**Sentiment Analysis**

La *Sentiment Analysis* (o analisi del sentimento) è una tecnica che ha l’obiettivo di analizzare grandi volumi di dati testuali per determinare il tipo di sentimento espresso: **positivo**, **negativo** o **neutrale**.

Questa analisi si basa su tecnologie di **elaborazione del linguaggio naturale** (Natural Language Processing, NLP) e su **tecniche di apprendimento automatico** (Machine Learning, ML).

Immagine che contiene testo, biglietto da visita, schermata, Carattere

Descrizione generata automaticamente

**Approcci alla Sentiment Analysis**

**Approccio Rule-Based**

L'approccio *rule-based* si basa su insiemi di regole grammaticali e semantiche predefinite. In questo metodo, le parole chiave in un testo vengono classificate secondo un *lessico positivo* e un *lessico negativo*, attribuendo un punteggio di sentimento complessivo al testo e alle sue singole componenti.

Questo sistema, pur essendo semplice e interpretabile, ha una bassa flessibilità e non si adatta facilmente a contesti nuovi o ambigui.

**Approccio Machine Learning**

L’approccio basato sul *machine learning* utilizza algoritmi addestrati su grandi quantità di dati etichettati per apprendere automaticamente i modelli linguistici associati ai sentimenti. Questo metodo tiene conto delle parole presenti nel testo **e dell’ordine in cui compaiono**, permettendo un’analisi più accurata del contesto.

Tra gli algoritmi più comuni troviamo:

* **Regressione lineare**
* **Naive Bayes**
* **Support Vector Machine (SVM)**
* **Deep Learning** (reti neurali profonde, come LSTM o Transformer)

Questo approccio è più flessibile e generalmente più preciso rispetto al metodo rule-based, ma richiede un ampio dataset per l’addestramento.

**Approccio Ibrido**

Molti sistemi adottano un approccio **ibrido**, che combina la robustezza delle regole con l’adattabilità del machine learning, cercando di ottenere il meglio da entrambi i mondi.

**Tipologie di Sentiment Analysis**

Esistono varie tipologie di sentiment analysis, tutte fondate sul concetto di **polarità**, cioè la direzione e l’intensità del sentimento espresso. La polarità può essere misurata su una scala continua (ad esempio da 0 a 100), dove 0 rappresenta neutralità assoluta e 100 un’emozione fortemente positiva o negativa.

**1. Fine-Grained Sentiment Analysis**

Tale approccio raggruppa il testo in diverse emozioni e nel livello di emozione espressa, detta “*analisi a grana fine”*, consente di distinguere diversi livelli di intensità del sentimento:

* Molto positivo
* Positivo
* Neutrale
* Negativo
* Molto negativo

È utile quando si desidera una classificazione dettagliata delle opinioni espresse.

**2. Aspect-Based Sentiment Analysis (ABSA)**

Questa tecnica si concentra su specifici **aspetti** o **caratteristiche** di un prodotto o servizio. Ad esempio, una recensione può essere positiva per il prezzo ma negativa per il servizio clienti.

È particolarmente utilizzata in ambito aziendale per ottenere un feedback dettagliato sull’esperienza dei clienti.

Il sistema inizialmente identifica nel testo **le entità o categorie** (aspetti) su cui l’utente sta esprimendo un’opinione.

Successivamente, estrae tali aspetti mediante tecniche:

-Rule-based: dizionari di termini predefiniti per ogni categoria

-Machine Learning: modelli supervisionati (SVM, Random Forest, CRF) o unsupervised (LSTM,CNN).

Per ogni aspetto trovato, il sistema valuta se l’opinione espressa è **positiva, negativa o neutrale** mediante dei metodi di classificazione (Naive Bayes, SVM, LSTM, o Transformer (BERT, RoBERTa).

Talvolta è possibile che venga valutata l’intensità del sentimento.

**3. Emotion-Based Sentiment Analysis**

Questo approccio cerca di identificare le **emozioni specifiche** espresse nel testo, come rabbia, gioia, tristezza, sorpresa, paura, ecc. Non si limita alla polarità ma indaga lo stato d’animo e le intenzioni dell’autore.

È un metodo più complesso e spesso utilizza modelli avanzati basati su *deep learning*.

**4.Comparative Sentiment Analysis**

L’obiettivo di questa analisi è identificare l’entità preferita all’interno di ciascuna opinione, soprattutto quando il sentiment viene espresso in forma relativa, piuttosto che assoluta.

**5.Sentiment Lexicon Acquisition**

Il lessico rappresenta una risorsa fondamentale per gli algoritmi di sentiment analysis. Questo approccio mira all’acquisizione di un nuovo dizionario di parole utili per analisi future. Esistono tre principali metodi per costruire tale lessico:

* approccio manuale;
* approccio basato su dizionari preesistenti (dictionary-based);
* approccio automatico tramite apprendimento da dati (corpus-based).

**Limiti della Sentiment Analysis**

Nonostante i grandi progressi, la sentiment analysis presenta ancora delle **criticità**:

* **Ironia e sarcasmo:** sono difficili da rilevare automaticamente, poiché il significato ironico spesso contraddice le parole letteralmente usate.
* **Negazioni implicite:** frasi come *“non è proprio il massimo”* possono risultare ambigue e difficili da classificare correttamente.
* **Espressioni idiomatiche:** modi di dire o linguaggi figurati possono essere interpretati erroneamente dai modelli.
* **Contesto culturale o emotivo:** la stessa frase può avere connotazioni diverse in base al contesto culturale, sociale o personale.
* **Ambiguità linguistica:** alcune parole possono avere sentimenti opposti a seconda del contesto ("caldo" può essere positivo riferito a un piatto, negativo per il clima).

RISULTATI ANALISI

E’ stata eseguita la Sentiment Analysis per i due dataset al fine di comprendere i sentimenti che emergono dal dibattito sulla “Nutrizione”.

Inizialmente sono stati considerati i due dataset separatamente considerati ciascuno come corpus unico, al fine di determinare qual è il sentimento prevalente.

Sia nel corpus composto dai threads che da quello composto dai commenti è emerso un sentimento prevalentemente positivo, mentre l’emozione registrata maggiormente risulta essere la “fiducia”.

Immagine che contiene testo, diagramma, linea, Diagramma

Descrizione generata automaticamente

Esaminando la distribuzione delle emozioni è possibile affermare che quelle più frequenti sono trust, anticipation e negative, seguite da fear e joy.

Le emozioni di disgust, anger e surprise sono meno numerose.

Immagine che contiene testo, diagramma, Diagramma, linea

Descrizione generata automaticamente

Nella distribuzione delle emozioni dei threads il pattern è simile, ma con percentuali inferiori per ciascuna emozione (dato il minor numero di token)

La prevalenza di trust e anticipation indica un tono costruttivo e di attesa di risultati, mentre la quota significativa di negative e fear riflette anche dubbi o ansie tipiche dei processi di consulenza nutrizionale.

Immagine che contiene testo, diagramma, Diagramma, linea

Descrizione generata automaticamente

Il grafico a barre affiancate mostra che, in termini percentuali, **trust** è più alto nei commenti, mentre **positive**, **anticipation** e **joy** sono relativamente più presenti nei thread.

Le emozioni negative (anger, disgust, fear) mostrano proporzioni simili in entrambe le fonti, con un leggero calo nei threads.

È possibile affermare che sia nei commenti che nei threads emergono più espressioni positive che negative, ma nei threads l’equilibrio è più sbilanciato verso il positivo, forse perché gli utenti nei threads tendono a focalizzarsi su feedback di supporto o di condivisione di soluzioni.

È possibile concludere affermando che i commenti (forse più diretti e personali) tendono ad esprimere maggiore fiducia, mentre nei threads (più orientati a discussioni aperte) spiccano l’ottimismo e la gioia legati allo scambio di consigli.

Valutazione delle opinioni contrastanti

Successivamente è stata eseguita una valutazione delle opinioni contrastanti presenti nei due set di dati.

E’ stato calcolato il sentiment per ciascun commento e ciascun threads e creata una funzione che determinasse il contrasto tra due elementi a confronto.

Le frasi contrastanti evidenziano utenti indecisi o che riconoscono pregi e limiti contemporaneamente, utile per individuare punti critici di intervento.

Vengono riportati degli estratti dei set di dati in cui sono state raccolte le opinioni contrastanti.

Threads:

Immagine che contiene testo, Carattere, schermata, bianco e nero

Descrizione generata automaticamente

Commenti:

Immagine che contiene testo, bianco e nero, Carattere, carta

Descrizione generata automaticamente

WORDCLOUD DELLE PAROLE PIÙ EMOTIVE

Commenti:

Immagine che contiene testo, Carattere, schermata, design

Descrizione generata automaticamente

Threads:

Immagine che contiene testo, Carattere, grafica, Elementi grafici

Descrizione generata automaticamente

ABSA

L’Aspect based sentiment analysis è basata sulla polarità, che indica la direzione emotiva espressa in un testo: Positive, Neutral, Negative.

È stato definito un dizionario manuale con parole chiave.

Ogni commento/testo è stato diviso in frasi per analizzare in modo più preciso le emozioni per ciascun aspetto.

Sono state selezionate solo le frasi contenenti almeno uno degli aspetti del dizionario.

Mediante syuzhet: get\_sentiment() è stato assegnato un punteggio numerico di sentiment a ciascuna frase che successivamente è stato trasformato in polarità.

Immagine che contiene testo, schermata, Carattere, diagramma

Descrizione generata automaticamente

L’ABSA evidenzia quali temi (aspetti) generano emozioni più forti e in quale direzione, fornendo indicazioni su dove focalizzarsi per migliorare l’esperienza utente.

Gli aspetti con sentiment medio più positivo includono probabilmente “rapport”, “goal”, “plan”; quelli più negativi appaiono “history” o “supplement” (in base al valore medio).

Immagine che contiene testo, schermata, diagramma, Diagramma

Descrizione generata automaticamente

Dal confronto della polarità dei sentimenti nei due set di dati viene confermata l’idea secondo cui i threads tendono ad essere maggiormente ottimisti rispetto ai commenti individuali.

COMPARATIVE SENTIMENT ANALYSIS

In questo tipo di analisi vengono comparati i commenti “padre” e i commenti “figli” ad esso associati, ossia le risposte.

Per fare ciò è stata ricostruita la gerarchia dei commenti tramite la colonna “comment\_id”.

Successivamente è stata calcolata la polarità (positivo/negativo) per ogni commento e sono stati messi in relazione i commenti con i loro “padri” (commenti cui rispondono).

Infine, è stata valutata la concordanza emotiva tra i due.

Immagine che contiene testo, schermata, Carattere, numero

Descrizione generata automaticamente

Le risposte tendono a seguire l’intonazione emotiva del commento originario, segno di coerenza nella discussione e di un ambiente di supporto piuttosto che conflittuale.

Immagine che contiene testo, schermata, diagramma, linea

Descrizione generata automaticamente

Nell’heatmap della variazione emotiva è uno strumento potentissimo per identificare i punti della discussione in cui il dialogo si blocca o diverge emotivamente.

Gli archi sono colorati in base ad “accordo” (verde), “disaccordo” (rosso) o “neutro” (grigio), mentre i nodi mantengono la colorazione secondo la polarità.

E’ un ottimo segnale la quasi totalità di archi verdi, con poche linee rosse che segnalano divergenze emotive.

LEXICON ACQUISITION

Il lexicon è il prodotto finale: un dizionario con parole e punteggi sentimentali.

La lexicon acquisition è il processo per generare quel dizionario, in modo automatico, partendo da dati con annotazioni (ad esempio commenti con score).

Mediante i log-odds ratio i commenti vengono etichettati (positivi vs negativi), e successivamente si calcola per ciascun sentimento quali sono le parole più rappresentative generando un lexicon personalizzato.

Immagine che contiene testo, schermata, Carattere, numero

Descrizione generata automaticamente

Immagine che contiene testo, schermata, Carattere, linea

Descrizione generata automaticamente

E’ possibile concludere che la Sentiment analysis nel complesso mostra un contesto fortemente positivo e di supporto, con alto livello di fiducia e anticipazione, poca ostilità e bassa incidenza di disaccordo. L’ABSA identifica i temi chiave (es. rapport, session, follow-up) con sentiment più elevato, mentre il lexicon personalizzato e le reti gerarchiche riflettono le dinamiche emotive e i termini distintivi del dominio nutrizionale.