

Università degli Studi di Salerno Dipartimento di Informatica

Corso di Laurea Triennale

Progettazione e sviluppo di un algoritmo per individuare i Dark Pattern

Relatori Candidato

Prof. Fabio Palomba Paolo Petta

Matricola: 0512105913

Anno Accademico 2021/2022

Ringraziamenti

Vorrei dedicare qualche riga per ringraziare tutte le persone che mi hanno supportato in questo percorso molto importante per me.

Un grazie speciale al Professore Palomba che mi ha fornito un grosso aiuto per lo sviluppo di questa tesi, con la sua professionalità e organizzazione impeccabile ha reso tutto più semplice.

Grazie immensamente ai miei genitori che hanno sempre creduto in me, affrontando al mio fianco i bassi del percorso e festeggiando ancor più vicini le soddisfazioni dei traguardi raggiunti. Grazie a Papà che senza dire nulla mi fa capire tanto (a volte). In particolare grazie a mia madre, nonostante qualche incomprensione è stata una roccia al mio fianco capace di smuovermi e di farmi affrontare nel migliore dei modi ogni insidia.

Grazie a mio fratello Lorenzo che seppur distante, con un semplice messaggio mi ha fatto sentire la sua vicinanza, tenendosi sempre aggiornato sul percorso.

Grazie a Matteo e Teresa che con molta scaramanzia sono stati al mio fianco, seguendo ma non chiedendo... perchè si sa "la data dell'esame non si chiede".

Grazie alla piccola Giorgia che mi ha rovinato tante giornate di studio soltanto per darmi la giusta ricarica e dandomi poi la possibilità di studiare al meglio.

Grazie a Zia Teresa, Zio Aldo e ai miei cugin* che ci sono sempre per me e ci sono sempre stati.

Grazie a tutta la mia squadra di volley dove ormai non trovo più compagni ma fratelli, indirettamente hanno permesso che io oggi sia qui dandomi valvola di sfogo con gli allenamenti e facendomi vivere emozioni incredibili.

Grazie ai miei compagni di viaggio Mariachiara, Antonio, Roberta, Salvatore, Giuseppe, Alessandro, Andrea, che mi hanno fatto affrontare le giornate universitarie in un modo diverso e spensierato dandomi sempre la giusta carica per seguirli nello studio. Grazie anche a Mario, per tanti Pocio con cui ho condiviso oltre al volley anche quest'ultimo anno universitario, spalla a spalla ce l'abbiamo fatta.

Grazie ai Bikazz, il gruppo di uscite in moto che mi hanno spesso fatto perdere

nottate di riposo per fare qualche giro, le rinunce allo studio a causa vostra sono state più che produttive, mi fate vivere dei bellissimi momenti. In particolare grazie ad Enrico Fabbricatore, non ci conosciamo da anni ma sappi che sei un grande amico, ogni volta che ci vediamo mi carichi a molla.

Grazie mille ai miei colleghi e titolari che mi hanno reso il luogo di lavoro un vero e proprio ambiente familiare e facendomi crescere tanto. Grazie a Francesco che ha dato inizio al tutto e mi permette di crescere in ambito lavorativo. Grazie in particolare Franco, che mi ha dato una grossa opportunità, dimostrandomi da sempre fiducia e comprendendo al massimo le mie esigenze.

Grazie agli amici di sempre, qualcuno in classe con me alle superiori, qualcuno di vecchia data, mi basta un'uscita con voi che tutto cambia.

Grazie anche a chi non c'è più, Nonna che da lassù sorvegli... spero di averti reso fiero.

Grazie alla donna che è sempre al mio fianco, Denise. Ha vissuto tutto, dalle volte in cui credevo di non farcela e volevo abbandonare fino ad oggi, il giorno in cui completo questo percorso e corono il mio sogno. Sei stata per me forza e motivazione, hai gestito momenti delicati come nessuno poteva fare accompagnandomi passo dopo passo verso giorni migliori. Mi hai donato indissolubili momenti di svago, senza mai farmi perdere di vista l'obbiettivo, per quanto io cerchi di apparire sempre sicuro di me sono sicuro che senza di te tutto questo non sarebbe stato possibile. Non potrò mai ringraziarti abbastanza.

Infine un grazie speciale a me, ho dimostrato a me stesso che non contano gli ostacoli... se voglio posso.

Indice

1	Intr	oduzio	one	7
	1.1	Contes	sto Applicativo	7
	1.2	Motiva	azioni ed obiettivi	9
	1.3	Risulta	ati ottenuti	10
	1.4	Strutt	ura della Tesi	11
2	Ana	disi de	llo stato dell'arte	12
	2.1	Tassor	nomia di Brignull	12
		2.1.1	Trick Questions	12
		2.1.2	Sneak into basket	14
		2.1.3	Roach motel	15
		2.1.4	Privacy Zuckering	16
		2.1.5	Price comparison prevention	17
		2.1.6	Misdirection	18
		2.1.7	Hidden costs	20
		2.1.8	Bait and switch	22
		2.1.9	Confirmshaming	23
		2.1.10	Disguised ads	24
		2.1.11	Forced continuity	25
		2.1.12	Friend spam	26
	2.2	Tassor	nomia dell'EDPB	27
		2.2.1	Overloading (Sovraccarico)	27
		2.2.2	Skipping (Saltare)	27
		2.2.3	Stirring (Stimolare)	27
		2.2.4	Hindering(Ostacolare)	27
		2.2.5	Fickle (Mutare)	28
		2.2.6	Left in the dark (Lasciare all'oscuro)	28
	2.3	Tasson	nomia di Colin M. Gray	29
		2.3.1	Nagging	29

		2.3.2	Obstruction	30
		2.3.3	Interface interference	31
		2.3.4	Forced Action	32
3	Def	inizion	ne dei meccanismi di identificazione	33
	3.1	Ricero	ea ed Analisi del dataset	33
	3.2	Svilup	opo degli algoritmi di machine learning	34
		3.2.1	Algoritmo che determina l'appartenenza alla famiglia dei Dark	
			Pattern	34
		3.2.2	Algoritmo di classificazione della categoria del dark pattern	35
	3.3	Svilup	ppo dell'app python	36
	3.4	Svilup	ppo dell'estensione Chrome	37
		3.4.1	Installazione dell'estensione Chrome	37
4	Val	utazio	ne preliminare	39
5	Cor	ıclusio	ni	42
	5.1	Sitogr	afia	43

Lista delle Figure

1	Esempio di Trick Question	7
2	Interfaccia dell'estensione sviluppata	10
3	Esempio di un risultato dato dall'estensione sviluppata	10
4	Esempio di Trick Questions	12
5	Test di Stroop	13
6	Altro esempio di Trick Question	13
7	Esempio di Sneak into Basket	14
8	Esempio di Roach Motel	15
9	CEO di Cambridge Analytica in conferenza	۱7
10	Esempio di Price Comparision Prevention	18
11	Esempio di Misdirection	19
12	Altra tipologia di esempio di Misdirection	20
13	Esempio di Hidden Costs	21
14	Esempio di Bait and Switch	22
15	Esempio di Confirmshaming	23
16	Esempio di Disguised Ads	24
17	Esempio di Forced Continuity	25
18	Caso Linkedin di Friend Spam	26
19	Esempio di Nagging	30
20	Esempio di Obstruction	31
21	Esempio di Interface Interference	31
22	Esempio di Forced Action	32
23	Panoramica del lavoro svolto per la creazione del dataset	33
24	Grafico prima e dopo dei record null	35
25	Confronto delle metriche	36
26	Interfaccia dell'estensione sviluppata	37
27	Estensione all'interno della lista	38
28	Esempio di stopwords	39

29	Esempio di stemming [19]	40
30	Metriche finali	40
31	Matrice di confusione	41

1 Introduzione

1.1 Contesto Applicativo

L'obiettivo delle Interfacce utente (UI) e dell'esperienza utente (UX) é quello di rendere l'utilizzo di applicativi software quanto più semplice possibile e coerenti. Spesso però possono racchiudere delle tecniche di manipolazione per guidare l'utente ad effettuare operazioni che non ha intenzione di svolgere, queste tecniche sono state denominate dark pattern. I dark pattern se ben progettati riescono nel migliore dei casi soltanto ad infastidire l'utente finale, ma nel peggiore ad apportare danni economici, condividere dati personali in maniera involontaria o indurre gli utenti ad avere comportamenti compulsivi. Sono stati categorizzati in base al tipo di meccanismo psicologico su cui prendono di mira l'uomo. Un esempio molto intuitivo di dark pattern è il seguente:

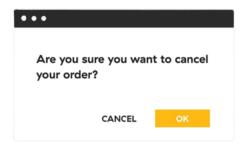


Figure 1: Esempio di Trick Question

Nell'immagine precedente é presente un Dark Pattern di tipo *Trick Questions*, i pulsanti di risposta sono due: uno evidenziato di giallo e ben visibile mentre l'altro no. Automaticamente il comportamento dell'uomo induce a cliccare sul pulsante evidenziato per rispondere in maniera affermativa ma per raggiungere l'obiettivo finale l'utente dovrebbe cliccare sull'altro pulsante.

Questa tipologia di manipolazione si può dire che è nata con la diffusione di internet in tutte le case ma è stata riconosciuta ed approfondita intorno al 2010 da Harry Brignull (UX designer) che ha dato un nome a questa pratica e ne ha definito una tassonomia. Sul suo sito deceptive design prova a spiegare nel dettaglio i trucchi utilizzati dai malintenzionati per rendere le persone sempre più coscienti dei pericoli in cui possono incorrere ed esplicita le ben dodici categorie identificate. [13] Al nostro fianco troviamo però il regolamento generale sulla protezione dei dati (GDPR) che richiede vengano rispettati alcuni fondamentali principi per il trattamento dei dati personali rendendo alcuni Dark pattern "non compilant", in particolare quelli riguardanti la privacy imponendo per esempio il principio di trasparenza che enuncia:

"il principio della trasparenza impone che le informazioni destinate al pubblico o all'interessato siano concise, facilmente accessibili e di facile comprensione e che sia usato un linguaggio semplice e chiaro, oltre che, se del caso, una visualizzazione."[1]

Nonostante questo affrontiamo i dark pattern quasi quotidianamente essendo utilizzati anche da colossi internazionali ed è impossibile non nominare Amazon, una piattaforma online che offre principalmente servizi di e-commerce ma non solo e occupa il primato in classifica riguardo alla vendita online, in quanto nel 2020 diversi utenti hanno segnalato la complessità nell'annullare l'iscrizione in quanto erano richiesti diversi passaggi molto macchinosi, a supporto degli utenti anche il Norwegian Consumer Council (NCC) ha portato uno studio che ha confermato le segnalazioni degli utenti inoltre sono arrivati alla scoperta di molti altri abbonamenti digitali con pratiche di disiscrizione estremamente complesse, in particolare ben il 26% degli utenti hanno affermato di avere problemi ad annullare l'iscrizione. Per fortuna, la diffusione

1.2 Motivazioni ed obiettivi

Nel corso degli anni con lo sviluppo esponenziale di internet è diventato sempre più difficile mantenere al sicuro la propria privacy e navigare in sicurezza sul Web, complici di questo andamento sono sicuramente i Dark Pattern che rendono l'utilizzo di questo immenso strumento una pratica spesso intrinseca e ricca di insidie. Strettamente collegati ai dark pattern troviamo i bias cognitivi, questi ultimi sono usati spesso per spingere il malcapitato utente verso decisioni o scelte che sembreranno logiche ed obiettive ma in realtà saranno scelte fortemente condizionate. Un esempio nella pratica può essere l'effetto scarsità e urgenza, evidente quando il consumatore viene spinto ad acquistare un determinato prodotto in tempi brevi poiché sta per esaurirsi oppure sta terminando l'offerta. Ciò induce spesso l'utente ad acquistare lo stesso il prodotto anche se non ne ha bisogno, soltanto per sfruttare la convenienza dell'offerta. Questo fenomeno in particolare non può essere considerato illegale ed è oggigiorno utilizzato tantissimo per smaltire vecchie giacenze rimaste in deposito tuttavia si può considerare una vera e propria manipolazione. [9]

"Il "segreto" dei dark pattern è proprio questo: convincere gli utenti di essere padroni delle loro azioni, quando invece ci sono dei fili invisibili che li muovono come se fossero dei burattini." [9]

Tutto questo ci permette di comprendere quanto sia facile sfruttare le vulnerabilità del cervello umano per scopi poco leciti e trasparenti. Questi fenomeni in parte si possono combattere rendendo noti alla comunità i dark pattern riscontrati sui portali dedicati e al fine di rendere la navigazione quanto più sicura possibile si è pensato a sviluppare un algoritmo di machine learning in grado di rilevare questi pattern per poi integrarlo in un'estensione del browser Google Chrome che una volta attivata effettua una scansione della pagina web e, nel caso fossero presenti pattern manipolatori, li mette in risalto all'utente finale.

1.3 Risultati ottenuti

L'obiettivo era quello di rendere la navigazione dei siti web più sicura e cosciente da parte dell'utente medio, per questo è stato sviluppato un algoritmo di machine learning in grado di riconoscere i dark pattern testuali presenti nella pagina effettuando una scansione ed analizzando il testo contenuto, il tutto é stato integrato in un'estensione del noto browser Google Chrome con una piccola interfaccia grafica, la quale permette di avviare la procedura di scansione e di visualizzare il numero di dark pattern riscontrati. Un'altra funzionalità è quella di evidenziare la presenza del



Figure 2: Interfaccia dell'estensione sviluppata

pattern all'interno della pagina vera e propria mettendo in risalto il testo aggiungendo uno sfondo giallo come di seguito.



Figure 3: Esempio di un risultato dato dall'estensione sviluppata

1.4 Struttura della Tesi

La tesi è strutturata nel seguente modo:

- Introduzione: Viene introdotto l'argomento ad alto livello menzionando la UI e la UX per poi approfondire il concetto di dark pattern definendo i rischi che creano per l'utente e mostrando un esempio arrivando poi ad un accenno del trattamento legislativo e del lavoro svolto da Brignull. È trattato poi l'argomento dei bias cognitivi e di come vengono sfruttati in unione ai dark pattern per poi definire infine gli obiettivi della tesi.
- Analisi dello stato dell'arte: Sono state analizzate le diverse categorie di dark pattern seguendo la tassonomia proposta dallo UX designer Brignull, spiegando nel dettaglio ogni tipologia per poi analizzare anche le tassonomie proposte dall'EDPB e da Colin M. Gray.
- Definizione dei meccanismi di identificazione: Sono spiegati i dettagli implementativi dello sviluppo dell'estensione Chrome, analizzando tutte le procedure svolte per completare il progetto.
- Valutazione preliminare: Analizzati i diversi test effettuati e i vari risultati ottenuti durante lo sviluppo con le relative soluzioni adottate.
- Conclusioni: Viene effettuata un'analisi dei risultati ottenuti nel complessivo e sono descritti i possibili sviluppi futuri al fine di migliorare ulteriormente il progetto.

2 Analisi dello stato dell'arte

In precedenza abbiamo introdotto il concetto di dark pattern rimanendo ad un livello alto di astrazione, scendendo più nel dettaglio è doveroso illustrare le diverse tipologie definite dalla tassonomia più importante, quella di Brignull poi accenneremo anche quelle più recenti proposte dall'EDPB (European Data Protection Board) e da Colin M. Gray.

2.1 Tassonomia di Brignull

Harry Brignull è un consulente indipendente per la UX che fornisce servizi di consulenza a grandi colossi internazionali come Spotify o Vodafone, prima di intraprendere la strada della consulenza ha conseguito un dottorato di ricerca in Scienze Cognitive. Circa nel 2010 ha approfondito il lavoro sulle esperienze utente non etiche coniando poi il termine "dark pattern" da cui è nato il suo sito web darkpattern.org [2] Con i suoi studi è riuscito a definire una tassonomia molto dettagliata:

2.1.1 Trick Questions

Il dark pattern trick questions appunto racchiude delle domande trabocchetto che riescono spesso ad ingannare l'utente facendogli fornire una risposta che non voleva davvero fornire. I componenti che possono indurre in inganno possono essere diversi: Per esempio possiamo trovare una grafica fuorviante come la seguente



Figure 4: Esempio di Trick Questions

Nel caso in cui l'utente volesse rifiutare la richiesta di donare del denaro viene raggirato da alcune convenzioni ormai radicate nel nostro cervello, infatti per rifiutare automaticamente cliccherà sul pulsante rosso questo perchè il dark pattern va a creare un'interferenza cognitiva, come quella sfruttata dal test psicologico di Stroop dove vengono proposte delle parole scritte con colori diversi e l'utente deve pronunciare a voce alta il colore dell'inchiostro cui è scritta la parola, di conseguenza solo il colore è l'informazione rilevante.[10]

GIALLO	MARRONE	ARANCIONE
VERDE	GIALLO	GRIGIO
ROSSO	ROSA	NERO

Figure 5: Test di Stroop

Nel nostro caso l'utente se non attento può rischiare di compiere un'azione involontaria che gli comporta anche un danno economico.

Un altro componente può forzare sulla sensibilità dell'utente come di seguito



Figure 6: Altro esempio di Trick Question

In questo caso il tutto si sofferma sulla psicologia umana che per non perdere l'offerta rifiuterà di cancellare l'abbonamento.

2.1.2 Sneak into basket

Il dark pattern Sneak into basket, letteralmente "intrufolarsi nel cestino" riguarda principalmente i siti web che vendono prodotti o servizi in quanto durante il processo di acquisto viene aggiunto un costo, attribuito a qualche servizio o prodotto aggiuntivo, automaticamente. Nel dettaglio Harry Brignull lo ha definito equivalente ad un lavoratore di un supermercato che mette le cose nel carrello a tua insaputa e questi oggetti attirano la tua attenzione soltanto una volta arrivato alla cassa o addirittura non l'attirano proprio.

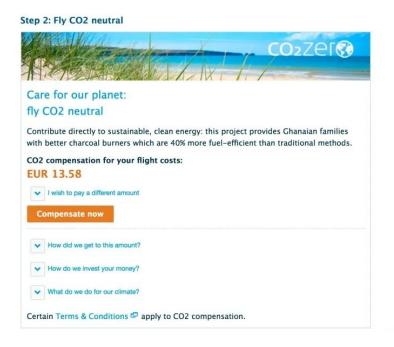


Figure 7: Esempio di Sneak into Basket

Talvolta viene sfruttato anche per far iscrivere un utente a qualche abbonamento o newsletter ne deduciamo che questi stratagemmi possono avere conseguenze fastidiose per utente poco attento che si ritroverà ad acquistare qualcosa inconsciamente oppure a ritrovarsi delle iscrizioni senza aver prestato coscientemente il consenso. Nella figura precedente possiamo notare come una nota compagnia aerea aggiunge automaticamente una donazione per compensare il consumo di carbonio, sebbene lo scopo sia molto nobile manca una richiesta esplicita forzando quindi l'utente a pagare ugualmente. [14]

2.1.3 Roach motel

Roach Motel si riferisce ad un design che rende molto facile entrare in una determinata situazione, ma allo stesso tempo ne rende molto difficoltosa l'uscita, andando più nel dettaglio spesso sono utilizzati nell'ambito degli abbonamenti, di fatti la procedura di iscrizione solitamente si conclude in massimo due passaggi, mentre l'annullamento della stessa richiede procedure faticose e complesse. [5] Un esempio, anche se un po' estremo, è il seguente:



Figure 8: Esempio di Roach Motel

nel quale per iniziare a ricevere la newsletter basta cliccare sul pulsante invia,

invece per essere rimosso dalla lista dei contatti bisogna inviare un messaggio in codice Morse in una finestra temporale ristretta e in un fuso orario molto inusuale.

2.1.4 Privacy Zuckering

Questo modello di design induce l'utente a condividere più informazioni personali di quanto vuole, prende il nome dal CEO di Facebook Mark Zuckerberg in quanto all'inizio, Facebook rendeva molto complicata la gestione dei dati personali e al contrario molto semplice la condivisione eccessiva per errore. [3] In particolare l'utilizzo di questo dark pattern è emerso nel 2018 con lo scandalo Cambridge Analytica, un'azienda specializzata nel raccoglimento dei dati personali tramite i social, analisi dei dati e manipolazione degli stessi la quale utilizzando algoritmi di machine learning è riuscita a creare dei profili psicologici in base ai dati estrapolati effettuando quasi un lavoro di psicometria. Già questo sembra molto preoccupante ma l'azienda non si è fermata qui, il passo successivo è stato creare un sistema di microtargeting comportamentale che in parole povere sarebbe una pubblicità altamente mirata in base al profilo psicologico dell'utente. Tutta questa tecnologia sarà stata utilizzata per svariati scopi poco chiari ma uno in particolare è stato quello delle elezioni presidenziali di Trump del 2016, di fatti il comitato dello stesso affidò la campagna elettorale a Cambridge Analytica la quale ha struttato tutte le sue tecnologie per indurre ogni singolo utente ad effettuare l'azione più comoda alla campagna elettorale anche tramite bot e account falsi. [17]



Figure 9: CEO di Cambridge Analytica in conferenza

Oggi Facebook sembra aver colmato le lacune presenti ed ha introdotto un'area di impostazioni della privacy più chiara ed intuitiva dalla quale si può avere il controllo delle proprie informazioni personali. Questo fenomeno però non è utilizzato soltanto da Facebook ma delle volte all'interno dei termini e condizioni, spesso vengono accettati dopo una lettura superficiale, si trovano delle clausole che permettono di vendere i dati personali a chiunque.

2.1.5 Price comparison prevention

Consiste nel rendere difficoltoso il confronto dei prezzi con prodotti della stessa tipologia oppure con gli stessi prodotti ma venduti da altri competitor. Questo design può essere sfruttato per esempio andando a manipolare le confezioni, di fatti confrontare una confezione contenente una mela con una confezione di una mela ed una banana risulta più complicato che confrontare due confezioni contenenti entrambe una mela.

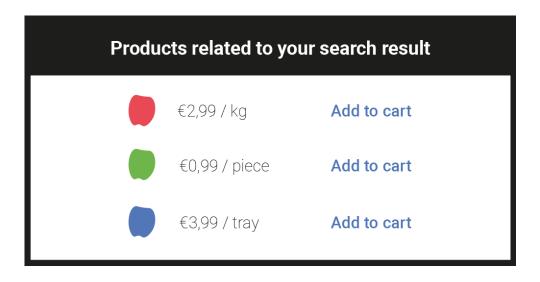


Figure 10: Esempio di Price Comparision Prevention

Nell'esempio precedente possiamo notare come vengano utilizzate tre unità di misura diverse tra loro rendendo la comparazione molto complicata se non impossibile.

2.1.6 Misdirection

Il pattern Misdirection, tradotto letteralmente come direzione sbagliata ha la potenzialità di giostrare e manipolare l'attenzione del cliente a suo piacimento, proprio come viene fatto dai maghi, mettendo in risalto le opzioni che si vogliono far scegliere all'utente finale. Possono esserci diversi esempi, il primo riguarda una richiesta di autorizzazione ai cookie



Figure 11: Esempio di Misdirection [7]

Si può notare che viene evidenziato in giallo l'opzione di accettare tutti i cookie mentre l'opzione contraria, di salvare la selezione viene quasi mimetizzata e resa poco visibile, questa grafica logicamente porterà il cliente a cliccare sul bottone in risalto, accettando tutti i cookie. Altro esempio possono essere le classiche pubblicità presenti nei giochi su mobile gratuiti che si presentano come nella figura seguente



Figure 12: Altra tipologia di esempio di Misdirection

È evidente che risulta molto difficile uscire da questa schermata senza scaricare il gioco proposto perché qualsiasi tocco sullo schermo porta al download del gioco mentre analizzando con un di attenzione si può notare la x in alto a sinistra, quasi mimetizzata, che ci permette di chiudere la pubblicità.

2.1.7 Hidden costs

Questo è uno dei tanti dark pattern utilizzati dalle aziende per aumentare i guadagni a discapito dell'utente. Consiste nell'aggiungere di nascosto alcuni costi all'acquisto che si sta effettuando con la speranza che il malcapitato non se ne accorga prima di effettuare il pagamento. Logicamente questo fenomeno può funzionare soltanto nella fase iniziale in quanto è molto controproduttivo riguardo la fidelizzazione del cliente che, una volta cosciente dell'accaduto, di certo ne rimarrà scontento

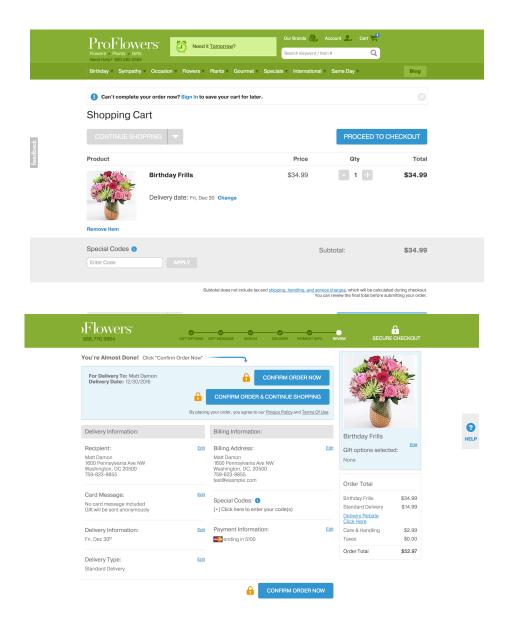


Figure 13: Esempio di Hidden Costs

Nella prima figura precedente si può notare come il costo sia di 34,99\$ non menzionando minimamente il costo di cura e gestione se non alla fine del processo (che ritroviamo nella seconda immagine), quando l'utente è pronto al pagamento. [4]

2.1.8 Bait and switch

Tradotto letteralmente significa esca ed interruttore, di fatti si attrae l'utente ad effettuare un'azione a lui familiare per poi far coincidere una conseguenza completamente diversa da quella che si aspettava



Figure 14: Esempio di Bait and Switch [8]

Nella figura precedente è rappresentato un esempio riguardante un avviso di un aggiornamento, l'utente ha tre scelte: cliccare sul pulsante di installazione, cliccare su "installa dopo" oppure cliccare sulla "X" in alto a destra che per convenzione ha il significato di chiudere la finestra, è proprio qui che sorge il problema in quanto la "X" non corrisponde alla chiusura, come da convenzione, ma farà partire automaticamente

l'aggiornamento. Questo esempio è ispirato al caso accaduto sul sistema operativo Windows, dove veniva forzato l'aggiornamento proprio in questo modo, raggirando i clienti finali.

2.1.9 Confirmshaming

Questo pattern si avvicina ancora di più alla vera e propria psicologia dell'uomo, sfruttando i suoi punti deboli e i suoi sensi di colpa. Nella pratica propone solitamente due opzioni, la prima evidenziata e a cui il pattern vuole far puntare e la seconda, meno evidenziata ma con una frase di psicologia inversa utilizzata con l'intento di manipolare i sentimenti dell'utente e incoraggiarlo a scegliere la prima opzione.

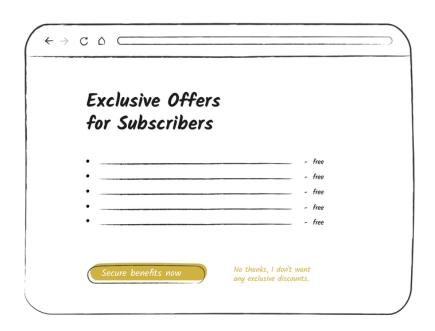


Figure 15: Esempio di Confirmshaming

In questo esempio viene proposto un'iscrizione e sono presenti le due opzioni, la prima che permette all'utente di approfittare dei benefici subito e la seconda invece che gli permette di rifiutare la proposta affermando che non vuole ricevere nessuno sconto, questa frase porta l'utente a ripensarci e lo incoraggia ad accettare la proposta iniziale.

2.1.10 Disguised ads

Rappresenta degli annunci pubblicitari mascherati, in quanto ti inducono a cliccare su di essi essendo molto simili a componenti comuni di una pagina con la differenza che al click scatenano diverse pubblicità molto fastidiose. È molto facile incontrare questo dark pattern all'interno di siti illegali che permettono il download pirata di contenuti a pagamento, come programmi, libri, film e streaming video. Un esempio che ne permette di capire molto facilmente la struttura è il seguente:



Figure 16: Esempio di Disguised Ads

Come possiamo notare in questa pagina, che permette di scaricare un libro in formato EPUB, è presente un link che dovrebbe rimandare al download "Easybytez", cliccando però su tale link si aprirà una nuova finestra con delle pubblicità.

2.1.11 Forced continuity

Spesso gli abbonamenti a servizi nascondono diverse insidie, una di queste può essere il dark pattern forced continuity, quest'ultimo si traduce in italiano come continuità forzata di fatti permette all'abbonamento di continuare forzatamente e uno dei modi è non avvertendo il cliente del rinnovo. Altro caso simile si riscontra al termine delle prove gratuite, quando automaticamente viene prelevato l'importo della quota senza chiedere un'ulteriore autorizzazione.



Figure 17: Esempio di Forced Continuity

Un esempio del caso precedentemente citato riguarda l'app di audiolibri Audible. Al momento della registrazione sono richiesti i dati della carta di credito e viene offerta una prova gratuita di 30 giorni, al termine senza nessun tipo di notifica viene prelevata la quota di \$14,95 e l'utente spesso se ne rende conto troppo tardi. [15]

2.1.12 Friend spam

Il dark pattern in fase preliminare prevede di avere l'accesso alla casella email o account social, in maniera tale da utilizzare poi questi strumenti per inviare messaggi ai contatti registrati o agli amici utilizzando proprio il malcapitato come mittente. L'esempio più famoso dell'utilizzo di tale pattern è stato da parte di Linkedin prima del 2015 dove nel processo di registrazione chiedevano l'accesso alla casella email per "rafforzare la tua rete"

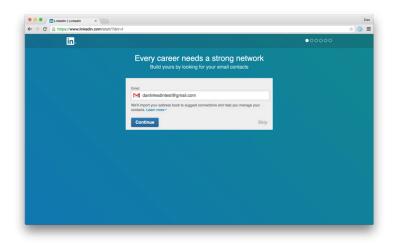


Figure 18: Caso Linkedin di Friend Spam

Una volta configurato l'account email ne è stato fatto un uso fraudolento inviando segretamente email ai contatti. Questo ha portato ad un'azione per cause collettive ed è risultata una pratica illegale ai sensi della legge della California.

2.2 Tassonomia dell'EDPB

Secondo l'ente europeo è evidente che alcuni dark pattern non rispettano i requisiti della GDPR come la trasparenza e la correttezza di fatti ha stipulato una sua sud-divisione in macro-categorie modellate sulla base degli effetti provocati agli utenti nonostante abbia precisato che non siano del tutto esaustive. [18]

2.2.1 Overloading (Sovraccarico)

Viene presentato agli utenti una grossa mole di richieste di informazioni, opzioni o possibilità per indurlo ad acconsentire a tutto e conseguentemente a condividere più dati involontariamente. Da questo gruppo nascono i sottotipi continuous prompting, privacy labyrinth e too many options.[18]

2.2.2 Skipping (Saltare)

Rappresenta una UI e UX predisposta a far prestare poca attenzione da parte dell'utente ad alcuni aspetti fondamentali della protezione dei dati. Questa categoria racchiude la tipologia deceptive confidence e look over there.[18]

2.2.3 Stirring (Stimolare)

Ha il compito di influenzare le scelte degli utenti sfruttando le loro emozioni e manipolandoli usando particolari interfacce studiate ad hoc. Raggruppa i sottotipi emotional driving e hidden in plain sight.[18] Si avvicina alla tipologia Confirmshaming proposta da Brignull.

2.2.4 Hindering(Ostacolare)

Appunto ostacola o blocca gli utenti verso la gestione delle informazioni personali rendendo la stessa molto difficile oppure talvolta impossibile. Da cui derivano le sottocategorie impasse, longer than necessary, misleading information.[18]

2.2.5 Fickle (Mutare)

Crea una struttura dell'interfaccia poco chiara con lo scopo di rendere difficoltoso navigare tra gli strumenti di tutela della privacy forniti. Da questa tipologia derivano anche lack of hierarchy, decontextualization.[18]

2.2.6 Left in the dark (Lasciare all'oscuro)

Questa tipologia nasconde informazioni all'utente, può addirittura nascondere gli strumenti per la gestione dei dati personali oppure lascia il tutto nell'incertezza. Le sotto categorie associate sono: linguistic discontinuity, contradictory information, ambiguous wording.[18]

2.3 Tassonomia di Colin M. Gray

Colin M. Gray è un professore associato alla Purdue University nel dipartimento di Tecnologia della Computer Grafica e Professore Associato di Learning Design & Technology nel Dipartimento di Curriculum e Istruzione. [20] Qualche anno più tardi anche lui ha approfondito l'argomento dei dark pattern affermando:

"Usiamo il termine dark pattern per definire i casi in cui i progettisti utilizzano la loro conoscenza del comportamento umano (ad esempio, la psicologia) e i desideri degli utenti finali per implementare funzionalità ingannevoli che non sono nel migliore interesse dell'utente" [6]

2.3.1 Nagging

Secondo Gray la categoria Nagging che tradotto vuole significare fastidioso include i comportamenti relativi a reindirizzamenti delle funzionalità sulla base di una o più interazioni. In particolare si riferisce a intrusioni ripetute durante il normale utilizzo dell'interfaccia, andando ad interrompere l'attività svolta dall'utente e rimandandolo su attività non sempre correlate alla principale, un esempio possono essere i popup che vanno ad oscurare l'interfaccia di utilizzo oppure avvisi audio che possono distrarre ed infastidire l'utente. [6]

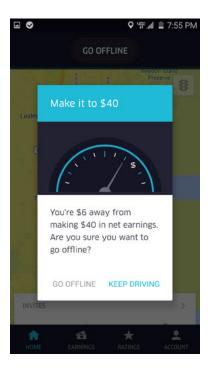


Figure 19: Esempio di Nagging

Un riscontro è stato trovato sull'app di Uber dove il driver che intende terminare il lavoro per la giornata viene infastidito da questo popup che lo invita a continuare per raggiungere un obbiettivo.[23]

2.3.2 Obstruction

Per Obstruction Gray intende un ostacolo posto all'interno di un flusso di lavoro rendendolo più difficile del dovuto e distogliendo l'attenzione dell'utente, può essere collegato alla tassonomia di Brignull ed in particolare ai tipi Roach Motel e Price Comparision Prevention.[6]



Figure 20: Esempio di Obstruction

Nell'immagine precedente si nota il caso di Apple relativo ad iOS 6 dove veniva nascosta l'impostazione per disattivare il tracciamento ai fini pubblicitari. [22]

2.3.3 Interface interference

Questa tipologia racchiude i casi in cui viene strutturata l'interfaccia utente al fine di manipolare l'utente finale per esempio andando a creare inganni visivi. [6]



Figure 21: Esempio di Interface Interference

Nella figura precedente notiamo come è stato nascosto intenzionalmente il link per annullare l'iscrizione, utilizzando il colore grigio chiaro su sfondo bianco. [24]

2.3.4 Forced Action

La categoria Forced Action identifica i casi in cui agli utenti è richiesto di eseguire un'azione specifica per continuare. Può essere un blocco per ricevere poi dei vantaggi oppure accedere a funzionalità specifiche [6]

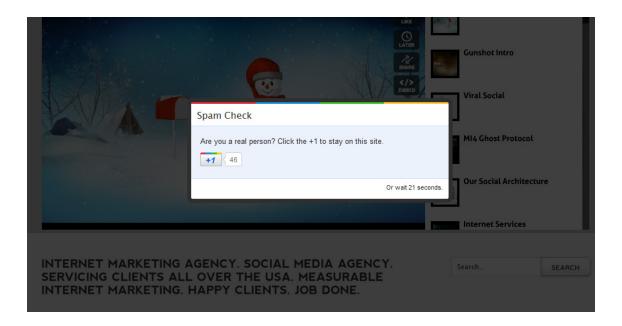


Figure 22: Esempio di Forced Action

Nel precedente esempio vediamo come viene richiesto all'utente di cliccare sul pulsante proposto oppure aspettare 21 secondi per verificare che sia davvero un umano.

3 Definizione dei meccanismi di identificazione

Si è pensato di sviluppare un'estensione del browser Chrome in grado di identificare i dark pattern presenti all'interno della pagina analizzata. Le fasi di lavoro sono state divise in:

3.1 Ricerca ed Analisi del dataset

Dopo svariate ricerche è stato identificato il dataset da utilizzare per lo sviluppo e l'addestramento dell'algoritmo. Il set di dati utilizzato è frutto di uno studio effettuato dall'università di Pricerton, USA per lo sviluppo di un web Crawler in grado di scansionare ed identificare i dark pattern presenti su diversi siti web di shopping. Dal loro studio sono emerse ben 1818 istanze di dark pattern che sono poi state suddivise in 15 categorie e a loro volta raggruppate in 7 rami, su circa 11 mila siti web scansionati almeno un migliaio hanno utilizzato dark pattern quindi l'11%, percentuale molto alta considerando che affermano di poter migliorare ulteriormente l'algoritmo. [16]

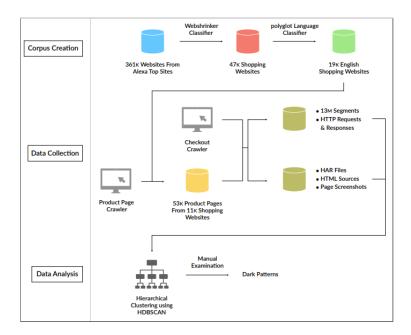


Figure 23: Panoramica del lavoro svolto per la creazione del dataset

Il dataset che conta circa 1800 record è strutturato con le seguenti colonne:

- Pattern String: Rappresenta il testo identificato come dark pattern
- Comment: Un commento sull'istanza del dark pattern
- Pattern Category: La categoria di cui fa parte il dark pattern identificato
- Pattern Type: Il tipo specifico a cui appartiene
- Where in website?: Posizione in cui è stato riscontrato il dark pattern
- Deceptive?: Determina se è ingannevole o meno
- Website Page: Link alla pagina dove è stato trovato il dark pattern

Di queste le feature più rilevanti sono sicuramente la Pattern String e la Pattern Type che sono state selezionate per lo sviluppo dell'algoritmo.

3.2 Sviluppo degli algoritmi di machine learning

In ottica dello sviluppo dell'app Python per identificare i dark pattern è stato necessario sviluppare due tipi di algoritmi, uno per analizzare se le porzioni di testo sono un dark pattern o meno e un altro invece per identificare a quale categoria appartengono.

3.2.1 Algoritmo che determina l'appartenenza alla famiglia dei Dark Pattern

Quest'ultimo ha il compito di agevolare il lavoro all'algoritmo di identificazione della categoria, andando ad effettuare una scrematura preliminare per poi permettere di analizzare soltanto i record identificati come veri e propri dark pattern.

È stato utilizzato anche un ulteriore dataset contenente delle stringhe estratte da siti web e per ognuna la relativa classificazione di apparteneza alla famiglia dei dark pattern. Inoltre è stato effettuato un lavoro di gestione per i record null, fino ad arrivare ad avere due dataframe: uno contenente le stringhe identificate come pattern relative al primo dataset e uno contenente invece quelle identificate come non dark pattern. Successivamente si è andati ad unire il tutto avendo un unico grande dataset contenente tanti esempi classificati come "dark" oppure come "non dark" nel dettaglio 1512 sono le istanze "dark", mentre 1894 quelle "non dark", arrivando poi a preparare il dataset per l'addestramento, suddividendolo nei gruppi di train e test. Sono poi stati trasformati i dati utilizzando uno schema di ponderazione dei termini, chiamato Tfidf per poi passare all'addestramento del classificatore binario di Bernoulli. Una volta completato l'algoritmo è stato esportato con il modulo joblib per essere utilizzato all'interno dell'applicazione vera e propria.

3.2.2 Algoritmo di classificazione della categoria del dark pattern

Per poter determinare a che categoria appartiene un dark pattern è stato sviluppato l'algoritmo in questione, seguendo i passi di seguito:

• Eliminazione dei record con dati mancanti utilizzando la funzione della libreria Pandas dropna, andando ad eliminare circa 300 record null presenti.

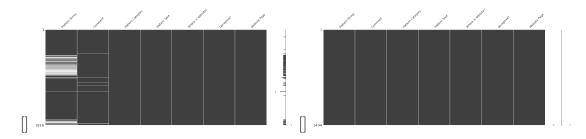


Figure 24: Grafico prima e dopo dei record null

- Effettuato Text Pre Processing, cioè l'attività di preparare il testo per renderlo quanto più comprensibile all'algoritmo di machine learning, andando ad utilizzare le stopword con la libreria nltk. Quest'ultime sono delle parole comuni che non aggiungono informazioni rilevanti alle frasi, per esempio "ancora", "altro".
- Implementato lo Stemming, una tecnica utilizzata per risalire alle parole nella loro forma basilare. Per esempio, la radice delle parole mangiare e mangiato

è mangiare. Questa tecnica è spesso utilizzata anche dai motori di ricerca per l'indicizzazione delle parole. [21]

• Sviluppato l'algoritmo vero e proprio di classificazione Multinomial Naive Bayes che permette di classificare in categorie. La scelta di questo algoritmo non è stata casuale, in quanto sono stati testati diversi algoritmi:

```
Multinomial Naive Bayes model accuracy: 0.9163879598662207
Multinomial Naive Bayes model f1: 0.9163879598662207
Multinomial Naive Bayes model precision: 0.9163879598662207
Multinomial Naive Bayes model recall: 0.9163879598662207
KNeighborsClassifier() accuracy: 0.9163879598662207
KNeighborsClassifier() f1 micro: 0.9163879598662207
KNeighborsClassifier() precision: 0.9163879598662207
KNeighborsClassifier() recall: 0.9163879598662207
```

Figure 25: Confronto delle metriche

Come si evince dall'immagine in alto il MultinomiaNB è risultato il più performante.

3.3 Sviluppo dell'app python

Lo sviluppo dell'applicazione Python è stato portato avanti utilizzando Flask, un micro-framework utilizzato per lo sviluppo di applicazioni web, che ha permesso di integrare i modelli di classificazione all'interno dell'estensione. Prima di tutto sono stati caricati i moduli joblib per avere a disposizione i modelli addestrati in precedenza, poi è stato predisposto per ricevere request in POST ed analizzare i dati contenuti nella richiesta. Per ogni token all'interno dei dati viene prima determinato se è classificato come dark pattern utilizzando l'algoritmo citato in precedenza, poi in caso di risposta positiva si passa a determinare la categoria con il secondo algoritmo.

3.4 Sviluppo dell'estensione Chrome

Un'estensione chrome è una funzionalità aggiuntiva del browser installata separatamente, con lo scopo di permettere all'utente di personalizzare al meglio l'ambiente di navigazione oppure di usufruire di funzionalità aggiuntive. Si è partiti dall'interfaccia grafica dell'estensione, strutturata come nel seguente modo:



Figure 26: Interfaccia dell'estensione sviluppata

Troviamo una prima sezione dove è mostrato il numero dei dark pattern rilevati, una seconda invece in cui è presente il pulsante per analizzare la pagina e nell'ultima parte un link per approfondire l'argomento.

È stato sviluppato poi uno script in grado di scansionare la pagina, estrapolandone soltanto il testo che viene passato ad un secondo script che ha il compito di riunire tutti i segmenti estratti ed inviarli, tramite metodo POST all'app predisposta in precedenza. Una volta ricevuto l'esito dall'app passa ad evidenziare tramite una funzione le sezioni in cui si è riscontrato un dark pattern.

3.4.1 Installazione dell'estensione Chrome

Per installare l'estensione è necessario:

- Recarsi all'interno del browser
- Navigare al link chrome://extensions/, che permette di gestire le estensioni

- Attivare in alto a destra la modalità sviluppatore
- Cliccare su "Carica estensione non pacchettizzata"
- Una volta seguiti questi passaggi bisogna caricare la cartella del progetto

Completati i passaggi elencati troveremo l'estensione all'interno della lista di gestione.

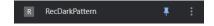


Figure 27: Estensione all'interno della lista

Ora cliccandoci verrà aperta l'interfaccia grafica come nella figura 25, che permette di analizzare la pagina e visualizzare il numero dei dark pattern rilevati.

4 Valutazione preliminare

Inizialmente lo sviluppo dell'algoritmo di categorizzazione dei dark pattern non era molto preciso e mostrava risultati di accuracy di circa il 60% sul set di dati proposto per il test, estratto selezionando circa il 25% del dataset. Per aumentare la percentuale di accuracy si è pensato di effettuare un text pre-processing considerevole, andando ad ignorare le stopwords (parole di arresto) che andavano ad influenzare l'addestramento del modello sfruttando la libreria NLTK (Natural language toolkit) in Python [11]

Sample text with Stop	Without Stop Words	
Words		
GeeksforGeeks – A Computer	GeeksforGeeks , Computer Science,	
Science Portal for Geeks	Portal ,Geeks	
Can listening be exhausting?	Listening, Exhausting	
I like reading, so I read	Like, Reading, read	

Figure 28: Esempio di stopwords

Successivamente è stata implementata anche la procedura di stemming, che come accennato in precedenze ha il compito di considerare soltanto la versione basilare delle parole

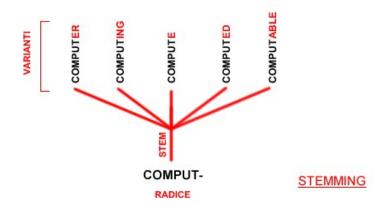


Figure 29: Esempio di stemming [19]

Nell'immagine precedente abbiamo un esempio: la parola Comput- può avere diverse varianti, come Computer, Computing, Compute, Computed, Computable. Infine queste implementazioni hanno aumentato vertiginosamente le metriche

```
Multinomial Naive Bayes model accuracy: 0.9163879598662207
Multinomial Naive Bayes model f1: 0.9163879598662207
Multinomial Naive Bayes model precision: 0.9163879598662207
Multinomial Naive Bayes model recall: 0.9163879598662207
```

Figure 30: Metriche finali

La matrice di confusione risultante ci mostra l'efficienza dell'algoritmo nelle diverse categorie

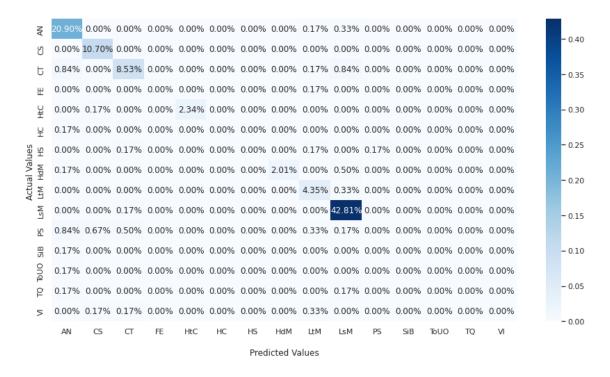


Figure 31: Matrice di confusione

Notiamo discreti risultati nelle categorie Activity Notification e Low Stock Message, degne di nota anche le categorie Confirmshaming e Count timer. La predominanza assoluta della tipologia Low Stock Message è sicuramente dovuta al numero di record presenti di quella tipologia, parliamo di circa 600 record su 1800 totali. Per gli altri risultati discreti abbiamo la tipologia Activity Notification che è presente ben 206 volte, così come Count Timer con 382 istanze. Sorprende la tipologia Confirmshaming che ha mostrato un sufficiente risultato con solo 31 esempi nel dataset. Tutte le altre tipologie hanno in totale soltanto 600 record, questo potrebbe essere motivo della scarsa efficienza generale.

5 Conclusioni

Lo sviluppo dell'estensione Chrome nella sua completezza ha dato risultati discreti in quanto quest'ultima è stata installata in locale e provata, riscontrando anche qualche risultato positivo. In particolare l'estrazione dei dati implementata funziona correttamente e invia tutto all'app Flask che sfrutta i modelli addestrati per determinare se appartiene alla famiglia dei dark pattern, in caso di esito positivo identifica a quale categoria appartengono.

L'app Flask infine invia all'estensione i risultati ottenuti in maniera tale da evidenziare i dark pattern riscontrati ed incrementare il contatore visualizzato.

Nonostante l'arduo sviluppo spesso i risultati non sono coerenti con quelli desiderati infatti si ipotizza in futuro la generazione di un dataset più vasto e variegato in aggiunta ad una modellazione più dettagliata dei dati.

5.1 Sitografia

- [1] Andrea Afferni. Dark pattern: cosa sono e il loro rapporto con il GDPR. https: //www.cybersecurity360.it/legal/privacy-dati-personali/darkpattern-cosa-sono-e-il-loro-rapporto-con-il-gdpr/#Dark_pattern_ e_GDPR. 2020.
- [2] Harry Brignull. *About me.* https://90percentofeverything.com/about/index.html.
- [3] Harry Brignull. *Privacy Zuckering*. https://www.deceptive.design/types/privacy-zuckering.
- [4] Harry Brignull. *Privacy Zuckering*. https://www.deceptive.design/types/hidden-costs.
- [5] Harry Brignull. Reach Motel. https://www.deceptive.design/types/roachmotel.
- [6] Bryan Battles Colin M. Gray Yubo Kou. The Dark (Patterns) Side of UX Design. https://dl.acm.org/doi/pdf/10.1145/3173574.3174108. 2018.
- [7] Dapde. Types and examples of dark pattern deception. https://dapde.de/en/dark-patterns-en/types-and-examples-en/irref%C3%BChrung2-en/.
- [8] Dapde. Types and examples of dark pattern deception. https://dapde.de/en/dark-patterns-en/types-and-examples-en/irref%C3%BChrung2-en/.
- [9] Deceptive design. What is deceptive design? https://www.deceptive.design/. 2010.
- [10] Francesca Fiore. L'Effetto Stroop in psicologia sperimentale. https://www.stateofmind.it/2018/06/effetto-stroop-psicologia/. 2018.
- [11] GeeksforGeeks. Removing stop words with NLTK in Python. https://www.geeksforgeeks.org/removing-stop-words-nltk-python/. 2022.

- [12] Susan Grant. Amazon's "Dark Patterns" How They Keep You From Cancelling Prime. https://consumerfed.org/amazons-dark-patterns-how-they-keep-you-from-cancelling-prime/. 2021.
- [13] Ictsviluppo. Cosa sono i dark pattern. https://www.ictsviluppo.it/ecommerce/cosa-sono-i-dark-pattern. 2021.
- [14] Interaction-design.org. How Designers Sneak Products into Users' Shopping Baskets. https://www.interaction-design.org/literature/topics/sneaking-into-basket#:~:text=Sometimes%2C%20dark%20patterns%20can%20be,adjust%20the%20amount%20to%20pay...2018.
- [15] David Martinson. *UX Dark Design Pattern* Forced Continuity. https://davidmartinsonnyc.medium.com/ux-dark-design-pattern-forced-continuity-6c7af78682ad. 2020.
- [16] Arunesh Mathur et al. "Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites". In: *Proc. ACM Hum.-Comput. Interact.* 1.CSCW (2019).
- [17] Emanuele Menietti. *Il caso Cambridge Analytica, spiegato bene.* https://www.ilpost.it/2018/03/19/facebook-cambridge-analytica/. 2018.
- [18] Andrea Michinelli. Dark pattern e dati personali: ecco le linee guida EDPB per il legal design dei social (e non solo). https://www.cybersecurity360.it/legal/privacy-dati-personali/dark-pattern-e-dati-personali-ecco-le-linee-guida-edpb-per-il-legal-design-dei-social-e-non-solo/. 2022.
- [19] Andrea Minini. Algoritmo di stemming. https://www.andreaminini.com/ir/stemming/.
- [20] Polytechnic Purdue. Colin M. Gray, PhD. https://polytechnic.purdue.edu/profile/gray42.

- [21] Stemming Lemmatization. https://www.tutorialspoint.com/natural_language_toolkit/natural_language_toolkit_stemming_lemmatization. htm#:~:text=Stemming%20is%20a%20technique%20used,stemming%20for%20indexing%20the%20words.. 2018.
- [22] darkpatterns uxp2. iOS 6: Hidden Ad Tracking. https://darkpatterns.uxp2.com/pattern/ios-6-hidden-ad-tracking/. 2018.
- [23] darkpatterns uxp2. *Uber: Driver Manipulation*. https://darkpatterns.uxp2.com/pattern/uber-driver-manipulation/. 2018.
- [24] darkpatterns uxp2. Zynga.com: Unsubscribe hidden as white on white. https://darkpatterns.uxp2.com/pattern/zynga-com-unsubscribe-hidden-as-white-on-white/. 2018.