

Etica nei dati: impatto sugli algoritmi di decisione

Progetto Ingegneria Informatica

Anno accademico 2023/2024

Leonardo Liparulo, Francesco Antonio Mazzola, Nicola Tummolo

985106 ,981413,986523

Tutor: Prof. Letizia Tanca

Introduzione

L'analisi dei dati basata su algoritmi di machine learning ha rivoluzionato il modo in cui prendiamo decisioni in svariati ambiti. Tuttavia, è fondamentale garantire che tali algoritmi non perpetuino discriminazioni già presenti nella società. Infatti nell'era digitale odierna, la gestione delle informazioni assume un ruolo sempre più centrale, intrecciandosi inevitabilmente con delicate questioni etiche. L'avvento dell'Intelligenza Artificiale (IA) e dei sistemi di apprendimento automatico amplifica ulteriormente questo scenario, ponendoci di fronte a nuove sfide e responsabilità e sollevando interrogativi etici di primaria importanza. Se ogni disciplina, sia essa scientifica o sociale, si basa su codici di comportamento ben definiti, monitorati da organismi professionali e dalla società civile, nell'ambito dell'IA tali codici risultano ancora in fase di sviluppo. Già negli anni '50 Norbert Wiener affrontava queste tematiche nell'ambito delle "cyber technologies", e numerosi documenti emanati da istituzioni professionali e regolatorie hanno formulato principi deontologici da seguire; tuttavia, la mancanza di un consenso unanime su tali regole etiche ostacola la loro completa applicazione. Un esempio emblematico riguarda la discriminazione di genere: se da un lato è ampiamente riconosciuto che tale fattore non dovrebbe influenzare le decisioni, dall'altro, situazioni come ad esempio i naufragi pongono dilemmi complessi. Infatti, dare priorità a donne e bambini, sebbene motivato dalla preservazione della specie, rappresenta pur sempre una discriminazione nei confronti degli uomini mentre, analogamente, le procedure di reclutamento nei posti di lavoro spesso svantaggiano le donne, basandosi sul pregiudizio che esse siano meno intelligenti degli uomini, o più portate per il lavoro di casalinga.

Come evidenziato dagli esempi sopracitati, le regole etiche non sono entità assolute, ma dipendono dal contesto e dalle circostanze specifiche. Le decisioni basate su algoritmi di machine learning o metodi di data science possono presentare criticità etiche intrinseche all'algoritmo stesso o ai dati utilizzati. Fondandosi su esperienze e dati passati, tali sistemi rischiano di perpetuare pregiudizi e concezioni obsolete, in contrasto con i valori etici odierni. Diventa quindi cruciale un attento vaglio dei dati impiegati, al fine di ottenere soluzioni "su misura" eticamente equilibrate *per il contesto specifico di utilizzo*. Ad esempio, un sistema per selezionare i migliori studenti per una borsa di studio, basato su dati degli ultimi 15 anni, potrebbe evidenziare una maggioranza di studenti caucasici tra i migliori candidati. Per garantire l'equità verso tutti i gruppi etnici, si potrebbe eliminare dai dati di input l'informazione sull'etnia. Ovviamente, a differenza degli esseri umani, i computer non possiedono la capacità innata di riconoscere e adattarsi al contesto; pertanto, diviene necessario sviluppare modelli in grado di rappresentare i molteplici contesti possibili e di "insegnare" ai computer il comportamento etico da adottare in ogni situazione.

Lo studio "Introducing Context-Aware, Ethical Data Transformation" [1] evidenzia la necessità di adottare un approccio etico nella trasformazione dei dati per prevenire distorsioni e discriminazioni. Il nostro progetto, partendo dalle considerazioni in esso contenute, ha l'obiettivo di prevenire la discriminazione nella selezione dei candidati a una promozione per il ruolo di manager. A tal fine,

abbiamo implementato in uno script Python algoritmi di selezione basati su diverse dimensioni etiche: *equality* – che consiste nel fornire a ciascuno le stesse opportunità -, *equity* – che consiste nel riconoscere a ciascuno la propria specificità, favorendo le persone più sfavorite - e *diversity* – cioè uniformare il grado con il quale diverse entità sono rappresentate nel dataset -. L'utente può scegliere la dimensione etica, l'attributo influenzato (ad esempio, genere, etnia) e il numero di candidati da promuovere. Il sistema seleziona i candidati da promuovere in base ai criteri scelti, generando un file .csv con i risultati; questi sono poi visualizzati tramite diagrammi a torta che evidenziano i trend a seconda della dimensione etica scelta.

E' interessante notare che i nostri risultati dimostrano che ogni scelta etica comporta un compromesso. Ad esempio, applicare il criterio della diversità garantisce una rappresentanza equilibrata di generi ed etnie tra i candidati selezionati, ma può comportare una riduzione del rendimento medio. Nel report vengono presentati in dettaglio i risultati ottenuti applicando le diverse dimensioni etiche, insieme a considerazioni personali sull'impatto di ogni scelta. Inoltre, vengono illustrati i diversi *Context Dimensions Tree* (CDT) e *Ethical Requirements Tree* (ERT) in base alla dimensione etica e all'attributo influenzato scelti. Il Context Dimension Tree è una struttura gerarchica che organizza e rappresenta le dimensioni e i valori del contesto. La radice del CDT è un nodo radice bianco, da cui si diramano nodi neri che rappresentano le dimensioni del contesto, ovvero le prospettive rilevanti per contestualizzare i dati. Ogni nodo di dimensione può avere come figli nodi bianchi, che rappresentano i valori associati a quella dimensione (nodi contesto). Ad esempio, una dimensione potrebbe essere il ruolo di una persona, con nodi bianchi associati come "impiegato" e "manager". I nodi concetto possono a loro volta contenere sotto-dimensioni, rappresentate anch'esse da nodi neri. Questa struttura gerarchica può essere ripetuta fino a raggiungere il livello di dettaglio desiderato. Sia i nodi di dimensione che quelli concetto possono avere attributi, rappresentati da piccoli nodi quadrati. Un attributo collegato a un nodo di dimensione serve come abbreviazione quando i possibili figli di quella dimensione sono molti. Ad esempio, se la dimensione "ruolo" ha molte istanze possibili, queste possono essere rappresentate da un singolo attributo che può assumere come valore uno dei possibili nodi bianchi. Gli attributi collegati a un nodo bianco rappresentano istanze specifiche all'interno dell'insieme di valori che il nodo bianco rappresenta. In sintesi, un Context Dimension Tree (CDT) permette di organizzare e rappresentare in modo gerarchico le dimensioni e i valori del contesto, consentendo una contestualizzazione dettagliata e strutturata dei dati. L'adozione di CDT nella preparazione dei dati da sottoporre ai sistemi di intelligenza artificiale e di apprendimento automatico rappresenta un passo fondamentale per garantire un loro sviluppo e utilizzo etico e responsabile.

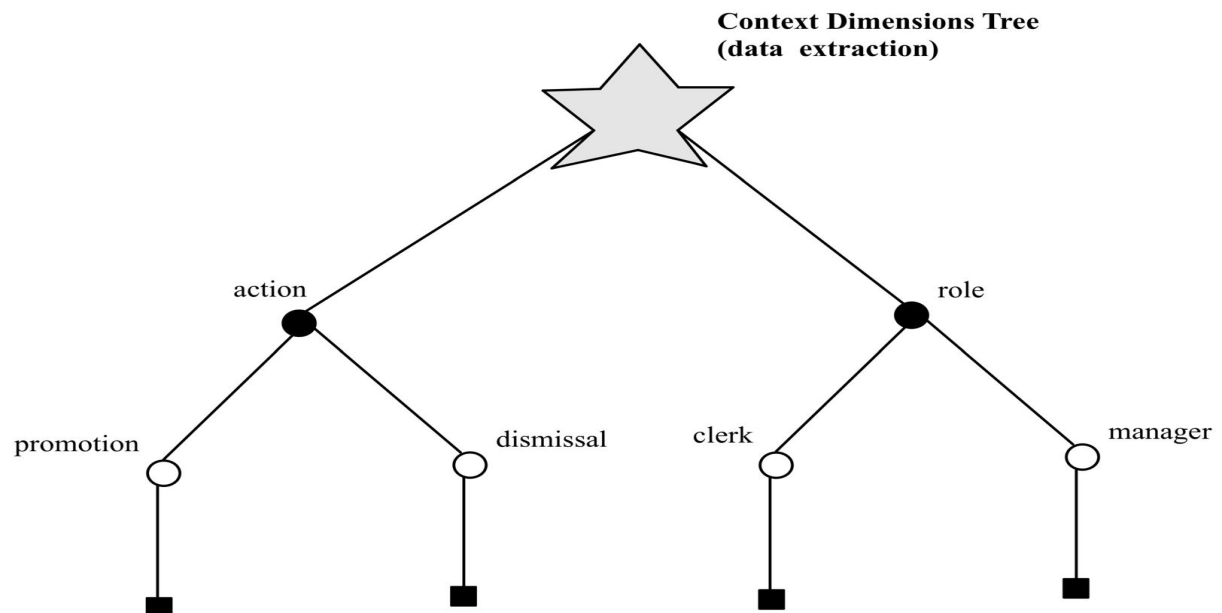
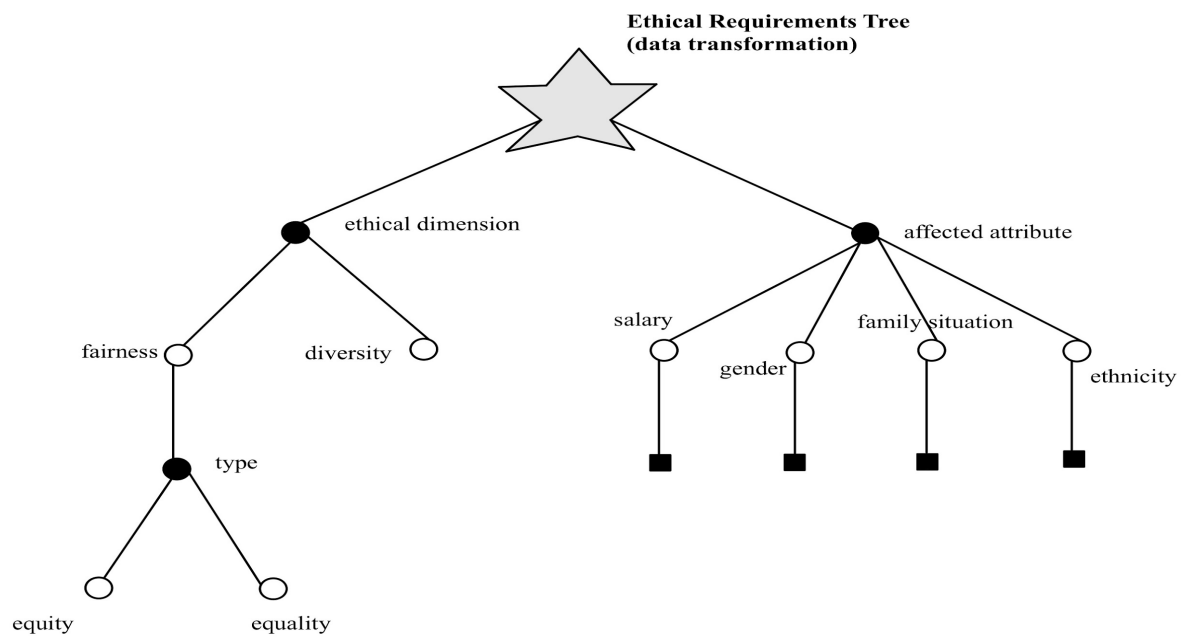
Per rappresentare le dimensioni etiche che guidano la trasformazione etica, utilizziamo un modello simile al CDT, l'*Ethical Requirements Tree* (ERT). L'ERT è un albero dei requisiti etici con due nodi neri che rappresentano:

- Ethical dimension: dimensioni etiche rilevanti per la situazione in questione. Nel nostro caso Fairness e Diversity.

- Affected attribute: l'attributo del dataset di input per il quale vogliamo garantire la conformità con la o le dimensioni etiche scelte.

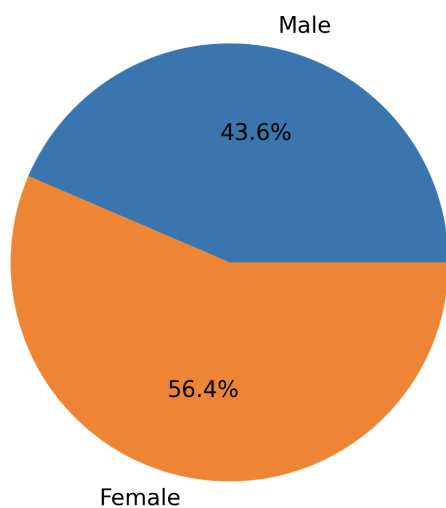
Il nostro progetto dimostra l'importanza di considerare le implicazioni etiche nella trasformazione dei dati per prevenire la discriminazione. La scelta di una dimensione etica comporta inevitabilmente dei

compromessi, e la valutazione di questi compromessi è fondamentale per prendere decisioni etiche e consapevoli.

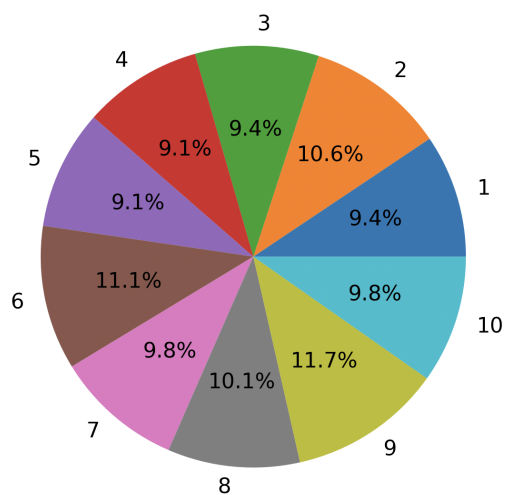


Di seguito una rappresentazione dei dati di partenza mediante grafico a torta della composizione dei **Clerk** in base a genere e indice di performance (Max=10):

Distribuzione per genere

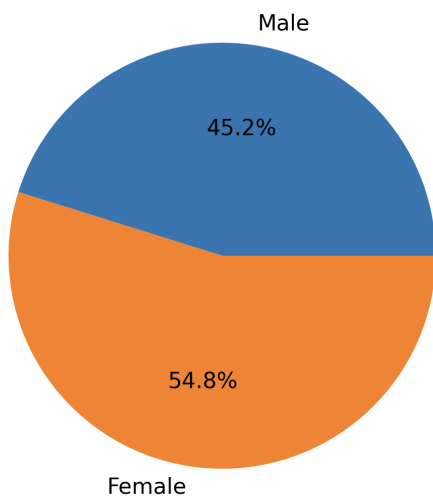


Distribuzione per indice performance.
La media è 5.59

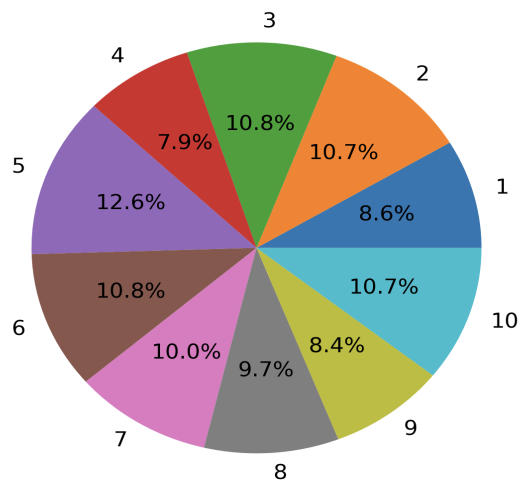


Di seguito una rappresentazione negli stessi dati della composizione dei **Manager** in base a genere e indice di performance:

Distribuzione per genere

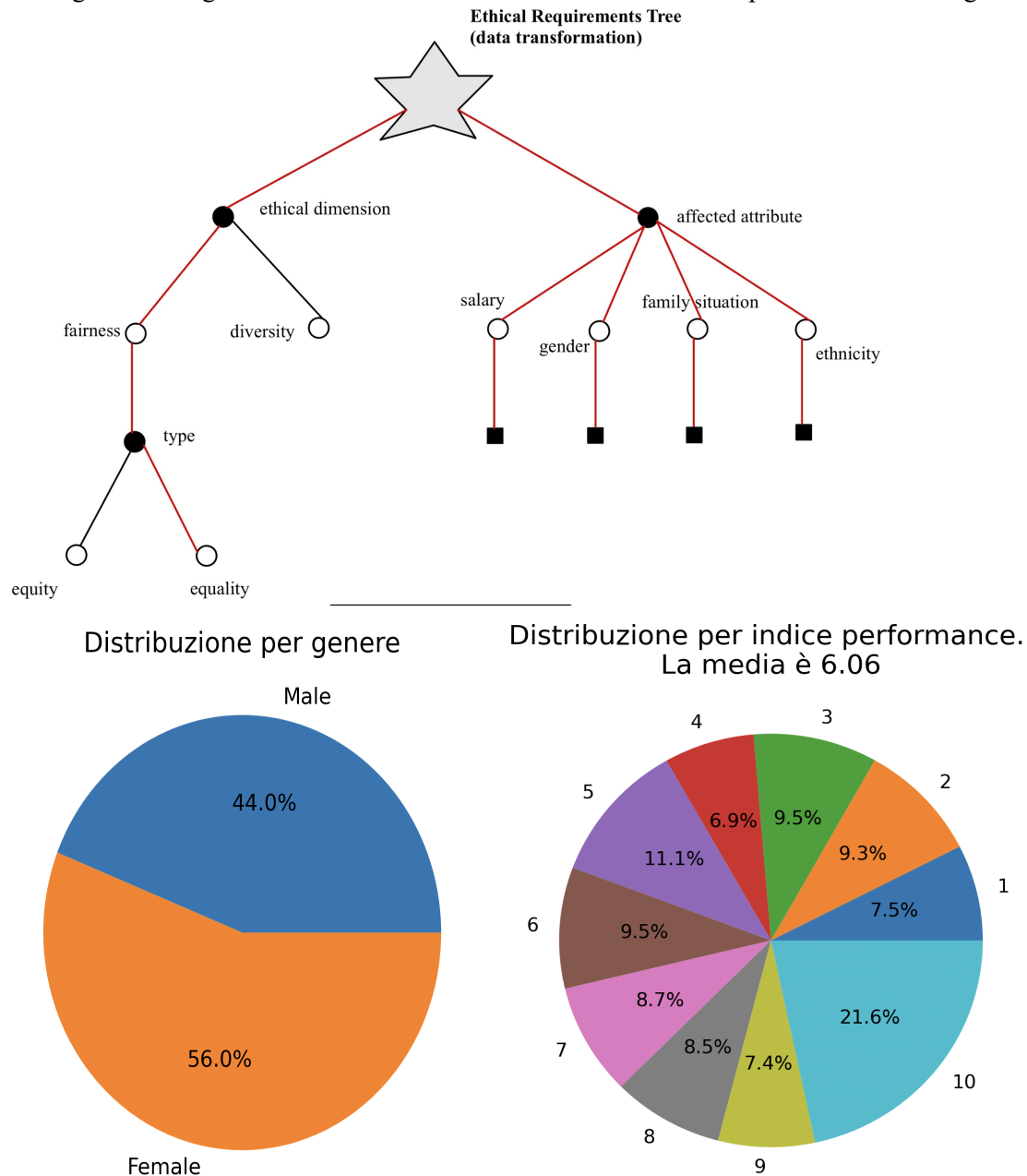


Distribuzione per indice performance.
La media è 5.51



Equality

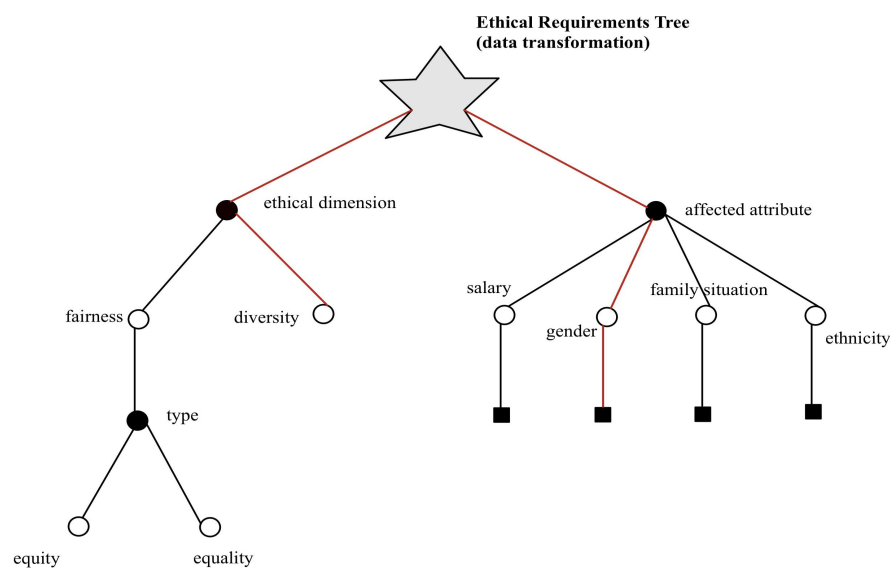
L'equality si basa sul principio di meritocrazia, selezionando i candidati con il rendimento più alto ("performance rate"). Questo approccio garantisce che i manager promossi siano i più qualificati per il ruolo, ma non tiene conto della rappresentanza di diversi gruppi. Ad esempio, se il dataset contenesse un numero significativamente maggiore di uomini rispetto a donne, l'approccio basato sulla meritocrazia potrebbe portare a una selezione prevalentemente maschile. Questo potrebbe avere un impatto negativo sulla diversità del team di manager e sulla sua capacità di comprendere e rappresentare le esigenze di un gruppo di lavoro eterogeneo. Nello script Python, per selezionare secondo il principio di equality, i clerk sono stati ordinati per performance rate, senza considerare etnia o genere. I migliori sono stati inseriti nella lista dei candidati da promuovere a manager.



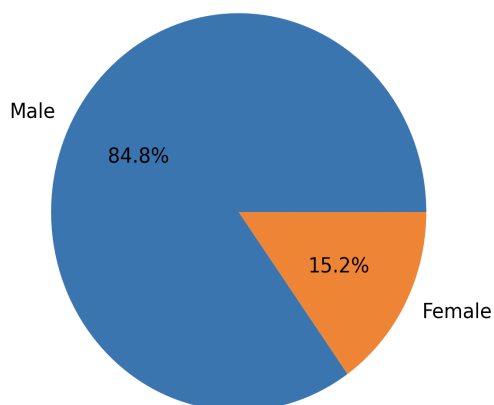
Diversity

La diversity si concentra sulla rappresentanza di diversi gruppi etnici all'interno del team di manager. Per raggiungere questo obiettivo, ci si assicura che il numero di candidati promossi per ogni gruppo etnico sia uguale. Ad esempio, se il dataset contiene il 30% di donne e il 70% di uomini, l'approccio basato sulla diversity garantirà che il 70% dei manager promossi sia donna e il 30% sia uomo. Questo metodo permette di creare un team di manager più eterogeneo, in grado di comprendere e rappresentare meglio le esigenze di un gruppo di lavoro diversificato. Tuttavia, questo approccio può comportare una diminuzione del rendimento medio dei manager promossi, in quanto non vengono selezionati esclusivamente i candidati con il rendimento più alto. Per ovviare a questo problema, si potrebbero utilizzare criteri di selezione ponderati che considerino sia il rendimento che la diversità, oppure si potrebbero implementare programmi di formazione per i manager promossi.

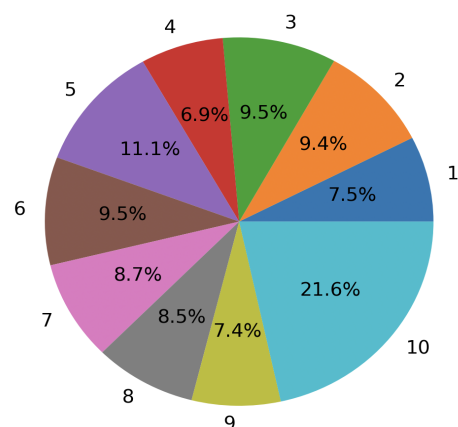
Nello script per selezionare secondo diversity abbiamo selezionato in parti uguali i migliori candidati per ogni possibile istanza dell'affected attribute.



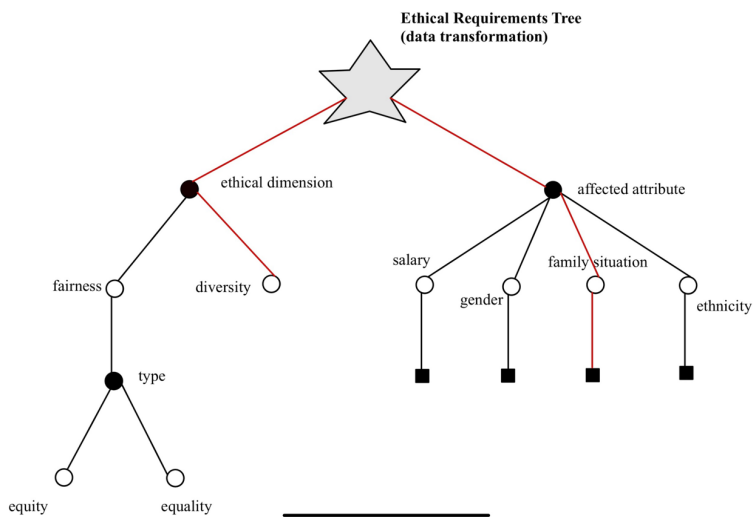
Distribuzione per genere



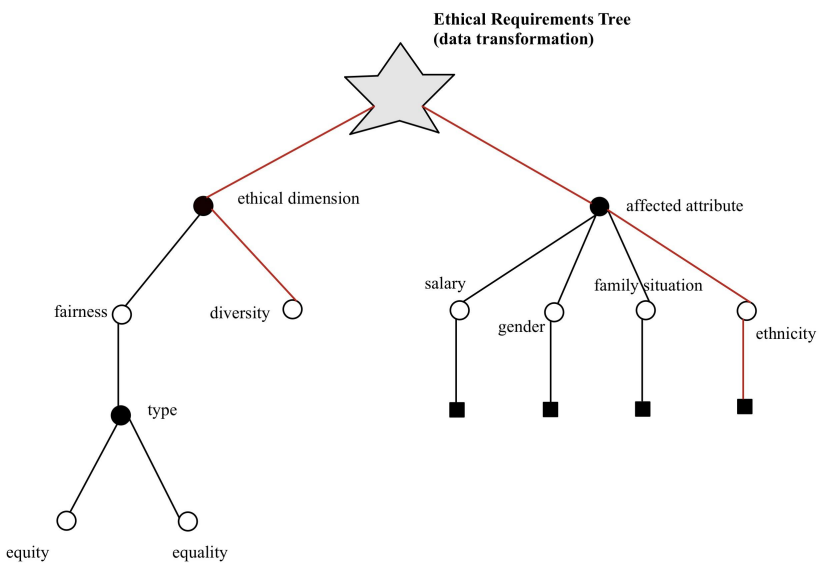
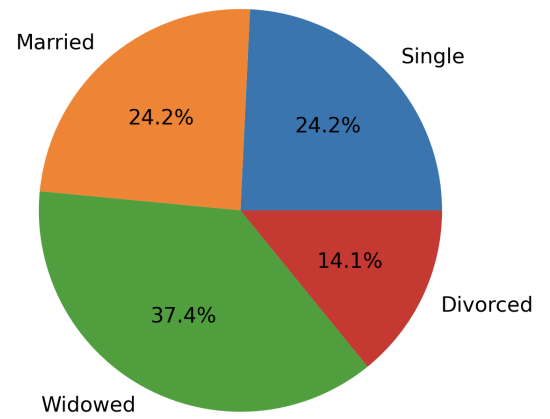
Distribuzione per indice performance.
La media è 6.06



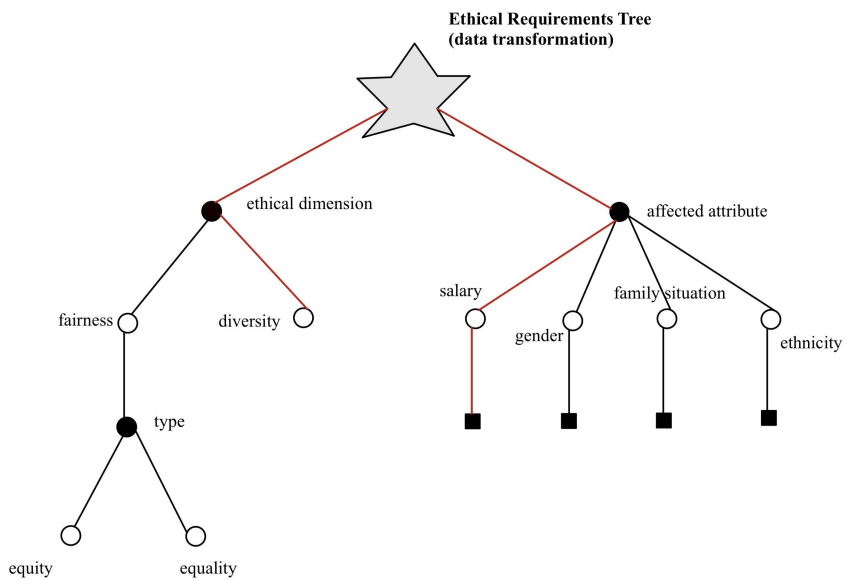
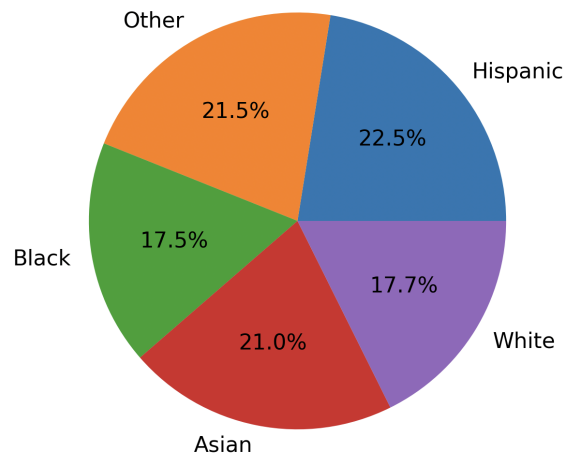
|



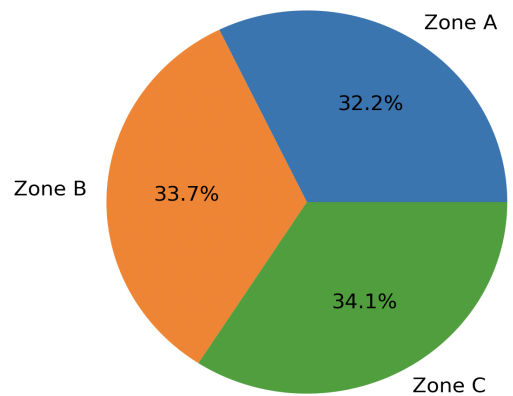
Distribuzione per Situazione Familiare



Distribuzione per Etnia



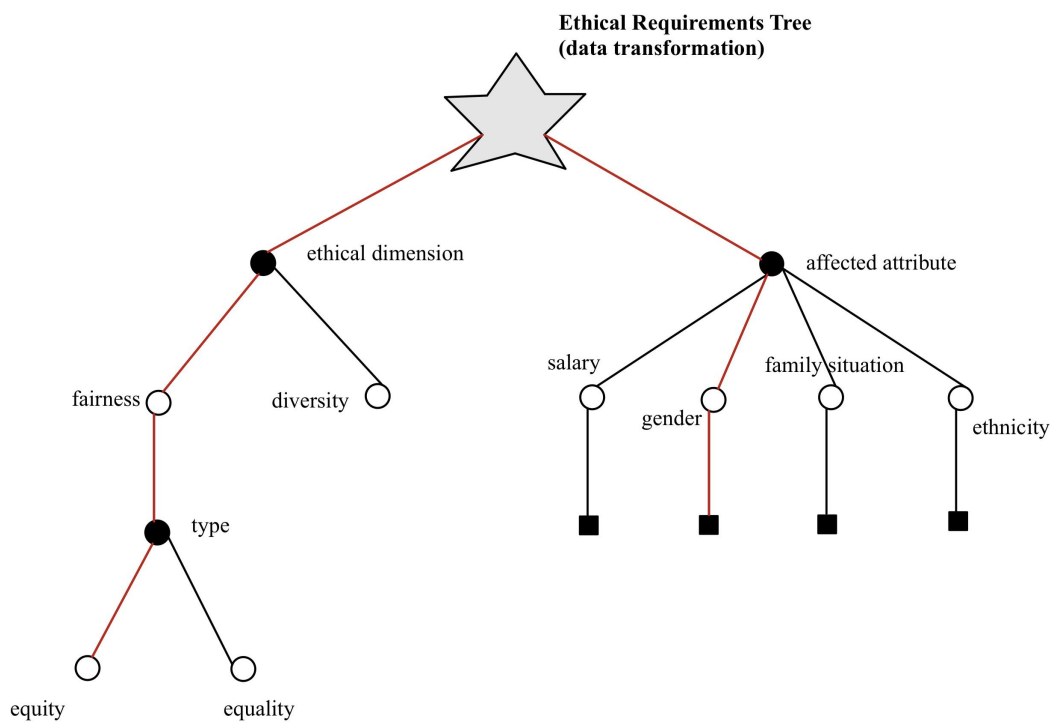
Distribuzione per Paga



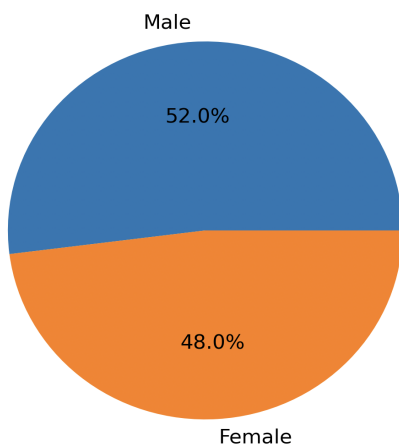
Equity

L'equity si propone di bilanciare equality e diversity, garantendo che ogni gruppo abbia le stesse opportunità di essere selezionato, pur mantenendo un approccio "fair": vengono aumentate le istanze dei clerk che fanno parte del gruppo meno presente, in modo da aumentare le loro possibilità di essere selezionati senza per questo abbassare il performance rate medio. In particolare nello script Python abbiamo implementato il seguente algoritmo per selezionare secondo equity:

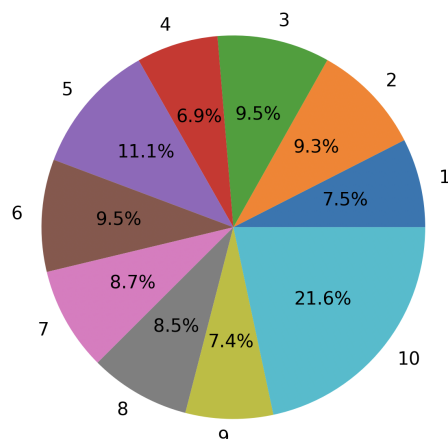
1. Valutare il disequilibrio D come differenza percentuale tra i generi: $D = (\%M - \%F) \cdot D = (W\%M - W\%F) \cdot D = (\%M - \%F)$ o $D = (\%Mm - \%Fm) \cdot D = (W\%Mm - W\%Fm) \cdot D = (\%Mm - \%Fm)$. Dove M rappresenta il numero totale di maschi, F il numero totale delle femmine, Mm i maschi manager, Fm le femmine manager.
2. Inserire D duplicati delle istanze del gruppo sotto-rappresentato nel dataset combinato, in modo da simulare la stessa percentuale di generi.



Distribuzione per genere



Distribuzione per indice performance.
La media è 6.06



Conclusioni

Il progetto di selezione etica dei manager si è rivelato un'esperienza estremamente arricchente sotto molteplici aspetti. La fase iniziale di studio ci ha permesso di approfondire un tema attuale e delicato, di cui non avevamo precedente conoscenza. In particolare, è stato stimolante scoprire come un algoritmo, seppur teoricamente corretto, possa generare risultati distorti a causa dei bias insiti nei dati storici. La fase pratica del progetto ci ha condotti a una riflessione profonda sul significato concreto dell'etica. Abbiamo constatato che l'etica presenta indubbiamente molteplici sfaccettature e che il contesto assume un ruolo fondamentale. Nel nostro caso, l'esempio applicativo è la selezione dei manager, ma a seconda delle circostanze, l'uso di diverse dimensioni etiche può rivelarsi più o meno appropriato. Inoltre, è emersa la crucialità di valutare il livello di bias o squilibrio del dataset di base rispetto a determinati gruppi etnici. Nel nostro caso, il dataset risultava già piuttosto equilibrato, pertanto l'applicazione di una dimensione etica rispetto a un'altra non ha prodotto differenze significative, sebbene si siano comunque osservate variazioni nel performance rate. Questa esperienza ha sottolineato l'importanza di adattare le strategie etiche al contesto specifico e alle caratteristiche del dataset utilizzato.

Per quanto riguarda i grafici, essi sono riferiti solo alla distribuzione dei clerk selezionati (100), ma tengono conto della distribuzione dei manager mostrata inizialmente. Per questo nella diversity ad esempio si osserva che i maschi selezionati sono 84.8 % : bisogna tenere conto dell'iniziale disparità a discapito degli uomini presente tra i manager.

E' importante sottolineare che la tecnologia da sola non può risolvere le questioni etiche. Occorre un impegno collettivo da parte di ricercatori, sviluppatori, professionisti, cittadini e istituzioni per definire e attuare principi etici chiari e condivisi che guidino lo sviluppo e l'utilizzo dell'IA. Solo attraverso un dialogo aperto e costruttivo tra diverse discipline e competenze potremo costruire un futuro digitale in cui l'intelligenza artificiale sia al servizio dell'umanità e del bene comune.

L'intelligenza artificiale rappresenta un potente strumento che può apportare benefici significativi alla società. Tuttavia, il suo sviluppo e utilizzo devono essere guidati da principi etici solidi per evitare conseguenze negative. La costruzione di un futuro digitale etico e responsabile richiede un impegno collettivo da parte di tutti gli attori coinvolti. La consapevolezza delle sfide etiche poste dall'IA è il primo passo per affrontarle con lungimiranza e costruire un mondo migliore per tutti.

Ringraziamo il Professore Fabio Alberto Schreiber e la Professoressa Letizia Tanca per il supporto e per i consigli.

Bibliografia:

- [1] E. Quintarelli, F. A. Schreiber, K. Stefanidis, L. Tanca, B. Oliboni: *Introducing Context-Aware, Ethical Data Transformation*, submitted for publication.
- [2] C. Bolchini, C. Curino, G. Orsi, E. Quintarelli, R. Rossato, F. A. Schreiber, and L. Tanca. And what can context do for data? *Comm. ACM*, 52(11):136–140, 2009.
- [3] H. Jagadish, J. Stoyanovich, and B. Howe. The many facets of data equity. *theWorkshop Proceedings of the EDBT/ICDT 2021 Joint Conference*.
- [4] J. Stoyanovich, S. Abiteboul, and G. Miklau. Data, responsibly: Fairness, neutrality and transparency in data analysis. In *EDBT*, 2016.