

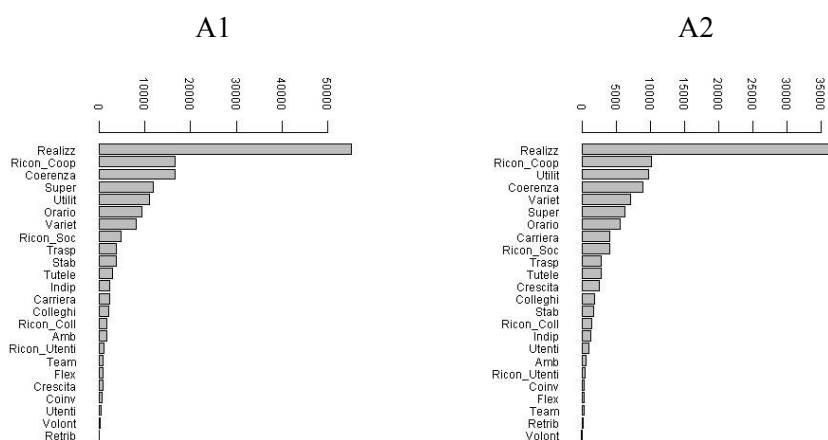


Figura 1. Graduatorie dei predittori secondo la misura M2 depurata dal *bias*.

È molto interessante notare che la Realizzazione Personale (Realizz) costituisce il *driver* largamente più importante in entrambe le analisi, seguito a una certa distanza dal Riconoscimento della Cooperativa per il lavoro svolto (Ricon_Coop). Si tratta in entrambi i casi di *driver* che fanno riferimento alla sfera relazionale. Sui *driver* successivi emergono le differenze tra il campione complessivo e il campione ristretto a chi si relaziona con gli utenti: nel primo caso viene data importanza alla Coerenza tra lavoro e formazione complessiva (Coerenza) e alle Relazioni con i superiori (Super), nel secondo caso acquista importanza, come prevedibile, prima di tutto l'Utilità del lavoro per utenti e loro familiari (Utilit) e successivamente la Coerenza tra lavoro e formazione complessiva. La Retribuzione complessiva figura tra i *driver* meno importanti, coerentemente con quanto affermato dalle teorie economiche sul lavoro nel contesto nonprofit.

È importante sottolineare che non vi è relazione tra l'entità della soddisfazione riguardo un certo *item* e la sua importanza nel determinare la soddisfazione complessiva. Un *item* poco importante potrebbe indifferentemente essere largamente soddisfacente o insoddisfacente per tutti i lavoratori e ciò non va interpretato pensando che essi si disinteressino completamente di tale aspetto del lavoro. Considerando, ad esempio, proprio la Retribuzione complessiva, tale elemento tende ad essere scarsamente soddisfacente per i lavoratori, ciononostante l'*item* figura tra gli ultimi in ordine di importanza. Questo non è un'incongruenza, e neppure va inteso come disinteresse completo dei lavoratori riguardo il riconoscimento monetario del proprio lavoro, ma significa semplicemente che l'insoddisfazione riguardo la Retribuzione non pregiudica l'eventuale soddisfazione complessiva per il lavoro. Normalmente, per come sono strutturati il modello e il calcolo delle relative misure di importanza, i predittori poco importanti sono caratterizzati da una delle due seguenti proprietà:

- scarsa variabilità (cioè largo accordo tra lavoratori nei riguardi dell'*item*: i soggetti tendono ad essere per la maggior parte soddisfatti o insoddisfatti)
- elevata variabilità ma indipendenza dalla JS complessiva (esiste eterogeneità tra i lavoratori nella soddisfazione riguardo l'*item*, tuttavia non si rileva connessione tra il livello di soddisfazione per l'*item* e la JS complessiva).

Per contro, gli *item* più importanti sono caratterizzati da eterogeneità tra i lavoratori e, al tempo stesso, connessione tra il livello di soddisfazione per l'*item* e la JS complessiva.

L'analisi può essere approfondita esaminando la variazione della JS complessiva in funzione dei predittori più importanti. Per quanto riguarda la Realizzazione personale e il Riconoscimento della Cooperativa per il lavoro svolto, non emergono differenze di rilievo tra l'analisi A1 e A2. Per questo motivo vengono riportati i grafici relativi solamente alla prima delle due (Fig. 2).

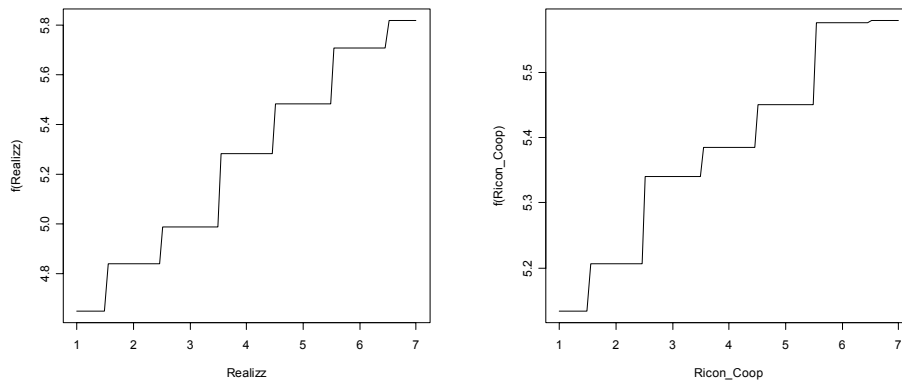


Figura 2. Andamento della JS complessiva in funzione dei *driver* Realizz (sinistra) e Ricon_Coop (destra).

Entrambi i *driver* hanno un impatto uniforme sulla JS complessiva, nel senso che a livelli di soddisfazione crescente riguardo il *driver* corrispondono livelli gradualmente crescenti di JS complessiva, con incrementi pressoché costanti.

Le analisi A1 e A2 si differenziano invece per quanto riguarda l'Utilità del lavoro svolto (Fig. 3)

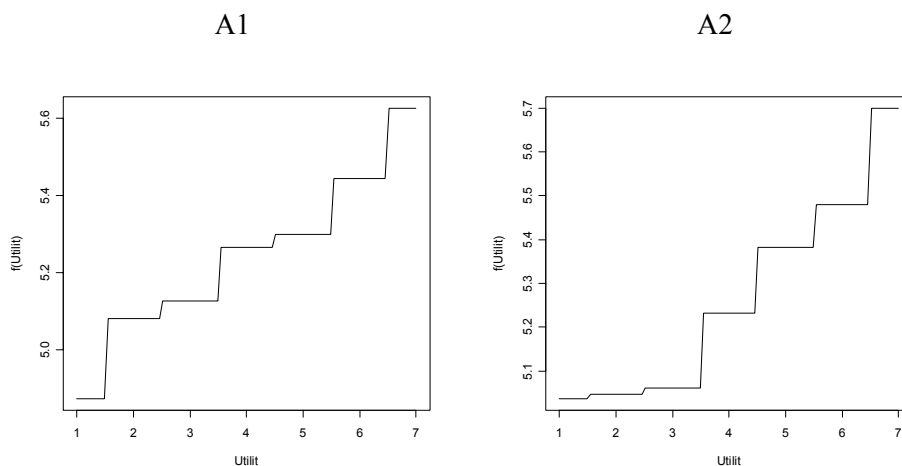


Figura 3. Andamento della JS complessiva in funzione del *driver* Utilit.

Nell'analisi A1 il *driver* Utilit presenta incrementi non costanti e determina livelli di JS complessiva più contenuti rispetto a quelli dell'analisi A2. Anche in quest'ultima la

relazione non è caratterizzata da incrementi costanti: il *driver* inizia ad avere influenza sulla JS complessiva a partire dal livello di soddisfazione 3 ed ha un forte incremento in corrispondenza del livello 7.

4. La relazione tra *Job Satisfaction* ed *Effort*

Con il termine *Effort* si intende l'impegno complessivamente profuso dal lavoratore nello svolgimento del proprio lavoro. Nel questionario sono state previste alcune domande deputate alla rilevazione di tale grandezza. Anche in questo caso, tra queste, era presente una domanda che chiedeva direttamente al lavoratore di valutare, in una scala crescente da 1 a 7, quanto si impegna nel proprio lavoro (variabile *Effort*). Come facilmente prevedibile, la variabilità delle risposte è risultata piuttosto scarsa: il 62,6% dei rispondenti ha autovalutato il proprio impegno con il punteggio massimo (7), il 27,8% e il 7,4% rispettivamente con uno e due punti in meno (6 e 5) e solo il 2,2% si è attribuito un punteggio più basso, senza comunque scendere al di sotto del 3. Per valutare l'esistenza di un'influenza, postulata da alcune teorie economiche (Borzaga e Musella, 2006), della JS sull'*Effort*, è stata effettuata una regressione con la tecnica del TB, utilizzando come variabile dipendente l'autovalutazione dell'impegno complessivo da parte del lavoratore. La ridotta variabilità delle risposte fa sì che il modello sia difficilmente in grado di cogliere i rari valori estremi, quindi bisogna aspettarsi una globale bontà di adattamento piuttosto ridotta. Poiché non sono emerse differenze di rilievo tra le due analisi A1 e A2, di seguito saranno presentati i risultati relativi all'intero campione (Fig. 4 e Tab. 3).

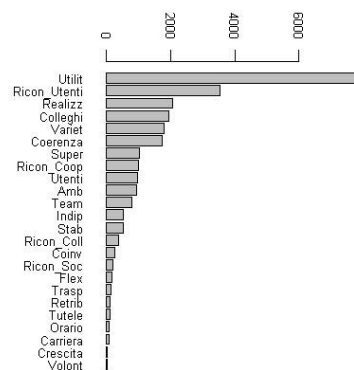


Figura 4. Graduatorie dei predittori secondo la misura M2 depurata dal *bias*.

Tabella 3. Risultati della regressione con la tecnica TB

ANALISI	NUMERO DI ALBERI	VARIANZA SPIEGATA (%)
A1	1175	23,32

Come atteso, la varianza spiegata indica che il modello si presta in modo scarsamente soddisfacente a spiegare il fenomeno in questione. In ogni caso risulta interessante, a scopo indicativo, notare che i due *item* di JS largamente più importanti nel determinare

L'Effort risultano essere l'Utilità del lavoro per utenti e loro familiari (Utilit) e il Riconoscimento degli utenti (Ricon_Utenti). Anche in questo caso si tratta di due *driver* di tipo relazionale. Le misure di importanza dei predittori risultano scarsamente correlate con quelle ottenute nell'analisi precedente (il coefficiente di correlazione lineare è pari a 0,26) e questo potrebbe indicare che l'Effort è collegato più che alla dimensione complessiva della JS, ad alcuni suoi aspetti specifici, rappresentati nella fattispecie dall'utilità del proprio lavoro per gli utenti e dal riconoscimento ottenuto da parte degli stessi.

Anche qui può essere interessante esaminare la variazione dell'Effort complessivo in funzione dei predittori più importanti (Fig. 5)

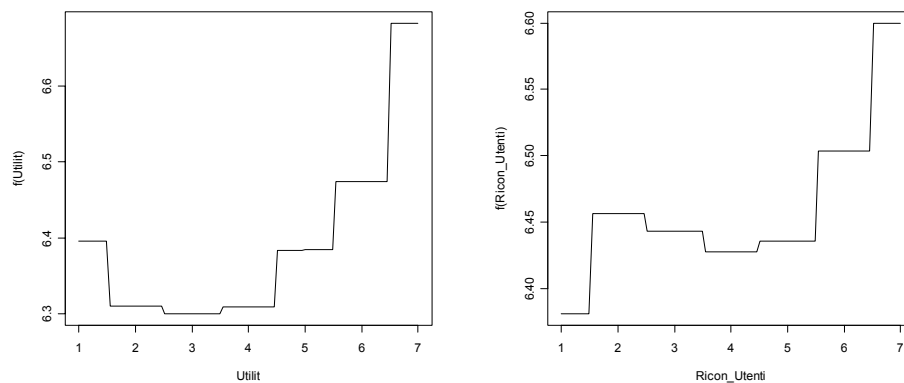


Figura 5. Andamento dell'Effort in funzione dei *driver* Utilit (sinistra) e Ricon_Utenti (destra).

Al di là di alcune anomalie nei valori della coda sinistra (livelli di soddisfazione 1 e 2), dovute essenzialmente alla scarsità di soggetti che presentano tali valori, si nota sostanzialmente che i due *driver* iniziano ad avere un impatto sull'Effort a partire rispettivamente dai livelli di soddisfazione 5 e 6. Ciò significa che l'impegno del lavoratore migliora significativamente quando la soddisfazione che egli ricava relativamente ai due elementi considerati è molto elevata.

5. Differenziali di Effort a parità di Job Satisfaction

Per integrare l'analisi è possibile analizzare più in profondità la relazione tra l'Effort e le caratteristiche del lavoratore e del tipo di Cooperativa in cui opera. Una volta evidenziato che esiste un effettiva influenza dei *driver* di JS sull'Effort, si può prendere in considerazione quanto, e in quale direzione, l'Effort profuso dai singoli lavoratori si scosta dal livello medio corrispondente al loro grado di soddisfazione. In altre parole si tratta di analizzare le relazioni esistenti tra i residui della regressione descritta nel paragrafo precedente (che qui chiameremo "differenziali di Effort") e alcune caratteristiche dei lavoratori e delle Cooperative. Tale analisi può essere svolta utilizzando l'Analisi delle Corrispondenze Multiple (ACM), una tecnica ben nota di Analisi dei Dati, che si applica quando si dispone di variabili qualitative tra le quali sussiste un certo grado di connessione, per evidenziare quali sono le modalità maggiormente collegate tra loro (per maggiori det-

tagli si veda ad esempio Bolasco, 1999). Uno degli *output* dell'analisi consiste nella rappresentazione grafica, di solito bidimensionale, di uno spazio cartesiano in cui sono localizzati congiuntamente i soggetti analizzati e le modalità delle variabili oggetto di studio. Le caratteristiche dei soggetti e i legami tra le modalità si interpretano osservando la vicinanza tra i punti che li rappresentano. In questo tipo di analisi accade comunemente che molti soggetti siano localizzati nello stesso punto dello spazio in questione, quindi per mettere in luce la presenza di più soggetti sovrapposti, si usa rappresentarli anziché con un punto, con una bolla di dimensioni proporzionali alla numerosità dei soggetti che vi ricadono.

Nel caso in questione sono state analizzate le relazioni tra i differenziali di *Effort* e svariate combinazioni delle caratteristiche dei lavoratori e delle Cooperative. I risultati più significativi sono stati ottenuti utilizzando le quattro variabili riportate in Tab. 4, per le quali l'inerzia spiegata in due dimensioni è risultata pari al 56,2%.

Tabella 4. Variabili qualitative analizzate con ACM

VARIABILE	MODALITÀ
Diff_Effort_qual (Differenziali di <i>Effort</i>)	<i>Effort</i> molto inferiore alla media (residui ≤ -1) <i>Effort</i> leggermente inferiore alla media ($-1 < \text{residui} \leq 0$) <i>Effort</i> leggermente superiore alla media ($0 < \text{residui} \leq 1$) <i>Effort</i> molto superiore alla media (residui > 1)
Sex (<i>Sesso</i>)	Maschio Femmina
Tipo_Coop (<i>Tipo di Cooperativa</i>)	A (servizi socio-assistenziali) B (formazione e inserimento nel mondo del lavoro)
Num_dip (Numero di dipendenti della <i>Cooperativa</i>)	15 lavoratori o meno da 16 a 49 lavoratori 50 lavoratori o più

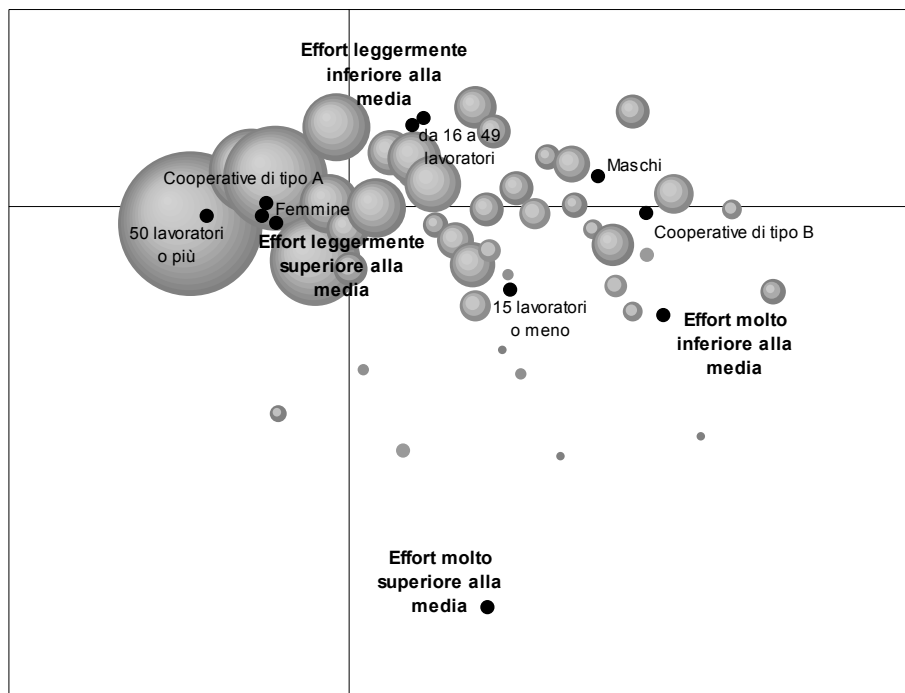


Figura 6. Rappresentazione bidimensionale ottenuta con ACM

La rappresentazione bidimensionale di Fig. 6 indica prima di tutto che la maggior parte dei lavoratori è localizzata nella zona caratterizzata da un *Effort* leggermente superiore alla media. Tale caratteristica trova elevata connessione con le modalità A, Femmina, 50 lavoratori o più, rispettivamente delle variabili Tipo_Coop, Sex, Num_dip. Gruppi residuali di lavoratori si trovano dispersi nelle zone caratterizzate da livelli di *Effort* leggermente o molto inferiori alla media. Qui si trovano localizzate anche le modalità che si riferiscono a Cooperative di tipo B, di dimensione più contenuta (fino a 49 dipendenti) e a lavoratori maschi.

6. Considerazioni conclusive

Scopo di questo lavoro era esplorare i diversi ambiti e le differenti sfaccettature della soddisfazione sul lavoro, individuando quali aspetti avessero maggior impatto sulla soddisfazione complessiva del lavoratore e sul suo impegno nello svolgimento del proprio lavoro. L'idea di base è derivata da alcuni postulati della teoria economica, secondo i quali i lavoratori del settore nonprofit sarebbero influenzati da fattori di tipo relazionale e legati a un certo livello di impegno sociale, più che ad aspetti meramente retributivi. Inoltre, sempre secondo alcune teorie economiche, a più elevati livelli di soddisfazione corrisponderebbe un maggior impegno sul lavoro ed un miglioramento complessivo delle prestazioni. L'evidenza empirica ha confermato in parte le attese, in quanto i fattori di tipo retributivo non sembrano avere un significativo impatto sulla soddisfazione complessiva dei lavoratori, a differenza di altri, legati alla realizzazione personale sul lavoro, che sono risultati predominanti. Di sicuro interesse è notare che i lavoratori che operano a stretto contatto con gli utenti e con le loro famiglie tendono ad attribuire grande importanza an-

che all'utilità del loro lavoro per i beneficiari del servizio. Dal punto di vista della relazione tra soddisfazione ed impegno, essa ha trovato conferma nei dati esaminati, anche se sembra essere prevalente, da questo punto di vista, non tanto la soddisfazione complessiva, quanto la percezione di utilità del proprio lavoro e il riconoscimento dello stesso da parte degli utenti. Anche questa considerazione appare di sicuro interesse nell'ambito degli studi del settore nonprofit.

Bibliografia

- Aldag R.J., Brief A.P. (1978), Examination of alternative models of job satisfaction. *Human Relations*, 31, 1, 91-98.
- Allen R.I., Lambert E.G., Pasupuleti S., Cluse-Tolar T., Ventura L.A. (2004), The impact of job characteristics on social and human service workers. *Social Work & Society*, 2, 2: 173-188.
- Archer K.J., Kimes R.V., (2008), Empirical characterization of Random Forest variable importance measures, *Computational Statistics & Data Analysis*, 52, 2249-2260.
- Bolasco S., *Analisi multidimensionale dei dati*, Carocci Editore, 1999.
- Borzaga C., Musella M., (2006), Sistemi di incentivi e soddisfazione per il lavoro nel settore dei servizi di pubblica utilità: una prospettiva di teoria economica, in Carpita M., D'Ambra L., Vichi M., Cittadini G. (editors), *Valutare la Qualità – I servizi di pubblica utilità alla persona*, Guerini Studio.
- Borzaga C. (editor) (2007), Quando le risorse umane fanno la differenza: il modello imprenditoriale delle cooperative sociali. Primi risultati di ICSI 2007: la nuova Indagine sulle Cooperative Sociali Italiane, *Impresa Sociale*, 3, http://www.issan.info/doc/lugset07Impresasociale_colori.pdf.
- Breiman L., (1996a), The heuristic of instability in model selection, *Annals of Statistics*, 24, 6, 2350-2383.
- Breiman L., (1996b), Bagging Predictions, *Machine Learning*, 24, 2, 123-140.
- Breiman L., (2001a), Random Forests, *Machine Learning*, 45, 1, 5-32.
- Breiman L., (2001b), Statistical Modeling: The Two Cultures, *Statistical Science*, 16, 3, 199-231.
- Breiman L., (2002), Manual on setting up, using, and understanding random forests v3.1, <http://oz.berkeley.edu/users/breiman>.
- Breiman L., Friedman J.H., Olshen R.A., Stone C.J., (1984), *Classification and Regression Trees*, Chapman & Hall.
- Brentari E. e Golia S. (2008), Measuring job satisfaction in the social service sector with the Rasch model, *Journal of Applied Measurement*, 9, 45-56.
- Carpita M., Zuccolotto P. (2008) Mining the Drivers of Job Satisfaction using Algorithmic Variable Importance Measures, in D'Ambra L., Rostirolla P., Squillante M. (editors), *Metodi, Modelli, e Tecnologie dell'Informazione a Supporto delle Decisioni*, parte I: Metodologie, FrancoAngeli.
- Ferratt T.W. (1981), Overall job satisfaction: it is a linear function of facet satisfaction? *Human Relations*, 34, 6: 463-473.
- Friedman J.H., (2001), Greedy function approximation: a gradient boosting machine, *Annals of Statistics*, 29, 1189-1232.

- Friedman J.H., Hastie T. and Tibshirani R., (2001), *The Elements of Statistical Learning; Data Mining, Inference and Prediction*, Springer, New York.
- Friedman J.H., and Popescu B.E., (2005), *Predictive Learning via Rule Ensembles*, Technical report, Stanford University.
- Manisera M. (2005) *Measuring job satisfaction by mean of nonlinear Principal Component Analysis*, PhD Thesis.
- Sandri M., Zuccolotto P., (2006), *Variable Selection Using Random Forests*, in Zani S., Cerioli A., Riani M. e Vichi M. (editors), *Data Analysis, Classification and Forward Search*, Springer, Berlino.
- Sandri M., Zuccolotto P., (2008) *A bias correction algorithm for the Gini measure of variable importance*, accettato per la pubblicazione su *The Journal of Computational & Graphical Statistics*.
- Strobl, C. (2005) *Statistical Sources of Variable Selection Bias in Classification Trees Based on the Gini Index*. Technical Report, SFB 386, http://epub.ub.uni-muenchen.de/archive/00001789/01/paper_420.pdf.
- Strobl, C., Boulesteix, A.-L., Zeileis, A. and Hothorn, T. (2007a) *Bias in Random Forest Variable Importance Measures: Illustrations, Sources and a Solution*, *BMC Bioinformatics*, 8-25, doi:10.1186/1471-2105-8-25.
- Strobl, C., Boulesteix, A.-L. and Augustin, T. (2007b), *Unbiased split selection for classification trees based on the Gini Index*, *Computational Statistics & Data Analysis*.
- Scarpello V., Campbell J.P. (1983), *Job satisfaction: are all the parts there?*, *Personnel Psychology*. 36, 577-600.
- Smith P.C., Kendall L.M., Hulin C.L. (1969), *The measurement of satisfaction in work and retirement*. Rand McNally, Chicago.