

UNIVERSITÀ DEGLI STUDI DI SALERNO

DIPARTIMENTO DI INFORMATICA



CORSO DI RETI GEOGRAFICHE: STRUTTURA, ANALISI E PRESTAZIONI

Twitter Sentiment Analysis: il caso Greta Beccaglia

Docente:

Prof.essa Delfina Malandrino

Studente:

Di Genio Gaetano 0522501339

Anno Accademico 2021/22

Indice

1	Introduzione ed Obiettivi	1
1.1	Violenza contro le donne, definizioni	1
1.2	Il caso Greta Beccaglia	2
1.3	Sentiment Analysis e Twitter	3
1.4	Costruzione del Dataset	3
1.5	Pre-elaborazione del dataset	5
1.6	Sentiment Analysis	7
1.7	Risultati e conclusioni	8

Capitolo 1

Introduzione ed Obiettivi

1.1 Violenza contro le donne, definizioni

La violenza contro le donne rappresenta un importante problema di sanità pubblica, oltre che una violazione dei diritti umani. Perfino in un'epoca che si professa civilizzata come la nostra ogni giorno figurano casi di violenza fisica o sessuale, stalking, minacce e peggio ancora, casi di femminicidio.

Ma cosa si intende davvero per "violenza contro le donne"? Tecnicamente si intende ogni atto di violenza fondata sul genere che provochi un danno o una sofferenza fisica, sessuale o psicologica per le donne, incluse le minacce, la coercizione o la privazione arbitraria della libertà. Così recita l'art 1 della dichiarazione Onu sull'eliminazione della violenza contro le donne.

Nel dettaglio:

- **VIOLENZA PSICOLOGICA**: si manifesta in forma indiretta, ad esempio mediante comportamenti, come non ascoltare, fraintendere volutamente, minacciare lesioni o vendetta, disprezzare la partner, trattarla come una domestica, intimidirla, colpevolizzarla, offenderla, controllarla e/o isolarla.

- **STALKING**: comportamento persecutorio messo spesso in atto quando la donna cerca di allontanarsi da una relazione violenta. Il maltrattatore perseguita l'ex-partner seguendola negli spostamenti, aspettandola sotto casa, al lavoro, telefonandole continuamente a casa, in ufficio, sul telefonino. Gli effetti possono essere devastanti: viene minato il senso dell'autonomia e dell'indipendenza della donna facendola sentire "in trappola"; molte donne riportano anche disturbi del sonno, difficoltà a concentrarsi fino ad arrivare, nei casi più estremi, a depressioni.

- **VIOLENZA ECONOMICA**: la violenza economica è caratterizzata dal legame o dalla dipendenza economica dalla persona che la esercita; per esempio vietando alla donna di svolgere un lavoro o un percorso formativo, sfruttando la donna come forza lavoro, ricoprendola di debiti, limitando o privando la donna del denaro per le spese domestiche, se non lavora, non rendendola partecipe al reddito familiare, o non corrispondendo gli alimenti dopo la separazione.

- **VIOLENZA FISICA**: la violenza fisica si esprime in un'aggressione diretta contro una persona, ad esempio mediante spintoni, tirate di capelli, schiaffi, pugni, ferite con un coltello, fino all'uccisione in casi estremi.

- **VIOLENZA SESSUALE**: la violenza sessuale definisce ogni atto sessuale attivo o passivo, imposto alla vittima mediante violenza fisica, minacce o abuso di autorità.

1.2 Il caso Greta Beccaglia

Empoli-Fiorentina del 27/11/2021 è stato un anticipo del sabato per la 14esima giornata di Serie A, turno nel quale la massima serie ha aderito all'iniziativa di sensibilizzazione contro la violenza sulle donne esibendo un segno rosso sul viso come in Figura 1.1 .



Figura 1.1

Le buone intenzioni manifestate sul campo hanno trovato un grave contraltare in un episodio avvenuto dopo il fischio finale del match quando l'inviata di Toscana TV, Greta Beccaglia, impegnata nel raccogliere le reazioni a caldo dei tifosi che defluivano dallo stadio, ha ricevuto da un uomo che passava di fianco una pacca sul sedere, mentre altri le urlavano frasi inopportune per cui nessuno dei presenti in studio o allo stadio è intervenuto.

L'episodio ha scatenato l'indignazione della giornalista che ha subito denunciato l'accaduto e, individuato il responsabile dell'atto, è stato accusato di aver commesso una **violenza sessuale** per cui rischierebbe fino a 12 anni di reclusione (la legge punisce la violenza sessuale con la reclusione da 6 a 12 anni).

L'episodio ha causato non poche polemiche e si discute costantemente su quanto sia giusto o meno condannare un atto del genere come una vera e propria violenza sessuale (o vederla come una molestia) e se la pena è giusta o assolutamente esagerata ed inopportuna.

1.3 Sentiment Analysis e Twitter

Cosa ne pensano quindi le persone sui social di tutto ciò? Si prenderà come riferimento il social network **Twitter**, che permette di ricercare particolari temi attraverso hashtag (ad. esempio 'violenzasulledonno') e visualizzare i tweet postati dagli utenti riguardo lo specifico argomento così da poter valutare il loro pensiero generale.

Il progetto si pone quindi l'obiettivo di analizzare attraverso il Machine Learning e la Sentiment Analysis, l'hate speech nei tweet che parlano di argomenti quali sessismo e violenza sulle donne concentrandosi sul caso della giornalista sportiva Greta Beccaglia.

Nello specifico si definiscono

- **Sentiment Analysis**: l'uso dell'elaborazione del linguaggio naturale, dell'analisi del testo, della linguistica computazionale e della biometria per identificare, estrarre, quantificare e studiare sistematicamente gli stati affettivi e le informazioni soggettive. Viene ampiamente applicata a campi quali recensioni e risposte a sondaggi, media online, social e materiali sanitari per applicazioni che vanno dal marketing al servizio clienti fino alla medicina clinica
- **Hate speech**: (incitamento all'odio), evento purtroppo comune su Internet. Spesso i social media come Facebook e Twitter affrontano il problema di identificare e censurare i post problematici mentre soppesano il diritto alla libertà di parola. L'importanza di rilevare e moderare l'incitamento all'odio è evidente dalla forte connessione tra esso e i crimini d'odio effettivi

1.4 Costruzione del Dataset

Il dataset viene costruito mediante le API offerte da Twitter e strutturato nel modo seguente:

- id**: ID univoco del tweet.
- retweet**: numero di condivisioni del tweet.
- likes**: numero di 'mi piace' del tweet.
- text**: testo del tweet.
- target**: polarità del tweet (positivo, neutrale o negativo).

Gli hashtag presi in considerazione sono i seguenti (Figura 1.2)

```
#hashtags
hashtags = ["#violenzasulledonno", "#GretaBeccaglia", "#nontelaprendere", "#violenzadigenere", "#molestiesulledonno",
            "#violenzamaschile", "#stopviolenzasulledonno", "#nonunadimeno", "violenza sulle donne"]
```

Figura 1.2

Si creano due funzioni, una per raccogliere i tweet relativi all'hashtag dato in input e creare una lista, l'altra per unire diverse liste (Figura 1.3)

Si invocano le funzioni passando i parametri e si crea un dataset .csv (Figura 1.4), l'output è mostrato dalla (Figura 1.5)

```
# funzione che prende in input un hashtag e ritorna una lista di tweet correlati
def create_tweets_list(htag):
    # raccolta dei tweet in base all'hashtag
    iterator = tweepy.Cursor(api.search_tweets, q= htag, tweet_mode="extended", lang="it").items(1000)

    lista = list(iterator)
    # seleziona solo i campi che ci interessano
    output = []
    for tweet in lista:
        text = tweet._json["full_text"]
        retweet_count = tweet.retweet_count
        likes = tweet.favorite_count
        line = {'text':text, 'retweet':retweet_count, 'likes': likes}
        output.append(line)
    return output;

# Unisce più liste
def Union(lst0, lst1, lst2, lst3, lst4, lst5, lst6, lst7, lst8):
    final_list = lst0 + lst1 + lst2 + lst3 + lst4 + lst5 + lst6 + lst7 + lst8
    return final_list
```

Figura 1.3

```
# Per ogni hashtag crea la relativa lista
# Unisce tutte le liste in un'unica lista comune

list0 = create_tweets_list(hashtags[0])
list1 = create_tweets_list(hashtags[1])
list2 = create_tweets_list(hashtags[2])
list3 = create_tweets_list(hashtags[3])
list4 = create_tweets_list(hashtags[4])
list5 = create_tweets_list(hashtags[5])
list6 = create_tweets_list(hashtags[6])
list7 = create_tweets_list(hashtags[7])
list8 = create_tweets_list(hashtags[8])

final_list = Union(list0, list1, list2, list3, list4, list5, list6, list7, list8)

# creazione dataset .csv
df = pd.DataFrame(final_list)
df.to_csv('tweet_dataset.csv', encoding='utf-8')

# aggiunta colonna per polarità
# 2 per positivo
# 1 per neutrale
# 0 per negativo
df["target"] = 0
df.to_csv("tweet_dataset.csv", index=False)
df.head()
```

Figura 1.4

	text	retweet	likes	target
0	** A #Majori in tanti per dire 'No' alla viole...	0	0	0
1	#14marzo giornata mondiale contro la #violenza...	0	0	0
2	#13marzo giornata mondiale contro la #violenza...	0	0	0
3	RT @ComunedìBergamo: #ComunedìBergamo #viole...	3	0	0
4	"L'urlo delle donne" è un libro che parla di v...	0	0	0

Figura 1.5

Prima di addestrare il modello sono stati eseguiti alcuni passaggi per rimuovere articoli, emoji, link, in generale pulire il dataset. (Figura 1.6) e (Figura 1.7) Il documento di testo è stato poi convertito in minuscolo per una migliore generalizzazione. Infine, abbiamo eseguito lo Stemming (riducendo le parole ai loro steli derivati) e la Lemmatizzazione (riducendo le parole derivate alla loro forma radice conosciuta come lemma) per ottenere risultati migliori (Figura 1.8), output in (Figura 1.9)

Figura 1.6

Figura 1.7

```
# stemming
st = nltk.PorterStemmer()
def stemming_on_text(data):
    text = [st.stem(word) for word in data]
    return data
df['text'] = df['text'].apply(lambda x: stemming_on_text(x))
df.head()
```

	text	retweet	likes	target
0	maiori tanti dire no violenza sulle donne s t ...	0	0	0
1	marzo giornata mondiale contro violenzasulledo...	0	0	0
2	marzo giornata mondiale contro violenzasulledo...	0	0	0
3	comunedibergamo violenzasulledonno nuova panch...	3	0	0
4	I urlo delle donne libro parla violenza sulle ...	0	0	0

Figura 1.8

```
#lemmatizzazione
lm = nltk.WordNetLemmatizer()
def lemmatizer_on_text(data):
    text = [lm.lemmatize(word) for word in data]
    return data
df['text'] = df['text'].apply(lambda x: lemmatizer_on_text(x))
df.head()
```

	text	retweet	likes	target
0	maiori tanti dire no violenza sulle donne s t ...	0	0	0
1	marzo giornata mondiale contro violenzasulledo...	0	0	0
2	marzo giornata mondiale contro violenzasulledo...	0	0	0
3	comunedibergamo violenzasulledonno nuova panch...	3	0	0
4	I urlo delle donne libro parla violenza sulle ...	0	0	0

Figura 1.9

1.6 Sentiment Analysis

Sono state individuate le radici delle parole con accezioni positive o negative (Figura 1.10) ed effettuata la sentiment analysis (Figura 1.11)

```
radiciPositive = ("ama", "amo", "affett", "allegri", "amabil", "apprezz", "armon", "affet",
                 "applaus", "abbracc", "ador",
                 "bell", "ben", "beat", "brav", "buon", "benef", "brill",
                 "cuor", "coeren", "celebr",
                 "dolc", "divert",
                 "evviva", "emoz", "elog",
                 "felic", "fest", "facil",
                 "gentil", "god", "grazi", "generos", "gioi",
                 "innamor", "interes", "insieme",
                 "libert",
                 "maestos", "miglior",
                 "pace", "passion", "perfe", "piac", "pura", "purezz", "prezios", "promuov",
                 "rilass", "riabbracc",
                 "solida", "spero", "speran", "success", "sì", "sacr", "stupend", "spettacol",
                 "viv", "vin", "valor", "vale", "vera", "vittor")
radiciNegative = ("accus", "amaro", "amarez", "arm", "ammazz",
                 "brut", "boicott", "boh", "bho",
                 "condann", "cazz", "crisi", "critic", "coglion",
                 "decent", "depress", "detest", "disgr", "delir", "damn", "drog",
                 "fumo", "fuma",
                 "esorcis",
                 "fascis",
                 "guai",
                 "immat", "insult", "inuman", "impone",
                 "lent",
                 "mor", "merd", "male", "maial",
                 "no", "nega", "ncazz", "negr",
                 "od", "oscur",
                 "perde", "preoccup", "pusillanim", "porc",
                 "rovina",
                 "schif", "satan", "sprof", "soffri", "soffer", "scandal", "scars", "spor", "spar", "stalk",
                 "trist", "trash", "tarocc",
                 "vergogn",
                 "zitt")
radiciDaEscludere = ("now", "nom", "not", "noz", "amp", "nor", "veramente", "imponent")
```

Figura 1.10

```
def define_sentiment(string):
    val = 0
    list_ = string.split()
    for word in list_:
        if word in emoticonsPositive:
            return 2 # Tweet considerato positivo
        elif word in emoticonsNegative:
            return 0 # Tweet considerato negativo
        else:
            if not word.startswith(radiciDaEscludere):
                if (word.startswith(radiciPositive)):
                    val = val + 1
                if (word.startswith(radiciNegative)):
                    if word not in ("non", "nonostante"):
                        val = val - 1
                # Il 'non' cambia dinamicamente la polarita' del tweet
                if word == "non":
                    val = val * -1
                # Tutte le parole prima di 'ma' e 'però' non vengono considerate
                if word in ("ma", "però"):
                    val = 0
    if (val > 0):
        label = 2
    elif (val < 0):
        label = 0
    else:
        label = 1
    return label

df['target'] = df['text'].apply(lambda tweet: pd.Series(define_sentiment(tweet)))
```

Figura 1.11

1.7 Risultati e conclusioni

La data visualization (Figura 1.12) mostra come l'opinione generale riguardo l'episodio è perlopiù neutrale.

Data Visualization

```
# distribuzione del dataset
ax = df.groupby('target').count().plot(kind='bar', title='Distribuzione dei dati', legend=False)
ax.set_xticklabels(['Negative', 'Neutral', 'Positive'], rotation=0)
# conserva i dati in liste.
text, sentiment = list(df['text']), list(df['target'])

sns.countplot(x='target', data=df)

<AxesSubplot:title={'center': 'Distribuzione dei dati'}, xlabel='target', ylabel='count'>
```

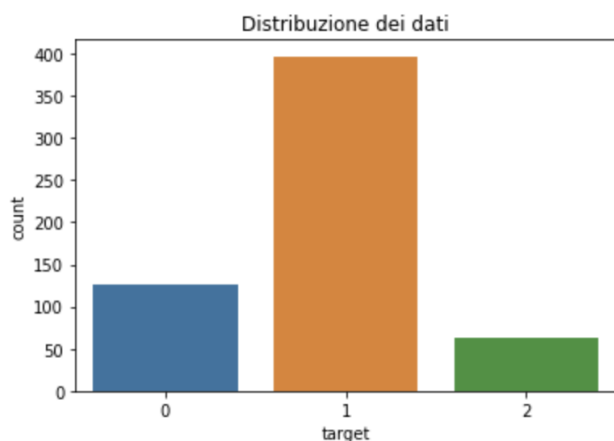


Figura 1.12

La Corte di Cassazione Penale (sentenza 27042/2010) ha stabilito che si ha violenza sessuale (art. 609 bis c.p.) ogni volta che viene compiuto un qualsiasi atto che consiste in un contatto corporeo (anche se fugace ed estemporaneo) tra soggetto attivo e soggetto passivo del reato, o comunque in un coinvolgimento della sfera fisica di quest'ultimo, e pone quindi in pericolo la libera autodeterminazione della persona offesa nella sua sfera sessuale.

Ad esempio il tocco non casuale dei glutei, ancorché sopra i vestiti, configura violenza sessuale e non la semplice molestia sessuale (art. 660 c.p.), che è invece integrata solo in presenza di espressioni volgari a sfondo sessuale ovvero di atti di corteggiamento invasivo ed insistito diversi dall'abuso sessuale.

Ciononostante in generale le persone sui social credono che non ci siano gli estremi per poter essere completamente a favore delle pene imposte (asserendo che si debba distinguere più meticolosamente un caso del genere da un caso di violenza sessuale molto più grave, e di conseguenza attribuire una pena corrispondente alla gravità dell'atto).

Qualcun altro asserisce che l'interesse mediatico attorno a questo avvenimento abbia contribuito ad attribuire una pena di questo genere mentre invece a telecamere spente le conseguenze sarebbero state diverse (forse totale silenzio?), fatto sta che chi non ha preso una posizione specifica a riguardo ha comunque contribuito a divulgare e sensibilizzare le masse su questi temi, consentendo di iniziare a

prendere in più seria considerazione una questione quale la violenza sulle donne, argomento che non è mai stato trattato a dovere.

In altre parole l'episodio potrebbe aver gettato basi più solide per poter affrontare la problematica della violenza sulle donne con maggiore serietà.

Il problema dunque resta aperto, ci si augura che presto si arrivi ad una soluzione definitiva e che episodi come questi diventino soltanto un brutto ricordo.

